

# 神经网络基础



# 前言

---

- 机器学习是实现人工智能的必经路径，而深度学习(Deep Learning,DL)作为机器学习(Machine Learning,ML)领域中一个新的研究方向，在计算机视觉、语音、自然语言处理等方向有着重要影响。深度学习源于神经网络的研究，因此本章节将介绍神经网络基础知识。

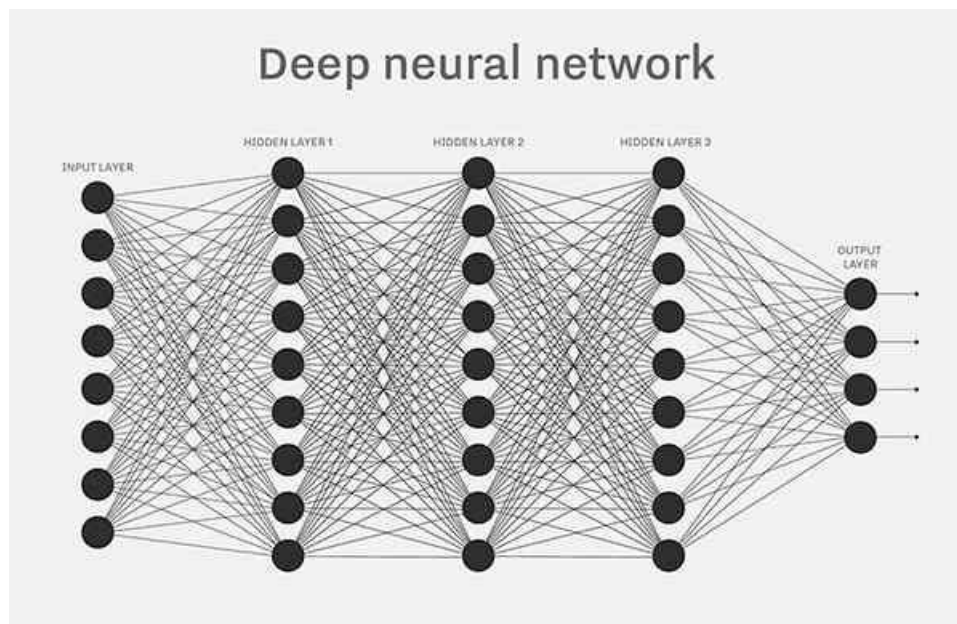
# 目录

---

1. 人工神经网络的定义与结构
2. 感知器与多层感知器
3. 激活函数的作用与种类
4. 神经网络的训练流程

# 人工神经网络定义

- **人工神经网络，简称神经网络（Artificial Neural Network, ANN）**：是由人工神经元互连组成的网络，它是从微观结构和功能上对人脑的抽象、简化，是模拟人类智能的一条重要途径，反映了人脑功能的若干基本特征，如并行信息处理、学习、联想、模式分类、记忆等。





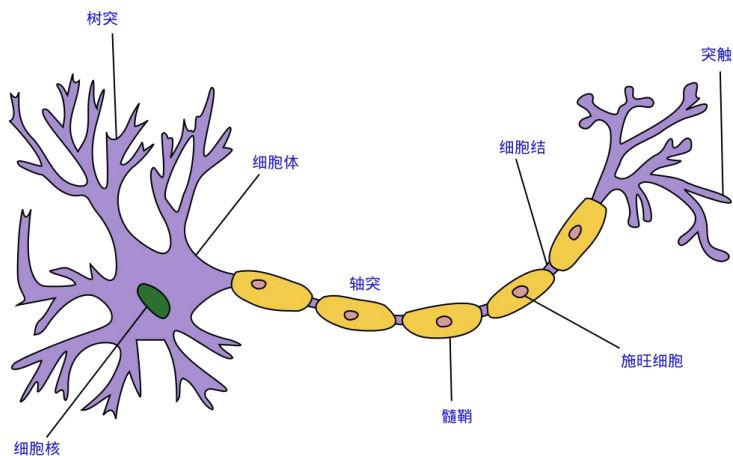
# 生物神经元与人工神经元

- 神经元即神经细胞，是神经系统结构和功能的基本单位。人脑中有数1000亿个神经元，其功能是接受（树突）、整合（细胞体）、传导（轴突）和输出（轴突末梢）信息实现信息转换。
- 人工神经元单元由线性函数和激活函数构成。对于输入 $n$ 维的特征向量 $X = [x_0, x_1, \dots, x_n]^T$ 为输入向量， $W = [\omega_0, \omega_1, \dots, \omega_n]^T$ 为权重， $b$ 为偏置，则：

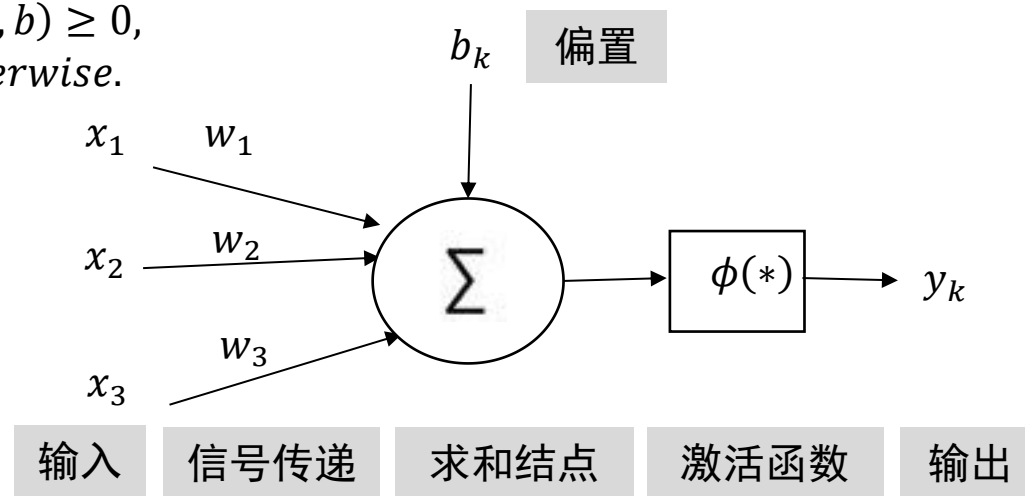
线性函数：

$$f(X, W, b) = WX + b = \sum_n w_i x_i + b = [W; b][X; 1]$$

$$\text{激活函数: } y_k = \phi(f(X, W, b)) = \text{sign}(f(X, W, b)) = \begin{cases} 1, & f(X, W, b) \geq 0, \\ -1, & \text{otherwise.} \end{cases}$$



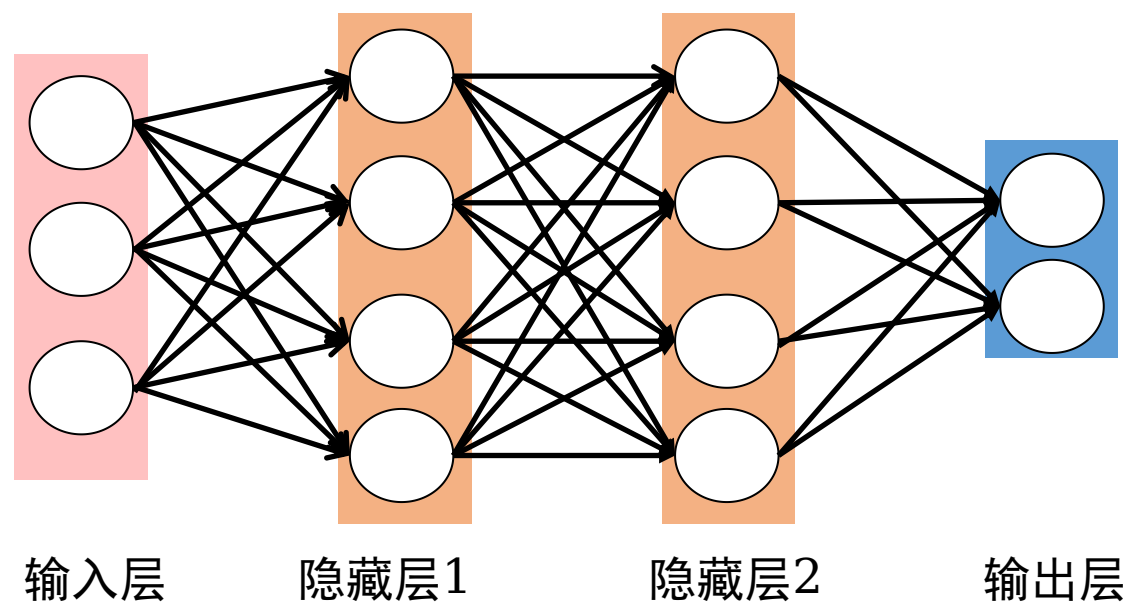
生物神经元



人工神经元

# 人工神经网络的结构

- 单个神经细胞功能比较简单，需要通过很多神经元一起协作完成复杂功能，通过一定的连接方式或信息传递方式进行协作的神经元可以看做是一个网络，就是人工神经网络。  
人工神经网络包括输入层、隐藏层和输出层。



