

人工神经网络笔记

杨哲涵

1. 感知机

1.1. 符号

u_j weighted sum of the input to neuron j

y_j output of neuron j

w_{kj} weight between neuron k (layer l) to neuron j (layer $l + 1$)

t_k or l_k golden label

$E(n)$ error or loss on one sample n

1.2. 激活函数

[activation functions](#)

1.3. 梯度下降

η 称为学习率(learning rate)

$$x' = x - \eta f'(x) \quad (1)$$

1.4. 错误传播

定义 1.4.1 (Mean Square Error):

$$E_k = \frac{1}{2}(t_k - y_k)^2$$
$$E = \sum_k E_k \quad (2)$$

定义 1.4.2 (Cross Entropy):

$$E_k = -t_k \log(y_k)$$
$$E = \sum_k E_k \quad (3)$$

1.5. 反向传播

Steps of BP is:

1. compute the local gradients
2. compute the upstream gradients (Gradient Output) from next operator unit
3. apply chain rule to compute downstream gradients (Gradient Input) of local inputs
4. repeat the same procedure on the back operator units

定义 1.5.1 (Gradient Output): Gradient output of neuron j in layer l is

$$\delta_j^l = \frac{\partial E}{\partial u_j^l} \quad (4)$$

定义 1.5.2 (Gradient Input): Gradient input of neuron j in layer l is

$$\frac{\partial E}{\partial w_{kj}^l} = \frac{\partial E}{\partial u_j^l} \frac{\partial u_j^l}{\partial w_{kj}^l} = \delta_j^l y_k^{l-1} \quad (5)$$