# 1.3

因为根据NFL定理，各种学习算法的期望性能相同。所以学习算法自身的归纳偏好与问题是否匹配往往起到决定性的作用。而在题目中，假设数据包含噪声，那么我们可以将数据中的噪声去除。先确认属性，忽略除确定属性以外的其他属性，从而达到去除数据当中的噪声的目的。

# 2.4

TPR：真正类率，又叫真阳率，代表预测是异常实际也是异常的样本数，占实际总异常数的比例——值越大 性能越好

FPR：假正类率，又叫假阳率，代表预测是异常但实际是正常的样本数，占实际正常总数的比例——值越小 性能越好

P：精确率，代表预测是异常实际也是异常的样本数，占预测是异常的总数的比例——值越大 性能越好

R：召回率，意义同TPR——值越大 性能越好

TPR = TP/(TP+FN)

FPR = FP/(FP+TN)

R = TPR

P = TP/(TP+FP)

# 2.6

错误率=错误分类个数/总体=FN+FP/TP+TN+FN+FP

ROC实质是把所有的分类阈值从0-1都在曲线上展示了一下，然后就可以看出整个模型的好坏。