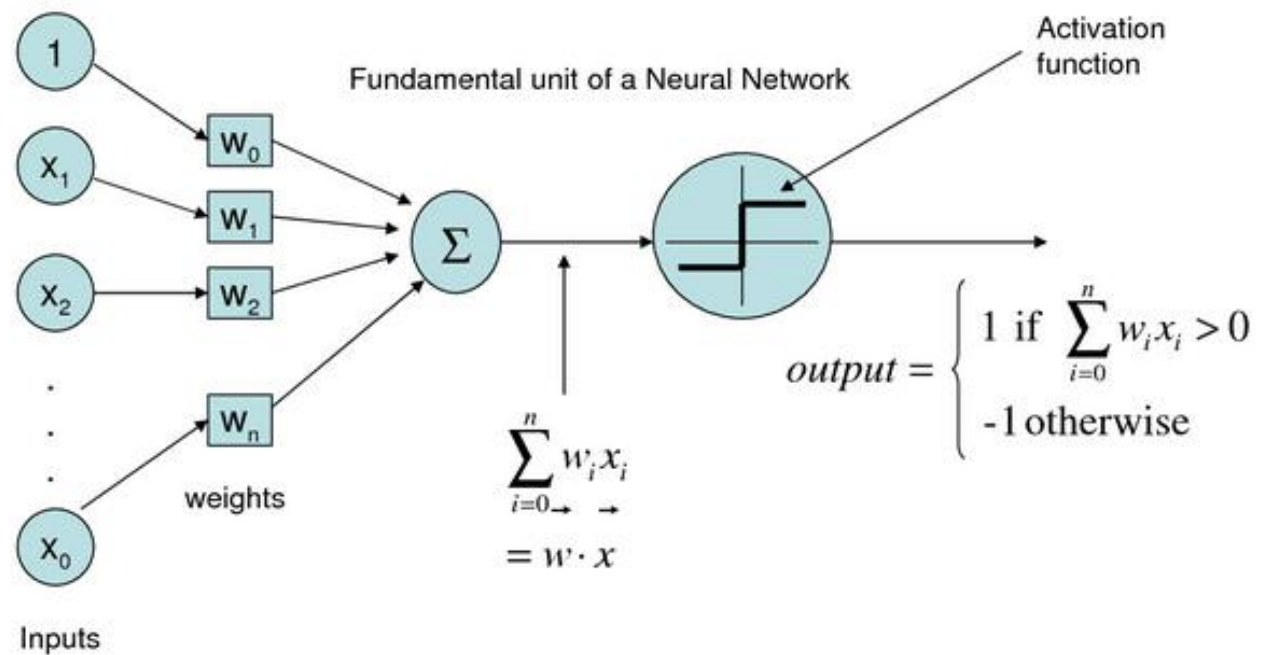
# 目录

---

1. 人工神经网络的定义与结构
- 2. 感知器与多层感知器**
3. 激活函数的作用与种类
4. 神经网络的训练流程

# 感知器

- 感知器（Perceptron）是Frank Rosenblatt在1957年所发明的一种最简单的人工神经网络，即单层神经网络。

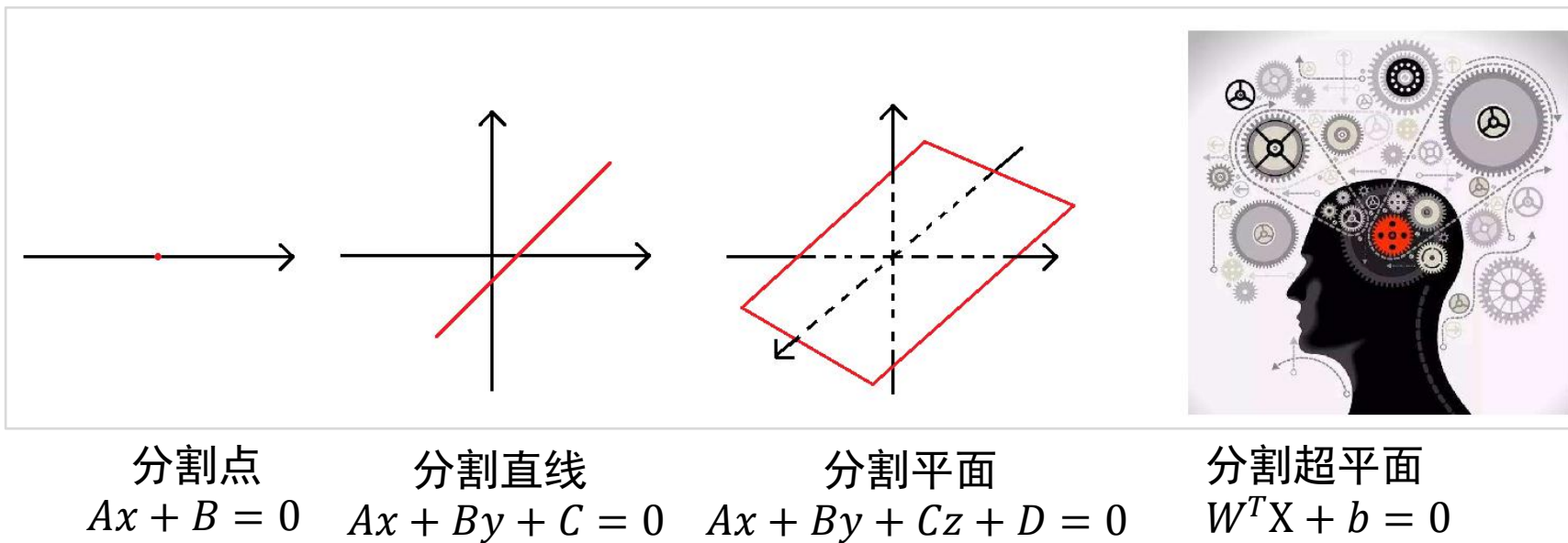


单层感知器

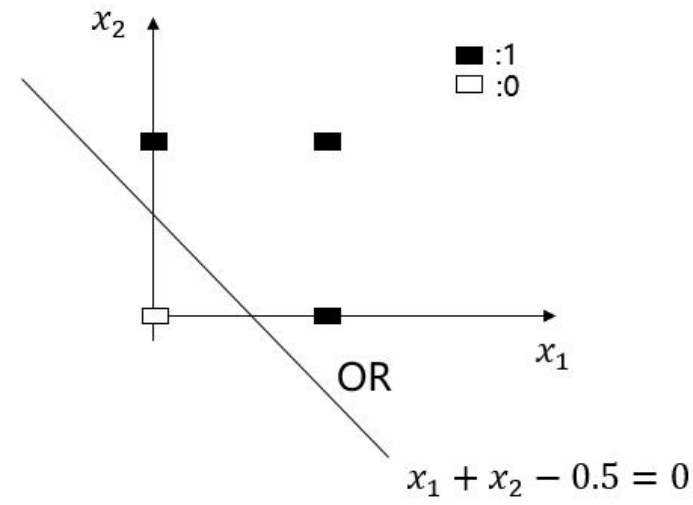
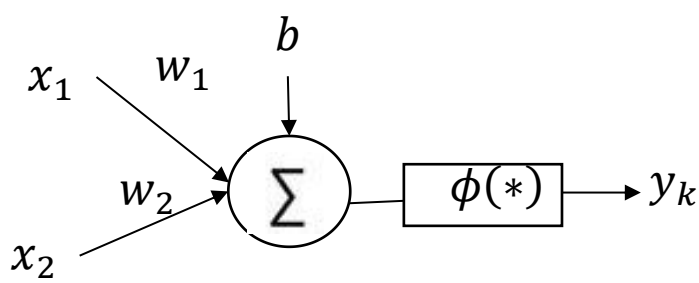
