第4次作业\_分类问题

电 25 吴晨聪 2022010311

# （1） Try to find the best classification results (by SVM, DT, and so on).

基本思路是建立分类树模型。通过改变训练集和测试集的占比、最大深度、超参数等方式来寻找最优的模型。同时，为了减少模型的训练参数，本次作业中建立了一个新的训练模型-Transformed，是对原来模型-Raw的输入参数进行了一定的处理，将线路的有功功率和无功功率结合为线路的视在功率，同时忽略了线路的功率损耗。

1. 改变训练集和测试集的占比

测试集中样本数量占总样本数量的比重为test\_size，从小到大改变test\_size得到关于训练集误差和测试集误差的两条曲线，如图所示，横轴为1-test\_size即训练集的大小Training set size）。可见当trainingsize大于0.5时两条曲线趋于稳定，故取testsize=0.4（即trainingsize=0.6）是较为合理的值。



图1 分类树模型的learning curve

1. 改变最大训练深度

如图2所示，先取testsize=0.4，再改变不同的最大训练深度得到其关于训练误差以及测试误差的一族曲线。

可以发现在分类树深度小于4的时候，训练误差及测试误差都会随着深度的提高而降低；而当深度大于5的时候，随着深度的提高，测试集的误差不再明显减小而训练集的误差仍在减小，即出现了过拟合的现象。因而最大深度应该在5左右较为合适。

此外，输入参数进行了一定处理的训练模型Transformed在训练参数上比Raw更小但是在预测的精度上也没有落后于Raw。

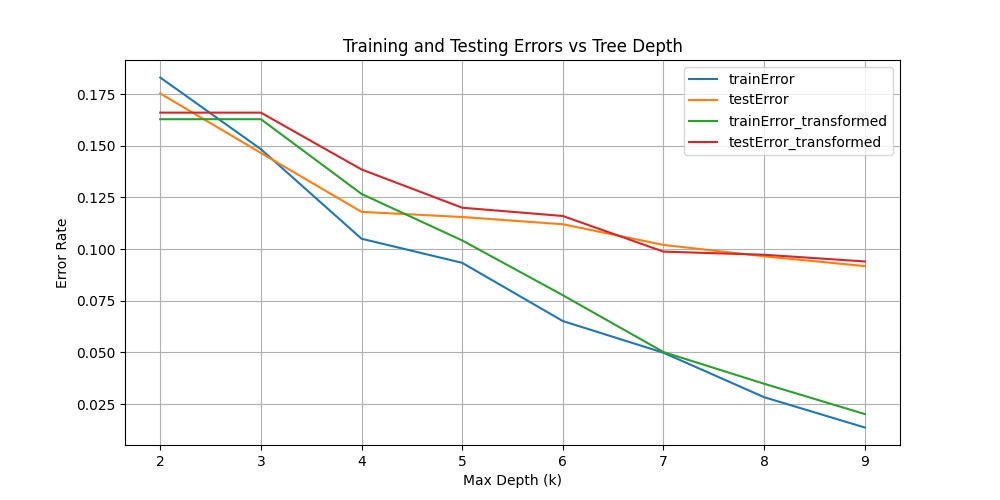


图2 两个分类树模型训练误差及测试误差关于分类树深度变化曲线

# （2） The precise awareness of insecurity power system state is usually critical for power system operation, which results in different tolerances of false-negative and false-positive samples. Try to consider this situation inyourmodel.

当考虑到实际应用背景时可知这里的False Positive和False Negative的权重是不一样的，一个系统并不安全但被误判为安全（False Positive）的危害要大于本来安全但被误判为不安全（False Negative），因而利用其预测结果的混淆矩阵来观测两个模型的precision的大小，进而判断何种情况下有更少的False Positive的概率。

图3所示是这两种模型的预测精度关于分类树深度的变化曲线，可见其预测的精度基本随着分类树的最大深度的提高而提高，因此为确保预测精度还是要在合适的范围内选择较大的深度。

一張含有 行, 圖表, 繪圖, 斜率、斜坡 的圖片

自動產生的描述

图3 两个分类树模型预测结果precision关于分类树深度变化曲线