

1st Exercise on Medical Image Processing

LU 183.630 – 2024 SoSe

Nora Roch, Julianna Mangold, Peter Sefcisk, Lukas Haider

Covariance matrix

The covariance was defined in the calculation as follows.

$$C_{ij} = \frac{1}{N-1} \sum_{n=1}^N (x_{i,n} - \bar{x})(x_{j,n} - \bar{x})$$

Figure 1 shows the 4 datasets, with the x coordinate of each point in blue and the y coordinate in red. There are 50 points in each data set.

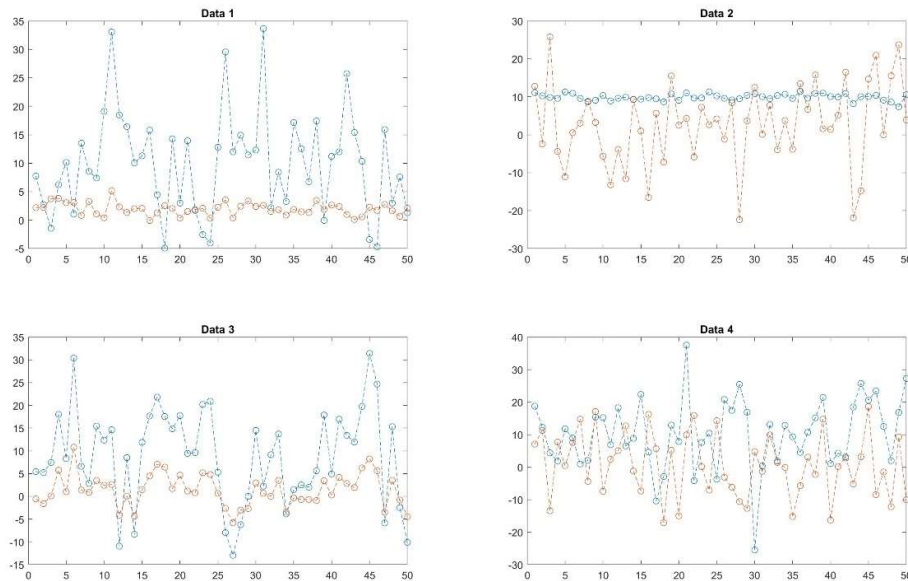


Figure 1 - Correlations in the Datasets

- Data1: There is almost no correlation between the red and the blue graphs, as shown by the correlation of 1.783, which is low in terms of variance. Furthermore, we can identify the blue graph with the higher variance (81,425) and the red graph with the lower (1,189).

$$C = \begin{bmatrix} 81,425 & 1,783 \\ 1,783 & 1,189 \end{bmatrix} = \begin{bmatrix} \text{Variance blue graph} & \text{Covariance red/blue graph} \\ \text{Covariance red/blue graph} & \text{Variance red graph} \end{bmatrix}$$

- Data2: This data set shows similar behaviour, except that y now has a higher variance.

$$C = \begin{bmatrix} 0,740 & 1,203 \\ 1,203 & 116,313 \end{bmatrix}$$

- Data3: In this correlation matrix we can observe a high correlation (36.101) in reference to the variance. This relationship can also be seen in the graph, as both graphs are decreasing/increasing together.

$$C = \begin{bmatrix} 110,493 & 36,101 \\ 36,101 & 12,753 \end{bmatrix}$$

- Data4: Once again, there is no significant correlation.

$$C = \begin{bmatrix} 115,843 & -4,522 \\ -4,522 & 96,587 \end{bmatrix}$$

PCA

In the following 2D plots it can be seen again that there is only a significant correlation in dataset 2.

