[Music]

Welcome to the eighth week of the machine learning course the final week it will be dedicated primarily to preparations for example but also to gives examples of applications to further illustrate what can be done with techniques from this area. And so this is the first lecture and we will here look at back at some assignment related tasks and my purpose is to try to give you a picture of what your exam will look like. The general outline of the upcoming exam will be as follows, there will be in the order of 50 tasks, these tasks can give you hundred marks in total, on the actual exam all tasks will be given in multiple choice type form, so this is what it looked like so a disclaimer for now for today for this lecture is I'm going to exemplify tasks here but not using a multiple choice type, I do this because it's more compact than the multiple choice time it's pretty awkward when I'm just want to simplify which kind of tasks you can wait for. And so there will be as for the assignments like a balance between recall type tasks which are will only be worth one mark and but then there also will be problem solving tasks which are worth more typically two to three marks, there will be a few more challenging tasks four and five marks, and what I will focus on very much today is how these tasks will be spread over the course, so the idea is that the exam should cover most part of the course, so I actually try to model the course now into 21 thematic categories and covering in the content and it's not necessarily so that an exam all these categories will be represented it could be so there are more questions or tasks on one such and something or it could be so that there are no question no tasks at all and so that will be up to the next exam when you see it. Tasks will be well aligned that's the ambition to the type you have seen at the assignments they can be slightly different in style but aim is not to move too far from the tasks you have seen so far. Also I want to ensure you that the tasks will be tightly coupled to the lectures and of course the recommended readings you should consider as useful but if the recommended readings are more to enhance your understandings all matters discussed on the lectures, so you do not have to fear tasks on topics only mention in the recommended materials even though the recommended material can illuminate or explain better or deeper something mentioned lectures. So that's the general outline and let's turn now to my proposed thematic categories, as you understand this kind of pretty complex course can be structured in many ways, so there is no actually the right or wrong way of structuring your course, it's up to your taste more or less. So here on this slide you can see it is 21 categories. At the at the top left you can see very general topics like elements from basic mathematics and statistics elements from logic and computer science elements from theory improving an artificial intelligence, I mean these are not the primary topics for the course but they will more or less always somehow come into play in the various corners of the rest of the course, so therefore they are here and therefore there will be a few elements on the exam that relates to that, but that is it's not a major part. So after that the course continues and it goes into the realm of learning scenarios, content about how categorization works, a few different approaches to inductive learning maybe as slightly side topics and some material on Bayesian networks and genetic algorithm they are part of the picture but they are not super central as you may have recognized for the course, then we come to instance based learning which is an important topic with many connections and that's follow up by cluster analysis as an example of supervised learning from that the course moves into learning scenarios where one can say that in one form or another there are few available domain knowledge that can be used as a strong basis for the learning process and that is more or less ended with reinforcement learning which is a natural starting point for moving into the realm of neural networks, and as you can see here there are and we also I guess you already realized from the material of week six, there is a strong focus on neural networks and the main reason with that is that is one of the sub areas of machine learning today there is a lot of limelight on and what you can the next slide is just to illustrate how the flow of the course has moved more or less but exactly it's a linearization actually I ambition has the course has developed more or less following the path just orally described.

What I will do now is simply have a pretty fast walk through these subtopics or I would say content categories, there is absolutely no ambition here to present anything new so if there happened to be a little detail that is new it's more of an error rather than something that was aimed for, I'm merely tried to collect on each of these slides some keywords that would give you the right to lace and look when you want to recapitulate what was articulated and focused on in the earlier lectures in the course. So the first slide here is on mathematics and statistics and as I hope you understood vectors matrices and tensors in general a lot of mathematical sub topics one can say related to some type of race is important and operations of those they will occur over and over again within the course, also some geometrical concepts relevant for visualizing state spaces as some optimization techniques occurring in particularly in relation to the enforcement learning graphs occur pretty often, so the very basis of graph theory, properties of graphs are important and also to some extent simple aspects of calculus as integration and differentiation. So these are the key mathematic topics needed, some statistics is also touched primarily inferential statistics where we draw conclusions about the population from sample I mean the core is actually the mathematical base is the probability theory, particular with stochastic processes, Bayesian methods and Monte carlo methods. So this is the picture for math, so very simple questions or recall and also very simple questions problem-solving character can occur for these topics.

