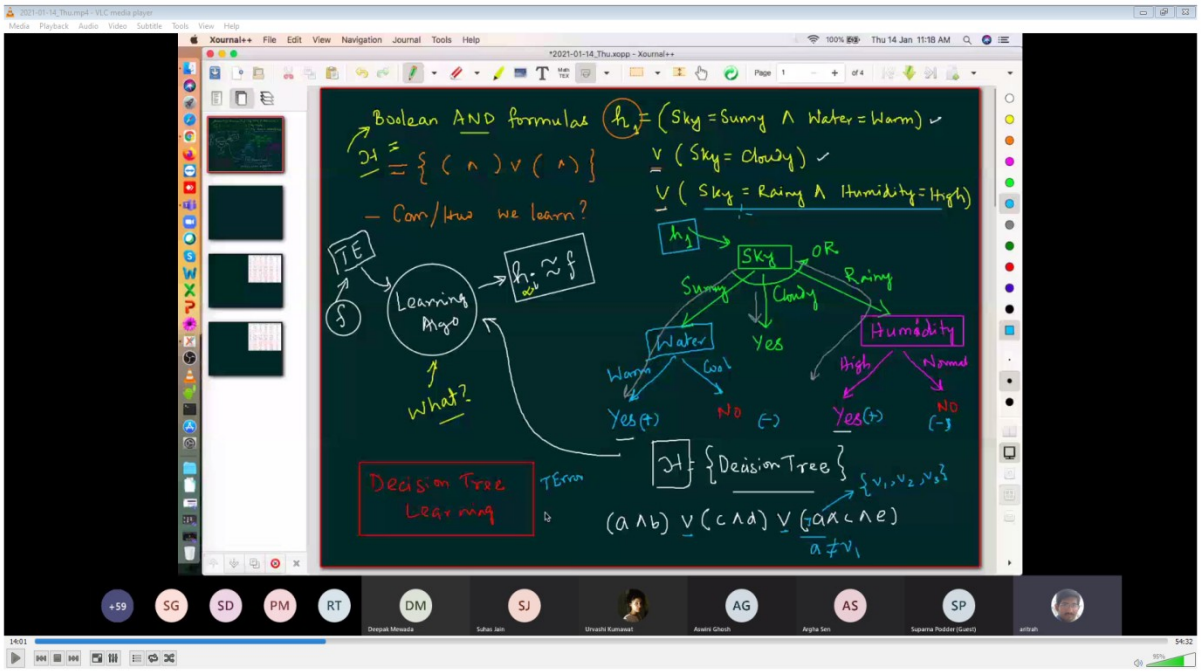
Lecture Scribe For Machine Learning(CS60050),

Spring 2020-2021

Decision Tree

Date: 14th January, 2021

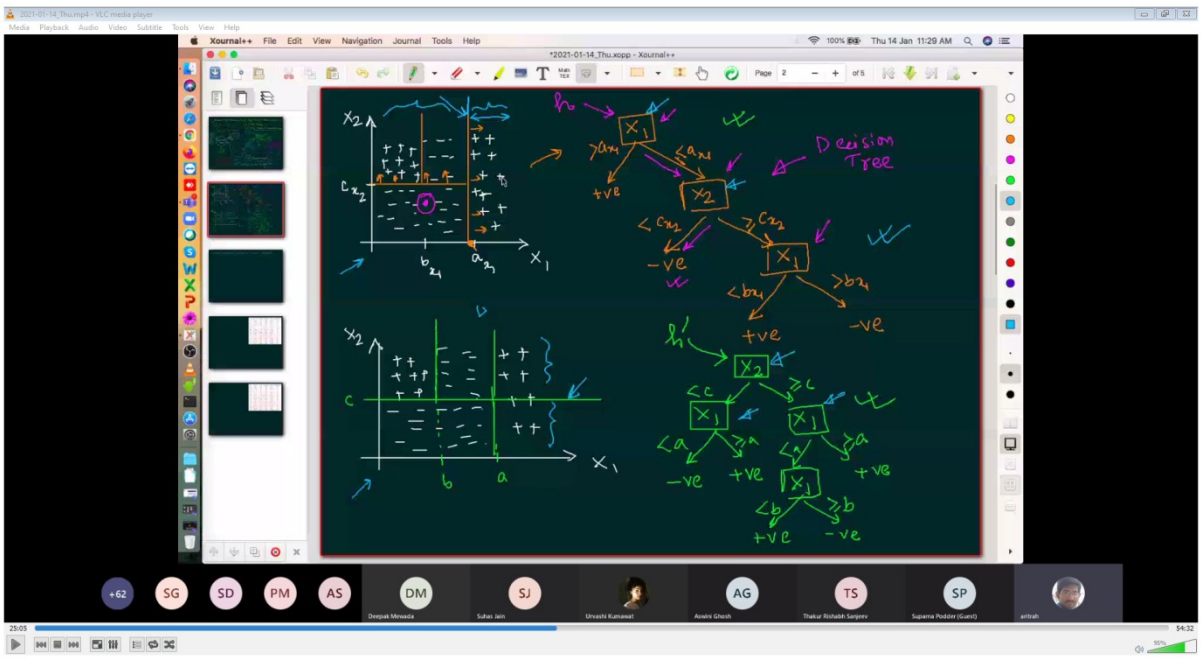


Decision Tree Learning

Decision tree learning is a method for approximating discrete-valued target functions. The learned function is represented by a decision tree. Decision tree learning is one of the most widely used and practical methods for inductive inference. Decision tree learning method searches a completely expressive hypothesis.

