



Министерство науки и высшего образования Российской Федерации
Калужский филиал
федерального государственного бюджетного
образовательного учреждения высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(КФ МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИУК «Информатика и управление»

КАФЕДРА ИУК4 «Программное обеспечение ЭВМ, информационные технологии»

ЛАБОРАТОРНАЯ РАБОТА №7

«Наивный байесовский классификатор. Виды классификаторов»

ДИСЦИПЛИНА: «Технологии анализа данных»

Выполнил: студент гр. ИУК4-82Б _____ (Карельский М.К.)
(Подпись)

Проверил: _____ (Ерохин И.И.)
(Подпись)

Дата сдачи (защиты):

Результаты сдачи (защиты):

- Балльная оценка:

- Оценка:

Калуга, 2024

Цель: формирование практических навыков проведения классификации с использованием наивного Байесовского классификатора.

Задачи:

1. Составить модель на основе наивного Байесовского классификатора.
2. Изучить виды наивного Байесовского классификатора

Задание:

Для всех вариантов необходимо провести классификацию с помощью наивного Байесовского классификатора. Составить модель для классификации. Сравнить данные, полученные моделью, с реальными данными. Данные для вариантов необходимо брать из набора `sklearn.datasets`

Вариант 5

Используя дополненный (`complement`) наивный Байес, построить модель (на обучающей выборке) и произвести классификацию (на тестовой выборке) цветков ириса. Визуализировать полученную классификацию. Построить матрицу несоответствий и интерпретировать результаты. Оценить точность модели (для оценки точности необходимо взять среднее арифметическое оценки 10 испытаний).

Листинг:

```
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import ComplementNB
from sklearn.metrics import confusion_matrix, accuracy_score

iris = load_iris()
X = iris.data
y = iris.target

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

model = ComplementNB()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

conf_matrix = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:")
print(conf_matrix)
```

```
plt.figure(figsize=(8, 6))
plt.scatter(X_test[:, 0], X_test[:, 1], c=y_pred, cmap='viridis', s=50,
alpha=0.8)
plt.xlabel('Sepal Length (cm)')
plt.ylabel('Sepal Width (cm)')
plt.title('Classification of Iris Flowers')
plt.colorbar(label='Predicted Class')
plt.show()
```

Результат:

```
Accuracy: 0.7
Confusion Matrix:
[[10  0  0]
 [ 0  0  9]
 [ 0  0 11]]
```

Рис. 1. Оценка точности и матрица несоответствий

Из матрицы видно, что модель имеет трудности с классификацией образцов второго класса и чаще ошибочно относит их к третьему классу. Модель хорошо справляется с классификацией образцов первого и третьего классов.

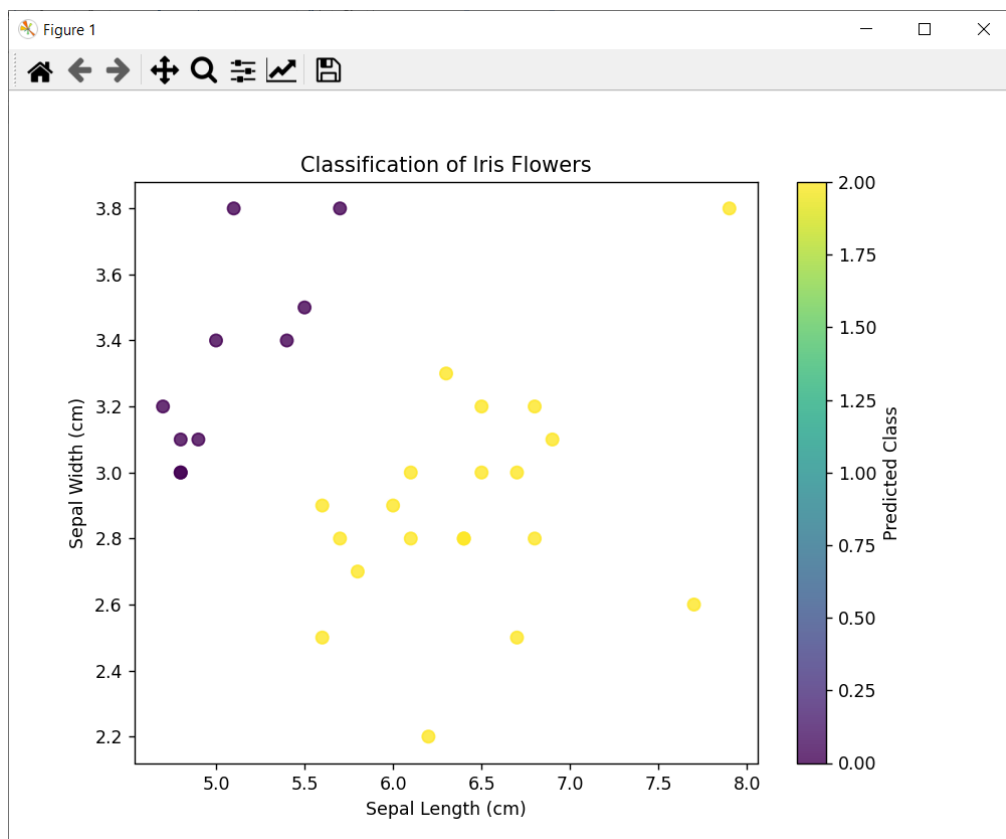


Рис. 2. Классификация

Вывод: в ходе выполнения лабораторной работы были получены практические навыки проведения классификации с использованием наивного Байесовского классификатора.