Herkansing wiskunde

Gradient descent met 3 voorbeelden en 2 gewichten

1e Iteratie

Stap 1: Voorspelde waarden en kosten

Initiële gewichten: $w_1 = 1$ en $w_2 = 0.5$

Voorbeeld 1:

- $x_1 = 1$, $x_2 = 2$, $y = 3$
- Voorspelling: $\hat{y}_1 = w_1 \cdot x_1 + w_2 \cdot x_2 = 1 \cdot 1 + 0.5 \cdot 2 = 2$
- Kosten: $(2 - 3)^2 = 1$

Voorbeeld 2:

- $x_1 = 2$, $x_2 = 3$, $y = 5$
- Voorspelling: $\hat{y}_2 = w_1 \cdot x_1 + w_2 \cdot x_2 = 1 \cdot 2 + 0.5 \cdot 3 = 3$
- Kosten: $(3 - 5)^2 = 4$

Voorbeeld 3:

- $x_1 = 3$, $x_2 = 4$, $y = 7$
- Voorspelling: $\hat{y}_3 = w_1 \cdot x_1 + w_2 \cdot x_2 = 1 \cdot 3 + 0.5 \cdot 4 = 4.5$
- Kosten: $(4.5 - 7)^2 = 6.25$

Gemiddelde kosten (MSE):
$\text{MSE} = \frac{1+4+6.25}{3} = 3.0833$

Stap 2: Berekening van de gradient

Voor w_1 berekenen we de afgeleide:
$\frac{\partial \text{MSE}}{\partial w_1} = \sum_{i=1}^{3} (\hat{y}_i - y_i) x_{1i}$

Hierbij:

- Voor $i = 1$: $\hat{y}_1 = 2$, $y_1 = 3$, $x_{11} = 1$
- Voor $i = 2$: $\hat{y}_2 = 3$, $y_2 = 5$, $x_{12} = 2$
- Voor $i = 3$: $\hat{y}_3 = 4.5$, $y_3 = 7$, $x_{13} = 3$

Dus,
$\frac{\partial \text{MSE}}{\partial w_1} = (2 - 3) \cdot 1 + (3 - 5) \cdot 2 + (4.5 - 7) \cdot 3$
$= -1 \cdot 1 + (-2) \cdot 2 + (-2.5) \cdot 3$
$= -1 - 4 - 7.5$
$= -12.5$

Voor w_2 berekenen we de afgeleide:
$\frac{\partial \text{MSE}}{\partial w_2} = \sum_{i=1}^{3} (\hat{y}_i - y_i) x_{2i}$

Hierbij:

- Voor $i = 1$: $\hat{y}_1 = 2$, $y_1 = 3$, $x_{21} = 2$
- Voor $i = 2$: $\hat{y}_2 = 3$, $y_2 = 5$, $x_{22} = 3$
- Voor $i = 3$: $\hat{y}_3 = 4.5$, $y_3 = 7$, $x_{23} = 4$

Dus,
$\frac{\partial \text{MSE}}{\partial w_2} = (2 - 3) \cdot 2 + (3 - 5) \cdot 3 + (4.5 - 7) \cdot 4$
$= -1 \cdot 2 + (-2) \cdot 3 + (-2.5) \cdot 4$
$= -2 - 6 - 10$
$= -18$

Stap 3: Update de gewichten

Gebruik een leersnelheid $\alpha = 0.01$ om de gewichten aan te passen:

Voor w_1:
$w_1 \text{ nieuw} = w_1 - \alpha \frac{\partial \text{MSE}}{\partial w_1} = 1 - 0.01 \cdot (-12.5) = 1.125$

Voor w_2:
$w_2 \text{ nieuw} = w_2 - \alpha \frac{\partial \text{MSE}}{\partial w_2} = 0.5 - 0.01 \cdot (-18) = 0.68$

Na de update zijn de nieuwe gewichten $w_1 = 1.125$ en $w_2 = 0.68$.

2e Iteratie

Nu met de nieuwe gewichten $w_1 = 1.125$ en $w_2 = 0.68$.

Voorbeeld 1:

- $x_1 = 1$, $x_2 = 2$, $y = 3$
- Voorspelling: $\hat{y}_1 = w_1 \cdot x_1 + w_2 \cdot x_2 = 1.125 \cdot 1 + 0.68 \cdot 2 = 2.485$
- Kosten: $(2.485 - 3)^2 = 0.290225$

Voorbeeld 2:

- $x_1 = 2$, $x_2 = 3$, $y = 5$
- Voorspelling: $\hat{y}_2 = w_1 \cdot x_1 + w_2 \cdot x_2 = 1.125 \cdot 2 + 0.68 \cdot 3 = 4.19$
- Kosten: $(4.19 - 5)^2 = 0.6889$

Voorbeeld 3:

- $x_1 = 3$, $x_2 = 4$, $y = 7$
- Voorspelling: $\hat{y}_3 = w_1 \cdot x_1 + w_2 \cdot x_2 = 1.125 \cdot 3 + 0.68 \cdot 4 = 5.795$
- Kosten: $(5.795 - 7)^2 = 2.612025$

Gemiddelde kosten (MSE):
$\text{MSE} = \frac{0.290225+0.6889+2.612025}{3} = 1.197717$

Berekening van de gradient voor de tweede iteratie

Voor w_1 berekenen we de afgeleide:
$\frac{\partial \text{MSE}}{\partial w_1} = \sum_{i=1}^{3} (\hat{y}_i - y_i) x_{1i}$

Hierbij:

