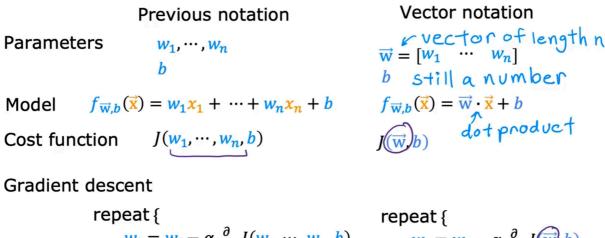
C. Gradient Descent

Friday, 5 September 2025

12:45 PM

A. Introduction



 $\begin{aligned} \text{peat} \, \{ & & \text{repeat} \, \{ \\ w_j &= w_j - \alpha \frac{\partial}{\partial w_j} J(\underbrace{w_1, \cdots, w_n, b}) & & w_j &= w_j - \alpha \frac{\partial}{\partial w_j} J(\underbrace{w_j, \cdots, w_n, b}) \\ b &= b - \alpha \frac{\partial}{\partial b} J(\underbrace{w_1, \cdots, w_n, b}) & & b &= b - \alpha \frac{\partial}{\partial b} J(\underbrace{w_j, \cdots, w_n, b}) \\ \} & & \} \end{aligned}$

Now,

Gradient descent

One feature
$$n \text{ features } (n \ge 2)$$
 repeat $\{ w = w - \alpha \left(\frac{1}{m} \sum_{i=1}^{m} (f_{w,b}(\mathbf{x}^{(i)}) - \mathbf{y}^{(i)}) \mathbf{x}^{(i)} \right) - \mathbf{y}^{(i)} \mathbf{x}^{(i)} \right) = 1$ repeat $\{ w_1 = w_1 - \alpha \left(\frac{1}{m} \sum_{i=1}^{m} (f_{\overline{w},b}(\overline{\mathbf{x}}^{(i)}) - \mathbf{y}^{(i)}) \mathbf{x}^{(i)} \right)$ \vdots $w_n = w_n - \alpha \left(\frac{1}{m} \sum_{i=1}^{m} (f_{\overline{w},b}(\overline{\mathbf{x}}^{(i)}) - \mathbf{y}^{(i)}) \mathbf{x}^{(i)} \right)$ simultaneously update w, b $\}$
$$b = b - \alpha \left(\frac{1}{m} \sum_{i=1}^{m} (f_{\overline{w},b}(\overline{\mathbf{x}}^{(i)}) - \mathbf{y}^{(i)}) \mathbf{x}^{(i)} \right)$$
 simultaneously update w, b $\}$

B. Alternative to Gradient Descent

An alternative to gradient descent

- Only for linear regression
- Solve for w, b without iterations

Disadvantages

- Doesn't generalize to other learning algorithms.
- Slow when number of features is large (> 10,000)

wilat you lieed to know

- Normal equation method may be used in machine learning libraries that implement linear regression.
- Gradient descent is the recommended method for finding parameters w,b