



机器学习导论

第3章 神经网络

3.1 神经网络与脑科学

谢茂强

南开大学软件学院

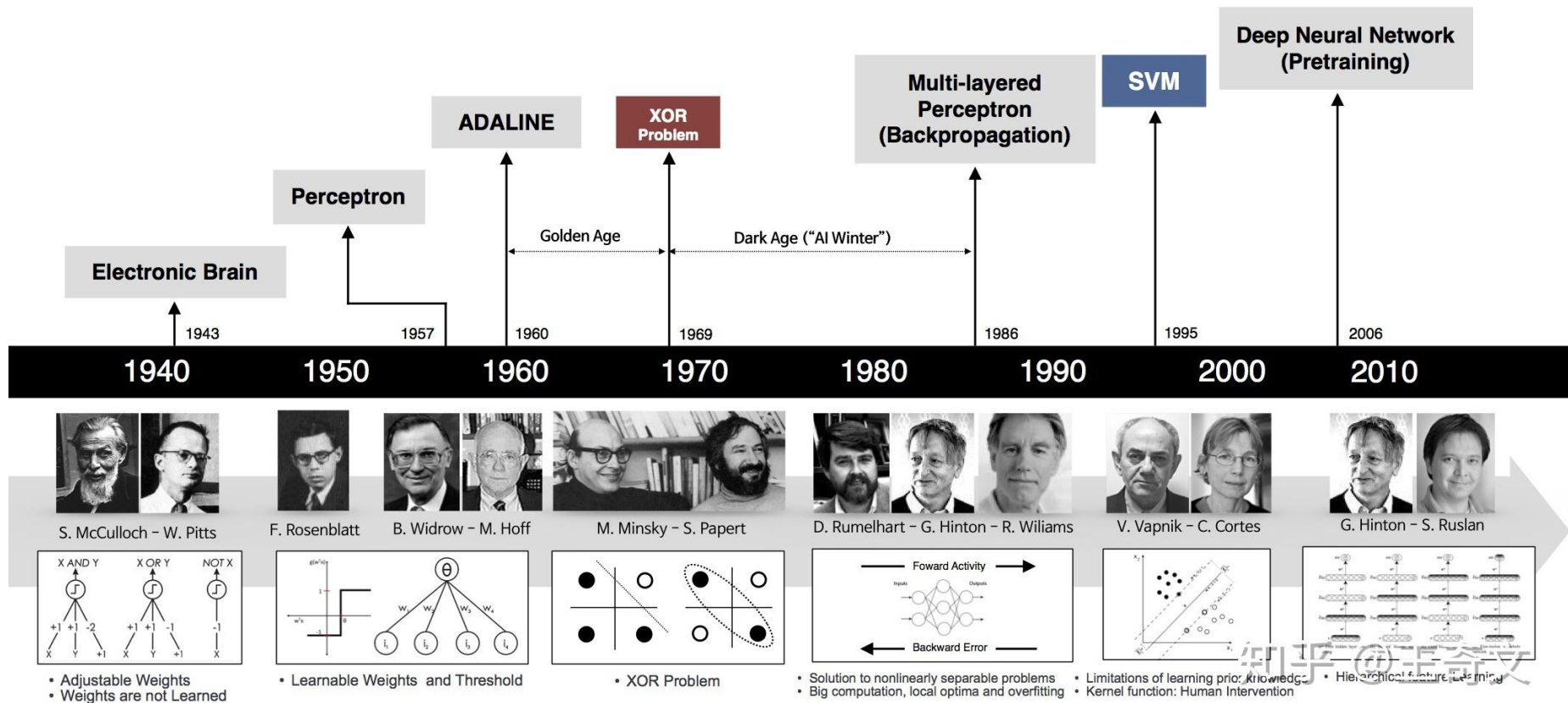
本节内容主要来自于吴恩达Coursera网课

人类历史上对智能的期望

- 诸葛亮：木牛流马、家庭机器人（让机械具备基本智能）
- 古代和近代的计算装置、现代电子计算机
- 图灵测试
- 1950~1970, General Problem Solving, 证明了《数学原理》中的38个定理。以及之后的知识工程（逻辑推理、知识表达）
- 1950s 开始，神经学、电子装置、软件（抽象神经元）三方面参与到神经网络的研究上，2015年开始复苏，近期蓬勃发展。（记忆、预测）
- 1970s开始的从样例中学习（统计学+计算）
- 人工智能（机器学习）：软件、计算机、加速卡、芯片、脑科学（神经学）
- 模拟人类大脑是人工智能研究的一大重要流派

- 软件、计算机、加速卡、芯片、脑科学（神经学）
- 在上世纪80年代到90年代初期很兴盛，后衰退
- 当时线性描述只能解决小部分问题，机器学习能解决的问题很少
- 使用特征二次组合和三次组合，对于当时的算力过于复杂

研究和模拟神经网络是人工智能研究的一大流派



- 人脑一直是智能研究领域的重要方向，尽管进展慢于各类计算装置
- 从信息处理的角度，大脑包括以下功能
 - 感觉、知觉、记忆、思维、想像和语言等
- 人工智能也可以看做是在实现人脑的功能
 - 实现手段有直接模拟
 - 大多数与大脑无关



