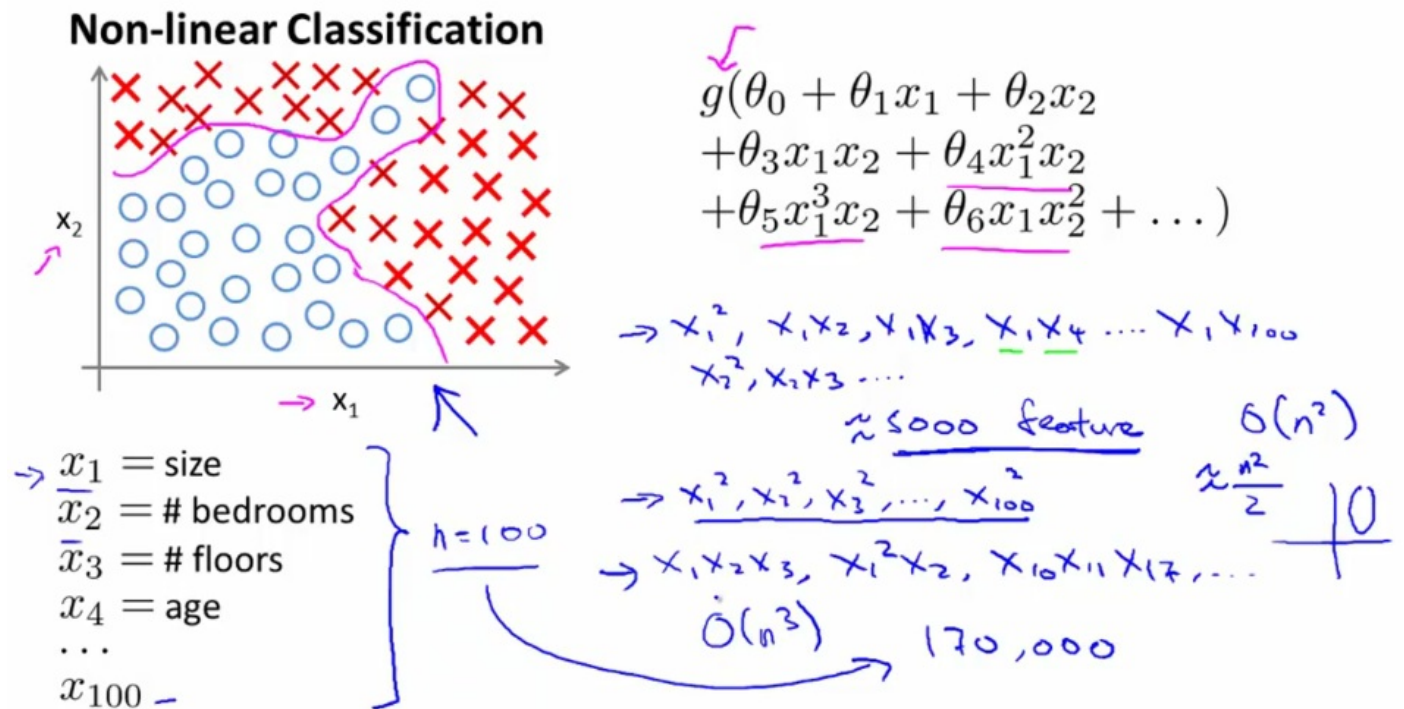
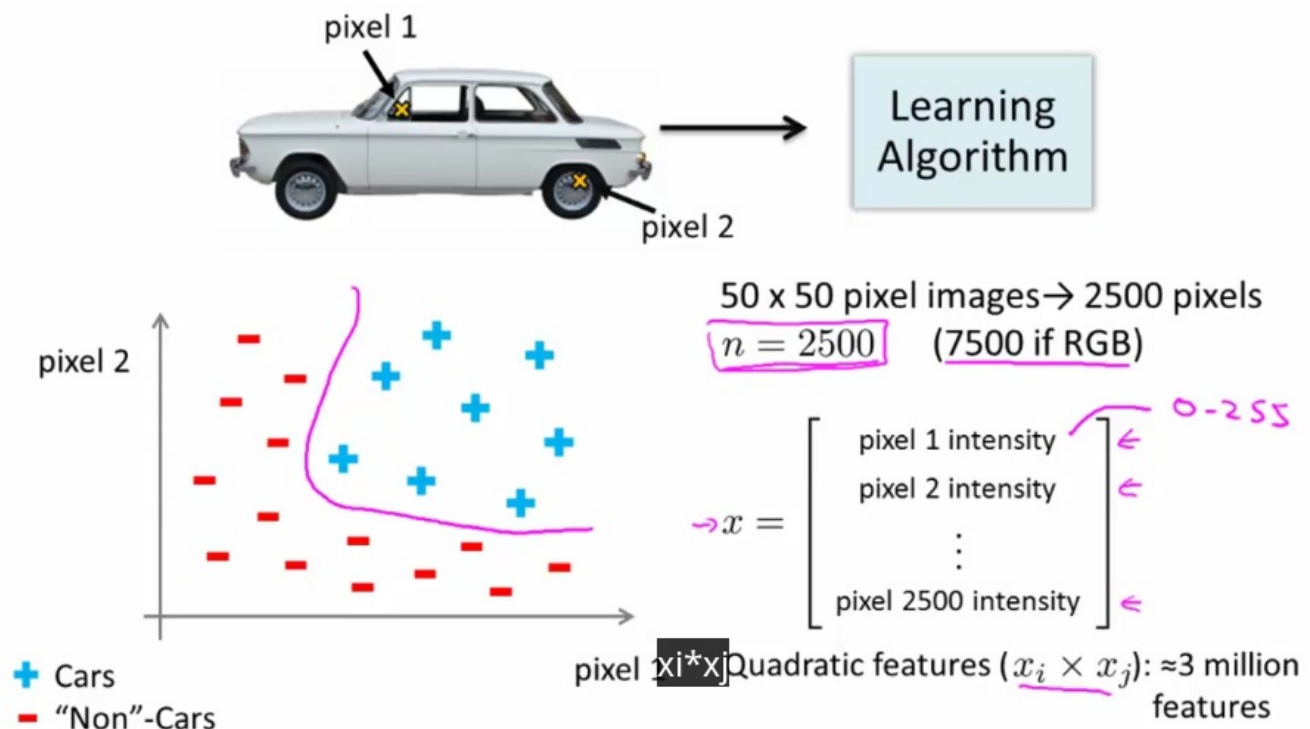


