Gregory Powers

Gxp145@case.edu

PQHS 471 Homework 1

Chapter 2 Q. 1 & 3

1. A. A flexible statistical learning method will likely perform better. Since the sample size is large, a model with more degrees of freedom can be estimated (i.e. less bias) without increasing variance (i.e. overfitting). That there are but few predictors will also tend to make overfitting less of a problem for a flexible model with a large number of observations.

   B. A flexible statistical learning method will likely perform worse. With a large number of predictors and a small number of observations, a flexible method will overfit. Further, flexible methods require more degrees of freedom (as they require more parameters to be estimated) compared to less flexible methods, something a small sample size may not be able to support.

   C. A flexible method will perform better. As the relationship is nonlinear, a model with many degrees of freedom is potentially going to better approximate the "true" data generating function.

   D. Flexible methods will perform worse than an inflexible method. Flexible methods are more apt to "chase the noise" in highly variable data and thus result in overfitting.
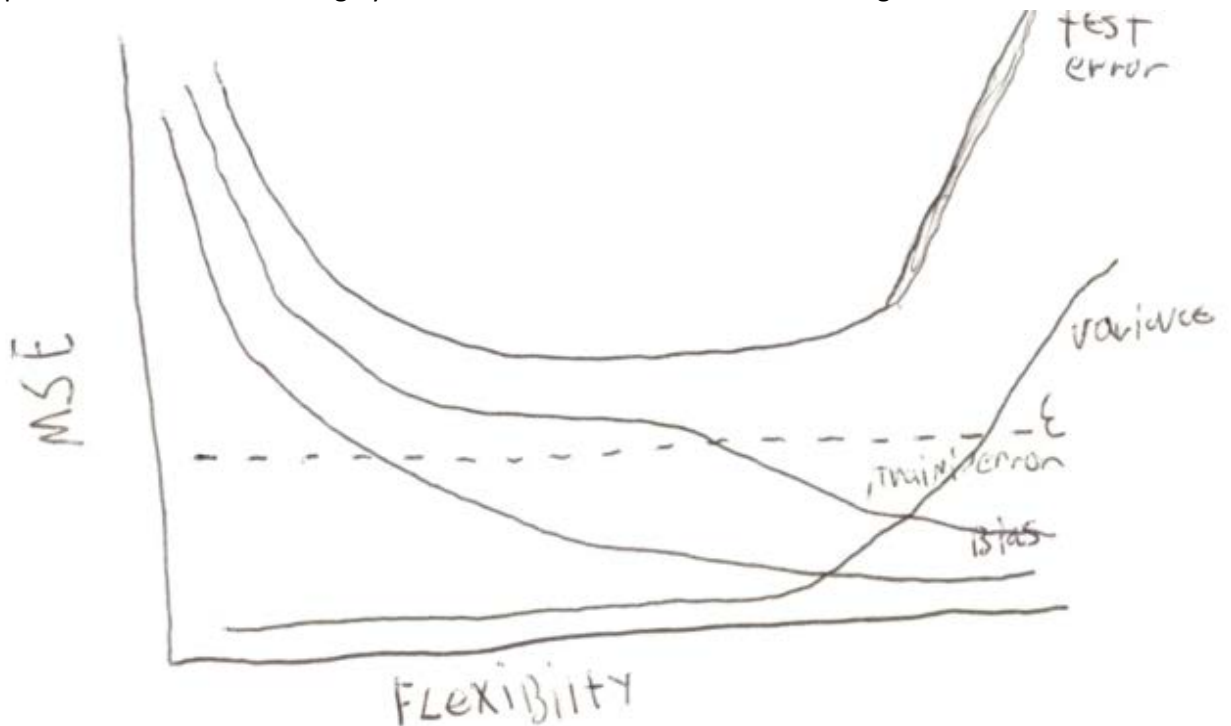
3.

A. Irreducible error (ϵ) is a straight line with 0 as its slope: it doesn't change regardless of model choice.

B. Test error is a concave up, U-shaped curve. Test error decreases as flexibility increases to the point where overfitting begins. When the model starts to become overfit, test error starts to increase.

C. Variance increases with higher flexibility. As models become more flexible, they also overfit decreasing generalizability (variance being differential fit of the same model across different datasets).

D. Training error decreases with higher flexibility: more flexible models are able to fit the data more exactly.

E. Bias will decrease with higher flexibility: models with more degrees of freedom can more closely approximate the real world data generating function.