Applied Natural Language Processing
Prof. Ramaseshan Ramachandran
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture - 30
Introduction to Machine Learning

In this lecture, we going to be talking about an Artificial Neural Network. We have been looking at the corpus from the perspective of the numbers first and then we started looking at some context related to the words and so on, later we started predicting the next keyword in a given sentence and then we started looking at what could be the probability of a given sentence by using the history that we have captured through the corpus.

Or we able to relief following the kind of transaction that we are getting into, right, slowly and sturdily from the numbers to the context and then to understanding the sentence and so on, right. We are going to be looking at another tool in this which would be called ANN or Artificially Neural Network which would also help us in terms of understanding whatever we had seen so far. So, this is yet another mechanism to process the corpus and then get some insights out of that, ok

So, our journey has been so far in terms of first identifying what could be the meaning of the word, right and then finding out what are the similarities or similar terms associate with the with every word in the corpus. I am also sure that you have now captured the idea of where we are going in terms of identifying the word, meaning and so on, right. The word meaning is related to the context that we are talking about, you know it is not about the exact dictionary meaning or trying to figure out what could be the synonyms and so on. So, even though we sketch they're using the context available, all right. So, here what I am going to be doing is I am going to be talking about why we need neural networks, ok.

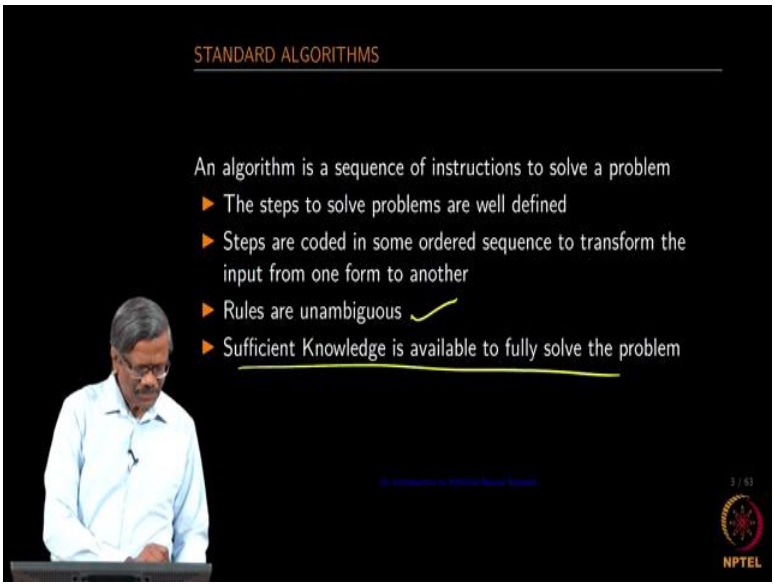
(Refer Slide Time: 02:21)

So, why cannot we just write, the way we write algorithms and then solve problems in natural language processing? So, why a no tool? Right. Then we will talk about what is missing in these standard algorithms, right in terms of creating new rules or in solving certain problems, ok.

And then, we will get into the definition of classification. We have seen this earlier again I will get into the definition of classification and then I show you how we can classify set of points you know in a linear fashion, so that is where we will start in terms of the classification. And then, we will talk about the linear morals of classification and then we will find out what a decision boundary is. And then later get into the artificial neural because we record some kind of a fundamental to get into the artificial neural networks, so we will talk about these elements before we get into the actual artificial neural network.

We will talk about perception, it is not listed here. The basic element of the neural network that is how it all started. And then, we will define what is a what is meant by learning in perception and then we will define a loss function, cost function, what are the limitations of the perceptron and then how do we iterate to really get into the solution by descending down the slope using gradient descent and then later we talk about feed-forward network and some application related to feed-forward network, right.

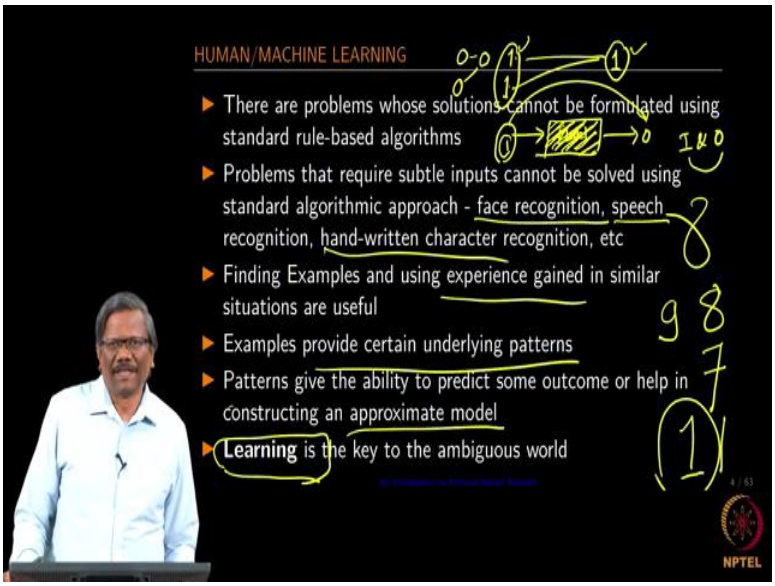
(Refer Slide Time: 04:19)



