Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет компьютерных систем и сетей

Кафедра информатики

Дисциплина: Машинное обучение (МО)

Отчет о лабораторной работе №7

«Метод главных компонент»

|  |
| --- |
| **Выполнил:** |
| **Студент гр. 858341** |
| **Немкович А. В.** |

|  |
| --- |
| **Проверил:** |
| **Стержанов М. В.** |

**Минск 2019**

Содержание

[Постановка задачи 3](#_Toc25275643)

[Ход выполнения 4](#_Toc25275644)

[Выводы 10](#_Toc25275645)

# Постановка задачи

Набор данных **ex7data1.mat** представляет собой файл формата \*.mat (т.е. сохраненного из Matlab). Набор содержит две переменные X1 и X2 - координаты точек, для которых необходимо выделить главные компоненты.

Набор данных **ex7faces.mat** представляет собой файл формата \*.mat (т.е. сохраненного из Matlab). Набор содержит 5000 изображений 32x32 в оттенках серого. Каждый пиксель представляет собой значение яркости (вещественное число). Каждое изображение сохранено в виде вектора из 1024 элементов. В результате загрузки набора данных должна быть получена матрица 5000x1024.

**Задание.**

1. Загрузите данные **ex7data1.mat** из файла.
2. Постройте график загруженного набора данных.
3. Реализуйте функцию вычисления матрицы ковариации данных.
4. Вычислите координаты собственных векторов для набора данных с помощью сингулярного разложения матрицы ковариации (разрешается использовать библиотечные реализации матричных разложений).
5. Постройте на графике из пункта 2 собственные векторы матрицы ковариации.
6. Реализуйте функцию проекции из пространства большей размерности в пространство меньшей размерности с помощью метода главных компонент.
7. Реализуйте функцию вычисления обратного преобразования.
8. Постройте график исходных точек и их проекций на пространство меньшей размерности (с линиями проекций).
9. Загрузите данные **ex7faces.mat** из файла.
10. Визуализируйте 100 случайных изображений из набора данных.
11. С помощью метода главных компонент вычислите собственные векторы.
12. Визуализируйте 36 главных компонент с наибольшей дисперсией.
13. Как изменилось качество выбранных изображений?
14. Визуализируйте 100 главных компонент с наибольшей дисперсией.
15. Как изменилось качество выбранных изображений?
16. Используйте изображение, сжатое в лабораторной работе №5.
17. С помощью метода главных компонент визуализируйте данное изображение в 3D и 2D.
18. Соответствует ли 2D изображение какой-либо из проекций в 3D?
19. Ответы на вопросы представьте в виде отчета.

# Ход выполнения

**Задания 1-2**: Загрузка данных ex7data1.mat из файла. Построение графика загруженного набора данных.

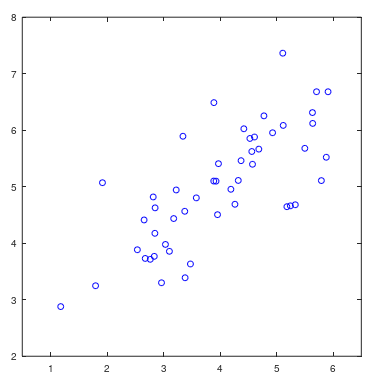


Рис 1. – График загруженного набора данных (ex7data1.txt)

**Задание 3**: Реализация функции вычисления матрицы ковариации данных. Вычисление координатов собственных векторов для набора данных с помощью сингулярного разложения матрицы ковариации.

Формула вычисления ковариационной матрицы:



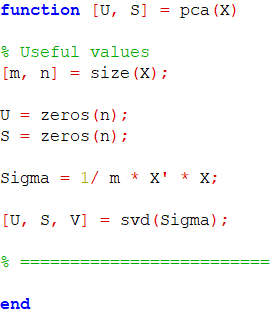


Рис 2. – Код функции pca

**Задание 5**: Построение на графике из пункта 2 собственных векторов матрицы ковариации.

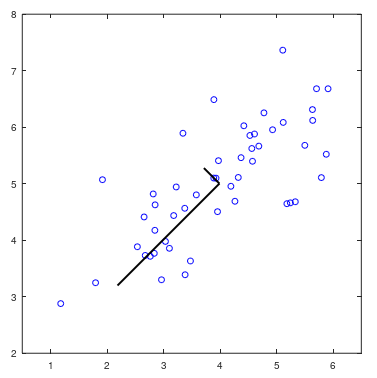


Рис 3. – График собственных векторов матрицы ковариации

**Задание 6**: Реализация функции проекции из пространства большей размерности в пространство меньшей размерности с помощью метода главных компонент.