Let's move on to logic and computer science. As you understand what we will only scratch the surface here but also elements from logic and computer science have occurred in the course and central for learning is the relation to the classical inference styles like Deduction, Induction Abduction and so on and so that's the key part and with respect to computer science which is an extremely wide area of course we mostly related to some basic knowledge about algorithms and data structures and to some extent the databases in the week 7 we also talked about supporting tools, so of course we also have touched programming languages to the distributed computing API and programming libraries and Technical computing systems. Then going on directly I think to artificial intelligence and Theory improving and there are a lot of couplings between theory improving and artificial intelligence but and we have talked about that in various corners of the course, we also talked about knowledge representation schemes that's necessary we can talk about learning and we talk about knowledge representation and one particular scheme that was highlighted was logic programming. Also such techniques occur in various corners of the course and so now we are moving more into what a machine learning, so as you remember very early in the course that was again some general lectures about the different types of learning scenarios whether we look at supervised learning with pre-sorted example or unsupervised learning or the very specific situation of reinforcement learning where the behavior of the system is evaluated and graded by some external environment, we also discussed distinction like batch learning contra online learning, batch learning being roughly that you collected all the examples all the instances makes the learning process based of them, in principle you have all the instances at the same time while online learning works more dynamically in a situation where new instances come up over time. Of course important is the relation between classification and regression that also occurred systematically during the course, the relation between symbolic and some symbolic learning different to situations where we want to adapt the system more of a Data analysis scenario where we don't really have a system in a narrow sense that's the behavior which is changing and then a number of separates more specific concepts always important when we discussed the various algorithmic overfitting, underfitting, linear sparability etc. Also in the same part of the course we talked about categorization and categorization also being close to semantic networks for the reason that many times in artificial intelligence categories have been represented in terms of synaptic networks. So what we did here was to look at basic principle for object oriented descriptions, the description of objects in terms of features and the building up structures among categories and some interdisciplinary basis for how we do categorize a world and finally a little more detailed information about the format of the so called Semantic networks the relations model in terms of edges in the network.

Okay leaving that we come I would say two more to the core of the machine learning course, the kind of classical core algorithms for doing induction more or less without a pre-existing domain theory. So here we talk about the difference kind of language that's needed to represent the instances and the hypotheses, how to define relations in the hypothesis space, typically in the form of mathematically or partial ordering also distinguishing between different approaches to do this either data driven starting from the examples or by generate and test which is it essentially means the opposite which we start top down building up and test hypotheses and see how well they fit. So data driven, bottom up, generate and test strategies. Then the focus of this was more on comparisons of the areas data-driven search strategies, like depth first, breadth first and version space.

Let's move on after that we looked on decision trees and how to build up decision trees, both for classification and regression, and the method is essentially top-down so we talked about various properties of the relation between the instances and the built up trees such as Purity/ Homogeneity we spent some time on various information theoretic measures that can help us to guide us, in how we build up the trees and doing some calculations of the simple calculations on those. We also looked at this particular algorithm which I would say is the mother algorithm in this round called ID3, some simple examples of using that. Then we also touch but not so much in detail how we can handle various situations that can occur like underfitting and overfitting and that's basically overfitting how to handle that by pruning of the tree. Finally we looked at alternative approaches, how to be a laugh parallel reasoning (19:05) in parallel trees and then used the best tree or the average result from using all the trees together, one example is random forest etc.

And so after decision trees we went into Bayesian networks and I would say what we did in Bayesian networks was not so merely to look at the learning of Bayesian networks even if this is an area itself but rather the basic properties of this networks and at the core there we have the Bayes theorem which helps us to infer causes from evidences and we also had an assignments in simple calculations so based on the use of Bayes theorem and also we looked at very very small examples where we built this very small Bayesian network for particular examples and at the key of those exercises is the built off of the conditional probability tables coupled to the network nodes. So this is the focus here.

