

HW #9 due Thursday, March 28

In Content I have placed two files: “parkinsons_updrs.data” and “parkinsons_updrs.names”. The first file contains the data and the second file explains it. Use multiple linear regression to predict “motor_UPDRS” as a function of the columns numbered 7 to 22.

Hint: If you save the data as “park” then the following command will give you the appropriate formula without manually writing the column names,

```
paste("motor_UPDRS~",paste(colnames(park[7:22]),collapse="+"))
```

Write code to perform 10-fold cross validation. Use this code to compute the average mean squared error of the test sets for several linear models. The linear models to consider include different numbers of variables: the first includes only the single most significant variable, the second includes the two most significant variables, the third includes the four most significant variables, the fourth includes the eight most significant variables, and the fifth includes all 16 variables. Make a single figure showing the cross validation results for these five models. Turn in your code, this figure, and explain in words your interpretation of this figure.