Bias/Variance and Cross-Validation

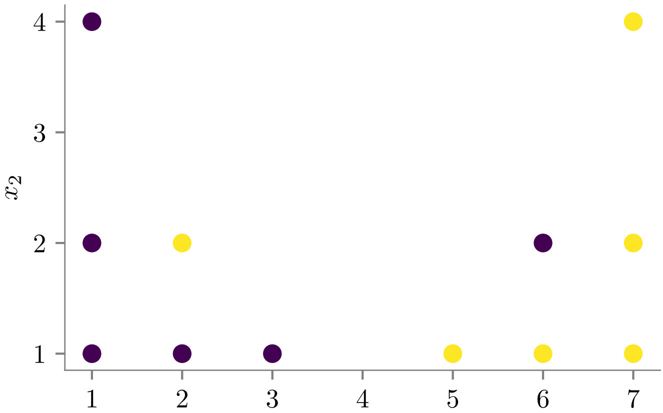
Nipun Batra and teaching staff

January 9, 2024

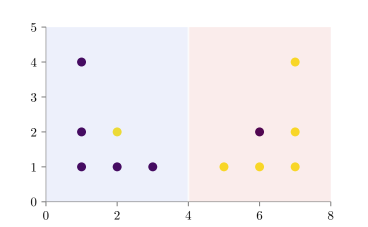
IIT Gandhinagar

A Question!

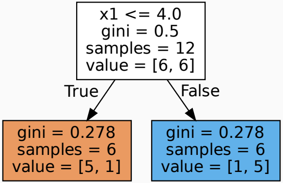
What would be the decision boundary of a decision tree classifier?



Decision Boundary for a tree with depth 1



(a) Decision Boundary



x1 <= 4.0
gini = 0.5
samples = 12
value = [6, 6]

True

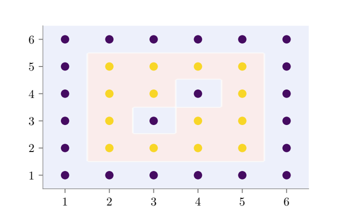
False

gini = 0.278
samples = 6
value = [5, 1]

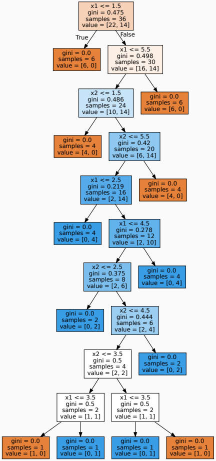
gini = 0.278
samples = 6
value = [1, 5]

(b) Decision Tree

Decision Boundary for a tree with no depth limit



(a) Decision Boundary



(b) Decision Tree

Are deeper trees always better?

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

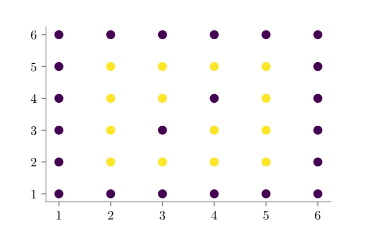
Are deeper trees always better?

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

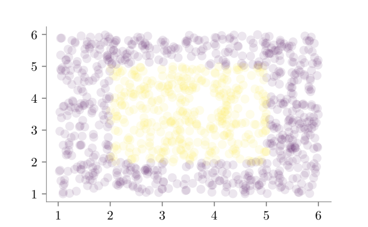
But, sometimes this can lead to *poor generalization*

An example

Consider the dataset below



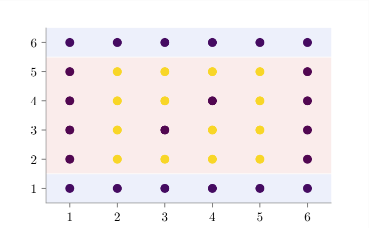
(a) Train Set



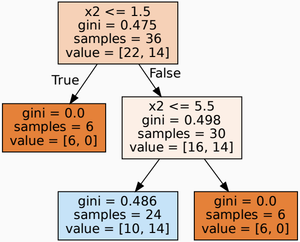
(b) Test Set

Underfitting

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



(a) Decision Boundary

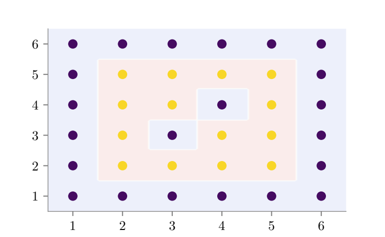


(b) Decision Tree

Overfitting

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

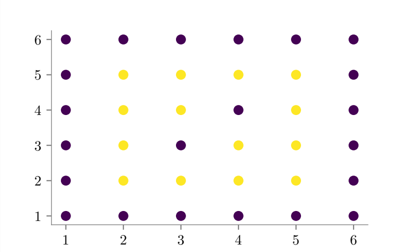
Decision tree learned has depth of 10.



