

# 逻辑回归的实现——学习报告——Day 6

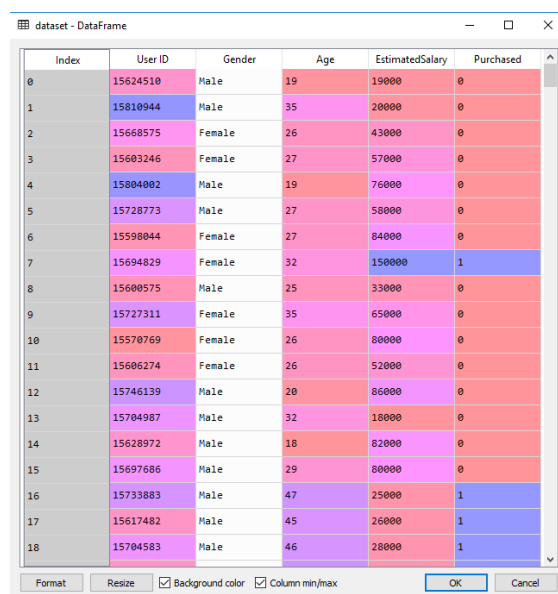
2019 年 7 月 15 日

## 1 数据预处理

- 导入库

和之前一样，导入numpy和pandas库，不过本次需要导入绘图库matplotlib.pyplot。

- 导入数据集



Index	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0
5	15728773	Male	27	58000	0
6	15598044	Female	27	84000	0
7	15694829	Female	32	150000	1
8	15600575	Male	25	33000	0
9	15727311	Female	35	65000	0
10	15570769	Female	26	80000	0
11	15606274	Female	26	52000	0
12	15746139	Male	20	86000	0
13	15704987	Male	32	18000	0
14	15628972	Male	18	82000	0
15	15697686	Male	29	80000	0
16	15733883	Male	47	25000	1
17	15617482	Male	45	26000	1
18	15704583	Male	46	28000	1

图 1: 数据集

- 将数据集分成训练集和测试集

- 特征缩放

将不同特征值化到同一区间，使每个特征中的数值平均变为0(将每个特征的值都减掉原始数据中该特征的平均值)、标准差变为1。

## 2 逻辑回归模型

昨天我们认识到逻辑回归是一个线性分类器，因此购买SUV和不购买SUV的两类用户将被一条直线分割。导入逻辑回归类，接着创建该类的对象，其将作为我们训练集的分类器。

- 讲逻辑回归应用于训练集

## 3 预测

- 预测测试集结果

```
[0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0  
0 0 1 0 0 0 0 1 0 0 1 0 1 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0  
0 0 1 0 1 1 1 1 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1]
```

## 4 评估预测

我们预测了测试集。现在我们将评估逻辑回归模型是否正确的学习和理解。因此这个混淆矩阵将包含我们模型的正确和错误的预测。

- 生成混淆矩阵

```
[[65 3]  
[8 24]]
```

- 可视化

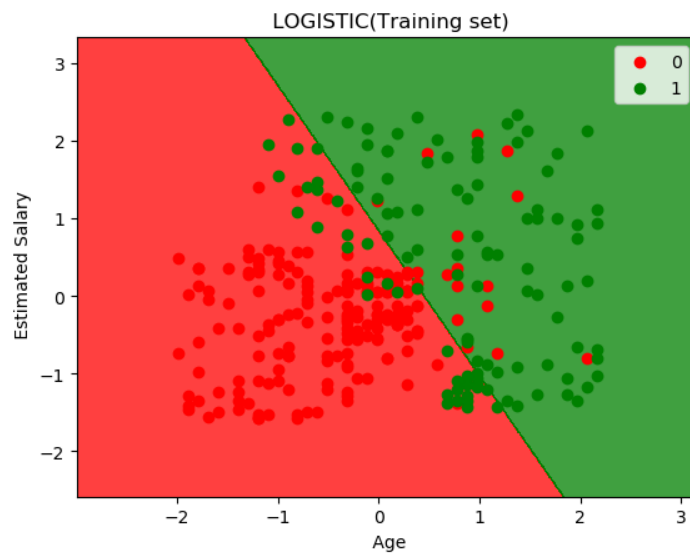


图 2: 训练集

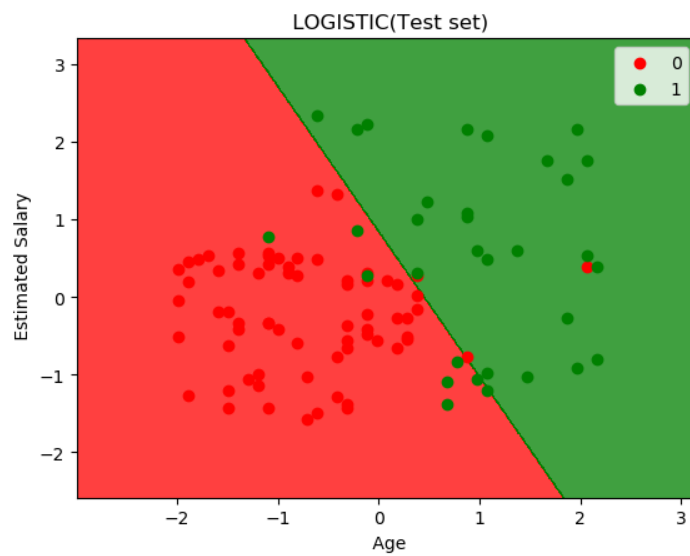


图 3: 测试集