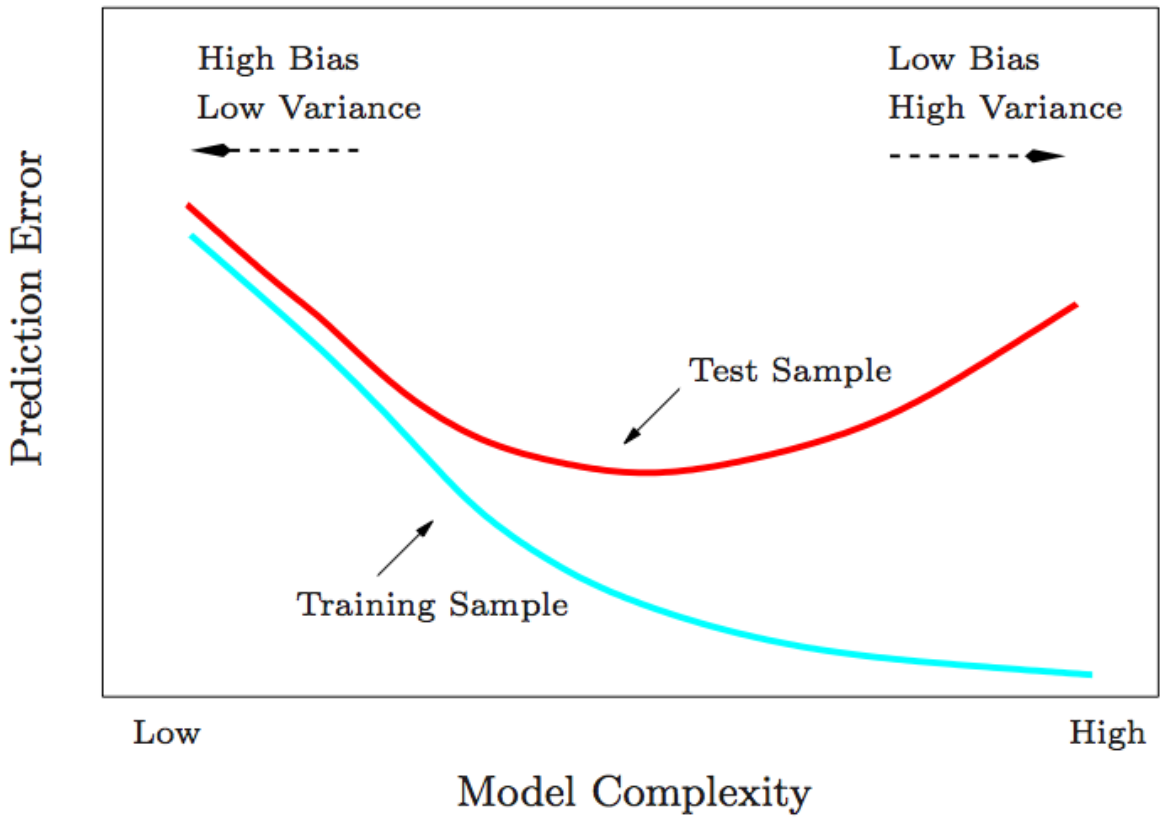
HW2 REPORT

Osman Başkaya

Cross validation is a common practice when one wants to see the system performance on unknown data. Cross validation avoids system overfits the data. If overfit occurs, system accuracy seems to be high but generalization power of the system will most likely be low. The figure below summarizes all these overfit-underfit scheme. Overfitting as known as high variance is in right side of the figure. As you seen, training error of the system is really low, on the other hand, test error is pretty high. Cross validation provides us more reliable performance scores of a system. We also use cross validation data parameter tuning like we did in this homework.



Grid search is for parameter tuning. SVM has one or more hyperparameters depending on what kernel you're using. Parameter tuning is very important for SVM. You may find useful

this practical guide for SVM. They shows real-world example to indicate how important parameter tuning is for SVM. Grid search is one the dumbest method we have for parameter tuning. It's actually line search in two or more dimension. In our example we try to find the best **cost** and **gamma** parameters by using grid search.

Some details on my implementation:

You can replicate my all work by writing **hw2** in the terminal window of Matlab. It creates all jpeg files for grid search heatmap figures, file that contains number of support vector for each different kernel and different chunk (cross validation data (question1)), writes all model files for each chunk. We make 5-fold cross validation so, for each different test chunk, I wrote 5 different SVM model for each kernel.

I made feature normalization which is very important for SVM (please see the practical guide).

I use no gamma parameter for Linear kernel since there is no gamma parameter for it. The reason why heatmap of linear kernel is a bit weird looking.

Important to note that, I tuned the hyperparameters according to all validation chunks for each different cost and gamma parameters. As you seen in the path, there should be 5 different heatmap for each kernel. When the estimate the best parameter setting, we need to consider all chunks scores. If this explanation is not enough, please ask me the details of my implementation.

Moreover, I also create a file shows which data points are correctly classified and which data points did not. For instance, **RBF-Accuracy-Figure.jpg** demonstrates the RBF kernel performance.