

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ
им. Н.Э. Баумана

Факультет «Информатика и системы управления»
Кафедра «Систем обработки информации и управления»

ОТЧЕТ

Рубежный контроль №__1__
по дисциплине «Методы машинного обучения
»

Тема: «Методы обработки данных»

ИСПОЛНИТЕЛЬ: _____ Ван Тяньшо _____

ФИО

группа ИУ5 И-23М _____

подпись

"28"__05__2024 г.

"__" _____ 2024 г

Москва - 2024

Задание

Необходимо подготовить отчет по рубежному контролю и разместить его в Вашем репозитории. Вы можете использовать титульный лист, или в начале ноутбука в текстовой ячейке указать Ваши Ф.И.О. и группу.

Тема: Методы обработки данных.

Мой вариант: 15+1=16

Задача №16.

Для набора данных проведите нормализацию для одного (произвольного) числового признака с использованием преобразования Бокса-Кокса (Box-Cox transformation).

Код задачи

```
import pandas as pd
from scipy import stats
import matplotlib.pyplot as plt

# 下载数据集并加载到pandas数据框中
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-red.csv"
data = pd.read_csv(url, sep=',')

# 查看数据集的前5行并打印前5行数据
print(data.head())

# 选择一个数值特征进行Box-Cox转换, 例如'fixed_acidity'
feature = 'fixed_acidity'

# 确保所有值都是正数
data[feature] = data[feature] + 1e-6

# 进行Box-Cox转换并填充转换后的值
data[feature + '_boxcox'], fitted_lambda = stats.boxcox(data[feature])

# 绘制转换前后的直方图进行比较
plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 1)
plt.hist(data[feature], bins=30, color='blue', alpha=0.7)
plt.title(f'Original {feature}')

plt.subplot(1, 2, 2)
plt.hist(data[feature + '_boxcox'], bins=30, color='green', alpha=0.7)
plt.title(f'Box-Cox Transformed {feature}')

plt.show()

print(f'Box-Cox transformation lambda: {fitted_lambda}')
```

Рисунок 1

Результат выполнения задачи:

⇒	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	\
0	7.4	0.70	0.00	1.9	0.076	
1	7.8	0.88	0.00	2.6	0.098	
2	7.8	0.76	0.04	2.3	0.092	
3	11.2	0.28	0.56	1.9	0.075	
4	7.4	0.70	0.00	1.9	0.076	

	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	\
0	11.0	34.0	0.9978	3.51	0.56	
1	25.0	67.0	0.9968	3.20	0.68	
2	15.0	54.0	0.9970	3.26	0.65	
3	17.0	60.0	0.9980	3.16	0.58	
4	11.0	34.0	0.9978	3.51	0.56	

	alcohol	quality
0	9.4	5
1	9.8	5
2	9.8	5
3	9.8	6
4	9.4	5

Рисунок 2

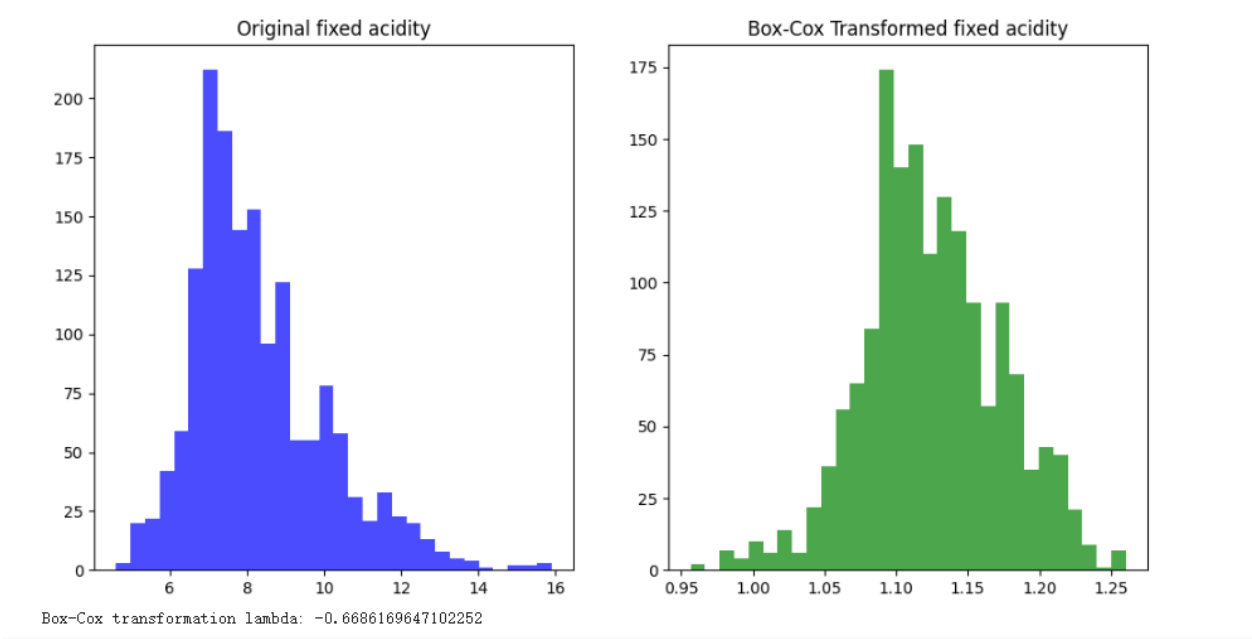


Рисунок 3