COMP ENG 4SL4: Machine learning

Assignment 1 - Trade-off between Overfitting and Underfitting

Instructor: Dr. Sorina Dumitrescu

TA: Ruilin Ji

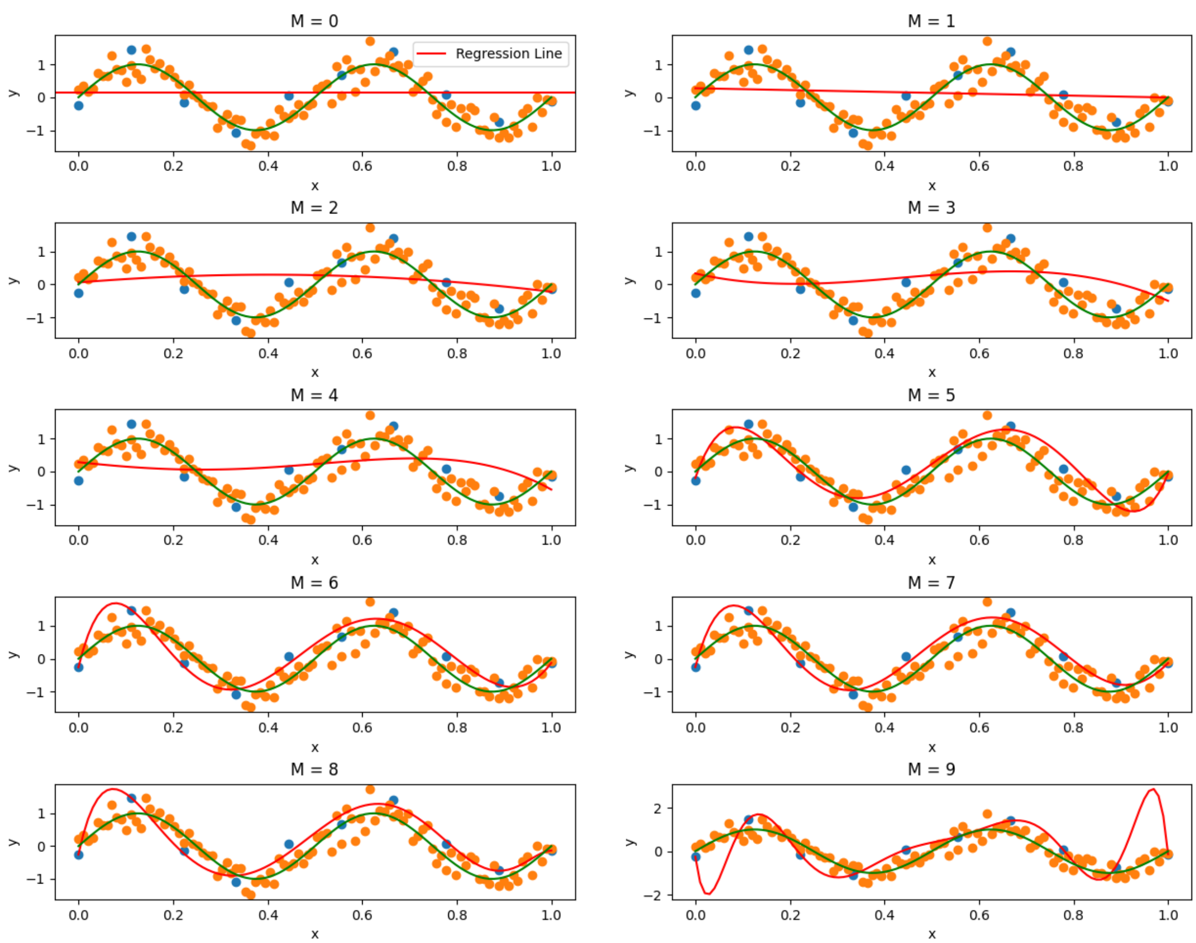
Richard Qiu – qiur12 – 400318681

Simulation without regularization

To train the model with the provided X_train and t_train samples over 10 different model capacities (denoted as *M*) from 0 – 9, the least squares linear regression method is used to find the best weights (denoted as *w*) of the training models. After finding out the w of each M, calculate the prediction points with below formula for each X_valid points and plot the curve.

$$f_M(x) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M .$$

Below graph display the information where orange points are the valid points, blue points are the training points, green line is the actual function f_true(x) and the red line is the predictor function fM(x).

