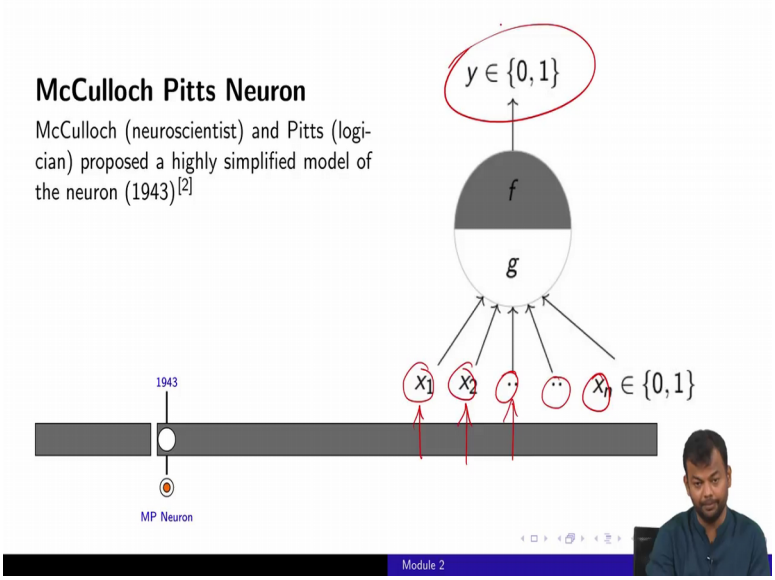
Deep Learning
Mitesh M. Khapra
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 01
Chapter 2: From Spring to Winter of AI

And now, what we will get into in the next chapter is, we will start talking about artificial intelligence. And this is titled as From the Spring to the Winter of AI. So, I am going to talk about when was this boom in AI started? OR when is that people started thinking and talking about AI seriously. And what eventually happened to the initial boom and so, on right.

(Refer Slide Time: 00:34)



So, let us start with 1943. Whereas, I saying that there was a lot of interest in understanding, how does a human brain work? And then come up with a computational or a mathematical model of that. So, Mcculloch Pitts, one of them was a neuroscientist and the other one was a logician right, no computer scientists or anything at that point of time.

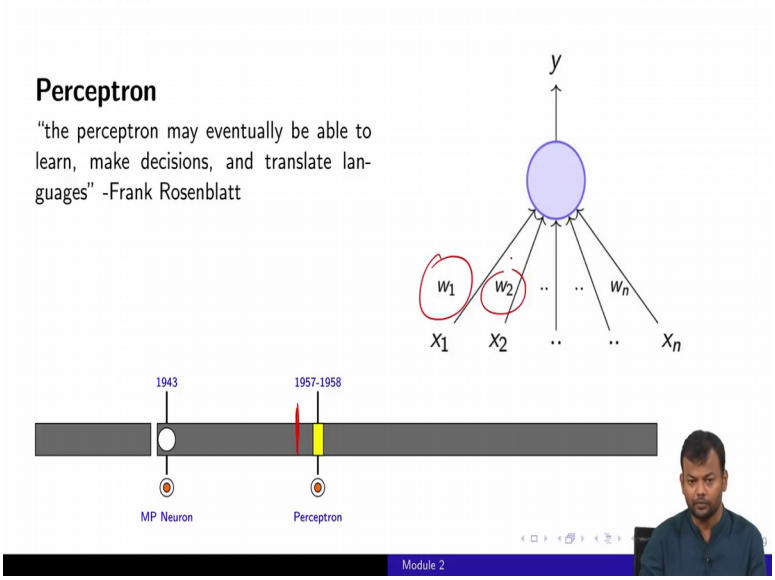
And they came up with this extremely simplified model, that just as a brain takes a input from lot of factors right. So now, suppose you want to decide whether you want to go out for a movie or not. So, you would probably think about, do you really have any exams coming up that could be our factor X 1. You could think about is a weather good to go

out is it raining would it be difficult to go out at this point. Would there be a lot of traffic is it a very popular movie and hence tickets may not be available and so on right.

So, being kind of presses all this information you might also look at things like the reviews of the movie, or the IMDB rating of the movie and so on. And based on all these complex inputs, it applies some function and then takes a decision yes or no that, I want to probably go for a movie. .

