
人工神经网络

认识数据集

数据集/样本集：事件或对象的集合(下图17个西瓜即称为数据集)

样本：数据集中每一个对象，即每个西瓜

样本特征或属性：每个样本的一些特点(色泽、根蒂、纹理...)

样本的维度：特征的数量(8)

样本的结果：好瓜，不是好瓜(标记)，**决定算法是否为监督学习**

编号	色泽	根蒂	敲声	纹理	脐部	触感	密度	含糖率	好瓜
1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	0.697	0.460	是
2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	0.774	0.376	是
3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	0.634	0.264	是
4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	0.608	0.318	是
5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	0.556	0.215	是
6	青绿	稍蜷	浊响	清晰	稍凹	软粘	0.403	0.237	是
7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	0.481	0.149	是
8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	0.437	0.211	是
9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	0.666	0.091	否
10	青绿	硬挺	清脆	清晰	平坦	软粘	0.243	0.267	否
11	浅白	硬挺	清脆	模糊	平坦	硬滑	0.245	0.057	否
12	浅白	蜷缩	浊响	模糊	平坦	软粘	0.343	0.099	否
13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	0.639	0.161	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	0.657	0.198	否
15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	0.360	0.370	否
16	浅白	蜷缩	浊响	模糊	平坦	硬滑	0.593	0.042	否
17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	0.719	0.103	否

1. 样本集合: $X = (X_1, X_2, \dots, X_n)$, 含有 n 个样本: $n = 7$
2. 每个样本含有 m 个特征: $X_i = (a_1, a_2, \dots, a_m)$: $m = 3$ 个特征
3. 类别集合 (k 类): $Y = (y_1, y_2, \dots, y_k)$: $k = 2$
4. 例子: 西瓜样本, 分类西瓜是好是坏

西瓜样本集合							
	1	2	3	4	5	6	7
纹理	清晰	模糊	模糊	清晰	清晰	模糊	清晰
色泽	清绿	乌黑	清绿	乌黑	清绿	乌黑	乌黑
敲声	清脆	浊响	浊响	沉闷	浊响	沉闷	清脆
类别	好瓜	坏瓜	坏瓜	好瓜	好瓜	坏瓜	好瓜

学号	性别	学生干部	综合成绩	毕业论文	就业情况
1	男	是	70-79	优	已
2	女	是	80-89	中	已
3	男	不是	60-69	不及格	未
4	男	是	60-69	良	已
5	男	是	70-79	中	已
6	男	不是	70-79	良	未
7	女	是	60-69	良	已
8	男	是	60-69	良	已
9	女	是	70-79	中	未
10	男	不是	60-69	及格	已
11	男	是	80-89	及格	已
12	男	是	70-79	良	已
13	男	不是	70-79	及格	未
14	男	不是	60-69	及格	已
15	男	是	70-79	良	已
16	男	不是	70-79	良	未
17	男	不是	80-89	良	未
18	女	是	70-79	良	已
19	男	不是	70-79	不及格	未
20	男	不是	70-79	良	未
21	女	是	60-69	优	已
22	男	是	60-69	良	已

- 1981年生物学家格若根（W. Grogan）和维什（W. Wirth）发现了两类蚊子(或飞蠓midges).



他们测量了这两类蚊子每个个体的翼长和触角长，数据如下：

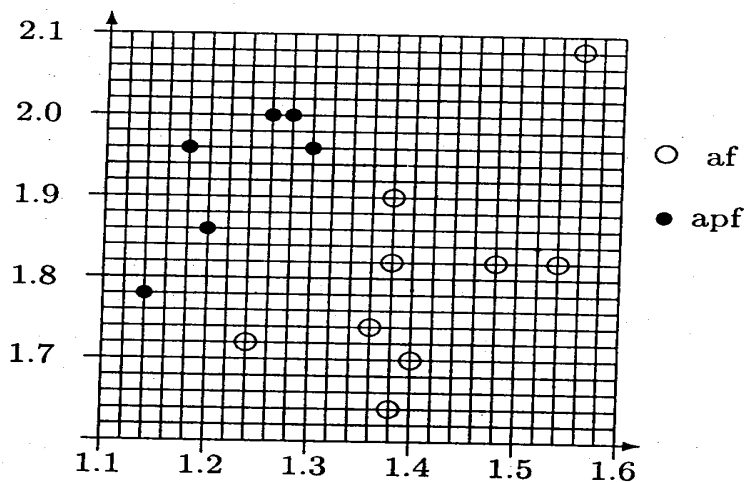
- | 翼长 | 触角长 | 类别 |
|------|------|-----|
| 1.78 | 1.14 | Apf |
| 1.96 | 1.18 | Apf |
| 1.86 | 1.20 | Apf |
| 1.72 | 1.24 | Af |
| 2.00 | 1.26 | Apf |
| 2.00 | 1.28 | Apf |
| 1.96 | 1.30 | Apf |
| 1.74 | 1.36 | Af |

- | 翼长 | 触角长 | 类别 |
|------|------|----|
| 1.64 | 1.38 | Af |
| 1.82 | 1.38 | Af |
| 1.90 | 1.38 | Af |
| 1.70 | 1.40 | Af |
| 1.82 | 1.48 | Af |
| 1.82 | 1.54 | Af |
| 2.08 | 1.56 | Af |

- 问：如果抓到三只新的蚊子，它们的触角长和翼长分别为(1.24, 1.80); (1.28, 1.84); (1.40, 2.04) .
问它们应分别属于哪一个种类？

解法一：

- 把翼长作纵坐标，触角长作横坐标；那么每个蚊子的翼长和触角决定了坐标平面的一个点. 其中 6个蚊子属于 APf类；用黑点 “.” 表示；9个蚊子属 Af类；用小圆圈 “。” 表示.



