Exercise 3

1. Load data **jain.txt**, first two columns are coordinates of datapoints, and third column is the true cluster label.

   Implement and apply following clustering methods to find clusters of this data. Stop iterations after pre-set maximal number of iterations. Assume there is only two clusters in the data. Do not use existing implementations of these methods.

   - K-means clustering
   - Mountain clustering
   - Subtractive clustering

   Compare the performance of these methods (i.e. number of correctly clustered datapoints, total loss/error, runtime). Examine the effects of initial locations of clusters and different parameters: What kind of parameters are required for each of these methods to work.

2. Load data **butterfly.txt** demonstrated in the lecture.

   Implement and apply both k-means and fuzzy c-means clustering methods with number of clusters set to two. Examine clustering results of point (3,2) in different trials. Do the clustering results change between different trials? Which one would be more appropriate clustering method for this data and setup?

   Set the number of clusters to three and repeat above experiments.