1. 当特征数量很多的时候，如果想包括其中所有的二次项是很难的。当然可以只包括其中部分的二次项，但是忽略太多的二次项会影响模型效果，另外还存在一个问题，除了二次项，三次项的数量更多。



检测汽车的学习算法特征数量：



因此简单的增加二次项或者三次项不是一个解决非线性回归问题的办法。

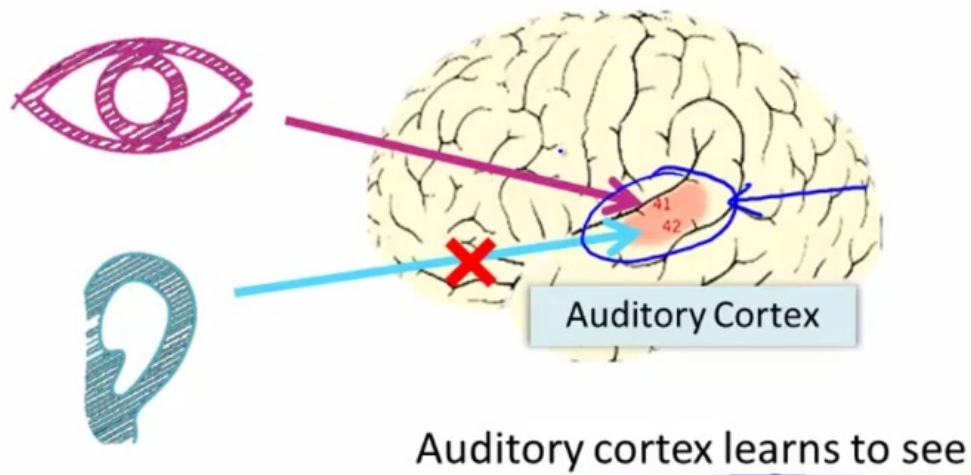
2. 神经网络。历史：

Neural Networks

- Origins: Algorithms that try to mimic the brain.
- Was very widely used in 80s and early 90s; popularity diminished in late 90s.
- Recent resurgence: State-of-the-art technique for many applications