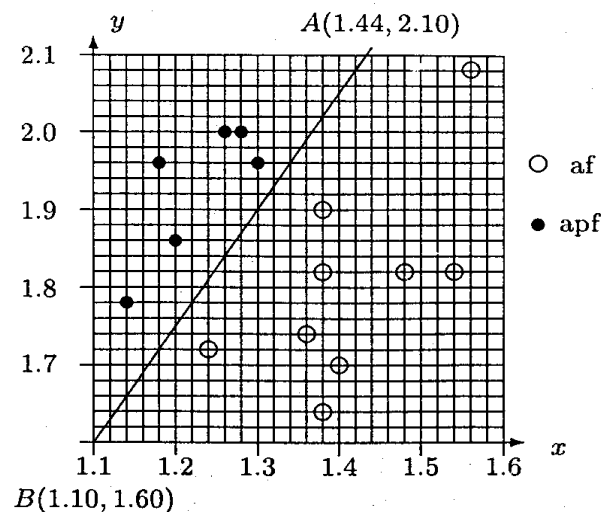
• 图1飞蠓的触角长和翼长

思路：作一直线将两类飞蠓分开

- 例如，取 $A = (1.44, 2.10)$ 和 $B = (1.10, 1.60)$ ，过 A B 两点作一条直线：

$$y = 1.47x - 0.017$$

其中 X 表示触角长； y 表示翼长。



分类规则： 设一个蚊子的数据为 (x, y)

如果 $y \geq 1.47x - 0.017$ ，则判断蚊子属 Apf 类；

如果 $y < 1.47x - 0.017$ ；则判断蚊子属 Af 类。

分类结果： $(1.24, 1.80)$, $(1.28, 1.84)$ 属于 Af 类；

$(1.40, 2.04)$ 属于 Apf 类。

拟合存在问题

□ 一元拟合问题：x为一维数据

①最小二乘拟合方法

②线性拟合，二次曲线拟合，三次曲线拟合及多项式拟合。

$$y = ax + b; y = ax^2 + bx + c$$

③a, b, c 称为参数,

④x, x^2 称为拟合所需的特征(feature)

□ 多元拟合问题：即 x为高维数据

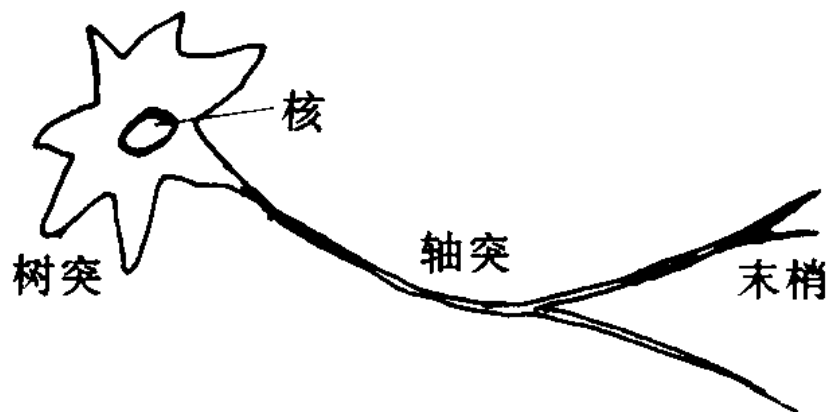
二次拟合

$$y = ax^2 + bx + c$$

$$h_{\theta}(x_1, x_2) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2$$

- 50*50像素的灰度图片：大约包含上百万个特征，
- 彩色图片，特征会增加至上千万
- 当数据库包含上万甚至上亿张图片时，**拟合方法**很难去拟合这些数据
- 要寻找一种数学模型，在此基础上不断减少特征，降低维度
- **人工神经网络即是在这样的背景下产生**

