1. Read Sections 3.6 and 3.8 in the ESL textbook by yourself.

2. Repeat the procedure of Problem 2 in Homework 1 and compare four methods as follows: (I) full linear model; (II) reduced linear model with lcavol, lweight, lbph, svi; (III) principal components regression (PCR); (IV) partial least squares (PLS). That is, randomly partition the prostate cancer data into 67 training data points and 30 testing data points; fit/tune your model on the training data and estimate the mean (absolute) prediction error and mean (squared) prediction error using the testing data; and repeat the procedure for 100 times.

   (1) Do pairwise comparison of the four methods in terms of mean (absolute) prediction error and mean (squared) prediction error, respectively. Announce all significant differences at 5% level.

   (2) Does your conclusion on model selection change across the 100 different random partitions of 67 training data points and 30 testing data points?

   (3) Do you conclude that PCR and PLS perform significantly better than the full linear model and/or reduced linear model? If not, why do we still need those dimension reduction methods?