# Bias-Variance and Cross Validation

Nipun Batra and teaching staff

July 26, 2025

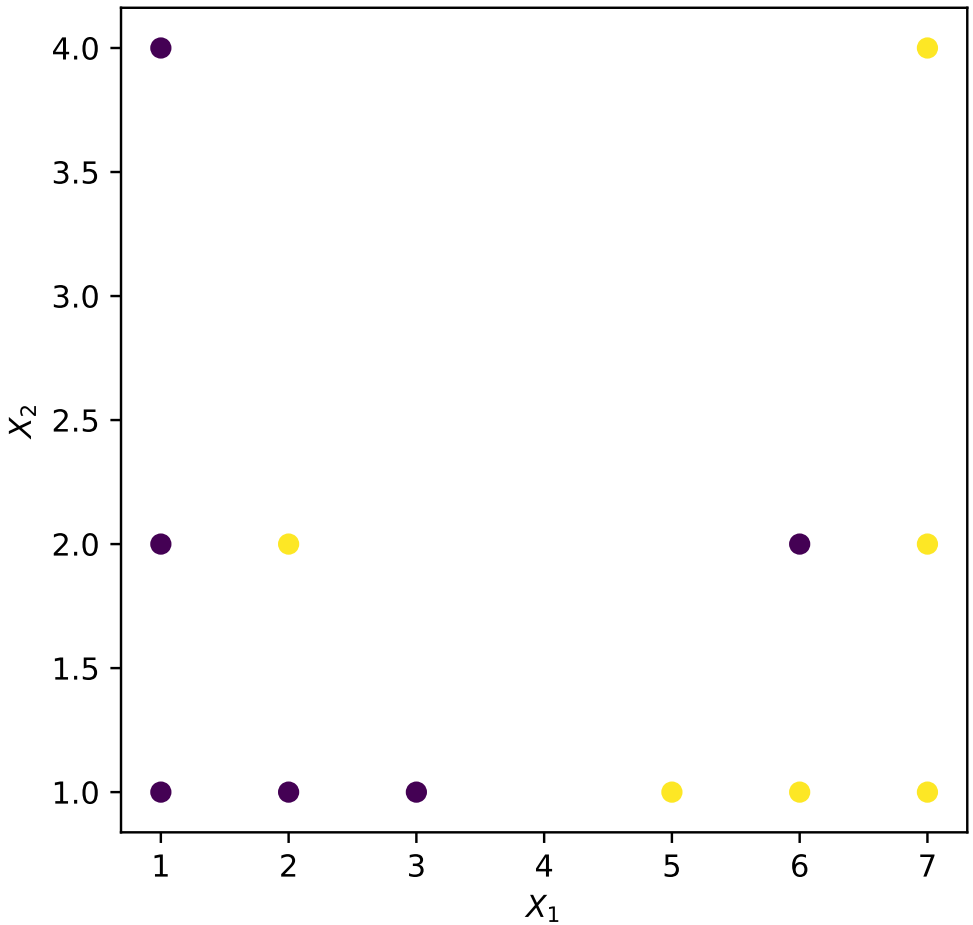
IIT Gandhinagar

# Table of Contents

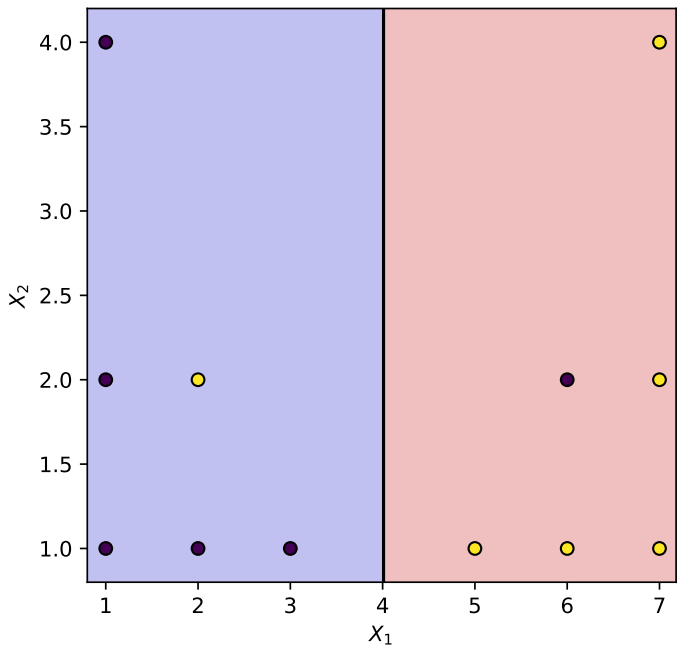
# Introduction to Bias-Variance

# A Question!

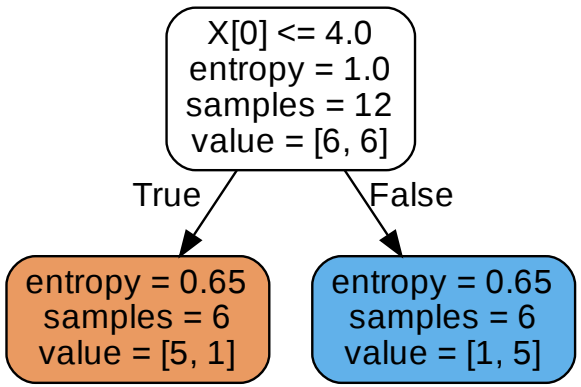
What would be the decision boundary of a decision tree classifier?



# Decision Boundary for a tree with depth 1

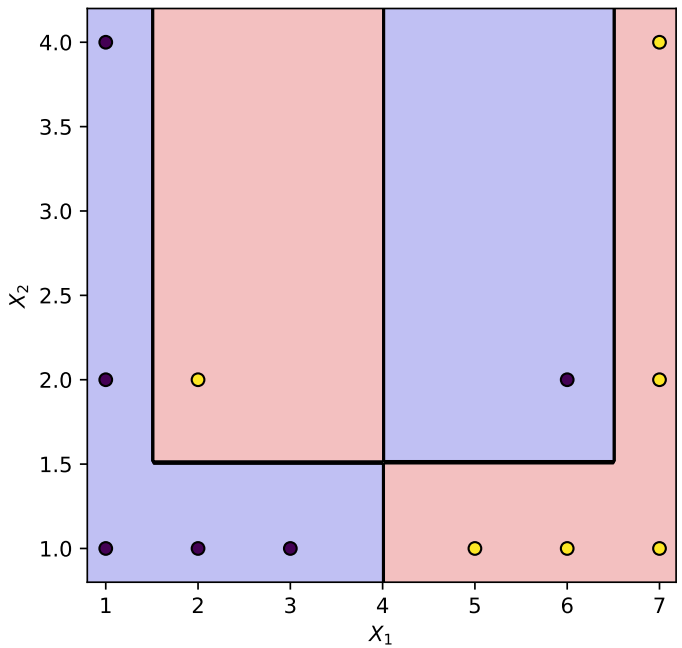



Decision Boundary

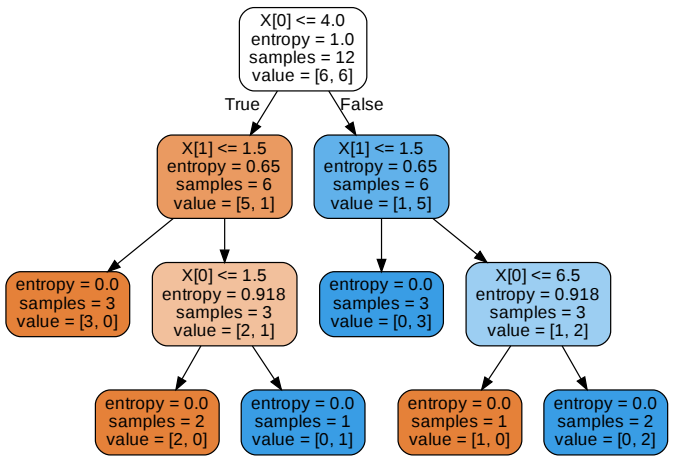



Decision Tree

# Decision Boundary for a tree with no depth limit




Decision Boundary




Decision Tree

# Are deeper trees always better?

As we saw, deeper trees learn more complex decision boundaries.

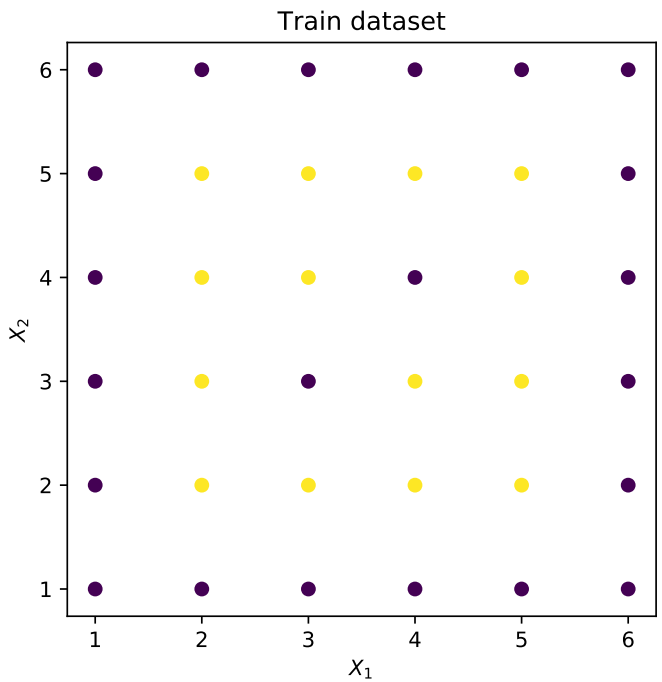
# Are deeper trees always better?

