ISL Fall 2015 Assignment VI. 100 pts.

PIN NUMBER(s):

**LOKESHWAR REDDY INJA (16207239/ LIRK9)**

**SANTHOSH KUMAR GATTU (16211118/ SG6N6)**

**SANTOSH KUMAR (16207971/ SK7Z9)**

You may submit this assignment in groups of upto three each. Write your **PIN numbers** on this sheet and include it as the cover page for your submission.

The objective of this assignment is to practice using R, and gain a fundamental understanding of linear regression. Your submission should include both your code as well your answers to the questions.

Electronic submission on Blackboard is due **latest by 11 pm on Wed, Nov 11th**. You may upload upto **three** submissions **before** the deadline – only the last submission will be graded. Submissions received after the deadline will be graded only for effort for a maximum of 70% of the total grade (Refer to class syllabus for detailed grading policy).

**State any assumptions you make, justify your answers, show intermediate steps and explain your results for maximum credit**. All answers should be in your own words with any sources you refer to cited at the appropriate places. Any knowledge you acquire from the Internet should be written in your own words and be appropriately referenced. Copying and pasting from the Internet, each other or any other source will not count as your effort (Refer to class syllabus for detailed policy on plagiarism).

**Remember that answers need to be word-processed (NOT handwritten) and should use R. Submit all your R code as a single merged file for all the assignments.**

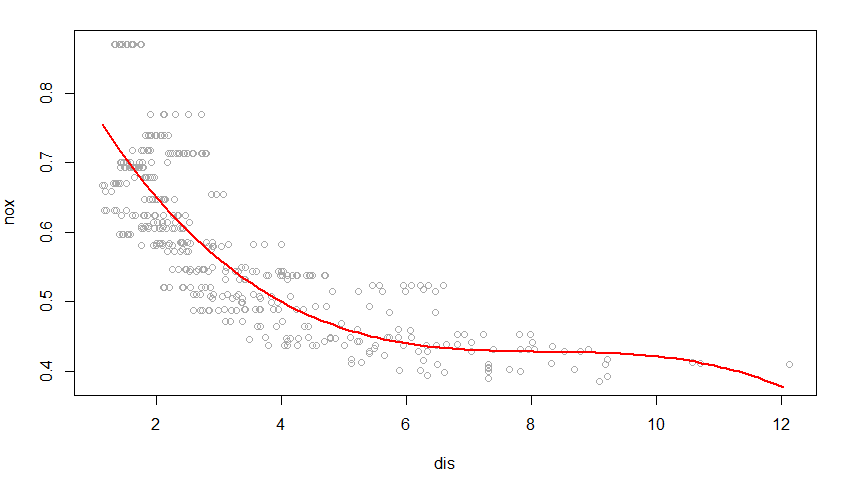
Answer the following questions from Chapter 7.

Q9

Q11 (Grad students only)

9.a)

Sol



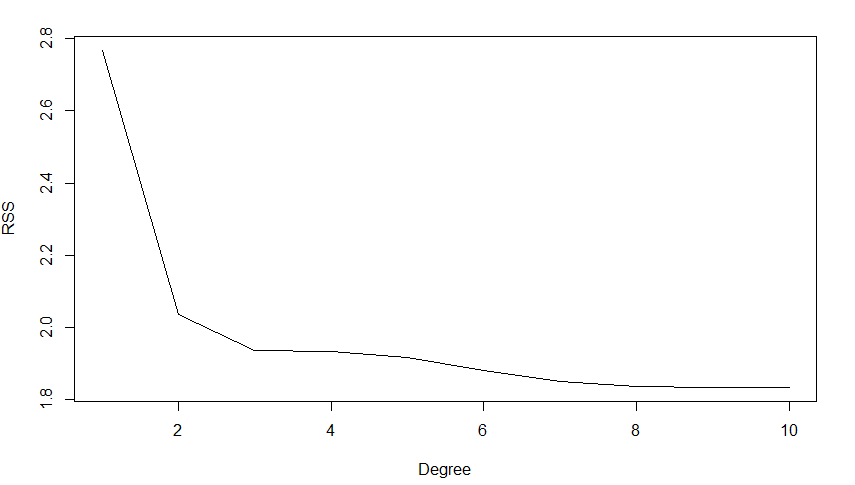
From the above plot we can observe that the polynomial terms are significant.

R code is attached separately.

9.b)

Sol.

R code is attached separately.

