Overfitting and Underfitting

Introduction

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

Underfitting

Underfitting occurs when a model is too simple to capture patterns in data, resulting in poor performance on both training and test data.

1 High bias

2 Low model complexity

3 Poor accuracy

```
Data → Simple Model → Poor Fit
```

Overfitting

Overfitting occurs when a model learns training data too well, including noise, and fails to generalize to new data.

1 High variance

2 Very complex model

3 High training accuracy, low test accuracy

```
Training Data → Complex Model → Perfect Fit
```

Bias-Variance Tradeoff

The bias-variance tradeoff balances simplicity and complexity to achieve optimal performance.

```
High Bias ←■■ Optimal Model ■■→ High Variance
```

How to Avoid Underfitting

1 Increase model complexity

2 Add relevant features

3 Reduce regularization

How to Avoid Overfitting

1 Simplify model

2 Use more training data

3 Apply regularization

4 Use cross-validation

Summary

Balancing underfitting and overfitting is crucial for building robust machine learning models.