[Music]

Welcome to the ninth lecture of the six week of the machine learning course. This lecture will be about deep learning and it's further developments. So the question I want to pose here is the following. Deep learning is a very hot topic at the moment if you open any paper, look at TV and so on, it's a high probability that somebody mentioned this term. So what is it and if I try to tell you what it is will it be a climax or an anticlimax for you. The are answer there I would say is that result wise it's a climax because what it's normally referred to as deep learning today has produced more success stories than other parts of machine learning so far. However from the purpose of this course and rhetorically it's an anti-climax because we have this week already talked about almost everything that is of relevance for the concept of deep learning in my opinion. So for this final lecture on deep learning is not much more time because the constituents we have already covered. Even if I just told you that there is not much more to be said about deep learning than what we already covered this week, I like to give you a few facts about the timeline of artificial neural networks and deep learning for that sake. Some of these facts have already been mentioned but I have now collected them on one slide to give you my picture of how things as developed. So let's start from the beginning, so in 1959 Arthur Samuel coined the term Machine Learning, so since the beginning of artificial intelligence machine learning has developed hand in hand in other sub areas of artificial learning, it's been a very much integrated Development. Also very early that were pioneering work on artificial neural networks as you have heard this week, for example the perceptron. But artificial neuron network has always formed a kind of separate stream in artificial intelligence so not always it's been a tight integration of the symbolic approaches in the artificial neural network. And this dream has been weaker or stronger depending on the time, it was very strong in the beginning like people weren't very optimistic but then as you see from the timeline here it was a long break from early 1960s to 1986 it's almost like it's like in the order of 25 years where very little happen for various reasons. But in the mid nineteen eighties things happen so on one hand there were very important work by Rumelhart and Hinton when they reintroduce multi-layer feed-forward networks and learning of those networks through back propagation. And also not much later Rumelhart took the initiative to the precursors of recursive neural networks such then developed hand-in-hand with artificial neural networks in general and also not much later from in the late 80s there was this the introduction of convolution neural networks in particular for image processing. So in the later half of the 80s the key ingredients of deep learning what is now called deep learning was developed. A little ironically the owner at the same time a researcher in machine learning Rina Dechter in one article introduced the term deep learning for multi-layered symbolic machine not in the context of artificial neural network. So at this point people like Rumelhart rather introduced the term connectionist learning for his improvements of artificial neural Networks, and published a well-known book on with that title. So actually it took considerable time until in in around 2000 the term deep learning started to be used for this combination of ANN, RNN and CNN by one researcher Igor and his colleagues. So around 2000 this terminology in a way what was adopted by the ANN community of course it's a good terminology because if you have multiple layer networks and you have many layers is natural to call them deep, and as you have understood this week especially recurrent neural networks as the effect that the number of layers grow and that is also of course the case in convolutional neural network of approaches. So all these things tend to make the number of layers larger but then there is another development that is very important because of course you may in theory create very powerful systems by creating systems with many layers which was well done through the combination in ANN, RNN and CNN. Okay but those complex networks also demands a lot of computation power and that computational power was essentially not there at that point in the beginning. So then in the first decade of the 2000s people started to experiment with using specialized hardware GPUs and we will come back to the use of these things in the coming week, so the application of those Hardware made it possible to run this more huge networks in an efficient way. So this development happened in the first detailed twenties made this combination of ANN, RNN and CNN that we know from now termed deep learning more useful. And an example of the usefulness happen actually in 2012 where one such system actually a deep learning system in the sense of combining ANN, CNN & RNN made a great breakthrough in this Imagenet challenge that I described earlier this week with the system called Alexnet. So that was one of the first real success stories and of course the term deep learning has been adopted in the decade before so with a breakthrough also the use of this term spread and actually today 2019 the term deep learning is a dominating term for the work of artificial neural networks in machine learning so this is my story to you. To convince you that the story I told you on the last slide is not just my invention, I will now will point on a really good survey article and I given you the link it's open access so you can click on this link and you can find this article. And it's a review article from nature volume 521 in 2015. This review article is are in the region by three of the key researchers so Geoffrey Hinton you can call him I would say the still active key person in artificial neural network as such and also more specifically in what's now called deep learning. Yoshua Benglo who has really continued to work of Hinton and developed it further. Yann LeCun who is a key person in the development of convolutional neural networks. So this is a kind of what you find in this our article is I would say the words from the horse month it's not what anybody else imagined or suggested it's the key person of this area have wrote this. Of course the key parts are not always objective totally because these were the persons coined the term deep learning so of course they want to promote it but they are very serious researchers, they are very instrumental in developing the field so you should listen to anybody you should listen to them actually. So what you find on this slide is actually the quotation from the summary of that article where they say the following so they say that deep learning allows computational models that are composed of multiple processing layers which is rather trivial statement start with but with the important property to learn representation of data with multiple levels of abstraction. So and this should be interpreted that these systems are able to themselves build up the representations not only use precomposed representations. They also say that these methods have been dramatically improved the state of the art in speech recognition, visual object recognition which are maybe the two very important tests errors, doesn't mean that these technology cannot be use the others it's as if these two have been the focal point of the development of this area. But today there are other things that are also important drug discovery in genomics and so on which are very important. So what deep learning does it discovers integral structure in large datasets by using back propagation algorithm to indicate how a machine should change these internal parameters that are used to compute representation in each