



Universidad  
**Tecmilenio**®





# Matemáticas computacionales para inteligencia artificial

Optimización continua





Según Luaces (2018), se plantea que actualmente estamos en una tercera generación, donde los métodos de optimización utilizan metamodelos para representar las complejas agrupaciones de los modelos de inteligencia artificial.

Los retos que estos métodos enfrentan actualmente abarcan desde simulaciones rigurosas del rendimiento de un automóvil, hasta el cálculo de la resistencia y deformación en una simulación de mecánica de sólidos. Gracias a la computación, los resultados son cada vez más perfectos y seguramente en muy poco tiempo se verán nuevas y extraordinarias aplicaciones.

Toda la formulación matemática que utilizan los algoritmos de aprendizaje automático se expresa de alguna forma por un método de optimización numérica.



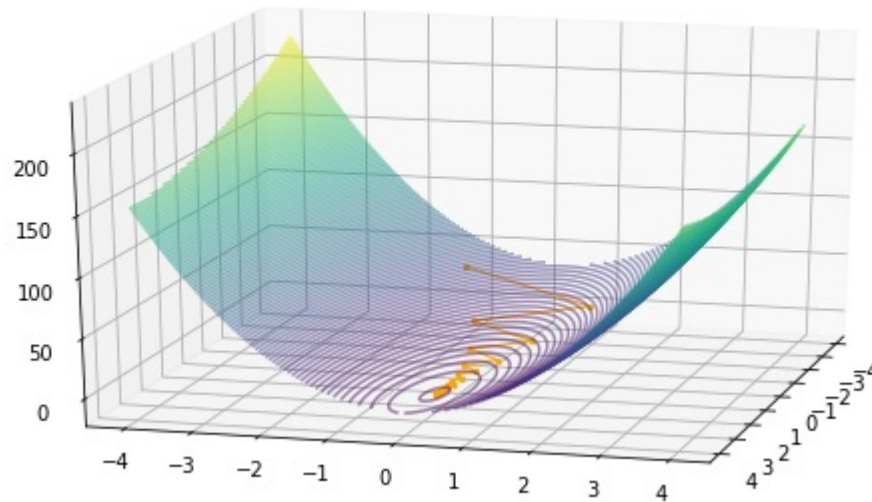


El **proceso de optimización** consiste en la selección de un conjunto de elementos posibles, la opción que más se aproxime a la solución, aplicando algún criterio antes de invertir demasiado tiempo y recursos de cálculo que nos entreguen una respuesta precisa.





La idea principal en la optimización es moverse hacia abajo, de manera opuesta al gradiente, con la esperanza de encontrar el punto más profundo lo antes posible, conocido como **mínimo**.





Imagina el siguiente ejemplo:

Eres el consultor de una empresa de tecnología que está considerando lanzar una campaña publicitaria para un nuevo producto. El departamento de finanzas te ha indicado que el presupuesto máximo para invertir en publicidad es de \$25,000. Se conoce que los dos medios más importantes que impactan a los clientes son las redes sociales y las campañas en *streaming* de video. El costo promedio de un anuncio en redes sociales es de \$25 y el del video es de \$200.





El número mínimo y máximo de anuncios en las redes sociales es que se pudieran realizar con el presupuesto es (segunda restricción):

$$s_{min} = 0, \quad s_{max} = \frac{\text{presupuesto}}{\text{costo del anuncio}} = \frac{25000}{25} = 1000$$

Mientras que la cantidad mínima y máxima de anuncios de video correspondería a (tercera restricción):

$$v_{min} = 0, \quad v_{max} = \frac{\text{presupuesto}}{\text{costo del anuncio}} = \frac{25000}{200} = 125$$

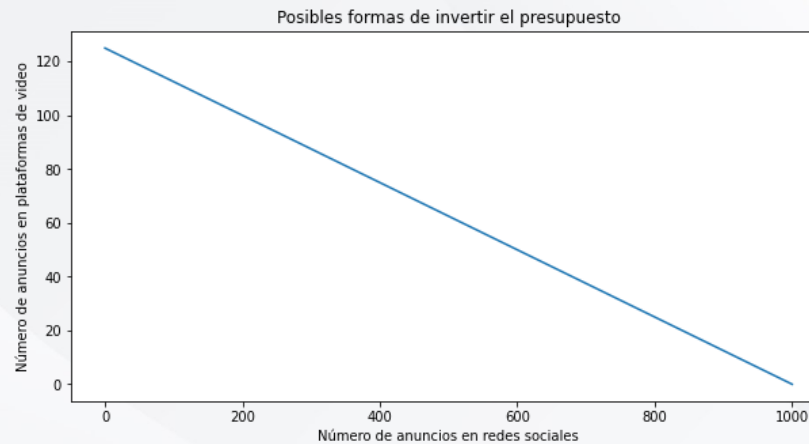






Si graficamos las restricciones del problema se pueden realizar varias observaciones:

