INF 4300 2016
Classification 3

Exercise 1.

You are given data from two classes with means and covariance as given below:

$$\mu_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix} \quad \Sigma_1^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \Sigma_1^{-1} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

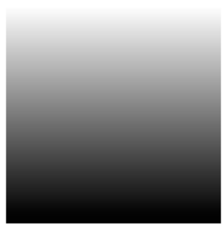
a) Compute the eigenvectors and eigenvalues of the covariance matrices (use Matlab if you want) and use them to sketch the ellipses for the covariance matrices in a plot.

b) Show that the decision boundary (g1(x)=g2(x)) in this case with features x_1 and x_2 can be expressed as $x_2=3.514-1.125x_1+0.1875x_1^2$

c) Plot the resulting decision boundary in Matlab.

d) Create a synthetic image with 2 bands with samples that span the feature space from e.g. -10 to 10 for both features. For simplicity, let us just consider a coarse grid of samples on integer values (-10,-9,….,0,….9,10). Feature image 1 should look like a horisontal ramp from -10 to 10, and feature image 2 like a vertical ramp from 10 to -10:


Feature 1


Feature 2

This corresponds to creating feature vectors that span the entire feature space (from -10 to 10). If we later classify all these feature vectors, the resulting classification map should have the same decision boundary as the plot we computed in b) and plotted c). This is just a way to create a visualization of the decision boundary without computing it analytically.

e) Classify this image, and verify that the shape of the decision boundary you got in c) is the same as you get after classifying this image.

Exercise 2: Implement a kNN-classifier and test it on the Landsat TM-image from last week. Experiment with K, and compare the classification accuracy to a Gaussian classifier.

The following exercises are from Exam 2015

Exercise 3: Classification

a) Consider a two-dimensional feature vector and a set of points in 2D feature space:
(-3,6) (-2,4) (-1, 2) (0,0) (1, -2) (2, -4) (3, -6)
Show that the covariance matrix between the two features is :

$$\Sigma = \begin{bmatrix} 4 & -8 \\ -8 & 16 \end{bmatrix}$$

Show all your calculations.

b) Given the two features defined above, would you base you classification on 1 or 2 features? Justify your answer.

c) The discriminant functions for a multivariate Gaussian classifier are given as:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

Consider two classes with equal prior probability and

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \Sigma_1^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \quad \Sigma_2^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \quad |\Sigma_1| = 1 \quad |\Sigma_2| = 1$$

Can the discriminant function be simplified in this case?

d) Classify the point $x = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ by computing the value of the discriminant functions and assign it to the class corresponding to the highest probability.

e) Explain how classifier sensitivity and specitivity are computed, and discuss their importance for a medical classification problem.

For the PhD-students only:

f) Let us assume that we have a 2-class classification problem with a 1-dimensional feature vector f(x) which is exponentially distributed given the class-conditional parameter λ_i:

$$f(x) = \lambda_i \exp^{-\lambda_i x}$$

Find an expression for the decision boundary for this classification problem.

Exercise 4: Clustering

a) Describe how the K-means clustering algorithm works and which parameters it has.

b) K-means clustering with K=2 is done on the data points given in the scatter plot below. Three different strategies for initializing the clustering is to assign the initial cluster centers to
   a. minimum/maximum among the points in the data set
   b. the first K points
   c. K random data points.

   In the table below the resulting cluster centers after each iteration in K- means clustering are given for two of these methods.