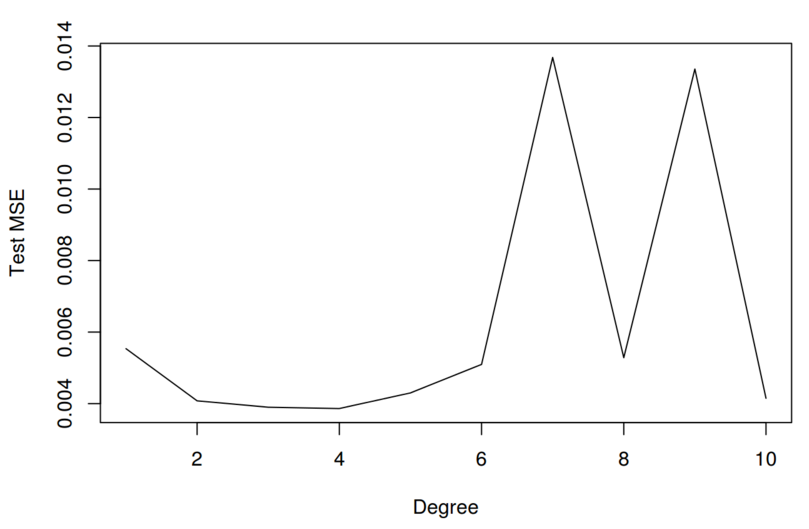


From the above plot we can observe that as the degree of polynomial is increases RSS decreases and it becomes minimum for polynomial of degree 10.

9.c)

Sol

R code is attached separately.



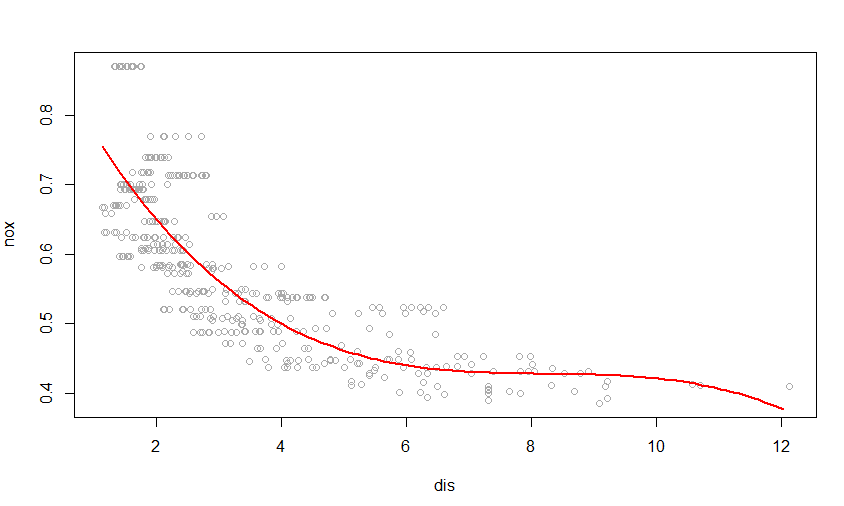
From the above plot we can observe that Test MSE is minimum for degree of polynomial 4.

9.d)

Sol

From the below fit we can observe that all the terms of spline fit are significant.

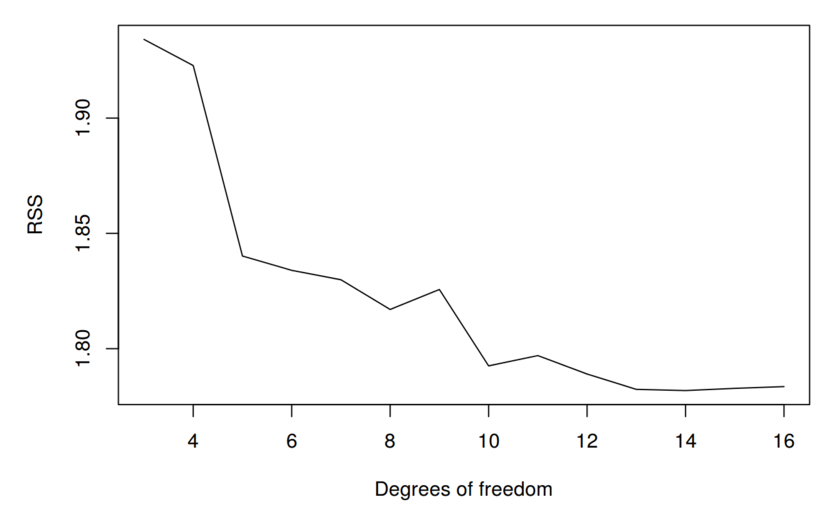
R code is attached separately.