So, this is an overly simplified model of, how the brain works is? And what this model says is that, you take inputs from various sources and based on that you come up with the binary decision right. So, this is what they proposed in 1943. So now, we have come to an artificial neuron. So, this is not a biological neuron this is how, you would implement it as a machine right. So, that was in 1943.

(Refer Slide Time: 02:01)



Then along and then this kind of led to a lot of boom in our interest in artificial intelligence and so on. And I guess around 1956, in a conference the term artificial intelligence was a formally coined. And within a 1 or 2 years from there Frank Rosenberg came up with this Perceptron model of doing computations or what Perceptron model of what an artificial neuron could be?

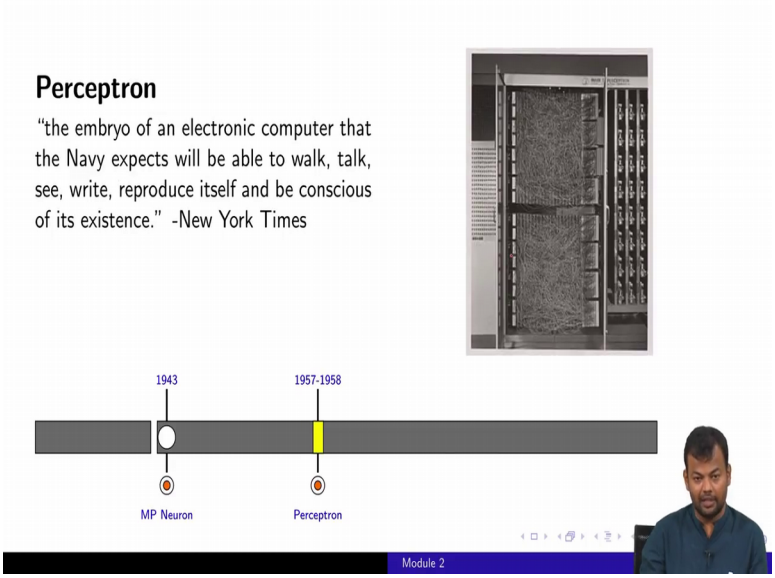
And we will talk about this in detail later on. The course and not tell you what these things are as of? Now, just think of the a new model was proposed and this is what he

had to say about this model right. So, he said that, the Perceptron may eventually be able to learn, make decisions and translate languages. Do you find anything odd about this statement yeah? So, learn and make decisions make sense, but why translate languages? Why is so specific? Why such a specific interest in languages? Right.

So, that you have to connect back to history right. So, this is also the period of the cold war and there was all those. Always a lot of interest; A lot of research and translation was actually fueled by the world war and evens that happened after that where, these countries which were at loggerheads with each other.

They wanted to understand what the other country is doing? But they did not speak each other's language. That is why, there was a lot of interest from espionage point of view or from spying and so on to be able to translate languages. And hence, that specific require and lot of this research would have been funded from agencies, which are interested in these things right, and the defence or war or something.

(Refer Slide Time: 03:36)



So, and this work was largely done for the navy and this is an this is an extract from the article written in New York times; Way back in 1957 or 58, where it was mentioned that the embryo often this Perceptron is an embryo of an electronic computer. That the navy expects will be able to walk, talk, see, write, reproduce itself and be conscious of it is existence. .

