

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ
УНИВЕРСИТЕТ
им. В.Г.Шухова»
(БГТУ им. В.Г.Шухова)

Кафедра технической кибернетики

Дисциплина: Теория матриц

Практическая работа № 8

Тема: «Метод главных компонент»

Выполнил:

Студент группы МТК-233

Орлов-Курेशи М. Н.

Проверил:

Кариков Е. Б.

Белгород 2023

Цель работы: изучить метод главных компонент. Реализовать метод главных компонент на языке программирования Python.

Метод главных компонент

Метод главных компонент и сингулярное разложение. Сингулярное разложение удобно использовать для нахождения главных компонент. В сингулярном разложении исходная матрица X представляется как произведение трех матриц

$$X = U\Sigma V^T$$

Здесь U – матрица, образованная левыми сингулярными векторами – ортонормированными собственными векторами u_r матрицы XX^T , соответствующим значениям λ_r ;

$$XX^T u_r = \lambda_r v_r;$$

V – матрица, образованная правыми сингулярными векторами – ортонормированными собственными векторами v_r матрицы $X^T X$;

$$X^T X v_r = \lambda_r v_r;$$

Σ – положительно определенная диагональная матрица, элементы которой являются сингулярными числами $\sigma_1 \geq \dots \geq \sigma_R \geq 0$ равными квадратным корням из собственных значений λ_r

$$\sigma_r = \sqrt{\lambda_r}$$

Связь между методом главных компонент и сингулярным разложением определяется следующими простыми соотношениями

$$T = U\Sigma, P = V.$$

Счета. Матрица счетов T дает нам проекции исходных образцов (J – мерных x_1, \dots, x_i) на подпространство главных компонент (A – мерное). Строки t_1, \dots, t_i матрицы T – это координаты образцов в новой системе координат. Столбцы t_1, \dots, t_A матрицы T – ортогональны и представляют проекции всех образцов на одну новую проекцию координатную ось.

Реализация алгоритма на языке программирования Python

```
X = Matrix(45, 5, [random.randint(1,5) for _ in range(45*5)])

sum_col = [0] * 5

for i in range(X.row):
    for j in range(X.column):
        sum_col[j] += X[i,j]

for i in range(5):
    sum_col[i] /= 45

base_data = X.copy()

for i in range(X.row):
    for j in range(X.column):
```

```

        X[i,j] -= sum_col[j]

class PCA:
    def __init__(self, n_components=None):
        self.n_components = n_components
        self.P = None

    def fit(self, data:Matrix, n_components=None):
        U, s, V = data.svd()
        if n_components is None:
            self.P = U.get_column(range(self.n_components), self.n_components)
        else:
            self.P = U.get_column(range(n_components), n_components)

    def transform(self, data:Matrix):
        return data * self.P

pca = PCA(2)
pca.fit(X)
T = pca.transform(X)

x = [point[0] for point in T._elements]
y = [point[1] for point in T._elements]

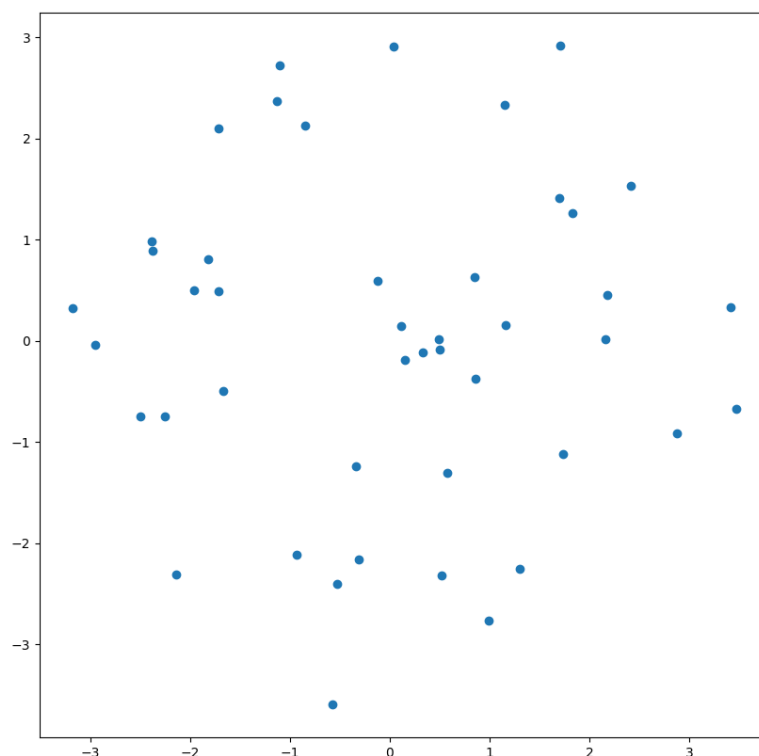
plt.figure(figsize=(10,10))
plt.scatter(x, y)

plt.show()

```

Скриншоты работы программы

Проекция исходных данных на двумерное пространство



Вывод: в ходе работы был изучен и реализован метод главных компонент. При помощи метода главных компонент была снижена размерность данных.

Список литературы

1. Юдин Д.А. Прикладные аспекты теории матриц: учебное пособие / Д.А. Юдин. - Белгород: Изд-во БГТУ, 2016.