Assignment 1 - Trade-off between Overfitting and Underfitting

Question 1: Train ten regression models of increasing capacity (corresponding to M from 0 to 9) and record and compare their training and validation errors.

Table 1 and figure 1 showcase how training and validation losses change with different model capacities (M). Initially, both losses were very high, indicating an **underfitting.** As the value of M increased, both losses slowly decreased. At the value of (M = 4) there was a rapid decrease in losses. At (M = 7) the validation losses started increasing again while the training losses kept decreasing, indicating an **overfitting**. It can be seen that the optimal value of M is at 7, having a minimum validation loss of 0.14.

Table 1: Training and validation errors for predictors with different capacities

Model Capacity (M)	Training loss (RMSE)	Validation Loss (RMSE)
0	0.6838	0.7136
1	0.6615	0.6715
2	0.6615	0.6715
3	0.6581	0.6572
4	0.6579	0.6574
5	0.242	0.2622
6	0.2419	0.2621
7	0.0766	0.14
8	0.0616	0.1843
9	0.0002	0.3592

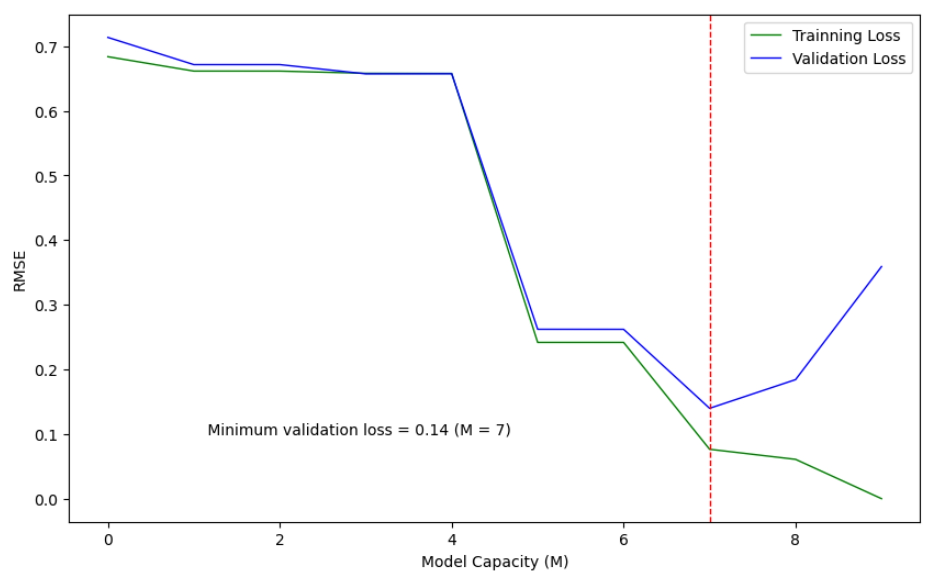

Figure 1: Comparing RMSE of different predictor capacities (M)

Question 2: For each M, plot the prediction $f_m(x)$ function and the curve $f_{true}(x)$ versus x from [0:1].

