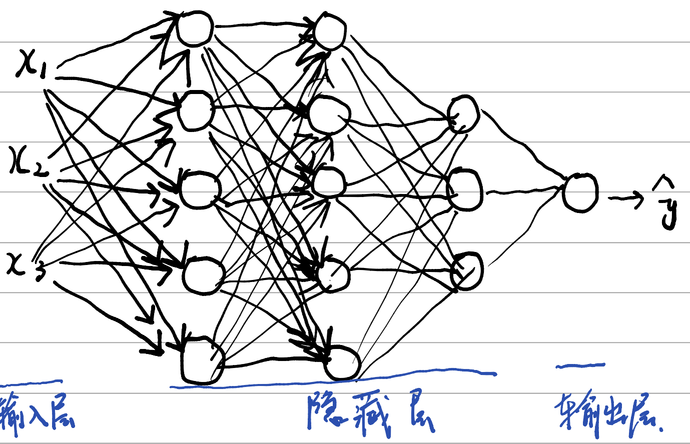
深层神经网络



假设有209张 64×64 的 rgb 格式图片
3个隐藏层，神经元个数分别为 20，7，5

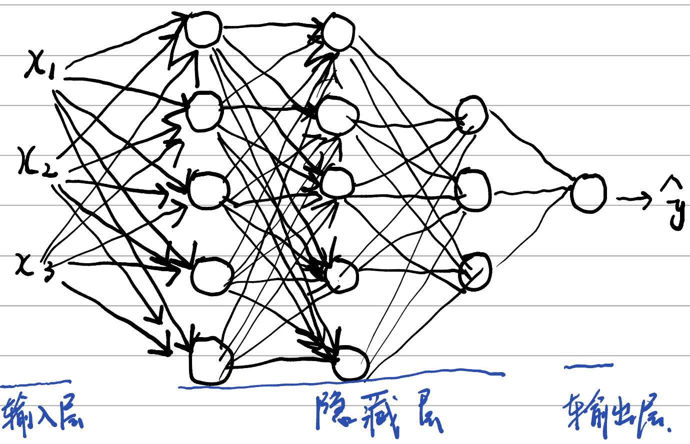
则可知输入 X 为 $(\underbrace{64\times64\times3}_{12288}, 209)$

将一张图片上所有像素点的rgb值堆叠成一列

一共有209张图片，相当于有209个样本

第一步

初始化 0~1，1~2，2~3，3~4 层计算使用的 W 和 b



第0层　第1层　第二层　第三层　第四层

$W^{[1]}=(20,12288)$ $W^{[2]}=(7,20)$ $W^{[3]}=(5,7)$ $W^{[4]}=(1,5)$

$b^{[1]}=(20,1)$ $b^{[2]}=(7,1)$ $b^{[3]}=(5,1)$ $b^{[4]}=(1,1)$

W：随机初始化

b：初始化为0

第二步

定义迭代次数，比如3000次，迭代逻辑

前向传播

0~1，1~2，2~3层激活函数使用 relu 函数

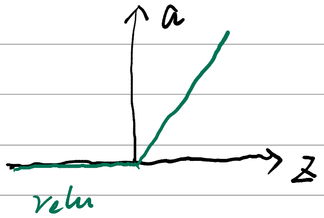
0~1层

$$z^{[1]} = W^{[1]} \cdot X + b^{[1]}$$

(20, 12288) (12288, 209) (20, 1)

(20, 209)

$$A^{[1]} = relu(z^{[1]})$$



$a = max(0, z)$

1~2层

$$z^{[2]} = W^{[2]} \cdot A^{[1]} + b^{[2]}$$

(7, 20) (20, 209) (7, 1)

(7, 209)

$$A^{[2]} = relu(z^{[2]})$$

2~3层

$$z^{[2]} = W^{[3]} \cdot A^{[2]} + b^{[3]}$$

(5, 7) (7, 209) (5, 1)

(5, 209)

$$A^{[3]} = relu(z^{[3]}$$

3~4层激活函数使用 sigmoid 函数

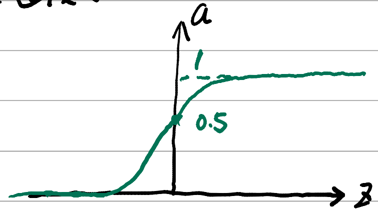
前向传播

3~4层

$$z^{[4]} = W^{[4]} \cdot A^{[3]} + b^{[4]}$$

(1, 5) (5, 209) (1, 1)

