

UNIVERSIDAD DE CONCEPCIÓN

Facultad de Ingeniería

Departamento de Ingeniería Metalúrgica

Profesor Patrocinante

Sr. Roberto Fustos T.

**ESTIMACIÓN DE ATRIBUTOS CATEGÓRICOS
DEL MODELO GEOLÓGICO CON HERRAMIENTAS
DE DATA MINING**



VALENTINA ESPERANZA NEIRA VIDAL

Informe de Memoria de Título

para optar al Título de

Ingeniera Civil de Minas

Mayo 2020

RESUMEN

El modelo geológico es esencial para la estimación de recursos como también para procesos aguas abajo, desde el diseño de mina hasta la selección de parámetros metalúrgicos. Actualmente, los modelos geológicos son construidos mediante interpretación determinística y métodos geoestadísticos. Si bien estas técnicas son utilizadas ampliamente, nuevas tecnologías entre ellas el *Data Mining*, no han sido introducidas a la industria minera a pesar de su capacidad para encontrar patrones en los datos y predecir observaciones futuras o no muestreadas.

El objetivo del trabajo es probar la aplicabilidad de algoritmos del *Data Mining*, en específico árboles de decisión y redes neuronales, en la estimación de variables del modelo geológico. Para esto, a partir de un conjunto de datos se construyeron tres modelos geológicos, dos con cada uno de los algoritmos a probar y otro con un método utilizado en minería que es el *co-kriging* indicador, comparándose los modelos obtenidos. Además, se evaluó el impacto de utilizar estos modelos en la estimación de recursos, utilizando separadamente cada modelo geológico predicho como un *input* en la estimación de la ley mediante *kriging* universal, añadiendo otra estimación de ley con *kriging* ordinario (sin utilizar el modelo geológico) para ser comparada con los demás casos. Lo anterior es aplicado en primer lugar a un escenario simulado y posteriormente a un caso real.

En el caso de estudio simulado, las redes neuronales logran un mejor desempeño en términos de precisión en comparación con el *co-kriging* indicador y los árboles de decisión, no obstante, los tres algoritmos son capaces de reproducir los principales patrones del modelo original, pudiendo concluir que las herramientas del *Data Mining* son capaces de igualar e incluso superar la capacidad predictiva de una herramienta geoestadística dentro del modelamiento geológico. En el caso de estudio real se sostiene la misma idea para el modelo geológico de texturas, pero no se pueden hacer conclusiones respecto al modelo de tipo de roca debido a diferencias significativas entre los modelos estimados y la baja precisión de validación alcanzada por los tres algoritmos, lo cual se arrastra por la alta complejidad del yacimiento y la poca información disponible.

Respecto al uso del *kriging* universal que utiliza los modelos geológicos como *inputs*, en los dos casos de estudio se disminuye considerablemente la varianza de *kriging*, y bajo un contexto de baja correlación espacial de la ley se consigue una disminución en el error de estimación de esta. Por lo tanto, la estrategia de utilizar *kriging* universal junto con un modelo geológico estimado por los algoritmos estudiados resulta favorable para la estimación de recursos minerales.

ABSTRACT

Geological modeling is essential for resource estimation as well to downstream processes, since mining design to metallurgic parameter selection. At present geological models are built by deterministic interpretation and geostatistical methods. Although these techniques are widely used in the mining industry, new tools coming from Data Mining, they've not been introduced to mining industry despite their ability to find patterns in data and predict future or unsampled observations.

The objective of this work consists in test the applicability of Data Mining's algorithms, specifically decision trees and neural networks, in the estimation of variables of the geological model. For this, three geological models were modeled from a data set, two with each of the algorithms under review, and other with a method used in mining, that is the indicator co-kriging, then these models are compared. In addition, the impact of using these models on resource estimation is assessed, using each geological model predicted separately as an input in the grade estimation through universal kriging. Another grade estimation with ordinary kriging is added (without using the geological model) to be compared with the other cases. All this is firstly applied to a case study and then to a case with real information.

For the simulated case study, neural networks achieve a better performance, in terms of accuracy, in contrast with indicator co-kriging and decision trees, however, all three are capable of replicate the main patterns from the original model, so one could conclude that Data Mining tools are able to equal and even exceed the predictive capacity of a geostatistical tool within geological modeling. For the real case study, the previous idea is sustained for the texture model, but one cannot make conclusions about type of rock model due significant differences between the estimated models and the low validation accuracy achieved by the three algorithms, which is explained by the complexity of the model.

In relation to the use of universal kriging in combination with the estimated geological models, in both cases of study the variance of kriging is considerably reduced, and in the context of low spatial correlation of the mineral grade, a decrease in the estimation error is achieved. Therefore, the strategy of using universal kriging in combination with a geological model estimated by the algorithms studied is favorable for the mineral resource estimation.

ÍNDICE DE CONTENIDOS

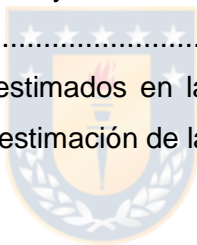
1. INTRODUCCIÓN	1
1.1. Generalidades	1
1.2. Motivación del trabajo	3
1.3. Planteamiento del problema	4
1.4. Objetivos.....	4
1.4.1. Objetivo general	4
1.4.2. Objetivos específicos.....	4
2. MARCO CONCEPTUAL	5
2.1. Conceptos y definiciones básicas de variables regionalizadas.....	5
2.1.1. Variable regionalizada	5
2.1.2. Función aleatoria.....	5
2.1.3. Momentos de primer orden.....	6
2.1.4. Momentos de segundo orden	6
2.2. Herramientas geoestadísticas.....	7
2.2.1. Variograma experimental	7
2.2.2. Variograma: caso multivariable	8
2.2.3. Modelo lineal de correogionalización.....	9
2.2.4. Kriging.....	10
2.3. Kriging ordinario.....	11
2.4. Kriging universal	12
2.5. Estimación de la función de distribución	13
2.5.1. Función indicadora	13
2.5.2. Análisis estructural de una función indicadora	14
2.6. Kriging indicador	16
2.6.1. Kriging indicador ordinario	17
2.6.2. Co-Kriging indicador.....	17
2.7. KKD: Descubrimiento de conocimiento en bases de datos	18
2.8. Data Mining.....	18
2.8.1. Técnicas de clasificación	19
2.8.2. Medidas de desempeño	21
2.8.2. Sobreajuste	23
2.9. Árboles de decisión.....	24

2.9.1. Algoritmo de Hunt.....	24
2.9.2. Medidas para seleccionar la mejor división	25
2.9.3. Algoritmo ID3	26
2.9.4. Algoritmo C4.5.....	26
2.9.5. Algoritmo C5.0.....	30
2.10. Redes neuronales artificiales	31
2.10.1. Mecanismos de aprendizaje	33
2.10.2. Arquitectura de redes neuronales.....	36
2.10.3. Algoritmo de retropropagación	37
2.11. Recursos minerales	39
3. METODOLOGÍA GENERAL	40
3.1. Procedimiento conjunto de datos simulado	40
3.2. Procedimiento conjunto de datos reales	42
4. CASO SIMULADO	43
4.1. Metodología de simulación	43
4.2. Aplicación	47
4.3. Análisis de resultados	48
4.3.1. Estimación de modelos geológicos.....	48
4.3.2. Estimación de leyes y categorización de recursos.....	55
5. CASO REAL	58
5.1. Descripción de la base de datos	58
5.1.2. Mapas	58
5.1.2. Estadísticas básicas	60
5.2. Aplicación	63
5.3. Análisis de resultados	65
5.3.1. Estimación de modelos geológicos.....	65
5.3.2. Estimación de leyes y categorización de recursos.....	69
6. CONCLUSIONES Y DISCUSIONES	72
7. REFERENCIAS	75
8. ANEXOS	78

ÍNDICE DE FIGURAS

Figura 1: Ejemplo de Variograma (Rossi, 2013)	8
Figura 2: Proceso KKD.....	18
Figura 3: Enfoque general para la construcción de un modelo de clasificación	20
Figura 4: Matriz de confusión para $m = 2$	21
Figura 5: Espacio ROC	22
Figura 6: Estructura básica de un árbol de decisión	24
Figura 7: Pseudocódigo del algoritmo ID3.....	26
Figura 8: Tipos de operaciones de post-poda.....	28
Figura 9: Pseudocódigo del algoritmo de obtención de reglas de C4.5	30
Figura 10: Modelo simplificado de una red neuronal biológica	31
Figura 11: Estructura de una neurona artificial	32
Figura 12: Ejemplo de red neuronal multicapa	33
Figura 13: Influencia de la salida de la neurona N_i en la entrada de la neurona N_j	35
Figura 14: Pseudocódigo algoritmo Back-propagation	38
Figura 15: Resumen procedimiento caso de estudio simulado.....	41
Figura 16: Resumen procedimiento caso de estudio real	42
Figura 17: Vista en planta de grilla de simulación.....	43
Figura 18: Histogramas Ley de Cu, Tipo de Roca y Alteración para simulación n° 20.....	44
Figura 19: Vista en planta de secuencia de extracción de sondajes de las primeras 6 bases simuladas para formar conjuntos de entrenamiento	45
Figura 20: Vistas bases simuladas n° 1, 5 y 10 (Color rojo representa conjunto de entrenamiento y color verde el conjunto de validación)	46
Figura 21: Histogramas Ley de Cu, Tipo de Roca y Alteración para conjunto de entrenamiento – simulación n° 20.....	46
Figura 22: Precisiones modelos de clasificación para Tipo de Roca	48
Figura 23: Precisión modelos de clasificación para Alteración	49
Figura 24: Secciones transversales del modelo original de tipo de roca versus modelos estimados para simulación n° 25.....	52
Figura 25: Secciones transversales del modelo original de alteración versus modelos estimados para simulación n° 25.....	53
Figura 26: Tiempos empleados en la construcción y utilización de modelos de clasificación para tipo de roca según algoritmo utilizado	54

Figura 27: Tiempos empleados en la construcción y utilización de modelos de clasificación para alteración según algoritmo utilizado	55
Figura 28: Porcentaje de tonelaje (Medido e Indicado) respecto al tonelaje total por metodología	56
Figura 29: Categorización de Recursos para simulación n°25.....	57
Figura 30: Proyecciones del yacimiento para Leyes de FE	58
Figura 31: Proyecciones del yacimiento para Tipo de Roca	59
Figura 32: Proyecciones del yacimiento para Texturas	59
Figura 33: Histograma Ley de Fe [%]	61
Figura 34: Box-Plots de ley de Fe [%] por tipo de roca (izquierda) y textura (derecha).....	62
Figura 35: Ley media vs ejes coordenados (este, norte, cota) desglosado por tipo de roca .	62
Figura 36: Ley media vs ejes coordenados (este, norte, cota) desglosado por textura.....	62
Figura 37: Secciones transversales para modelo de tipo de roca según clasificador.....	67
Figura 38: Secciones transversales para modelo de texturas según clasificador	68
Figura 39: Nubes de dispersión entre leyes estimadas y reales para kriging ordinario (izquierda) y kriging universal (derecha)	69
Figura 40: Cantidad de bloques estimados en las categorías medido, indicado e inferidos según metodología aplicada en la estimación de la ley	70



ÍNDICE DE TABLAS

Tabla 1: Mesetas de variogramas directos y cruzados para simulación	44
Tabla 2: Criterio de clasificación de recursos para caso simulado.....	48
Tabla 3: Resumen de resultados en estimación de variables del modelo geológico.....	50
Tabla 4: Sensibilidad y Especificidad según variable y modelo de clasificación para simulación n° 25	51
Tabla 5: RMSE y RKV promedio para 25 primeras simulaciones según kriging empleado... 56	
Tabla 6: Estadísticas básicas por Tipo de Roca	60
Tabla 7: Estadísticas básicas por Textura	60
Tabla 8: Criterio de clasificación de recursos para caso real	64
Tabla 9: Precisiones de validación en construcción de modelos geológicos por algoritmo... 65	
Tabla 10: Áreas bajo la curva (ROC) según categorías de modelos de tipo de roca y alteración y algoritmo utilización como clasificador	66
Tabla 11: Estadísticas validación kriging ordinario y kriging universal	69
Tabla 12: Tonelaje y ley media para bloques medidos e indicados según metodología	71



CAPÍTULO 1

INTRODUCCIÓN

1.1. Generalidades

La estimación de recursos es el proceso de estimación del tonelaje, ley, tamaño, forma y ubicación de depósitos minerales [1], además es una parte fundamental en la evaluación de un proyecto minero ya que, a fin de cuentas, determina el valor que posee la empresa. El objetivo de la estimación de recursos es generar una representación convincente, a través de modelos, que describen cómo la fracción económica de la mineralización se distribuye en el espacio mediante la combinación de: interpretación geológica, recolección de información mediante sondeos y análisis estadísticos.

La importancia de la estimación de recursos se ve reflejada una vez que se comienza con la explotación del mineral y los nuevos datos provenientes de pozos de producción se reconcilian¹ con los datos predichos en la estimación. Si la estimación es incorrecta, la valorización del yacimiento también lo es, y una inversión que inicialmente parece menor puede terminar siendo mucho mayor [2].

Aunque, la calidad de la estimación de un recurso mineral depende en primera instancia de la información disponible y la complejidad geológica del yacimiento, depende también fuertemente de las habilidades técnicas y la experiencia del equipo de estimadores, de cómo estos resuelven los problemas, el nivel de atención a los detalles, y la calidad de la documentación que justifica las decisiones tomadas en cada etapa [3].

Si bien en la estimación de recursos es fundamental el estudio de la ley de él o los elementos de interés para la valorización del yacimiento, la entrada para una adecuada estimación de recursos es el modelo geológico. Este modelo corresponde a una representación, matemática o conceptual, de las variables que controlan la mineralización, interpretando la ubicación y geometría de las unidades geológicas, las cuales están definidas principalmente por información proveniente de variables categóricas como lo son la alteración, tipo de roca, textura, estructuras, entre otras [3, 4].

El modelamiento geológico tiene una importante repercusión en etapas aguas abajo, por ejemplo, la litología y alteración afecta directamente a los procesos metalúrgicos, desde la selección

¹ *Reconciliación* es el proceso de comparación entre la producción explotada y lo estimado.

de parámetros de conminución hasta la elección del método de recuperación, trascendiendo en el plan de envío a planta/botadero en la etapa de planificación minera. También, los diferentes tipos de roca se asocian a diferentes unidades geotécnicas del yacimiento, los que afectan a la selección del método de explotación y a los diferentes parámetros del diseño de la mina, ya sea de una explotación subterránea o rajo abierto. Por otro lado, un estudio previo de las variables categóricas del modelo geológico provee subdivisiones que tienen mayor homogeneidad estadística, importante para la asignación de variables continuas al modelo, dado que algunas características geológicas están relacionadas con la distribución de leyes minerales en el yacimiento [1, 4, 5].

Las variables observadas dentro de un depósito mineral no son independientes entre sí, de hecho, existe una dependencia espacial producto de la génesis del depósito, o sea, todos los procesos geológicos que contribuyeron a su formación [6]. Luego, como se desea tener una descripción clara de la continuidad espacial de las variables a modelar, se vuelve frecuente introducir el enfoque científico desarrollado para estudiar la distribución espacial de los fenómenos correlacionados, es decir, la Geoestadística [7].

La Geoestadística consiste en un set de herramientas estadísticas para el estudio de variables distribuidas en el espacio, dependientes entre sí, más conocidas como variables regionalizadas (como lo son las variables de la mineralización) [8], y tiene por objetivo obtener la “mejor” estimación del valor de una variable en un punto en el espacio en función de los datos disponibles.

En el ámbito minero, como es conocido, por razones principalmente económicas, una pequeña fracción del depósito mineral es muestreado [3], por lo tanto, para obtener un valor estimado de cada bloque del modelo geológico, es necesario utilizar técnicas de interpolación. El enfoque tradicional del modelamiento de bloques es estimar un único valor para cada bloque del modelo, obteniendo la mejor predicción posible, en cierto sentido estadístico. Esta estimación se puede realizar usando métodos determinísticos, fundamentados en el conocimiento del estimador en el espacio o usando técnicas geoestadísticas, que toman en cuenta la variabilidad espacial, como son por ejemplo las diferentes versiones del *kriging* [9].

Dado que esta memoria de título se centra en la estimación de atributos categóricos del modelo geológico, una de las herramienta que se abordarán es el *kriging* indicador, método que codifica la variable geológica en diferentes categorías y calcula la probabilidad de que cada categoría se encuentre en una ubicación no muestreada utilizando los datos cercanos [10].

1.2. Motivación del trabajo

Si bien las técnicas Geoestadísticas e incluso los modelos determinísticos son utilizados ampliamente en la industria minera nuevas disciplinas capaces de abordar la estimación de atributos categóricos se han dejado de lado debido a su reciente conocimiento.

El rápido avance computacional y el abaratamiento en los sistemas de almacenamiento de datos ha dado surgimiento a un conjunto de tecnologías [11], denominado *Data Mining*, despertando el interés de una diversidad de áreas, que van desde el *marketing* y negocios hasta la medicina, ciencia e ingeniería [12], incluyendo áreas que trabajan con datos espaciales como la geografía y la astronomía [13, 14]; todas con el objetivo común de descubrir patrones, dentro de sus enormes bases de datos, que expliquen cómo funcionan sus objetos de estudio, para así, poder predecir situaciones futuras o contextos inesperados [15].

Las tareas del *Data Mining* se han desarrollado para explorar grandes bases de datos con el fin de encontrar patrones novedosos y útiles, que de otro modo podrían permanecer desconocidos, estas también proporcionan la capacidad de predecir el resultado de una observación futura [12].

Como se ha mencionado anteriormente, las variables que controlan la mineralización describen importantes patrones y estructuras de correlación espacial, fundamentales para la evaluación y planificación de minas. Por otro lado, los diferentes modelos desarrollados a lo largo de la vida de la mina, desde los de estimación hasta los de producción, requieren almacenar numerosas variables, como por ejemplo ley de mineral, mineralización, recuperación, etc., lo cual es posible gracias a las capacidades de los *hardwares* actuales, creándose grandes bases de datos, tamaño que sigue en aumento [3]. De lo anterior, resulta sugerente utilizar herramientas del *Data Mining* para buscar patrones dentro de un depósito mineral que proporcionen una aplicabilidad dentro del área de estimación de recursos.

Dada la importancia mencionada de estimar correctamente las variables categóricas del modelo geológicos, la memoria de título se enfoca en los procedimientos de clasificación del *Data Mining*, los que ofrecen un enfoque sistemático para construir modelos de clasificación a partir de un conjunto de datos, empleando algoritmos de aprendizaje que identifiquen aquel modelo que mejor ajuste la relación entre el conjunto de variables de la base de datos con categorías ya predefinidas [12].

1.3. Planteamiento del problema

Con lo anterior, se plantea la siguiente hipótesis, la cual se busca verificar en el desarrollo de la memoria de título presente:

“La estimación de atributos categóricos pertenecientes al modelo geológico puede realizarse utilizando algoritmos provenientes del área del *Data Mining*, para posteriormente, incorporar este conocimiento en las etapas de estimación y categorización de los recursos. De esta forma, sería posible incrementar la cantidad de información disponible y tener un mejor manejo del riesgo en la toma de decisiones en proyectos mineros”.

1.4. Objetivos

1.4.1. Objetivo general

El objetivo general de esta memoria de título es probar la aplicabilidad de diferentes algoritmos provenientes del área del *Data Mining* en la estimación de atributos categóricos del modelo geológico, de tal forma de incrementar la cantidad de información a la hora de realizar la estimación y categorización de recursos, y ver cómo esto impacta en la toma de decisiones en proyectos mineros.

1.4.2. Objetivos específicos

- Aplicar algoritmos del *Data Mining*, específicamente árboles de decisión y redes neuronales, para la identificación de patrones en datos pertenecientes a un depósito mineral, los cuales posteriormente permitan estimar atributos categóricos del modelo geológico.
- Evaluar y comparar los resultados obtenidos a través de estos algoritmos con un método actualmente empleado en la industria minera que corresponde al *co-kriging* indicador.
- Evaluar el impacto que genera esta información extra al utilizarla para la estimación de la ley en el depósito y posterior categorización de recursos.

CAPÍTULO 2

MARCO CONCEPTUAL

A continuación, se describen los antecedentes necesarios para el desarrollo del trabajo.

2.1. Conceptos y definiciones básicas de variables regionalizadas

2.1.1. Variable regionalizada

Una variable regionalizada se define como una variable aleatoria $Z(\mathbf{x})$, donde \mathbf{x} representa la posición. Un conjunto de variables aleatorias se denomina función aleatoria. Así, el conjunto de variables aleatorias, indexados sobre un dominio D , es una función aleatoria que denotaremos $\{Z(\mathbf{x}), \mathbf{x} \in D\}$ [8]. Luego se puede decir que una variable regionalizada es una realización de una función aleatoria.

Una variable regionalizada mide propiedades o atributos relacionados a un fenómeno regionalizado, como por ejemplo la ley mineral que es una característica de un fenómeno geológico denominado mineralización [16], y que presenta cierta dependencia espacial (espacio geográfico). Intuitivamente, esto indica que mientras más cercanas estén situadas dos leyes, más correlacionadas están y mientras más separadas, menos relación existe entre ellas. Una variable regionalizada puede ser de naturaleza continua o categórica.

2.1.2. Función aleatoria

Como se menciona en la sección anterior, una variable regionalizada es una realización de una función aleatoria $\{Z(\mathbf{x}), \mathbf{x} \in D\}$. Al considerar una serie de sitios $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ el grupo de variables aleatorias $\{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)\}$ es caracterizado por una función de distribución multivariada:

$$F_{x_1, \dots, x_k}(z_1, \dots, z_k) = Prob\{Z(\mathbf{x}_1) < z_1, \dots, Z(\mathbf{x}_k) < z_k\} \quad \forall z_1, \dots, z_k \in \mathbb{R}^k. \quad (1)$$

La ecuación (1) constituye la distribución espacial de una función aleatoria, esta reúne las funciones de probabilidad para todo k en todas las elecciones posibles de $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ en D^k , por lo que se recurre a simplificaciones, resultando así las distribuciones univariable y bivariable con $k = 1$ y $k = 2$ respectivamente.

La caracterización de una función aleatoria se puede simplificar aún más considerando sólo algunos parámetros descriptivos o “momentos” [9]. Considerando dos variables $Z(\mathbf{x}_1)$ y $Z(\mathbf{x}_2)$ se tiene:

2.1.3. Momentos de primer orden

Valor Esperado: Representa la media alrededor de la cual se distribuyen los valores tomados por las realizaciones de la función aleatoria.

$$m_1 = E[Z(\mathbf{x}_1)], \quad m_2 = E[Z(\mathbf{x}_2)]. \quad (2)$$

2.1.4. Momentos de segundo orden

Varianza: Medida de dispersión de $Z(\mathbf{x})$ respecto a su valor medio $m(\mathbf{x})$, su raíz cuadrada corresponde a la **desviación estándar**.

$$\sigma_1^2 = var[Z(\mathbf{x}_1)] = E\{[Z(\mathbf{x}_1) - m_1]^2\} = E[Z(\mathbf{x}_1)^2] - m_1^2. \quad (3)$$

Covarianza: La covarianza centrada entre dos variables aleatorias $Z(\mathbf{x}_1)$ y $Z(\mathbf{x}_2)$ corresponde a una función que mide la relación entre dos variables y se define como:

$$C_{12} = cov[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] = E[(Z(\mathbf{x}_1) - m_1)(Z(\mathbf{x}_2) - m_2)] = E[Z(\mathbf{x}_1)Z(\mathbf{x}_2)] - m_1 m_2. \quad (4)$$

El termino $E[Z(\mathbf{x}_1)Z(\mathbf{x}_2)]$ corresponde a la **covarianza no centrada** entre las variables $Z(\mathbf{x}_1)$ y $Z(\mathbf{x}_2)$.

Variograma: Mide la desviación cuadrática media entre dos variables, da un indicio de qué tan distintas son dos variables aleatorias realizadas en dos lugares distintos.

$$\gamma(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2} var[Z(\mathbf{x}_1) - Z(\mathbf{x}_2)]. \quad (5)$$

Es necesario recurrir a la **Hipótesis de Estacionaridad** para realizar inferencia estadística, es decir, estimar los momentos a partir de un conjunto de datos experimentales. Esta hipótesis postula que la distribución espacial de la función aleatoria es invariante por traslación, es decir, las propiedades de un conjunto de datos no dependen de su posición absoluta en el espacio, sino que sólo de sus posiciones relativas. Luego, la esperanza y la varianza son constantes (independientes de su ubicación \mathbf{x}), mientras que la covarianza y el variograma sólo dependen del vector de separación $\mathbf{h} = \mathbf{x}_1 - \mathbf{x}_2$. Aplicando la hipótesis de estacionaridad se obtienen las siguientes relaciones:

$$\gamma(\mathbf{h}) = Var[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] = E\{[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2\}, \quad (6)$$

$$C(0) = var[Z(\mathbf{x})] = \sigma^2 \quad \text{con } \mathbf{h} = 0, \quad (7)$$

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \quad \text{o} \quad C(\mathbf{h}) = C(0) - \gamma(\mathbf{h}). \quad (8)$$

2.2. Herramientas geoestadísticas

Un aspecto fundamental en el modelamiento geoestadístico es establecer medidas cuantitativas de la variabilidad o continuidad espacial para ser utilizadas posteriormente en una estimación o simulación [3]. Para realizar la estimación de una variable regionalizada no muestreada, la geoestadística opera en dos etapas [9]:

1. Análisis estructural, en esta etapa se describe la correlación de la variable de interés entre puntos en el espacio con la ayuda del variograma.
2. Estimación de la variable de interés en los sitios no muestreados por medio del método de *kriging*.

2.2.1. Variograma experimental

En la ecuación (6) se encuentra la definición de variograma que se interpreta como el valor esperado de la diferencia al cuadrado de dos variables separadas por una distancia \mathbf{h} . El variograma no depende de la ubicación de las muestras, y como es una medida de variabilidad, crece a medida que las muestras se vuelven más disímiles entre sí [3].

Es posible estimar el variograma con la información disponible con el estimador llamado **variograma experimental** para un vector de separación \mathbf{h} dado, el cual queda definido por:

$$\gamma^*(\mathbf{h}) = \frac{1}{2 |N(\mathbf{h})|} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2. \quad (9)$$

donde $N(\mathbf{h}) = \{(\alpha, \beta) \text{ tal que } \mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h}\}$ y $|N(\mathbf{h})|$ es el número de pares contenidos en el conjunto $N(\mathbf{h})$. Los $|N(\mathbf{h})|$ pueden ser reunidos usando tolerancias razonables de distancia y dirección [9].

Entre las principales características del variograma, **Figura 1**, se encuentran [16]:

- **Meseta:** Cota superior del variograma donde ya no existe correlación lineal. También se puede definir como el límite del variograma cuando h tiende a infinito.
- **Rango:** Distancia a la cual se alcanza la no-correlación.
- **Efecto pepita:** Representa una discontinuidad puntual del variograma en el origen. Puede deberse a errores de medición o a la escala de la misma.

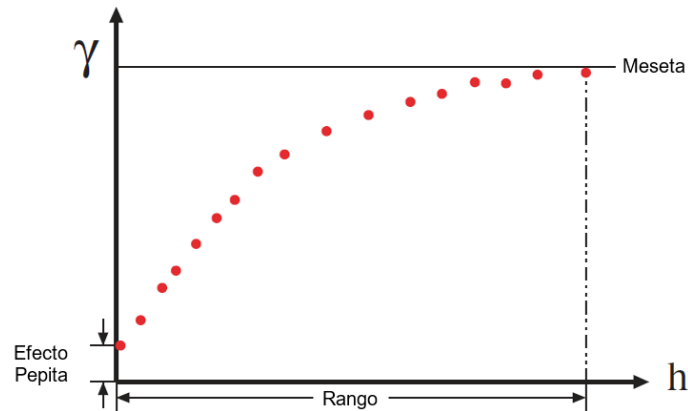


Figura 1: Ejemplo de Variograma (Rossi, 2013)

El variograma puede ser calculado en diferentes direcciones. El patrón de continuidad espacial es anisotrópico cuando la variable es más continua en una dirección que en otra, en caso contrario se dice isotrópico [3].

Los puntos del variograma experimental no se usan directamente en la estimación subsiguiente, más bien, una función paramétrica se ajusta a esos puntos para obtener un modelo tridimensional [3]. Para que una función sea un variograma debe cumplir con varias propiedades matemáticas, entre ellas, paridad, nulidad en el origen, positividad y ser de tipo negativo condicional, esta última propiedad se resume en la ecuación (10) y es una condición necesaria y suficiente para que γ sea el variograma de una función aleatoria, al mismo tiempo es una propiedad restrictiva y difícil de controlar, por lo que, en general se elige un modelo de variograma entre funciones de las cuales se conoce que cumplen tal condición [9].

$$\forall k \in \mathbb{N}^*, \forall \lambda_1, \dots, \lambda_k \in \mathbb{R} \text{ tales que } \sum_{i=1}^k \lambda_i = 0, \forall \mathbf{x}_1, \dots, \mathbf{x}_k \in D, \sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \leq 0. \quad (10)$$

Entre las funciones más utilizadas están los **modelos esféricos**, **gaussianos** y **exponenciales**. Además, el variograma puede ser ajustado como la suma de modelos válidos, denominados, **estructuras anidadas** [8].

2.2.2. Variograma: caso multivariable

Es frecuente que varias variables regionalizadas se refieran a un mismo fenómeno, por ende, para resaltar la relación estructural entre variables, mejorar la estimación de una variable gracias a otra(s) variable(s) o mejorar la consistencia entre las estimaciones de diferentes variables, es utilizada la Geoestadística multivariable [9, 17].

Se llama **corregionalización** al conjunto de variables regionalizadas que denotaremos z_1, \dots, z_N . Estas variables están definidas sobre un mismo dominio acotado de R^d , llamado campo de la corregionalización y denotado D . Las funciones aleatorias asociadas a las variables regionalizadas se denotan como Z_1, \dots, Z_N .

Bajo la hipótesis de estacionaridad de segundo orden, la **covarianza cruzada** entre dos funciones aleatorias Z_i y Z_j para un vector \mathbf{h} se define como:

$$C_{ij}(\mathbf{h}) = E\{Z_i(\mathbf{x})Z_j(\mathbf{x} + \mathbf{h})\} - E\{Z_i(\mathbf{x})\} \times E\{Z_j(\mathbf{x} + \mathbf{h})\}. \quad (11)$$

Cuando $i = j$, se encuentra la función de **covarianza simple o directa**. Se define la matriz $\mathbf{C}(\mathbf{h})$ de funciones de covarianzas simples y cruzadas como:

$$\mathbf{C}(\mathbf{h}) = [C_{ij}(\mathbf{h})]_{i,j=1,\dots,N}. \quad (12)$$

El **variograma cruzado** entre dos funciones aleatorias Z_i y Z_j para un vector \mathbf{h} se define como:

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2}E\{[Z_i(\mathbf{x}) - Z_i(\mathbf{x} + \mathbf{h})][Z_j(\mathbf{x}) - Z_j(\mathbf{x} + \mathbf{h})]\}. \quad (13)$$

El **variograma simple o directo** corresponde al caso $i = j$. La matriz de variogramas se define de la siguiente manera:

$$\mathbf{\Gamma}(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]_{i,j=1,\dots,N}. \quad (14)$$

La relación del variograma cruzado con la función de covarianza se define en la siguiente relación:

$$\gamma_{ij}(\mathbf{h}) = C_{ij}(\mathbf{0}) - \frac{1}{2}[C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})], \quad (15)$$

matricialmente:

$$\mathbf{\Gamma}(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \frac{1}{2}[\mathbf{C}(\mathbf{h}) + \mathbf{C}(-\mathbf{h})]. \quad (16)$$

Esta relación muestra que el variograma cruzado toma el promedio de la función de covarianza cruzada en los valores $+\mathbf{h}$ y $-\mathbf{h}$. Como en el caso univariable, se puede calcular las covarianzas y variogramas simples y cruzados experimentales para un determinado vector \mathbf{h} .

