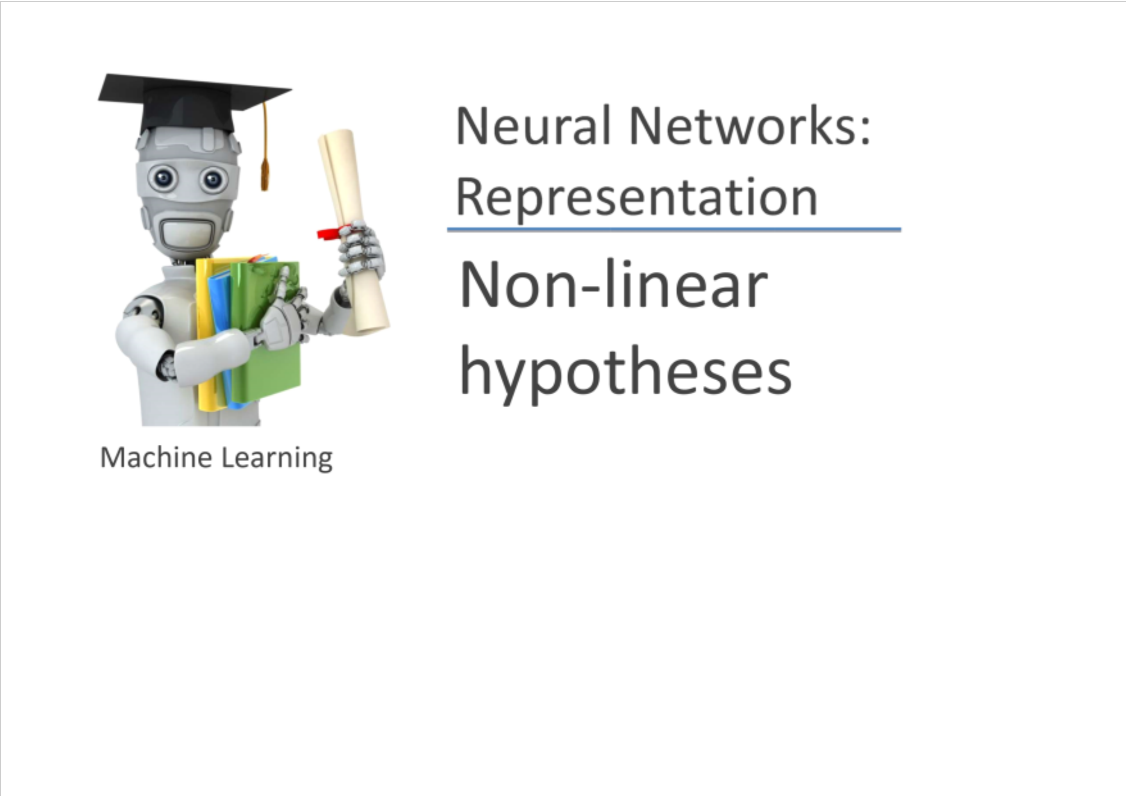
Lecture8

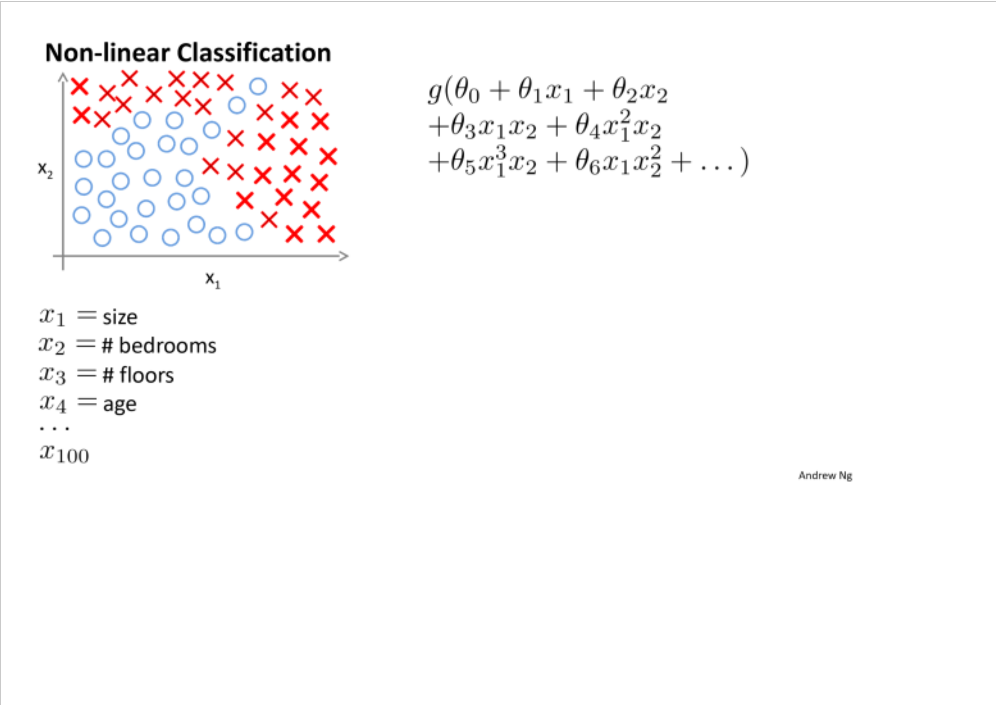
2017年7月31日 18:20



神经网络入门
非线性假设

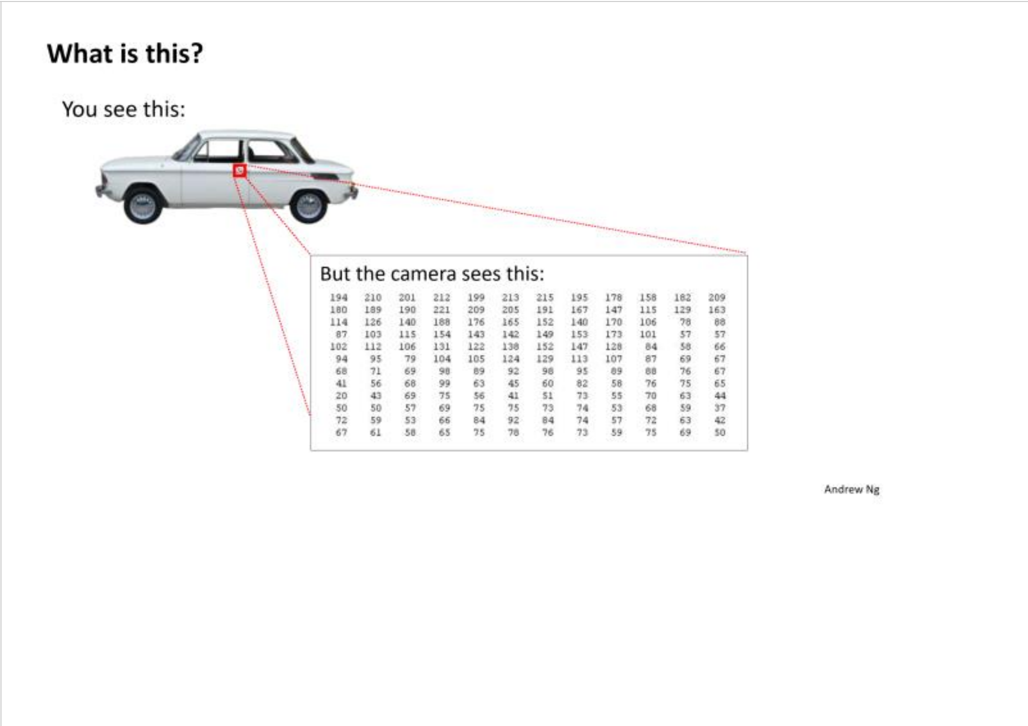
第 2 页

2017年7月31日　18:20



通过不断增加变量x以及次方，来实现复杂的曲线，用来划分。

这里只需要明白，就是不断增加特征量还有参数theta。



视觉处理的例子，计算机是处理图像的通过矩阵形式数据

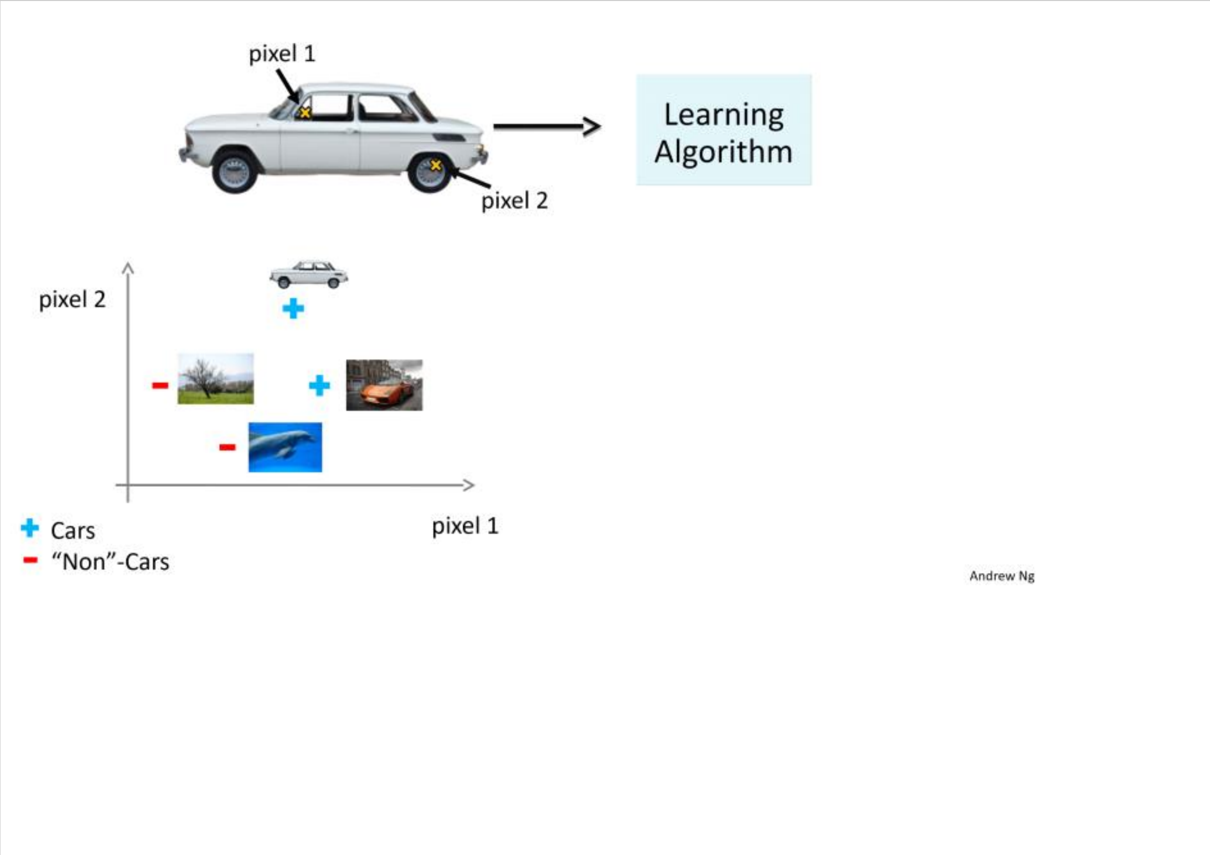
第 4 页

2017年7月31日 18:20

0/1分类

第 5 页

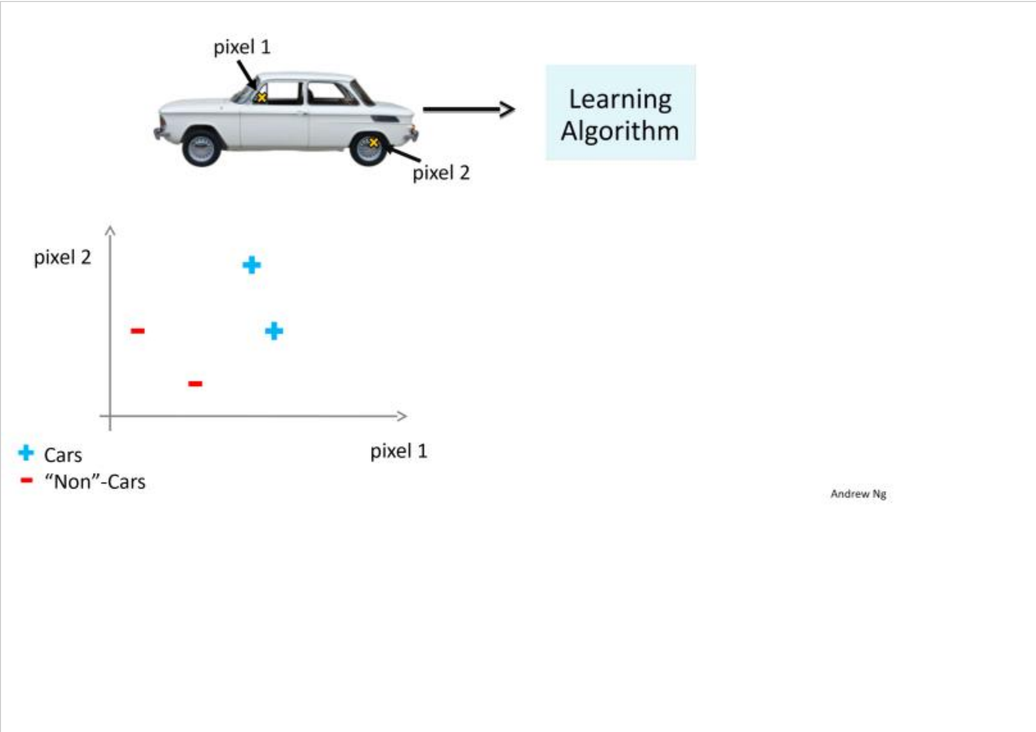
2017年7月31日 18:20



汽车的例子

第 6 页

2017年7月31日 18:20



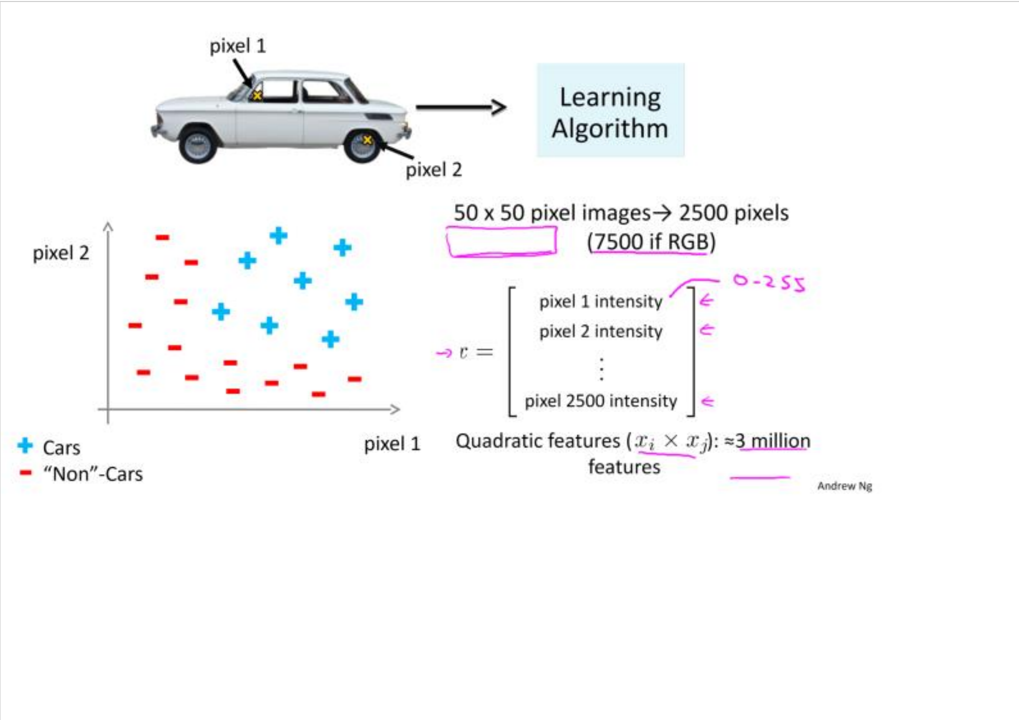
将汽车抽象化为坐标对应的点

加号代表是汽车

减号代表不是汽车

这样就可以使用逻辑回归分割了

2017年7月31日 18:20

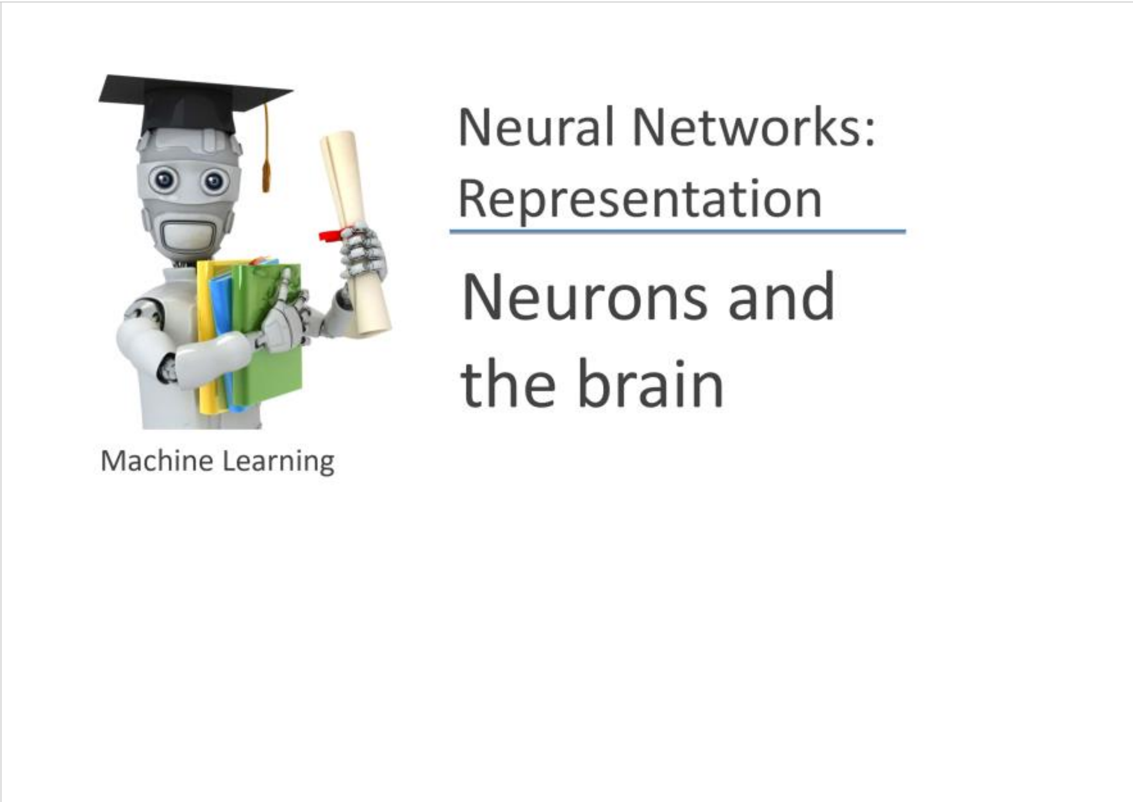


上述是简单的例子
在这个复杂的坐标图中
就可以运用之前ppt上的增加特征量和参数的方法
形成曲线来分类

同时，这也引出向量和矩阵相关的计算

第9页

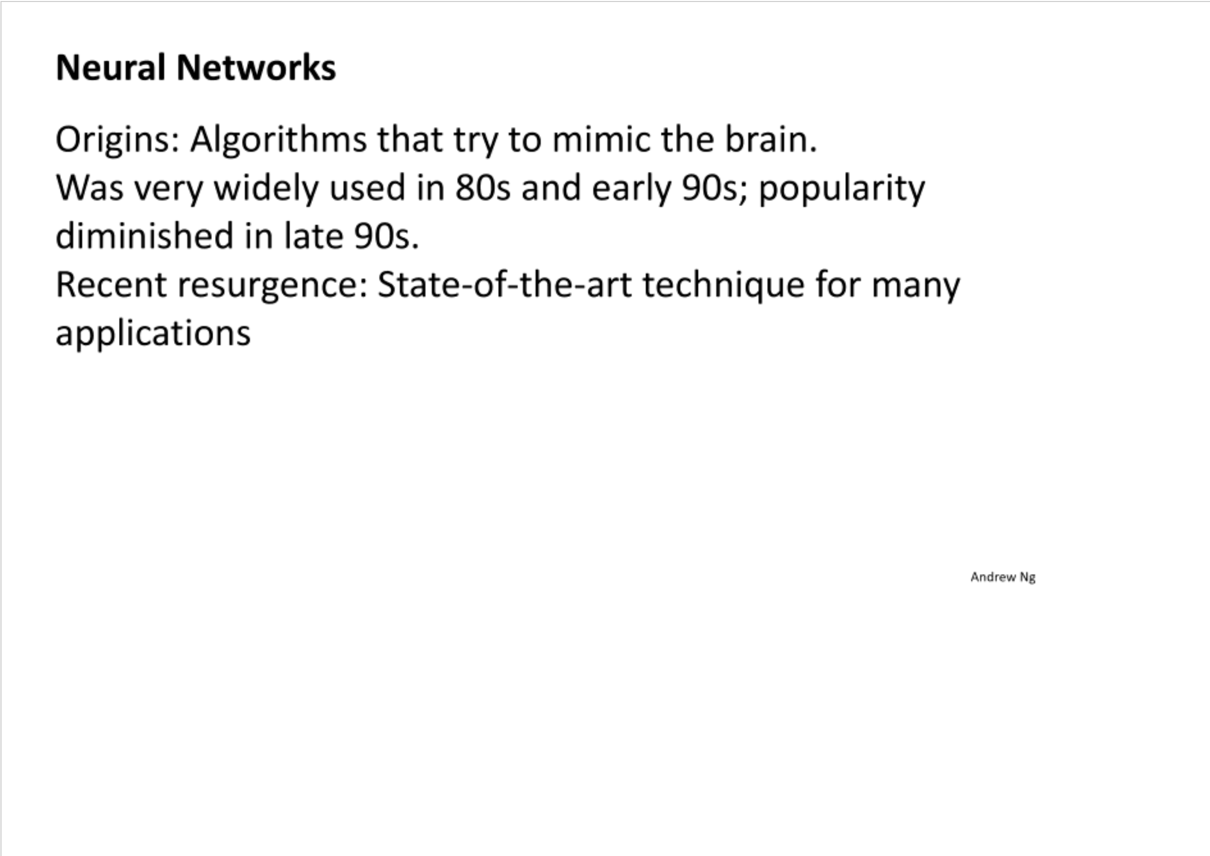
2017年7月31日 18:20



神经网络和大脑
讲解神经网络的来源

第 10 页

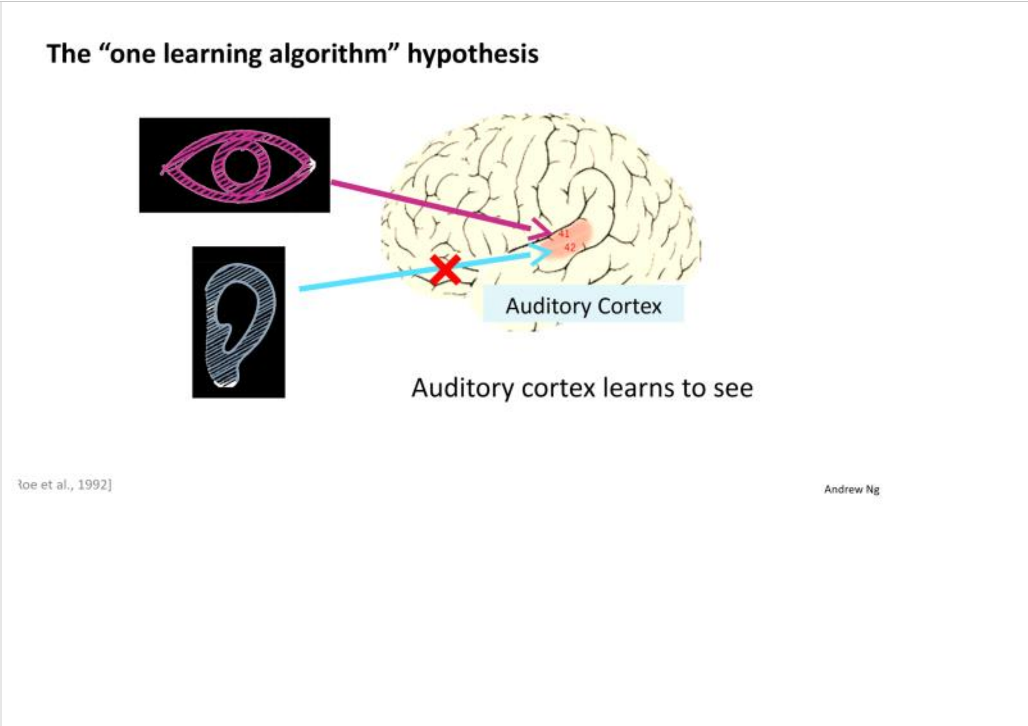
2017年7月31日 18:20



自行翻译吧

第 11 页

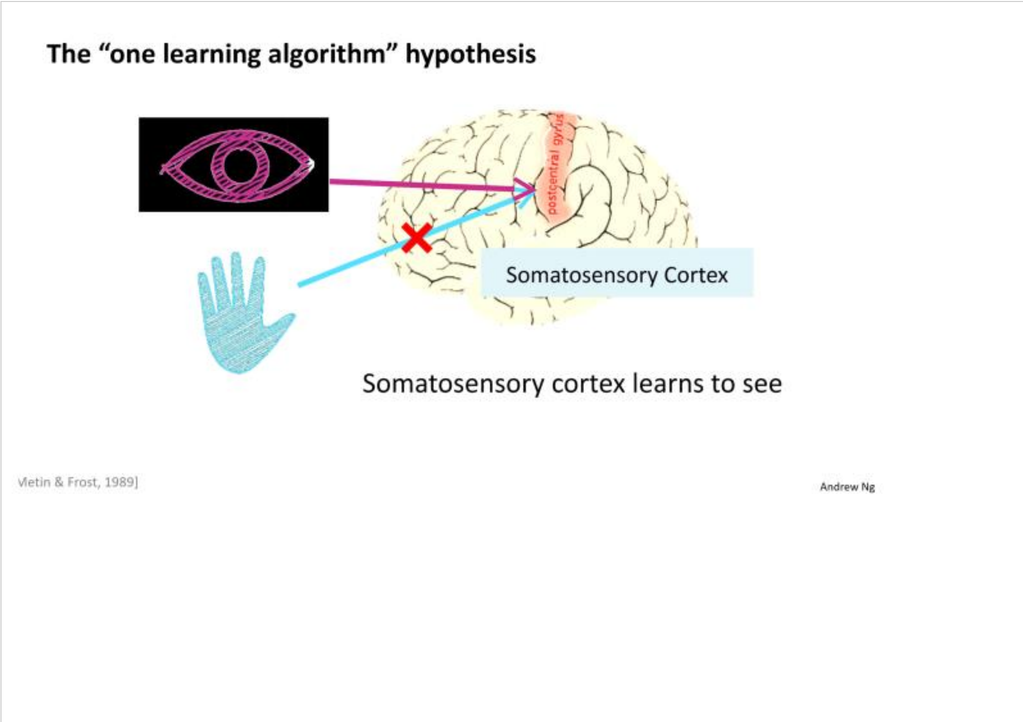
2017年7月31日 18:20



这里大脑划分了一个区域
这个区域就是大脑对视觉和听觉处理的区域
在这个区域里，大脑学会如何看和如何听
也就是说看懂和听懂

第 12 页

2017年7月31日 18:20