As we saw, deeper trees learn more complex decision boundaries.

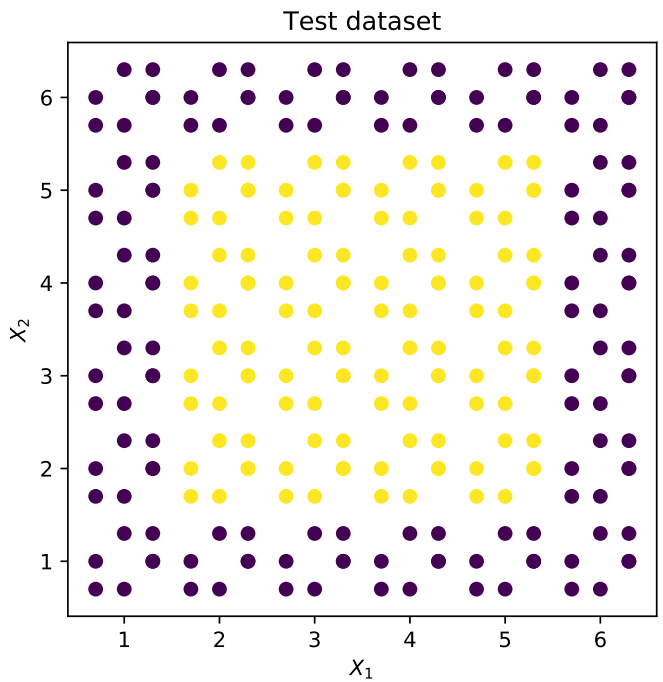
But, sometimes this can lead to poor generalization

# An example

Consider the dataset below



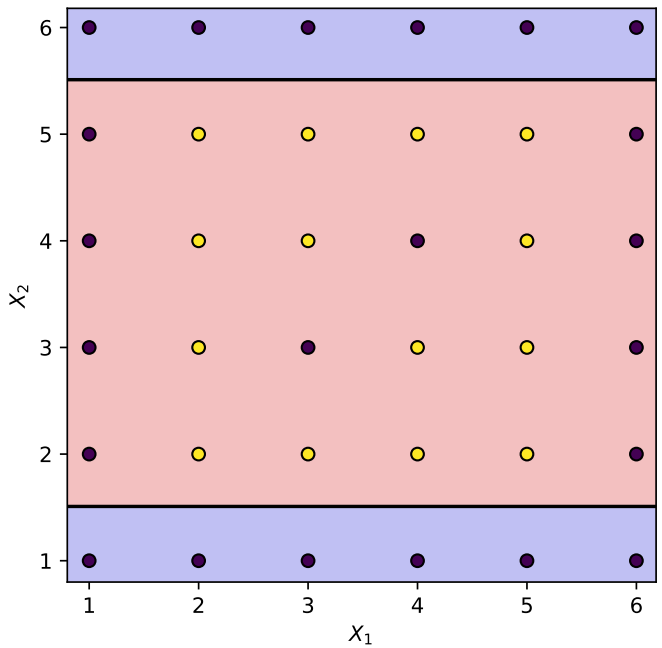

Train Set



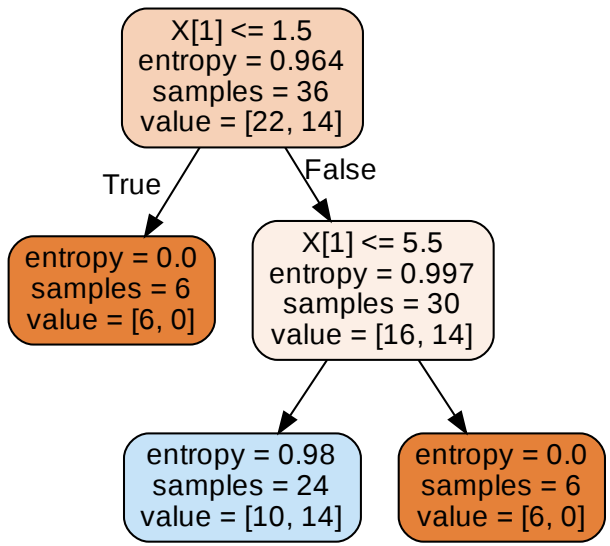

Test Set

# Underfitting

Underfitting is also known as high bias, since it has a very biased incorrect assumption.




Decision Boundary

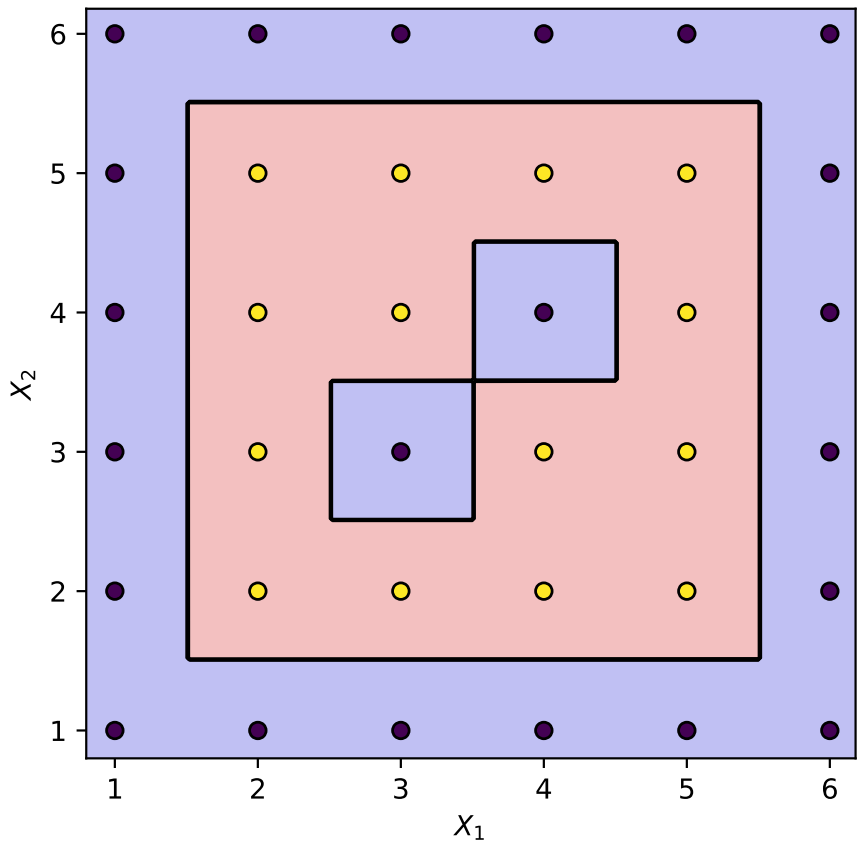



Decision Tree

# Overfitting

Overfitting is also known as high variance, since very small changes in data can lead to very different models.

Decision tree learned has depth of 10.



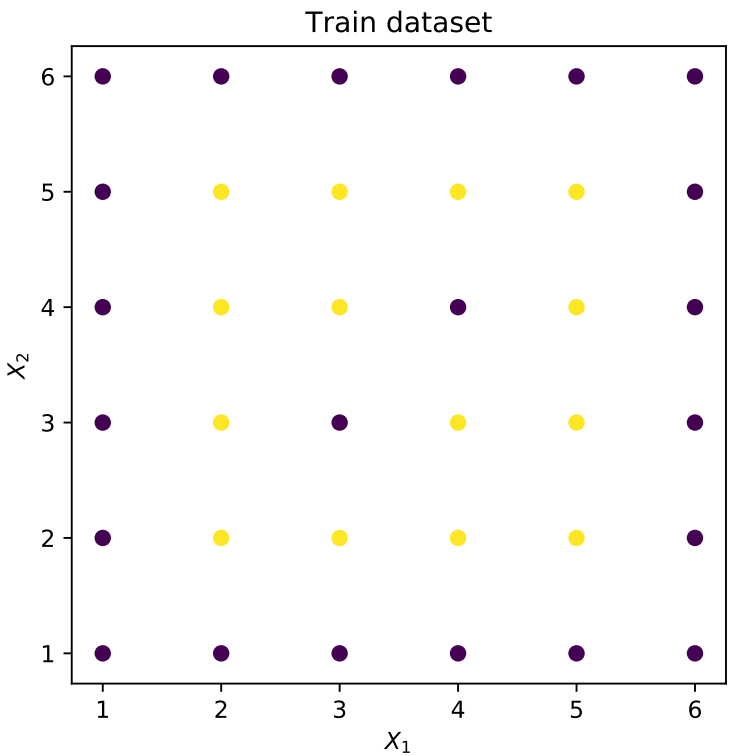
# Intuition for Variance

A small change in data can lead to very different models.

# Intuition for Variance

A small change in data can lead to very different models.

