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:

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

Explicarea Detaliată a Matematicii din Spatele Gradient Descent

3. **Calculul Gradientului:**
   - **"Gradientul funcției de cost $\nabla J(\theta)$ este un vector de derivate parțiale în raport cu fiecare parametru θ_j. Acest gradient indică direcția de creștere maximă a funcției de cost."**
   - **"Matematic, gradientul pentru $J(\theta)$ este dat de:** $\nabla J(\theta) = \left[\frac{\partial J(\theta)}{\partial \theta_1}, \frac{\partial J(\theta)}{\partial \theta_2}, \ldots, \frac{\partial J(\theta)}{\partial \theta_n} \right]$**"**

4. **Actualizarea Parametrilor:**
   - **"Parametrii θ sunt actualizați iterativ folosind regula:** $\theta := \theta - \alpha \nabla J(\theta)$
     - **unde α este rata de învățare care controlează mărimea pașilor de actualizare."**

5. **Regresia Liniară:**

   - **"În contextul regresiei liniare, funcția de cost $J(\theta)$ este adesea definită ca eroarea pătratică medie (MSE). Gradientul pentru θ în acest caz este calculat ca: $\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$**
   - **"Aplicând Gradient Descent, parametrii sunt actualizați conform formulei: $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$**

Exemplu concret

Să luăm un exemplu simplu cu un model liniar $h_\theta(x) = \theta_0 + \theta_1 x$ și să vedem cum se aplică Gradient Descent:

1. **Funcția de cost**:
   $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$

2. **Gradientul**:
   $\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})$
   $\frac{\partial J(\theta)}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)}) x^{(i)}$

3. **Actualizarea parametrilor**:
   $\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})$
   $\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)}) x^{(i)}$

Implementare pas cu pas

1. **Inițializează parametrii**: θ_0 și θ_1 la valori inițiale, de exemplu 0.

2. **Setează rata de învățare**: α (de exemplu, 0.01).

3. **Repetă până la convergență**:

   - **Calculează predicțiile**: $h_\theta(x) = \theta_0 + \theta_1 x$
   - **Calculează gradientul**:
     $\text{grad}_{\theta_0} = \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})$
     $\text{grad}_{\theta_1} = \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)}) x^{(i)}$
   - **Actualizează parametrii**:
     $\theta_0 := \theta_0 - \alpha \cdot \text{grad}_{\theta_0}$
     $\theta_1 := \theta_1 - \alpha \cdot \text{grad}_{\theta_1}$