**Machine Learning**

**Session 1**

1. Introduction to the module, recommended books
2. **Machine Learning (ML)**: Using computational and numerical techniques for analysis and processing of signals and data.
3. **Example techniques of ML**:
   1. *Classification*: Categorization of variables. Examples of binary classification include spam filtering which is based on text analysis on the content of an email and face detection which allows a system to decide if there is a detectable face in a sample picture.
   2. *Regression*: Regression is a statistical measurement that attempts to determine the strength of the relationship between one dependent variable and a series of other variables (known as independent variables). Examples include stock market prediction which uses historic market data to make predications on the prices of stocks in the future.
   3. *Clustering*: Clustering is the task of dividing a set of data points into several groups such that data points in the same groups are similar and dissimilar to the data points in other groups. It is a collection of objects based on similarity and dissimilarity between them. When plotted on a graph, the data points in each group are closely coupled together and separated from other groups. For example, clustering a collection of photos in a data set based on the colour schemes would result in various categories of photos with similar colour schemes. Will be explained further in the specific lecture on clustering.
   4. *Ranking*: Given a query *q* and a collection *D* of documents that match the query, the problem is to rank, that is, sort, the documents in *D* according to some criterion so that the "best" results appear early in the result list displayed to the user. Ranking in terms of information retrieval is an important concept in computer science and is used in many different applications such as search engine queries and recommender systems. Many search engines use ranking algorithms to provide users with accurate and relevant results.
   5. *Recommendation systems*: The purpose of a recommender system is to suggest relevant items to users. Examples include Amazon eCommerce and Netflix.
4. **Types of learning:**
   1. *Supervised Learning*: Learning a function when the result is known to exist in the training data. For example, house price prediction and stock market prediction.
      1. *Regression*: Regression is a supervised machine learning technique which is used to predict continuous values. The goal of the regression algorithm is to allow the future prediction of one of the variables, if given the other values. For example, learn to predict the price of a house given the size of the house.
      2. *Classification*: Classification is a supervised learning concept which categorizes a set of data into classes. The most common classification problems are – speech recognition, face detection, handwriting recognition, document classification, etc.
   2. *Unsupervised learning*: Learning a function when the result is NOT known to exist in the training data. Identification of patterns in data sets containing data points that are neither classified nor labelled. For example, association mining, where algorithms find associations among data points, a capability that retailers, for example, can use to identify what products are often bought together.
   3. *Reinforcement Learning*: Reinforcement learning is the training of machine learning models to make a sequence of decisions. Taking actions in an environment to maximize the notion of cumulative reward.

**Summary**:

ML workflow:

1. Feature Engineering/Design: Pre-process the data based on domain specific/expert knowledge.
2. Model: Choose a machine to make the prediction using these features.
3. Cost Function: Write an equation that describes how well or badly your model is doing.
4. Minimize cost function: Write an algorithm to minimize/maximize your cost function, preferably provably so.
5. Generalization: Check that you have not overfitted or underfitted the data.