

Course	Python ML (Python for Machine Learning)
Term	
Week	
Date	
Topic	Decision Trees aka Classification and Regression Trees (CART)

Decision Trees

Classification and Regression Trees

Siva R Jasthi

Computer Science and Cybersecurity
Metropolitan State University

Classification and Regression Trees (CART)

- Build a tree by splitting on independent variables
 - To predict the outcome for an observation, follow the splits, making a prediction based on the final node
 - CART models can be applied to predict:
 - A continuous outcome – called a regression tree
 - A discrete (classification) outcome – called a classification tree
- Hence the name “Classification And Regression Tree”
- <https://medium.com/geekculture/applying-7-classification-algorithms-on-the-titanic-dataset-278ef222b53c>

Some sample data

NO	Interview	GPA	Experience
1	0	3.27	1.93
2	0	3.37	0.07
3	0	3.57	1.91
4	0	3.91	4.35
5	0	3.2	1.7
6	1	3.9	2.41
7	1	3.94	3
8	0	3.66	2.47
9	0	3.63	0.93
10	0	3.06	4.14
11	0	3.21	3.34
12	0	3.18	3.97
13	0	3.69	0.54
14	0	3.38	3.62
15	1	3.77	2.06
16	0	3.5	4.1
...
17	1	3.31	3.46
26	0	3.78	0.29
27	0	3.87	1.21
28	0	4	0.49
29	1	3.87	2.11

GPA, Experience: are independent variables

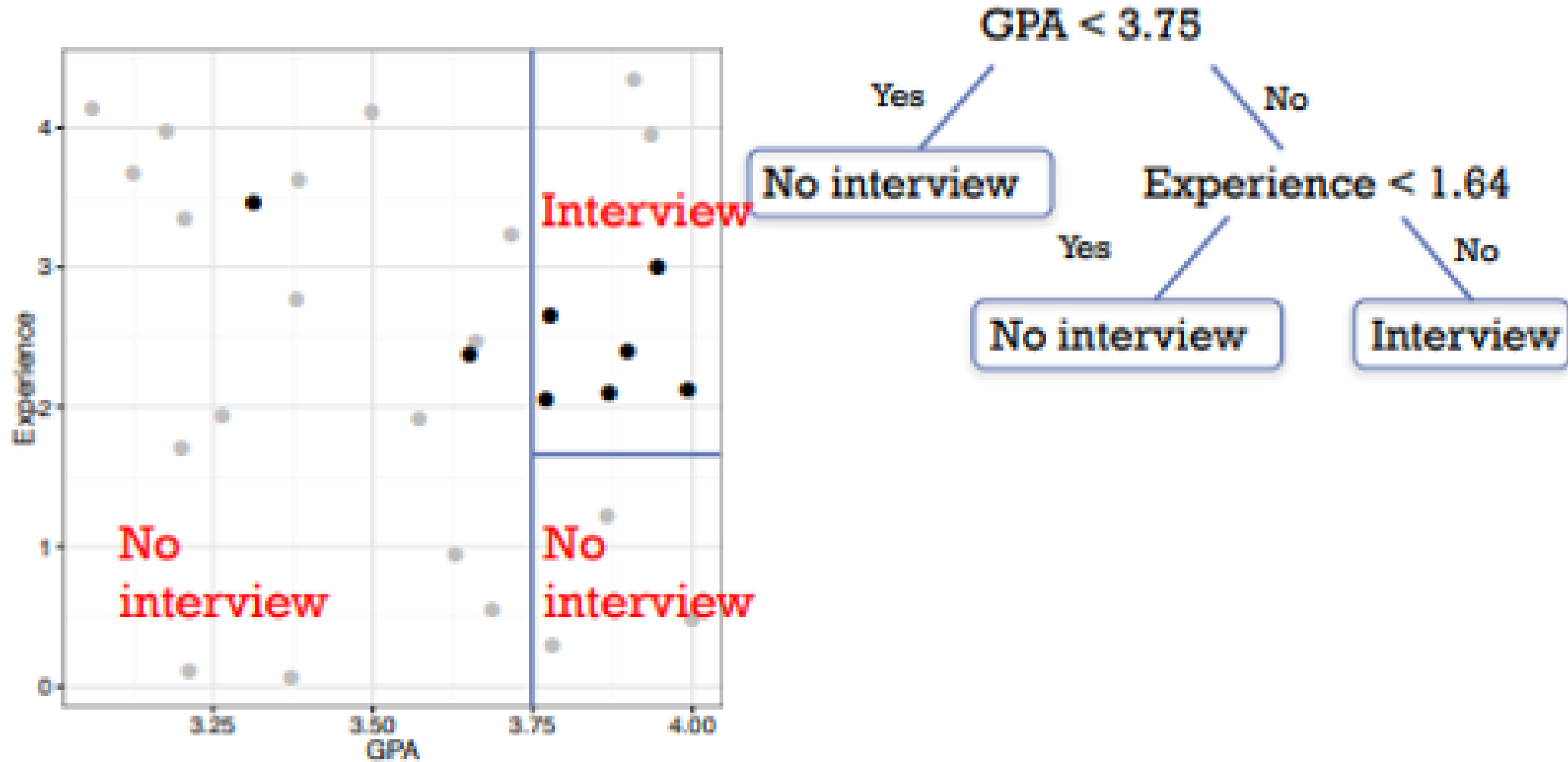
Interview: Whether a candidate gets an interview or not depends on GPA and Experience.