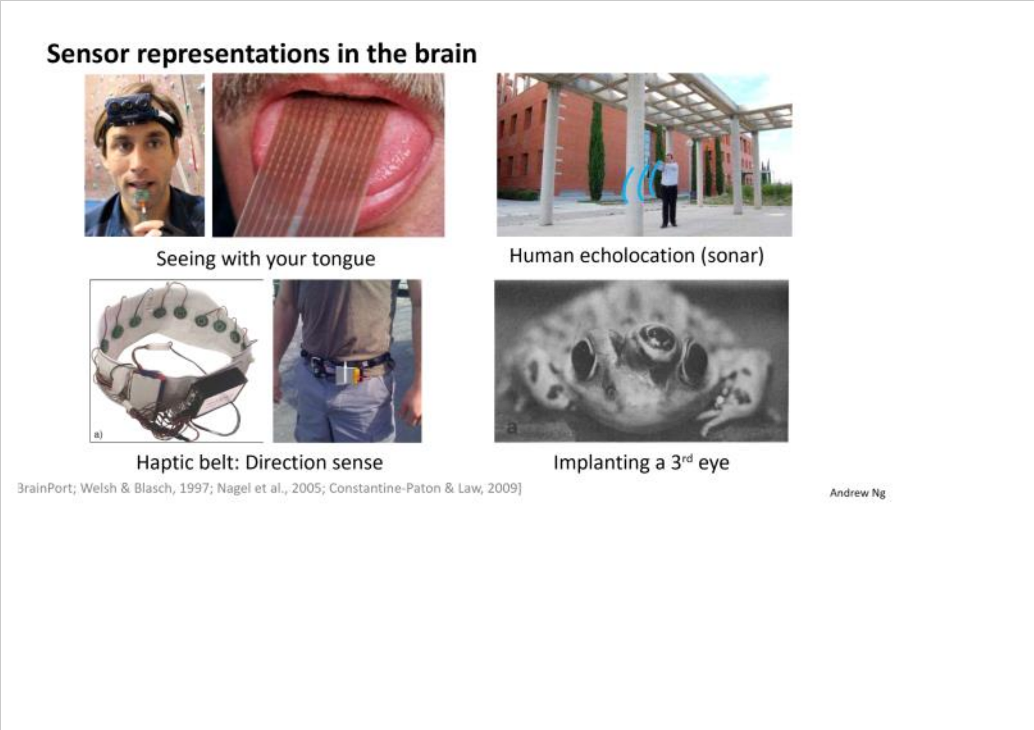


类似上面的ppt

这是处理其他感觉的区域

第13页

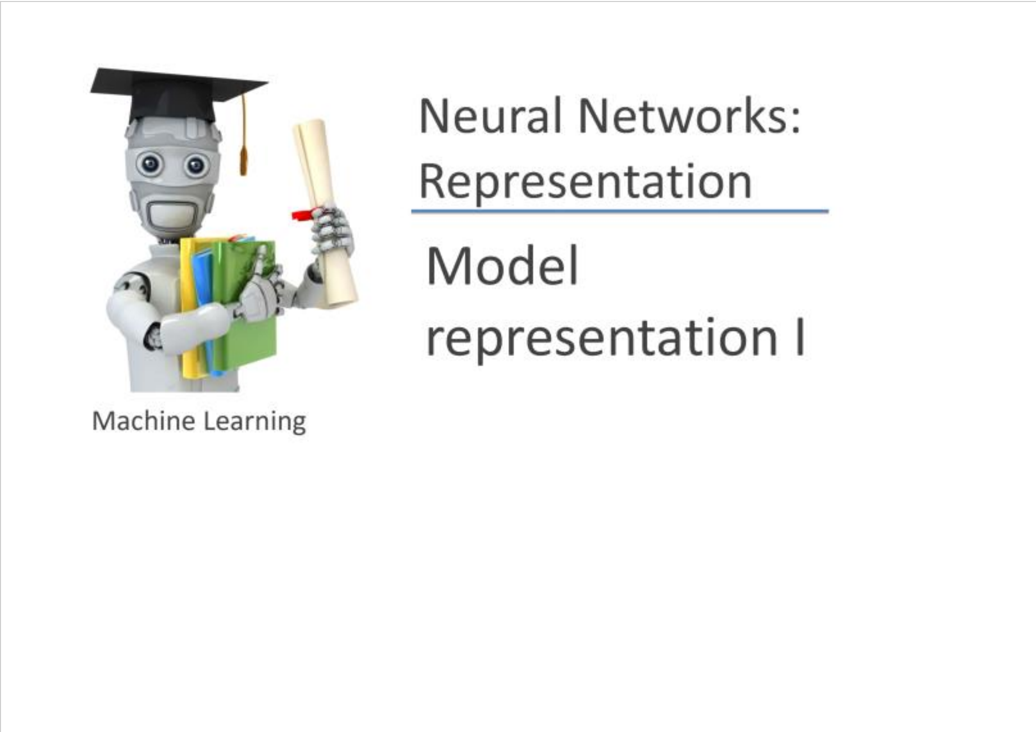
2017年7月31日 18:20



这里举了处理各个感觉的相关例子作为扩展

第15页

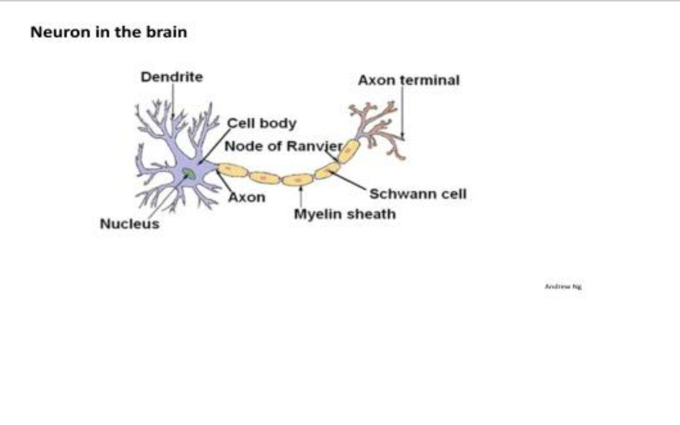
2017年7月31日 18:20



这里开始讲解神经网络这个模型

第 16 页

2017年7月31日 18:20



这是神经元核心结构，同时也抽象化为

输入，输出和隐藏层

输出/轴突（output/Axon）
处理单元/神经核（processing unit/ Nucleus）
输入/树突（input/Dendrite）

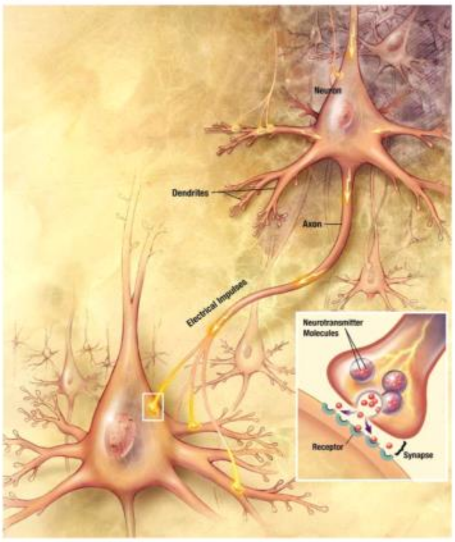
有这个三个功能即可，开始参数化，进行一些基本的运算

疑问：到底如何计算？
计算的法则是什么？

第 17 页

2017年7月31日 18:20

Neurons in the brain



[Credit: US National Institutes of Health, National Institute on Aging]

Andrew Ng

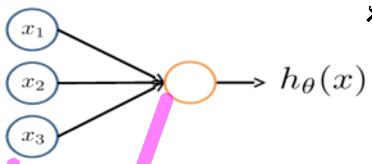
第 18 页

2017年7月31日 18:20

Neuron model: Logistic unit

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

特征量　参量



Sigmoid (logistic) activation function.

Andrew Ng

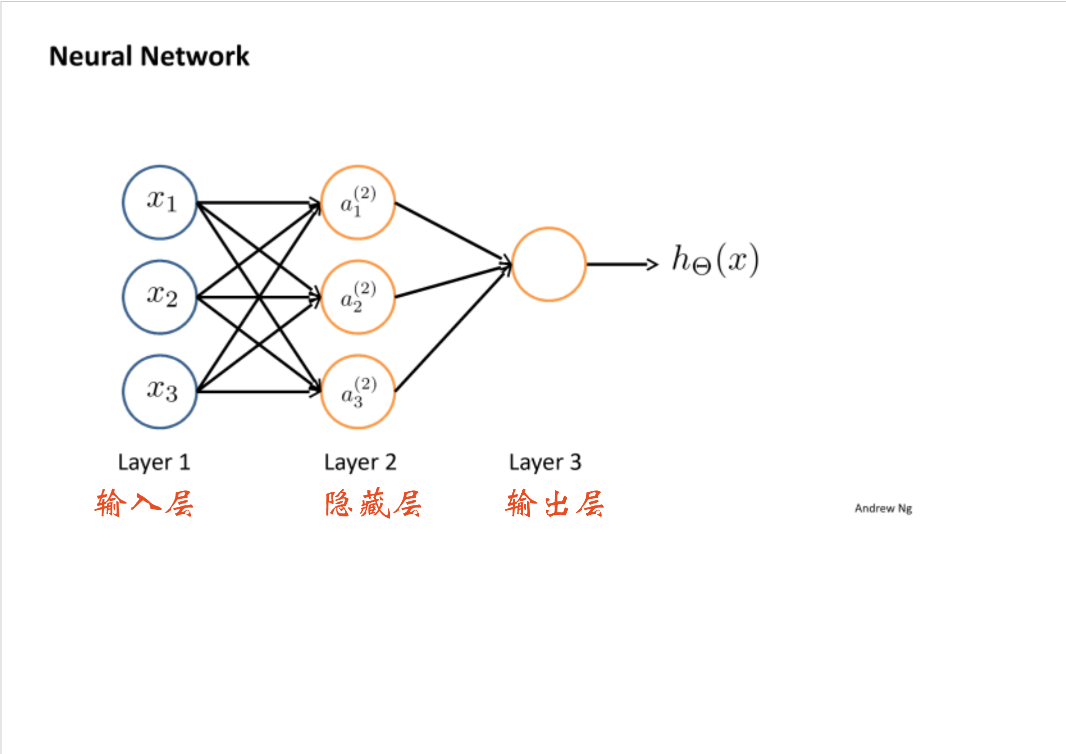
接下就给出具体如何计算，通过神经元

第一层，X1,X2,X3
第二层，就是输出层，输出我们想要的结果

我们从图中可以列出特征量和参数量
X0和theta0是常量。后面会讲到作用

第 19 页

2017年7月31日 18:20



这里有三层

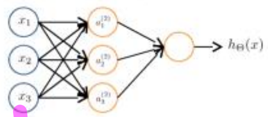
