LINEAR REGRESSION MODELS W4315

HOMEWORK 2 QUESTIONS

January 28, 2010

Instructor: Frank Wood

1. (20 points) In the file "problem1.txt"(accessible on professor's website), there are 500 pairs of data, where the first column is X and the second column is Y. The regression model is $Y = \beta_0 + \beta_1 X + \epsilon$

a. Draw 20 pairs of data randomly from this population of size 500. Use MATLAB to run a regression model specified as above and keep record of the estimations of both β_0 and β_1. Do this 200 times. Thus you will have 200 estimates of β_0 and β_1. For each parameter, plot a histogram of the estimations.

b. The above 500 data are actually generated by the model $Y = 3 + 1.5X + \epsilon$, where $\epsilon \sim N(0, 2^2)$. What is the exact distribution of the estimates of β_0 and β_1?

c. Superimpose the curve of the estimates' density functions from part b. onto the two histograms respectively. Is the histogram a close approximation of the curve?

2. (20 points) Use the same data set in the last problem, we will estimate β_0 and β_1 using Newton-Raphson method.

a. Draw a 3d plot using MATLAB(check "surf" command for example) to illustrate how the SSE varies according to different combinations of estimates of β_0 and β_1. So to speak, draw a 3d plot where x and y axes represent different values of slope and intercept of the regression line respectively, while z axis is the SSE.

b. Use Newton-Raphson method to minimize the SSE and give estimates of the parameters(slope and intercept) of the regression line. Give a geometrical interpretation of the method and explain how it works.

3. (10 points) a. In simple linear regression setting $y = \beta_0 + \beta_1 x + \epsilon$, write out the explicit form the error function.

b. Prove this function is convex with respect to its variables(β_0 and β_1).