Computational Statistics II
Assignment 3: a simple EM

Alessia Pini

Aim of this assignment is to write in a simple case, the code for performing an EM algorithm on a mixture of bi-variate Gaussians with $K = 3$ components of the mixture, and to apply it for performing a simple clustering.

1. Describe the EM algorithm applied to a mixture of Gaussians. Specify the properties of the method (without proving them), and underline the differences with the k-means algorithm (max. 2 pages).

2. Consider the `iris` dataset (`data(iris)`). Perform a 3-means clustering using as variables the sepal length and petal length only.

3. Initialize the parameters of the EM using $\pi_1 = \pi_2 = \pi_3 = 1/3$, $\Sigma_k = I$, $\mu_1 = (4.5, 1)$, $\mu_2 = (6, 3)$, and $\mu_3 = (8, 7)$. Perform an E-step and an M-step. Visualize the results of this first iteration.

4. Appky the EM to this example, iterating through E-step and M-step. Use the convergence of the log likelihood as stopping criterion.

5. Compare the results with what was obteined from the k-means.

6. **Bonus.** Perform the EM algorithm using all the data in the `iris` dataset.