Machine Learning,ML

Prof. Carl Gustaf Jansson
Prof. Henrik Boström
Prof. Fredrik Kilander
Department of Computer Science and Engineering
KTH Royal Institute of Technology, Sweden

Lecture 6
Characterization of Learning Problems

Welcome to the second week of the machine learning course. The theme for this week is characterization of learning problems. As you know this course of many weeks and we will gradually become more concrete regarding various algorithms and the application of those algorithm into a specific problem. However this week has the purpose to give you a general overview on machine learning and its varieties, also various situations and scenarios for use of machine learning and crucial issues to consider in those situations.

So we will start discuss one major distinction regarding the role of machine learning. So a lot of this course will be focused on called data analysis, which means analyzing large data sets to create distractions these abstractions can then be used for various purposes. The second context though is what is termed here adaptive systems and as you know, we will are surrounded by more and more of autonomous systems of various kinds that duest services and in general these autonomous systems in order to be useful need to be adaptive. Of course also here machine learning as a growing role. The distinction between data analysis as such and use of machine learning in adaptive systems it's still a useful distinction it's still very relevant for sorting out and understanding various approaches to machine learning but of course the borderline between the two shiners tend to become increasingly blurred ten-twenty years ago, it was a clear demarcation line between the areas but it not so today. Of course in general for data analysis the major reason for all learning is to ultimately act better in future situations. So let's turn at first to say something about machine learning in the context of adaptive systems, so there is one major paradigm in machine learning called reinforcement learning you may have heard about it and I'm sure you will hear about it more examples of the use of reinforcement learning are in the context of game playing but also in the context of robots that means how to adapt robots so at the movements of the robot get more optimum. So there's a pretty abstract model of reinforcement learning in the sense that it's viewed as an agent that I've seen in an environment and the agents perform a series of actions and the environment as it is supposed to react and give feedback based on those actions and in this case we term this kind of feedback the reward and of course in the long run the idea is that the series of actions and the reward given from the environment to the agents is supposed to iteratively improve the functionality of the agent and in the long run create some kind of optimal performance. Reward of course it is a general word it can be negative can be positive. So we can say reward can be credit or can be blamed and of course in any concrete application of this abstract model it is assumed both that the environment in question is able to provide reward in a concrete form and that the agent of course is also able to internalize this concrete reward and adapt the internal behavior of the system typically this kind of system has a number of internal parameters and what happens when the reward is considered is that these parameters are modified but it could be in general any change of the internal structure of the agent that can take place in this context. So to continue a little about reinforcement learning as I just said it's absolutely important that the agent can concisely manage the handling of the reward and as a consequence update the internal structure of the agent itself. It is further assumed that the typical scenario for reinforcement learning is a strong theory based hardware and software system so we are not talking about here a system where we learn in the absence of a theory we rather learn in presence of a very strong theory so just conceived. A robot normally, a robot is a very advanced design and if we used reinforcement learning just to optimize that the movement of the arm, we can say that we actually this kind of robot system learn on the margin of its already existing behavior. We will look into different kinds or ways of normalizing this later in the course like already now can say that one very commonly used way of modeling the environment in this kind of case is, using what's called the markov decision process and also then in turn a very general applied mathematical technique called dynamic programming, so these are very typical tools in the toolbox of engineer in this respect, but as you will see later reinforcement learning can be realized in many different shapes. So now I want to turn to data analysis, and the first thing I want to do is to discuss with you a little about what I call the end to end process for application of machine learning to real-world problems. So as you see on the slide there is long list of steps to consider and rather far down the line in this list you find the core analysis face. So many things I assume when we think about machine learning in its application we talked about the core analysis face when we have an algorithm and we apply that algorithm to a dataset and then we get some output from that process but as you see from the whole list this is absolutely not the only saying. There are so many other steps needed in order to solve real problems and of course the first thing is to harvest the data and for realistic application is absolutely not certain what the data are, what form it has, it can be dating in many forms from very many sources, and potentially then there is a lot of work to harvest that from all these sources and bring it together and then of course because the forms very different then need to be some pre-processing of this data, I mean a typical case in which the input and the data is in image form or like just sound or something else, then that has to be transformed into a digital form that can be reasonably used for a machine learning algorithm and then of course one also will consider whether we are learning in the absence of a theory more or less we have a lot of data and we really don't have a theory or the contrast is that we have a very strong theory so what as I said earlier enforcement learning we learn on the margin on prior knowledge. Exactly where we put the border between machine learning and some other engineering skills, so little unclear but I assume when we come to feature engineering it's clearly with the realm of machine learning and feature engineering is of course the skills that we will come back to that is needed to express the data items or data set in a form that is manageable for the algorithm. So again since we've given with the devised and fabricated the right kind of data set, we have to choose an algorithm that is suitable for the situation but it's not normally not just to choose an algorithm because an algorithm is not a very seldom a black box, an algorithm can be adapted and should be adapted probably in order to give a good performance and so typically whenever to adjust the out briefs when talking about hyper-parameters of algorithm, language bias, complexity, management issues. Okay, so when then if all of this is done then the core analysis can take case, but when that face happened normally there are more things to do because it's not certainly so that the output from the application of the other create data in the form that you require so that needed some thought processing and finally of course in order for the result to be used either in a system directly, online updating and you need to prepare the form of the output the other case is of course if you want them use it for decision-making you need to prepare a material that are useful for decision-makers.