- Cada combinación de anuncios sin importar la plataforma y que está dentro de nuestro presupuesto se encuentra bajo la línea.
- Cualquier combinación sobre la línea supera nuestro presupuesto
- Si queremos gastar completamente el presupuesto (algo que casi siempre es cierto) la combinación perfecta se encuentra sobre la línea.

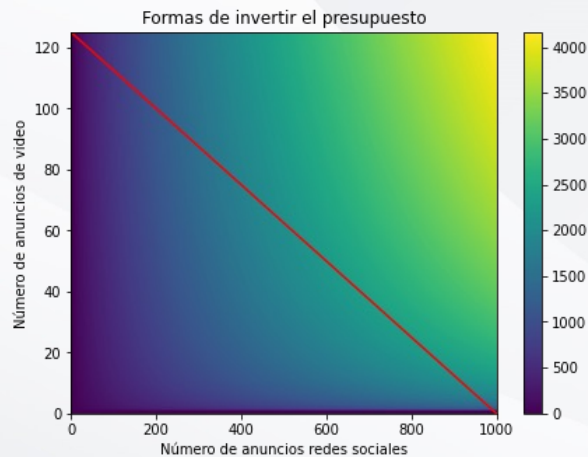




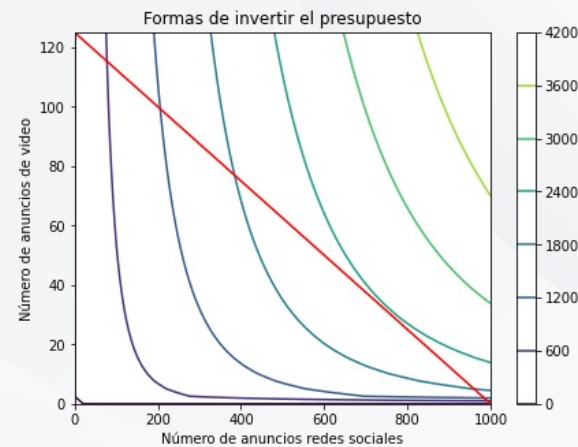


Es posible analizar este tipo de problemas empleando:

- a) Un mapa de color.
- b) Un mapa de contorno del gradiente.



(a)



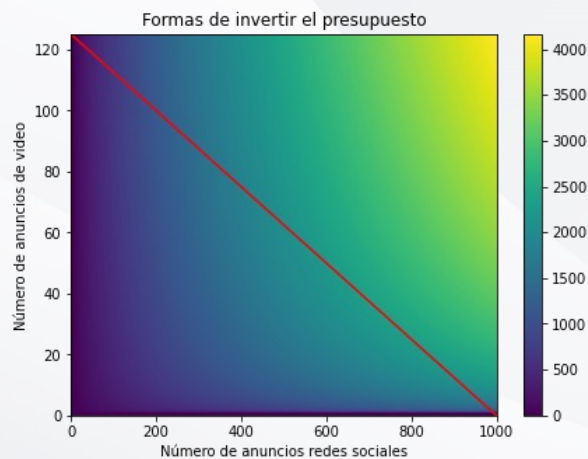
(b)

Con el apoyo de estas gráficas se puede ver que la solución buscada está sobre los 700 anuncios en las redes sociales y los 30 anuncios en las plataformas de *streaming*. La solución matemática es el punto específico donde la función de retorno es tangente a la línea roja y la herramienta perfecta para calcularlo es el multiplicador de Lagrange.

