

实战案例3-1：手机价格预测(1)

作者：Robin

日期：2018/03

提问：[小象问答](#)

数据集来源：[kaggle](#)

声明：[小象学院](#)拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散布。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利

1. 案例描述

该项目的目的主要是基于手机的一些基本配置信息（如：电池容量、是否支持双卡等）作出价格区间（等级）的预测。该预测有助于手机公司制定合理的价格，给其他手机制造商带来强硬竞争；另一方面，对于消费者来说，可以帮助他们验证是否为目标手机支付了合理的价格。

2. 数据集描述

- Kaggle[提供的数据集](#)
- 数据字典
 - **battery_power**: 电池容量，整型
 - **blue**: 是否为蓝色，整型（0和1）
 - **clock_speed**: CPU时钟频率，浮点型
 - **dual_sim**: 是否支持双卡，整型（0和1）
 - **fc**: 前置摄像头像素，整型，单位（百万）
 - **four_g**: 是否支持4G，整型（0和1）
 - **int_memory**: 内存容量，整型，单位（GB）
 - **m_dep**: 手机厚度，浮点型，单位（cm）
 - **mobile_wt**: 手机重量，浮点型
 - **n_cores**: 处理器核数，整型
 - **pc**: 主摄像头像素，整型，单位（百万）
 - **px_height**: 分辨率高度，整型
 - **px_width**: 分辨率宽度，整型
 - **ram**: RAM内存，整型，单位（MB）
 - **sc_h**: 屏幕长度，整型，单位（cm）
 - **sc_w**: 屏幕宽度，整型，单位（cm）
 - **talk_time**: 最长通话时间，整型
 - **three_g**: 是否支持3G，整型（0和1）
 - **touch_screen**: 是否为触屏，整型（0和1）
 - **wifi**: 是否支持无线，整型（0和1）
 - **price_range**: 价格等级，整型（0, 1, 2, 3）

3. 任务描述

- 使用scikit-learn建立不同的机器学习模型进行手机价格等级预测

4. 主要代码解释

- 代码结构

```
lect04_proj
├── data
│   └── data.csv    # CSV数据文件
├── output
│   └── model_comparison.csv    # 模型比较结果的CSV文件(程序的输出)
├── config.y        # 配置文件
├── main.py          # 主程序
└── lect04_proj_readme.pdf    # 案例讲解文档
```

- main.py

使用的模型及相关参数配置。该项目中使用了4个机器学习模型，并为不同的学习模型指定了参数空间。如：kNN，指定了3个k值用于比较对结果的影响：5, 10, 15。

```
def main():
    ...
    model_name_param_dict = {'kNN':    [5, 10, 15],
                             'LR':     [0.01, 1, 100],
                             'SVM':    [0.01, 1, 100],
                             'DT':     [50, 100, 150]}
    ...
```

- main.py

选择最优模型，`np.argmax(list)`返回list中最大值的索引号，如`np.argmax([5, 10, 2])`的返回值是1，因为第1个位置上的值最大，为10。之后根据该索引号获取最优的准确率及模型。

```
def train_model(X_train, y_train, X_test, y_test, param_range, model_name='SVM'):
    ...
    # 记录最优模型
    best_idx = np.argmax(scores)
    best_acc = scores[best_idx]
    best_model = models[best_idx]
    ...
```

5. 案例总结

- 该项目通过预测手机价格等级，巩固并实践了使用scikit-learn搭建简单的预测模型：
 - 机器学习流程
 - scikit-learn中常用预测模型的使用
 - 参数的选择

6. 课后练习

- 参考随堂代码，试着使用决策树输出特征的重要性。

参考资料

1. [scikit-learn官方教程](#)
2. [通过scikit-learn理解机器学习](#)