What is MSE?

\_\_The average squared difference between the estimated values and the actual value.

\_\_(how close a fitted line is to data point)

Training MSE

\_\_The MSE resulted from computing the training data is training MSE

\_\_We only care if the method doesn’t work on the training data. However, it is more important to care if the method works on test data.

\_\_This means we are interested in the accuracy of the predictions that we obtain when we apply our method to the previously unseen test data.

Test MSE

\_\_The squared difference between the estimate values and the actual test response.

Example

\_\_Goal is to find a method that can accurately predict diabetes risk for *future* patients based on their clinical measurements

\_\_Not interested in whether or not the method accurately predicts diabetes risk for patiens in training data, because we know they already have diabetes

The Bias-Variance Trade-Off

\_\_Expected test MSE is the average test MSE that we would obtain if we repeatedly estimated *f* using a large number of training sets and tested each at .

\_\_Three fundamental quantities: variance of , the squared bias of , and the variance of the error term epsilon.

\_\_This tells us that in order to minimize the expected test error, we need to select a statistical learning method that achieves low variance and low bias

\_\_variance inherently a nonnegative quantity and bias is also nonnegative. Therefore, expected test MSE can never lie below variance of the error term (epsilon)

\_\_Variance refers to the amount of which would change if we estimate it using a different training data set

\_\_different training data set used to fit the statistical learning method results in different .

\_\_if the method has high variance then small changes in the training data can result in large changes in .

\_\_more flexible statistical methods have higher variance.

\_\_Bias refers to the error that is introduced by approximating a real-life problem, which may be extremely complicated, by a much simpler model

\_\_Example: linear regression assumes that there is a linear relationship between predictors and response.

\_\_It is unlikely that a real-life problem truly has a simple linear relationship. This means using linear regression will result in some bias in the .

\_\_More flexible methods result in less bias.

In general:

\_\_When using a more flexible method, the bias decrease, but variance increase.

\_\_The relative rate of change of variance and bias determines whether the test MSE increase or decrease

\_\_As flexibility of the methods increases, the bias tends to initially decrease faster than the variance increase. Expected test MSE declines.

\_\_However, at some point, the derivative of bias will be close to 0 because the flexibility of the model can fit the training data easily.

\_\_At the same time, variance will rapidly increase due to the model being overfitting. expected test MSE increase.

Conclusion

\_\_This is a trade-off because it is easy to obtain a method with extremely low bias but high variance or vis versa

\_\_To conclude chapter 2.2, it is critical to select the correct statistical learning method that achieves *low* bias and *low* variance. It is also important to select correct level of flexibility.