Let's move to the next so genetic algorithms, it's an important area but there are been many things covered on this course and not everything can be treated equally thoroughly, so of course my judgment is we had a pretty shallow descriptions of genetic algorithms focusing on the basic ideas the way you referred need to represent things with the terminology inspired by genetics and also looked at a typical algorithm scheme including various faces such selection, the best gene, evaluation of the generation of new generated creation of new generations and so on. And finally and there we also had some exercises coupled to that very small exercises related to the key operations in particular the crossover relation which is the main operation for reproduction and finally we also shortly described how one can implement rule-based systems in terms of genetics and thereby by doing that one get a learning property of the rule-based system, oh you have a rule-based system you can apply it you can look how it works, you accumulate evidence did you use that evidence for evaluating the fitness of the rules in the sense how much did the rule contributed to a good outcome of the problem solving and when it then you essentially create a new generation of a rule base where some of the rules are taken away some of the rules are reproduced changed and you get a new generation and then you can apply them again, and also strange enough I think it which I also highlighted as a parallel, I mean when we come to neural networks later it's important to have a mechanism also to back propagate the results or return given from the outside world and essentially grade the performance of the various components neural network case it's the performance of the various weights in the system but in this case with a classifier system implemented as a genetic algorithm it's essentially becomes a grading of the specific rules but one need is similar back propagating algorithm as neural network case and in here it's called the Bucket Brigade algorithm. So nothing is new under this one.

After that we turn to instance-based learning which is a lot of content in that section, so where we discussed instance based learning in general, the structure of the instance space, we focus on the K nearest neighbor algorithm, we look at distance and similarity matrices, we also looked at the so called weighted nearest neighbor algorithm, we did them discussed binary linear classifiers, we also go more went more in-depth on to support vector machines and finally we discussed the situation where we could buy smart mappings from a two dimensional space to high dimensional space typically maybe three, could enable the use of support vector machines also for nonlinear cases. And as a natural extensions of that we went into cluster analysis I think a lot of the focus there what are looking at the large variety of different kinds of clustering techniques and I think I would say we mostly know Partitioning based clustering, K-means clustering is the main approach, we looked at hierarchical based clustering and also density based clustering. There was some focus there on we had some focus on the distance similarity matrices which are more or less the same topic as for instance based learning and actually we didn't do much calculation in this realm there have been some examples on the assignments on how to calculate proximity matrices and that should be distinguished from the similarity matrix, a similarity matrix is concerns the distance between the instances while the proximity matrix concerns the distance quotation mark between the clusters.

After that we went into the realm of learning in the context of prior knowledge and there were two parts initially, that are more closely related to two to the existence of theories in symbolic form. So one is inductive logic programming and what we did here was to discuss all some logic programming fundamentals but also the basically the two main ways of doing learning in the total in a inductive logic programming it's one based on specialization, the other based on generalizations, and we showed some examples on these two variants.

After that we went into explanation based learning which is essentially the situation where you have an almost complete domain knowledge or domain theory that you want to kind of complete or finalize and also explanation based learning in many cases doesn't mean that totally new knowledge is created, but rather that the available knowledge is compiled in more efficient forms.

The next big block is Reinforcement learning, we talked about the terminology, we talked about the concept of value function, we talked about the most common way of modeling reinforcement learning situations which is the Markov decision process model, and then we talked to various ways of attacking the solution to such problems, we talked about dynamic problem of programming, we talked about the Carlo simulations and finally time difference models with Q-learning as our main examples, and of course we also discussed some more principle distinction between various approaches, here passive versus active, on policy versus off-policy and so on. So this became I would say pretty substantial part so even if it's just one sub theme, here I would say yeah the real comment is that's not always of total balance between these parts, some are much more context dense than the others, if we look at this what I call here subtopics.

And before we went into neural network we had a pretty short lecture on Case Based Reasoning. Obviously this topic which is an area of itself has not been so well covered in the course either but let's move now to, I was able the most heavy the subtopic set of sub topic is related to neural networks, so first there was this lectures based focused on the single atomic parts, the neurons and of course the precursor to the current neural model called the perceptron. And of course important at the core here are the weight updating scheme as for as we will see four most of the artificial neural network practice, and there was also some talk about various kind of activation functions, that can be used and the properties of those.

