PS02

Melih Karaköse, Fethi Kahvecioğlu

March 20, 2024

1)

a) Extremely Large Sample Size (n) and Small Number of Predictors (p)

For an extremely large sample size n and a small number of predictors p, we would generally expect a flexible statistical learning method to perform better than an inflexible method.

Justification

When the sample size is large, flexible methods can capture complex patterns and relationships in the data without overfitting. The large sample size provides enough information for the flexible method to estimate the complex model accurately. On the other hand, inflexible methods may not be able to capture the true underlying relationship, leading to higher bias.

b) Extremely Large Number of Predictors (p) and Small Number of Observations (n)

For an extremely large number of predictors p and a small number of observations n, we would generally expect an inflexible statistical learning method to perform better than a flexible method.

Justification

When the number of predictors is much larger than the number of observations, flexible methods are prone to overfitting. With a small sample size, flexible methods may capture noise and random fluctuations in the data, leading to poor generalization performance. Inflexible methods, which have fewer parameters, are less likely to overfit in this scenario.

c) Highly Non-linear Relationship between Predictors and Response

If the relationship between the predictors and the response is highly non-linear, we would generally expect a flexible statistical learning method to perform better than an inflexible method.

Justification

Flexible methods, such as non-parametric models or neural networks, can capture complex non-linear relationships more effectively than inflexible methods like linear regression. Inflexible methods assume a specific form for the relationship (e.g., linear), which may not be appropriate for highly non-linear relationships, leading to higher bias.

d) Extremely High Variance of Error Terms ($\sigma^2 = \mathbf{Var}(\epsilon)$)

If the variance of the error terms is extremely high, we would generally expect an inflexible statistical learning method to perform better than a flexible method.

Justification

When the variance of the error terms is high, the signal-to-noise ratio in the data is low, making it challenging to estimate complex models accurately. Flexible methods may capture noise and random fluctuations in the data, leading to overfitting. Inflexible methods, which have fewer parameters, are less likely to overfit and may perform better in such scenarios by capturing the underlying signal more robustly.

1)

a)

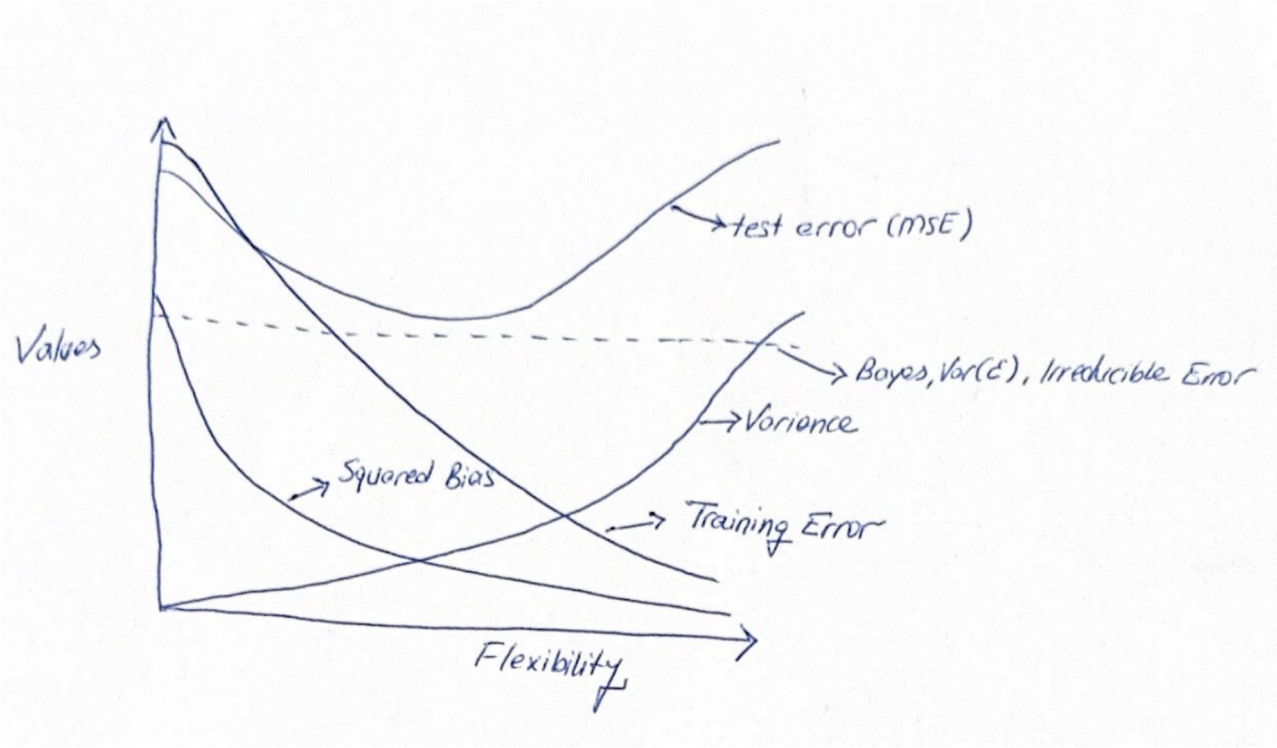

Figure 1: Statistical Curves

1. **Training Error**
   - **Low Flexibility:** With a simple model, the training error is high because the model cannot capture the complexity of the data.
   - **Increasing Flexibility:** As the model becomes more complex, it fits the training data better, and the training error decreases.
   - **High Flexibility:** At high complexity, the training error is very low because the model can fit almost perfectly to the training data.
2. **Test Error (MSE)**
   - **Low Flexibility:** Initially, the test error (MSE) is high due to underfitting; the model is too simple and has high bias.