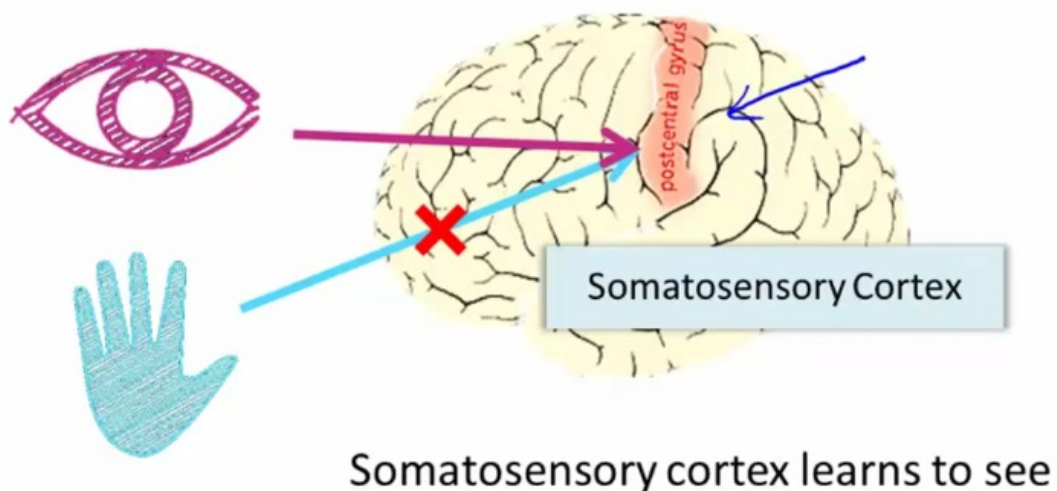
一个学习算法来模仿大脑假设：

The “one learning algorithm” hypothesis



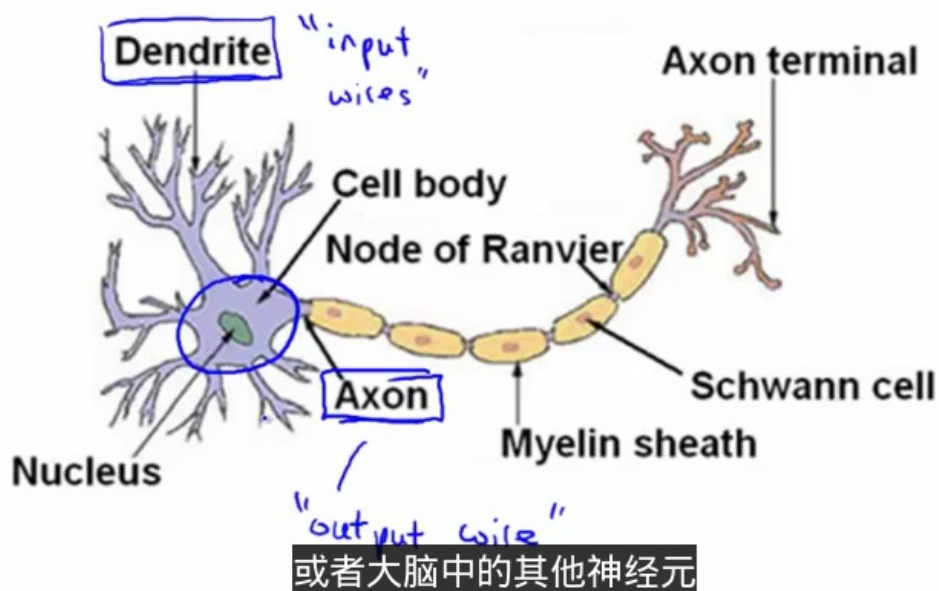
大脑里面处理听觉的部分转接到眼睛，就可以学会视觉。处理触觉的部分转接到眼睛也能完成视觉辨认任务。这称为神经重接实验。

