

开始之前的准备工作

磨刀不误砍柴工，在开始各种算法之前，我们还是要先准备点东西（俗称：qian - xi）

一.经验误差与过拟合

1.错误率：

$$\text{错误率} = \frac{\text{错误样本}}{\text{样本数}}$$

2.精度：

$$\text{精度} = 1 - \text{错误率}$$

3.训练误差(training error)(或者经验误差：empirical error)：

- 预测输出与真实输出的差异

4.泛化误差：

- 在新样本上的误差

5.过拟合（overfitting）：

- 学习器把训练样本自身的特点当成潜在样本的一般性质，导致泛化能力下降（训练的太好了）

6.欠拟合（underfitting）：

- 学习器训练样本一般性质未学到（学的不好）

二.评估方法

1.模型的好坏影响因素:

- 1). 算法;
- 2). 数据;
- 3). 任务需求;

2.评估方法：

- 1). 留出法:
直接将数据集 D 划分为两个互斥的集合,其中一个集合作为训练集 S ，另一个作为测试集 T
- 2). 交叉验证:

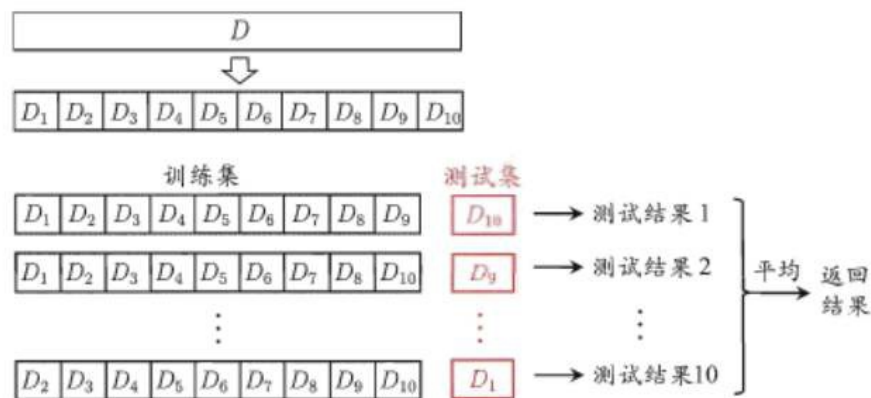


图 2.2 10 折交叉验证示意图

知乎 @爬山的小码农

- 3). 自助法(重要):
从给定 m 个样本的数据集 D 中,从 D 中每次随机采样一个样本拷贝到 D' 中再将该样本放回 D (使得下次仍有概率采样到),重复 m 次 (D 中有部分样本在 D' 多次出现,也有部分从未出现,概率为: $(1 - \frac{1}{m})^m$ 取极限: $\lim_{m \rightarrow \infty} (1 - \frac{1}{m})^m \rightarrow \frac{1}{e} \approx 0.368$ 即 D 中有 36.8% 未被采样过,于是, D' 作为训练集, D 与 D' 的差集作为训练集

三.性能度量（将在下一篇详细讲解）：

1. 错误率与精度
2. 查全率、查准率与F1
3. ROC与AUC
4. 代价敏感错误率与代价曲线

四.比较检验（后续详解）：

1. 假设检验
2. 交叉验证t检验
3. McNemar检验
4. Friedman检验与Nemenyi后续检验

五.偏差与方差（后续详解）：

人们不仅要通过算法实验估计泛化性能，而且要了解为什么有这样的性能，偏差和方差是解释算法泛化性能的重要工具