Objective of the Project

The first goal of the project is to develop machine learning models with uncertain knowledge extended by mathematical structures and combined with theoretical models in the form of differential equations. The second goal is to develop methods of incorporating prior knowledge to support vector machines (SVM) based on proposed models.
Specifically, we consider classification and regression problems.

Basic Research

The planned work is divided into two parts, designing machine learning models and methods with incorporated mathematical structures, and designing methods for combining differential equations with machine learning methods. The proposed approach for the first part of incorporating mathematical structures consists of three sequential activities:
1. Axiomatization of classification and regression problems for the knowledge sets model with incorporated basic mathematical structures.
2. Derive theoretical justification for the models defined in the previous step.
3. Develop novel methods and SVM extensions based on defined models and theory.

The proposed approach for the second part of combining differential equations with machine learning methods consists of two sequential activities:
1. Incorporating initial and boundary conditions for differential equations to SVM.
2. Develop methods for incorporating knowledge about solutions of systems of differential equations to SVM.

Reasons for Choosing the Research Topic

There are two main approaches to modeling the world. The first one is to create strictly theoretical models based on for example differential equations. This approach has been successfully used in many domains, like physics, economy. The second approach is to use general methods like for example machine learning methods for finding models based on empirical data. The latter has become popular recently for the reason of availability of more and more data however without the additional knowledge about the structure of the problem. We think that some tasks could be better modeled by the fusion of both approaches then apart. The idea of this project is to use general methods, like SVM and to incorporate additional knowledge extracted from the data about mathematical structure of the problem and about a theoretical model in the form of differential equations. We can consider such modeling in terms of human intelligence. We use two types of systems of thinking System 1 and System 2 according to Nobel Prize laureate Daniel Kahneman [Kahneman2011]. The first one is intuition, the second one is rational thinking. The general methods mimic the System 1, while the additional knowledge and theoretical models mimic the System 2. Intuition thinking is less accurate, but faster. Combining it with strict rules from the System 2 could improve the performance of solving some tasks. General methods learn from limited number of examples and are often based on statistical assumptions about unseen examples, which are generated independently from the same unknown probability distribution. For this reason, the statistically motivated general methods work better with more examples. The idea of this project is to improve performance of the general methods by instead of increasing the number of examples which could be impossible, adding some prior knowledge about mathematical structure of the problem and about theoretical models in the form of differential equations. The long term goal of artificial intelligence field is to create an artificial intelligent system. It is not possible to build artificial intelligence system with only statistical methods. We need also strict rules [Kisielewicz2011]. This is the reason of incorporating additional knowledge to statistical models. The next idea is to combine statistical models with theoretical models which are usually differential equations. Such combination of models could potentially improve the performance of any of the systems alone. We can consider such combinations in two ways: 1. statistical models have additional knowledge about structure in the form of differential equations 2. we improve accuracy of a model based on differential equations by empirical data represented by a statistical model.

Consider the comparison of a knowledge sets model to statistical models. Statistical models based on probability theory are a very successful approach to modeling the world. But the question arises if the methods like SVM can be derived from more general models not assuming that data are independent and identically distributed (i.i.d.). There are some research findings about possibility of weakening the assumptions of statistical machine learning models like i.i.d. data. While in machine learning the concept of using some kind of generalized probability theory is mostly absent, in other fields there are some signs that the probability theory with axioms defined by Kolmogorov is not enough to describe the world and we need more generalized theory. For this reason, in cosmology, in order to model black holes the free probability is used which is based on non-commutative von-Neumann algebra [Heller2014]. The idea of the free probability is becoming more popular recently. The theory is applied also to describing random matrices. It is more focused on a mathematical structure that is the algebra, with less focus on randomness. All general theorems of probability theory, like the law of large numbers have been redesigned. This is an early sign, that we need generalizations of probability theory to describe the world. Another reason is purely speculative. The probability theory has been axiomatized by Kolmogorov in 1933, almost 100 years ago and is rather fully exploited. The more general question is how to replace uncertainty in the form of randomly generated samples according to some unknown distribution with another type of uncertainty in machine learning models. We did the first step by defining a knowledge sets model with different types of uncertainty, like prediction bands in [Orchel2015].

The key concept in statistical learning theory is the Vapnik--Chervonenkis (VC) dimension. The VC dimension of a class C of hypotheses is the largest number N such that some set of N points is shattered by rules in C [Kulkarni2011]. For example, the class of linear hypotheses in d-space has a VC dimension of d+1. The similar concept is the Popper dimension describing difficulty of falsifiability of a class and is defined as the largest number N such that every set of N points is shattered by rules in C. For the class of linear hypotheses in d-space, the Popper dimension is equal to 2, because three colinear points cannot be shattered. In this project, we focus on limiting the possible point configurations by using prior knowledge about points. For example, if we add knowledge about configurations with p > d points that the line segments between points with the same classification do not cross, then the VC dimension is p. The next question regards how to choose the best hypothesis for the set of points. Different general rules have been invented such as Occam's razor, minimum description length principle, structural risk minimization (SRM). We generally take into account best fit to the given data, and we favor hypotheses with lower complexity.

References

[1] D. Kahneman, Thinking, Fast and Slow. Farrar, Straus and Giroux, 2011.
[2] A. Kisielewicz, Sztuczna inteligencja i logika. WNT, 2011.
[3] M. Heller, Granice nauki. Copernicus Center Press, 2014.
[4] M. Orchel, “Solving classification problems by knowledge sets,” Neurocomputing, vol. 149, pp. 1109–1124, 2015.
[5] S. Kulkarni and G. Harman, An Elementary Introduction to Statistical Learning Theory, ser. Wiley Series in Probability and Statistics. Wiley, 2011.