

Machine learning: stochastic gradient descent

Stochastic gradient descent (SGD) es un algoritmo de machine learning que entrena modelos minimizando errores en modelos predictivos. Es un algoritmo popular por entrenar modelos como máquinas de vectores de soporte, regresión logística y redes neuronales artificiales.

Cómo funciona SGD:

- Comienza con un conjunto inicial de parámetros.
- Selecciona aleatoriamente un punto o lote de datos del conjunto de entrenamiento.
- Calcula el gradiente de la función de pérdida con respecto a los datos seleccionados.
- Ajusta los parámetros en la dirección opuesta al gradiente.
- Repite hasta que el modelo converja o alcance un criterio de parada.

La función de pérdida total se calcula sobre todo el conjunto de datos:

$$\text{TrainLoss}(w) = \frac{1}{|D_{\text{train}}|} \sum_{(x,y) \in D_{\text{train}}} \text{Loss}(x,y,w)$$

En cada iteración, se requiere recorrer todos los ejemplos de entrenamiento para actualizar los pesos, lo cual es costoso en términos computacionales.

Algoritmo de Descenso de Gradiente Estocástico (SGD):

Se actualizan los pesos tras evaluar cada muestra en lugar de esperar todo el conjunto de datos.

Pseudocódigo:

Initialize $w = [0, \dots, 0]$

For $t = 1, \dots, T$:

For $(x, y) \in D_{\text{train}}$:

$w \leftarrow w - \eta \nabla \text{Loss}(x, y, w)$

Elección del Tamaño de Paso (η)

- Pequeño ($\eta \rightarrow 0$) -> Mas estable pero mas lento.
- Grande ($\eta \rightarrow 1$) -> Mas rapido pero menos estable

Diferencias entre GD y SGD:

Descenso de Gradiente (GD)	Descenso de Gradiente Estocástico (SGD)
Actualiza pesos con toda la data	Actualiza con cada ejemplo individual
Computacionalmente costoso	Más eficiente para grandes datasets
Suaviza la convergencia	Convergencia más ruidosa