输入层，隐藏层，输出层

$$x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

第 20 页

2017年7月31日 18:20

Neural Network



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

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

$$a_1^{(2)} = g(\Theta_{10}^{(1)}x_0 + \Theta_{11}^{(1)}x_1 + \Theta_{12}^{(1)}x_2 + \Theta_{13}^{(1)}x_3)$$
$$a_2^{(2)} = g(\Theta_{20}^{(1)}x_0 + \Theta_{21}^{(1)}x_1 + \Theta_{22}^{(1)}x_2 + \Theta_{23}^{(1)}x_3)$$
$$a_3^{(2)} = g(\Theta_{30}^{(1)}x_0 + \Theta_{31}^{(1)}x_1 + \Theta_{32}^{(1)}x_2 + \Theta_{33}^{(1)}x_3)$$
$$h_\Theta(x) = a_1^{(3)} = g(\Theta_{10}^{(2)}a_0^{(2)} + \Theta_{11}^{(2)}a_1^{(2)} + \Theta_{12}^{(2)}a_2^{(2)} + \Theta_{13}^{(2)}a_3^{(2)})$$

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

Andrew Ng

$$x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \Theta^1 = \begin{bmatrix} \Theta_{10}^{(1)} & \Theta_{11}^{(1)} & \Theta_{12}^{(1)} & \Theta_{13}^{(1)} \\ \Theta_{20}^{(1)} & \Theta_{21}^{(1)} & \Theta_{22}^{(1)} & \Theta_{23}^{(1)} \\ \Theta_{30}^{(1)} & \Theta_{31}^{(1)} & \Theta_{32}^{(1)} & \Theta_{33}^{(1)} \end{bmatrix} \qquad a^2 = \Theta^1 * x$$