So what we will do now is to slowly convert on the key topic of this week which is classification so I can say that two main scenarios in data analysis one important area which we actually will not spend so much time on is was in statistics is called a regression. An regression is essentially to use the analysis of all states to establish a means of prognosis for future states and in contrast to that classification is to look at descriptions of objects and abstract from those objects trying to define concepts or classes so that the definition of those concepts can serve as a basis for classifying new objects in the future. So regression is about prognosis for states and classification is about the base for being able to classify objects not seen so far. So saying if you work more about regression being as the technique from statistics used to predict values of a target quantity when the target quantity is typically continuous and you see to the right an example how it could look like a very simple example which is a linear dependency between two numerical variables and I mean such variable there are like millions of examples of that situation so the numerical values could be the length of person as a function of the person's age, it can be the price of a property as a function of a point in time and so on, but as you understand I mean this is just the simplest kind of example many cases you state can be something more complex it can be a bundle of variables can be some structured set of valuables and regression of course makes sense as a concept also for those more complex cases. So from now on we only focus mostly on classification or as we will synonymously call it concept learning and we will have four lectures with more this week three lectures on the keys used and the fourth lecture on the tutorial for the assignments. So the three lectures of the first will be on objects categories and features, primarily sort out the terminologies, yes because it's a pretty complicated situation and part of the difficulty of reading about this area is there are so many terms around they're often synonymous terms. The second lecture we will talk about features and the rollover feature engineering and finally in lecture 2.4 we will turn again to what is here called scenarios for learning where we will talk about a number of important distinctions such as supervised contra unsupervised, learning online versus offline, instance base versus abstraction based learning and so on. But before we leave this lecture when I'm going to introduce a simple example that will be referred to both in the lecture series but maybe in the in forth coming so of course over week we will introduce the few examples that will be the base for various discussions. As you may remember from the first week something that exists at the moment is a growing set of repositories for datasets of different kinds many of them can be opened access, some are more restricted and then and so on. So what I have done here is, I tried to select the reasonably limited them dataset from on a research repository in the UCI ML repository and essentially the name of the dataset is a zoo dataset. I mean if you look at the whole repository TAS 351 dataset some datasets with a huge number of data. So my selection is just was very practical here I need something simple example to use as a basis for some discussions and so this is a very naïve and partial classification of animals and there are 107 objects or data items and they are each sector at characterized by 18 features, 18 is not so small but not so large either and essentially these hundred and thousand objects are actually categorized in pre-categorized labeled in seven categories so that that's the example we will look at. So for this data set the zoo data set and we have a class structure or category structure which is essentially focusing on the classification of animals on a certain level so actually this whole data set starts from single examples of specific kinds of animals and the task is to classify those in terms of seven categories which are all animals and shows and categories are mammal, bird, reptile, fish, amphibians, insects and invertebrates. That's how the data sets it up so it means that every single data item in the data set is labeled with one of these seven items. So next we come to the features in this data set, so as you will soon see every single data item is characterized by one of these, all of these features and most of these features are Boolean, so they are either 0 1 so an animal as hair or it has not hair, it has feathers or not feathers, it lays eggs or not, so they are Boolean. And the only exceptions are the only I would say predictive feature, that is not a boolean is legs because legs is an integer and tells how many legs that animal has. So yeah we will discuss the engineering of features and in one of the later lectures, but you can see here of course that one issue that is important with the choice of features is that the feature set should be rich enough to cover all the seven categories of animals that we are interested in classifying. If we choose too few features or choose the features with some bias, it may be is that the feature set would be perfect to classify birds and useless to classify mammals and so on, of course if we one could want to have hundreds and hundreds of features it's not a problem because we could bring everything but as you will see later it's often desirable to keep the feature set down and then one have an issue of choosing the appropriate features. There are two kinds of features here that are a little special because they characterize or classify the right item and the important one for the task is class type which is then always one of these seven types so all that it is them are pre classified and given a class time. But then of course you also have something a called an animal name and that's essentially saying that this animal is an ape, this is a horse, this is a dog and so on and actually this data set only have one instance, only one data item with the same name so now are not several examples of dogs not several examples of cats it's a single well-defined characterization of each of these kinds of animals on that conceptual level. So here you'll finally see the whole data set, all the 107 or our objects it's a kind of handy that we can one can show them all data set on one slide as I said earlier there is only one example of each animal almost a kind of basic level, so it's one dog, one crab, one penguin and as on duplicates of animals on this level. And so the data set as is it given focus on a task where the analysis is from basic animal upwards towards animal and as we will discuss a little more later of course every domain has many levels I mean in zoology a lot of people have spent years and years of finding the appropriate classification from the very lowest level to the highest level and it's not always relevant to discuss things on all levels normally you won't choose a level which is relevant for the problem at hand and this is just an example of such a choice. So this is the end of this first lecture and thanks so much for your attention I hope you start to get a feeling for the whole area now and we will go into some more details and what we will do on the next lecture we will look little more about the terminologies here in particular the terminologies for objects, categories and features to sort out what is what.