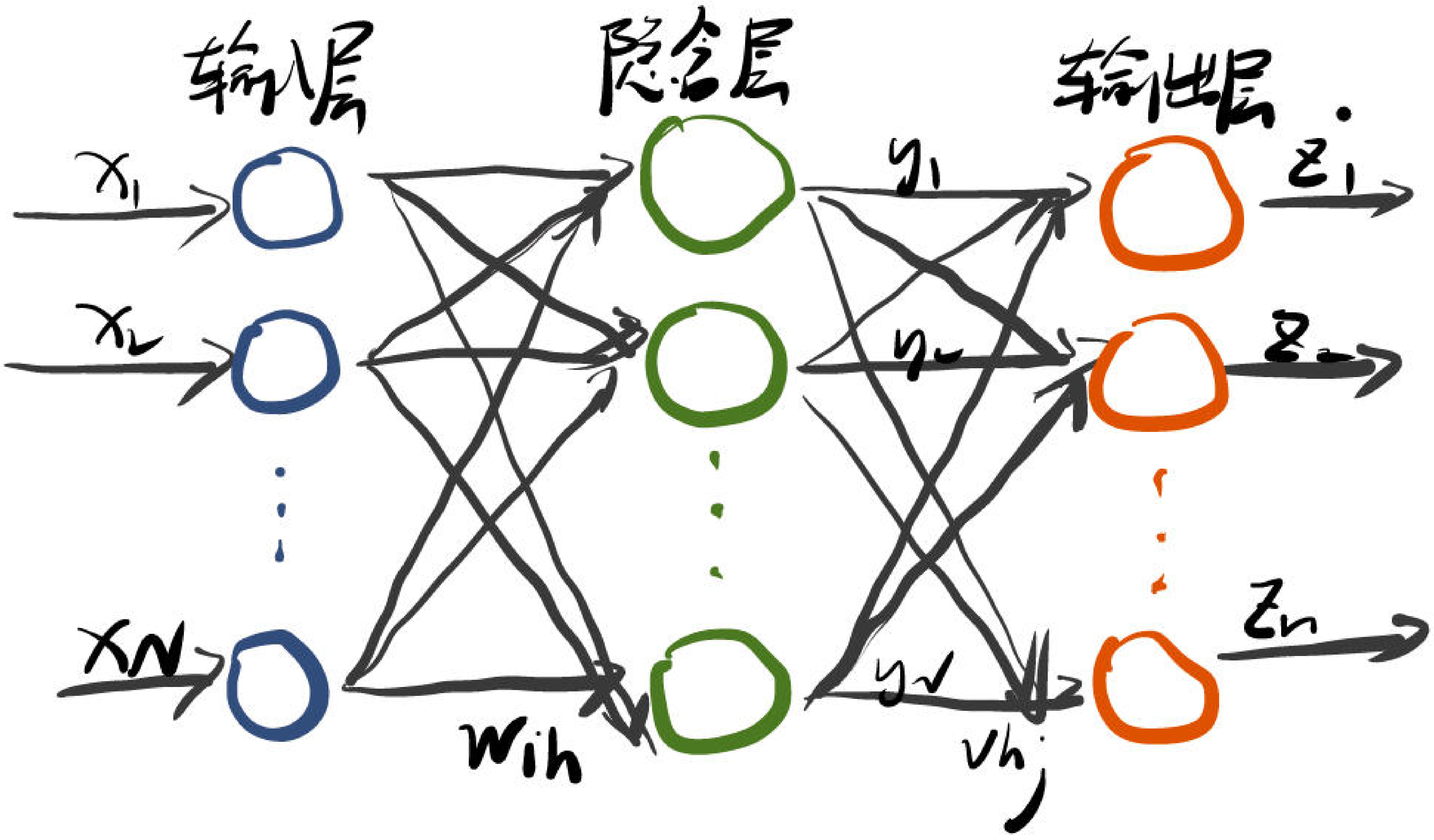
