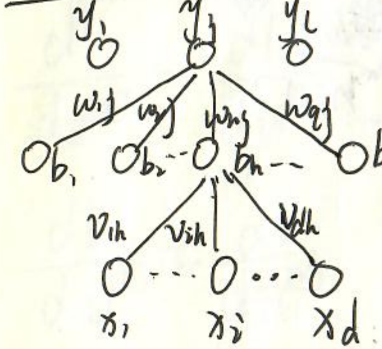
2019/10/02 单个样本 BP (BackPropagation)



输出层. 第j个输出神经元的输入 $\beta_j = \sum_{h=1}^{q} w_{hj} b_h$

隐层 第h个隐层神经元的输入 $\alpha_h = \sum_{i=1}^{d} v_{ih} x_i$

输入层 (x_k, y_k)

输出: $\hat{y}_k = (\hat{y}_1^k, \hat{y}_2^k, \ldots \hat{y}_l^k)$

$\hat{y}_j^k = f(\beta_j - \theta_j)$ (sigmoid) $\beta_j = \sum_{h=1}^{q} w_{hj} b_h$

$\alpha_h = \sum_{i=1}^{d} v_{ih} x_i$

网络在 (x_k, y_k) 的均方误差:

$E_k = \frac{1}{2}\sum_{j=1}^{l} (\hat{y}_j^k - y_j^k)^2$ [认定网络可以表示映射关系的前提下, 计算方差来优化参数]

需确定的参数有:

$d \times q + q \times l + q + l$
输入 输出 隐层阈值 输出层阈值.

使梯度下降: $v \leftarrow v + \Delta v$

更新参数: $\Delta w_{hj} = -\eta \frac{\partial E_k}{\partial w_{hj}}$ 误差函数对权重求导 使偏导为0

(单个训练, 样本更新)

$$\frac{\partial E_k}{\partial w_{hj}} = \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial w_{hj}}$$

$$\frac{\partial \beta_j}{\partial w_{hj}} = b_h$$

$\frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j}$ 由于 $\hat{y}_j^k = f(\beta_j - \theta_j)$ 是sigmoid函数

所以 $\frac{\partial \hat{y}_j^k}{\partial \beta_j} = (\hat{y}_j^k)' = f(x) \cdot (1 - f(x))$

$$g_j = -\frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} = -(\hat{y}_j^k - y_j^k) \cdot \hat{y}_j^k \cdot (1 - \hat{y}_j^k)$$

$$= \hat{y}_j^k \cdot (1 - \hat{y}_j^k) \cdot (y_j^k - \hat{y}_j^k)$$

(权重) $\Delta w_{hj} = \eta g_j b_h$ (η 学习率)

$\Delta \theta_j = -\eta g_j$ (阈值) $\Delta \gamma_h = -\eta e_h$ (阈值)

(权重) $\Delta v_{ih} = -\frac{\partial E_k}{\partial b_h} \cdot \frac{\partial b_h}{\partial \alpha_h} \cdot \frac{\partial \alpha_h}{\partial v_{ih}} \cdot \eta$

$$= -\sum_{j=1}^{l} \frac{\partial E_k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot f'(\alpha_h - \gamma_h) \cdot \eta$$

$$= -\sum_{j=1}^{l} -g_j w_{hj} \cdot b_h (1 - b_h) = b_h (1 - b_h) \sum_{j=1}^{l} g_j w_{hj}$$