

Assignment 3 Solutions

Ques 1.

(A)

Maxdiff = 0.405

(i) No scaling

The estimated regression matrix is **0.1514 0.6973 0.7970 -0.5943 0.9837 0.0325** and the maxdiff value is **0.405**. Singular values are obtained accordingly.

(ii) Scaling using the true standard deviation values (MLPCA)

Here the data was scaled using the true standard deviation values, [0.1, 0.08, 0.15, 0.2, 0.18].

The estimated regression matrix is **0.0058 0.9891 0.9748 -0.9503 0.9470 0.1059** and the maxdiff value is **0.1059**. The singular values obtained are **[237.69, 2.605, 0.9987, 0.9920, 0.9447]**.

It can be seen that the maxdiff value is much less than the case with no scaling and that the least m values ($m=3$ corresponding to number of dependent variables) are close to the value 1, indicating that the number of dependent variables are chosen correctly.

(B)

After applying IPCA, we get the following results

Singular values = **[239.3557 2.5577 1.0139 0.9995 0.9849]**

Maxdiff value assuming F3 and F5 as independent variables = **0.1087**

Variance-

0.0979

0.0845

0.1474

0.1878

0.1786

(C)

We can see from the singular values that the estimate of m is wrong. If we have chosen the value of m correctly, the least m singular values be close to 1 which is not the case here. But if we consider the previous case, we can notice that at least 3 singular values are close to 1, thus confirming the hypothesis that there are 3 independent variables

(D)

Large Condition Number corresponds to Singular Matrix

Low condition number corresponds to Invertible matrix

We want our matrix to have low condition number

Best case – F1, F2 as independent variables

Worst Case - F1, F5 as independent variables

Ques2.

(A)

The maximum absorbance for Co, Ni and Chromium are

510 406 394

RMSE = 0.0026, values around that have been marks too

(B)

It can be seen that the RMSE values reduce as the number of PCs increase. This may be because more principal components capture more information thus reducing the RMSE value. But we are not able to estimate the number of species correctly using the RMSE values, as they reduce in an almost constant trend.

(C)

Here we can see that the RMSE value suddenly drops after we take three principal components. We also note that the RMSE values are much lower than the case where the data was not scaled. We can conclude that the MLPCA can model the system much better than normal PCA and that at least three principal components must be taken to get a good result. Hence we were able to capture the right amount of pure species components (which is three) by conducting MLPCA but not using PCA

(D)

From RMSE clearly, we can say that the number of species is 3 because the slope changes drastically in the RMSE Plot.