**Липецкий государственный технический университет**

Факультет автоматизации и информатики

Кафедра автоматизированных систем управления

ЛАБОРАТОРНАЯ РАБОТА №2

по дисциплине «Прикладные интеллектуальные системы и экспертные

системы»

Бинарная классификация

Студент Крутских А.Ю.

Группа М-ИАП-22

Руководитель Кургасов В.В.

Липецк 2022 г.

Цель работы

Получить практические навыки решения задачи бинарной классификации данных в среде Jupiter Notebook. Научиться загружать данные, обучать классификаторы и проводить классификацию. Научиться оценивать точность полученных моделей.

Задание кафедры

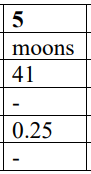
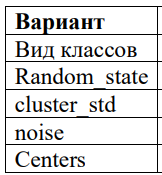
1. В среде Jupiter Notebook создать новый ноутбук (Notebook);
2. Импортировать необходимые для работы библиотеки и модули;
3. Загрузить данные в соответствие с вариантом;
4. Вывести первые 15 элементов выборки (координаты точек и метки класса);
5. Отобразить на графике сгенерированную выборку. Объекты разных классов должны иметь разные цвета;
6. Разбить данные на обучающую (train) и тестовую (test) выборки в пропорции 75% - 25% соответственно;
7. Отобразить на графике обучающую и тестовую выборки. Объекты разных классов должны иметь разные цвета;
8. Реализовать модели классификаторов, обучить их на обучающем множестве. Применить модели на тестовой выборке, вывести результаты классификации:

* Истинные и предсказанные метки классов
* Матрицу ошибок (confusion matrix)
* Значения полноты, точности, f1-меры и аккуратности
* Значение площади под кривой ошибок (AUC ROC)
* Отобразить на графике область принятия решений по каждому классу В качестве методов классификации использовать:

1. Метод к-ближайших соседей (n\_neighbors = {1, 3, 5, 9})
2. Наивный байесовский метод
3. Случайный лес (n\_estimators = {5, 10, 15, 20, 50})
4. По каждому пункту работы занести в отчет программный код и результат вывода;
5. По результатам п.8 занести в отчет таблицу с результатами классификации всеми методами и выводы о наиболее подходящем методе классификации ваших данных.

Ход работы

Вариант 5.



Для всех вариантов, использующих для генерации make\_classification, дополнительные параметры: n\_features=2, n\_redundant=0, n\_informative=1, n\_clusters\_per\_class=1.

1. в среде Jupiter Notebook создать новый ноутбук (Notebook);
2. импортировать необходимые для работы библиотеки и модули;
3. загрузить данные в соответствие с вариантом;
4. вывести первые 15 элементов выборки (координаты точек и метки класса);
5. отобразить на графике сгенерированную выборку. Объекты разных классов должны иметь разные цвета;
6. разбить данные на обучающую (train) и тестовую (test) выборки в пропорции 75% - 25% соответственно;
7. отобразить на графике обучающую и тестовую выборки. Объекты разных классов должны иметь разные цвета;
8. реализовать модели классификаторов, обучить их на обучающем множестве. Применить модели на тестовой выборке, вывести результаты классификации:

* Истинные и предсказанные метки классов
* Матрицу ошибок (confusion matrix)
* Значения полноты, точности, f1-меры и аккуратности
* Значение площади под кривой ошибок (AUC ROC)
* Отобразить на графике область принятия решений по каждому классу В качестве методов классификации использовать:

1. Метод к-ближайших соседей (n\_neighbors = {1, 3, 5, 9})
2. Наивный байесовский метод
3. Случайный лес (n\_estimators = {5, 10, 15, 20, 50})
4. по каждому пункту работы занести в отчет программный код и результат вывода;
5. по результатам п.8 занести в отчет таблицу с результатами классификации всеми методами и выводы о наиболее подходящем методе классификации ваших данных.