3.2 (exp) GPA 3.8

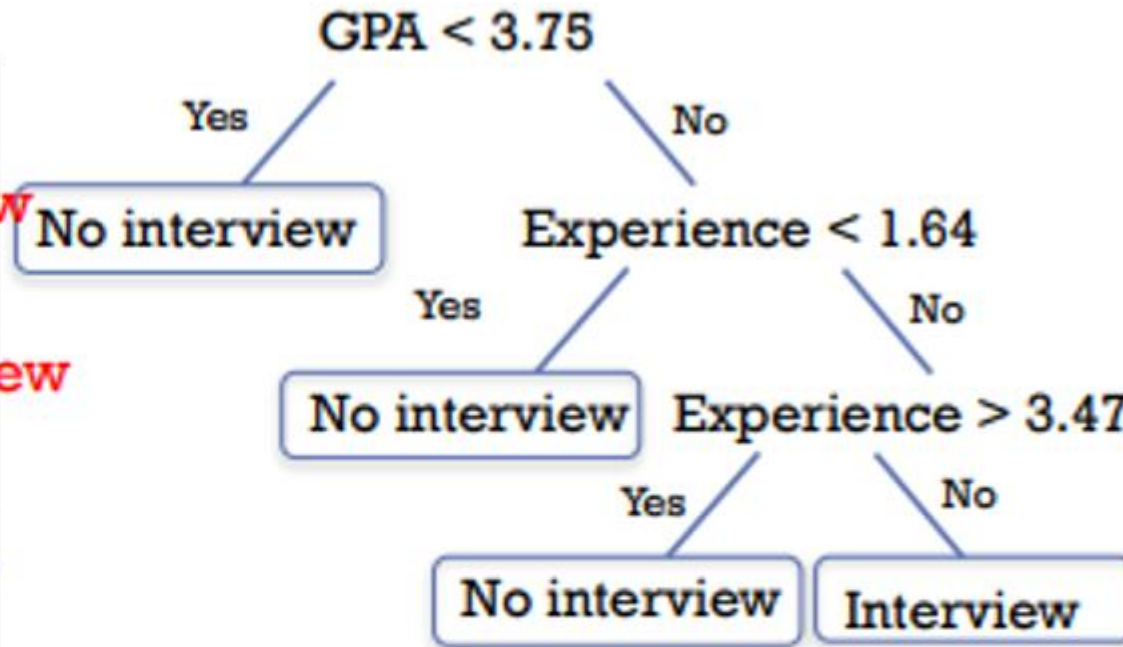
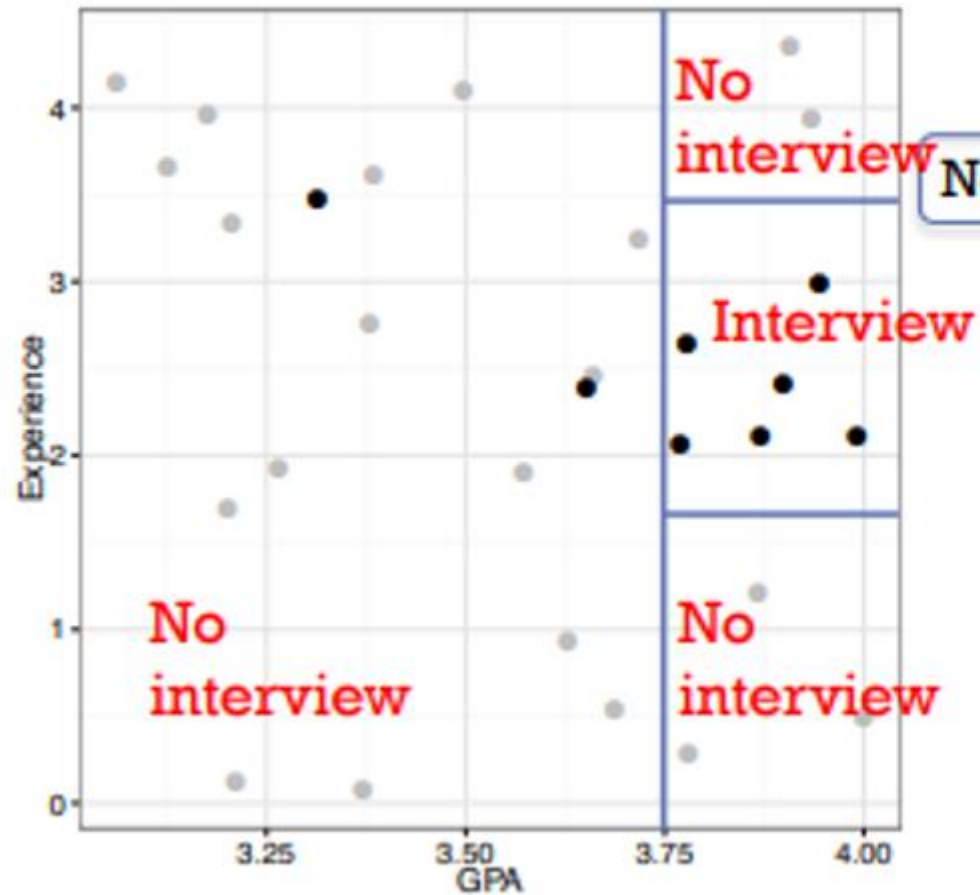
Decision Tree based on GPA



