

课后作业：广义线性模型

作者：欧新宇 (Xinyu OU)

【作业提交】

将测试结果Excel电子表格形式进行提交，同时提交源代码。

1. 测试结果命名为: ex04-结果-你的学号-你的姓名.xls (.xlsx)
2. 源代码命名为: ex04-01-你的学号-你的姓名.py

结果文件，要求每小题标注题号，两题之间要求空一行

分别使用线性回归、岭回归及套索回归对"波士顿房价"数据集进行回归分析，要求：

1. 在进行数据集划分的时候，训练集占70%，测试集占30%
2. 分别测试岭回归 $\alpha = [0, 0.1, 2, 5]$ 时的准确率
3. 分别测试套索回归 $\alpha = [0.001, 0.2, 1]$ 时的准确率
4. 分别给出训练集和测试机的得分，要求结果保留4位小数 (ex04-01)
5. 导入相关库

```
1 # 1. 导入基本运算库
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 # 配置参数使 matplotlib 绘图工具可以显示中文
6 plt.rcParams['font.sans-serif'] = [u'Microsoft YaHei']
7
8 # 2. 导入线性回归模型
9 from sklearn.linear_model import LinearRegression
10 from sklearn.linear_model import Ridge
11 from sklearn.linear_model import Lasso
12 from sklearn.model_selection import train_test_split
13
14
15 # 3. 载入糖尿病数据集，并使用 train_test_split() 函数对数据集进行划分
16 from sklearn.datasets import load_boston
17 X = load_boston().data
18 y = load_boston().target
19
20 X_train, X_test, y_train, y_test = train_test_split(X, y, random_state =
21 83, test_size = 0.3)
22
23 # 4. 训练模型
24 lr = LinearRegression().fit(X_train, y_train)
25 ridge0 = Ridge(alpha = 0).fit(X_train, y_train)
26 ridge01 = Ridge(alpha = 0.1).fit(X_train, y_train)
27 ridge2 = Ridge(alpha = 2).fit(X_train, y_train)
28 ridge5 = Ridge(alpha = 5).fit(X_train, y_train)
29 lasso001 = Lasso(alpha = 0.001, max_iter = 100000).fit(X_train, y_train)
30 lasso02 = Lasso(alpha = 0.2, max_iter = 100000).fit(X_train, y_train)
31 lasso1 = Lasso(alpha = 1, max_iter = 100000).fit(X_train, y_train)
```

```

32 # 4. 性能分析
33 print("线性回归: 训练集得分: {0:.4f}, 测试集得分:
    {1:.4f}".format(lr.score(X_train, y_train), lr.score(X_test, y_test)))
34 print("岭回归(alpha=0): 训练集得分: {0:.4f}, 测试集得分:
    {1:.4f}".format(ridge0.score(X_train, y_train), ridge0.score(X_test,
    y_test)))
35 print("岭回归(alpha=0.1): 训练集得分: {0:.4f}, 测试集得分:
    {1:.4f}".format(ridge01.score(X_train, y_train), ridge01.score(X_test,
    y_test)))
36 print("岭回归(alpha=2): 训练集得分: {0:.4f}, 测试集得分:
    {1:.4f}".format(ridge2.score(X_train, y_train), ridge2.score(X_test,
    y_test)))
37 print("岭回归(alpha=5): 训练集得分: {0:.4f}, 测试集得分:
    {1:.4f}".format(ridge5.score(X_train, y_train), ridge5.score(X_test,
    y_test)))
38 print("套索回归(alpha=0.001): 训练集得分: {0:.4f}, 测试集得分:
    {1:.4f}".format(lasso001.score(X_train, y_train), lasso001.score(X_test,
    y_test)))
39 print("套索回归(alpha=0.2): 训练集得分: {0:.4f}, 测试集得分:
    {1:.4f}".format(lasso02.score(X_train, y_train), lasso02.score(X_test,
    y_test)))
40 print("套索回归(alpha=1): 训练集得分: {0:.4f}, 测试集得分:
    {1:.4f}".format(lasso1.score(X_train, y_train), lasso1.score(X_test,
    y_test)))
41
42

```

```

1 线性回归: 训练集得分: 0.7469, 测试集得分: 0.7129
2 岭回归(alpha=0): 训练集得分: 0.7469, 测试集得分: 0.7129
3 岭回归(alpha=0.1): 训练集得分: 0.7468, 测试集得分: 0.7138
4 岭回归(alpha=2): 训练集得分: 0.7404, 测试集得分: 0.7143
5 岭回归(alpha=5): 训练集得分: 0.7362, 测试集得分: 0.7123
6 套索回归(alpha=0.001): 训练集得分: 0.7469, 测试集得分: 0.7131
7 套索回归(alpha=0.2): 训练集得分: 0.7232, 测试集得分: 0.7069
8 套索回归(alpha=1): 训练集得分: 0.6793, 测试集得分: 0.6261

```

```

1 线性回归: 训练集得分: 0.7469, 测试集得分: 0.7129
2 岭回归(alpha=0): 训练集得分: 0.7469, 测试集得分: 0.7129
3 岭回归(alpha=0.1): 训练集得分: 0.7468, 测试集得分: 0.7138
4 岭回归(alpha=2): 训练集得分: 0.7404, 测试集得分: 0.7143
5 岭回归(alpha=5): 训练集得分: 0.7362, 测试集得分: 0.7123
6 套索回归(alpha=0.001): 训练集得分: 0.7469, 测试集得分: 0.7131
7 套索回归(alpha=0.2): 训练集得分: 0.7232, 测试集得分: 0.7069
8 套索回归(alpha=1): 训练集得分: 0.6793, 测试集得分: 0.6261

```