Ход работы

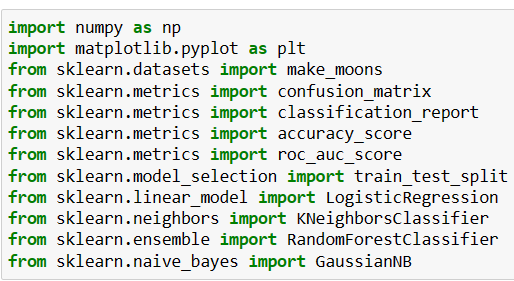
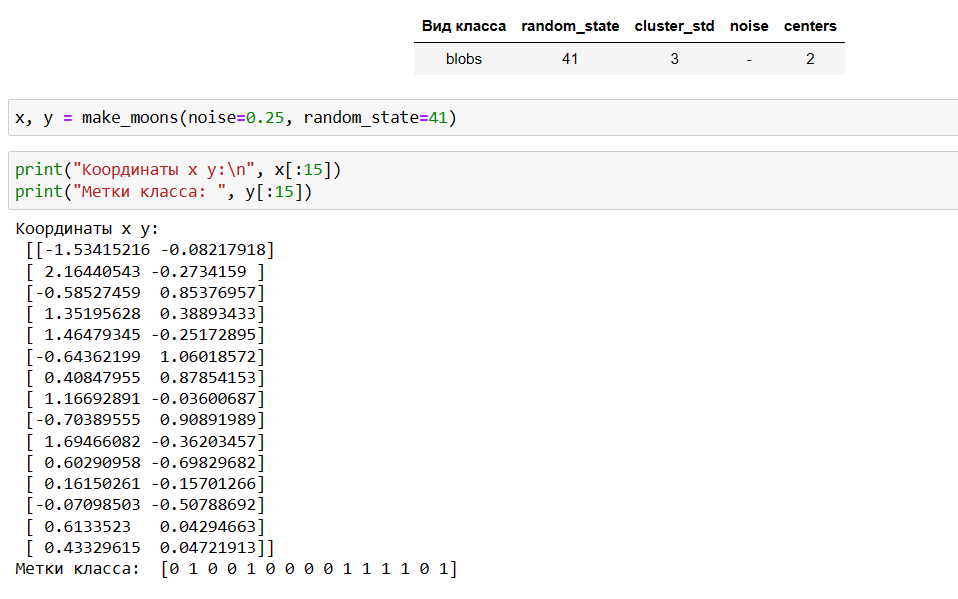


Рисунок 1 - Импортируем библиотеки

  
Рисунок 2 – Генерируем выборку

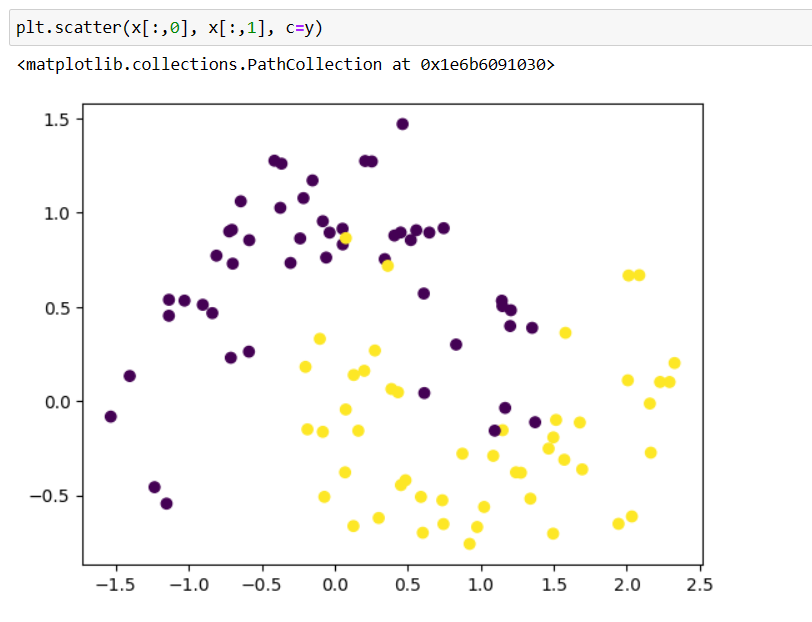


Рисунок 3 – График сгенерированной выборки

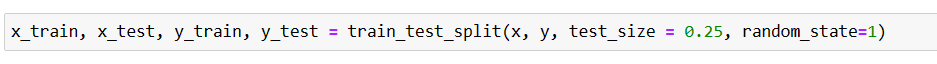


Рисунок 4 - Разбиваем данные на обучающую (train) и тестовую (test) выборки в пропорции 90% - 10% соответственно