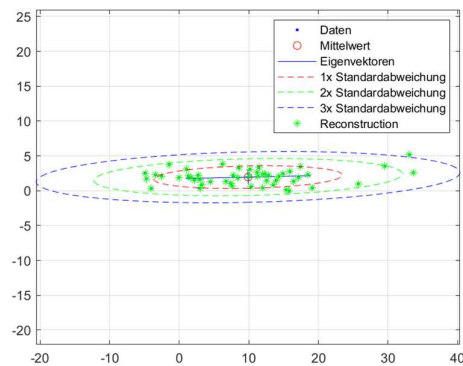


Figure 2 - PCA with Data1

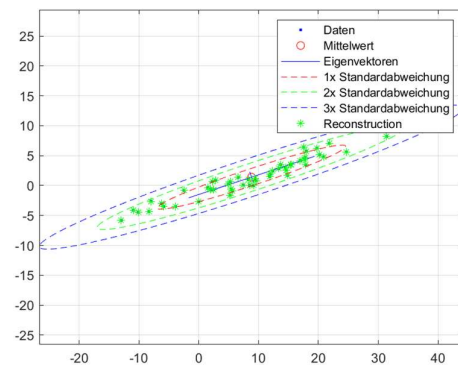


Figure 4 - PCA with Data3

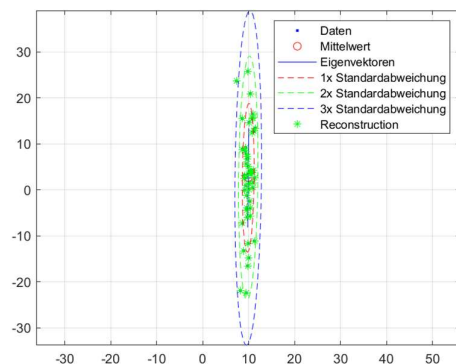


Figure 3 - PCA with Data2

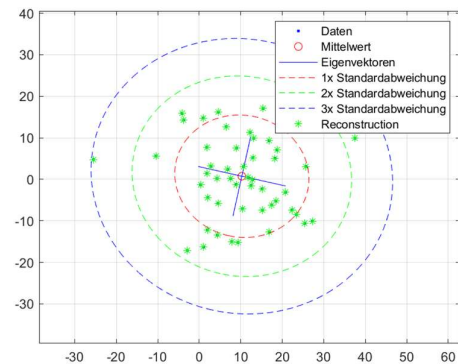


Figure 5 - PCA with Data4

The eigenvectors are represented by the blue lines in the figures. They represent the intrinsic value system. The length of these eigenvectors corresponds to the corresponding eigenvalues.

The eigenvalue system is fundamentally used to decouple the different variables, as there is no correlation in it and its components are therefore independent from each other. In such a system, the eigenvalues represent the variance to be expected. The total variance of the data can also be thought of as a Pythagorean high-dimensional triangle, and the total variance can be calculated by summing the individual variances.

Failure to subtract the mean will result in an incorrect covariance matrix and therefore an incorrect eigenvalue system.

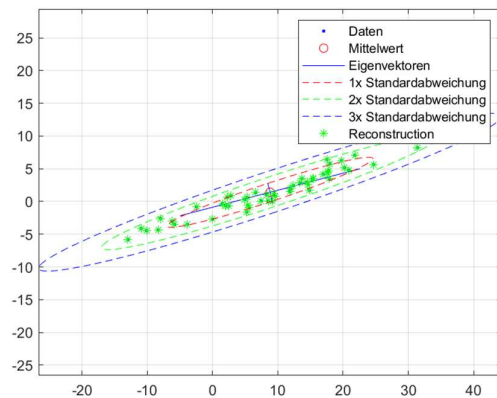


Figure 6 - Missing mean subtraction

Subspace projection

- By projecting the data on the main vector, we want to have the coordinates of these data on this main vector. The main vector is selected by using the highest eigenvalues. Here the first eigenvalue is the highest, so the main vector is the first column of the eigenvectors' matrix. Before the projection the data dimensionality was about 2×50 and now after projecting on the main vector it is reduced to 1×50 (1 row, 50 columns). The projection on the main vector is reconstructed in the initial data space and all points are on the vertical axis. The data points are projected on the vertical axis of the initial data space.

