Previous exam exercises on classification

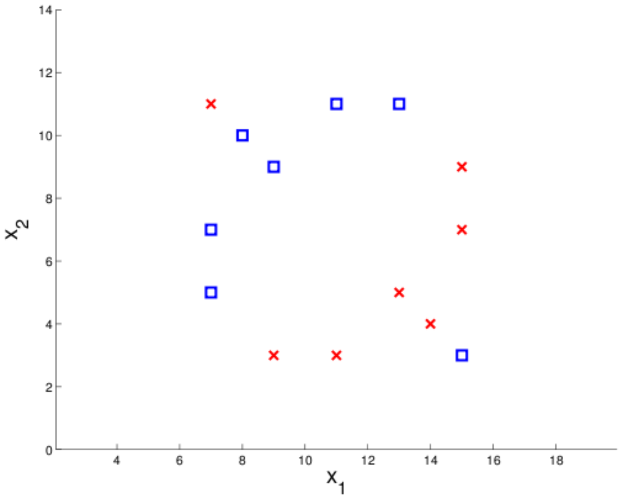
Exercise 4 2012: Classification with 2 features

You are given feature vectors $\mathbf{x}_1$ from class ω_1 and feature vectors $\mathbf{x}_2$ from class ω2.
The training set consists of the following points:
Class 1 points: {(11,11), (13,11), (8,10), (9,9), (7,7), (7,5), (15,3)}
Class 2 points: {(7,11), (15,9), (15,7), (13,5), (14,4), (9,3), (11,3)}

$$\mu_1 = \begin{bmatrix} 10 \\ 8 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 9.67 & -1.0 \\ -1.0 & 9.67 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 12 \\ 6 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 9.67 & -1.0 \\ -1.0 & 9.67 \end{bmatrix}$$

A scatter plot of the features are given below:



a) Are the two classes linearly separable? Please explain.

b) In the scatter plot given in the enclosure, please indicate the decision boundary that you would get using a Gaussian classifier with equal diagonal covariance matrices. Make sure to label the regions with the proper class label.

c) In the same scatter plot, also sketch the decision boundary that you would get using a 1NN classifier.

d) Classify sample (6,11) using the Gaussian classifier.

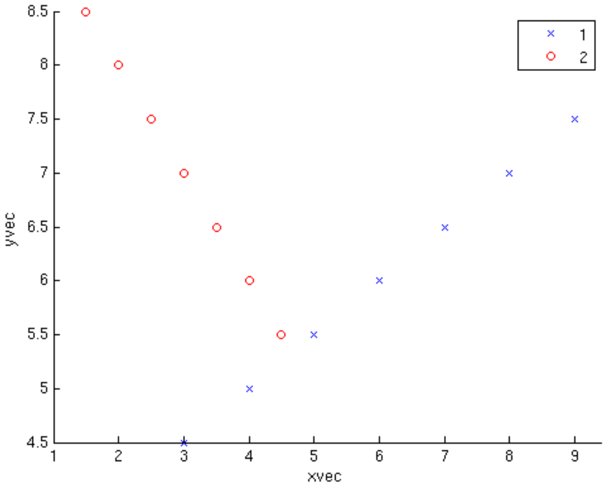
e) Classify sample (14,3) using the nearest neighbor classifier.

f) Which of the two classifiers do you think is most robust? Justify your answer.

Exercise 5 2012: Classification

a) Explain briefly the three special cases we have for a Gaussian classifier.

b) A scatter plot of two features for two classes is given.



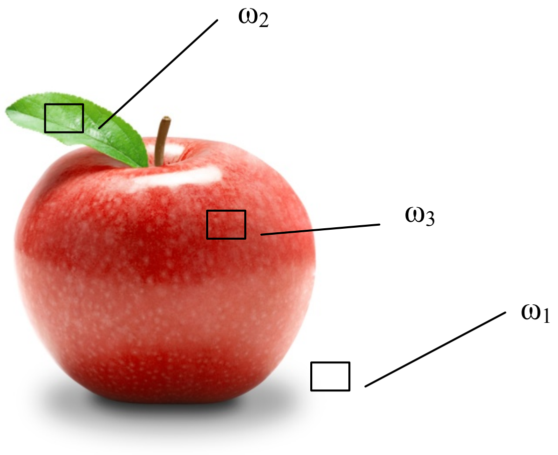
The statistics class statistics are:

$$\mu_1 = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 4.667 & 2.333 \\ 2.333 & 1.167 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 3 \\ 7 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 1.167 & -1.167 \\ -1.167 & 1.167 \end{bmatrix}$$

On this dataset, would you use a linear or a quadratic classifier? Justify your answer.

