Statistical Machine Learning HW1 20170240 박기범

Question a.

Let’s set y\_hat = Xw, where X is data set and w is weight set. Then, we can express the loss function like this

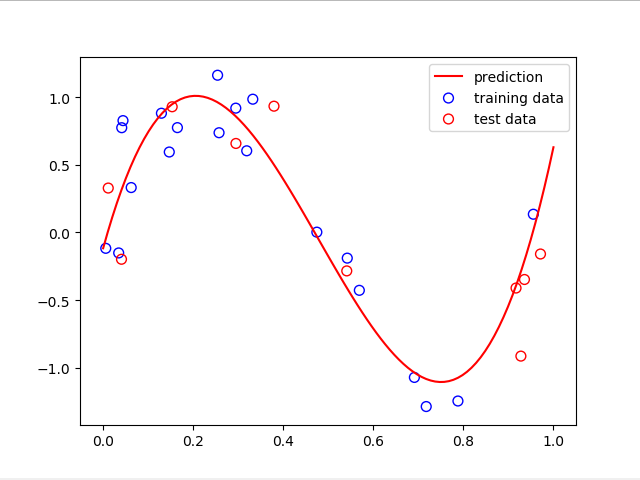
Since , we can rewrite the loss function like this.

We want to find the weights to minimize this loss function, so let’s derivate the loss function by w.

Then if is invertible,

Question b.

I chose the 3 dimension since it has the best average and std. Also if we look at the graph, the graph plots the trend of training data very well and also it has small bias between test data.



Question c.

For 3 dimension this result comes out and it was the best case among the other dimensions.

[Average MSE] : 0.0947

[MSE\_std] : 0.0396

Question d.

It follows the training data very well, however, there are some test data point which has large bias, in other words it is overfitted. Also, average of MSE and MSE standard deviation were 0.6764 and 0.8805 each which are very large compared to the best case above. It means that the 9 dimension has large error and also large variances in model. Therefore, if we use high dimension when training model, we always have to consider the overfitting. For example, we should evaluate the models using average MSE and standard deviation of MSE at the same time such as by using (average MSE+std).