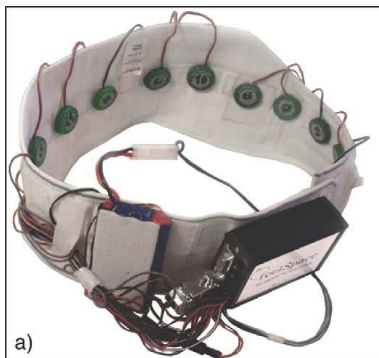
使用听觉、触觉替代“视觉”功能



Seeing with your tongue



Human echolocation
(sonar)

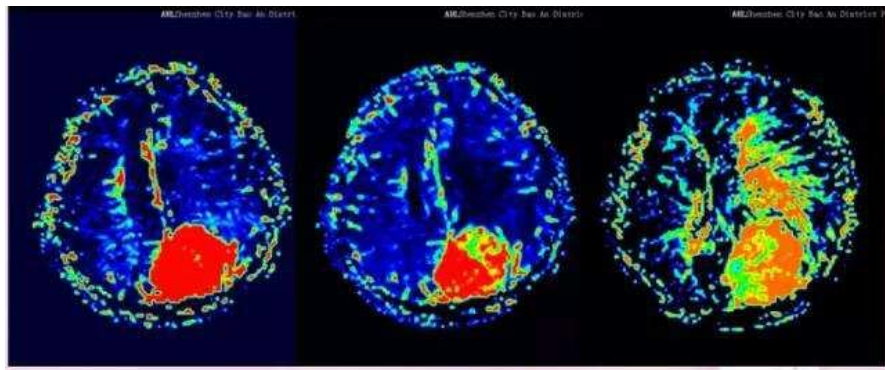


Haptic belt: Direction sense



Implanting a 3rd eye Artificial arm

通过脑电信号和机械装置替代猴子的进食



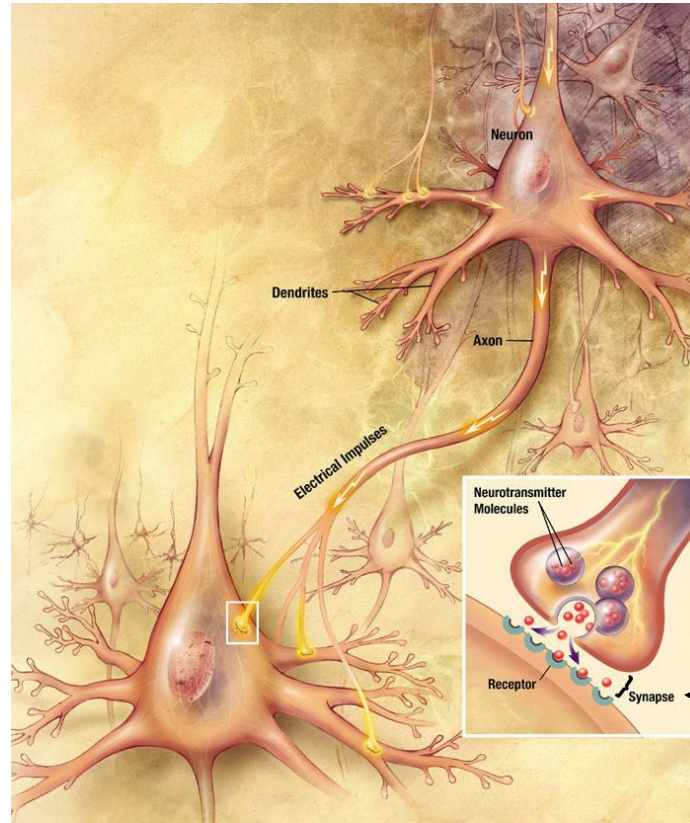
- 2013年4月2日，美国宣布启动脑科学计划（BRAIN Initiative）
- 欧洲脑计划（The Human Brain Project）
- 日本脑计划（Brain/Minds Project）
- 2016年，“脑科学与类脑研究”列为“科技创新2030—重大项目”。
这被解读为“中国脑计划”也正式启动。



- 脑认知原理解析
- 脑介观神经联接图谱绘制
- 脑重大疾病机理与干预研究
- 儿童青少年脑智发育
- 类脑计算与脑机融合技术研究

专栏2 科技前沿领域攻关	
01 新一代人工智能	前沿基础理论突破，专用芯片研发，深度学习框架等开源算法平台构建，学习推理与决策、图像图形、语音视频、自然语言识别处理等领域创新。
02 量子信息	域域、城际、自由空间量子通信技术研发，通用量子计算原型机和实用化量子模拟机研制，量子精密测量技术突破。
03 集成电路	集成电路设计工具、重点装备和高纯靶材等关键材料研发，集成电路先进工艺和绝缘栅双极型晶体管（IGBT）、微机电系统（MEMS）等特色工艺突破，先进存储技术升级，碳化硅、氮化镓等宽禁带半导体发展。
04 脑科学与类脑研究	脑认知原理解析，脑介观神经联接图谱绘制，脑重大疾病机理与干预研究，儿童青少年脑智发育，类脑计算与脑机融合技术研发。
05 基因与生物技术	基因组学研究应用，遗传细胞和遗传育种、合成生物、生物药等技术创新，创新疫苗、体外诊断、抗体药物等研发，农作物、畜禽水产、农业微生物等重大新品种创制，生物安全关键技术研究。
06 临床医学与健康	癌症和心脑血管、呼吸、代谢性疾病等发病机制基础研究，主动健康干预技术研发，再生医学、微生物组、新型治疗等前沿技术研发，重大传染病、重大慢性非传染性疾病防治关键技术研究。
07 深空深地深海和极地探测	宇宙起源与演化、透视地球等基础科学研究，火星环绕、小行星巡视等星际探测，新一代重型运载火箭和重复使用航天运输系统、地球深部探测装备、深海运维保障和装备试验船、极地立体观测平台和重型破冰船等研制，探月工程四期、蛟龙探海二期、雪龙探极二期建设。