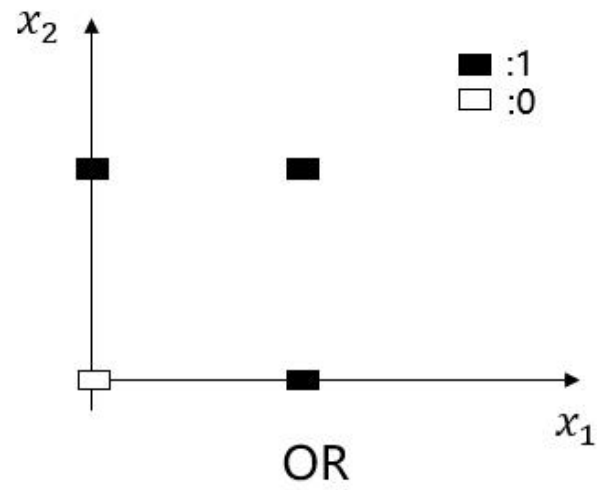


# 感知器的作用

- 单层感知器本质上相当于一个二分类器。



# 感知器处理线性分类



X1	X2	OR
1	0	1
1	1	1
0	1	1
0	0	-1

将样本代入感知器

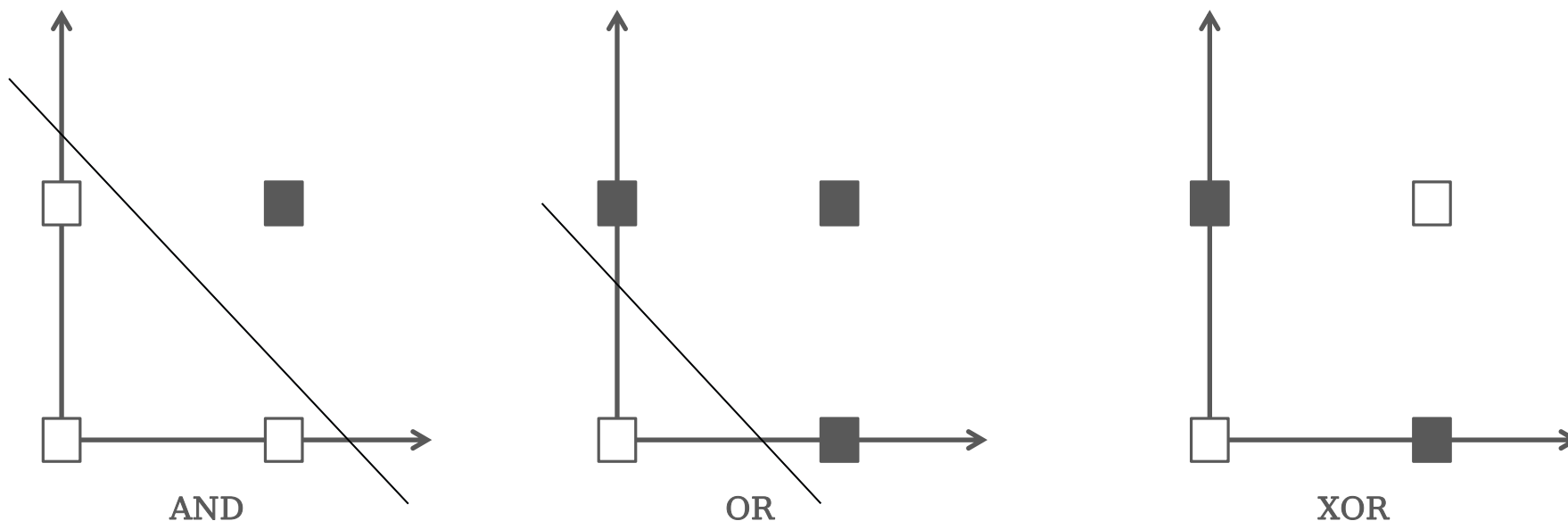
$1 = \phi(w_1 * 1 + w_2 * 0 + b)$   
 $1 = \phi(w_1 * 1 + w_2 * 1 + b)$   
 $1 = \phi(w_1 * 0 + w_2 * 1 + b)$   
 $-1 = \phi(w_1 * 0 + w_2 * 0 + b)$

