### 02 集合模型 - AdaBoost

###### 集合算法

单个模型得到的预测结果总是片面的，所以可以使用不同的条件构建多个不同模型，给出的多个预测结果，利用平均（回归）或者投票（分类）的方法，得到最终预测结果。

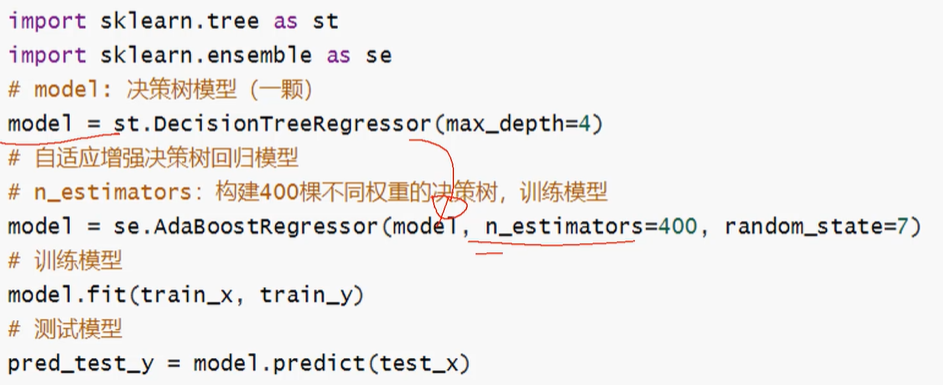
注：基于决策树的集合算法，就是按照某种规则，构建多棵彼此不同的决策树模型，分别给出针对未知样本的预测结果，最后通过平均或投票得到相对综合的结论。

###### AdaBoost回归

1. 首先为样本矩阵中的样本随机分配初始权重，由此构建一棵带有权重的决策树，在该预测进行输出时，通过加权平均的方式产生预测值。
2. 将训练样本代入模型，预测其输出，对那些预测值和实际值误差大的样本，提高其权重，由此形成第二课决策树。
3. 重复以上过程，构建出多个不同权重的决策树。
4. 最终由得到的这些决策树一同对未知输出的样本进行预测，通过平均的方式产生预测值。

问：为什么是提高其权重。

###### AdaBoost回归实现



###### 特征重要性

作为决策树模型训练过程的副产品，决策树划分子表时所使用的特征先后顺序也标志了相应特征的重要程度，此即为该特征重要性指标。

训练得到的模型对象提供了属性：feature\_importances\_来存储每个特征的重要性。

