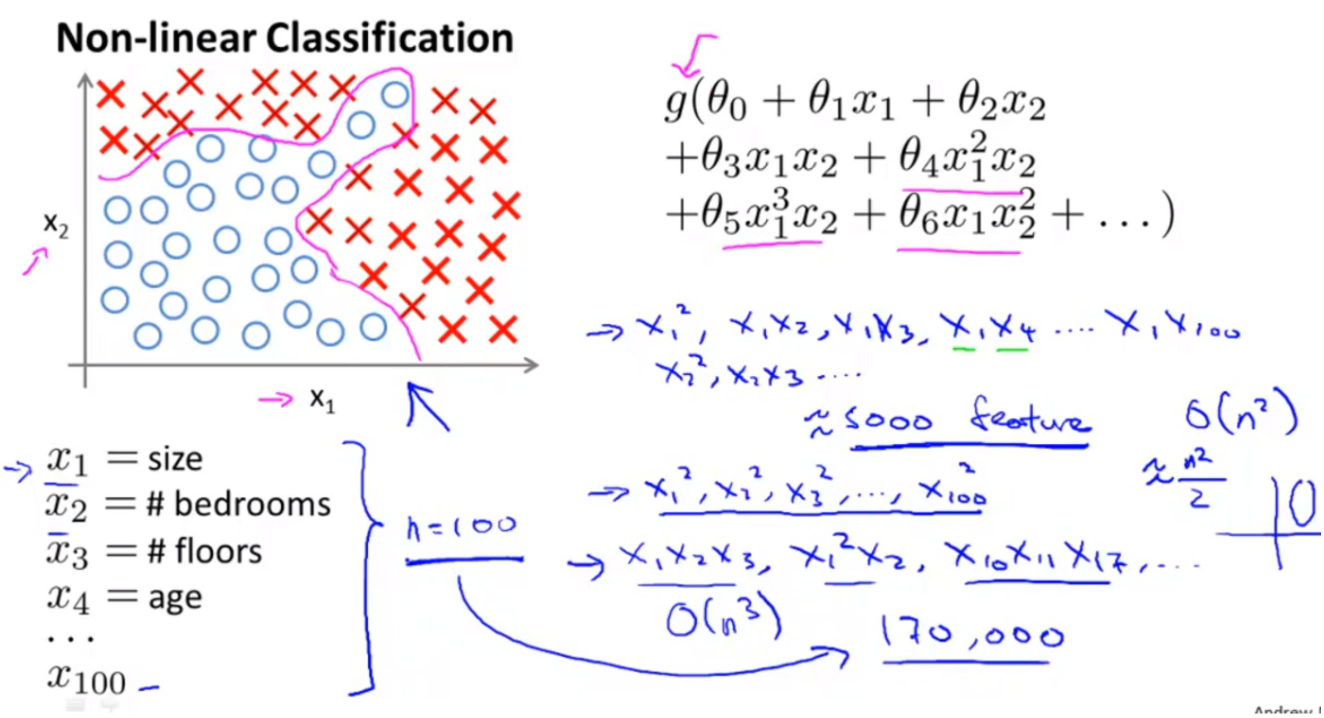
Week_4 神经网络

Non-linear Hypotheses(非线性假设)

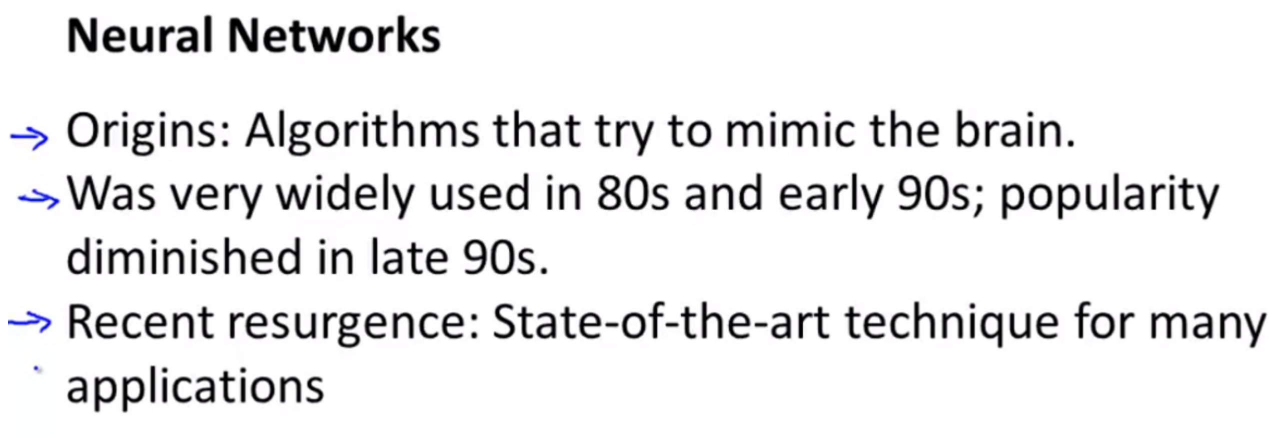
上一节中我们曾经的提到的一个监督学习分类的问题,如果我们利用逻辑回归算法来解决这样一个问题,我们首先需要构造一个包含很多非线性项的逻辑回归函数即$g(\theta^T X)$,当我们的特征向量比较少的时候这种算法确实能得到不错的结果比如只有x_1和x_2两项,但是在许多复杂的机器学习问题中,涉及的特征向量往往是多于两项的,但特征向量变多,由特征向量带来的组合形式也会变多,如下:



因此,只是简单的增加 二次项或者三次项之类的逻辑回归算法并不是一个解决 复杂非线性问题的好办法 因为当n很大时 将会产生非常多的特征项,比如我们在计算机视觉中所需要的非线性分类器来判断图片上的内容

Neurons and the Brain

- 神经网络算法的背景



- 我们大脑中有一块区域可以同时处理视觉,听觉与触觉.这使我们联想到实现大脑处理的方法是不是只需要一个单一的算法就可以解决,而不是需要运行成千上万个不同的程序或算法来做这些判断,也许我们需要做的就是找出一些近似的或实际的大脑学习算法然后通过大脑自学掌握如何处理不同类型的数据

Model Representation I(模型表示1)

在神经网络中,是由树突即输入神经(接收来自其他神经元的信息),神经元的输出神经叫做轴突(用来给其他神经元传递信号或者传递信息的),简而言之,神经元可以看作是一个计算单位

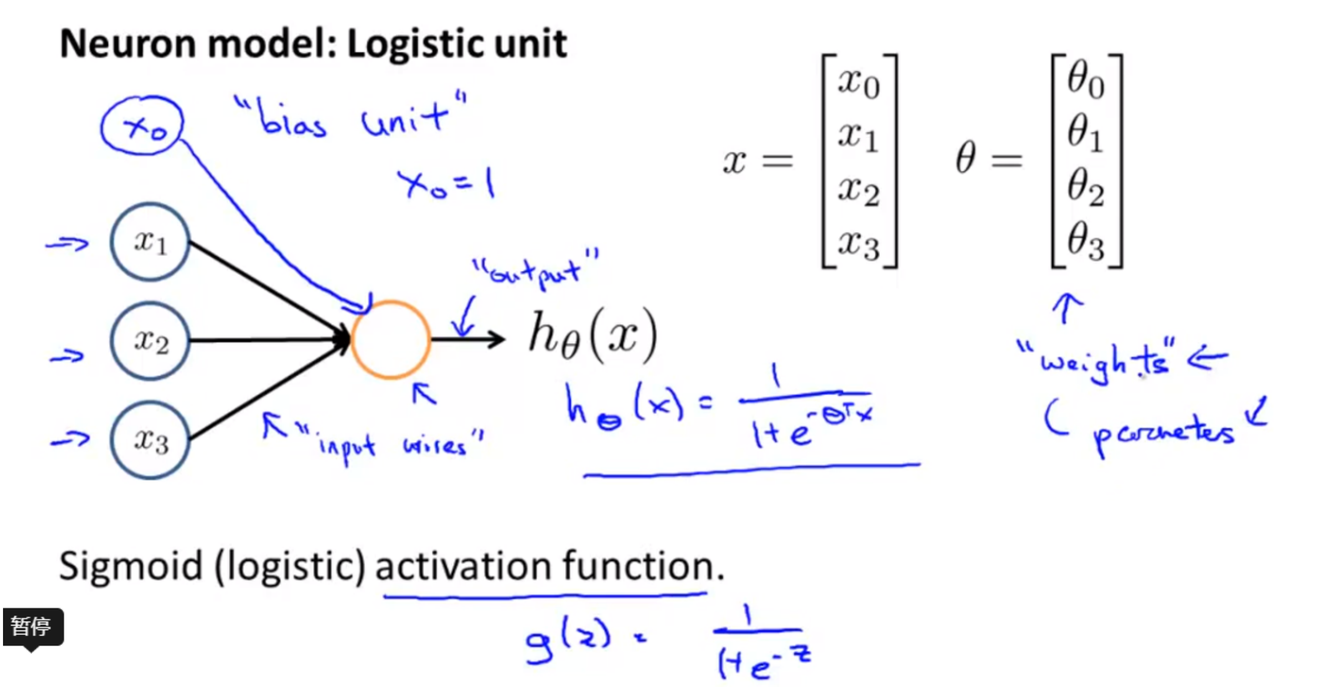
- 它从输入神经接受一定数目的信 并做一些计算然后将结果通过它的轴突传送到其他节点或者大脑中的其他神经元