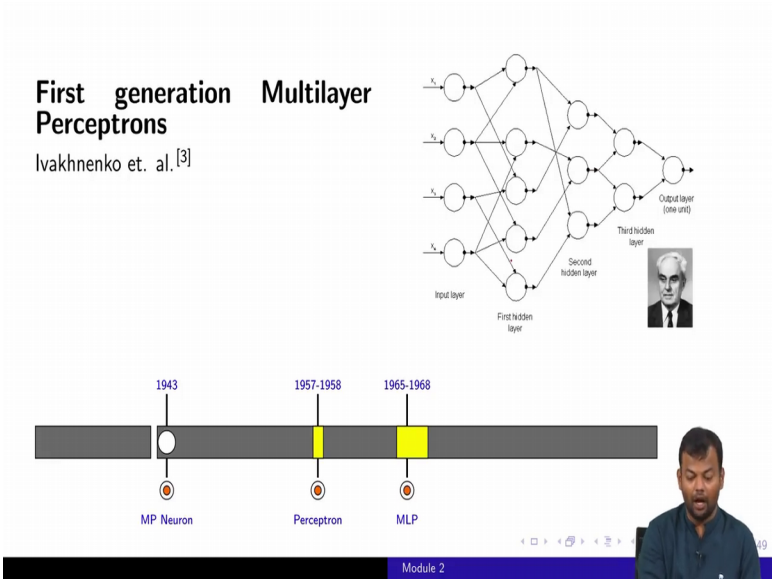
So, I am not quoting something from 2017 or 18. This is way back in 1957 58. Why I am that is why I like the history part of it. So, recently there is a lot of boom or a lot of hype around, AI that AI will take over a lot of things will take our jobs. It might eventually, we might be colonized by AI agents and so on.

So, I just want to emphasize that, I do not know whether, that will happen or not, but this is not something new we have been talking about the promise of AI as far back. Since, 1957 1958 right, this not something new that people are talking about. Now, it is always been there and to what extend, this promise will be fulfilled is yet to be seen.

And of course, as compared to 1957 58, we have made a lot of progress in other fields which have enabled. AI to be much more successful than it was earlier for example, we have much better compute power. Now we have lots of data now and all thanks to the internet, and other things that you can actually crawl tons and tons of data and then try to learn something from a data or try to make the machine learn something from it right.

So, we have made a lot of progress in other aspects where, which AI is now at a position. Where, it can really make a difference? But just wanted to say that, these are not things which I have not been said in the past; it has always been the day has always been considered to be very promising and perhaps a bit hyped also right. So, that is about 1957 58.

(Refer Slide Time: 05:25)

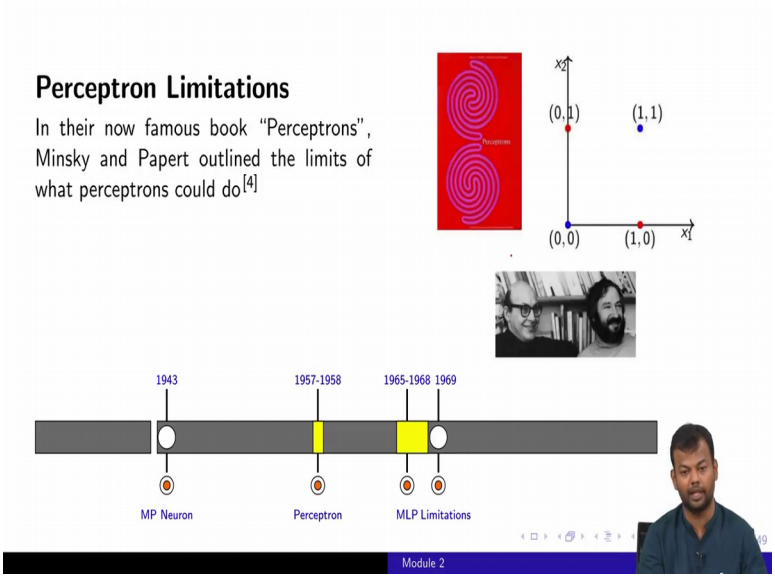


Then now, what we talk about? What is all the for the past 8 to 10 years? At least when we talk about AI talking about deep learning? And that is what this course is about largely about deep learning. I am not saying that other and what deep learning is largely about? If I want to tell you in a very layman nutshell term is. It is about a large number of artificial neurons connected to each other in layers and functioning towards achieving certain goal right.

So, this is like a schematic of what a deep neural network or a feed forward neural network would look like. Now this is again not something new which is up in the last 8 to 10 years; although, people have started discussing it a lot in the last 8 to 10 years. Look at it way back in 1965 68 opposed something which looked very much like a modern deep neural network or a modern feed forward neural network.

And in many circles, he is considered to be one of the founding fathers of modern deep learning right.

