IC272 – Data Science 3 *Aryan Garg*

REPORT *B19153*

Lab Assignment – 3 *+91-8219383122*

1. (a) Output of the min-max normalization of data after outlier replacement by non-outlier median of each attribute:

```
                    ***** Lab Assignment - 3 *****

Q1
a) Min-Max Normalization:

        Attributes                              BEFORE (Min, Max)  AFTER (Min, Max)
              temperature                    (10.08511, 31.375)       (3.0, 9.0)
              humidity                        (80.53333, 99.72)       (3.0, 9.0)
              pressure       (999.7183333333, 1029.3185106383)       (3.0, 9.0)
              rain                                 (0.0, 45.0)       (3.0, 9.0)
              lightavgw/o0                  (0.0, 10565.3523)       (3.0, 9.0)
              lightmax                         (2259, 54612)       (3.0, 9.0)
              moisture                          (0.0, 100.0)       (3.0, 9.0)
```

Figure 1: Q1(a) Min-Max Normalization

Output of z-score normalization:

```
b) Standardization with μ=0, σ=1:
                                          BEFORE (μ, σ)                                AFTER (μ, σ)
        temperature    (21.369383661375664, 4.123161231575466)   (-7.443781041296288e-16, 0.9999999999999999)
        humidity        (93.48644115343915, 4.768865753965264)   (1.5940217987422353e-15, 1.0000000000000002)
        pressure      (1014.9019655627992, 5.6916420561171055)   (-7.669350163759811e-15, 0.9999999999999999)
        rain            (10.28678306878307, 9.526541815555257)   (-1.2030353198054607e-16, 0.9999999999999999)
        lightavgw/o0  (2237.8998394708997, 2205.2551219743423)                (-1.5037941497568257e-16, 1.0)
        lightmax       (21788.62328042328, 22053.315399022667)    (3.759485374392064e-17, 0.9999999999999999)
        moisture       (32.38605259259259, 33.635433988151505)                                (0.0, 1.0)
```

Figure 2: Q1(b) Z-score normalization

It is evident that we have to deal with very small values using z-score normalization which will improve running time of the training.

However, human interpretation of data after z-score normalization is impractical.

2. Synthetic data (2 X 1000) created using NumPy's random class' multivariate normal function. Here's the scatter plot of data:

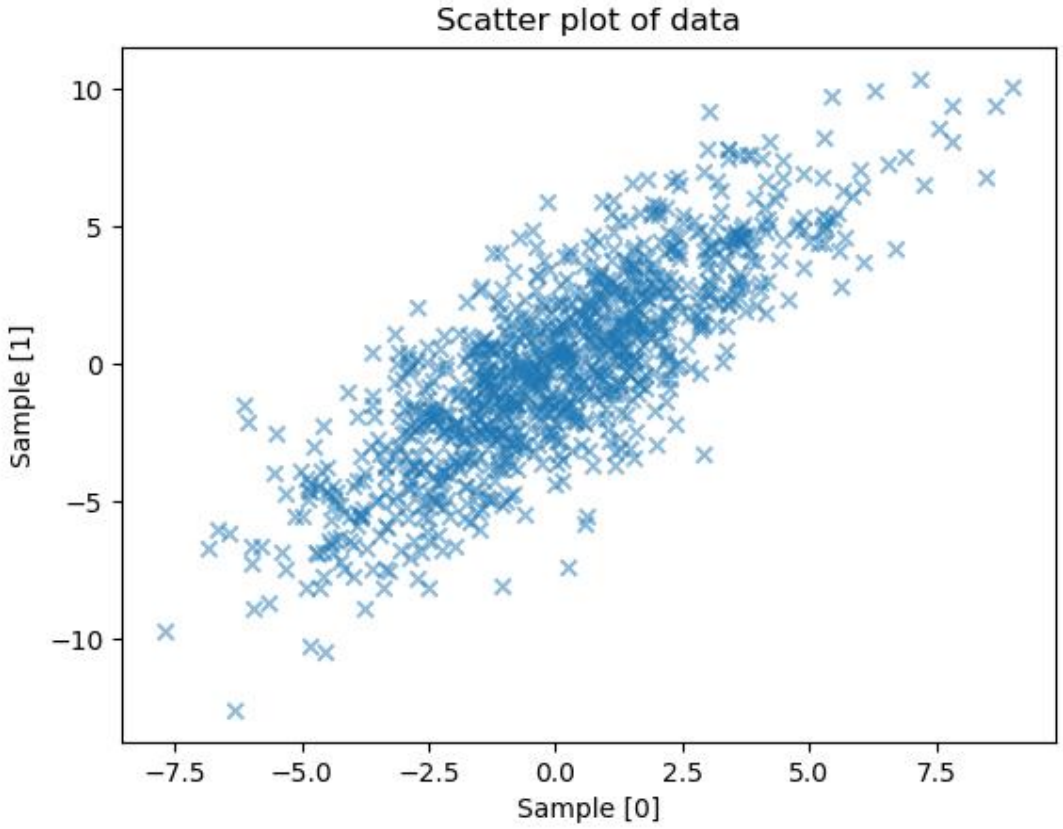

Figure 3: Q2(a) Synthetic Data

```
b) Eigenvalues and Eigenvectors of covariance matrix:

                Eigenvalues: 1.6194967085312015, 17.544561636819374
                Eigenvectors: [-0.83167507  0.55526261], [-0.55526261 -0.83167507]
```

Figure 4: Q2(b)

Then we project the data on both eigenvectors individually and then draw the data alongside the original eigen vectors. Here are the scatter plots:

