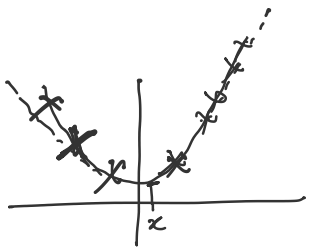
Gradient descent

Monday, 9 February 2026 12:19 AM

Gradient Descent



If slope = -ve → ↑ b
slope = +ve → ↓ b

$b_{new} = b_{old} - slope$

$= b_{old} - \eta \cdot slope$

↓
learning rate.

for eg. $\eta = 0.0001$

$\Rightarrow b_{old} - 0.0001 \, m$

$\Rightarrow 100 - 1 \times 90$

$\Rightarrow 100 - 180$

-80

when to stop
1) steps fixed - epocs
2) $b_{new} - b_{old} = 0.0001$

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Gradient Descent

- Batch
  - ↓
  - all data
  - - full data
- Stochastic
  - ↓
  - Each row
  - - 1 row
- Mini batch GD
  - ↓
  - Define a batch
  - - (eg - bs - 30)

rows

$$m_n = m_0 - \eta \times (\text{slope})_{m=0}$$

$$m_n = m_0 \quad \eta \times \left(\frac{\delta L}{\delta m}\right)$$

$$b_n = b_0 - \eta \times (\text{slope})_{b=0}$$

$$b_n = b_0 - \eta \times \left(\frac{\delta L}{\delta b}\right)_{b=0}$$

$$\frac{\delta L}{\delta \beta_m} = \underbrace{-\frac{2}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)}\, x_{im}$$

we have to multiply the difference to the particular feature x value.