

COMP 2019 Workbook Exercises Week 8 – ML Validation Answers

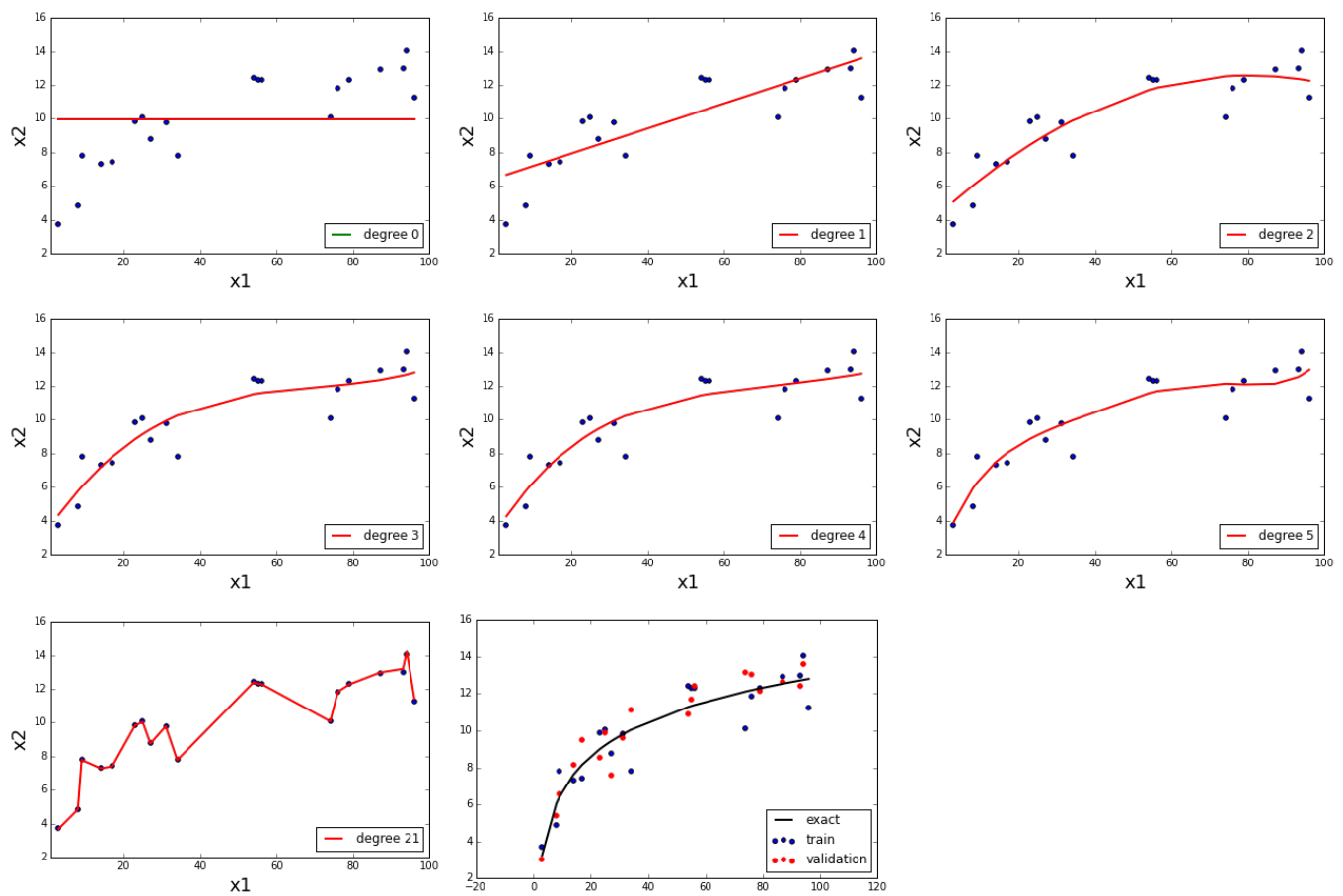
Question 1

M0-M1: both training and dev errors are high. High Bias. Add extra features.

M7-M21: training error is low, but dev error is high. High Variance. Add more training data, reduce features, or impose regularization.

Overall, M5 exhibits the least dev error. However, there is little difference between M2,3,4,5. We would probably choose M2, as this is the simplest model that exhibits good behaviour.

Here are plots of a few selected models that we can inspect, and a plot showing the training and dev data:



Question 2

We use cross-validation, as the degree of the best model is unknown. We first reserve samples to be used for testing of the final selected model (~20% of samples). The remaining samples will be used for training and dev set.

We loop to construct model of increasing complexity, starting with degree $d=0$. Since the dataset contains few samples, we use cross-validation to train each model and estimate its performance on unseen data. Let us use 5-fold cross-validation, so that we train 5 models for each degree, each trained and tested on a different partition of the data. In each split of the cross-validation, the training set includes 28 samples and the development set comprises 8 samples. Each time, we train on the training set and estimate the performance on the development set. We then average the estimates over all 5 models to find an estimate of the performance for each degree. (We repeat this 4 times, once for each degree d . Each time we obtain an estimate.)

We then select the degree of model based on these estimates. Once we know which degree yields the best performance, we then train a model of that degree on the entire training set.

Finally, we estimate the performance of the resulting model on unseen data using the test set.