

模型评估与选择

2019年6月20日 21:25

基础概念：

错误率，精度：错误/正确样本数所占比例

误差：训练误差（实际训练），泛化误差（针对学习器）

过拟合（将训练样本本身的特点视为一般性质），欠拟合（学习器导致）
关键问题

混淆矩阵：真实情况与预测结果为两段组成的表格

机器学习的目的是：选择泛化误差最小的模型（线性，非线性，概率.....）

一、模型评估方法（设置测试集，NFL定理表明没有一定最好的）

1、留出法：进行分层采样；7：3；多次划分取平均。



2、交叉验证法：将数据分成若干子集，每次选择一个子集作为测试集，轮流。



3、留一法：数据含 m 个样本，分为 m 个子集。

D'

能力差，一般训练样本过

换取平均。

3、留一法：数据含m个样本，分为m个子集。

D'

4、自助法：对于数据集D，每次放回进行采样得一个事件放入 D' 中，重复
$$\lim_{m \rightarrow \infty} \left(1 - \frac{1}{m}\right)^m = \frac{1}{e} \approx 0.368$$

可得，数据中存在0.368的数据未在采样数据集中

5、调参：算法的参数(超参数)；模型的参数（如函数的系数）

二、性能度量：评估模型的泛化能力/一般使用误差进行度量

1、错误率与精度：错误/正确样本数占样本的个数；

2、查准率、查全率、F1：针对“检索出来的信息有多少是用户感兴趣这一”
查准率，又称召回率：P=抽取正确在正例的比例，表正确率

查全率：R=抽取的正确的占总的比例，表竟可能多的取出好瓜

PR曲线：若某曲线包住一曲线，则该曲线好于内部曲线。，也可更具平

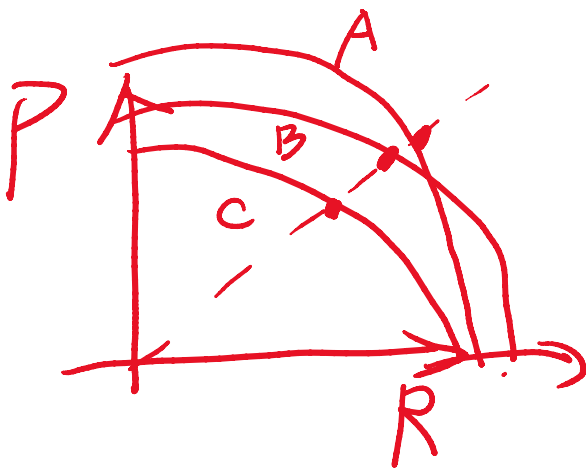
F1：查准率与查全率的调和平均数。（可用来衡量查全率与查准率的相

$$\frac{1}{F_1} = \frac{1}{2} \left(\frac{1}{P} + \frac{1}{R} \right)$$
$$\frac{1}{F_\beta} = \frac{1}{1 + \beta^2} \left(\frac{1}{P} + \frac{\beta^2}{R} \right)$$

3、ROC 与 AUC

重复m次，得到测试集

标准”



平衡点 (BEP) 选择

(对重要性)

ROC:

AUC: ROC的面积, , 越大越好。