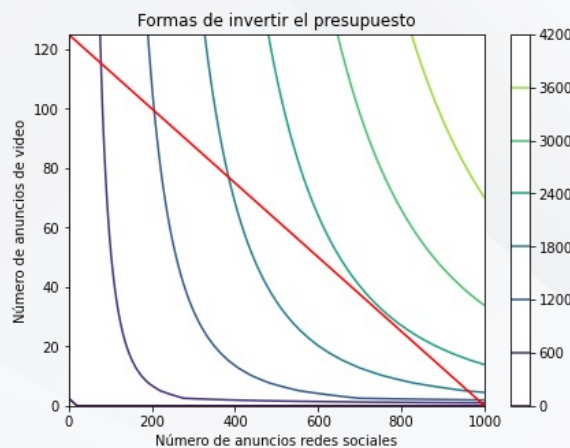




Una vez obtenidos los resultados, se puede comprobar que están cercanos al valor estimado visualmente, quedando de manera final que, para obtener el mayor rendimiento con el presupuesto asignado, la mejor estrategia es realizar 750 anuncios en las redes sociales y colocar 25 videos publicitarios en las plataformas de transmisión (*streaming*).



(a)



(b)





Piensa en qué otro tipo de problema podrías aplicar los algoritmos de optimización vistos en esta lección.





En este tema recordamos que los métodos de optimización numérica están en todas partes, ya que son la forma principal mediante la cual las computadoras y los dispositivos electrónicos realizan los cálculos matemáticos que necesitan para funcionar.

En el aprendizaje automático los métodos como el gradiente descendente se utilizan constantemente para evaluar la función de coste o la función de pérdida, la cual, a partir de su valor mínimo, nos indica que tan buena es la predicción que se está haciendo sobre el valor esperado. En este tipo de aplicaciones comprender la importancia de los gradientes, la diferenciación y la forma de encontrar los máximos o mínimos de una función es una habilidad imprescindible.







# Matemáticas computacionales para inteligencia artificial

Modelos matemáticos  
para aprendizaje  
automático



En este tema se dará un recorrido por los tres pilares de un sistema de inteligencia artificial: los datos, los modelos y el aprendizaje, indicando cuáles son los recursos matemáticos que más se utilizan en cada uno. Además, se presenta una breve introducción a los principales criterios que se usan para seleccionar un buen modelo de aprendizaje automático.





Una vez que los datos están representados como vectores, entonces se pueden construir las funciones de predicción (predictores). Según Peter (2020), un predictor es una función que dada una entrada particular (en este caso, un vector de características) produce una salida (puede ser un simple número). Dicha función puede representarse como:

$$f: \mathbb{R}^D \rightarrow \mathbb{R}$$



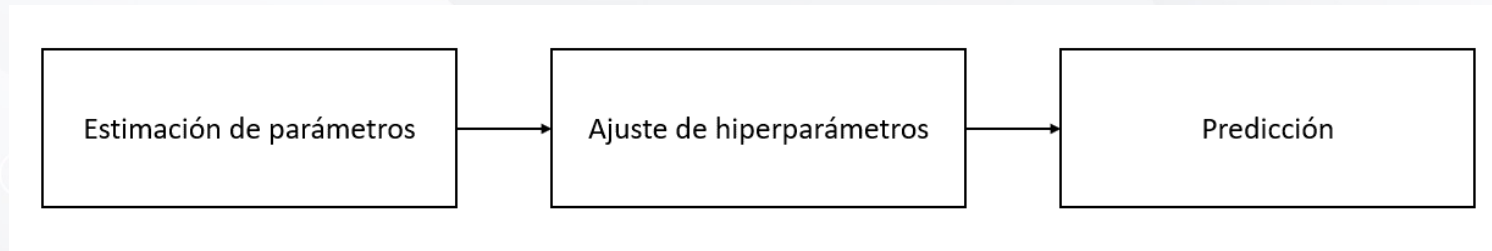
Donde el vector de entrada  $x$  tiene  $D$  dimensiones y la función  $f(x)$  luego de ser aplicada a este vector, retorna un valor real.

Un modelo de aprendizaje automático no es más que una función de predicción que recibe como entradas los datos en forma vectorial de la situación que se está analizando.





El aprendizaje es el proceso de obtener los parámetros adecuados, por tanto, el objetivo principal del aprendizaje es encontrar el mejor modelo y sus respectivos parámetros. Donde el adjetivo mejor, es directamente proporcional al comportamiento del predictor sobre datos desconocidos.