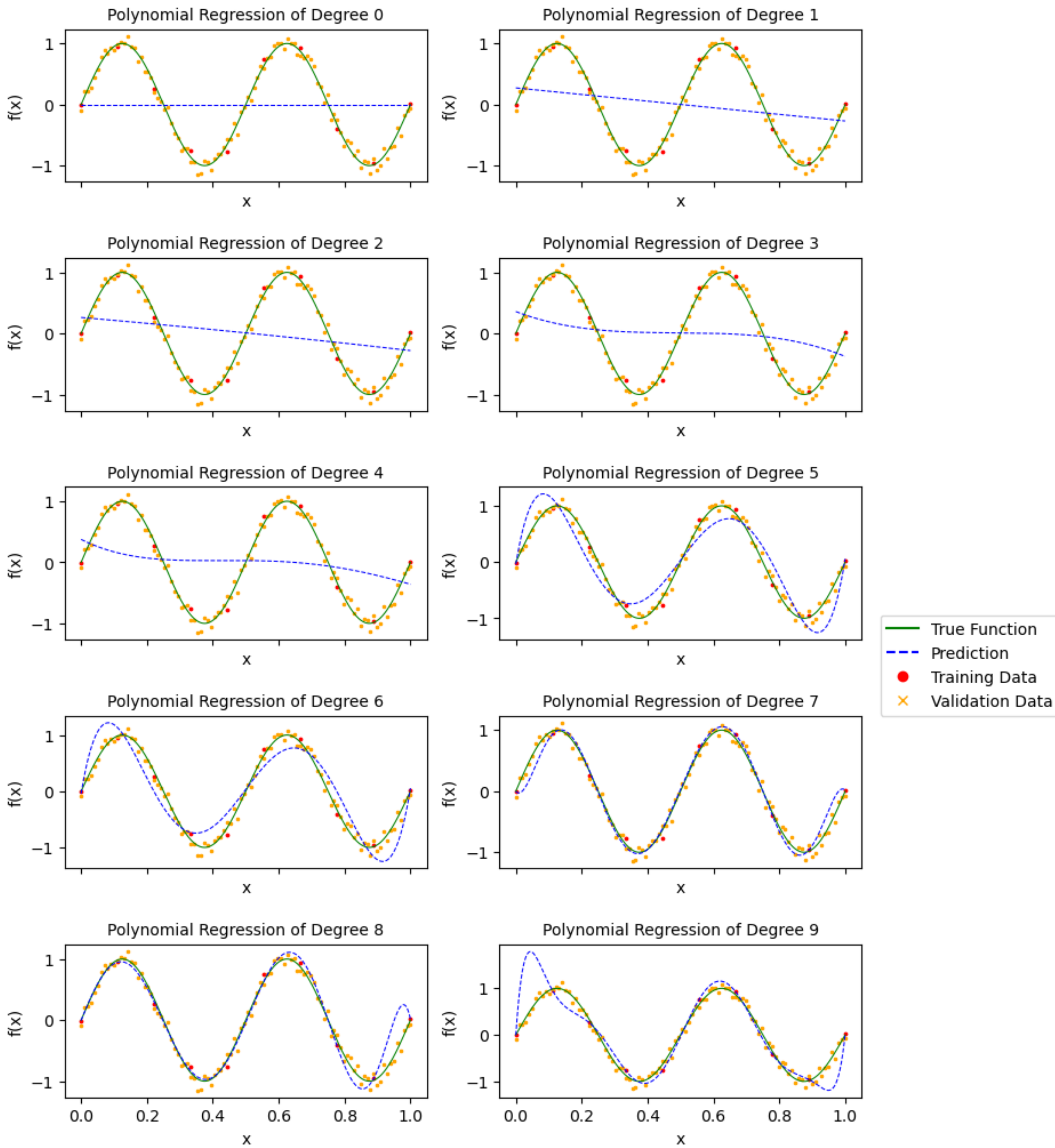

Figure 2: Graphing different capacity predictors against the true

Figure 2 shows that, as the value of M increases, the predictor starts to follow the behavior of the true function up until M = 7 where the predictor started following the noise of the training data, diverging from the true function.

Question 3: For each M = 9, retrain the model with regularization. Try more values for λ until you find an "optimal" value λ_1 that eliminates overfitting. Plot the training and validation errors versus different λ values and identify in the plot the underfitting and the overfitting regions. Identify a value λ_2 for which underfitting occurs.

The below figure 3 shows how different values of λ affected the training and validation losses while fitting a ninth-degree polynomial (M=9). Notice the x-axis uses a logarithmic scale. It can be seen that using small values of lambda log(λ) < -12.5 (blue region) didn't cause a noticeable regularization effect, keeping the original model overfitted with high validation loss and low training loss as have been in previous section. The meaning of overfitting is that as the model capacity is so high that it learned the noise within the sample data and deviated from the actual true function.

Beginning from around log(λ) = -12.5 to log(λ) = -7.69 (red region), the training loss started increasing and the validation loss started decreasing which means the overfitting started is vanishing, reaching the minimum validation loss of 0.134 at around log(λ) = -7.69. Starting from there and above (yellow region), both losses started to increase back again, indicating an underfitting. Underfitting means that the model capacity is so small to be able to mimic the distribution of the data samples.

From the previous findings, it can be concluded that the optimal λ_1 that eliminate overfitting is 2.048e-08 ($\log(\lambda_1) = -7.69$). There are many values for λ_2 for which underfitting occurs which starts from log(λ) > -7.69. Thus, I will pick an obvious one to do the next comparison which is $\lambda_2 = 1$ ($\log(\lambda_2) = 0$).

