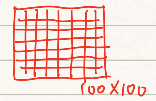
Neural Network

Motivation

当数据的维度n太大时，使用线性回归就需要很多参数。

eg：输入100×100的灰度图片：对每个像素点视为一个维。

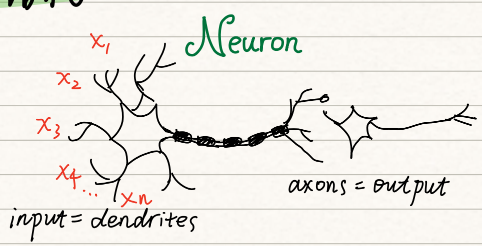


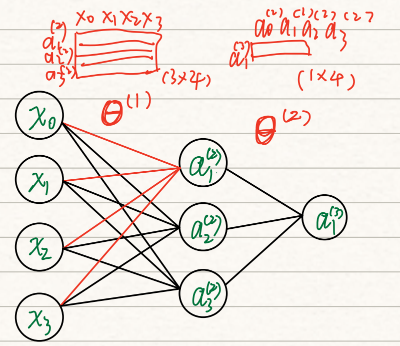
若只考虑一次项式：需要 100×100 个参数

若考虑二次项式，需要 $\frac{(10000)^2}{2}$ 个参数，而且二次式还往往表达能力有限

神经网络能很好地处理高维数据.

模型结构





"线性回归可视做单层神经网络" $h_\theta(x)=\frac{1}{1+e^{-\theta x}}$

input layer　hidden layer　output layer

$[x_0,x_1,x_2,x_3]\to[a_1^{(2)}\ a_2^{(2)}\ a_3^{(2)}]\to h_\theta(x)$

符号解释：

$a_i^{(j)}$ = 第j层的第i个激活单元

$\Theta^{(j)}$ = 第j层到第j+1层的权重；$\Theta^{(j)}$的大小为 $(\underbrace{s_{j+1}}_{输出层} \times \underbrace{s_j}_{输入层} \underbrace{+1}_{偏置层})$ → eg: $\Theta^{(1)}: 3\times(3+1)$

$\Theta^{(2)}: 1\times(3+1)$

$$\begin{cases} a_1^{(2)} = g(\Theta_{10}^{(1)}x_0+\Theta_{11}^{(1)}x_1+\Theta_{12}^{(1)}x_2+\Theta_{13}^{(1)}x_3) \\ a_2^{(2)} = g(\Theta_{20}^{(1)}x_0+\Theta_{21}^{(1)}x_1+\Theta_{22}^{(1)}x_2+\Theta_{23}^{(1)}x_3) \\ a_3^{(2)} = g(\Theta_{30}^{(1)}x_0+\Theta_{31}^{(1)}x_1+\Theta_{32}^{(1)}x_2+\Theta_{33}^{(1)}x_3) \end{cases}$$

3×4　3个输出，(3+1)个输入！

$$h_\Theta(x) = \underbrace{a_1^{(3)}}_{因为只有一个} = g(\underbrace{\Theta_{10}^{(2)}a_0^{(2)}+\Theta_{11}^{(2)}a_1^{(2)}+\Theta_{13}^{(2)}a_2^{(3)}+\Theta_{14}^{(2)}a_3^{(2)}}_{z^{(3)}})$$

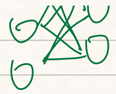
更简洁的表达：$h_\theta(x)=a_1^{(3)}=g(z_1^{(3)})=g(\Theta^{(2)}a^{(2)})$

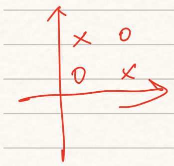
$$\begin{cases} z^{(j)} = \Theta^{(j-1)}a^{(j-1)} \\ a^{(j)} = g(z^{(j)}) \\ h_\theta(x) = a^{(j+1)} = g(z^{(j+1)}) \end{cases}$$

由其决定

理解：线性回归就相当于从输入层直接到输出层的映射
而神经网络是从隐藏层到输出层，
隐藏层是输入层的映射。
众多的隐藏层就提高了特征的表示能力。

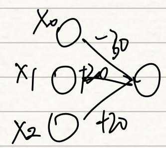
想法： 这样一个单层神经网络就是softmax回归.

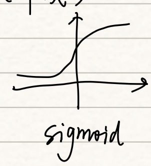




XOR异或

神经网络表示 OR(异或) ⟷ 线性回归很难表示.





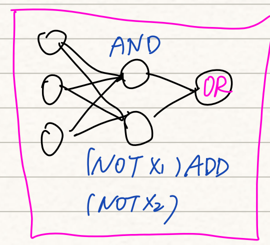
$$g(-30+20x_1+20x_2)$$

x_1	x_0	$h_\theta(x)$
0	0	$g(-30)\approx 0$
1	0	$g(-10)\approx 0$
0	1	$g(-10)\approx 0$
1	1	$g(10)\approx 1$

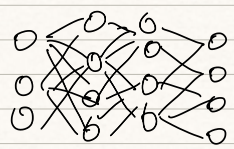
可表示 AND, OR, XOR, NOR, (NOT X1) AND (NOT X2) 等.

中间一些隐藏层可计算之一.
各单元

如: XOR:



Multi-Class Classification



$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \rightarrow \begin{bmatrix} a_1^{(2)} \\ a_2^{(2)} \\ \vdots \end{bmatrix} \rightarrow \begin{bmatrix} a_1^{(3)} \\ a_2^{(3)} \\ \vdots \end{bmatrix} \rightarrow \begin{bmatrix} h_\theta(x)_1 \\ h_\theta(x)_2 \\ \vdots \\ h_\theta(x)_n \end{bmatrix}$$

一个 $h_\theta(x)$ 代表 $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$.

第三类.