

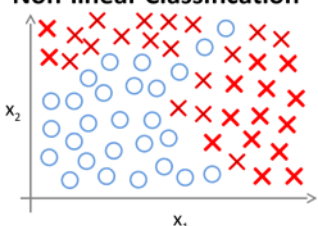


Machine Learning

Neural Networks: Representation Non-linear hypotheses

神经网络入门
非线性假设

Non-linear Classification



x_1 = size
 x_2 = # bedrooms
 x_3 = # floors
 x_4 = age
...
 x_{100}

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 x_2 + \theta_5 x_1^3 x_2 + \theta_6 x_1 x_2^2 + \dots)$$


Andrew Ng

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。

What is this?

You see this:



But the camera sees this:

| | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 194 | 210 | 201 | 212 | 199 | 213 | 215 | 195 | 178 | 158 | 182 | 209 |
| 180 | 189 | 190 | 221 | 209 | 205 | 191 | 167 | 147 | 115 | 129 | 163 |
| 114 | 126 | 140 | 188 | 176 | 165 | 152 | 140 | 170 | 106 | 78 | 88 |
| 87 | 103 | 115 | 154 | 143 | 142 | 149 | 153 | 173 | 101 | 57 | 57 |
| 102 | 112 | 106 | 131 | 122 | 138 | 152 | 147 | 128 | 84 | 58 | 66 |
| 94 | 95 | 79 | 104 | 105 | 124 | 129 | 113 | 107 | 87 | 69 | 67 |
| 68 | 71 | 69 | 98 | 89 | 92 | 98 | 95 | 89 | 88 | 76 | 67 |
| 41 | 56 | 69 | 99 | 63 | 45 | 60 | 82 | 58 | 76 | 75 | 65 |
| 20 | 43 | 69 | 75 | 56 | 41 | 51 | 73 | 55 | 70 | 63 | 44 |
| 50 | 50 | 57 | 69 | 75 | 75 | 73 | 74 | 53 | 68 | 59 | 37 |
| 72 | 59 | 53 | 66 | 84 | 92 | 84 | 74 | 57 | 72 | 63 | 42 |
| 67 | 61 | 58 | 65 | 75 | 78 | 76 | 73 | 59 | 75 | 69 | 50 |

Andrew Ng

视觉处理的例子，计算机是处理图像的通过矩阵形式数据

Computer Vision: Car detection



Cars



Not a car

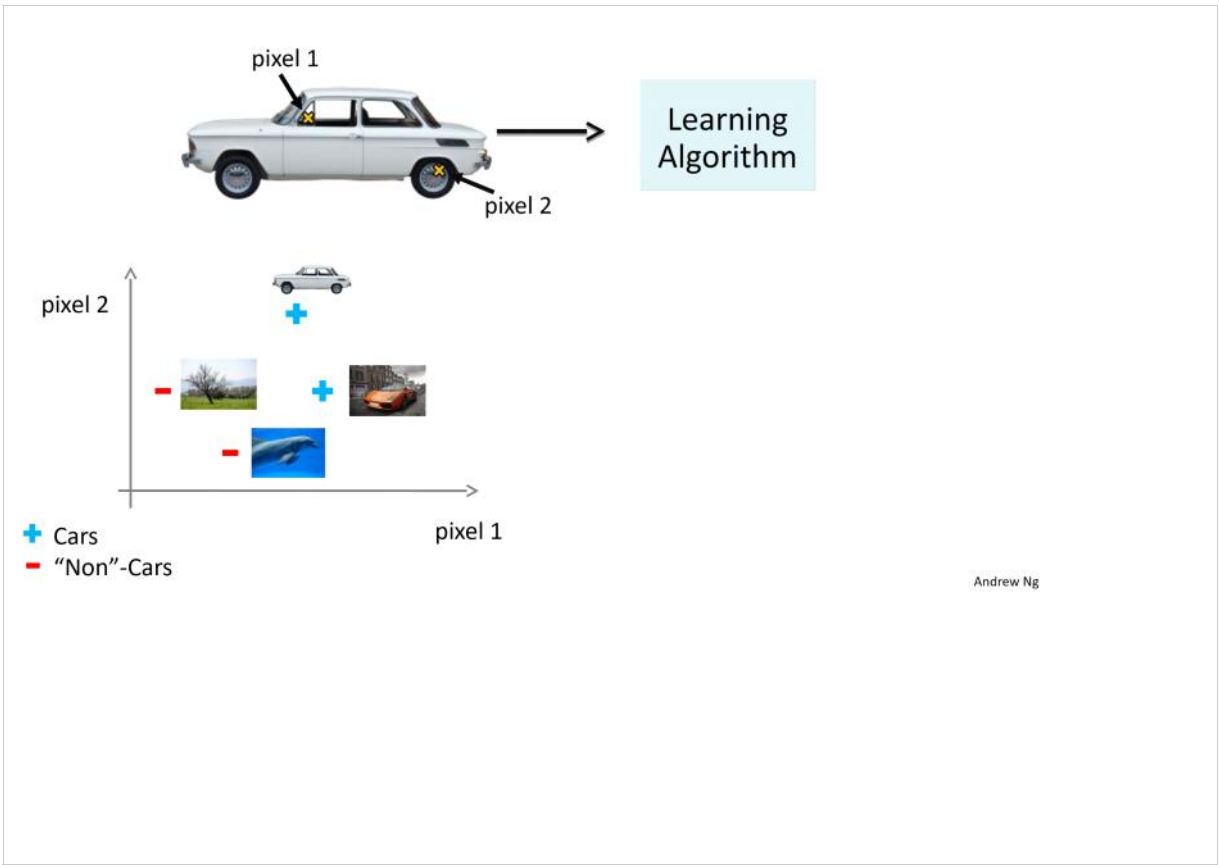
Testing:



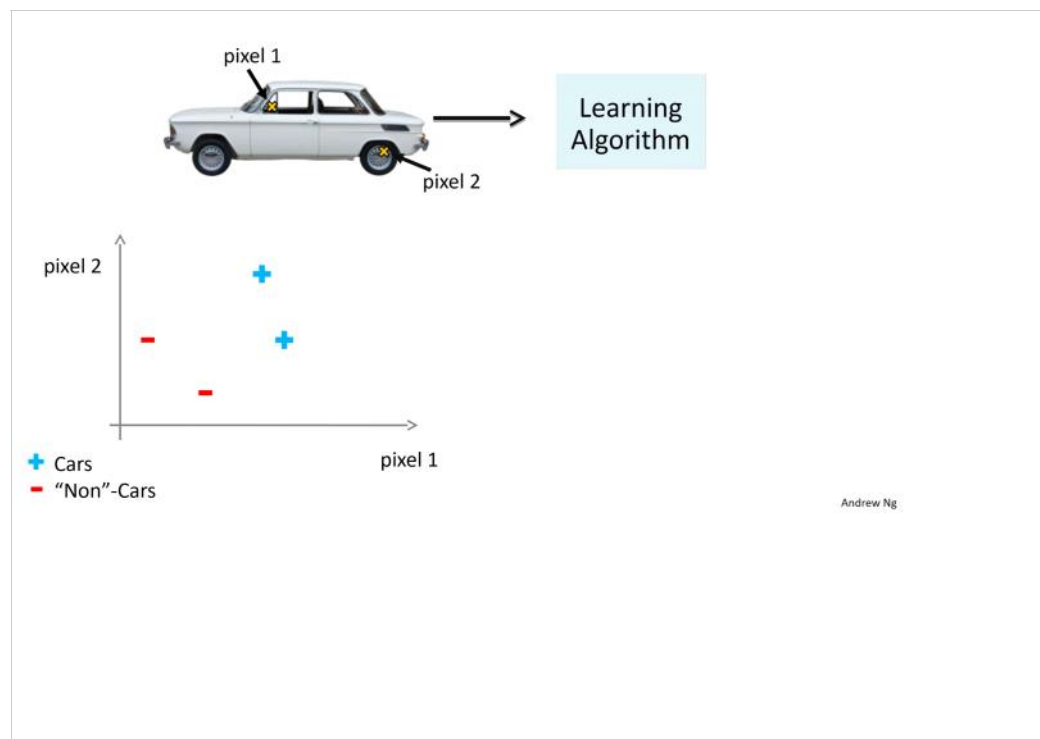
What is this?

