

Министерство цифрового развития, связи и массовых коммуникаций
Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Сибирский государственный университет телекоммуникаций и
информатики» (СибГУТИ)

Отчёт по лабораторной работе №3

Регрессионный анализ

Выполнил:

студент гр. ИП-111

Кузьменок Д.В.

Проверил:

Старший преподаватель кафедры ПМиК

Дементьева К.И.

Новосибирск, 2024 г.

Задание

Данные необходимо рассматривать как три набора. Данные для красного вина, данные для белого, общие данные вне зависимости от цвета. Необходимо построить модель для каждого из наборов, обучить её и сравнить полученные при помощи модели результаты с известными. Для обучения использовать 70% выборки, для тестирования 30%. Разбивать необходимо случайным образом, а, следовательно, для корректности тестирования качества модели, эксперимент необходимо провести не менее 10 раз и вычислить среднее значение качества регрессии.

Вариант задания: использовать классическую модель `LinearRegression`.

Имя файла: `winequalityN.csv`.

Результаты

Для проверки полученных результатов я запускаю 30 раз подсчет точности для вычисления средней точности угадывания класса.

В случае всех вин:

Все вина:

1)	Точность:	82.667%
2)	Точность:	83.949%
3)	Точность:	82.154%
4)	Точность:	82.821%
5)	Точность:	83.436%
6)	Точность:	82.974%
7)	Точность:	83.436%
8)	Точность:	82.821%
9)	Точность:	83.333%
10)	Точность:	83.487%
11)	Точность:	83.282%
12)	Точность:	83.026%
13)	Точность:	82.462%
14)	Точность:	82.718%
15)	Точность:	83.436%
16)	Точность:	83.026%
17)	Точность:	84.000%
18)	Точность:	83.077%
19)	Точность:	83.282%
20)	Точность:	82.667%
21)	Точность:	82.051%
22)	Точность:	82.821%
23)	Точность:	82.205%
24)	Точность:	83.333%
25)	Точность:	82.462%
26)	Точность:	82.718%
27)	Точность:	83.282%
28)	Точность:	82.821%
29)	Точность:	82.667%
30)	Точность:	83.179%

Средняя точность: 82.986% за 30 проходов

В случае белых вин:

Белые вина:

- 1) Точность: 84.694%
- 2) Точность: 84.762%
- 3) Точность: 85.782%
- 4) Точность: 85.170%
- 5) Точность: 84.966%
- 6) Точность: 85.374%
- 7) Точность: 84.354%
- 8) Точность: 84.830%
- 9) Точность: 83.129%
- 10) Точность: 85.442%
- 11) Точность: 84.762%
- 12) Точность: 85.034%
- 13) Точность: 85.238%
- 14) Точность: 83.197%
- 15) Точность: 83.673%
- 16) Точность: 85.170%
- 17) Точность: 84.354%
- 18) Точность: 84.286%
- 19) Точность: 84.898%
- 20) Точность: 84.626%
- 21) Точность: 84.626%
- 22) Точность: 84.286%
- 23) Точность: 84.694%
- 24) Точность: 84.898%
- 25) Точность: 84.898%
- 26) Точность: 84.490%
- 27) Точность: 85.034%
- 28) Точность: 85.102%
- 29) Точность: 84.762%
- 30) Точность: 84.490%

Средняя точность: 84.701% за 30 проходов

В случае красных вин:

Красные вина:

- 1) Точность: 87.917%
- 2) Точность: 87.292%
- 3) Точность: 86.667%
- 4) Точность: 88.333%
- 5) Точность: 86.875%
- 6) Точность: 90.208%
- 7) Точность: 87.917%
- 8) Точность: 88.958%
- 9) Точность: 88.958%
- 10) Точность: 90.417%
- 11) Точность: 88.125%
- 12) Точность: 90.625%
- 13) Точность: 88.125%
- 14) Точность: 87.708%
- 15) Точность: 89.375%
- 16) Точность: 86.667%
- 17) Точность: 88.958%
- 18) Точность: 88.125%
- 19) Точность: 87.292%
- 20) Точность: 89.375%
- 21) Точность: 88.125%
- 22) Точность: 87.500%
- 23) Точность: 88.125%
- 24) Точность: 88.542%
- 25) Точность: 90.417%
- 26) Точность: 87.292%
- 27) Точность: 88.542%
- 28) Точность: 88.750%
- 29) Точность: 89.375%
- 30) Точность: 90.833%

Средняя точность: 88.514% за 30 проходов

Код программы

```
import pandas as pd
from sklearn import preprocessing
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split

def preprocess_data(data):
    data_X = preprocessing.normalize(data[:, :-1])
    data_Y = data[:, -1]
    train_x, test_x, train_y, test_y = train_test_split(data_X, data_Y,
test_size=0.3, stratify=data_Y)
    return train_x, test_x, train_y, test_y

def train_linear_regression(train_x, train_y):
    linear_regression = LinearRegression()
    linear_regression.fit(train_x, train_y)
    return linear_regression

def evaluate_model(model, test_x, test_y):
    predicted = model.predict(test_x)
    success = 0
    for i in range(len(test_x)):
        if abs(test_y[i] - predicted[i]) < 1:
            success += 1
    return success / len(test_x) * 100

def run_evaluation(data, n_iterations):
    total_accuracy = 0
    for i in range(n_iterations):
        train_x, test_x, train_y, test_y = preprocess_data(data)
        model = train_linear_regression(train_x, train_y)
        accuracy = evaluate_model(model, test_x, test_y)
        print(f"{i + 1}) Точность: {accuracy :.3f}%")
        total_accuracy += accuracy
    return total_accuracy / n_iterations

def main():
    file_name = "winequalityN.csv"
    n_iterations = 30

    data = pd.read_csv(file_name, header=0).fillna(0)
    data.loc[data.type == 'white', 'type'] = 0
    data.loc[data.type == 'red', 'type'] = 1
    data = data.to_numpy()

    print(f'\nВсе вина:')
    average_accuracy = run_evaluation(data, n_iterations)
    print(f'Средняя точность: {average_accuracy:.3f}% за {n_iterations}
проходов\n\n')
```

```
print(f'Белые вина:')
white_data = data[data[:, 0] == 0]
average_accuracy = run_evaluation(white_data, n_iterations)
print(f'Средняя точность: {average_accuracy:.3f}% за {n_iterations}
проходов\n\n')

print(f'Красные вина:')
red_data = data[data[:, 0] == 1]
average_accuracy = run_evaluation(red_data, n_iterations)
print(f'Средняя точность: {average_accuracy:.3f}% за {n_iterations}
проходов')

if __name__ == "__main__":
    main()
```