(Refer Slide Time: 06:26)

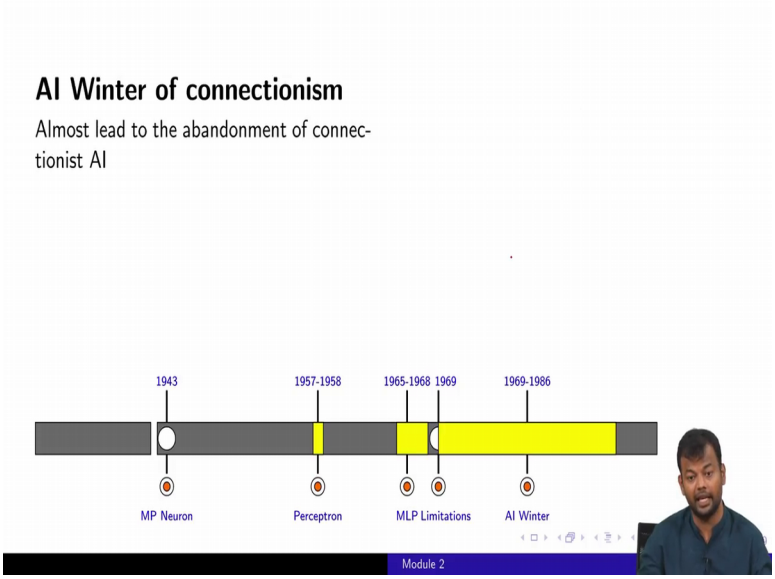


So, that is about that (Refer Slide Time: 6:30) right. From 1943 to 1968, it was mainly about the springtime for AI and what I mean by that? That everyone was showing interest in that. The government was funding a lot of research in AI and people really thought that AI could deliver a lot of things on a lot of (Refer Slide Time: 6:46) for various applications, health care defence and so on.

And then around, 1969 an interesting paper came out by these 2 gentlemen Minsky and Papert. Which essentially outlined some limitations of the Perceptron model and we will talk about these limitations later on in the course. In the second or third lecture, but for now I will not get into a details of that, but what it is said that it is possible that a Perceptron cannot handle some very simple functions also.

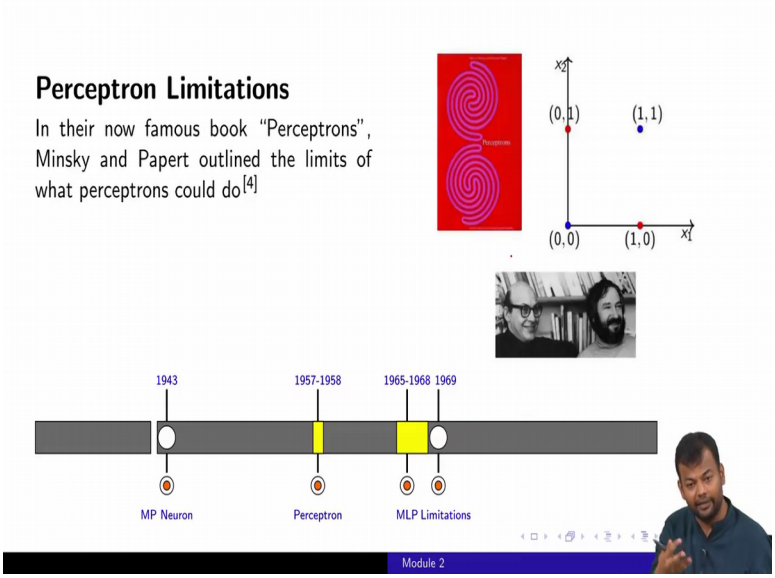
So, you are trying to make the Perceptron learn some very complex functions. Because, the way we decide how to watch a movie is a very complex function of the inputs that we considered, but even a simple function like x or is something which a Perceptron cannot be used to model that is what this paper essentially showed. And this led to severe criticism for AI and then, people started losing interest in AI and lot of government funding actually subsided after 1969.