求得权重和偏重的范围

$w_1 > 0$   
 $w_2 > 0$   
 $b < 0$

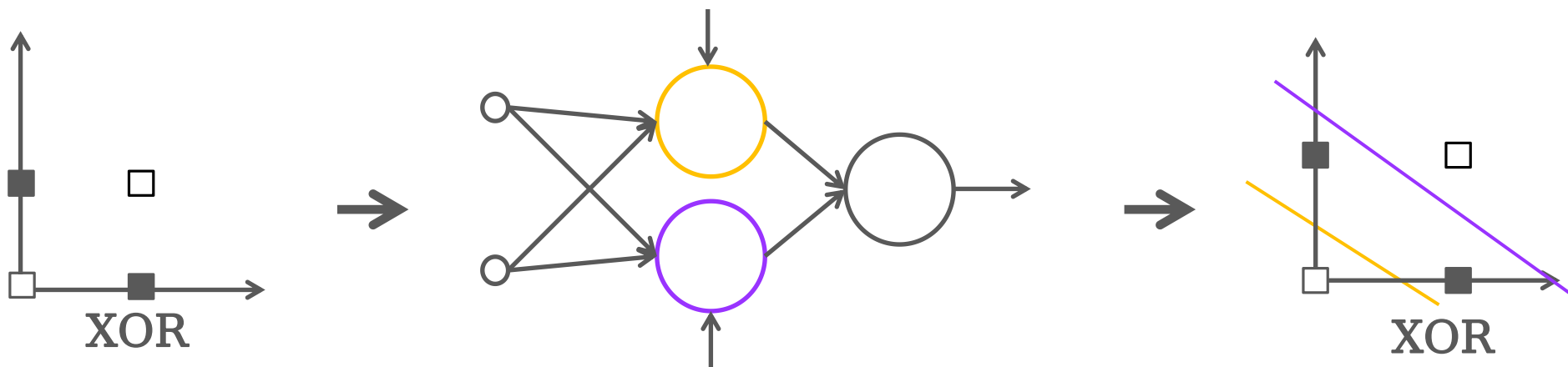
# XOR问题

- 1969年，美国数学家及人工智能先驱Minsky在其著作中证明了感知器本质上是一种线性模型，只能处理线性分类问题，无法处理非线性数据。



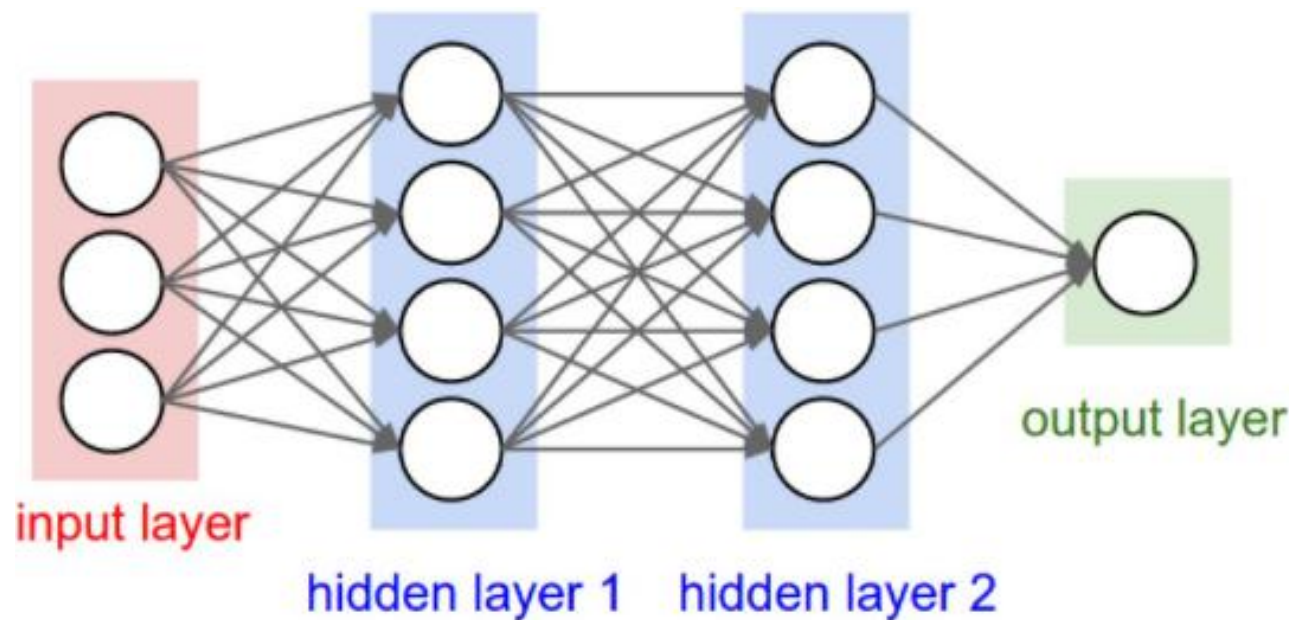
# 感知器的局限性

- 感知器的局限性：无法处理非线性分类。但是发现可以通过添加**隐藏层**来实现非线性分类。如下所示：



# 多层感知器

- 多层感知器（Multilayer Perceptron,MLP）是一种前向结构的人工神经网络，映射一组输入向量到一组输出向量。MLP可以被看作是一个有向图，由多个的节点层所组成，每一层都全连接到下一层。除了输入节点，每个节点都是一个带有非线性激活函数的神经元。





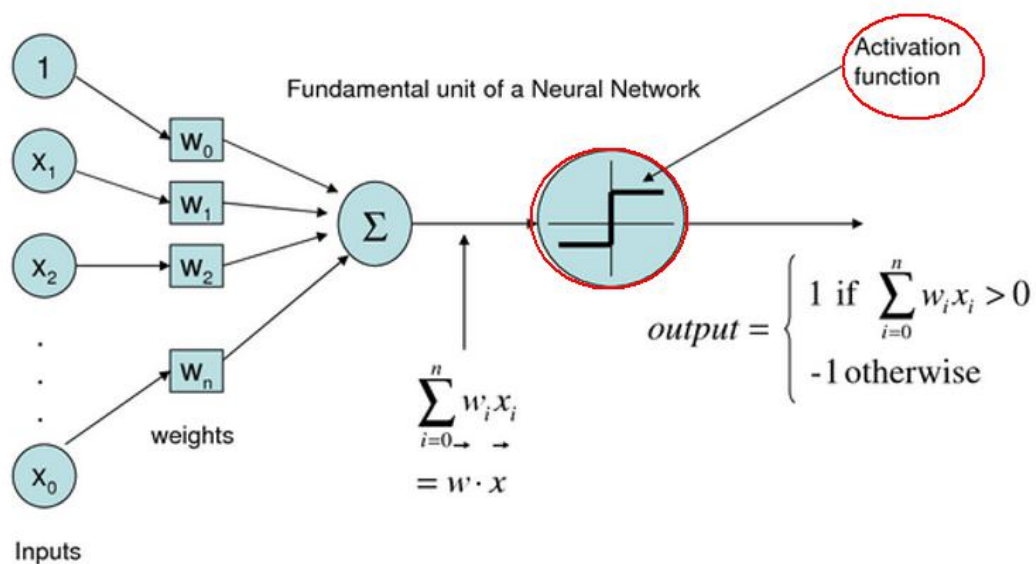
# 目录

---

1. 人工神经网络的定义与结构
2. 感知器与多层感知器
- 3. 激活函数的作用与种类**
4. 神经网络的训练流程

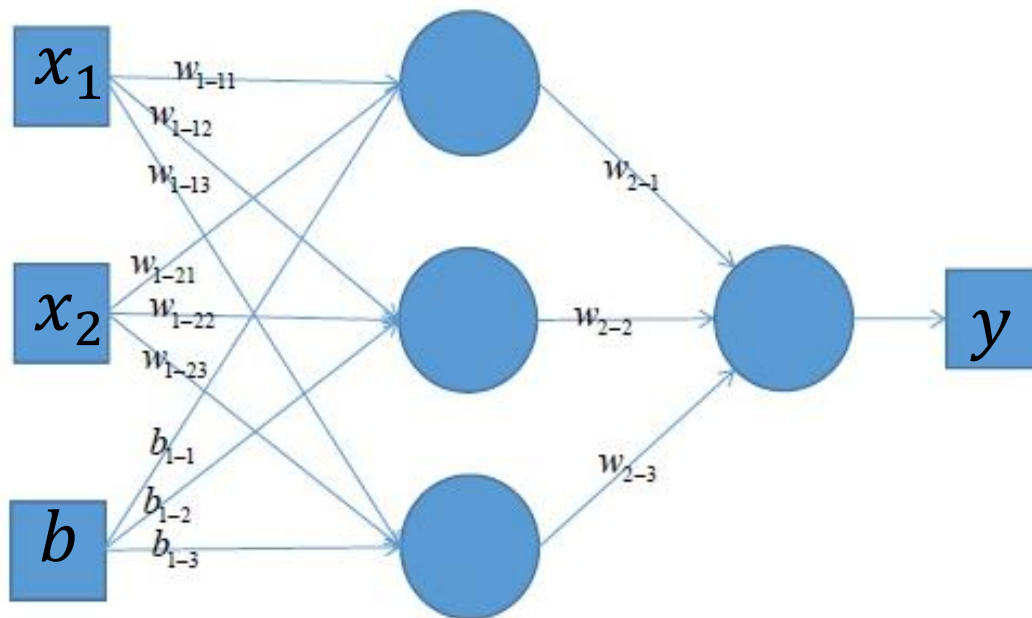
# 激活函数的概念

- 神经网络中的每个神经元节点接受上一层神经元的输出值作为本神经元的输入值，并将输入值计算后传递给下一层，输入层神经元节点会将输入属性值直接传递给下一层（隐层或输出层）。在多层神经网络中，上层节点的输出和下层节点的输入之间具有一个函数关系，这个函数称为**激活函数**（又称**激励函数**）。

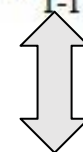


单层感知器

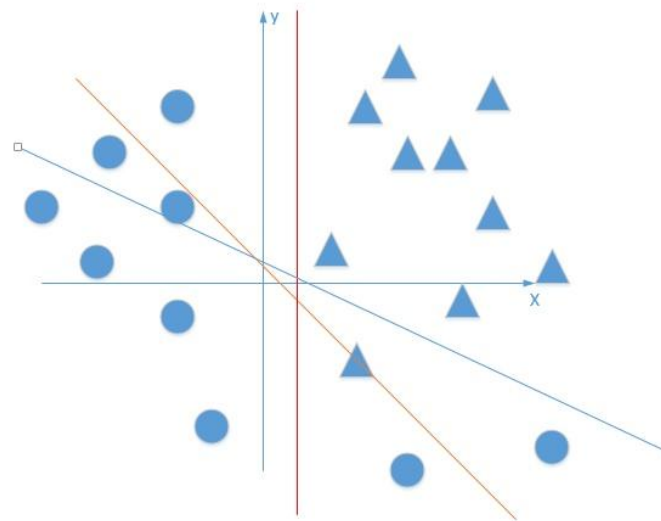
# 激活函数的作用 (1)