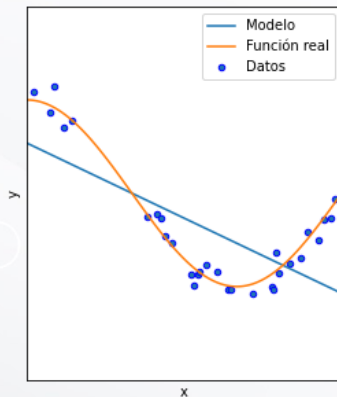
Conceptualmente es posible distinguir tres fases algorítmicas cuando se construye un sistema de aprendizaje automático: el entrenamiento o estimación de parámetros, el ajuste de hiperparámetros o selección del modelo y la predicción o inferencia.



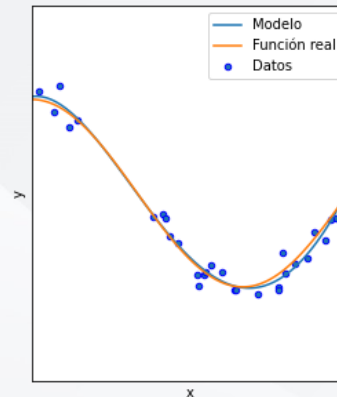




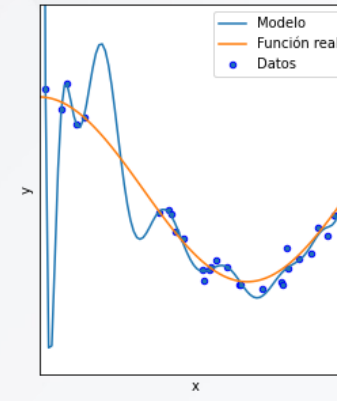
La fase de entrenamiento o estimación de parámetros es cuando ajustamos nuestro modelo predictivo en función de los datos de entrenamiento. En esta fase el objetivo es encontrar buenos predictores dados los datos que se conocen.



(a)



(b)



(c)

Cuando se estiman los parámetros del predictor se pueden dar tres situaciones muy comunes:

- a) Infraajuste.
- b) Un ajuste adecuado.
- c) Sobreajuste.





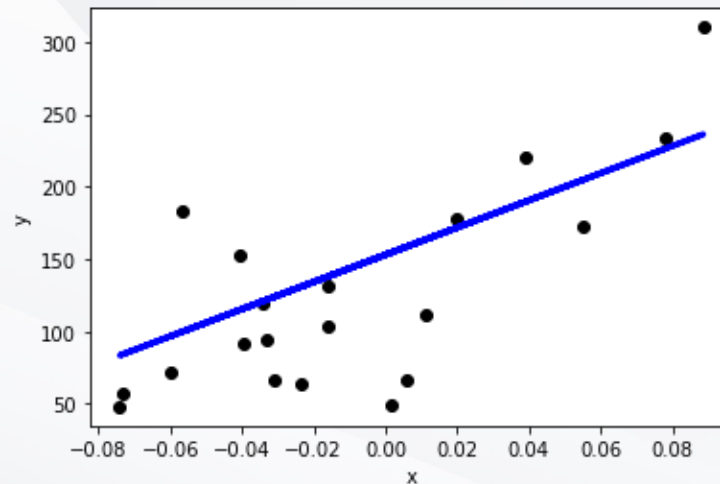
**Ajuste de los hiperparámetros o selección del modelo:** los parámetros son las variables que definen el predictor y que son aprendidos de forma autodidacta por el algoritmo de aprendizaje. Estos parámetros son modificados directamente por el propio algoritmo basado en los datos de entrenamiento y son calculados para lograr que el modelo sea óptimo en cierto sentido.

Los hiperparámetros no son aprendidos por el propio algoritmo a partir de los datos, por lo tanto, deben ser establecidos por el analista de datos antes de ejecutarlo. La elección del número de componentes es un ejemplo de hiperparámetro, y esta elección puede afectar significativamente el rendimiento del modelo.





**Predicción o inferencia:** la fase de predicción es cuando usamos un predictor entrenado con nuevos datos de prueba. En esta fase los parámetros y la elección del modelo ya están determinados, por lo tanto, el predictor está listo para recibir nuevos vectores de datos y generar una salida adecuada.

