**Week 2: Introduction to Classification**

This week we continue with supervised learning and focus on **classification**

in which one’s targets/labels can have only a discrete set of values. Classification forms the basis of any complex decision making system, where we often have to decide between (yes=1/no=0) [binary decision making] or decide the category a data sample fits in, such as grocery item belonging to dairy, or breakfast, or frozen or produce etc. [multi-class classification].

1. *Classification*

This week we introduce the concept of hypothesis testing and classification. We introduce several models to perform classification.

Linear: Logistic regression, Support vector Machine with linear kernel

Non linear: Neural Network, Support vector machine with Gaussian Kernel, Decision Trees (Random Forest).

Next, we learn data wrangling and feature selection methods on a hands-on exercise.

Additionally, we look at real instances when the percentage of data samples belonging to the different classes are highly skewed , i.e. one class contains 80-95% of samples while the other class contains much fewer samples. This would be the situation in any digital shopping portal, where the number of views per product are way more than the number of purchases. We learn how to deal with classification in practical setups to cater to the skewness in data in order to avoid biases in data modeling.

*The coding assignments are much more involved this week, so group study sessions and use of Gathertown is highly recommended.*

*B. Week 2 Learning Objectives:*

* To correlate the concept of hypothesis testing with classification.
* To grasp working knowledge of Logistic regression, Neural Networks, Support Vector Machines and Decision Tree classification models (Linear and Non-Linear Classifiers).
* To extend classification performance from binary to multiple classes.
* To understand the practical implementation of models for unbalanced data sets.
* To learn classification output metrics and visualization methods.
* To apply data wrangling to classification on larger, realistic problem settings.