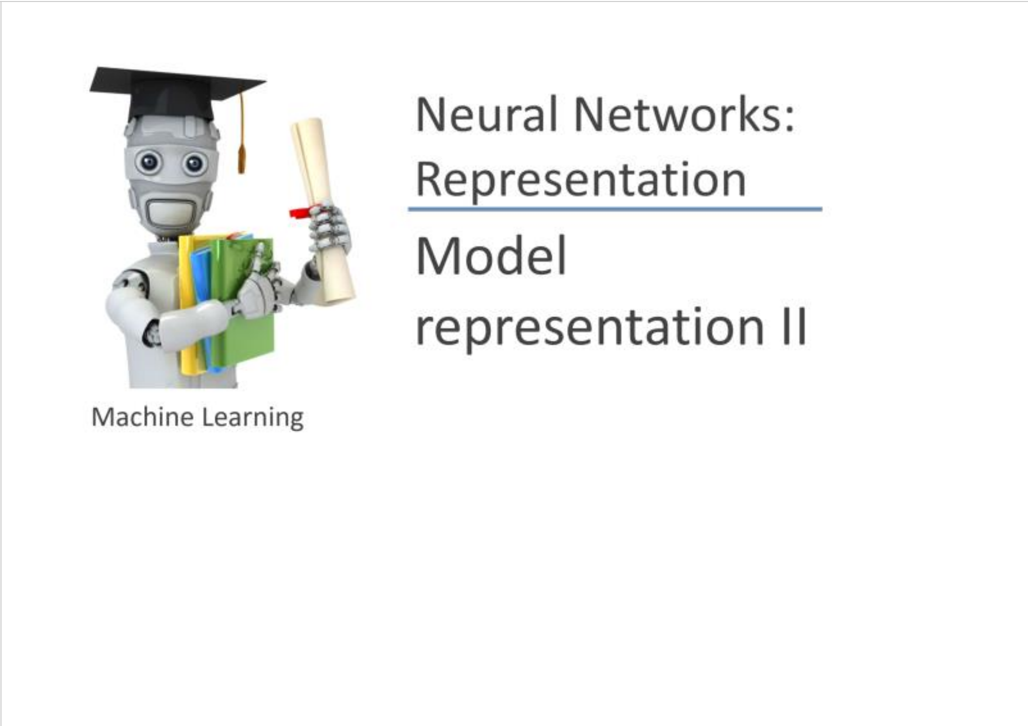
根据之前有个简单的列举特征量和参量

我们可以先列出变量矩阵，然后再公式化，便于理解

类似的第二层的列举，可以列出第三层，最终得到最后一个式子

第 22 页

2017年7月31日 18:20

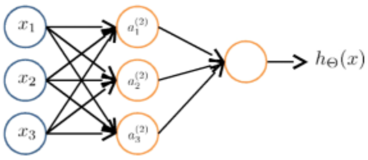


有了上一个章节的基本知识
上一章节主要说明，如何从
神经网络提取变量
这一章节，就对提取的变量
进行计算

第 23 页

2017年7月31日 18:20

Forward propagation: Vectorized implementation



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

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

$$\Theta^1 = \begin{bmatrix} \Theta_{10}^{(1)} & \Theta_{11}^{(1)} & \Theta_{12}^{(1)} & \Theta_{13}^{(1)} \\ \Theta_{20}^{(1)} & \Theta_{21}^{(1)} & \Theta_{22}^{(1)} & \Theta_{23}^{(1)} \\ \Theta_{30}^{(1)} & \Theta_{31}^{(1)} & \Theta_{32}^{(1)} & \Theta_{33}^{(1)} \end{bmatrix}$$

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

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

此处为简化公式
引出变量Z

Add $a_0^{(2)} = 1$.