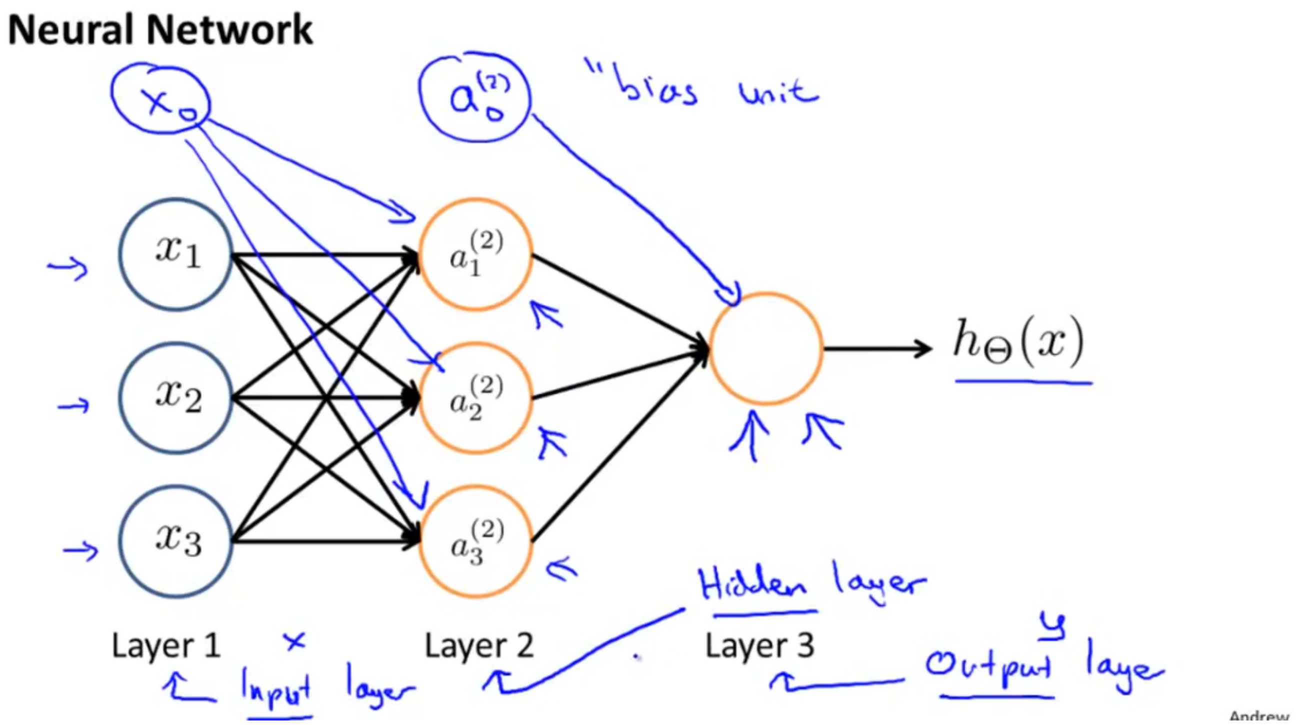
由此可以看出人类思考的模型:我们的神经元把自己的收到的消息进行计算并向其他神经元 传递消息

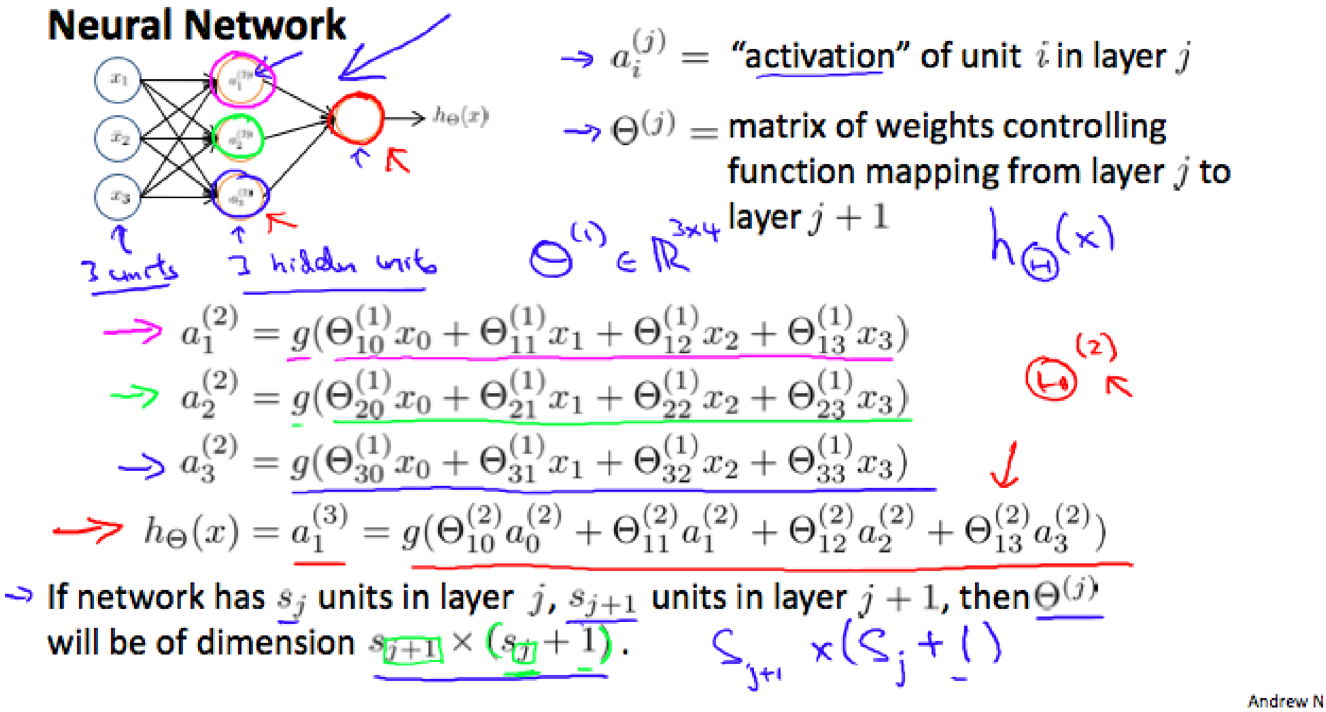
在神经网络中,我们可以利用在分类问题中使用的逻辑函数即$g(\theta^T X) = \frac{1}{1+e^-\theta^T X}$