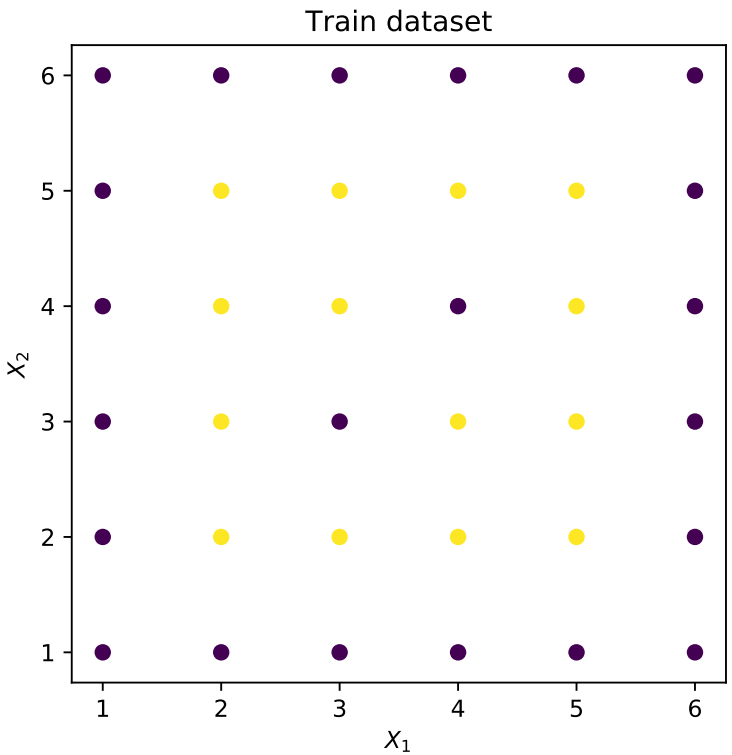
Dataset 1



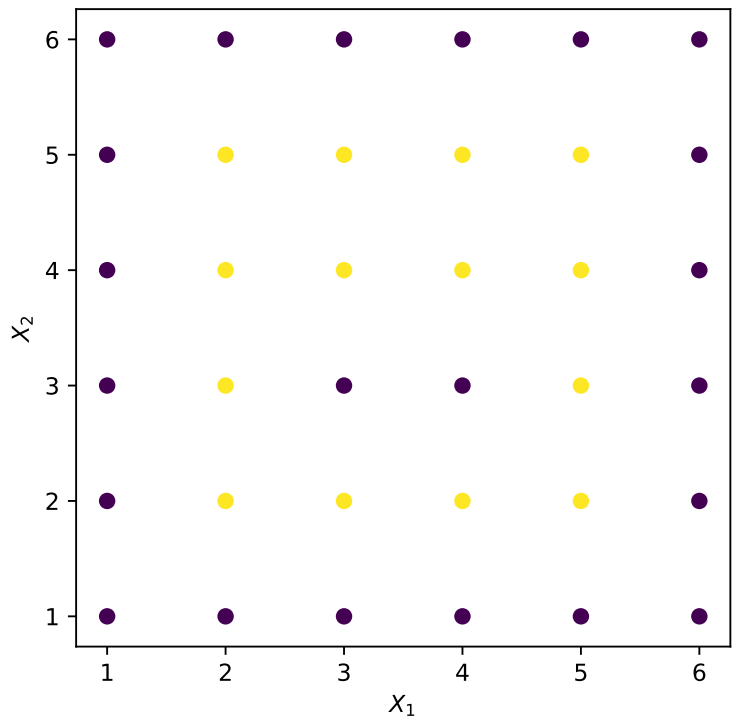
# Intuition for Variance

A small change in data can lead to very different models.