(Refer Slide Time: 07:38)



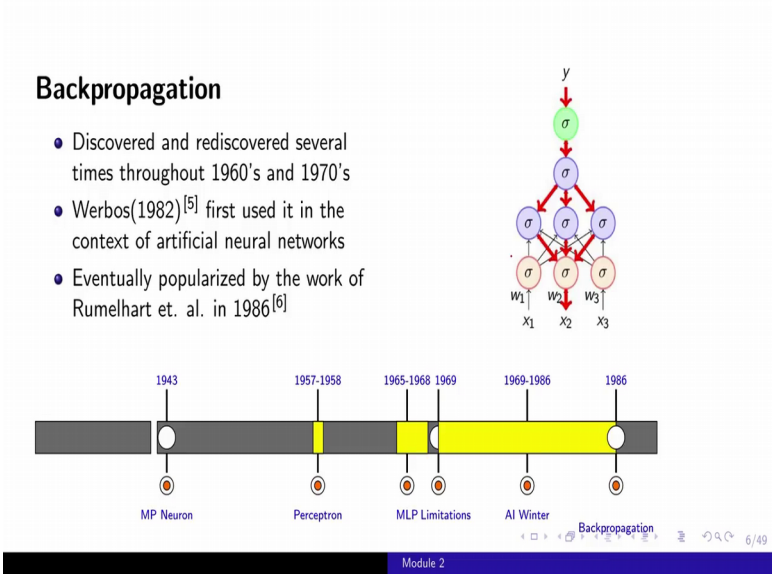
All the way to 1986 actually this was the AI winter of connectionism. So, there was very little interest in connectionist AI. So, there are 2 types of AI one is symbolic AI and the other is connectionist AI. So whatever, we are going to study in this course about neural networks and all. That probably falls in connectionist AI paradigm and there was no interest in this and people. I mean hard to get funding and so on for these 17 to 18 years.

(Refer Slide Time: 08:02)



And that was largely triggered by this study that was done by Minsky and Papert and interestingly they were also often misquoted and what they had actually said in that papers. So, they had said a single Perceptron cannot do it. They In fact, said that a multi-layer network of Perceptrons can do it, but no one focused on the second part that, a multi-layer network of Perceptron people started pushing the idea that a Perceptron cannot do it. And hence, we should not be investigating it. And so, on right, so that is what happened for a long time and this known as the winter the first winter.

(Refer Slide Time: 08:32)

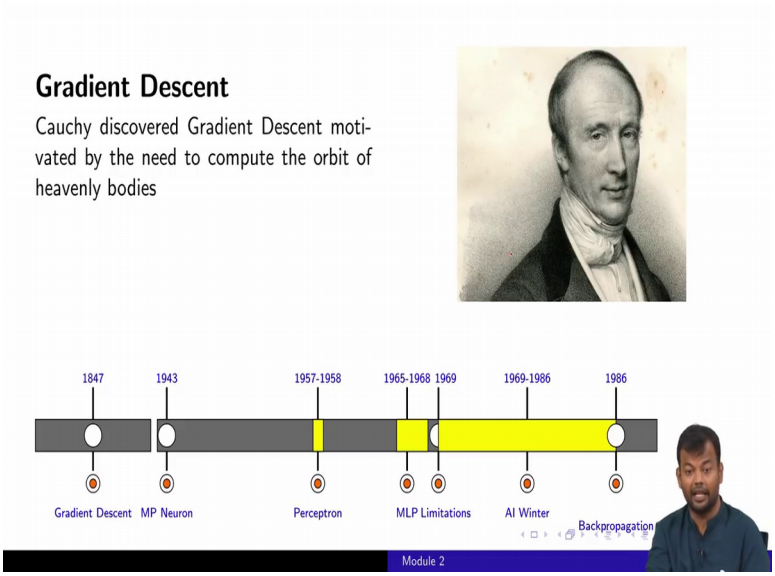


Then around 1986 actually came this algorithm which is known as back propagation. Again this is an algorithm which we are going to cover in a lot of detail in the course in the 4th or 5th lecture. And this algorithm actually enables to train a deep neural network right. So, deep network of neurons is something that you can train using this algorithm.

Now, this algorithm was actually popularized by at Rumelhart and others in 1986, but it is not completely discovered by them, this was also around in various other fields. So, it was there in I think in systems analysis or something like that. It was being used for other purposes in a different context and so on. And Rumelhart other and others in 1986 were the first to kind of popularize it in the context of deep neural networks.

