Липецкий государственный технический университет

Факультет автоматизации и информатики

Кафедра автоматизированных систем управления

Лабораторная работа № 1

# По дисциплине: «Прикладные интеллектуальные системы и

# экспертные системы»

Бинарная классификация фактографических данных

Студент Коретников Н.И.

Группа М-ИАП-23-1

Руководитель Кургасов В.В.

к.т.н. доцент

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:

a) Метод к-ближайших соседей (n\_neighbors = {1, 3, 5, 9})

b) Наивный байесовский метод

c) Случайный лес (n\_estimators = {5, 10, 15, 20, 50})

9) По каждому пункту работы занести в отчет программный код и результат вывода.

10) По результатам п.8 занести в отчет таблицу с результатами классификации всеми методами и выводы о наиболее подходящем методе классификации ваших данных.

11) Изучить, как изменится качество классификации, если на тестовую часть выделить 10% выборки, 35% выборки. Для этого повторить п.п. 6 – 10.

Вариант 6: Вид классов: moons; Random\_state: 23; noise: 0.3.

Ход работы:

Импорт библиотек и прописываем исходные данные лабораторной работы по 6 варианту. Весь процесс работы лабораторной работы иллюстрируется на рисунках.

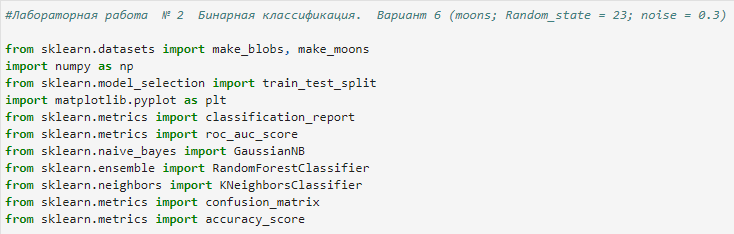


Рисунок 1 – Импорт библиотек

Импортируем в код готовую функцию принятия решений.

