**Lecture 6 – Large Scale Machine-Learning**



Algorithm = Machine learning

If { algorithm extract information from data + produce summary of data & decision is made from it}

Machine learn a model from the data.

**Unsupervised Learning**

* Means that the input data does not tell the clustering algorithm what the clusters should be
* The clustering algorithms allow us to classify future data into of the clusters.

**Supervised Learning**

* The available data includes information about the correct
* The data classified already is called the training set

**Type of y is Arbitratry**

* y is a real number. ML problem is called regression.
* y is a boolean value true or false (+1 and -1). The problem is binary classification.
* y is a member of some finite set (classes). The problem is multiclass classification.

**RMSE –** root mean square error

**Minimizing RMSE -**

**Large Scale Machine Learning**

**Method**

* Decision trees
* The form of function *f* is a tree.
* Each node of the tree has a function of x that determines to which child or children the search must proceed.
* Decision trees are suitable for binary and multiclass classification.
* Perceptrons
* suitable for binary classification.
* slide 34 – Estimating W
* slide 35 – Convergence
* Properties of perceptron: Separability, Convergence
* Neural nets(Neural Networks)
* Suitable for binary or multiclass classification.
* Support Vector Machines
* an advance over the algorithms traditionally used to select the weights and threshold
* The result is a classifier that tends to be more accurate on unseen data
* Instance based learning