9.e)

Sol

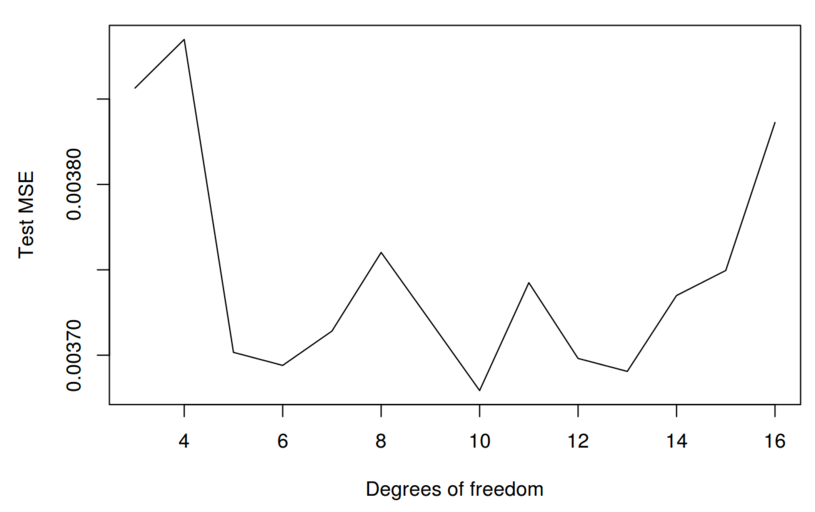
R code is attached separately.



From the above fit we can observe that upto Degree of freedom 14 RSS is decreasing and it increses again from 14.

9.f)

Sol. R code is attached separately.



Test MSE is less for 10 degrees of freedom.

11.a)

Sol R code is attached separately.

We create variables according to the equation Y=−2.1+1.3X1+0.54X2.

set.seed(1)

X1 = rnorm(100)

X2 = rnorm(100)

eps = rnorm(100, sd = 0.1)

Y = -2.1 + 1.3 \* X1 + 0.54 \* X2 + eps

11.b)

Sol. R code is attached separately.

beta0 = rep(NA, 1000)

beta1 = rep(NA, 1000)

beta2 = rep(NA, 1000)

beta1[1] = 10

11.c)

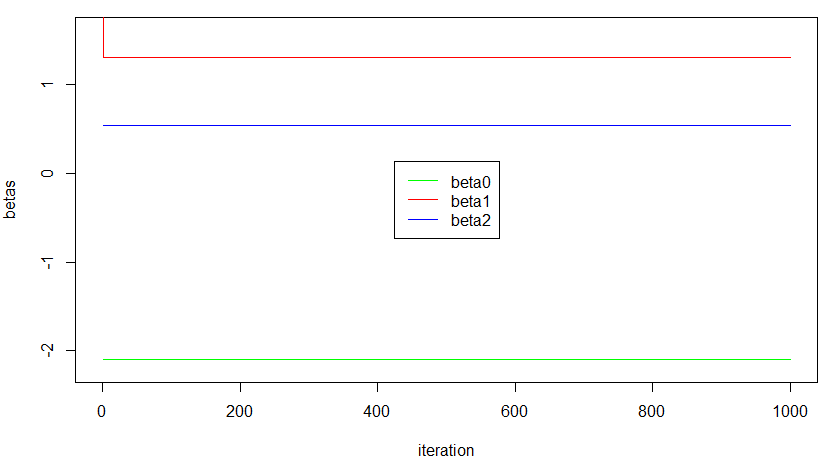
Sol. R code is attached separately

11d)

Sol. R code is attached separately

11e)

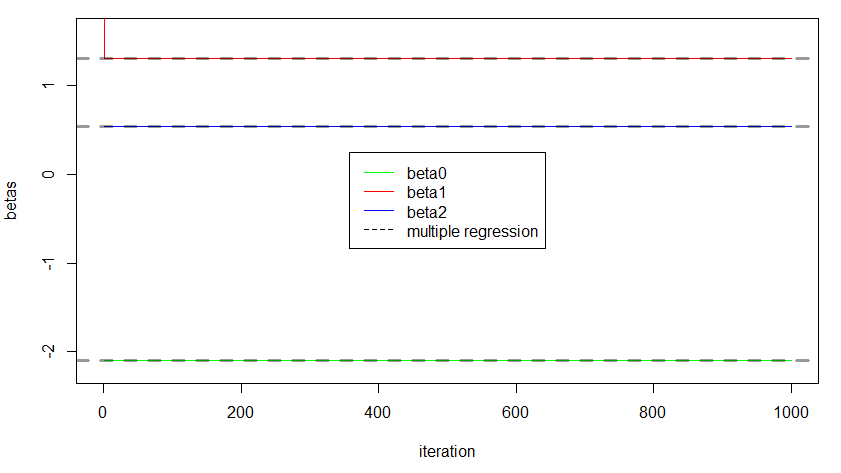
Sol. R code is attached separately.



11f)

Sol. R code is attached separately.

From the below fit we can observe that estimated multiple regression coefficients exactly matches with the coefficients which are obtained using backfitting, this is represented by the dotted lines.



11g)

Sol.

Only one backfitting iteration is sufficient to attain a good approximation to the multiple regression estimates.