STA314 Homework 1

student number: 1003942326

Yulin WANG

26/09/2019

Question 1

(a)

Flexible statistical learning method would perform better,
because sample size is large enough to fit more parameters and small number of predictors limit the model variance. And flexible models can fit the data closer without worrying about overfitting.

(b)

Flexible statistical learning method would perform worse,
because it will be more likely to overfit small number of observations.

(c)

Flexible statistical learning method would perform better,
because it is less restrictive(with higher degrees of freedom) on the shape of the fit.

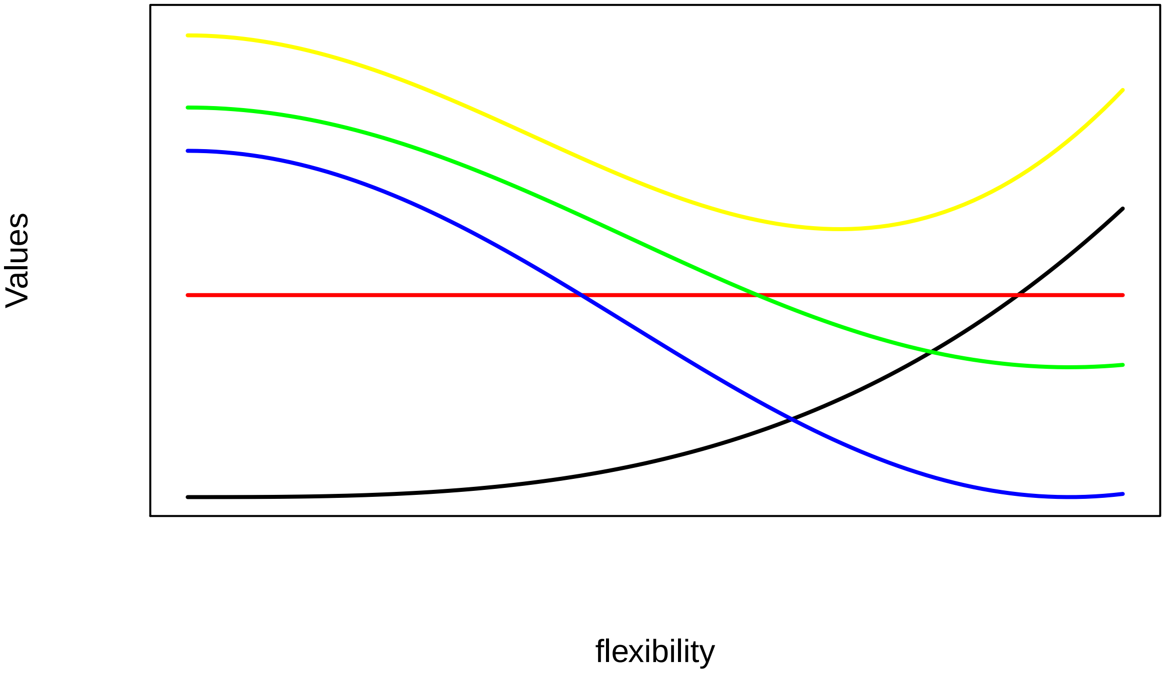
(d)

Flexible statistical learning method would perform worse,
because it will be more likely to overfit (since it will follow the high variance closely).

Question 2

(a)

Here are the five curves:



Here are the corresponding labels:

- `yellow` = test error
- `green` = training error
- `blue` = typical(squared) bias
- `red` = (Bayes) irreducible error
- `black` = variance

(b)

1) Test error will have a concave up(U-shaped) curve,
   because it reflects the interaction between variance and bias.

2) Trainig error will decrease monotonically as flexibility increases,
   because increasing flexibility will result in overfitting which will produce lower MSE on the training data.

3) Typical(squared) bias will decrease monotonically as flexibility increases,
   because there are fewer assumptions made about the shape of the fit.

4) (Bayes) irreducible error will stay constant regardless of the model fit,
   because we cannot avoid the systematic error.

5) Variance will increase monotonically as flexibility increases,
   because changing data points will have more effect on the parameter estimates and increasing flexibility will result in overfitting.

Question 3

Advantages:

A very flexible approach will fit the data better with fewer prior assumptions.

Disadvantages:

It would be hard to interpret the very flexible models and the prediction accuracy may decrease due to overfitting.

Circumstances:

1) A more flexible approach might be preferred when the training data is very complex or we mainly care about the result rather than the inference.

2) A less flexible approach might be preferred when the data has a simple shape or the inference is very important.

Question 4

Differences:

For parametric approach, we assume the functional form of f while the non-parametric methods do not make explicit assumptions about the functional form of f.

Advantages:

The problem of estimating f reduces to one of estimating a set of parameters without fitting an entirely arbitrary function f and do not need as many observations as non-parametric approach.

Disadvantages:

Choosen model might not match the true unknown form of f and it is too far from thr true f then our estimate will be poor.