Dataset 1

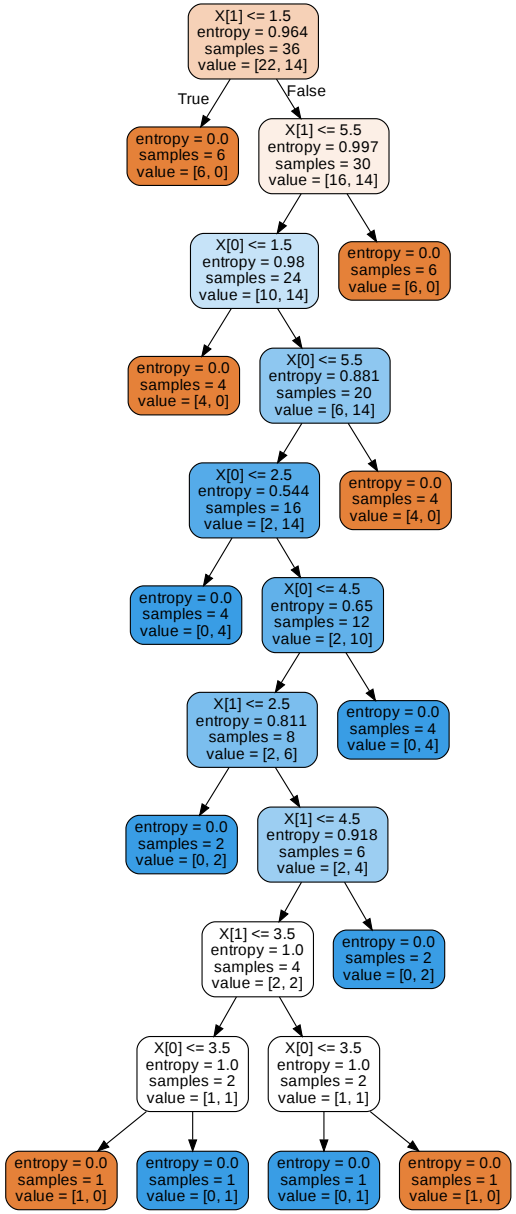



Dataset 2

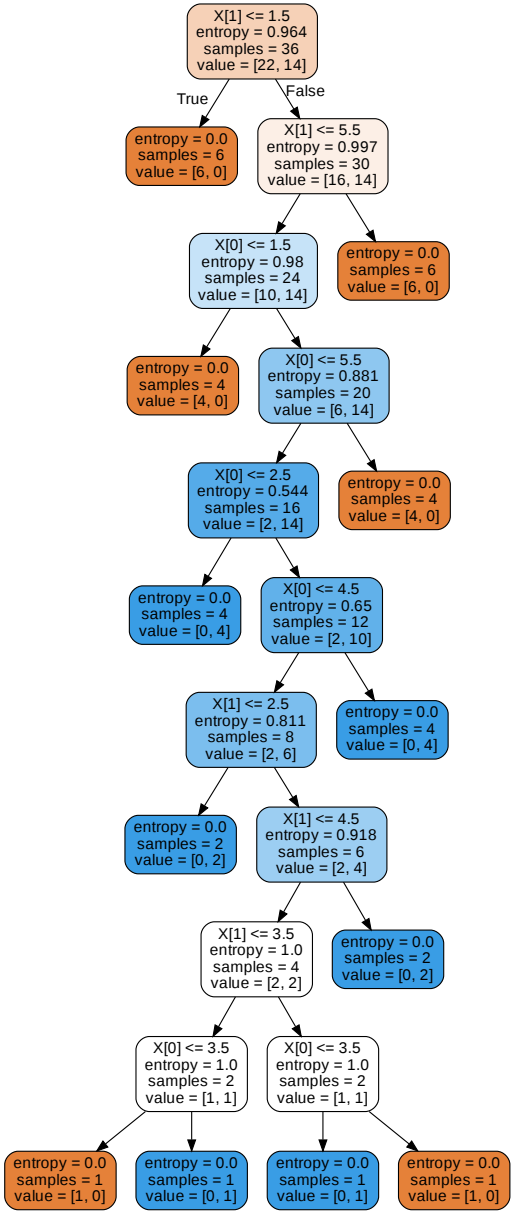


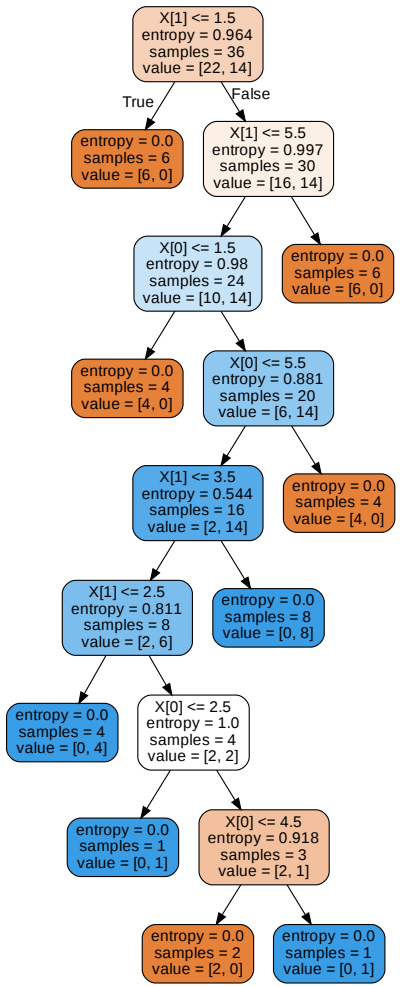
# Intuition for Variance

# Intuition for Variance



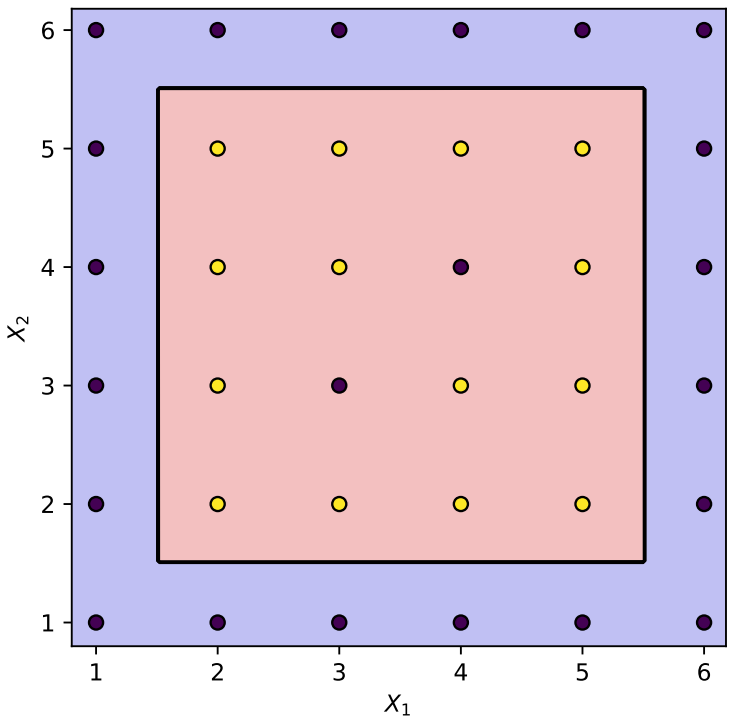
# Intuition for Variance



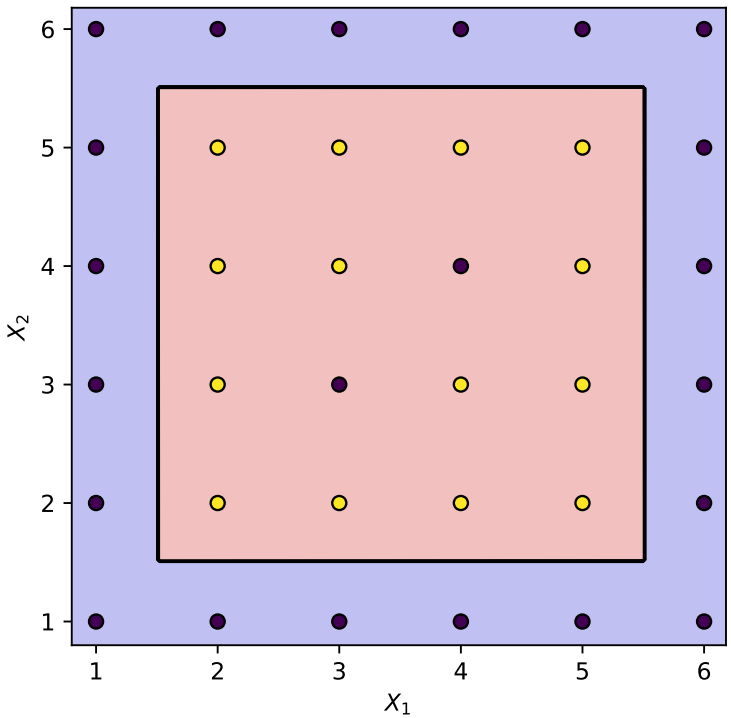



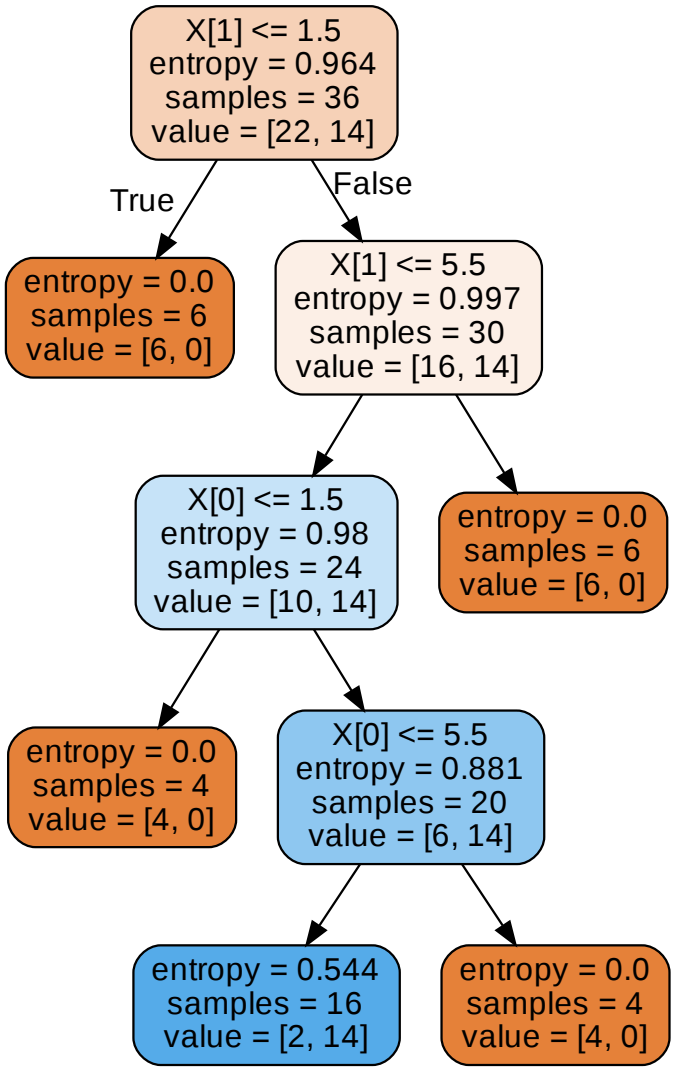
# A Good Fit

# A Good Fit

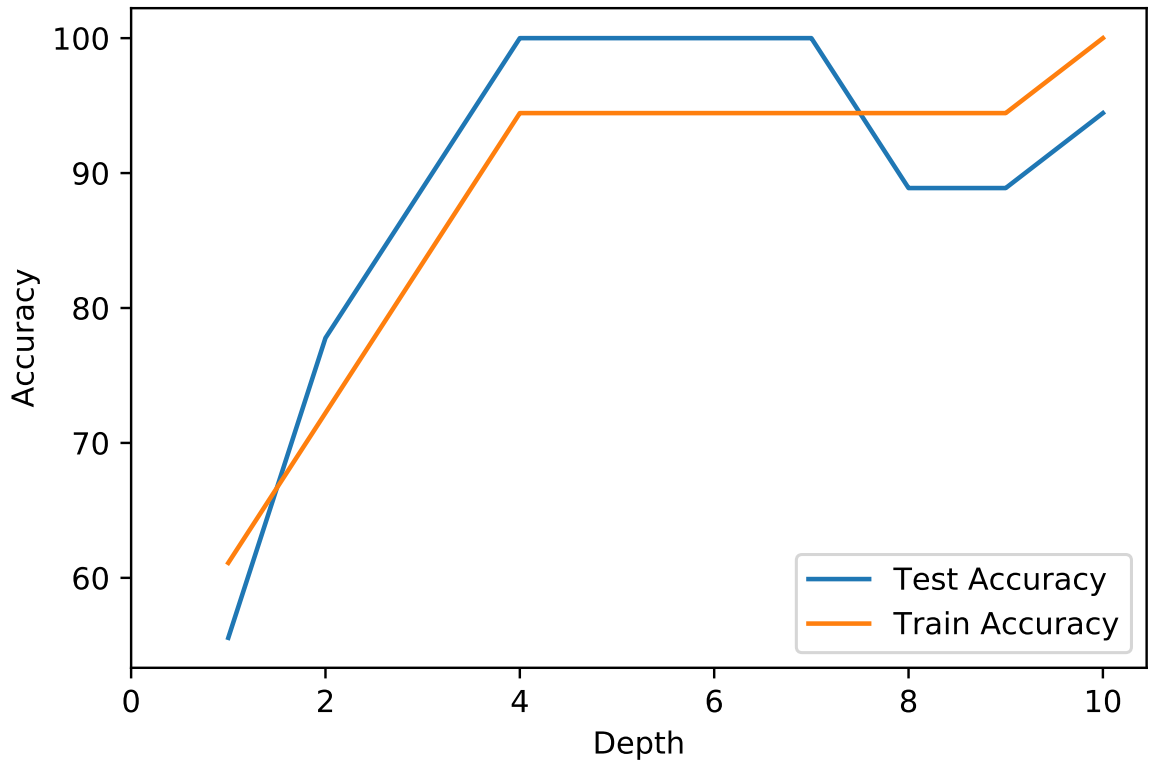


# A Good Fit

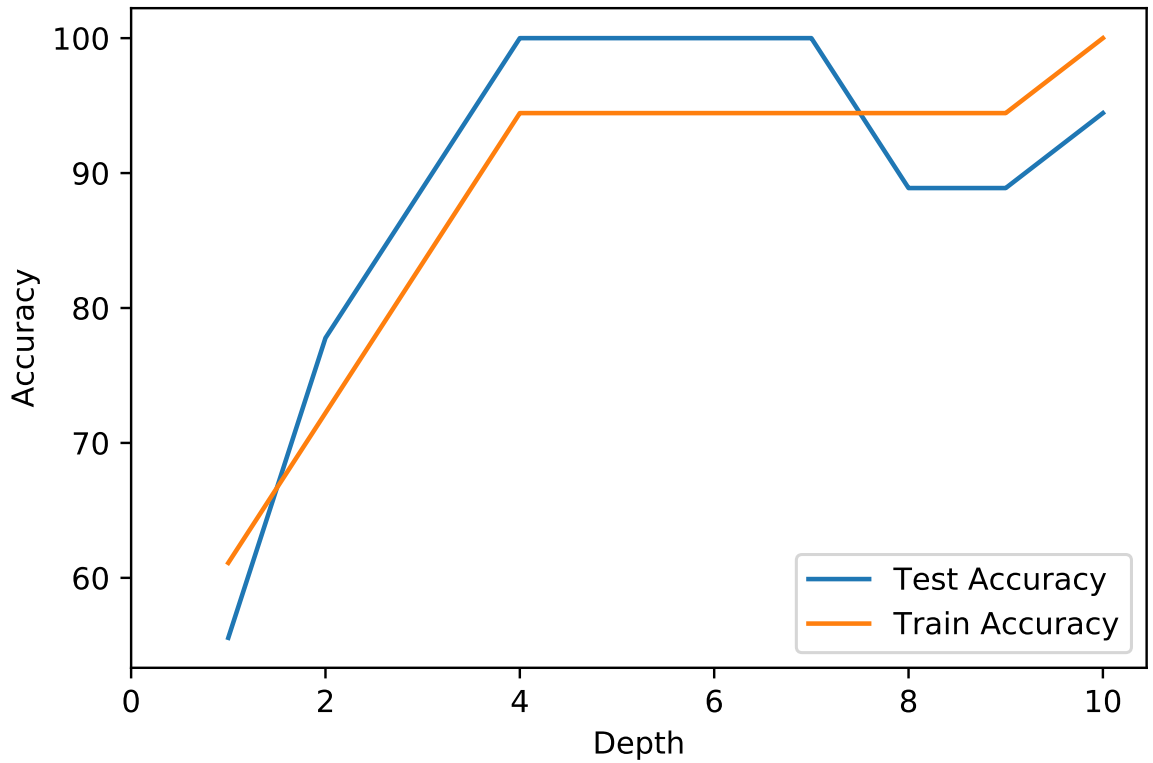

6
5
4
3
2
1
$X_2$
1
2
3
4
5
6
$X_1$


X[1] <= 1.5
entropy = 0.964
samples = 36
value = [22, 14]
True
False
entropy = 0.0
samples = 6
value = [6, 0]
X[1] <= 5.5
entropy = 0.997
samples = 30
value = [16, 14]
X[0] <= 1.5
entropy = 0.98
samples = 24
value = [10, 14]
entropy = 0.0
samples = 6
value = [6, 0]
entropy = 0.0
samples = 4
value = [4, 0]
X[0] <= 5.5
entropy = 0.881
samples = 20
value = [6, 14]
entropy = 0.544
samples = 16
value = [2, 14]
entropy = 0.0
samples = 4
value = [4, 0]