- 大脑可视作为1000多亿神经元组成的神经网络

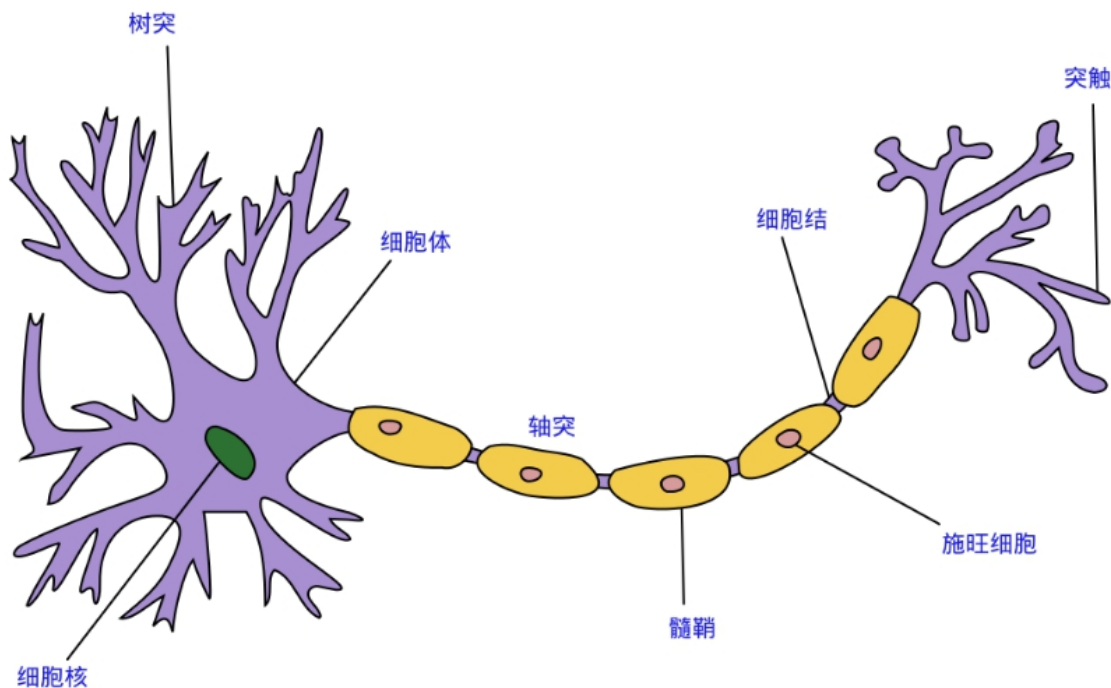


- 图3 神经元的解剖图

1.1 神经元

1904年生物学家就已经知晓了神经元的组成结构

- 多个**树突**，主要用来接受传入信息；
- 一条**轴突**，轴突尾端有许多轴突末梢，给其他多个神经元传递信息。
- 突触**，轴突末梢跟其他神经元的树突产生连接，从而传递信号。
- 人脑中的神经元形状：



- 神经元的信息传递和处理是一种电化学活动.
- 树突由于电化学作用接受外界的刺激；通过胞体内的活动体现为轴突电位，当轴突电位达到一定的值则形成神经脉冲或动作电位；
- 再通过轴突末梢传递给其它的神经元.
- 从控制论的观点来看；这一过程可以看作一个多输入单输出非线性系统的动态过程

神经网络研究的两个方面

- 从生理上、解剖学上进行研究
- 从工程技术上、算法上进行研究

1.0 神经网络

□ 人工神经网络

①Artificial Neural Network, ANN, 简称神经网络 (Neural Network, NN) 。

②是一种模仿生物神经网络（动物的中枢神经系统，特别是大脑）的结构和功能的数学模型或计算模型，用于对函数进行**估计或近似**。

③模拟人类的大脑，造出会思考的机器

④神经元是构成神经网络的最基本的单元

神经网络的作用

(1) 蚂蚁群

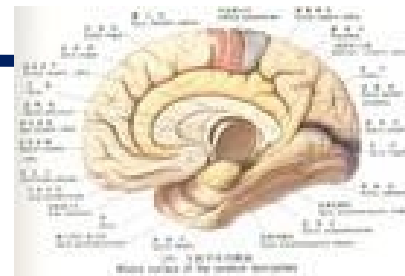
一个蚂蚁有50个神经元，单独的一个蚂蚁不能做太多的事；甚至于不能很好活下去。但是一窝蚂蚁；设有 10 万个个体，那么这个群体相当于500万个神经元（当然不是简单相加，这里只为说明方便而言）；那么它们可以觅食、搬家、围攻敌人等等。



(2) 网络说话

人们把一本教科书用网络把它读出来（当然需要通过光电，电声的信号转换）；开始网络说的话像婴儿学语那样发出“巴、巴、巴”的声响；但经过BP算法长时间的训练竟能正确读出英语课本中 90%的词汇。

人工神经网络的基本特点



- (1) 可处理非线性
- (2) 并行结构. 对神经网络中的每一个神经元来说; 其运算都是同样的. 这样的结构最便于计算机并行处理.
- (3) 具有学习和记忆能力. 一个神经网络可以通过训练学习判别事物; 学习某一种规律或规则. 神经网络可以用于联想记忆.
- (4) 对数据的可容性大. 在神经网络中可以同时使用量化数据和质量数据 (如好、中、差、及格、不及格等).
- (5) 神经网络可以用大规模集成电路来实现. 如美国用 **256** 个神经元组成的神经网络组成硬件用于识别手写体的邮政编码.