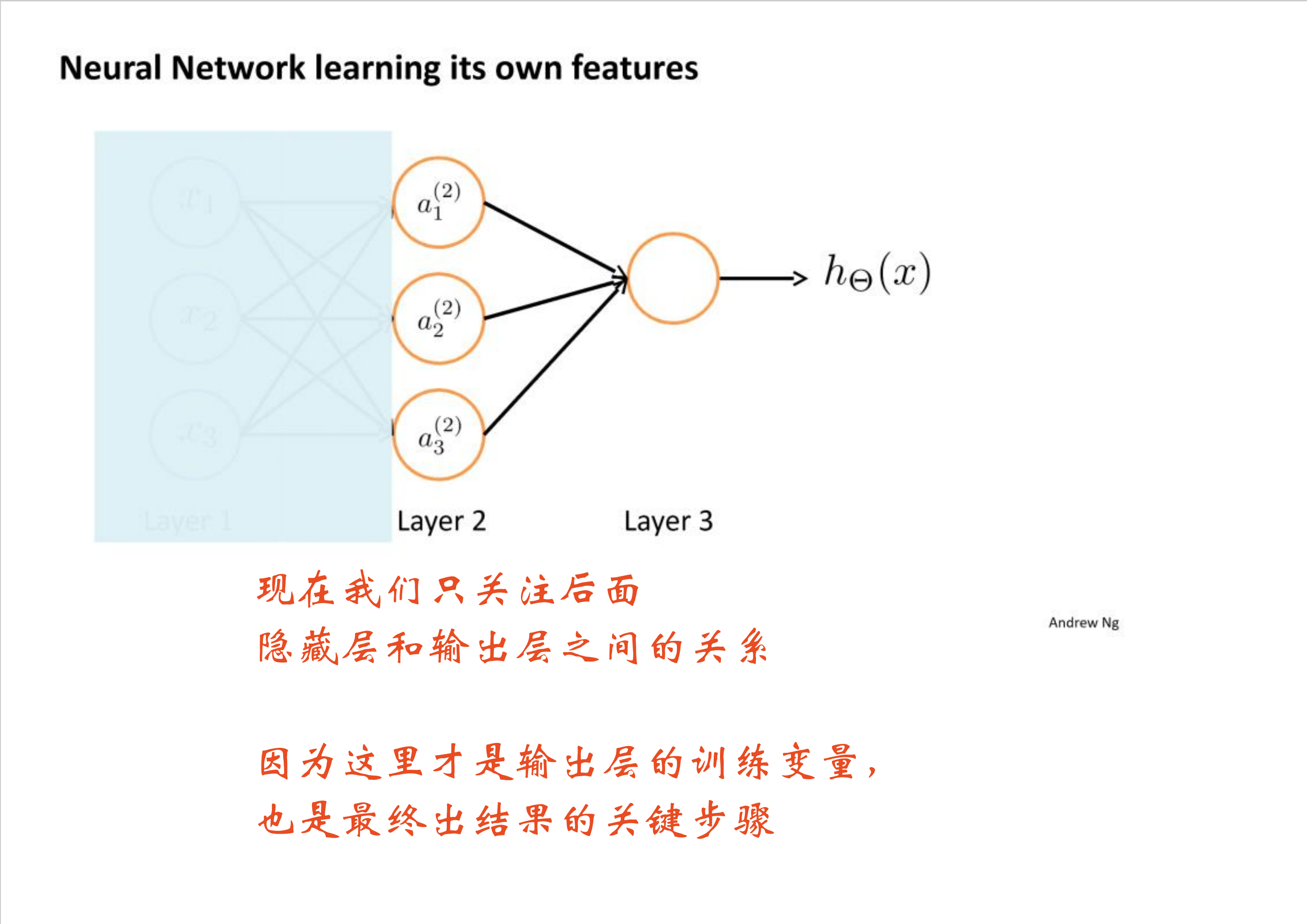
$$z^{(3)} = \Theta^{(2)}a^{(2)}$$
$$h_\Theta(x) = a^{(3)} = g(z^{(3)})$$