And this was a very important discovery because, even today all the neural network, so most of them are trained using back propagation right. And of course, there have been several other advances, but the core remains the same that you use back propagation to train a deep neural network right. So, something this was discovered almost 30 years back is still primarily used for training deep neural networks. That is why, this was a very important paper or breakthrough at that time.

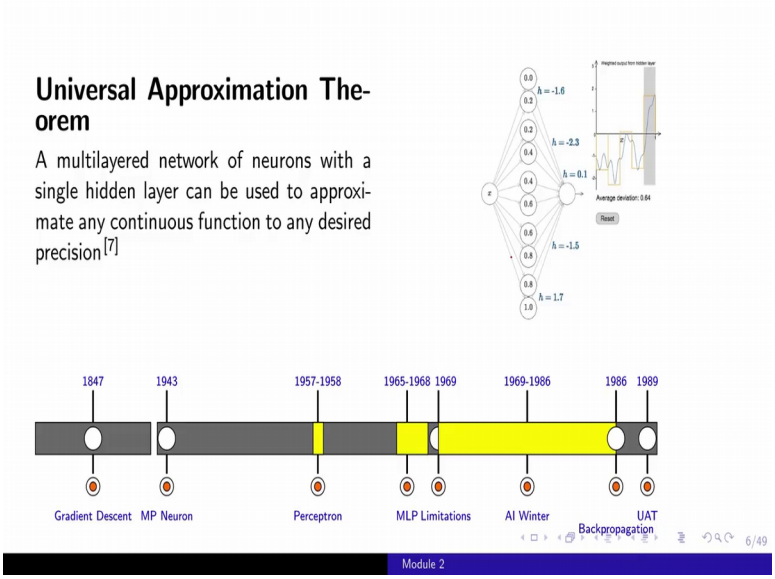
(Refer Slide Time: 09:58)



And around the same time so, again interestingly, so back propagation, is used in conjunction with something known as gradient descent. Which was again discovered way back in 1847 by Cauchy and he was interested in using this to compute the orbit of heavenly bodies rights.

That is something that people care about at that time today of course, we use it for various other purposes one of them being discovering cats and videos or even for medical imaging or for describing. Whether, certain have of cancer is being depicted in a X ray or things like that there is a lot of other purposes for which, deep neural networks and hands. And hence, back propagation gradient descent and other things are being used for it, but again these are not very modern discoveries these are dated way back 30 years and even gradient descent is almost 150 years and so on right. So, that is what I wanted to emphasize.

(Refer Slide Time: 10:43)



And around the same time in 1990 or 1989 there is this another interesting theorem which was proved which is known as the universal approximation theorem. And this is again something that we will cover in the course in the third lecture or something like where we will talk about the power of a deep neural network right.

So, again the importance of this or why this theorem was important will become clear later and when we cover it in detail. But for now it is important to understand that, what this theorem said is that if you have a deep neural network you could basically model all types of functions continuous functions to any desired precision.

So, what it means in very layman terms is that. If the way you make decisions using a bunch of inputs is a very, very complex function of the input. Then you can have a neural

network, which will be able to learn this function right in many laymen terms, that is what it means.

And if I have to type it up a bit or I have to say it in a very enthused and excited manner. I would say that basically it says that, deed neural networks can be used for solving all kinds of machine learning problems. And that is roughly what it says, but with a pinch of salt and a lot of caveats, but that is what it means at least in the context of this course.

So, this is all around 1989 and despite this happening right some important discoveries towards the late end of 80's, which was back propagation universal approximation theorem. People were still not being able to use deep neural networks for really solving large practical problems right. And a few challenges there was of course the compute power at that time was not at a level where, it could support deep neural networks.

We do not have enough data for training deep neural networks and also in terms of techniques. While back propagation is a sound technique, it is to fail when you have really deep neural network. So, when people try it training a very deep neural network they found that the training does not really converge the system does not really learn anything and so on. And there were certain issues with using back propagation of the shelf at that time because of which it was not very successful right.

So, again despite these slight boom around 86 to 90 where some important discoveries were made. And even follow up in 92 93 and so on. There is still not a real big hype around deep neural networks or artificial neural networks and at time again a slump a slow winter right up till 2006.