Exercise 1 of Medical Image Processing SS2014

By: Dr. René Donner

Team: Stephan Spindler 0828742, S. Ehsan Hosseini 1028660,

---

1. Covariance Matrix:

a. <u>ourCov.m</u> , <u>Run.m</u>

b. Welche Informationen stehen an welcher Stelle von C?

Variance and covariance matrix is a symmetric n x n matrix that the diagonal elements show the variance of Dataset and the off-diagonal elements show the covariance of Dataset. We calculate the covariance matrix C between the different variables in Data set to measure the relation and show the relationship between the Variables by:

- Mean of the data X and Y
- Deviation sums of squares and cross product matrix
- The number of scores in each column of the original matrix

$$\begin{bmatrix} V_A & \mathrm{cov}_A \\ \mathrm{cov}_A & V_A \end{bmatrix}$$

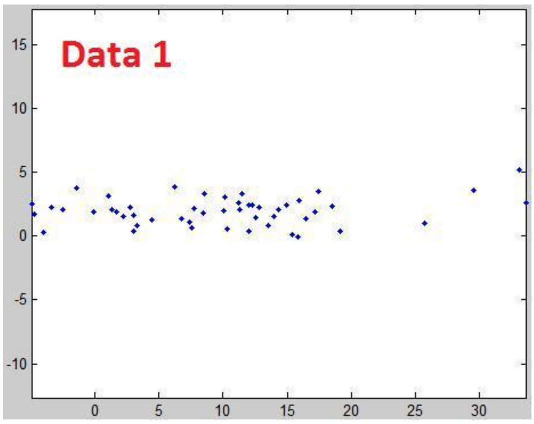
<u>Interpret the results</u>: (in linear situation)

- Sign of Covariance: (nature of relationship)
  If positive then X and Y move in the same direction. (Positive correlation)
  If negative then X and Y move in the opposite directions. (Negative correlation)

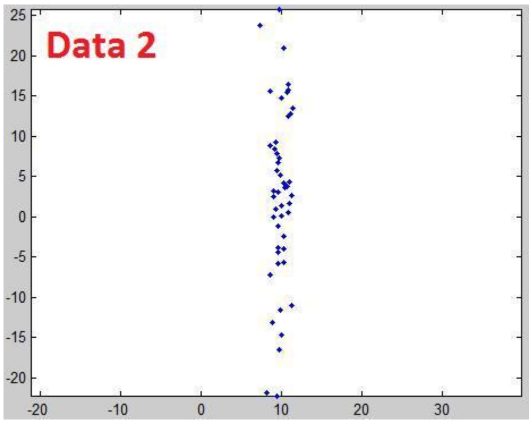
- Size of Covariance: (strength of relationship)
  If large, then there is a strong relationship (correlation).
  If small, there is a weak or no relationship (correlation).

Interpretieren Sie die unterschiedlichen C zwischen den Datensets!

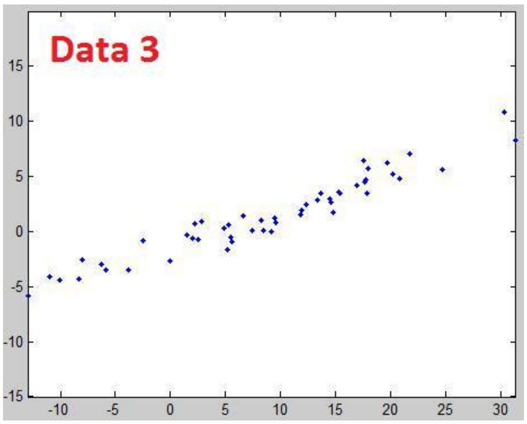
- Covariance = 1.7825 means a positive relationship between variables of Dataset 1.
  The variables of dataset 1 are correlated but not really high and non-linear.
- The data are around the Y-mean value. (2,5)
- Variance of 81.4252 (X) shows that the variables of these dataset are more variable than the Dataset with variance of 1.1887 (Y)



