[Music]

Welcome to this first lecture of the sixth week in on the course machine learning and this lecture will divide it in two parts, but the whole week will be focused on the area of artificial neural networks and their role in machine learning. So you heard a little amount neural networks earlier in the course but I will repeat a few important things before we go into the more detailed descriptions of this approach. So regarding the inspiration for this approach, the main inspiration is the human or even animal nervous systems, this means the nervous systems of the body but maybe from the side of inspiration particularly the brain, so the current model in neuroscience is that nervous activity is based on very small atoms called neurons, and these neurons they work in a electrochemical way and they transmit signals to each other and the terminology is pretty straightforward there is a cell body, that's the center of a new world there are some input channels they're called dendrites there is an output channel it's called an axon, and the process where this unit sends a signal is called firing. There is also another word use that is pretty important to know and it was called synapse actually it's synapse is the points where the axons sending out the signals connect to the dendrites of other, so it is the connection points between the communication elements so to say. One very important point here is that in the animal system or human system neurons are not homogeneous, there are a lot of different kinds of neurons in the body, so they are tailored to specific purposes but if you abstract of course you can say that more or less they work in the way I just described.

So furthermore all these atomic elements, the neurons are connected in huge network. So to give you the order of complexity here you can say there are 100 billion of neurons in the human brain, and 10,000 times as many connections, so actually you can see the million or billion neural connections in the brain and you can see if you have an analogy with parameters in another system you can say that the property of that connection so you have so many parameters that you adjust in order to define the functionality of the system. While the real thing as I described in the earlier slides is a very complex heterogeneous system. We now turn to computer science again and define what we call the computational model for artificial neural networks which is of course inspired by the real but very much simpler and very much abstract more very much more homogeneous, so an artificial neural network or ANN is a network of nodes or units that we call artificial neurons and these are connected by edges so the corresponding graph is directed typically there is one layer of input nodes, on layer output nodes and an arbitrary number of layers in between and we call those layers in it, this is not the total truth because there may be some systems where you have edges that are by bi-directional and where there is no big distinction between input & output, but those are exceptions. So edge is typically have a weight that can be adjusted and this way it increases or decreases strength of a signal coming via that connection. Actually the weights are very important because when we took we're going to talk about learning one can say that the functionality of the system are very much decided by the setting of the weights and therefore by changing the weights can somehow learn. So the output of each neuron computed by some typically nonlinear function of the sum of the inputs given that that sum exceeds a threshold. Potentially all the neurons can fire in parallel so it is not if potentially it is a totally parallel system but many times there are some temporal constraint put on this which means that there are sequential set of events going on. Many times in real applications the layers are given certain functions so this means that we don't have a network that is totally homogeneous that a lot of things can happen anywhere is the design of the network is such that in certain layers certain things happened in other layers such as other things happen but that's not very much after the detailed engineering and solving a particular problem. So also typically not always the signal travels from the input level to the output which is natural but it could be so that there are loops which means that signals are reprocessed. However if we run a data item through the system we get some output and then of course we have a data set and we run all the ideas so the whole process of running all our data items or data sets through the particular network we have that is normally called in the alien terminology for an Epoque.

The ANN model is an abstract model but hopefully you already understood that there can be very variants on these networks, so the purpose of this slide is just to exemplify to you and give you some of the important keywords to get a flavor of that there are many different networks and during this week we will look at a not all but a representative subset of these networks and I will try to explain the purpose behind the different designs. Let's for a moment look at a single neuron, and look at what actually happens when the neuron fires. So we have a set of inputs x1 to xn, to this particularly Nueron and for each of those connections input connections there is a weight, and then we have a threshold and the role of the threshold is that you can say that there is an amount of signals coming in which is calculated as the weighted sum of the of the inputs, but the neuron will only fire if that the total amount of signals coming in is exceeding this this threshold. Finally even if the neuron fires is just not an normally a binary output, but the output is actually then a function of this sum of signals and there are made of different choices for that function and which also is an important characteristics of a particular network. So we want to learn eventually we will look at that later how to adjust the weights thereby changing the functionality of network and thereby learning. However it's not only the weight that's important the threshold also have some importance for the functionality, so normally when you some kind of trick to homogenize the system so in a way you try to get it very objectively and rid of the threshold and anyway convert it to a kind of nominal weight. So you can see on this slide in the end one can take away the threshold, introduce a zero input always having the signal one and then you can set the weight of that additional zero input to minus the threshold that case we call the value of that weight for a bias. But this is just a technical transformation to also allow that we can learn the bias or threshold through the same learning mechanism as a useful for the weights.