# Accuracy vs Depth Curve

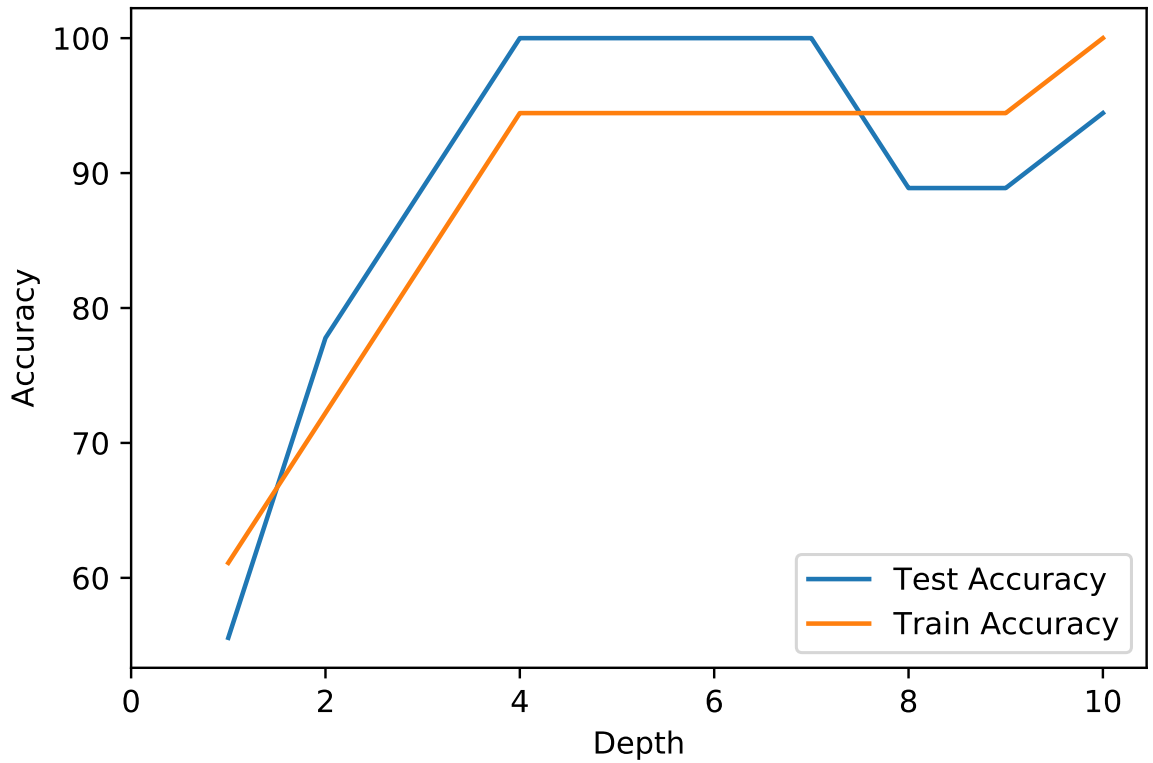


# Accuracy vs Depth Curve




As depth increases, train accuracy improves

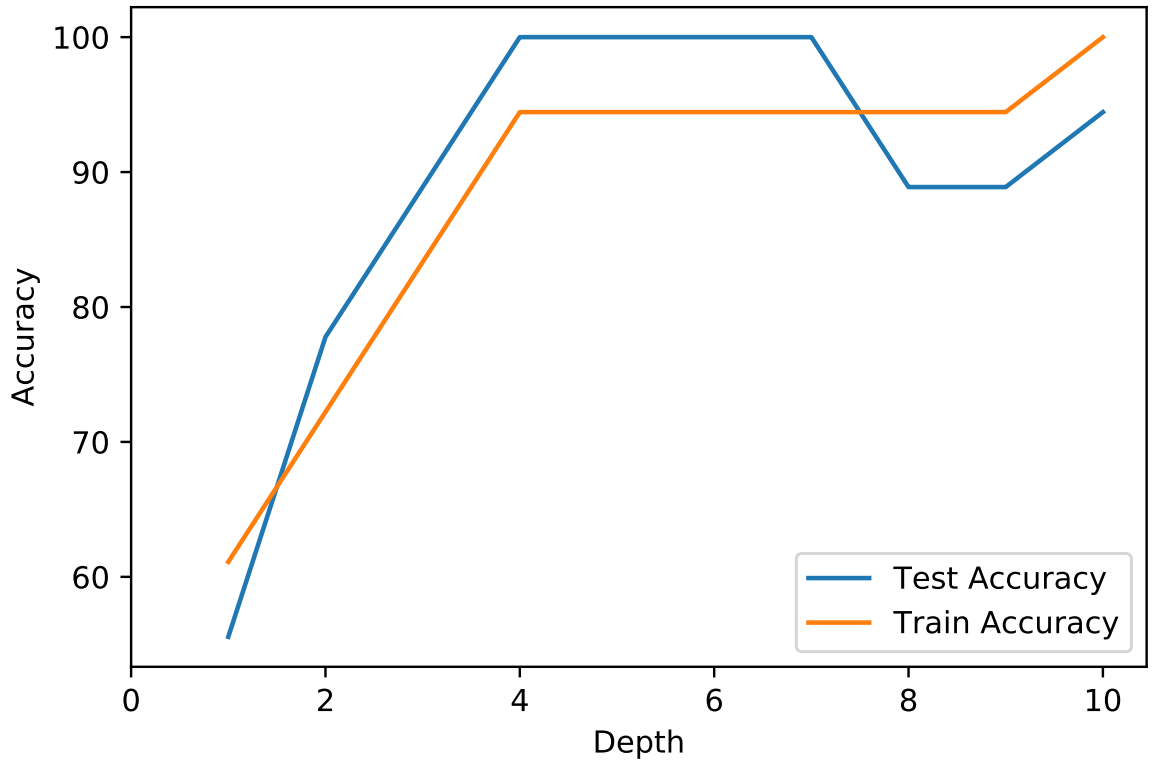
# Accuracy vs Depth Curve




As depth increases, train accuracy improves

As depth increases, test accuracy improves till a point

# Accuracy vs Depth Curve




As depth increases, train accuracy improves

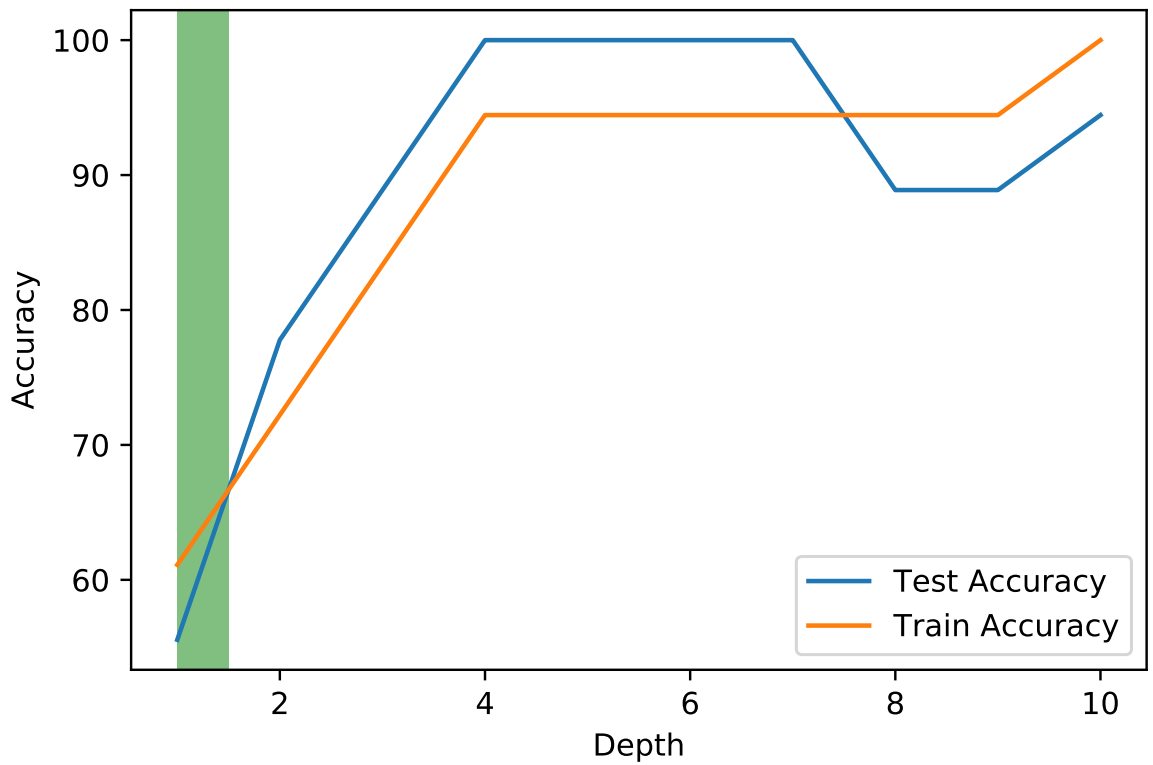
As depth increases, test accuracy improves till a point

At very high depths, test accuracy is not good (overfitting).

# Accuracy vs Depth Curve : Underfitting

The highlighted region is the underfitting region.

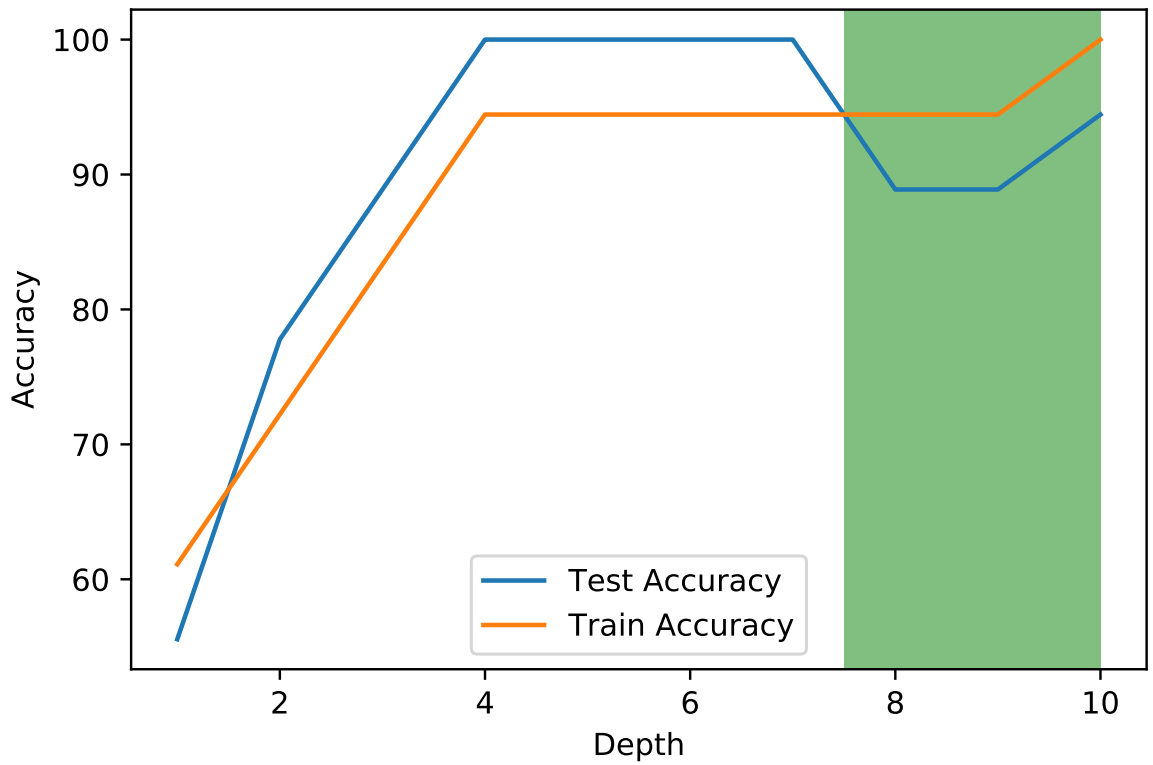
Model is too simple (less depth) to learn from the data.



# Accuracy vs Depth Curve : Overfitting

The highlighted region is the overfitting region.

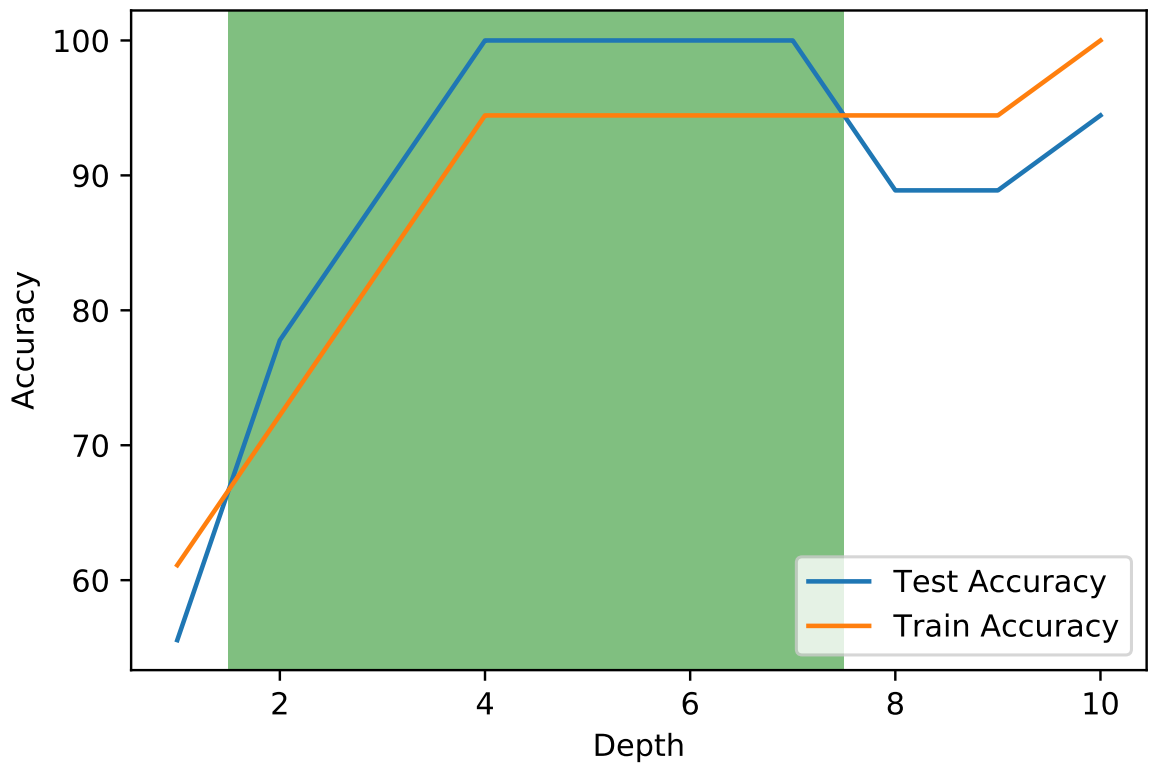
Model is complex (high depth) and hence also learns the anomalies in data.



# Accuracy vs Depth Curve

The highlighted region is the good fit region.

We want to maximize test accuracy while being in this region.