Andrew Ng

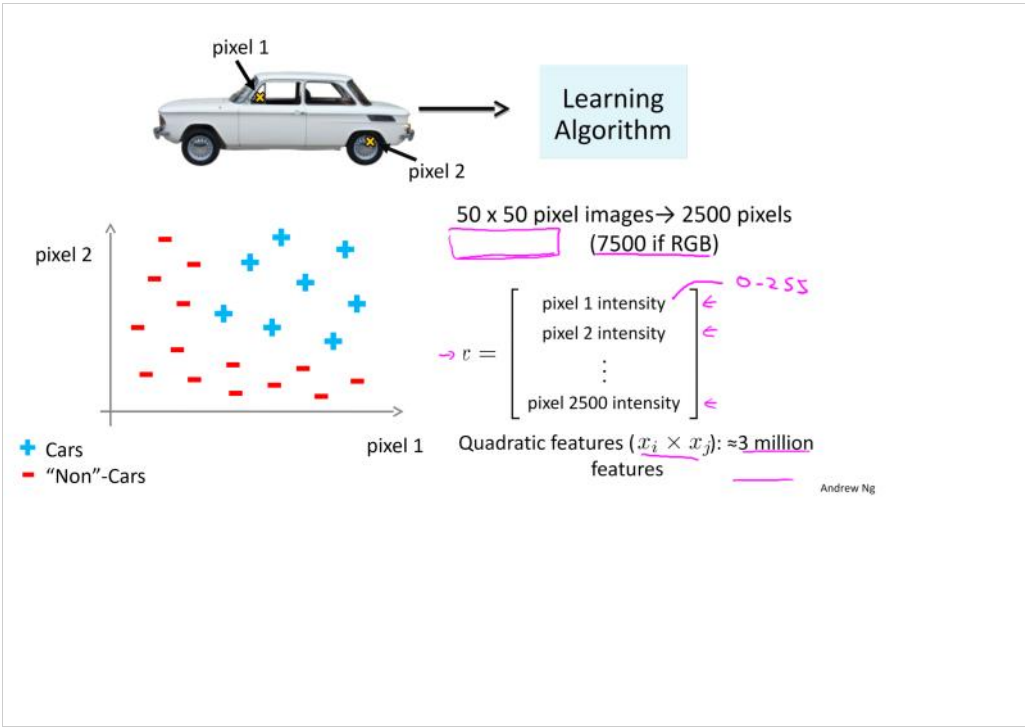
0/1分类



汽车的例子



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



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



Machine Learning

Neural Networks: Representation

Neurons and the brain

神经网络和大脑
讲解神经网络的来源

Neural Networks

Origins: Algorithms that try to mimic the brain.

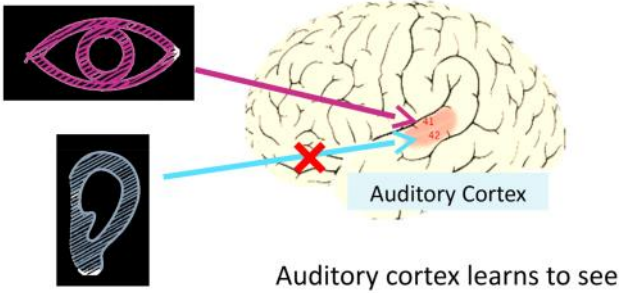
Was very widely used in 80s and early 90s; popularity diminished in late 90s.

Recent resurgence: State-of-the-art technique for many applications

Andrew Ng

自行翻译吧

The “one learning algorithm” hypothesis



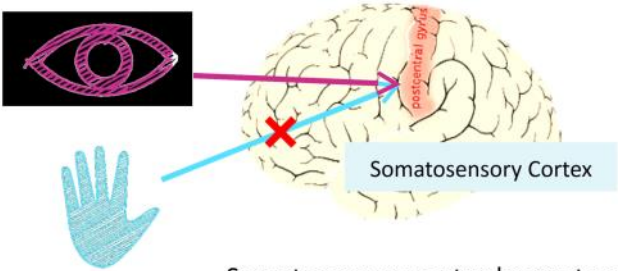
Auditory cortex learns to see

[Joel et al., 1992]

Andrew Ng

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

The “one learning algorithm” hypothesis



Somatosensory cortex learns to see

Vietin & Frost, 1989]

Andrew Ng

类似上面的ppt
这是处理其他感觉的区域

Sensor representations in the brain



Seeing with your tongue



Human echolocation (sonar)



