

## ASSIGNMENT 12 - BIAS VARIANCE TRADEOFF IN R

IS 605 FUNDAMENTALS OF COMPUTATIONAL MATHEMATICS - 2014

Using the *stats* and *boot* libraries in R perform a cross-validation experiment to observe the bias variance tradeoff. You'll use the auto data set from previous assignments. This dataset has 392 observations across 5 variables. We want to fit a polynomial model of various degrees using the *glm* function in R and then measure the cross validation error using *cv.glm* function.

Fit various polynomial models to compute *mpg* as a function of the other four variables *acceleration*, *weight*, *horsepower*, and *displacement* using *glm* function. For example:

```
glm.fit=glm(mpg~poly(dis+hp+wt+acc,2), data=auto)
cv.err5[2]=cv.glm(auto,glm.fit,K=5)$delta[1]
```

will fit a 2nd degree polynomial function between *mpg* and the remaining 4 variables and perform 5 iterations of cross-validations. This result will be stored in a *cv.err5* array. *cv.glm* returns the estimated cross validation error and its adjusted value in a variable called *delta*. Please see the help on *cv.glm* to see more information.

Once you have fit the various polynomials from degree 1 to 8, you can plot the cross-validation error function as

```
degree=1:8
plot(degree,cv.err5,type='b')
```

For your assignment, please create an R-markdown document where you load the auto data set, perform the polynomial fit and then plot the resulting 5 fold cross validation curve. Your output should show the characteristic U-shape illustrating the tradeoff between bias and variance.