

Министерство образования Республики Беларусь  
Учреждение образования  
«Брестский государственный технический университет»  
Кафедра ИИТ

Лабораторная работа №2  
По дисциплине: «ОМО»  
Тема:» Линейные модели для задач регрессии и классификации»

Выполнил:  
Студент 3-го курса  
Группы АС-66  
Пекун М.С.  
Проверил:  
Крощенко А.А.

Брест 2025

Цель: Изучить применение линейной и логистической регрессии для решения практических задач. Научиться обучать модели, оценивать их качество с помощью соответствующих метрик и интерпретировать результаты.

#### Вариант 8

##### Регрессия (Прогнозирование качества вина)

1. Датасет: *Wine Quality*
2. Цель: предсказать оценку качества вина (quality) как непрерывную величину.
3. Задания:
  - загрузить данные;
  - обучить модель линейной регрессии на всех доступных признаках;
  - рассчитать MSE и  $R^2$ ;
  - визуализировать зависимость quality от alcohol с линией регрессии.

##### Классификация (Определение "хорошего" вина)

1. Датасет: *Wine Quality*
2. Цель: классифицировать вино как
  - “хорошее” — если  $quality \geq 7$ ,
  - “плохое” — если  $quality < 7$ .
3. Задания:
  - создать новый бинарный целевой столбец на основе столбца quality;
  - обучить модель логистической регрессии;
  - рассчитать Accuracy, Precision и Recall для класса “хорошее”;
  - построить матрицу ошибок.

Код:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression,
LogisticRegression
from sklearn.metrics import mean_squared_error, r2_score,
accuracy_score, precision_score, recall_score, confusion_matrix,
ConfusionMatrixDisplay

# --- 1. РЕГРЕССИЯ: прогноз качества вина ---
# Используем датасет Wine Quality (красное вино)
url = "https://archive.ics.uci.edu/ml/machine-learning-
databases/wine-quality/winequality-red.csv"
data = pd.read_csv(url, sep=';')

print("Первые строки данных:")
print(data.head())

# Разделение признаков и целевой переменной
```

```

X = data.drop(columns=['quality'])
y = data['quality']

# Разделение на обучающую и тестовую выборки
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Обучение модели линейной регрессии
lin_reg = LinearRegression()
lin_reg.fit(X_train, y_train)

# Предсказание
y_pred = lin_reg.predict(X_test)

# Метрики
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("\n=== РЕГРЕССИЯ ===")
print(f"MSE: {mse:.4f}")
print(f"R²: {r2:.4f}")

# Визуализация зависимости quality от alcohol
plt.figure(figsize=(6, 4))
plt.scatter(data['alcohol'], data['quality'], alpha=0.5,
label='данные')

# Линия регрессии только по признаку alcohol
x_line = np.linspace(data['alcohol'].min(),
data['alcohol'].max(), 100)
y_line = lin_reg.intercept_ +
lin_reg.coef_[X.columns.get_loc('alcohol')] * x_line

plt.plot(x_line, y_line, color='red', label='линия регрессии')
plt.xlabel('Alcohol')
plt.ylabel('Quality')
plt.legend()
plt.title('Зависимость качества вина от содержания алкоголя')
plt.show()

# --- 2. КЛАССИФИКАЦИЯ: определение "хорошего" вина ---
data['good'] = (data['quality'] >= 7).astype(int)

X = data.drop(columns=['quality', 'good'])

```

```
y = data['good']

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

log_reg = LogisticRegression(max_iter=1000, solver='liblinear')
log_reg.fit(X_train, y_train)

y_pred = log_reg.predict(X_test)

# Метрики
acc = accuracy_score(y_test, y_pred)
prec = precision_score(y_test, y_pred)
rec = recall_score(y_test, y_pred)

print("\n=== КЛАССИФИКАЦИЯ ===")
print(f"Accuracy: {acc:.4f}")
print(f"Precision: {prec:.4f}")
print(f"Recall: {rec:.4f}")

