MATH4432 Homework 1 [Conceptual Parts]

Zhang Zhe 20866321

September 24, 2024

Chapter 2

1. For each of parts (a) through (d), indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an infexible method. Justify your answer.

(a) The sample size n is extremely large, and the number of predictors p is small.

(b) The number of predictors p is extremely large, and the number of observations n is small.

(c) The relationship between the predictors and response is highly non-linear.

(d) The variance of the error terms, i.e. $\sigma^2 = \text{Var}(\varepsilon)$, is extremely high.

Answer
(a) (b)

(c) The performance for flexible statistical learning method will in general be better, it is because a flexible method is able to fit into more data points, whilst an inflexible method is unable to fit most of the data points, and this brings underfitting issue.

(d) If the variance of the error terms is high, it indicates that the sample contains a lot of noises, and performance of a flexible statistical learning method will be worse, since the model may fit into these unwanted noises.

3. We now revisit the bias-variance decomposition.

(a) Provide a sketch of typical (squared) bias, variance, training error, test error, and Bayes (or irreducible) error curves, on a single plot, as we go from less fexible statistical learning methods towards more fexible approaches. The x-axis should represent the amount of fexibility in the method, and the y-axis should represent the values for each curve. There should be five curves. Make sure to label each one.

(b) Explain why each of the five curves has the shape displayed in part (a).