**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра МОЭВМ**

отчет

**по лабораторной работе №2**

**по дисциплине «Машинное обучение»**

Тема: Понижение размерности пространства признаков

|  |  |  |
| --- | --- | --- |
| Студент гр. 8304 |  | Холковский К.В |
| Преподаватель |  | Жангиров Т. Р. |

Санкт-Петербург

2021

**Цель работы**

Ознакомиться с методами понижения размерности данных из библиотеки Scikit Learn

**Ход работы**

**Загрузка данных**

Были загружены и нормированы данные и построена диаграмма.

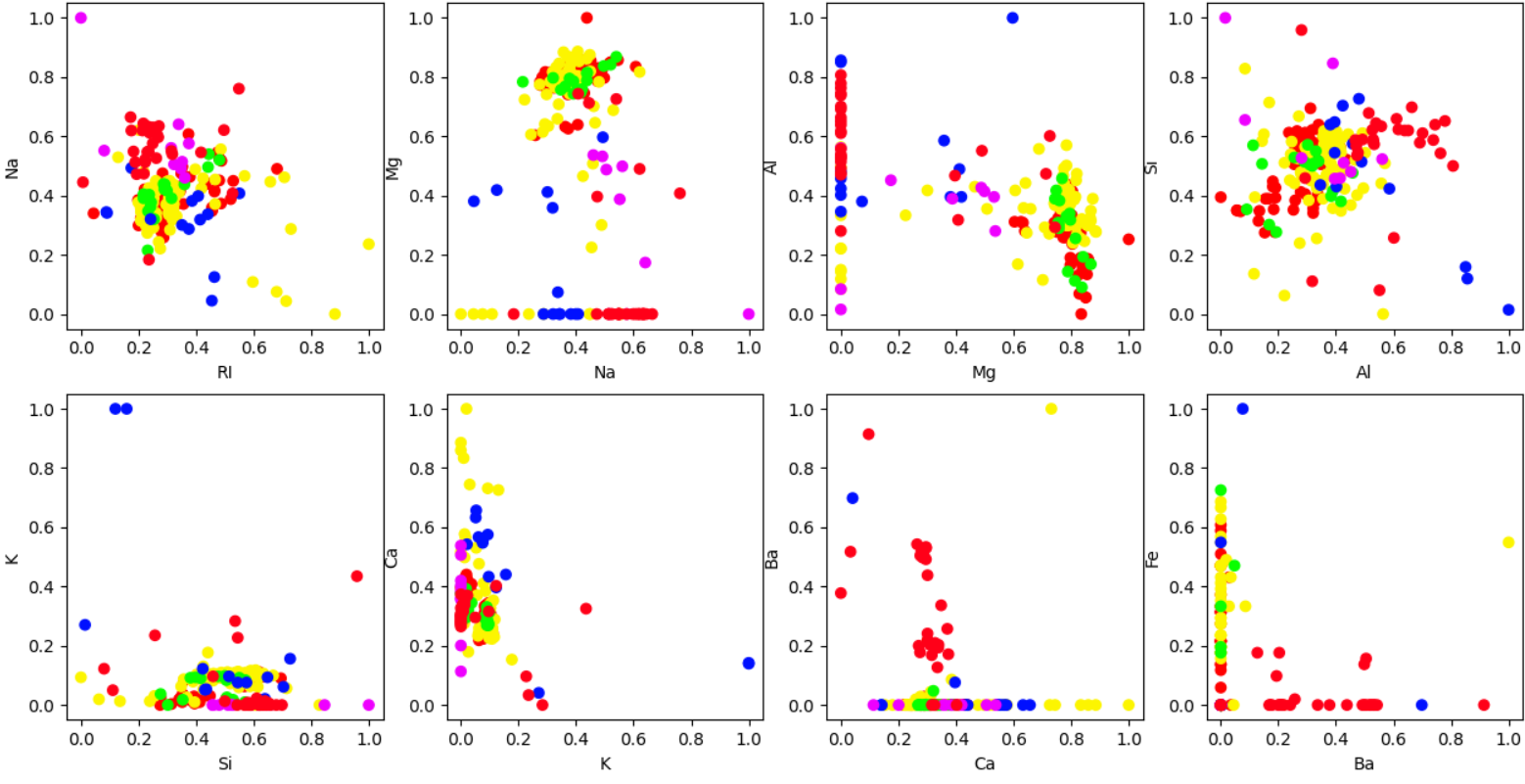


Рис 1 – Диаграмма рассеяния

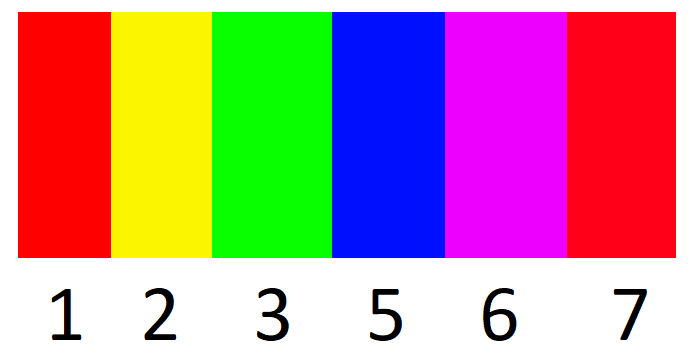


Рис 2 – Соответствие цвета на диаграмме и класса в датасете

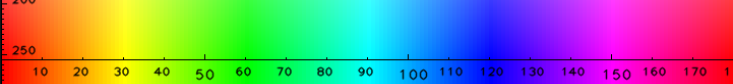


Рис 3 – HSV colormap

**Метод главных компонент**

Для нашего набора данных были получены следующие значения объясненной дисперсии и собственные числа.

[0.45429569 0.17990097]

[5.1049308 3.21245688]

Была построена диаграмма для 2-х компонент