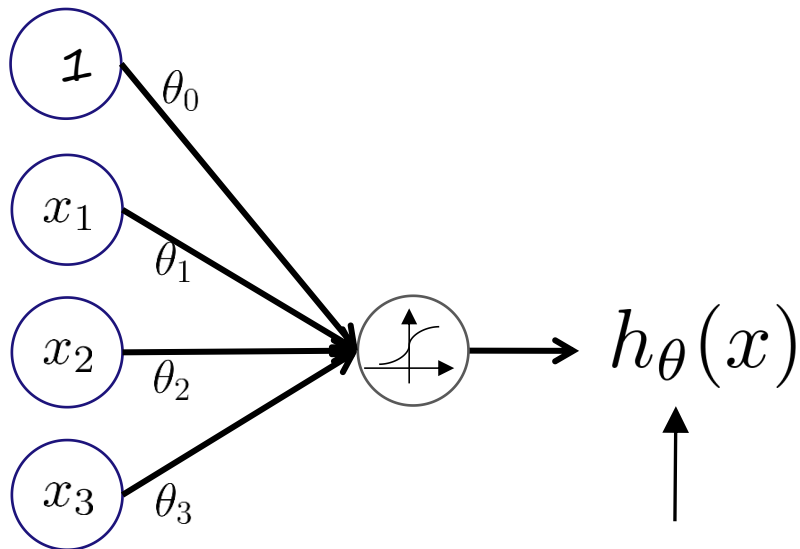
Neurons in the brain (信号传导)



轴突末端

突触

Neuron model: Logistic unit



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

Sigmoid (logistic)激活函数

1943年, [McCulloch and Pitts, 1943]
M-P神经元模型

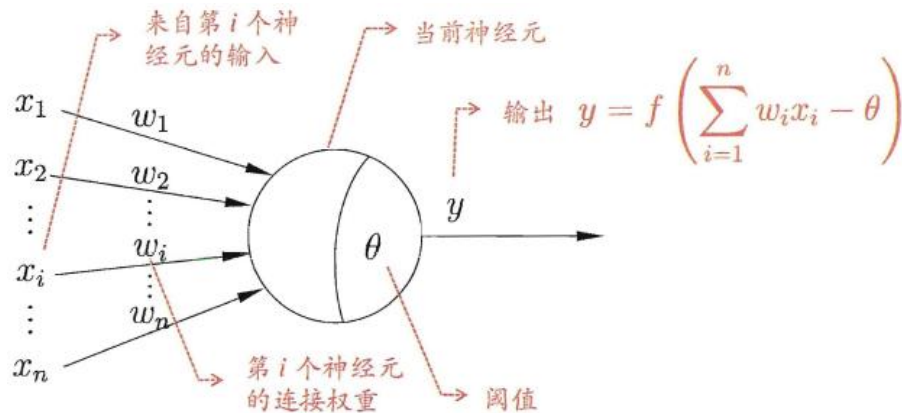
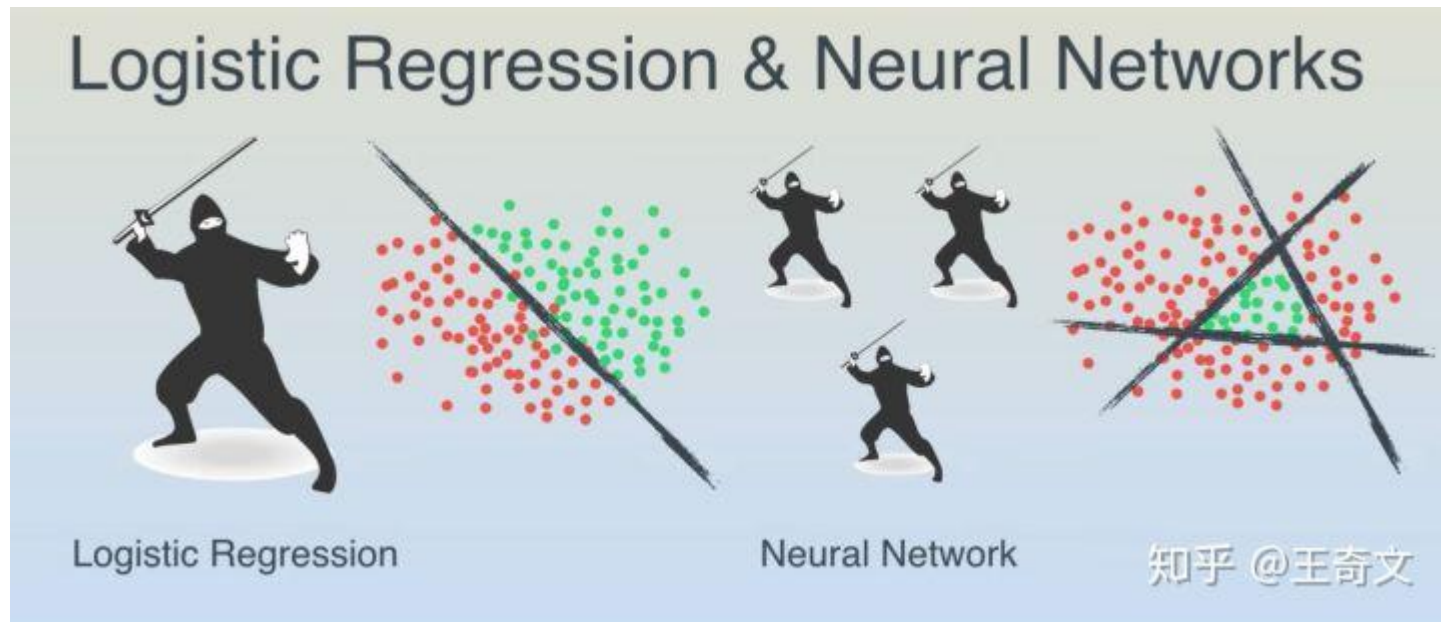