Exercise 5 2009 . Classification using Bayes rule



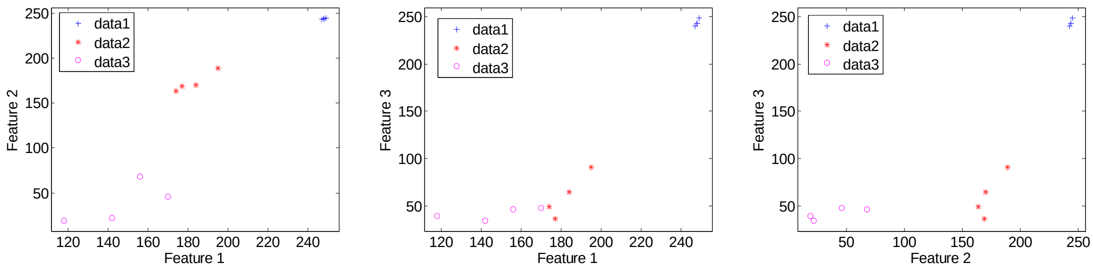
pixel	class	channel 1	channel 2	channel 3
1	ω1	248	244	243
2	ω1	247	243	240
3	ω1	248	244	243
4	ω1	249	245	249
MEAN	ω1	248	244	243.75
VAR	ω1	0.67	0.67	14.25
5	ω2	174	164	49
6	ω2	177	169	36
7	ω2	184	170	65
8	ω2	195	189	91
MEAN	ω2	182.5	173	60.25
VAR	ω2	87	120.67	560.92
9	ω3	170	46	48
10	ω3	156	68	46
11	ω3	142	22	34
12	ω3	118	19	39
MEAN	ω3	146.5	38.75	41.75
VAR	ω3	491.67	526.25	41.58

Figur 1: Training image

For the multispectral image (3 channels) above, we have the training data picked from the indicated regions as given in the table

Your task is to classify the pixels in the image below into class ω1 (background), ω2 (leaf), and ω3 (apple) based on the information from the image in Figur 1. You are only allowed to use two channels.

a) Evaluate by visual inspection (using 2D feature spaces) which two channels are the most suited.



b) Based on this you are designing a classifier using Bayes rule. The data is modeled with gaussian distributions having the same covariance matrix $\mathbf{\Sigma}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$. Write the generic form of the discriminant function.

c) Classify the pixels below using the functions as in (b). Note that just giving the class for the pixels will not give any score even if correct. Whether you answer the classification problem graphically or numerically is equivalent.

pixel	channel 1	channel 2	channel 3
1	200	45	50
2	169	136	55
3	231	218	201
4	203	176	131

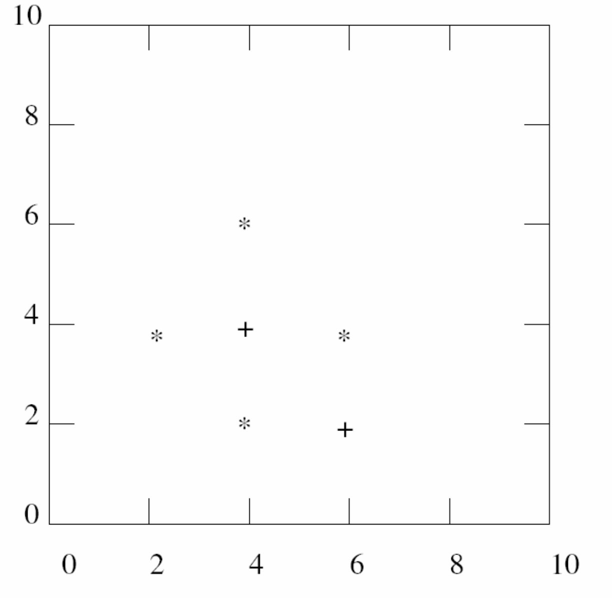
Remember that, when assuming a 2D Gaussian distribution, the point probability of a point $x=[x_1, x_2]^T$ can be written on the form

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \frac{1}{(2\pi)^{1/2} \, |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \right)^T \Sigma^{-1} \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \right) \right)$$

where the probability distribution function takes as parameters the vector $\mu=[\mu_1, \mu_2]^T$ and a matrix $\sum$.

Exercise 6 2009. K-nearest neighbor classification

In the following questions you will consider a *k*-nearest neighbor classifier using Euclidean distance metric on a binary classification task. We assign the class of the test point to be the class of the majority of the *k* nearest neighbors. Note that a point can be its own neighbor.



