Homework 2

1. MNIST data referred to below is included in mnist.

Problem 2.7 There is a link to a very large dataset of handwritten figures on the book website (the MNIST dataset). Download it and use a Perceptron to learn about the dataset.

2. The data set required is in http://archive.ics.uci.edu/ml/datasets/Auto+MPG. It is also included in auto_mpg.

(a) You are asked to run the program below for regression. You need to complete it by downloading the data and separating it in to a training part and a testing part. How to do this will be explained in class. You also need the regression routines for chapter 2 of the textbook. After getting the output for testing data, discuss what you understand about this output.

(b) plot the testtgt and testout signals on the same graph to see how they compare.

```
# Code from Chapter 2 of Machine Learning: An Algorithmic Perspective
# by Stephen Marsland
(http://seat.massey.ac.nz/personal/s.r.marsland/MLBook.html)

# You are free to use, change, or redistribute the code in any way you
wish for
# non-commercial purposes, but please maintain the name of the original
author.
# This code comes with no warranty of any kind.

# Stephen Marsland, 2008

# This is the start of a script for you to complete
from pylab import *
from numpy import *
import linreg

auto = loadtxt('/Users/srmarsla/Book/Datasets/auto-mpg/auto-
mpg.data.txt',comments='"')

# Separate the data into training and testing sets

# Normalise the data

# This is the training part
beta = linreg.linreg(trainin,traintgt)
testin = concatenate((testin,-ones((shape(testin)[0],1))),axis=1)
testout = dot(testin,beta)
error = sum((testout - testtgt)**2)
print error
```

3. The prostate data referred to below is included in prostate_data.

Problem 2.8 For the prostate data available via the website, use both the Perceptron and logistic regressor and compare the results.