- **Optimal Flexibility:** As flexibility increases, the model generalizes better, and the test error decreases to its lowest point, achieving the best generalization.
- **High Flexibility:** Further increasing flexibility leads to overfitting; the model starts capturing noise as if it were a true signal, causing the test error to rise again.

3. **Squared Bias**
   - **Low Flexibility:** Squared bias is high because a simple model makes strong assumptions about the data's shape that do not match reality.
   - **Increasing Flexibility:** As the model becomes more complex, these assumptions are relaxed, and the squared bias decreases.
   - **High Flexibility:** At very high complexity, the squared bias is very low because the model no longer relies on strong assumptions about the data's shape.
4. **Variance**
   - **Low Flexibility:** The variance is low because a simple model doesn't change much with different training data.
   - **Increasing Flexibility:** As the model becomes more complex, it becomes more sensitive to fluctuations in the training data, so the variance increases.
   - **High Flexibility:** At high complexity, the variance is high because the model is over-sensitive to the training data, capturing random noise.
5. **Bayes(Var(e)), Irreducible Error**
   - This represents the error that cannot be reduced regardless of the model's complexity. It's the noise inherent in the data, and even the best model cannot reduce this error.

b)

1. **Training Error**
   - **Shape:** Decreasing curve.
   - **Reason:** As model complexity increases, the model fits the training data more closely, leading to a decrease in training error. This is because a more complex model has more parameters to adjust to the data.
2. **Test Error (MSE)**
   - **Shape:** U-shaped curve.

- **Reason:** Initially, increasing model complexity reduces the test error because the model generalizes better. However, past a certain point, the model starts to overfit, learning the noise in the training data as if it were true signal, which increases the test error.

3. **Squared Bias**
   - **Shape:** Decreasing curve.
   - **Reason:** Bias is the error from erroneous assumptions in the learning algorithm. With low complexity, the model makes strong assumptions, leading to higher bias. As complexity increases, the model makes fewer assumptions, thus reducing the squared bias.

4. **Variance**
   - **Shape:** Increasing curve.
   - **Reason:** Variance measures how much the model's predictions would change if we used a different training dataset. A more complex model is more sensitive to the data it's trained on, so its variance increases with model complexity.

5. **Bayes(Var(e)), Irreducible Error**
   - **Shape:** Horizontal line.
   - **Reason:** This is the error that cannot be reduced by the model, no matter how complex it is. It represents the noise inherent in the data, so it remains constant regardless of model complexity.

c)

As can be seen in the figure above, when the model is made more complex by increasing the number of parameters, the test error initially decreases slightly but then increases. However, when this parameter and hence complexity increases excessively (billions of parameters), this test error decreases after a certain threshold value. Since it is not known what causes this drop and there is no proof, it remains a phenomenon.

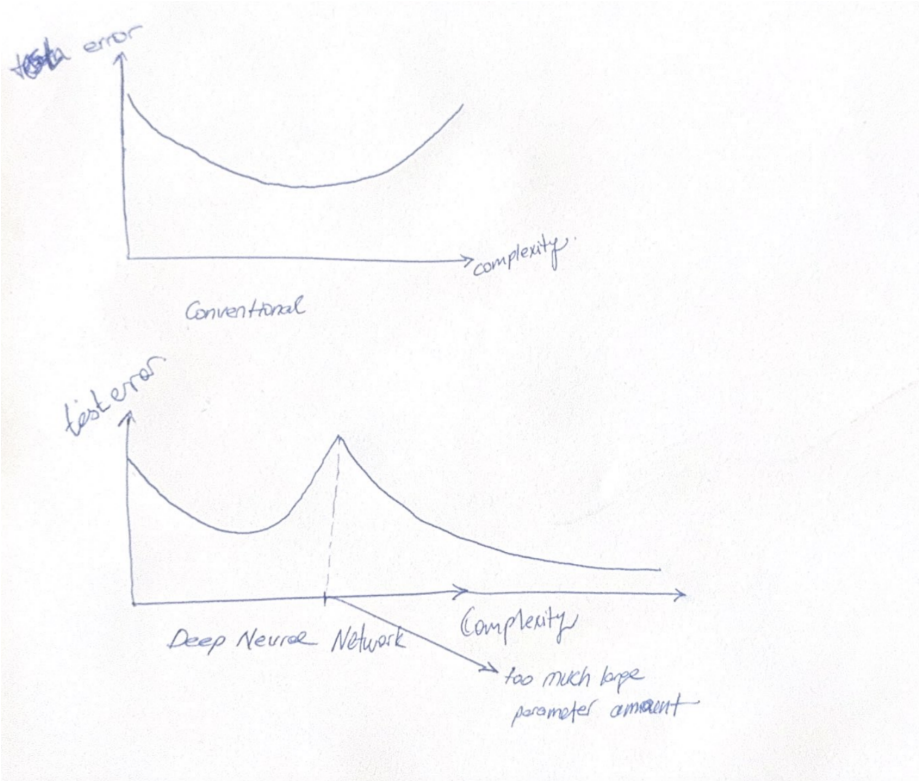

Figure 2: Statistical Curves

This phenomenon is used today by LLM models such as GPT3.5, GPT4. Since these models have billions of data to use, they can train their models and reduce the test error considerably even if they have billions of parameters.