Haptic belt: Direction sense



Implanting a 3rd eye

BrainPort; Welsh & Blasch, 1997; Nagel et al., 2005; Constantine-Paton & Law, 2009]

Andrew Ng

这里举了处理各个感觉的
相关例子作为扩展



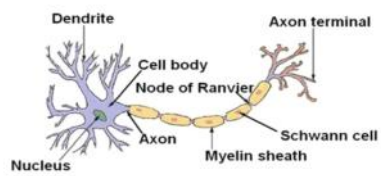
Machine Learning

Neural Networks: Representation

Model representation I

这里开始讲解神经网络这个模型

Neuron in the brain



Andrew Ng

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)

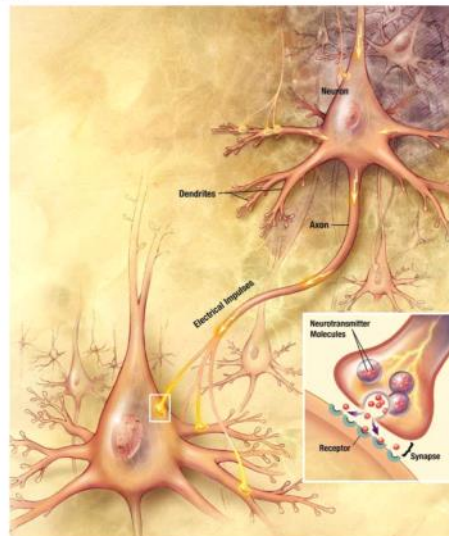
输入/树突 (input/Dendrite)