The “one learning algorithm” hypothesis



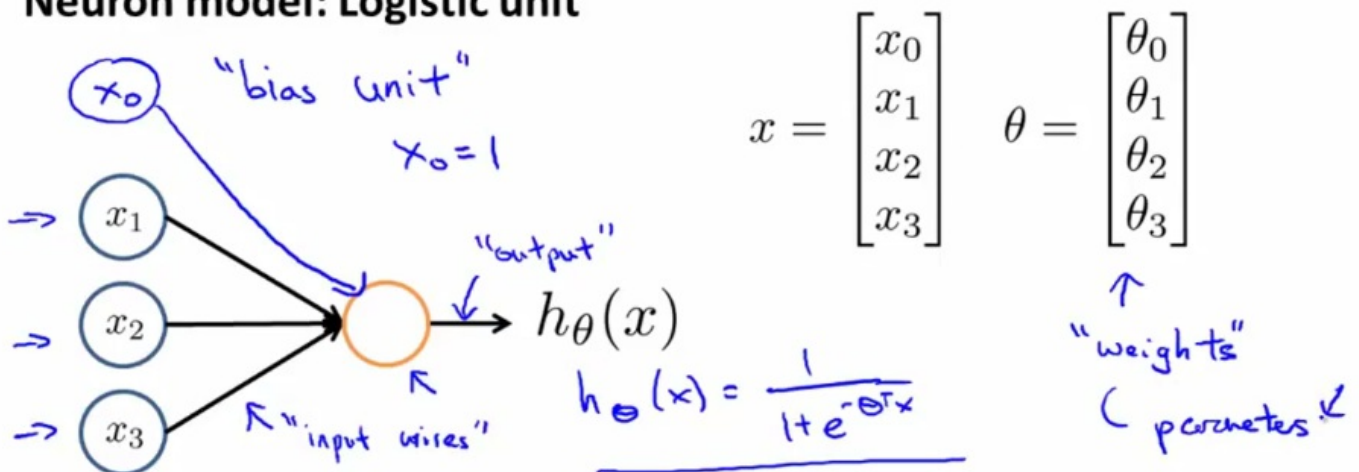
神经元结构：细胞，传入神经元，传出神经元。

Neuron in the brain



人工神经网络用逻辑单元来模拟神经元活动。(x0-偏置单元) 下面是单一神经元：

Neuron model: Logistic unit

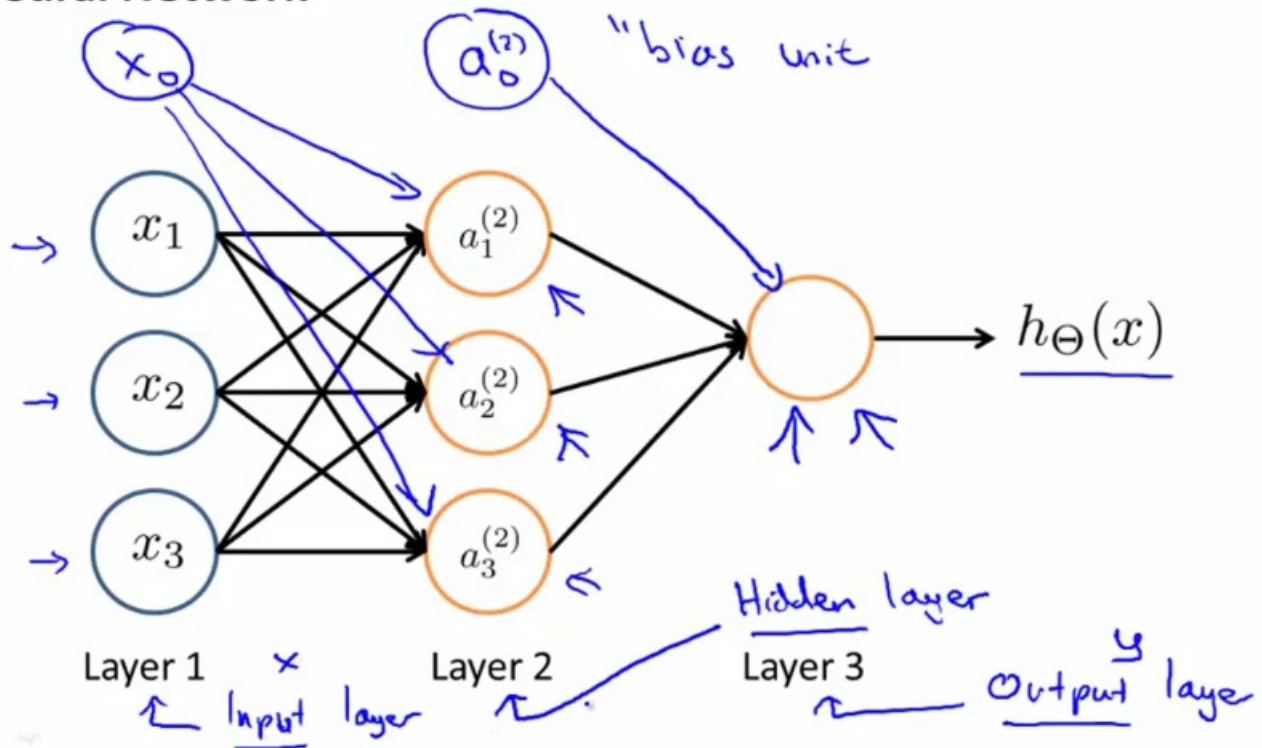


Sigmoid (logistic) activation function.

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

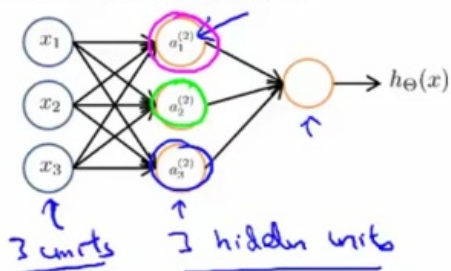
输入层，隐藏层，输出层：

Neural Network



记号表示及映射矩阵的维度：

Neural Network



$\rightarrow a_i^{(j)}$ = "activation" of unit i in layer j

$\rightarrow \Theta^{(j)}$ = matrix of weights controlling function mapping from layer j to layer $j+1$

$$\Theta^{(1)} \in \mathbb{R}^{3 \times 4}$$

$$\rightarrow a_1^{(2)} = g(\Theta_{10}^{(1)} x_0 + \Theta_{11}^{(1)} x_1 + \Theta_{12}^{(1)} x_2 + \Theta_{13}^{(1)} x_3)$$

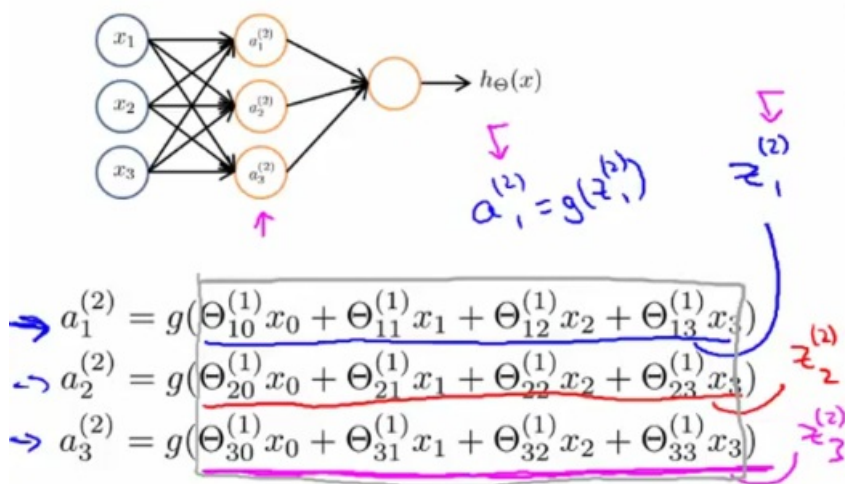
$$\rightarrow a_2^{(2)} = g(\Theta_{20}^{(1)} x_0 + \Theta_{21}^{(1)} x_1 + \Theta_{22}^{(1)} x_2 + \Theta_{23}^{(1)} x_3)$$

$$\rightarrow a_3^{(2)} = g(\Theta_{30}^{(1)} x_0 + \Theta_{31}^{(1)} x_1 + \Theta_{32}^{(1)} x_2 + \Theta_{33}^{(1)} x_3)$$

$$h_{\Theta}(x) = a_1^{(3)} = g(\Theta_{10}^{(2)} a_0^{(2)} + \Theta_{11}^{(2)} a_1^{(2)} + \Theta_{12}^{(2)} a_2^{(2)} + \Theta_{13}^{(2)} a_3^{(2)})$$