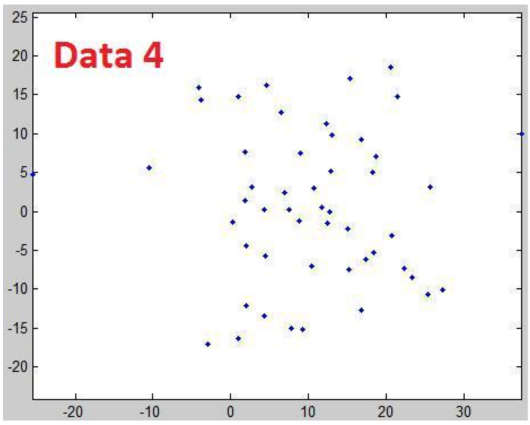
- Data in dataset 2 are highly correlated with the positive value of covariance and they are around the X-mean value. (10)
- The scatter plot shows a non-linear of data set with a small variance 0.7404 for the X-values.
- In this dataset the Y-values are more variable than the X-values.



- The highest value of covariance among the all datasets is in this case. The value of 36.1014 shows a positive correlation and of course a linear relationship between the variables.



- In this dataset both of the variances are too high and it means the variables are too far from the mean value.
- A negative covariance shows an un-correlated data in dataset.



2. PCA:

a. Plotten Sie mit <u>plot2DPCA.m</u> Ihre Ergebnisse für die Daten aus <u>daten.mat</u>.
b. Was geben die Eigenvektoren an? Wo sieht man das im Plot?

- Eigenvectors are used to obtain a more stable set of descriptors for viewing the data. The eigenvectors of the covariance matrix constitute the principal components to decreasing systematic variation, and increasing non-systematic variation (noise). If the number of selected eigenvectors is adequate, the data can be reconstructed from the chosen set of eigenvectors.
- To see the eigenvectors we are looking for orthogonal vectors that cut each other around the mean-value so we can see them clearly in plot 4 from Dataset 4.

c. **Was geben die Eigenwerte an? Wo sieht man das im Plot? In welcher Relation stehen sie zur Gesamtvarianz?**

- Eigenvalue determine the arrangement of the eigenvectors. Therefore, the most significant eigenvector represents the main component and the first two components (with the two highest eigenvalues) represent almost 98% of the total variance.
- In the plot, the length of the eigenvectors is defined by the eigenvalues. We can see in plot some ellipses those are defined by the PCA cloud of points. The ellipses are located around the eigenvectors and their size is defined by the eigenvalues.
- Therefore, the eigenvectors with the smallest eigenvalues are in the area of the smallest variances.

d. **Welchen Einfluss hat ein fehlender Mittelwertabzug (bei D) auf die Berechnung?**

- Since the eigenvalues derived from the covariance matrix which based on a fracture of the mean-value, the eigenvectors and eigenvalues could not put the ellipse exact and ideal on the data cloud.
- Therefore, the eigenvectors stretch an orthonormal basis which is not on the actual basis of the data.

3. Unterraum-Projektion:

a. **Welche Dimension haben Ihre Daten jetzt?**

1-D
Existing data set in d-dim. A space with n points by Projection in q-dim.
Space with d > q.

Beschreiben Sie den Effekt von Projektion und Rekonstruktion auf die Datenpunkte.

- The projection is used for dimension reduction -as projected onto the first principal component- and compression of the data points.
- The reconstruction will multiply the compressed data points with the transposed eigenvector and will add up the previously subtracted mean values again.
- The reconstructed data points are now on the main vector.

Wie groß ist der Durchschnittliche Fehler zwischen Rekonstruktion und Originaldaten?

- The average error corresponds to the variance of the original data and reconstruction or the sum of the unused eigenvalues respectively.

b. **Welche Eigenvektoren werden Sie verwenden, um eine Datenmatrix mit möglichst wenig Fehler mit möglichst wenig Eigenvektoren (in diesem Fall 1) darzustellen?**

- The eigenvectors with the highest eigenvalues are used because they include the most of the information. The smaller the eigenvalues, the less loss of information. Thus, in a 2x2 matrix choose the first major vector.

