Exercise 1 30.03.21

Show that the correlation matrix using Pearson's correlation coefficient is identical to a covariance matrix computed from standarised data.

given: 2 random variables $x \in \mathbb{R}^n$ and $y = \mathbb{R}^n$, n = number of samples

Pearson's correlation coefficient $r = \dfrac{\frac{1}{n-1}\sum_{i=1}^{n}[(x_i-\mu_x)(y_i-\mu_y)]}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_x)^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\mu_y)^2}}$

where μ_x and μ_y are the means of x and y

x and y have been standarised before the covariance matrix was computet

$x' = \dfrac{x-\mu_x}{\sigma_x}$ $\quad y' = \dfrac{y-\mu_y}{\sigma_y}$ where σ_x and σ_y are the standard deviations of x and y

covariance matrix $\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y)$

answer

$$r = \frac{\frac{1}{n-1}\sum_{i=1}^{n}[(x_i-\mu_x)(y_i-\mu_y)]}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_x)^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\mu_y)^2}} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

$$\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y)$$

$$\sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_x)^2}$$

$$\sigma_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\mu_y)^2}$$

standarized data x' and y' (mean-centered)

$$\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x'-0)(y'-0)$$

insert definition of x' and y'

$$\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)$$

simplify

$$\sigma_{xy} = \frac{1}{n-1}\frac{1}{\sigma_x\sigma_y}\sum_{i=1}^{n}(x-\mu_x)(y-\mu_y) = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

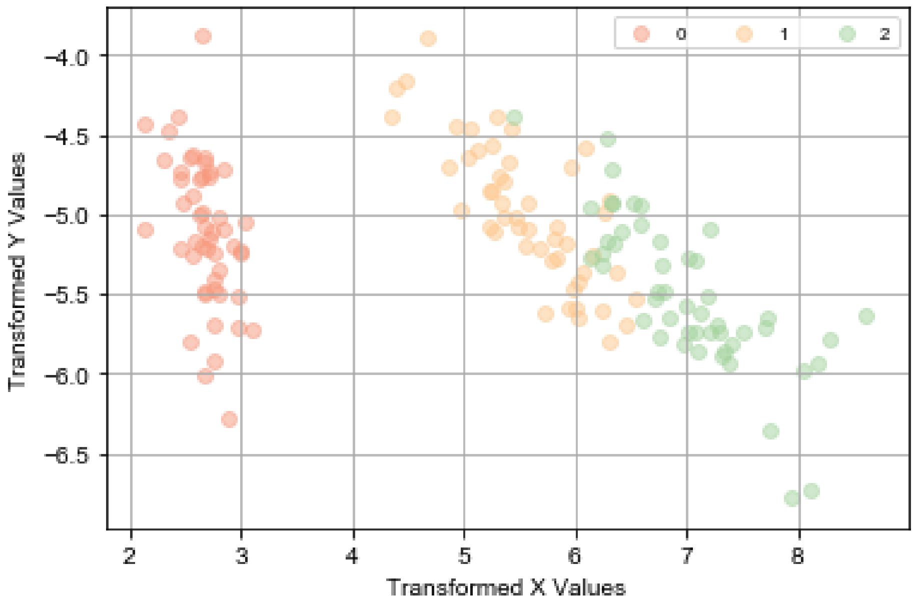
Exercise 1 30.03.21

Deviding the covariance matrix by the product of the standard deviation of x and y scales the values down to a limited range from -1 to +1 which is the range of the correlation matrix

