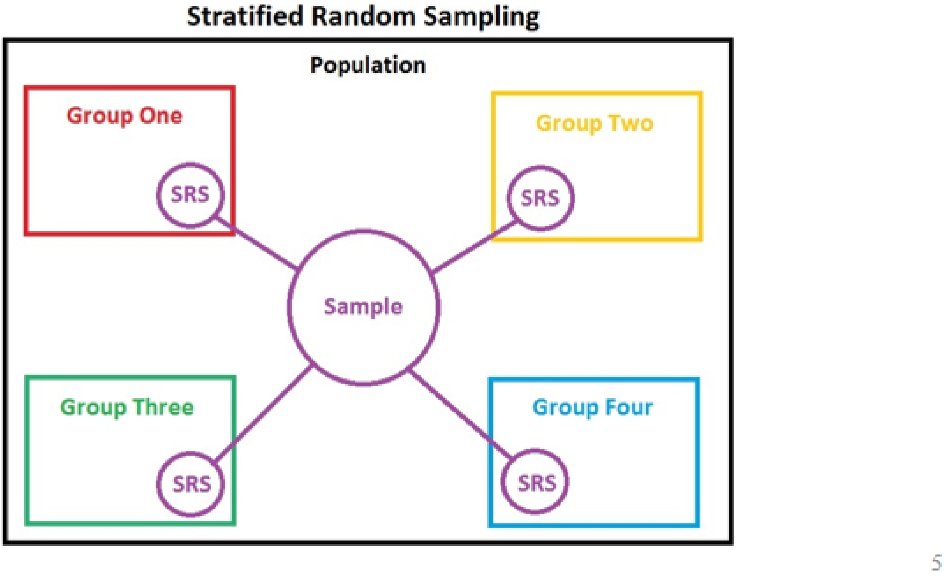
留出法-分层采样

- To keep the same data distribution of the training and testing set and avoid the extra error， we can adopt the stratified random sampling (分层采样).
- Keep the same class distribution.



局限性：

　　模型评估具有很大的随机性，效果取决于数据被选入训练集还是测试集

　　可用来训练的标记数据较少----要预留测试集

重复随机子采样

留出法重复多次来提高性能

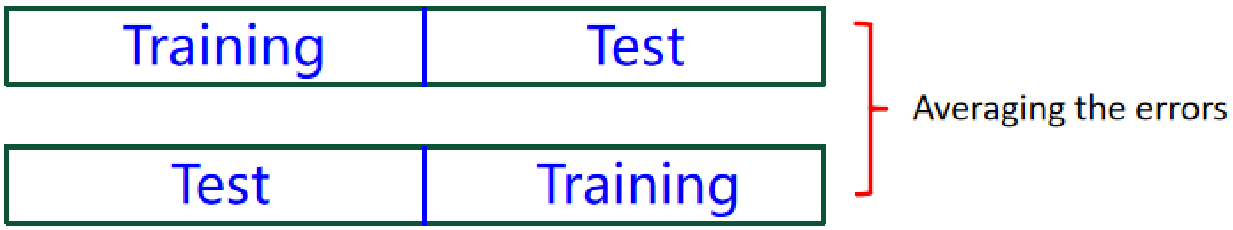
缺点：

没有使用尽量多的数据进行训练

不知道每个数据用于训练或者测试的次数

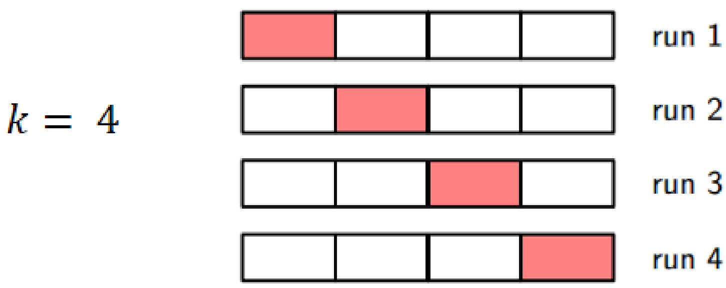
交叉验证

总误差是两次训练模型误差的均值



每份数据用作测试集一次

- **More general:** the k-fold cross-validation k重交叉验证
- Segments the data into k equal-sized partitions.
- During each run, one partition is chosen for testing, while the rest are used for training.
- This procedure is repeated k times so that each partition is used for testing exactly once.每份数据用作测试集一次
- The total error is found by averaging the errors for all k runs.



Similar to holdout method, the k-fold cross-validation can be repeated several times, e.g., p.

10

留一法

测试集只包含一个数据

Booststrap自助法

假设对训练数据进行采样而不是替换

训练集和测试集中没有重复数据

- The probability a record is chosen by a bootstrap sample is

$$1-\left(1-\frac{1}{N}\right)^N \rightarrow 1-e^{-1}=0.632$$

Records that are not included in the bootstrap sample become part of the test set. 没有被自助法选中当做训练集数据的当做测试集

- Bootstrap can generate many different training sets from the same original dataset, which can benefit the ensemble learning. 可以从同一个原始数据集生成许多不同的训练集，有利于集成学习

- As the bootstrap would change the data distribution of the resulted training set (different from that of the original dataset), and involve extra error, if we have enough data, holdout and cross-validation would be more widely used. 自助法会改变结果训练集的数据分布，造成额外误差，如果有足够多的数据集尽量不采用该方法