2.2.3. Modelo lineal de corregionalización

No es posible modelar los variogramas cruzados y directos de un conjunto de variables independientemente, debido a que la matriz $\mathbf{\Gamma}(\mathbf{h})$ debe ser simétrica y de tipo positivo para todo

vector \mathbf{h} . Para satisfacer lo anterior se recurre al modelo lineal de correogionalización (generalización del concepto de modelo anidado en caso univariable).

Se supone que los variogramas simples y cruzados son combinaciones de un mismo conjunto de modelos base:

$$\forall i, j \in [1, N], \gamma_{ij}(\mathbf{h}) = \sum_{u=1}^S b_{ij}^u g_u(\mathbf{h}), \quad (17)$$

matricialmente:

$$\mathbf{\Gamma}(\mathbf{h}) = \sum_{u=1}^S \mathbf{B}_u g_u(\mathbf{h}), \quad (18)$$

donde: $\mathbf{\Gamma}(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]_{i,j=1,\dots,N}$ es la matriz de variogramas directos y cruzados,

$g_u(\mathbf{h})$ es un modelo básico de variograma, y

$\mathbf{B}_u = [b_{ij}^u]_{i,j=1,\dots,N}$ es la matriz de correogionalización lineal.

Una condición suficiente para que el modelo así definido sea matemáticamente válido es que cada una de las matrices de correogionalización sea simétrica de tipo positivo, es decir, todos los valores propios de \mathbf{B}_u deben ser positivos o nulos. En el caso de dos variables Z_1 y Z_2 , la condición anterior se cumple si:

$$\forall u \in [1, S], |b_{12}^u| \leq \sqrt{b_{11}^u b_{22}^u}, \quad (19)$$

la desigualdad implica que un modelo básico que aparece en el variograma cruzado con una meseta no nula presenta forzosamente una meseta no nula en los dos variogramas directos. En el caso de tres o más, la desigualdad anterior ya no es suficiente para asegurar que las matrices de correogionalización sean de tipo positivo, sino que se debe calcular sus valores propios para comprobar si son positivos o no.

2.2.4. Kriging

Es una herramienta de estimación local de una variable regionalizada. Se utiliza en casos de estimación univariable, y provee una medida de certeza para sus estimaciones. El estimador, digamos Z^* , es una combinación lineal de los n valores observados, esto es:

$$Z^*(\mathbf{x}_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}), \quad (20)$$

donde x_0 es el sitio de la variable a estimar, x_α con $\alpha = 1, \dots, n$ son los sitios con datos. Los ponderadores λ_α junto con el coeficiente μ son las incógnitas del problema de *kriging*.

Los n coeficientes λ_α son calculados para obtener el predictor Z^* , tal que [9]:

$$\text{a) Sea insesgado, es decir,} \quad E[Z^*(x_0) - Z(x_0)] = 0, \text{ y} \quad (21)$$

$$\text{b) Minimice el error de predicción,} \quad \text{Var}[Z^*(x_0) - Z(x_0)] \text{ es mínima.} \quad (22)$$

Los predictores del *kriging* son interpoladores exactos y se calculan de forma que son los mejores predictores lineales insesgados. Una propiedad importante del *kriging* es la de suavizamiento que implica una dispersión de los valores estimados menor a la dispersión de los valores reales.

Existen varios tipos de *kriging*, que difieren principalmente por las suposiciones acerca de la media de la variable de interés [8]. Estos pueden ser clasificados según su forma, lineales o no lineales, el soporte, puntual o en bloques, y si es paramétrico o no [18]. Para efectos del presente trabajo, son de particular interés el *kriging* ordinario, universal e indicador.

2.3. Kriging ordinario

Se supone que la variable regionalizada es la realización de una función aleatoria $Z(x)$, estacionaria, con media constante pero desconocida.

A partir de la ecuación (20) y las propiedades de insesgo y mínima varianza, se obtiene el siguiente sistema de ecuaciones que resuelve el problema de *kriging*:

$$\begin{cases} \sum_{\beta=1}^n \lambda_\beta \gamma(x_\alpha - x_\beta) - \mu = \gamma(x_\alpha - x_0), \quad \forall \alpha = 1, \dots, n, \\ \sum_{\alpha=1}^n \lambda_\alpha = 1, \end{cases} \quad (23)$$

donde la variable μ se denomina multiplicador de Lagrange.

La varianza del *kriging* ordinario queda dada por:

$$\sigma_{KO}^2 = \sigma^2 - \sum_{\alpha=1}^n \lambda_\alpha C(x_1) - \mu = \sum_{\alpha=1}^n \lambda_\alpha \gamma(x_\alpha - x_0) - \mu \quad (24)$$

donde $\sigma^2 = C(0)$ es la varianza a *priori* de la función aleatoria $Z(x)$, o sea, la meseta del variograma.

2.4. Kriging universal

En los supuestos hechos hasta ahora respecto al *kriging* se ha asumido que la variable regionalizada es estacionaria, pero en muchos casos, la variable no satisface estas condiciones y se caracteriza por exhibir una tendencia [16]. Para tratar este tipo de variables es frecuente descomponer la variable $Z(x)$ como la suma de la tendencia, tratada como una función determinística $m(x)$, más una componente estocástica estacionaria $R(x)$:

$$Z(x) = m(x) + R(x), \quad (25)$$

donde $E[Z(x)] = m(x)$ y $E[R(x)] = 0$.

La tendencia puede expresarse como una suma ponderada de funciones conocidas $f_l(x)$ y coeficientes desconocidos a_l con $l = 0, \dots, L$, mediante:

$$m(x) = \sum_{l=0}^L a_l f_l(x), \quad (26)$$

donde, por convenio, $f_0(x) = 1$, para todo x .

El predictor *kriging* universal se define como:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i), \quad (27)$$

y los ponderadores λ_i son determinados por el siguiente sistema de ecuaciones [19]:

$$\begin{cases} \sum_{j=1}^n \lambda_j(x) C_R(x_j - x_i) + \sum_{k=1}^K \mu_k(x) f_k(x_i) = C_R(x_j - x_i), & i = 1, \dots, n \\ \sum_{j=1}^n \lambda_j(x) f_k(x_j) = f_k(x), & k = 0, \dots, K \end{cases} \quad (28)$$

La varianza del *kriging* universal queda dada por:

$$\sigma_{KU}^2 = \sum_{i=1}^n \lambda_i Z(x_i) + \sum_{l=1}^L \mu_l f_l(x_0), \quad (29)$$

Cabe notar que sí $L = 0$ y $f_1(x) = 1$, el sistema de ecuaciones del *kriging* universal y su varianza coinciden con las del *kriging* ordinario.

2.5. Estimación de la función de distribución

En ocasiones interesa estimar la probabilidad de que la variable $Z(x)$ sea mayor, menor o igual a un valor determinado $z_k \in \mathbb{R}$, para lo cual se requiere estimar la función de distribución del proceso espacial [20]. La función de distribución condicional acumulada describe el rango de posibles valores que una variable puede tomar y se expresa como:

$$F(z_k, x | (n)) = Prob\{Z(x) \leq z_k | (n)\} \quad (30)$$

donde “|(n)” significa condicional a los datos cercanos.

Una manera de llevar a cabo la estimación de la distribución de una función aleatoria, es utilizar el *kriging* indicador, técnica que se basa en la transformación del conjunto de variables originales de acuerdo a un indicador $I(x, z_k)$.

2.5.1. Función indicadora

La función indicadora consiste en una transformación binaria de la variable aleatoria original $Z(x)$ que, para un determinado valor $z_k \in \mathbb{R}$, toma la siguiente forma [8, 21]:

$$I_Z(x; z_k) = I(Z(x) \leq z_k) \begin{cases} 1, & \text{si } Z(x) \leq z_k \\ 0, & \text{si } Z(x) > z_k \end{cases} \quad (31)$$

La función indicadora consiste en discretizar el rango de valores de la variable continua Z , R_Z , en una serie de K valores de corte z_k , $k = 1, \dots, K$.

En el presente trabajo, como se ha mencionado, se evaluará la estimación de variables categóricas, las cuales permiten codificar un conjunto de dominios que subdividen el espacio, como por ejemplo, dominios por tipos de roca, o representar variables con un número limitado de categorías [9].

Sea Z una variable categórica o categorizable, en K categorías, s_k , $k = 1, \dots, K$. Estas categorías deben ser exhaustivas y mutuamente excluyentes. Análogo a la forma de la ecuación (31), la función indicadora de una variable categórica es:

$$I(x; s_k) = I(Z(x) = s_k) \begin{cases} 1, & \text{si } Z(x) = s_k \\ 0, & \text{si } Z(x) \neq s_k \end{cases} \quad (32)$$

El indicador en este caso se interpreta como la probabilidad de que una categoría predomine en una ubicación particular. Las propiedades de exclusividad y exhaustividad implican las siguientes relaciones:

$$I(x; s_k) \cdot I(x; s_{k'}) = 0, \forall k \neq k', \text{ y} \quad (33)$$

$$\sum_{k=1}^K i(x; s_k) = 1. \quad (34)$$

2.5.2. Análisis estructural de una función indicadora

Con la variable indicadora $I(x; s_k)$, se procede a trabajar de manera análoga a lo planteado en las secciones anteriores con la variable $Z(x)$, tanto en el análisis estructural como para la estimación de valores en posiciones no muestreadas. Dado un valor s_k y considerando que se está trabajando en termino de datos categóricos, el valor esperado de la función aleatoria $I(x; s_k)$ queda dado por:

$$E[I(x; s_k)] = 1 \cdot Prob\{Z(x) = s_k\} + 0 \cdot Prob\{Z(x) \neq s_k\} = Prob\{Z(x) = s_k\} = F_x(s_k), \quad (35)$$

la relación anterior sigue siendo válida en el caso de una distribución condicional.

Los momentos de segundo orden, entre dos variables $I(x; s_k)$ e $I(x+h; s_k)$ separadas por un vector h , se definen como:

- **Varianza**

$$var[I(x; s_k)] = Prob\{Z(x) = s_k\}[1 - Prob\{Z(x) = s_k\}] \quad (36)$$

- **Covarianza no centrada**

$$K_I(h; s_k) = E[I(x+h; s_k) \cdot I(x; s_k)] = Prob\{Z(x+h) = s_k \text{ y } Z(x) = s_k\} \quad (37)$$

- **Covarianza centrada**

$$\begin{aligned} C_I(h; s_k) &= cov[I(x+h; s_k), I(x; s_k)] = E[I(x+h; s_k) \cdot I(x; s_k)] - E[I(x+h; s_k)] \cdot E[I(x; s_k)] \\ &= Prob\{Z(x+h) = s_k \text{ y } Z(x) = s_k\} - F_{x+h}(s_k) \cdot F_x(s_k) \end{aligned} \quad (38)$$

- **Variograma**

$$\begin{aligned} 2\gamma_I(h; s_k) &= E\{[I(x+h; s_k) - I(x; s_k)]^2\} = E[I(x+h; s_k)] + E[I(x; s_k)] - 2K_I(h; s_k) \\ &= C_I(0; s_k) - C_I(h; s_k) \end{aligned} \quad (39)$$

- **Covarianza no centrada cruzada**

$$K_I(h; s_k, s_{k'}) = E[I(x+h; s_k) \cdot I(x; s_{k'})] = Prob\{Z(x+h) = s_k \text{ y } Z(x) = s_{k'}\} \quad (40)$$

- **Covarianza centrada cruzada**

$$\begin{aligned}
C_I(h; s_k, s_{k'}) &= cov[I(x+h; s_k) \cdot I(x; s_{k'})] \\
&= E[I(x+h; s_k) \cdot I(x; s_{k'})] - E[I(x+h; s_k)] \cdot E[I(x; s_{k'})] \\
&= Prob\{Z(x+h) = s_k \text{ y } Z(x) = s_{k'}\} - F_{x+h}(s_k) \cdot F_x(s_{k'})
\end{aligned} \tag{41}$$

- **Variograma cruzado**

$$2\gamma_I(h; s_k, s_{k'}) = E\{[I(x+h; s_k) - I(x; s_{k'})][I(x+h; s_{k'}) - I(x; s_k)]\}. \tag{42}$$

Dado que $Z(x)$ es una función aleatoria para la cual se conocen los valores $Z(x_1), \dots, Z(x_n)$ en las localidades x_1, \dots, x_n la estimación del semivariograma se hace de manera análoga a la sección anterior, pero aplicando los resultados sobre $I(x, s_k)$. Los variogramas indicadores $\gamma_I(h, s_k)$ son estimados para cada categoría dada $s_k \in \mathbb{R}$, así el estimador del variograma según Matheron queda de la siguiente forma:

$$\hat{\gamma}_I(h; s_k) = \frac{1}{2|N(t)|} \sum_{(i,j) \in N(t)} [I(x_i, s_k) - I(x_j, s_k)]^2, \tag{43}$$

donde $N(t) = \left\{ \frac{i,j}{x_i} - x_j = h \right\}$ y $|N(t)|$ es su cardinal.

Un procedimiento para la estimación de la función de distribución $F_x(s_k)$, además de los métodos por *kriging* que serán evaluados en la sección siguiente, puede derivarse de la relación existente entre la meseta del variograma indicador, digamos $S(\cdot)$, y la función de distribución, $F(\cdot)$:

$$S(s_k) = \lim_{\|h\| \rightarrow \infty} \gamma_I(h; s_k) = F_x(s_k) - F_x(s_k)^2, \tag{44}$$

los valores de la función de distribución $F_x(s_k)$ se obtiene al resolver la ecuación de segundo grado:

$$F_x(s_k) - F_x(s_k)^2 + S(s_k) = 0, \tag{45}$$

la meseta, $S(s_k)$, toma valores en el intervalo $[0, 0.25]$, alcanzando su máximo para la mediana de la distribución, que denotamos con s_M , es decir $F_x(s_k) = 0.5$. además, es una función creciente $(-\infty, s_M]$ y decreciente en $[s_M, \infty)$.

Una vez conocida la meseta del variograma se puede calcular la distribución con:

$$F_x(s_k) = 0.5(1 + \text{sign}(s_k - s_M) \sqrt{1 - 4S(s_k)}), \tag{46}$$

esta puede ser interpretado como una probabilidad bivariada:

$$\gamma_I(h, s_k) = 0.5\{Pr[Z(x) = s_k, Z(x+h) \neq s_k]\} + 0.5\{Pr[Z(x+h) = s_k, Z(x) \neq s_k]\}. \tag{47}$$

2.6. Kriging indicador

El *kriging* indicador es un método no lineal y no paramétrico en que los valores son convertidos a 0 y 1 dependiendo de su relación con una categoría dada. Utilizando la relación de la ecuación (20) se puede construir una función de distribución condicional acumulada mediante la unión de K estimadores tipo *kriging* indicador [3]. Esta función representa un modelo probabilístico sobre la incertidumbre de los valores $Z(x)$ no muestreados.

Para estimar $E\{I(x; s_k)|n\}$ usamos el siguiente ponderador lineal:

$$I^*(x; s_k) = E^*\{I(x; s_k)|n\} = \sum_{\alpha=1}^n \lambda_{\alpha}(x; s_k)I(x; s_k) \quad (48)$$

Cuando se utilizan varias categorías k , el sistema es llamado usualmente *kriging* indicador múltiple. Los ponderadores y la función de distribución acumulada condicional son dependientes tanto de la ubicación como del número de categorías s_k , con $k = 1, \dots, K$. Luego, hay un variograma indicador $\gamma_1(h; s_k)$ y un sistema *kriging* por categoría.

Las etapas a seguir para utilizar el estimador *kriging* son:

1. Elegir las categorías s_k .
2. Para $k = 1, \dots, K$
 - i. Codificar los datos en indicadores $I(x, s_k)$, ecuación (32),
 - ii. Realizar análisis variográfico,
 - iii. Realizar *kriging* del indicador, estimando probabilidad condicional de que $Z(x) = s_k$,
 - iv. Procesar estimaciones para obtener distribución condicional válida.

El *kriging* indicador presenta ciertas ventajas y desventajas [22]:

- Ventajas:
 - Toma en cuenta la estructura de cada indicador.
 - No requiere un modelamiento a priori de la distribución teórica de $F_x(s_k)$.
 - No requisito que exista estacionaridad global, pero si local.
- Desventajas:
 - Se tiene la misma cantidad de variogramas que umbrales o categorías definidas y, por tanto, también la misma cantidad de sistemas *kriging* por resolver, lo que lo hacer consumidor de tiempo.
 - *El kriging* no garantiza que los ponderadores sean no negativos.

2.6.1. Kriging indicador ordinario

El valor esperado del transformador indicador para cada categoría se asume desconocido pero constante dentro de una vecindad local [23]. Considerando la ecuación (48) y sujeto a $\sum_{\alpha=1}^n \lambda_{\beta} = 1$. El resultado del sistema de ecuaciones para el *kriging* indicador ordinario es:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C_I(x_{\alpha} - x_{\beta}; z_k) + \mu = C_I(x_{\alpha} - x; z_k); & \alpha = 1, \dots, n, \\ \sum_{\beta=1}^n \lambda_{\beta} = 1. \end{cases} \quad (49)$$

2.6.2. Co-Kriging indicador

El *kriging* indicador toma en cuenta la posición de un valor respecto a una categoría, pero no su proximidad: un valor ligeramente por encima de la categoría no se distingue de un valor muy grande. Una forma de incorporar más información es considerar el valor de las variables asociadas a una serie de umbrales s_n . Argumentando en un contexto multivariado, es lógico tratar de estimar $I(s_n)$ mediante *co-kriging* utilizando no solo los indicadores $I_{Z(x_{\alpha})=s_n}$ sino también los indicadores de los demás niveles $I_{Z(x_{\alpha})=z_m}$ con el *co-kriging* indicador. Luego el estimador queda con la forma [22]:

$$I^* = \lambda_0 + \sum_m \sum_{\alpha=1}^N \lambda_{m\alpha} I_{Z(x_{\alpha}) < z_m} \quad (50)$$

Considerando el *co-kriging* con una media desconocida, el *co-kriging* indicador ordinario se resuelve con las siguientes relaciones:

$$\lambda_0 = 0, \quad \sum_{\alpha=1}^N \lambda_{n\alpha} = 1, \quad \sum_{\alpha=1}^N \lambda_{m\alpha} = 0 \quad \text{si } m \neq n \quad (51)$$

Desde un punto de vista teórico, este método es claramente más poderoso que el *kriging* indicador ordinario visto en la sección anterior, ya que incorpora la información de que un valor de una muestra que está ligeramente por encima de un umbral s_n ($I_{Z(x_{\alpha}) < s_{n+1}} = 1$) o bien sobre el umbral ($I_{Z(x_{\alpha}) < s_m} = 0$ para todos los niveles considerados), mientras que en ambos casos $I_{Z(x_{\alpha}) < s_n} = 0$. Pero tiene un inconveniente: requiere la estimación y el modelado de covarianzas, directas y cruzadas, de todos los niveles, sin mencionar la resolución de sistemas muy grandes.

El correspondiente sistema de *co-kriging* indicador implica una matriz de K^2 covarianzas indicadoras directas y cruzadas del tipo:

$$C_I(h; z_k, z_{k'}) = Cov\{I(x; z_k), I(x + h; z_{k'})\} \quad (52)$$

2.7. KKD: Descubrimiento de conocimiento en bases de datos

De la necesidad de descubrir conocimiento a partir de los datos, nace el proceso llamado “Descubrimiento de Conocimiento en Bases de Datos” o KKD por sus siglas en inglés (*knowledge discovery in databases*). El KKD se define como: “Proceso no trivial de identificar patrones en los datos que sean válidos, novedosos, potencialmente útiles y comprensibles” [24]. El proceso KKD envuelve una serie de pasos en los cuales el usuario debe ir tomando decisiones. En la **Figura 2**, se presenta un esquema general de los pasos que abarca este proceso.

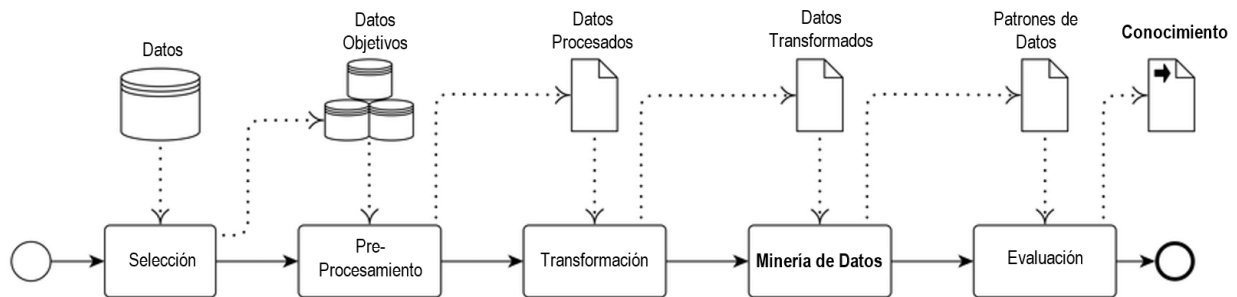


Figura 2: Proceso KKD

El *Data Mining*, es la etapa central del proceso *KKD*, y muchos usan esta expresión para referirse al proceso *KKD* completo [25], aunque el resto de las etapas siguen siendo mayoría e implican mayores costos de tiempo. El objetivo del pre-procesamiento es transformar los datos de entrada a un formato apropiado para un posterior análisis, esto implica limpiar los datos para eliminar posible ruido, quitar datos duplicados y seleccionar los registros y/o características que sean relevantes para el *Data Mining*. Por otro lado, la evaluación de los resultados de la minería de datos o post-procesamiento asegura que los resultados sean válidos y útiles para sacar conclusiones que proporcionen un nuevo conocimiento [12], que es el objetivo final del proceso *KKD*.

2.8. Data Mining

Actualmente la tecnología es barata y los sistemas de gestión de bases de datos son capaces de trabajar con cantidades masivas de datos. Las organizaciones grandes disponen de cantidades enormes de datos almacenados, análogos al tamaño de una mina, llena de información en bruto, que sí es extraída y procesada se transforma en información útil, lo que es igual a conocimiento. Para abordar esta problema podemos recurrir a una tecnología reciente denominada *Data Mining* o Minería de Datos [26].

No existe una definición única del término *Data Mining*, una de estas expresa: “integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión” [11], en otras palabras es el proceso de analizar datos desde diferentes perspectivas con el objetivo de resumir los datos en segmentos de información útil [15]. El *Data Mining* es de carácter multidisciplinar, es decir, se basa en ideas de diferentes áreas como la estadística, la inteligencia artificial, algoritmos de reconocimiento de patrones, *Machine Learning*, etc. y usa como materia prima las bases de datos [12].

Los modelos del *Data Mining* se clasifican en función de su propósito general en, **Modelos descriptivos** que buscan obtener patrones (correlaciones, tendencias, *clusters* y anomalías) que resuman la subyacente relación de los datos, y **Modelos Predictivos** que tienen el objetivo de predecir el valor de atributos, conocidos como variables objetivo, basados en los valores de otros atributos mejor conocidos como variables explicativas [27]

Los modelos predictivos realizan tareas que se separan según la naturaleza del resultado que se quiere obtener, en **Tareas de Clasificación** usadas para predecir variables categóricas, y **Tareas de Regresión** para predecir variables continuas. Ambos modelos tiene como objetivo minimizar el error entre los valores predichos y los valores reales de la variable objetivo [12]. Dado el objetivo de esta memoria de título, el enfoque se centra en las técnicas de clasificación de los modelos predictivos.

2.8.1. Técnicas de clasificación

La clasificación es la tarea de aprender una función objetivo f que, a partir de un conjunto de entrada, del cual se conoce a que categoría pertenece cada variable, sea capaz de predecir a que categoría pertenecen nuevas variables sin una categorización previa. La función objetivo se conoce informalmente como un **Modelo de Clasificación** [12].

Una tarea de clasificación es un enfoque sistemático para construir modelos de clasificación a partir de un conjunto de datos de entrada. Algunas de estas técnicas incluyen: árboles de clasificación, clasificadores basados en reglas, redes neuronales, etc. Cada técnica emplea algoritmos de aprendizaje para identificar el modelo que mejor se ajuste a la relación entre el conjunto de variables objetivo y la categoría predefinida del conjunto de entrada [12]. El modelo

además de ajustarse bien a la base de entrada, debe predecir correctamente a que clase, definida previamente, pertenecen registros nuevos no clasificados.

En la **Figura 3**, se ilustra el enfoque general para resolver problemas de clasificación. Un set de entrenamiento consiste en registros con categoría conocida, este es usado para construir un modelo de clasificación, el cual subsecuentemente es aplicado a un set de prueba, que consiste en registros con categorías desconocidas.

La evaluación del rendimiento de un modelo de clasificación se basa en el conteo de variables dentro del set de prueba que fueron predichas correcta e incorrectamente por el modelo.

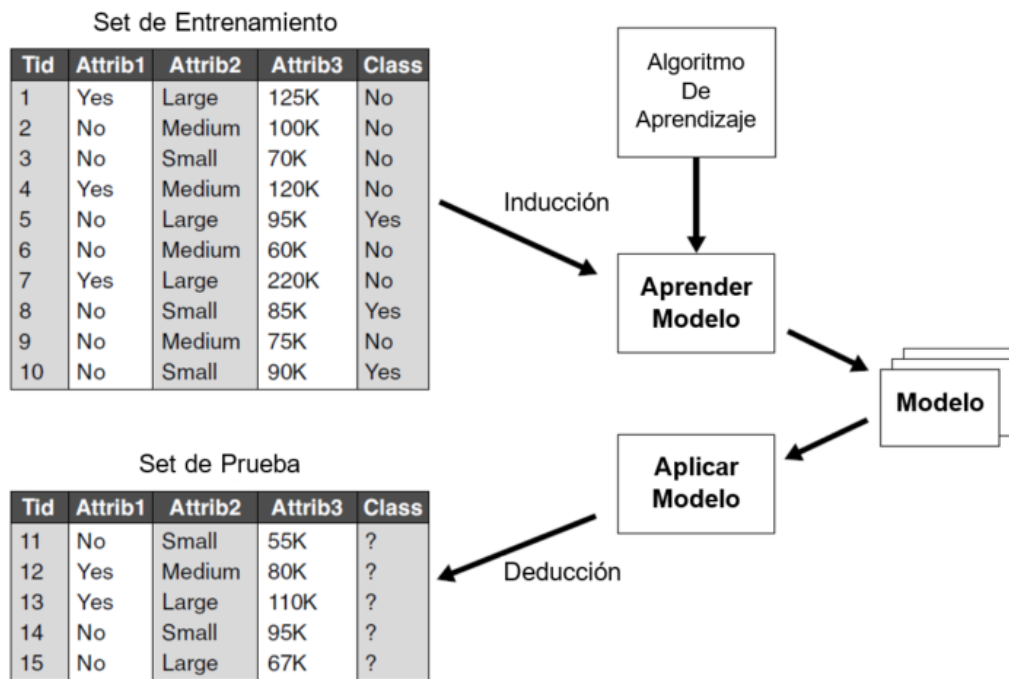


Figura 3: Enfoque general para la construcción de un modelo de clasificación

Las predicciones del *Data Mining* deben ser inspeccionadas por profesionales familiarizados con la materia de trabajo, de manera que comprendan y verifiquen lo que se ha producido. Hay que encontrar un término medio entre la capacidad explicativa del modelo y su poder de predicción, conforme aumenta el poder de predicción del modelo baja su capacidad de interpretar el fenómeno objeto de estudio y viceversa [15]. Por ejemplo, los árboles de decisión conducen a modelos de fácil interpretación, pero tienen un bajo poder predictivo porque las decisiones son tomadas al contestar a preguntas de tipo binario si-no. Por otro lado, las redes neuronales tienen un gran poder predictivo, pero resulta difícil asignar una interpretación a su funcionamiento.

2.8.2. Medidas de desempeño

La problemática al utilizar diferentes algoritmos de *Data Mining* es escoger entre diferentes modelos que buscan describir el mismo conjunto de datos.

Se debe seleccionar la mejor combinación de parámetros para el algoritmo de aprendizaje, para esto la meta es optimizar una medida de desempeño del algoritmo en un conjunto de datos independientes. Las medidas de desempeño permiten evaluar de manera cuantitativa si uno de los modelos ajustados es mejor que otro. Estas medidas son calculadas en función de la matriz de confusión asociada al modelo.

Una **matriz de confusión** es una herramienta que permite la visualización del desempeño de un algoritmo de clasificación. Contiene información acerca de la clasificación real y la predicha, es de tamaño $m \times m$, donde m es el número de diferentes categorías. En la **Figura 4** se muestra una matriz de confusión para $m = 2$ [28]:

Observación \ Predicción	Positivo	Negativo
	Positivo	Verdaderos Positivos [VP]
Negativo	Falsos Positivos [FP]	Verdaderos Negativos [VN]

Figura 4: Matriz de confusión para $m = 2$

Las entradas en la matriz de confusión tienen el siguiente significado:

- Verdaderos positivos (VP): Casos positivos que el modelo identifica como tales.
- Verdaderos negativos (VN): Casos negativos que el modelo identifica como tales.
- Falsos positivos (FP): Casos negativos que el modelo identifica como positivos.
- Falsos negativos (FN): Casos positivos que el modelo identifica como negativos.

En base a la matriz de confusión se definen las siguientes medidas de desempeño:

- Precisión: Es la razón de predicciones que fueron correctamente clasificadas por el modelo.

$$Precisión = \frac{VP + VN}{VP + FP + VN + FN} \quad (53)$$

- Sensibilidad: Probabilidad de obtener un resultado positivo cuando la observación es efectivamente positiva.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (54)$$

- Especificidad: Probabilidad de obtener un resultado negativo cuando la observación es efectivamente negativa.

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (55)$$

La sensibilidad también suele llamarse como razón de verdaderos positivos (RVP), mientras que al término 1-especificidad se llama razón de falsos positivos (RFP).

Otra herramienta útil para evaluar el desempeño de modelos de clasificación es mediante la utilización de una **curva ROC** (acrónimo de *Receiver Operating Characteristic*, o *Característica Operativa del Receptor* en español). Una curva ROC es una representación gráfica de la sensibilidad (o VPR) frente al valor FPR (1 – especificidad) para un sistema clasificador. Donde al variar alguno de los parámetros del modelo, es posible obtener una curva en el espacio ROC [29].

Así, un espacio ROC se define por RFP y RVP como ejes X e Y respectivamente, y representa los intercambios entre verdaderos positivos y falsos positivos. Donde cada resultado de predicción o instancia de la matriz de confusión representa un punto en el espacio ROC. En la **Figura 5** se muestra el espacio ROC y diferentes puntos dentro de este a modo de ejemplo.

