Homework 2
(due Tuesday, Feb 9th 2016)

1. **Least squares implementation:** Implement and execute a least squares classifier on Fisher's Iris flower dataset (`http://en.wikipedia.org/wiki/Iris_flower_data_set`). The dataset contains 50 samples from each of the three classes (*setosa, virginica, versicolor*). For the choices of 10%, 30% and 50% as training data, run 10 random trials in each case. (For example, if you choose 10% for training, in each of the 10 trials, randomly pick 10% for training and the remainder for testing.) To document performance, evaluate and report the training and test set misclassification errors in each case (10%, 30% or 50% training data chosen and for each random trial). Try and balance the percentages of data from each class in the training set. Document all choices including percentages of each class in each random trial. Given the training set, form the $N \times (D+1)$ matrix $\tilde{X}$ (with an extra column of ones for the bias). When training the classifier, use the following equation, where the matrix $\tilde{X}^T \tilde{X} + \lambda I_{(D+1)}$ is positive definite:

$$\tilde{\Theta} = \left(\tilde{X}^T \tilde{X} + \lambda I_{(D+1)} \right)^{-1} \tilde{X}^T Y. \tag{1}$$

where $\tilde{\Theta} = \{\Theta, \Theta_0\}$. The free parameter λ can be chosen for example to minimize the misclassification error on the training set. Document your choice of λ in each case.

2. Theodoridis MATLAB exercise 7.20.