Figur 2: k-nn classification task, a plus indicates a positive example and a star a negative example.

a) In the Figur 2, sketch the *1*-nearest neighbor decision boundary for this dataset.

b) How would the point (8,1) be classified using 1-nn?

Exercise 5 2008. Classification using Bayes rule

Remember that, when assuming a 2D Gaussian distribution, the point probability of a point $x=[x_1, x_2]^T$ can be written on the form

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \frac{1}{(2\pi)^{1/2}\,|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)^T \Sigma^{-1} \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)\right)$$

where the probability distribution function takes as parameters the vector $\mu=[\mu_1, \mu_2]^T$ and a matrix $\sum$. Classification can be done by assigning a point to the class having the highest probability (i.e., the highest function *f(x)*). Any function based on a monotonous transform from the probability will give the same classification result, and is called discriminant function. By taking the logarithm of the above expression, a discriminant function can be created.

Assume two normally distributed classes with parameters

$$\mu^1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \Sigma^1 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$
$$\mu^2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \Sigma^2 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

The a priori probabilies for both classes are equal.

a) Sketch the class means in a plot

b) Sketch the covariance matrices in the same plot

c) You are given the data points listed below. Insert the points in your plot. Classify each point according to the classifier specified above.

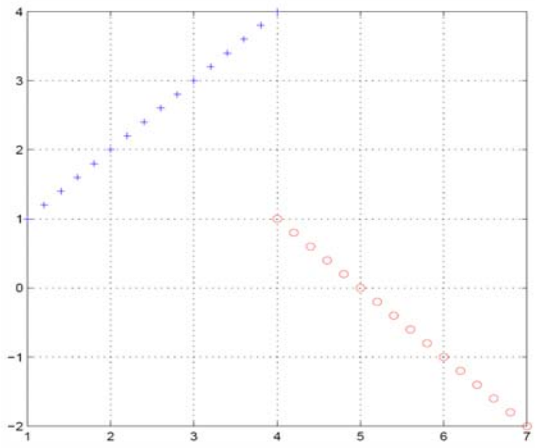
$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 7 \end{pmatrix}$$

d) Calculate the decision boundary.

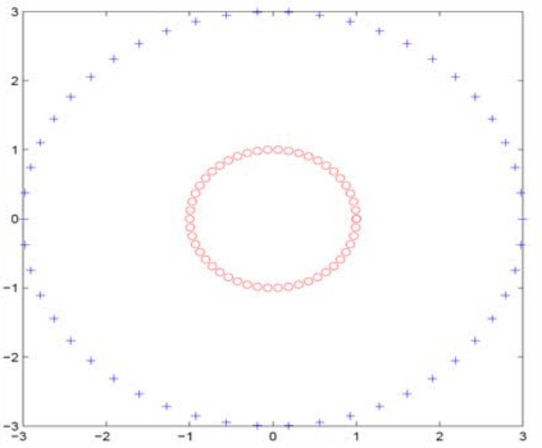
e) Sketch the decision boundary in the plot.

f) Consider the datasets in figures below, A and B. In each of these datasets there are two classes, ’+’ and ’o’. Each class has the same number of points. Each data point has two real valued features, the X and Y coordinates. For both of these datasets, we want to design a Bayes classifier assuming Gaussian distributed data . The covariance matrices are not equal across classes, but they are diagonal, on the form $\Sigma_j=\sigma^2 I$.

Given these assumptions, sketch the resulting decision boundaries in both cases. If the classifier breaks down, explain.