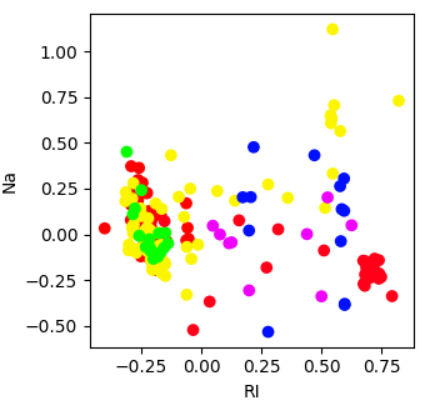


Рис 4 – Диаграмма PCA

Было выявлено количество компонент, объясняющих не менее 85% - 4.

Данные восстановленные при помощи inverse\_transform соответствуют исходным данным пропорционально размеру дисперсии данных.

Параметр svd\_solver:

|  |  |
| --- | --- |
| auto | Если входные данные больше 500x500, а количество извлекаемых компонентов меньше 80% наименьшего измерения данных, тогда более эффективнее использовать 'randomized'. В противном случае вычисляется full SVD и, возможно, впоследствии усекается. |
| full | Полное вычисление SVD. |
| arpack | Запускает SVD, усеченный до n\_components. Требуется строго 0 <n\_components <min (X.shape). |
| randomized | Вычисление рандомизированного SVD. |

**Модификации метода главных компонент**

По диаграммам видно, что linear соответствует обычному PCA .

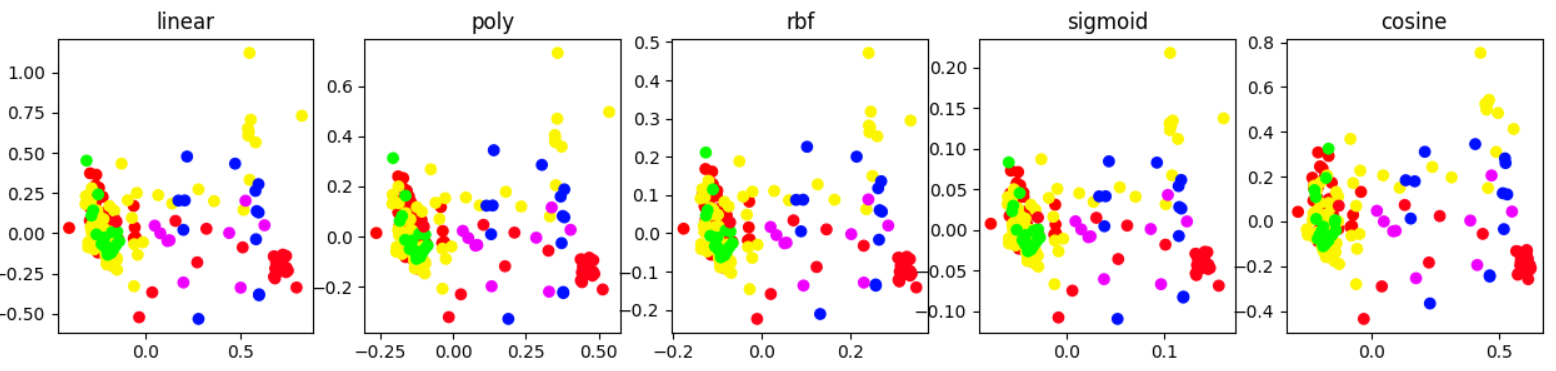


Рис 5 – Диаграммы KPCA

По диаграммам видно, что при alpha=0 данные SparsePCA совпадают с данными PCA.