- 1943 提出神经元模型
- 1949 提出学习算法及规则
- 1957 提出感知器-人工神经元模型
- 1969 <感知器>发表,串行计算机全胜
- 1982 建立人工神经网络模型
- 1986 提出BP神经网络算法

□ MP神经元模型

1943年，参考了生物神经元的结构，
Warren McCulloch和Walter Pitts
提出MP神经元模型

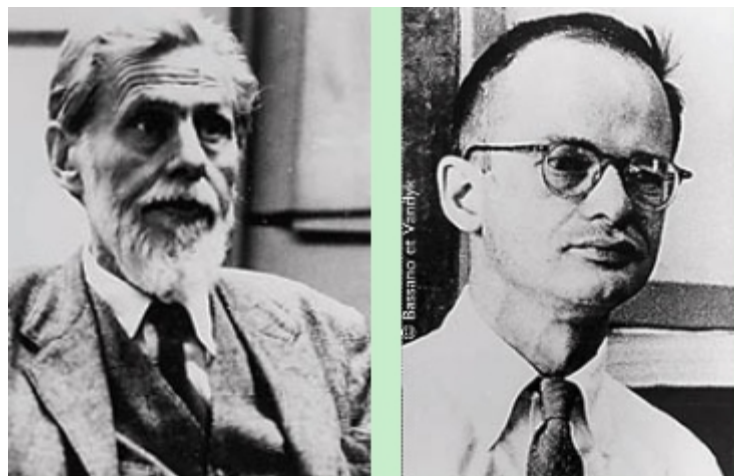


图5 Warren McCulloch (左) 和 Walter Pitts (右)

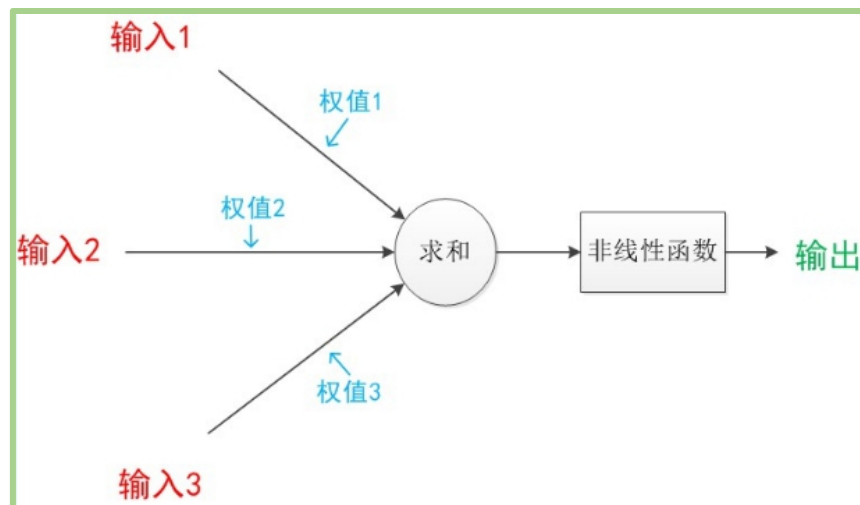
□ MP神经元结构:

输入: 可以类比为神经元的树突,

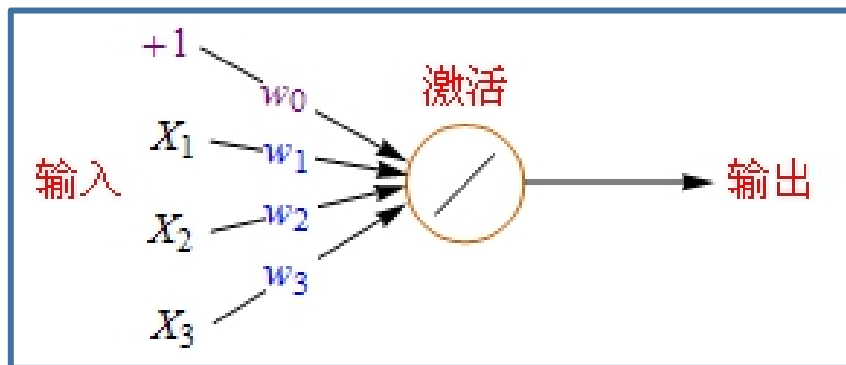
输出: 可以类比为神经元的轴突,

计算: 则可以类比为细胞核

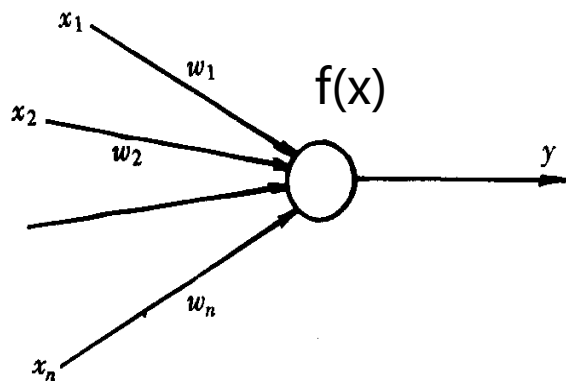
□ 箭头线: “连接”具有‘权值’。



□ 神经元构成:



- +1代表偏移值(偏置项, Bias Units);
- X_1, X_2, X_2 代表初始特征;
- w_0, w_1, w_2, w_3 代表权重(Weight), 即参数, 是特征的缩放倍数; 特征经过缩放和偏移后全部累加起来,
- 此后还要经过一次激活运算然后再输出



- 其中 $x = (x_1, \dots, x_m)^T$ 输入向量, y 为输出, w_i 是权系数;
- 输入与输出具有如下关系:

$$y = f\left(\sum_{i=1}^m w_i x_i - \theta\right)$$

θ 为阈值, $f(X)$ 是激活函数; 它可以是线性函数, 也可以是非线性函数.