(1, 209)



sigmoid：$a = \frac{1}{1+e^{-z}}$

$$A^{[4]} = sigmoid(z^{[4]})$$

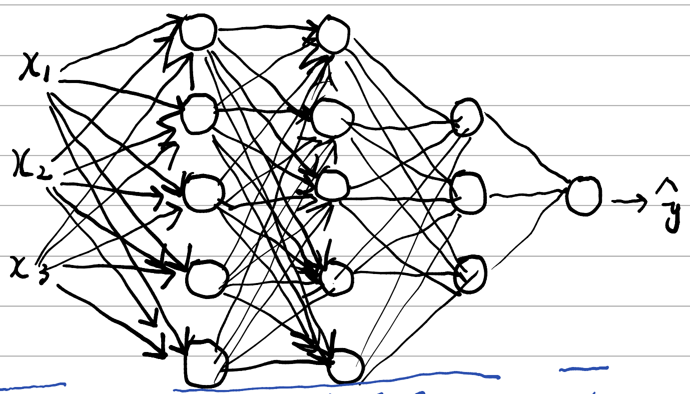
(也是所有样本的最后一层(输出层)的预测值

计算 cost

利用代价函数：

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

反向传播



计算第4层 dW, db, dA_prev

计算代价函数对输出层激活值 $A^{[L]}$ 的梯度 dAL
(1,209)

根据单个样本损失函数

$$L_i=-[y_i\log(A_i^{[L]})+(1-y_i)\log(1-A_i^{[L]})]$$

对 $A_i^{[L]}$ 求导

$$\frac{\partial L_i}{\partial A_i^{[L]}}=\frac{\partial}{\partial A_i^{[L]}}[-y_i\log(A_i^{[L]})-(1-y_i)\log(1-A_i^{[L]})]$$

$$=-y_i\cdot\frac{1}{A_i^{[L]}}+\frac{1-y_i}{1-A_i^{[L]}}$$

整理：$\frac{\partial L_i}{\partial A_i^{[L]}}=-\left(\frac{y_i}{A_i^{[L]}}-\frac{1-y_i}{1-A_i^{[L]}}\right)$

对所有样本的 $\frac{\partial L_i}{\partial A_i^{[L]}}$ 向量化表示,

$$dAL=-\left(\frac{Y}{AL}-\frac{1-Y}{1-AL}\right)$$

(1,209)

计算dZ

单样本推导

$$\frac{\partial L_i}{\partial Z_i^{[L]}}=\frac{\partial L_i}{\partial A_i^{[L]}}\cdot\frac{\partial A_i^{[L]}}{\partial Z_i^{[L]}}=$$

∵ 3~4层使用的激活函数是 sigmoid

即 $A_i^{[L]}=\sigma(Z_i^{[L]})=\frac{1}{1+e^{-Z_i^{[L]}}}$

$$\frac{\partial A_i^{[L]}}{\partial Z_i^{[L]}}=\frac{0-(0+e^{-Z_i^{[L]}}\cdot(-1))}{(1+e^{-Z_i^{[L]}})^2}$$

$$= \frac{e^{-z_i^{[L]}}}{(1+e^{-z_i^{[L]}})^2}$$

$$= \frac{1}{1+e^{-z_i^{[L]}}} \cdot \frac{1+e^{-z_i^{[L]}}-1}{1+e^{-z_i^{[L]}}}$$

$$= A_i^{[L]} \cdot (1 - A_i^{[L]})$$

$$\therefore \frac{\partial L_i}{\partial z_i^{[L]}} = \frac{\partial L_i}{\partial A_i^{[L]}} \cdot \frac{\partial A_i^{[L]}}{\partial z_i^{[L]}}$$

对所有样本向量化表示

$$\underset{(1,209)}{dz} = \underset{(1,209)}{dAL} \cdot \underset{(1,209)}{A^{[L]}(1 - A^{[L]})}$$

这里是逐元素相乘

计算dw

根据 $z = W \cdot A_{prev} + b$

$$dw = \frac{\partial L}{\partial W} = \frac{1}{m} \cdot \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w}$$

$$\frac{\partial L}{\partial z} \to \underset{(1,209)}{dz}, \quad \frac{\partial z}{\partial w} \to \underset{(\ ,209)}{A_{prev}^T}$$

计算db

$$db = \frac{\partial L}{\partial b} = \frac{1}{m} \cdot \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial b}$$

$$= \frac{1}{m} \cdot dz$$

计算 dA_prev

$$\frac{\partial L}{\partial A_{prev}} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial A_{prev}}$$

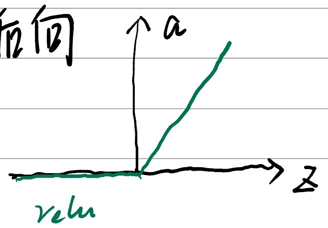
$$= W^T \cdot \underset{(1,209)}{dz}$$

dw, db 用于更新参数 w, b

dA_prev 返到第3层 继续进行梯度下降.

第3,2,1层 使用的激活函数为 Relu函数

∴在求dz时要使用Relu函数,其它后向传播和之前一致



$a=max(0,z)$

更新隐藏层 w, b 参数

$w = w - \alpha dw$

$b = b - \alpha db$

直到到达最大迭代次数

预测

输入图片，将图片存储成模型所需的输入格式，使用上面获取到的隐藏层最佳的 w, b 进行一次前向传播得到预测值，用预测值与0.5进行比较，如果概率>0.5表示与标签为1类型一致，否则与标签为0类型一致