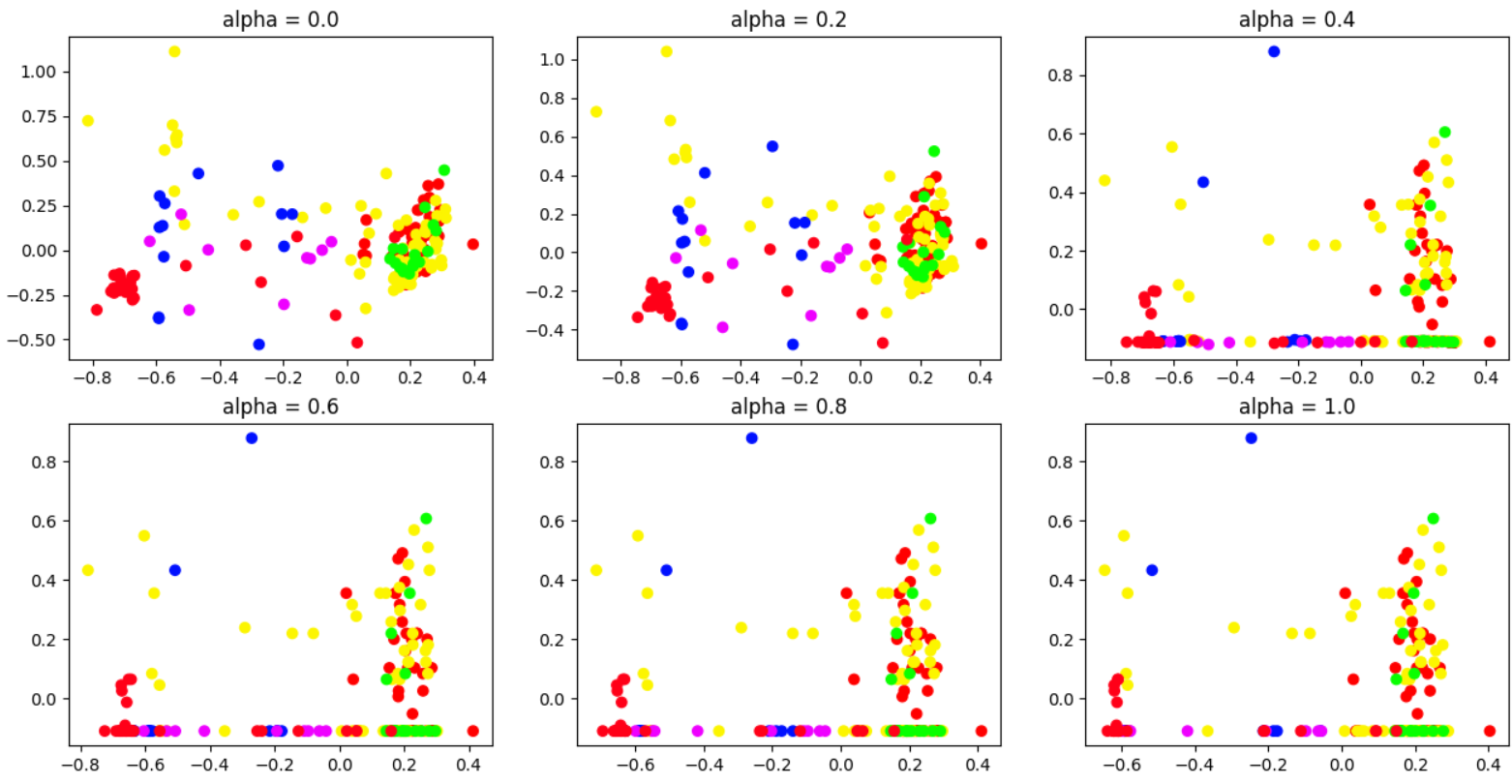


Рис 6 – Диаграммы SPCA

**Факторный анализ**

По диаграмме видно, что данные сильно отличаются.

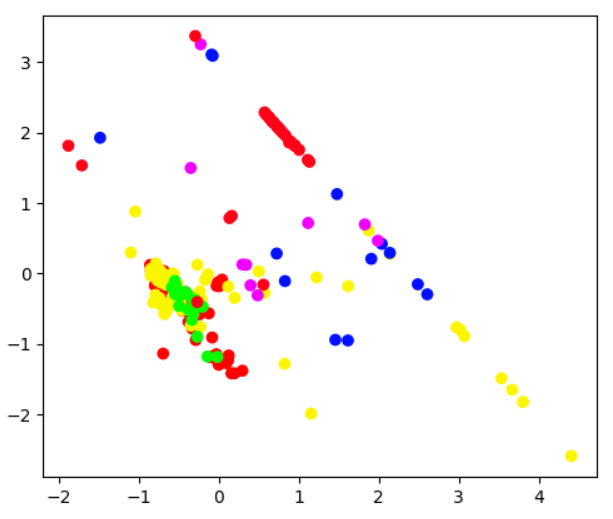


Рис 6 – Диаграмма факторного анализа

Математика факторного анализа и анализа главных компонент (PCA) различна. Факторный анализ явно предполагает наличие скрытых факторов, лежащих в основе наблюдаемых данных. Вместо этого PCA стремится идентифицировать переменные, которые являются составными частями наблюдаемых переменных. Хотя эти методы могут давать разные результаты, они схожи с тем, что ведущее программное обеспечение, используемое для проведения факторного анализа (SPSS Statistics), использует PCA в качестве алгоритма по умолчанию.