Course syllabus «Statistical Learning Theory»

Approved by
Programme Academic Council
Protocol Nr. 05 from 27.08.2018

Author	Dr. Vladimir Batagelj
Number of credits	4
Contact hours	48
Self-study hours	104
Course	1,2
Educational format	Without use of online course

I. Goals and Results of Mastering the Discipline; Prerequisites

Statistical learning is the field that sets the framework for machine learning drawn from statistics and functional analysis. The goal of statistical learning theory is to study, in a statistical framework, the properties of learning algorithms. This study serves a two-fold purpose. On one hand it provides strong guarantees for existing algorithms, and on the other hand suggests new algorithmic approaches that are potentially more powerful.

In this course we will go in detail into the theory and methods of statistical learning, and in particular complexity regularization (i.e., how do you choose the complexity of your model when you have to learn it from data). This issue is at the heart of the most successful and popular machine learning algorithms today, and it is critical for their success. In this course you'll learn some of the techniques needed for the analysis and near-optimal tuning of such algorithms. This is an elective course, offered to MASNA students, and examples used in class may differ depending on students' interests.

As a result of taking the course, students should:

Know:

- the basic concepts from statistical learning theory.
- theoretical foundation of why some machine learning algorithms are successful in a large range of applications, with special emphasis on statistics.
- several paradigms in statistical learning theory to select models (Structural risk minimization, Maximal likelihood, Minimal Description Length, etc.)
- the link between cryptography and computational limitations of statistical learning

Be able to:

- Calculate sizes of training sets for several machine learning tasks in the context of PAC-learning (and hence calculate VC-dimensions)
- develop and/or foster critical reviewing skills of published empirical research using applied statistical methods.
- to criticize constructively and determine existing issues with applied linear models in published work

Have:

- a training of mathematical skills such as abstract thinking, formal thinking and problem solving;
- in-depth understanding of boosting algorithms and a few other algorithms for machine learning.
- theoretical understanding of several online learning algorithms and learning with expert advice.

Basic knowledge of introductory statistics are required for this course. The basics of this discipline should be used in all other program related courses.

The course is strongly related and complementary to other compulsory courses provided in the first year (e.g. Applied Linear Models II, Contemporary Data Analysis) and sets a crucial prerequisite for later courses and research projects as well as for the master thesis. The course gives students an important foundation to develop and conduct their own research as well as to evaluate research of others.

II. Content of the Course

Please note: due to their complexity, some of the sessions will run over multiple class periods.

SESSION ONE: Probably approximately correct learning

Several examples of simple learning algorithms are studied such as for learning of axis aligned rectangles. The general problem is formulated. Instance classes that are not learnable are discussed.

SESSION TWO: VC-dimensions

We define VC-dimensions and prove the fundamental theorem of statistical learning theory: if the VCdimension of the set of classifiers is finite, then we can learn a correct classifier with any precision using finitely many samples. Moreover, the theorem gives us upperbound for the samples in terms of the VCdimension and the parameters of our precision measure.

SESSION THREE: Structural risk minimization

We consider a class of problems that is learnable in the sense above, and relax our criteria for learning to what is called nonuniform learnability. We provide a characterization of classifier spaces that are learnable in this sense and again this theorem gives us sample bounds. We discuss a special framework called Minimum Description Length principle and discuss some applications.

SESSION FOUR: The time complexity of learning and cryptography

A typical cryptographic encoding of a string is an example of a an object with a clear structure, but of which no learning algorithm can find the structure of the object in a reasonable time. In this part we discuss some learning tasks that are impossible for computational reasons. Under a plausible computational complexity assumption (which is required for secure RSA encryption) one can show that neural networks of small depth and regular languages can not be learned.

SESSION FIVE: Boosting

A weak learning algorithm can generate from a train set a model that is slightly better than random guessing. We observe that are weak learnable coincide with the classes that are PAC-learnable and present some efficient algorithms to transform weak learners to the stronger PAC-learning algorithms. We study AdaBoost and DeepBoost, two algorithms that are currently very popular in machine learning. Furthermore we prove performance guarantees and given an alternative explanation of their success using game theory.

SESSION SIX: Online learning

In online learning there is initially no training set. After each prediction, the class of the label is declared and the learning algorithm can use this information to imporve the prediction model. We study 1) prediction with expert advice, 2) linear classification algorithm such as the perceptron algorithm and 3) the connection between online learning and game theory.

SESSION SEVEN: Probabilistic formulations of prediction problems

Plug-in estimators, empirical risk minimization; linear threshold functions, perceptron algorithm, risk bounds (concertation inequalities, uniform convergence; convex surrogate losses for classification).

SESSION EIGHT: Game-theoretic formulations of prediction problems

Minimax strategies for log loss, linear loss, and quadratic loss, universal portfolios and online convex optimization.