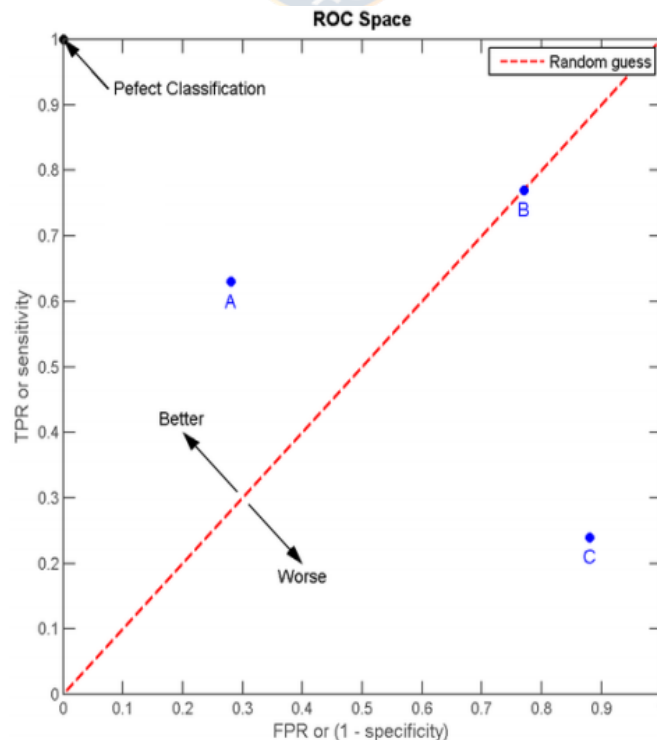


Figura 5: Espacio ROC

En la **Figura 5** el mejor método posible de predicción se situaría en un punto en la esquina superior izquierda, de coordenadas (0,1) que representa un 100% de sensibilidad (ningún falso negativo) y un 100% también de especificidad (ningún falso positivo). Por el contrario, una clasificación totalmente aleatoria daría un punto a lo largo de la línea diagonal (B), desde el extremo inferior izquierdo hasta la esquina superior derecha, que se llama también línea de no-discriminación. La diagonal divide el espacio ROC. Los puntos por encima de la diagonal (A) representan los buenos resultados de clasificación, y los puntos por debajo de la línea representan los resultados pobres.

Las curvas ROC pueden usarse para valorar y comparar cuantitativamente un modelo, calculando el área bajo la curva conseguida, mejor conocida como AUC (*area under the curve*).

2.8.2. Sobreajuste

Un problema que pueden exhibir los modelos del Data Mining es el Sobreajuste o *overfitting*. Se dice que un modelo está sobreajustado cuando predice muy bien el conjunto de entrenamiento, pero tiene baja capacidad de predicción al introducir información nueva. Las principales razones de sobreajuste se deben a que el modelo objetivo es demasiado complejo o que se tienen muchas características pero un número bajo de registros de entrenamiento [30].

Algunas soluciones posibles al problema de sobreajuste son:

- Probar un modelo más simple (red neuronal con menos capas o número de neuronas).
- Reducir la dimensionalidad de los registros en el conjunto de datos.
- Si es posible añadir más registros de entrenamiento.
- Regularizar el modelo.

2.9. Árboles de decisión

Los árboles de decisión son modelos con estructura jerárquica en forma de árbol, que representan conjuntos de decisiones que generan reglas para la clasificación de un conjunto de datos. Cuando la variable objetivo es categórica, como en nuestro caso, se habla de árboles de clasificación [15]. En la **Figura 6** se presenta la estructura básica de un árbol de decisión, donde cada nodo del árbol encaja con un atributo y de este parten tantas ramas como valores distintos tiene ese atributo. En los nodos terminales u hojas, se encuentran las distintas categorías de la variable objetivo.

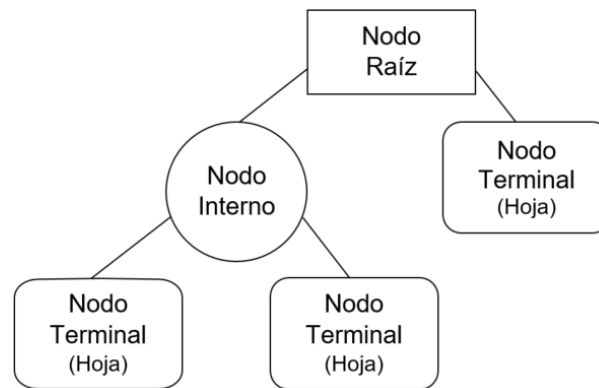


Figura 6: Estructura básica de un árbol de decisión

Dado un árbol de decisión, para clasificar un dato nuevo, este se inspecciona desde la raíz hasta llegar a un nodo terminal. Los nodos raíz e internos, representan interrogantes sobre un atributo y el resultado correspondiente indica la rama del árbol que se debe recorrer hasta alcanzar un nodo terminal, el cual proporciona la categoría a la que pertenece el registro [31].

Son muchos los árboles de decisión que se pueden construir a partir de un conjunto de datos dado, unos más precisos que otros, por lo que se han desarrollado algoritmos eficientes para introducir una precisión razonable, aunque sub-óptima, de los árboles de decisión, en un tiempo razonable [12].

2.9.1. Algoritmo de Hunt

En el algoritmo de Hunt, el árbol de decisión crece de manera recursiva mediante la partición sucesiva del conjunto de entrenamiento en subconjuntos denominados “puros”. Sea S_t el conjunto de datos de entrenamiento asociados al nodo t dado y $\{y_1, \dots, y_c\}$ el conjunto de categorías que puede tomar el registro a clasificar:

Paso 1: Si todos los datos de S_t pertenecen a la misma categoría y_i , entonces t es un nodo terminal que se etiqueta como y_i .

Paso 2: Si S_t contiene datos que pertenecen a más de una categoría, se selecciona un test que conduce a dividir los datos en subconjuntos más pequeños. Un nodo hijo es creado para cada resultado del test y los registros en S_t son distribuidos en los nodos hijos basándose en los resultados del test. El algoritmo es aplicado recursivamente en cada nodo hijo.

Se deben tener en cuenta las siguientes consideraciones:

- Si para alguno de los nodos hijos creados en el paso 2 no existen registros asociados, el nodo se convierte en un nodo terminal con valor igual al de la categoría más frecuente.
- Si todos los registros asociados a S_t tienen valores de los atributos idénticos, pero de distinta categoría, el nodo se convierte en un nodo terminal con valor igual al de la categoría más frecuente.

2.9.2. Medidas para seleccionar la mejor división

Existen diferentes medidas que pueden ser utilizadas para determinar la mejor manera de dividir el conjunto de datos en la construcción del árbol de decisión. Estas medidas se definen en términos de las distribuciones de los datos antes y después de realizar la división. Las medidas desarrolladas para seleccionar la mejor división de los datos son basadas en la minimización del “grado de impureza” de los nodos hijos, es decir, cuan desordenados o mezclados quedan los nodos una vez divididos. La medición del grado de impureza se puede realizar a través de:

$$Entropía(t) = - \sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t), \quad (56)$$

$$Gini(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2, y \quad (57)$$

$$Error de Clasificación (t) = 1 - \max_{1 \leq i \leq c} [p(i|t)], \quad (58)$$

donde $p(i|t)$ es la fracción de datos pertenecientes a la categoría i en el nodo t , c es el número de categorías y, $0 \log_2 0 = 0$ en los cálculos de entropía. Entre menor sea el grado de impureza, más sesgada es la distribución de la categoría [12].

Para seleccionar el atributo con el que se subdivide el conjunto de datos, se debe comparar el grado de impureza del nodo padre (antes de la subdivisión) con el de los nodos hijos (después de la subdivisión). Entre mayor sea esta diferencia, mejor es test que condiciona los atributos. La ganancia Δ , es un criterio que puede ser usado para determinar la bondad de la partición [31]:

$$\Delta(S, A) = I(S) - \sum_{j=1}^k \frac{N(S_j)}{N} I(S_j), \quad (59)$$

donde A es un atributo con k valores distintos que dividen el conjunto de registros de entrenamiento S en los subconjuntos de registros S_1, \dots, S_k . N y $N(S_j)$, con $j = 1, \dots, k$, son los tamaños asociados con el nodo padre y los nodos hijos respectivamente.

Luego, se selecciona el atributo con el que se obtiene la mayor ganancia tras la división. Si se utiliza la entropía como medida de impureza en la ecuación (59) se habla de ganancia de información, Δ_{info} [12].

2.9.3. Algoritmo ID3

El sistema ID3 es un algoritmo, desarrollado por Ross Quinlan [32], para la elaboración de un árbol de decisión. El procedimiento consiste en seleccionar un atributo como raíz del árbol y crear una rama con cada uno de los posibles valores de dicho atributo. Con cada rama resultante (nuevo nodo del árbol), se realiza el mismo proceso, esto es, se selecciona otro atributo y se genera una nueva rama para cada posible valor del atributo. Este procedimiento continúa hasta que los registros se clasifiquen a través de uno de los caminos del árbol. En la **Figura 7** se esquematiza el algoritmo ID3 [33]:

1. *Seleccionar* el atributo A_i que maximice la ganancia $\Delta(A_i)$.
2. *Crear* un nodo para ese atributo con tantos nodos hijos como categorías tenga.
3. *Introducir* los datos de entrenamiento en nodos hijos según la categoría que tenga el atributo A_i .
4. Por cada nodo hijo:
 - a. Si sólo hay ejemplos de una clase, C_k , entonces etiquetarlo con C_k .
 - b. Si no, llamar a ID3 con una tabla formada por los datos de entrenamiento de ese nodo, eliminando la columna del atributo A_i .

Figura 7: Pseudocódigo del algoritmo ID3

2.9.4. Algoritmo C4.5

Quinlan propone una mejora al algoritmo anterior con el algoritmo C4.5 este permite [33, 34] :

1. Empleo del concepto razón de ganancia.

2. Construir árboles de decisión cuando algunos de los registros presentan valores desconocidos para algunos de los atributos.
3. Trabajar con atributos que presenten valores continuos.
4. La *poda* de los árboles de decisión.
5. Obtención de Reglas de Clasificación.

i. Razón de Ganancia

El test basado en el criterio de maximizar la ganancia tiene como sesgo la elección de atributos con muchos valores. C4.5 modifica el criterio de selección del atributo empleando la “*razón de ganancia*”, evitando así que las variables con mayor número de categorías salgan beneficiadas en la selección [35], así:

$$\text{Razón de Ganancia} = \frac{\Delta_{info}}{\text{Split Info}} \quad (60)$$

Con:

$$\text{Split Info} = - \sum_{j=1}^n \frac{|N(S_j)|}{N} \log_2 \left(\frac{|N(S_j)|}{N} \right), \quad (61)$$

donde n es el número total de particiones y Split Info representa la información potencial generada al dividir S en n subconjuntos, mientras que la ganancia de información, ecuación (59), mide la información relevante para la clasificación que surge de la misma división. Por lo tanto, la ***Razón de Ganancia*** expresa la proporción de información generada por la partición que es útil para la clasificación.

ii. Valores Desconocidos

C4.5 admite datos con atributos desconocidos tanto en el proceso de aprendizaje como en el de validación. Para calcular la razón de ganancia de un atributo con valores desconocidos, durante el proceso de aprendizaje, se redefinen las ecuaciones (59) y (61):

$$\Delta(S, A) = \frac{N_c(S_j)}{N} \left(I(S) - \sum_{j=1}^k \frac{N(S_j)}{N} I(S_j) \right), \quad (62)$$

$$\text{Split Info} = - \left(\sum_{i=1}^n \frac{|D_i|}{D} \log_2 \left(\frac{|D_i|}{D} \right) \right) - \frac{N_d(S_j)}{N} \log_2 \left(\frac{N_d(S_j)}{N} \right). \quad (63)$$

aquí N_c y N_d son el número de datos con atributo conocido y desconocido respectivamente. Para el cálculo de las entropías se tienen en cuenta únicamente los datos con atributos que tengan un valor conocido.

En el entrenamiento, los datos con valores desconocidos se distribuyen con pesos de acuerdo a la frecuencia de aparición de cada posible valor del atributo en el resto del conjunto de entrenamiento. El peso ω_{ij} con que un registro i se distribuiría desde un nodo clasificado con el atributo A hacia el hijo con valor j en dicho atributo se calcula mediante la ecuación (64), en la que ω_i es el peso del dato i al llegar al nodo, esto es, antes de distribuirse, y $p(A = j)$ la suma de pesos de todos los datos del nodo con valor j en el atributo A entre la suma total de pesos de todos los datos del nodo (ω).

$$\omega_{ij} = \omega_i p(A = j) = \omega_i \frac{\omega_{A=j}}{\omega}. \quad (64)$$

iii. Atributos Continuos

Si un atributo continuo A_i presenta los valores ordenados v_1, v_2, \dots, v_n , se comprueba cuál de los valores $z_j = (v_j + v_{j+1})/2$; $1 \leq j < n$, supone una ruptura del intervalo $[v_1, v_n]$ en dos subintervalos $[v_1, z_j]$ y $(z_j, v_n]$ con mayor ganancia de información. El atributo continuo, ahora con dos únicos valores posibles, entrará en competencia con el resto de los atributos disponibles para expandir el nodo.

iv. Poda del árbol de decisión

El árbol de decisión ha sido construido a partir de un conjunto de datos, por tanto, refleja correctamente este grupo. Sin embargo, como esos ejemplos pueden ser muy diferentes entre sí, el árbol resultante puede llegar a ser bastante complejo. Para facilitar la comprensión del árbol puede realizarse una *poda*. El algoritmo C4.5 realiza la poda después de haber desarrollado el árbol (*post-poda*). El proceso comienza en los nodos hoja y recursivamente continúa hasta llegar al nodo raíz. Se consideran dos operaciones de poda ilustradas en la **Figura 8** [33].

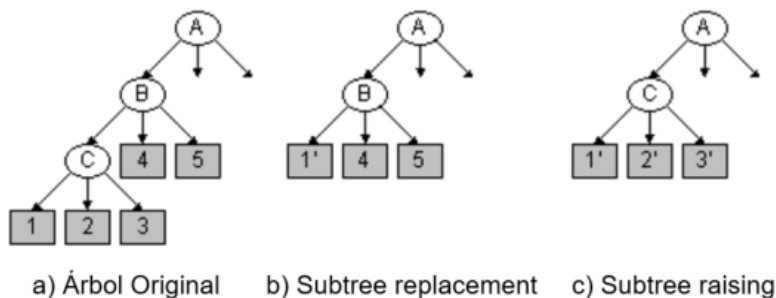


Figura 8: Tipos de operaciones de post-poda

En la **Figura 8.b** se realiza *subtree replacement*, donde el nodo C es reemplazado por uno de sus subárboles. En **Figura 8.c** se realiza *subtree raising*, donde el nodo B es sustituido por el subárbol con raíz C, en este caso hay que reclasificar los registros de entrenamiento a partir del nodo C. El proceso de podado comienza en las hojas y continúa hacia la raíz, pero el dilema es cómo reemplazar un nodo interno por una hoja (*replacement*) o reemplazar un nodo interno por uno de sus nodos *hijo* (*raising*). Para esto se compara el error estimado de clasificación en el nodo en cuestión y se compara con el error en cada uno de sus hijos y en su padre para realizar alguna de las operaciones o ninguna.

El error de clasificación se estima basándose en los propios datos de entrenamiento. Para ello, en el nodo donde se quiere estimar el error de clasificación, se toma la clase mayoritaria de sus ejemplos como clase representante. Esto implica que habrá E errores de clasificación de un total de N ejemplos que se clasifican a través de dicho nodo. El error observado será $f = E/N$, siendo q la probabilidad de error de clasificación del nodo y $p = 1 - q$ la probabilidad de éxito. Se supone que la función f sigue una distribución binomial de parámetro q . Lo que se desea obtener es el error e , probabilidad del extremo superior con un intervalo $[f - z, f + z]$ de confianza c . Dado que se trata de una distribución binomial, e se obtiene mediante las siguientes ecuaciones:

$$P\left[\frac{f - q}{q(1 - q)/N} \leq z\right] = c, \quad (65)$$

$$e = \left(\frac{f + \frac{z^2}{2N} + z\sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}}\right). \quad (66)$$

Como factor de confianza c se suele emplear en C4.5 el 25%, dado que es el que mejores resultados suele dar y que corresponde a un $z = 0.69$ [33].

v. **Obtención de Reglas de Clasificación**

Cualquier árbol de decisión se puede convertir en reglas de clasificación. El algoritmo de generación de reglas consiste en, por cada rama del árbol, las preguntas y sus valores estarán en la parte izquierda de las reglas y la etiqueta del nodo hoja correspondiente en la parte derecha. El algoritmo C4. 5 [34] realiza un podado de las reglas obtenidas. Ver **Figura 9**.

En cuanto a la estimación del error, se realiza del mismo modo que para realizar el podado del árbol de decisión.

```

ObtenerReglas (árbol) {
  Convertir el árbol de decisión (árbol) a un conjunto de reglas, R
  error = error de clasificación con R
  Para cada regla Ri de R Hacer
    Para cada precondición pj de Ri Hacer
      nuevoError = error al eliminar pj de Ri
      Si nuevoError <= error Entonces
        Eliminar pj de Ri
        error = nuevoError
    Si Ri no tiene precondiciones Entonces
      Eliminar Ri
}

```

Figura 9: Pseudocódigo del algoritmo de obtención de reglas de C4.5

2.9.5. Algoritmo C5.0

C5.0 es el algoritmo sucesor de C4.5, el cual incluye todas las funcionalidades de C4.5 con ciertas características adicionales. Se destacan las siguientes mejoras [36]:

- Conjunto de reglas que ocupan menos memoria y entrenan con mayor rapidez.
- Incremento en velocidad para entrenamiento con la misma precisión que C4.5.
- “*Boosting*”: técnica para generar y combinar múltiples clasificadores con el objetivo de mejorar la precisión predictiva. Reducción de la tasa de error para el conjunto total de datos.
- C4.5 trata a todos los errores de la misma forma, mientras que C5.0 permite la definición de un coste distinto para cada pareja de valores {predicho, real} de una clase. Si usamos esta opción, C5.0 construye un clasificador que minimiza los costes esperados de una clasificación errónea, en vez de minimizar las tasas de error.
- Una ventana de referencia-cruzada permite enlazar un registro con un componente del clasificador que sea relevante.

2.10. Redes neuronales artificiales

Una **Red Neuronal Artificial** (RNA) es un modelo matemático o computacional inspirado en la red neuronal biológica, en otras palabras, es una emulación del sistema nervioso central, el cual se relaciona con la manera en que el cerebro responde a los estímulos del mundo exterior y la experiencia que se obtiene de estos. Consiste en un sistema interconectado de neuronas artificiales que funciona de manera adaptativa, cambiando su estructura en base a información que fluye a través de la red durante una fase de aprendizaje, hasta alcanzar una salida deseada [37, 38].

La neurona biológica es la unidad funcional y estructural del sistema nervioso y, como se observa en el modelo simplificado de la **Figura 10**, se encuentra formado por el cuerpo celular o soma, una prolongación llamada axón, un conjunto de ramificaciones terminales llamadas dendritas y zonas de conexión entre una neurona y otra, conocidas como sinapsis. La función principal de las neuronas es la transmisión de impulsos nerviosos. Estos viajan por la neurona comenzando por las dendritas hasta llegar a las terminaciones de axón, donde pasan a otra neurona por medio de la conexión sináptica [39].

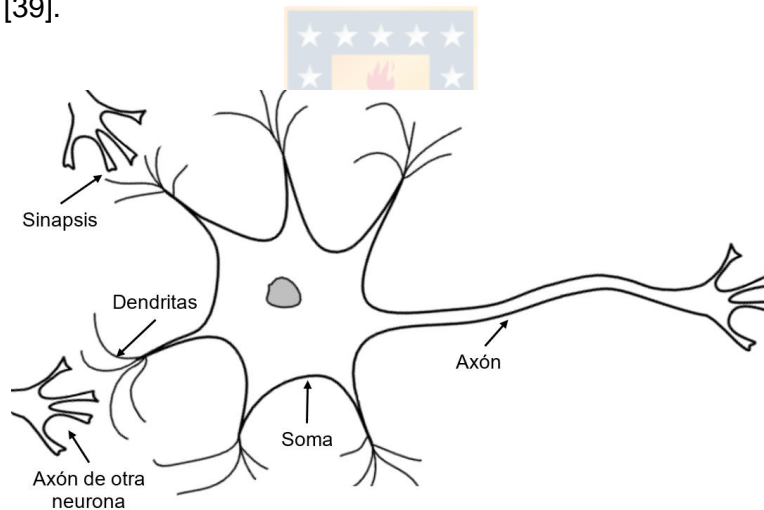


Figura 10: Modelo simplificado de una red neuronal biológica

La principal similitud entre una neurona biológica y una neurona artificial es que ambas adquieren conocimiento a través del aprendizaje. En la **Figura 11** se muestra un modelo de neurona artificial i junto con la analogía existente con una neurona biológica, esta consta de:

- Un conjunto de entradas x_1, \dots, x_n ,
- Pesos sinápticos w_1, \dots, w_n correspondientes a cada entrada,
- Una función de agregación, Σ ,
- Una función de activación, f , y
- Una salida y .

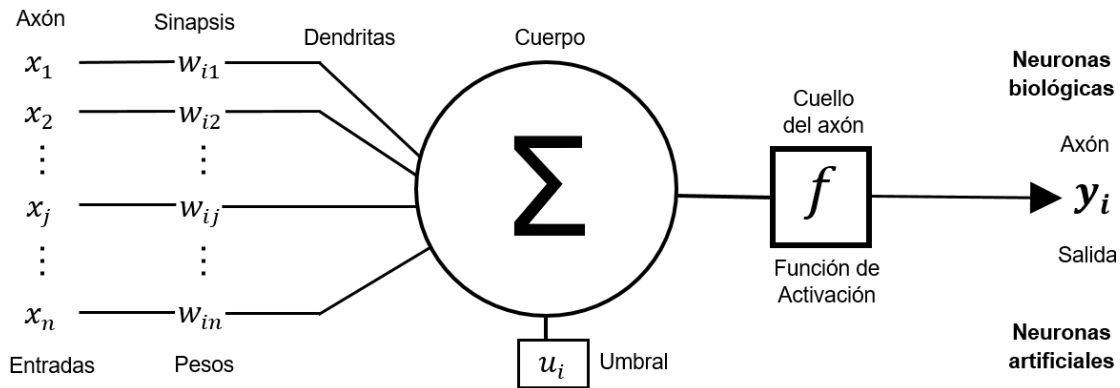


Figura 11: Estructura de una neurona artificial

La neurona artificial puede ser vista como un diagrama donde la neurona se representa por un nodo y las líneas junto con los pesos son las conexiones entre las entradas y la salida de la neurona. La neurona recibe información en forma de vector, cada entrada x_n , donde n representa el número de entradas, es multiplicada por un correspondiente peso w_n , estos pesos representan la fuerza de las interconexiones de la red. Las entradas ponderadas por sus respectivos pesos son incorporadas a la neurona mediante la función de agregación Σ , que comúnmente corresponde a la sumatoria de estas [40]. Al igual que las neuronas biológicas, las neuronas artificiales tienen diferentes estados de activación, los cuales son calculados con la función de activación, transformando la entrada global, menos el umbral u_i , en un valor de activación, cuyo rango normalmente va de (0 a 1) o de (-1 a 1), pudiendo estar totalmente inactiva (0 o -1) o totalmente activa (1) [41]. Algunos ejemplos de funciones de activación son:

1. Logística:

$$f(a) = \frac{1}{1 + e^{-a}}, \quad (67)$$

con un rango de salida que va de 0 a 1.

2. *Softmax*:

$$f(a) = \frac{e^{-a}}{\sum_i a_i}, \quad (68)$$

con un rango de salida que va de 0 a 1.

Finalmente, la salida de neurona i queda dada por:

$$y_i = f\left(\sum_j x_j \cdot w_{ij} + u_i\right). \quad (69)$$

Las neuronas se distribuyen dentro de la red formando capas, con un número determinado de dichas neuronas en cada una de ellas [41]. Como se ilustra en la **Figura 12** se distinguen 3 tipos de capas dependientes de su posición en la red:

1. Capa de entrada: es aquella que recibe información proveniente de las fuentes externas de la red.
2. Capas ocultas: son internas a la red y no tienen contacto directo con el entorno exterior. El número de capas ocultas puede estar entre cero y un número elevado. Las neuronas de las capas ocultas pueden estar interconectadas de distintas maneras, lo que determina las distintas arquitecturas de redes neuronales.
3. Capa de salida: transfieren información de la red hacia el exterior.

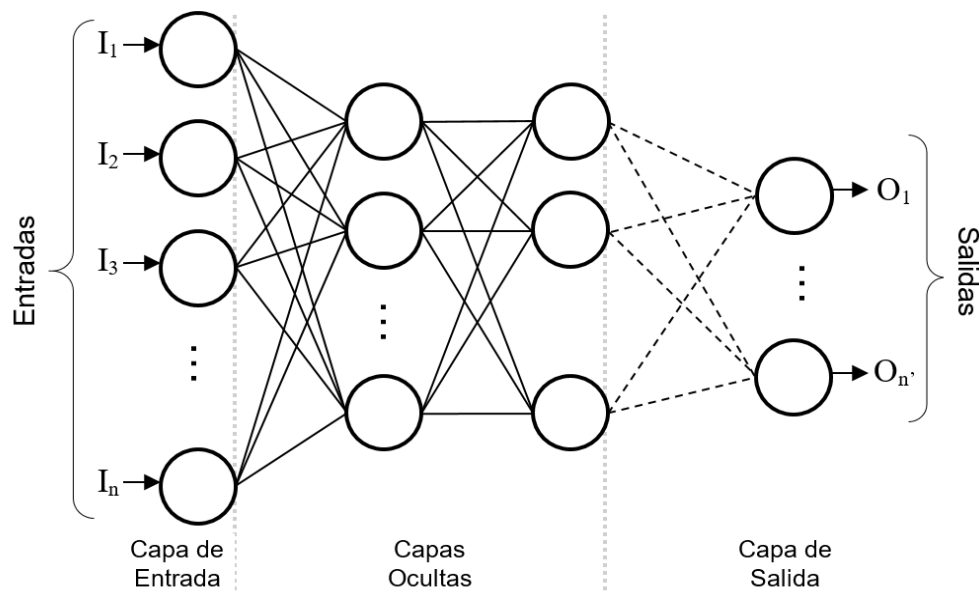


Figura 12: Ejemplo de red neuronal multicapa

2.10.1. Mecanismos de aprendizaje

El propósito de que las redes neuronales procesen datos de entrada es que se obtenga una salida deseada. Semejante a una red neuronal biológica, las redes neuronales artificiales son capaces de detectar y aprender patrones complejos y características dentro de los datos, aprendiendo de la experiencia y aplicando tal conocimiento a la resolución de problemas nuevos, en otras palabras, una red neuronal debe aprender a calcular la salida correcta para cada vector de entrada de un conjunto de registros. El proceso de aprendizaje se conoce como entrenamiento y consiste en alimentar la red con un conjunto de datos de entrenamiento y adaptar los pesos de

acuerdo a una regla de aprendizaje. Una vez entrenadas las redes neuronales son capaces de hacer predicciones, clasificaciones y segmentaciones [33, 38, 41].

Durante el proceso de aprendizaje, los pesos de las conexiones de la red sufren modificaciones, y se puede afirmar que este proceso ha terminado, o sea que la red ha aprendido, cuando los valores de los pesos permanecen estables ($d w_{ij}/dt = 0$). Una generalización de la regla para cambiar los pesos de la red es:

$$\text{Peso Nuevo} = \text{Peso Viejo} + \text{Cambio de Peso}, \quad (70)$$

matemáticamente:

$$w_{ij}(t + 1) = w_{ij}(t) + \Delta w_{ij}, \quad (71)$$

donde t hace referencia a la etapa de aprendizaje, $w_{ij}(t + 1)$ al peso nuevo y $w_{ij}(t)$ al peso viejo.

Existen dos métodos de aprendizaje importantes:

1. **Aprendizaje supervisado:** Se caracteriza por realizar el entrenamiento controlado por un supervisor, quien determina la respuesta que debería generar la red a partir de una entrada determinada. En caso de que esta no coincida con la deseada, se procede a modificar los pesos de las conexiones, con el fin de conseguir que la salida obtenida se aproxime a la deseada. Este tipo de aprendizaje considera, a su vez, tres formas de llevarlo a cabo:
 - a. Aprendizaje por corrección del error.
 - b. Aprendizaje por refuerzo.
 - c. Aprendizaje estocástico.

2. **Aprendizaje no supervisado:** El entrenamiento no requiere de influencia externa para ajustar los pesos, ni tampoco recibe información que le indique si la salida generada en respuesta a una determinada entrada es correcta o no. La red debe encontrar características estadísticamente destacadas dentro de la población de entrada. Existen varias posibilidades de interpretación de la salida de la red, que dependen de su estructura y el algoritmo de aprendizaje empleado, como, por ejemplo, en algunos casos la salida representa el grado de similitud entre la información que se le está presentando en la entrada y las informaciones que se le han presentado en el pasado, en otros casos se podría realizar *clustering* o establecimiento de categorías.

Para efectos de este trabajo, es de preciso interés los mecanismos de aprendizaje supervisado, particularmente los efectuados por corrección del error.

Aprendizaje por corrección de error

Consiste en ajustar los pesos de las conexiones de la red en función de la diferencia entre los valores deseados y los obtenidos a la salida de la red, es decir, en función del error cometido en la salida.

Un ejemplo de este tipo de algoritmos es la **regla de aprendizaje del Perceptrón**², en la cual a cada neurona de la capa de salida se le calcula la desviación a la salida deseada o el error δ , el cual se usa para cambiar los pesos sobre la conexión de la neurona precedente. Considérese las neuronas N_i e N_j , mostradas en la **Figura 13**, con entradas x_i e x_j y salidas y_i e y_j , respectivamente y con w_{ij} representando el peso de la conexión entre ambas neuronas.

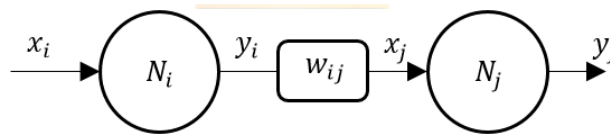


Figura 13: Influencia de la salida de la neurona N_i en la entrada de la neurona N_j

Sea \bar{y} la salida deseada, el cambio de los pesos por medio de la regla de aprendizaje del Perceptrón se realiza según la siguiente regla:

$$\Delta w_{ij} = \sigma \cdot y_i \cdot \delta_j \quad \text{con } \delta_j = \bar{y} - y_j \quad (72)$$

con σ representando el aprendizaje de la red.

Otro algoritmo perteneciente a esta clasificación es la **regla de aprendizaje Delta** o **regla del mínimo error cuadrado**, tal como el algoritmo anterior utiliza la desviación a salida deseada, pero además considera todas las neuronas predecesoras que tiene la neurona de salida, permitiendo cuantificar el error global cometido en cualquier momento durante el proceso de entrenamiento de la red, entregando así más información sobre el error y por ende acelerando el proceso de aprendizaje. Luego el error δ calculado es igualmente repartido entre las conexiones de las neuronas predecesoras.