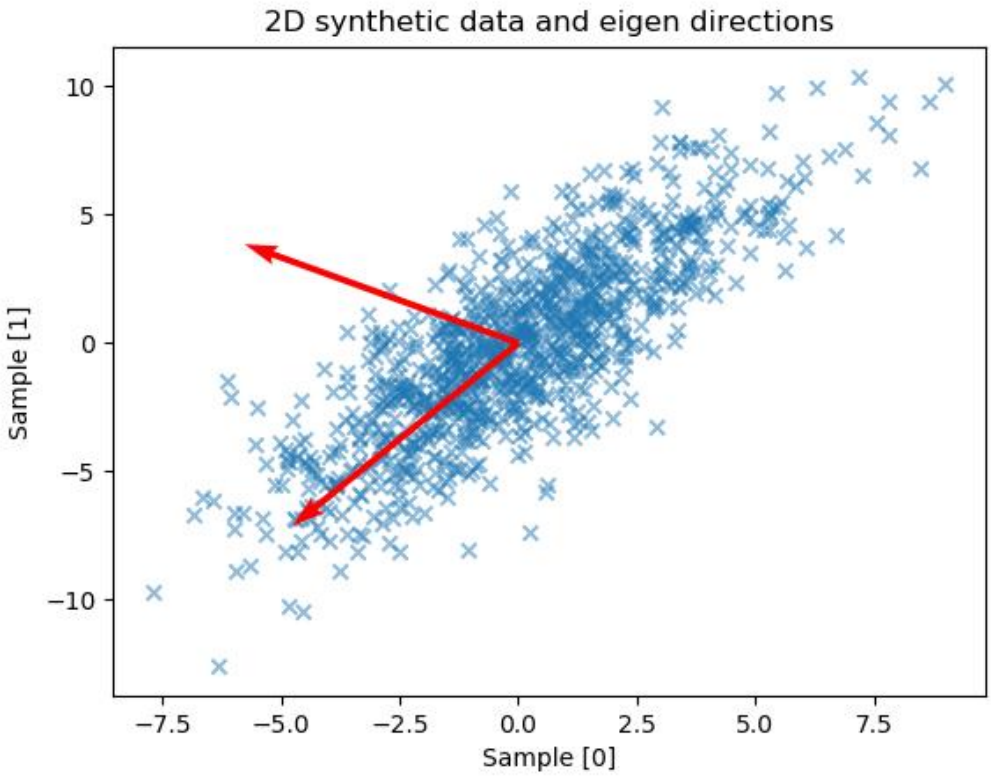

Figure 5: Q2(c) Projection on eigenvectors

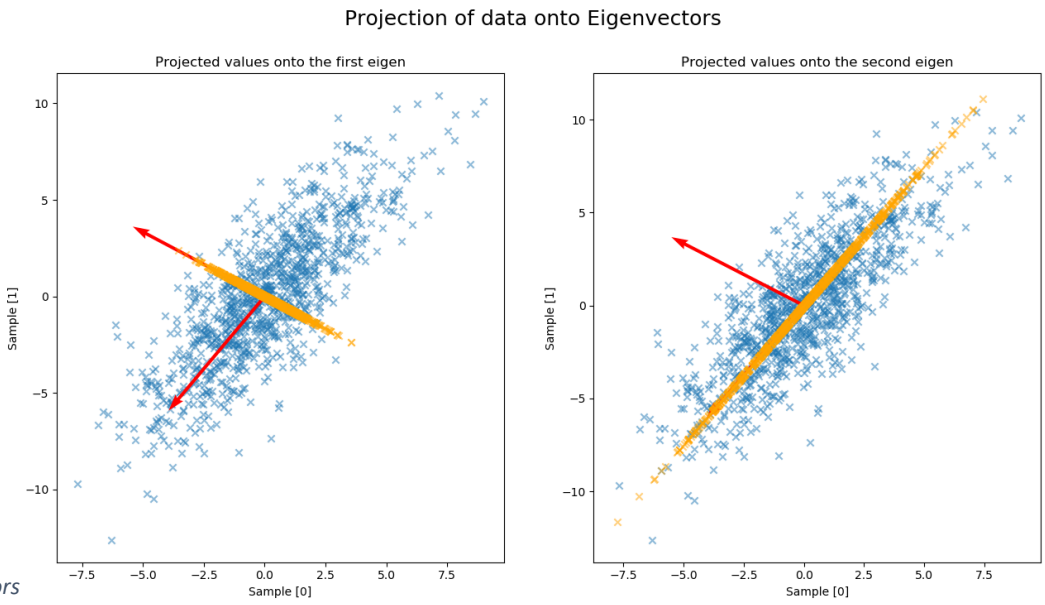

Figure 6: Eigenvectors

Mean-square-error on projection:

```
d) Reconstruction error between reconstructed data and original data:

                MSE for Sample [0]: 5.480079582358328e-32
                MSE for Sample [1]: 9.295654950980545e-32
```

Figure 7: Q2(d) Reconstruction Error

3. After dimensionality reduction, from l = 7 to l = 2:

```
Q3

a) Variance and Eigenvalue of projected data:

        Variance, Eigenvalue at Dimension [0]: 488574184.8009496, 489091742.20010304
        Variance, Eigenvalue at Dimension [1]: 2637699.7383154533, 2640493.911767059
```