Decision Tree

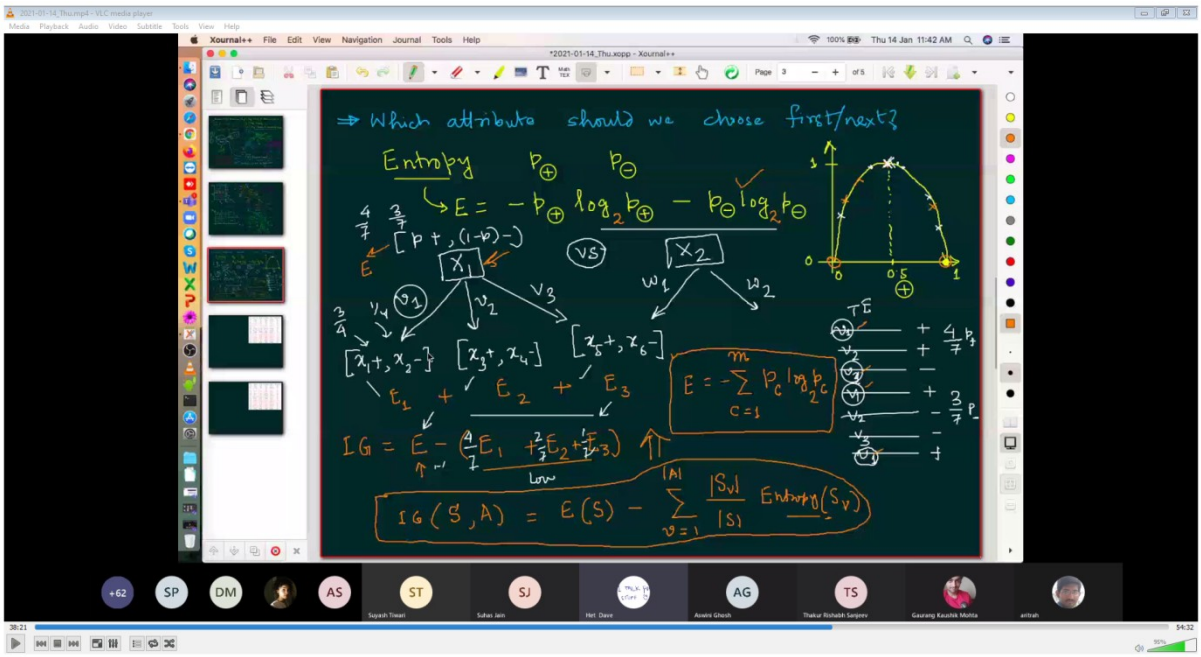
Decision trees represent a disjunction of conjunctions of constraints on the attribute values of instances. Each path from the tree root to a leaf corresponds to a conjunction of attribute tests. The tree itself is a disjunction of these conjunctions. Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. Each node in the tree specifies a test of some attribute of the instance. Each branch descending from a node corresponds to one of the possible values for the attribute. Each leaf node assigns a classification.



Which Attribute is most suitable

We would like to select the attribute that is most useful for classifying Examples. Information gain measures how well a given attribute separates the training examples according to their target classification. In order to define information gain precisely, we use a measure commonly used in information theory, called entropy.

Entropy



Entropy characterizes the impurity of an arbitrary collection of examples.

Given a collection S, containing positive and negative examples of some target concept, the entropy of S relative to this Boolean classification is:

Entropy(S) = -p+ log2p+ - p- log2p-

- S is a sample of training examples
- p+ is the proportion of positive examples
- p- is the proportion of negative examples

Entropy – Non-Boolean Target Classification

If the target attribute can take on c different values, then the entropy of

S relative to this c-wise classification is defined as

Entropy(S) = summation(i=1 to i=c) -pi log2pi

- pi is the proportion of S belonging to class i.
- The logarithm is still base 2 because entropy is a measure of the expected encoding length measured in bits.
- If the target attribute can take on c possible values, the entropy can be as large as log2 c.

Information Gain

- entropy is a measure of the impurity in a collection of training examples
- information gain is a measure of the effectiveness of an attribute in classifying the training data.

- **information gain measures the expected reduction in entropy by partitioning the examples according to an attribute.**

IG(S,A) = Entropy(S) – summation(v=1 to v= |A|) (|Sv| / |S|) Entropy(Sv)

- **S – a collection of examples**
- **A – an attribute**
- **Sv – the subset of S for which attribute A has value v**

Gini Impurity

Gini Impurity is a measurement used to build Decision Trees to determine how the features of a data set should split nodes to form the tree. More precisely, the Gini Impurity of a data set is a number between 0-0.5, which indicates the likelihood of new, random data being miss classified if it were given a random class label according to the class distribution in the data set.

Entropy vs Gini Impurity

The maximum value for entropy is 1 whereas the maximum value for Gini impurity is 0.5.

As the Gini Impurity does not contain any logarithmic function to calculate it takes less computational time as compared to entropy.

Conclusion

we have covered a lot of details about decision tree, how it works and maths behind it, attribute selection measures such as Entropy, Information Gain, Gini Impurity with their formulas, and how machine learning algorithm solves it.

Scribe prepared by Sudhanshu Sharma(18CS30041)