网络结构 激励函数（每个神经元最重要的就是激励函数） 损失函数（结果评价） 梯度下降

学习目标：

理解伸进网络结构

理解逻辑回归是一种简化的网络结构

理解神经网络的激励函数

理解神经网络的损失函数

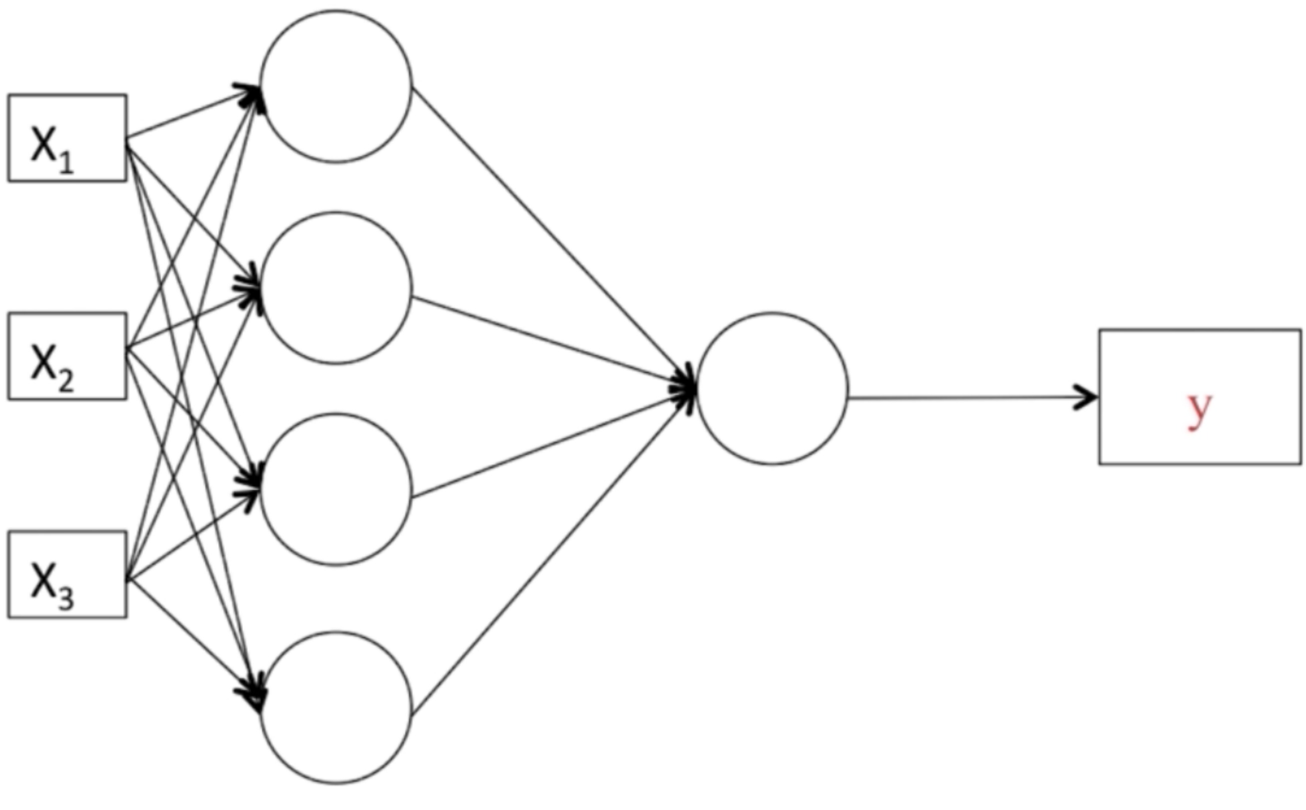
理解逻辑回归中的梯度下降

理解神经网络的网络向量化

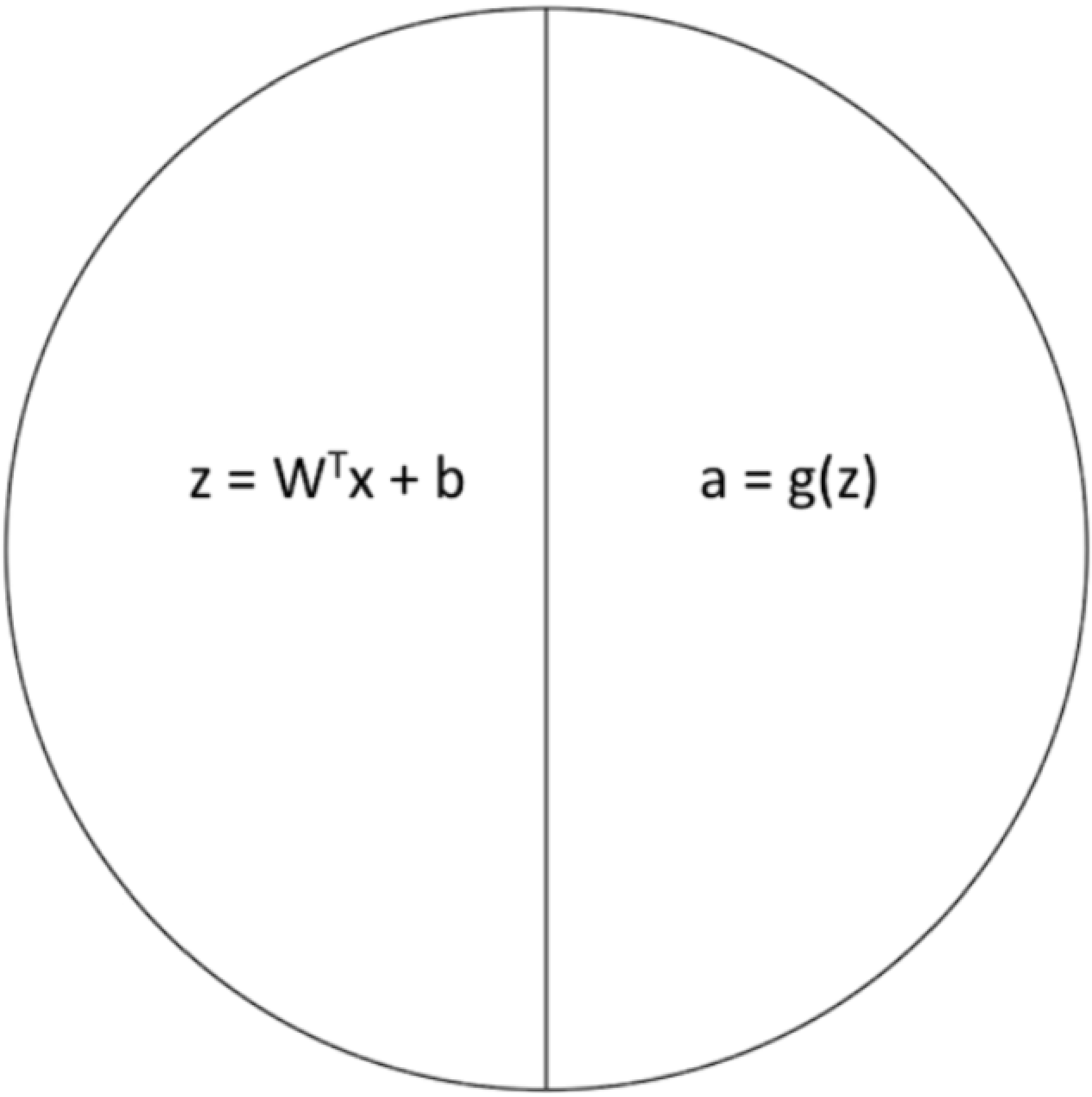
理解神经网络的梯度下降

通过观察神经网络的学习过程形成较为直观的理解