# Матрица ошибок
cm = confusion_matrix(y_test, y_pred)
disp = ConfusionMatrixDisplay(confusion_matrix=cm,
display_labels=['Плохое', 'Хорошее'])
disp.plot(cmap='Blues')
plt.title('Матрица ошибок: классификация качества вина')
plt.show()
```

График регрессия:

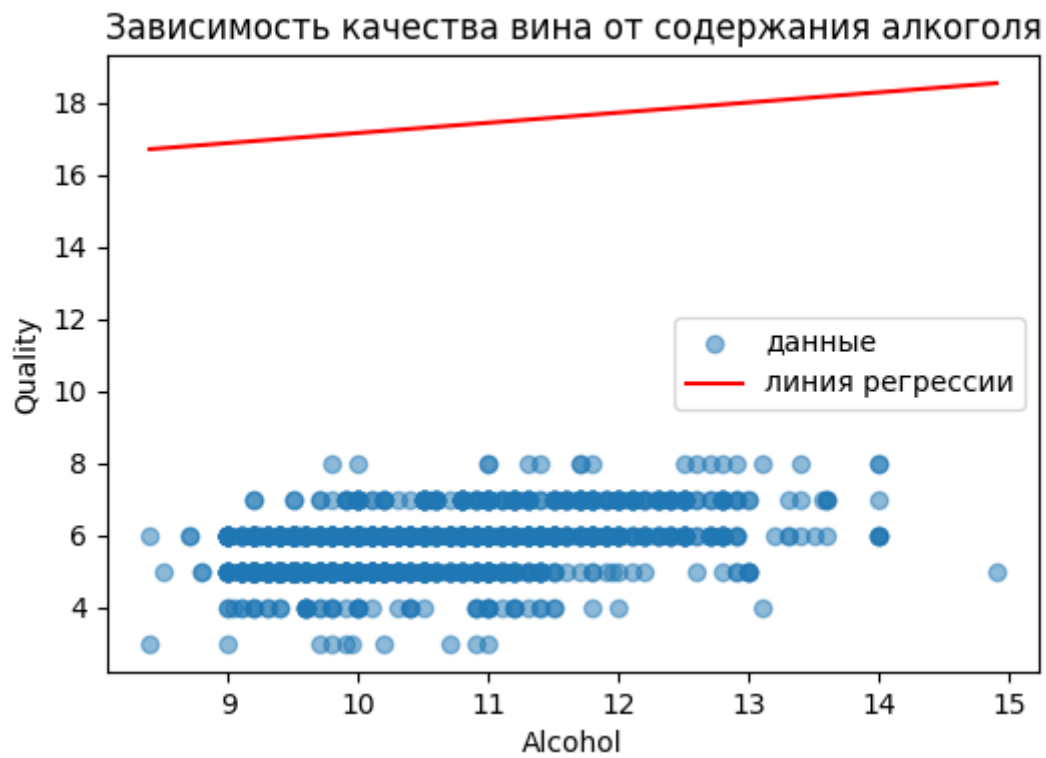


График классификация:



## Результаты:

Первые строки данных:

	fixed acidity	volatile acidity	citric acid	...	sulphates	alcohol	quality
0	7.4	0.70	0.00	...	0.56	9.4	5
1	7.8	0.88	0.00	...	0.68	9.8	5
2	7.8	0.76	0.04	...	0.65	9.8	5
3	11.2	0.28	0.56	...	0.58	9.8	6
4	7.4	0.70	0.00	...	0.56	9.4	5

[5 rows x 12 columns]

=== РЕГРЕССИЯ ===

MSE: 0.3900

R<sup>2</sup>: 0.4032

=== КЛАССИФИКАЦИЯ ===

Accuracy: 0.8656

Precision: 0.6250

Recall: 0.2128

Process finished with exit code 0

**Вывод:** в результате выполнения данной лабораторной работы изучили применение линейной и логистической регрессии для решения практических задач. Научились обучать модели, оценивать их качество с помощью соответствующих метрик и интерпретировать результаты.