图 5.1 M-P 神经元模型



Layer 1

Input Layer

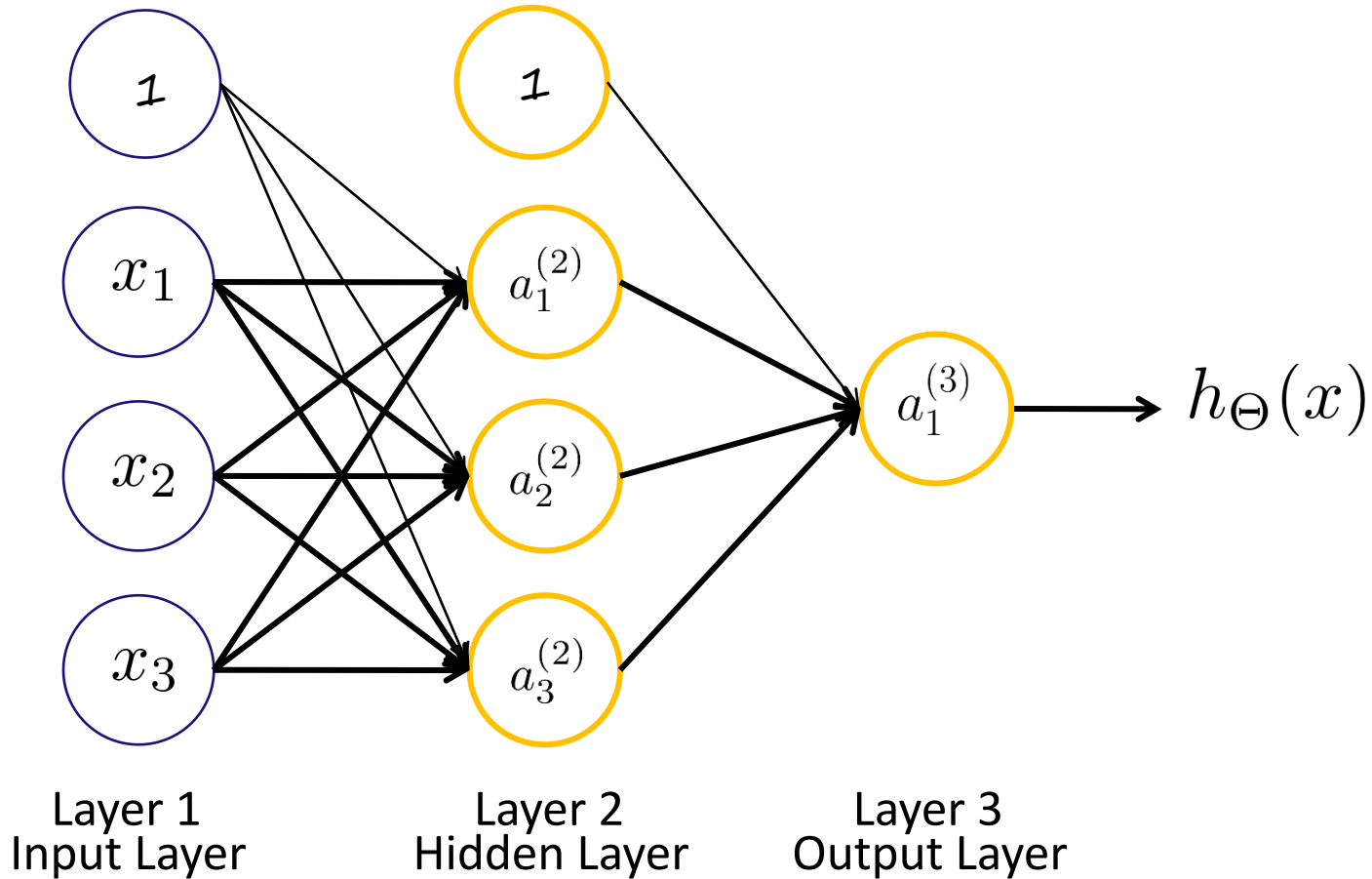
Layer 2

Hidden Layer

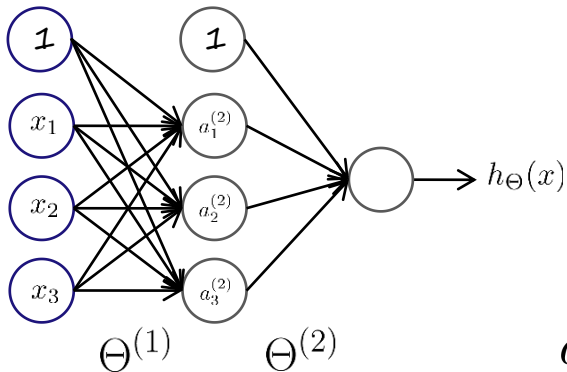
Layer 3

Output Layer

Neural Network



Neural Network



When $j = 1$, $\Theta^{(j)}$ will be of dimension

$$s_2 \times (s_1 + 1) = 3 \times 4$$

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

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

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

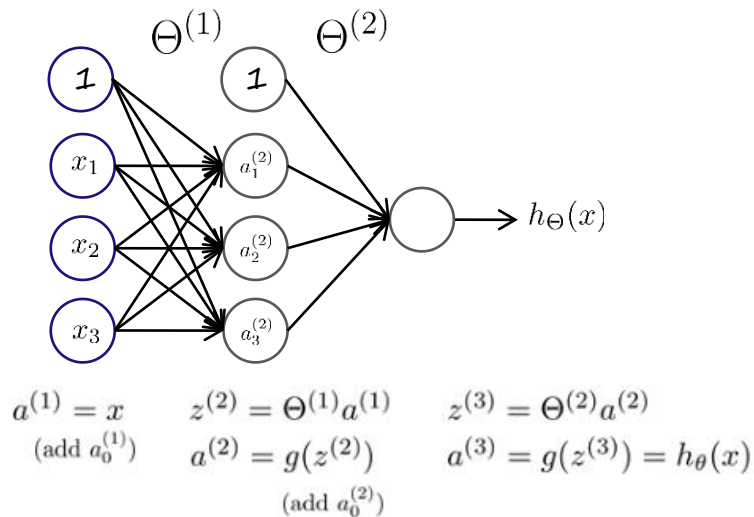
$$a_2^{(2)} = g(\Theta_{20}^{(1)} x_0 + \Theta_{21}^{(1)} x_1 + \Theta_{22}^{(1)} x_2 + \Theta_{23}^{(1)} x_3)$$

