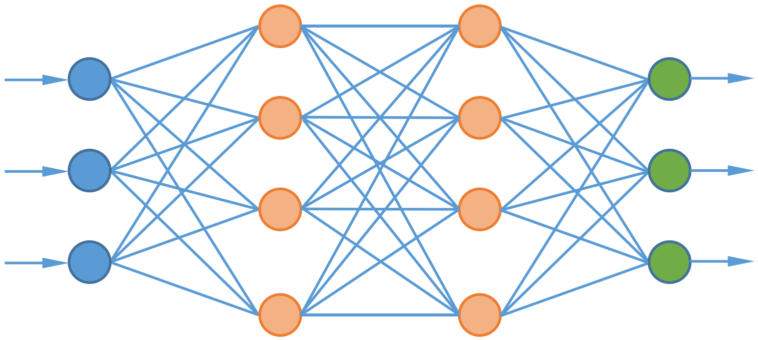
12.5 激活函数

中国大学MOOC

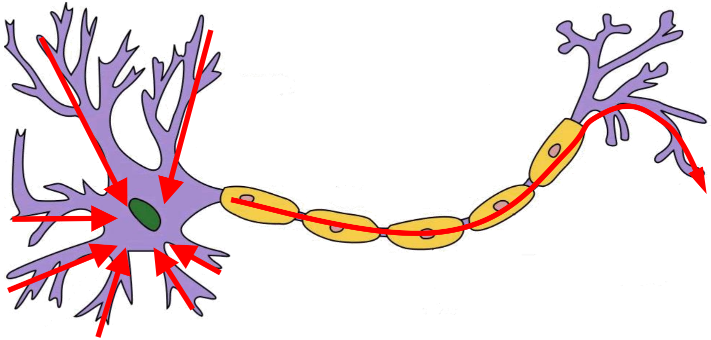
生物神经网络

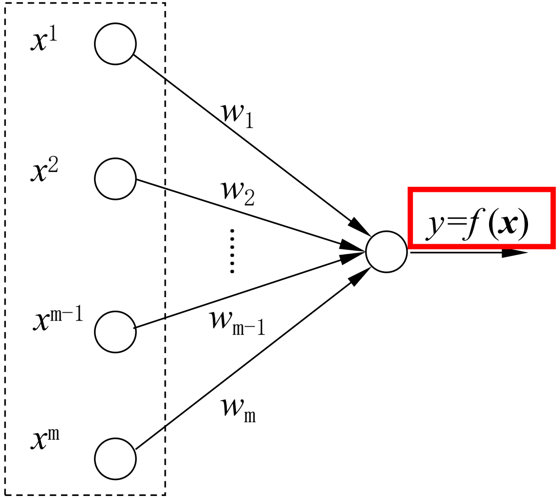
人工神经网络

由人工**神经元**和神经元之间的**连接**构成的一种**数学模型**



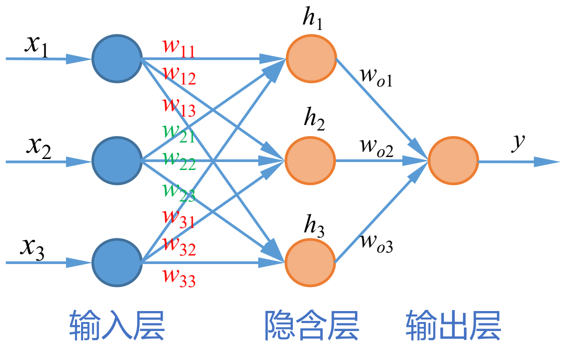
□ 激活函数的作用





线性函数

无论有多少层隐含层，输出都是输入特征的**线性组合**



隐含层

$$h_1 = \sum_{j=1}^{3} w_{1j} x^j + w_{10} = W_1^T X$$

$$h_2 = \sum_{j=1}^{3} w_{2j} x^j + w_{20} = W_2^T X$$

$$h_3 = \sum_{j=1}^{3} w_{3j} x^j + w_{30} = W_3^T X$$

输出层

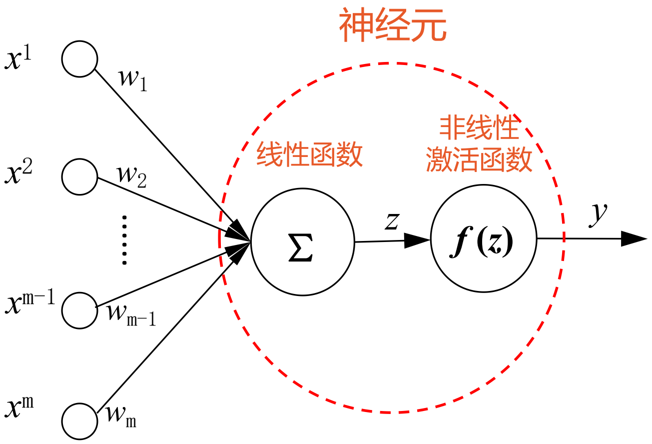
$$y = \sum_{j=1}^{3} w_{oj} h_j + w_o = W_o H_j$$

$$= W_o (W_1^T X + W_2^T X + W_3^T X)$$