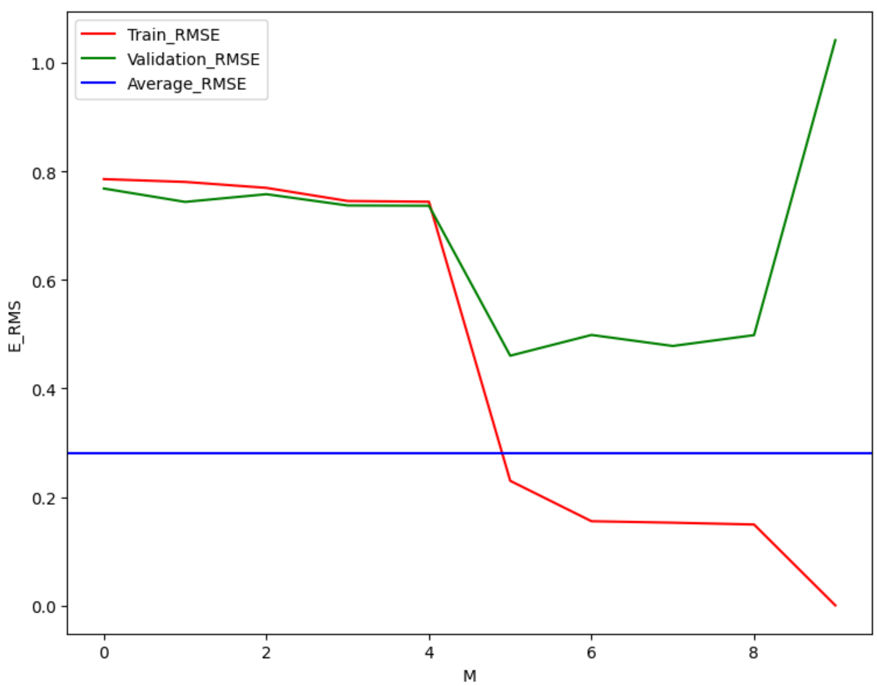


And to calculate the root mean squared errors, below formula is used (PRML Book 1.1 Example: Polynomial Curve Fitting) use the squared of predictor points minus actual points and calculate the squared root over the N samples.

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 \tag{1.2}$$

$$E_{\text{RMS}} = \sqrt{2E(\mathbf{w}^\star)/N} \tag{1.3}$$

The graph is obtained below. Where the blue line is a constant function were calculated by the average RMSE between the valid points and the true function.



As we can seen from the plot where as M increases, both training and validation errors get improved(decrease) and reach stable between 6 and 8, but the validation errors increase significantly after 8. In terms of underfitting/overfitting trade-off, below observations are found.

Overfitting region: 8 – 9 (RMSE starts increasing significantly for the validation error)

Underfitting region: 0 – 4 (RMSE is around 0.8 for both train validation error)

Optimal capacity region: 5 – 8 (RMSE is stable and around the minimum point of the graph)

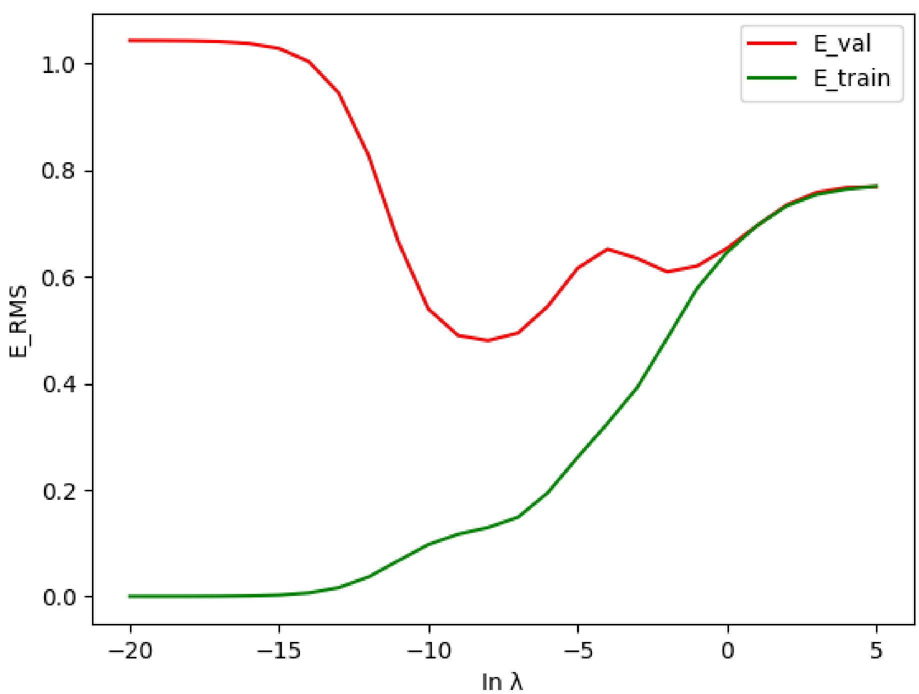
Simulation with regularization when M = 9

As we observed that the overfitting will happen when M = 9, a regularization term is applied to overcome the large magnitudes. After standardizing the linear model with provide code, the updated least squares linear regression method is used to calculate the w.