layer from representation in the previous. So this little paragraph and in their summary that really refers to the key machinery of feed-forward multiple layer neural networks with learning through back propagation so this is a key point here in what they think is deep learning. And then there is a final paragraph that articulates the two other components, some deep convolutional nets have brought about breakthroughs in processing images videos speech and audio, with recurrent networks have some light on sequential data such as text and speech, which were the two other components in the lectures this week. So as a summary to the to the right you can see my own little graph on what constitutes deep learning, is actually the combination of the basic machinery in feed forward multiple layer artificial neural networks ANN with backpropagation with the developments in convolutional neural networks and recurrent neural networks. So these are the triples of elements that constitute this area. Summarizing might attempt to characterize deep learning, I give you my preferred terminology. So computer science is a broad field and it has a subfield called artificial intelligence and artificial intelligence in turn as a subfield called machine learning. As part of machine learning we have this combination of ANN, RNN and CNN that we took some efforts to talk about this week and this combination is what is should be deducted as deep learning I also included there's a little bigger framework relating to machine learning so actually machine learning is today considered part of this umbrella concept data science which actually also cover other areas of computer science outside artificial intelligence like Big Data and actually also other areas outside computers science and of course centrally statistics, so this is the big picture I want to convey to you. Finally a few words then about the recent development in this area of deep learning so one important stream of research going on and development going on at the moment is going further on what was highlighted in this summary of the deep learning article I refer to, is the learning of structures and features in a presentation so the automated learning of structures and features and this is like on the top of the traditional categories of machine learning, it's supervised learning unsupervised learning, and feature learning, you can all do that without learning any new structures because you can predefine the structural features but the ambition here is the learn structures and features themselves. So this seems to be still a very important endeavor in deep learning not in all necessary in all other parts of a machine learning, but there is this inherent problem with ANN is that it's at the bottom sub symbolic which means that if you create a very complex artificial neural network system and let it learn over a long time you build lot of structures and so far is not being trivial to understand what is build up, you establish some knowledge in this system and you can use the system for solving problems and you can do that efficiently. However it's not trivial to say actually what has been learned what is the new knowledge created in the system that enables better performance. So and of course people have realized that we should useless in large scales, it's also important that not more creating efficient systems will better performance but also being able to in a symbolic form extract what has been learned in a way that we as human can read it and understand it. So there are a lot of work I would say at the moment on that particular issue, so one of these new terms coming up is Disentangle representations which is not the whole story but part of these attempts. The other things going on because I hope you have noted that in my characterization of deep learning associative memory didn't play a big role. So I talked about this combination of a ANN, RNN and CNN which are the key deep learning still associative memory techniques is an important stronger (17:13) within the neural networks as such but obviously traditionally not the Box been (17:16) driven development of deep learning. But today there are a number of attempts to actually integrate these areas, so integration of associative memory approaches with this mainstream over ANN, RNN and CNN techniques is an issue key issue and there are examples of that and something called Deep Belief network as an example. Also there are some work to connect the ANN, RNN and CNN technology to the learning of Bayesian system Bayesian network related systems. There is also this terminology that we haven't really talked so much about there is this distinction in machine learning and related fields between generative and discriminative classification, so in discriminative classification we start from the purely inductive approach so we look at the examples and we'll try to infer which abstractions or classes connect to the various examples we have while in a generative approach we actually starts from our hypothesis or classes from that partly I would say through deductions come up with probabilities for which kind of examples would belong to the classes and of course it had to be validated by looking at the examples but it's more of a top-down approach and actually now there are more focus at the moment I would say on developing this generative approaches then the discriminative ones which were the classical. So also we can also see an enhanced use of the ANN, RNN and CNN a bunch of techniques for the purpose of reinforcement learning. So I hope you it's been clear that all work in artificial neural network is kind of orthogonal to these genres in machine learning supervised, unsupervised reinforced, I mean this can be used for any of these, so it's not so that ANN is particularly useful in one of the song (20:01)can be used for depends on what you want and there are more work going on now in using this technique for reinforcement learning. Another trend at the moment is scaling up of applications in time critical and safety critical applications like self-driving vehicles. I mean self-flying vehicles or similar things is a very hot technological area at the moment, machine learning systems are needed in those contexts but putting a machine learning system into such contexts expose you to issues of elective time criticality and safety criticality. And it's not evident that of all the solutions of machine learning are already to cope with those challenges. Also we can see here a shift towards really being an able to handle really Big Data, I mean traditionally machine learning started with small data sets I think for a long time they've been developing managing larger data set but still today we can see we have huge data sets, we are continuously generation of data from many sources and if machine learning or deep learning should be able to be useful it also have to extend its ability to handle these larger data sets. Also of course there is a continuous work on utilizing special hardware this is they're just here to stay. A self-driving car has a special GPU on board that can be used for machine learning. So things go on here and a new hardware architectures are trying and this we will see for a long time be a development area. Finally there is also a consolidation and enhanced open access to toolboxes and software support systems. For some time now I mean the support for an engineer that want to really apply deep learning as it has increased a lot but there are still improvements to be made and this is also an active. So this is my summary what's going on in this area. So by this I will conclude this lecture and I will conclude this week on our artificial neural networks. So the only thing remaining is the last lecture which has always be a tutorial on the assignments for this week so thank you very much.