Foundations of Machine Learning and AI

INSEAD PhD Course
P4 (March-April)

Instructors: Nicolas Vayatis, Theodoros Evgeniou

Course Outline

AI and Machine Learning have become central topics of discussion in the popular press after being developed for over 50 years in Academia – by computer scientists and, in more recent years, by mathematicians and statisticians. These fields are expected to have a major impact in potentially every aspect of research as well as business: from basic science fields such as life sciences, to Decision Sciences, Finance, but also areas like Sociology, Economics, and other Social Sciences.

However, while one can be a "reasonable" user of some popular machine learning and AI methods, gaining an edge in terms of innovation in research and practice but also taking full advantage of the capabilities offered by these technologies requires a more fundamental understanding of the principles behind these booming fields.

The goal of this course is to:

- Provide the foundations of Machine Learning and AI, so that students can better understand these methods, use them, and potentially develop their own custom based ones that can also use to advance their respective fields;
- Provide an overview of some of the most important machine learning methods used in research and practice;
- Provide students not only with a historical perspective of these fields, but also with a view of the state-of-the-art methodologies and research advances as well as views on future directions;
- Help students use machine learning methods appropriately in their research fields, with the aim of developing insights that are only feasible due to the usage of these new "microscopes".

The course will be run as a combination of lectures, discussions of important papers, exercises, coding (in R or Python), and a class project. Participants are required to have knowledge of the core Probability and Statistics (I and II) courses.

Grading

20% Class Participation and Paper Presentation
30% Exercises: two exercise sets, combining mathematical and hands-on application exercises
50% Class Project: "*Develop Your Own Machine Learning Method and Share the Code on Github*". Details to be provided in class.

Recommended Books

While we will not follow any specific book, the following books are some of the "classics" in the field. We will also use a few chapters from them.

- V. N. Vapnik, Statistical Learning Theory, Wiley, 1998.
- L. Devroye, L. Gyorfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, Springer, 1996.
- T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, 2nd Ed., Springer, 2009.

Course Sessions

Sessions 1-2: Introduction and Set Up: AI and the Machine Learning Problem

In this session we will first provide a brief history of AI and Machine Learning, and outline the fundamental problems these fields aim to solve. We will then shift to the theoretical foundations of Machine Learning and provide an overview of the field, of some popular machine learning methods, of application of Machine Learning and AI, as well as a summary of this course.

Main concepts: Symbolic AI, Connectionism, Statistical Learning, Approximation Theory, Bias-Variance, Empirical Risk Minimization, Hypothesis Spaces, Loss Functions, Generalization Error, Learnability, Consistency Properties.

Background Readings:

- Chapter 0 ("*Introduction*") and Section 3.10 ("*Kant's Problem of Demarcation and Popper's Theory of Non-Falsifiability*") of V. N. Vapnik, Statistical Learning Theory, Wiley, 1998
- Poggio, T. and F. Girosi, Regularization Algorithms for Learning that are Equivalent to Multilayer Networks, *Science*, 247, 978-982, 1990.
- Chapter 2 of L. Devroye, L. Gyorfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, Springer, 1996.
- Nature Insights, Machine Intelligence, *Nature*, Vol. 521 No. 7553, pp. 435-482, 2015 (a collection for reference to skim through)

- D. Donoho, High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality, Stanford University, 2000

Sessions 3-4: From Classical Statistics to Machine Learning

In this session we will develop and analyze some of the most common machine learning methods that are also the closest to classical statistical/econometric methods. We will also discuss about relations between Machine Learning and other important fields such as optimization theory, regularization theory for ill-posed problems, and signal processing.

Main concepts: Regularization theory, Ridge Regression, Lasso, Support Vector Machines, Kernels, Sparsity, Model Selection, Cross-Validation, Matrix Completion, Recommender Systems.

Background Readings:

- "The Learning Problem and Regularization", Lecture Notes, MIT course 9.520 on Statistical Learning Theory and Applications
- T. Evgeniou, M. Pontil and T. Poggio, Regularization networks and support vector machines, Advances in Computational Mathematics, 2000.
- Sections 1.7 and 1.8 of V. N. Vapnik. Statistical Learning Theory. Wiley, 1998.
- R Packages: ElasticNet, glmnet

Sessions 5-6: Theoretical Foundations of Machine Learning

In this session we will introduce the main mathematical tools and intuitions that can help us better understand why and when machine learning methods work. We will also discuss some of the main theorems that explain the predictive performance of machine learning methods. It is these theorems, together with advances in computing power, storage, and availability of (big) data, which led to the recent important breakthroughs of AI and Machine Learning in all scientific and business areas.

Main concepts: Concentration Inequalities, Complexity Measures, Learning Rates and bounds, VC-dimension, Structural Risk Minimization, Stability, Rademacher Complexity, Estimation and Generalization/Prediction Error, Approximation Theory.