And after that comes to say the classical core mythology part of artificial neural networks which is the multi-layer networks where the neurons are the part and the methodology for feeding forward signals through the network for estimating the error at the output level and then for back propagation of the outcome of that error estimation as a basis of the back propagated error constituents updating weights. And a natural follow-up is then an extension of the basic ANN architectures which are the recurrent neural networks which is essentially a way to mostly most cases artificially creating a possibility to handle States by introducing these special units that can be unfolded to handle the multiple states of a specific item in the model. So we discussed the kind of standard version of that we call it Vanilla recurrent neural network but of course Illustrated how it can be extended with putting these elements in multiple layers and also how there can be communication back and forth in the same layer, and so where we really did what looked more in detail of this works more on the vanilla version and the unfolding on that. We also talked very shortly about short and long-term memory which is a very successful variant of this successor in the sense that it's being used with a win great success in certain applications areas.

And then more or less wait for a little while we moved from the mainstream ANN into the realm of associative memory, of course also still modeling associative memory by an artificial neural network and the focus there was mostly on the approach by Donald Hebb as it inspired by cognitive models of the brain.

We moved on to further realizations of associative memory two versions of that one called the Hopfield networks and another called the Boltzmann machine. So we discussed the Hopfield networks, we discussed the energy concepts that it is used to guide us in the search for optimal solutions and also we in the Boltzmann machine more in particular on the process called Annealing inspired from metallurgy by essentially heat up that's an analogy of course heat up the system and then slowly cool it down with a hope that is heating and cooling sure in some cases get her out of undesirable local maxima and bring us to the global minima. So then actually the end the core content of the course was focused on convolutional neural network, as you may remember convolutional neural language primarily targeting applications in image recognition, so we both look a little about the source of inspiration for the basic operation here convolution also local at the interdisciplinary inspiration sources for this whole architecture which is actually the special organization of the convolution system. So we looked at various terminologies and various items occurring in a convolution neural network architecture and we showed very very simple examples how these central operations could be done primarily convolution and Pooling. So that's it we reach the end of the sub-themes.

So that was all the subthemes. So what I've done now and I this I will not go through in detail actually it's here for your service, is here to just give you examples of the kind of questions that can come up on your exam relating it's very important to now I want to stress that this is just inspirational examples to make you focus easier when you now go back and rehearse based on the other material you have from the earlier lectures, so what you find actually here is a list of questions that I hopefully makes you can guide your preparation studies in the right way, there are many questions that can be put obviously with each little category, also the other disclaimer is that for space sake here I didn't want to give this question in multiple choice look but I want to stress that on your exam all your tasks will be formulated in multiple choice form, they will not be open questions like here but as you understand every open question can be re-engineered into a multiple choice for which happened on your exam. So please have a look and use this source of information so it may be so that some of these questions have occurred in the assignment, it may be something that day but mostly I don't think they have but it could be so not that these are questions examples are exclusive. So this were the simple recall questions one mark and then you got a similar list of example questions with the same disclaimers for the more problems or your own solving oriented one which give a couple of marks each depending actual number may depend on the judged complexity of the small task. And finally I also gave you some example spread out across the sub themes of the little more complex tasks, that you will have not many not so many as this, so I mean if you want to give the have the picture of the balance between tasks go back to the first slide I gave you, so don't believe you we get ten hard questions this will not happen we will get a handful or fewer hard question majority of the question will be mark one or two three questions.

Okay so good luck with that I hope this lecture has been useful for you and it can guide your preparation for the exam, so thanks for your attention we will follow up with another lecture, however I want to state very clearly that this lecture application and demo example is only there for yourselves it's an absolutely no bearing on the exam. So I advise you now to focus on the preparation for exam so and hopefully what I've given you today is useful, this very final lecture it's just to give you some illustrations of the usefulness of the area and as I said absolutely no bearing on your exam thank you so much.