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？

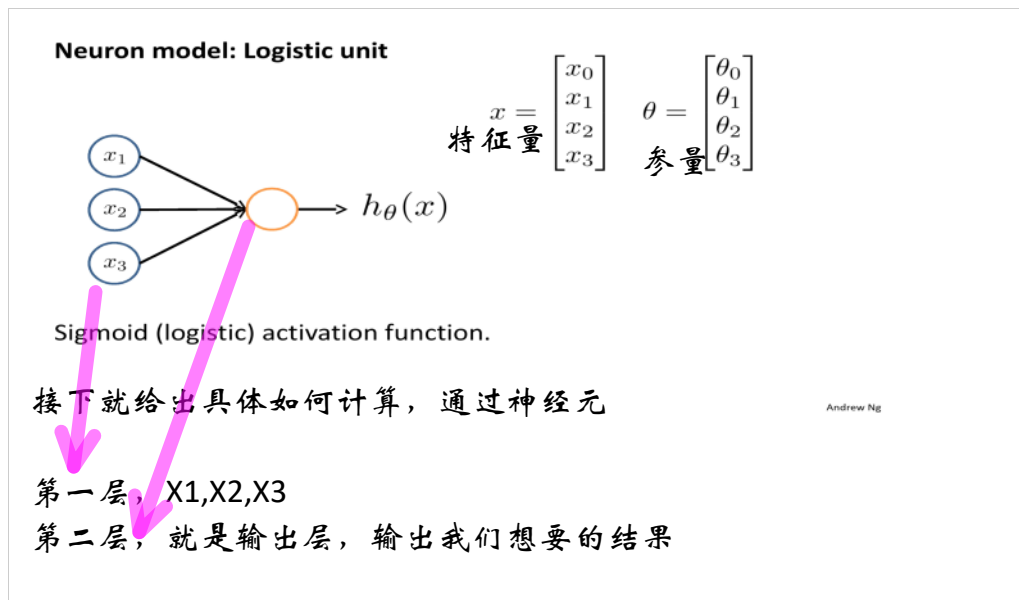
计算的法则是什么？

Neurons in the brain



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

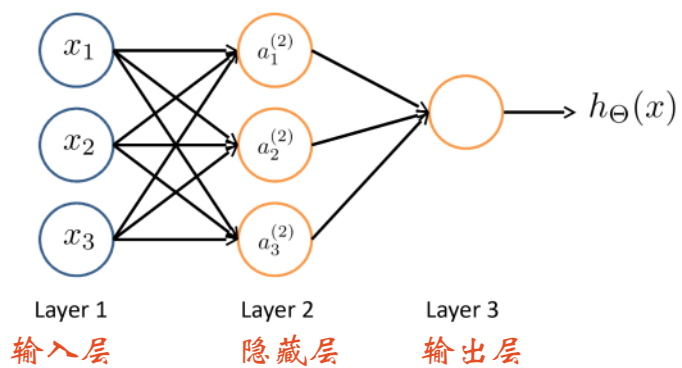
Andrew Ng



我们从图中可以列出特征量和参数量

x_0 和 θ_0 是常量。后面会讲到作用

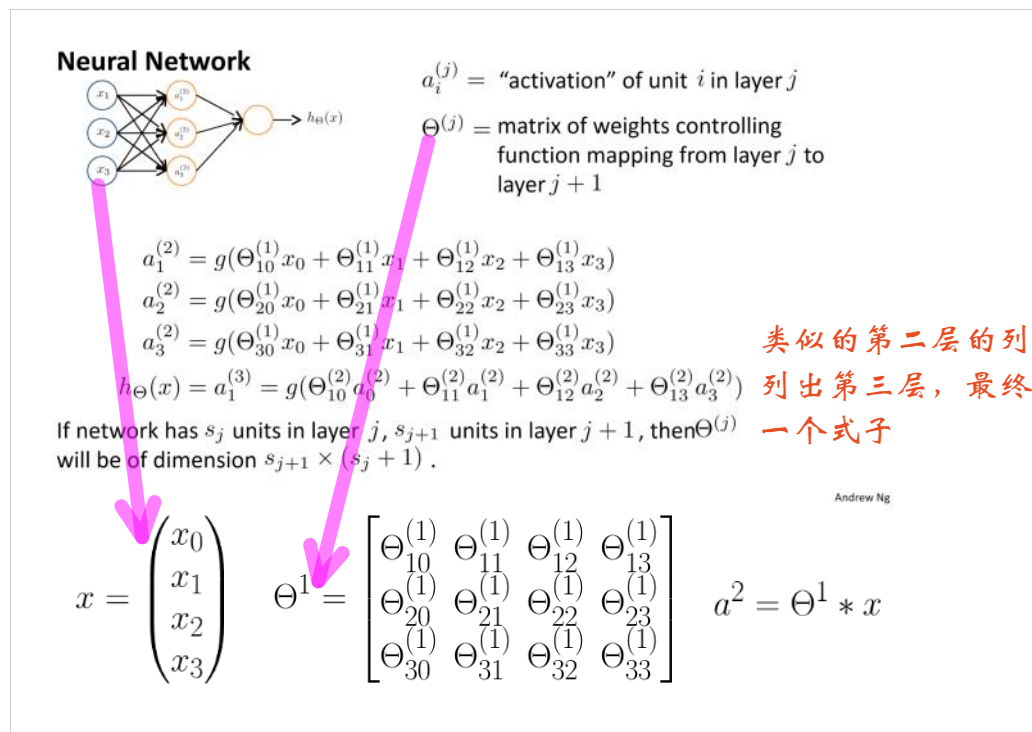
Neural Network



Andrew Ng

这里有三层
输入层，隐藏层，输出层

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



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



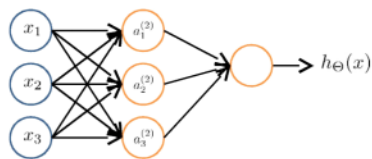
Machine Learning

Neural Networks: Representation

Model representation II

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

Forward propagation: Vectorized implementation



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

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

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

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

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

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

$$z^{(2)} = \Theta^{(1)} x \quad \text{此处为简化公式}$$

$$a^{(2)} = g(z^{(2)}) \quad \text{引出变量Z}$$

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

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

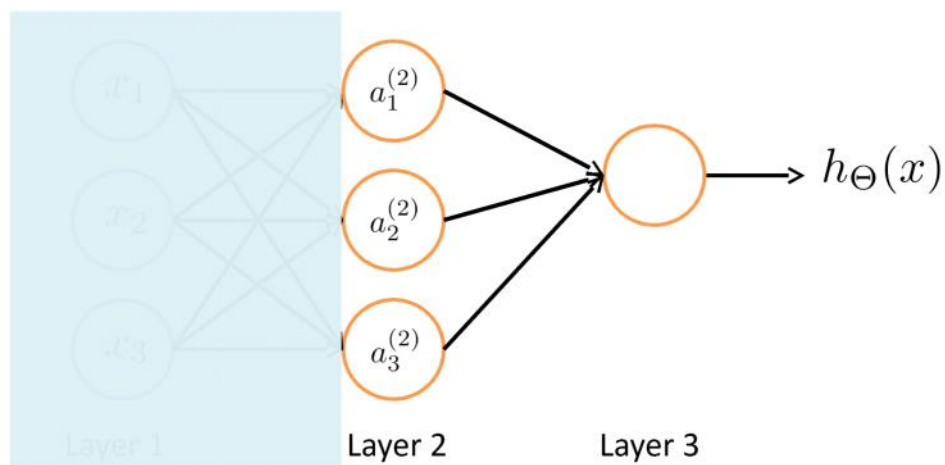
$$h_{\Theta}(x) = a^{(3)} = g(z^{(3)})$$

