AI in Marketing
Prof. Zillur Rahman
Department of Management Studies
Indian Institute of Technology, Roorkee

Week 1
Lecture-4 Introduction to AI Algorithms

Welcome to this NPTEL online certification course on Artificial Intelligence in Marketing. And now we are discussing module 4. So, we are talking about introduction to AI algorithms, and we are in chapter 1 and module 4. So, this is what we are talking about, that is introduction to artificial intelligence algorithms. So, to introduce the module, we will talk about what are AI algorithms and how they work. What are the various types of commonly used AI algorithms under the different kind of learning patterns. And we have seen that the various types of learning patterns are supervised learning, unsupervised learning, and reinforcement learning. And then we will talk briefly on the Algorithmic Bill of Rights.

So now to start with what is an AI algorithm. The definition of algorithm is a set of instructions to be followed in calculation or other operations. So, it is a pure simple set of instructions. This applies to both mathematics and computer science. So thus, at the essential level an AI algorithm is the programming that tells the computer how to learn to operate on its own. An AI algorithm is much more complex than what most people learn about in algebra of course. A complex set of rules drive AI program. Determine their steps and their ability to learn without an algorithm AI would not exist.

While a general algorithm can be simple, AI algorithms are by nature more complex. AI algorithms work by taking in training data that helps the algorithm to learn how the data is acquired and is labeled marks key difference between different types of AI algorithm. So, keep in mind how that data is acquired and is labeled. That gives the clear key difference between different types of AI. At the core level an AI algorithm takes in training data labeled or unlabeled supplied by developers or acquired by the program itself and uses that information to learn and grow. Then it completes its task using the training data as a basis. So that training data is so important for this kind of AI. Some types of AI algorithms can be taught to learn on their own and take a new data to change and refine their processes. Others will need the intervention of programmers in order to streamline.

Now let us look at the various types of AI algorithms. So, there are three major categories of AI algorithms that we have already learned in the previous module namely one was supervised learning, the second was unsupervised learning and the third was reinforcement learning. The key difference between these algorithms are in how they are

trained and how they function. So, their differences come from A, how they are trained and B, how they function. Under those categories there are dozens of different algorithms. We will discuss about the most popular and commonly used from each category as well as where they are commonly used.

So supervised learning algorithms. The first and the most commonly used category of algorithm is supervised learning. These work by taking in clearly labeled data while being trained and using that to learn and grow. It uses the labeled data to predict outcomes for other data. The name supervised learning comes from the comparison of a student learning in the presence of a teacher or expert. So that is why it is called as supervised learning. Building a supervised learning algorithm that actually works takes a team of dedicated experts to evaluate and review the results, not to mention data scientists to test the models. The algorithms created to ensure their accuracy against the original data and catch any errors from the artificial intelligence.

So, this is the decision tree. One of the most common supervised learning algorithm, decision trees get their name because of their tree like a structure even though the tree is inverted. So, it is the inverted tree. The roots of the tree are the training data sets and they lead to specific nodes which denote a text attribute. Nodes often lead to another node and a node that does not lead onward is called a leaf. So, this is the decision node that is the root node then we have a subtree decision node, leaf node. This is again leaf node because nothing flows from them. Here in this decision node again another decision node and then this is leaf node, this is leaf node, and this is leaf node.

Decision trees classify all the data into decision nodes. It uses a selection criterion called attribute selection measures which takes into account various measures, some examples would be entropy, gain ratio, information gain etc. Using the root data and following the ASM the decision tree can classify the data it is given by following the training data into sub nodes until it reaches the conclusion. A decision tree diagram with root node, decision node and leaf node for better understanding. So, this is how root node with friends then yes windy, cold yes/no below par, so this is the branch no splitting into walk or cart decision nodes walk above par cart cold etc. So, this is the demonstration of this decision tree diagram.

Another type is random forest. The random forest algorithm is actually a broad collection of different decision trees leading to its name. The random forest builds different decision trees and connects them to gain more accurate results. So that is the main advantage that it gives more accurate results. This can be used for both classification and regression type of supervised learning.

