MG 265 August 3:0

Data Mining

Instructor

Parthasarathy Ramachandran
Email: parthar@iisc.ac.in

Teaching Assistant

Email:

Department: Management Studies

Course Time: Tue The 9-10:30 AM

Lecture venue: Management Studies Classroom 2

Detailed Course Page:

Announcements

Brief description of the course

The objective of this course is to read about various data mining algorithms and implement some of them using the MapReduce framework. Specifically we will discuss association rules, clustering, and classification.

Prerequisites

Probability and statistics, any programming language

Syllabus

Module 1 - Python programming

Introduction to python programming

List, tuple, dictionaries

Subscripting

Function definitions. Iterators

Importing modules

File handling

Classes, sub-classes and inheritance

Module 2

Hadoop and map reduce framework

Mapreduce algorithms for some useful programming constructs

Module 3

Association rule mining

Apriori property

FP-tree building and conditional FP trees

Module 4

Clustering problem

K-means and Hierarchical clustering

Mixture models and EM algorithm

DBSCAN algorithm

Module 5

Classification

Decision tree building with ID3 and C4.5

Naive Bayes classification

Bayes Net building and interpretation

Module 6

Data Summarisation using Principal component analysis

Course outcomes

After taking this course a student should be able to

1. Write non-trivial programs in Python

2. Work on a hadoop system and work with MapReduce framework

3. Perform market basket analysis using Apriori and FP tree

4. Perform clustering analysis

5. Understand the advantages and limitations of some of the popular clustering methods

6. Build a decision tree for classification

7. Work with BayesNet

8. Use PCA for data summarisation

Grading policy

Midterm - 25%

Final - 25%

Programming assignments and project - 50%

Assignments

Resources