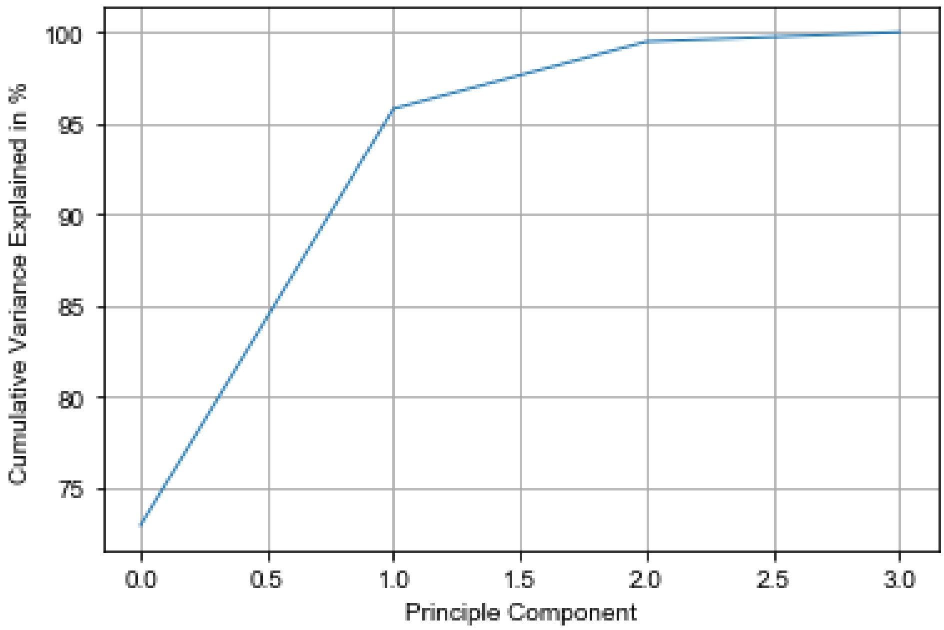
Exercise 2a

PCA on Iris data, plot of transformed data with highlighted samples according to their classes.

Input data projected onto a 2 dimensional subspace



Cumulative plot of variance explained values per principal component



Variance Explained PCA:
PC 1: 0.73
PC 2: 0.23
PC 3: 0.04
PC 4: 0.01

Exercise 2b 30.03.21

Pseudocode for PCA using SVD

Require: A matrix X of n examples $\in \mathbb{R}^{d \times n}$

1.) Normalisation of data X by substracting the mean

for $i = 1, \dots, n$ set $x_i \leftarrow x_i - \frac{1}{n} \sum_{i=1}^{n} x_i$

2.) Computation of singular value decomposition (SVD) from normalised data

$X = L \Delta R^T$

where L is an $n \times n$ matrix containing the left singular vectors
Δ is an $n \times d$ diagonal matrix containing the singular values
R is an $d \times d$ matrix containing the right singular vectors

3.) Computation of eigenvalues and eigenvectors of SVD

the columns of matrix L, which are also the left singular vectors, are the orthonormal eigenvectors of XX^T

$$XX^T = L\Delta(R^T R)\Delta^T L^T = L\Delta\Delta^T L$$

The diagonal entries of $\Delta\Delta^T$, that is, the squared nonzero singular values, are the non-zero eigenvalues of XX^T

→ The squared singular values of SVD are n times the eigenvalues of PCA.

→ The eigenvectors of PCA are the same as the left singular vectors of SVD for mean-centered data.

4

Exercise 2b 30.03.21

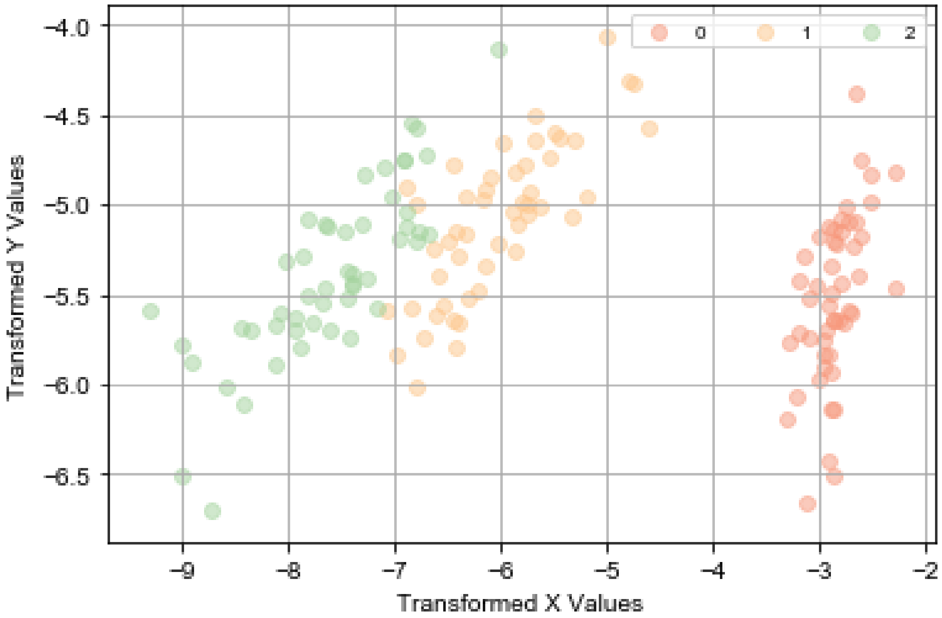
4.) Transformation of input data X onto 2-dimensional subspace using the first 2 PCs inferred through SVD.

