No.

Date.

模式识别第三周作业

何东阳

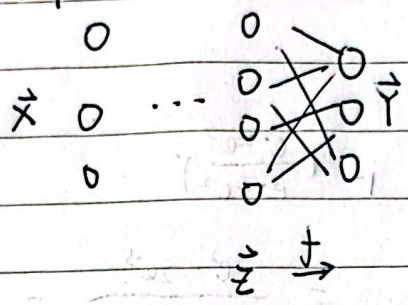
4.1 证明：假设对于一个网络F

映射关系是 $\vec{y}=F(\vec{x})$

其中$\vec{y}$是m维向量$(y_1, y_2 \cdots, y_m)$，$\vec{x}$是n维向量$(x_1, x_2, \cdots, x_n)$

假设网络最后一层的映射关系是 $\vec{y}=f(\vec{z})$

其中$\vec{z}$是倒数第二层的输出向量，示意图如下：



假如激活函数是线性函数，相当于每一层都是线性函数

则 $\vec{y}=\vec{w}_n \cdot \vec{z}_n$，$\vec{w}$是最后一层权重，同理

$\vec{z}_n=\vec{w}_{n-1}\cdot\vec{z}_{n-1}, \cdots \Rightarrow \vec{y}=\vec{w}_n\cdot\vec{w}_{n-1}\cdots\vec{w}_1\cdot\vec{x}$

其中$\vec{w}_1, \cdots, \vec{w}_n$是每一层权重，因此$\vec{y}$和$\vec{x}$是线性关系

如果多层感知机节点的激活函数采用线性函数，网络无法实现非线性映射

4.2 (1)：证明：(i)：$\theta(s)=\frac{1}{1+e^{-s}}=\frac{e^s}{(1+e^{-s})e^s}=\frac{e^s}{e^s+1}$

(ii)：$\theta(-s)=\frac{1}{1+e^s}=1-\frac{e^s}{1+e^s}=1-\theta(s)$ (由(i)可知)

(iii)：$\theta'(s)=-\frac{(-e^{-s})}{(1+e^{-s})^2}=\frac{e^{-s}}{(1+e^{-s})^2}=(1-\frac{1}{1+e^{-s}})(\frac{1}{1+e^{-s}})$

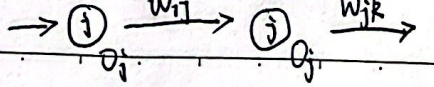
$=\theta(s)(1-\theta(s))$

(2)：$f(s)=\tanh(s)=\frac{e^s-e^{-s}}{e^s+e^{-s}}=\frac{e^s(e^s-e^{-s})}{e^s(e^s+e^{-s})}=\frac{e^{2s}-1}{e^{2s}+1}=(\frac{e^{2s}}{e^{2s}+1}-\frac{1}{e^{2s}+1})\frac{2}{2}=2\theta(2s)-1$

$f'(s)=2\theta'(2s)=4\frac{\partial\theta(s)}{\partial s}\Big|_{2s}=4\frac{e^{2s}}{(1+e^{2s})^2}=1-\tanh(s)^2$

(3)：使用tanh()后：使用以下神经网络模型

本题参考的ppt对应部分的sigmoid反向传播证明



No.

Date.

其中 对于节点 j, $O_j = f(net_j) = f(\sum_i w_{ij} O_i)$

对于输出 k, ~~$O_k = \Sigma$~~ $\hat{y}_j = O_j$, 设期望输出为 y_j

使用 MSE 评估有: $E = \frac{1}{2}\sum_j (y_j - \hat{y}_j)^2$

梯度为 $\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} = \delta_j O_i$, 其中 $\delta_j = \frac{\partial E}{\partial net_j}$

又 $\delta_j = \frac{\partial E}{\partial O_j} \frac{\partial O_j}{\partial net_j} = -(y_j - \hat{y}_j) f'(net_j)$, 对于输出节点,

$\delta_j = \sum_k \frac{\partial E}{\partial net_k} \cdot \frac{\partial net_k}{\partial O_j} \cdot \frac{\partial O_j}{\partial net_j} = \sum_k \delta_k w_{jk} f'(net_j)$, 对于隐层节点,

对于 $f(x) = \tanh(x)$ 其中 $f'(net_j) = 1 - \tanh(net_j)^2$, 代入得

$$\delta_j = \begin{cases} -(y_j - \hat{y}_j)(1 - \tanh(net_j)^2) & \text{输出节点} \\ \sum_k \delta_k w_{jk} * (1 - \tanh(net_j)^2) & \text{输隐层节点} \end{cases}$$

$\Rightarrow$ 更新公式: $w_{ij}(t+1) = w_{ij}(t) - \eta \delta_j(t) O_i(t)$

- 为何多层感知机不用 tanh: 因为 tanh 是饱和激活函数, x 不断增大或减小时, y 会逐渐趋于固定, 接近 -1 或 1, 导致梯度很小, 学习效果不好。

notebook