Figure 8: Q3(a) Variance and Eigen values

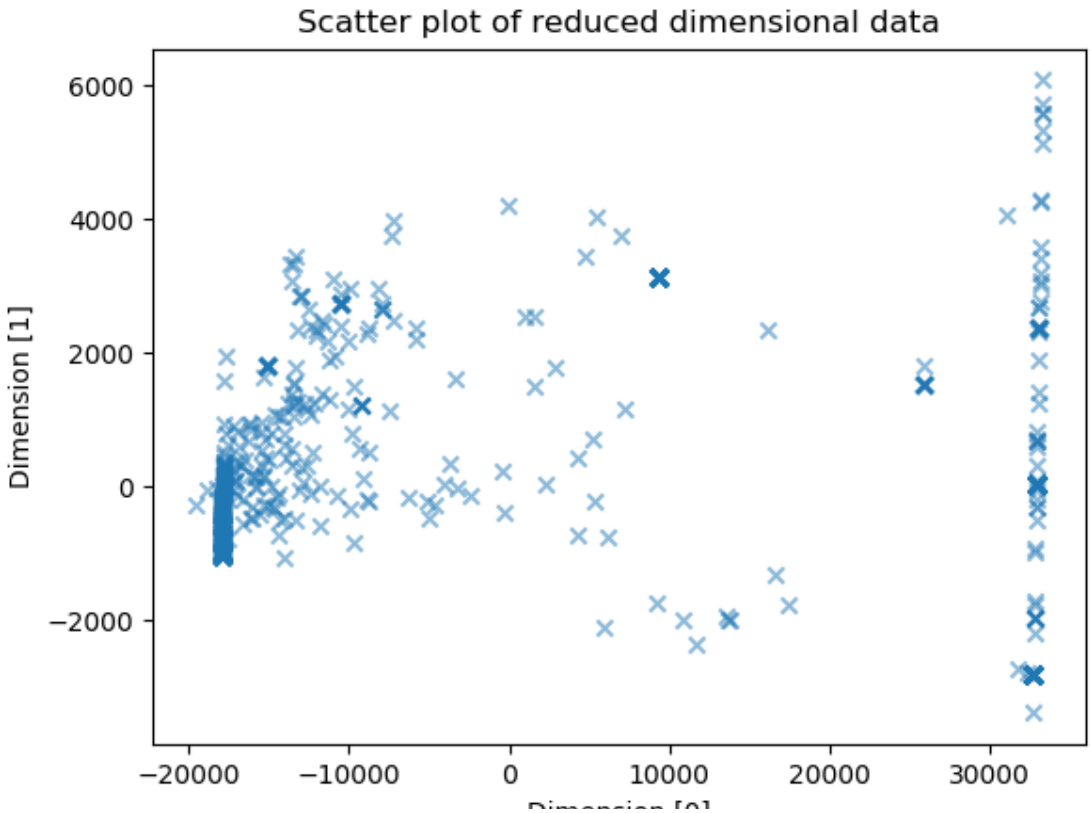

Figure 9: Q3(a) Scatter plot of data

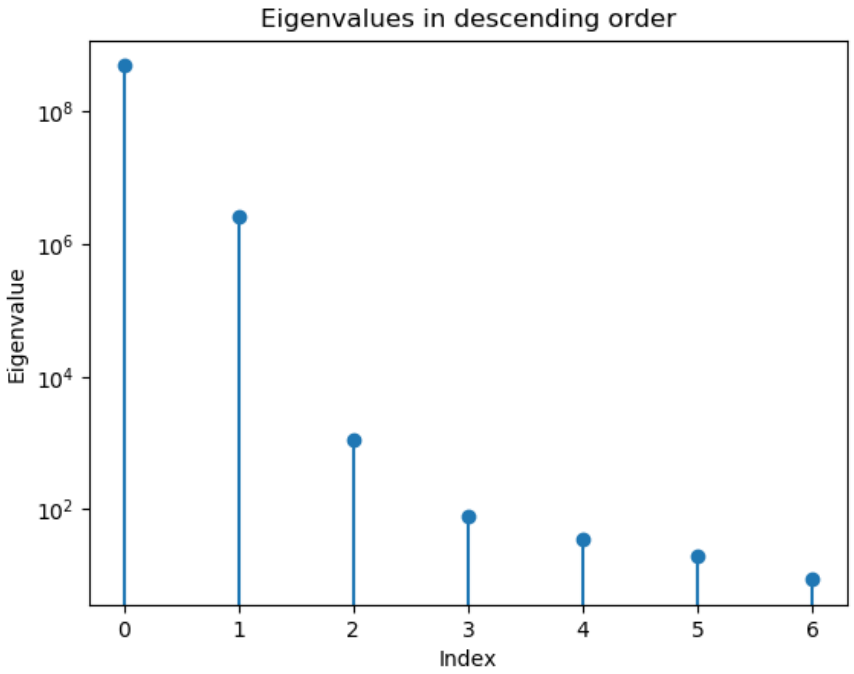