,在此也可以叫激励函数,这里的θ参数可以被称做权重

其逻辑单元表示如下:

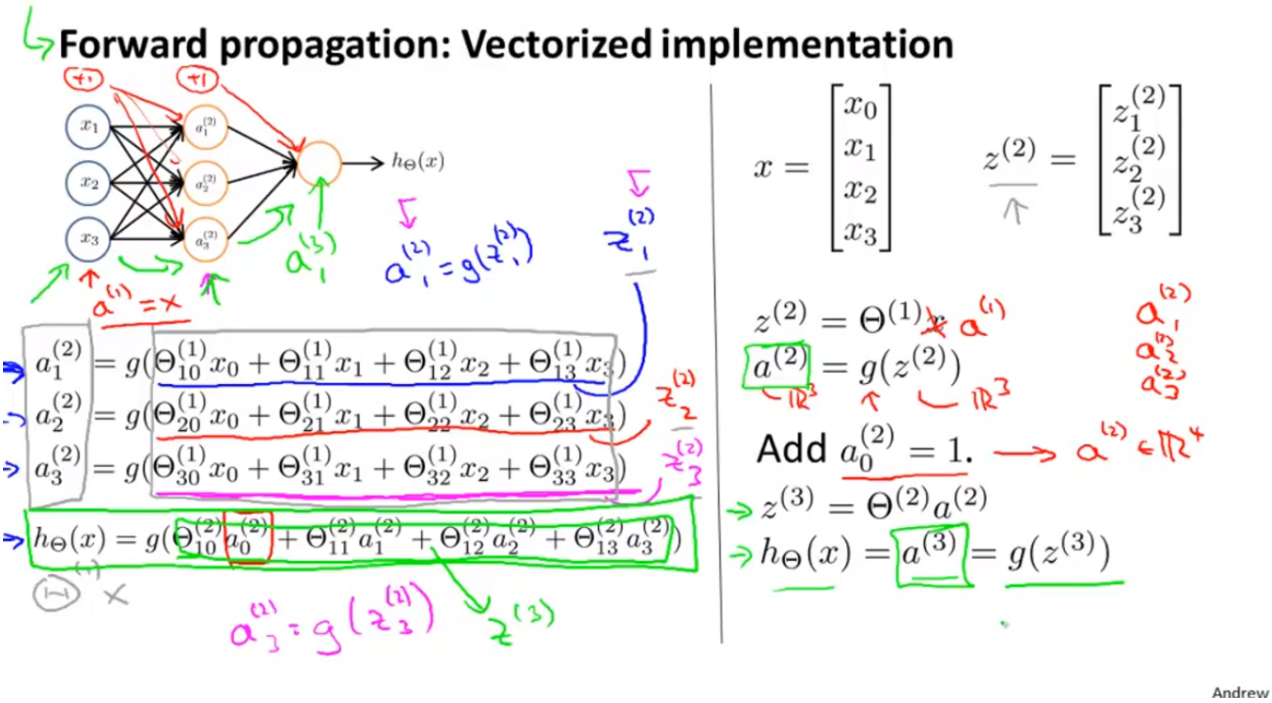




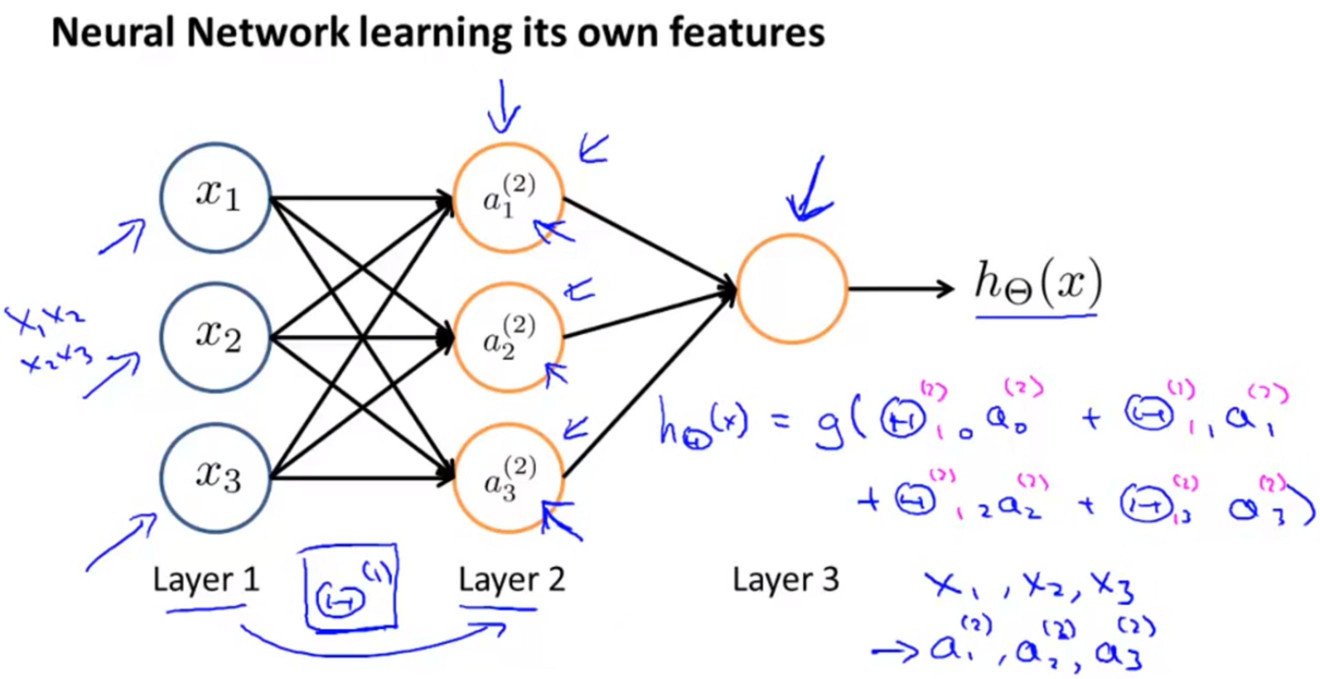


Model Representation II(模型表示2)

