deeplearning.ai

Basics of Neural Network Programming

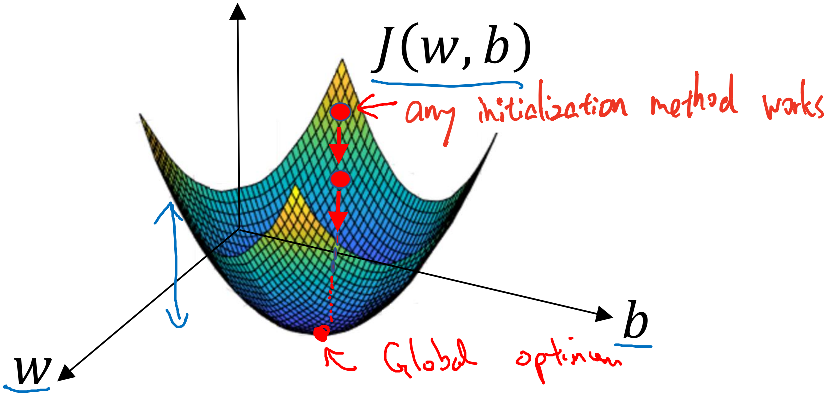
Gradient Descent

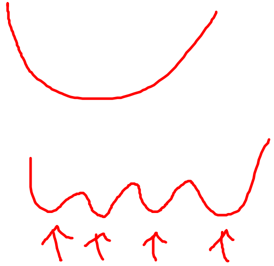
Gradient Descent

Recap: $\hat{y} = \sigma(w^T x + b)$, $\sigma(z) = \frac{1}{1+e^{-z}} \leftarrow$

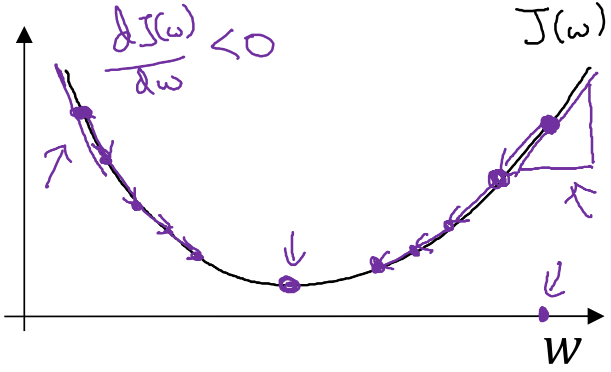
$$\underline{J(w,b)} = \underline{\frac{1}{m}\sum_{i=1}^{m}\mathcal{L}(\boxed{\hat{y}^{(i)}}, \boxed{y^{(i)}})} = -\frac{1}{m}\sum_{i=1}^{m}\Big(y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})\Big)$$

Want to find w, b that minimize $J(w,b)$





Gradient Descent



learning rate

Repeat {

$$\left(w := w - \alpha \underbrace{\boxed{\frac{dJ(w)}{dw}}}_{\text{"dw"}} \right.$$

}

$$w := w - \alpha \underline{dw}$$

$$\underbrace{\frac{dJ(w)}{dw}}= ?$$

$$J(w,b)$$

$$\left\{ \begin{array}{l} w := w - \alpha \boxed{\frac{dJ(w,b)}{dw}} \quad \boxed{\frac{\partial J(w,b)}{\partial w}} \rightarrow dw \\ b := b - \alpha \boxed{\frac{dJ(w,b)}{db}} \quad \boxed{\frac{\partial J(w,b)}{\partial b}} \rightarrow db \end{array} \right.$$

∂ "partial derivative"

d J

Andrew Ng