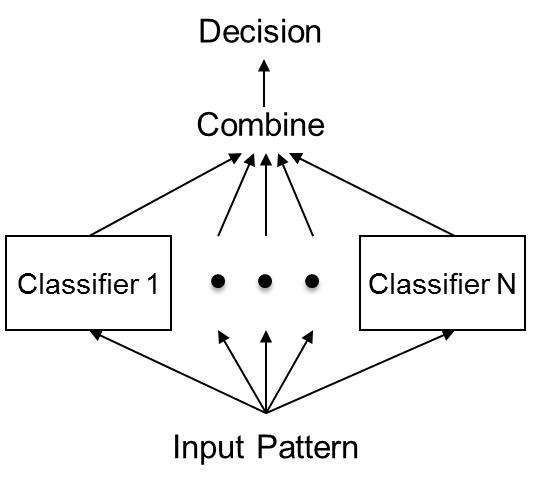
# 集成学习

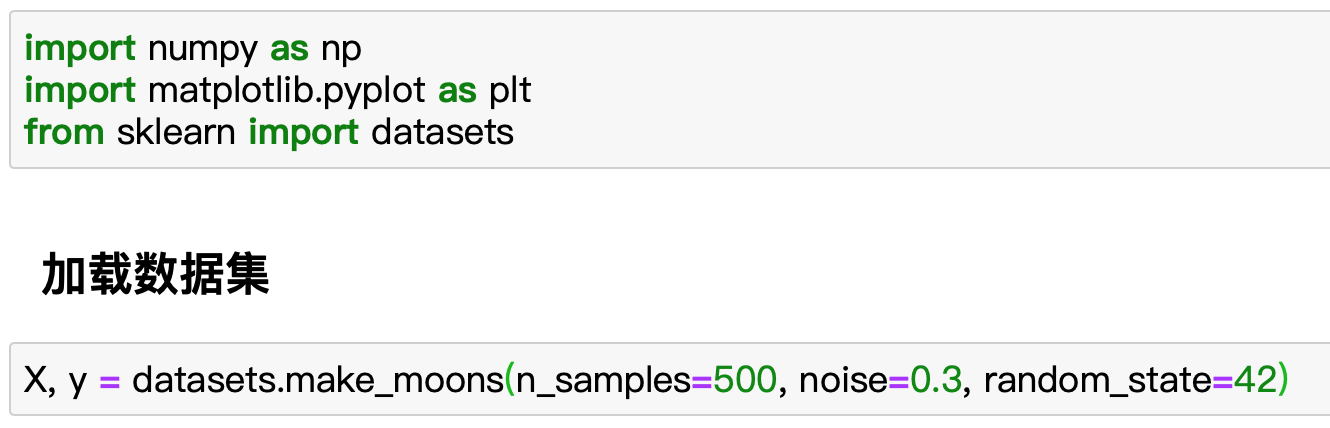
* 概述

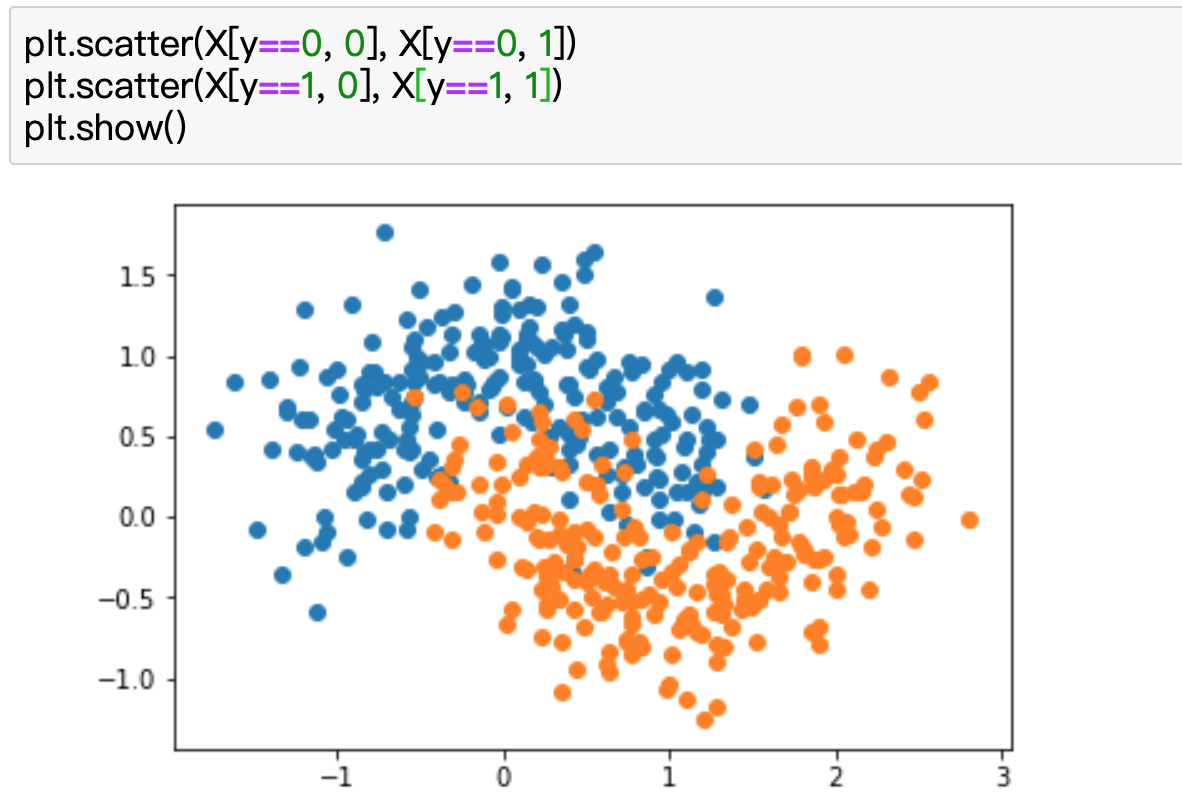
**集成学习是使用一系列学习器进行学习，并使用某种规则把各个学习结果进行整合从而获得比单个学习器更好的学习效果的一种机器学习方法。**

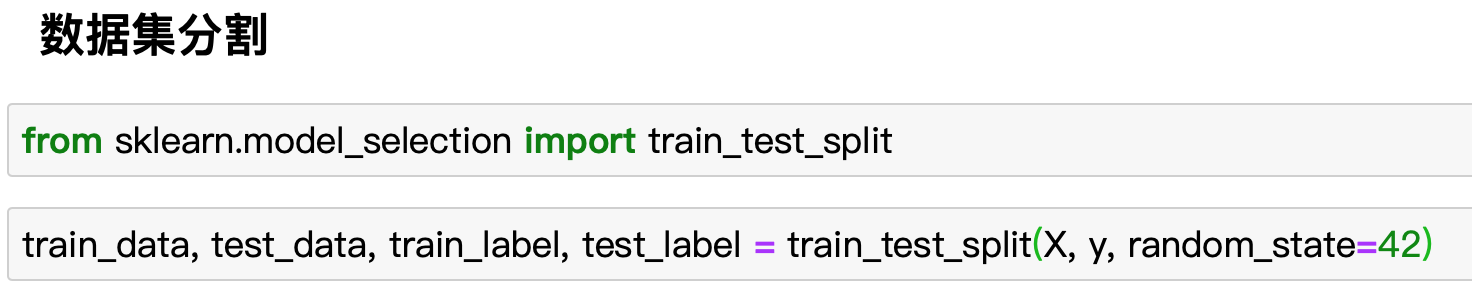


* Voting Classifier

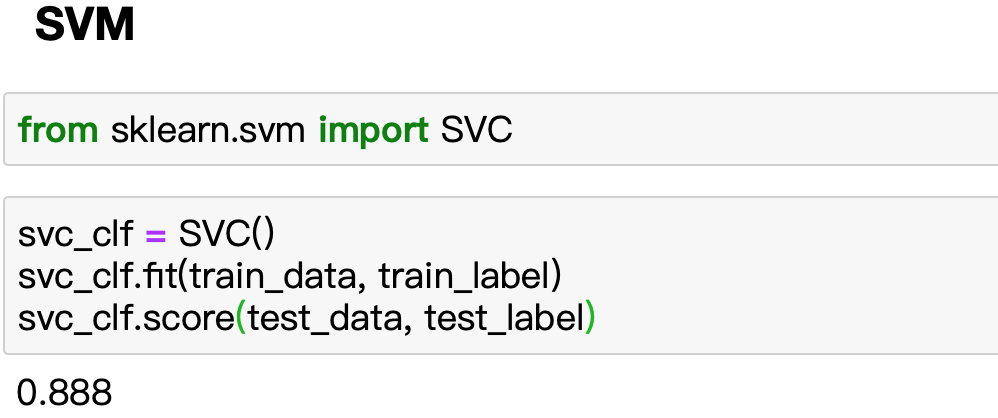
**Voting Classifier是一种投票决策方式，采用少数分从多数的方式来选择最终的输出结果。**