例,



$$z = \sum_{i=1}^m w_i x_i - \theta$$

取激活函数为符号函数

$$\text{sgn}(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

则

$$y = f(z) = \begin{cases} 1, & \sum_{i=1}^m w_i x_i > \theta, \\ 0, & \sum_{i=1}^m w_i x_i \leq \theta, \end{cases}$$

- 如果没有激活函数,
多层神经网络退化为一个多层的线性回归模型,
难以学习如图像、音频、文本等复杂数据的特征。
- 神经网络要引入激活函数来给神经网络增加一些非线性的特性,
- 常见的激活函数大多是非线性函数

Sigmoid(S形曲线)

$$y = 1 / (1 + e^{-x})$$

双曲正切函数

$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

□ 影响

①1943年发布的MP模型，简单，建立了神经网络的地基。

然而，MP模型中，权重的值都是预先设置的，因此不能学习

②1949年心理学家Hebb提出了Hebb学习率，

人脑神经细胞的突触（也就是连接）上的强度上可以变化的。

于是科学家们开始考虑用调整权值的方法来让机器学习。

③限于当时的计算机能力，近10年后，第一个真正意义的神经网络才诞生。

□ 1958年，计算科学家Rosenblatt提出了由两层神经元组成的神经网络。

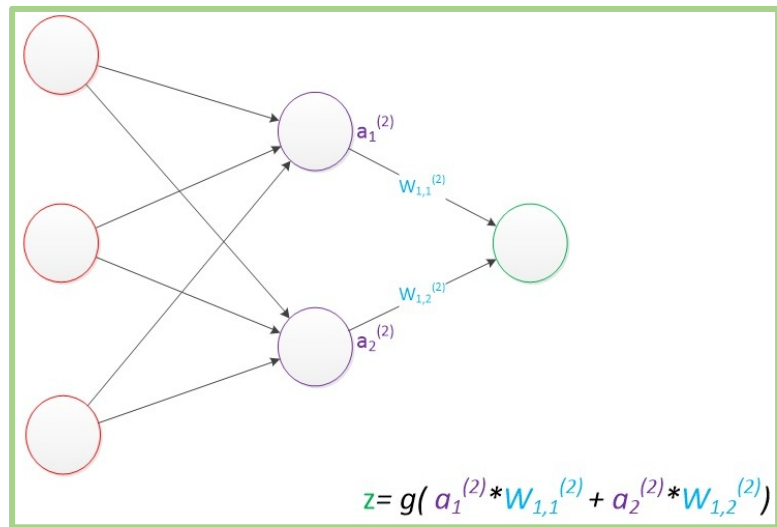
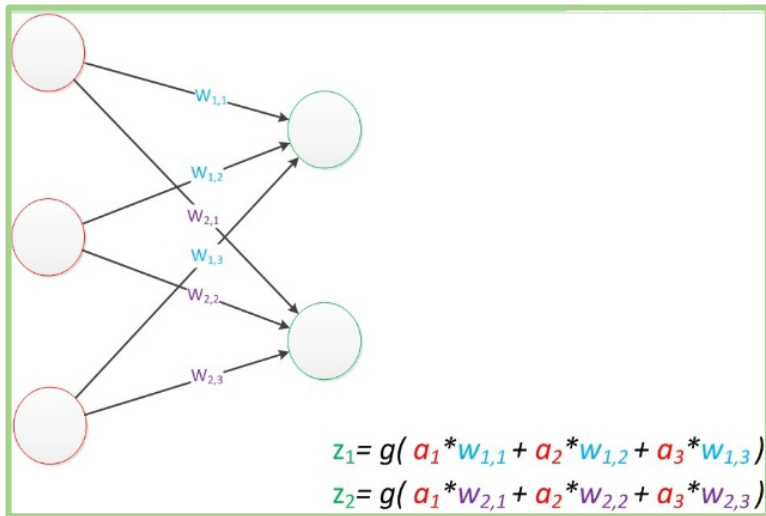
“感知器”（Perceptron）

2 感知器: Perceptron

- 感知器: 多层神经元组成的神经网络

首个可以学习的人工神经网络

- 感知器的结构:
 - 单层感知器(左): 多个M-P模型的累叠
 - 多层感知器(右): 其中一层是固定权值的



感知机分类例子：如何分辨香蕉和苹果

- **香蕉、苹果特征：颜色 p_1 和形状 p_2 , 1代表苹果, -1代表香蕉**

输入 p_1 : 苹果红色(1),香蕉黄色(-1); p_2 苹果圆形(1),香蕉弯形(-1)

预设权重: $w_1=w_2=1$, $b=0$

输出 z : $z=w_1*p_1+w_2*p_2+b$

对苹果的鉴别结果: $z = 1*1+1*1 = 2$;

对香蕉的鉴别结果: $z = -1*1+1*(-1) = -2$;

- 对结果 z 进行处理, 即可实现对二者进行归类

- **问题:**

这里的权重为事先指定的, 如果换一个其它值, 则影响分类结果,

对于随意选取的参数, 如何使输出值依旧正确?