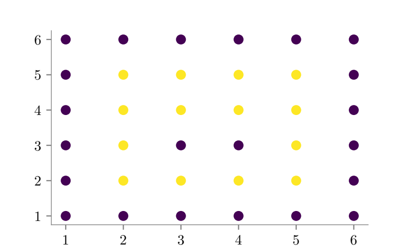
Intution for Variance

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

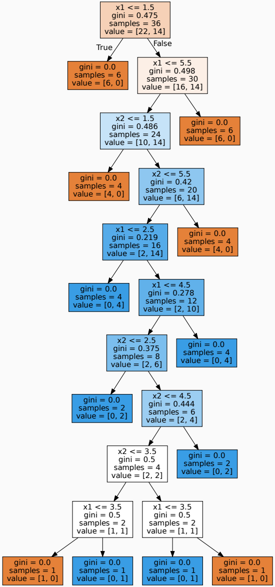
Dataset 1

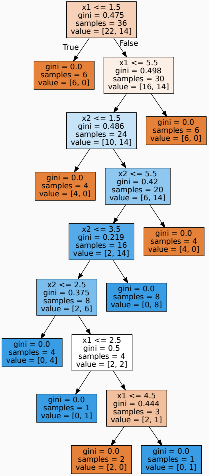


Dataset 2

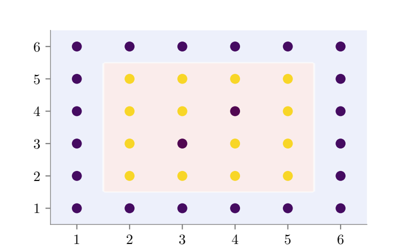


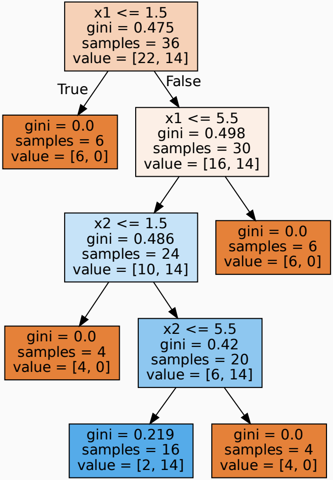
Intution for Variance





A Good Fit





x1 <= 1.5
gini = 0.475
samples = 36
value = [22, 14]

True

False

gini = 0.0
samples = 6
value = [6, 0]

x1 <= 5.5
gini = 0.498
samples = 30
value = [16, 14]

x2 <= 1.5
gini = 0.486
samples = 24
value = [10, 14]

gini = 0.0
samples = 6
value = [6, 0]

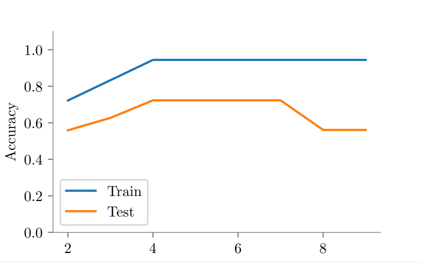
gini = 0.0
samples = 4
value = [4, 0]

x2 <= 5.5
gini = 0.42
samples = 20
value = [6, 14]

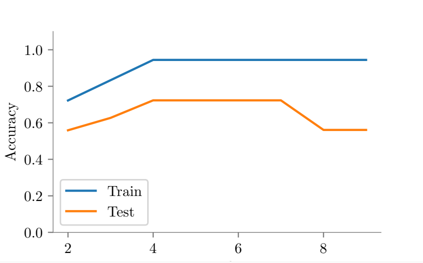
gini = 0.219
samples = 16
value = [2, 14]

gini = 0.0
samples = 4
value = [4, 0]