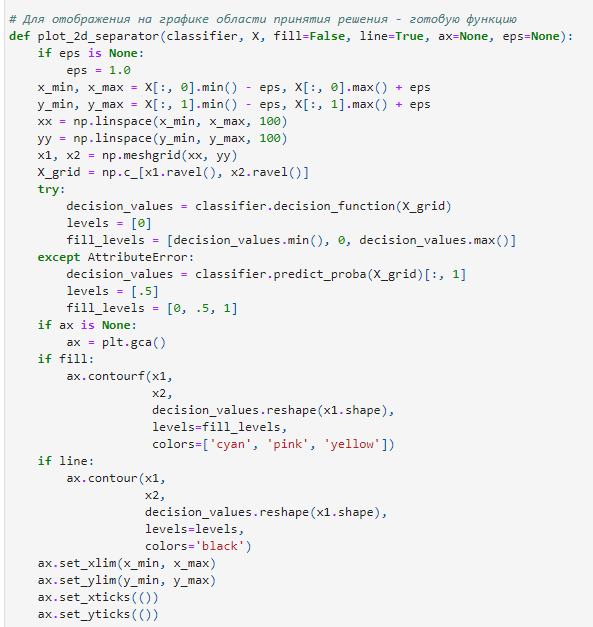


Рисунок 2 – Вставка в код готовой функции

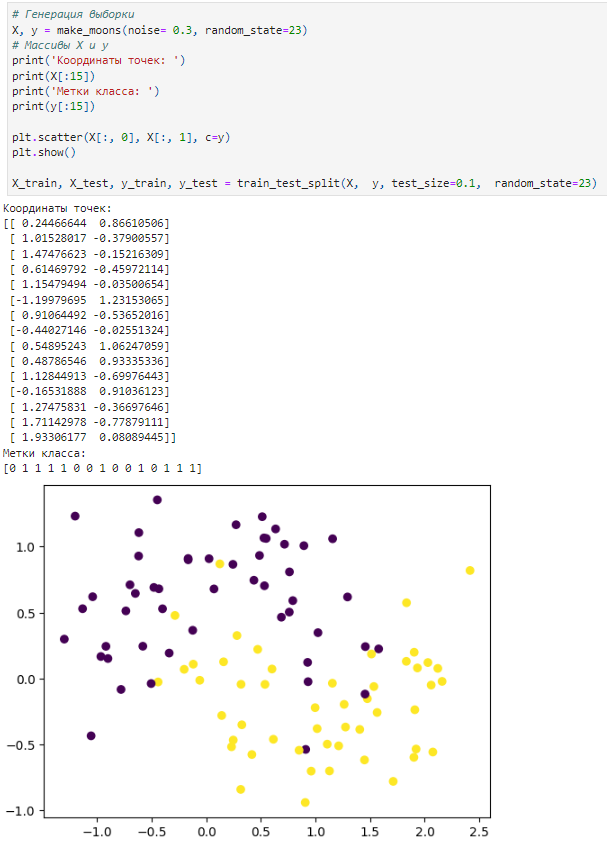


Рисунок 3 – Генерация выборки 90% обучающегося и 10% тестового множества

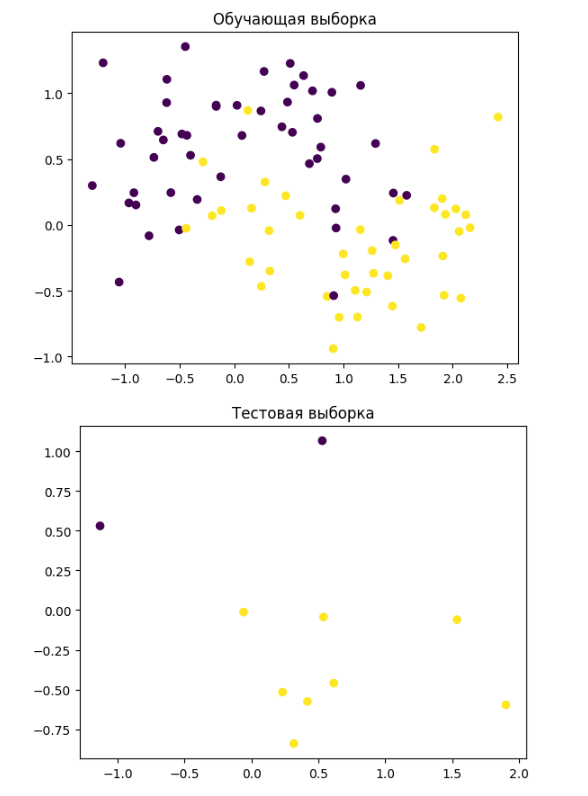


Рисунок 4 – Выборки 90% обучающегося и 10% тестового множества изображены отдельно

Создание моделей ближайших соседей (1,3,5,9).

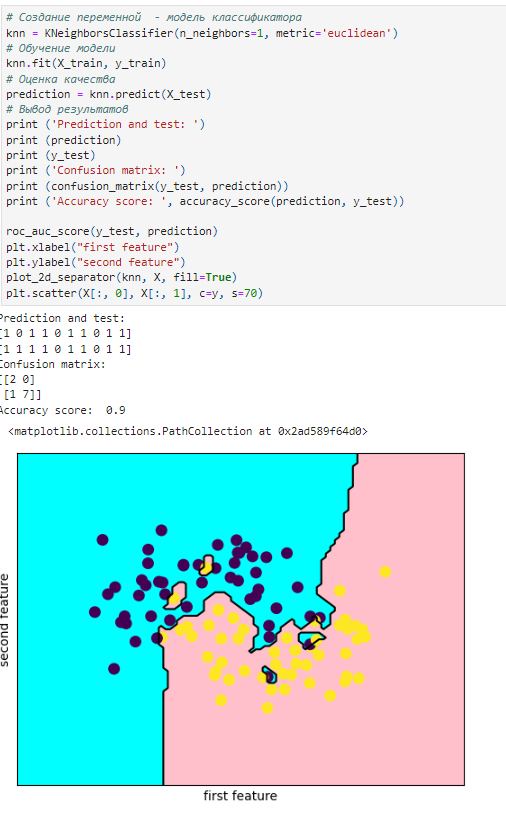


Рисунок 5 –Метод ближайших соседей (1)

