

### Lab #2: Polynomial Regression

Due: 3/16/2011 11:59 PM

In this homework, your task is filling out the given function skeletons. You will learn how to compute polynomial regression and visualize data using R. This homework will help you understand the basics of regression and the concept of overfitting and underfitting.

#### Function **polyfit(data, order)**

Read the data frame given as argument “data” and compute polynomial regression of order “order”. If order = 0, it should return a vector with one element (the sample mean). If order = 1, it should return a vector with two elements (from  $w_1 \cdot x_1 + w_0$ , the vector should be  $[w_1 \ w_0]$ ). Your **polyfit** function should be able to handle up to arbitrary order as far as there is enough number of training instances.

#### Function **predvalue(w, x)**

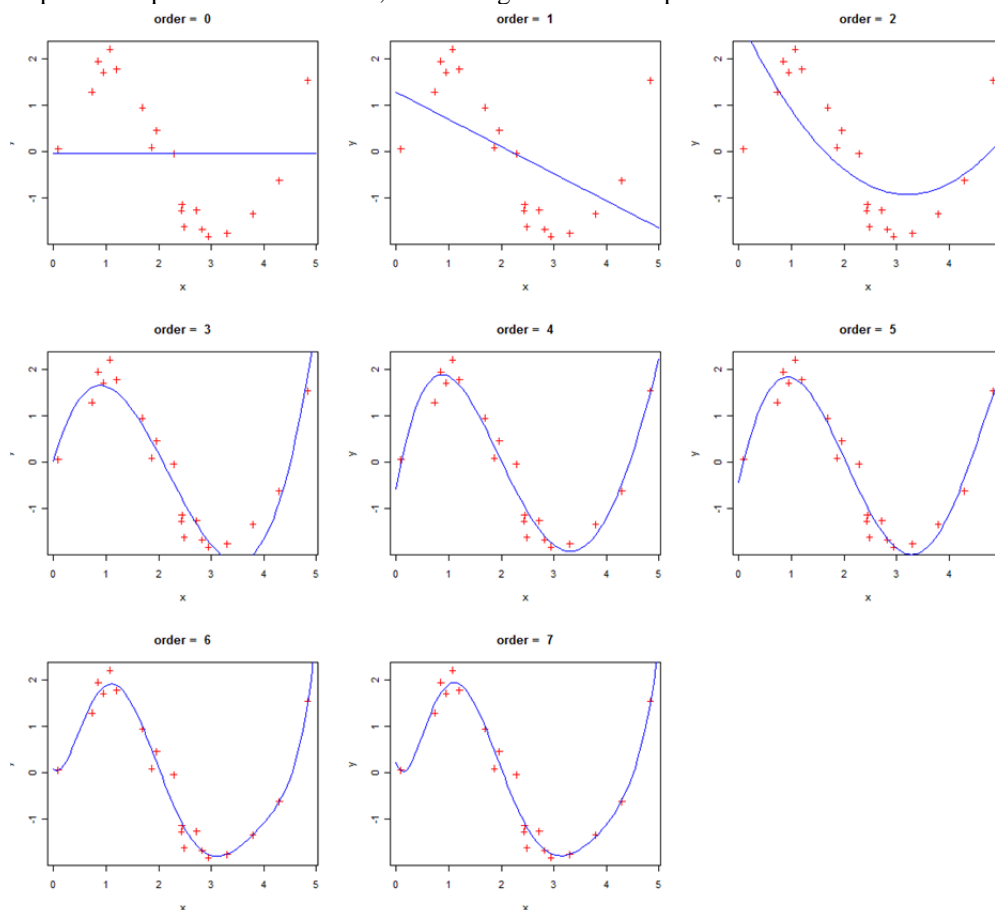
This function will return a vector containing the predicted values for each element in the vector “x” by using “w” calculated from **polyfit** function. This function will be used everywhere in this homework.