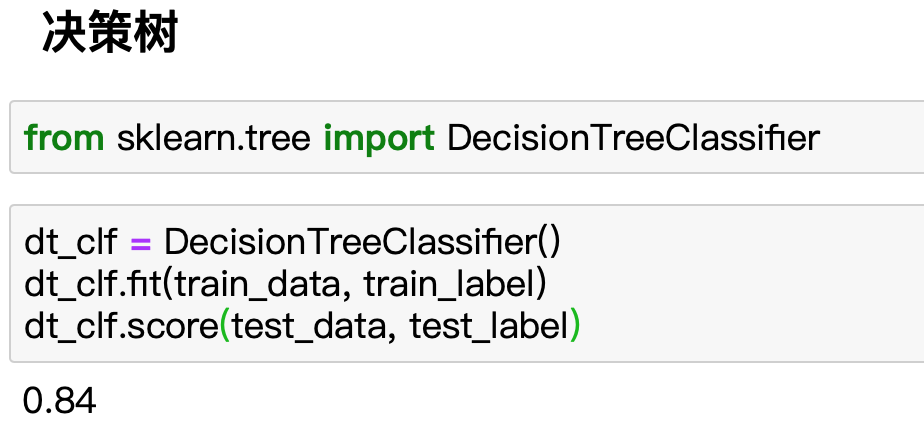
****

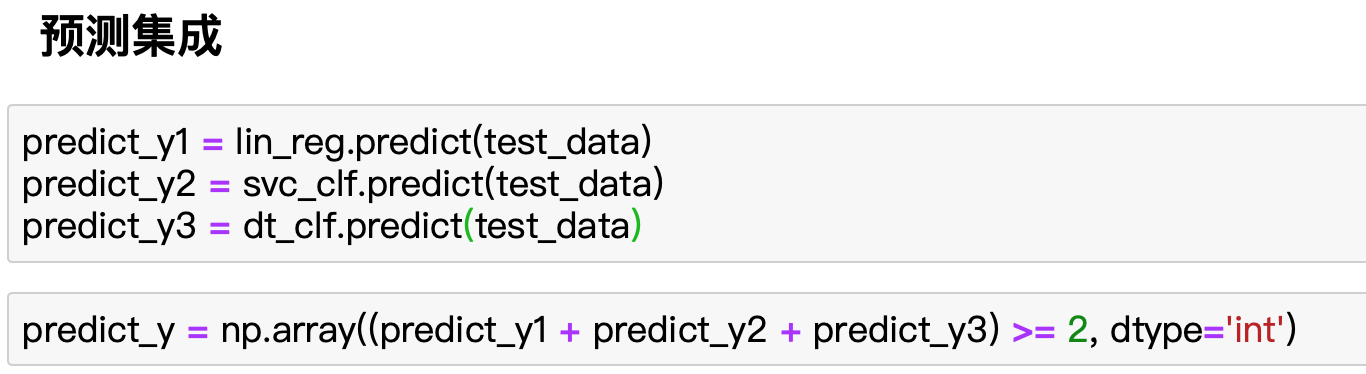
****

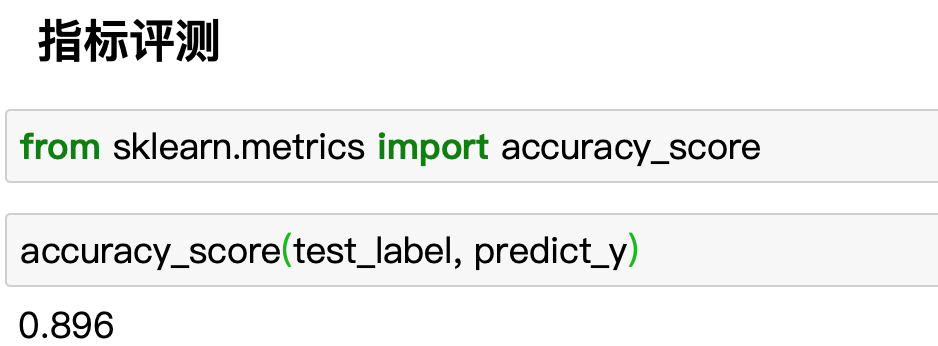
****

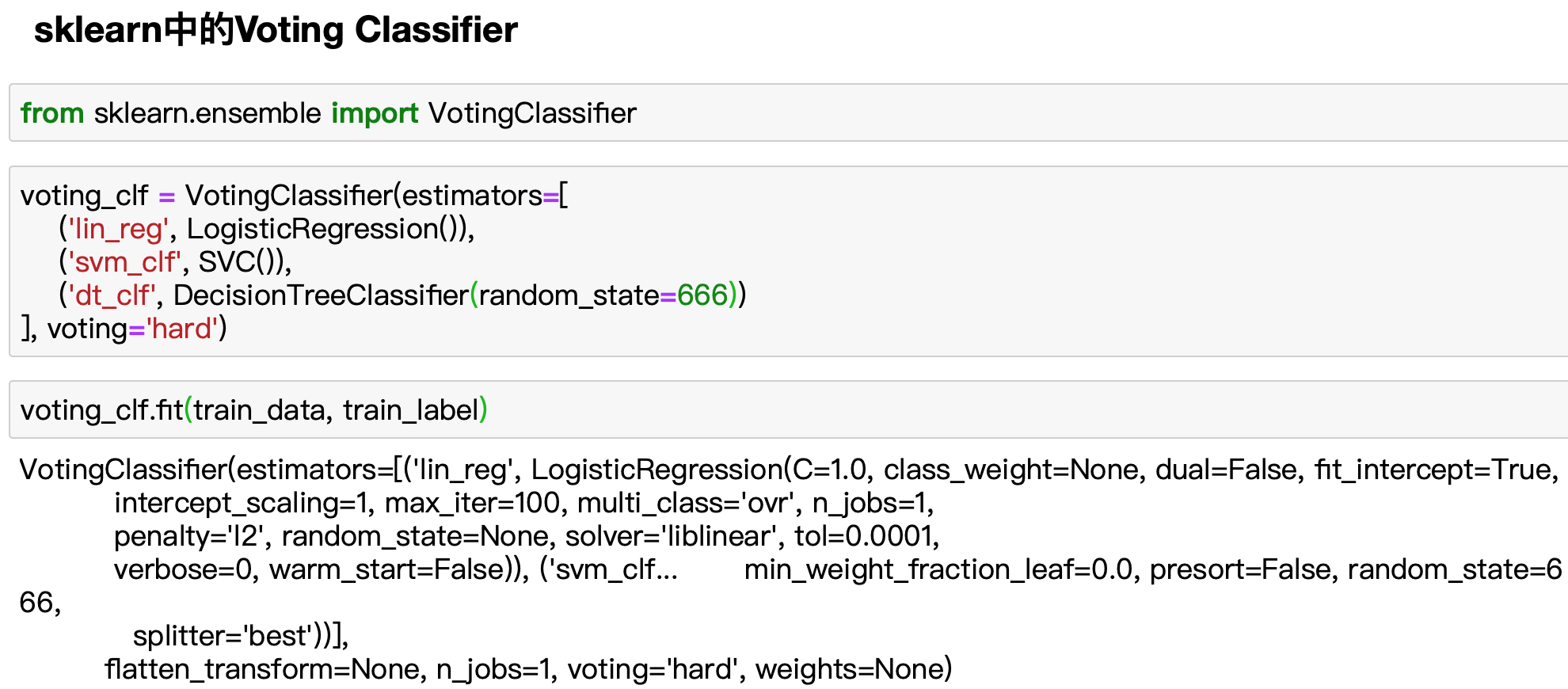
****

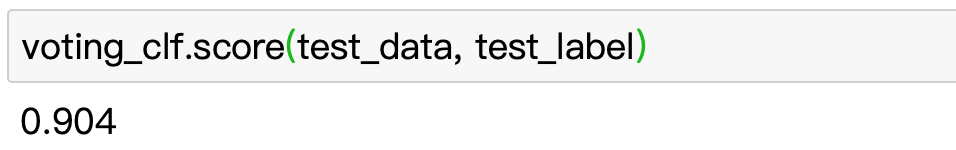
****

****

****

****

****

****

* Soft Voting

