Assignment 5 (Polynomial Regression)

- 1. Generate 50 real number for the variable X from the uniform distribution U [0,1]
- 2. Construct the training set T = { $(x_1,y_1),(x_2,y_2),...,(x_{20},y_{20})$ } using the relation Yi = $\sin(2\pi x_i) + \epsilon_i$ where $\epsilon_i \sim N(0,0.25)$
- 3. In the similar way construct a testing set of size 100

i.e. Test = {
$$(x'_1, y'_1), (x'_2, y'_2), \dots, (x'_{50}, y'_{50})$$
}

4. Estimate the Least Square polynomial regression model of order M= 1,2, 3, 9, using the training set T.

For example for M=1, we need to estimate $f(x) = \beta_1 x + \beta_0$

For M = 2,
$$f(x) = \beta_2 x^2 + \beta_1 x + \beta_0$$
.

- 5. List the value of coefficients of estimated polynomial regression models for each case.
- 6. Obtain the prediction on testing set and compute the RMSE for polynomial regression models for order M =1,2,3 and 9.
- 7. Plot the estimate obtained by polynomial regression models for order M =1,2,3 and 9 for training set along with y_1, y_2 , , y_{100} . Also plot our actual mean estimate $E(Y/X) = \sin(2\pi x_i)$.
- 8. Plot the estimate obtained by polynomial regression models for order M =1,2,3 and 9 for testing set along with y'_{1}, y'_{2} , y'_{100} . Also plot the sin(2 π x'_{i}).
- 9. What happens when we increase the value of M. Note down your observations.
- 10. Also, try to find the statistical reasons behind your observation.