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

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

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

Forward propagation: Vectorized implementation



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

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

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

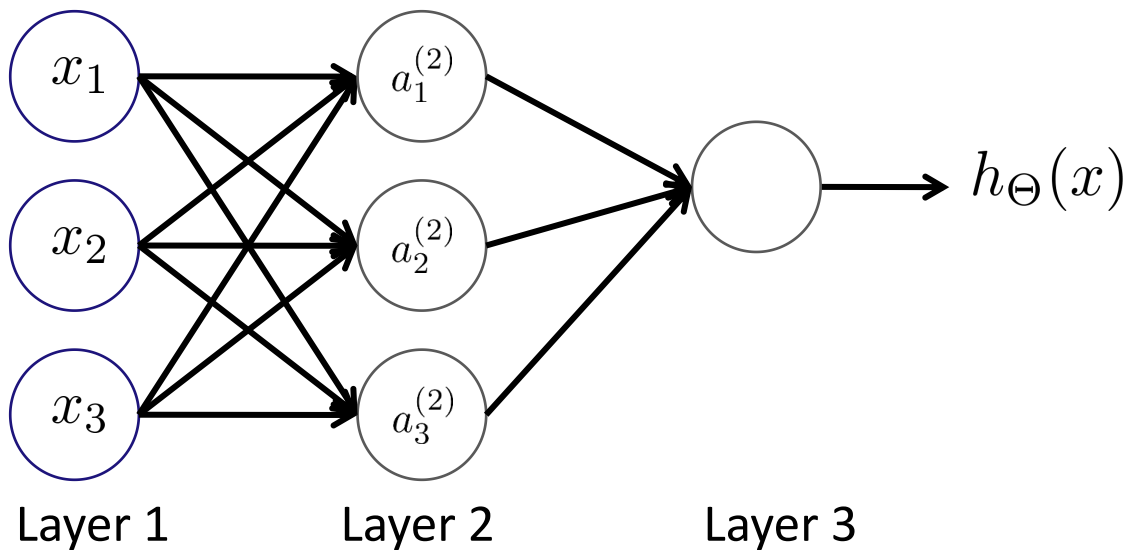
$$\begin{aligned}
 z^{(2)} &= \Theta^{(1)} x \\
 a^{(2)} &= g(z^{(2)})
 \end{aligned}$$

