Документация по sklearn <https://scikit-learn.org/stable/index.html>

**1 Импорт датасета**

1.1 Найти заданный датасет. Описать его в отчете (к какой предметной области относится, в чем состоит задача, сколько объектов каких классов содержит и какими признаками характеризуется)

1.2 Импортировать датасет в python в том формате, с которым работает sklearn.

1.3 Нормализовать датасет <https://scikit-learn.org/stable/modules/preprocessing.html>

Важно: то, что мы понимаем нормализацией – в документации называется standardization (6.3.1.), а normalization – это нормализация векторов и матриц. Также нужно привести категориальные переменные к численным (6.3.4).

**2 Метод главных компонент**

<https://scikit-learn.org/stable/modules/decomposition.html>

2.1 Заполнить таблицу с собственными векторами и собственными значениями

2.2 Снизить размерность датасета до 2 и вывести двухмерную карту датасета.

2.3 Снизить размерность датасета до 3 и вывести двухмерную карту датасета.

2.4 Снизить размерность датасета до некоторого количества измерений, которое покажется оптимальным исходя из собственных значений.

**3. Метод ближайших соседей**

<https://scikit-learn.org/stable/modules/neighbors.html>

Оценить работу метода ближайших соседей на изначальном датасете в режиме Leave-One-Out (один из экземпляров датасета классифицируется на основе остальных). Посчитать правильно и неправильно классифицированных соседей. Посчитать время выполнения классификации 100 объектов.

Сравнить точность и производительность метода для разных настроек:

- Количество ближайших соседей (1,3,5,7,9,...)

- Использование весов (uniform – равные весы, distance – веса зависят от расстояния)

- Размерность пространств (исходное, либо разные пространства, полученные методом главных компонент).

- Nearest neighbors vs Radius Neighbors (c разными радиусами) vs Nearest Centroid

Результаты занести в таблицы.

Сделать содержательные выводы о применимости

В отчете привести написанные скрипты.

Варианты датасетов с UCI <https://archive.ics.uci.edu/datasets>

Heart disease

Wine

Heart Failure Clinical Records

Energy Efficiency

Wholesale customers

Glass Identification

Seeds