² El Perceptrón es una red neuronal compuesta por dos capas, una entrada y una de salida, sin capas ocultas.

Un último ejemplo para esta clasificación es la **regla de propagación hacia atrás**, más conocida como ***back propagation***, o como regla Delta multicapa, es una generalización de la regla de aprendizaje Delta pero que permite realizar cambios sobre los pesos en las conexiones de las capas ocultas. Se discutirá sobre este mecanismo en la sección subsiguiente.

Antes de comenzar el proceso de entrenamiento se debe determinar el estado inicial de la red, lo que implica escoger un conjunto inicial de pesos para las diferentes conexiones entre las neuronas de la red. Esto se puede realizar por varios criterios, siendo el más utilizado el otorgar un peso aleatorio a cada conexión, encontrándose los mismos dentro de cierto intervalo.

Para determinar cuándo se detendrá el proceso de aprendizaje, es necesario establecer una condición de detención. Normalmente el entrenamiento se detiene cuando el cálculo del error cuadrático sobre todo el conjunto de entrenamiento ha alcanzado un mínimo o cuando para cada uno de los ejemplos dados, el error observado está por debajo de un determinado umbral, otra condición de detención puede ser cuando se han corrido un cierto número de pasos de entrenamiento completamente.

Luego de alcanzarse la condición de detención, los pesos no se volverán a cambiar, así la transformación de los datos de entrada a los de salida está completa, interpretándose la red como una función f oculta, que funciona como la “receta” de cómo la salida será calculada a partir de un vector de entrada.

El paso siguiente al entrenamiento consiste en comprobar si la red neuronal puede resolver la predicción de registros nuevos para los cuales ha sido entrenada. Con el propósito de validar la red neuronal se requiere de otro conjunto de datos, denominado conjunto de validación. Cada ejemplo del conjunto de validación contiene los valores de las variables de entrada, con su correspondiente solución, pero en esta instancia no se le será otorgada a la red neuronal. Luego se compara la solución calculada para cada registro de validación con la solución conocida.

2.10.2. Arquitectura de redes neuronales

La arquitectura de una red neuronal consiste cómo se organizan las neuronas para formar capas. Los parámetros fundamentales de la red son: el número de capas, el número de neuronas por capa y el tipo de conexiones entre neuronas.

Redes Monocapa: Se establecen conexiones entre neuronas que pertenecen a una única capa que constituye la red. Se utilizan en tareas relacionadas con la autoasociación (regenerar información de entrada que se presenta de forma incompleta).

Redes Multicapa: Se disponen de conjuntos de neuronas agrupadas en varias capas. Normalmente, todas las neuronas de una capa reciben señales de entrada de la capa antecesora y envían señales de salida a la capa inmediatamente posterior, estas conexiones se denominan conexiones hacia adelante o *feedforward*. Sin embargo, existe la posibilidad de conectar la salida de neuronas de capas posteriores a la entrada de capas anteriores, estas conexiones se denominan conexiones hacia atrás o *feedback*.

Para la finalidad de este trabajo se consideran redes multicapa con conexiones hacia adelante, las cuales son entrenadas por medio de mecanismos de aprendizaje supervisado, específicamente el algoritmo de retropropagación que se detalla en la siguiente sección.

2.10.3. Algoritmo de retropropagación

Como se ha mencionado anteriormente, el algoritmo de retropropagación o *back-propagation*, es un método común que utiliza el mecanismo de aprendizaje supervisado para enseñar a una red neuronal, *feedforward* multicapa, a realizar una tarea específica.

Se distinguen dos etapas dentro del algoritmo, primero se aplica un patrón de entrada por la red y la salida se compara con la salida deseada calculándose el error cometido por cada neurona de salida, estos errores se transmiten hacia atrás, partiendo de la capa de salida, hacia todas las capas intermedias. Cada neurona recibe un error proporcional sobre el error total de la red. Basándose en el error recibido, se ajustan los errores de los pesos sinápticos de cada neurona [42]. La deducción del algoritmo se encuentra en el anexo A, y se resume en la **Figura 14** [33].

El éxito y velocidad de convergencia de este mecanismo de entrenamiento tiene un alto grado de dependencia de la configuración del mismo. Por ello se han realizado una serie de métodos que permiten mejorar significativamente la performance del algoritmo.

- **Presentación de los ejemplos.** La aleatorización del orden en que se presentan los ejemplos en los distintos ciclos de entrenamiento evita que los resultados se vean distorsionados por el orden de los ejemplos.

- **Función de activación.** El uso de una función de activación adecuada puede acelerar notoriamente el tiempo de entrenamiento.
- **Normalización de las entradas.** Si bien el algoritmo no exige que los valores de entrada se encuentren normalizados, esta es una buena técnica para acelerar los tiempos de entrenamiento. La normalización de las entradas debe realizarse de manera tal que el valor medio de la misma se encuentre cercano a cero.
- **Pre-procesamiento de los ejemplos.** Si un atributo sólo puede tomar N valores diferentes, la entrada de la red puede subdividirse en N entradas, cada una de las cuales representa a una clase. Cada una de estas entradas ahora puede tomar dos valores, verdadero o falso. Esta técnica puede ayudar a mejorar los tiempos de entrenamiento de la red.

Paso 1: *Inicialización aleatoria de los pesos.*

Paso 2: *Dado un patrón del conjunto de entrenamiento $(x, \bar{y}(x))$, se presenta el vector x a la red y se calcula la salida de la red para dicho patrón, $y(x)$.*

Paso 3: *Se evalúa el error $e(x)$ cometido por la red.*

Paso 4: *Se modifican todos los parámetros de la red utilizando:*

$$w_{ij}(t+1) = w_{ij}(t) + [\alpha \delta_{pj} y_{pj} + \beta \Delta w_{ij}(t)]$$

Donde:

$$\delta_{pj} = \begin{cases} (\bar{y}_{pj} - y_{pj}) f'_j(\sum_j x_j(t) \cdot w_{ij}) & \text{si } j \text{ es una neurona de salida} \\ (\sum_k \delta_{pk} \cdot w_{kj}) f'_j(\sum_j x_j(t) \cdot w_{ij}) & \text{si } j \text{ es una neurona oculta} \end{cases}$$

**Notaciones explicadas en anexo A.*

Paso 5: *Se repiten los pasos 2, 3 y 4 para todos los patrones de entrenamiento, completando así un ciclo de aprendizaje.*

Paso 6: *Se realizan n ciclos de aprendizaje (paso 2, 3, 4 y 5) hasta que se verifique el criterio de detención establecido.*

Figura 14: Pseudocódigo algoritmo Back-propagation

2.11. Recursos minerales

Un **Recurso Mineral** es una concentración u ocurrencia de material de interés económico intrínseco en o sobre la corteza de la Tierra en forma y cantidad en que haya probabilidades razonables de una eventual extracción económica. Los recursos minerales se subdividen en orden de confianza geológica ascendente, en categorías de Inferidos, Indicados y Medidos [43].

Un **Recurso Mineral Inferido** es aquella parte de un Recurso Mineral en la que el tonelaje, ley y contenido de mineral se puede estimar con un bajo nivel de confianza.

Un **Recurso Mineral Indicado** es aquella parte de un Recurso Mineral en la que el tonelaje, densidad, forma, características físicas, ley y contenido de mineral pueden estimarse con nivel razonable de confianza.

Un **Recurso Mineral Medido** es aquella parte de un Recurso Mineral para el cual puede estimarse con un alto nivel de confianza el tonelaje, su densidad, forma, características físicas, ley y contenido de mineral.

Existen varios criterios a considerar para definir a qué categoría pertenece un determinado bloque, los cuales se clasifican en criterios geológicos, geométricos y geoestadísticos, los cuales idealmente deberían combinarse de manera de obtener una clasificación más acertada [44]. En el presente trabajo se utiliza un criterio geoestadístico, específicamente la varianza relativa de *kriging* que consiste en dividir la varianza de cada bloque por el cuadrado de la ley estimada, proporcionando una medida de incertidumbre estandarizada y sin dimensiones que se utiliza para la clasificación de cada bloque [45].

$$\text{Varianza Relativa} = \frac{\text{Varianza de kriging del bloque estimado}}{(\text{ley media del bloque estimado})^2}. \quad (73)$$

Asumiendo que la distribución de los valores $[Z(x) - Z^*(x)]$ es una distribución normal, donde $Z(x)$ es el valor de la ley real del bloque y $Z^*(x)$ es el valor estimado, la expresión (73) puede ser expresada en función de un nivel de confianza seleccionado. Por ejemplo, para un 95% de confianza la expresión anterior queda expresada como:

$$\text{Varianza Relativa} = \frac{1.96 \times \text{Varianza de kriging del bloque estimado}}{(\text{ley media del bloque estimado})^2}. \quad (74)$$

CAPÍTULO 3

METODOLOGÍA GENERAL

El estudio de la aplicabilidad de los algoritmos provenientes del *Data Mining* en el modelamiento geológico es aplicado en dos escenarios, un caso de estudio simulado, en el que se conoce en detalle toda la extensión del yacimiento, y un caso de estudio real a partir de un sondeo de un yacimiento de hierro. En cada escenario se construyen modelos de clasificación a partir de un conjunto de entrenamiento con atributos del modelo geológico, utilizando redes neuronales y árboles de decisión, que corresponden a los algoritmos del *Data Mining*, y también con *co-kriging indicador* (herramienta ya utilizada en minería). Los modelos resultantes con cada técnica empleada son evaluados y comparados cualitativamente mediante la inspección de cortes transversales en cada uno de estos y además cuantitativamente midiendo el desempeño de cada herramienta a través de la precisión de predicción.

Para estudiar el impacto que genera incorporar los distintos modelos geológicos obtenidos en la estimación y categorización de recursos, cada modelo es utilizado como una variable de entrada en la estimación de la ley mediante *kriging* universal, obteniéndose tres modelos de leyes distintos (uno por cada técnica empleada en la construcción del modelo geológico), a los cuales se añade un modelo estimado con *kriging* ordinario sin tomar en consideración el modelo geológico. Lo anterior es replicado en cada escenario en estudio.

Los análisis se efectúan a través de RStudio, interfaz que permite acceder de manera sencilla al lenguaje de programación orientado a objetos R [46]. R consta de un sistema base y de paquetes adicionales que extienden la funcionalidad. Aquellos utilizados en el desarrollo de esta memoria de título se presentarán oportunamente.

A continuación, se sintetizan los procedimientos utilizados en cada caso de estudio.

3.1. Procedimiento conjunto de datos simulado

- Se generó una base de datos simulados, con dos variables categóricas y una variable continua, distribuidas en 3D, correspondientes a un muestreo intensivo de una zona de estudio, que imita la variabilidad y estructura espacial real de las variables regionalizadas presentes en un depósito mineral.

- Dada la densidad del muestreo se particiona la base en dos conjuntos, uno de entrenamiento y otro de validación.
- El conjunto de entrenamiento representa muestras recolectadas en una eventual campaña de exploración geológica que, junto con los algoritmos del *Data Mining* y el método *co-kriging* indicador, es utilizado en la construcción de modelos de estimación de las variables categóricas presentes.
- El conjunto de validación se utiliza para calibrar los parámetros de los algoritmos y medir el desempeño de estimación de estos al comparar los valores estimados en los puntos del conjunto de validación con los valores conocidos.
- Una vez obtenidos y validados los modelos de estimación de variables categóricas se utiliza cada uno de ellos como tendencia para la predicción de la variable continua (ley) mediante el uso de *kriging* universal cuyos resultados son comparados con los estimados mediante *kriging* ordinario.
- Con cada una de las predicciones se genera un plan de estimación y categorización de recursos, reportando las medidas estadísticas principales.
- Finalmente se realiza un análisis de los diferentes escenarios de riesgo asociados a los modelos geológicos predichos según la técnica empleada y cómo se diferencian de los datos reales.

El procedimiento mencionado se resume en la **Figura 15**.

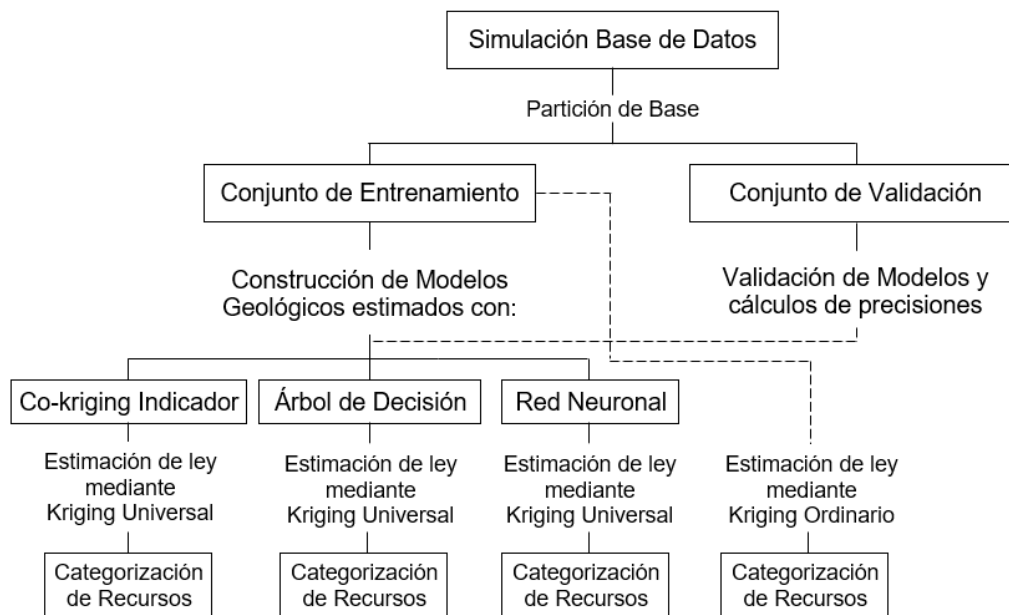


Figura 15: Resumen procedimiento caso de estudio simulado

3.2. Procedimiento conjunto de datos reales

- Se realiza un análisis descriptivo completo de las variables presentes en la base de datos.
- Con los algoritmos de *Data Mining* y el método *co-kriging* indicador se generan los modelos de clasificación para las variables categóricas presentes.
- A diferencia del caso de estudio simulado, el conjunto de datos reales no puede ser particionado en un conjunto de entrenamiento y validación debido a la baja cantidad de datos, por lo tanto, para medir el desempeño de los modelos generados se realiza validación cruzada del tipo *leave-one-out*³.
- Se realiza el plan de estimación y categorización de recursos incorporando los modelos geológicos predichos a la estimación de la ley mediante *kriging* universal, también se realiza el plan de estimación con *kriging* ordinario para comparar.
- Finalmente se realiza una evaluación de cada predicción y se entrega el peor, el mejor y el caso más favorable.

A continuación, en la **Figura 16** se ilustra el procedimiento explicado:

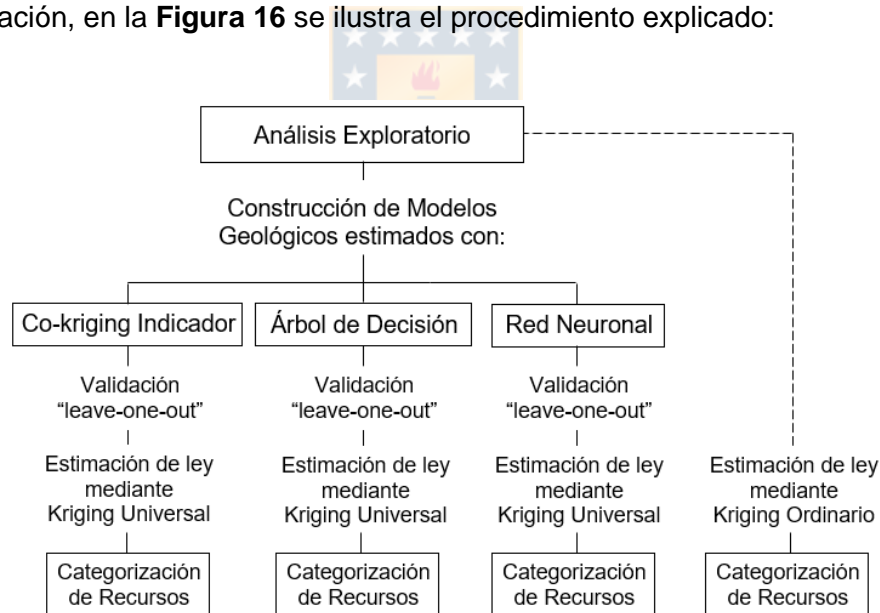


Figura 16: Resumen procedimiento caso de estudio real

En cada uno de los conjuntos de datos, se realiza un análisis crítico de los resultados y se elaborarán conclusiones y discusiones parciales acerca de la aplicabilidad de los algoritmos estudiados en la estimación de recursos.

³ La validación cruzada *Leave-one-out* consiste en separar los datos dejando sólo una muestra para el conjunto de prueba mientras que todo el resto conforma el conjunto de entrenamiento.

CAPÍTULO 4

CASO SIMULADO

4.1. Metodología de simulación

El objetivo de la simulación es generar 100 modelos que representen la variabilidad espacial de variables contenidas en un depósito de cobre (ley, tipo de roca y alteración), los cuales son utilizados para probar la aplicabilidad de los algoritmos del *Data Mining* en diferentes escenarios. Lo anterior es realizado mediante el paquete de R *gstat*, el cual proporciona una amplia gama de funciones de modelamiento, predicción y simulación geoestadística univariable y multivariable [47].

En primer lugar, es necesario crear una grilla sobre la cual se realizan las simulaciones. Se generan las coordenadas geográficas (x, y, z) , correspondientes a los centroides de los bloques de $10 \times 10 \times 10 \text{ m}^3$, y que se extienden entre 0 a 200, 0 a 100 y 0 a 300 metros respectivamente. En la **Figura 17** se ilustra una vista en planta de la grilla de simulación en la cual se cuentan 231 puntos, considerando los puntos a través de z se tiene un total de 7161 puntos o bloques sobre los que se realiza cada simulación.

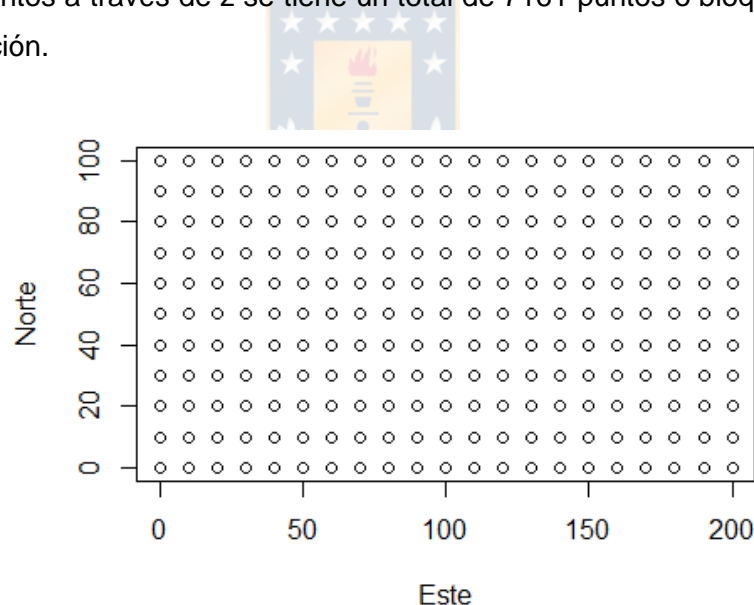


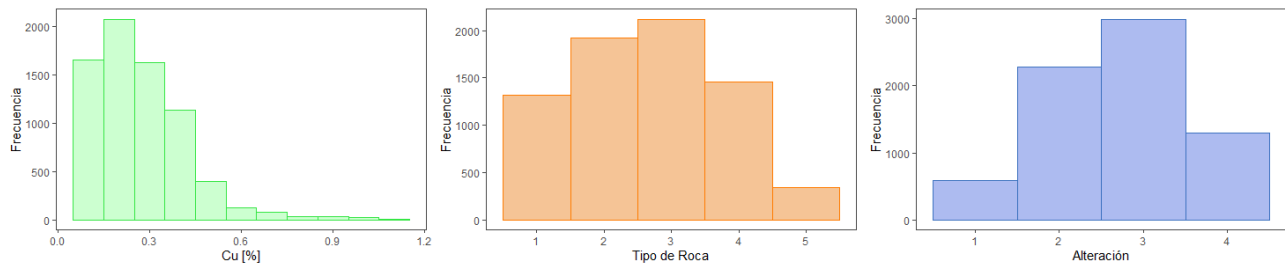
Figura 17: Vista en planta de grilla de simulación

Las simulaciones son de tipo Gaussiano no condicional, por lo cual se le debe proporcionar al programa los parámetros de los variogramas directos y cruzados de las variables a estudiar, cuyas mesetas se indican en la **Tabla 1**. Todos los variogramas se ajustan a un modelo Gaussiano con alcance igual a 100 metros.

Tabla 1: Mesetas de variogramas directos y cruzados para simulación

	Ley Cu	Roca	Alteración
Ley Cu	0.40	-0.20	0.10
Roca	-0.20	0.30	0.05
Alteración	0.10	0.05	0.20

El primer resultado de la simulación son variables continuas que no representan los valores reales de un depósito mineral por lo tanto es necesario transformarlas. La ley de cobre es llevada a una escala con una distribución sesgada positivamente y con valores que van desde 0% a 1.2% de Cu fino. Las variables tipo de roca y alteración son categorizadas asignándole 5 y 4 clases respectivamente. En la **Figura 18** se enseñan los histogramas de las variables ley de cobre, tipo de roca y alteración correspondientes a la simulación n°20.

**Figura 18: Histogramas Ley de Cu, Tipo de Roca y Alteración para simulación n° 20**

Obtenidas las bases simuladas se procede a particionar cada base en un conjunto de entrenamiento y otro de validación. Para determinar qué datos seleccionar como parte del conjunto de entrenamiento se simuló una campaña de exploración, en la cual se extraen muestras de la base con la forma de sondajes verticales, con largo variable y ubicación en el plano XY aleatoria. Cabe señalar que la selección aleatoria de los datos influye en la calidad de estos ya que no se puede asegurar de antemano que cada variable del conjunto de entrenamiento se encuentre bien representada.

Por otro lado, y con el objetivo de estudiar cómo se comportan los algoritmos según la cantidad de información proporcionada en el conjunto de entrenamiento, a cada base simulada se le extrae una cantidad diferente de datos, empezando en la simulación n°1 con 11 sondajes (273 muestras) y agregando un sondaje a medida que crece el índice de la base simulada, es decir, 12 sondajes en la simulación n°2 (303 muestras), 13 sondajes en la simulación n°3 (326 muestras) y así sucesivamente hasta terminar con 110 sondajes en la simulación n°100 (2570 muestras). En la siguiente figura se muestra una secuencia de extracción de sondajes para las primeras 6

simulaciones, en esta se observa que se va agregando un sondaje a medida que aumenta el índice de la simulación.

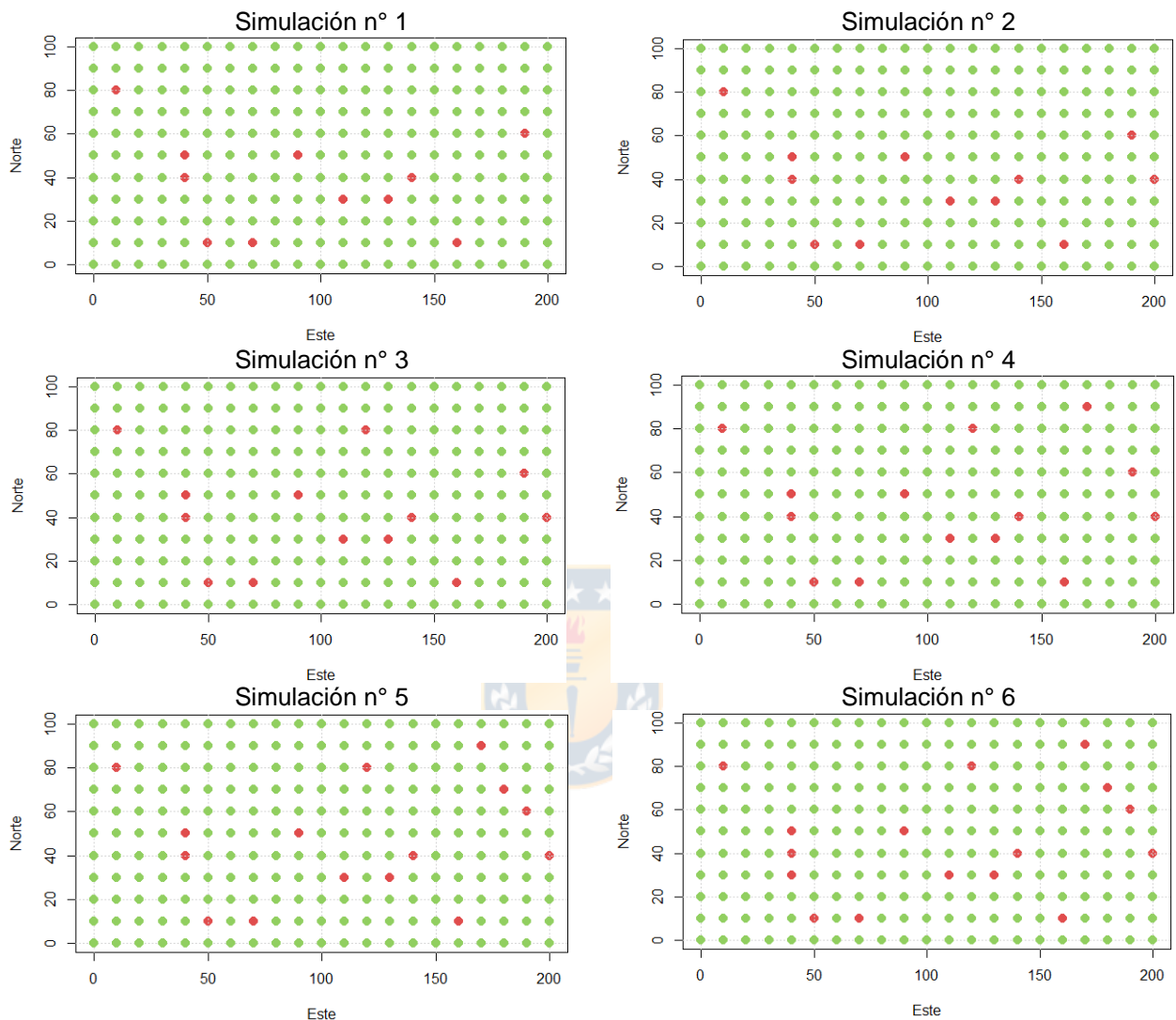


Figura 19: Vista en planta de secuencia de extracción de sondajes de las primeras 6 bases simuladas para formar conjuntos de entrenamiento

En la **Figura 20** se puede observar vistas de las bases simuladas n° 1, 5 y 10, en rojo se representa las muestras del conjunto de entrenamiento (las cuales representan sondajes), se aprecia que las simulaciones de mayor índice poseen más muestras y además se puede notar que los sondajes tienen diferentes largos.

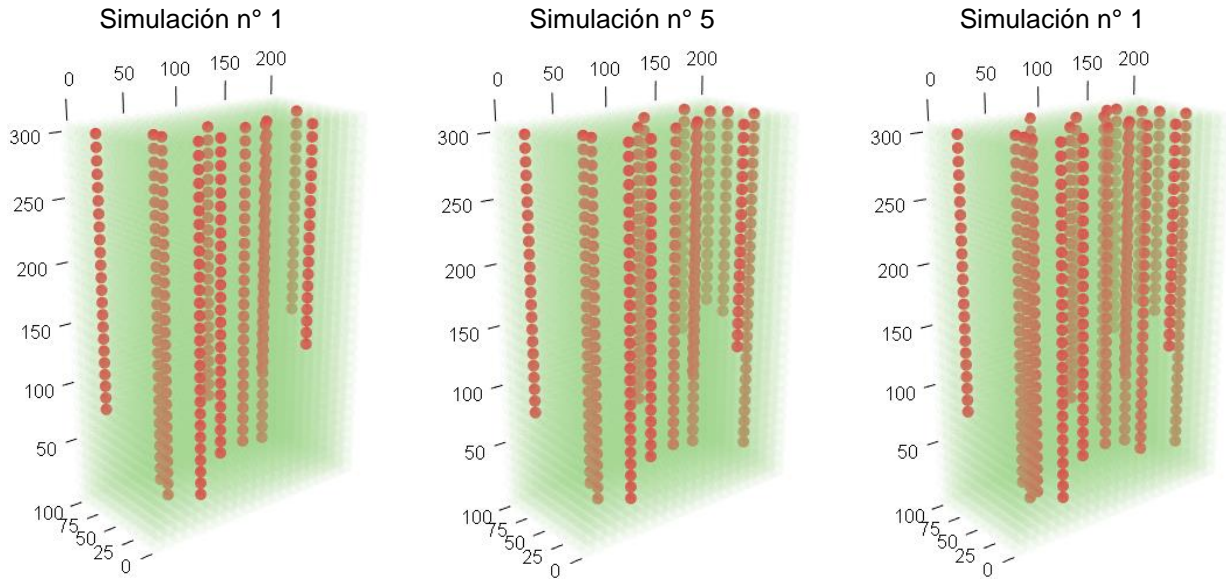


Figura 20: Vistas bases simuladas n° 1, 5 y 10 (Color rojo representa conjunto de entrenamiento y color verde el conjunto de validación)

En la **Figura 21** se enseñan los histogramas de las variables del conjunto de entrenamiento para la simulación n° 20, el cual contiene 30 sondajes y un total de 725 muestras para estimar los modelos geológicos del depósito completo, se puede apreciar a simple vista, comparando con la **Figura 18**, que el conjunto de entrenamiento en este caso mantienen una distribución semejante a la de la base completa, por lo que cada variable está bien representada.

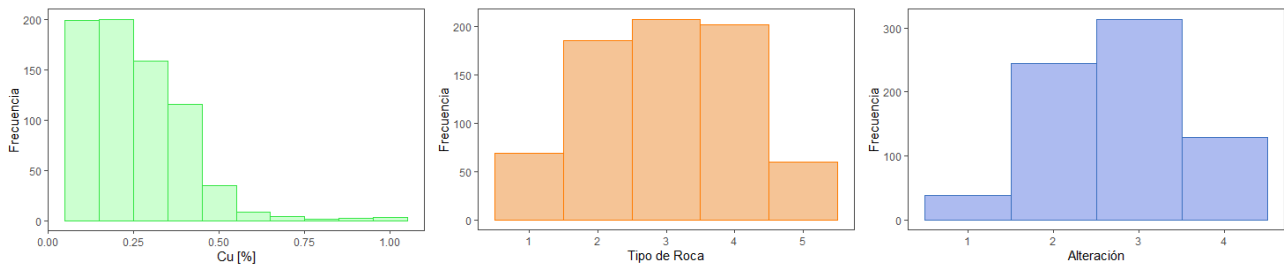


Figura 21: Histogramas Ley de Cu, Tipo de Roca y Alteración para conjunto de entrenamiento – simulación n° 20

4.2. Aplicación

Particionada las bases simuladas, se aplican los métodos a estudiar sobre el conjunto de entrenamiento, para obtener modelos de clasificación de las variables geológicas presentes. Para esto se utilizaron los siguientes paquetes disponibles en R.