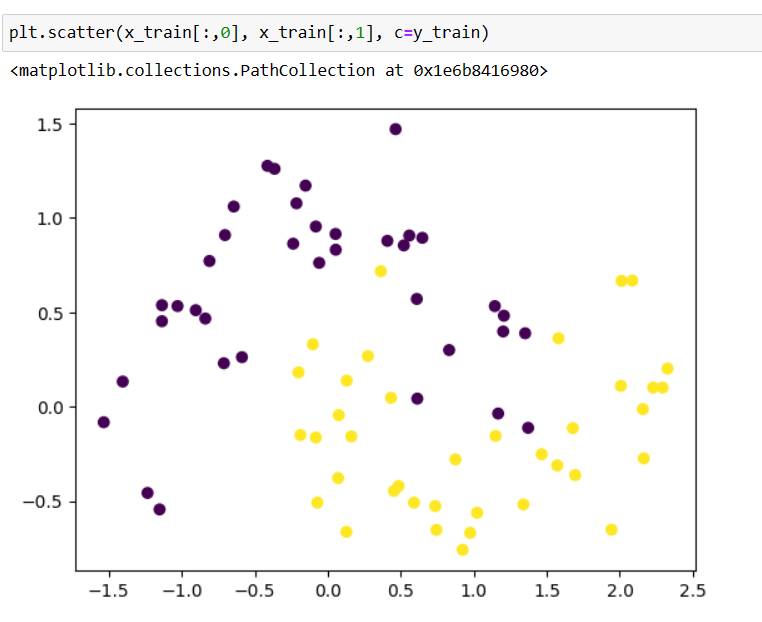
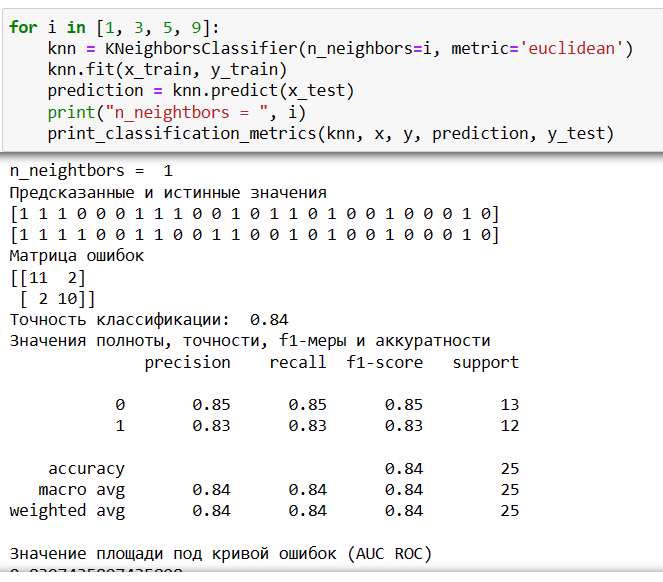
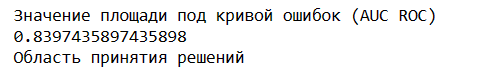


Рисунок 5 - График обучающей выборки



Рисунок 6 - График тестовой выборки





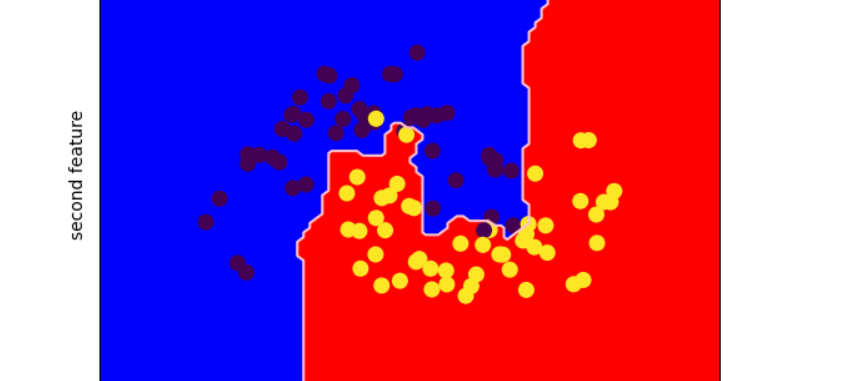
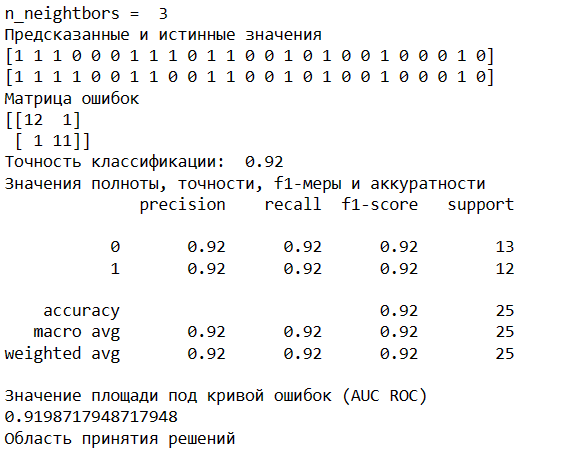


Рисунок 7 – Метод k-ближайших соседей (n=1)