新的公式

换汤不换药

Andrew Ng

Neural Network learning its own features

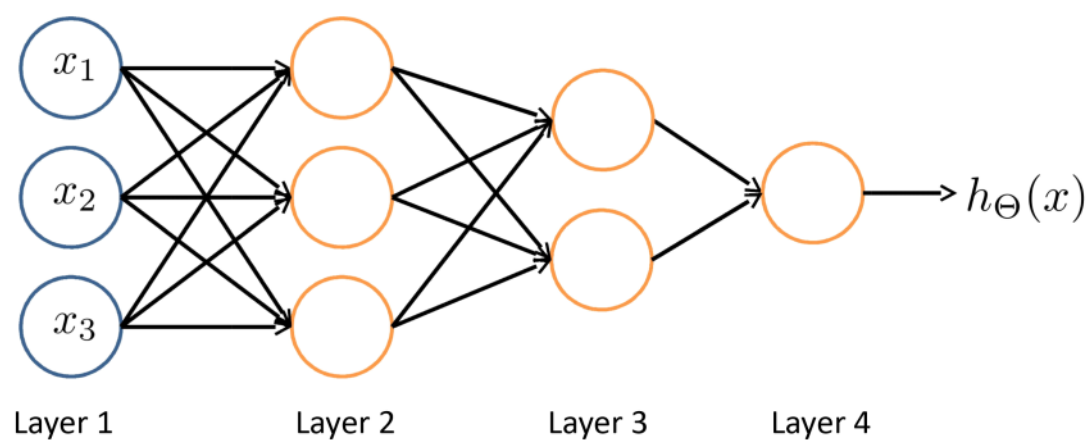


现在我们只关注后面
隐藏层和输出层之间的关系

Andrew Ng

因为这里才是输出层的训练变量，
也是最终出结果的关键步骤

Other network architectures



这里给出类似的神经网络的结构
基本上只要抓住核心结构
输入，隐藏和输出这三点即可

Andrew Ng



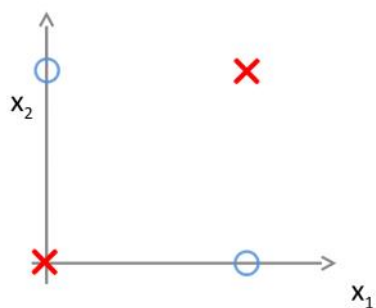
Machine Learning

Neural Networks: Representation

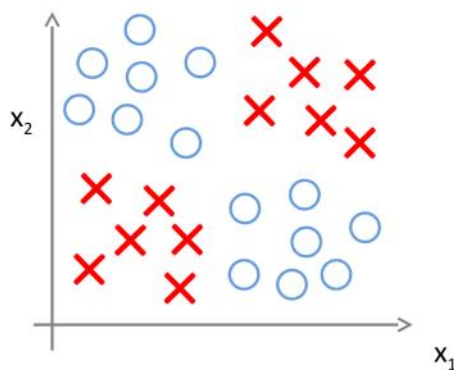
Examples and intuitions I

Non-linear classification example: XOR/XNOR

x_1, x_2 are binary (0 or 1).