# The big question!?

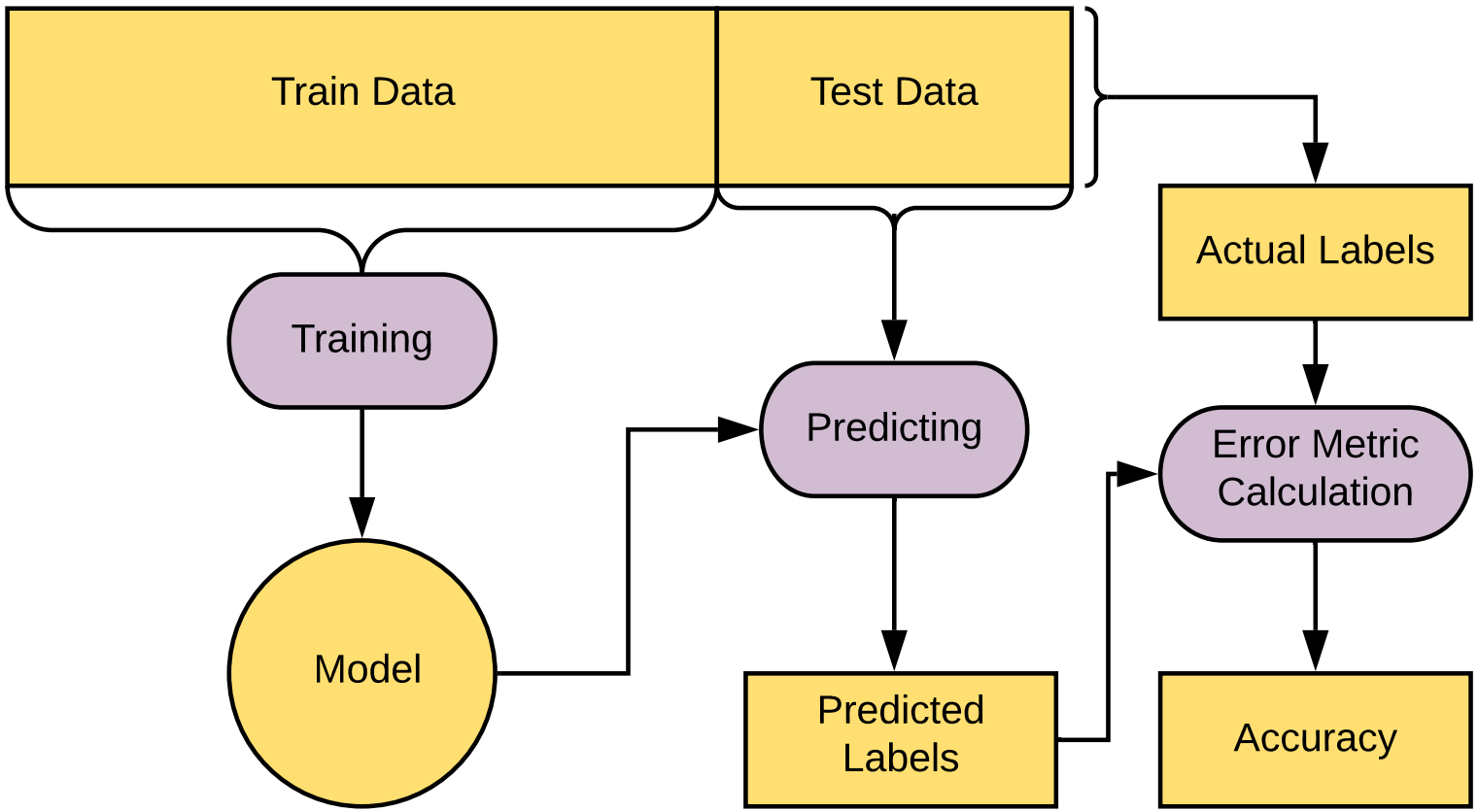
How to find the optimal depth for a decision tree?

# The big question!?

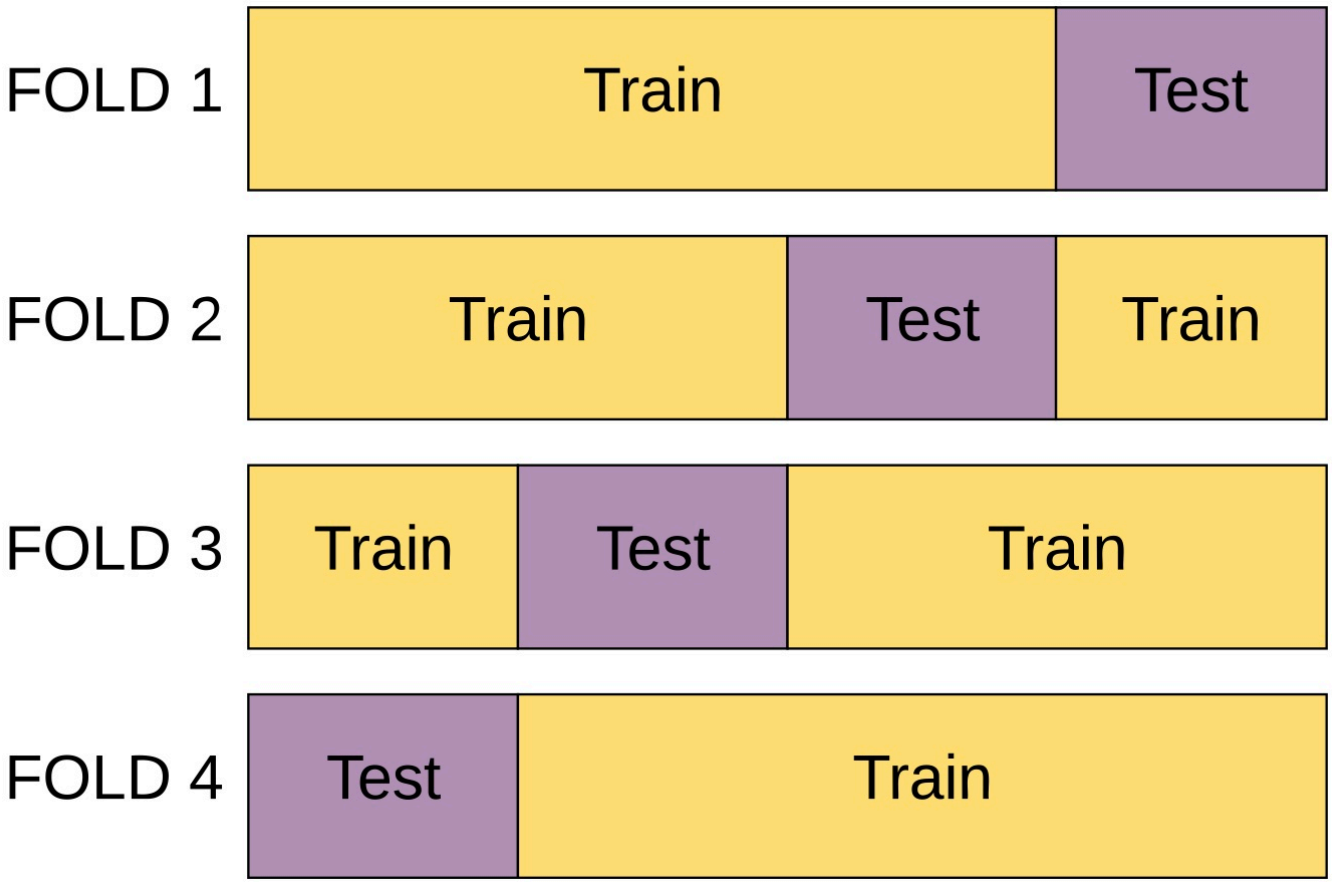
How to find the optimal depth for a decision tree?

Use cross-validation!

