

## Лабораторная работа №4

### Классификация. Метод $k$ – ближайших соседей ( $k$ -NN)

$k$ -Nearest Neighbors (KNN) - это контролируемый алгоритм машинного обучения, который можно использовать как для задач регрессии, так и для задач классификации. KNN является непараметрическим, что означает, что алгоритм не делает предположений о базовых распределениях данных. Это в отличие от такой техники, как линейная регрессия, которая является параметрической и требует от нас нахождения функции, описывающей взаимосвязь между зависимыми и независимыми переменными.

Преимущество KNN в том, что он интуитивно понятен. При использовании для классификации точка запроса (или тестовая точка) классифицируется на основе  $k$  помеченных обучающих точек, которые находятся ближе всего к этой точке запроса.

### Задание

1. Создать симулированный набор данных и записать его на диск в виде csv файла со следующими параметрами:

- продукт;
- сладость;
- хруст;
- класс.

Продукт	сладость	хруст	класс
Яблоко	7	7	Фрукт
салат	2	5	Овоц
бекон	1	2	Протеин
банан	9	1	Фрукт
орехи	1	5	Протеин

рыба	1	1	Протеин
сыр	1	1	Протеин
виноград	8	1	Фрукт
морковь	2	8	Овощ
апельсин	6	1	Фрукт



- Запрограммировать метрический классификатор по методу k-NN. Для проверки решить ту же задачу методом k-NN библиотеки sklearn.
- Прочитать сгенерированный набор данных. Настроить классификатор. Провести эксперимент по классификации с контролем для подготовленных примеров.
- По возможности результаты визуализировать.
- Ввести в набор данных и примеры продукты еще одного класса (возможно изменив набор параметров) и повторить эксперимент.
- По результатам эксперимента подготовить отчет по лабораторной работе.

## Порядок выполнения работы

1. Был сгенерирован csv файл, в котором находятся данные по продуктам, в соответствии с заданием (с полями «продукт», «сладость», «хруст», «класс»).

Класс «0» - Овощи, «1» – Фрукты, «2» – Протеин.

1	<u>продукт</u> , <u>сладость</u> , <u>хруст</u> , <u>класс</u>
2	яблоко, 7, 6, 1
3	салат, 2, 5, 0
4	бекон, 1, 2, 2
5	банан, 9, 1, 1
6	рыба, 1, 1, 2
7	сыр, 1, 1, 2
8	виноград, 8, 1, 1
9	морковь, 2, 7, 0
10	апельсин, 6, 1, 1
11	груша, 8, 4, 1
12	картофель, 1, 7, 0
13	творог, 2, 1, 2
14	киви, 7, 2, 1
15	фасоль, 2, 2, 2
16	кукуруза, 4, 4, 0
17	манго, 7, 5, 1
18	хлеб, 1, 3, 2
19	лук, 1, 6, 0
20	дыня, 9, 4, 1
21	нектарин, 6, 4, 1
22	

2. Далее был запрограммирован метрический классификатор по методу k-NN. Его необходимо было настроить, для более точного прогнозирования класса. Методом подбора значений, получилось добиться результата работы классификатора, со 100% точностью. При тестовой выборке в 20% и K=3, результатом программы следующий:

Точность метрического классификатора с 3 классами: 1.0, при тестовой выборке 20% от всех данных и k = 3

3. Далее был запрограммирован классификатор по методу k-NN с использованием библиотеки sklearn. При тестовой выборке в 20% и K=3, результатом программы следующий:

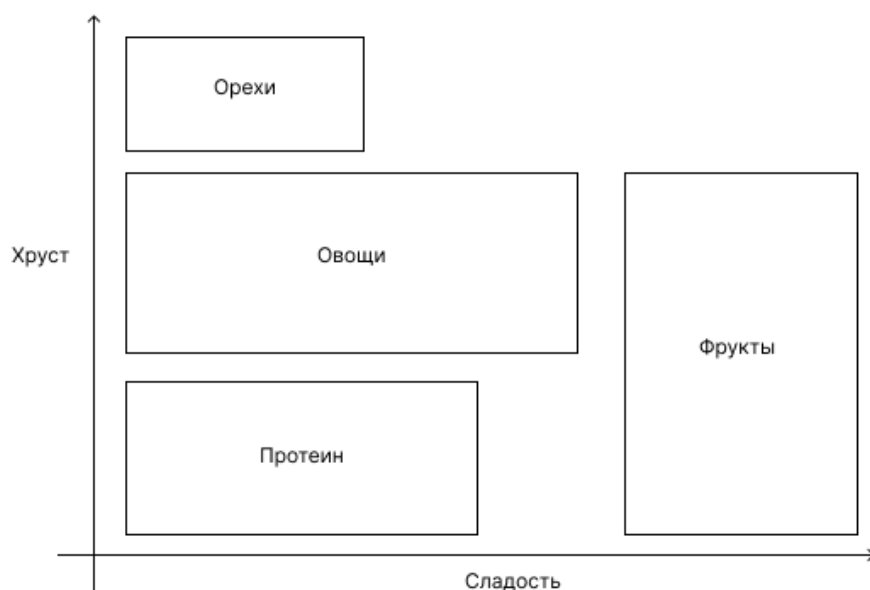
Точность классификатора sklearn с 3 классами: 1.0, при тестовой выборке 20% от всех данных и k = 3

4. Далее был сгенерирован второй dataset, с добавлением нового класса продуктов: 3 – Орехи.

