General feedback for Assignment 5

For Q1 and Q2, the most common mistake is that the value of the training errors is not included in the figure title, which is required in the questions.

For Q3, it is very important to notice that the validation data cannot be used for fitting a model (Also important for Q4, 5 and 6). For generalization, what we want to know is how a model performs on “unseen” data. The validation will be meaningless if the model already knew the answer. In addition, most of the students computed the correct errors for linear and quadratic models. However, it is also important to give your conclusion formally. For example, you should answer “The quadratic model performed better because it obtained a lower validation error”.

Nearly all students plotted the training errors and validation errors for different polynomial degrees in Q4,5 and 6. We did not mark the plots. However many students gave their answer according to their observations of the plots. Our suggestion here is plotting the validation errors in the logarithmic scale. Because the variation range of validation errors is large, the differences between validation errors among lower degrees are not clear in those figures. As a result, some considered that the performance of lower degree polynomial basis (e.g. 2-7) are equal. In fact, the validation errors are different if they were printed out. Therefore, plotting the validation errors in the logarithmic scale can help to visualize the change of validation errors clearer (as you can see in the sample answers we provided).

For Q6, it is important to randomize the (indices of) data points. So every time the partitions for the k-fold cross-validation can be different, and therefore the model selected can be different. It is good to use a loop to repeat 5-fold cross validation several runs to generate different results. Alternatively, answer in the text is also acceptable, e.g., “I run 5-fold cross validation several times, and it did not always select the same model. It is because of the random permutation. Different training, validation combinations can lead to different results.”

For normalizing the error in Q5 and 6, the overall error should be divided by the number of all data points. Using np.mean (or other mean function) is fine in Q5. However, it is incorrect in Q6 because the overall error will be divided by k, which equals 5.

--

This feedback was written by Chao Han, Shuo Zhou, Fariba Yousefi and Mauricio A Alvarez