

Overfitting and Underfitting in Machine Learning

Examples and Mitigation Strategies

Kazeem Adesina DAUDA

AI4WIA BOOTCAMP

September 9, 2023

Table of Contents

- 1 Introduction
- 2 Overfitting Example
- 3 Underfitting Example
- 4 Good fit
- 5 Mitigation Strategies
- 6 Conclusion

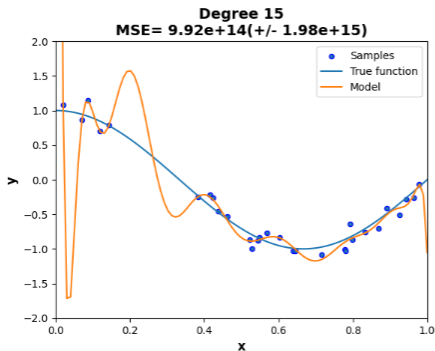
Overfitting and Underfitting

Overfitting and underfitting are common problems in machine learning that affect model performance.

- Overfitting: When a model learns the training data too well but fails to generalize to new, unseen data.
- Underfitting: When a model is too simple to capture the underlying patterns in the data, leading to poor performance on both training and test data.

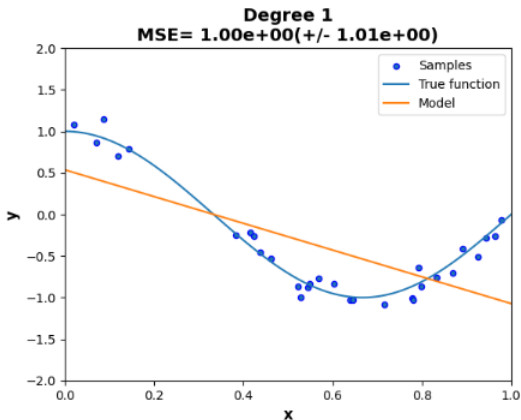
Overfitting Example

- Let's consider an example where we fit a high-degree polynomial to a small dataset.
- R Code for Overfitting:



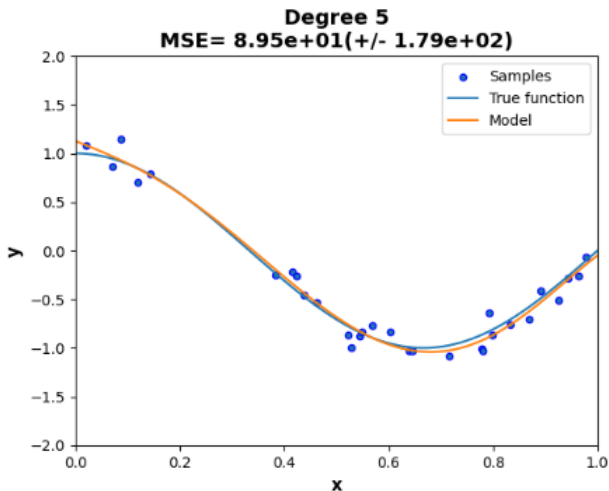
Underfitting Example

- In this example, we use a linear model for a complex dataset.
- R Code for Underfitting:



Good fit Example

A good fit implies that the model captures the underlying patterns in the data without overfitting or underfitting



How to Address Overfitting and Underfitting

- Cross-Validation: Use techniques like k-fold cross-validation to evaluate model performance on different data subsets.
- Regularization: Apply L1 or L2 regularization to penalize complex models.
- Feature Selection: Choose relevant features and eliminate irrelevant ones.
- Ensemble Methods: Combine multiple models to reduce overfitting and improve generalization.

Conclusion

- Overfitting and underfitting are common challenges in machine learning.
- Understanding these issues is crucial for model selection and performance optimization.
- By using appropriate strategies, we can strike a balance between model complexity and generalization.

Questions?