新的公式
换汤不换药

Andrew Ng

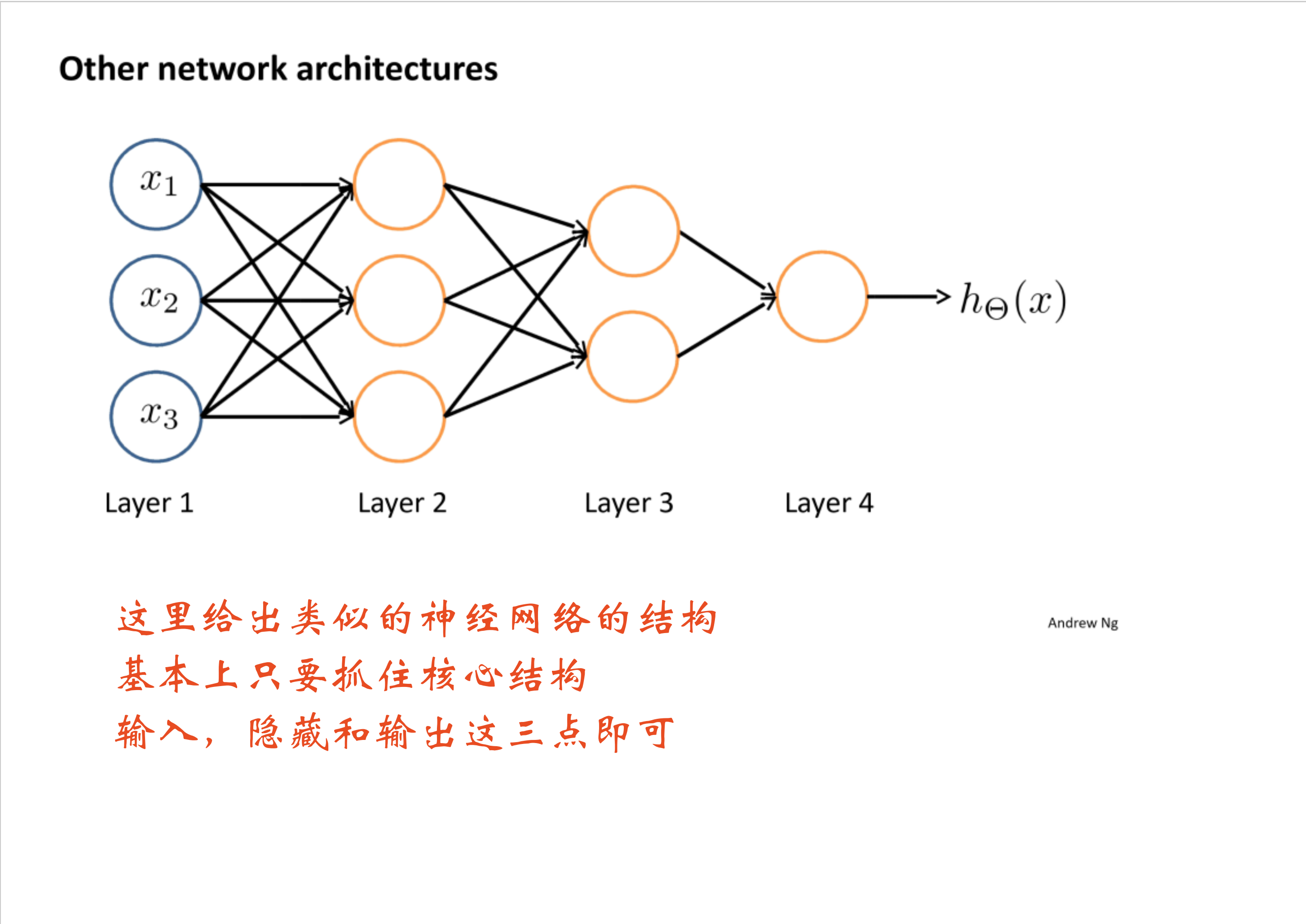
第 24 页

2017年7月31日 18:20



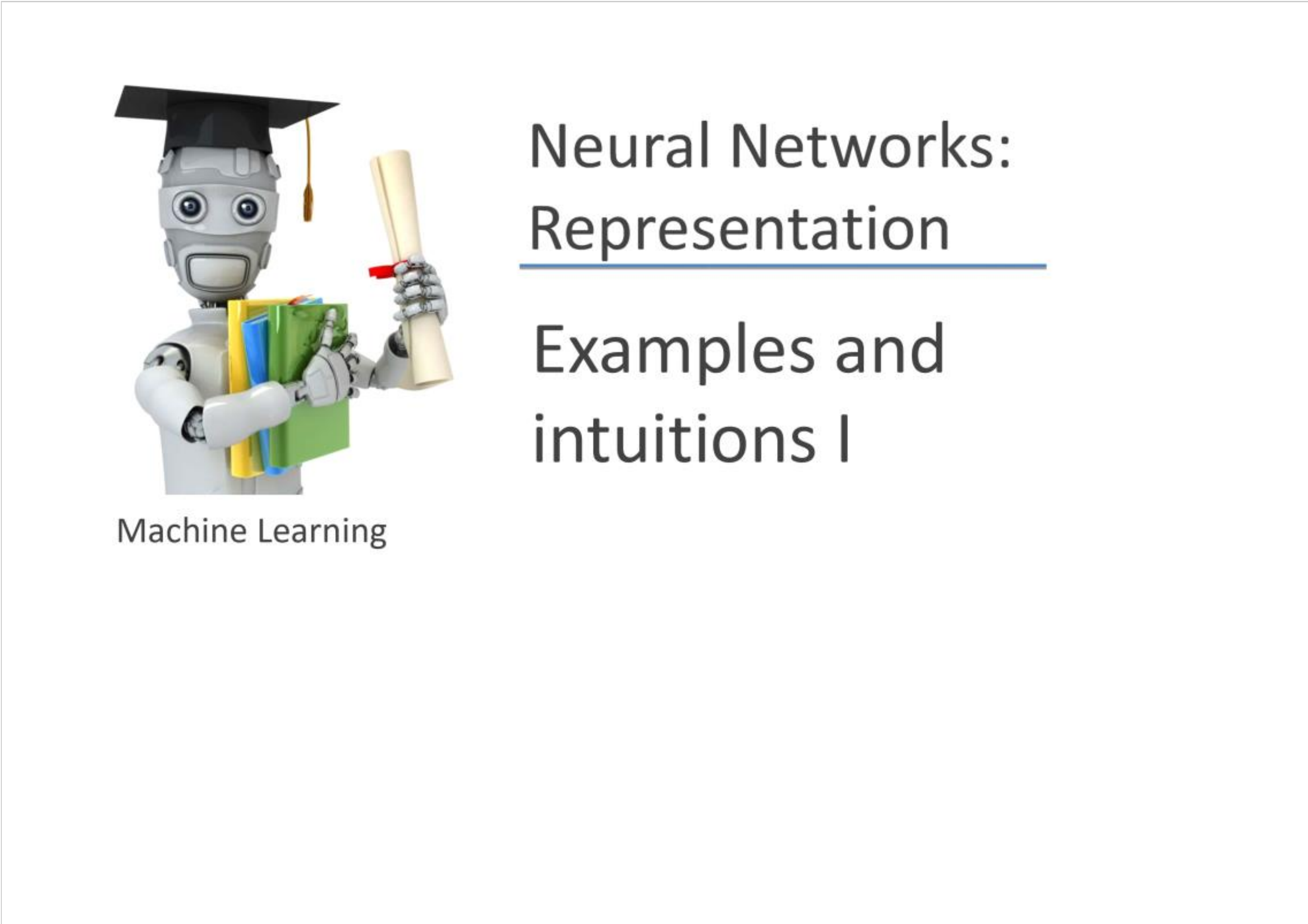
第 25 页

2017年7月31日 18:20



第 27 页

2017年7月31日 18:20

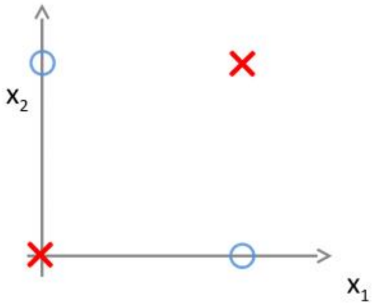


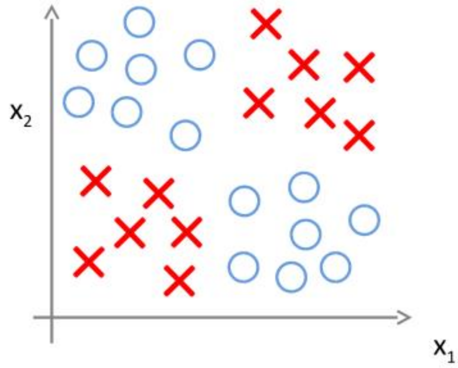
第 28 页

2017年7月31日 18:20

Non-linear classification example: XOR/XNOR

x_1, x_2 are binary (0 or 1).





$$
\begin{aligned}
y = {} & x_1 \text{ XOR } x_2 \\
& x_1 \text{ XNOR } x_2 \\
& \text{NOT } (x_1 \text{ XOR } x_2)
\end{aligned}
$$

Andrew Ng

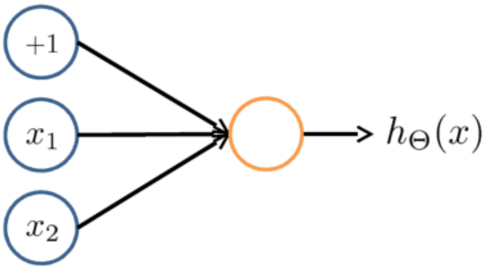
第 29 页

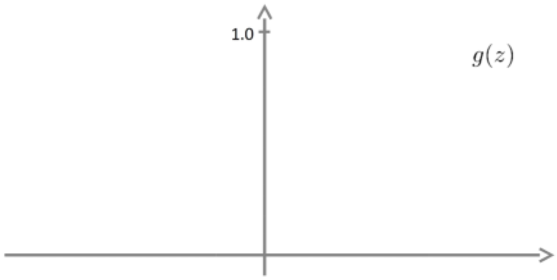
2017年7月31日 18:20

Simple example: AND

$x_1, x_2 \in \{0, 1\}$

$y = x_1 \text{ AND } x_2$





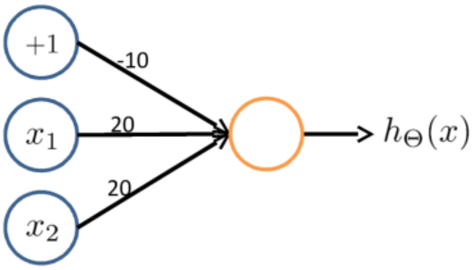
x_1	x_2	$h_\Theta(x)$
0	0	
0	1	
1	0	
1	1	

Andrew Ng

第30页

2017年7月31日 18:20

Example: OR function



x_1	x_2	$h_\Theta(x)$
0	0	
0	1	
1	0	
1	1	

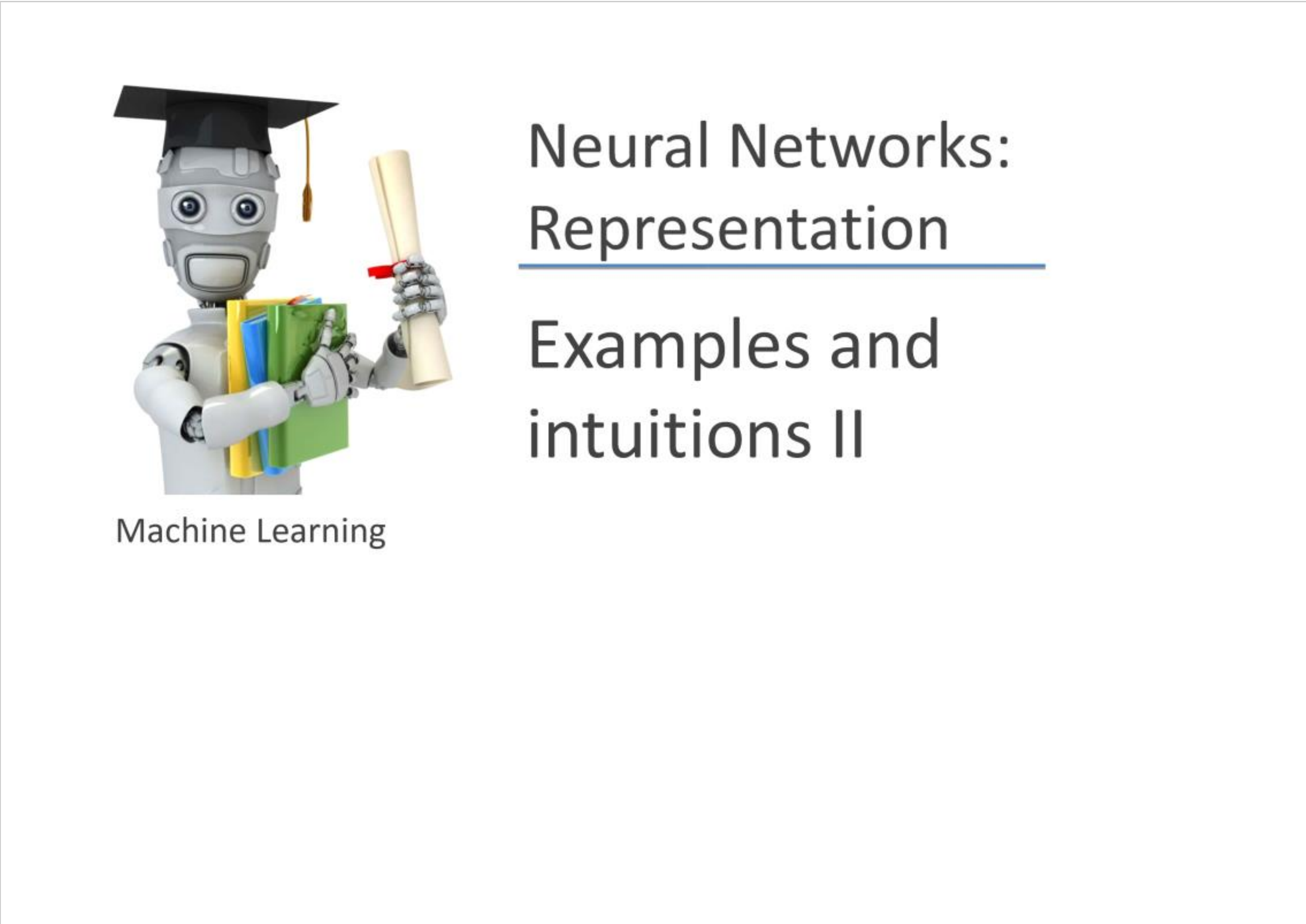
Andrew Ng

第 31 页

2017年7月31日 18:20

第 32 页

2017年7月31日 18:20

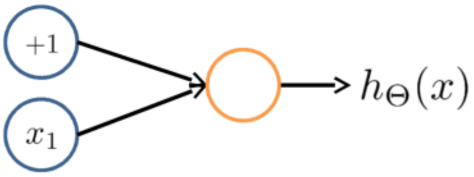


第 33 页

2017年7月31日 18:20

x_1 AND x_2 x_1 OR x_2

Negation:



x_1	$h_\Theta(x)$
0	
1	

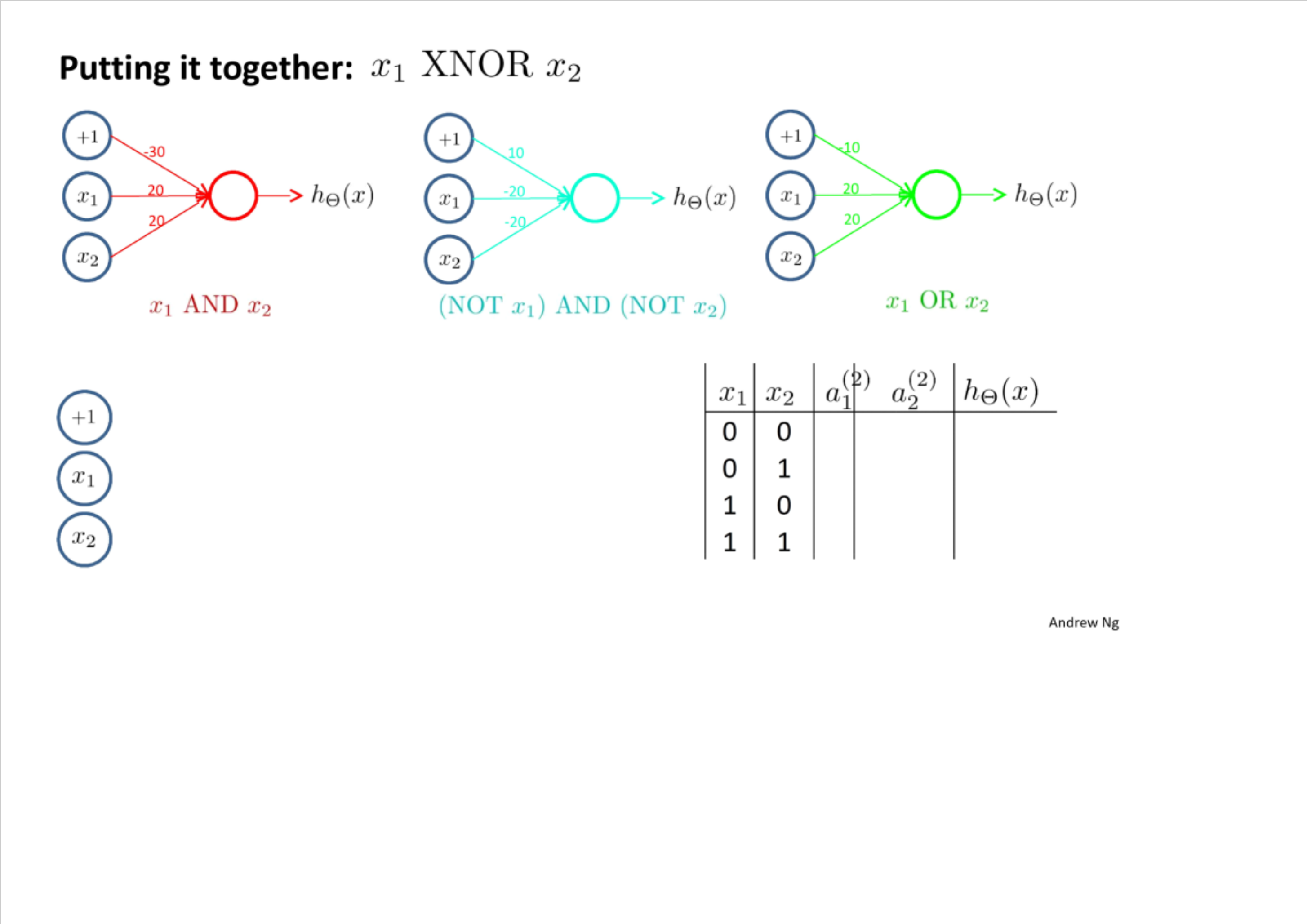
$$h_\Theta(x) = g(10 - 20x_1)$$

(NOT x_1) AND (NOT x_2)

Andrew Ng

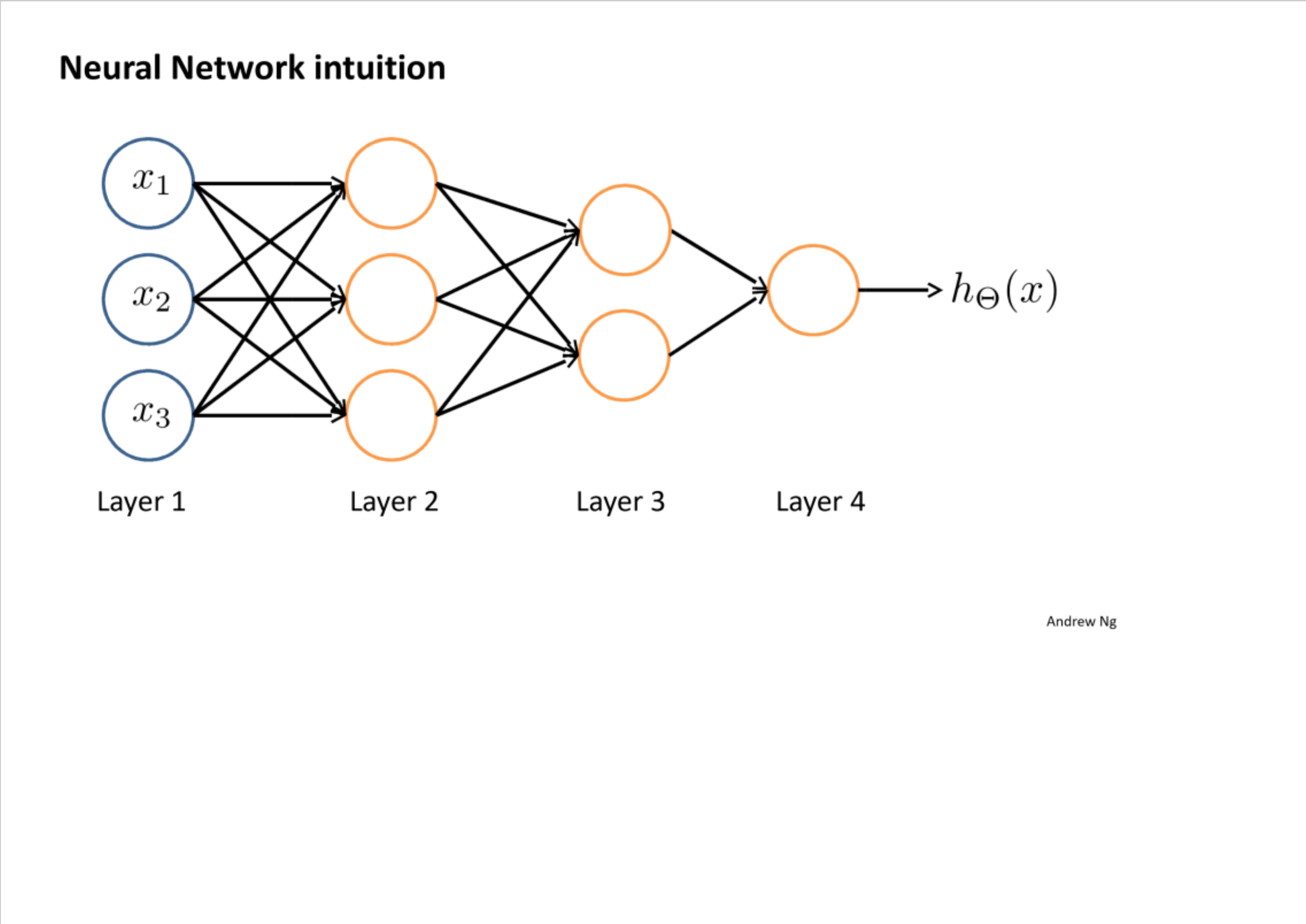
第 34 页

2017年7月31日 18:20



第 35 页

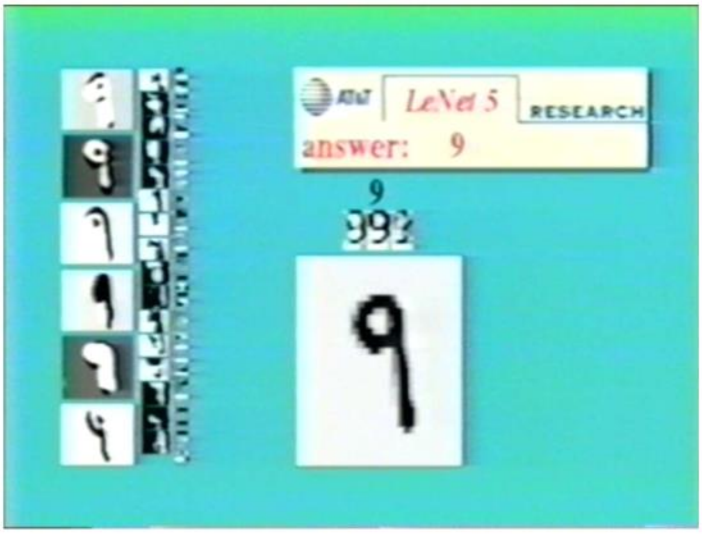
2017年7月31日 18:20



第 36 页

2017年7月31日 18:20

Handwritten digit classification



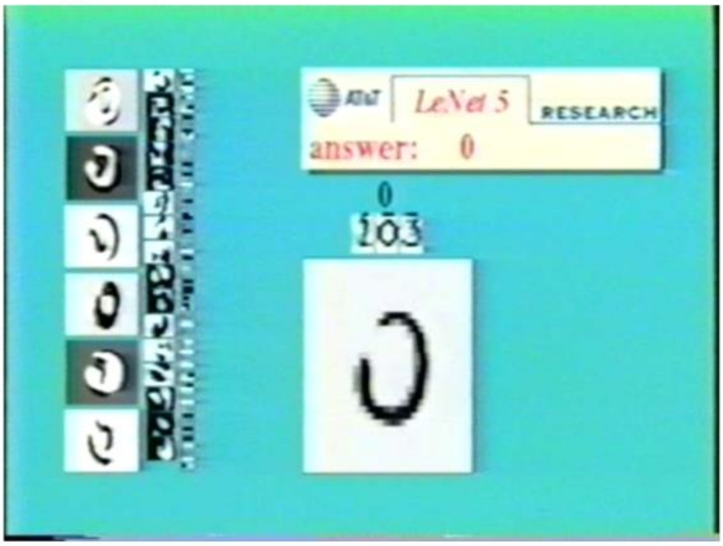
[Courtesy of Yann LeCun]