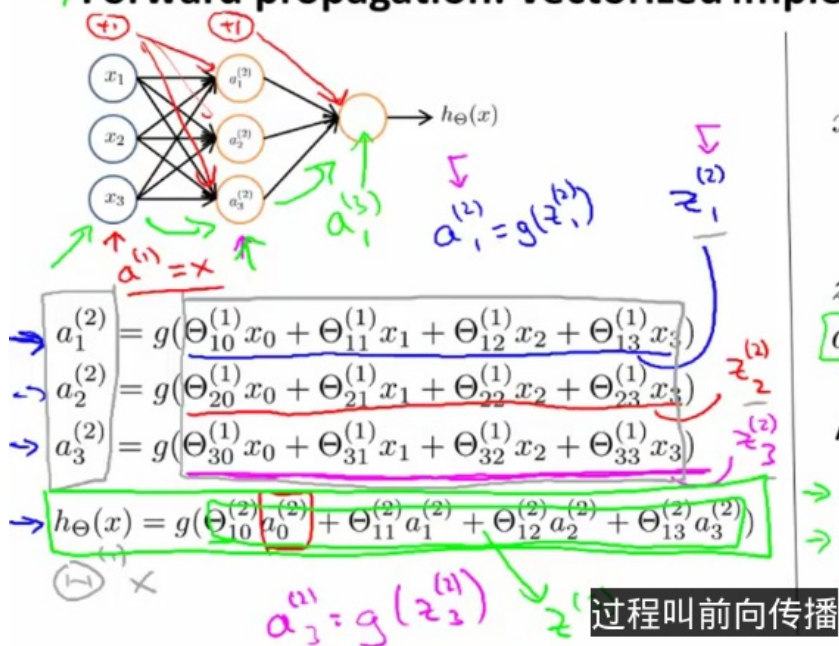
\rightarrow If network has s_j units in layer j , s_{j+1} units in layer $j+1$, then $\Theta^{(j)}$ will be of dimension $s_{j+1} \times (s_j + 1)$.

参数简化表示：z



向量化：->前向传播 (Forward Propagation)

Forward propagation: Vectorized implementation



$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad z^{(2)} = \begin{bmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \end{bmatrix}$$

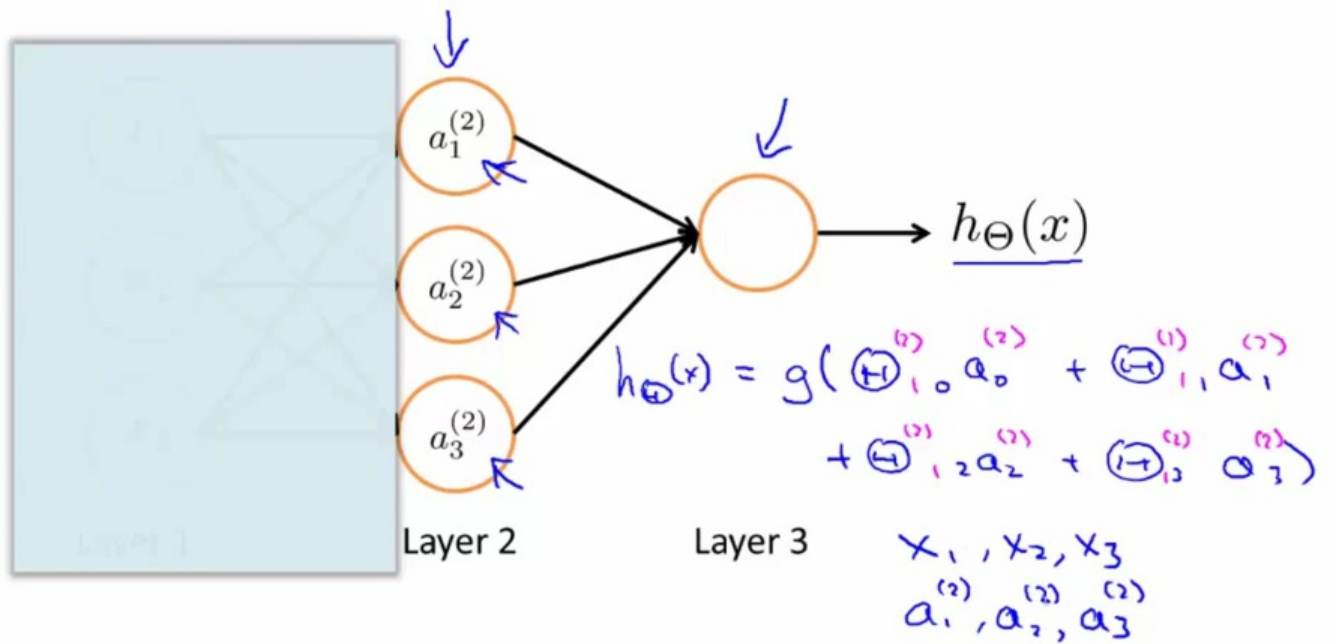
$$z^{(2)} = \Theta^{(1)}a^{(1)} \quad a^{(2)} = g(z^{(2)})$$

