
DATA SCIENCE PCA LAB 5

Roll no: 20K-0409

Screen Shots

DATAs#1

```
30 |         return FinalSelectedVectors, matrix_rank
31 |
32 |     # path of dataset
33 |     dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\Country-data.csv'
34 |
35 |     principal_components_housing, rank_housing = perform_pca(dataset_path)
36 |
37 |     # first 5 principal components and their rank
38 |     print("First 5 Principal Components for the housing dataset:")
39 |     print(principal_components_housing[:, :5])
40 |     print("Matrix Rank for the housing dataset:", rank_housing)
41 |
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS AZURE

Shape of the dataset before PCA: (167, 10)
First 5 Principal Components for the housing dataset:
[[-5.09077639e-03+0.00000000e+00j -3.93196540e-03+0.00000000e+00j
 -2.79611783e-03-6.88621257e-03j -2.79611783e-03+6.88621257e-03j
 -5.86583601e-03+0.00000000e+00j]
 [2.33696329e-02+0.00000000e+00j -5.55072178e-04+0.00000000e+00j
 -1.83273883e-02-1.04200908e-02j -1.83273883e-02+1.04200908e-02j
 1.41289706e-02+0.00000000e+00j]
 [3.57057863e-02+0.00000000e+00j 1.79815974e-03+0.00000000e+00j]

DATAs#2

```
32 |     # path of dataset
33 |     dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\bike-sharing.csv\bike-sharing.csv'
34 |
35 |     principal_components_housing, rank_bike = perform_pca(dataset_path)
36 |
37 |     # first 5 principal components and their rank
38 |     print("First 5 Principal Components for the bike sharing dataset:")
39 |     print(principal_components_housing[:, :5])
40 |     print("Matrix Rank for the bike sharing dataset:", rank_bike)
41 |
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS AZURE

PS C:\Users\Mukand\Desktop\data set for pca> & "C:/Program Files/Python310/python.exe" "c:/Users/Mukand/Desktop/
Shape of the dataset before PCA: (17379, 17)
First 5 Principal Components for the bike sharing dataset:
[[0.+0.j 0.+0.j 0.+0.j 0.+0.j 0.+0.j]
 [0.+0.j 0.+0.j 0.+0.j 0.+0.j 0.+0.j]
 [0.+0.j 0.+0.j 0.+0.j 0.+0.j 0.+0.j]
 ...
 [0.+0.j 0.+0.j 0.+0.j 0.+0.j 0.+0.j]
 [0.+0.j 0.+0.j 0.+0.j 0.+0.j 0.+0.j]
 [0.+0.j 0.+0.j 0.+0.j 0.+0.j 0.+0.j]]
Matrix Rank for the bike sharing dataset: 15
PS C:\Users\Mukand\Desktop\data set for pca> |

```
33 # dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\bike-sharing.csv\bike-sharing.csv'
34 # dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\bike-sharing.csv\bike-sharing.csv'
35 dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\autos.csv'
36 # dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\Country-data.csv'
37 # dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\customer.csv\customer.csv'
38 # dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\abalone.csv'
39
40 principal_components_housing, rank_bike = perform_pca(dataset_path)
41
42 # first 5 principal components and their rank
43 print("First 5 Principal Components for the dataset:")
44 print(principal_components_housing[:, :5])
45 print("Matrix Rank for dataset:", rank_bike)
46
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS AZURE

[Done] exited with code=1 in 5.15 seconds

[Running] python -u "c:\Users\Mukand\Desktop\data set for pca\pca lab # 5.py"

Shape of the dataset before PCA: (167, 10)

First 5 Principal Components for the dataset:

```
[[ -5.09077639e-03+0.00000000e+00j -3.93196540e-03+0.00000000e+00j
  -2.79611783e-03-6.88621257e-03j -2.79611783e-03+6.88621257e-03j
  -5.86583601e-03+0.00000000e+00j]
 [ 2.33696329e-02+0.00000000e+00j -5.55072178e-04+0.00000000e+00j
  -1.83273883e-02-1.04200908e-02j -1.83273883e-02+1.04200908e-02j
  1.41289706e-02+0.00000000e+00j]
 [ 3.57057863e-02+0.00000000e+00j 1.79815974e-03+0.00000000e+00j
  1.82111746e-02-4.78167975e-02j 1.82111746e-02+4.78167975e-02j]
```

DataSet # 3

DATASet#4

```
36 # dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\Country-data.csv'
37 dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\customer.csv\customer.csv'
38 # dataset_path = r'C:\Users\Mukand\Desktop\data set for pca\abalone.csv'
39
40 principal_components_housing, rank_bike = perform_pca(dataset_path)
41
42 # first 5 principal components and their rank
43 print("First 5 Principal Components for the dataset:")
44 print(principal_components_housing[:, :5])
45 print("Matrix Rank for dataset:", rank_bike)
46
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS AZURE Code

principal_components_housing, rank_bike = perform_pca(dataset_path)

File "c:\Users\Mukand\Desktop\data set for pca\lab #5.py", line 15, in perform_pca

AAT = np.matmul(A, AT)

numpy.core._exceptions._ArrayMemoryError: Unable to allocate 17.8 GiB for an array with shape (48895, 48895) and data type float64

[Done] exited with code=1 in 6.089 seconds

[Running] python -u "c:\Users\Mukand\Desktop\data set for pca\lab #5.py"

Shape of the dataset before PCA: (9134, 25)

First 5 Principal Components for the dataset:

```
[[ 2.95088290e-03+3.75495775e-03j 2.95088290e-03-3.75495775e-03j
  1.90630748e-03-7.67307485e-04j 1.90630748e-03+7.67307485e-04j
  -2.19007450e-03+0.00000000e+00j]
 [-6.09544185e-05+4.96169503e-04j -6.09544185e-05-4.96169503e-04j
  6.15558735e-04-2.82162584e-04j 6.15558735e-04+2.82162584e-04j
  3.86333132e-05+0.00000000e+00j]
 [ 3.71595403e-03-7.28122793e-03j 3.71595403e-03+7.28122793e-03j
  -6.43476455e-03+5.95359061e-04j -6.43476455e-03-5.95359061e-04j
  6.52768438e-03+0.00000000e+00j]
```

DataSet #5

```
[Done] exited with code=1 in 1.011 seconds
```

[Running] python -u "c:\Users\Mukand\Desktop\7th sem 21st Aug - 2023\DS\lab 5\data set for pca\lab.py"

Shape of the dataset before PCA: (2930, 79)

First 5 Principal Components for the dataset:

```
[[ 0.0097941 +0.00000000e+00j -0.00383068-1.72529485e-04j
  -0.00383068+1.72529485e-04j 0.00172079+2.41449258e-03j
  0.00172079-2.41449258e-03j]
 [ 0.01704646+0.00000000e+00j -0.00020056+2.20056734e-05j
  -0.00020056-2.20056734e-05j -0.00034567+3.19054536e-04j
  -0.00034567-3.19054536e-04j]
 [ 0.01811726+0.00000000e+00j -0.01111475-2.05183456e-03j
  -0.01111475+2.05183456e-03j -0.01018182-6.61870288e-03j
  -0.01018182+6.61870288e-03j]
 ...
 [ 0.00598477+0.00000000e+00j -0.00706216+2.49251454e-03j
  -0.00706216-2.49251454e-03j 0.00313746+5.31285548e-03j
  0.00313746-5.31285548e-03j]
```