感知器的学习功能

感知器的学习规则:

- 修改感知器的权值 w_i 与偏置 b

$$w_{new} = w_{old} + e * p, \quad b_{new} = b_{old} + e$$

其中 e 误差, $e = t - a$, t 为期望值, a 为实际输出

- 例子, 取 $w_1=1, w_2=-1, b=0$, 苹果的属性 $p = 1$

$$z = p_1 w_1 + p_2 w_2 + b = 0$$

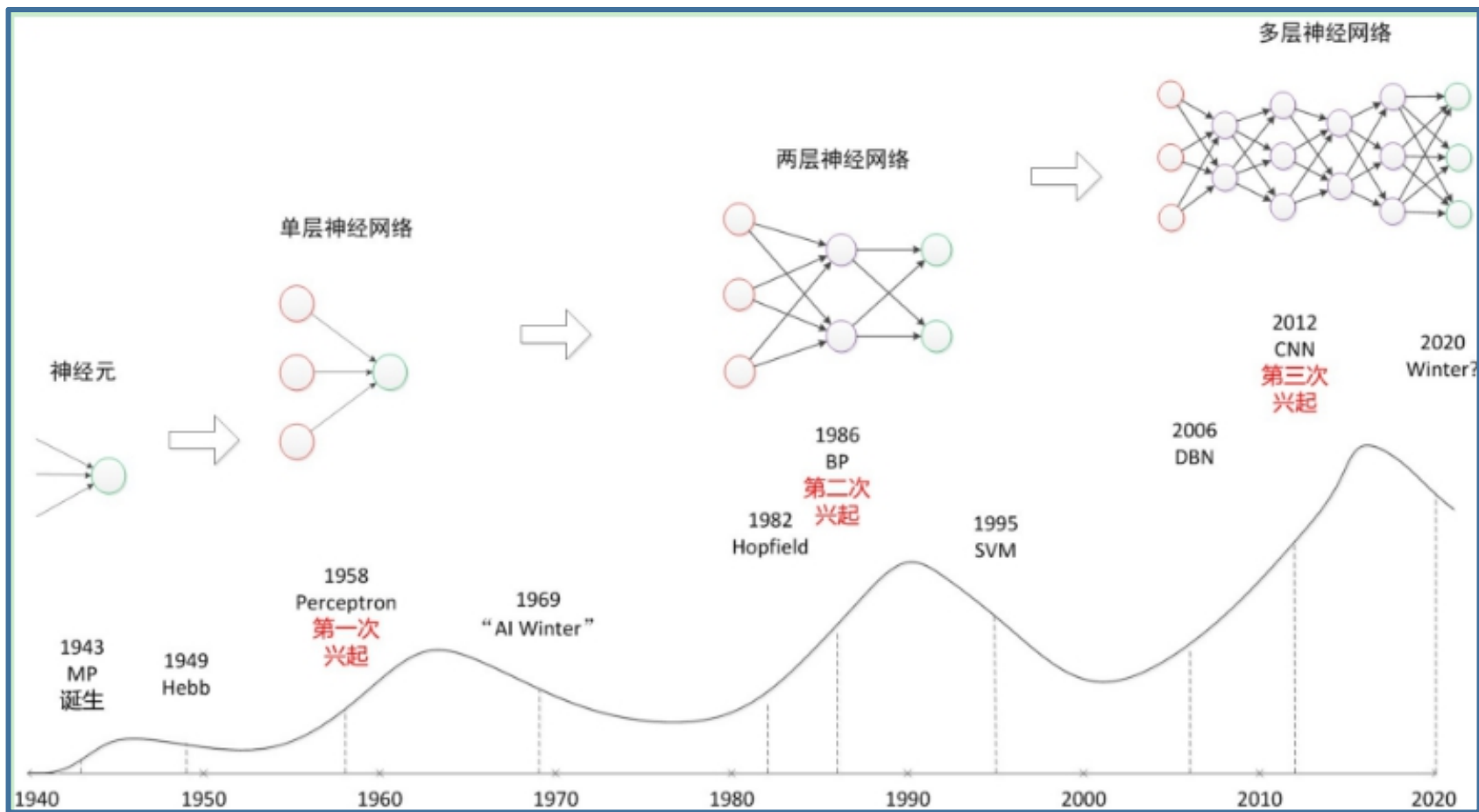
期望结果1, 实际为0, 输出值出现错误

- 学习规则: 计算误差, $e = t - z = 1$
- 代入以上公式重新计算权重 w_i 与偏置 b
- 重新计算输出值
- 依次迭代

- **M-P模型：一个神经元结构，但是没有参数学习的过程**
- 单层感知机引入损失函数，并提出了学习的概念，学习能力有限
- **可以解决简单的线性可分问题，无法处理非线性问题**
- 多层感知机通过增加层数解决非线性问题
需要人为固定一层参数，只能训练其中一层。
- 1986年Hinton提出了反向传播算法，使得训练多层网络成为可能。