Add $a_0^{(2)} = 1$. $\rightarrow a^{(2)} \in \mathbb{R}^4$

$$z^{(3)} = \Theta^{(2)}a^{(2)} \quad h_{\Theta}(x) = a^{(3)} = g(z^{(3)})$$

从某种角度看，神经网络就像是逻辑回归模型，只不过模型特性不是 x ,而是它自己的特性，即隐藏层：

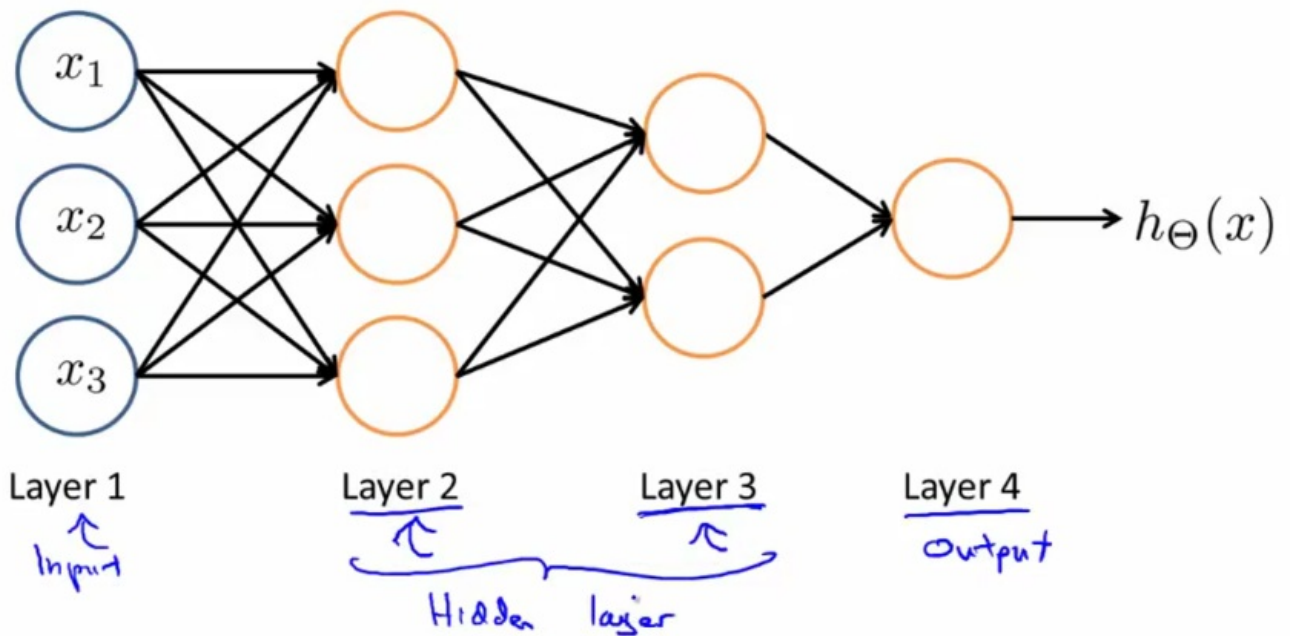
Neural Network learning its own features



用不同的参数值(theta)来学习到隐藏层，然后把隐藏层作为特征来训练逻辑回归模型。

更复杂的网络架构：

Other network architectures

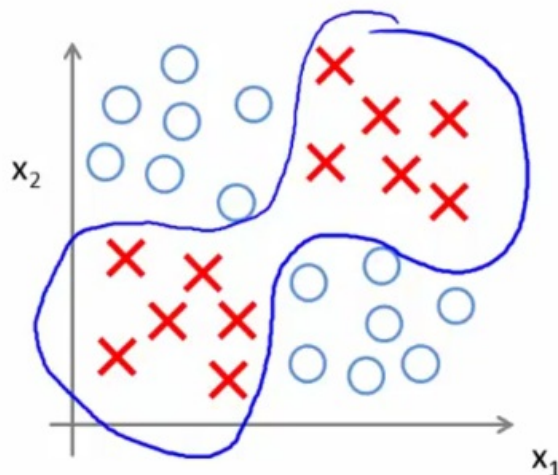
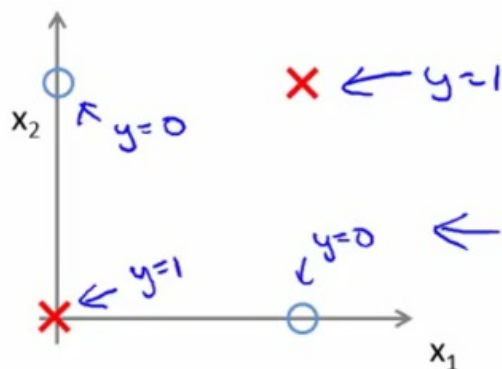


Andrew

非线性分类例子：异或。

Non-linear classification example: XOR/XNOR

→ x_1, x_2 are binary (0 or 1).



