## Objectives

Decision tree methods are a common baseline model for classification tasks due to their visual appeal and high interpretability. This module walks you through the theory behind decision trees and a few hands-on examples of building decision tree models for classification. You will realize the main pros and cons of these techniques. This background will be useful when you are presented with decision tree ensembles in the next module.

* Describe and use decision trees and decision-tree ensemble models for classification
* Identify and implement common ensemble models for classification, including bagging, boosting, stacking, and random forest
* Become familiarized with the pros and cons of decision tree methods
* Build decision trees models with sklearn

Intro to decision trees

* Want to predict whether customers will play tesnnis based on temperature, humidity, wind, outlook
* Segement data based on features to predict results.
* Trees that predict categorical results are decision trees

Example: use slope and elevation in himalayas

Predict average precipitiation (continuous value)

Values at leaves are averages of members

Building a decision tree

* Select a feature and split data into binary tree
* Continue splittting with available features.

Until:

* Leafe nodes are pure )only one class remains)
* A maximum depth is reached.
* A performanc emetric is achieved

Building the best decision tree

* Use greedy search: find the best split at each step
* What defines the best split
* One that maximizes the info gained form the split
* How is info gained defined

Splitting based on classification error

* Classification error equation

Entropy-based splitting

Splitting based on ntropy allows further splits to occur

Classification error vs entropy

* Classification erorr is a flat funciton with a maximum at center
* Center represents ambiguity 50/50spsli
* Splitting meterics favor results that are furtherst away from the center

Entropy has the same maximum but is curved

Curvatiure allows splitting to continue to curve

With classification error, the function is flat

Final avverage classificaiton error can be identical to parent

With entropy gain, the function has a bulge

Allows avg. Info. Of children to be less than parent

Results in info gain and continued gain

Gini index

Often used for splitting

Function is similar to entropy (buldge)

Doesn't contain logarithm.

Decision trees are high variance

* Problem decision trees tend to overfit
* Small changes in data greatly affect prediction (high variance
* Solution prune trees

How to ddecide what leaves to prune

Solution prune based on classification error threshold

Easy to interpret and implement

(if then else – logic)

Handle any data category

No preprocessing or scaling required

Decision tree classifier syntax

From sklearn.tree import DecisionTreeClassifier

Decision trees notebook Pt. 1

Stratefied shuffle split

Pt.2

Pd.Series ??

Dt.preict???

Pt.3

Grid search with cross validation

Mean squared error

Plotting out decision tree

## End of module review: Decision Trees

### **Decision Trees**

Decision trees split your data using impurity measures. They are a greedy algorithm and are not based on statistical assumptions.

The most common splitting impurity measures are Entropy and Gini index. Decision trees tend to overfit and to be very sensitive to different data.

Cross validation and pruning sometimes help with some of this.

Great advantages of decision trees are that they are really easy to interpret and require no data preprocessing.