3 神经网络

- 大量感知器进行组合
- 本质：通过参数与激活函数来拟合特征与目标之间的真实函数关系
- 神经网络发展历史



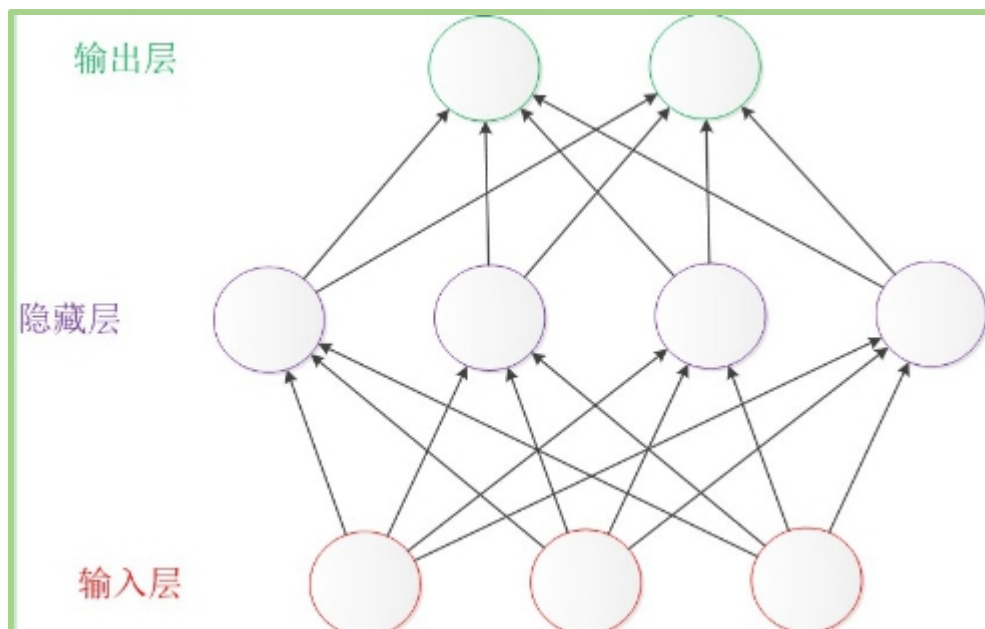
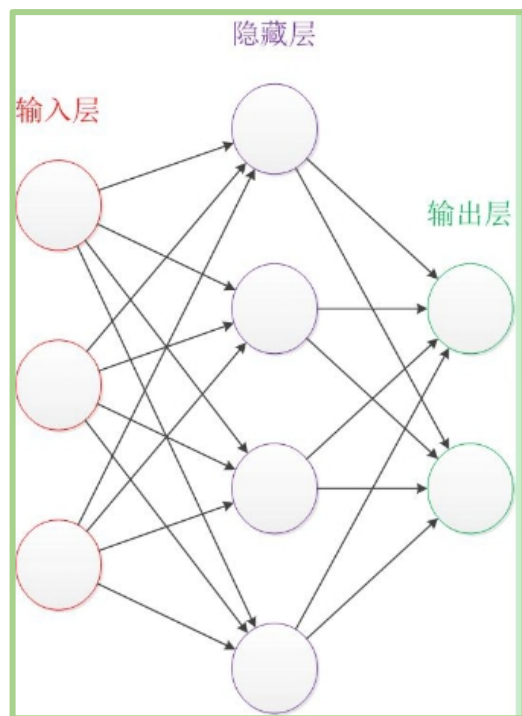


□ 经典神经网络的构成：三个部分(从左至右，从下至上两种结构)

①红色的输入层，3个输入单元

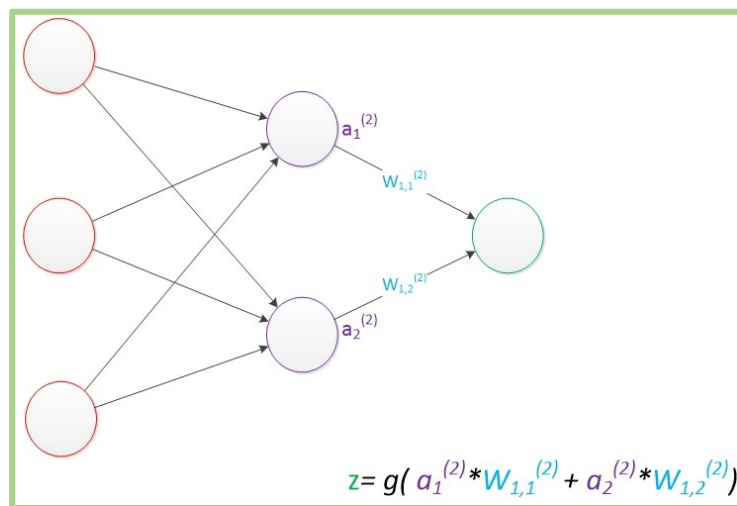
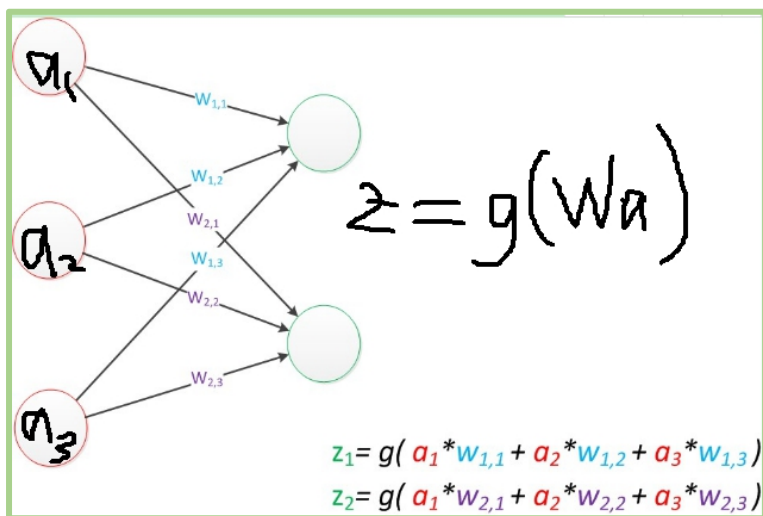
②绿色的是输出层，2个单元

