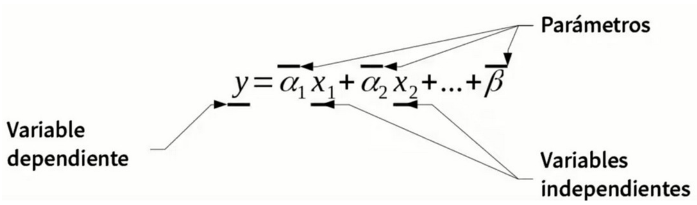
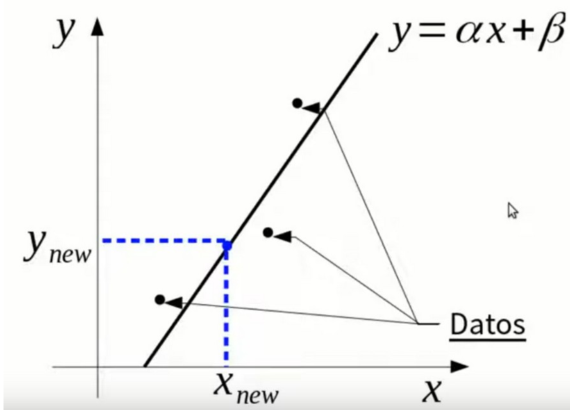
Funciones de error en Machine Learning

Desarrollo matemático para Regresión Lineal

Modelo matemático para la Regresión Lineal

La regresión lineal es un modelo matemático que relaciona de forma lineal una variable dependiente con una o más variables independientes.



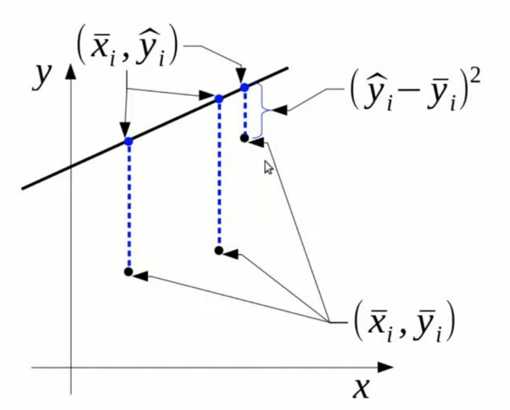


Los valores reales diferirán del modelo creado:

Modelo

$$y=\alpha x+\beta$$

Datos	Predicciones
$(\bar{x}_0, \bar{y}_0)$	$(\bar{x}_0, \hat{y}_0=\alpha\bar{x}_0+\beta)$
$(\bar{x}_1, \bar{y}_1)$	$(\bar{x}_1, \hat{y}_1=\alpha\bar{x}_1+\beta)$
$(\bar{x}_2, \bar{y}_2)$	$(\bar{x}_2, \hat{y}_2=\alpha\bar{x}_2+\beta)$



Definimos la función de error como la media de la diferencia (distancia en el eje y) entre el valor real y el predicho.

Usamos el valor al cuadrado para que errores de signo contrario no se compensen. Además eso contribuye a dar más peso en el cálculo del error a los valores más alejados.

$$L=\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-\bar{y}_i)^2$$

Dato predicho

Dato real

***Se busca que sea continua y diferenciable.**

Esta función se puede diferenciar respecto de los parámetros y encontrar donde se minimiza el error

$$L(\alpha,\beta)=\frac{1}{3}[(\underbrace{\alpha\bar{x}_0+\beta}_{\hat{y}_0}-\bar{y}_0)^2+(\underbrace{\alpha\bar{x}_1+\beta}_{\hat{y}_1}-\bar{y}_1)^2+(\underbrace{\alpha\bar{x}_2+\beta}_{\hat{y}_2}-\bar{y}_2)^2]$$

Podemos hallar el error **de forma analítica**:
Hallamos la derivada parcial de esta función respecto de cada una de las variables independientes

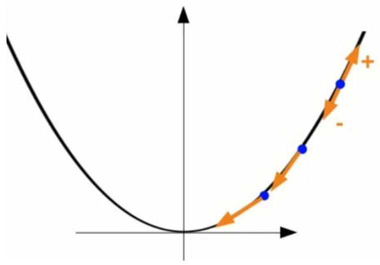
Igualamos esta derivada parcial a cero para averiguar el mínimo

$$\frac{\partial L(\alpha,\beta)}{\partial\alpha}=0 \qquad \frac{\partial L(\alpha,\beta)}{\partial\beta}=0$$

O a través del **Descenso del gradiente**:
Vamos paso a paso iterando el cálculo de la derivada. El gradiente.
Después de cada iteración hay que recalcular el mínimo de la función de error.
Según sea máxima la pendiente de bajada en una dirección avanzamos en esa dirección.
Cuando el gradiente es cero, no es posible más bajada (el error no puede descender más)

$$\alpha^{(k+1)}=\alpha^{(k)}-\eta\frac{\partial L}{\partial\alpha} \qquad \beta^{(k+1)}=\beta^{(k)}-\eta\frac{\partial L}{\partial\beta}$$

$$\underbrace{\alpha^{(k+1)}}_{\text{Valor actualizado}}=\underbrace{\alpha^{(k)}}_{\text{Valor anterior}}-\overbrace{\eta}^{\text{Factor de aprendizaje}}\underbrace{\frac{\partial L}{\partial\alpha}}_{\text{Gradiente}}$$



Cálculo analítico	Descenso del gradiente
• Es más rápida. • Da soluciones exactas. • Halla el mínimo absoluto. • Es posible que haya que resolver ecuaciones no lineales. • Se complica fuera del modelo de regresión lineal.	• La iteración lo vuelve un método más lento. • Puede quedarse atascado en mínimos locales. • Puede aplicarse a diferentes funciones de error y modelos.

[Modelos para entender una realidad caótica | DotCSV]()

[Regresión Lineal y Mínimos Cuadrados Ordinarios | DotCSV]()

[¿Qué es el Descenso del Gradiente? Algoritmo de Inteligencia Artificial | DotCSV]()