- **spMC:** Es utilizado para predecir variables categóricas mediante *co-kriging* indicador. A diferencia otros *softwares* utilizados para predecir variables categóricas, spMC implementa modelos de Cadena de Markov para estimar probabilidades de transición experimentales de los datos categóricos en vez covarianzas/variogramas [48]. Para optimizar los resultados se probaron diferentes tolerancias de búsqueda en la estimación del transiograma seleccionando aquel que entrega una mayor precisión.
- **C5.0:** Permite la construcción de árboles de clasificación, con capacidad para generar árboles de predicción simples modelos basados en reglas, *ensembles* basados en *boosting* y asignación de distintos pesos a los errores.
- **Neuralnet:** Es utilizado para el entrenamiento de redes neuronales utilizando el mecanismo de aprendizaje *back-propagation* [49]. Permite configuraciones flexibles, entre las cuales para optimizar los resultados se probaron redes con 2 y 3 capas ocultas con un máximo de 20 y 10 neuronas respectivamente, y *thresholds* entre 0.01 a 0.1 como criterio de detención. Se utilizó una función logística como función de activación.

Luego los distintos modelos de clasificación se utilizan para predecir las variables geológicas presentes sobre los puntos del conjunto de validación, es decir los tipos de roca y de alteración. Para comparar el poder predictivo de los modelos obtenidos por los diferentes algoritmos se utiliza como medida de desempeño la precisión, ecuación (53).

La información de los modelos geológicos estimados se incorpora (separadamente) como tendencia externa para estimar la ley de cobre mediante *kriging* universal, también se estima la ley mediante *kriging* ordinario con el fin de comparar. Ambos casos se realizan utilizando el paquete de R, introducido anteriormente, *gstat*.

Se realiza una categorización de recursos para los 4 modelos de leyes estimados, en la **Tabla 2** se indica el criterio de clasificación basado en la varianza relativa de *kriging*.

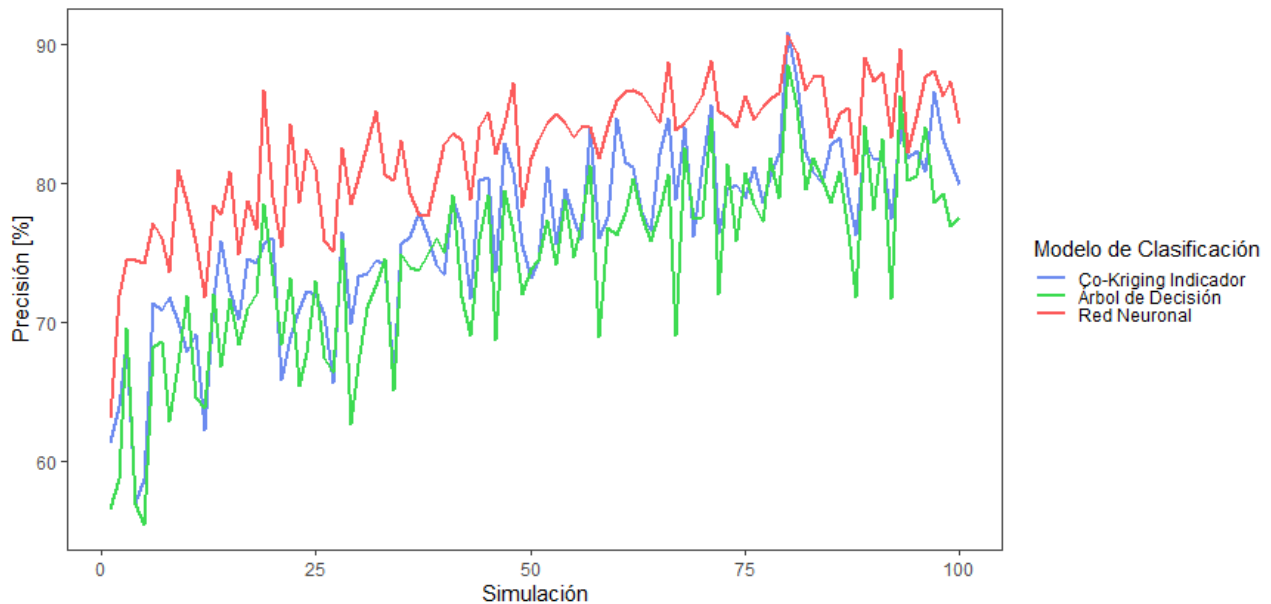
Tabla 2: Criterio de clasificación de recursos para caso simulado

Recurso Medido	varianza relativa de kriging < 15%
Recurso Indicado	varianza relativa de kriging 15% – 25%
Recurso Inferido	varianza relativa de kriging > 25%

4.3. Análisis de resultados

4.3.1. Estimación de modelos geológicos

En la **Figura 22** se resumen las precisiones obtenidas por los diferentes métodos en la estimación del tipo de roca para las 100 simulaciones. Se observa que los modelos estimados por redes neuronales logran mejores precisiones en la mayoría de los casos (98%), seguidos por el *co-kriging* indicador y finalmente por los modelos de árboles de decisión.

**Figura 22: Precisiones modelos de clasificación para Tipo de Roca**

Por otro lado, en la **Figura 23** se presenta las precisiones obtenidas para los modelos de alteraciones, donde se aprecia un comportamiento similar al anterior, pero observándose una disminución en las diferencias de precisión entre los modelos estimados por redes neuronales contra los estimados por *co-kriging* indicador y árboles de decisión. Además, al comparar la **Figura 22** con la **Figura 23** se observa que en general se obtienen mayores precisiones en los modelos asociados a la alteración lo que puede deberse al número de categorías que deben diferenciar los algoritmos,

teniendo mayor complejidad aquel con más clases a estimar. En las **Tabla B.1** y **Tabla B.2** del anexo B se encuentran las precisiones de los modelos de tipo de roca y alteración detallados.

Si bien, tanto en los modelos de tipo de roca como en los de alteración, se percibe una tendencia de aumento de la precisión a medida que se incrementa la información del conjunto de entrenamiento (número de sondajes), el ruido presente en ambos gráficos refleja que los tres algoritmos son afectados por la calidad de información obtenida en el muestreo, esto resalta la importancia de trabajar con un muestreo representativo. Asimismo, en ambos casos se puede observar que las diferencias en precisión entre los modelos estimados por redes neuronales versus los demás algoritmos son mayores para las primeras simulaciones (las cuales poseen una menor cantidad de muestras para el entrenamiento) y luego comienzan a estrecharse a medida que aumenta el número de sondajes, lo que sugiere inclinarse por modelos estimados por redes neuronales en contextos de poca información.

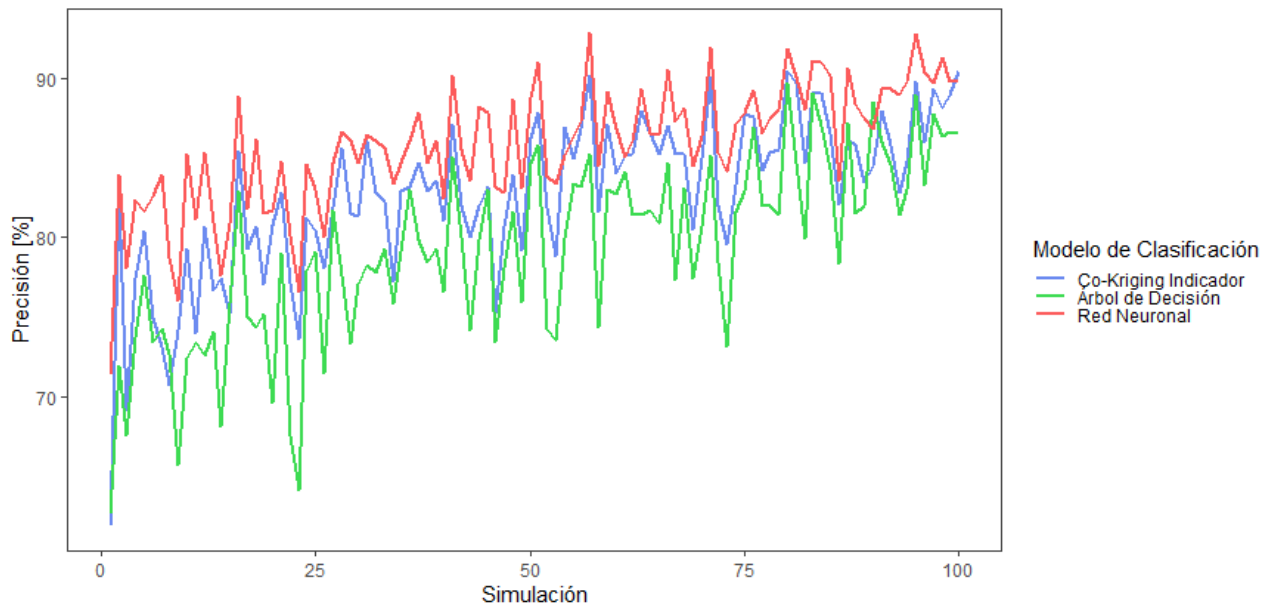


Figura 23: Precisión modelos de clasificación para Alteración

En la **Tabla 3** se sintetizan las precisiones obtenidas por los algoritmos al estimar las variables del modelo geológico para las 100 simulaciones, se destaca que las redes neuronales tienen una menor desviación respecto a la media en comparación a los demás métodos, por ende, son menos dependientes de la cantidad de información.

Tabla 3: Resumen de resultados en estimación de variables del modelo geológico

	Tipo de Roca			Alteración		
	Co-Kriging Indicador	Árboles de Decisión	Redes Neuronales	Co-Kriging Indicador	Árboles de Decisión	Redes Neuronales
Precisión promedio	76.42	74.31	82.37	82.74	79.23	85.84
Desviación estándar	6.27	6.60	4.78	5.04	5.78	3.90
Precisión mínima	57.01 (sim 4)	55.38 (sim 5)	63.07 (sim 1)	61.85 (sim 1)	62.65 (sim 1)	71.37 (sim 1)
Precisión máxima	90.75 (sim 80)	88.41 (sim 80)	90.59 (sim 80)	90.46 (sim 80)	89.69 (sim 80)	92.82 (sim 57)

Basándose únicamente en la precisión para escoger un modelo de clasificación, y considerando los resultados presentados se determinaría que las redes neuronales tienen un mayor desempeño al estimar las variables geológicas en términos generales, pero esto no es siempre el objetivo al recurrir a un modelo geológico. En ocasiones puede ser necesario analizar los modelos de clasificación de acuerdo a su capacidad de discriminar una o ciertas clases puntuales las cuales pueden ser de mayor interés o tener una mayor repercusión en etapas aguas abajo, esto puede llevarse a cabo utilizando como indicador la sensibilidad o especificidad por categoría, ecuaciones (54) y (55) respectivamente presentadas en la sección 2.8.2.

En la **Tabla 4**, se recopilan las sensibilidades y especificidades obtenidas por cada método según las categorías de los modelos de tipo de roca y alteración para la simulación n° 25, si bien las redes neuronales consiguen mayores precisiones (81.06% y 82.98% para los modelos de tipo de roca y alteración respectivamente) en comparación con el *co-kriging* indicador (71.96% y 80.40%) y los árboles de decisión (72.94% y 79.10%), se observa por ejemplo que para el tipo de roca 1 y 2 el *co-kriging* indicador alcanza mejores sensibilidades. Por otro lado, pese a que el *co-kriging* indicador y los árboles de decisión alcanzan precisiones semejantes se aprecian diferencias significantes entre sus sensibilidades para el tipo de roca. Por lo tanto, antes de tomar una decisión sobre cual modelo es mejor es necesario identificar cuál o cuáles son las categorías que tienen mayor trascendencia en el proceso para el cual se construye el modelo. Se destaca además la baja sensibilidad que se alcanza por todos los métodos para la alteración 4.

Tabla 4: Sensibilidad y Especificidad según variable y modelo de clasificación para simulación n° 25

		Co-Kriging Indicador		Árbol de Decisión		Red Neuronal	
		Sensibilidad	Especificidad	Sensibilidad	Especificidad	Sensibilidad	Especificidad
Tipo de roca	1	0.688	0.990	0.332	0.998	0.654	0.996
	2	0.934	0.823	0.833	0.878	0.915	0.911
	3	0.709	0.804	0.708	0.820	0.793	0.876
	4	0.523	0.971	0.691	0.912	0.733	0.953
	5	0.613	1.000	0.765	0.995	0.848	0.990
Alteración	1	0.966	0.899	0.898	0.914	0.933	0.922
	2	0.780	0.892	0.778	0.893	0.813	0.905
	3	0.762	0.921	0.809	0.894	0.827	0.928
	4	0.327	0.998	0.235	0.994	0.432	0.995

En las **Figuras 24 y 25** se comparan los modelos originales de tipo de roca y alteración con los modelos estimados por los algoritmos en estudio, por medio de secciones transversales cortadas a través del eje Y. A una escala macro se aprecia que todos los algoritmos utilizados logran representar, a diferente nivel de detalle, los patrones del modelo original. En general es el modelo estimado por redes neuronales el que más se asemeja al modelo original, pero en ocasiones el *co-kriging* indicador logra representar detalles que se le escapan a las redes neuronales. Tanto los modelos estimados por redes neuronales como los estimados por *co-kriging* indicador consiguen en gran medida dibujar bordes suavizados, mientras que los modelos estimados por árboles de decisión suelen dibujar ángulos rectos, propio de modelos más simplificados. Junto con la **Tabla 4** se comprueba por ejemplo la baja sensibilidad para el tipo de roca 1 en el modelo estimado por árbol de decisión ya que en la **Figura 24** prácticamente no hay presencia de esta categoría en el modelo.

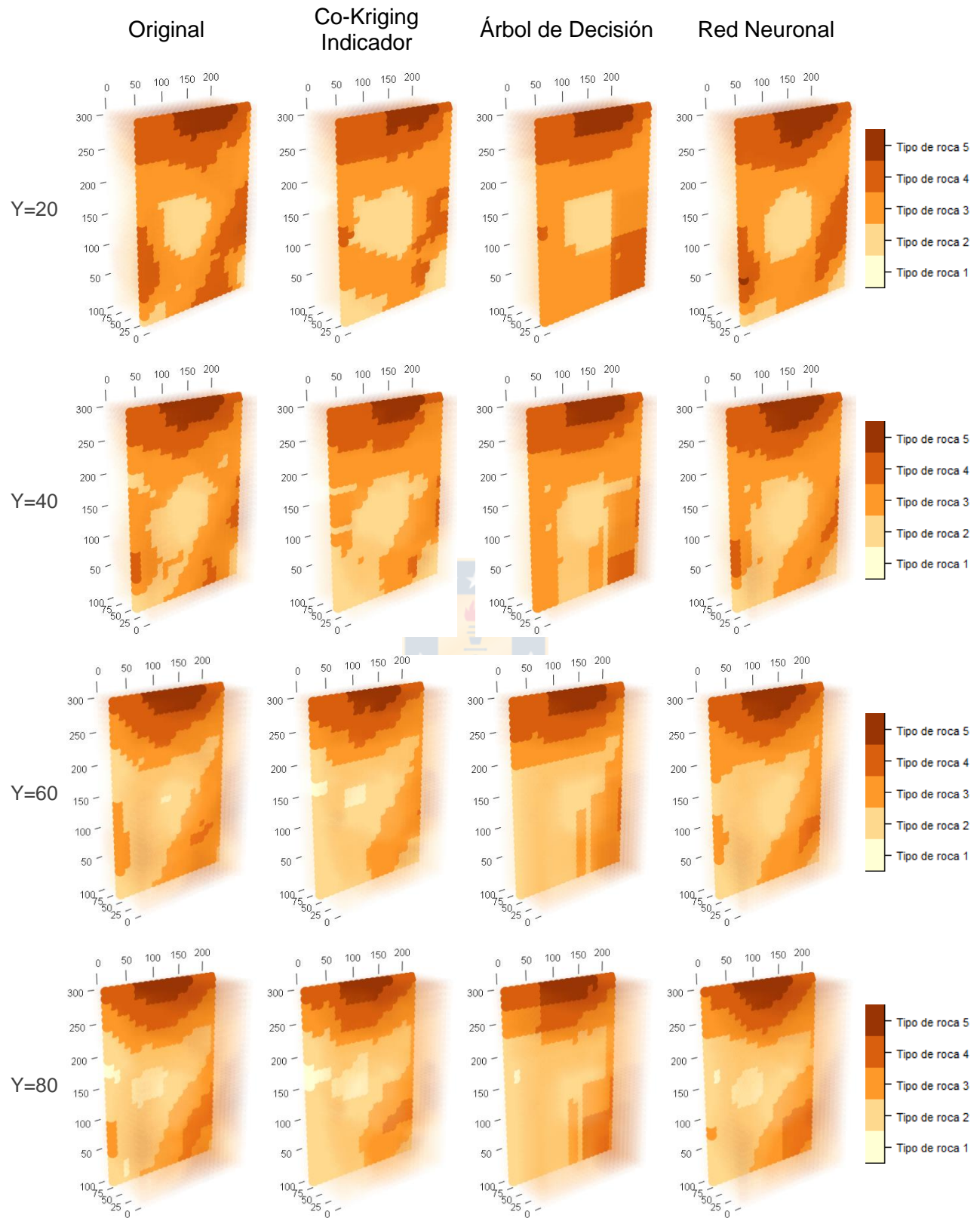


Figura 24: Secciones transversales del modelo original de tipo de roca versus modelos estimados para simulación n° 25

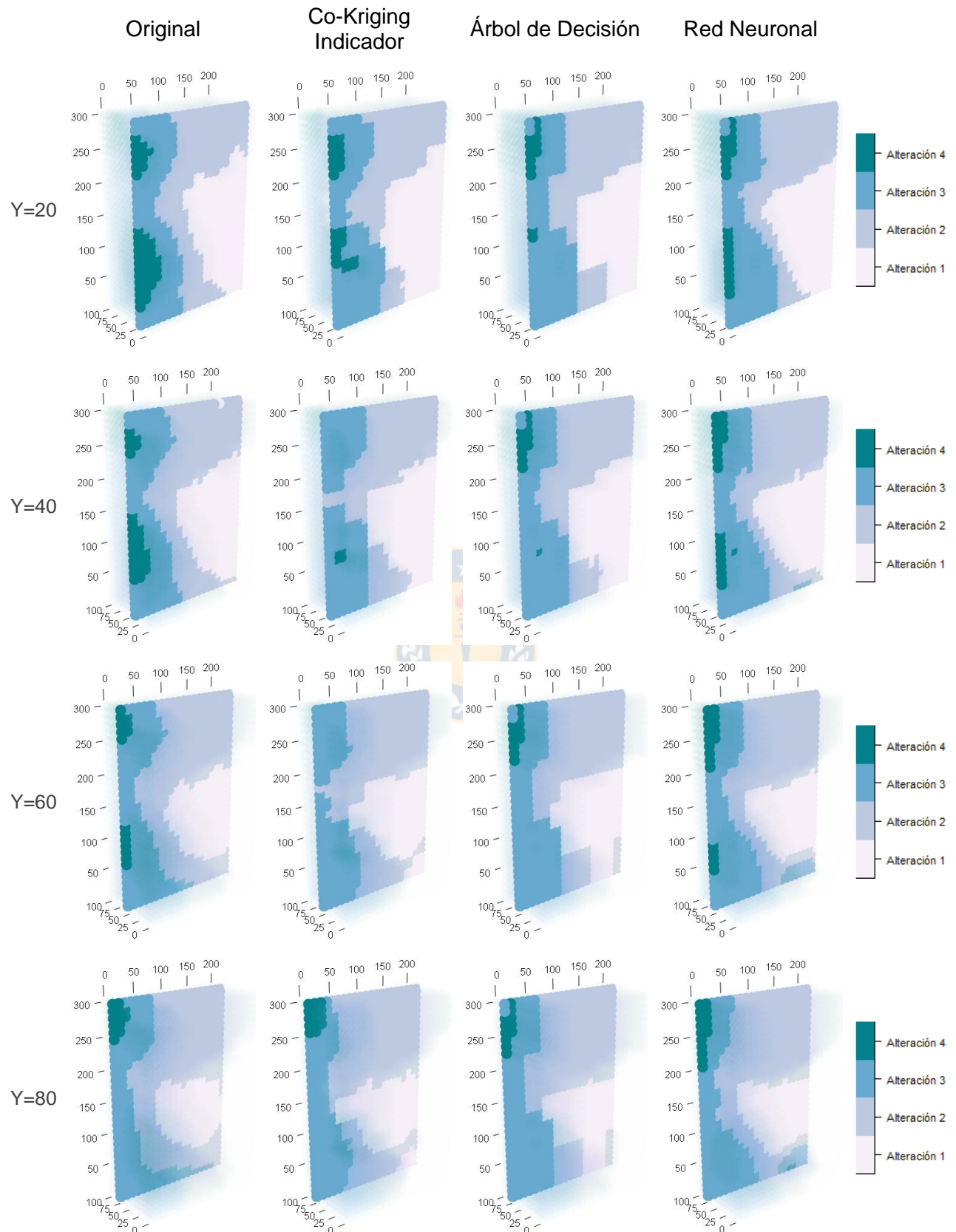


Figura 25: Secciones transversales del modelo original de alteración versus modelos estimados para simulación n° 25

En las **Figuras 26 y 27** se gráfica el tiempo consumido por los diferentes algoritmos en la construcción y utilización de los modelos de clasificación para tipo de roca y alteración respectivamente.

Se destaca en primer lugar la velocidad de los árboles de decisión que promedian un tiempo menor a 0.04 segundos entre los dos modelos. En cuanto al algoritmo de redes neuronales su mayor dificultad se encuentra en la construcción del modelo, ya que la red debe adaptarse a los datos de entrenamiento a partir de pesos iniciales aleatorios y mediante reiteradas iteraciones, hasta cumplir un criterio de detención establecido, en las gráficas se confirma que la dificultad aumenta cuando el modelo tiene más información que aprender (recordar que aumenta la cantidad de sondajes a medida que aumenta el índice de la simulación), por lo tanto, aumenta el consumo de tiempo empleado en la construcción del modelo, el ruido presente puede deberse a la complejidad de la base o a la complejidad de la arquitectura de la red ya que entre mayor es el número de neuronas y/o capas, mayor es la cantidad de pesos que el algoritmo debe modificar. Finalmente, para el *co-kriging* indicador, se observa que el tiempo es relativamente constante, pero tiende a disminuir a medida que aumenta la cantidad de datos de entrenamiento y disminuye la cantidad de bloques a estimar, por lo tanto, se deduce que este algoritmo realiza un mayor esfuerzo al realizar la estimación del modelo que en estimar los parámetros del transiograma.

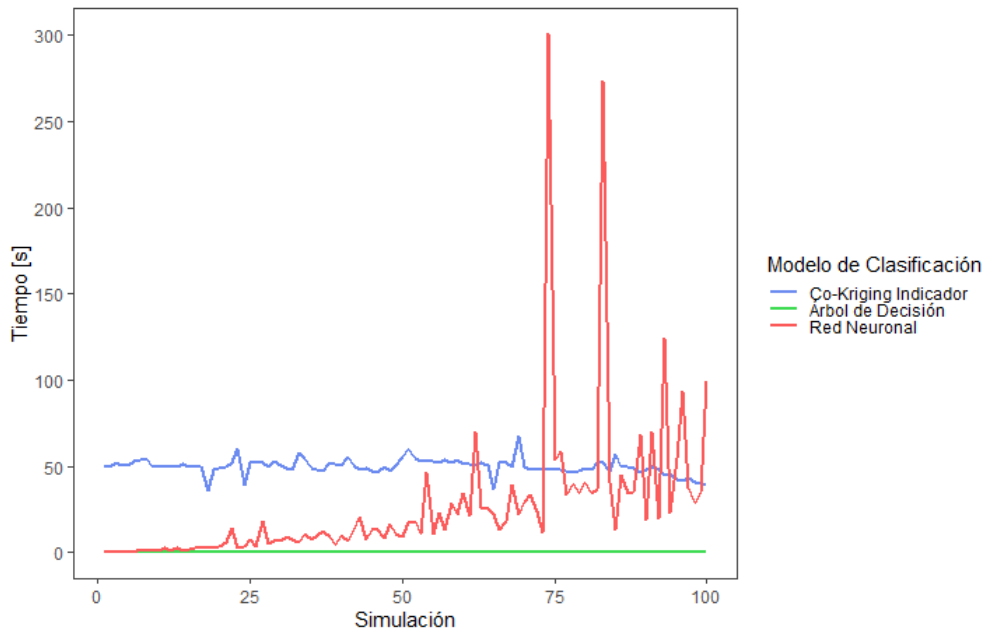


Figura 26: Tiempos empleados en la construcción y utilización de modelos de clasificación para tipo de roca según algoritmo utilizado

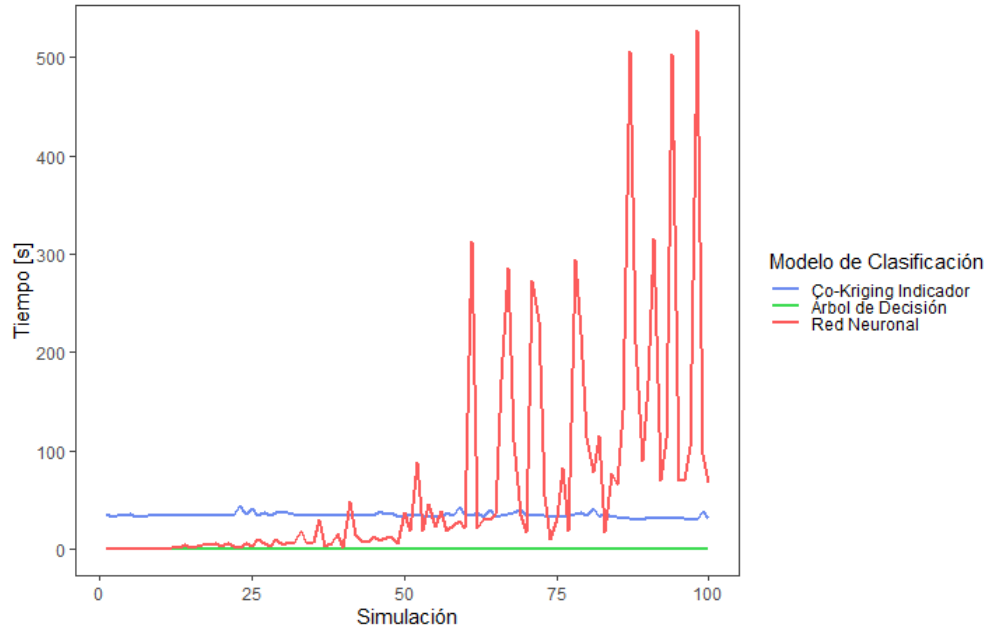


Figura 27: Tiempos empleados en la construcción y utilización de modelos de clasificación para alteración según algoritmo utilizado

Se destaca que hasta cierto punto las redes neuronales son más veloces que el *co-kriging* indicador pero estos gráficos solo evidencian la construcción de un modelo por simulación. El *co-kriging* indicador requiere probar diferentes tolerancias y/o direcciones de búsqueda en la construcción del transiograma, por lo que el tiempo reflejado en los gráficos se multiplicaría por la cantidad de parámetros probados. Para encontrar una red neuronal óptima es necesario construir varios modelos hasta encontrar el mejor, por lo que a fin de cuentas el tiempo empleado aumenta considerablemente. Pero en los árboles de decisión, el algoritmo C5.0 optimiza en una iteración la construcción del modelo, lo que presenta una amplia ventaja en comparación con los otros métodos.

4.3.2. Estimación de leyes y categorización de recursos

Realizada la estimación de leyes mediante *kriging* universal a las primeras 25 simulaciones⁴, incorporando, independientemente, la información geológica estimada por *co-kriging* indicador, árboles de decisión y redes neuronales, no se observan diferencias significativas respecto al error cuadrático medio (RMSE), ni tampoco al compararlos con los valores estimados por *kriging* ordinario, esto probablemente por la buena correlación espacial de la ley de cobre. Sin embargo, se advierte una disminución de la varianza relativa de *kriging* (RKV) al utilizar *kriging* universal versus *kriging* ordinario, independiente de la herramienta utilizada para estimar el modelo geológico, de hecho, no

⁴ Se escogen las primeras 25 simulaciones por ser las que poseen menor cantidad de sondajes, condición más usual en la realidad.

hay grandes diferencias entre las metodologías que usan *kriging* universal (referirse a **Tabla 5**). A partir de esto se espera obtener mayor cantidad de recursos medidos e indicados al utilizar *kriging* universal con tendencias externas.

Tabla 5: RMSE y RKV promedio para 25 primeras simulaciones según kriging empleado

Promedio	Kriging Ordinario	Kriging Universal + Co-kriging Indicador	Kriging Universal + Árbol de Decisión	Kriging Universal + Red Neuronal
RMSE	0.0413	0.0413	0.0415	0.0412
Varianza relativa de kriging	0.2943	0.2270	0.2312	0.2279

En la **Tabla C.1** del anexo C se detallan la cantidad de bloques por categoría medidos, indicados o inferidos, según la metodología utilizada para la estimación de leyes y utilizando el criterio de la **Tabla 2**. En la **Figura 28** se comparan los porcentajes de tonelaje medido e indicado en relación al tonelaje total del depósito, se observa que siempre se consigue un mayor tonelaje de recursos medidos e indicados utilizando *kriging* universal por sobre el *kriging* ordinario, en cambio al comparar los modelos estimados por *kriging* universal según el método usado para la construcción del modelo geológico se encuentran resultados semejantes entre sí.