#### Function **polygraphs(filename)**

Read the file given as argument “filename” and repeatedly call **polyfit** function with orders from 0 to 7. For each order of polynomial regression, plot data points from the dataset and the regression line computed from the dataset on a 3 by 3 table. For the regression, use the regression weights computed by **polyfit** function, and draw regression lines using **predvalue** function.

#### Example

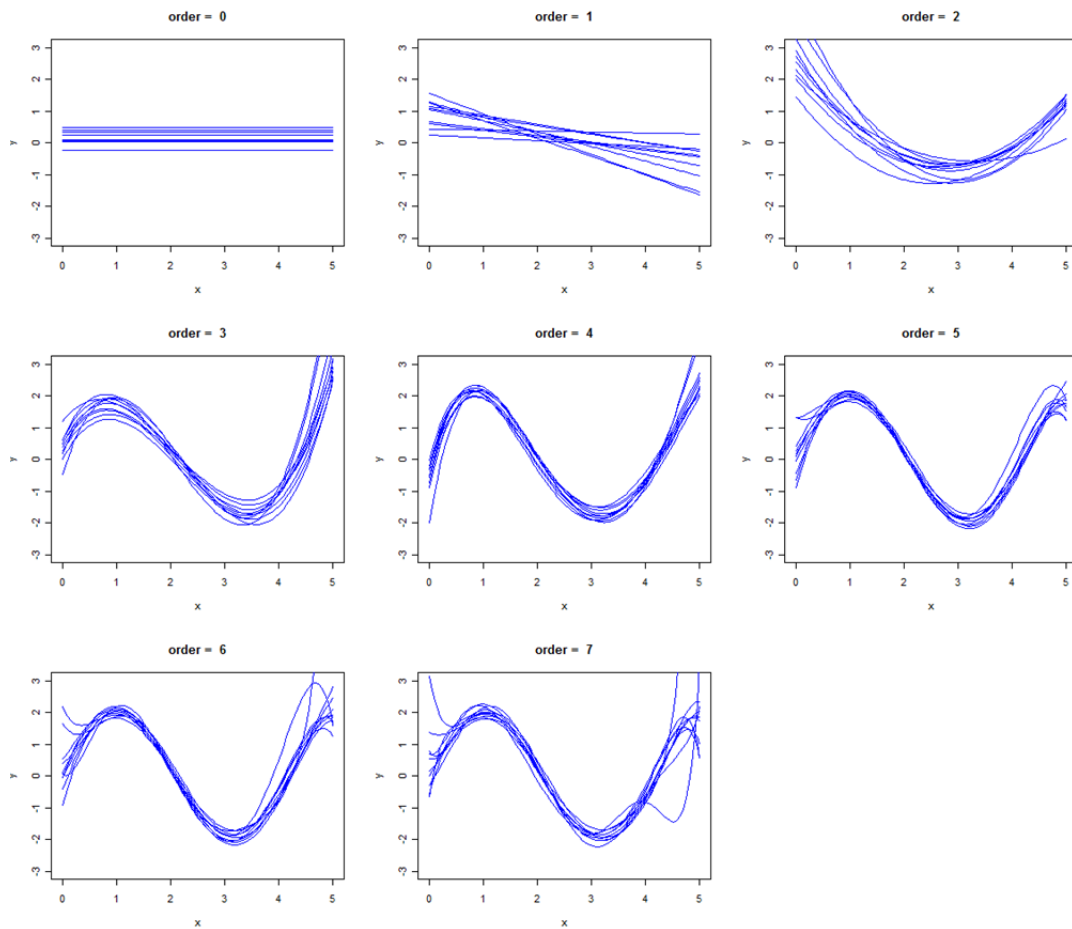
Given data points are plotted as red crosses, and the regression line is plotted as a blue line.