- Forward propagation(前向传播):Vectorized implementation

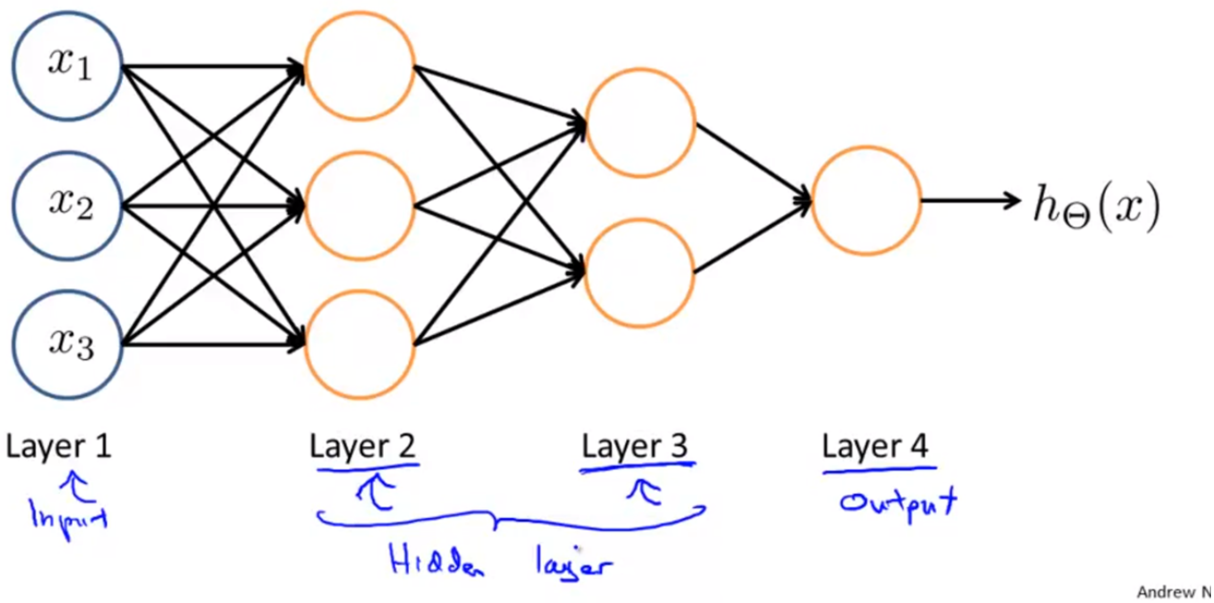


- 与先前的逻辑回归不同的是,这个算法可以灵活低快速学习任意的特征项,并把计算后的特征项输入到最后的单元,即利用这样一个隐藏层计算更复杂的特征并输入到最后的输出层