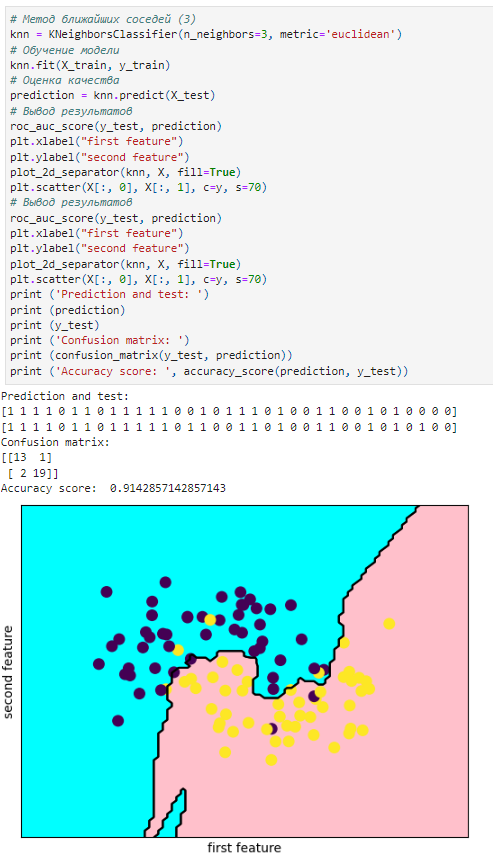


Рисунок 6 –Метод ближайших соседей (3)

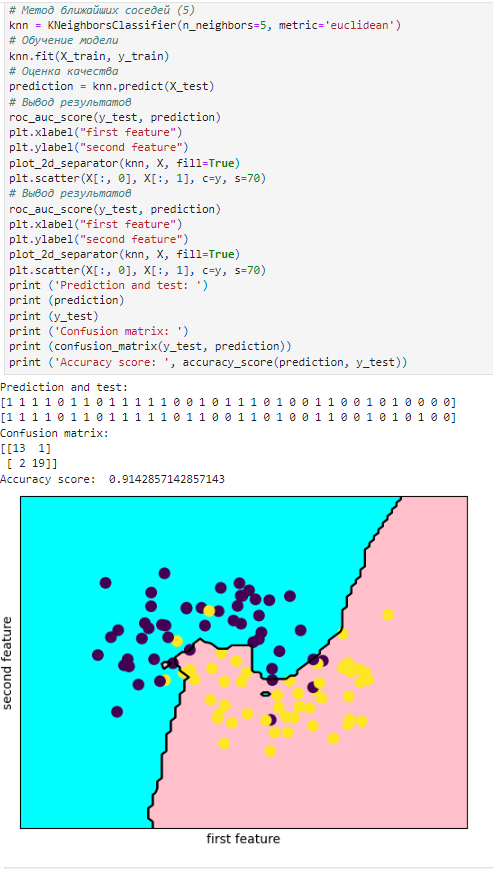


Рисунок 7 –Метод ближайших соседей (5)

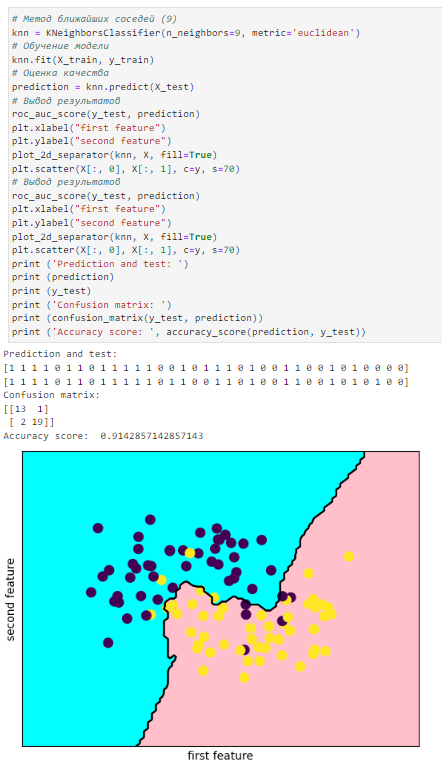


Рисунок 8 –Метод ближайших соседей (9)

Создаем новую модель, методом Наивного Байесовского классификатора.