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

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

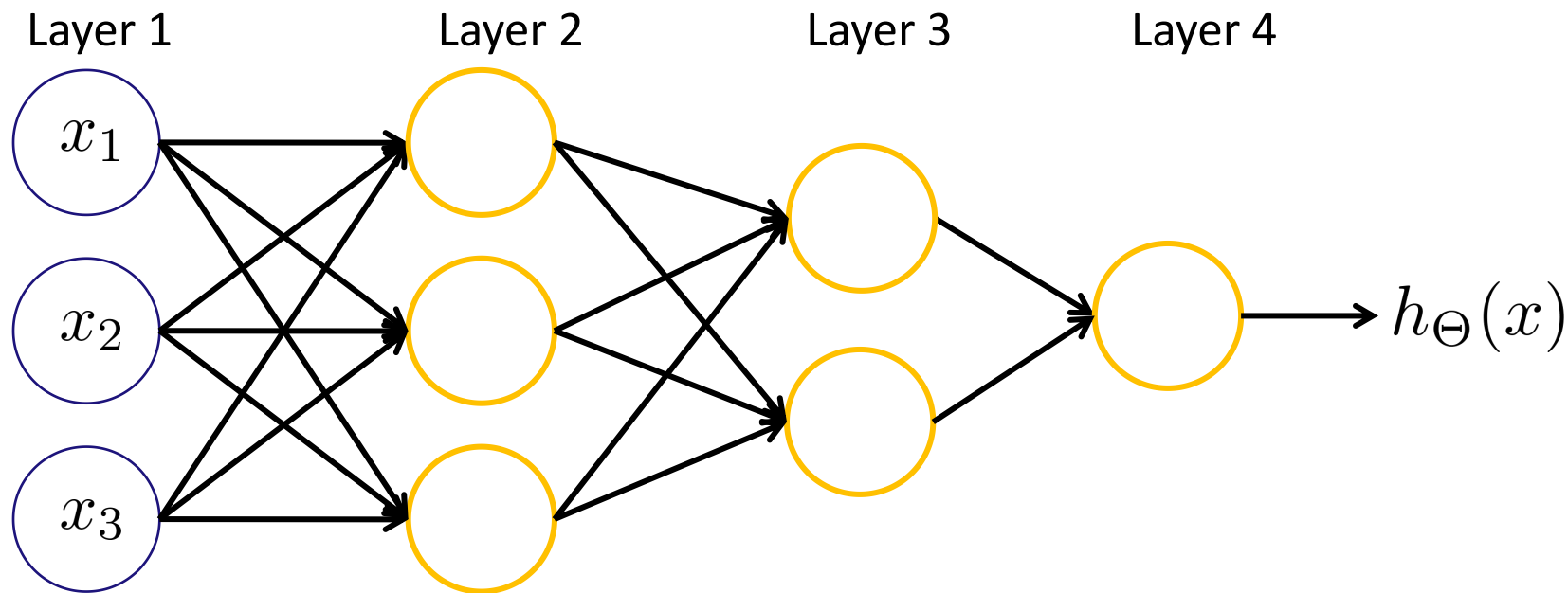
$$h_{\theta}(x) = a^{(3)} = g(z^{(3)})$$

Neural Network learning its own features



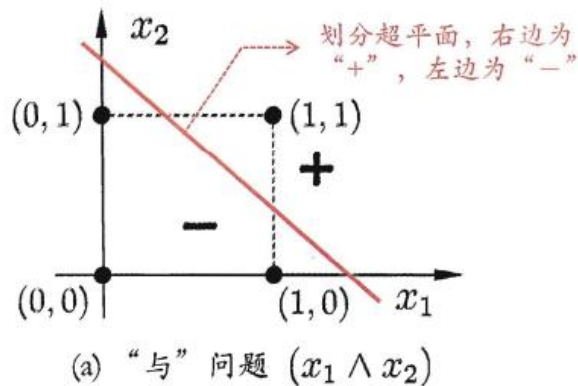
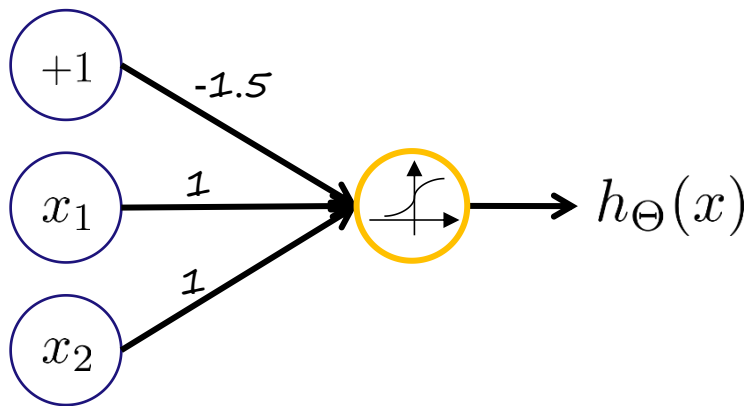
- 万能近似定理 (universal approximation theorem) (Hornik et al., 1989;Cybenko, 1989) 表明，一个前馈神经网络如果具有线性输出层和至少一层具有任何一种 ‘ ‘挤压’ ’ 性质的激活函数（例如logistic sigmoid激活函数）的隐藏层，只要给予网络足够数量的隐藏单元，它可以以任意精度来近似任何从一个有限维空间到另一个有限维空间的Borel 可测函数（定义在紧集上的连续函数）。（《Deep learning》英文版P194，中文版P171）

