**Лабораторная работа № 1. PCA**

|  |
| --- |
| **Цель:** научиться применять метод PCA для осуществления визуализации данных |

**Общее задание**

1. Используя выборку по варианту, осуществить проецирование данных на плоскость первых двух и трех главных компонент (двумя способами: 1. вручную через использование numpy.linalg.eig для вычисления собственных значений и собственных векторов и 2. с помощью sklearn.decomposition.PCA для непосредственного применения метода PCA – два независимых варианта решения);

2. Выполнить визуализацию полученных главных компонент с использованием средств библиотеки matplotlib, обозначая экземпляры разных классов с использованием разных цветовых маркеров;

3. Используя собственные значения, рассчитанные на этапе 1, вычислить потери, связанные с преобразованием по методу PCA. Сделать выводы;

4. Оформить отчет по выполненной работе, загрузить исходный код и отчет в соответствующий репозиторий на github.

**Задание по вариантам**

| **№ варианта** | **Выборка** | **Класс** |
| --- | --- | --- |
| 1 | seeds.zip | Последняя колонка |
| 2 | hcv+data.zip | Category |
| 3 | exasens.zip | Diagnosis ID |
| 4 | heart+failure+clinical+records.zip | death\_event |
| 5 | wholesale+customers.zip | Region |
| 6 | seeds.zip | Последняя колонка |
| 7 | hcv+data.zip | Category |
| 8 | exasens.zip | Diagnosis ID |
| 9 | heart+failure+clinical+records.zip | death\_event |
| 10 | wholesale+customers.zip | Region |
| 11 | seeds.zip | Последняя колонка |
| 12 | hcv+data.zip | Category |
| 13 | exasens.zip | Diagnosis ID |
| 14 | heart+failure+clinical+records.zip | death\_event |
| 15 | wholesale+customers.zip | Region |
| 16 | seeds.zip | Последняя колонка |
| 17 | hcv+data.zip | Category |
| 18 | exasens.zip | Diagnosis ID |
| 19 | heart+failure+clinical+records.zip | death\_event |
| 20 | wholesale+customers.zip | Region |

**Важные замечания**

1. Рекомендуется использовать библиотеку pandas для загрузки и манипуляции с данными;

2. Указание класса дается только для визуализации, при применении метода PCA данную колонку следует удалять;

3. В случае отсутствия данных следует применить какую-либо стратегию для их замещения.

**Руководства, литература**

См. файл PCA\_explanation.pdf