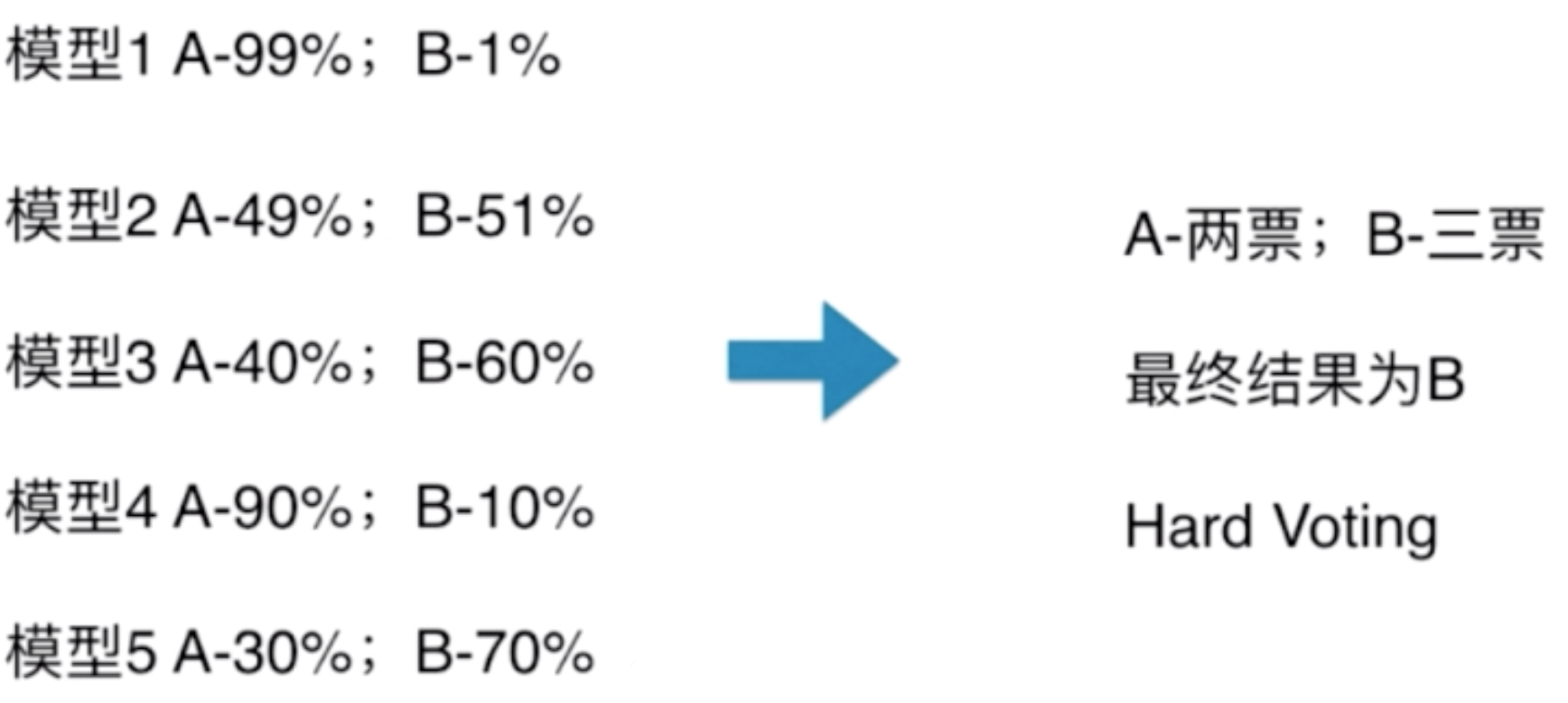
**少数服从多数的投票方式在很多时候并不合理。**

**更合理的方式，应该是有权值。**

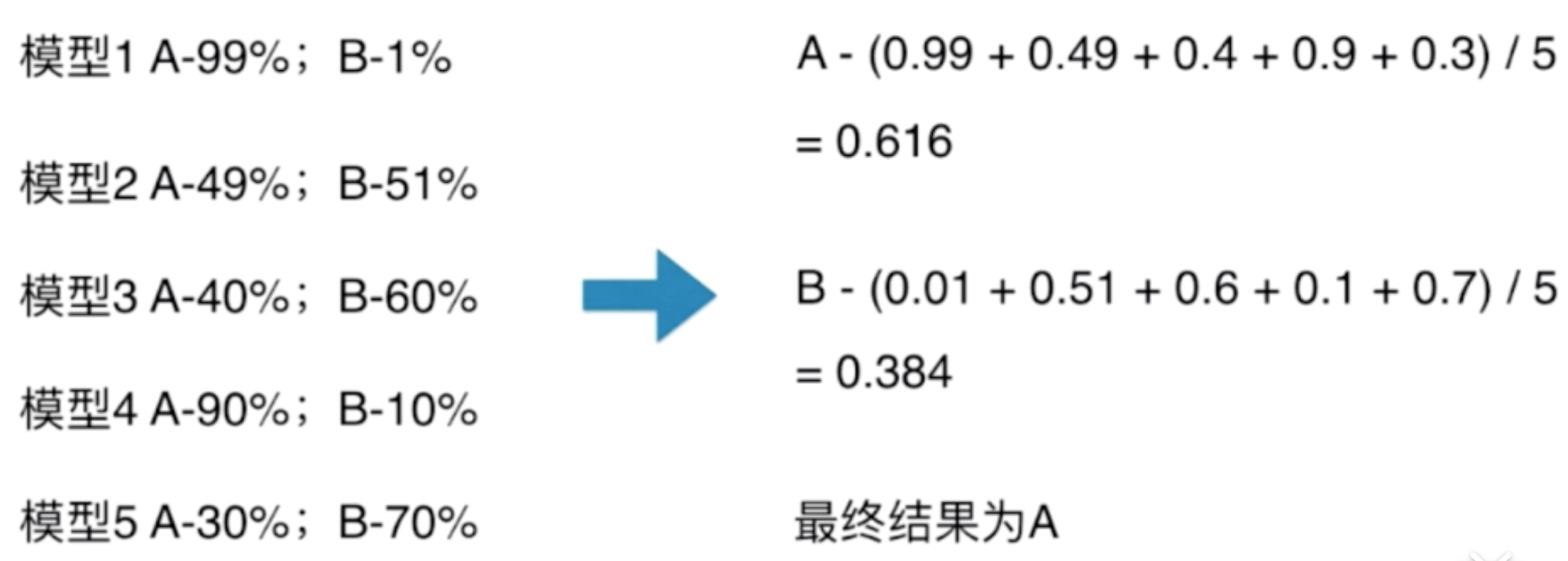
**例如**

**唱歌比赛中，专业评委的投票比观众的投票比重更大**

**Hard Voting**

****

**Soft Voting**

****

**注：使用Soft Voting，必须要求每个模型都能够估计概率**

****

* Bagging和Pasting