4. Untersuchungen in 3D:

a. **Beschreiben Sie die Relation von Kovarianz Matrix (Varianzen), Eigenwerten und Eigenvektoren und den Ellipsoiden der Standardabweichungen.**

- The boundaries of the ellipsoids of the standard deviations are formed from the roots of the variances, which form the diagonal of the covariance matrix. The sum of the eigenvalues formed the total variance.
- For more information please see the answers of 2nd question.

b. **Projizieren Sie auf den Unterraum, der durch die ersten beiden Eigenvektoren aufgespannt wird. Welche Dimension haben Ihre Daten?**

- Projection of R^d to R^q with d = 3 & q = 2 => R^2, since q is stretched by Lambda 1 and 2 (2 main components).

Rekonstruieren Sie die Punkte im Originalraum und plotten Sie das Ergebnis. Welche Information ist verloren gegangen?

- The variance of the 3rd eigenvalue which includes ~ 2.5% of the total variance.

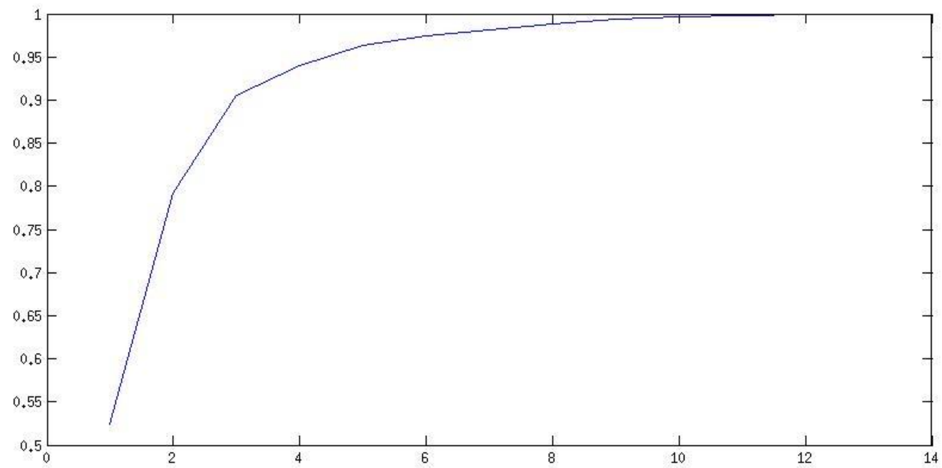
5. Shape Modell:

a. **<u>generateshape.m</u> , <u>Run.m</u>**
b. **Schreiben Sie eine Funktion <u>plotShape</u>, die die Shapes in blau darstellt und plotten und interpretieren Sie die Einzelnen Modes (d.h. b ist 0 bis auf einen Wert) im Bereich von ±3λ, wobei λ die Standardabweichung des entprechenden Modes bezeichnet. Die Funktion soll gleichzeitig auch das mean shape (d.h. b gleich dem Nullvektor) in rot darstellen. Beschreiben und interpretieren Sie.**

- For each of the 14 models are plotted n shape contours, corresponding to the deviations from the mean shape in the range ± 3λ. By plotting the eigenvector matrix (imagesc (VECs); figure (gcf)) is seen, what relevance have the first 14 eigenvectors over the rest.
- If we set the vector b = 0, the result is the mean shape, since shape = eigenvectors * 0 + mean shape = mean shape.

c. **Setzen Sie nun <u>b=randn (1,nEigenvectors).*stddeviations</u>. Beschränken Sie nun wie in den 2D und 3D Beispielen die Zahl der Eigenvektoren, dementsprechend die Länge von b, plotten Sie die resultierenden Shapes und interpretieren Sie. Beschränken Sie so, dass das Shape Modell 100%, 95%, 90% und 80% der Gesamtvarianz beinhaltet.**

- The vector b is depending on the length and the chosen λ-range filled with random numbers from the respective pooled standard deviation.
- As an approach to a λ for a total variance of 100% runs to infinity, the values from around the 10th components have 100% (at least a 99.99…%)



On the X axis is the index of the component or eigenvalue, the Y-Axis contains the cumulative variance.