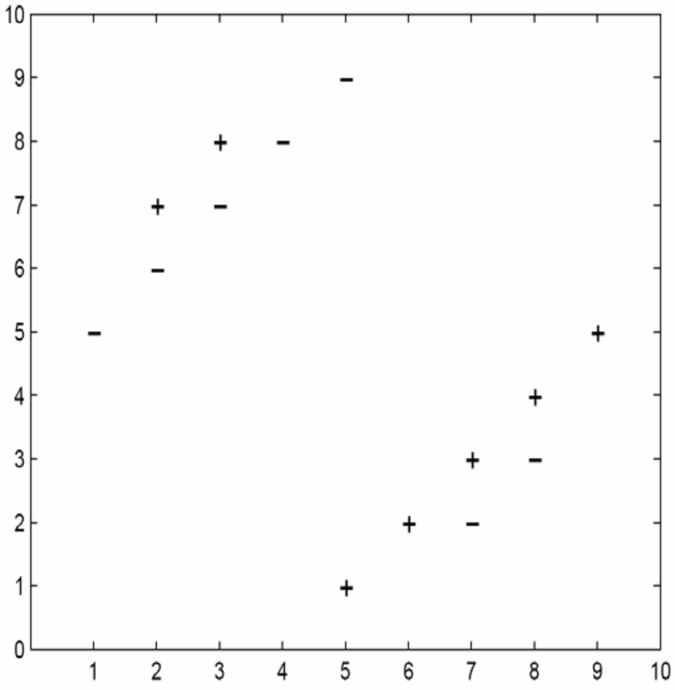
A



B

Exercise 6 2008. K-nearest neighbor classification

In the following questions you will consider a *k*-nearest neighbor classifier using Euclidean distance metric on a binary classification task. We assign the class of the test point to be the class of the majority of the *k* nearest neighbors. Note that a point can be its own neighbor.



c) In the figure, sketch the *1*-nearest neighbor decision boundary for this dataset.

d) If you try to classify the entire training dataset using a kNN classifier, what value of *k* will minimize the error for this dataset? What is the resulting training error?

e) What happens if you use a very large *k* on this dataset? Why might too small values of *k* also be bad?

What value of *k* minimizes leave-one-out cross-validation error for this dataset? For any reasonable choice of *k*, the resulting minimal error is 4/14.

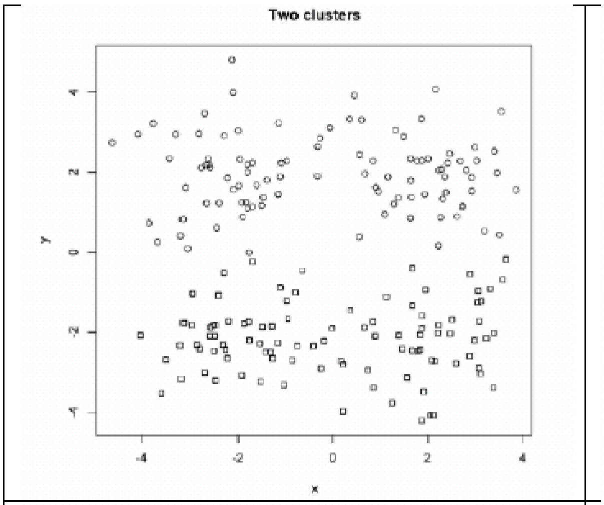
Exercise 2 2010 : k-means clustering

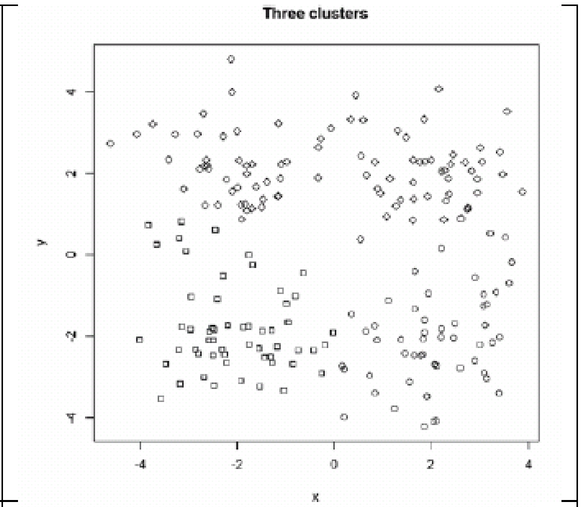
a) Explain what the *k*-means clustering algorithm is. You do not need to write code, but give a precise verbal description which someone could turn into code.

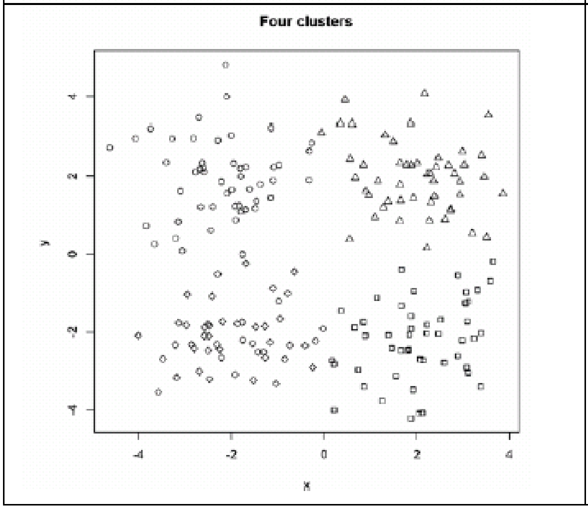
b) Can *k*-means ever give results which contain more or less than *k* clusters?