**虽然有很多的机器学习算法，但从投票的角度看，投票的数量任然不够多。**

**解决方案**

**1.创建更多的子模型，集成更多子模型的意见**

**2.子模型之间不能一致，子模型之间要存在差异化**

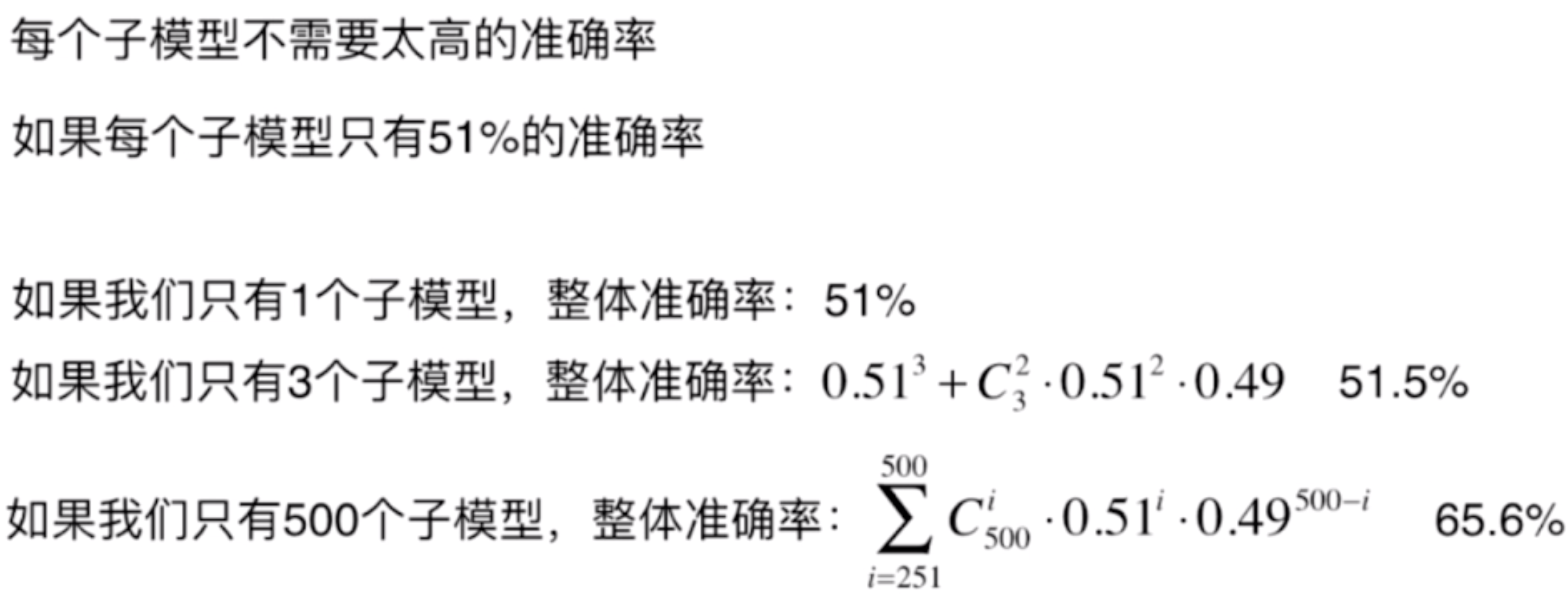
**如何让子模型之间存在差异化?**

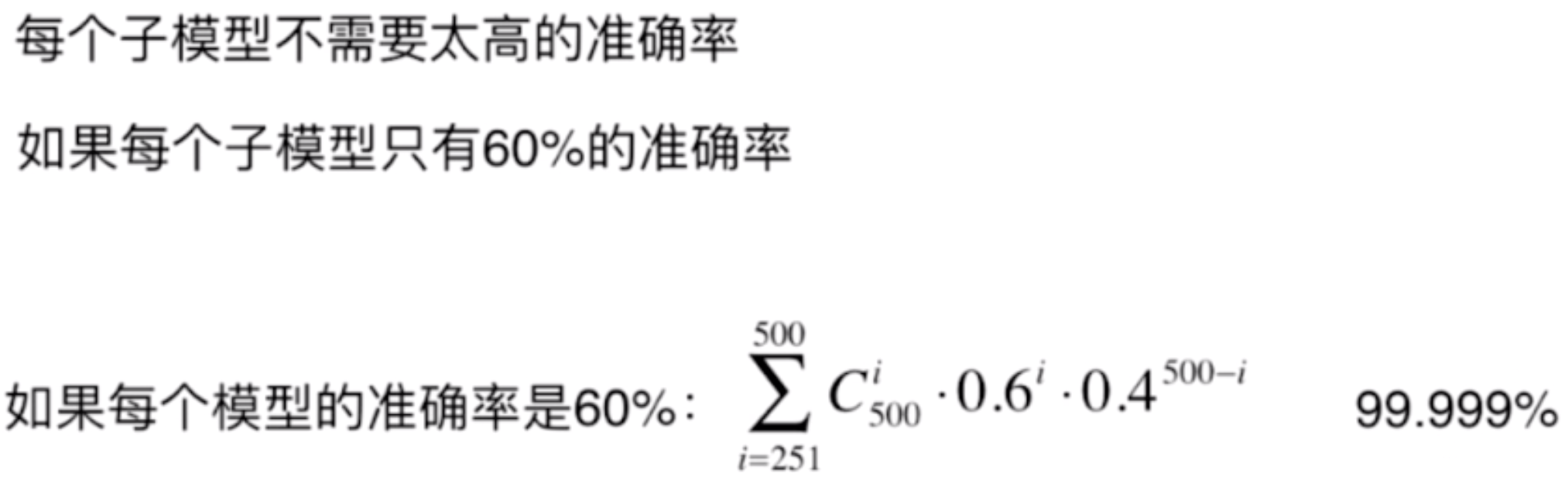
**每个子模型只看样本数据的一部分**

**例如：总共有500个样本，每个子模型只看100个样本数据。**

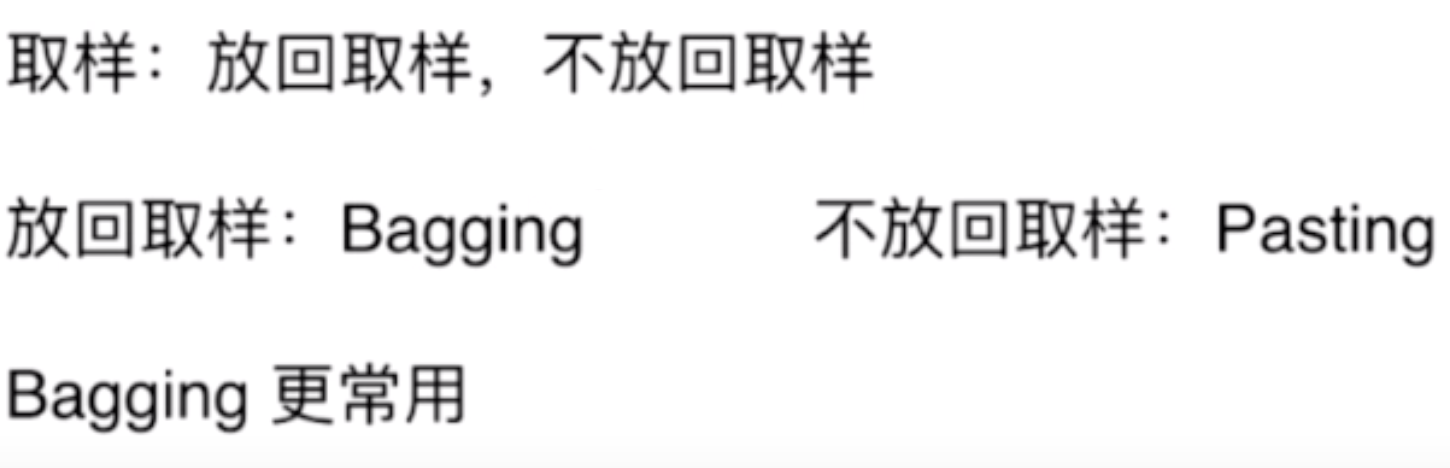
