1. We want to predict whether a person likes a movie. We know whether the person is male or female. From a population survey we know that 60% of people like the movie and of these 20% are males. You may assume that 50% of the population is male. Use Bayes Rule to construct a simple movie prediction classifier.

2. Consider now a different movie prediction task. Suppose training data is available which consists of pairs (x_i,y_i), j=1,2,…,n, where y_i=1 if the i^{th} person likes the move and 0 otherwise, and x_i=1 if the i^{th} person is male and 0 otherwise. Describe how to construct a logistic regression classifier for this task.

3. Logistic regression classification is often said to be most effective for linearly separable data. Explain what this means for a classifier with two inputs. What can happen is the data is <u>not</u> linearly separable ?

4. Download the file chdage.dat from:

https://www.scss.tcd.ie/doug.leith/ST3009/chdage.dat

This contains data from a medical study. Each row corresponds to a different person and consists of three columns. The first is the person ID. The second is their age. The third is whether they have heart disease or not (1 or 0). Our aim is to use age as an input/feature to predict whether a person has heart disease. Use logistic regression to construct a Matlab classifier for this data, and write a short report on your results (to include a plot of the original data overlaid with the classifier predictions and a report on the fraction of predictions which are correct).