- 我们从其他更复杂的神经网络架构也可以看出神经网络算法的原理:从输入层的激励开始向前传播到第一隐藏层然后传播到第二 隐藏层或者更多的隐藏层计算出更复杂的特征项最终到达输出层.并且通过向量化这些计算使的整个算法变得容易理解

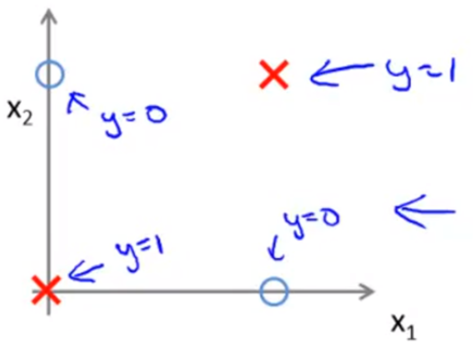
Other network architectures

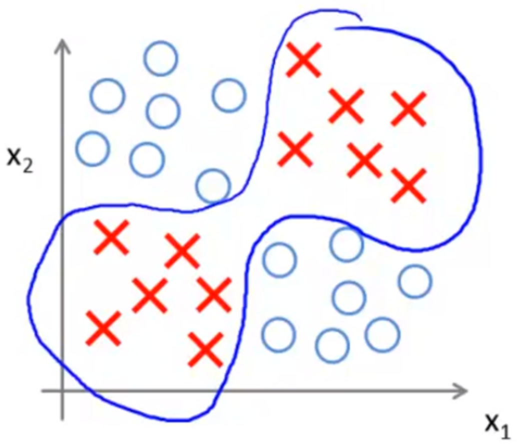


Examples and Intuitions I

Non-linear classification example: XOR/XNOR

→ x_1, x_2 are binary (0 or 1).





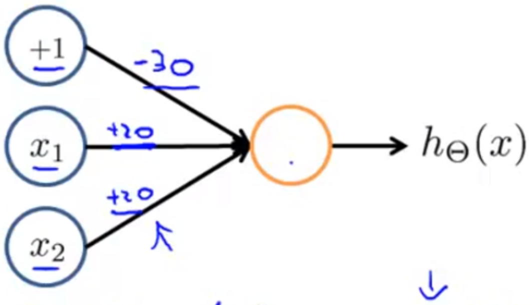
$$y = x_1 \text{ XOR } x_2$$

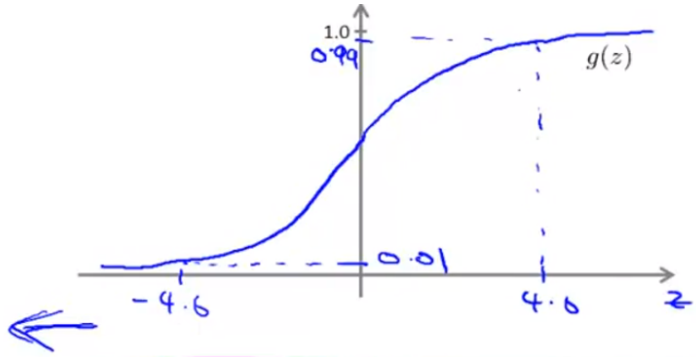
$$x_1 \text{ XNOR } x_2$$

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

Simple example: AND

- $x_1, x_2 \in \{0, 1\}$
- $y = x_1 \text{ AND } x_2$



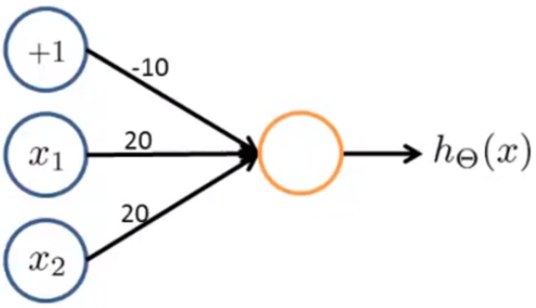


$$h_\Theta(x) = g(\underset{\Theta^{(1)}_{10}}{-30} + \underset{\Theta^{(1)}_{11}}{20}x_1 + \underset{\Theta^{(1)}_{12}}{20}x_2)$$

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	$g(10) \approx 1$

$h_\Theta(x) \approx x_1 \text{ AND } x_2$

Example: OR function



$g(-10 + 20x_1 + 20x_2)$

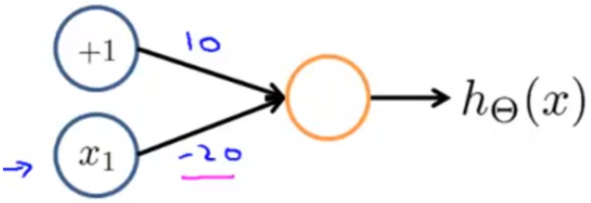
x_1	x_2	$h_\Theta(x)$
0	0	$g(-10) \approx 0$
0	1	$g(10) \approx 1$
1	0	≈ 1
1	1	≈ 1

Examples and Intuitions II

→ x_1 AND x_2 → x_1 OR x_2 $\{0,1\}$.