So, this is something that you all know very well. I do not have to explain to you in detail what we do in the standard algorithmic approach, correct. So, I just briefly mention what it is even though you know all of these. This is nothing but a sequence of instructions to solve a problem very simple, correct. The steps to solve the problems are very well defined. So, you know what is the input, we know what are the rules to manipulate the input and then we know what we expect out of the program, as an output. And then the steps are coded in some ordered sequence to transform the input from one form to the other.

Rules are very unambiguous, right. So, this is something we do very well, you know without specifying the rule is not possible to solve any problem in the algorithmic fashion. So, we need to be very clear about the domain, we need to be very clear about how to define the rules and then how we can manipulate the input using the rules and then later get some output based on those. So, we know very well about the domain to solve the given problem, ok.

(Refer Slide Time: 05:43)



So, why do we need machine learning? First of all we should understand what we learned, right. As a human being, we have been learning right throughout, right from the day we have born we start learning things, right. For example, a newborn child, right after 10 days old or 12 days old, the family starts introducing the family members to the newborn baby, right, ok. This is your dad, this is your grandpa, this is your grandma, this is your sister, brother and so on.

You know poor kid, the right has to listen to what you are saying. I am not sure what it really understands, it only looks at you and then you keep feeding that info every time reinforcing that without giving any input to the child, correct. So, we do not tell the child that your father looks like this with all features and you do not describe any of those. You just mention that this is your farther, this is your mother this is your sister and so on.

If you think about what we are doing with the machine, right. So, we very clearly mention what is the input, what are the roles to manipulate the input and what we expect the output and so on. For the poor newborn baby we do not give anything, but we just feed the outcome, right. We just show what is the output and then ask the child to interpret everything and figure out what are the features of the mother, father and so on. So, we care about similar problems even in real-life here.

For example, you are looking at a car you know from above, you are still able to recollect I am sorry, you are still able to find out and say that, the car is standing there.

But if you are close enough you will be able to even recognize what that car model is and so on. And if you look at the car from any direction top, side, front, back and so on you still able to recognize that, right. So, how do we do that? So, what kind of features that we feed in into your brain, so that we are able to say that this is a car and this car model is, right. So, these are certain kinds of problems that we want to solve.

For example, face recognition is one example. So, if we are looking at the camera like I am doing right now, without turning my face in this fashion. And every time when I come in if I have a face recognition software and I am standing at the right distance from the camera it should be able to recognize my face clearly, right, based on the distance between the eyes, maybe head a distance between my eyebrows and the mouth and lots of features that you feed in order for you to really recognize whether that person is a human or not, right.

And then we have speech recognition. We keep talking and there is an application that recognizes your continuous speech, right. For example, when I am talking, I am talking about this speech recognition, right. So, when I am doing it quickly you know recognize the beach, recognize speech, recognize beach, right. It is so fast that it is difficult to distinguish between whether it recognizes speech or recognize the beach.

So, based on the context we will be able to figure out whether it is the first one or the second one. So, for that we require some kind of features to extracted, which are not going to be provided as a rule in terms of the algorithm. So, the system has to really pick up based on the features that are available and then figure out the rules, construct the rule, and then based on the rules it has captured from the data it should be able to give you some output.

The same way the handwritten character recognition, right. So, there are so many ways you can write 8 you know. Some people write 8 this way, some write this way and so on. See for example, some people write 7 like this. It is 1 or you can is this 1 or 7 underlined, ok. So, there is a way in which this could also be 1, right, g or 9 like this, right. So, this requires certain intelligence and rule that needs to be extracted from the data which are not going to be available as part of the fixed rule set that you input in your algorithm. So, these sorts of applications, you need to build a rule-based on the examples there are

available based on the data that you have captured so far and so on, ok. For that you cannot go for the standard algorithmic approach, ok.

The examples are really helpful in this case. So, if keep feeding lots and lots of examples of this type and then keep reinforcing you know during the training time that what I am feeding is 7 and not 1 and what I am feeding here is 1 and a not a 1 underline. So, you keep reinforcing during the, training process, ok. So, during the training process there is experience gained and those experiences would be useful in real-life situations. The same thing happened to us as well as, right.

So, we keep looking at certain examples before we solve a similar type of problem let say maths or physics or any other engineering subjects. So, we require some examples to start. So, in the same fashion we might be able to provide a good example, so that system learns from the examples. So, they give certain patterns which are captured by the system.