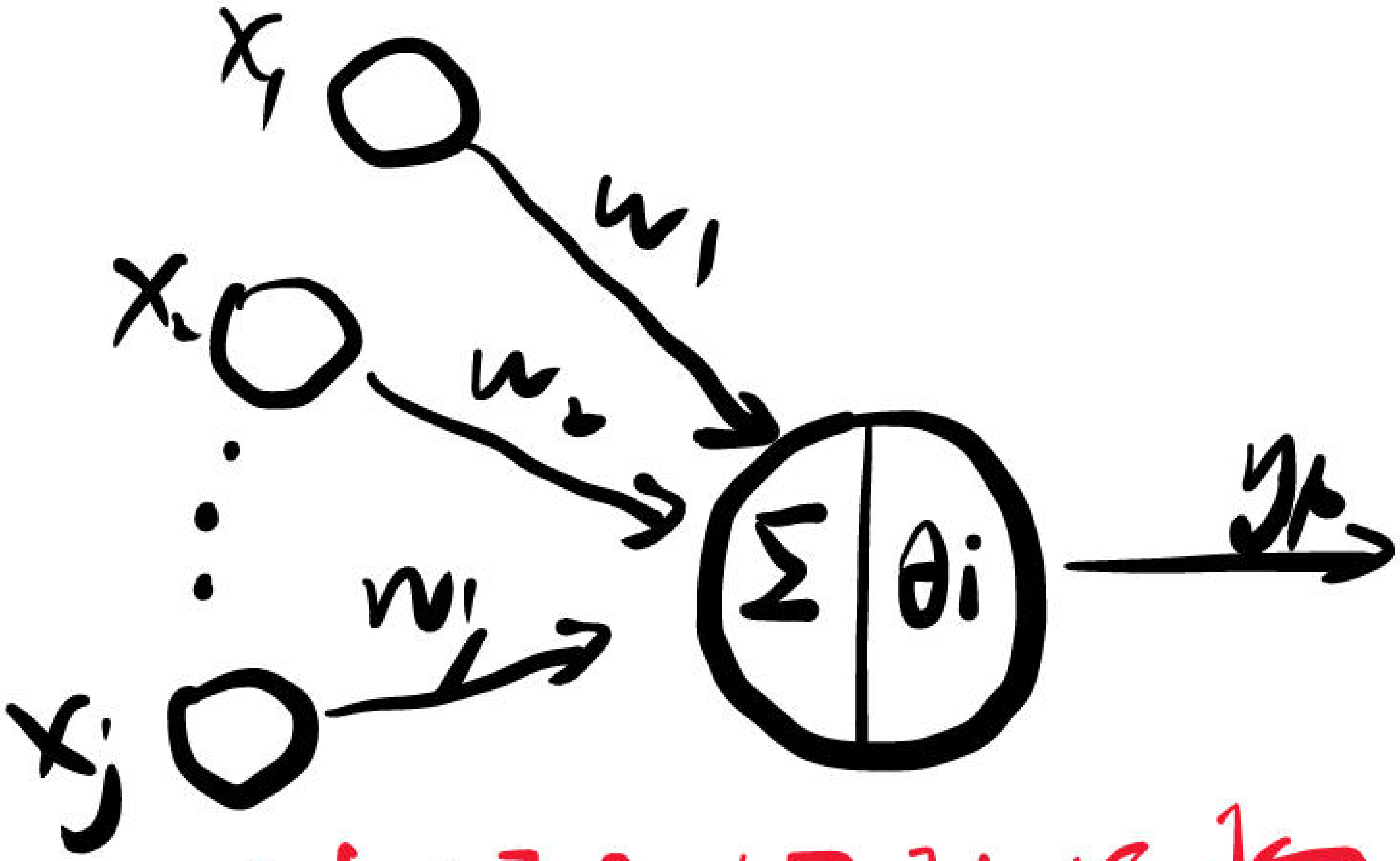
BP神经网络

