**1. Cos’è il gradiente?**

Il **gradiente** è un vettore che rappresenta la direzione e l’intensità di variazione di una funzione. Viene calcolato come la derivata parziale della funzione rispetto a ciascuna delle sue variabili.

**Matematicamente:**

Immagine che contiene testo, Carattere, bianco, schermata

Descrizione generata automaticamente

**Intuizione geometrica:**

• Il gradiente punta nella direzione di **massima crescita** della funzione.

• La **norma** (lunghezza) del gradiente indica quanto velocemente la funzione cresce in quella direzione.

**Applicazione nel machine learning:**

Nel contesto dell’ottimizzazione, il gradiente di una funzione di perdita ( *J(w)* ) indica come modificare i parametri  *w*  (come i pesi del modello) per ridurre l’errore.

**2. Cos’è il Gradient Descent (Discesa del Gradiente)?**

La **discesa del gradiente** è un algoritmo iterativo per trovare il minimo di una funzione. Si utilizza per ottimizzare una funzione di perdita nel machine learning.

**Come funziona:**

1. Si calcola il gradiente della funzione di perdita rispetto ai parametri ( *w* ).

2. Si aggiornano i parametri  *w*  muovendosi nella direzione opposta al gradiente (perché il gradiente punta verso la crescita, ma vogliamo minimizzare):

Immagine che contiene testo, Carattere, ricevuta, schermata

Descrizione generata automaticamente

3. Si ripete fino a raggiungere la convergenza (quando i cambiamenti nei parametri diventano trascurabili o si raggiunge un numero massimo di iterazioni).

**Vantaggi:**

• Semplice ed efficace per molte funzioni.

• Adatto per problemi di ottimizzazione con un numero elevato di variabili.

**Svantaggi:**

• Può rimanere bloccato in minimi locali.

• La scelta di un tasso di apprendimento troppo grande o troppo piccolo può compromettere la convergenza.

**3. Cos’è il gradiente stocastico (SGD)?**

Il **gradiente stocastico** (Stochastic Gradient Descent) è una variante del Gradient Descent che aggiorna i parametri usando **un singolo campione alla volta**, invece dell’intero dataset.

**Differenza principale con Gradient Descent tradizionale:**

• **Gradient Descent standard:** Calcola il gradiente usando **tutti** i campioni del dataset (gradiente esatto).

• **Gradiente stocastico:** Calcola il gradiente usando **un solo campione** scelto casualmente (gradiente approssimato).

**Formula dell’SGD:**

Immagine che contiene testo, Carattere, bianco, tipografia

Descrizione generata automaticamente

Immagine che contiene testo, Carattere, bianco, tipografia

Descrizione generata automaticamente

**Vantaggi:**

• Molto più veloce per dataset di grandi dimensioni, poiché aggiorna i pesi dopo ogni campione.

• Introduce rumore nei passi di aggiornamento, il che può aiutare a sfuggire ai minimi locali.

**Svantaggi:**

• Più rumoroso rispetto al Gradient Descent standard, il che può causare oscillazioni attorno al minimo globale.

• Richiede un numero maggiore di iterazioni per convergere.

Immagine che contiene testo, schermata, Carattere, ricevuta

Descrizione generata automaticamente

Immagine che contiene testo, Carattere, algebra

Descrizione generata automaticamente