
Practice 10

EXERCISE 1: DIMENSIONALITY REDUCTION (PRINCIPAL COMPONENT ANALYSIS - PCA)

- a) Assign the seed of the rand and randn function to 0.
- b) Generate 1000 random data from a normal 2D centered at [3,4] and covariance matrix [1 0.8 0.8 1], using the function of the pattern library *randNorm*.
- c) Subtract the average for each data item, that are centered at (0,0).
- d) Calculate the conversion matrix PCA using $W = \text{pca}(x)$;
- e) Draw the original data and the converted, calculated as $x1 = W * x$;
- f) The data can be reset using $x2 = \text{pinv}(W) * x1$; Draw the recomposed data and compare it with the originals.
- g) Evaluate the loss of information (using *sumsqr*) by projecting the data in the first dimension of 2D space converted by PCA, and evaluate the loss information to reconstruct the data.
- h) Modify W slightly and evaluate the loss of information obtained. Is it always higher?

EXERCISE 2: DIMENSIONALITY REDUCTION (FISHER METHOD)

- a) Assign the seed of the rand and randn function to 0.
- b) Generate 2 classes:
 - o CLASS 1: 1000 random data from a normal 2D centered on [3,4] and covariance matrix [1 0.8 0.8 1]
 - o CLASS 2: 1000 random data from a normal 2D centered at [5.0] and covariance matrix [1 0.8 0.8 1]
- c) Calculate the matrix of PCA, convert to PCA space the original data and draw the converted data
- d) Calculate the matrix of Fisher, convert to FISHER space the original data and draw the converted data

EXERCISE 3: DIMENSIONALITY REDUCTION (DRAWING DATA)

- a) Use the DNATRAN Statlog data base.
 - Display data using PCA and FISHER, projected onto two dimensions
 - What method is most suitable for classification and / or display of information?
- b) Use the data base DIABETES Statlog data base.
 - You want to predict the second component from the others. To do this do the following:
 - o Reduce the dimension of 8-2
 - o Calculate the linear model
 - o Determine the error by applying the linear model

- Perform the above steps twice, one reducing the dimension with PCA and another with Fisher. Which is better?