"万能的模型＋误差修正函数"

每次根据训练得到的结果与预想的结果进行误差分析，进而修改$\boxed{\text{权值}}$和$\boxed{\text{阈值}}$，一步一步得到能输出和预想一致的模型，此即为BP神经网络的核心。



神经元是以生物研究及大脑的响应机制而建立的拓扑结构网络，模拟神经冲突的过程.



神经元拓扑结构

对于第i个神经元，x_1，x_2…x_j为神经元的输入，输入常为对系统模型关键影响自变量，w_1，w_2，…w_j为连接权值调节各个输入量的占重比。由此得到神经元净输出：

$$Net_{in} = \sum_{i=1}^{n} w_i * x_i$$

θ_i表示该神经元的阈值，根据生物学中的知识，只有当神经元接收到的信息达到阈值才会被激活。

※激活函数：Sigmond函数->一般

当 $Net_{in}-\theta_j>0$ 时，神经元被激活：

$$y_i=f(Net_{in}-\theta_j)$$

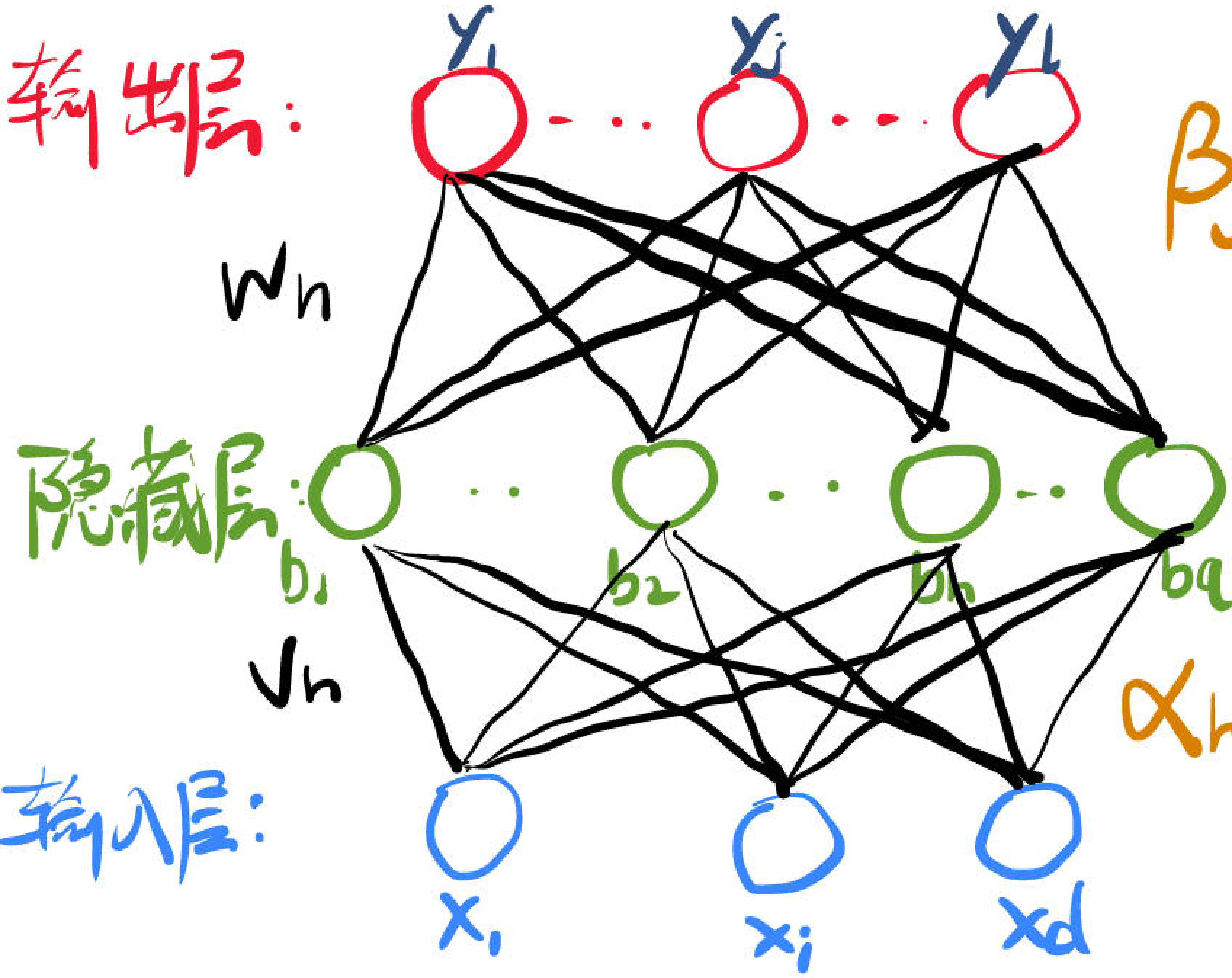
已知在BP神经网络模型中，我们有三层结构：输入层、隐藏层、输出层。

设：V_{ih}为输入层到隐藏层的权值

γ_h为第h个神经元的阈值 隐藏层

W_{hj}为隐藏层到输出层的权值

θ_j为第j个神经元的阈值 输出层.



