greatlearning
Learning for Life

DATA MINING (Week 2)

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

greatlearning
Learning for Life

DSBA CURRICULUM DESIGN

FOUNDATIONS

Data Science Using Python

Statistical Methods for Decision Making

CORE COURSES

Advanced Statistics

Data Mining (Week-2/6)

Predictive Modelling

Machine Learning

Time Series Forecasting

Data Visualization

DOMAIN APPLICATIONS

Financial Risk Analytics

Web & Social Media Analytics

Marketing Retail Analytics

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

greatlearning
Learning for Life

LEARNING OBJECTIVE OF THIS MODULE

- Clustering
- CART & Model Performance Measures
- Random Forest
- Neural Network

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

greatlearning
Learning for Life

LEARNING OBJECTIVES OF THIS SESSION -

- CART
- Model Performance Measures

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

greatlearning
Learning for Life

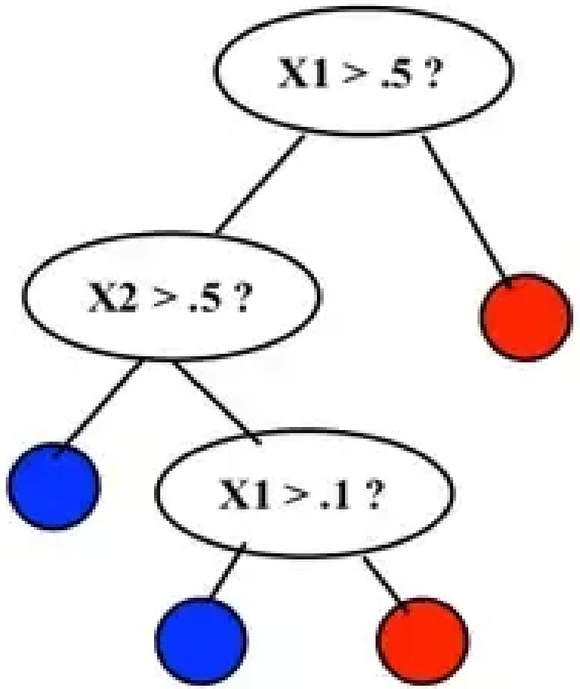
TRY ANSWERING THE FOLLOWING

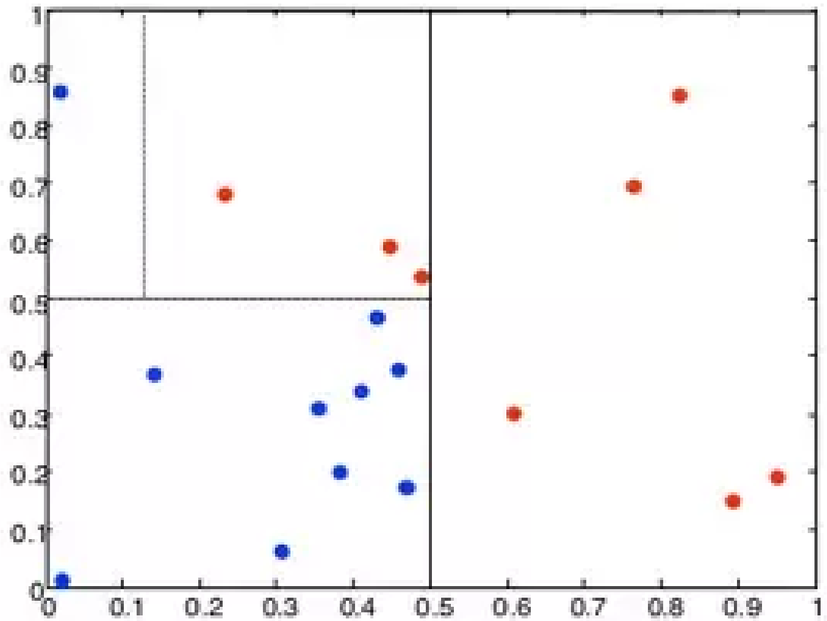
- Does CART classification tree technique makes use of Euclidean Distance?
- Is it preferable to do PCA before CART?
- Name a few Model Performance techniques.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

greatlearning
Learning for Life

BROAD OVERVIEW- Classification Tree





Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

greatlearning
Learning for Life

Industry Application - Predicting Supreme Court decisions using CART

In the US, the legal system operates at three levels - District Courts (makes initial decision), Circuit Courts (deals with appeals against the District Level decision) and Supreme Court (if still not resolved, it makes final decision on the appeal).

In the year 2002, Prof Andrew Martin, a professor of Political Science at Washington University decided to use statistical modelling techniques to predict the outcome of Supreme Court against a panel of Legal Experts. The data was collected from 1994 to 2001 in order to make predictions for the year 2002. The dependent variable was 0 (if the lower court decision was affirmed) or 1 (if the lower court decision was reversed).
Using 628 cases from 1994 to 2005 predictions were made for 68 cases that were to be decided in October 2002.

A set of 83 legal experts recruited to perform the same predictions were able to deliver only 59% accuracy whereas the CART model delivered 75% accuracy.

Reference: http://www.wusct.wustl.edu/media/man1.pdf

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

greatlearning
Learning for Life

CASE STUDY - Loan Delinquent

Based on the given loan data can we understand the major factors or characteristics of a borrower which makes them to get into delinquent stage.

- Delinquency is a major metric in assessing risk as more and more customers getting delinquent means the risk of customers that will default will also increase.
- The main objective is to minimize the risk for which you need to build a decision tree model using CART technique that will identify various risk and non-risk attributes of borrower's to get into delinquent stage.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

greatlearning
Learning for Life

ANY QUESTIONS

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

greatlearning
Learning for Life

HAPPY LEARNING

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.