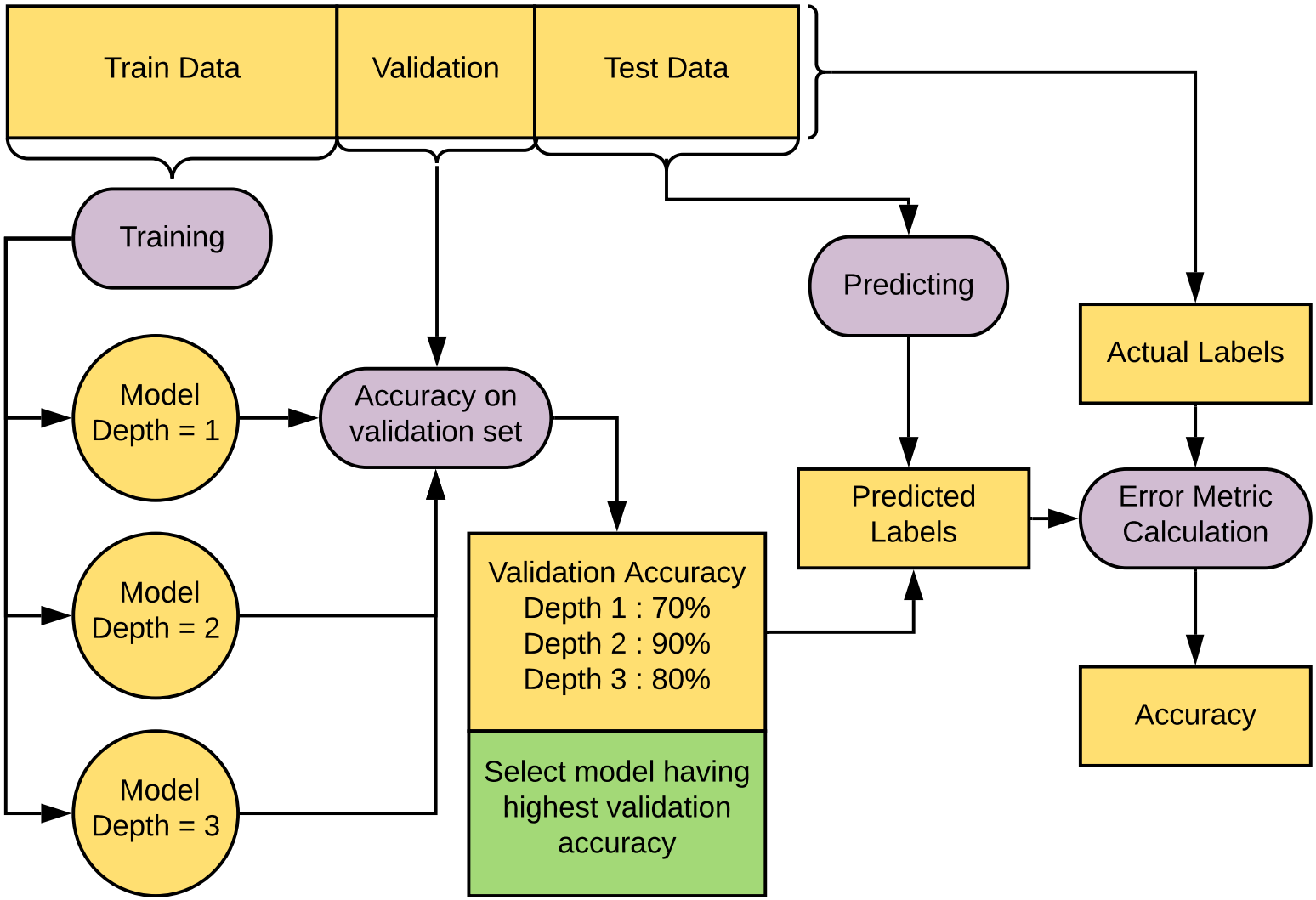
# Our General Training Flow



# K-Fold cross-validation: Utilise full dataset for testing



# The Validation Set

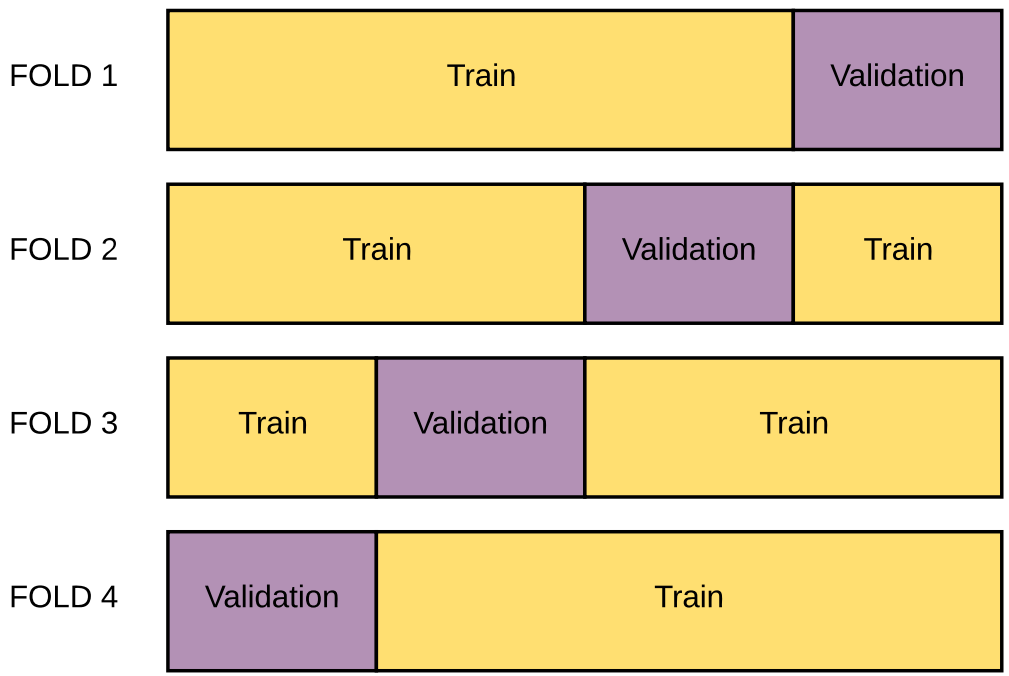


# Nested Cross Validation

Divide your training set into $k$ equal parts.

Cyclically use 1 part as "validation set" and the rest for training.

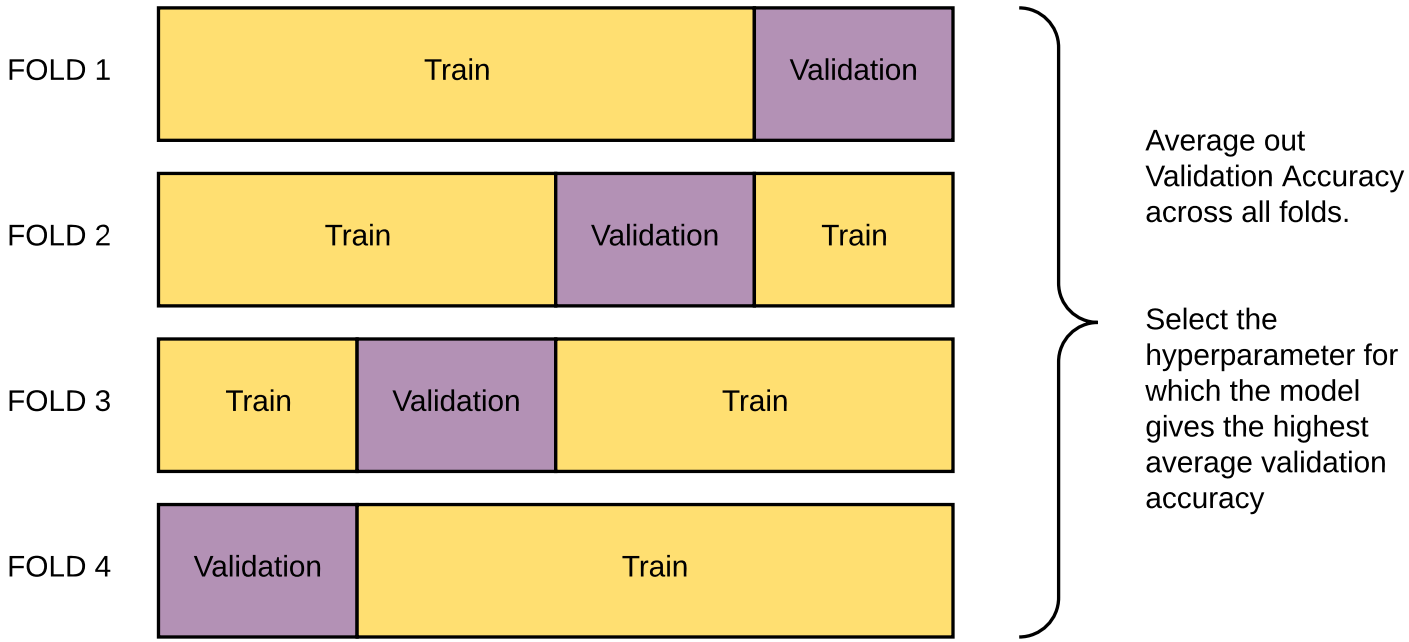
Here $k = 4$



# Nested Cross Validation

Average out the validation accuracy across all the folds

Use the model with highest validation accuracy



# Practice and Review

# Pop Quiz: Bias-Variance Concepts

1. What causes high bias in a model? Give an example.

# Pop Quiz: Bias-Variance Concepts

1. What causes high bias in a model? Give an example.

# Pop Quiz: Bias-Variance Concepts

1. What causes high bias in a model? Give an example.
2. What causes high variance in a model? Give an example.

# Pop Quiz: Bias-Variance Concepts

1. What causes high bias in a model? Give an example.
2. What causes high variance in a model? Give an example.

# Pop Quiz: Bias-Variance Concepts

1. What causes high bias in a model? Give an example.
2. What causes high variance in a model? Give an example.
3. How does cross-validation help in model selection?

# Pop Quiz: Bias-Variance Concepts

1. What causes high bias in a model? Give an example.
2. What causes high variance in a model? Give an example.
3. How does cross-validation help in model selection?

# Pop Quiz: Bias-Variance Concepts

1. What causes high bias in a model? Give an example.
2. What causes high variance in a model? Give an example.
3. How does cross-validation help in model selection?
4. Why can't we directly optimize for test error?

# Key Takeaways

- **Bias-Variance Decomposition**: Total error = $\text{Bias}^2$ + Variance + Noise

# Key Takeaways

- **Bias-Variance Decomposition**: Total error = $\text{Bias}^2$ + Variance + Noise
- **High Bias**: Underfitting, model too simple

# Key Takeaways

- **Bias-Variance Decomposition**: Total error = $\text{Bias}^2$ + Variance + Noise
- **High Bias**: Underfitting, model too simple
- **High Variance**: Overfitting, model too complex

# Key Takeaways

- **Bias-Variance Decomposition**: Total error = $\text{Bias}^2$ + Variance + Noise
- **High Bias**: Underfitting, model too simple
- **High Variance**: Overfitting, model too complex
- **Cross-Validation**: Essential for proper model evaluation

# Key Takeaways

- **Bias-Variance Decomposition**: Total error = $\text{Bias}^2$ + Variance + Noise
- **High Bias**: Underfitting, model too simple
- **High Variance**: Overfitting, model too complex
- **Cross-Validation**: Essential for proper model evaluation
- **Model Selection**: Choose complexity that balances bias and variance

# Key Takeaways

- **Bias-Variance Decomposition**: Total error = $\text{Bias}^2$ + Variance + Noise
- **High Bias**: Underfitting, model too simple
- **High Variance**: Overfitting, model too complex
- **Cross-Validation**: Essential for proper model evaluation
- **Model Selection**: Choose complexity that balances bias and variance
- **No Free Lunch**: Cannot reduce both bias and variance simultaneously

# Next time: Ensemble Learning

- How to combine various models?

# Next time: Ensemble Learning

- How to combine various models?
- Why to combine multiple models?

# Next time: Ensemble Learning

- How to combine various models?
- Why to combine multiple models?
- How can we reduce bias?

# Next time: Ensemble Learning

- How to combine various models?
- Why to combine multiple models?
- How can we reduce bias?
- How can we reduce variance?