Accuracy vs Depth Curve

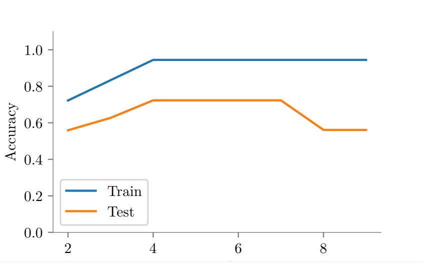


Accuracy vs Depth Curve



As depth increases, train accuracy improves

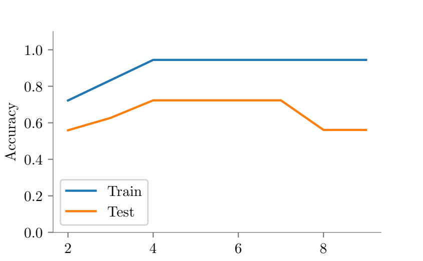
Accuracy vs Depth Curve



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

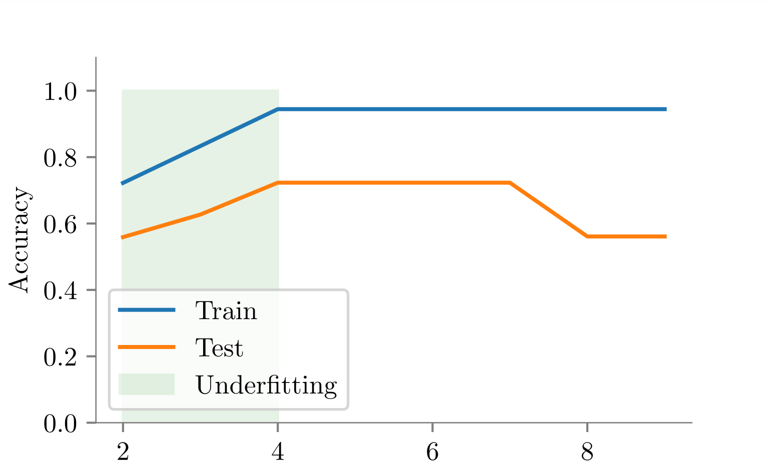
As depth increases, test accuracy improves till a point

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

Accuracy vs Depth Curve : Underfitting

The highlighted region is the underfitting region.

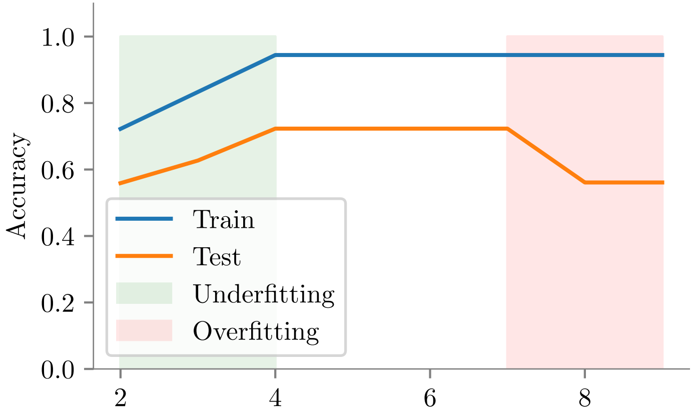
Model is too simple (less depth) to learn from the data.



Accuracy vs Depth Curve : Overfitting

The highlighted region is the overfitting region.

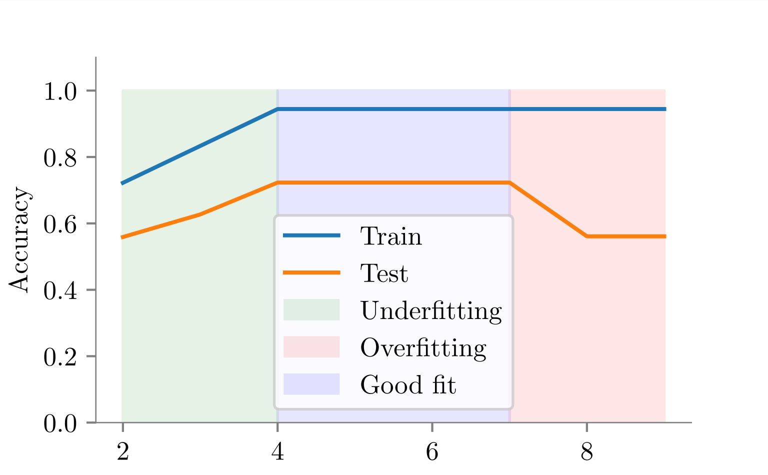
Model is complex (high depth) and hence also learns the anomalies in data.



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!?

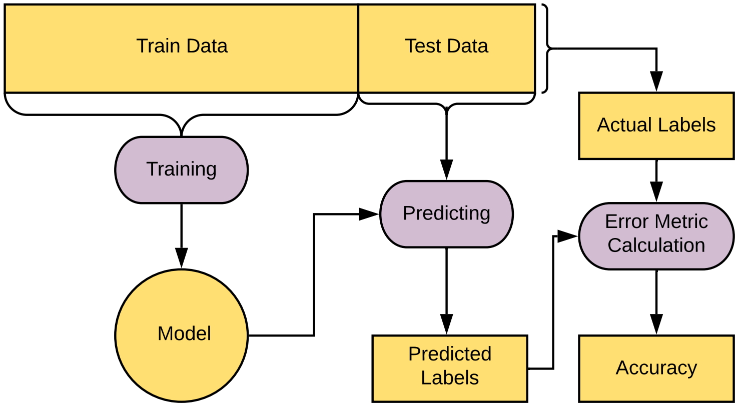
How to find the optimal depth for a decision tree?

The big question!?