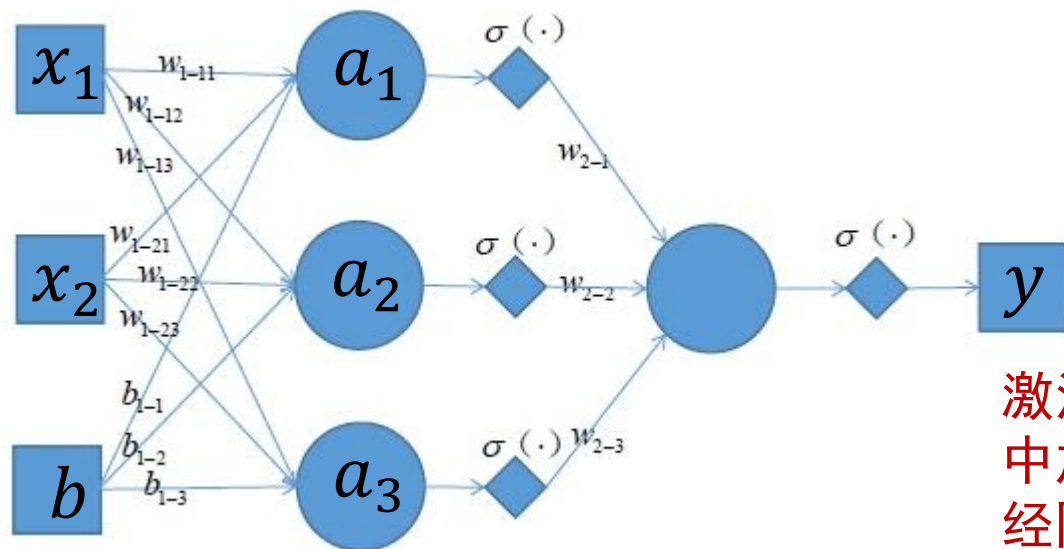
$$y = w_{2-1}(w_{1-11}x_1 + w_{1-21}x_2 + b_{1-1}) \\ + w_{2-2}(w_{1-12}x_1 + w_{1-22}x_2 + b_{1-2}) \\ + w_{2-3}(w_{1-13}x_1 + w_{1-23}x_2 + b_{1-3})$$



不带激活函数的多个感知机等价于线性函数

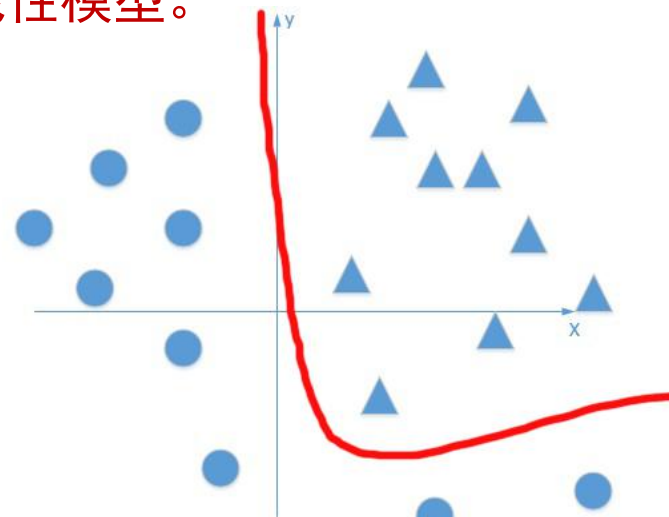


## 激活函数的作用 (2)

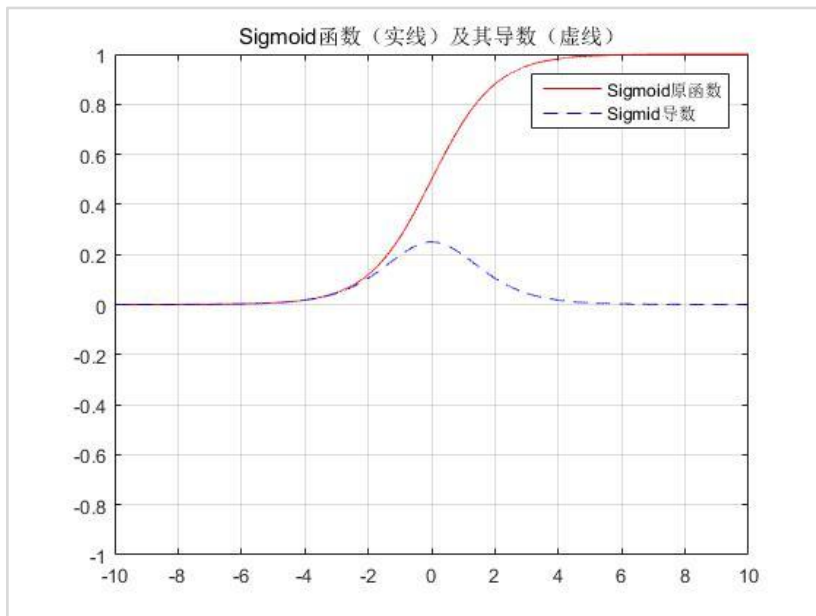


$$\begin{aligned}a_1 &= w_{1-11}x_1 + w_{1-21}x_2 + b_{1-1} \\a_2 &= w_{1-12}x_1 + w_{1-22}x_2 + b_{1-2} \\a_3 &= w_{1-13}x_1 + w_{1-23}x_2 + b_{1-3} \\y &= \sigma(w_{2-1}\sigma(a_1) + w_{2-2}\sigma(a_2) + w_{2-3}\sigma(a_3))\end{aligned}$$