Decision Tree based on GPA and some Experience



Decision Tree based on GPA and Experience



Decision Tree : Model Summary

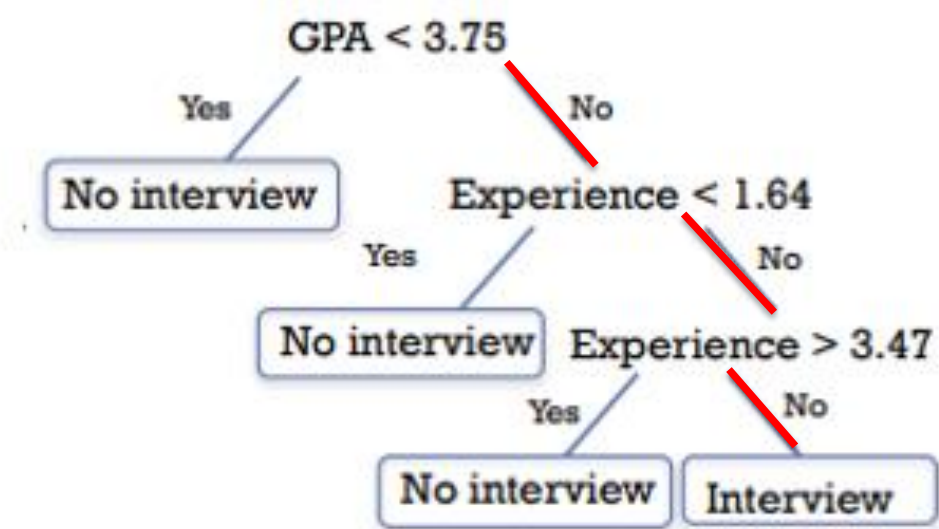
- Model predicts a student will get an interview if they have GPA of 3.75+ and 1.65- 3.47 years of experience
- Others are predicted not to get interviews



Predictions follow a simple path

- Consider Stephanie, who has a 3.91 GPA and 0.8 years of experience
- Is she predicted by the model to get an interview?

3.2 (exp) GPA 3.8



NO	Interview	GPA	
1	0	3.27	
2	0	3.37	
3	0	3.57	
4	0	3.91	
5	0	3.2	
6	1	3.9	
7	1	3.94	
8	0	3.66	
9	0	3.63	
10	0	3.06	
11	0	3.21	
12	0	3.18	
13	0	3.69	
14	0	3.38	
15	1	3.77	
16	0	3.5	
...	
17	1	3.31	
26	0	3.78	
27	0	3.87	
28	0	4	
29	1	3.87	

Some sample data

- Trees divide the feature space (the space of possible features) into k distinct regions R_1, R_2, \dots, R_k
- The same prediction is made for every X that falls into a particular region R_j