Background Readings:

- T. Poggio, R. Rifkin, S. Mukherjee, and P. Niyogi, General conditions for predictivity in learning theory, *Nature*, Vol. 428, 419-422, 2004.
- F. Cucker and S. Smale, On the mathematical foundations of learning, Bulletin of the American Mathematical Society, 2002.
- S. Bucheron, O. Bousquet, G. Lugosi, Theory of Classification: A Survey of Some Recent

Advances, ESAIM, Probability and Statistics, 2005

Sessions 7-8: Ensemble Methods and Other Algorithms

In this session we will discuss some well known approaches to combining machine learning methods. Combinations of methods, much like combinations of diverse expert opinions, is known to improve the accuracy of models/groups. We will discuss some theoretical underpinnings of ensemble methods as well as some further machine learning methods such as Classification and Regression Trees, Random Forests, Bagging and Boosting, and Neural Networks. We will also start exploring machine learning software packages.

Main concepts: Bagging, Boosting, Random Forests, Boosted Trees, Neural Networks.

Background Readings:

- Chapters 9.2, 10.1-10.9 and glance through the remaining of Chapter 10 of T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, 2nd Ed., Springer, 2009.
- Hibon, M., T. Evgeniou, To Combine or Not to Combine: Selecting among Forecasts and their Combinations, International Journal of Forecasting, 2005
- R Packages: randomForest, rpart

Sessions 9-10: Deep Learning and Recent Mysteries in AI

In this session we will discuss some of the most common Deep Learning methods, and also touch upon some current open problems in Machine Learning and AI. A more general framework of machine learning and AI will also be discussed, and some recent applications of these tools will be presented.

Main concepts: Perceptron, Feed-forward Neural Networks, Convolutional Neural Networks, Stochastic Gradient Descent, Back-propagation, Hierarchical Learning, Feature Learning.

Background Readings:

- I. Goodfellow, Y. Bengio and A. Courville, Deep Learning book, MIT Press, 2016. Glance through the book for a general idea.
- H. Mhaskar, Q. Liao, T. Poggio, Learning Functions: When Is Deep Better Than Shallow, 2016 (Skim through)
- L. Bottou, Stochastic Gradient Descent Tricks, Neural Networks: Tricks of the Trade p. 421-436, 2012

Sessions 11-12: Data Representations, Feature Learning, and Applications

In this session we will revisit the problem of machine learning, this time from the point of view of finding good data (“world”) representations. We will revisit and discuss topics like sparse representations, kernels, and learning data representations using deep learning methods. We will then discuss a number of applications of machine learning, ranging from text mining to time series prediction and analysis of network and graph data.

Main concepts: Sparsity, Variable Selection, Feature Learning, Kernels, Sparse PCA, Low Rank Representations, Dictionary Learning, Text Mining, Time Series, Network Data.

Background Readings:

- Chapters 1 and 2.1-2.2 of T. Hastie, R. Tibshirani and M. Wainright, Statistical Learning with Sparsity: The Lasso and Generalizations, 2016
- B. Olshausen and D. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature* volume 381, pages 607–609 (13 June 1996)
- H. Zhou, T. Hastie, and R. Tibshirani, Sparse Principal Component Analysis,

Sessions 13-14: Other Topics and Paper Presentations

In this session participants will present a number of papers that will be selected during the course. We will also discuss other topics not covered in this course. More online resources will be shared during the course. Participants are also expected to contribute some of these resources on the course website throughout the course.

Example concepts: Deep Reinforcement Learning, Fairness in AI, Independent Component Analysis, Generative Adversarial Networks, Compressed Sensing, Random Matrix Theory, Wavelets, High Dimensional Statistics, Information Theory, Compression, Gaussian Processes, Graphical Models, Approximation Theory, Splines, Reproducing Kernel Hilbert Spaces, Bootstrap, Clustering, Matrix Estimation, Matrix Completion, Low Rank, Active Learning, Experimental Design, Change Point Detection, Natural Language Processing, Text Mining, etc.

Example papers:

- Argyriou, A., T. Evgeniou, M. Pontil, Convex Multi-task Feature Learning, Machine Learning 73 (3), 243-272, 2008 (will be discussed)
- J. R. Hauser, O. Toubia, T. Evgeniou, R. Befurt, D. Dzyabura, Disjunctions of conjunctions, cognitive simplicity, and consideration sets, Journal of Marketing Research, Vol. 47, No. 3, pp. 485-496, June 2010 (will be discussed)
- O. Toubia, E. Johnson, T. Evgeniou, P. Delquie, Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters, Management Science, March 2013 (will be discussed)
- Clémençon, S. , G. Lugosi, N. Vayatis, Ranking and Empirical Minimization of U-statistics, The Annals of Statistics 36 (2), 844-874, 2008 (will be discussed)
- J. Li, P. Rusmevichientong, D. Simester, J. N. Tsitsiklis, S. I. Zoumpoulis, The value of field experiments, Management Science, Vol. 61(7), pp. 1722-1740, 2015
- Hardt, N., E. Price, N. Srebro, Equality of Opportunity in Supervised Learning, NIPS 2016
- Neural Information Processing Systems (NIPS), Conference
- Knowledge Discovery and Data Mining (KDD), Conference
- Fairness, Accountability, and Transparency in Machine Leaning (FAT/ML), Conference