Weekly report of lessons

Name: Ravi Pratap Singh

Roll No: 20CS60R60

The week : 16sept , 17sept ,18sept

The topics covered :Handling Noise in data,Occam's Razor,Model Selection, Representation of concept ,Decision tree ,Applications of decision tree,Decision tree structure, Principle of decision tree construction, Choosing the best attribute ,Gain Function , ID3 ,Overfitting.

Summary topic wise :

- The major sources of noise in data are : Not very accurate measurement of features, Error in labelling,Missing additional attributes. Zero error is not always desirable and also not feasible.
- Occam's Razor states that a simple model will generalize better than a complex model, it helps to shave off unnecessary complexities.
- A model should be selected depending upon the number of parameters,degree of polynomial for regression. Divide data set into 3 parts, Training ,Validation and test.
- A concept could also be represented as disjunction of conjunction of attributes
- A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute ,each branch represents the outcome of the test, and each leaf node represents a class label .
- Instances described by a fixed set of attributes , Target function is discrete valued etc.
- Decision tree can be seen as the set of positive and negative values ona graph and can be partitioned with the help of axis parallel lines.
- Try to form pure leaves as early as possible , For each data set choose the split that reduces impurity and then split the set in which highest reduction of impurity is there.
- The best attribute to split is chosen which gives more purity , information content and less uncertainty . It is done with the help of Entropy measure and Information gain .
- The information gain is based on the decrease in entropy after a dataset is split on an attribute. Constructing a decision tree is all about finding the attribute that returns the highest information gain .
- ID3 algorithm iteratively (repeatedly) dichotomizes(divides) features into two or more groups at each step.
- Training data gived more error than testing data.

Concepts challenging to comprehend : Calculating Entropy.

Interesting and exciting concepts : Building the decision tree.

Concepts not understood: None

Any novel idea of yours out of the lessons : None

Difficulty level of the Quiz :Fair

Did the quiz questions enhance your understanding of the topics covered : Yes ,the questions asked in the quiz were a bit on the challenging side and forced me to think deep about what I have learned . Moreover some of the questions that I couldn't answer also enhanced the knowledge about that topic and realize what I have missed out.