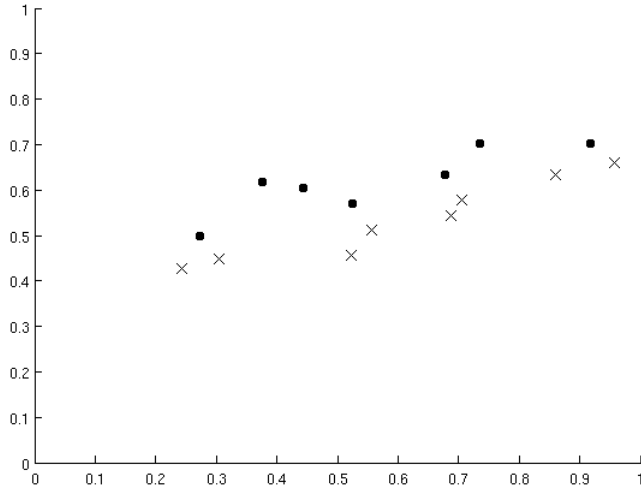


Exam Computational Methods for Data Analysis (14/06/2013)

Choose any two of the following questions:

1. Briefly describe k-NN classification. Will k-NN classification work on the data displayed in the figure below? Or not? Explain why.



2. You have just performed PCA on some large dimensional data to obtain the eigenvectors- what do the eigenvectors represent/denote?; When you attempt to reconstruct the original data with the eigenvectors, how does the reconstruction error vary when you use more or less eigenvectors? Explain.
3. What is the difference between empirical error and risk (or true error)? How do they relate to the problem of overfitting ?
4. How does Decision Tree algorithm work? What are the differences with other Vector Space Models such as Support Vector Machines?