While a solo decision tree has no outcome and a narrow range of groups the forest assures a more accurate result with the bigger number of groups and decisions. It has the

added benefits of adding randomness to the model by finding the best feature among a random subset of features. Overall, these benefits create a model that has wide diversity that many data scientists favor. So, as we can see from the diagram the results of decision tree 1 2 & 3 are combined which is then averaged out or the majority is considered as the final result. So, these are the data sets, so this is decision tree 1, this is decision tree 2 this is decision tree 3. So, with the 3 with the same type of data sets there are 3 decision trees and the 3 results majority voting averaging and then we get the final results.

Another is vector, support vector machines. The support vector machine algorithm is another common AI algorithm that can be used for either classification or regression but is most often used for classification. The support vector machine works by plotting each piece of data on a chart in n dimensional space where n is the number of data points. Then the algorithm classifies the data point by finding the hyperplane that separates each class there can be more than one hyperplane.

The main objective of a support vector machine is to segregate the given data sets in the best possible way. The distance between either nearest point is known as the margin. The objective is to select a hyperplane with the maximum possible margin between support vectors in the given data set. SVM searches for the maximum marginal hyperplane. So these are the two axes X1 and X2 and here it is a negative hyperplane then they are support vector this is positive hyperplane so that is maximum margin and this is maximum margin hyperplane.

So generate hyperplanes which segregates the classes in the best way. The figure on the top shows three hyperplanes black, blue, and orange so these are the three hyperplanes. Here the blue and orange have high classification errors, but the black is separating the two classes correctly. So, this black is separating these two. Select the right hyperplane with the maximum segregation from the either nearest data points as shown in figure at the bottom.

Then there is Naive Bayes. The reason this algorithm is called Naive Bayes is that it is based on Bayes theorem and also relies heavily on a large assumption that the presence of one particular feature is unrelated to the presence of other features in the same class. That major assumption is the Naive aspect in the name. So Naive Bayes is useful for large data sets with different classes. It like many other supervised learning algorithms is a classification algorithm. It is an algorithm that learns the probability of every object, its features and which groups they belong to. It is also known as a probabilistic classifier. For example, you cannot identify a word based on its feature and color as there are many words with similar attributes, but you make a probabilistic prediction about the same and that is where Naive Bayes algorithm comes in. So, these are the three classifiers 1 2 3 and this is Naive Bayes classifiers, and this is how they are classified into three different

categories. Naive Bayes use the following equation that is P (H upon E) is equal to P (E upon H) times P (H) divided by P (E). So, P (H upon E) denotes how event H happens when event E takes place. Then P (E upon H) represents how often event E happens when event H takes place first. P (H) represents the probability of event X happening on its own and P (E) represents the probability of event Y happening on its own.

So then comes linear regression is a supervised learning AI algorithm used for regression modeling. It is mostly used for discovering the relationship between data points predictions and forecasting. Much like support vector machines it works by plotting pieces of data on a chart with the x axis as the independent variable and the y axis as the dependent variable. The data points are then plotted out in a linear fashion to determine their relationships and forecast possible future data.

Linear regression is one of the easiest and most popular machine learning algorithms. So that is the best part that is it is the easiest. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous real or numeric values such as the sales salary age product prices etc. The linear regression algorithm shows a linear relationship between a dependent that is y and one or more independent that is x variables hence that is called as a linear regression. Since the linear regression shows the linear relationship which means it finds how the value of the dependent variable is changing according to the value of the independent variable. The linear regression model provides a sloped straight line representing the relationships between the variables. So, this is these are the independent variables on the x axis and here we have the dependent variable and then we have all these data points and in between is the line of regression. So now it tells that how it will happen here what what will be the dependent variable look like at this level of independent variable.

Then comes logistic regression. A logistic regression algorithm usually uses the binary value 0, 1 to estimate values from a set of independent variables. The output of logistic regression is either 1 or 0, yes or no. An example of this would be a spam filter in email. The filter uses logistic regression to mark whether the incoming mail is spam 0 or not 1. Logistic regression is only useful when the dependent variable is categorical either yes or no.

So if it is not, if it is somewhere in between then logistic regression will not work. The logistic regression model is based on logistic function which is the type of S shaped curve that maps any continuous input to the probability value between 0 and 1. The logistic function allows us to model the relationship between the independent variables and the probability of dependent variable taking on the value of 1. The logistic regression model estimates the coefficient of the independent variables that are most predictive of the dependent variable. These coefficients are used to create a linear equation that is then

transformed by the logistic functions to produce a probability value for the dependent variable taking on the value 1.