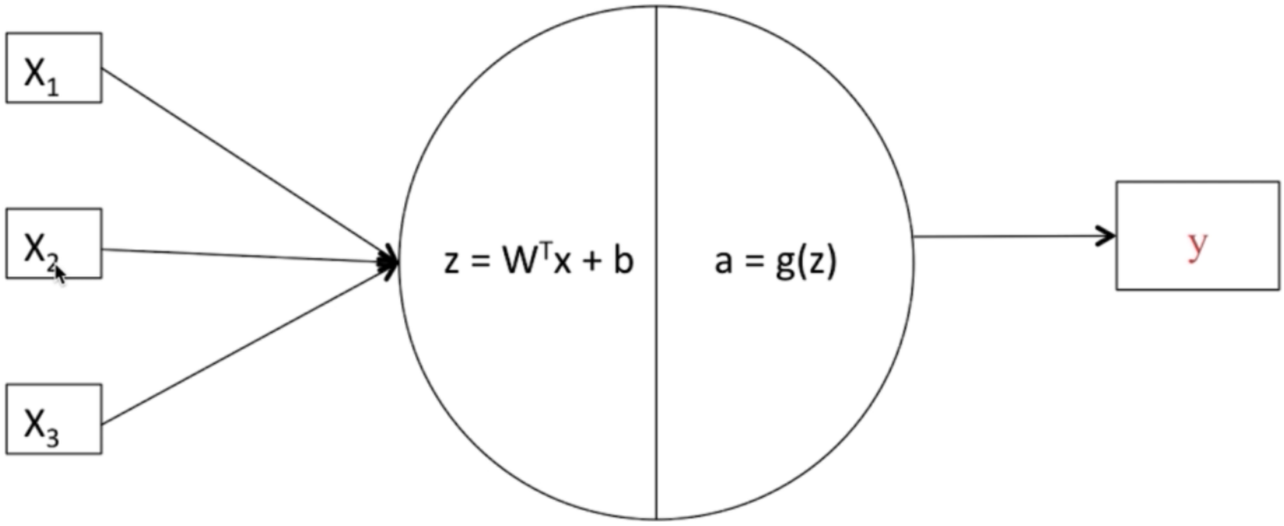
网络结构



网络结构



逻辑回归

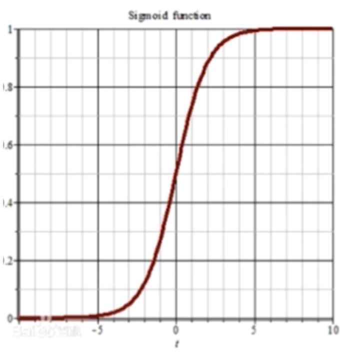


激励函数

激励函数的作用是提供规模化的非线性化能力

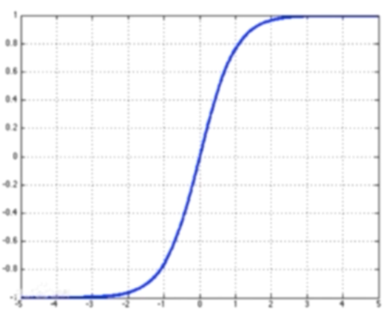
激励函数

Sigmoid



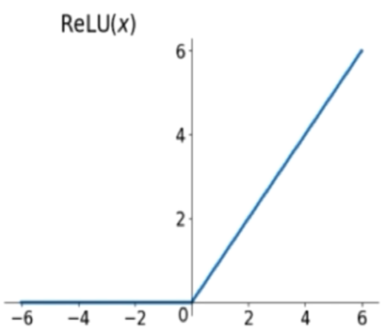
$$f(x)=\frac{1}{1+e^{-x}}$$

tanh