激活函数相当于在神经网络中加入了非线性因素，使神经网络变为非线性模型。

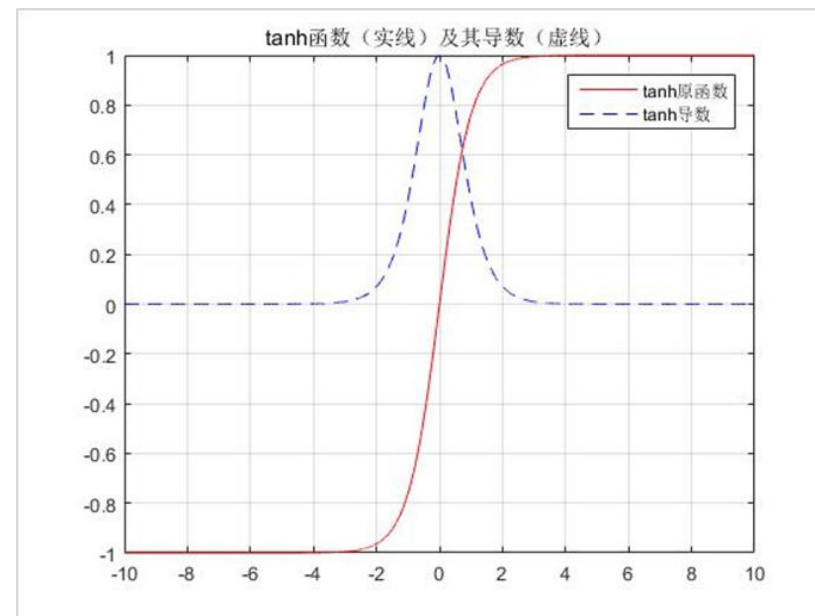


# 激活函数 (1)



Sigmoid函数

$$f(x) = \frac{1}{1 + e^{-x}}$$

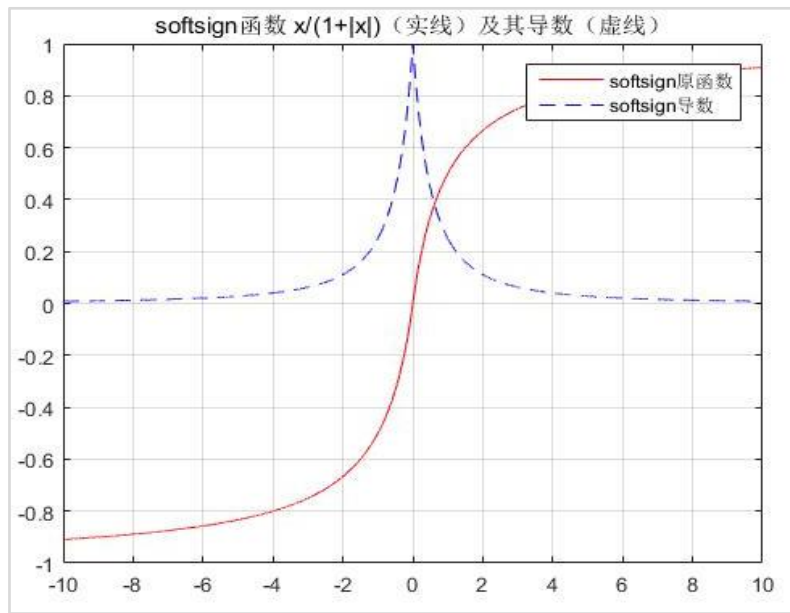


Tanh函数

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

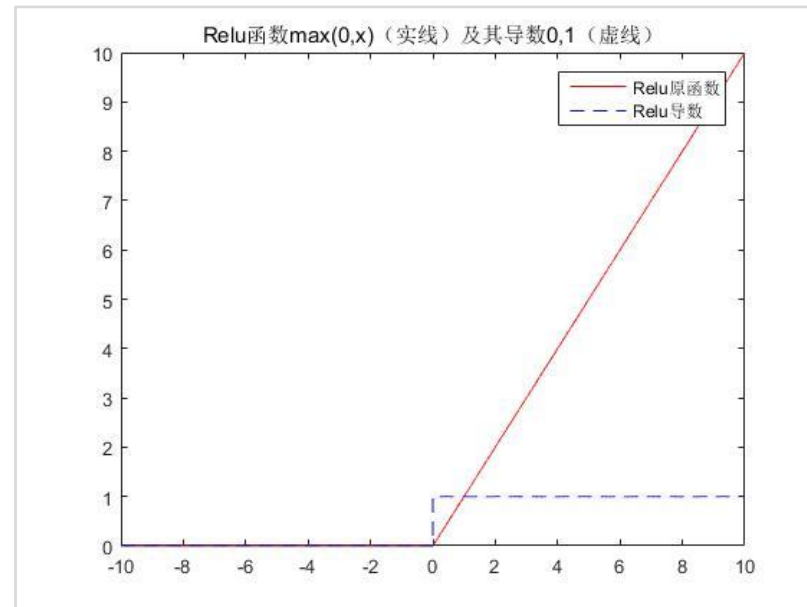


## 激活函数 (2)



Softsign函数

$$f(x) = \frac{x}{1 + |x|}$$



ReLU函数

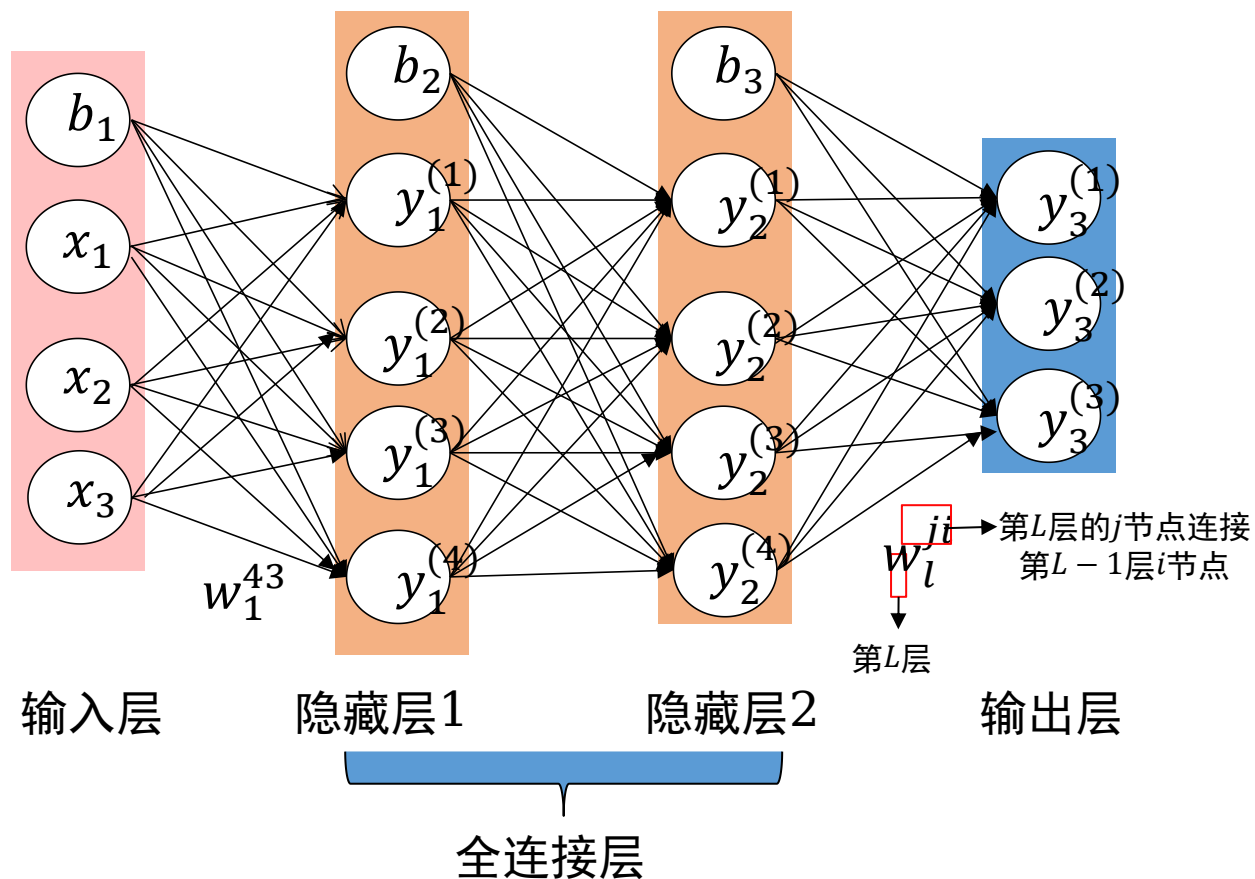
$$f(x) = \max(0, x)$$

# 目录

---

1. 人工神经网络的定义与结构
2. 感知器与多层感知器
3. 激活函数的作用与种类
- 4. 神经网络的训练流程**

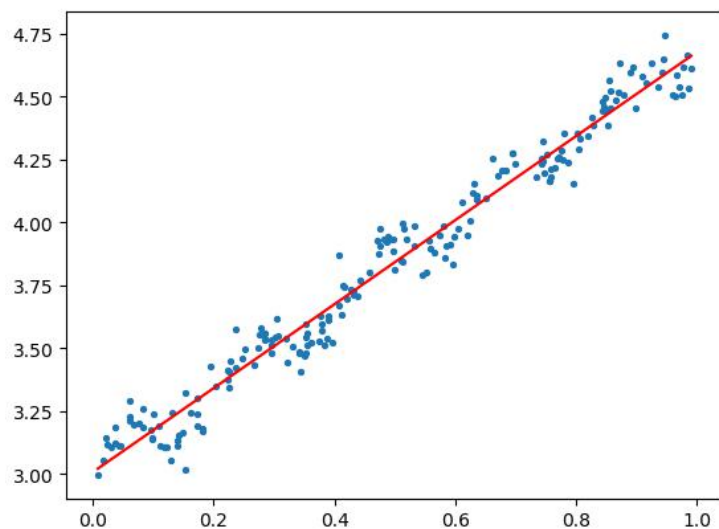
# 前馈神经网络的结构



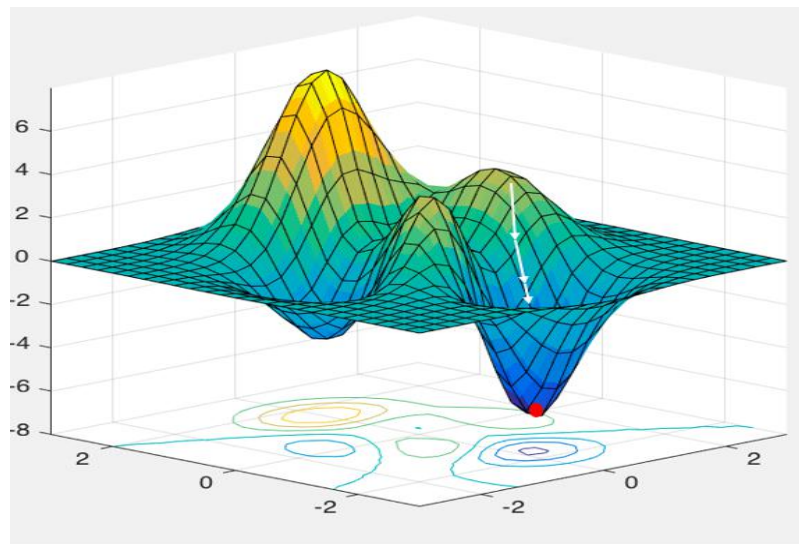
- 前馈神经网络（Feedforward Neural Network）是典型的深度学习模型。具有以下特点：
  - 输入节点并无计算功能，只是为了表征输入矢量各元素值。
  - 每个神经元只与前一层的神经元相连；接收前一层的输出，并输出给下一层。采用一种单向多层结构；整个网络中无反馈，信号从输入层向输出层单向传播，可用一个有向无环图表示。

