**网格搜索法与交叉验证**

1. **网格搜索**

网格搜索算法是一种通过遍历给定的参数组合来优化模型表现的方法。

在使用使用学习器对数据进行预测时，通常需要对模型的参数进行调整，以使得学习器能够更好地拟合和预测数据，网格搜索算法能够遍历所有给定的参数组合，有利于参数调优。

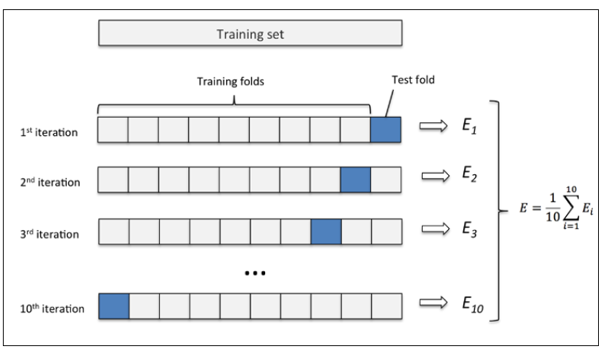
但对于这些参数组合，如何评价孰优孰劣，交叉验证能够在这方面发挥作用。

**2.交叉验证**

交叉验证的基本思想是将原始数据进行分组,一部分做为训练集,另一部分做为验证集,首先用训练集对分类器进行训练,再利用验证集来测试训练得到的模型,以此来做为评价分类器的性能指标。

交叉验证的过程：

常见的交叉验证为10折交叉验证，起示意图如下：



* 第一步，不重复抽样将原始数据随机分为 k 份。
* 第二步，每一次挑选其中 1 份作为测试集，剩余 k-1 份作为训练集用于模型训练。
* 第三步，重复第二步 k 次，这样每个子集都有一次机会作为测试集，其余机会作为训练集。
* 在每个训练集上训练后得到一个模型，
* 用这个模型在相应的测试集上测试，计算并保存模型的评估指标，
* 第四步，计算 k 组测试结果的平均值作为模型精度的估计，并作为当前 k 折交叉验证下模型的性能指标。

交叉验证的优点：

* 交叉验证用于评估模型的预测性能，尤其是训练好的模型在新数据上的表现，可以在一定程度上减小过拟合。
* 还可以从有限的数据中获取尽可能多的有效信息。