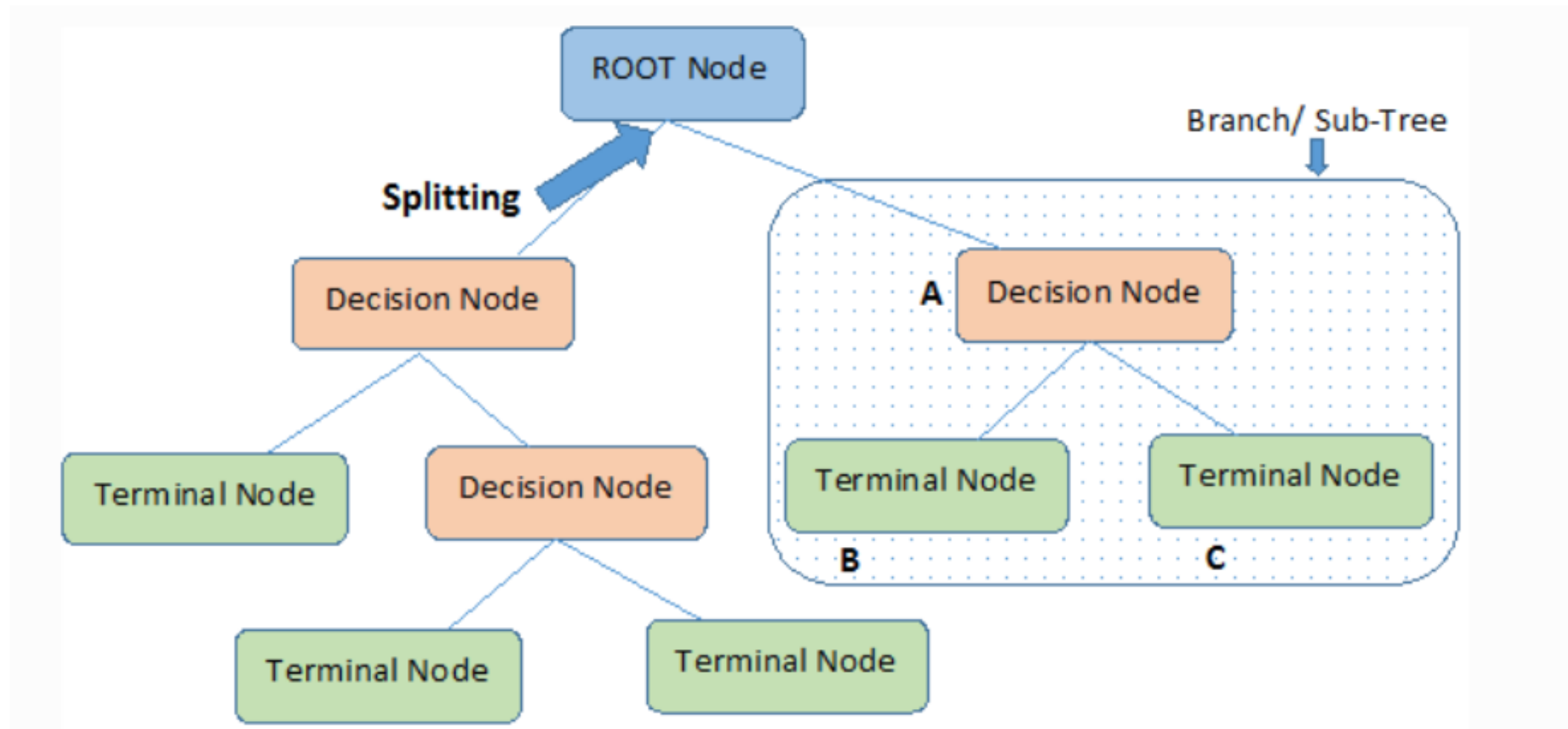


Predictions (regression and classification)

- For regression trees, make a prediction based on the mean response value (w.r.t. to the training data) in the corresponding bucket
- For classification trees, make a prediction based on the most commonly occurring class in the corresponding bucket n Class proportions also informative



Decision Nodes, Leaf Nodes



- <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>

Splitting Decisions and Model Performance

Gini impurity: This is a measure of the probability of misclassifying a randomly chosen element from the node. It is defined as the sum of the probabilities of each class label squared, subtracted from one. The Gini impurity is minimized when all the elements in the node belong to the same class.

