TINONS1 EXERCISES WEEK 5

Exercise 1 :

Create an artificial data set that is drawn from a Gaussian Mixture Model (GMM) in 2 dimensions, that is, a probability distribution $p(x) = \sum p(k)\, p(x|k)$ where $p(x|k)$ is a gaussian distribution. As a starting point, "generate_Gaussian_data.m" from week 3 can be used. Use eg. K = 2 mixtures with mixture coefficients $p(1) = \pi_1 = 0.3$ and $p(2) = \pi_2 = 0.7$. Choose the mean vectors for the two mixtures as [0 0] and [4 4], respectively. The covariance matrices are 1*I and 2*I, respectively, where I is the unity matrix ([1 0; 0 1]). Create e.g. N=1000 samples. Although not strictly correct, you can simply create 300 samples from mixture 1 and 700 from mixture 2, to achieve (roughly) correct mixture coefficients.

Plot the samples.

Calculate the probability p(x = [2 0]), that is, the value of the probability distribution function at the point x=[2 0]. Does the result correspond to the plot of the samples ?

Calculate the so-called *responsibilities* for the point x=[2 0] from the two mixtures, that is, calculate p(k=1 | x=[2 0]) and p(k=2 | x=[2 0]). Do similarly for the point x=[6 0] and explain the result from the knowledge of the GMM parameters (means, covariances and mixture priors).

Exercise 2 :

Apply the Expectation-Maximization (EM) algorithm for GMMs to the artificial data set from exercise 1. This is the "training/learning" phase where the parameters of the model are estimated (inferred). The file "GMMs.m" can be used as a starting point.

Try first with 2 mixtures and see if the correct parameters are found. Next try with a larger number of mixtures - e.g. 10 mixtures. Do you get any problems ?

Exercise 3 :

Apply the Expectation-Maximization (EM) algorithm for GMMs to your own case. For instance, train a GMM for each class and find test/training errors - compare with other trained classifiers. Otherwise, the model can be used unsupervised (not using class information) to explore your data - for instance, to find possible subgenres in music.