$$y = x_1 \text{ XOR } x_2$$

$$x_1 \text{ XNOR } x_2$$

$$\text{NOT } (x_1 \text{ XOR } x_2)$$

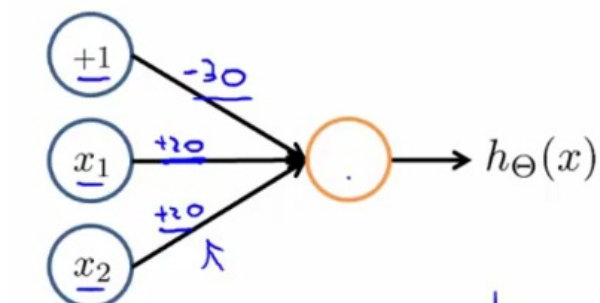
的神经网络 我们先

可以实现与运算的神经网络：

Simple example: AND

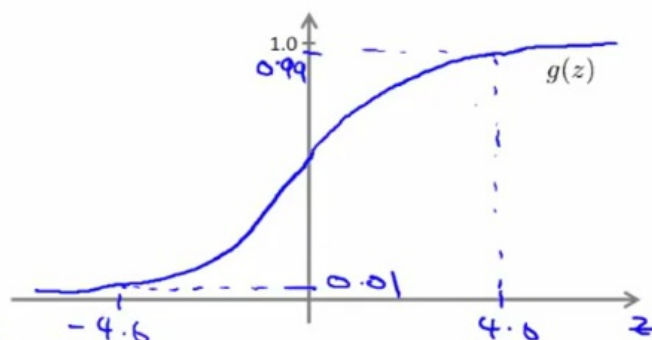
→ $x_1, x_2 \in \{0, 1\}$

→ $y = x_1 \text{ AND } x_2$



$$h_{\Theta}(x) = g(-30 + 20x_1 + 20x_2)$$

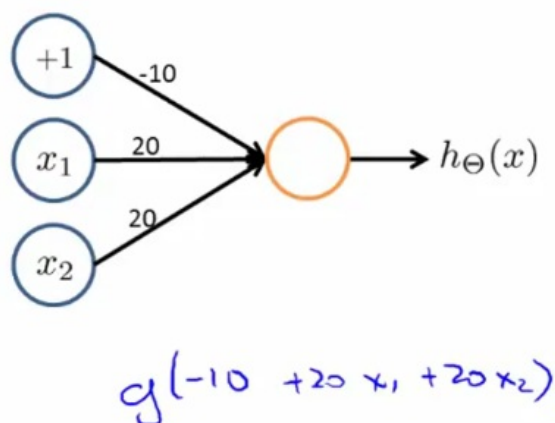
逻辑“与”的计算结果



x_1	x_2	$h_{\Theta}(x)$
0	0	$g(-30) \approx 0$
0	1	$g(-10) \approx 0$
1	0	$g(-10) \approx 0$
1	1	$g(10) \approx 1$

可以实现逻辑或的神经网络：

Example: OR function



x_1	x_2	$h_{\Theta}(x)$
0	0	$g(-10) \approx 0$
0	1	$g(10) \approx 1$
1	0	≈ 1
1	1	≈ 1

实现逻辑或的神经网络：

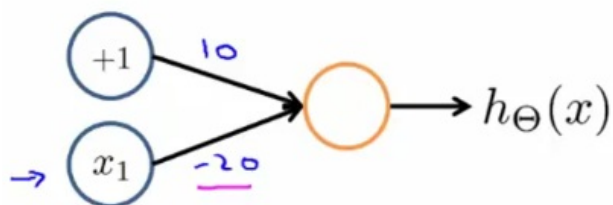
$\rightarrow x_1 \text{ AND } x_2$

$\rightarrow x_1 \text{ OR } x_2$

$\{0, 1\}$

Negation:

NOT x_1



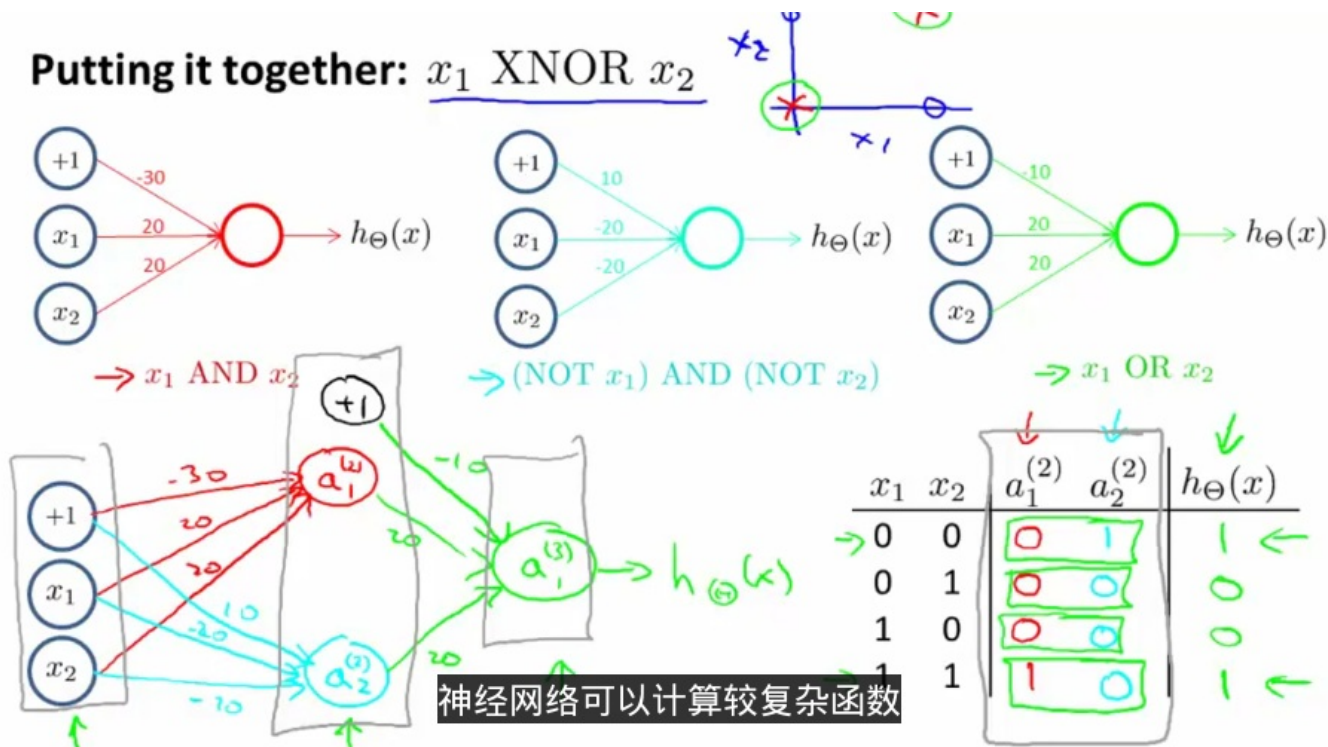
x_1	$h_{\Theta}(x)$
0	$g(10) \approx 1$
1	$g(-10) \approx 0$