③紫色的是中间层（也叫隐藏层），4个单元。



4 两层神经网络

- 结构：输入层，输出层，中间层



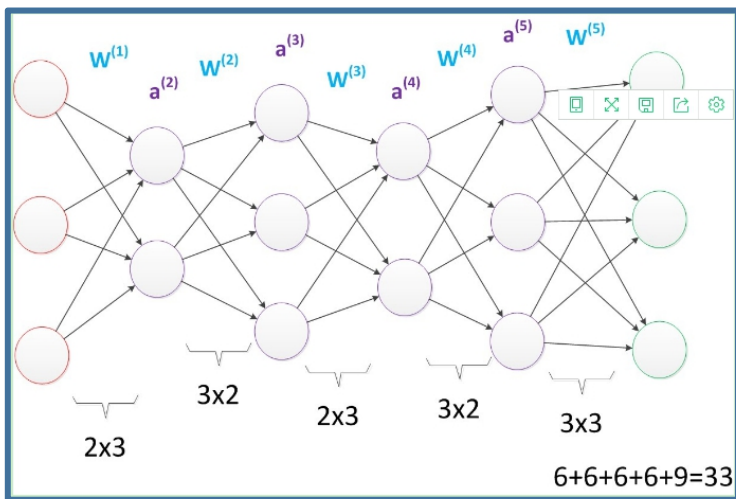
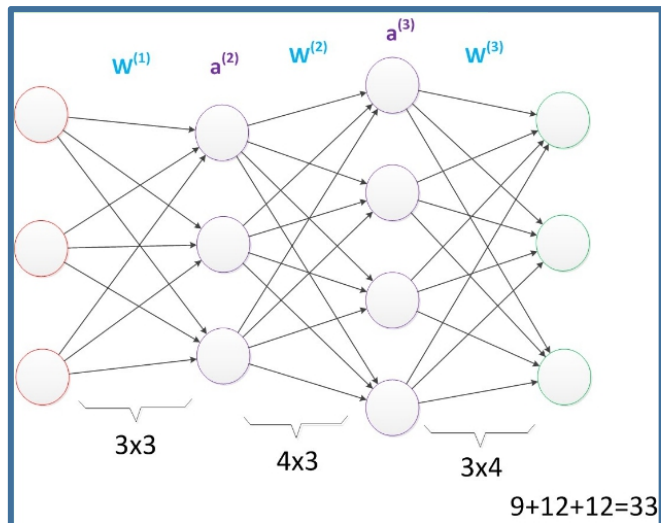
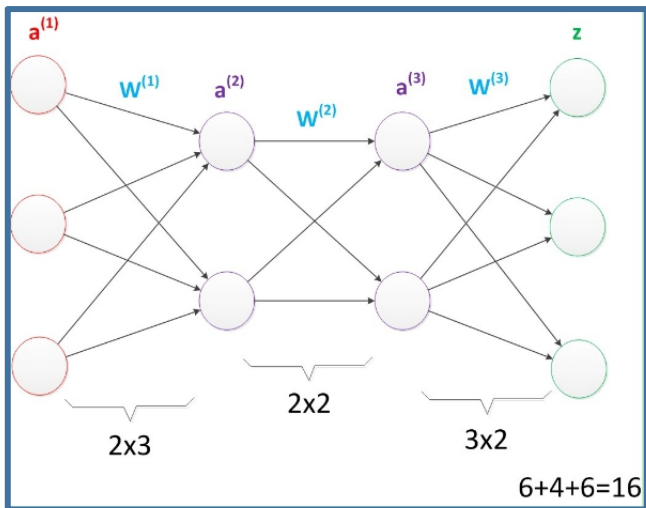
权重矩阵, $a^2 = g(W^1 a)$,

$$z = g(W^2 a^2)$$

5: 多层神经网络与参数

- 多层神经网络中的层数增加了很多。

具有更深入的表示特征，以及更强的函数模拟能力。



参数的意义及求解方案

- 神经网络：模拟特征与目标之间的真实关系的方法，
更多的参数意味着其模拟的函数可以更加的复杂，可以有更多的容量（capacity）去拟合真正的关系
在参数数量一样的情况下，更深的网络往往具有比浅层的网络更好的识别效率
- 学习过程：
不断的修改 w 、 b 两个参数值，使最终的误差达到最小。
- 如何有效的修改这些参数，使误差最小化，是一直在不断尝试解决的问题。
- 80年代，误差反向传播算法(BP算法)的提出，提供了真正有效的解决方案