Assignment 2

Artificial Dataset

a) Generate 20 real number for the variable X from the uniform distribution U [0,1]
b) Construct the training set T = { $(x_1,y_1),(x_2,y_2),\ldots\ldots,(x_{20},y_{20})$} using the relation
    1. $Yi = \sin(2\pi x_i) + \epsilon_i$ where $\epsilon_i \sim N(0,0.25)$
c) In the similar way construct a testing set of size 50
    a. I,e. Test = { $(x'_1,y'_1),(x'_2,y'_2),\ldots\ldots,(x'_{50},y'_{50})$}
d) Estimate the regularized least squared polynomial regression model of order M= 1,2, 3, 9, using the training set T.
    i. For example for M=1 , we need to estimate
    ii. $F(x) = \beta_1 x + \beta_0$
    iii. For M = 2
    iv. $F(x) = \beta_2 x^2 + \beta_1 x + \beta_0$.

e) List the value of coefficients of estimated regularized least squared polynomial regression models for each case.
f) Obtain the prediction on testing set and compute the RMSE for regularized least squared polynomial regression models for order M =1,2,3 and 9.
g) Plot the estimate obtained by regularized least squared polynomial regression models for order M =1,2,3 and 9 for training set along with $y_1, y_2, \ldots, y_{20}$. Also plot our actual mean estimate $E(Y/X) = \sin(2\pi x_i)$.
h) Plot the estimate obtained by regularized least squared polynomial regression models for order M =1,2,3 and 9 for testing set along with $y'_1, y'_2, \ldots, y'_{50}$. Also plot the $\sin(2\pi x'_i)$.
i) Study the effect of regularization parameter λ on testing RMSE and flexibility of curve and list your observations.