So, certain patterns are not clearly visible from the data. So, the system should be able to really figure out those lantern patterns which are hidden inside based on the examples that you have provided. So, we expect the machine to really do these jobs. So, you know well that the machine is capable of doing the computation really fast, whereas the thinking part is not that great when compared to the human.

But based on the examples if it is able to really extract the patterns and then using those patterns somehow it is able to match the pattern with the incoming new set of data and then say that it may belong to this, that is a great improvement and that is what exactly we want to do in the machine learning. The pattern gives the ability to predict some outcomes or help in constructing an approximate model.

So, what all happens? You know if you look at these normal or this standard programming model we do like this, right. So, let us say that this is your input and this is your output and this is your let us say your model that contains all the rules and other processing capabilities, right. And what we are going to be done in machine learning is this is not known this is the backbox. What all we know is a certain type of inputs that we have and then these inputs have some correspondence to the output.

So, we know that the input and the output are related in some fashion and we want to find out what is the relationship between these and that is what we call the model, ok. That model is going to be estimated based on the examples that we start providing to the system. So, that is where we start looking at the pattern. The system starts looking at the patterns, you know we are also very good in terms of identifying patterns the same fashion system also identifies a lot of lantern patterns inside and start building some approximate model that is what I mentioned in this particular bullet point.

So, when it starts looking at the new examples and then start figuring out that this could be one of the patterns that I had already seen. So, I can approximately club this particular input along with the existing patterns that I have stored as my model and slightly update the model, so that next time in the pattern of this type comes I can provide the output related to that pattern. So, that is what is called learning. Learning is an important aspect of machine learning as well as for human rights. We are all lifelong learns, so in the same way machine also you know go to keep learning the new pattern that is coming in. So, it is very similar to what we do, right.

Suppose, if you have learned some subject and then you are very thorough with that and you have not updated yourself, but you are able to answer most of the queries related to what you have learned, assuming that the particular subject had moved ahead in terms of research and content and you have not updated. And if somebody asks you a question related to the new field you will not be able to answer clearly, right. So, you may understand that it belongs to the subject that you know very well, but you will not be able to give a very clear answer. So, what you do? You start learning reading about it and so on and then looking at various examples of the new topic in the same subject it starts learning one after the other. So, you keep learning and updating yourself every day, right.

In the same fashion, the models are not going to remind static. So, they going to keep changing based on the environment. Supposing, if the model has leaned every pattern related to a certain object, then there is no need for you to really update that model. But I think that scenario own exists, there will be always be something new and the model will be updated on a regular basis and that is what we call as the learning. When you keep changing the model that we are building we call that as the learning process.

So, in the case of the machine learning what happens is since we know the input and the output are somehow related and we start with the known samples, ok. So, let us assume that there are some binary examples. If I feed in 1 and 1 the output will always be 1. If I feed 0 and 0 the output will always be 0. Suppose, if we know these kinds of patterns, right. So, we know that this is the input and this is my output and we since we do not know the model that transforms the input into the output we start providing certain weights or variables into the model and that particular set of parameters will keep changing depending on what you are trying to teach the system.

For example, when you provide these numbers 1 and 1, initially and if the model parameters are not known initially you just set some random values, but you know what the outcome is going to be. Based on the random value it just creates a rough model and outputs some value which is not let us say it is equal to 1. Then you tell the system that a hey, what you are computed is not close enough and it is equal to 1 and you have compared something greater or less than 1, so correct this. And based on the correction the model parameters are updated and so on.

So, once you fed all the inputs and the corresponding outputs related to that and trained this, the system is going to be in a state where when you give the input that you have used for training will always give the expected output. So, that time you claim that the model has been built and it is ready to do the real-time, real-life job, ok. So, this is what is machine learning at a very high level.

So, how is it useful in natural language processing? So, what is that we are going to be doing in the natural language processing with machine learning and so on? We already have a model where we can estimate the next word, we can figure out what could be the probability of a new sentence based on whatever corpus that I have used, I already estimated all the parameters and so on. I already have a model. What new that we can do?

So, we will talk about that as we move along. So, one thing that we are definitely going with machine learning is to look at the classification aspect. I going to be looking at the word embedding, you remember we spoke about the word embedding in the LSI model, where we used SVD to figure out the latent patterns inside the term-document matrix, right.

In the same fashion, we will do the word embedding extraction using a neural network. And then later we will also get into the aspects of learning a sentence, how can I encode a sentence which we are not seen in the probabilistic model? How could I encode a paragraph? How do I encode a problem statement? How do I translate from one language to another? Again, you can use the statistical machine translation to do that, but what is so great about machine translation. So, we will see in the later lectures. We will also talk about how conversations can be modeled and so on.