分类的评估指标

模型选择

超参数：超参数的值不受学习算法本身的调整。

正则化超参数控制模型的容量。适当控制模型容量可以防止过度拟合。

另一种类型的超参数来自训练过程本身。

例如，随机梯度下降（SGD）优化需要学习率和批量大小。

一些优化方法需要收敛阈值。

这些也需要设置为合理的值，以便在训练过程中找到一个好的模型。

超参数调整的**验证集**

请注意，测试示例并没有以任何方式用于对模型进行选择，包括其超参数

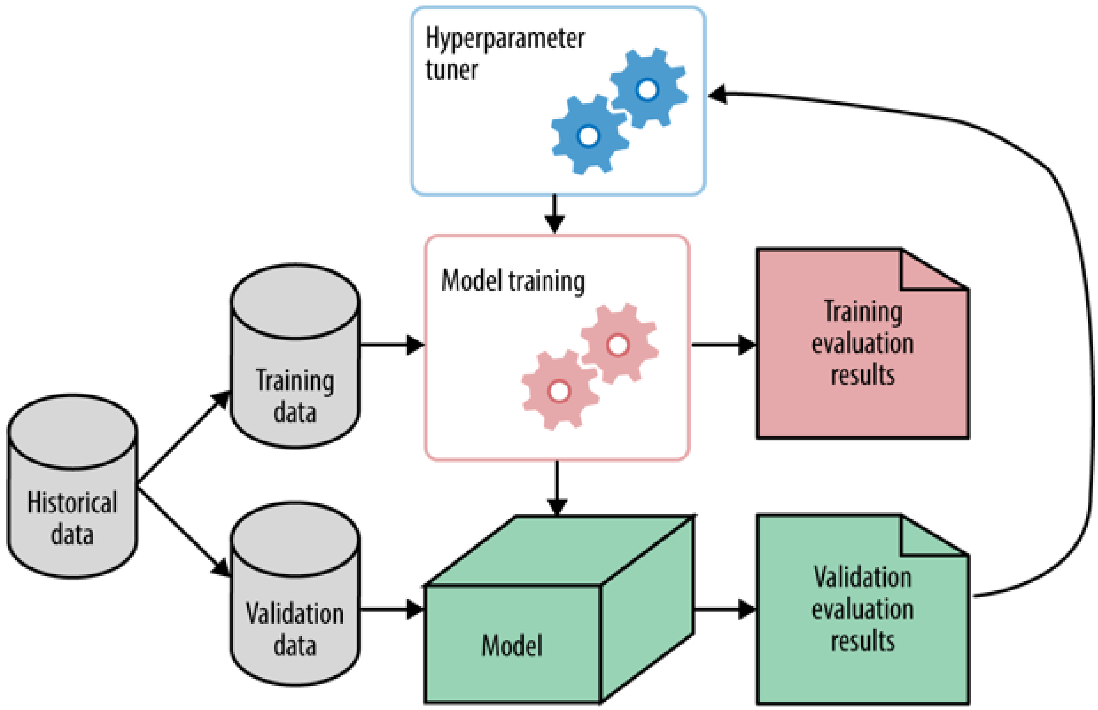
因此，测试集中的任何示例都不能用于验证集中。

具体来说，我们将训练数据划分为两个不相交的子集

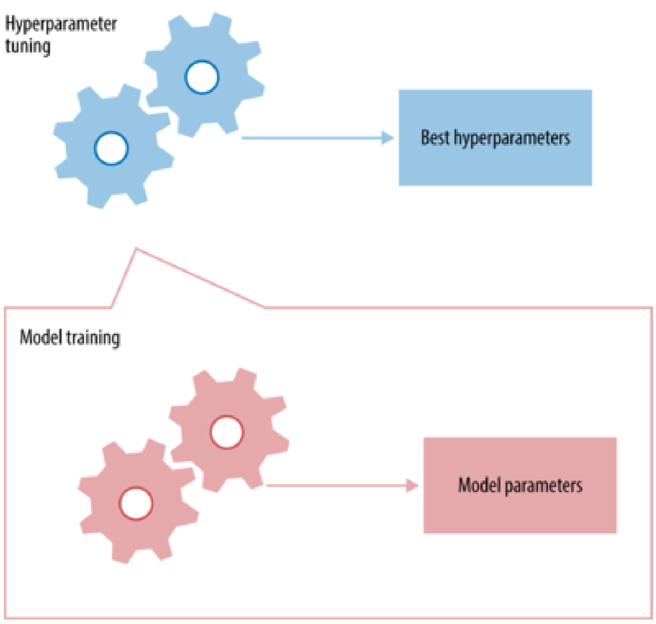
一个（训练集）用于训练模型参数。

一个（验证集）用于**估计训练期间或训练后的泛化误差，从而允许相应地更新超参数**。

模型选择是指选择适合数据的正确模型（或模型类型）的过程。



对特定超参数设置的每次试验都涉及训练模型——一个内部优化过程



The outcome of hyperparameter tuning is the best hyperparameter setting.

The outcome of model training is the best model parameter setting.

超参数车削算法网格搜索

网格搜索挑选出一个超参数值网格，对其中的每一个进行评估，并返回获胜者。

- If the hyperparameter is the number of leaves in a decision tree, then the grid could be 10, 20, 30, ..., 100.
- For regularization parameters, it is common to use exponential scale: 10^{-5}, 10^{-4}, 10^{-3}, ..., 1.