$$f(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$

ReLU



$$f(x)=\max(0,x)$$

损失函数

单次训练损失

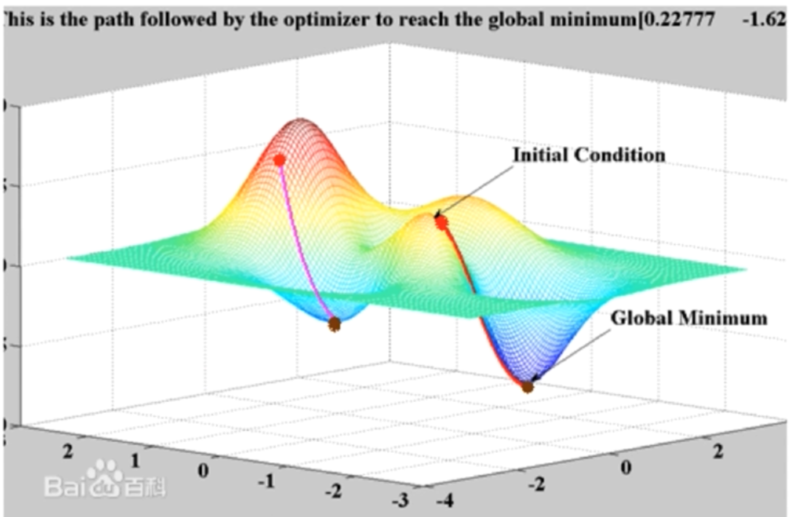
$$L\left(\hat{y}^{(i)}, y^{(i)}\right) = -\left(y^{(i)} \log\left(\hat{y}^{(i)}\right) + \left(1 - y^{(i)}\right)\log(1 - \hat{y}^{(i)}\right)$$