**每个子模型不需要太高的准确率**

**例如**

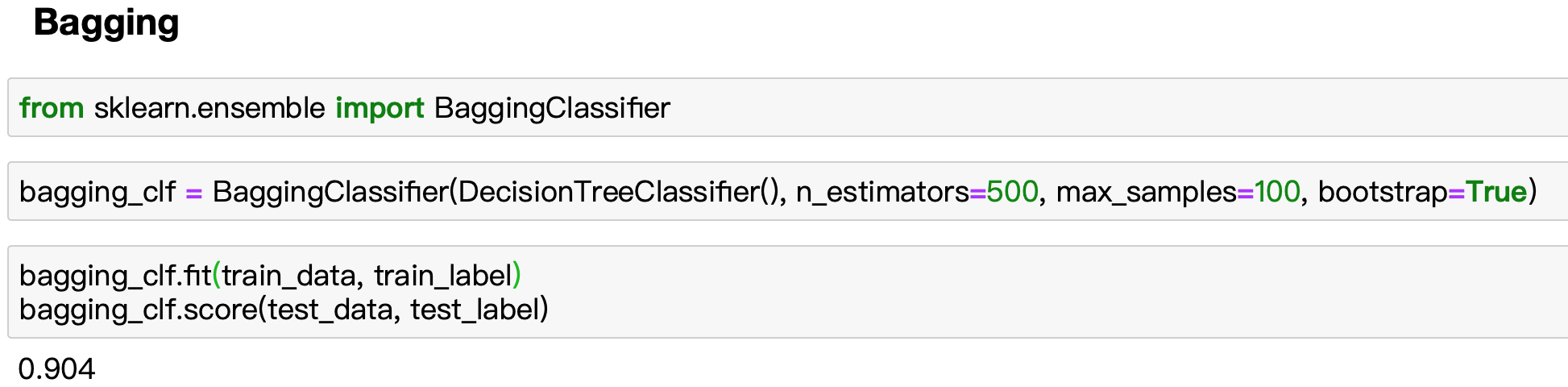
****

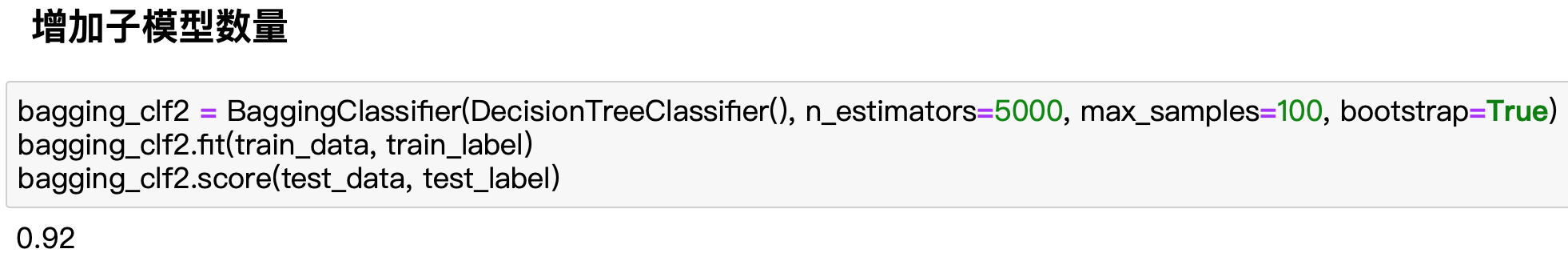
****

**创建子模型差异性**

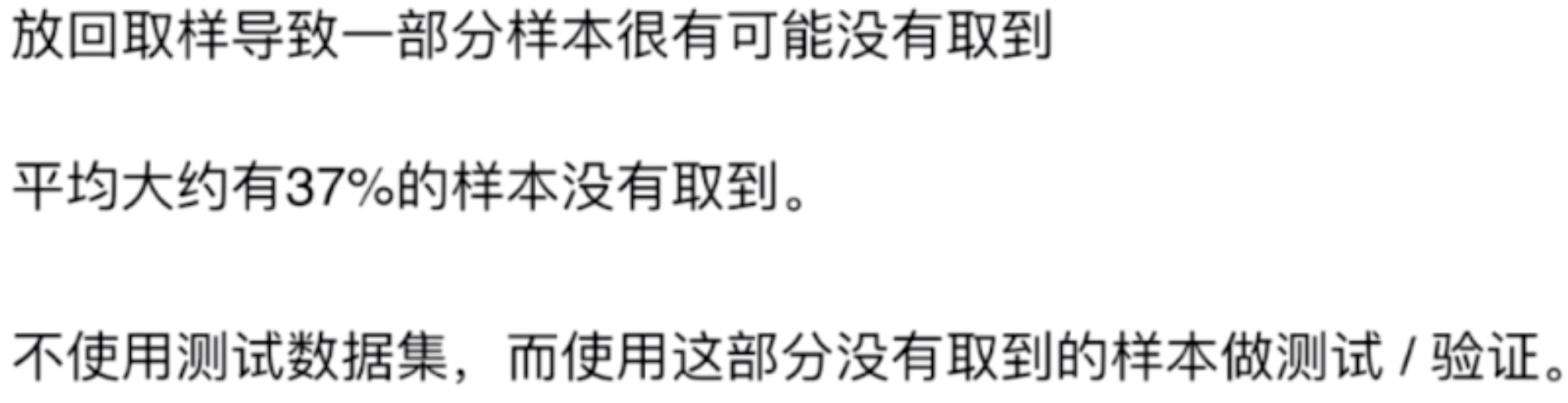
****

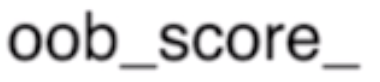
****

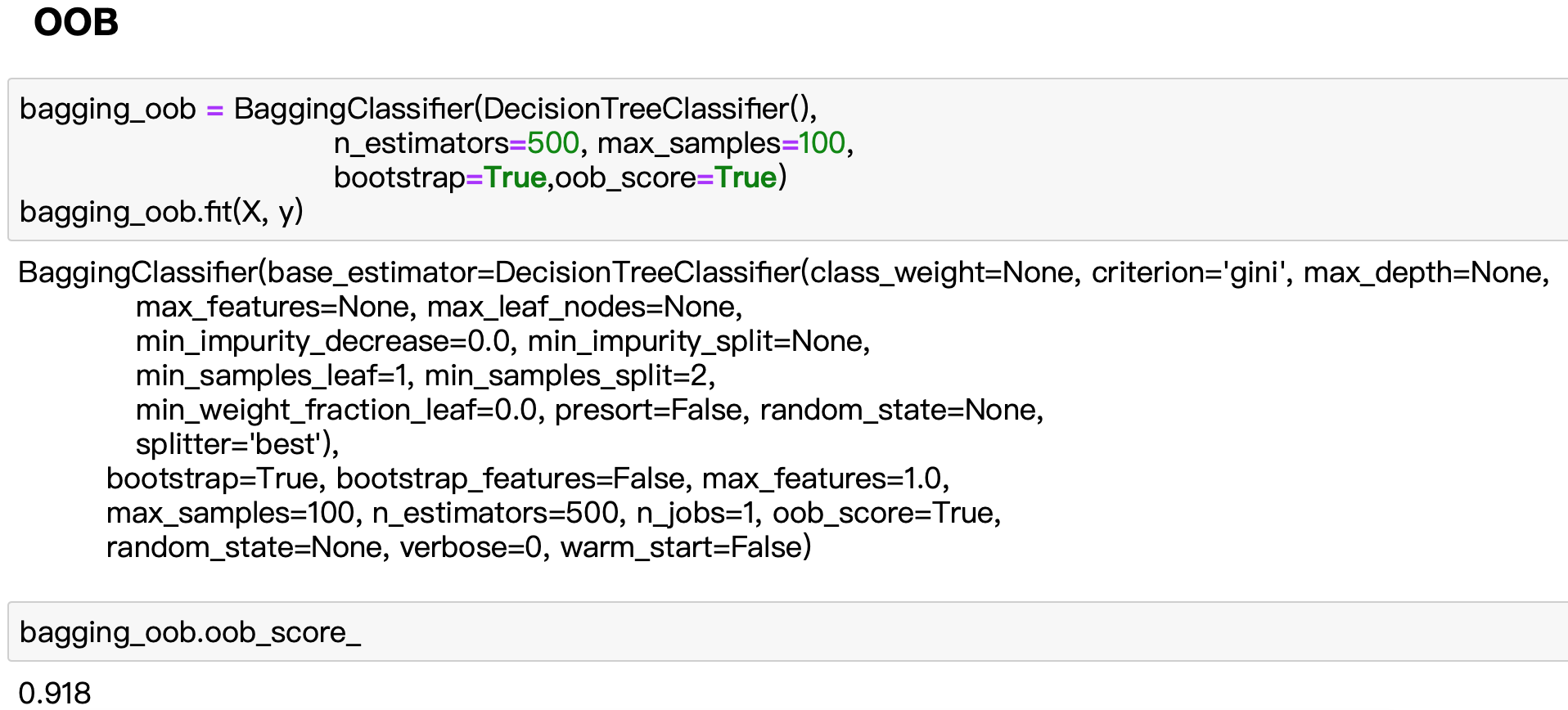
****

****

* OOB(Out-of-Bag)

