

极限学习机预测原理

极限学习机（Extreme Learning Machine, ELM）的结构网络如图 1 所示，假设有 N 个不同的样本 (X_i, t_i) ，其中 $X_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ ， $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$ ，对于隐含层的节点个数为 L 的单隐层前馈神经网络，其网络表达式可表示为：

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) = o_j, j = 1, 2, \dots, N \quad (1)$$

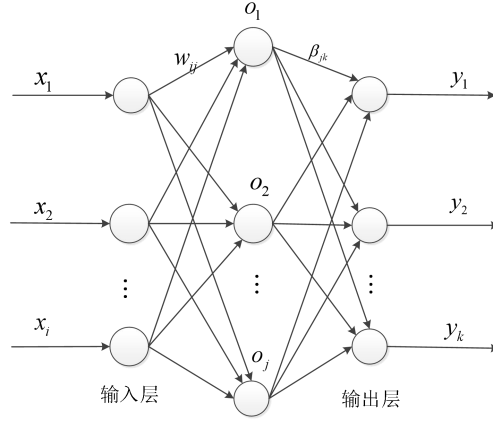


图 1 ELM 结构网络图

其中， $g(x)$ 为激励函数； $W_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ 为输入层神经元与隐含层神经元的连接权值，即输入权重； b_i 为隐含层节点偏移值； β_i 为隐含层神经元与输出层神经元的连接权值，即输出权重； o_j 为预测输出。

ELM 样本训练以网络输出值与样本值间的误差最小为目标，即

$$\sum_{i=1}^N \|o_j - t_j\| = 0 \quad (2)$$

存在系数矩阵 β_i ， W_i ， b_i 使得

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) = t_j, j = 1, 2, \dots, N \quad (3)$$

即

$$H\beta = T \quad (4)$$

其中： $H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_L \cdot x_1 + b_L) \\ \vdots & & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_L \cdot x_N + b_L) \end{bmatrix}$ ，为隐含层的输出矩阵， β 为输出权重矩阵， T 为 ELM 的期望输出。

对于已产生输入权重值 w_i 和隐含层节点偏移值 b_i 的 ELM 而言，训练一个

ELM，就是求解表达式 $H\beta=T$ 中的 $\hat{\beta}$ 的值。在 H 矩阵中， L 是隐含层中神经元数量， N 是训练样本数。当 $L=N$ ，则隐含层的输出矩阵 H 为方阵，且为可逆矩阵；当 $L<N$ ，则隐含层的输出矩阵 H 不是方阵，这时若要得到 $\hat{\beta}$ ，就要求 H 的伪逆矩阵，综上，输出矩阵 $\hat{\beta}$ 的求解方法如下：

$$\hat{\beta} = H^+T = (H^TH)^{-1}H^TT \quad (5)$$

根据其预测原理，ELM 算法的预测流程图如图 2。

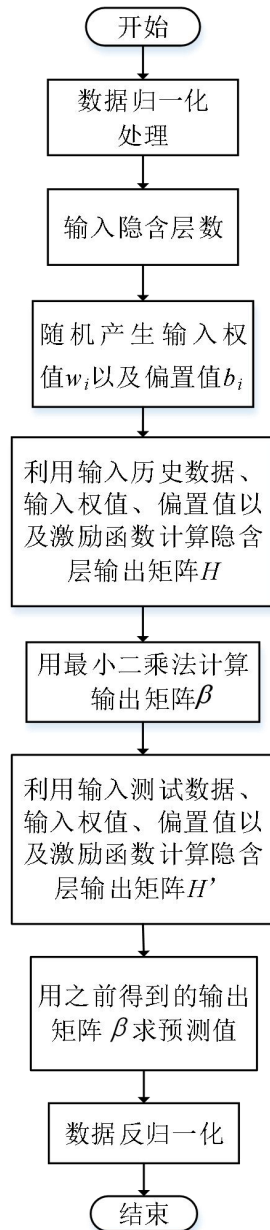


图 2 ELM 的预测流程图