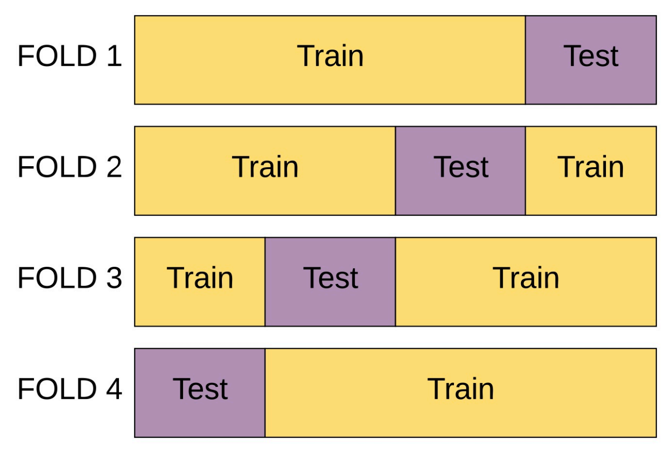
How to find the optimal depth for a decision tree?

Use cross-validation!

Our General Training Flow



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



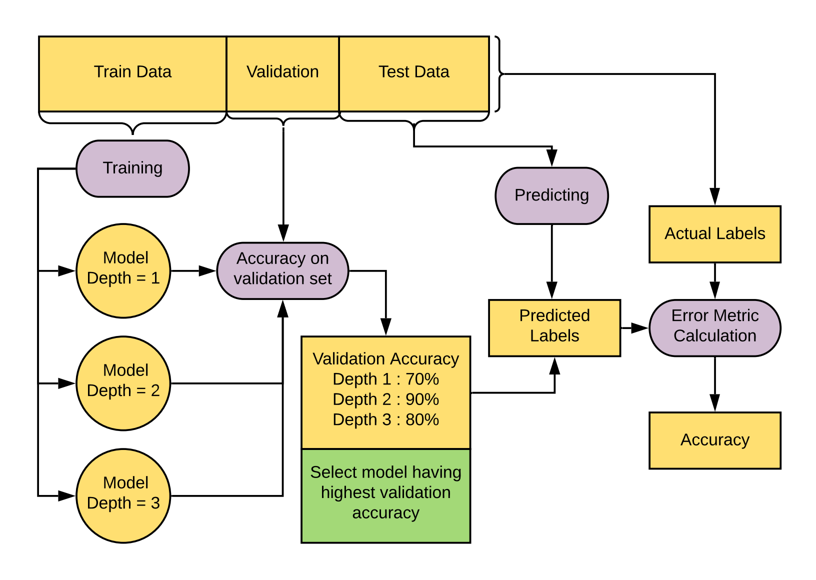
FOLD 1 Train Test

FOLD 2 Train Test Train

FOLD 3 Train Test Train

FOLD 4 Test Train

The Validation Set

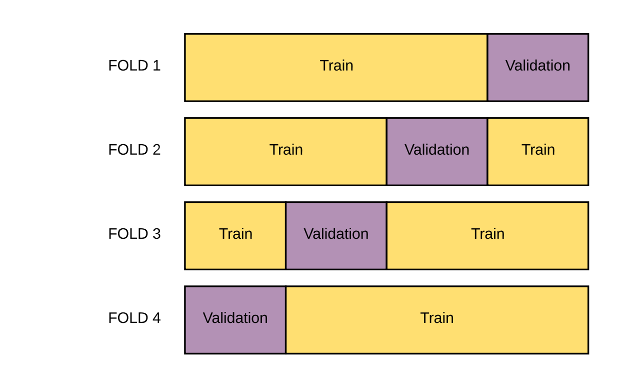


Nested Cross Validation

Divide your training set into K equal parts.

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

Here $K = 4$



FOLD 1 Train Validation

FOLD 2 Train Validation Train

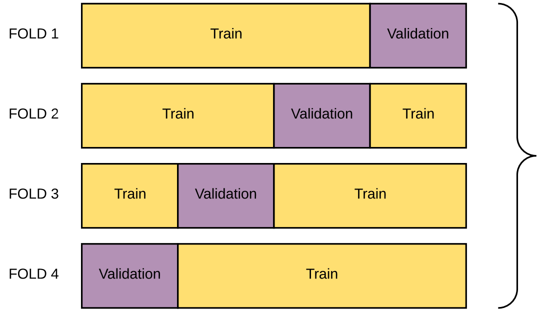
FOLD 3 Train Validation Train

FOLD 4 Validation Train

Nested Cross Validation

Average out the validation accuracy across all the folds

Use the model with highest validation accuracy



Average out Validation Accuracy across all folds.

Select the hyperparameter for which the model gives the highest average validation accuracy

Next time: Ensemble Learning

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