Homework 4
Statistical Learning, Spring Term 2019
STAT760

The website for the Statistical Learning book has a repository of data:

https://web.stanford.edu/~hastie/ElemStatLearn/

The South African Heart Disease data set has the predictors

sbp   systolic blood pressure
tobacco   cumulative tobacco (kg)
ldl   low densiity lipoprotein cholesterol
adiposity
famhist   family history of heart disease (Present, Absent)
typea   type-A behavior
obesity
alcohol   current alcohol consumption
age   age at onset

and one outcome

chd   response, coronary heart disease

Problem 1

   a) Divide the data into a train a and a test set. Train a logistic regression predictor for chd
       using from one to eight predictors (write your own linear regression function). Graph
       the test error for the eight possible regressions.
       (15 points)

Program the classifier using Matlab, R, Python, or any other language of your choice.

Problem 2 (5 points)

Solve the problems 1 and 2 from Chapter 4 of the ISL book (Section 4.7).
http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf