**Problem 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.

1. The sample size is extremely large, and the number of predictors is small.

**Ans:** In this case, a flexible statistical learning method will work better because it will be able to handle the fluctuations and trend in the data better than its inflexible counterpart, such as a simple linear model. Also, due to the large sample size, there are less chances of overfitting and with a flexible method we will be able to reduce the bias as well.

1. The number of predictors is extremely large, and the number of observations is small.

**Ans:** Here, an inflexible method will work better (in general) because the low sample size may cause overfitting issues for a flexible method.

1. The relationship between the predictors and response is highly non-linear.

**Ans:** In general, flexible methods work best when high non-linearity is present between the predictors and response. This is because inflexible methods would fail to capture the data patterns resulting in high bias.

1. The variance of the error terms, i.e. , is extremely high.

**Ans:** In this case, an inflexible method will perform better because a flexible method will tend to capture extra noise in the data resulting from the high variance in the error terms.

**Problem 5:** What are the advantages and disadvantages of a very flexible (versus a less flexible) approach for regression or classification? Under what circumstances might a more flexible approach be preferred to a less flexible approach? When might a less flexible approach be preferred?

**Ans:***Advantages of very flexible methods:*

1. Capability to handle large observations and make accurate predictions or classifications (low bias).
2. These methods can handle high non-linearity between the predictor and response variables.

*Disadvantages of very flexible methods:*

1. These methods often act like black boxes and can be difficult to interpret.
2. These tend to overfit data when the sample size is small.

*Advantages of less flexible methods:*

1. Easy to interpret.
2. Can handle datasets having large variances in the error terms.

*Disadvantages of less flexible methods:*

1. Leads to large bias when sample size is large.
2. Cannot capture non-linearity between predictor and response variables.

In general, very flexible methods work best when:

1. The sample size is extremely large, and the number of predictors is small.
2. The relationship between the predictors and response is highly non-linear.

Similarly, less flexible methods work best when:

1. The number of predictors is extremely large, and the number of observations is small.
2. The variance of the error terms is extremely high.

**Problem 7:** The table below provides a training data set containing six observations, three predictors, and one qualitative response variable.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Obs. |  |  |  |  |
| 1 | 0 | 3 | 0 | Red |
| 2 | 2 | 0 | 0 | Red |
| 3 | 0 | 1 | 3 | Red |
| 4 | 0 | 1 | 2 | Green |
| 5 | -1 | 0 | 1 | Green |
| 6 | 1 | 1 | 1 | Red |

Suppose we wish to use this data set to make a prediction for when using -nearest neighbors.

1. Compute the Euclidean distance between each observation and the test point, .

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Obs. |  |  |  |  | Distance from test point (rounded to 3 decimal places) |
| 1 | 0 | 3 | 0 | Red | 3.000 |
| 2 | 2 | 0 | 0 | Red | 2.000 |
| 3 | 0 | 1 | 3 | Red | 3.162 |
| 4 | 0 | 1 | 2 | Green | 2.236 |
| 5 | -1 | 0 | 1 | Green | 1.414 |
| 6 | 1 | 1 | 1 | Red | 1.732 |

1. What is our prediction with ? Why?

**Ans:** With , the predicted class will be Green because if single neighbors are considered then observation 5 is the nearest point having = ’Green’.

1. What is our prediction with ? Why?

**Ans:** Here, the three nearest neighbors belong to observations 5, 6, and 2 respectively. Since Red occurs more frequently in this cluster, so the prediction will be Red.

1. If the Bayes decision boundary in this problem is highly nonlinear, then would we expect the best value for to be large or small? Why?

**Ans**: In this problem, if the Bayes decision boundary is highly nonlinear, then the best value for is expected to be small. Otherwise, if is large, we would obtain a smoother boundary resulting in possible misclassification.