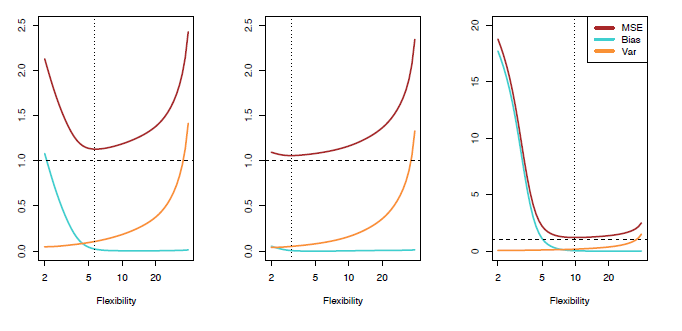
**CSCE 623 Spring 2019 Machine Learning In-Class Work, Day 4 (4 Apr 2019)**

From Chapter 2: Statistical Learning (to be completed when instructor indicates)

1. What is *n* in this equation? Why do we used squared error instead of mean error?



1. Give one or more sources of irreducible error
2. In your own words, what is flexibility? Give an example of a low flexibility model, and a high flexibility model:
3. **Training** MSE is often lower than **test** MSE, especially when the model is very flexible. Why?
4. In your own words, when discussing models, what is Variance? What is the relationship between Variance and Flexibility?
5. Why does increased variance lead to good performance on the training set but poor performance on the test set?
6. Without referring to your text, indicate which of these diagrams represent the performance of fitting various models to data which comes from a linear phenomenon… and which diagram represents the performance comes from a highly non-linear phenomenon?



From Chapter 2: K-Nearest Neighbors (to be completed when instructor indicates)

1. Consider the following training and test error rates on KNN as a function of 1/K. Describe why training performance gets better as K is reduced to 1, but Test performance gets better and then starts to get worse as K is reduced to 1

