Assignment 1_conceptual

Pong Yui Yi Monica

2024-10-7

question 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 inflexible 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. $\sigma^2 = Var(\epsilon)$, is extremely high.

answer:

This question is about the bias-variance trade-off. Denoted as

$$E[(Y - \hat{f}(x))^2 | X = x_0] = \sigma_\varepsilon^2 + \text{Bias}^2[\hat{f}(x_0)] + \text{Var}[\hat{f}(x_0)]$$

A flexible method is expected to perform with higher variance but lower bias, while an inflexible method is expected to perform with lower variance but higher bias. Our target is to minimize the test error, hence we want to balance the bias and variance.

1. **Flexible method is expected to be better.** when the sample size n is large, the data is complicated, and the flexible method can capture the complexity of the data and fit well, since small changes in the data would not greatly affect the whole dataset, so the error due to variance is low, which may hard to lead to overfitting. Meanwhile, since the number of predictors p is small, the computational cost of using a more flexible method would not be a big issue, so the flexible method is expected to be better.
2. **Inflexible method is expected to be better.** Opposite to the situation in part (a), having a large number of predictors makes flexible methods more computationally costly. At the same time, the small number of n means that small changes in the data could result in large changes in f, which may lead to high variance and overfitting. Therefore, in order to reduce the computational cost and also reduce the potential for error due to variance, inflexible method is expected to be better.
3. **Flexible method is expected to be better.** Since the relationship between the predictors and response is highly non-linear, a flexible method would be able to capture and fit to the non-linearity and reduce the bias, and error introduced by approximating a real-life problem. If we choose inflexible method, it would be more likely to underfit the data becuase of its non-linearity, and the error introduced by bias would be higher than the variance which leads to higher test error.
4. **Inflexible method is expected to be better.** Under the given information, we cannot determine whether the bias is high or low, however, since the variance is extremely high, we assume the variance is far more higher than the bias (and perhaps it's low), a more flexible model will fit the noise and lead to a much more higher variance, and finally turn out to be overfitting, hence, we expected to first use

the inflexible method to obtain the balance of variance and biase, and then do the further analysis and obtain more information.

question 2.3

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 flexible statistical learning methods towards more flexible approaches. The x-axis should represent the amount of flexibility 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).

answer:

(a).

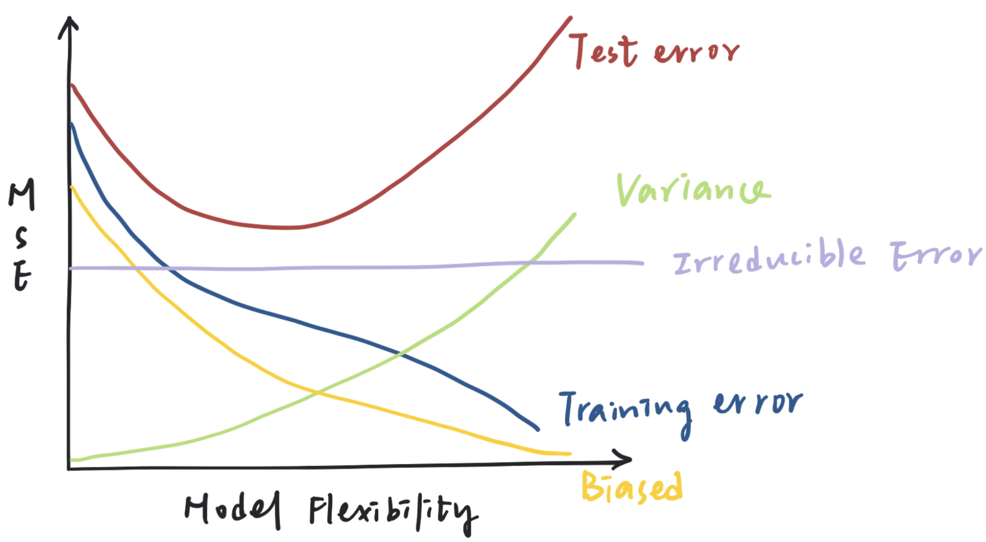

Figure 1: Bias-Variance Decomposition Curves

(b).
The bias generally starts off high for methods with low flexibility (unless the true f is close to linear) and then decreases down to zero as flexibility increases. On the other hand, variance starts out at zero (since the least flexible approach is assuming f takes a single constant value) and then increases as flexibility increases, since flexible methods are susceptible to changing considerably when any single data point changes. Training error starts out at some non-zero value, with magnitude depending on the linearity of the true f . As flexibility increases, it decreases down to zero. Similar to the training error, test error starts out at some non-zero value, with magnitude depending on the linearity of the true f . As flexibility increases, it then decreases down to a minimum (but never below the irreducible error) before increasing again. Lastly, the irreducible error is a constant non-zero value which is the absolute minimum possible test error (though that minimum need not be achieved).

question 3.1

question 3.5