- Voor $i = 1$: $\hat{y}_1 = 2.485$, $y_1 = 3$, $x_{11} = 1$
- Voor $i = 2$: $\hat{y}_2 = 4.19$, $y_2 = 5$, $x_{12} = 2$
- Voor $i = 3$: $\hat{y}_3 = 5.795$, $y_3 = 7$, $x_{13} = 3$

Dus,
$\frac{\partial \text{MSE}}{\partial w_1} = (2.485 - 3) \cdot 1 + (4.19 - 5) \cdot 2 + (5.795 - 7) \cdot 3$
$= (-0.515) \cdot 1 + (-0.81) \cdot 2 + (-1.205) \cdot 3$
$= -0.515 - 1.62 - 3.615$
$= -5.75$

Voor w_2 berekenen we de afgeleide:
$\frac{\partial \text{MSE}}{\partial w_2} = \sum_{i=1}^{3} (\hat{y}_i - y_i) x_{2i}$

Hierbij:

- Voor $i = 1$: $\hat{y}_1 = 2.485$, $y_1 = 3$, $x_{21} = 2$
- Voor $i = 2$: $\hat{y}_2 = 4.19$, $y_2 = 5$, $x_{22} = 3$
- Voor $i = 3$: $\hat{y}_3 = 5.795$, $y_3 = 7$, $x_{23} = 4$

Dus,
$\frac{\partial \text{MSE}}{\partial w_2} = (2.485 - 3) \cdot 2 + (4.19 - 5) \cdot 3 + (5.795 - 7) \cdot 4$
$= (-0.515) \cdot 2 + (-0.81) \cdot 3 + (-1.205) \cdot 4$
$= -1.03 - 2.43 - 4.82$
$= -8.28$

Update de gewichten voor de tweede iteratie

Gebruik opnieuw een leersnelheid $\alpha = 0.01$:

Voor w_1:
$w_1 \text{ nieuw} = w_1 - \alpha \frac{\partial \text{MSE}}{\partial w_1} = 1.125 - 0.01 \cdot (-5.75) = 1.183$

Voor w_2:
$w_2 \text{ nieuw} = w_2 - \alpha \frac{\partial \text{MSE}}{\partial w_2} = 0.68 - 0.01 \cdot (-8.28) = 0.763$

Na de tweede iteratie zijn de nieuwe gewichten $w_1 = 1.183$ en $w_2 = 0.763$.

Uitleg

Uitleg variabelen

x1 en x2: In het voorbeeld van een huis verkopen stellen x1 en x2 de kenmerken of eigenschappen van het huis voor die worden gebruikt om de verkoopprijs te voorspellen. Specifiek kan x1 de grootte van het huis in vierkante meters voorstellen, terwijl x2 het aantal slaapkamers vertegenwoordigt. Deze kenmerken zijn de inputvariabelen die het model gebruikt om voorspellingen te doen.

y: y vertegenwoordigt de uitvoer van het model, oftewel de voorspelde verkoopprijs van het huis. Dit is wat we proberen te voorspellen met behulp van ons model.

w1 en w2: w1 en w2 zijn de gewichten van het model die de impact van de kenmerken (x1 en x2) op de voorspelde verkoopprijs bepalen. In het voorbeeld van een huis kan w1 bijvoorbeeld de impact van de grootte van het huis (x1) op de verkoopprijs vertegenwoordigen, terwijl w2 de impact van het aantal slaapkamers (x2) op de verkoopprijs vertegenwoordigt. Deze gewichten worden aangepast tijdens het trainingsproces om het model in staat te stellen nauwkeurige voorspellingen te doen.

Uitleg stappen

Stap 1: Bereken de voorspelde waarden en de kosten

We beginnen met het voorspellen van de uitvoerwaarden ($\hat{y}$) voor elk trainingsvoorbeeld met behulp van de huidige gewichten. Vervolgens berekenen we de kosten voor elk voorbeeld door het verschil te nemen tussen de voorspelde waarde en de werkelijke waarde, het kwadraat van dit verschil te nemen, en vervolgens het gemiddelde van deze kwadratische verschillen te berekenen. Dit geeft ons de Mean Squared Error (MSE), een maatstaf voor hoe goed onze huidige gewichten presteren.

Stap 2: Bereken de gradient van de kostenfunctie

De gradient van de kostenfunctie geeft aan hoe snel de kosten veranderen met betrekking tot veranderingen in de gewichten. We berekenen de partiële afgeleiden van de MSE ten opzichte van elk gewicht. Deze afgeleiden vertellen ons in welke richting en hoeveel we elk gewicht moeten aanpassen om de kosten te verlagen.

Stap 3: Werk een update stap uit

We bepalen hoeveel we de gewichten moeten aanpassen door de gradiënten te vermenigvuldigen met een leersnelheid (α), die aangeeft hoe groot de stappen moeten zijn. Vervolgens passen we de gewichten aan door de leersnelheid vermenigvuldigd met de bijbehorende gradient van elke parameter af te trekken van de huidige waarden van de gewichten.

Stap 4: Herhaal de stappen

Door de stappen 1 tot 3 te herhalen, passen we de gewichten iteratief aan om de kosten te verlagen. We blijven dit herhalen totdat we een punt bereiken waarop de kosten niet meer significant dalen of een vooraf ingesteld aantal iteraties hebben bereikt.

Dit iteratieve proces van het aanpassen van de gewichten op basis van de gradients van de kostenfunctie is wat gradient descent in essentie doet. Het uiteindelijke doel is om de gewichten te vinden die leiden tot minimale kosten, waardoor ons model de beste voorspellingen kan doen voor nieuwe gegevens.