Decision Trees

A lesson by: Mortaja Alqasab

Objectives

By the end of this lesson students will be able to

- Implement a decision tree model
- Predict new events based on a decison tree
- Manipulate and change the parameters of a decision tree

Decision Trees

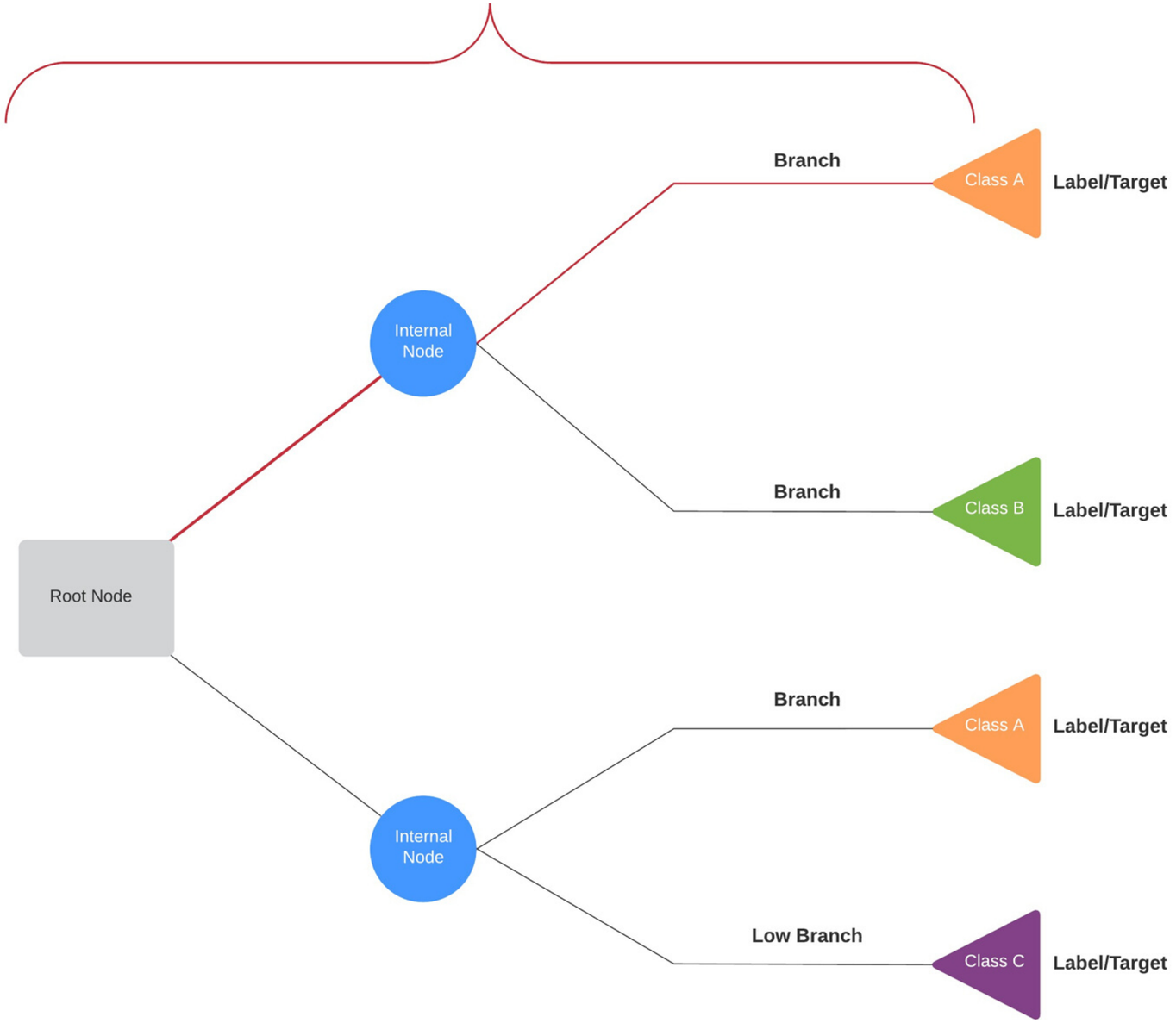
What is a Decision Tree?

- A Decision Tree is a type of a supervised machine learning model used both for classfication and regression problems
- They use tree-like structure to represent decisions by splitting features to construct nodes and branches to represent decisions
- The objective of constructing the tree is to predict a class label or a continues variable

Decision Trees

What do they consist of?

- Branch: represents an outcome from a test on a feature (Yes/No)
- Internal Node: represents a test on a feature (Is it cloudy?)
- Leaf Node: Label/Target (Raining/Dry)
- Classification Rule: The path from the root node to a leaf node.


Root Node
Internal Node
Internal Node
Branch
Class A
Label/Target
Branch
Class B
Label/Target
Branch
Class A
Label/Target
Low Branch
Class C
Label/Target

Metrics

How can we evaluate a decision tree?

- Gini Imputiry: is the likelihood a chosen sample is incorrectly labled if randomly chosed in subset.

- Information Gain: is the most commonly used metric and it is represents how much information is gained from splitting the data when a Internal Node is tested

- Variance Reduction: usually used for Regression Trees and the is the reduction of the total variance of the target when splitting a feature

Hyperparamters

What common parameters used to tune a Decision Tree?

- criterion: the metric used to evaluate the quality of a split ('gini', 'entropy')
- splitter: decides how to choose a split ('best'; for best split, 'random'; for best random split)
- max_depth: the maximum depth of a tree
- min_samples_split: the minimum samples required to split a node

Contact Me

https://www.linkedin.com/in/mortajaalqassab

moqassa@gmail.com