

一、概述

本文主要讲述机器学习中的监督学习的实例，房屋价格的预测，一个房屋有很多现实条件，如面积、地理位置、城市。这些条件对房屋的价格都会有影响，我们将从已有数据中得出大致关系。这里分别使用拟合决策树和AdaBoost算法。主要参考《Python机器学习经典实例》，以及各类博客。

二、内容

首先我们说一下回归树与分类树的区别：树模型有分类树与回归树，分类树是面向分类的，在每个叶子节点出来的是一个分类标签；回归树是面向回归的，和拟合函数一样，最终输出的是一个连续值。这里我们用的就是面向回归的决策树来进行训练。

- 1、导入数据
- 2、分别创建决策树回归模型，与AdaBoost算法的决策树回归模型
- 3、开始训练
- 4、得出结果、计算误差，评价两算法的区别

数据中每个房屋有13个特征影响价格，哪个特征更重要？在sklearn库中构建模型时，feature_importances_函数反馈了这一问题。两算法中对价格影响的前三位中特征相同、排名却不相同，如普通的拟合决策树第一位为RM，而AdaBoost中第一位为LSTAT。根据计算出的均方误差（mse）和解释回归模型的方差得分（evs）来看，AdaBoost模型得到的结果更为接近现实。

指标	含义
CRIM	城镇人均犯罪率
ZN	住宅用地比例
INDUS	非零售商业用地比例
CHAS	CHAS变量，0或者1
NOX	一氧化氮浓度
RM	每个住宅的平均房间数
AGE	1940年以前自用房屋的比例
DIS	距离五个波士顿就业中心的加权距离
RAD	距离高速公路的便捷指数
TAX	该地区每一万美元的不动产税率
PRTATIO	该地区教师学生比例
B	该地区黑人比例
LSTAT	该地区中低收入阶层比例

AdaBoost算法，是将同一个训练集训练不同的分类器(弱分类器)，然后把这些弱分类器集合起来，构成一个更强的最终分类器（强分类器）。

以我的理解AdaBoost可将同特征的数据集成若干个弱分类器，最终集成**强分类器**，这样AdaBoost可将特征对房价的影响放大，从而筛选出不同比重的特征。

三、心得体会

这是机器学习中的一个很经典的监督学习的例子。从数据导入、模型搭建，计算结果都有一个完整的流程，并且过程简洁，便于理解

参考博客：https://blog.csdn.net/huanfeng_AI/article/details/107675987
<https://baike.baidu.com/item/adaboost/4531273?fr=aladdin>