TINONS1 EXERCISES WEEK 3

Exercise 1 :

Create artificial data (eg. N=1000 samples) from a 2D gaussian distribution with specific mean and covariance matrix - use the "generate_gaussian_data.m" file with default values.

Estimate the marginal distributions (ie. $P(x_1)$ and $P(x_2)$) by making histograms of the two dimensions independently - do they look gaussian ? should they be ?

Estimate the conditional distribution $P(x_2 \mid x_1 > 5)$ - ie. the probability distribution of x_2 given that x_1 is larger than 5. In Matlab, this can be done by first finding the indexes where x_1 is larger than 5 (eg. "id = y(: , 1) > 5"), then copying corresponding values of x_2 into a new variable (eg. "ynew = y(id, 2)") and finally finding histogram of ynew.

Estimate the covariance matrix and the mean vector from the data - do they correspond to the expected values (the values used to generate the data..) ?

Exercise 2 :

Apply the gaussian classifier (as a generative model) to the 2D dataset given in "classification_2D_dataset_large.mat".

In the generative modelling process, the first step is to model the class-conditional probabilities for each class - ie. $P(x \mid C_i)$ where x is the feature vector and C_i is class i. This is done on the training set. Next, estimate $P(C_i \mid x)$ for each x in the test set. The relation between $P(x \mid C_i)$ and $P(C_i \mid x)$ is found by the use of Bayes' theorem.

Exercise 3 :

Apply the gaussian classifier to your own case. Some dimensionality reduction will often be necessary as a first step - why is that ?

What should $P(C_i)$ be in your case ?