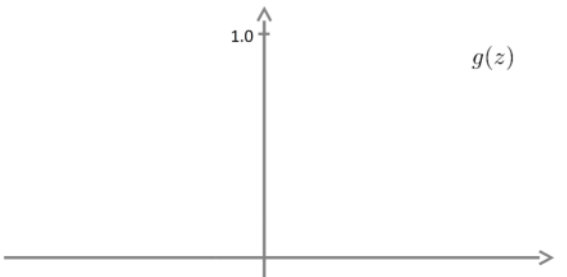
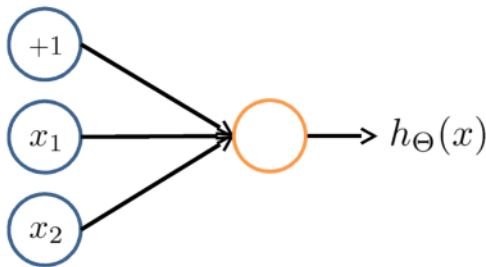
$$\begin{aligned} y &= x_1 \text{ XOR } x_2 \\ &= x_1 \text{ XNOR } x_2 \\ &= \text{NOT } (x_1 \text{ XOR } x_2) \end{aligned}$$



Andrew Ng

Simple example: AND

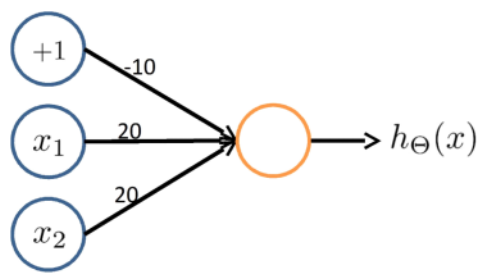
$x_1, x_2 \in \{0, 1\}$
 $y = x_1 \text{ AND } x_2$



| x_1 | x_2 | $h_{\Theta}(x)$ |
|-------|-------|-----------------|
| 0 | 0 | |
| 0 | 1 | |
| 1 | 0 | |
| 1 | 1 | |

Andrew Ng

Example: OR function



| x_1 | x_2 | $h_{\Theta}(x)$ |
|-------|-------|-----------------|
| 0 | 0 | |
| 0 | 1 | |
| 1 | 0 | |
| 1 | 1 | |

Andrew Ng

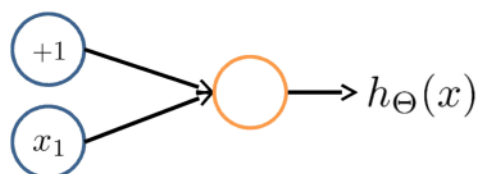




Machine Learning

Neural Networks: Representation

Examples and intuitions II

x_1 AND x_2 x_1 OR x_2 **Negation:**

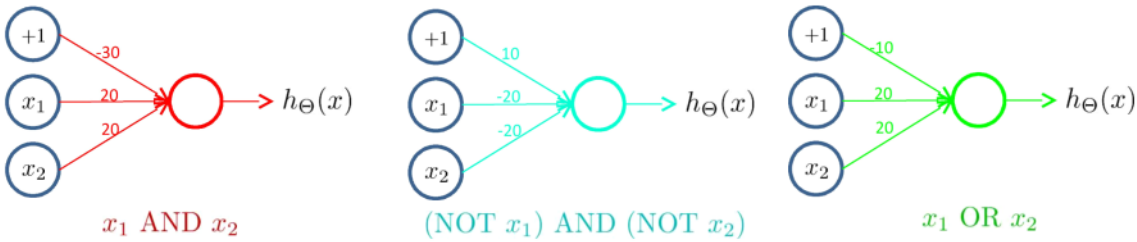
| x_1 | $h_{\Theta}(x)$ |
|-------|-----------------|
| 0 | |
| 1 | |

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

(NOT x_1) AND (NOT x_2)