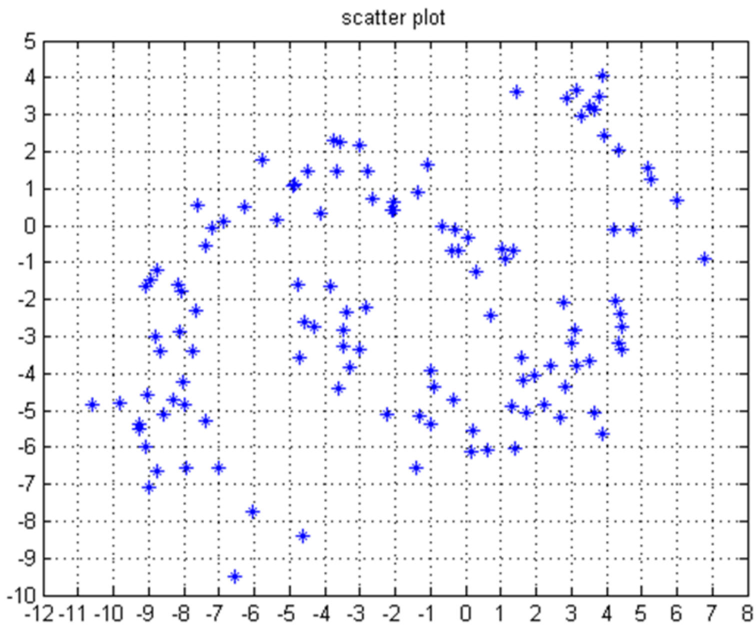
   Discuss which of these two methods you would choose for this particular data set, and how different the clustering result would be.

Initialization	Cluster 1 mean	Cluster 2 mean
Iteration 0	(-10.6, -4.8)	(6.8, -0.9)
Iteration 1	(-6.4, -2.9)	(1.8, -1.4)
Iteration 2	(-6.4, -2.9)	(1.7, -1.5)
Iteration 3	(-6.4, -2.9)	(1.7, -1.5)
Iteration 4	(-6.4, -2.9)	(1.7, -1.5)
Iteration 5	(-6.4, -2.9)	(1.7, -1.5)
Iteration 6	(-6.4, -2.9)	(1.7, -1.5)

Table 1 Clustering, initialization using minmax points

Initialization	Cluster 1 mean	Cluster 2 mean
Iteration 0	(1.6, -3.6)	(-1.3, -5.1)
Iteration 1	(2.2, -0.9)	(-5.6, -3.1)
Iteration 2	(2.1, -1.5)	(-6.0, -2.8)
Iteration 3	(2.0, -1.6)	(-6.2, -2.7)
Iteration 4	(1.9, -1.7)	(-6.3, -2.6)
Iteration 5	(1.9, -1.7)	(-6.3, -2.6)
Iteration 6	(1.9, -1.7)	(-6.4, -2.6)

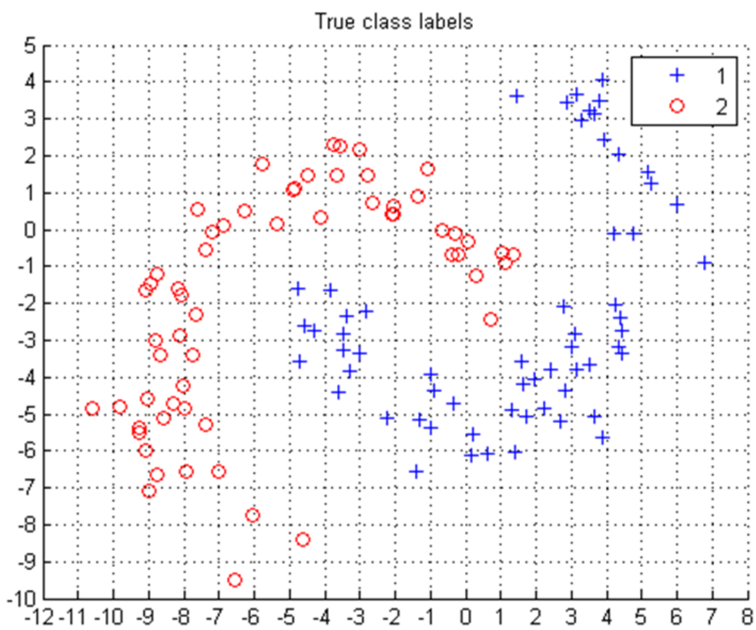
Table 2 - Cluster means after initialization by the first points in the data set



c) Select one of the initializations and indicate the decision boundaries the clustering would result in on the scatter plot in enclosure 1.

d) This data originates from a classification data set with known class labels. A scatter plot with class labels is given in the figure below.

Plot your estimated decision boundary from the clustering on the labelled scatter plot. Use the scatter plot to compute the confusion matrix for the clustering result.



e) Discuss if a Gaussian classifier with full class-conditional covariance matrix would perform well on this data set given the known class labels.