# **14 May – Introduction to Machine Learning (part 1)**

## What is the difference between AI, machine learning and deep learning?

* AI is nothing but the automation of cognition that humans have (humans can see, read, listen, speak, think, touch, feel etc.). Think, touch and feel are part of AGI (artificial general intelligence) as these are higher cognition activities. The other activities are relatively easy to automate and have been done to some competent level.

AI is an umbrella term used for both cognitive and data automation, but data automation is not strictly AI (they can be ML / DL which are subsets of AI).

There are 3 categorizations of AI: Reinforcement learning, deep learning and machine learning.

Reinforcement learning provides an elements of human learning, wherein there is constant feedback whenever we learn. Deep learning and machine learning both can do data automation. However deep learning has a capability to go beyond data automation to cognitive automation whereas machine learning doesn’t.

Another important difference is that reinforcement learning has the shortest time delta in terms of feedback while machine learning has the highest.

## Why is there a hype in AI in recent years?

* This is due to two reasons – i) reduced cost of computing, storage and networking, and ii) availability of huge amounts of data with organizations collected from personal laptops and mobiles.

## What is data science?

* Science involves formulating a hypothesis and conducting experiments that prove or disprove the hypothesis. Data science involves taking a data problem, formulating a business hypothesis, experimenting using different tools (python, spark, statistics, deep learning, machine learning, reinforcement learning etc.) and record observations until you can conclusively to prove or disprove the hypothesis.

## Given a use case, can this be solved using machine learning or do we need deep learning?

* Depending on the use case, deep learning can be overkill in some areas or be required to solve the problem.

## What is the difference between statistics and machine learning?

* There are three concepts here: statistics, statistical modelling and machine learning.

Firstly, statistics and statistical models are not the same.

Statistics is the mathematical study of data. You cannot do statistics unless you have data. A statistical model is a model for the data that is used either to infer something about the relationships within the data or to create a model that is able to predict future values. Often, these two go hand-in-hand.

Machine learning is built upon a statistical framework but also draws upon a large number of other fields of mathematics and computer science.

There are lots of statistical models that can make predictions, but predictive accuracy is not their strength. Machine learning models provide various degrees of interpretability but they generally sacrifice interpretability for predictive power. The purpose of machine learning is obtaining a model that can make repeatable predictions.

Machine learning is all about results, whereas statistical modelling is more about finding relationships between variables and the significance of those relationships, whilst also catering for prediction.

## What is ETP cycle? What do experience (E), task (T) and performance measure (P) mean in the context of a computer or machine learning algorithm?

* Experience means the X variables (i.e., independent variables). Task is the goal of the algorithm, which is the variable we are trying to predict. Performance measure is the measure of how good the machine learning algorithm performs the task.

## What are different names used for the X and Y variables in machine learning?

* X – Input / feature / experience.
* Y – Label / class / task.

## What are the two methods of performance measure of a computer or machine learning algorithm?

* Model algorithmic evaluation and business evaluation. Algorithmic evaluations are empirical methods whereas business evaluations depend on the use case.

## What are the two methods of performance measure of a computer or machine learning algorithm?

* Model algorithmic evaluation and business evaluation. Algorithmic evaluations are empirical methods whereas business evaluations depend on the use case.

## What are supervised and unsupervised learning?

* If the task T or Y variable (label or target variable) is removed from the problem Y = f(X), that is called unsupervised learning. We are trying to identify inherent patterns from the data. If the task T is known (i.e., Y is available and identified), it is called supervised learning. This distinction is important as there are specific algorithms corresponding to each of the two end-goals. Unsupervised learning is tougher because without a task, a performance measure is not available. These categorizations are present in both machine learning and deep learning.

## What are some example algorithms for supervised and unsupervised learning?

* Supervised: Linear regression, logistic regression, decision tree, random forest, k nearest neighbours, naïve bayes, support vector machines.

Unsupervised: K means clustering, association rules.