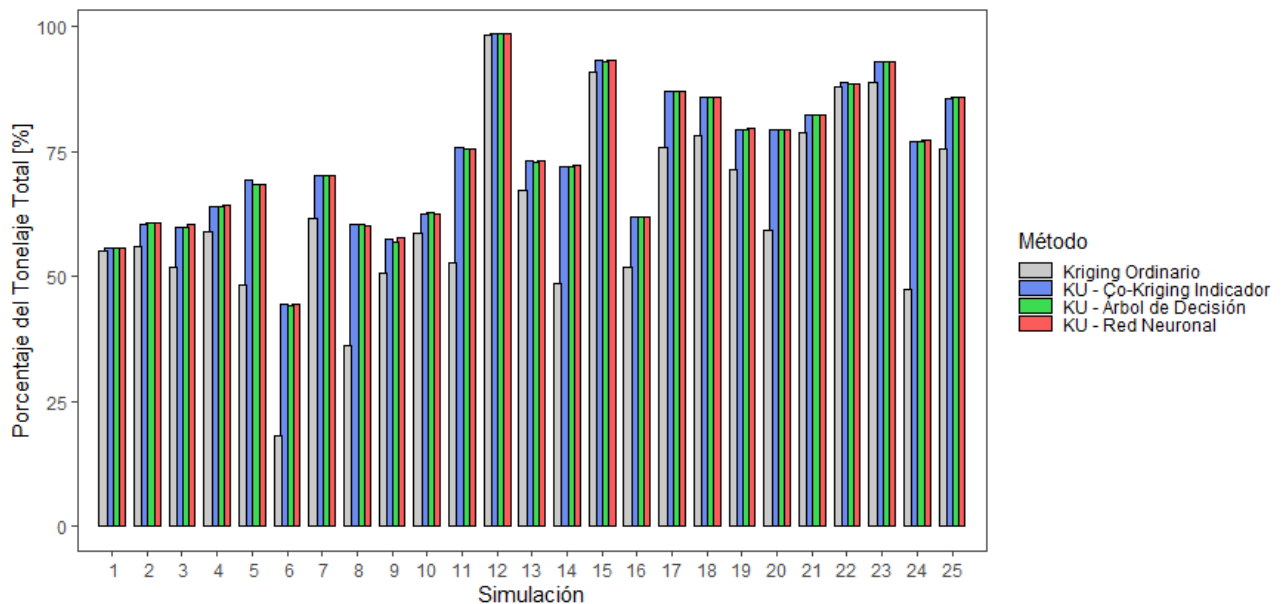


Figura 28: Porcentaje de tonelaje (Medido e Indicado) respecto al tonelaje total por metodología

Dependiendo del escenario se puede visualizar que sobresale levemente un método por sobre los demás sin obedecer a ningún patrón estudiado, lo que sugiere desarrollar un análisis de sensibilidad y especificidad por categoría, además de explicar la relación existente entre la ley y las categorías de las variables del modelo geológico, para interpretar las diferencias entre los modelos estimados por *kriging* universal. Lo anterior se explica porque los modelos geológicos estimados poseen un sesgo que puede estar relacionado a una mayor presencia de una categoría asociada a leyes altas o bajas, sumado a que al aplicar como criterio de categorización de recursos la varianza relativa de *kriging* se favorece los bloques con mayor ley, en otras palabras, los modelos asociados a mayores leyes se le asignaran un mayor número de bloques medidos e indicados y en consecuencia un mayor tonelaje.

En la **Figura 29** se resume la categorización de recursos para la simulación n° 25, en esta se observa que las diferencias entre los modelos que utilizan *kriging* universal junto con los diferentes modelos geológicos estimados son mínimas, pero al comparar con la categorización hecha junto al *kriging* ordinario se obtienen una mayor cantidad de bloques medidos y menos bloques inferidos.

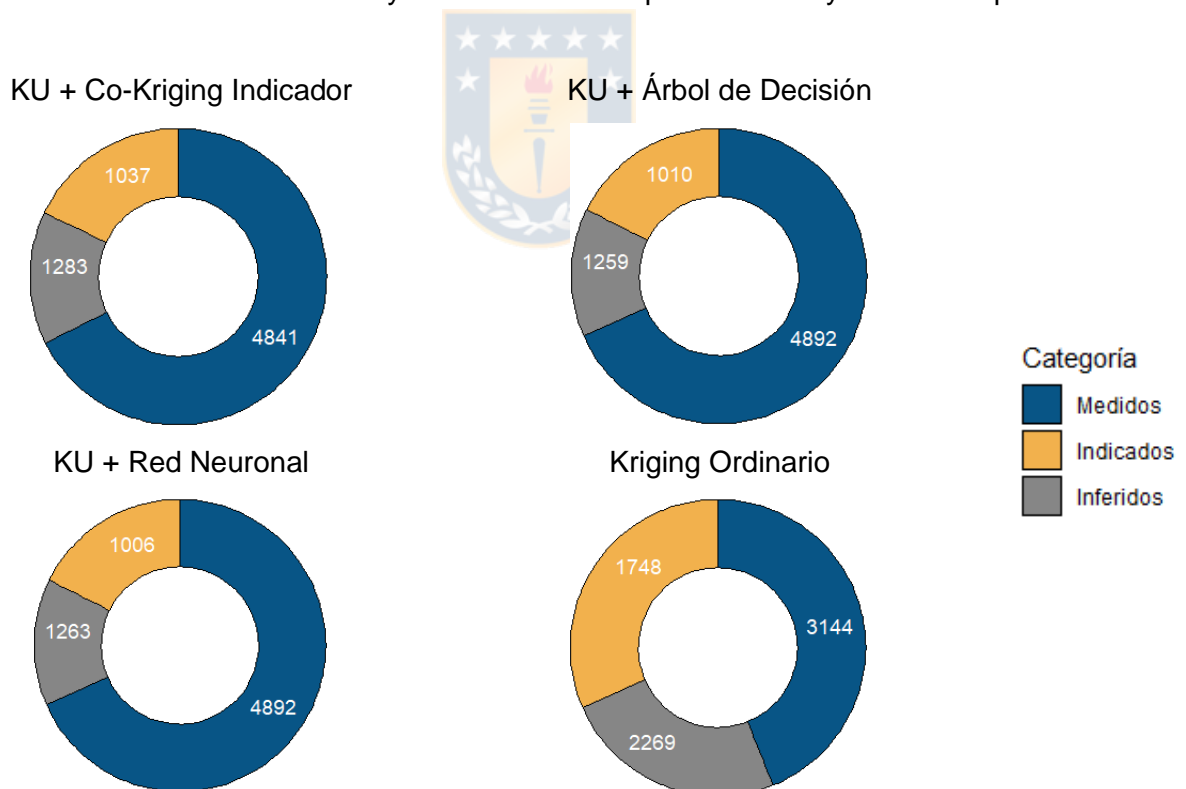


Figura 29: Categorización de Recursos para simulación n°25

CAPÍTULO 5

CASO REAL

5.1. Descripción de la base de datos

La base de datos consiste en 660 muestras provenientes de sondajes de un yacimiento de Hierro, en los cuales se encuentra información de leyes de hierro, tipo de roca y textura. Los tipos de roca presentes consisten en 5 grupos: Metandesitas (MET), Hierro de mena (HIE), Brechas (BRH), Andesita (AND) y Dioritas (DIO). Mientras que las texturas presentes se dividen en 4 grupos: Porfídica (POR), Afanítica (AFA), Maciza (MAC) y Brechosa (BRE).

5.1.2. Mapas

La base de datos tiene un número de 11 sondajes de diferentes largos, configurados irregularmente. La zona cubierta en el muestreo es de aproximadamente 740 metros en la dirección Este, 1360 metros en la dirección Norte y 550 metros en la vertical. En las siguientes figuras se muestran proyecciones en planta y perfiles de los sondajes indicando información sobre leyes de hierro, tipo de roca y textura.

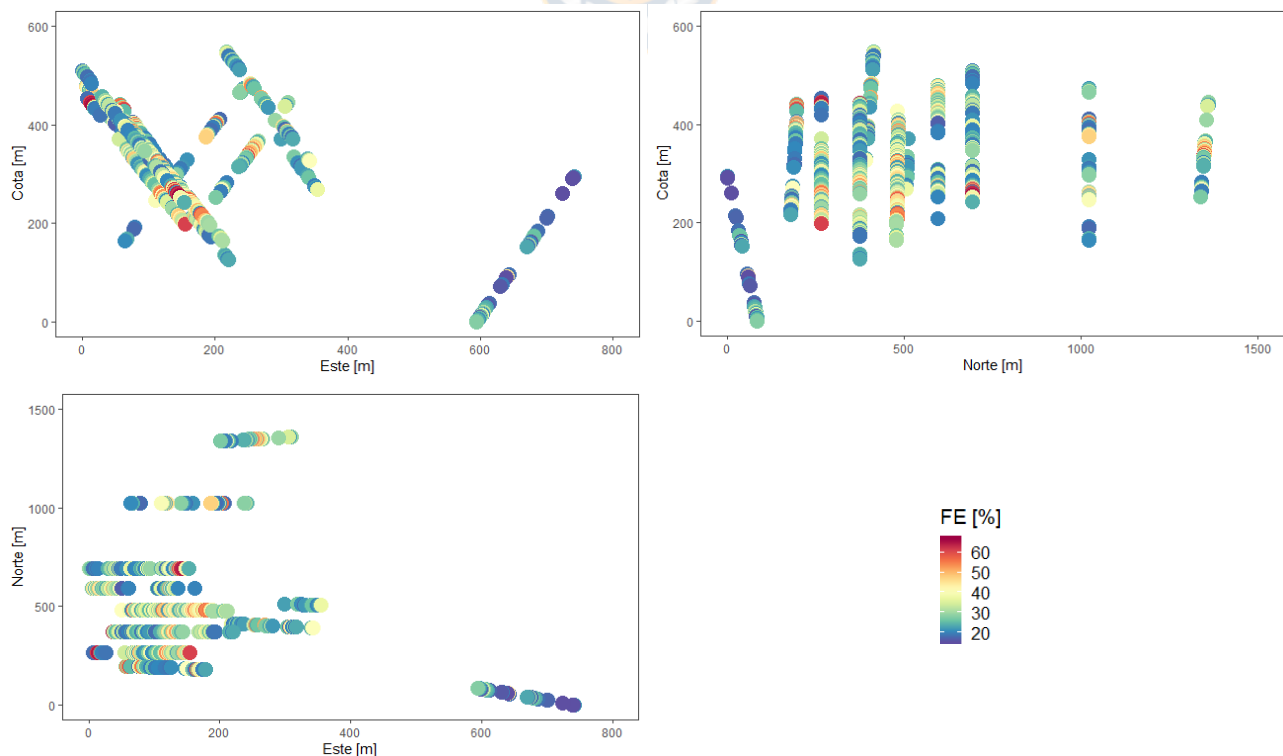


Figura 30: Proyecciones del yacimiento para Leyes de FE

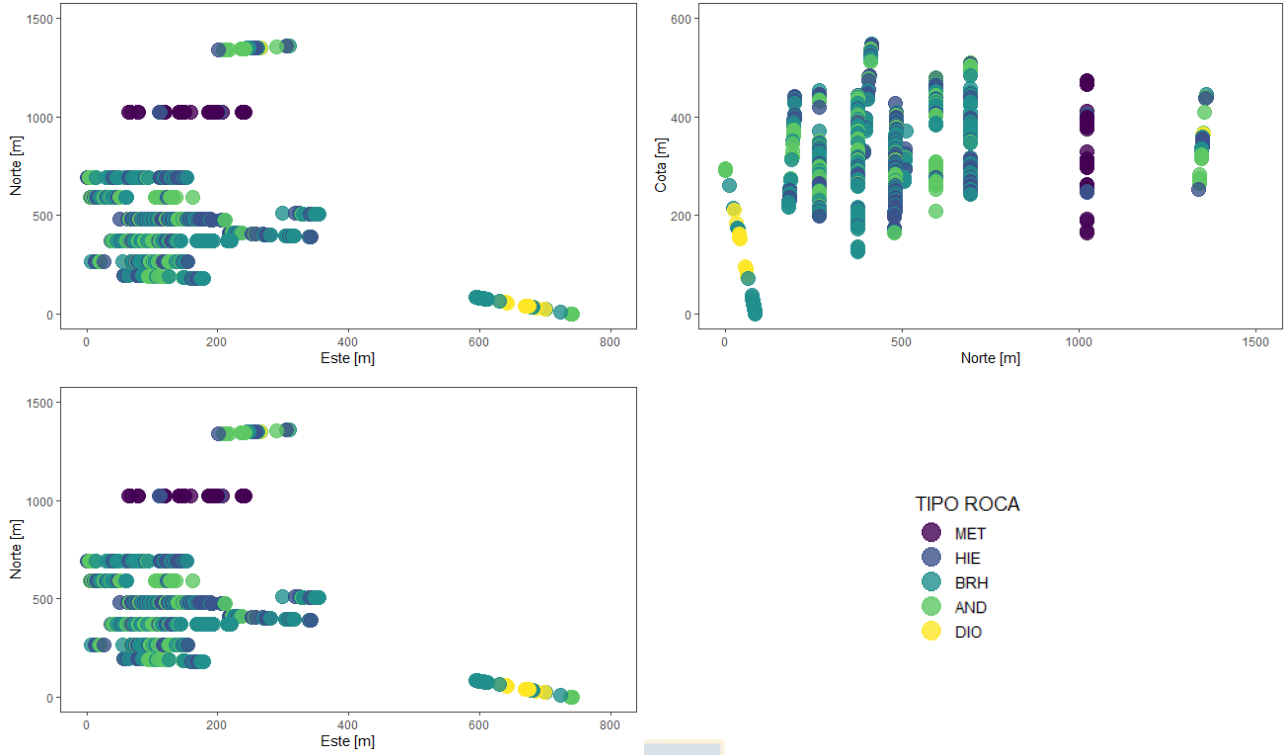


Figura 31: Proyecciones del yacimiento para Tipo de Roca

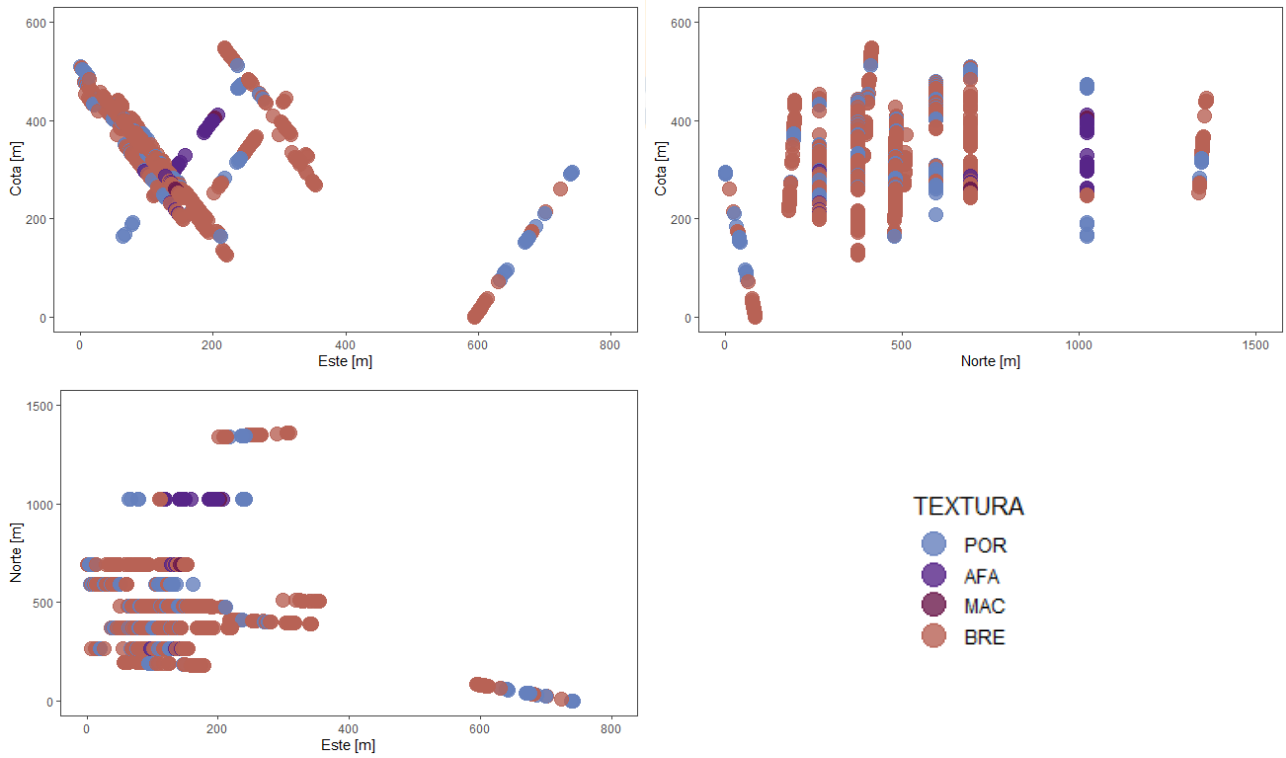


Figura 32: Proyecciones del yacimiento para Texturas

En las proyecciones anteriores se observa el muestreo con sondajes dispuestos en dos inclinaciones diferentes (45° y 135° aproximadamente), los que abarcan una extensión bastante amplia teniendo en cuenta el número de sondajes, además se aprecia un sondaje relativamente aislado del grupo principal. En la **Figura 30** se observa un comportamiento errático de la ley, principalmente en el centro del muestreo, mientras que los sondajes más alejados del centro muestran un comportamiento más bien continuo, pero con leyes bajas. En la **Figura 31**, se observa que predomina el hierro de mena (HIE), brechas (BRH) y andesitas (AND) en el centro del muestreo, hay presencia de metandesitas (MET) en sólo un sondaje, mientras que las dioritas (DIO) se pueden encontrar únicamente en el sondaje aislado. Finalmente, en la **Figura 32** no se destaca ningún patrón para la textura porfídica (POR) ni para la brechosa (BRE), que son las más predominantes dentro del muestreo, mientras que la textura afanítica (AFA) y maciza (MAC) se concentran principalmente en un sondaje.

5.1.2. Estadísticas básicas

A continuación, en las **Tablas 6 y 7**, se presentan las estadísticas básicas de las leyes de hierro, desglosadas por tipo de roca y textura, en conjunto con los valores globales.

Tabla 6: Estadísticas básicas por Tipo de Roca

Parámetro	FE	FE_MET	FE_HIE	FE_BRH	FE_AND	FE_DIO
Media [%]	32.70	26.52	43.36	28.49	24.99	23.91
Desviación Estándar [%]	11.66	8.48	11.09	7.45	6.18	9.01
Mínimo [%]	15.53	18.00	18.38	15.53	15.68	15.53
Máximo [%]	67.13	47.81	67.13	50.17	49.17	45.88
Coefficiente de Variación	0.36	0.32	0.26	0.26	0.25	0.38
Cuenta	660	33	223	273	117	14

Tabla 7: Estadísticas básicas por Textura

Parámetro	FE	FE_POR	FE_AFA	FE_MAC	FE_BRE
Media [%]	32.70	24.90	27.34	59.49	34.34
Desviación Estándar [%]	11.66	6.50	9.38	11.21	11.51
Mínimo [%]	15.53	15.53	18.00	35.27	15.53
Máximo [%]	67.13	49.17	47.81	67.13	66.04
Coefficiente de Variación	0.36	0.26	0.34	0.19	0.34
Cuenta	660	112	29	7	512

En la **Tabla 6** se destaca que las mayores leyes de hierro se asocian al hierro de mena (HIE) mientras que las menores corresponden a las dioritas (DIO) y andesitas (AND). El coeficiente de variación de la ley hierro es medianamente alto, característico de una distribución asimétrica como se muestra en el histograma de la **Figura 33**, al comparar los coeficientes de variación de ley por tipo de roca se encuentra un comportamiento más heterogéneo en las metandesitas (MET) y dioritas (DIO).

En la **Tabla 7** las leyes más altas se asocian a la textura maciza (MAC), donde a la vez presenta el coeficiente de variación más bajo en comparación a las demás texturas, indicando una distribución más homogénea de la ley, pero es poco representada dentro del muestreo. Por otro lado, la textura brechosa (BRE) se asocia a leyes medianas con un coeficiente de variación alto, pero representa la textura predominante dentro del muestreo.

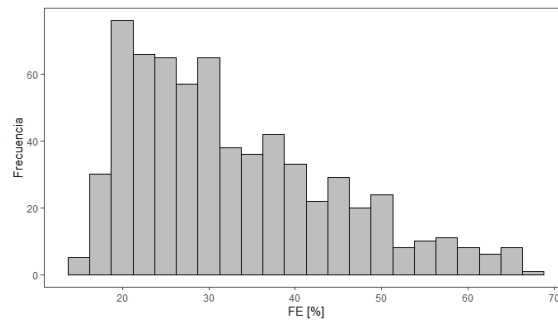


Figura 33: Histograma Ley de Fe [%]

En la **Figura 34** se presentan box-plots de la ley de hierro desglosada por tipo de roca y texturas. En el box-plot global de la ley de Fe se visualiza que los valores siguen una distribución asimétrica positiva, reuniéndose la mayor cantidad de muestras alrededor de una ley de 30%, los valores atípicos pueden explicarse por los altos valores asociados a la textura maciza (MAC) los cuales no difieren de un escenario común. En los box-plots desglosados se pueden observar varios valores atípicos por categoría, si bien son valores realistas, pueden dificultar la estimación de la ley. En el caso del tipo de roca se puede observar que la mayoría de las cajas se solapan, es decir, no existe una diferencia significativa entre sus medias, con excepción para el hierro de mena (HIE) que posee valores levemente más altos. También hay solapamiento entre la mayoría de texturas, pero hay una gran diferencia con la textura maciza (MAC), que como se menciona anteriormente es poco representada dentro del muestreo.

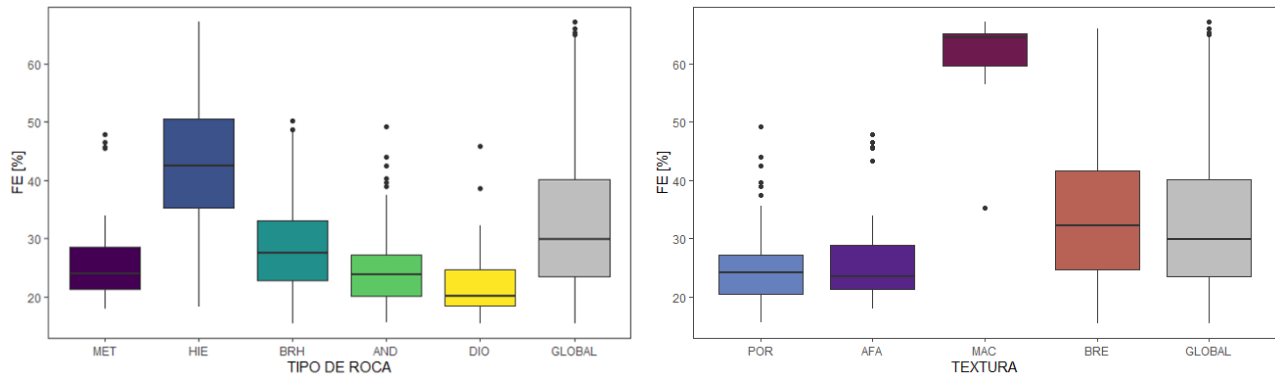


Figura 34: Box-Plots de ley de Fe [%] por tipo de roca (izquierda) y textura (derecha)

En las Figuras 35 y 36 se estudia el comportamiento de la media de hierro a través de los diferentes ejes coordenados con el fin de verificar la estacionaridad, desglosando los valores de ley según el tipo de roca y textura. Se encuentra una tendencia de disminución de la ley de Fe a mayor profundidad (bajo la cota 300 m) la cual se sigue en casi todos los tipos de roca y textura, exceptuando las andesitas (AND) y la textura brechosa (BRE). También hay una pequeña disminución en la ley en la dirección norte bajo los 400 m, la cual se reitera para las brechas (BRH) y andesitas (AND). En la dirección este no se encuentran tendencias significativas de cambio en el valor de la ley media de hierro en términos generales, pero si se puede observar cambios en la ley asociada a las metandesitas (MET) y dioritas (DIO). La dirección en la que se observa mayor dependencia es en la vertical por lo que se probara introducir esta tendencia en el *kriging* universal.

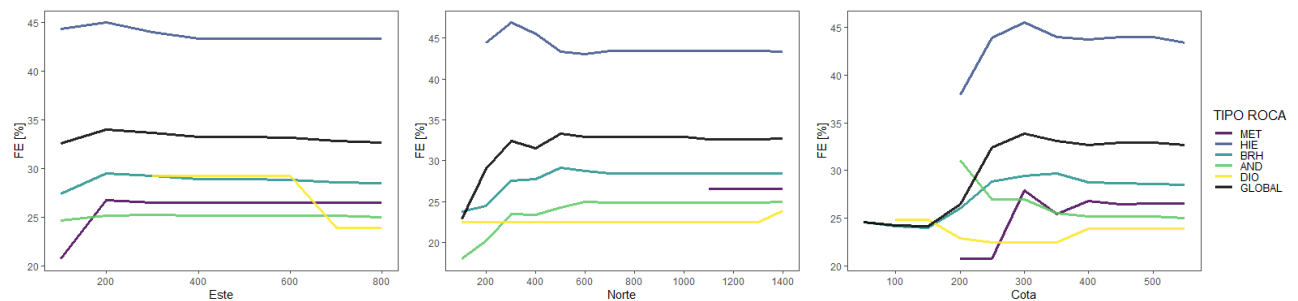


Figura 35: Ley media vs ejes coordenados (este, norte, cota) desglosado por tipo de roca

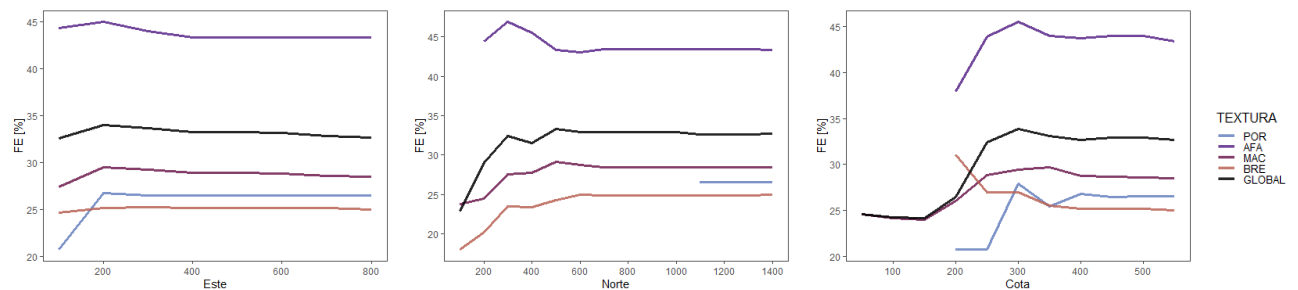


Figura 36: Ley media vs ejes coordenados (este, norte, cota) desglosado por textura

5.2. Aplicación

Para la construcción de los modelos de tipo de roca, texturas y posteriormente de leyes, es necesario en primer lugar una grilla de estimación, la cual es construida considerando un tamaño de bloque de $20 \times 20 \times 20 \text{ m}^3$ que contiene todos los sondajes de la base en su interior, resultando un modelo con 63313 bloques sobre el cual se realizan las diferentes estimaciones.

Los algoritmos en estudio son aplicados con los mismos paquetes de R utilizados en el caso de estudio simulado, spMC para el *co-kriging* indicador, C5.0 para los árboles de decisión, pero por motivos del costo de tiempo empleado en la búsqueda de una red neuronal adecuada con el paquete Neuralnet, se recurre al paquete nnet. El paquete nnet utiliza redes neuronales con conexiones hacia adelante con una sola capa oculta, a diferencia de Neuralnet que permite una mayor flexibilidad en la arquitectura de la red, no obstante, nnet permite utilizar la función de activación *softmax* (ecuación (68) sección 2.10) la cual es empleada con mucho éxito en problemas de clasificación. Con este paquete es necesario ir probando diferentes números de neuronas para la capa oculta y controlar el número de iteraciones para disminuir el sobreajuste [50].

No es posible contar con un conjunto exclusivo para la validación de los modelos, dado que se tiene un muestreo reducido en contraste con la extensión del modelo de bloques, y además se cuenta con categorías poco representadas, por lo tanto, se recurre a una validación cruzada del tipo *leave-one-out* con 660 iteraciones que coincide con el número de muestras, implicando un gran costo computacional. El algoritmo más costoso en términos computacionales son las redes neuronales seguido por el *co-kriging* indicador y finalmente el menos costoso corresponde a los árboles de decisión, este último tiene la ventaja de ser muy rápido (aproximadamente 0.5 segundos por iteración).

Es importante notar que para seleccionar los parámetros de búsqueda en la construcción del transiograma en el *co-kriging* indicador se realiza la validación cruzada iterando diferentes tolerancias de búsqueda y escogiendo para la posterior estimación sobre el modelo de bloques aquellos parámetros que maximicen la precisión de la validación. Del mismo modo, en las redes neuronales se validan distintas arquitecturas de red (número de neuronas) variando además el número de iteraciones para seleccionar finalmente aquellos parámetros que entreguen la mayor precisión de validación. En el caso de los árboles de decisión, dado que el algoritmo realiza por si solo el proceso de post-poda solamente se valida un árbol por modelo. Además de evaluar la

precisión obtenida por validación, se realiza un análisis de sensibilidad y especificidad por categoría con la ayuda de la curva ROC.

Para la estimación de leyes de FE, por medio de *kriging* ordinario y universal, también se realiza validación cruzada con la ayuda de la función “krige.cv” del paquete gstat para buscar los modelos de variograma más adecuados y comparar además la capacidad de predicción de ambos métodos en el contexto estudiado. En el caso del *kriging* universal la validación se utiliza también para elegir las variables que actúan como tendencia en el modelo (tipo de roca, textura y coordenadas), resultando la variable tipo de roca la que más contribuye a la descripción de la ley, aportando en menor medida el tipo de textura y la coordenada vertical.

Sobre el modelo de bloques se construyen los modelos geológicos con los diferentes algoritmos (con sus parámetros seleccionados en la validación), estos actúan como variables de entrada en la posterior estimación de leyes, la cual es utilizada en la categorización de recursos junto con el criterio detallado en la **Tabla 8**.

Tabla 8: Criterio de clasificación de recursos para caso real

Recurso Medido	varianza relativa de kriging < 30%
Recurso Indicado	varianza relativa de kriging 30% – 50%
Recurso Inferido	varianza relativa de kriging > 50%

5.3. Análisis de resultados

5.3.1. Estimación de modelos geológicos

Las precisiones obtenidas por los diferentes algoritmos en estudio al realizar la validación de la construcción de los modelos de tipo de roca y texturas, se encuentran en la **Tabla 9**. Se puede observar que, al igual como sucede en el caso de estudio simulado, las precisiones mayores corresponden a las redes neuronales para ambos modelos, y a la vez los árboles de decisión superan en precisión al *co-kriging* indicador en el modelo de tipo de roca, opuesto a lo que ocurre con el modelo de texturas. Es necesario recalcar que en los modelos de tipo de roca se alcanzan precisiones considerablemente bajas en comparación con el modelo de texturas, esto puede explicarse por la complejidad extra de trabajar con una mayor cantidad de categorías con alta variabilidad espacial, de hecho en la **Figura 31** en las zonas con mayor presencia de hierro de mena (HIE), brechas (BRH) y andesitas (AND) se aprecia un comportamiento más bien errático del tipo de roca.

Tabla 9: Precisiones de validación en construcción de modelos geológicos por algoritmo

Método / Modelo	Precisión [%]	
	Tipo de Roca	Texturas
Co-Kriging Indicador	55.91	83.64
Árboles de Decisión	58.18	82.88
Redes Neuronales	63.64	86.52

Para analizar el desempeño de los algoritmos para cada categoría del modelo se construyen curvas ROC y se comparan las áreas bajo la curva resultantes por método según la categoría en estudio, las curvas se encuentran ilustradas en la **Figuras D.1 y D.2** del anexo D, en tanto los resultados se resumen en la **Tabla 10**. Comparando las áreas bajo las distintas curvas ROC se observa que a pesar de que se obtienen mejores precisiones con las redes neuronales, el *co-kriging* indicador consigue una mayor capacidad discriminadora para las categorías del modelo de textura, por otro lado, para el tipo de roca las redes neuronales promedian un mayor AUC, pero el *co-kriging* indicador y el árbol de decisión obtienen un mayor rendimiento para la metandesitas (MET) en comparación con la red neuronal.