任意函数

- 灵活、功能强大；
- 过于复杂，难于训练；
- 通用性差

12.5 激活函数



- 所有输入的**线性组合**

$$z=\sum_{j=1}^{m} w_j x^j + w_0 = W^T X$$

- 使用一个**非线性**的**激活函数**

$$y=f(z)=f(W^T X)$$

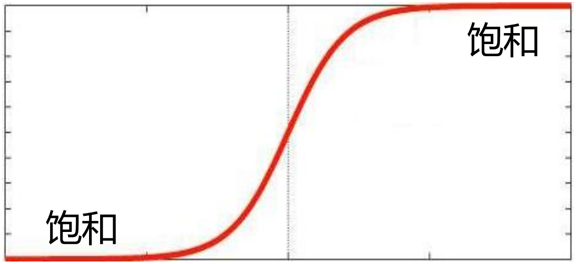
激活函数的性质

- 简单的非线性函数
- 连续并可导
- 单调函数

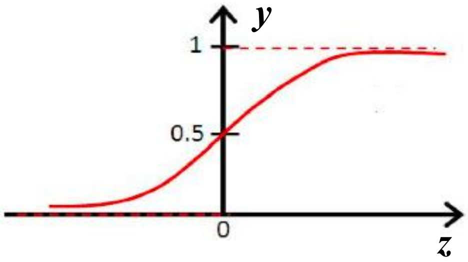
西安科技大学
计算机科学与技术学院

常用的激活函数

- sigmoid函数

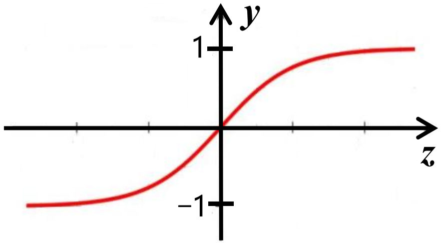


- 对数几率函数 logistic



$$\sigma(z)=\frac{1}{1+e^{-z}}$$

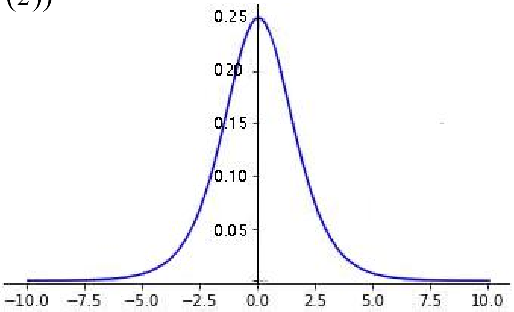
- 双曲正切函数 Tanh



$$tanh(z)=\frac{sinh\,z}{cosh\,z}=\frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}$$

