机器学习概述

http://blog.csdn.net/han\_xiaoyang/article/details/50469334

http://blog.csdn.net/han\_xiaoyang/article/details/52910022

**特征选择：**

scikit learn的预处理中比较常用的函数

http://scikit-learn.org/stable/modules/preprocessing.html

http://scikit-learn.org/stable/modules/classes.html#modulesklearn.feature\_extraction

特征选择 基于模型

http://scikit-learn.org/stable/modules/feature\_selection.html

 过滤型

Ø sklearn.feature\_selection.SelectKBest

 包裹型

Ø sklearn.feature\_selection.RFE

 嵌入型

Ø feature\_selection.SelectFromModel

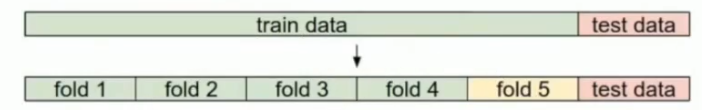
Ø Linear model，L1正则化

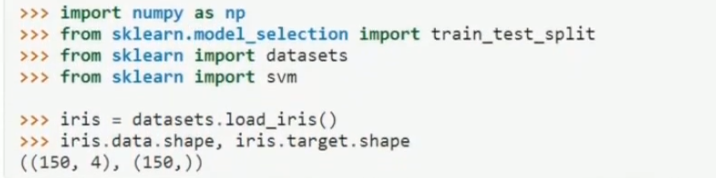
**模型选择：**

K折交叉验证（K-fold cross validation）

l http://scikit-learn.org/stable/modules/cross\_validation.html

把数据分成几个fold, 每次用其中一个fold





**超参数的选择:**

http://scikit-learn.org/stable/modules/grid\_search.html

http://scikit-learn.org/stable/modules/generated/sklearn.model\_selection.GridSearchCV.html



**plot learning curve**:绘制学习曲线

https://www.zybuluo.com/hanxiaoyang/note/545131

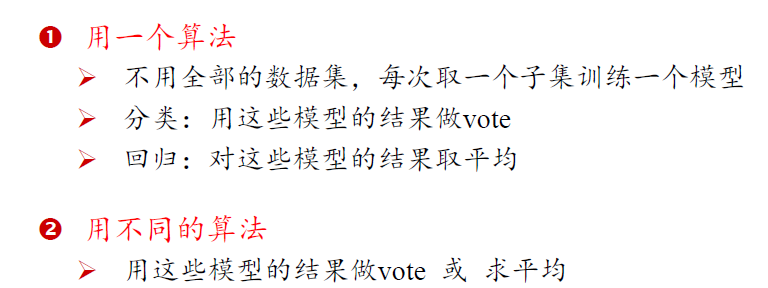
**模型集成：model assemble**

bagging / stacking(blending) / adaboost

http://scikit-learn.org/stable/modules/classes.html#module-sklearn.ensemble

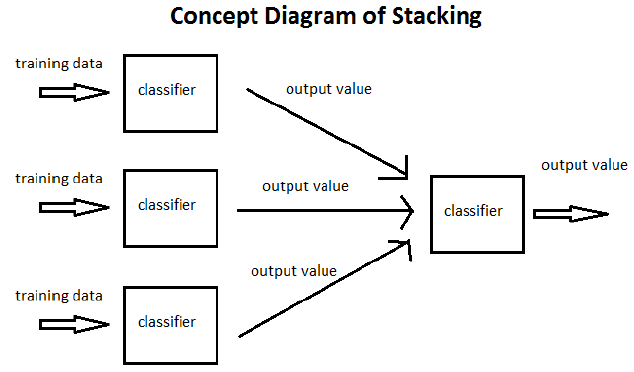
**bagging:**

http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html



**stacking:**





blending:

把 一同模型的输出线性组合