Negation:

NOT x_1



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

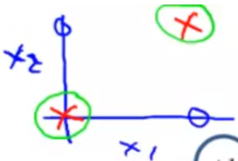
$$h_\Theta(x) = g(10 - 20x_1)$$

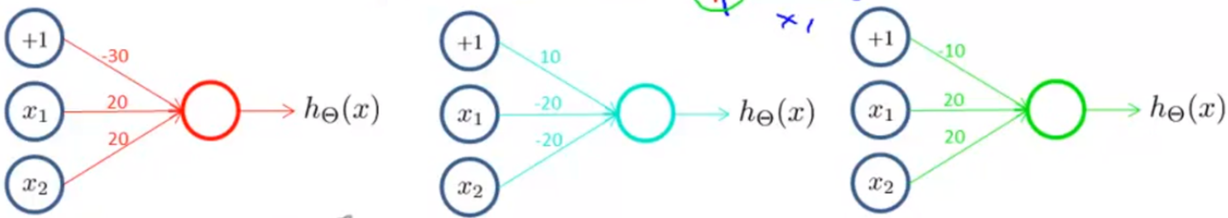
→ (NOT x_1) AND (NOT x_2)

(= 1 if and only if

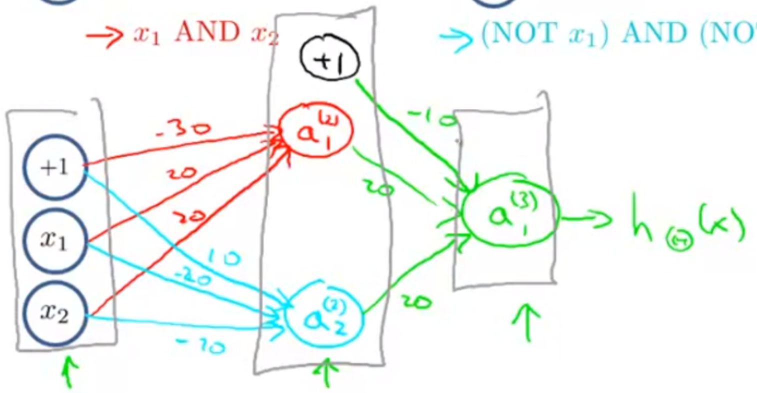
→ $x_1 = x_2 = 0$

Putting it together: x_1 XNOR x_2

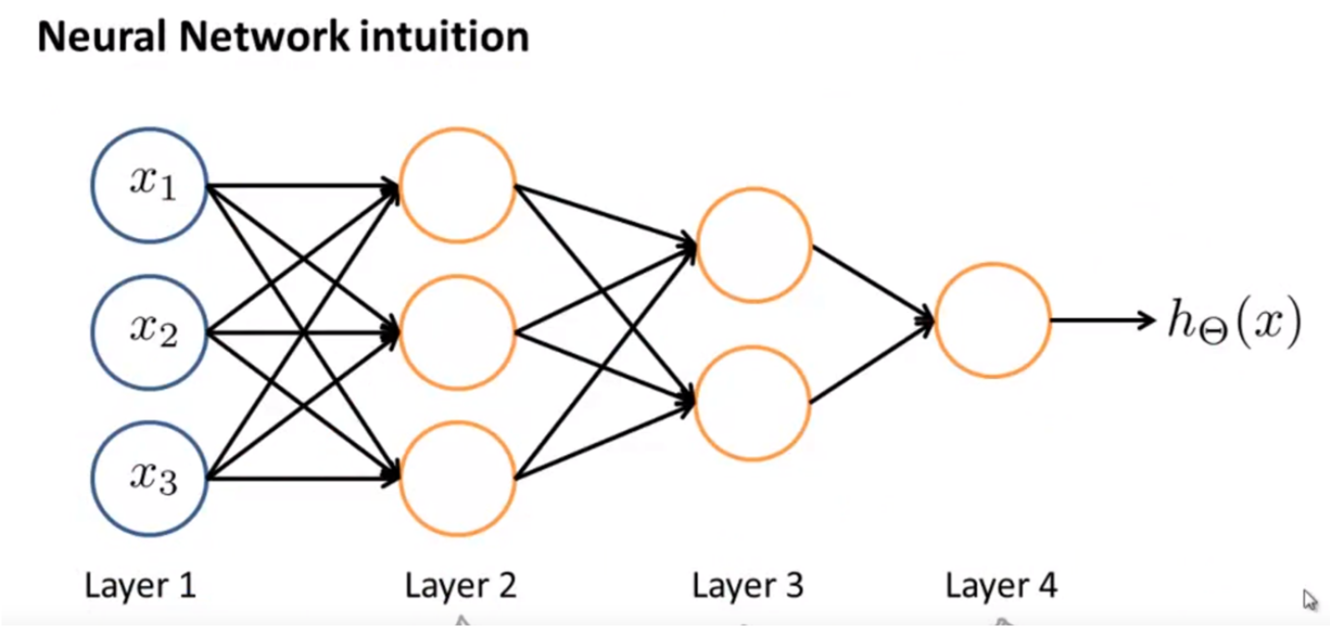




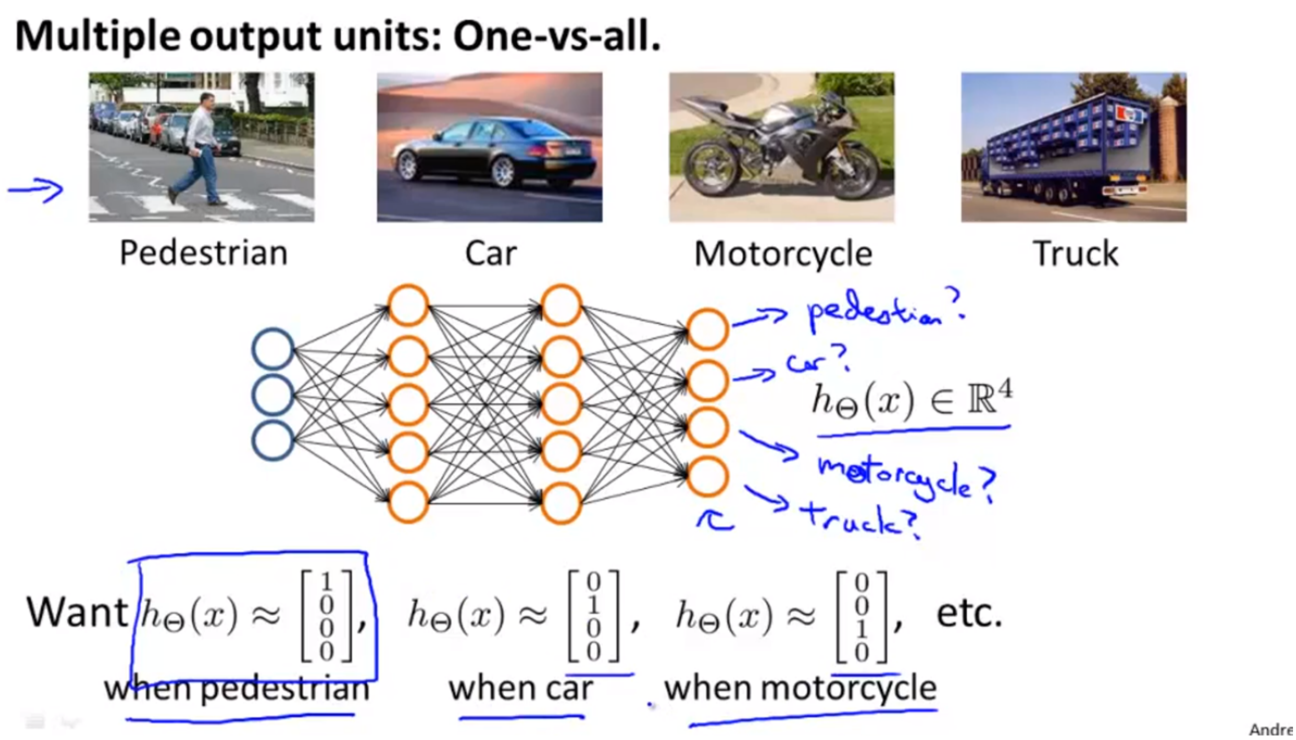
→ x_1 AND x_2 → (NOT x_1) AND (NOT x_2) → x_1 OR x_2



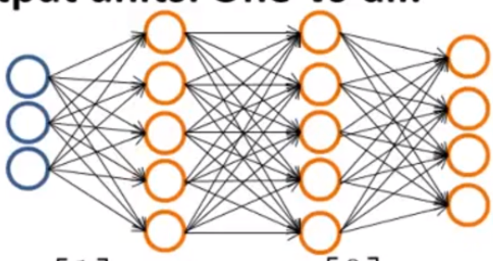
	x_1	x_2	$a_1^{(2)}$	$a_2^{(2)}$	$h_\Theta(x)$	
→	0	0	0	1	1	←
	0	1	0	0	0	
	1	0	0	0	0	
→	1	1	1	0	1	←



Multiclass Classification



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: $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})$

~~Previously $y \in \{1, 2, 3, 4\}$~~

$y^{(i)}$ 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

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

$\underline{h_\Theta(x^{(i)})} \approx \underline{y^{(i)}}$

$\in \mathbb{R}^4$