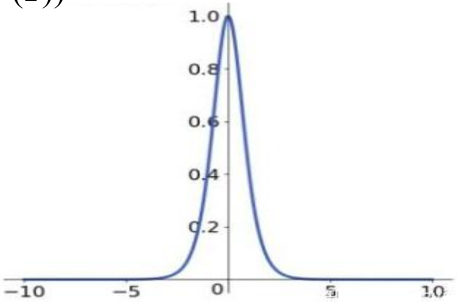
logistic函数的导函数

$$\sigma'(z) = \sigma(z)(1-\sigma(z))$$



Tanh函数的导函数

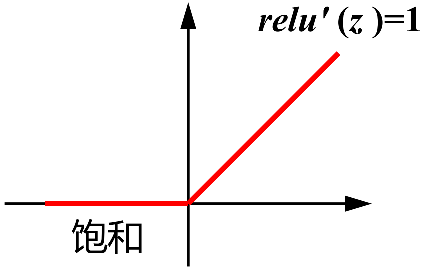
$$tanh'(z) = 1 - (tanh(z))^2$$



梯度消失问题

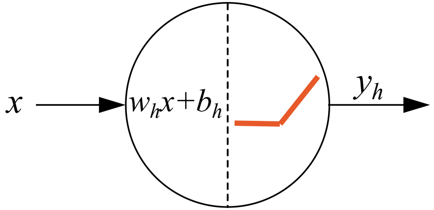
- 多层神经网络，误差反向传播
- 使用链式法则，误差经过每一层会不断衰减
- 网络层数较深时，梯度值趋近于0，参数更新几乎停滞

ReLU函数（Rectified Linear Unit，修正线性单元）



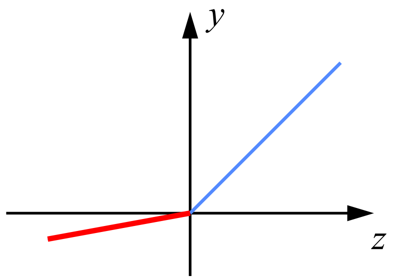
$$y = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases}$$

$$= max(0, z)$$



- z>0时，导数等于1，缓解了梯度消失问题
- 不存在幂运算，计算速度快
- 导数恒等于1，训练模型收敛速度快
- 输出不是以0为均值的，会影响收敛的速度
- z<0时，梯度为0，神经元死亡

Leaky-ReLU函数



$$y=\begin{cases} z & z\geq 0 \\ \dfrac{z}{\boxed{a}} & z<0 \end{cases} \qquad a\in(1,+\infty)$$

- 避免了ReLU神经元死亡
- 神经网络的计算和训练速度快
- 超参数a需要人工调整

PReLU函数 和 RReLU函数

PReLU（Parameteric Rectified Linear Unit，参数化修正线性单元）

$$y = \begin{cases} z & z \geq 0 \\ \boxed{\alpha} z & z < 0 \end{cases}$$

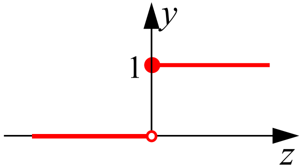
α：可训练参数

RReLU (Randomized Leaky Rectified Linear Unit，随机纠正线性单元)

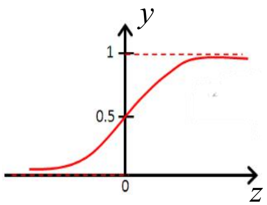
- 训练阶段，负值部分的斜率是随机分配的（均匀分布）
- 测试阶段，负值的斜率是固定的（训练阶段所有α的平均值）

□ 激活函数的特点

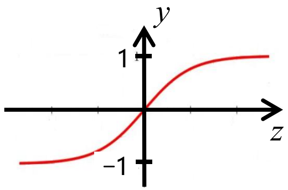
- 包含充分的梯度信息
- 能够识别阈值



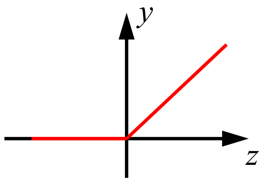
阶跃函数



Logistic函数



Tanh 函数



ReLU 函数