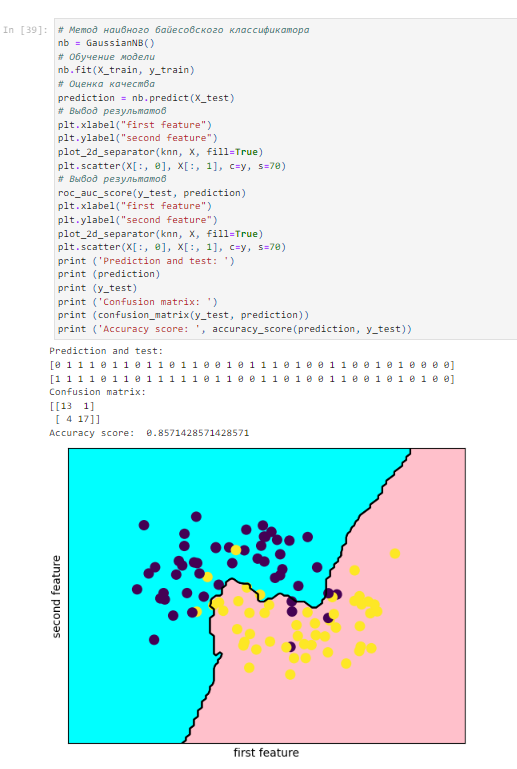


Рисунок 9 – Метод Наивного Байесовского классификатора

Создание новой модели методом случайных лесов (5,10,15,20,50).

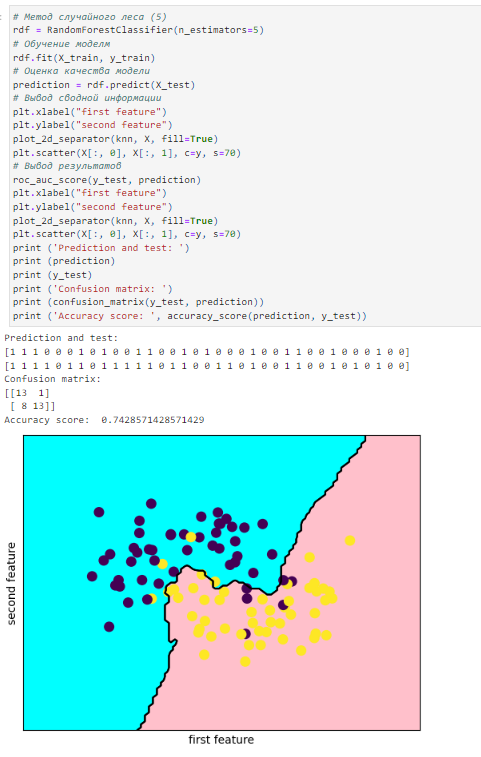


Рисунок 10 – Метод случайных лесов (5)

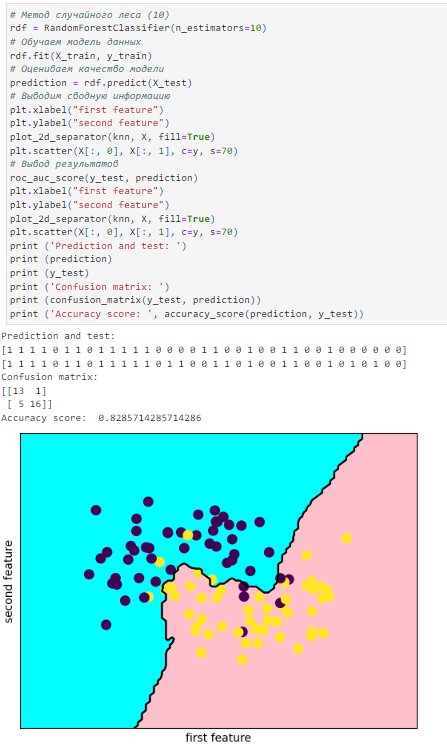


Рисунок 11 – Метод случайных лесов (10)