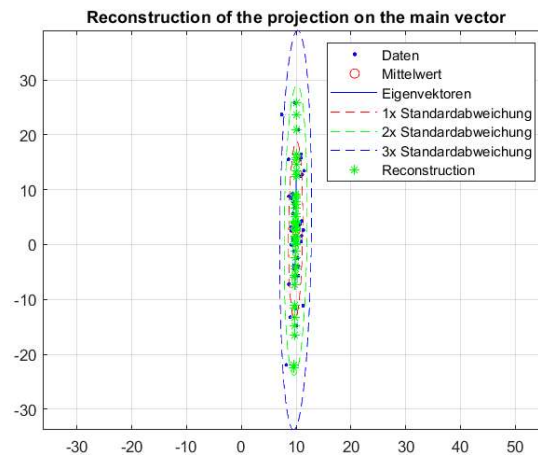


Figure 7 - Reconstructed data in the initial space from the projection on the main vector

To have the average error the difference of coordinates between each pair of points (one point from reconstructed associated to its original point from original data) is computed and next, the difference is divided by the number of points (i.e. the number of columns in the data). In this case the average error along the vertical axis (2nd component of each vector) is around zero because each pair of points (between original data and

reconstructed ones) has the same ordinate, the first component gives the average error. Here it is about 0.0126, so 1.26%.

If we look at the standard deviation using the difference between the reconstructed data and the original data, a standard deviation of ca. 0.853 is obtained for the first component (and near 0 for the second one which is logical).

- b. Instead of using the main vector here the side vector is used, that is to say the eigenvector corresponding to the lowest eigenvalue, so the 2nd vector (2nd column) of the eigenvectors' matrix. We proceed the same steps that the steps made for the main vector (question above a.). The same dimensionality as the projection on the main vector is obtained after projecting data on the side vector, so 1x50 (1 row, 50 columns).

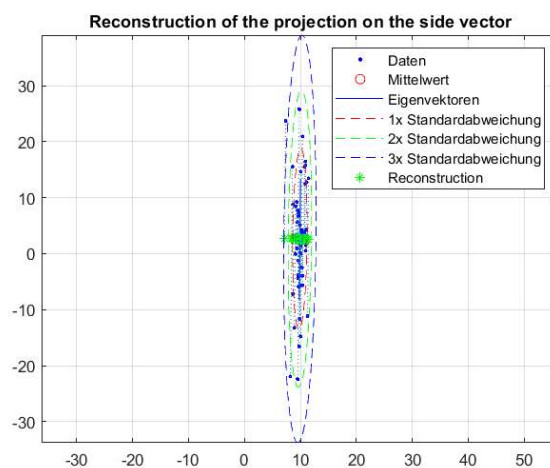


Figure 8 - Reconstructed data in the initial space from the projection on the side vector

In this case the average error along the horizontal axis (1st component of each vector) is around zero (0.003) because each pair of points (between original data and reconstructed ones) has the same abscissa, the second component gives the average error. Here it is about 0.0254, so 2.54%.

If we look at the standard deviation using the difference between the reconstructed data and the original data, a standard deviation of ca. 10.7848 is obtained for the second component (and near 0 (0.1123) for the first one which is logical).

If we compare the projection on the side vector with the projection on the main vector, it is observed that the average error is less important, lower for the data reconstructed by projecting on the main vector than the average error obtained for reconstructed data by projecting on the side vector. Same observation for the standard deviation.

Consequently, it is better to use the 1st vector of the eigenvectors' matrix (1st column corresponding in the highest eigenvalue) to have the reconstructed data with the minimal error.

Investigation in 3D

- a. Relation between covariance matrix, eigenvalues, eigenvectors and the ellipsoidal of the standard deviation:

The covariance matrix can be used to calculate the eigenvalues and eigenvectors of the data. The eigenvectors represent the directions of maximum variance, aligning with the principal axes of the ellipsoid. The extent of the spread of the ellipsoid in each direction is determined by the square root of the corresponding eigenvalues.

- b. What is the dimensionality of the projected data? What type of information has been lost after reconstruction?

