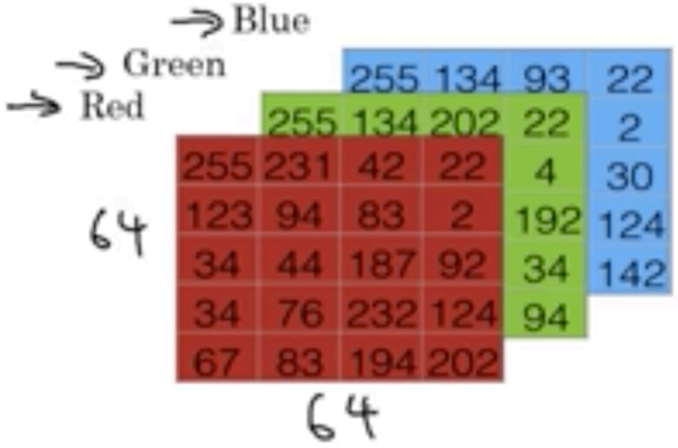
神经网络与深度学习

Binary Classification (二分类)

1 (cat) vs 0 (non cat)

y



$$X = \begin{bmatrix} 255 \\ 231 \\ 42 \\ 22 \\ \cdots \end{bmatrix}$$

N=64*64*3=12288

X $\xrightarrow{f(x)}$ y(0 or 1)

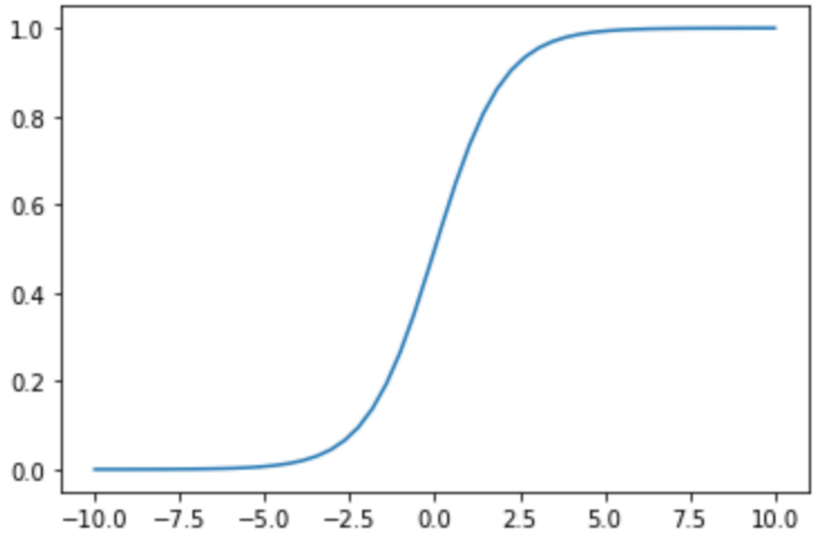
构建映射

$$W = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \cdots \end{bmatrix}$$

$$\sigma(W^T X + b) = y$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

激活函数：将z缩放至(0,1)



目的：找到一个合适的向量W，使得输入对应的X，能够输出正确的Y (1 or 0)

构建损失函数

约定　样本真实值：y

模型预测值：$\hat{y}$

$$\hat{y} = \sigma(w^T x + b)$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

损失函数(Loss Function) : L(y, $\hat{y}$)

用于衡量y和 $\hat{y}$ 的差别

$$|y - \hat{y}|$$

$$(y - \hat{y})^2$$

$$-(y log\hat{y} + (1-y) log(1-\hat{y}))$$ 二分类 ★

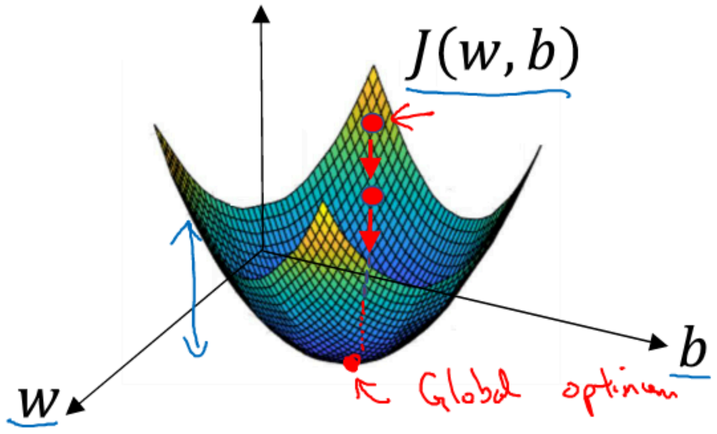
if y==1　　if y==0

最小化损失函数

m个样本下的损失函数

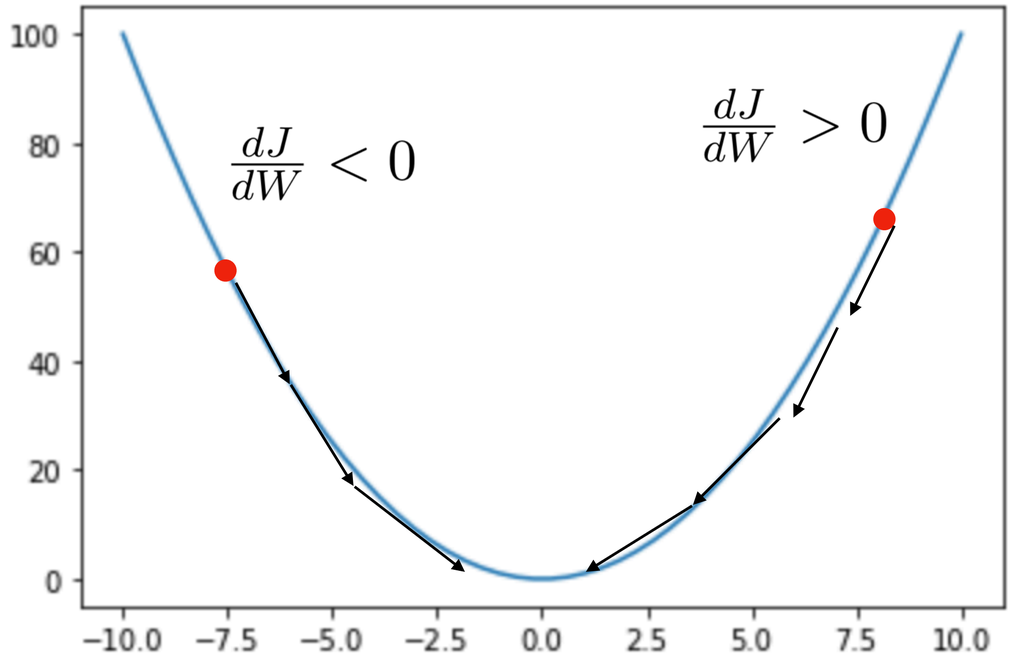
$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})$$

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



Andrew Ng

梯度下降法



学习率(learning rate)

迭代公式 $W^* = W - \alpha \frac{dJ}{dW}$

$$W = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \dots \end{bmatrix}$$

对于每一个W1,W2...
都需要用迭代公式优化