Machine Learning and Data Science: Fundamental and Challenges

Abstract:

Supervised and unsupervised pattern classification are important Machine Learning (ML) techniques which have wide range of applications. While supervised classification techniques use training samples to obtain the class information of test samples, unsupervised classification partitions the given data set into homogeneous groups based on some similarity/dissimilarity metric. The lecture is divided into two parts, where in the first part, we will discuss some of the commonly used supervised and unsupervised machine leaning algorithms like Decision Tree and Bayes classifier, as well as K Mean and DB Scan clustering algorithm with suitable examples.

In the second part of the lecture we will focus on the basic issues and challenges in data science, starting with the simple to complex data and finally to the big data. We will demonstrate how to use the open source architectures like Apache Spark, Hadoop and Cloud, to handle the voluminous datasets along