The logistic regression is commonly used in fields such as healthcare, marketing, finance and social sciences to predict the likelihood of an event occurring such as whether a patient has a certain disease or whether the customer will buy a product or not. So now as you can see from this figure, we have independent variable X and then we have dependent variable Y. Now this dependent variable varies from 0 to 1. Independent variable can take any value. But dependent variable can take values only between 0 to 1. So this predicts Y lies within the 0 to 1 range. So that is why it is an S shaped curve. So now you see at this level what will be the value of Y, at this level what will be the value of Y and at this level what will be the value of Y.

So, as we have already studied before unsupervised learning algorithms are given data that is not labeled. Unsupervised learning algorithm use that unlabeled data to create models and evaluate the relationships between different data points in order to give more insights to the data.

The next comes k-means clustering. k-means is an algorithm designed to perform the clustering function in unsupervised learning. It does takes this by taking in the predetermined clusters and plotting out all the data regardless of the cluster. It then plots a randomly selected piece of data such as the centroid of each cluster. Think of it as a circle around each cluster with that piece of data as the exact center point. From there it sorts the remaining data points into clusters based on their proximity to each other and the centroid data point for each cluster. The algorithms take the unlabeled data sets as input, divides the data set into k numbers of clusters and repeat the process until it does not find the best cluster. The value of k should be predetermined in this algorithm. The k-means clustering algorithm mainly performs two tasks. The first is to determine the best value of k center points or centroids by an iterative process. The second is to assign each data point to its closest k center. Those data points which are near to the particular k center creates a cluster.

So, the working of the k-means algorithm is as follows. First, select the number of k to decide the number of clusters. Second is select random k points or centroids. The third is to assign each data point to those closest centroids which means form which will form the predefined k cluster. The fourth is to calculate the variance and the place a new centroid of each cluster. The fifth is repeat the third steps which means reassigning each data point to the new closest centroid for each cluster. So, this is how it works before the k-means and this after k-means how neatly they are clustered.

The next comes Gaussian mixture model. Gaussian mixture models are similar to k-means clustering in many ways. Both are concerned with sorting data into predetermined

clusters based on proximity. However, Gaussian models are a little more versatile in the shapes of the clusters they allow. k-means clustering only allows data to be clustered in circles with the centroid in the center of each cluster. Gaussian mixtures can handle data that lands on the graph in more linear patterns allowing for oblong shaped structures. This allows for greater clarity in clustering of one data point lands inside the circle of another cluster.

The starting point and training process of the k-means and GMM are the same. However, k-means uses a distance-based approach and GMM uses a probabilistic based approach. There is one primary assumption in GMM that the data set consists of multiple Gaussians in other words a mixture of the Gaussians. It is used to forecast the sales of a product to understand customer churn through the lengths of different groups of customers. Some AI algorithms can use either supervised or unsupervised data input and still function. They might have slightly different applications based on their status.

The next comes k nearest neighbor algorithm. So k nearest neighbor that is k and n algorithm is a simplistic AI algorithm that assumes that all the data points provided are in proximity to each other and plots them onto a map to show the relationship between them. Then the algorithm can calculate the distance between data points in order to extrapolate the relationships and calculate the distance on the graph. Both supervised and unsupervised algorithms in supervised learning can be used for either classification or regression applications. In unsupervised learning it is popularly used for anomaly detection that is finding data that does not belong and removing it. k and n is a nonparametric algorithm which means it does not make any assumptions on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately. Instead at the training phase it just stores the data sets and when it gets new data it classifies that data into a category that is much similar to the new data. k and n algorithms can be used for regression as well as classification but mostly it is used for classification problems.

Suppose there are two categories category A and category B and we have a new data point x1 so this data point will lie in which of these categories. To solve this type of problem we need a k and n algorithm with the help of k and n we can easily identify the category or class of a particular data set. So now this was category A, this was category B and now we have this new data. Now after k and n this new data point is assigned to category 1. So now it becomes easier to deal with this this data point rather than just having one data point stand alone.

