输入输出因子的选取

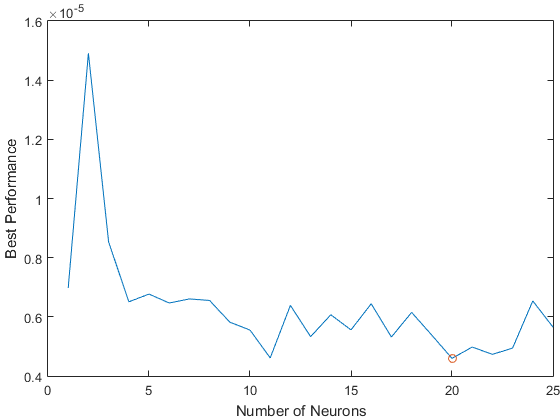
由于研究对象的特殊性，我们只能通过各类网站及报告获取到零碎的数据，经过数据规整化后，我们获得了~~仅包含N,P输入总量以及叶绿素a浓度的~~129组数据，由于水体叶绿素a浓度是表征水体中藻类现存量的最直接指标，故水体中总叶绿素a浓度是模型的输出因子，通过对叶绿素a浓度的预测可以间接实现对藻类引发的水华进行预测．在确定模型输出因子后，合理选择网络的输入因子对正确应用BP模型和保证模型预报精度非常重要。因为输入因子中，可能存在与输出因子关联弱的噪声因子，或者重叠反映系统信息的冗余因子．无论是噪音因子还是冗余因子，都会加大分析问题的难度，增加模型的复杂性，最终影响模型预测能力．筛选BP神经网络输入因子的基本原则是选择与输出因子相关而又彼此无关的环境因子网络结构。

最终，我们以N,P输入总量作为输入因子，去除了数据量过少的因子以及其他可能的噪声因子及冗余因子。

网络结构

早有理论证明：3层BP网络，当各层神经元均采用S型函数时，可满足任意复杂的非线性函数拟合逼近问题【7】．这个存在性结论对神经网络结构的设计具有重要的指导作用．参考该结论，确定水华预测模型为3层网络结构，即1个输入层、1个隐含层和1个输出层．然而，选取隐含层神经元数、层与层间的激活函数，虽然有规则可依，但所得结果差异较大，甚至完全不同【8】．因此，隐含层神经元数、层与层间的激活函数仍需根据具体的研究问题而定．

隐含层神经元数



通过对不同隐含层神经元数的模型进行训练，使用Mean squared normalized error performance function来分析隐含层神经元数对训练效果的影响，从分析结果发现：当隐含层神经元数大于3个时，训练效果开始随隐含层神经元数的增长在小范围波动，总体上效果很好，当隐含层神经元数为20时，得到最好结果,加上我们的数据集不是很大，因此我们以20作为隐含层神经元数.

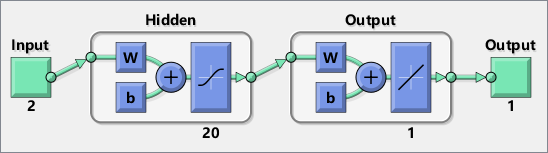
激活函数

我们使用MATLAB神经网络工具箱中的函数拟合网络，以a sigmoid transfer function in the hidden layer and a linear transfer function in the output layer

依照上述的网络结构，得到网络结构拓扑以及模型表达式

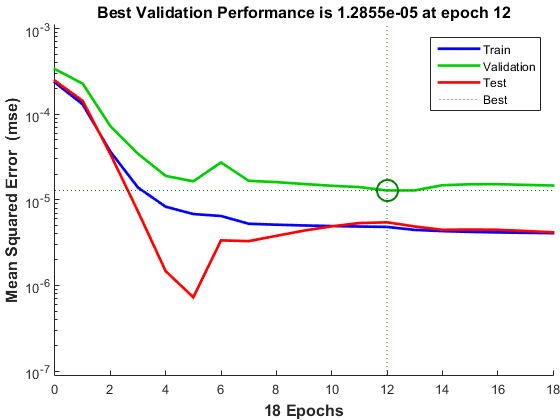
y=purelin(V·tansig(W·x+b1)+b2)．

式中：x为输入变量，即N浓度与P浓度；Y为输出变量，即当前N，P浓度下的叶绿素a浓度；w和b。为输入层与隐含层问连接权重和阈值；y和b。隐含层与输出层间的连接权重和阈值．

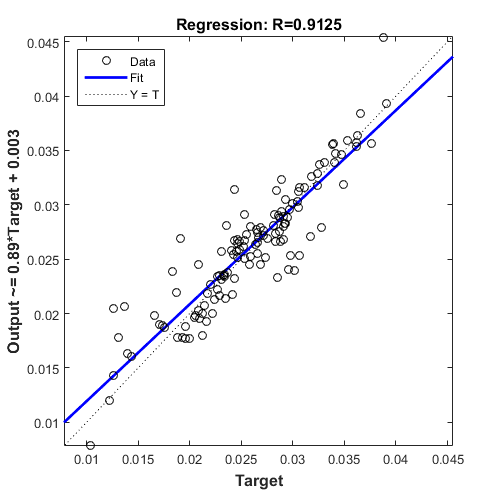


结果分析

我们对数据进行分配，其中91组作为训练数据，19组作为验证数据，19组作为测试数据，训练18代后结果收敛，得到的训练效果曲线如下



相关系数曲线如下



最终训练出来的模型与实际数据相关性达到了91%，说明神经网络拟合效果很好，符合我们对N,P含量对叶绿素a含量影响的预期表现.

也就是说，N,P含量之间影响到了水中藻类的生长,经过统计与预测，当N含量大于2.5mg/L,P含量大于0.3mg/L时便可能使叶绿素a超标,产生水华，并且在实际水藻的生长环境允许范围内，叶绿素a含量随着N，P含量的增长而增长,也就意味着水中藻类数量在增加。