Neural Network Intuition



也有增加多个隐含层的神经网络

感知器(Perceptron)示例：逻辑与的实现

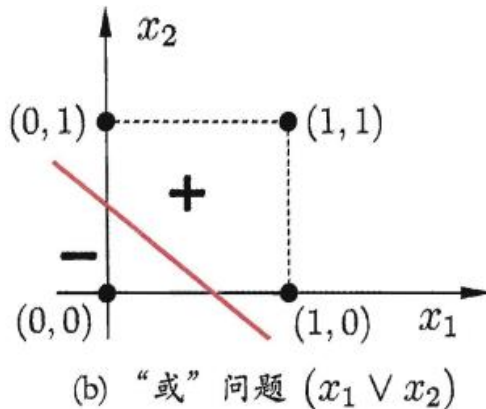
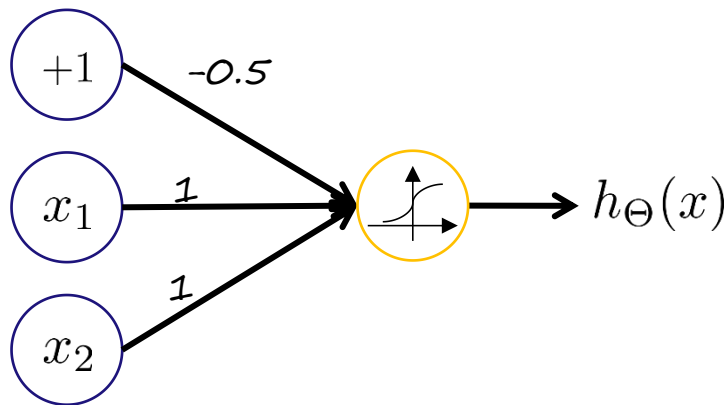


$$h_{\theta}(x) = \text{sigmoid}(x_1 + x_2 - 1.5)$$

$$\hat{y} = \begin{cases} 1 & h_{\theta}(x) \geq 0.5 \\ 0 & h_{\theta}(x) < 0.5 \end{cases}$$

x_1	x_2	$h_{\theta}(x)$
0	0	0
0	1	0
1	0	0
1	1	1

感知器示例：逻辑或的实现

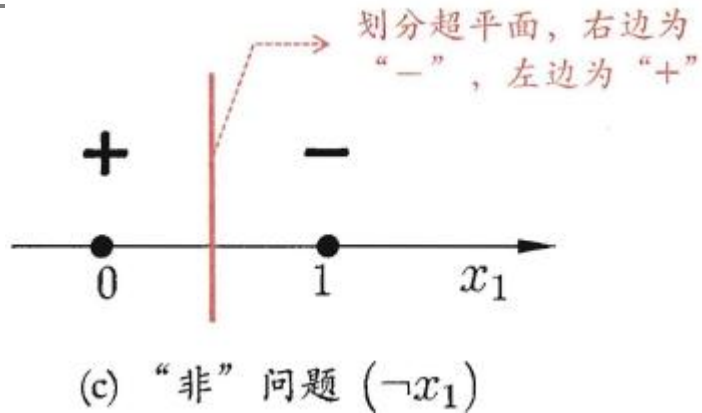
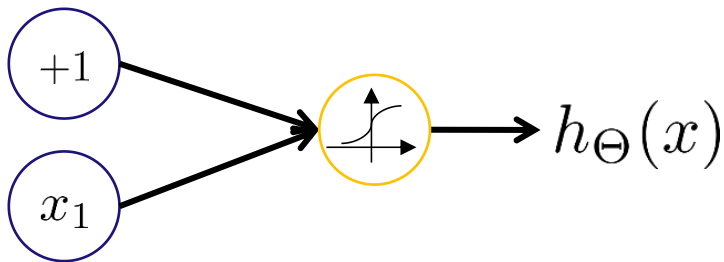


$$h_{\theta}(x) = \text{sigmoid}(x_1 + x_2 - 0.5)$$

$$\hat{y} = \begin{cases} 1 & h_{\theta}(x) \geq 0.5 \\ 0 & h_{\theta}(x) < 0.5 \end{cases}$$

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

感知器示例：逻辑非的实现

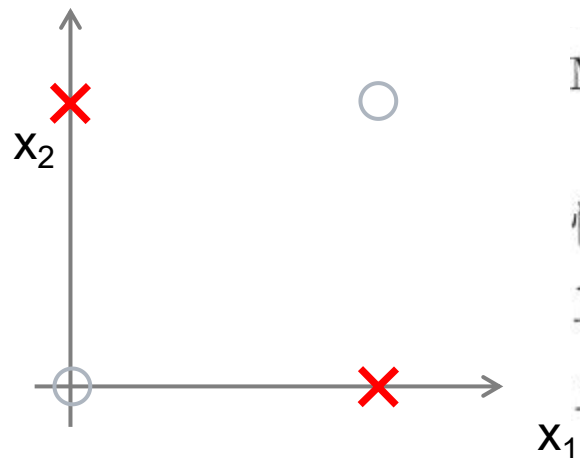


$$h_{\theta}(x) = \text{sigmoid}(-x_1 + 0.5)$$

$$\hat{y} = \begin{cases} 1 & h_{\theta}(x) \geq 0.5 \\ 0 & h_{\theta}(x) < 0.5 \end{cases}$$

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

逻辑异或 XOR 与 第一次神经网络的衰落



MIT 计算机科学研究的奠基人马文·闵斯基 (Marvin Minsky, 1927—) 与 Seymour Papert 在 1969 年出版了《感知机》一书, 书中指出, 单层神经网络无法解决非线性问题, 而多层网络的训练算法尚看不到希望. 这个论断直接使神经网络研究进入了“冰河期”, 美国和苏联均停止了神经网络研究的资助