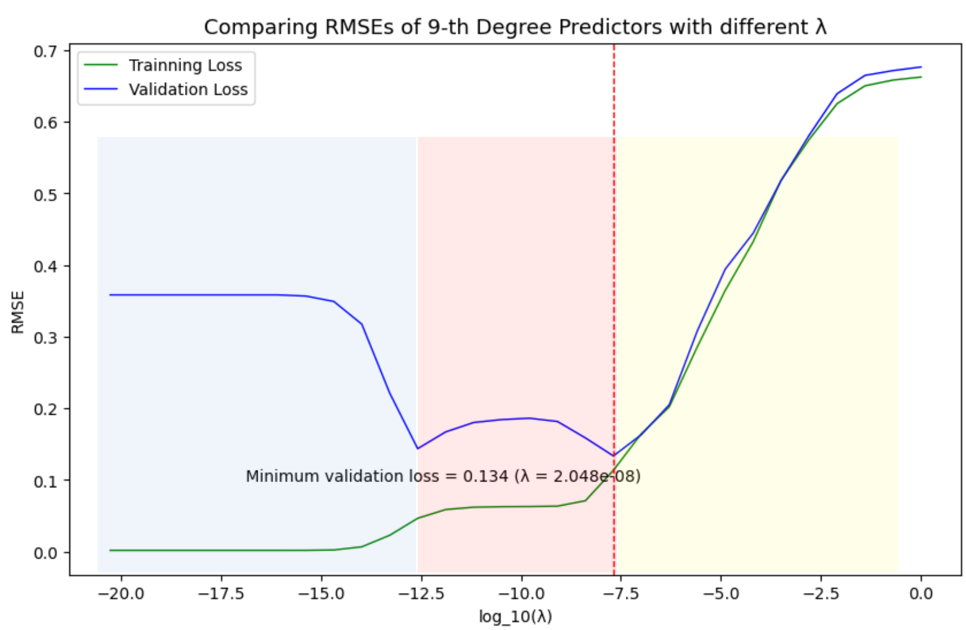

Figure 3: Comparing RMSE of 9-th degree predictors with different λ

Queston 4: For each of λ_1 and λ_2 plot the prediction function $f_m(x)$ and the curve $f_{true}(x)$ against x, and all the points in the training set and in the validation set with their true targets.

As can be seen in figure 4, the same capacity predictor can achieve astonishingly different results if the regularization parameter λ is tuned correctly. The graph on the left presents the optimal λ_1 that eliminates underfitting and on the right is a λ_2 that causes underfitting.

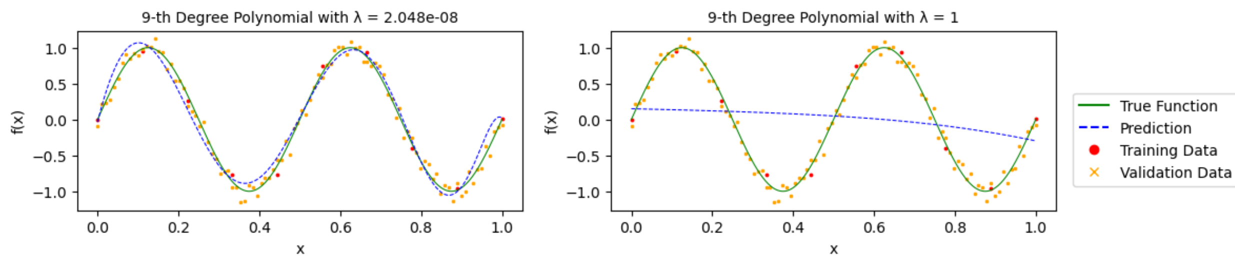

Figure 4: Graphing 9-th degree predictors with ridge regularization using λ_1 and λ_2

Queston 5: Among all models that you have trained select the best model. Then measure its performance on the test set.

From figure 1 and 3, it can be concluded that the best achieved model (the one with the least validation error) is 9-th degree polynomial with regularization parameter λ = 2.048e-08. To measure its performance on the test dataset, I standardized the test set and fed it to the model to generate the predictions and calculated the RMSE loss against the actual target.

Testing loss of best model: 0.14, against a previously calculated validation loss of 0.134.

Queston 6: Draw a plot of the training and validation root mean squared errors versus M, for all twelve predictor functions. Also include in this plot RMSE between the targets and the true function $f_{true}(x)$ for the examples in the validation set.