Andrew Ng

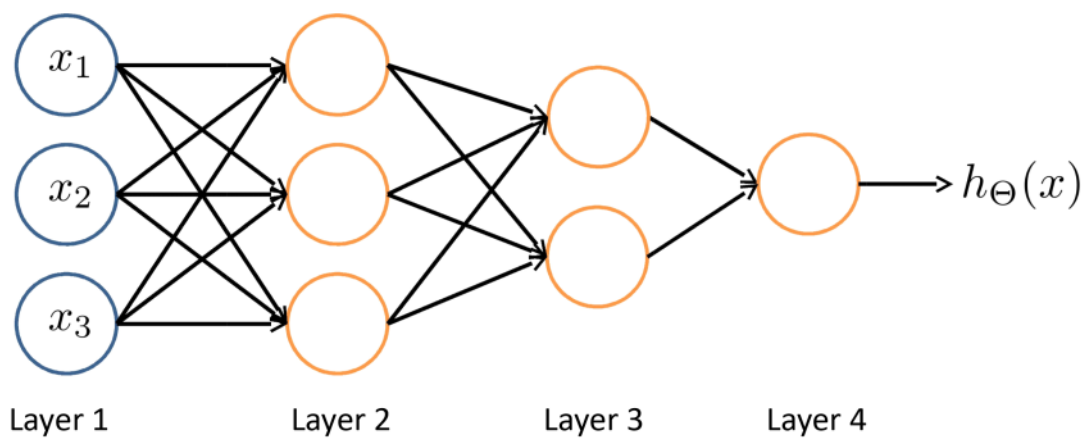
Putting it together: x_1 XNOR x_2



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

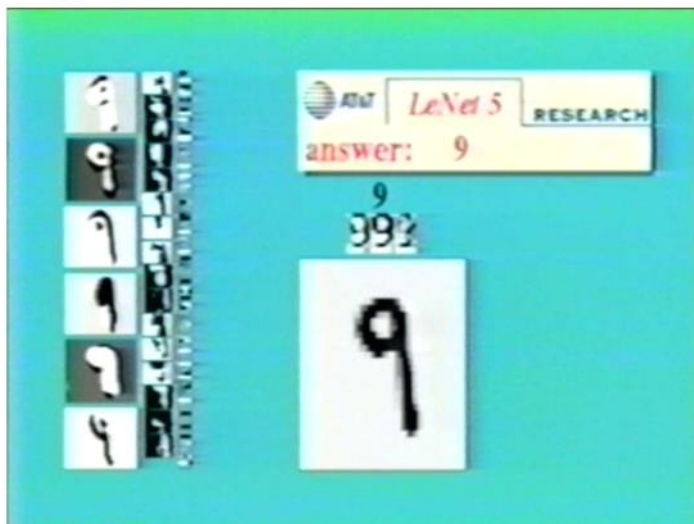
Andrew Ng

Neural Network intuition



Andrew Ng

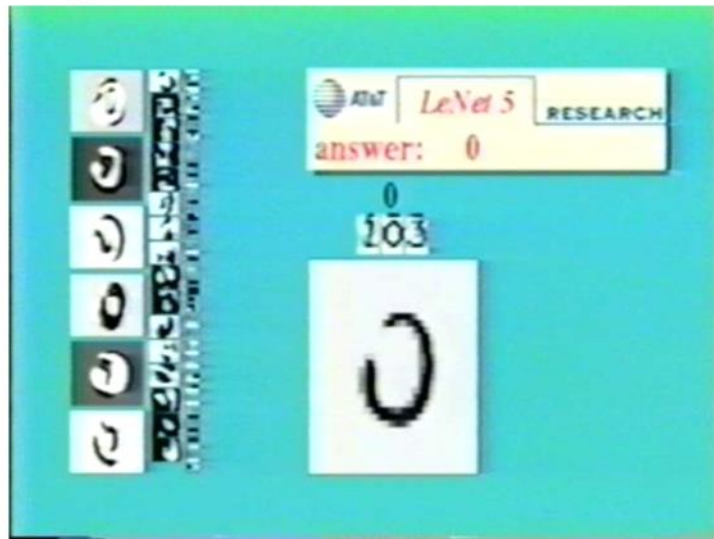
Handwritten digit classification



Courtesy of Yann LeCun]

Andrew Ng

Handwritten digit classification



Courtesy of Yann LeCun]

Andrew Ng

Andrew Ng



Machine Learning

Neural Networks: Representation

Multi-class classification

Andrew Ng

Multiple output units: One-vs-all.



Pedestrian



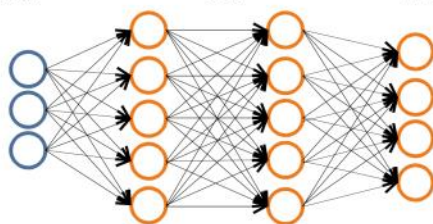
Car



Motorcycle



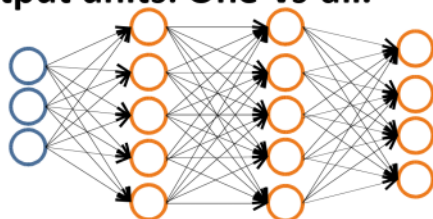
Truck



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

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

Andrew Ng

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

$y^{(i)}$ 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

Andrew Ng