$$h_{\Theta}(x) = g(10 - 20x_1)$$

$\rightarrow (\text{NOT } x_1) \text{ AND } (\text{NOT } x_2)$
 来计算 $= 1$ if and only if
 $\rightarrow x_1 = x_2 = 0$

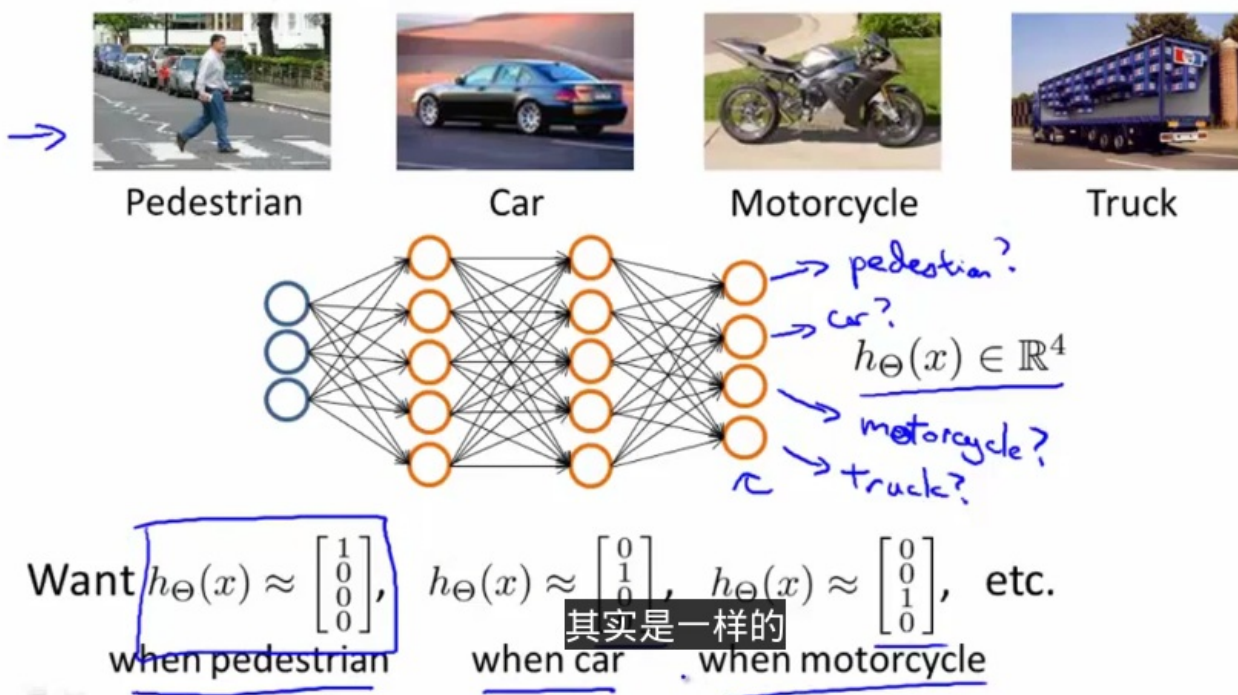
搭建人工神经网络实现异或：

Putting it together: x_1 XNOR x_2



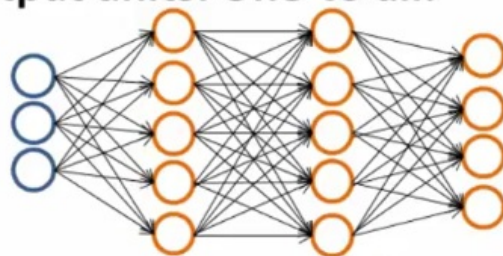
多输出单元：One-vs-All: 输出多元向量

Multiple output units: One-vs-all.



训练方法：

Multiple output units: One-vs-all.



$$h_{\Theta}(x) \in \mathbb{R}^4$$

Want $h_{\Theta}(x) \approx \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $h_{\Theta}(x) \approx \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $h_{\Theta}(x) \approx \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, etc.
 when pedestrian when car when motorcycle

Training set: $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})$

$\rightarrow y^{(i)}$ one of $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$
 pedestrian car motorcycle truck

$(x^{(i)}, y^{(i)})$

~~Previously~~
 ~~$y \in \{1, 2, 3, 4\}$~~
 $h_{\Theta}(x^{(i)}) \approx y^{(i)}$
 $\in \mathbb{R}^4$