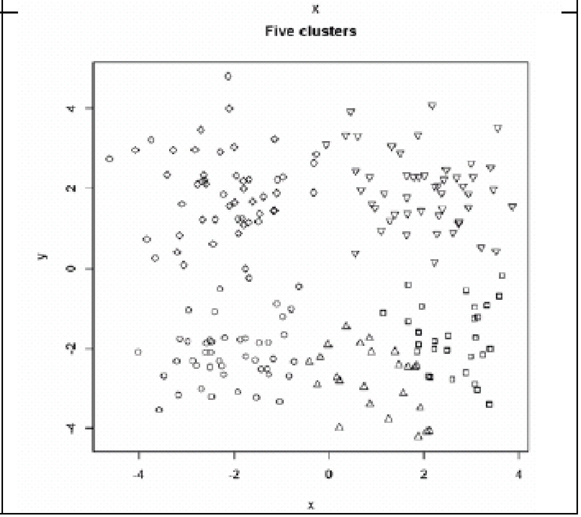
c) Explain what the sum-of-squares is for *k*-means.

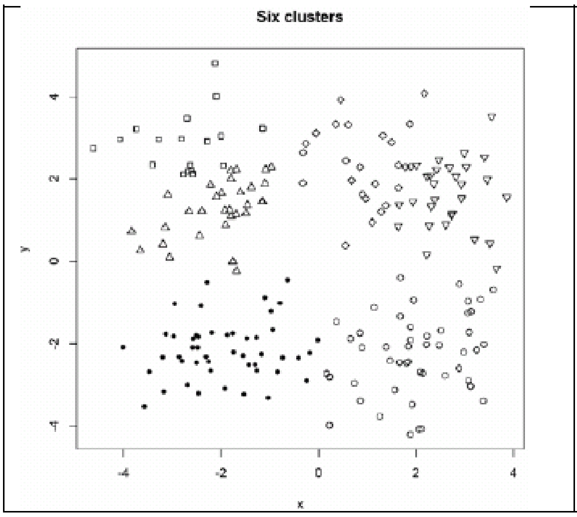
d) The next pages show the results of clustering the same data with *k* means, with *k* running from 2 to 6; also a plot of the sum-of-squares versus *k*. How many clusters would you guess this data has, and why? Does it matter whether the plot is an average over many runs of the algorithm?

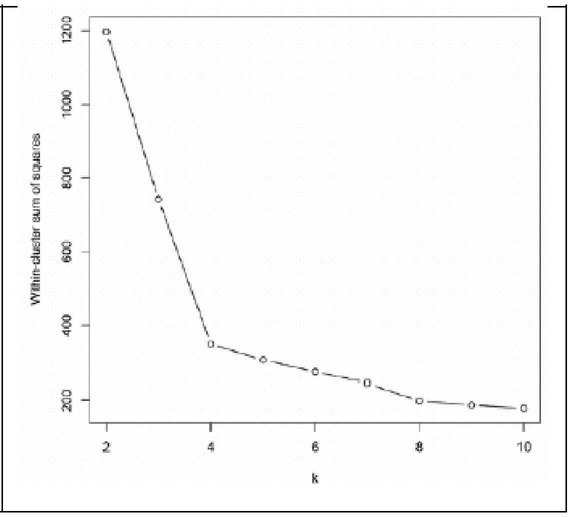

Two clusters


Three clusters


Four clusters


Five clusters


Six clusters



Oppgave 2 2011: Klassifikasjon

a) Gitt et klassifikasjonsproblem med 2 klasser.

$$\mu_1 = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Skisser middelverdiene og kovariansmatrisene i et plott.

b) Skisse desisjonsflatene (‘decision boundary’) for en multivariat normalfordeling når klassene har lik *a priori* sannsynlighet.

c) Hva vet vi om verdien på diskriminantfunksjonene for de to klassene på desisjonsflaten?

d) Diskriminantfunksjonen for en normalfordeling med felles kovariansmatrise er gitt ved uttrykket

$$g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + w_{i0}$$

$$\text{where} \quad \mathbf{w}_i = \frac{1}{\sigma^2}\boldsymbol{\mu}_i \text{ and } w_{i0} = -\frac{1}{2\sigma^2}\boldsymbol{\mu}_i^t\boldsymbol{\mu}_i + \ln P(\omega_i)$$

Anta at klassene fremdeles har lik *a priori* sannsynlighet og middelverdi og kovarians som gitt a). Gitt to egenskaper $\boldsymbol{x_1}$ og $\boldsymbol{x_2}$, finn et uttrykk for desisjonsflaten som en funksjon av $\boldsymbol{x_1}$ og $\boldsymbol{x_2}$.