Decisions with Machine Learning

What is Machine Learning?

- Machine learning is a set of techniques that use data and algorithms to predict results and gradually improve by gaining access to more data.
- In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labelled or unlabeled, your algorithm will produce an estimate about a pattern in the data.

Machine Learning Methods

- Supervised Learning: defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts itself until the model understands the nuances of the data.
- Unsupervised Learning: Uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention.

Supervised Learning

- Supervised Learning can be separated into two main types of problems namely, classification and regression
- **Classification**: use of an algorithm to accurately classify given data into specific categories. Eg. whether a loan gets approved or not
- **Regression**: used to predict the value of a continuous variable from known data. Eg. predicting the price of a home

Classification Algorithms

- Choosing the right classification algorithm is very important. An algorithm that performs classification is called a classifier. A classifier algorithm should be fast, accurate, and sometimes, minimize the amount of training data that it needs.
- We have many different algorithms available to choose from including decision trees, logistic regression, k-nearest neighbors etc
- Here, we will be focusing on decision trees.

Decision Trees

- In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions.
- An example of a decision tree can be seen in the picture

Decision Trees Example

Am I out of shampoo?
- Yes → Is it raining?
  - Yes → Do not go to the market
  - No → Go to the market
- No → Do not go to the market

Decision Trees Contd.

- A decision tree is a different type of classifier that performs a **recursive partition of the sample space**
- A decision tree typically starts with a single node, which branches into possible outcomes. Each of those outcomes leads to additional nodes, which branch off into other possibilities. This gives it a treelike shape.
- Leaf nodes are the final nodes of the decision tree after which, decision tree algorithm won't split the data.
- The other nodes in the tree are interchangeably called *split nodes*, *decision nodes* or *internal nodes*.

Decision Trees Contd.

- Advantages of Decision Trees:
  - Easy to read and interpret
  - Easy to build
  - Less preprocessing of data required
- Disadvantages of Decision Trees:
  - Can become complex for large amounts of data
  - Susceptible to small changes in data
  - Training time increases a lot with increase in data

Example of a decision tree

- dtree.png from the github repo

Interpreting decision trees

- Lets now switch over to the jupyter notebook to talk a little more about how decisions are made using the results of this algorithm