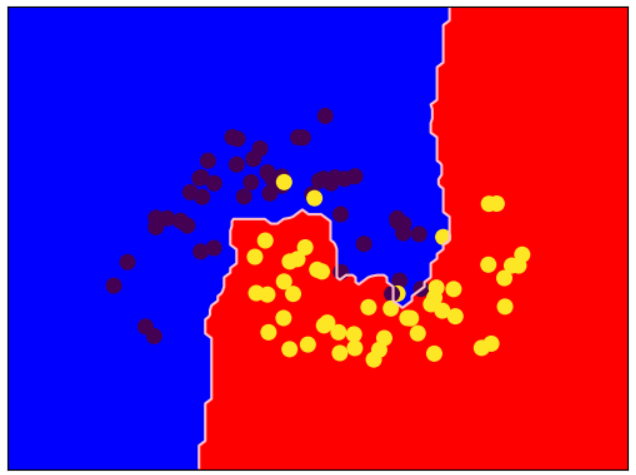
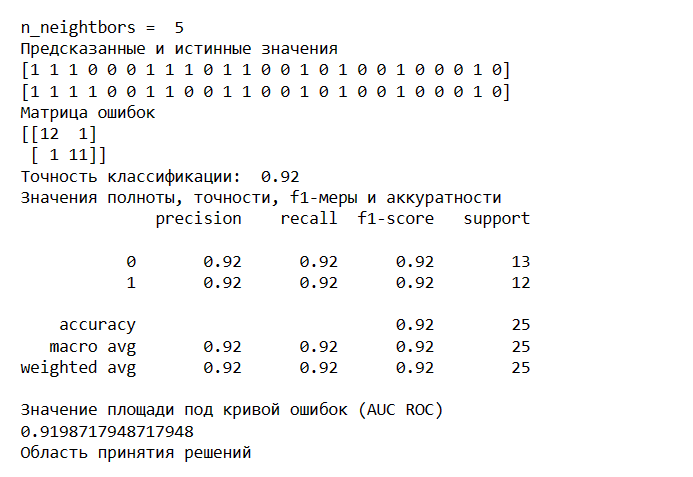


Рисунок 8 – Метод k-ближайших соседей (n=3)



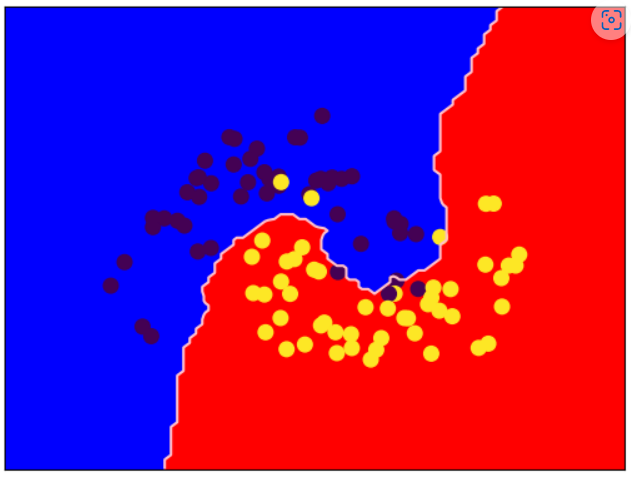
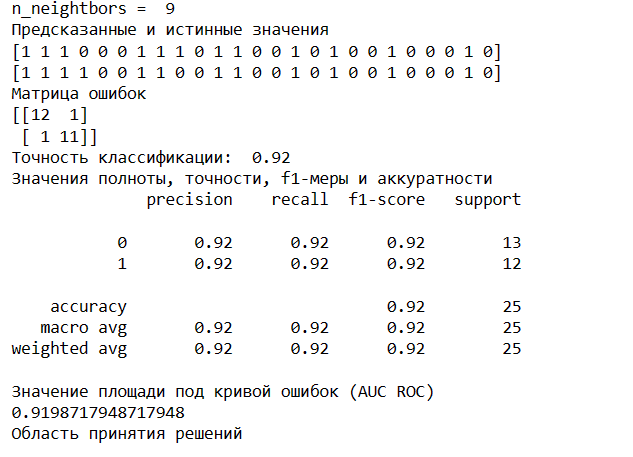


Рисунок 9 – Метод k-ближайших соседей (n=5)