After projecting the data in the subspace constructed by the first two eigenvectors, the data becomes 2 dimensional.

Reconstructing these points into the original space results in the loss of the variance of the points along the third axis defined by the third eigenvector.

Because we sort the eigenvalues in a descending order, the lost information is along the least significant axis corresponding to the smallest eigenvalue.

Plot of original and reconstructed data:

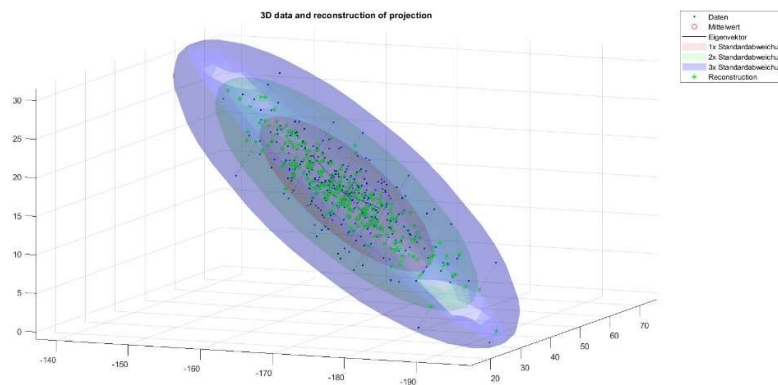


Figure 9 - 3D Reconstruction

Shape modeling

In the following figures, the vector b was varied, and new shapes were generated. Figure 10 shows a variation of one time the standard deviation of the first eigenvector of the pca, where the standard deviation was implicitly assumed to be the root of the corresponding eigenvalue. The somewhat unexpected effect that the new shape is smaller than the mean can be explained by the fact that the pca randomly decides in which direction the vector is considered positive when calculating the eigenvector.

In the other Figure 11 and Figure 12 the deviation was increased to 3 times the standard deviation, resulting in a larger deviation of the new shapes. The sign behaviour of the deviation can also be seen.

Finally, a random distribution was applied to the eigenvectors with the greatest influence on the total variance. Therefore, Figure 13 shows the distribution with 100% of the total variance, Figure 14 with 95%, Figure 15 with 90% and Figure 16 with 80% total variance. Reducing the total variance results in a better new shape towards the mean.