Figure 10: Q3(b) Eigenvalues in descending order

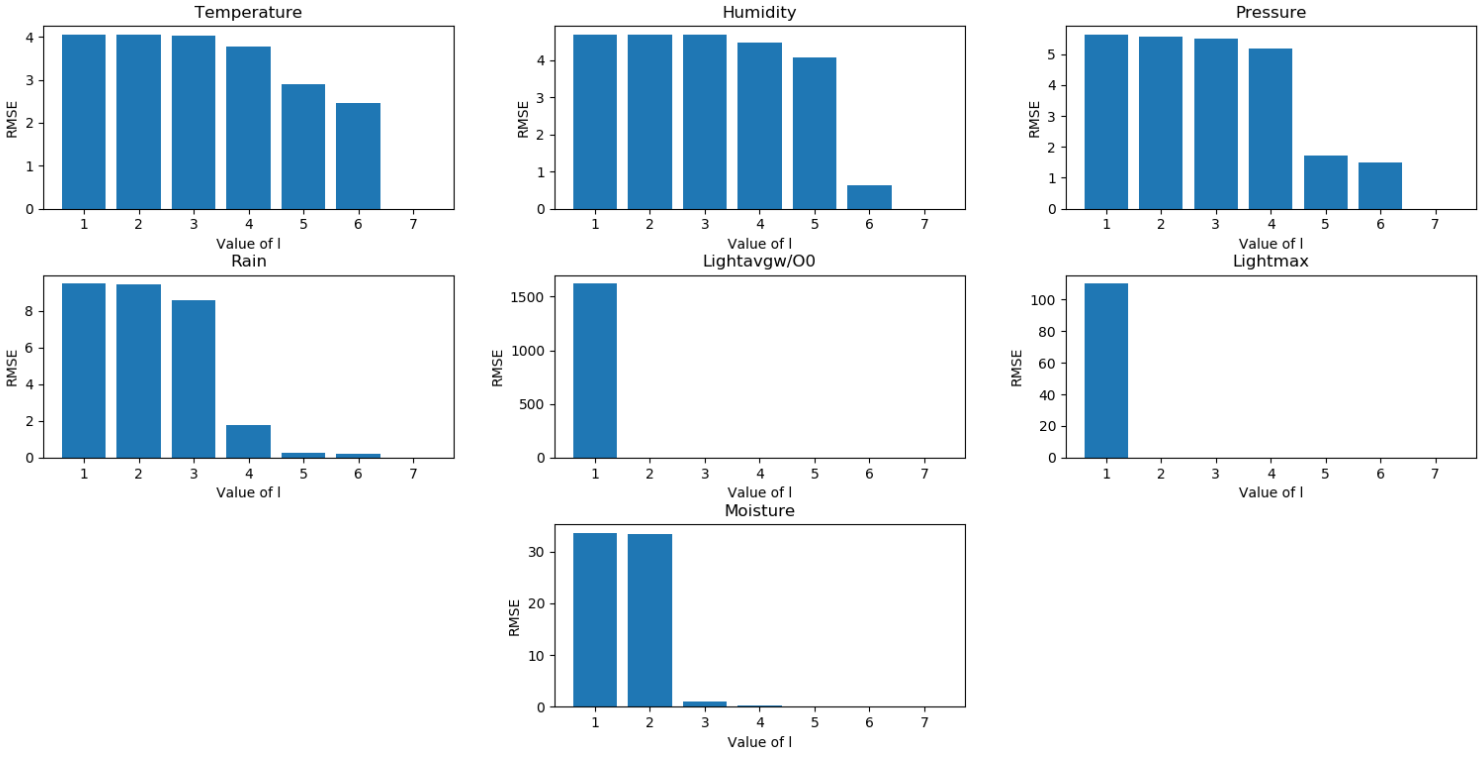

Figure 11: Q3(c) Reconstruction Error