****

****

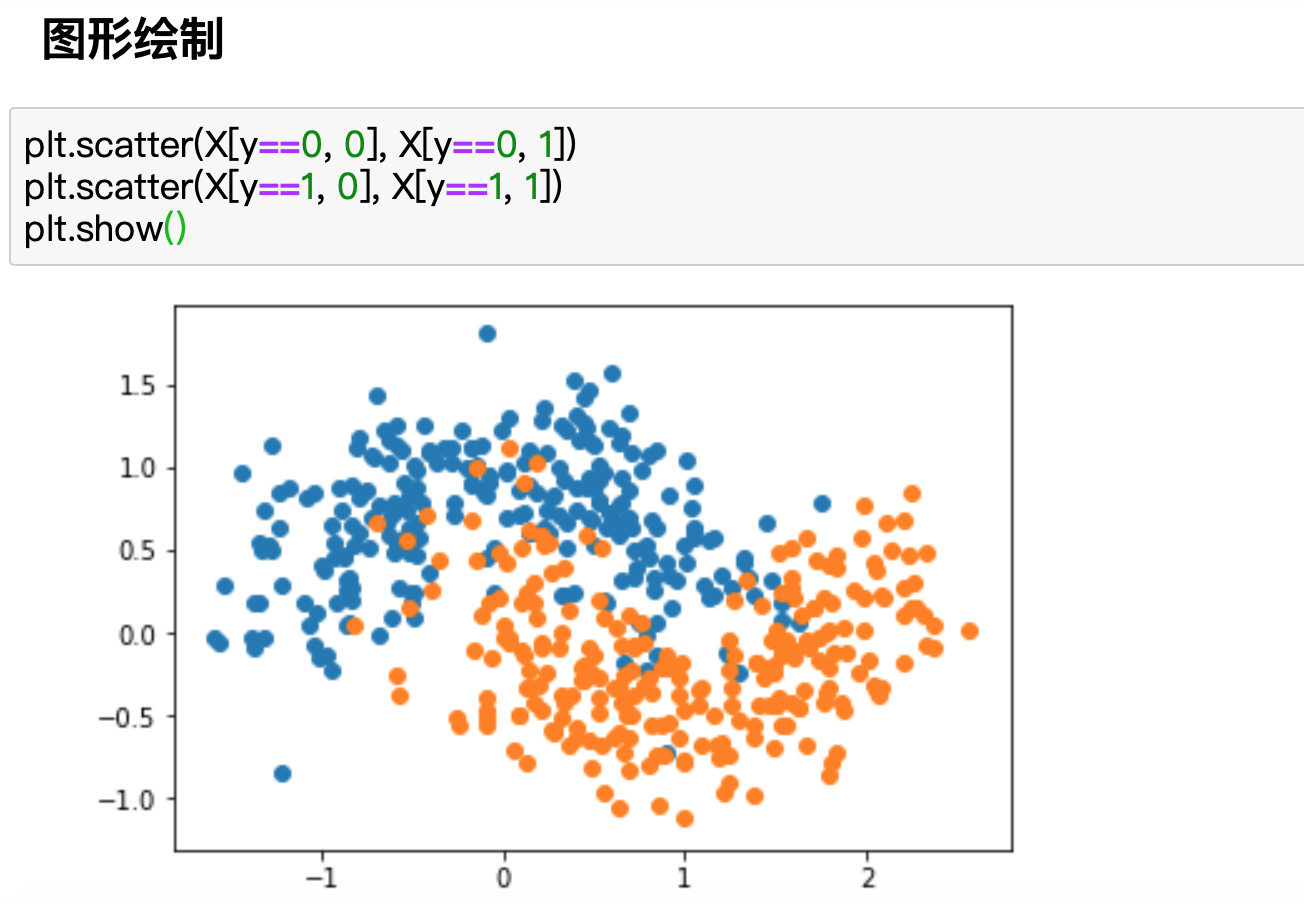
****

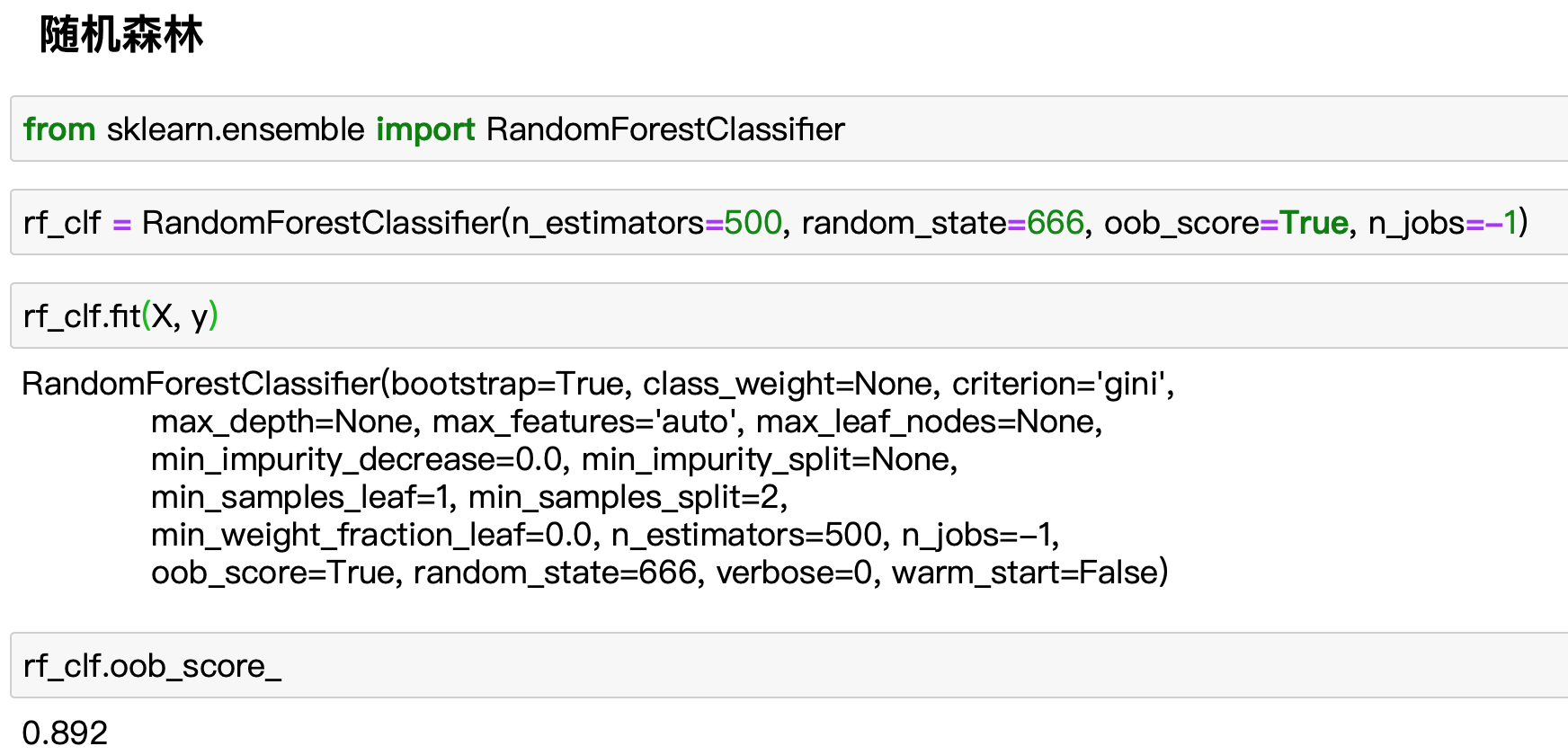
* 随机森林

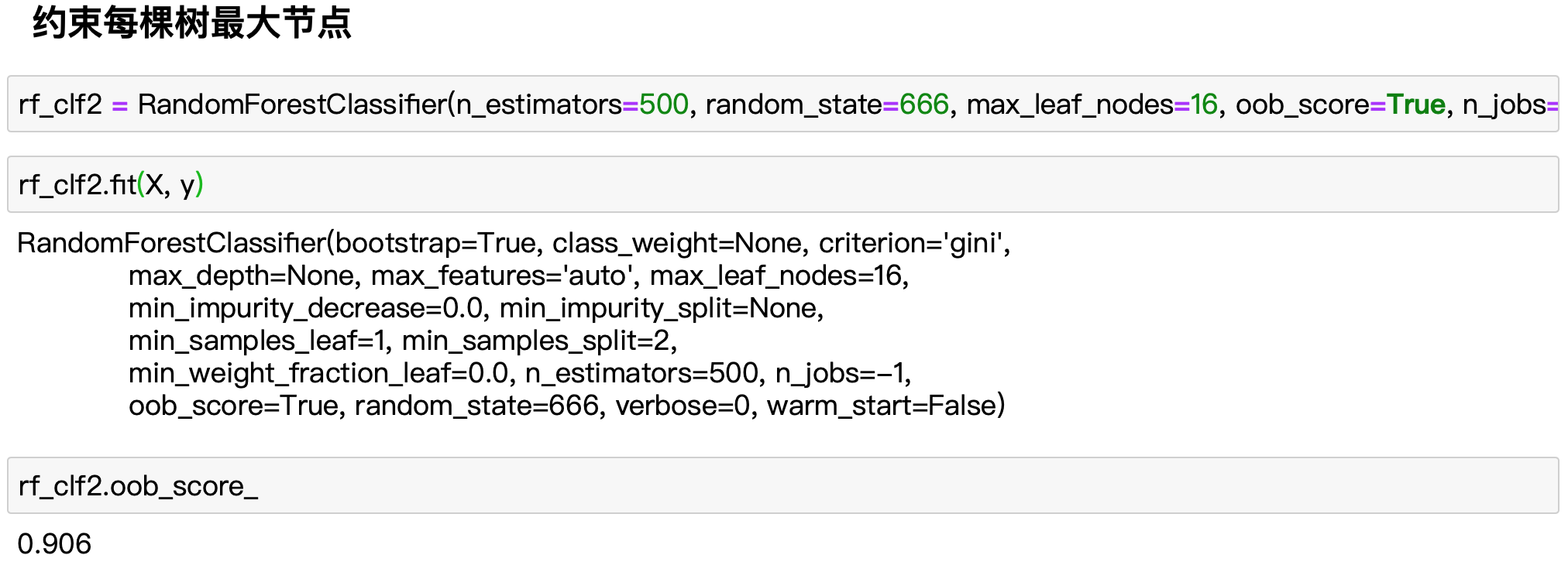
**随机森林，实际是由多棵决策树组成。**

**决策树在节点的划分上，在随机的特征子集上寻找最优划分特征。**

****

****

****

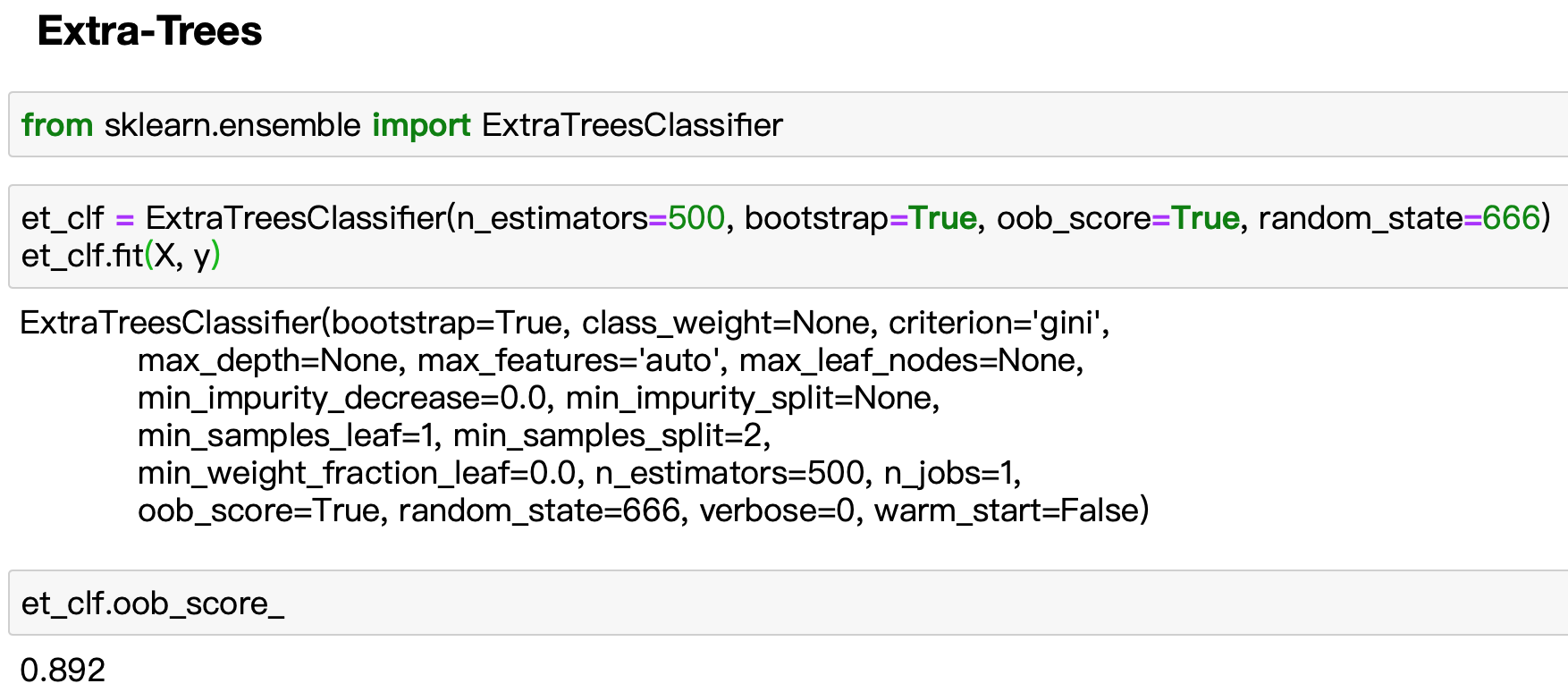
