Pattern Recognition
Lab week 2

Useful matlab functions:

`mean`, `var`, `cov`, `meshgrid`, `mvnpdf`, `mesh`, `rand`

Assignment 1: *covariance matrix.*

1. Compute the mean and the covariance matrix of the following set of feature vectors:

$$\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \\ 9 \end{bmatrix} \begin{bmatrix} 8 \\ 7 \\ 3 \end{bmatrix} \begin{bmatrix} 7 \\ 4 \\ 8 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \\ 5 \end{bmatrix}$$

hint we consider a feature vector to be a vector of values of measurements of different features of an object. So here we have five measurements (the five vectors) of three features; the first feature having values 4, 6, 8, 7 and 4, the second feature having values 5, 3, 7, 4 and 6 and the third feature having values 6, 9, 3, 8 and 5).

2. Using the computed mean and covariance vector, model the observed data by a normal distribution and compute the probability density in points: [5 5 6], [3 5 7] and [4 6.5 1]

Assignment 2: *covariance matrix, analytically.*

Consider vectors $\begin{bmatrix} a \\ b \end{bmatrix}$ and $\begin{bmatrix} c \\ d \end{bmatrix}$.

1. Compute the covariance matrix of these vectors (on paper and in terms of a,b,c, and d).

hint: we consider a feature vector to be a vector of values of measurements of different features of an object. So here we have two measurements (the two vectors) of two features (the first feature having values a and c, and the second feature having values b and d). Note also that the vector of means takes its means over the features, so we have as many means as we have features, where the means represent the average value taken over the different measurements.

2. What is the covariance matrix of $\begin{bmatrix} a+k \\ b+k \end{bmatrix}$ and $\begin{bmatrix} c+k \\ d+k \end{bmatrix}$?

3. What is the covariance matrix of $\begin{bmatrix} a*k \\ b*k \end{bmatrix}$ and $\begin{bmatrix} c*k \\ d*k \end{bmatrix}$?

Assignment 3: *2D Gaussian pdf, Mahalanobis distance.*

Generate a two-dimensional Gaussian pdf with a mean [3 4] and covariance matrix $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$.

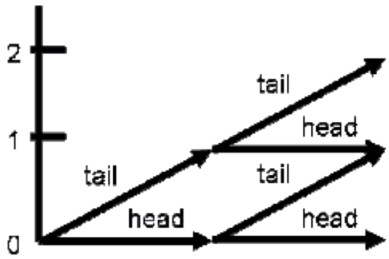
1. Plot this function on [-10 10] $\times$ [-10 10] using the `mesh` function.
2. Compute the Mahalanobis distance between the points [10 10]', [0 0]', [3 4]', [6 8]' and the mean of this density function.

hint: use the definition of the Mahalanobis distance from the lecture slides, Wikipedia or Mathworld

Assignment 4: *independent identically distributed random binary variables.*

We play a game of "random walk"

- Every player starts at 0.
- Each turn every player tosses a coin.
- Head → Lose and don't move.
- Tail → Advance one position.



1. Simulate a game where 1000000 people play for 100 turns. Make a plot of the number of people that end on a specific end-point. What does this distribution resemble? Explain why. What would be the *theoretical* mean and variance of this distribution?

Assignment 5: *multivariate normal density, discriminant functions, minimum error rate classification, unequal priors, dichotomizer.*

Consider two two-dimensional normal distributions with means $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and covariance matrices $\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$ and $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ repectively. Let the priors be $P(w_1) = 0.3$ and $P(w_2) = 0.7$.

1. Propose discriminant functions g1(x,y) and g2(x,y) that can be used for minimum error rate classification.

hint: see the lecture slides

2. Compute analytically the decision boundary and plot it in a 2D (x vs y) figure.

hint: solve g1(x,y) - g2(x,y) = 0. This will eventually give you a quadratic equation to solve; you might have to use MATLAB's `real` function to filter out imaginary results.

Assignment 6: *naïve Bayesian rule*

Consider the following table of probabilities for the occurrence of certain key words in an email:

	Spam	non-spam
Anti-aging	0.00062	0.000000035
Customers	0.005	0.0001
Fun	0.00015	0.0007
Groningen	0.00001	0.001
Lecture	0.000015	0.0008
Money	0.002	0.0005
Vacation	0.00025	0.00014
Viagra	0.001	0.0000003
Watches	0.0003	0.000004

Assume that the priors of receiving spam and non-spam are 0.9 and 0.1, respectively.

1. Using the naïve Bayes rule, classify the following email texts as spam or non-spam:
   a) "We offer our dear customers a wide selection of classy watches."
   b) "Did you have fun on vacation? I sure did!"