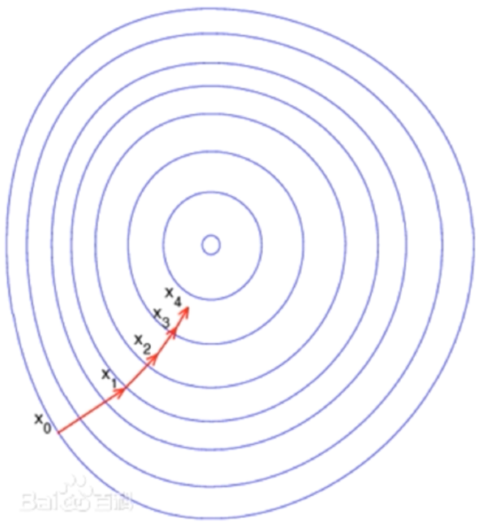
损失函数

全部训练损失

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

逻辑回归梯度下降





逻辑回归梯度下降

$$w := w - \alpha \frac{dJ(w,b)}{dw}$$

$$b := b - \frac{dJ(w,b)}{db}$$

网络向量化

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} -> \begin{bmatrix} a_1^1 \\ a_2^1 \\ a_3^1 \\ a_4^1 \end{bmatrix} -> [a^2] -> y$$

网络向量化

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} -(W_{3*4}^1, b_{4*1}^1) > \begin{bmatrix} a_1^1 \\ a_2^1 \\ a_3^1 \\ a_4^1 \end{bmatrix} -(W_{4*1}^2, b_{1*1}^2) > [a^2] -> y$$

网络向量化

$$a^1 = g(W^{1T} x + b^1)$$

$$y = a^2 = g(W^{2T} x + b^2)$$

网络向量化

$$a^n = g(W^{nT} x + b^n) \Leftrightarrow z^n = W^{nT} x + b^n; a^n = g(z^n)$$

网络梯度下降

$$z^n = W^{nT} a^{n-1} + b^n$$

$$a^n = g(z^n)$$

网络梯度下降

$$da^n, \boxed{dW^n, db^n, da^{n-1}, (dz^n)}$$

$$dz^n = da^n * g'(z^n)$$

$$dW^n = dz^n \bullet a^{n-1}$$

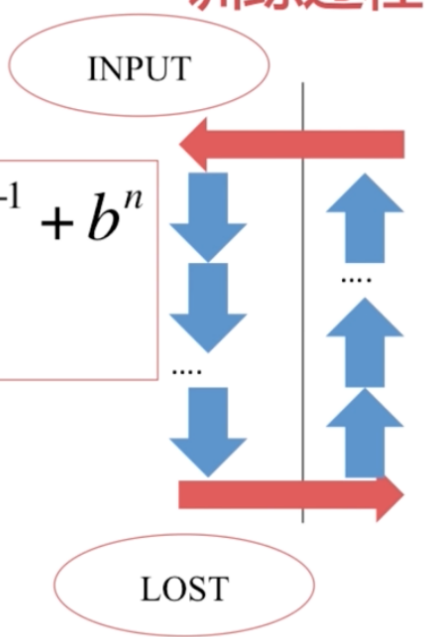
$$db = dz^n$$

$$da^{n-1} = W^{nT} \bullet dz^n$$

训练过程

$$z^n = W^{nT} a^{n-1} + b^n$$

$$a^n = g(z^n)$$



$$dz^n = da^n * g'(z^n)$$

$$dW^n = dz^n \bullet a^{n-1}$$

$$db = dz^n$$

$$da^{n-1} = W^{nT} \bullet dz^n$$