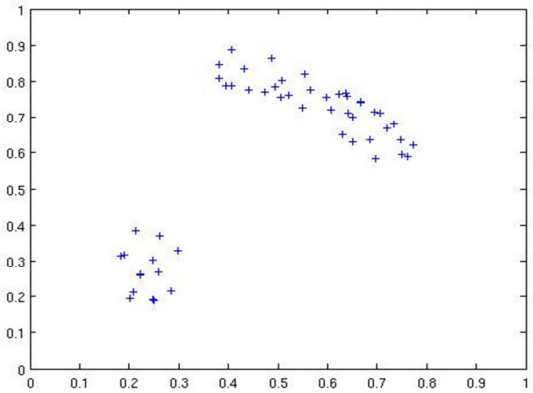
Exam Computational Methods for Data Analysis (19/07/2013)

Choose any two of the following questions:

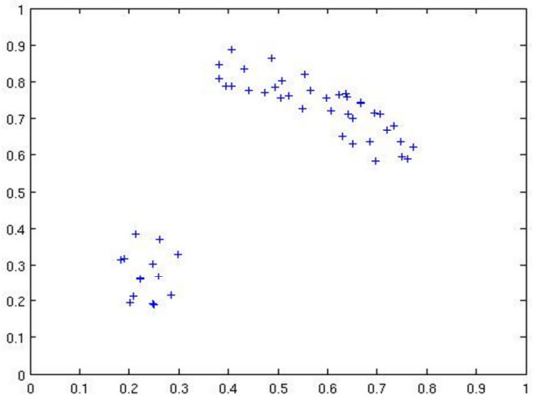
1)a Describe the difference in algorithm between k-Nearest Neighbor and Parzen classification.
1)b Give an example in which a linear classifier works while k-NN does not. Explain why.
1)c Give an example in which k-NN works but a linear classifier does not. Explain why.

2a) Briefly describe k-means clustering
2b) Which clusters do you expect to get for k=2 and k=3 in the figure below (draw the clusters)? Why?

K = 2



K = 3



3) Difference between linear and non linear classifier: the role of kernel functions.

4) What is overfitting? Describe the concept using Decision Tree, Naive Bayes classifiers and an abstract regression algorithm as examples