Andrew Ng

第 37 页

2017年7月31日 18:20

Handwritten digit classification

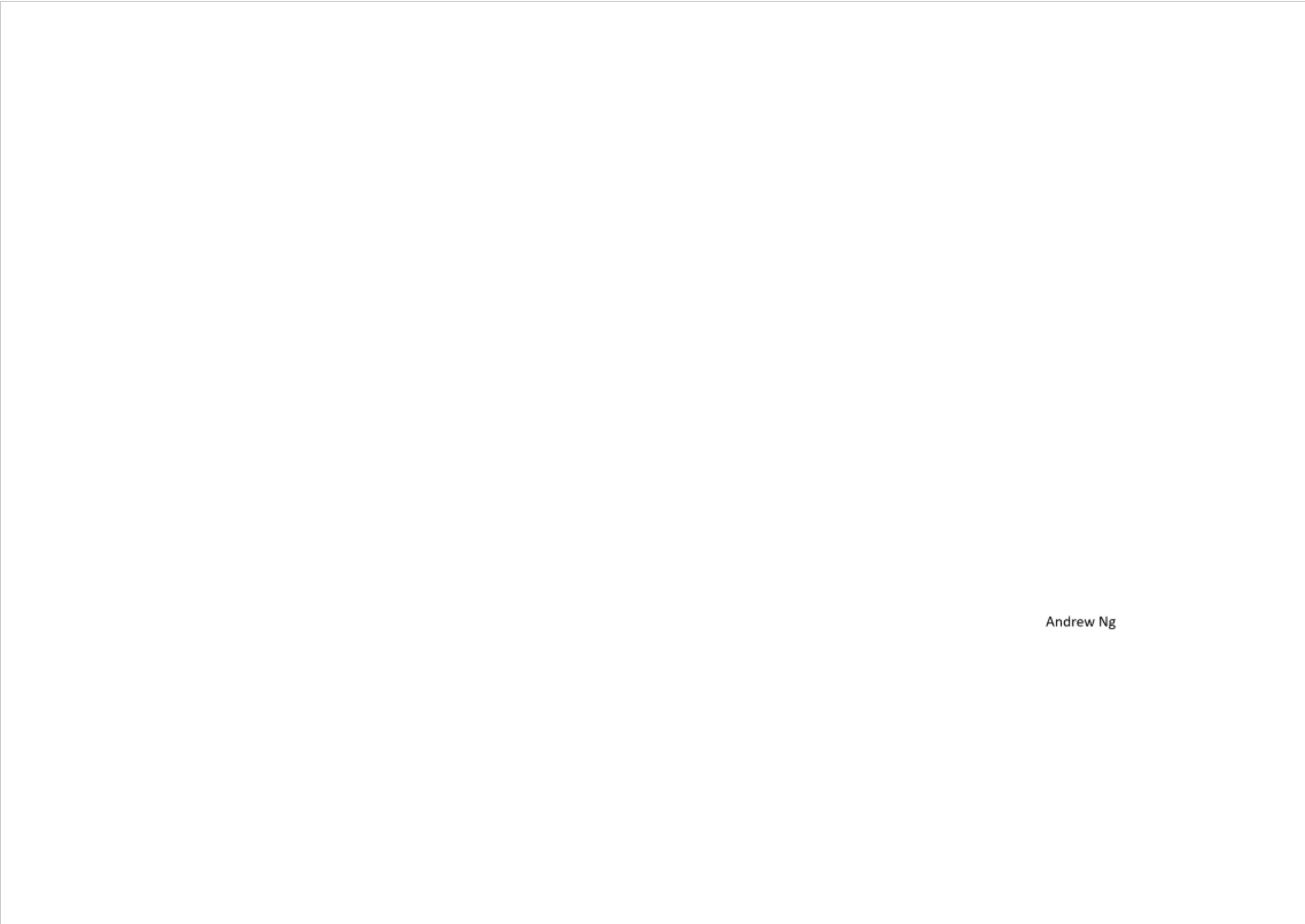


[Courtesy of Yann LeCun]

Andrew Ng

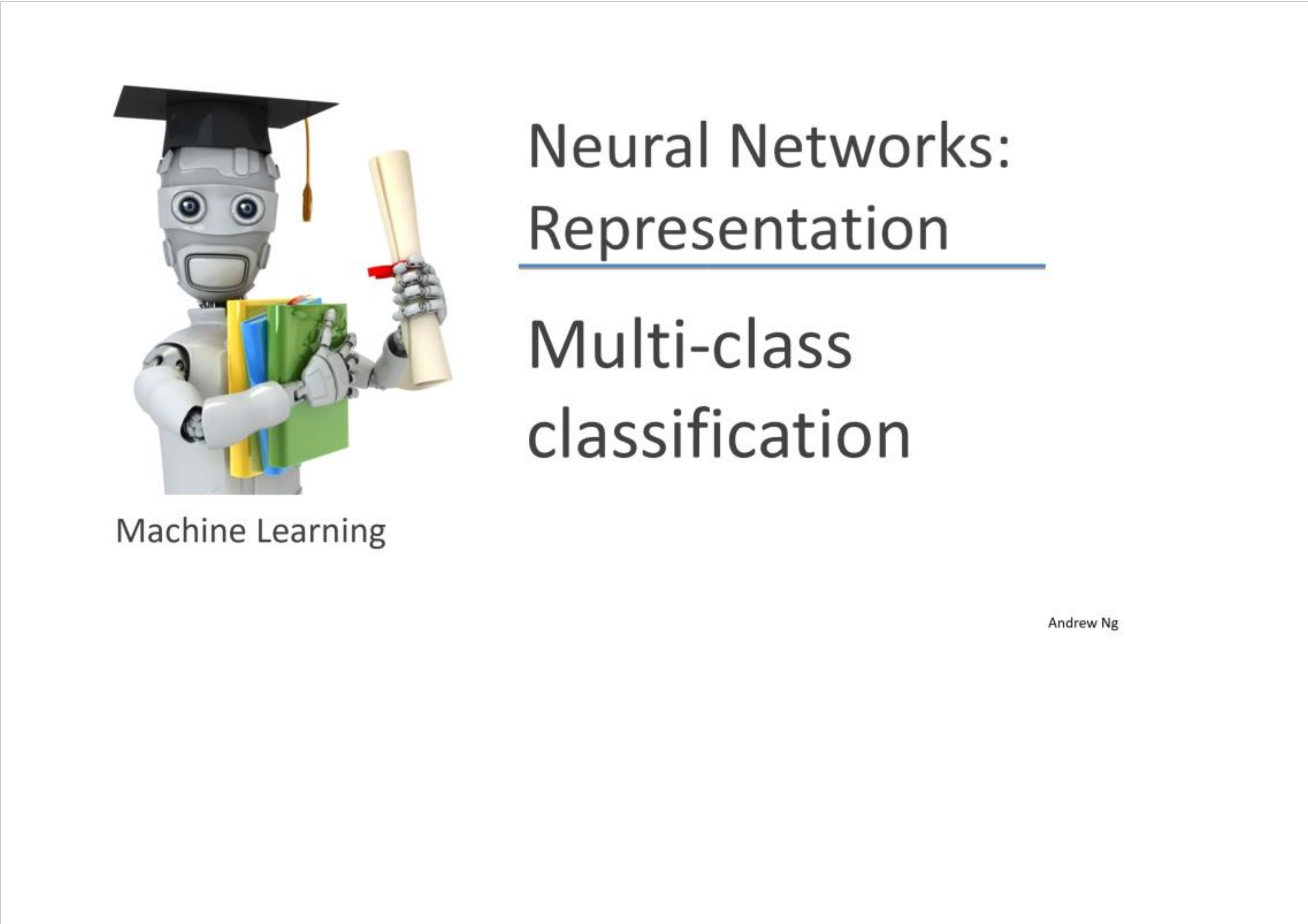
第 38 页

2017年7月31日 18:20



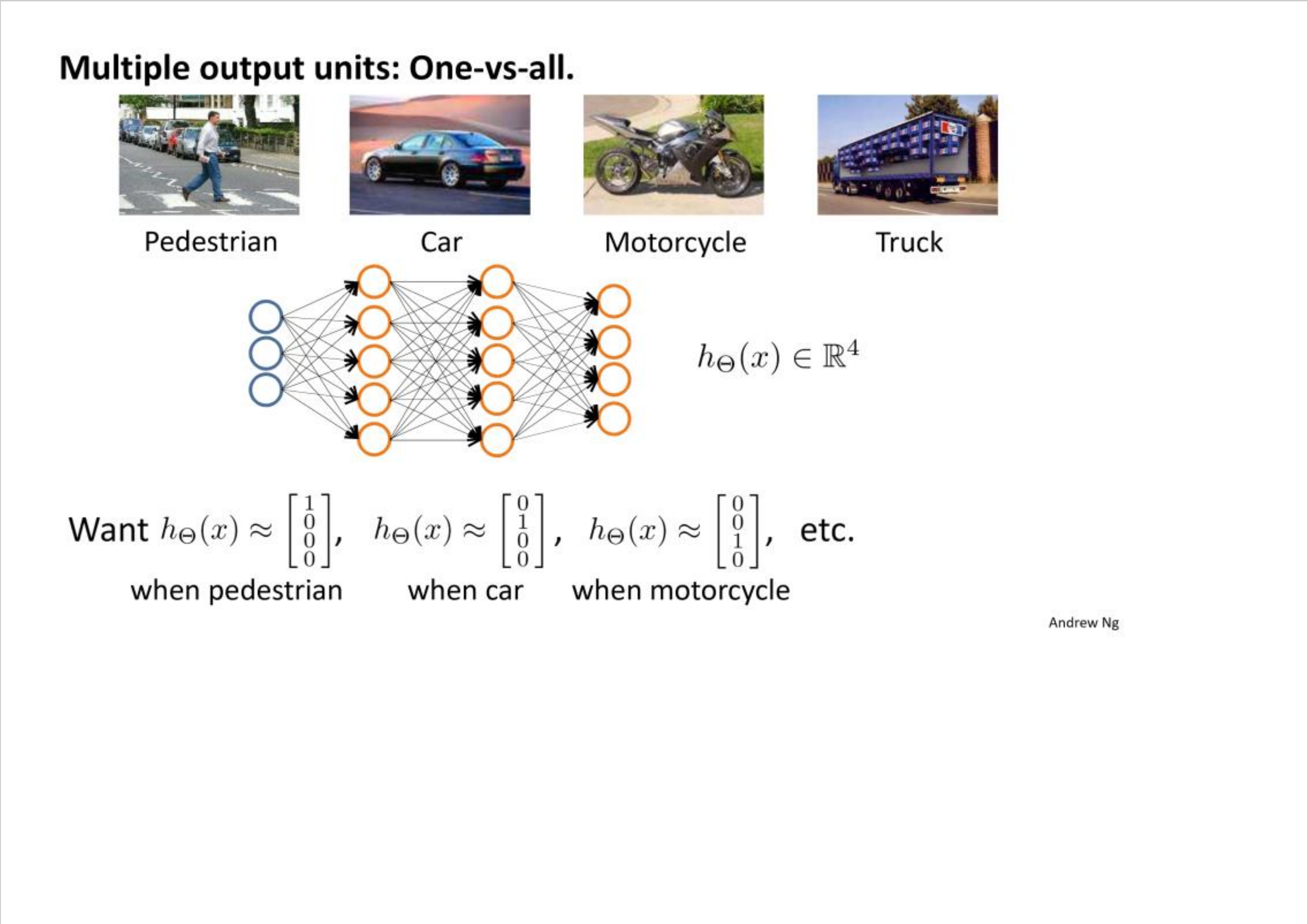
第 39 页

2017年7月31日 18:20



第 40 页

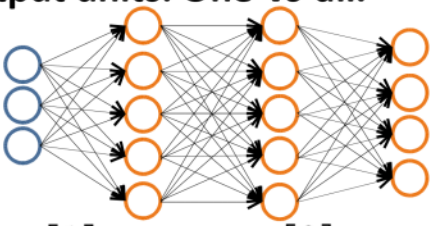
2017年7月31日 18:20



第41页

2017年7月31日　18:20

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

$y^{(i)}$ 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

Andrew Ng

第 42 页

2017年7月31日 18:20