### Function **ofpolygraphs()**

This function shows how the regression computed from one dataset does not fit into other datasets because of the overfitting problem. This function will draw 8 graphs as **polygraphs** function does. Draw 10 regression lines computed from data1.dat ~ data10.dat on each cell of the 3 by 3 table.

### Example



### Function **biassq(order)**

### Function **variance(order)**

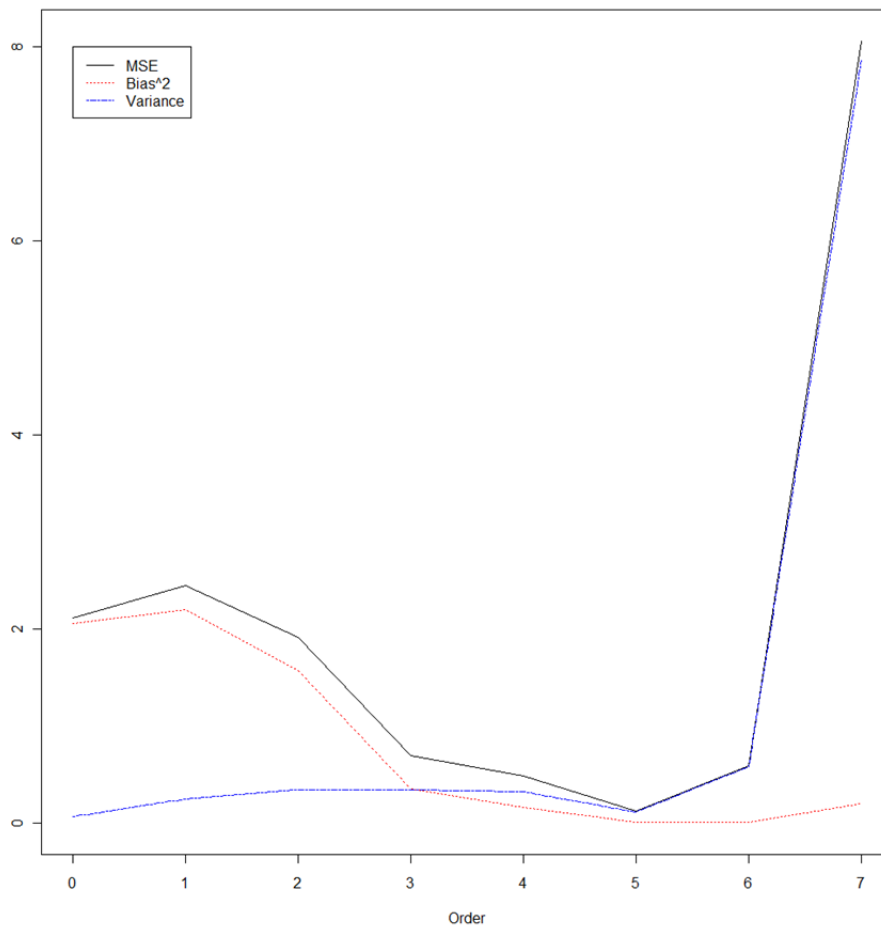
### Function **MSE(order)**

These functions will compute bias<sup>2</sup>, variance, and the mean square error of your polynomial regression of given “order”. They will execute your polynomial regression on datasets data1.dat ~ data10.dat, and store the results in the matrix w. The i-th row stores the parameters for your polynomial regression for the i-th dataset. The datasets were generated from the function **f** (given in lab2.R file), plus some noise. Use the function **f** and the dataset “data.dat” for calculating bias<sup>2</sup>, variance, and MSE.

### Function **MBVgraph()**

Plot a line graph using your **MSE**, **biassq**, and **variance** functions with orders from 0 to 7 (as the graph in figure 4.6 of the textbook). Make sure that  $MSE = Bias^2 + Variance$ .

### Example



### Submission Format

#ID\_FIRSTNAME\_LASTNAME\_lab2.R

For example, 20110101\_Ethem\_Alpaydin\_lab2.R