Класс «0» - Овощи, «1» – Фрукты, «2» – Протеин, «3» - Орехи.

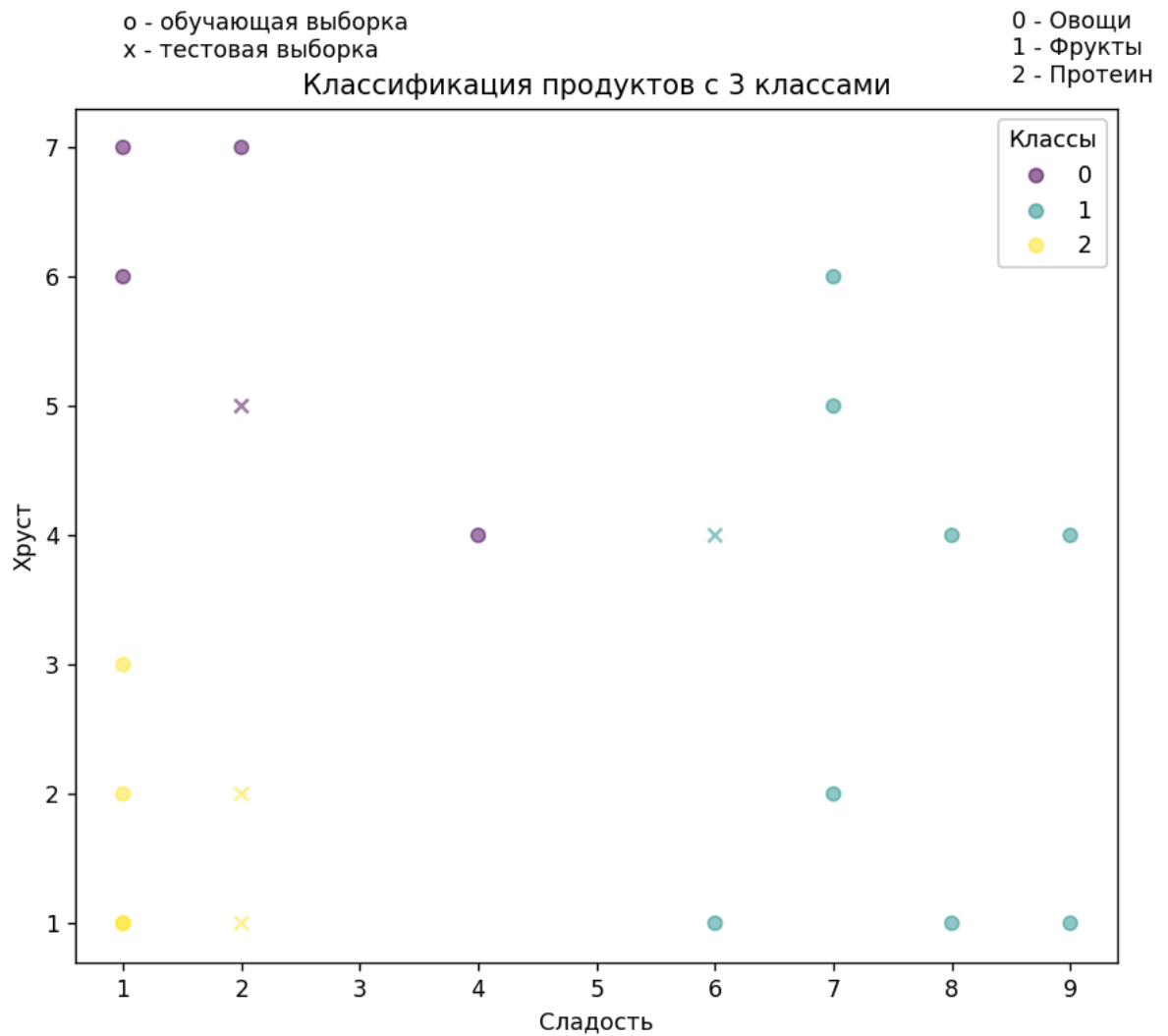
```
1  продукт, сладость, хруст, класс
2  яблоко, 7, 6, 1
3  салат, 2, 5, 0
4  бекон, 1, 2, 2
5  банан, 9, 1, 1
6  кедровый орех, 3, 7, 3
7  рыба, 1, 1, 2
8  сыр, 1, 1, 2
9  виноград, 8, 1, 1
10 морковь, 2, 6, 0
11 фундук, 2, 9, 3
12 груша, 8, 4, 1
13 картофель, 1, 6, 0
14 творог, 2, 1, 2
15 киви, 7, 2, 1
16 фасоль, 2, 2, 2
17 грецкий орех, 1, 8, 3
18 кукуруза, 4, 4, 0
19 манго, 7, 4, 1
20 дыня, 9, 4, 1
21 арахис, 3, 8, 3
```

Данный класс располагается по отношению к другим следующим образом:



5. Далее результаты работы классификаторов были визуализировать и по графикам видно, что результаты соответствуют графикам распределения для 3 классов и для 4 классов:

**Визуализация работы метрического классификатора с 3 классами при тестовой выборке в 20% и  $K=3$**



## Визуализация работы классификатора sklearn с 4 классами при тестовой выборке в 20% и K=3