Tabla 10: Áreas bajo la curva (ROC) según categorías de modelos de tipo de roca y alteración y algoritmo utilización como clasificador

		Co-Kriging Indicador	Árboles de Decisión	Redes Neuronales
Textura	POR	0.819	0.629	0.782
	AFA	0.853	0.857	0.823
	MAC	0.998	0.712	0.854
	BRE	0.818	0.679	0.794
	Promedio	0.872	0.719	0.813
Tipo de Roca	MET	0.997	0.997	0.967
	HIE	0.656	0.638	0.701
	BRH	0.616	0.651	0.688
	AND	0.711	0.667	0.728
	DIO	0.783	0.855	0.890
	Promedio	0.753	0.761	0.794

En las **Figuras 37** y **38** se ilustran cortes transversales resultantes de la estimación de los modelos de tipo de roca y textura respectivamente sobre el modelo de bloques.

Para el modelo de tipo de roca se aprecian diferencias considerables entre los modelos, lo cual es entendible por la baja precisión alcanzada por los tres métodos, se destacan algunas similitudes, por ejemplo, tanto el *co-kriging* indicador como los árboles de decisión aproximan una sección completa de metandesita (MET) a través de la mayoría de cortes, esto podría explicarse por el hecho de que en el sondeo de este sector predomina en cantidad este tipo de roca (**Figura 31**) y tanto los árboles de decisión como el *co-kriging* indicador utilizan la información más cercana para hacer predicciones, esto no ocurre en el modelo estimado por redes neuronales, ya que este algoritmo utiliza toda la información entregada para ajustar la red. En los modelos de redes neuronales y árboles de decisión predominan las brechas (BRH) y andesitas (AND) por sobre el hierro de mena (HIE), opuesto a lo que sucede en el modelo del *co-kriging* indicador y dado que el hierro de mena está asociado a leyes más altas se espera que el modelo estimado por *co-kriging* indicador este asociado también a leyes más altas. Dado las disimilitudes entre modelos es poco confiable inclinarse por un modelo en particular sin la opinión de expertos. Es recomendable aumentar la cantidad de sondajes para lograr representar correctamente las categorías con pocas muestras y si es posible conservar un conjunto de la información para validar y calibrar los modelos.

Para el modelo de texturas a grandes rasgos existen similitudes entre los modelos para las texturas brechosa (BRE) y porfídica (POR), las cuales resultan ser las más representadas dentro del muestreo, pero hay poca concordancia para las texturas afanítica (AFA) y maciza (MAC). Cabe notar que al igual que en el caso simulado se obtienen modelos más simplificados al estimar atributos del modelo geológico mediante árboles de decisión.

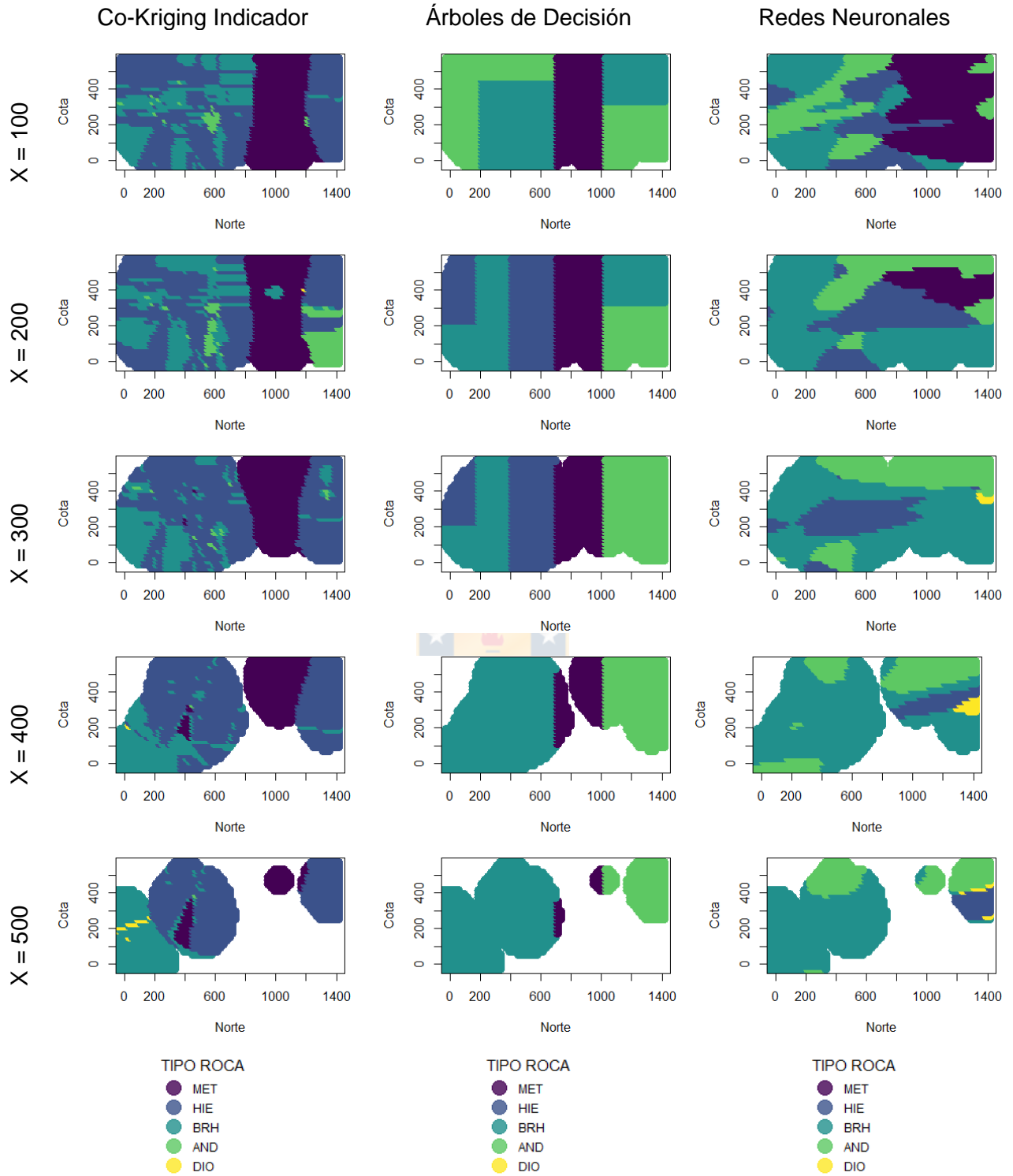


Figura 37: Secciones transversales para modelo de tipo de roca según clasificador

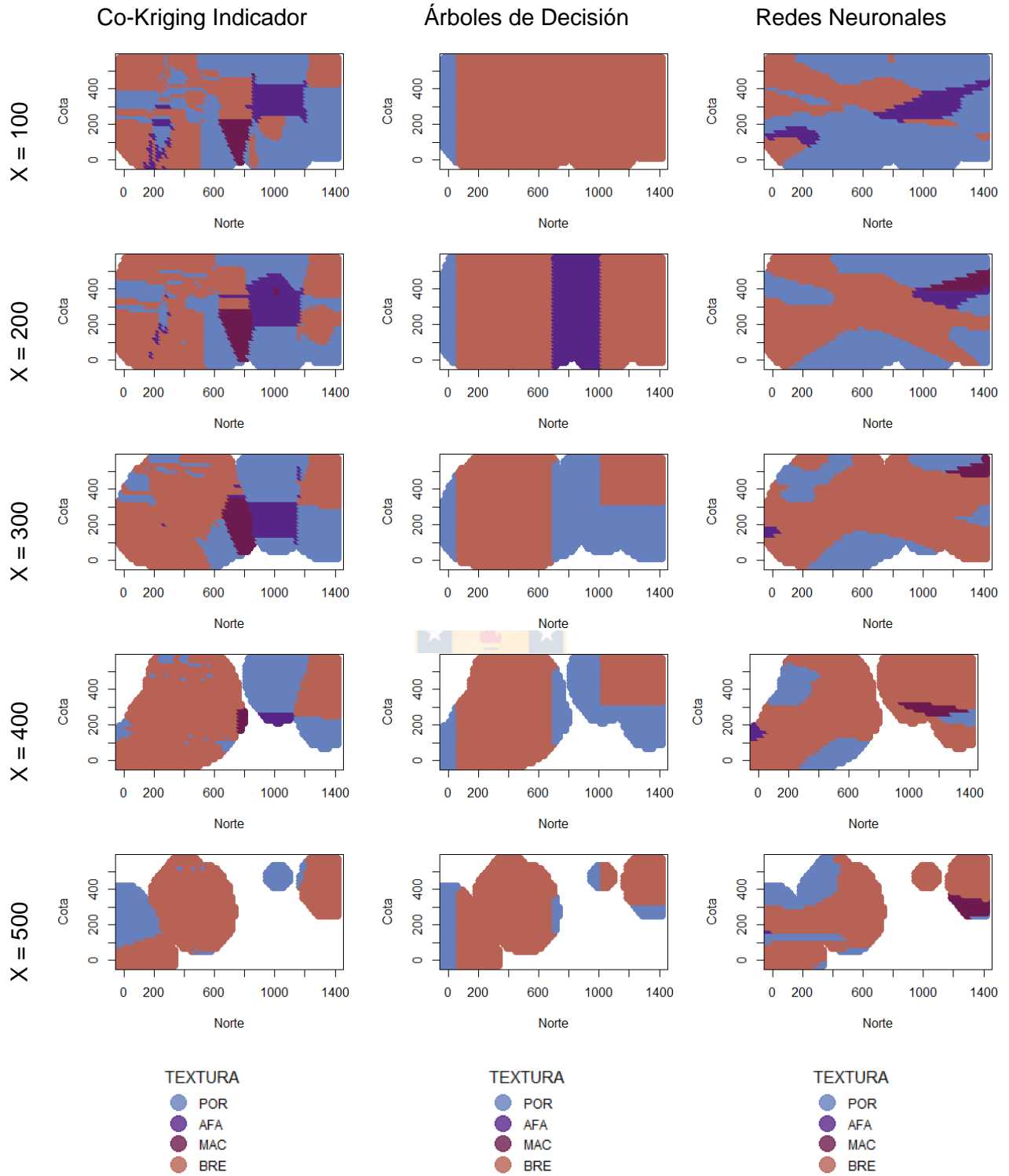


Figura 38: Secciones transversales para modelo de texturas según clasificador

5.3.2. Estimación de leyes y categorización de recursos

En primer lugar, se realiza la validación cruzada del *kriging* ordinario y *kriging* universal utilizando como tendencias los modelos de tipo de roca y textura, además de la coordenada vertical (en el anexo E se pueden comparar los variogramas obtenidos en cada caso). En la **Tabla 11** se comparan las estadísticas resultantes en las validaciones.

Tabla 11: Estadísticas validación *kriging* ordinario y *kriging* universal

Estadística	Kriging Ordinario	Kriging Universal
RMSE	9.46	7.33
Media del error absoluto	7.42	5.60
Ley media estimada [%]	32.89	32.79
Ley media real [%]	32.70	32.70
Correlación	0.60	0.78

El *kriging* universal disminuye el error cuadrático medio y la media del error absoluto en comparación con el *kriging* ordinario, no obstante, estos errores siguen considerándose altos y pueden explicarse por el alto coeficiente de variación en los valores de la ley en términos generales y desglosada por los diferentes atributos del modelo geológico que además han sido poco representados en el muestreo. Cabe notar que ambos métodos logran aproximarse a la ley media real.

En la **Figura 39** se presentan las nubes de dispersión entre leyes estimadas y las reales para el *kriging* ordinario y universal, en estas se evidencia que el *kriging* universal obtiene una mayor correlación entre la predicción y la observación por lo que este último resulta más confiable a la hora de estimar el modelo de leyes para el yacimiento en estudio.

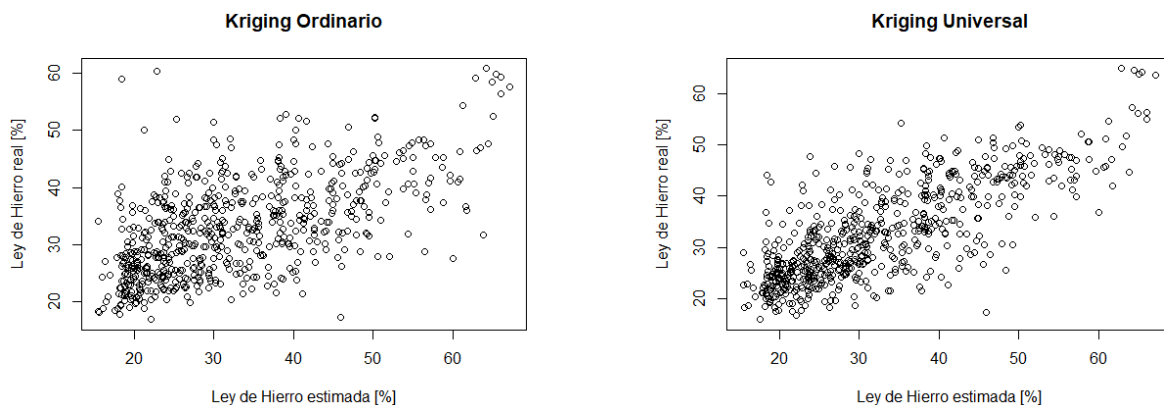


Figura 39: Nubes de dispersión entre leyes estimadas y reales para *kriging* ordinario (izquierda) y *kriging* universal (derecha)

Aplicando el criterio de la varianza relativa de *kriging* según los parámetros de la **Tabla 8** sobre los modelos de leyes estimados por *kriging* ordinario y *kriging* universal, se realiza la categorización de recursos cuyos resultados se resumen en la **Figura 40**, obteniéndose una cantidad significativamente mayor de bloques medidos e indicados al utilizar *kriging* universal en comparación con el *kriging* ordinario, es decir, que el *kriging* universal en este caso también disminuyó la varianza de *kriging*.

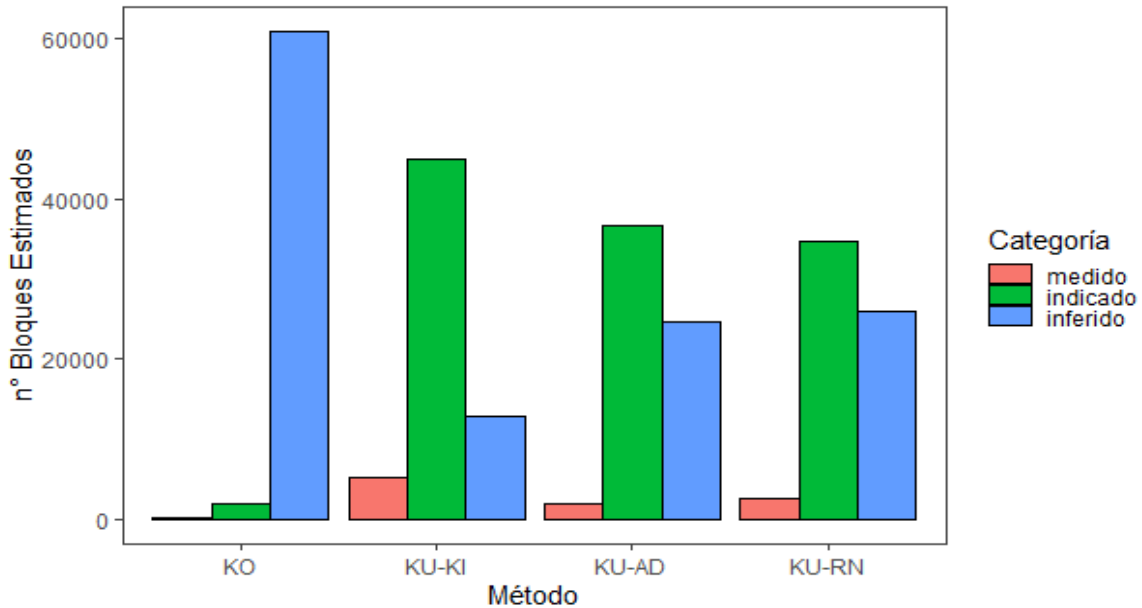


Figura 40: Cantidad de bloques estimados en las categorías medido, indicado e inferidos según metodología aplicada en la estimación de la ley

El escenario más pesimista, que a la vez resulta el menos confiable por lo demostrado en la validación, corresponde al que utiliza *kriging* ordinario, básicamente por tener una gran cantidad de bloques inferidos. El caso más optimista corresponde al que utiliza *kriging* universal junto con *co-kriging* indicador para la estimación del modelo geológico, esto se explica con que el modelo de tipo de roca está compuesto principalmente por hierro de mena (HIE), roca asociada a leyes altas en el muestreo, sumado a que el criterio de categorización empleado (varianza relativa de *kriging*) tiende a favorecer los bloques con mayor ley, resultando así un mayor tonelaje de bloques medidos e indicados. Por otro lado, en la utilización del *kriging* universal junto con redes neuronales o árboles de decisión se obtienen una cantidad de bloques medidos e indicados similar, mucho mayor a la cantidad del *kriging* ordinario, pero menor respecto al *kriging* universal con *co-kriging* indicador, esto por tener un modelo de tipo de roca asociado a leyes más bajas.

En la **Tabla 12** se resume el tonelaje y ley media para los bloques medidos e indicados según la metodología empleada, cabe señalar que para el cálculo de tonelaje la densidad en los yacimientos de hierro varía según la ley, por lo tanto, las leyes estimadas también afectan los tonelajes esperados. Para el cálculo de densidad se utilizó la siguiente relación⁵:

$$D(\text{ton}/\text{m}^3) = 2.2959 + 0.03793 \times \text{Ley de FE} [\%] \quad (75)$$

Tabla 12: Tonelaje y ley media para bloques medidos e indicados según metodología

Método	Medidos		Indicados	
	Tonelaje [MTON]	Ley Media [%]	Tonelaje [MTON]	Ley Media [%]
Kriging Ordinario	9.20	33.38	55.75	31.74
KU + Co-Kriging Indicador	169.81	44.23	1277.46	33.08
KU + Árbol de Decisión	56.11	37.67	994.04	28.77
KU + Red Neuronal	83.29	40.31	949.50	29.63

Si bien es justificable inclinarse por los métodos que utilizan *kriging* universal para la estimación de ley, es necesario resolver las discrepancias y mejorar el desempeño en la estimación del modelo de tipo de roca antes de introducirlo como información extra en la estimación de la ley, para así no caer en el error de sobreestimar o subestimar el valor del yacimiento.

⁵ Curva de correlación %FE/Densidad establecida para mina El Tofo

CAPÍTULO 6

CONCLUSIONES Y DISCUSIONES

En primer lugar, respecto al caso de estudio simulado se confirma que las herramientas provenientes del *Data Mining* son capaces de predecir atributos categóricos asociados al modelo geológico de un depósito mineral al mismo nivel e incluso mejor que una herramienta geoestadística ampliamente utilizada en la industria minera como lo es el *co-kriging* indicador. Respecto a este análisis se tienen las siguientes consideraciones:

- En términos de precisión las redes neuronales son superiores a los árboles de decisión y al *co-kriging* indicador, aun así, debido a la flexibilidad de las redes en cuanto a su configuración no se puede afirmar que la red seleccionada (luego del proceso de validación) sea la que mejor se adapta al problema, ya que dentro de una infinidad de configuraciones y debido al coste computacional que estas implican se prueba un rango acotado de número de neuronas y de criterios de detención, por consiguiente existe la posibilidad de lograr mejores resultados siempre y cuando se cuente con el tiempo y capacidad de *hardware* necesarios para iterar un mayor número de validaciones.
- Si bien los árboles de decisión no superan en precisión a las redes ni al *co-kriging* indicador y suelen construir modelos más simplificados, tienen la ventaja de ejecutarse en un tiempo considerablemente corto en comparación con los demás algoritmos independiente de la cantidad de información entregada, lo que puede resultar de gran ayuda en análisis preliminares, además por ser un algoritmo asequible y simple, siempre puede ser considerado como un modelo alternativo o ayudar a expertos en la interpretación geológica sin una demanda grande de tiempo.
- En cuanto al análisis concerniente a la cantidad de información entregada para la construcción de los modelos si bien se observó una tendencia del aumento de la precisión con el aumento de esta, no resulta ser el factor más importante, ya que la calidad de la información obtenida en el muestreo condiciona fuertemente el desempeño de los algoritmos. Se debe tener en consideración que las precisiones de los algoritmos comienzan a asemejarse al introducir cantidades exageradas de información. Además, las redes neuronales resultan tener una menor dependencia de la cantidad de información en comparación con los demás algoritmos empleados.
- La complejidad de los modelos también afecta el desempeño de los algoritmos, esto se comprueba con que el modelo con menos categorías, o sea el de alteración, logra siempre mejores precisiones que el modelo de tipo de roca que contiene más categorías.

Antes de discutir sobre el caso de estudio real es necesario recalcar las principales diferencias entre la base real y las bases simuladas para lograr comprender las diferencias en sus resultados.

- En el caso de estudio simulado se presenta una ley mineral con mayor continuidad espacial comparado con el caso real, incluso en este último hay presencia de leyes más bien erráticas.
- Igualmente, en el caso real se aprecian categorías con poca continuidad espacial, especialmente las asociadas al modelo de tipo de roca, lo que aumenta la complejidad en la construcción de este modelo.
- Si bien las primeras simulaciones del caso de estudio simulado poseen poca información en el conjunto de entrenamiento, la extensión de los bloques a estimar es considerablemente menor a la del caso de estudio real que por cierto posee pocos sondajes.

Además, la base real comprende otras dificultades observables a priori, como categorías con una representatividad baja, un coeficiente de variación respecto a la media de la ley medio-alto, sondajes aislados y por lo tanto espacios importantes sin muestrear. Por otro lado, la validación de los modelos estimados, debido a la baja densidad del muestreo no puede realizarse con un conjunto de validación exclusivo, obligando a utilizar validación cruzada midiendo la precisión media asociada a predecir correctamente puntos contenidos en los sondajes que posteriormente se utilizan en la construcción del modelo final.

Con esto una de las principales conclusiones que deja el caso de estudio real es la importancia de contar con un conjunto de datos exclusivo para la validación, que permita asegurar la correcta calibración de los parámetros de los algoritmos y/o tener una medida confiable del error de predicción, ya que la validación cruzada no es suficiente para detectar el sobreajuste y más aun no permite evaluar el modelo final.

Con el caso de estudio real se sostiene que las herramientas provenientes del *Data Mining* pueden predecir atributos categóricos asociados al modelo geológico, se logran resultados aceptables para el modelo de texturas, a pesar de la escasez de muestras y la poca representatividad de algunas categorías, pero no se puede afirmar lo mismo para el modelo de tipo de roca, las precisiones de la validación son bajas para los tres algoritmos y los modelos clasificados difieren significativamente entre ellos, esto debido a la complejidad que se ha mencionado anteriormente, por lo tanto se recomienda:

- Aumentar la cantidad de muestras que aseguren la representatividad de todas las categorías.

- Contar con un conjunto exclusivo para la validación que asegure la correcta calibración de los algoritmos (tolerancias de búsqueda en el *kriging*, número de neuronas y criterio de detención en las redes neuronales).
- En modelos complejos, como el de tipo de roca, considerar la unión de categorías en unidades geológicas que compartan características similares respecto a la ley.

Incorporar los modelos geológicos como *inputs* en la estimación de la ley mediante *kriging* universal no genera diferencias significativas en los valores estimados por *kriging* ordinario en el caso de estudio simulado, debido a que las simulaciones presentan una alta continuidad espacial de la ley, mientras que en el caso de estudio real, el cual presenta una baja continuidad espacial de la ley, resulta en una disminución considerable del error de estimación en comparación con el *kriging* ordinario, por lo tanto si se busca disminuir el error de estimación en contextos de baja autocorrelación de la ley es conveniente optar por el *kriging* universal.

Por otro lado, en ambos casos de estudio resultan en una disminución importante en la varianza de estimación, lo que implica un mayor tonelaje medido e indicado. En particular para el caso de estudio simulado los tonelajes alcanzados con *kriging* universal se asemejan más al tonelaje real de cobre fino contenido en los yacimientos simulados. Por lo tanto, para este caso también resulta que la mejor estrategia es aplicar *kriging* universal.

Con relación a la utilización de los modelos geológicos estimados con el *kriging* universal, se debe recalcar que todos tienen un sesgo asociado a sobrestimar o subestimar la presencia de ciertas categorías, de esto resulta interesante explicar la disposición que tiene cada algoritmo de predecir una categoría por sobre otra, ya que si el modelo tiende a exagerar por ejemplo categorías asociadas a leyes altas el modelo de leyes se inclinará a predecir leyes altas generando un modelo más optimista y viceversa. Por lo tanto, en los casos donde las precisiones de los algoritmos no posean diferencias significativas, se recomienda hacer un análisis por categorías y tomar en consideración los riesgos pertinentes.

En resumidas cuentas, las herramientas provenientes del *Data Mining* poseen un gran potencial de aplicación en la estimación de atributos categóricos del modelo geológico, lo cual puede utilizarse en la etapa de estimación de recursos como también en otras etapas del ciclo minero. Se recomienda seguir estudiando sobre otros casos reales y si es posible comparar los resultados de los algoritmos con información de producción.

CAPÍTULO 7

REFERENCIAS

- [1] **NOBLE, Alan C.** Mineral Resource Estimation. En: *SME MINING ENGINEERING HANDBOOK*. Third Edit. 2001, p. 203-216.
- [2] **MAZUMDAR, Joe y COOK, Brent.** The Abstract Art of Resource Estimation. *Exploration insights newsletter*. 2018, 1-13
- [3] **ROSSI, Mario E. y DEUTSCH, Clayton V.** *Mineral resource estimation*. Springer Science & Business Media, 2013.
- [4] **PÉREZ, Cristian M.** *Modelamiento Geológico Estocástico con Simulación Geoestadística*. 2011. Universidad de Chile.
- [5] **DEUTSCH, Clayton V.** A sequential indicator simulation program for categorical variables with point and block data: BlockSIS. *Computers and Geosciences*. 2006, vol 32(no 10), 1669-1681.
- [6] **DAVID, Michel y BLAIS, Roger A.** Geostatistical ore reserve estimation. *Elservier*. 1997.
- [7] **MATHERON, G.** Principles of geostatistics. *Economic Geology*. 1963, vol 58(no 8), 1246-1266
- [8] **JOURNEL, Andre G. y HUIJBREGTS, Charles J.** *Mining geostatistics*. London: Academic press, 1978.
- [9] **EMERY, Xavier.** Geoestadística. *Universidad de Chile*. 2013, vol 107.
- [10] **GOOVAERTS, Pierre.** *Geostatistics for Natural Resources and Evaluation*. Oxford University Press, 1997.
- [11] **MOLINA, Luis Carlos.** Data mining: torturando a los datos hasta que confiesen. *FUOC*. 2002.
- [12] **TAN, Pan-Ning, STEINBACH, Michael y KUMAR, Vipin.** *Introduction to data mining: pearson new international edition*. Pearson Higher Ed, 2013.
- [13] **CANGREJO, Denisse, SC, Aljure M. y ING, Juan G. Agudelo.** Minería de datos espaciales. *Universidad Nacional de Colombia*. ISSN 1909-0056. 2014, 8(3), 71-78.
- [14] **KOPERSKI, Krzysztof.** *A progressive refinement approach to spatial data mining*. Canada: Simon Fraser University, 1999.
- [15] *Introducción a Data Mining*. 2007. Recuperado de: <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/introduccion-DM.pdf>
- [16] **GIRALDO, Ramón.** *Introducción a la Geoestadística: Teoría y Aplicación*. Bogota: Universidad Nacional de Colombia. 2002.
- [17] **RIVOIRARD, Jacques.** *Course on Multivariate Geostatistics*. 2003.
- [18] **DÍAZ, Martín A.** *Geoestadística Aplicada*. Instituto de Geofísica, UNAM. 2002.
- [19] **JOURNEL, A.G. y ROSSI, M.E.** When Do We Need a Trend Model in Kriging? *Mathematical Geology*. 1989, 21(7), 715-739.

- [20] **IGLESIAS REY, Susana.** *Comparación de estimadores de la función de distribución para datos espaciales.* 2016. Universidad de Santiago de Compostela.
- [21] **JOURNEL, Andre G.** Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology.* 1983, 15(3), 445-468.
- [22] **CHILES, JEAN-PAUL y DELFINER, PIERRE.** *Geostatistics: Modeling Spatial Uncertainty.* John Wiley & Sons, 1999.
- [23] **DEUTSCH, Clayton V. y JOURNEL, Andre G.** *Geostatistical software library and user's guide.* New York: Oxford University Press, 1998.
- [24] **FAYYAD, Usama M., PIATETSKY-SHAPIO, Gregory y SMYTH, Padhraic.** Knowledge Discovery and Data Mining: Towards a Unifying Framework. *KKD.* 1996, 145(2), 82-88.
- [25] **BALCÁZAR, José L.** *Minería de Datos.* Dept. LSI, UPC. 2012.
- [26] **GUTIÉRREZ, José M.** *Data Mining: Extracción de Conocimiento en Grandes Bases de Datos.* 2001.
- [27] **BERZAL, Fernando.** *Introducción al Data Mining.* Departamento de Ciencias de la Computación e IA, Universidad de Granada. 2015.
- [28] **KOHAVI, Ron y PROVOST, Foster.** Glossary of Terms. En: *Machine Learning.* 1998, p. 271-274. ISSN 00138703. Disponible en: doi:10.1023/A.
- [29] **GARCÍA, Arno D.** *ANÁLISIS DE DATOS Y BÚSQUEDA DE PATRONES EN APLICACIONES MÉDICAS.* 2015. Universidad de Chile. Disponible en: doi:10.30798/makuiibf.323102.
- [30] **BURKOV, Andriy.** *The Hundred - Page Machine Learning.* Quebec City: Can.: Andriy Burkov, 2019. ISBN 199957950X.
- [31] **EXPÓSITO, Christopher.** *Arboles de Decisión.* Universidad de la Laguna: Departamento de Ingeniería Informática y de Sistemas. 2003.
- [32] **QUINLAN, J. Ross.** *Induction of Decision Trees.* Kluwer Academic Publisher. 1986.
- [33] **MOLINA LÓPEZ, José Manuel y GARCÍA HERRERO, Jesús.** *Técnicas de análisis de datos: Aplicaciones prácticas utilizando Microsoft Excel y Weka.* 2006, 266.
- [34] **QUINLAN, J. Ross.** *C4.5: programs for machine learning.* Morgan Kaufmann, 1993.
- [35] **LÓPEZ, Bruno.** *Inteligencia Artificial: Algoritmo C4.5.* 2005.
- [36] **NETTALETON, David.** Capítulo 7: Técnicas. En: *Técnicas para el análisis de datos clínicos.* 2012, p. 119-135.
- [37] **LARRAÑAGA, Pedro, INZA, Iñaki y MOUJAHID, Abdelmalik.** Tema 8. Redes Neuronales. 2015.
- [38] **SINGH, Yashpal y CHAUHAN, Alok Singh.** Neural Networks In Data Mining. *Journal of Theoretical & Applied Information Technology.* 2009, 5.