On this line you can just look at an example of how to use the mechanism on the last slide. So let's look at the network, look at the neuron with a certain set of inputs 5 5 3 then we can under set of weights that we predefined and then we can convey the weighted sum and we get an a result which is 7 then we have a threshold of 3 so we subtract and the outcome then is 4, which is about zero so that's fine and then before we send out the signal we also have to apply a function and we make a choice of function and the choice we are made in this case you can see to the right so you can see it as I do some look up there so you look at the value 4 and then you get the output 1 in that particular case. And what you see at the bottom is a circle it is the same thing but there you can also see how the threshold is converted into this extra input layer. There are many kinds of total activation functions or transfer function this is the function to transform the final output from the neuron and we will come back to this during during the week. My comment on this point is that if you have a nonlinear problem and then you have some problems solving model before happily at nonlinear a very naive common sense fact is that you have to include nonlinear elements in your model in order to be able to treat a nonlinear problem, so therefore and this goes of course also for neural networks so this means that if you want neural network to be able to handle nonlinear problems it's wise to include some elements in the model that is nonlinear and actually the choice of this activation function is one of the key points where you can see to that this happen. So if you only have linear activation function you may have a problem but if you have a nonlinear activation function and you could design your system so it also can potentially handle nonlinear problems.

As always for a particular approach it's important to understand how do you actually solve problems for this approach, so in the neural network case what you have to do is you have to model your problem and then you have to map that problem model on to typically the neurons in the input level and when you have done that then you start the computation and then you get some result in the output layer and but then of course you also have to see to that the form of solution you desire have to be possible to extract from the output layer. I mean the simplest case and in the usual case is that the input is a feature vector so you describe the object or situation you want to analyze in terms of a feature vector and then in the simplest case then you could simply map your feature vector onto neurons in the input layer. However there are two cases that we will come to back to during the week that the input elements are not as simple as it could be so that there are complex sequences in space or time and that demands some special treatment, but it could also be that the objects have a totally different form of representation, so every input in image and can also be like some speech profile that situation we'll also need special considerations, and actually these two cases is complex sequences and non-symbolic input items like images are the two important special cases that have to be treated.

This slide is intended to give you a picture of how the new artificial neural network area developed initially I call is here the childhood of the area, and essentially you can see that there is a history here for 40 years from the 40s actually up to the mid 80s and let's start with the end so actually the end point here is that two things happen in the mid 80s that the term deep learning was roster we will hear a lot about deep learning but actually it was coined as a time in 1986 by one researcher in machine learning but it was not used about for artificial neural and is it was used for some other kind of symbolic machine learning. At the same time more or less some researchers well-known researchers like Rumelhart and colleagues published a key article on how you can actually use artificial intelligence for solving a practical problem. And that article was kind of the starting point for the strong growth of this area as a key technique within machine learning. However hero ironically they didn't call it deep learning at the time they called it artificial neural network or connection in learning and so in 1986. Ok so let's see what led up to this so essentially the one started in the 40s and I would name two people who contribute it originally so McCullouh and Pitt, they actually do use the first model of this kind and if you really look at that model it has many of the characteristics of what people have done ever since, so it's actually a very key initial contribution. However more or less in parallel other people like Donald Hebb presented complementary theories for this, which is also has influenced we want also see it come back to that. And then actually like in the next decade there were some experiments people try to build actually some simple new machines, welcome the perceptron which is actually the first real implementation and then in the sixties there were some very important results in the neural science field on our vision systems and you will see in a later lecture this week that that work from the sixties has really inspired the approach that now more or less are taken in in this area to handle the learning based on images. So this is the initial story as for many areas it goes very slow in the beginning it took 40 years from the first ideas to something concrete could be demonstrated, but of course when it taken off it goes much fast.

Let's have a look at a few of these early works. So the really important work by a McCullouh Pitts for the first time they try to demonstrate on how a neural like unit can perform logical operations. So actually what they did that you just looked at a single neuron and they devised the first attempt for modeling that with a very simple model when your binary output you have no function that transforms that output you have inputs, positive inputs and what the negative inhibitory inputs and inputs all the same weights and of course as a signal says everyone it is pretty simple, also they have this rule that if you get an input one only inhibitory input, 0 only an inhibitory input it will have a veto power in the situation. So it's not actually identical to later approaches but it shows the basic architecture and what was also observed in their work which is kind of interesting that they very already in their original article observed that there are certain logical operations but like XOR that cannot be solved by a single unit can be solved by a system of unit but it cannot be solved by a single unit of this this kind. Another important early work was that by Donald Hebb and we will have treated separately during this week but I would say now is just a few words about the basic idea that happened so essentially the idea of Hebb was the following that if you assume that the that the neurons fire or act in parallel it is a fact that if two neurons fire at the same time and they are connected and that parallel firing of the two connected neural would normally strengthen the connection between, which means that the learning that should take place is to really increase the weight of that connection one is one and the other is zero and then one should typically lower the strength of that connection in in the learning process so actually then Hebbs devise the model a formal model for that kind of basic idea. So I can say it's another way of philosophy for updating ways than in some of the other systems. Finally a word about a kind of interesting system that was built in 1956 actually the same year as the area of artificial intelligence was coined, and it's built by one of the pioneers in artificial intelligence Marvin Minsky. So what actually he and his group tried to do is not a device a software system but actually build a machine, so they built a machine of 40 synapses which as you may remember then is the connection point between the output of one neuron an input of other. And as you can see on this image this is a picture of one of the synthesis so actually there was a physical machine consisting of 40 of these things and that was complicated and electrical mechanical connections among them, so as far as I know this is the first attempt to really physically implement the ideas of artificial intelligence. So this was the end of this part so we will continue in part two