

Machine, Data and Learning (CS7.301)

Spring 2022, IIIT Hyderabad
10 Mar, Thursday (Lecture 13)

Taught by Prof. Vikram Pudi

Overview of Data Analytics (contd.)

Classification

Here, we try to infer a model that classifies data points based on some training data.

An example of a classification method is *k-nearest neighbours* models, which classify records using their k nearest neighbours in the training data. Decision trees are another type of model used for this problem.

Clustering

This problem centres around finding groups of similar records. It is different from classification in that the classes (here called clusters) are not known *a priori*.

One algorithm to do this is the k -means algorithm. It starts with a random partition, and as long as clusters change, each record is added to the cluster to whose mean it is closest, and the means are recomputed.

Data Warehousing

Data warehousing involves extracting, transforming and loading data from multiple sources in an enterprise.