****

* Extra-Trees

**决策树在节点划分上，使用随机的特征和随机的阈值。**

**提供了额外的随机性，抑制了过拟合，但增大了偏差。**

**由于随机性的因为，Extra-Trees具有更快的训练速度。**

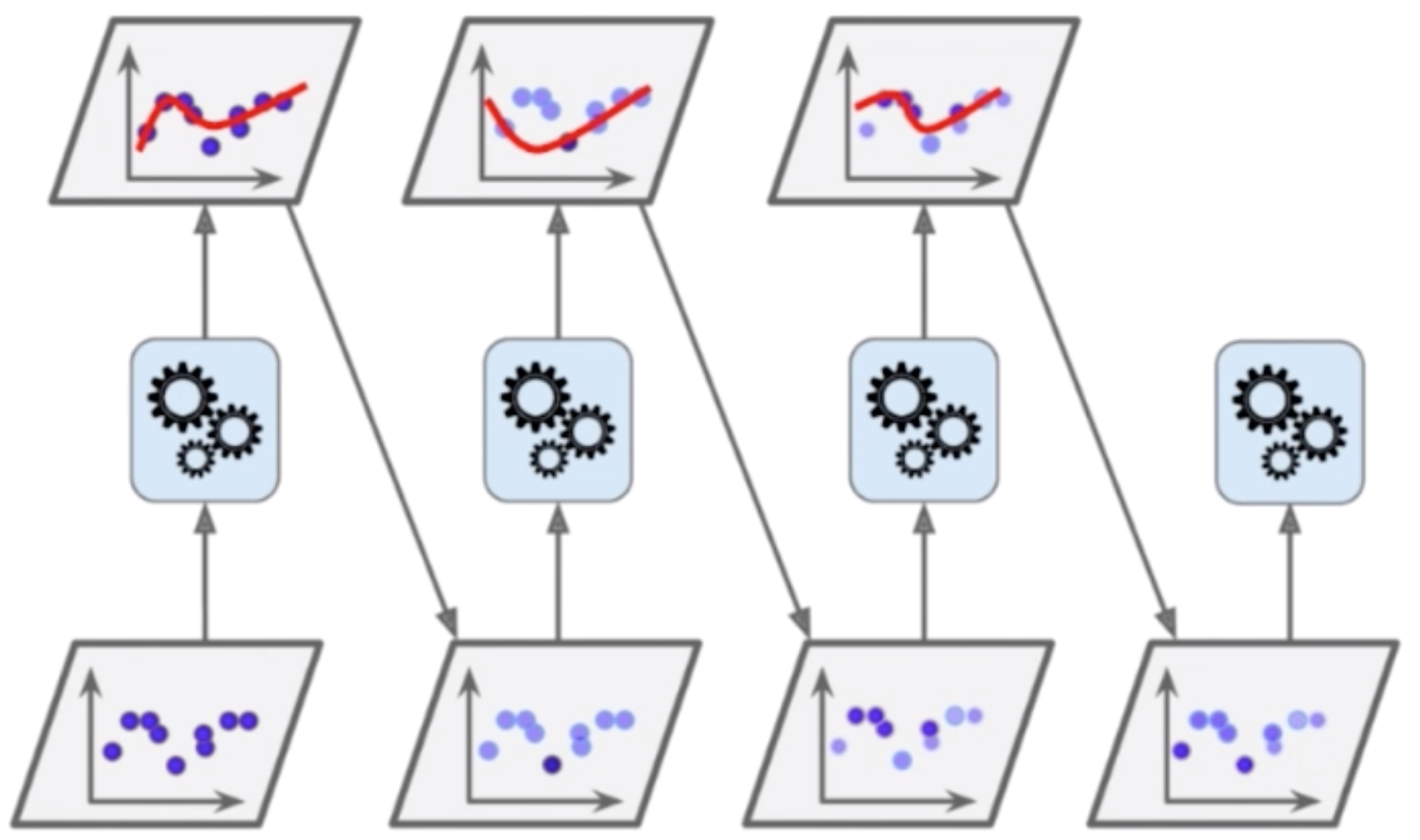
****

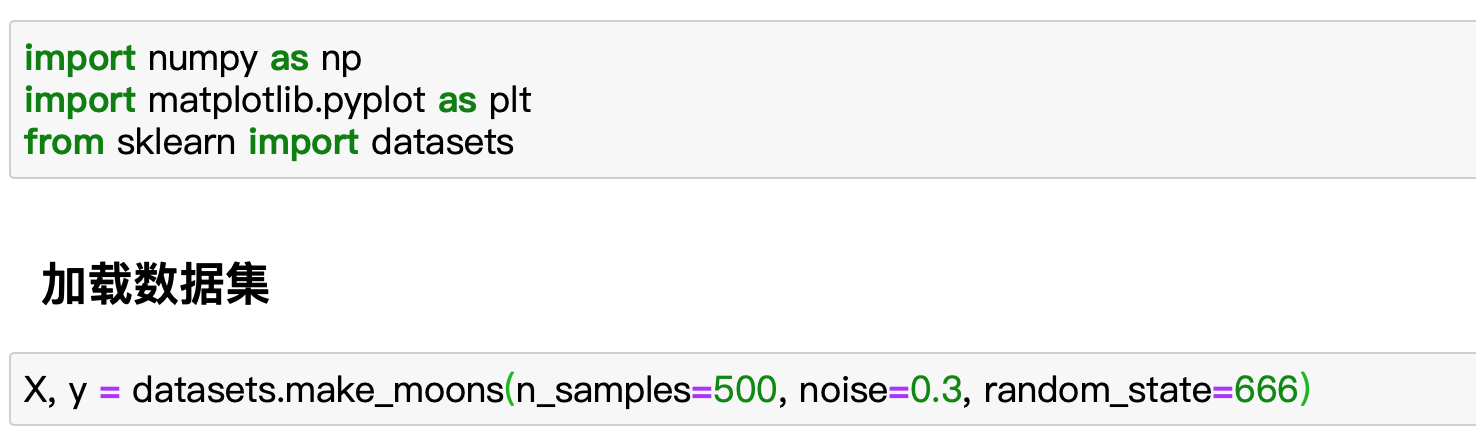
* Boosting