- [39] **CAPARRINI, Fernando.** *Redes Neuronales: una visión superficial.* 2018 [accedido. 19-05-2019]. Recuperado de: <http://www.cs.us.es/~fsancho/?e=72>
- [40] **SINGH GILL, Navdeep.** *Artificial Neural Networks Applications and Algorithms.* 2019 [accedido. 28-09-2019]. Recuperado de: <https://www.xenonstack.com/blog/artificial-neural-networkapplications/?fbclid=IwAR1EYWPIRvKgF8j8yogfjFZmmhaUcvrOa5JR6goQM4FailtE5wbFrdKlbs>
- [41] **MATICH, Damián Jorge.** *Redes Neuronales: Conceptos Básicos y Aplicaciones.* 2001.
- [42] **BERTONA, Luis Federico.** *Entrenamiento de redes neuronales basado en algoritmos evolutivos.* 2005. UNIVERSIDAD DE BUENOS AIRES.
- [43] **JORC.** *Australasian code for reporting of exploration results, mineral resources, and ore reserves.* Report of the Joint Committee of the Australasian Institute of Mining and Metallurgy, Australian Institute of Geoscientist and Minerals Council of Australia (JORC). 2004.
- [44] **ORTIZ, J.M. y EMERY, Xavier.** Categorización de recursos y reservas mineras. *1 st International Conference on Mining Innovation MININ.* 2004, 198-208.
- [45] **EMERY, Xavier, ORTIZ, Julián M. y RODRÍGUEZ, Juan J.** Quantifying uncertainty in mineral resources by use of classification schemes and conditional simulations. *Mathematical Geology.* ISSN 08828121. 2006, 38(4), 445-464.
- [46] **ANDINA, Matias.** *Introducción a estadística con R.* 2015. Recuperado de: <https://bookdown.org/matiasandina/R-intro/>
- [47] **PEBESMA, Edzer.** The meuse data set : a brief tutorial for the gstat R package. 2019
- [48] **SARTORE, Luca, FABBRI, Paolo y GAETAN, Carlo.** spMC: An R-package for 3D lithological reconstructions based on spatial Markov chains. *Computers and Geosciences.* 2016, 94, 40-47.
- [49] **FRITSCH, Stefan y GUENTHER, Frauke.** *Package 'neuralnet'.* 2019. Recuperado de: <https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf>
- [50] **BRIAN, Ripley y VENABLES, William.** *Package 'nnet'.* 2016.

CAPÍTULO 8

ANEXOS

8.1. Anexo A: Deducción algoritmo *back-propagation*

Como se menciona en la sección 3.7.2. el algoritmo de back-propagation es un método de aprendizaje supervisado de corrección del error para redes neuronales multicapa con conexiones hacia adelante. El algoritmo propone una actualización iterativa de los pesos tal como la ecuación (71). Tomando una variación proporcional al gradiente de una función de error $E(w)$ se tiene:

$$w(t + 1) = w(t) - \alpha \nabla E w[t(t)] \quad (\text{A.1})$$

El primer paso del algoritmo consiste en propagar hacia adelante un patrón de entrada X_p y obtener la salida Y_p . La salida de la neurona i viene dada según su estado de activación. Considerando una función de salida identidad se tiene:

$$y_i(t) = F(f_i(h_i(t))) \quad (\text{A.2})$$

siendo:

$$h_i(t) = \sum_j x_j(t) \cdot w_{ij} \quad (\text{A.3})$$

Comparando la salida obtenida Y_p con la salida deseada \bar{Y}_p se obtiene un error dado por:

$$e_p = \frac{1}{2} \sum_{k=1}^M (\bar{y}_{pk} - y_{pk})^2 \quad (\text{A.4})$$

donde k es el índice de neurona para las neuronas de la última capa, y M el total de neuronas de la misma. El error total de la red está dado por:

$$e = \frac{\sum_{p=1}^P e_p}{P} \quad (\text{A.5})$$

Siendo p el índice de ejemplo, y P el número total de ejemplos.

De acuerdo a la ecuación (A.1) la variación de los pesos sinápticos es proporcional al gradiente de la función error:

$$\Delta w_{ij} = -\alpha \frac{\partial e_p}{\partial w_{ij}} \quad (\text{A.6})$$

Aplicando la regla de la cadena a (A.6) se obtiene:

$$\frac{\partial e_p}{\partial w_{ij}} = \frac{\partial e_p}{\partial h_j} \frac{\partial h_j}{\partial w_{ij}} \quad (\text{A.7})$$

La ecuación (A.6) expresa la derivada del error en función de dos derivadas. La derivada del error respecto al potencial resultante h_j indica como varia el error al variar la entrada de la neurona j , mientras que la derivada respecto al peso sináptico w_{ij} indica como varia la entrada de la neurona j al variar el peso de la conexión que va desde la neurona i hasta la neurona j . El segundo término de la expresión (A.6) se puede expresar a partir de la ecuación (A.3) de la siguiente manera:

$$\frac{\partial h_j}{\partial w_{ij}} = \frac{\partial \sum_i y_{pi} \cdot w_{ij}}{\partial w_{ij}} = y_{pi} \quad (\text{A.8})$$

Escribiendo el primer término de la ecuación (A.7) como:

$$\frac{\partial e_p}{\partial h_j} = -\delta_{pj} \quad (\text{A.9})$$

se tiene:

$$\frac{\partial e_p}{\partial w_{ij}} = -\delta_{pj} y_{pi} \quad (\text{A.10})$$

luego la ecuación (A.6) queda expresada de la siguiente manera:

$$\Delta w_{ij} = -\alpha \delta_{pj} y_{pi} \quad (\text{A.11})$$

Para calcular el valor de delta se vuelve a aplicar la regla de la cadena:

$$\delta_{pj} = -\frac{\partial e_p}{\partial h_j} = -\left(\frac{\partial e_p}{\partial y_{pj}} \frac{\partial y_{pj}}{\partial h_j} \right) \quad (\text{A.12})$$

el cálculo del segundo término de la ecuación (A.12) es simple al observar las ecuaciones (A.2).

$$\frac{\partial y_{pj}}{\partial h_j} = \frac{\partial f_j(h_j)}{\partial h_j} = f'_j(h_j) \quad (\text{A.13})$$

sin embargo, para el cálculo del primer término de la ecuación (A.12) es necesario distinguir dos casos.

- **La neurona j es una neurona de salida**

En este caso se puede obtener el segundo término a partir de la ecuación (A.4) ya que el subíndice j es igual al subíndice k .

$$\frac{\partial e_p}{\partial y_{pj}} = \frac{\partial \frac{1}{2} \sum_{j=1}^M (\bar{y}_{pj} - y_{pj})^2}{\partial y_{pj}} = -(\bar{y}_{pj} - y_{pj}) \quad (\text{A.14})$$

Así, la variación de los pesos de una conexión que va hacia la capa externa de la red se calcula como:

$$\Delta w_{ij} = \alpha (\bar{y}_{pj} - y_{pj}) f'_j(h_j) y_{pj} \quad (\text{A.15})$$

- **La neurona j es una neurona oculta**

En este caso es necesario aplicar nuevamente la regla de la cadena:

$$\frac{\partial e_p}{\partial y_{pj}} = \sum_k \left(\frac{\partial e_p}{\partial h_k} \frac{\partial h_k}{\partial y_{pj}} \right) \quad (\text{A.16})$$

donde k es el subíndice de las neuronas que pertenecen a la próxima capa. La ecuación (A.16) se puede reescribir utilizando la ecuación (A.3).

$$\frac{\partial e_p}{\partial y_{pj}} = \sum_k \left(\frac{\partial e_p}{\partial h_k} \frac{\partial (\sum_j w_{kj} y_{pj})}{\partial y_{pj}} \right) = \sum_k \left(\frac{\partial e_p}{\partial h_k} w_{kj} \right) \quad (\text{A.17})$$

Y por la ecuación (A.9) se tiene que:

$$\frac{\partial e_p}{\partial y_{pj}} = - \sum_k \delta_{pk} w_{kj} \quad (\text{A.18})$$

Así, la variación de los pesos de una conexión que va desde una capa hacia otra de la red que no sea la externa se calcula como:

$$\Delta w_{ij} = \alpha \sum_k (\delta_{pk} w_{kj}) f'_j(h_j) y_{pj} \quad (\text{A.19})$$

En la implementación del algoritmo, se toma una amplitud de paso que viene dado por la tasa de aprendizaje α . A mayor tasa de aprendizaje el proceso será más rápido. Sin embargo, si la tasa de aprendizaje es muy alta puede dar lugar a oscilaciones en torno a un mínimo local. Es posible disminuir el impacto de dichas oscilaciones mediante la adición de un momento β , quedando la expresión (A.11) expresada de la siguiente manera:

$$\Delta w_{ji}(t+1) = \alpha \delta_{pj} y_{pj} + \beta \Delta w_{ji}(t) \quad (\text{A.20})$$

De esta manera el momento β determina el efecto en el instante $t+1$ del cambio de los pesos realizado en el instante t .

La aplicación resumida del algoritmo se encuentra en la **Figura 14** de la sección 3.7.

8.2. Anexo B: Precisiones según modelos de clasificación

Tabla B.1: Precisiones obtenidas según el método de clasificación empleado en la construcción de los modelos de tipo de roca y alteración por simulación (parte 1)

Simulación	Tipo de Roca			Alteración		
	Kriging Indicador	Árbol de Decisión	Red Neuronal	Kriging Indicador	Árbol de Decisión	Red Neuronal
1	61.41	56.53	63.07	61.85	62.65	71.37
2	63.91	58.66	71.84	82.49	71.89	83.89
3	68.63	69.54	74.44	69.11	67.51	78.05
4	57.01	56.9	74.47	77.40	73.66	82.39
5	58.66	55.38	74.18	80.44	77.57	81.58
6	71.33	68.15	77.06	74.93	73.39	82.60
7	70.81	68.64	75.98	73.13	74.21	83.91
8	71.74	62.83	73.63	70.68	72.69	78.73
9	70.02	67.02	80.93	74.20	65.65	76.06
10	67.8	71.93	78.54	79.30	72.42	85.20
11	69.14	64.51	75.72	73.97	73.38	81.11
12	62.19	63.77	71.73	80.66	72.55	85.32
13	71.55	72.02	78.37	76.67	74.10	80.95
14	75.82	66.77	77.68	77.45	68.15	77.59
15	72.23	71.64	80.78	75.15	75.87	80.82
16	70.23	68.42	74.87	85.44	82.89	88.90
17	74.49	70.95	78.67	79.30	75.02	81.79
18	74.19	71.97	76.62	80.68	74.36	86.17
19	75.65	78.48	86.69	77.01	75.15	81.53
20	76.01	73.34	78.98	80.78	69.56	81.71
21	65.81	68.37	75.41	82.82	79.00	84.80
22	68.93	73.2	84.18	77.13	67.58	80.12
23	70.94	65.36	78.54	73.61	64.14	76.61
24	72.21	67.73	82.39	81.27	77.90	84.60
25	71.96	72.94	81.06	80.40	79.10	82.98
26	70.51	67.37	75.86	78.08	71.43	79.98
27	65.66	66.36	75.08	81.85	81.66	84.68
28	76.43	75.88	82.52	85.57	77.61	86.67
29	69.82	62.64	78.42	81.49	73.32	86.05
30	73.27	66.99	80.4	81.30	77.00	84.67
31	73.47	71.07	82.74	86.10	78.27	86.47
32	74.43	72.6	85.17	82.78	77.80	86.11
33	74.11	74.53	80.55	82.29	79.29	85.65
34	66.05	65.1	80.12	77.23	75.88	83.36
35	75.63	74.86	83.07	82.93	79.09	84.65
36	76.12	73.88	79.21	83.14	82.96	86.04
37	77.85	73.73	77.71	84.70	79.82	87.84
38	76.11	74.85	77.57	82.91	78.46	84.70
39	74.11	76.06	80.37	83.60	79.27	86.09
40	73.39	74.96	82.74	81.02	76.61	82.47
41	78.79	79.09	83.55	87.13	85.02	81.02
42	76.91	71.73	83.08	82.02	80.02	87.13
43	71.62	69.04	78.79	80.04	74.19	82.02
44	80.18	75.91	84.02	82.02	79.85	80.04
45	80.4	79.08	85.05	83.18	82.96	82.02
46	73.58	68.68	82.13	75.17	73.41	83.18
47	82.83	79.43	84.22	80.66	78.08	75.17
48	80.83	76.56	87.15	83.93	81.64	80.66
49	75.5	72	78.28	79.16	75.90	83.93
50	73.17	73.77	81.73	86.29	84.70	79.16

Tabla B.2: Precisiones obtenidas según el método de clasificación empleado en la construcción de los modelos de tipo de roca y alteración por simulación (parte 2)

Simulación	Tipo de Roca			Alteración		
	Kriging Indicador	Árbol de Decisión	Red Neuronal	Kriging Indicador	Árbol de Decisión	Red Neuronal
51	74.44	74.41	83.00	87.83	85.78	90.97
52	81.13	77.29	84.30	81.89	74.24	83.85
53	75.63	74.10	84.96	78.85	73.51	83.33
54	79.57	78.79	84.29	86.89	79.94	85.17
55	77.36	74.62	83.26	84.97	83.36	86.35
56	75.87	77.06	83.97	86.82	83.17	87.24
57	83.85	81.23	83.96	90.15	85.22	92.82
58	76.02	68.94	81.75	81.65	74.32	84.51
59	77.52	76.73	84.14	87.12	82.99	89.11
60	84.61	76.19	85.89	84.04	82.72	86.8
61	81.44	77.86	86.60	85.14	84.11	85.05
62	81.18	80.31	86.65	85.25	81.44	86.11
63	77.90	77.59	86.34	87.89	81.4	89.37
64	76.55	75.84	85.36	86.55	81.72	86.52
65	82.11	77.89	84.34	85.26	80.85	86.4
66	84.68	80.59	88.63	86.98	84.69	90.53
67	78.80	69.06	83.76	85.19	77.28	87.3
68	84.00	82.50	84.28	85.19	83.09	88.16
69	76.13	77.42	85.12	80.53	77.45	84.5
70	81.22	77.52	86.32	84.97	80.64	86.32
71	85.62	84.61	88.78	90.09	85.15	91.95
72	76.31	71.95	85.18	82.02	78.45	85.29
73	79.52	81.34	84.71	79.6	73.16	84.12
74	79.88	75.85	83.96	83.31	81.62	87.11
75	78.91	80.72	86.20	87.74	82.94	87.84
76	81.14	78.52	84.57	87.6	86.93	89.26
77	78.55	77.22	85.51	84.24	81.93	86.54
78	80.51	81.78	86.05	85.36	81.93	87.43
79	82.11	78.92	86.40	85.48	81.39	88.02
80	90.75	88.41	90.59	90.46	89.69	91.87
81	87.42	85.48	89.29	89.69	84.71	90.03
82	82.20	79.52	86.67	84.64	79.96	88.02
83	80.82	81.77	87.60	89.01	89.05	91.13
84	79.99	80.33	87.72	89.04	87.00	90.97
85	82.68	78.61	83.26	86.52	84.29	90.15
86	83.26	80.83	84.96	82.08	78.33	83.52
87	79.04	76.39	85.36	86.19	87.21	90.62
88	76.26	71.82	80.64	85.76	81.5	88.35
89	83.03	84.12	89.01	83.46	82.02	87.43
90	81.69	78.01	87.27	84.48	88.45	86.79
91	81.74	83.18	87.92	87.9	85.98	89.35
92	77.37	71.64	83.25	85.82	84.56	89.29
93	83.94	86.22	89.65	82.8	81.39	88.91
94	81.75	80.19	82.24	84.79	82.96	89.77
95	82.29	80.44	84.70	89.79	88.96	92.75
96	80.85	83.98	87.56	86.00	83.3	90.35
97	86.50	78.63	88.00	89.3	87.74	89.69
98	83.13	79.19	86.27	88.08	86.34	91.29
99	81.65	76.91	87.31	88.95	86.63	89.75
100	79.90	77.48	84.16	90.46	86.54	89.83

8.3. Anexo C: Resultados categorización de recursos para caso de estudio simulado

Tabla C.1: Cantidad de bloques clasificados como medidos, indicados o inferidos según metodología de estimación y n° de simulación

SIM	Bloques											
	Kriging Ordinario			Kriging Universal + Co-Kriging Indicador			Kriging Universal + Árbol de Decisión			Kriging Universal + Red Neuronal		
	Medidos	Indicados	Inferidos	Medidos	Indicados	Inferidos	Medidos	Indicados	Inferidos	Medidos	Indicados	Inferidos
1	2708	1234	3219	2755	1229	3177	2757	1221	3183	2757	1225	3179
2	2625	1391	3145	3146	1185	2830	3159	1188	2814	3159	1193	2809
3	1889	1826	3446	3086	1197	2878	3083	1202	2876	3119	1201	2841
4	2644	1571	2946	3381	1211	2569	3367	1218	2576	3378	1220	2563
5	1795	1659	3707	3382	1586	2193	3315	1577	2269	3366	1529	2266
6	492	811	5858	1701	1475	3985	1690	1479	3992	1702	1475	3984
7	2737	1677	2747	4147	882	2132	4142	883	2136	4142	882	2137
8	1062	1528	4571	2719	1605	2837	2704	1617	2840	2705	1608	2848
9	2343	1290	3528	3238	874	3049	3179	900	3082	3241	887	3033
10	2761	1441	2959	3419	1052	2690	3466	1038	2657	3428	1039	2694
11	1994	1782	3385	4423	1003	1735	4383	1014	1764	4410	1006	1745
12	6683	357	121	6734	316	111	6732	317	112	6734	315	112
13	3376	1433	2352	4353	889	1919	4347	880	1934	4342	892	1927
14	2125	1359	3677	3827	1329	2005	3817	1341	2003	3822	1348	1991
15	5163	1338	660	5886	782	493	5890	776	495	5887	792	482
16	2527	1181	3453	3295	1131	2735	3294	1133	2734	3297	1128	2736
17	3009	2427	1725	4929	1305	927	4918	1310	933	4917	1320	924
18	4155	1441	1565	5053	1097	1011	5058	1097	1006	5054	1096	1011
19	3713	1393	2055	4889	796	1476	4890	794	1477	4900	795	1466
20	2535	1713	2913	4398	1283	1480	4392	1287	1482	4401	1284	1476
21	4469	1166	1526	5115	778	1268	5121	776	1264	5123	773	1265
22	5570	725	866	5687	674	800	5673	672	816	5677	665	819
23	4524	1837	800	5889	764	508	5886	768	507	5887	766	508
24	2095	1301	3765	3899	1617	1645	3880	1638	1643	3907	1624	1630
25	3144	2269	1748	4841	1283	1037	4892	1259	1010	4892	1263	1006

8.4. Anexo D: Curvas ROC por categorías (caso de estudio real)

En las **Figura D.1** y **Figura D.2** se encuentran las curvas ROC de las distintas categorías asociadas a los modelos de texturas y tipo de roca respectivamente. En el caso del modelo de textura se destaca la capacidad discriminatoria del *co-kriging* indicador sobre las redes neuronales a pesar de que esta última obtiene una mayor precisión. Para el modelo de tipo de roca sobresale mayoritariamente las redes neuronales con excepción para la metandesita (MET).

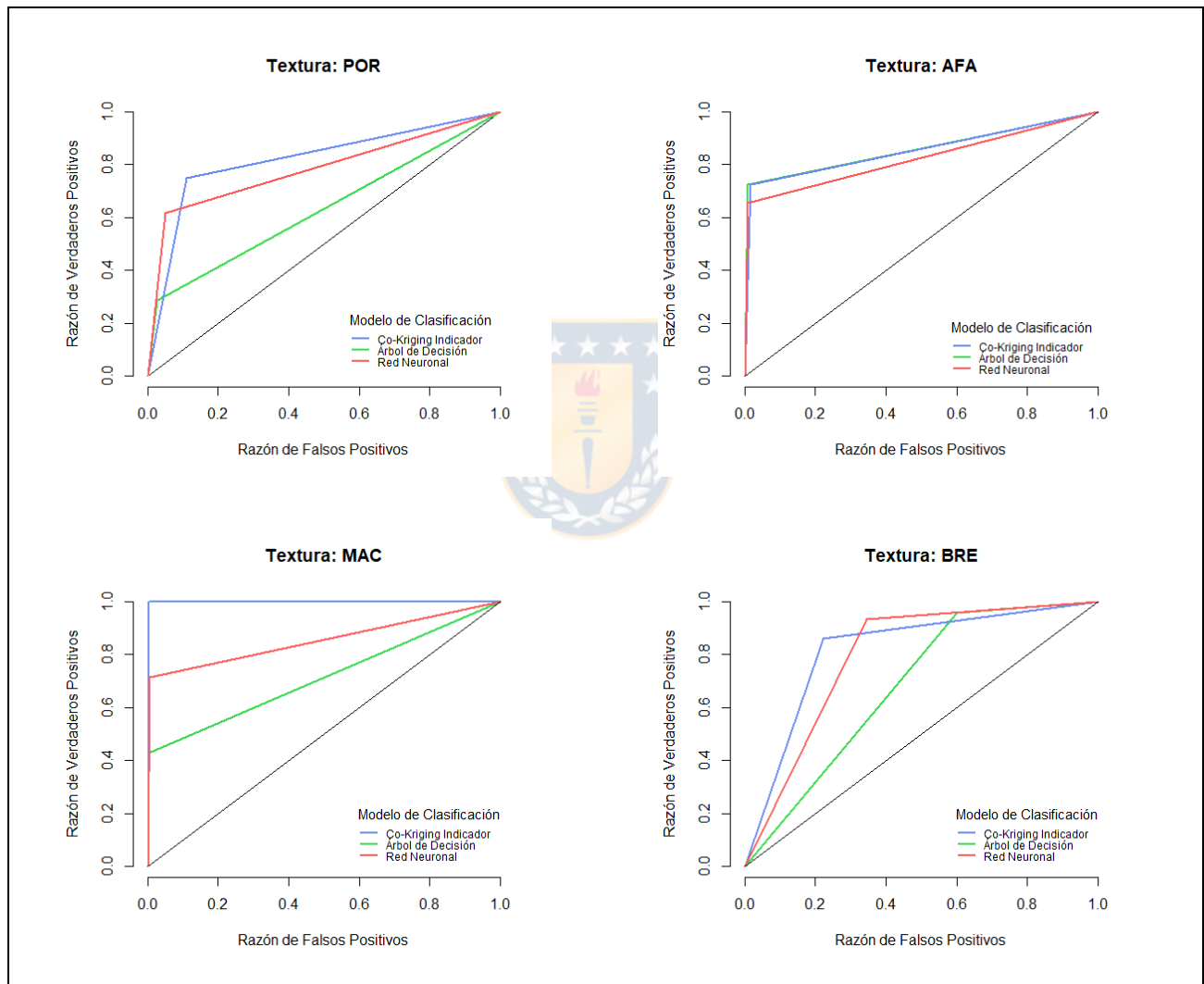


Figura D.1: Curvas ROC según tipo de roca y modelo de clasificación

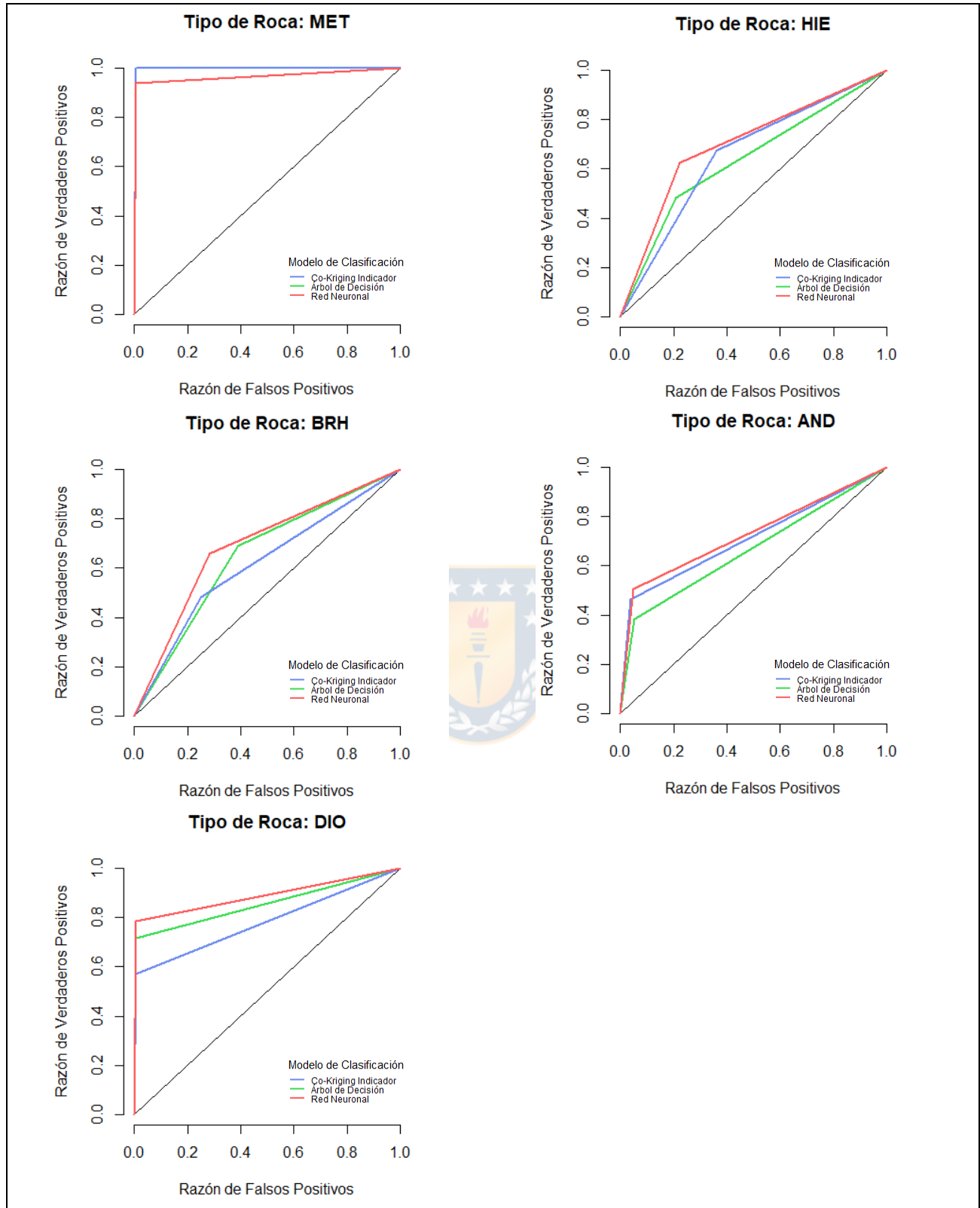


Figura D.2: Curvas ROC según textura y modelo de clasificación

8.5. Anexo E: Variogramas caso de estudio real

En las siguientes figuras se comparan los variogramas utilizados para la estimación de *kriging* ordinario y universal respectivamente. El variograma residual es construido mediante la dependencia de la ley de Hierro con el tipo de roca, textura y la coordenada vertical. Las principales diferencias se encuentran en una disminución de la meseta en el variograma residual junto con un aumento del efecto pepa en el origen del modelo.

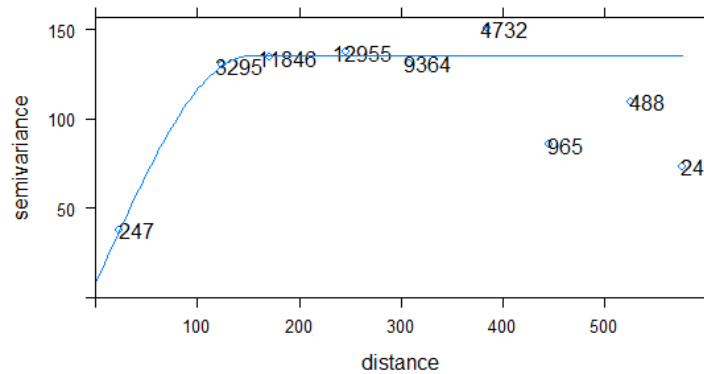


Figura E.1: Variograma omnidireccional de la ley de Fe

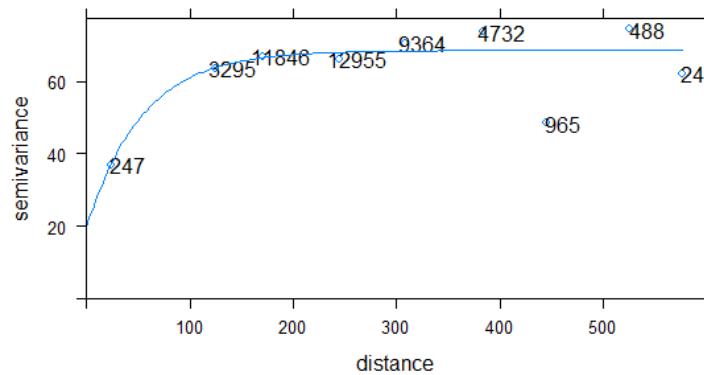


Figura E.2: Variograma residual de la ley de Fe en función del tipo de roca, textura y coordenada vertical

UNIVERSIDAD DE CONCEPCIÓN - FACULTAD DE INGENIERÍA
Departamento de Ingeniería Metalúrgica
 Hoja Resumen Memoria de Título

Título:	
Estimación de atributos categóricos del modelo geológico mediante herramientas de Data Mining	
Nombre Memorista: Valentina Esperanza Neira Vidal	
Modalidad	Profesor (es) Patrocinante (s)
Concepto	Prof. Roberto Fustos T.
Calificación	
Fecha	
28/05/2020	
Prof.	Ingeniero Supervisor:
	Institución: Universidad de Concepción
Comisión (Nombre y Firma)	
Prof. Bernardo Lagos A.	Prof. René Gómez P.
Resumen	
<p>El presente trabajo busca probar la aplicabilidad de algoritmos del Data Mining, en específico árboles de decisión y redes neuronales, en la estimación de variables del modelo geológico. Para esto, a partir de un conjunto de datos se construyeron tres modelos geológicos, dos con cada uno de los algoritmos a probar y otro con un método utilizado en minería que es el co-kriging indicador. Además, se evalúa el impacto de utilizar estos modelos en la estimación de recursos, utilizando separadamente cada modelo geológico predicho como un input en la estimación de ley mediante kriging universal, añadiendo otra estimación de ley con kriging ordinario (sin utilizar el modelo geológico) para ser comparada con los demás casos. Lo anterior es aplicado en primer lugar a un escenario simulado y posteriormente a un caso real.</p> <p>En el caso simulado, las redes neuronales logran mejor desempeño en términos de precisión comparado con el co-kriging indicador y los árboles de decisión, no obstante, los tres algoritmos son capaces de reproducir los principales patrones del modelo original, pudiendo concluir que las herramientas del Data Mining son capaces de igualar e incluso superar la capacidad predictiva de una herramienta geoestadística dentro del modelamiento geológico. En el caso de estudio real se sostiene la misma idea para el modelo geológico de texturas, pero no se pueden hacer conclusiones respecto al modelo de tipo de roca debido a diferencias significativas entre los modelos estimados y la baja precisión de validación alcanzada por los tres algoritmos, lo cual se arrastra por la complejidad del yacimiento y la poca información disponible.</p> <p>Respecto al uso del kriging universal que utiliza los modelos geológicos como inputs, en los dos casos de estudio se disminuye considerablemente la varianza de kriging, y bajo un contexto de baja correlación espacial de la ley se consigue una disminución en el error de estimación de esta. Por lo tanto, la estrategia de utilizar kriging universal junto con un modelo geológico estimado por los algoritmos estudiados resulta favorable para la estimación de recursos minerales.</p>	