Derivate Parțiale și Gradientul

Gradientul funcției de cost este un vector de derivate parțiale care indică direcția de creștere maximă a funcției. Dacă $J(\theta)$ este o funcție de mai multe variabile (parametri), gradientul este dat de:

$$\nabla J(\theta) = \left[\frac{\partial J(\theta)}{\partial \theta_1}, \frac{\partial J(\theta)}{\partial \theta_2}, \ldots, \frac{\partial J(\theta)}{\partial \theta_n}\right]$$

Algoritmul Gradient Descent

Algoritmul Gradient Descent începe cu o estimare inițială a parametrilor θ și iterativ își actualizează valorile pentru a reduce funcția de cost. Actualizarea parametrilor se face conform următoarei reguli:

$$\theta := \theta - \alpha \nabla J(\theta)$$

unde:

- α este rata de învățare, un parametru scalar care controlează pasul actualizării.
- $\nabla J(\theta)$ este gradientul funcției de cost.

Pași Detaliați

1. **Inițializarea Parametrilor:**

    - Alege o valoare inițială pentru θ.

2. **Calcularea Gradientului:**

    - Calculează gradientul funcției de cost în punctul curent θ:
      $\nabla J(\theta) = \left[\frac{\partial J(\theta)}{\partial \theta_1}, \frac{\partial J(\theta)}{\partial \theta_2}, \ldots, \frac{\partial J(\theta)}{\partial \theta_n} \right]$

3. **Actualizarea Parametrilor:**

    - Actualizează parametrii utilizând regula:
      $\theta := \theta - \alpha \nabla J(\theta)$

4. **Convergență:**

    - Repetă pașii 2 și 3 până când gradientul este suficient de mic sau până când schimbările în θ sunt neglijabile.

Exemplu de Regresie Liniară

Pentru a face lucrurile mai concrete, să luăm un exemplu simplu de regresie liniară cu o funcție de cost de formă:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

unde:

- $h_\theta(x) = \theta_0 + \theta_1 x$ este funcția ipotezei (predicția modelului).
- m este numărul de exemple de antrenament.
- $x^{(i)}$ și $y^{(i)}$ sunt valorile caracteristicii și respectiv, eticheta pentru exemplul i.

Gradientul pentru fiecare parametru θ_j este:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Algoritmul Gradient Descent pentru acest caz devine:

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$