Entropy: This is a measure of the amount of information needed to describe the distribution of the target variable in the node. It is defined as the negative sum of the probabilities of each class label multiplied by the logarithm of the probabilities. Entropy is minimized when all the elements in the node belong to the same class.

Classification error: This is a measure of the error rate of the most frequent class in the node. It is defined as one minus the maximum probability of any class label in the node. Classification error is minimized when all the elements in the node belong to the same class.

- Let us refer to “ML_Basic_Concepts_Glossary.pdf” (slides 32 – 44)

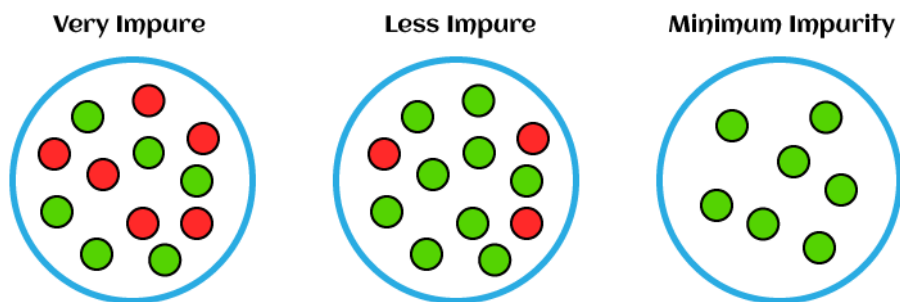
Gini Index

The Gini index is a measure of impurity or diversity commonly used to evaluate the quality of a split in a decision tree.