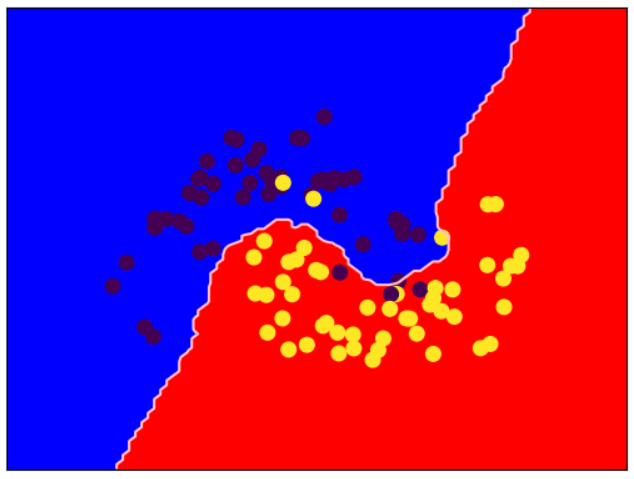
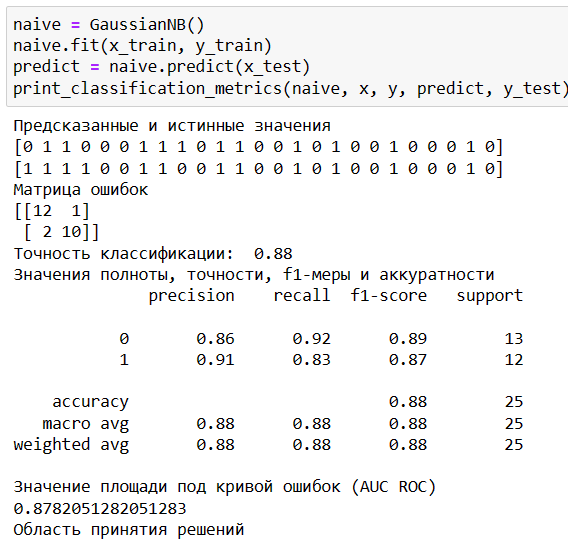


Рисунок 10 – Метод k-ближайших соседей (n=9)



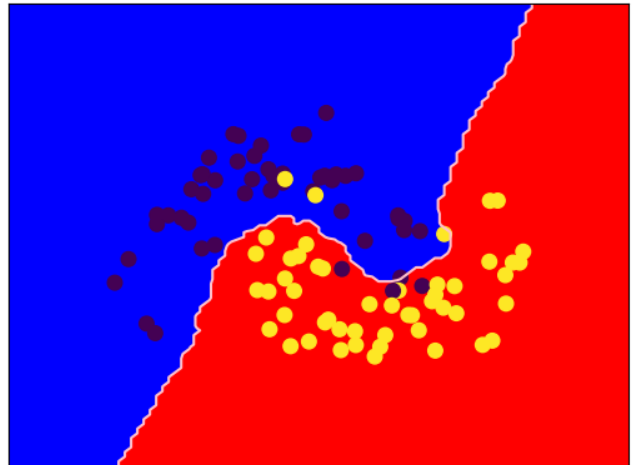


Рисунок 8 – Метод

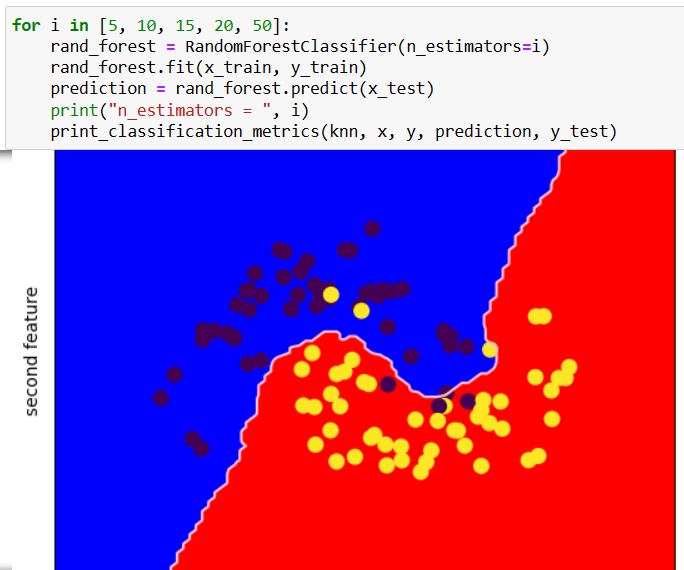


Рисунок 9 – Случайный лес

Вывод

В ходе выполнения данной лабораторной работы мы получили базовые навыки работы с языком python и набором функций для анализа и обработки данных. Получили практические навыки решения задачи бинарной классификации данных в среде Jupiter Notebook. Научились загружать данные, обучать классификаторы и проводить классификацию. Научились оценивать точность полученных моделей.