5.) Plotting transformed data

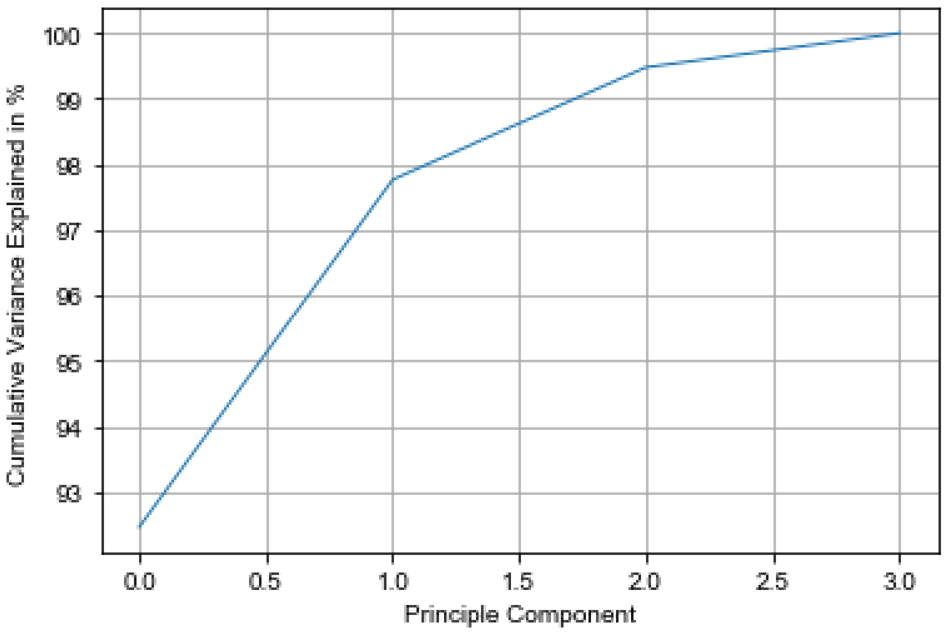
Exercise 2d

SVD based PCA on Iris data, plot of transformed data with highlighted samples according to their classes.

Input data projected onto 2 dimensional subspace.



Cumulative Plot of variance explained values per principal component



Variance Explained SVD:

PC 1: 0.92

PC 2: 0.05

PC 3: 0.02

PC 4: 0.01

Exercise 2d

Comparison of the results from PCA and PCA using SVD.

For both approaches, the first two PCs help to separate the three classes. Overall, the classes are not perfectly linearly separable. Especially, for cluster yellow and green (class 1 and 2) there are points on the wrong side of a linear decision boundary. Class 1 and 2 cluster closely together while class 0 is isolated.

By looking at the Variance explained of PCA using SVD, the first PC explains 92 % of the variance while the first PC of PCA explains 73 % of the variance. However, in order to explain at least 95 % of the variance, the first two PCs of both approaches are required.

Exercise 2e

The construction of the covariance matrix lies in $O(d^2n)$ and calculating the eigenvalues of the covariance matrix lies in $O(d^3)$ where d is the number of features and n the number of samples. Thus, the standard runtime is very expensive for large d. An advantage of PCA using SVD over PCA using eigen-value decomposition is that one can omit the calculation of the covariance matrix which is computationally hard for large d.