SESSION NINE: Neural networks

Stochastic gradient methods, combinatorial dimensions and Rademacher averages, hardness results for learning, efficient learning algorithms.

III. Grading

Course grade will be completed as follows:

Course Element	**% Towards Final Grade**
Final Exam	**50%**
Final In-Class or Take-home exam (at the discretion of the instructor)	*50%*
Participation and responsibility grade	**50%**
Homework Assignments (5 x Varied points)	*20%*
In-Class Labs (9-10 x Varied points)	*20%*
Quizzes (Best 9 of 10, Varied points)	*10%*
Total	100%

If the final grade is non-integer, it is rounded according to algebraic rules. If has a half (.5) at the end, we are rounding upward. Rounding of cumulative grades and other rounding issues are performed according to the HSE rules.

IV. Grading Tools

This class contains several assignments that test student knowledge and understanding throughout the course.

Quizzes

You cannot meaningfully participate in the seminar if you have missed my lecture and did not do any reading. Therefore, to encourage you to prepare for seminars, every seminar will have a quiz on the lecture material and all assigned readings for the week. This includes the very first seminar, which will focus on Lecture 1 material. You are allowed to miss any one quiz (skip a seminar, not prepare, etc.) – in other words, I will count the best 9 out of 10 quizzes that we will have. If you submit all ten, I will count best nine. All quizzes will be done online and submitted to me via SurveyMonkey (links will be given in class).

Important: I record IP addresses and only accept quizzes submitted from with the HSE IP address. Quizzes submitted from other locations are NOT counted towards your grade. In other words, to participate in a quiz, you have to be present in class.

In-class Labs

There will be a lab assignment in almost every seminar, depending on our progress. Since we will be learning SAS, and learning quickly, you will need to devote a substantial time to it. Seminar labs should help you with this task. At the end of the lab, you will submit your completed assignment for the day (or as much as you were able to complete) to me via LMS.

Homework assignments

There will be several homework assignments that will provide additional hands-on practice for the concepts we've learned in class and practiced during the seminar. Homeworks will be assigned as needed throughout the semester. All homework submissions must be done by the stated deadline via the LMS system.

V. Resources

5.1 Main Literature

1. Kulkarni, Sanjeev, and Gilbert Harman. An Elementary Introduction to Statistical Learning Theory, John Wiley & Sons, Incorporated, 2011. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=697570.
2. Harman, Gilbert, and Sanjeev Kulkarni. Reliable Reasoning : Induction and Statistical Learning Theory, edited by Tom Roeper, MIT Press, 2007. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=3338667.
3. Alpaydin, Ethem. Introduction to Machine Learning, MIT Press, 2014. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=3339851.
4. Mohri, Mehryar, et al. Foundations of Machine Learning, MIT Press, 2012. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=3339482.
5. Murphy, Kevin P.. Machine Learning : A Probabilistic Perspective, MIT Press, 2012. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=3339490.

5.2 Additional Literature

1. Schapire, Robert E., and Yoav Freund. Adaptive Computation and Machine Learning : Boosting - Foundations and Algorithms, MIT Press, 2012. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=3339451.
2. Lantz, Brett. Machine Learning with R, Packt Publishing Ltd, 2013. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=1343653.
3. Puneet Mathur. Machine Learning Applications Using Python. Cases Studies from Healthcare, Retail, and Finance. URL https://link.springer.com/book/10.1007/978-1-4842-3787-8. Springer Link
4. Ramasubramanian, Karthik, and Abhishek Singh. *Machine Learning Using R*. Apress, 2017. URL https://link.springer.com/book/10.1007/978-1-4842-2334-5. Springer Link.
5. Haroon, Danish. "Python Machine Learning Case Studies: Five Case Studies for the Data Scientist." (2017). URL https://link.springer.com/book/10.1007/978-1-4842-2823-4. Springer Link.
6. Sarkar, Dipanjan, Raghav Bali, and Tushar Sharma. "Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems." (2017). URL https://link.springer.com/book/10.1007/978-1-4842-3207-1. Springer Link.

5.3 Software

№ п/п	Name	Access conditions
1.	MicrosoftWindows 7 Professional RUS MicrosoftWindows 10 MicrosoftWindows 8.1 Professional RUS	*From the university's internal network (contract)*
2.	Microsoft Office Professional Plus 2010	*From the university's internal network (contract)*
3.	R, R studio	*Open access. URL: https://www.r-project.org/*

5.3 Material and technical support

Classrooms for lectures on the discipline provide for the use and demonstration of thematic illustrations corresponding to the program of the discipline, consisting of:

- PC with Internet access (operating system, office software, antivirus software);
- multimedia projector with remote control.