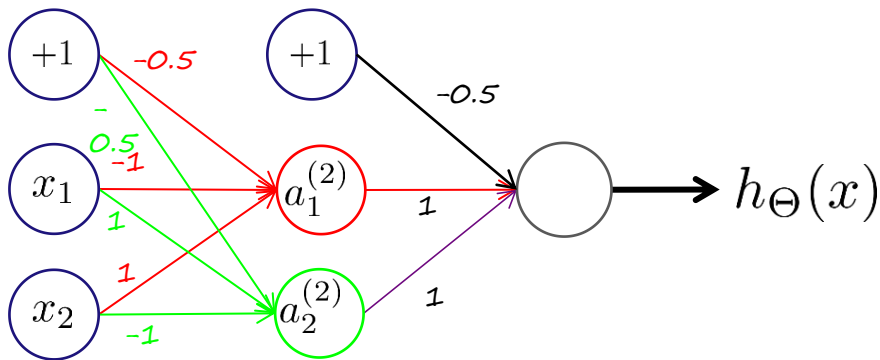
$$y = x_1 \text{ XOR } x_2$$

$$y = x_1 \oplus x_2 = (\bar{x}_1 \cap x_2) \cup (x_1 \cap \bar{x}_2)$$

逻辑异或的实现



$$y = x_1 \oplus x_2 = \underbrace{(\bar{x}_1 \cap x_2)}_{a_1^{(2)}} \cup \underbrace{(x_1 \cap \bar{x}_2)}_{a_2^{(2)}}$$



		$(\bar{x}_1 \cap x_2)$	$(x_1 \cap \bar{x}_2)$	
x_1	x_2	$a_1^{(2)}$	$a_2^{(2)}$	$h_{\Theta}(x)$
0	0	0	0	0
0	1	1	0	1
1	0	0	1	1
1	1	0	0	0

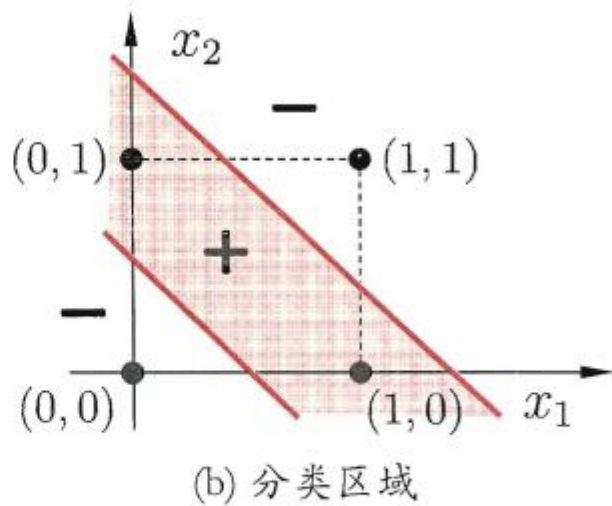
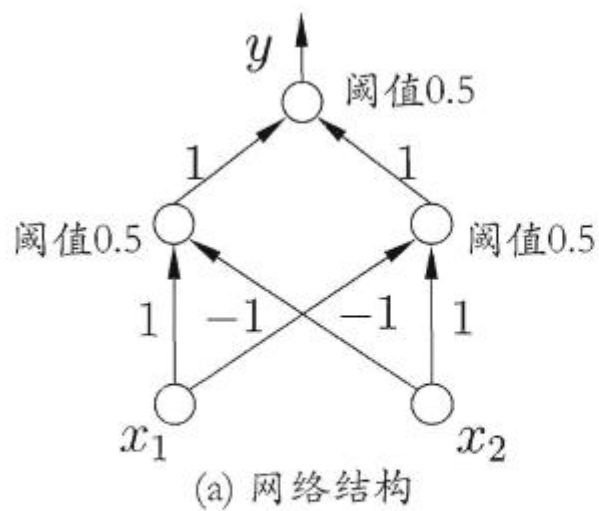


图 5.5 能解决异或问题的两层感知机

神经网络天然支持多类分类



Pedestrian



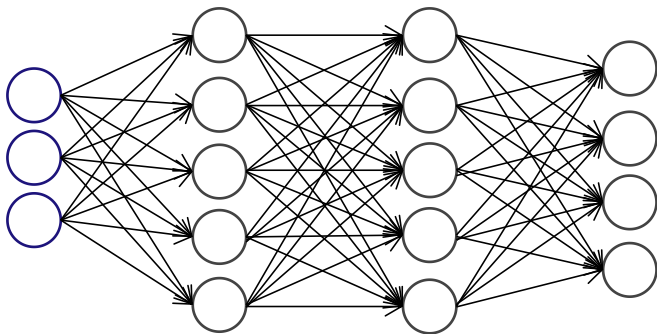
Car



Motorcycle



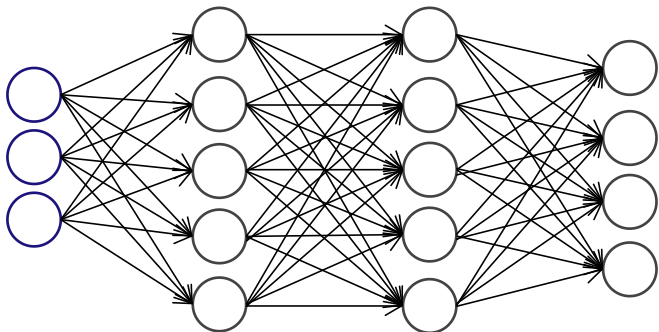
Truck



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

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

Multiple output units: One-vs-all.



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

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

$h_{\Theta}(x) \approx \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ when car

$h_{\Theta}(x) \approx \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ 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

- 计算机领域的神经网络与神经科学（脑科学）的神经网络的关联和区别
- 使用神经元模型实现逻辑运算
- 使用多层神经元模型实现异或运算
- 使用神经网络实现多类分类任务