The next comes neural networks. Neural network algorithm is a term for a collection of AI algorithms that mimic the functions of a human brain. So that mimics the functions of a human brain. These tend to be more complex than many of the algorithms discussed above and have applications which are still being discovered. So, all those applications

have not yet been discovered. They are still in the process of discovery. In unsupervised and supervised learning, it can be used for classification and pattern recognition.

It consists of interconnected nodes that is neurons organized into layers. Information flows through these nodes and the network adjusts the connection strengths, that is weights. During training to learn from data enabling it to recognize patterns make predictions and solve various tasks in machine learning and artificial intelligence. And there are three levels in the network architecture. The input layer, the hidden layer which can be more than one and the output layer. Because of the numerous layers it is sometimes referred to as the MLP that is a multi-layer perceptron.

It is possible to think of the hidden layer as a distillation layer which extracts some of the most relevant patterns from the inputs and send them on to the next layer for further analysis. It accelerates and improve the efficiency of the network by recognizing just the most important information from the inputs and discarding the redundant information.

So this is how it works. So, this is the hidden layer then there are input layers, network layers, feed forward. So, this is network output, and this is back propagation. Then comes reinforcement learning algorithm. The last major type of AI algorithm is reinforcement learning algorithm which learns by taking in the feedback from the results of its action.

This is typically in the form of a reward. The reinforcement algorithm is usually composed of two major parts. The first is an agent that performs an action that is one and the second is the environment in which the action is performed. So, these are the two major parts of this algorithm. The cycle begins when the environment sends a state signal to the agent that cues the agent to perform a specific action within the environment.

Once the action is performed the environment sends the reward signal to the agent informing it on what happens so that the agent can update and evaluate its last action. Then with that new information it can take the action again. That cycle repeats until the environment sends a termination signal. So, there are two types of reinforcement the algorithm can use. One is either a positive reward or a negative reward.

In reinforcement algorithms there are slightly different approaches depending on what is being measured and how it is being measured. Here are some definitions of different models and measures. So, one is Policy. The approach that agent takes to determine the next action taken by the agent. The second is Model. The situation and dynamics of the environment and the third is Value. The expected long-term results. This is different from the reward which is the result of a single action within the environment. The value is the long-term result of many actions.

The second is value based. In this value-based reinforcement algorithm the agent pushes towards an expected long-term return. So that is important here. Instead of just focusing on the short-term reward. The next is policy based. A policy-based reinforcement algorithm usually takes one of the two approaches to determine the next course of action. Either a standardized approach so that can be one approach. Where any state produces the same action or the dynamic approach where certain probabilities are mapped out and the probabilities calculated. Each probability has its own policy reactions.

The next is the model based in this algorithm the programmers create a different dynamic for each environment. So, the programmer creates different dynamics for each environment. One environment one dynamics. That way when the agent is put into each different model it learns to perform consistently under each condition.

Now let us look at the Algorithmic Bill of Rights. In January 2017 the US Public Policy Council of the Association for Computing Machinery which consists of educators, researchers and professionals in the world of information technology outlined a set of guiding principles that could serve as a precursor for an Algorithmic Bill of Rights. These principles cover seven general areas which we will be discussing one by one.

So, the first one is awareness. Those who design, implement, and use algorithms must be aware of their potential biases and possible harm and take these into account in their practices. The second is access and redress. Those who are negatively affected by algorithms must have systems that enable them to question the decisions and seek redress. The third is accountability. Organizations that use algorithms must take responsibility for the decisions. Those algorithms reach even if it is not feasible to explain how the algorithms arrive at those decisions.

The fourth is explanation. Those affected by algorithms should be given explanations of the decisions and the procedures that generated them. The fifth is data provenance. Those who design and use algorithms should maintain records of the data used to train the algorithm and make those records available to appropriate individuals to be studied by to be studied for possible biases. The sixth is auditability. Algorithms and data should be recorded so that they can be audited in case of possible harm. The seventh is validation and testing. Organizations that use algorithms should test them regularly for biases and make the results publicly available.

So, to conclude in this module, we have briefly introduced AI algorithms and how do they work. Then we have discussed about the commonly used AI algorithms under the various learning techniques that is supervised, unsupervised and reinforcement learning. And then finally, we have given a brief on the Algorithmic Bill of Rights. And these are the 5 references from which the material for this module was taken. Thank you.