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X} + \tfrac{N}{2}\mathbf{B})^{-1}\mathbf{X}^T\mathbf{t}.$$

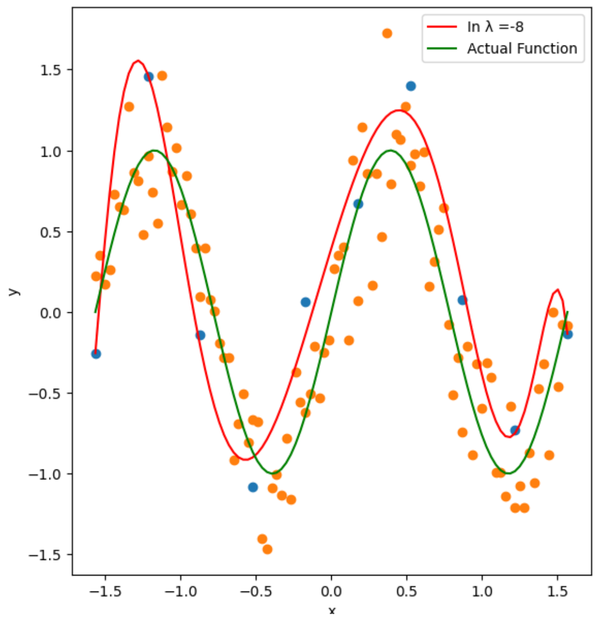
and $\mathbf{B}$ is the $(D+1)$-by-$(D+1)$ matrix where all nondiagonal elements are 0, $b_{00} = 0$ and the remaining diagonal elements are 2λ $(b_{11} = \cdots = b_{DD} = 2\lambda)$.

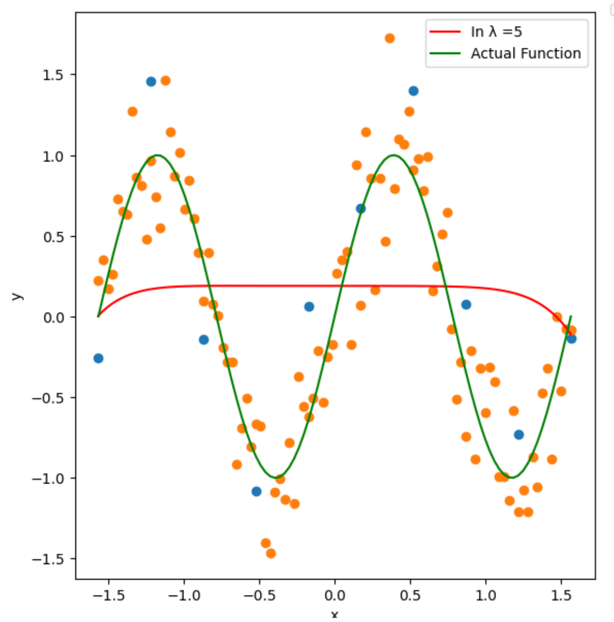
By tunning the In λ parameters, the best range to illustrate the underfitting/overfitting trade-off is from -20 to 5.



By observing the graph, the optimal capacity region (In λ) is between -10 and -7 where the validation error is around the minimum and keeps table. The overfitting region is between -10 and -oo where if the In λ is too small, there are no effects on decreasing the large weights and the RMSE will stay around 1.0. The underfitting region is between -7 and +oo where the regularized term affects too much on the original training data and the training error will increase and approach stable around 0.8.

The following parameters λ1 = -8 is chosen to eliminate the overfitting when M = 9 because it is at the minimum point of the above graph, and it is stable. λ2 = 5 is chosen where the underfitting occurs as the following trend lines after this point will reach the maximum and become stable around 0.8RMSE.





By plotting the respective λ graph, it is clearly to see that λ1 successfully eliminates the overfitting and provide a good predictor function while the λ2 is nearly a flat line that misses lots of the validation points which is a good example of the underfitting.

Model Selection and testing

To choose the best model, compare the two RMSE graph, when M = 7, it aligns with the same trend as the training error and the validation RMSE is at the minimum around 0.5 hence selected as the best model.

By generating 100 test points similar to the way generate validation points. Below graph is achieved when M = 7. As we can seen the predictor function has the same shape as the actual function and the RMSE is calculated equal to 0.5115211774171243 which verifies the validation result.