$$\beta_j = \sum_{h=1}^{q} w_{hj} * b_h$$

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

激活函数为Sigmond函数，则隐藏层的输出为：

$$b_h = f(Netin - \gamma_h) = f(\alpha_h - \gamma_h)$$

最终的输出y为：

$$y_i = f(Netin - \theta_j) = f(\beta_i - \theta_i)$$

在某个训练示例(x_k, y_k)中，假设神经网络训练输出为

$$y_k' = (y_1^{k'}, y_2^{k'}, \ldots, y_l^{k'})$$

那么这次预测结果的误差我们可以用最小二乘法表示：

$$E_k=\frac{1}{2}\sum_{j=1}^{l}(\hat{y}_j^k-y_j^k)^2$$

而我们现在要做的就是调整参数一步一步缩小E_k.

参数：① v_{ih} → 共有 $d\cdot q$ 个

② w_{hj} → 共有 $q\cdot l$ 个

③ 输出与隐藏神经元个数：$q+l$

故需要调整的参数共有：

$$(d+l+1)*q+l$$

梯度下降法调参.

假设有学习速率η，η太快可能导致越过最优解，η太小则会因为太慢而降低算法速率.

$$Param += -\eta\frac{\partial E_k}{\partial Param}$$

首先我们看看隐藏层到输出层的权重调整值：

$$\Delta w_{hj} = -\eta\frac{\partial E_k}{\partial w_{hj}}$$

由此即可根据梯度下降调整参数.

现列出每一个公式：

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

$$\begin{vmatrix} x_1 & x_2 & x_3 & \cdots & x_d \end{vmatrix} \cdot \begin{vmatrix} v_{11} & v_{12} & v_{13} & \cdots & v_{1q} \\ v_{21} & v_{22} & v_{23} & \cdots & v_{2q} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_{d1} & v_{d2} & v_{d3} & \cdots & v_{dq} \end{vmatrix}$$

$$b_h = f(\alpha_h - \gamma_h)$$

$$\beta_j = \sum_{h=1}^{q} w_{hj} * b_h$$

$$\begin{vmatrix} b_1 & b_2 & b_3 & \cdots & b_h \end{vmatrix} \cdot \begin{vmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1l} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2l} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{q1} & w_{q2} & w_{q3} & \cdots & w_{ql} \end{vmatrix}$$

$$y_j^{k'} = f(\beta_j - \theta_j)$$

$$E_k = \frac{1}{2} \sum_{j=1}^{l} (y_j^{k'} - y_j^{k})^2$$

综上我们可以得知 w_{hj} 先影响 β_j，再影响 $y_j^{k'}$，最后影响 E_k（一个 w 权值只会影响一个 β_j）所以我们可得：

$$\Delta w_{hj} = -\eta \frac{\partial E_k}{\partial w_{hj}} = -\eta \frac{\partial E_k}{\partial y_j^{k'}} \cdot \frac{\partial y_j^{k'}}{\partial \beta_j} \cdot \boxed{\frac{\partial \beta_j}{\partial w_{hj}}}$$

∵ ① $\beta_j = \sum_{h=1}^{q} w_{hj} * b_h$ $\quad \boxed{\frac{\partial \beta_j}{\partial w_{hj}}} = b_h$

② $E_k = \frac{1}{2} \sum_{j=1}^{l} (y_j^{k'} - y_j^{k})^2$ $\quad \boxed{\frac{\partial E_k}{\partial y_j^{k'}}} = \frac{1}{2} \cdot 2 (y_j^{k'} - y_j^{k}) = y_j^{k'} - y_j^{k}$

③ $y_j^{k'} = f(\beta_j - \theta_j)$ $\quad \boxed{\frac{\partial y_j^{k'}}{\partial \beta_j}} = y_j^{k'}(1 - y_j^{k'})$

$(f'(x) = f(x)(1 - f(x)))$

$$\therefore \Delta w_{hj} = \eta\, y_j^{k'} (1 - y_j^{k'}) (y_j^{k} - y_j^{k'}) \cdot b_h$$