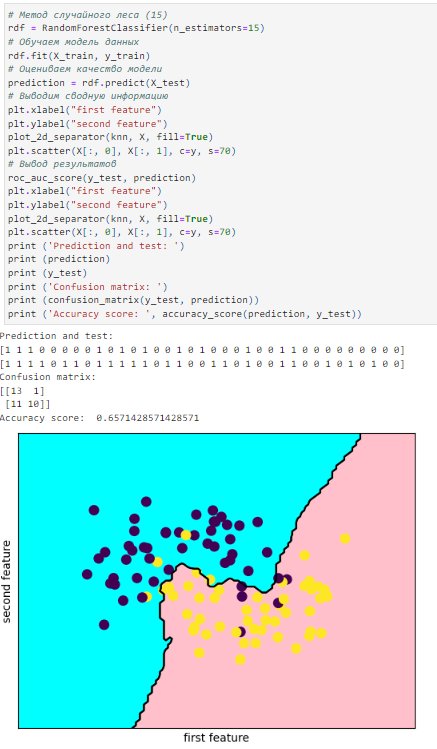


Рисунок 12 – Метод случайных лесов (15)

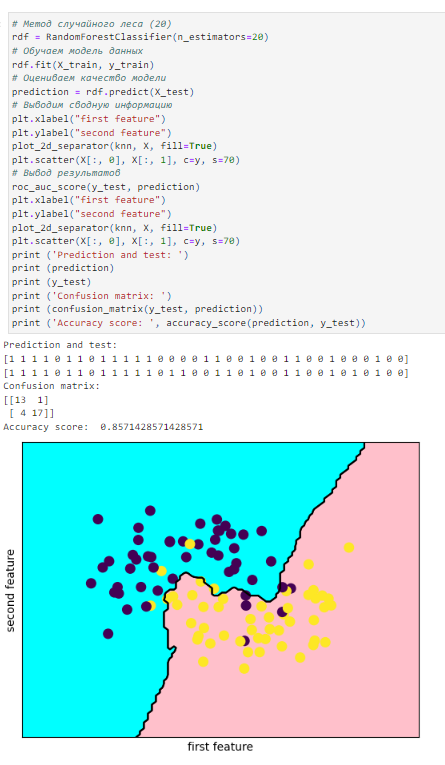


Рисунок 13 – Метод случайных лесов (20)

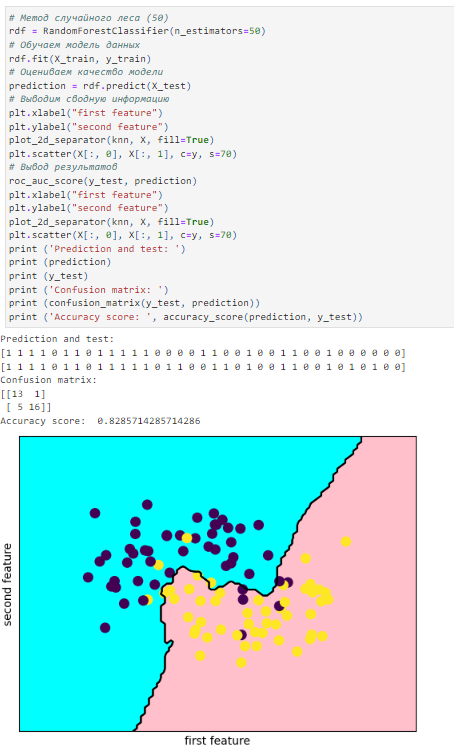


Рисунок 14 – Метод случайных лесов (50)

Если разобьём тестовую и обучающуюся выборки на 35% и 65%, то получим такой результат.