It measures how often a randomly chosen element from a set would be incorrectly labeled.

The Gini index ranges from 0 to 1, with 0 indicating perfect purity (all elements have the same label) and 1 indicating maximum impurity (an equal number of elements have each possible label).

The attribute with the lowest Gini index after the split is chosen as the root node of the next subtree, and the process is repeated recursively until a stopping criterion is met.



$$Gini = 1 - \sum_{i=1}^C (p_i)^2$$

- <https://www.learnbymarketing.com/481/decision-tree-flavors-gini-info-gain/>

Entropy

Entropy is a measure of the impurity or disorder of a set of examples or data points.

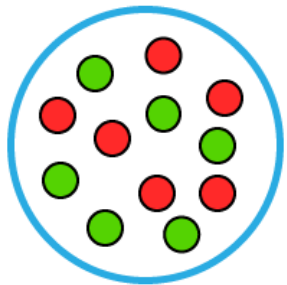
It is commonly used in decision trees and other classification algorithms to determine the optimal splitting criteria for a given set of data.

When all the data points in a set belong to the same class, the entropy is zero, indicating perfect purity. On the other hand, if the data points are evenly distributed among all the classes, the entropy is maximum, indicating complete disorder or randomness.

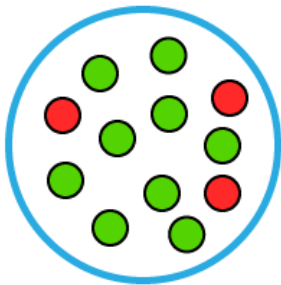
$$H = - \sum_i p_i (\log_2 p_i)$$

where $p(x)$ is the probability of a data point belonging to a particular class.

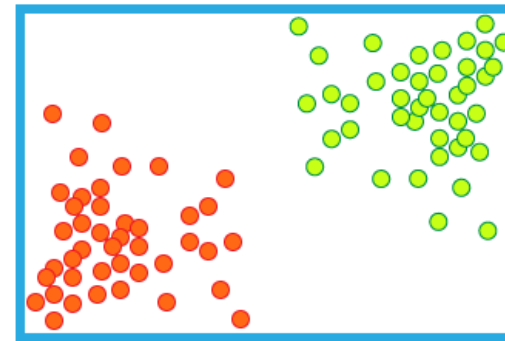
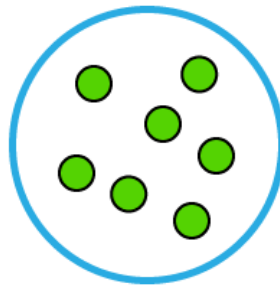
Very Impure



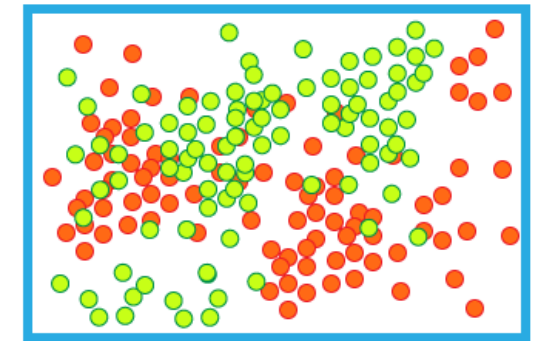
Less Impure



Minimum Impurity



Low Entropy



High Entropy

Advantages of CART

- A CART model is more easily interpretable than a linear regression or logistic regression model
 - Simple rules to determine a prediction
 - Provides transparency to decision process
 - Graphical display
- CART selects the significant variables for us
- CART can deliver nonlinear predictions

References

- <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>
- <https://medium.com/geekculture/applying-7-classification-algorithms-on-the-titanic-dataset-278ef222b53c>

Any Questions?

Thank You.