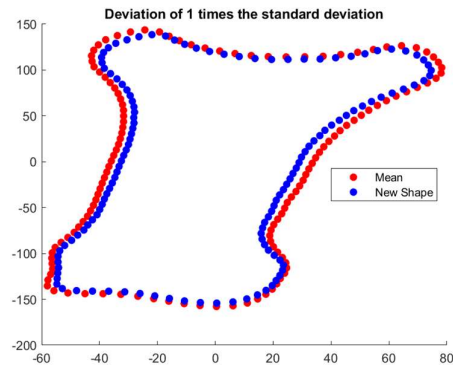


Figure 10 - Deviation of 1 times the standard deviation in the main vector

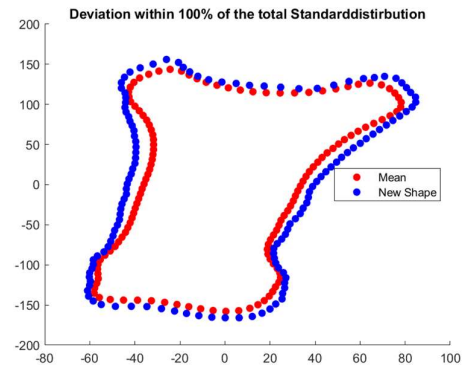


Figure 13 - Normal distribution with 100% of the total variance

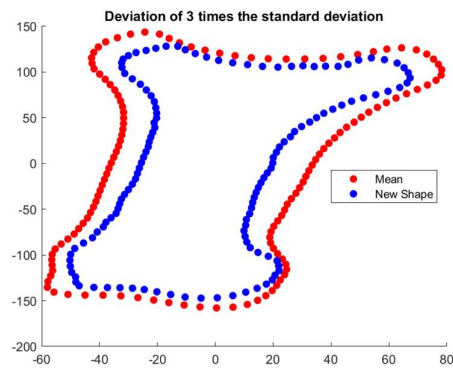


Figure 11- Deviation of 3 times the standard deviation in the main vector

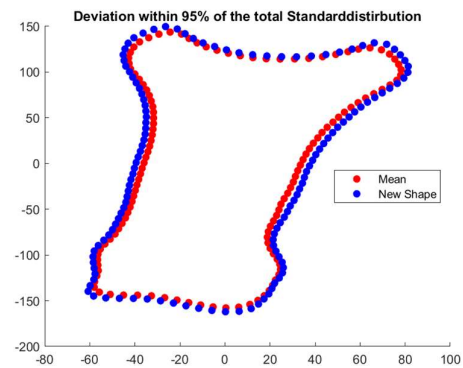


Figure 14 - Normal distribution with 95% of the total variance

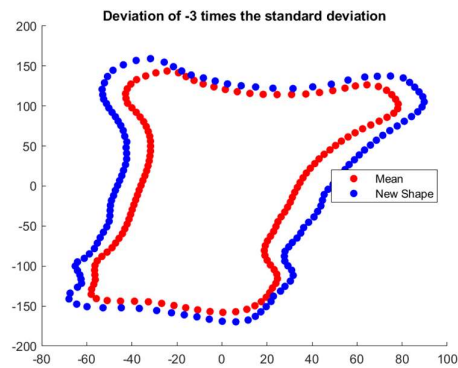


Figure 12 - Deviation of -3 times the standard deviation in the main vector

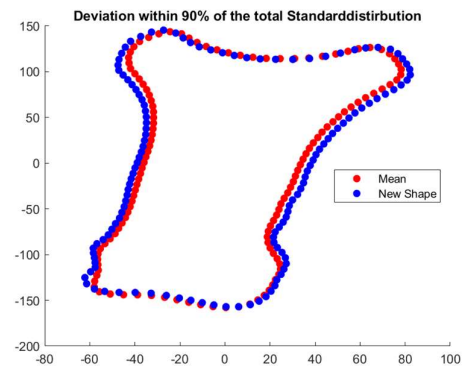


Figure 15 - Normal distribution with 90% of the total variance

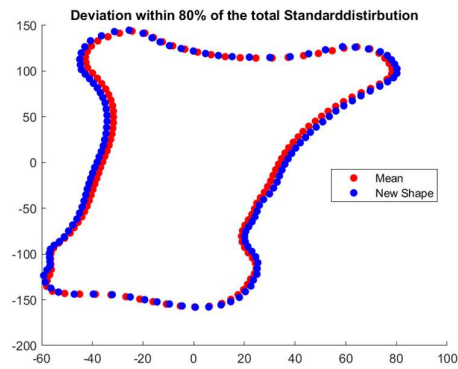


Figure 16 - Normal distribution with 80% of the total variance

List of illustrations

Figure 1 - Correlations in the Datasets	1
Figure 2 - PCA with Data1	2
Figure 3 - PCA with Data2	2
Figure 4 - PCA with Data3	2
Figure 5 - PCA with Data4	2
Figure 6 - Missing mean subtraction	3
Figure 7 - Reconstructed data in the initial space from the projection on the main vector	3
Figure 8 - Reconstructed data in the initial space from the projection on the side vector	4
Figure 9 - 3D Reconstruction	5
Figure 10 - Deviation of 1 times the standard deviation in the main vector.....	6
Figure 11 - Deviation of 3 times the standard deviation in the main vector.....	6
Figure 12 - Deviation of -3 times the standard deviation in the main vector.....	6
Figure 13 - Normal distribution with 100% of the total variance	6
Figure 14 - Normal distribution with 95% of the total variance	6
Figure 15 - Normal distribution with 90% of the total variance	6
Figure 16 - Normal distribution with 80% of the total variance	7