# 损失函数

- 机器学习中使用的代价函数被称作损失函数（Loss Function）。
  - 损失函数衡量了评分函数的预测与真实样本标签的吻合度。
  - Loss的值都会设置为和吻合程度负相关。如果算法公式是正相关，定义损失函数时候加负号，调整为负相关。
- 机器学习在数据集上学习，就是通过对训练数据集的学习，调整模型的参数，使得损失函数的值取得最低或者相对较低的值。



损失函数作用说明



损失函数



# 深度学习中常用的损失函数

- 二次代价函数：

$$E(W) = \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

- 交叉熵代价函数：

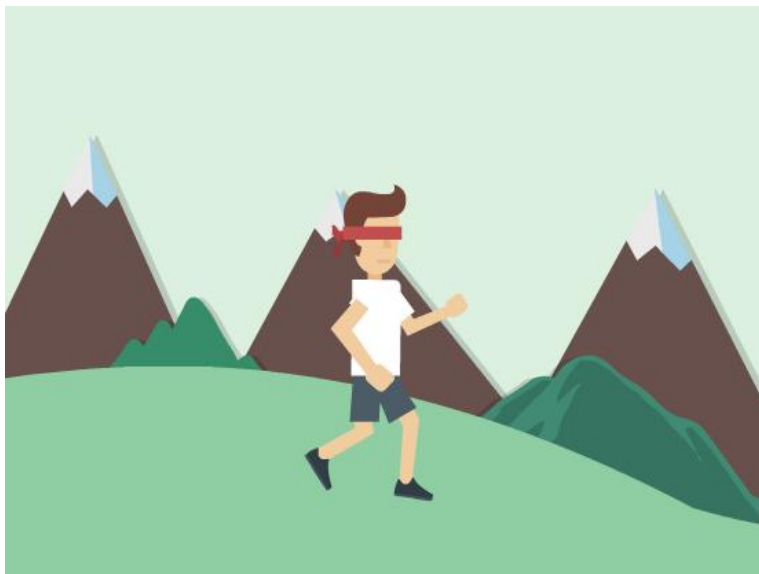
$$E(W) = -\frac{1}{n} \sum_x \sum_{d \in D} [t_d \ln o_d]$$

- 交叉熵误差刻画了两个概率分布之间的距离，是分类问题中使用较多的一种损失函数。
- 一般二次代价函数更多得用于回归问题，而交叉熵误差更多的用于分类问题。

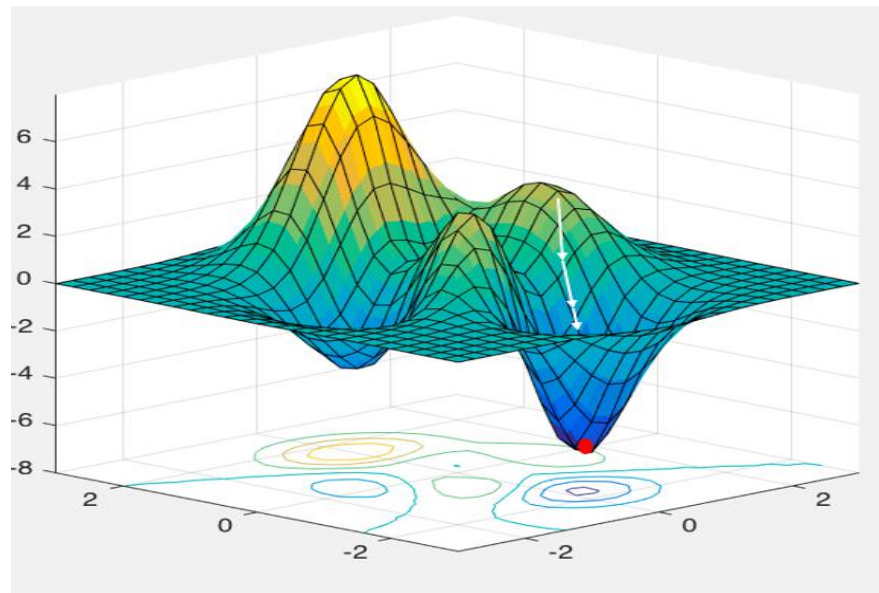


# 梯度下降法 (1)

- **梯度下降法的目的：** 让损失函数沿着负梯度的方向进行搜索，不断迭代更新参数，最终使得损失函数最小化。
- **解决思路：** 负梯度方向是函数下降最快的方向，那我们可以从某个点开始，沿着 $-\nabla E(W)$ 方向一路前行，期望最终可以找到 $E(W)$ 的极小值点，这就是梯度下降法的核心思想。



