18101246 이사빈

Exercises / Conceptual

1.

First of all, we have to define what the flexible statistical learning method is. Flexible statistical learning refers to the use of machine learning algorithms that are able to fit complex nonlinear models to data, while also avoiding overfitting by using regularization techniques.

Therefore, a flexible statistical learning method should typically perform better than an inflexible method on difficult problems.

The choice of whether to use a flexible or inflexible statistical learning method depends on the specific problem at hand and the tradeoff between bias and variance that is desired. If the goal is to achieve the best possible predictive accuracy and the dataset is large enough to support a complex model, then a flexible method may be preferred. However, If interpretability or simplicity is important, or if the dataset is small, an inflexible method may be appropriate.

Respect to the size of datasets which we could observe or sample and variance, (a), (c), (d) would be better using a flexible statistical learning method. on the other hands, (b) is preferred to use an inflexible statistical learning method due to the number of observation n is small.

2.

Classification and regression are two types of supervised learning tasks in machine learning. The main difference between classification and regression is the type of output they produce.

The Output types are different. regression models predict a continuous numerical value as the output, such as stock price or etc. however classification models predict a categorical class as the output, such as classifying images as either cats or dogs or etc.

Because of the reason, (a), (c) would be the regression tasks as well.

Furthermore, in (a), they are interested in understanding which factors affect CEO salary. When we try to figure something latent or invisible things out from an observed data. We could regard it as an inference. not prediction. and, The inference focuses on understanding the relationship between variables and drawing conclusions about the underlying mechanisms.

Therefore, in (b), the task would be classification and focus on prediction. in (c), they are interested in predicting the % change in the USD/Euro exchange rate in relation to the weekly changes in the world, stock markets.

6.

The parametric statistical learning approach lies in the assumption made about the underlying data distribution.

Parametric statistical learning methods assume that the data follows a specific parametric distribution. such as a normal distribution and the parameters of the distribution are estimated from the data. If the assumption of the underlying distribution is incorrect, the model may not fit the data well.

Non-parametric statistical learning methods do not assume a specific distribution for the data. Instead, they aim to learn the underlying relationship between the predictors and the response variable from the data itself. This method can be more flexible and can capture complex relationships between variables, but may require more data and ca be computationally intensive.

There are few advantages of a parametric approach to regression or classification.

First of all, the model can be simpler and more efficient computationally than a non-parametric model. Also parametric models typically have a clear mathematical form and can provide insight into the relationship between the predictor variables and the response variable.

On the other hand, what if the assumptions are incorrect, the model may not fit the data well and can result in biased or inaccurate predictions.

7.

1. 3, 2, , , , (for, Obs\_1,2,3,4,5,6)
2. Green, because the 5th observation is the nearest neighbor to the test point
3. Red, because the top 3 nearest neighbors are Obs2, 5, 6 which are red, green, red
4. Small, in a highly non-linear problem, we would expect the best value for K to be small to capture the local structure of the data and adapt to the non-linear decision boundary.