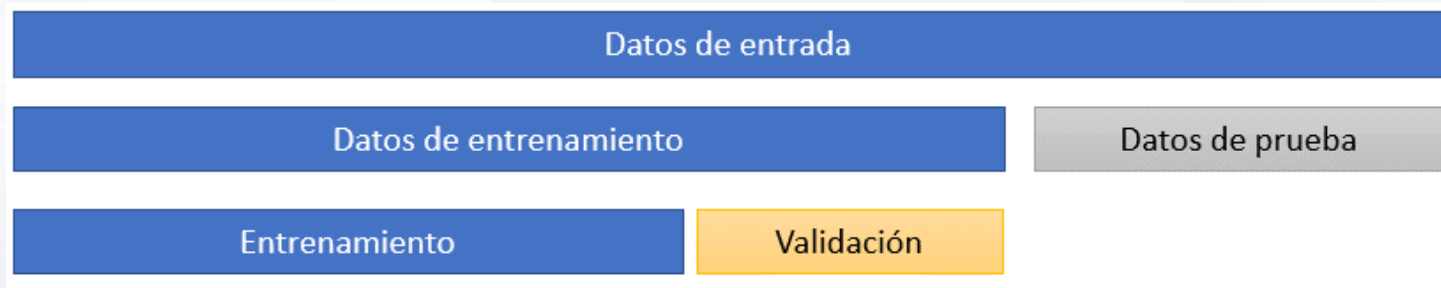


Considerando las dos aproximaciones del aprendizaje automático mencionadas anteriormente, correspondientes a si el predictor es una función o un modelo probabilístico, cuando el predictor representa esta segunda opción, la fase de predicción se suele llamar inferencia.





Una de las primeras aproximaciones que se utiliza para seleccionar un modelo es la validación cruzada, la cual proporciona una estimación del error de generalización dividiendo repetidamente el conjunto de datos inicial en conjuntos de entrenamiento y prueba.







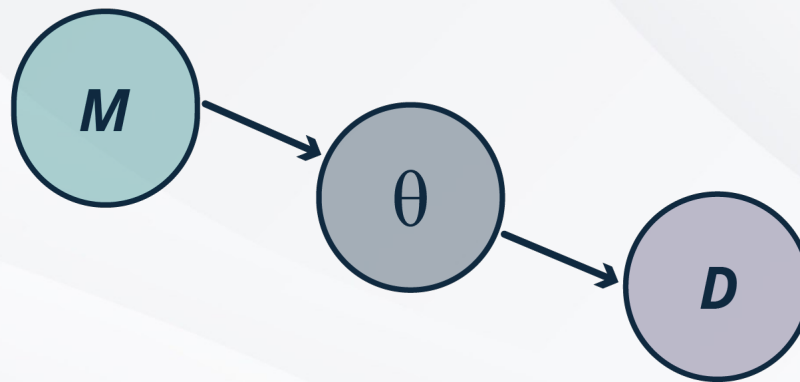
Si se considera un número finito de modelos  $M = \{M_1, \dots, M_k\}$ , donde cada  $M_k$  posee  $\theta_k$  parámetros, y se establece una probabilidad a priori  $p(M)$  de forma equivalente para todo el conjunto de modelos, se puede aplicar el método bayesiano para seleccionar el modelo adecuado analizando el correspondiente proceso generativo asociado.

En este caso:

$$M_k \sim p(M)$$

$$\theta_k \sim p(\theta | M_k)$$

$$D \sim p(D | \theta_k)$$





Investiga sobre los modelos matemáticos que más se utilizan para representar problemas resueltos con aprendizaje automático.





En este tema se abordaron los tres elementos principales de un sistema de aprendizaje automático: los datos, los modelos y el aprendizaje. La mayor interrogante que se hace un experto en inteligencia artificial consiste en determinar cuándo un modelo se puede considerar adecuado, pues el nivel de calidad de este no se mide con la evaluación sobre los datos conocidos, sino con su comportamiento ante valores totalmente nuevos.

Los datos que representan una problemática pueden tener diversas formas, pero aplicando algunas transformaciones se puede obtener su representación vectorial. A partir de esta estructura en forma de vectores se procede a encontrar un predictor, que no es más que un modelo matemático (una función o un modelo probabilístico) que describe adecuadamente dichos datos. El proceso de aprendizaje consiste en encontrar el mejor predictor y sus respectivos parámetros.

