
SPSS:交叉验证（简单交叉验证、k折交叉验证、留一法）

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/statistics/2477.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

SPSS:交叉验证（简单交叉验证、k折交叉验证、留一法）。针对经验风险最小化算法的过拟合的问题，给出交叉验证的方法，这个方法在做分类问题时很常用：

一：简单的交叉验证的步骤如下：

- 1、从全部的训练数据 S 中随机选择 s 的样例作为训练集 train ，剩余的作为测试集作为测试集 test 。
- 2、通过对测试集训练，得到假设函数或者模型。
- 3、在测试集对每一个样本根据假设函数或者模型，得到训练集的分类，求出分类正确率。
- 4、选择具有最大分类率的模型或者假设。

这种方法称为 hold-out cross validation 或者称为简单交叉验证。由于测试集和训练集是分开的，就避免了过拟合的现象

二：k折交叉验证 k-fold cross validation

- 1、将全部训练集 S 分成 k 个不相交的子集，假设 S 中的训练样例个数为 m ，那么每一个子集有 m/k 个训练样例，相应的子集称作 $\{s_1, s_2, \dots, s_k\}$ 。
- 2、每次从分好的子集中里面，拿出一个作为测试集，其它 $k-1$ 个作为训练集
- 3、根据训练训练出模型或者假设函数。
- 4、把这个模型放到测试集上，得到分类率。
- 5、计算 k 次求得的分类率的平均值，作为该模型或者假设函数的真实分类率。

这个方法充分利用了所有样本。但计算比较繁琐，需要训练k次，测试k次。

三：留一法 leave-one-out cross validation

留一法就是每次只留下一个样本做测试集，其它样本做训练集，如果有k个样本，则需要训练k次，测试k次。

留一发计算最繁琐，但样本利用率最高。适合于小样本的情况。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发