**Boosting也是用于集成多个模型。**

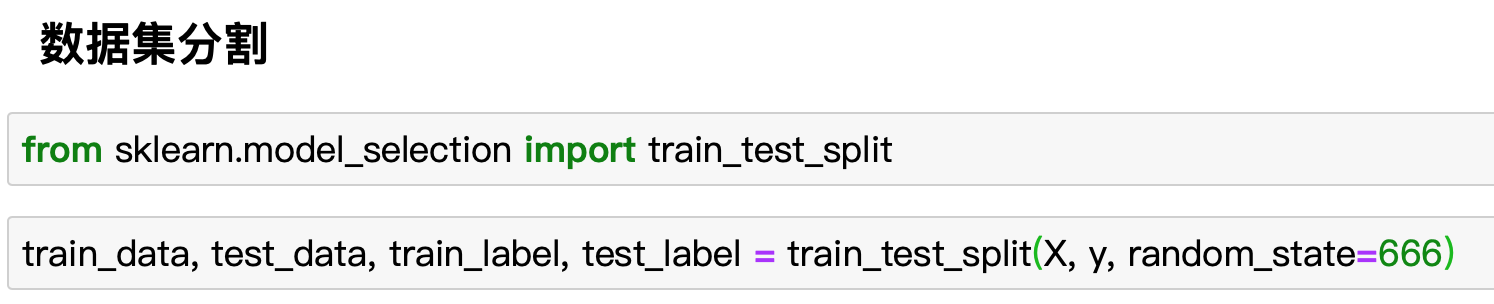
**但每个模型都是在尝试增强(Boosting)整体的效果。**

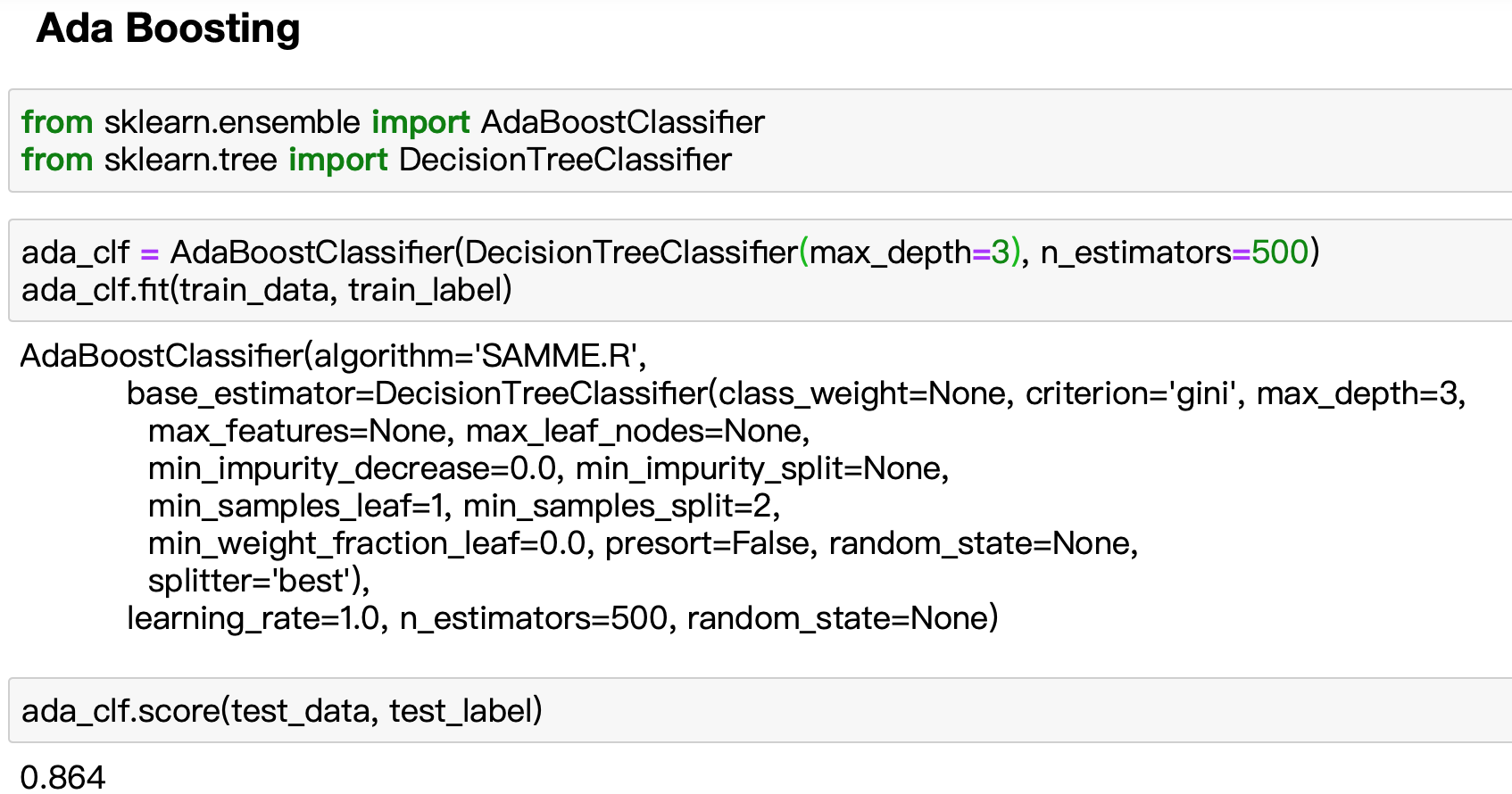
**Ada Boosting**

****

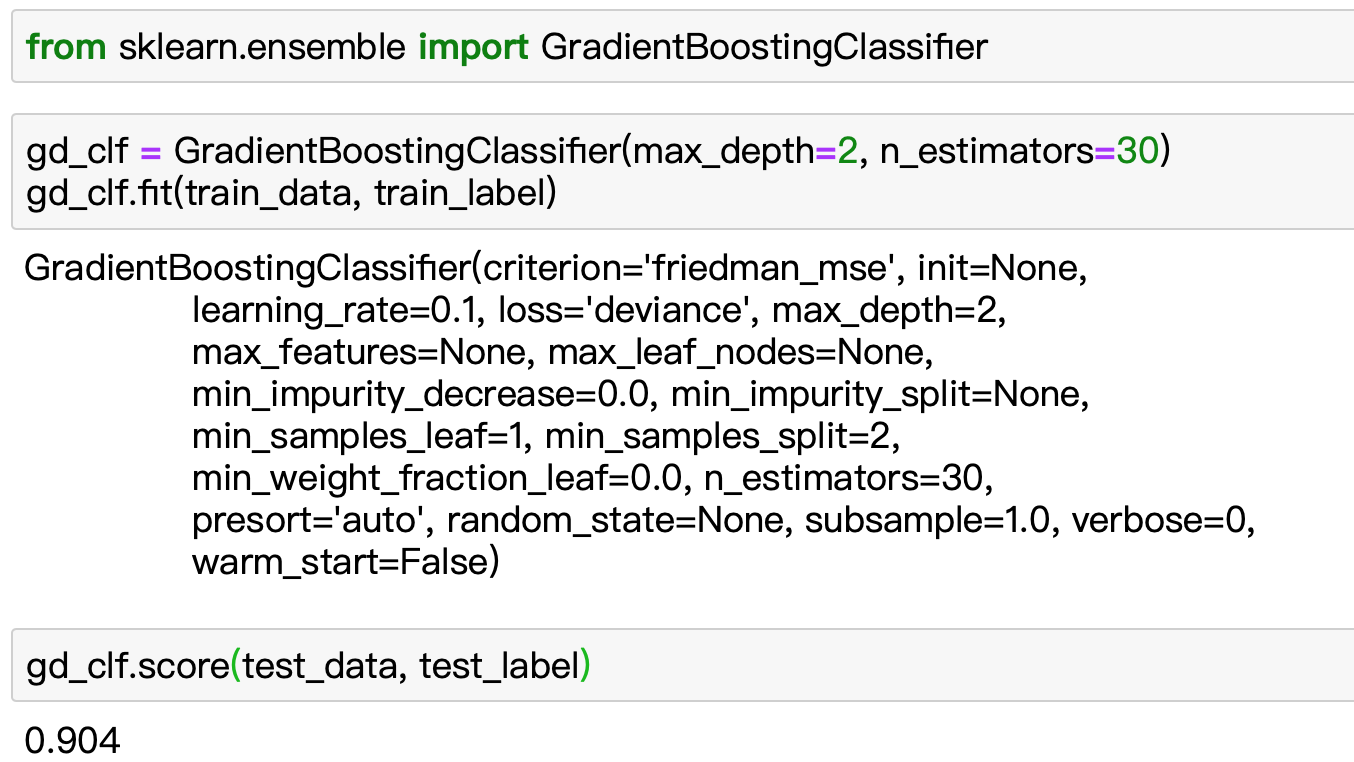
****

****

****

****

**Gradient Boosting**

****