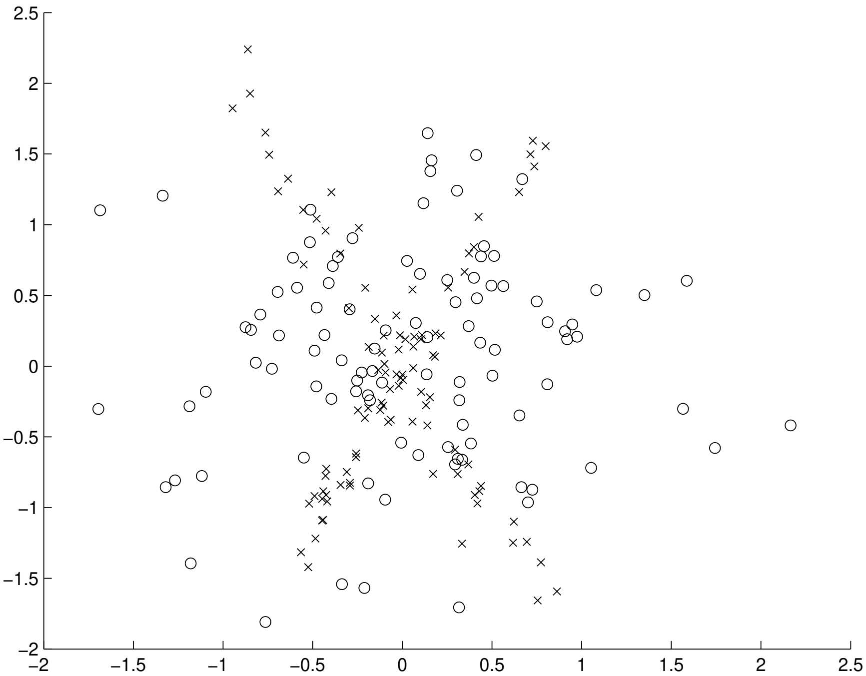
ELEC 531 Homework 1

Due Tuesday, Sept. 3, at class

Pattern Classification from Training Examples:

Suppose that we have a DNA test for detecting a gene that causes acne (aka "zits"). When the test is run on a person, the result is a two-dimensional feature vector. The scatter plot below depicts the two-dimensional features for a set of persons with and without acne. The o's are the measured features of cases of persons without acne, and the x's are persons known to have acne. These datasets are available on the course web page (`class1.mat` and `class2.mat`).



Based on this data, we would like to design a classification rule that takes the DNA results (the two-dimensional feature vectors) for new test cases and classifies each person as either +1 (having the acne gene) or -1 (not having the acne gene). Come up with a rule that takes a two-dimensional feature vector as an input and produces either +1 or -1. Obviously, the x's and o's are not easily separated, so its probably not possible to perfectly classify a person, but we'd like to be as accurate as possible. Once you have a classification rule, you can apply your method to 1000 test cases (in the file `test.mat` also on the webpage). You should produce a vector of dimension 1000×1, with each entry a +1 or -1 indicating your classification. Call your classification vector c, store it in a file called class.mat, and send it to Becca Willett (willett@rice.edu). She will compare your classifications to the "true" classifications (which are known to us; I generated the data synthetically according to a certain probability model and so I know the true class in each case). The person with the best accuracy on the test set will get 10 extra bonus points (in addition to the standard 10 points that this homework problem is worth).