蒙着眼睛下山



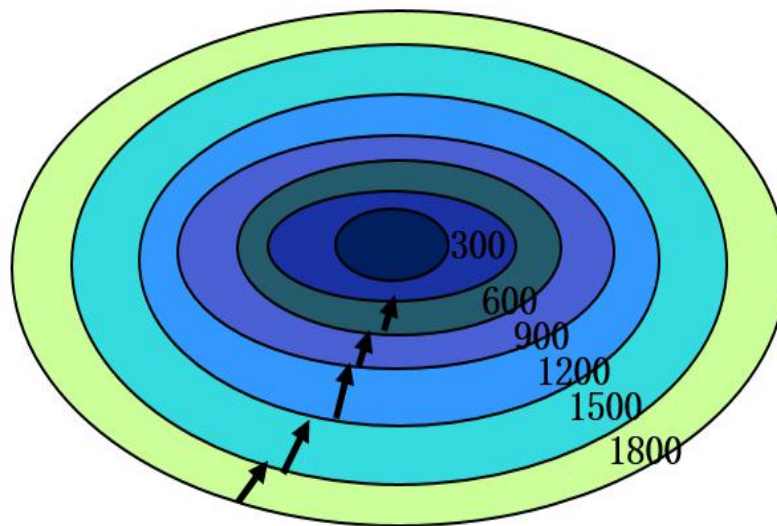
损失函数

## 梯度下降法 (2)

- 学习率(Learning Rate, LR), 根据误差梯度调整权重数值的系数, 通常记作 $\eta$ 。

$$w' = w - \eta * \frac{\partial Loss}{\partial w}$$

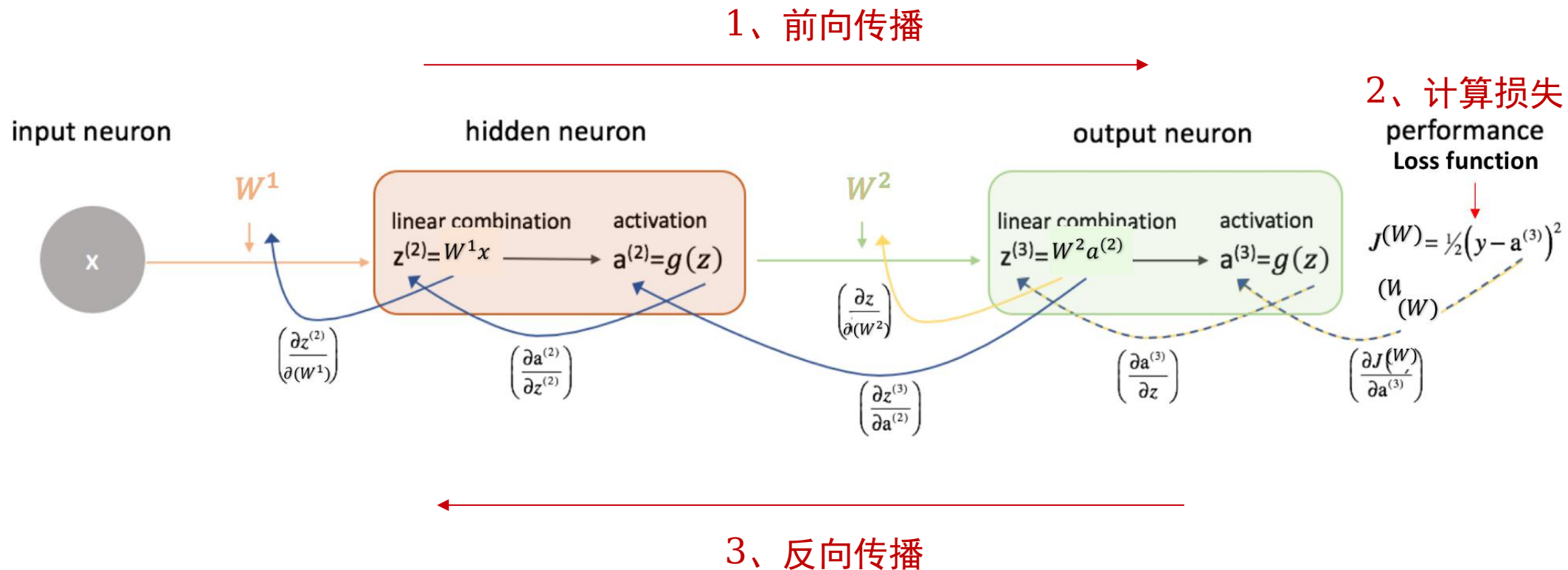
- 通过学习率和梯度值更新所有参数值使得网络的损失函数值降低。



# 网络训练流程 (1)



# 网络训练流程 (2)



# 反向传播算法

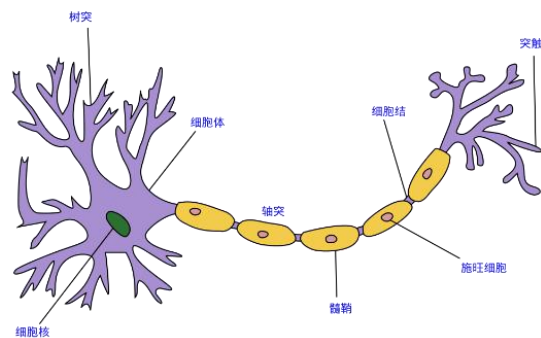
- 误差反向传播算法（Error Back Propagation）是神经网络的重要算法。它使用链式求导法则将输出层的误差反向传回给网络，使神经网络的权重有了较简单的梯度计算实现方法。其步骤为：
  - 将损失函数值反向传播给每个计算单元；
  - 每个计算单元根据获得的误差更新权重。
- 梯度下降算法的参数更新公式： $W_{t+1} = W_t - \eta \frac{dC}{dW}$ ，其中  $C = J(W)$  为损失函数。

$$\begin{aligned}\frac{\partial C}{\partial w^L} &= \frac{\partial z^L}{\partial w^L} \frac{\partial a^L}{\partial z^L} \frac{\partial C}{\partial a^L} = a^{L-1} \underbrace{\sigma'(z^L)(a^L - y)}_{\text{delta}} \\ \frac{\partial C}{\partial a^{L-1}} &= \frac{\partial z^L}{\partial a^{L-1}} \frac{\partial a^L}{\partial z^L} \frac{\partial C}{\partial a^L} = \underbrace{w^L \sigma'(z^L)(a^L - y)}_{\text{new delta : } w^L * \text{tmp, and NOT } w^L * \text{delta}} \\ \frac{\partial C}{\partial w^{L-1}} &= \frac{\partial z^{L-1}}{\partial w^{L-1}} \frac{\partial a^{L-1}}{\partial z^{L-1}} \frac{\partial C}{\partial a^{L-1}} = a^{L-2} \sigma'(z^{L-1}) \times \underbrace{w^L \sigma'(z^L)(a^L - y)}_{\text{new delta}} \\ &\quad \underbrace{\hspace{10em}}_{\text{new tmp}}\end{aligned}$$

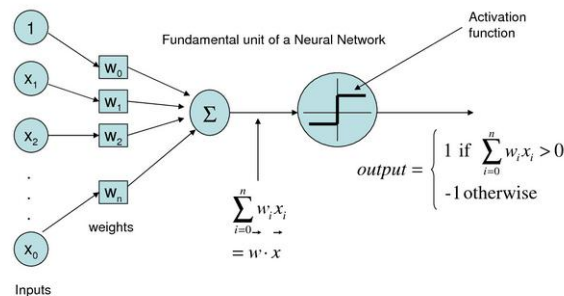


# 深度神经网络

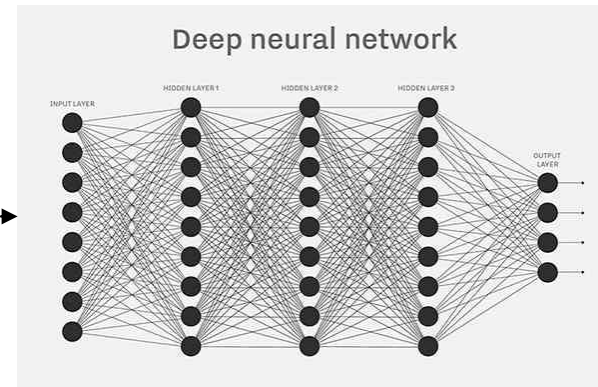
- 深度神经网络（Deep Neural Network, DNN）：隐藏层比较多（大于2）的神经网络。而深度学习，就是使用深度神经网络架构的机器学习方法。



生物神经元



单层感知器



深度神经网络

# 思考题

---

1. 【多选题】设计网络结构需要考虑哪些因素？（）
- A. 隐藏层激活函数
  - B. 损失函数
  - C. 隐藏层层数
  - D. 输出层激活函数

# 本章总结

---

- 本章节主要介绍了神经网络基础包括人工神经网络的定义与结构、感知器与多层感知器、激活函数的作用与种类、前馈神经网络、神经网络的训练流程。

# Thank you.

把数字世界带入每个人、每个家庭、  
每个组织，构建万物互联的智能世界。  
Bring digital to every person, home, and  
organization for a fully connected,  
intelligent world.

**Copyright©2021 Huawei Technologies Co., Ltd.  
All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

