**Data mining**

**Lab 7**

**Expectation Maximization Clustering**

Write a script that implements the Expectation-Maximization (EM) algorithm for clustering (see Algorithm 13.3 in Chapter 13). Run the code on the **iris.txt** dataset. Use the first attribute for clustering.

For EM initialization, use the first n/k points for cluster 1, the next n/k for cluster 2, and so on. For convergence testing, you can compare the sum of the euclidean distance between the old means and the new means over the k clusters. If this distance is less than ϵ=0.001 you can stop the method.

Your program output should consist of the following information:

1. The final mean for each cluster
2. Number of iterations the EM algorithm took to converge.
3. Final cluster assignment of all the points, where each point will be assigned to the cluster that yields the highest probability P(Ci|xj)
4. Final size of each cluster