Underfitting, Overfitting, and Regularization in Machine Learning

Bisma Fajar

August 2024

Abstract

This report explores the concepts of underfitting, overfitting, and regularization in machine learning. We will discuss the causes and consequences of both underfitting and overfitting, and explore various regularization techniques to mitigate these issues.

1 Introduction

In machine learning, the goal is to create models that generalize well to unseen data. A model that is too simple may not capture the underlying patterns in the data, leading to *underfitting*. Conversely, a model that is too complex may fit the noise rather than the signal, leading to *overfitting*. To address these issues, *regularization* techniques are employed to find a balance between underfitting and overfitting.

2 Underfitting

Underfitting occurs when a model is too simple to capture the underlying structure of the data. This usually happens when the model has insufficient capacity or is too constrained. An example of underfitting is using a linear model to fit non-linear data.

3 Overfitting

Overfitting occurs when a model learns not only the underlying pattern but also the noise in the data. This often happens when the model is too complex

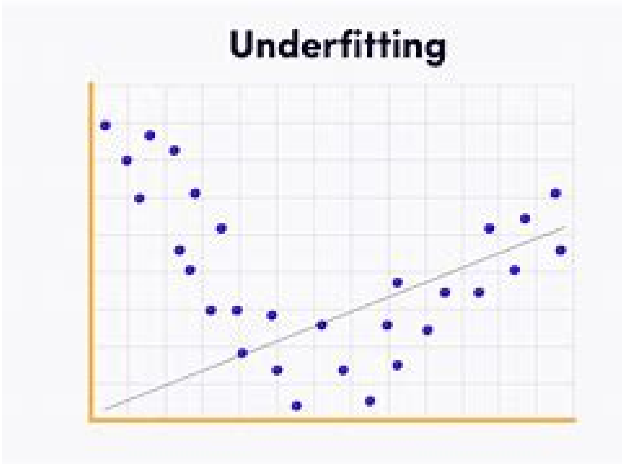

Figure 1: Example of Underfitting: A linear model on non-linear data.

relative to the amount of training data. An example of overfitting is using a high-degree polynomial to fit data that is essentially linear.

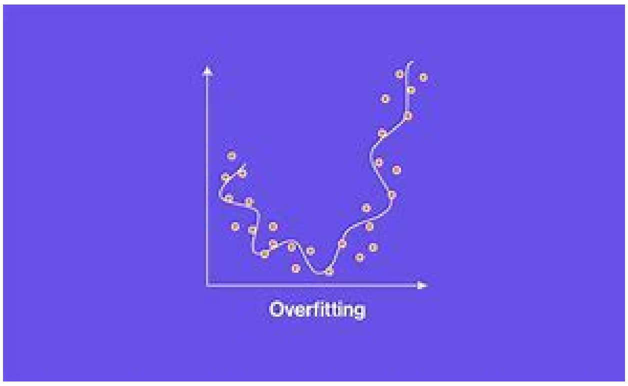

Figure 2: Example of Overfitting: A high-degree polynomial model fitting noise in the data.

4 Regularization Techniques

Regularization techniques are used to prevent overfitting by adding a penalty for model complexity. Two common regularization techniques are L1 (Lasso) and L2 (Ridge) regularization.

4.1 L1 Regularization

L1 regularization adds a penalty equal to the absolute value of the magnitude of coefficients, effectively shrinking some coefficients to zero. This can be useful for feature selection.

$$\text{L1 Loss} = \sum_{i=1}^{n} |w_i| \tag{1}$$

4.2 L2 Regularization

L2 regularization adds a penalty equal to the square of the magnitude of coefficients, which prevents the coefficients from becoming too large.

$$\text{L2 Loss} = \sum_{i=1}^{n} w_i^2 \tag{2}$$

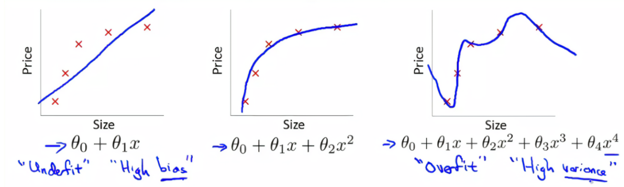

Figure 3: Effect of Regularization: The model complexity is reduced to prevent overfitting.

5 Conclusion

Understanding and addressing underfitting and overfitting are crucial for developing robust machine learning models. Regularization techniques provide a powerful tool to mitigate these issues by controlling model complexity.

References