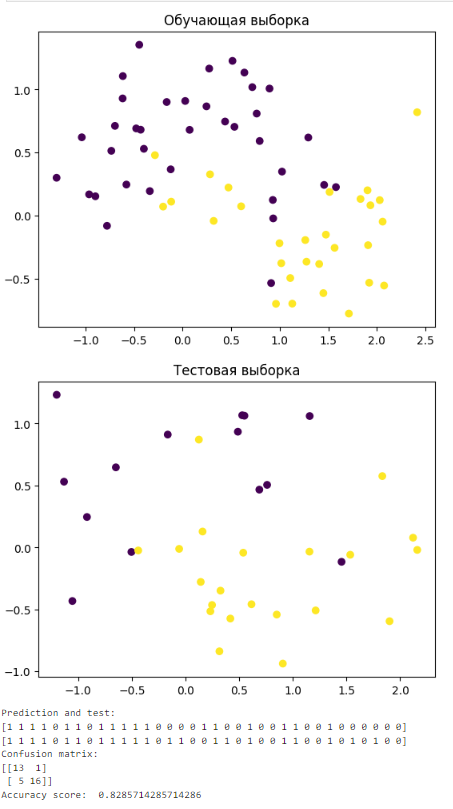


Рисунок 15 – Выборка тестовая 35% и обучающая 65%

Для каждой выборки сделаем таблицы с применением различных моделей.

Таблица 1. Результаты точности при тестовом множестве 10%

|  |  |
| --- | --- |
| Модели, методы  (параметры) | Результат |
| Метод k-ближайших  соседей (1) | Точность: 0,9  Площадь по кривой: 0,9 |
| Метод k-ближайших  соседей (3) | Точность: 0.914  Площадь по кривой:0.914 |
| Метод k-ближайших  соседей (5) | Точность: 0.914  Площадь по кривой:0.914 |
| Метод k-ближайших  соседей (9) | Точность: 0.914  Площадь по кривой: 0.914 |
| Наивный байесовский  классификатор | Точность:0.857  Площадь по кривой:0.857 |
| Случайный лес  (5) | Точность:0.74  Площадь по кривой:0.74 |
| Случайный лес  (10) | Точность:0.82  Площадь по кривой:0.82 |
| Случайный лес  (15) | Точность:0.657  Площадь по кривой:0.657 |
| Случайный лес  (20) | Точность:0.857  Площадь по кривой:0.857 |
| Случайный лес  (50) | Точность:0.828  Площадь по кривой:0.828 |

Таблица 2. Результаты точности при тестовом множестве 35%

|  |  |
| --- | --- |
| Модели, методы  (параметры) | Результат |
| Метод k-ближайших  соседей (1) | Точность: 0,828  Площадь по кривой: 0,828 |
| Метод k-ближайших  соседей (3) | Точность: 0.914  Площадь по кривой:0.914 |
| Метод k-ближайших  соседей (5) | Точность: 0.914  Площадь по кривой:0.914 |
| Метод k-ближайших  соседей (9) | Точность: 0.914  Площадь по кривой: 0.914 |
| Наивный байесовский  классификатор | Точность:0.857  Площадь по кривой:0.857 |
| Случайный лес  (5) | Точность:0.628  Площадь по кривой:0.628 |
| Случайный лес  (10) | Точность:0.742  Площадь по кривой:0.742 |
| Случайный лес  (15) | Точность:0.8  Площадь по кривой:0.8 |
| Случайный лес  (20) | Точность:0.82  Площадь по кривой:0.82 |
| Случайный лес  (50) | Точность:0.83  Площадь по кривой:0.83 |

Полученные результаты были выведены в таблицы.

Вывод:

В процессе выполнения лабораторной работы получили практические навыки решения задачи бинарной классификации данных в среде Jupiter Notebook.

А также была определена максимальная точность классификации 0,914, при модели методом ближайших соседей и при 10% и 35% тестовой выборки.