

模型评估与选择

评估方法

留出法：多次留出数据作为验证集，然后取误差的平均值。

交叉验证法：先将数据集D划分为k个大小相似的互斥子集，即 $D = D_1 \cup D_2 \cup \dots \cup D_k$ ，每个子集都互不相交，即从D中通过分层采用的方式得到。然后，每次用K-1个子集的并集作为训练集，余下的那个子集作为测试集。

为了减小因样本划分不同而引入的差别，K折交叉验证通常随机使用不同的划分重复p次，最终的评估结果是这p次k者交叉验证的均值。

缺点：计算复杂度高。

自助法：用于解决训练数据规模不同而引入的方法。其基本的操作如下：

给定包含m个样本的数据集D，每次随机地从D中取出一个样本，将其拷贝放入D1集合，然后将采样放回数据集，使得样本可以在下次采用中仍有可能被采样，采样m次后。D1作为训练集，D作为测试集。这种方法实际评估的模型与期望评估的模型都使用m个训练样本，而我们人有数据总量越1/3的，没在训练集的数据用于测试。

自助法在数据集较小，难以有效划分训练 / 测试集时有用。但自助法产生的数据集改变了初始数据集的分布，这会引入估计偏差。

性能度量的标准

错误率：错误结果/样本总数

精度：正确结果/样本总数

查准率：真正例占整个预测正例的比例。举例：检索的信息有多少比例是用户感兴趣的

查全率：真正例占整个样本数的比例。举例：用户感兴趣的信息有多少被检索出来。

注：查准率与查全率是一对矛盾的度量，一般来说，查准率高时，查全率往往偏低；查准率低时，查全率往往偏高。

F1：是基于查准率与查全率的调和平均。其定义如下：

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

P和R分别为查准率和查全率。