

### EXERCISE 1: PRINCIPAL COMPONENT ANALYSIS

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

### EXERCISE 2: FISHER METHOD

- Assign the seed of the rand and randn function to 0.
- Generate 2 classes:
  - CLASS 1: 1000 random data from a normal 2D centered on [3,4] and covariance matrix  $\begin{bmatrix} 1 & 0.8 & 0.8 & 1 \end{bmatrix}$
  - CLASS 2: 1000 random data from a normal 2D centered at [5.0] and covariance matrix  $\begin{bmatrix} 1 & 0.8 & 0.8 & 1 \end{bmatrix}$
- Calculate the matrix of PCA, convert to PCA space the original data and draw the converted data
- Calculate the matrix of Fisher, convert to FISHER space the original data and draw the converted data

### EXERCISE 3: 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:
    - Reduce the dimension of 8-2
    - Calculate the linear model
    - Determine the error by applying the linear model
- c) Perform the above steps twice, one reducing the dimension with PCA and another with Fisher. Which is better?