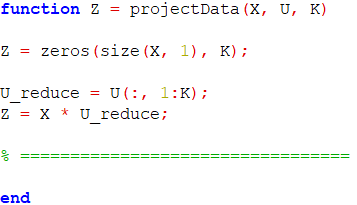


Рис 4. – Код функции функции проекции из пространства большей размерности в пространство меньшей размерности

**Задание 7**: Реализация функции вычисления обратного преобразования.

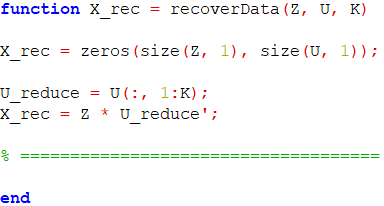


Рис 5. – Код функции вычисления обратного преобразования

**Задание 8**: Построение графика исходных точек и их проекций на пространство меньшей размерности (с линиями проекций).

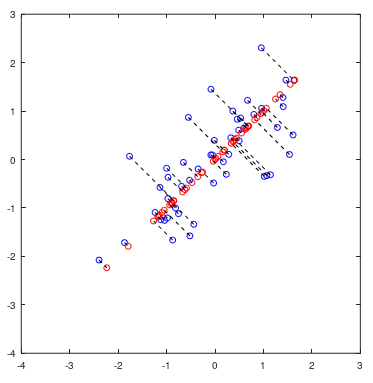


Рис 6. – График исходных точек и их проекций на пространство меньшей размерности

**Задания 9-10**: Загрузка данных ex7faces.mat из файла. Визуализация 100 случайных изображений из набора данных.



Рис 7. – Визуализация 100 случайных изображений из набора данных (ex7faces.mat)

**Задания 11-12**: Вычисление собственных векторов с помощью метода главных компонент. Визуализация 36 главных компонент с наибольшей дисперсией.



Рис 8. – Визуализация 36 главных компонент с наибольшей дисперсией

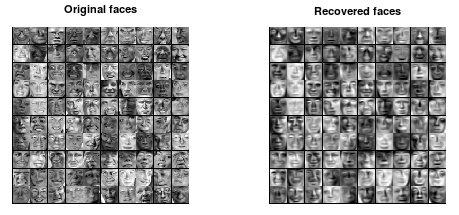


Рис 9. – Визуализация исходных изображений и восстановленных используя 36 главных компонент

**Задание 14**: Визуализация 100 главных компонент с наибольшей дисперсией.

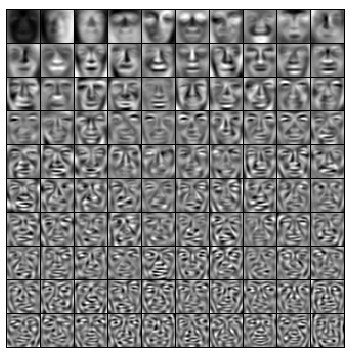


Рис 10. – Визуализация 100 главных компонент с наибольшей дисперсией

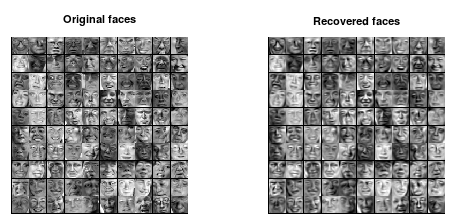


Рис 11. – Визуализация исходных изображений и восстановленных используя 100 главных компонент

**Задания 16-17**: Использование изображения, сжатое в лабораторной работе №5. Визуализация данного изображения в 3D и 2D с помощью метода главных компонент.



Рис 12. – Исходное изображение

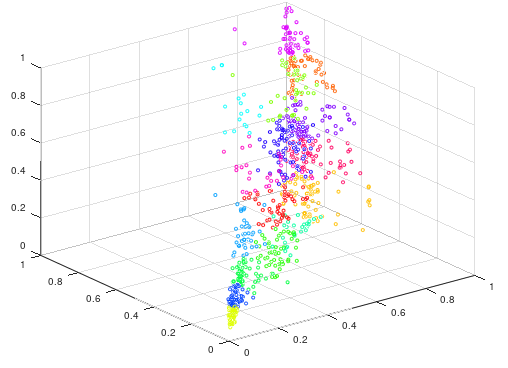


Рис 13. – Визуализация изображения в 3D

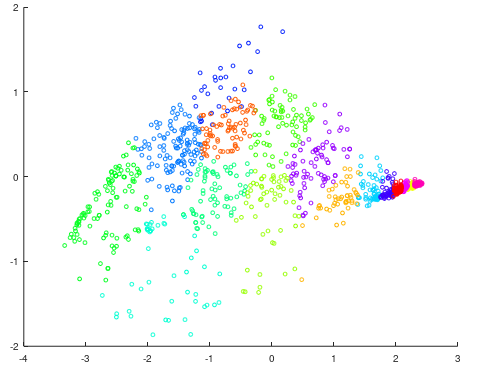


Рис 14. – Визуализация изображения в 2D

# Выводы

Мы научились использовать метод главных компонент для проекции из пространства большей размерности в пространство меньшей размерности с помощью вычисления матрицы ковариации данных. А также выполнять вычисления обратного преобразования.

Применение метода главных компонент:

1. Сжатие

* уменьшение необходимого объема памяти для хранения данных
* ускорение алгоритма обучения

2. Визуализация (использование в 2D и 3D)