Macadamia:
Master's Programme in Machine Learning and Data Mining

Tapani Raiko FIRSTNAME.LASTNAME@TKK.FI
Kai Puolamäki
Juha Karhunen
Jaakko Hollmén
Antti Honkela
Samuel Kaski
Heikki Mannila
Erkki Oja
Olli Simula
Helsinki University of Technology, Department of Information and Computer Science,
P.O. Box 5400, FI-02015 TKK, Espoo, Finland
http://www.cis.hut.fi/macadamia/

Keywords: master's programme, machine learning, data mining, course curriculum

Abstract

Macadamia is a two-year master's programme for machine learning and data mining given in the Department of Information and Computer Science at Helsinki University of Technology. This paper describes its curriculum and how the courses are organized. The emphasis is on our three machine learning courses.

1. Introduction

The Laboratory of Computer and Information Science founded by prof. Kohonen in 1965 at the Helsinki University of Technology has a long history for researching and teaching neural networks and pattern recognition. The main research topics have been self-organising maps (Kohonen, 1995), data mining (Hand et al., 2001), and independent component analysis (Hyvärinen et al., 2001). The research focus has since shifted from neural networks to machine learning, algorithmic data analysis and data mining. Many of our master's level courses were already given in English or their materials were in English during the last years. So it was a natural step further to install an international masters' programme for ensuring that our department gets a sufficient amount of qualified master's students. In 2007, the first international group of students started at our new Macadamia master's programme for machine learning and data mining (Oja et al., 2007).

Macadamia is a very focused master's programme concentrating on a well-defined topic. It is based on the strong existing curriculum in the department for Finnish students. Macadamia also gives a strong background for postgraduate studies in the area. It is an interesting question how focused the studies should be at this level.

Our laboratory has historically attracted students from several fields within our university, including computer science, electrical engineering, and engineering physics and mathematics. To continue this tradition, our aim has been to make Macadamia accessible to students with all these different backgrounds as well.

2. Our Course Assortment

The list of obligatory and relevant courses is given in Table 1. Each course is lectured once per year and the students have the freedom to take them when they wish.

Normal courses last half a year with two hours of lectures and two hours of exercise sessions per week. Attendance to these is not obligatory. The course is completed by taking an examination and submitting a project assignment. The project assignments typically involve programming and analysing a given data set.

Special courses are given as seminars with a varying topic each semester. The organisational details of these courses vary but typically the organiser of the seminar gives an introductory lecture and has a list of topics for presentations. Each student gives one or two seminar presentations and possibly selects a problem for each topic. The special course is completed by attending a certain percentage of sessions, giving the required amount of presentations, and by submitting written solutions to problems. On some courses, there can also be one or several project works. Usually there is no examination.

The text books used in the obligatory courses are as follows. The two courses on machine learning are based on (Alpaydin, 2004) and (Bishop, 2006), machine learning and neural networks is based on (Ham & Kostanic, 2001) and other material. The data mining course uses (Mannila & Toivonen, 1998; Hand et al., 2001). Information visualisation course is based on (Ware, 2004).

3. Our Machine Learning Courses

One of our objectives in constructing the curriculum of the Macadamia was to modify our previous curriculum to better fit the currently popular research and application areas. For this purpose two courses focusing on neural network courses were replaced by a single new course "Machine Learning and Neural Networks", and one advanced course on statistical learning methods was replaced by two new courses: the undergraduate introductory course "Machine Learning: Basic Principles", and the graduate level course "Machine Learning: Advanced Probabilistic Methods".

In the following, we describe in more detail these three machine learning courses that form the core of our Macadamia master's programme. More information on the other courses can be found on their homepages; see the web site `http://www.cis.hut.fi/teaching/index_en.shtml`.

3.1. First Course "Machine Learning: Basic Principles"

One of our objectives in constructing the curriculum of the Macadamia was to modify the previous curriculum to better fit the contemporary topics. For this purpose two courses focusing on neural networks were condensed to one, and one advanced course on statistical learning methods was replaced by two courses: the introductory course "Machine Learning: Basic Principles" and advanced course "Machine Learning: Advanced Probabilistic Methods".

After the introductory course, lectured and constructed by Kai Puolamäki, the student should be able to apply the basic methods to real world data, understand the basic principles of the methods and have necessary prerequisites to understand and apply new concepts and methods that build on the topics covered in the course. As a prior knowledge we require the basic mathematics and probability courses, basics of algorithms and the basic programming courses.

The course should also be sufficiently interesting to attract gifted individual for the more advanced courses, but at the same time, be useful also for those for whom it is the only course on machine learning.

We also wanted to have a text book to avoid the situation where the course material would be too scattered; after a long consideration, we chose Alpaydin (2004) as a text book. The purpose of the course was to give emphasis on the principles of machine learning and probabilistic reasoning, and avoid introducing all possible methods.

The course was designed modular, with lectures, problem sessions that took place during the lecturing periods and term projects. At each stage the students were required to apply the knowledge to the real world data sets.

The term project consisted of a non-trivial classification task where the students had to classify web sites as spam given the WEBSPAM-UK2006 collection data. The term project was organized in the form of challenge, with nominal prizes awarded to the best performing teams, who also presented their work during a mini-workshop that took place during one of the problem sessions. We advised the students to favour simple and understandable methods, and not even try fancy approaches that can be found from the literature.

We collected extensive course feedback. The term project was considered by the students challenging and it was appreciated, especially the fact that it was about a "real problem", not just some toy data. The term project made it possible to apply the principles and methods learned in the course into practice. Probably also the challenge format and the mini-workshop were successful as the students saw various approaches to the same problem they had been trying to tackle with.

One source of criticism was that the course had lots of content. One thing for the future is probably to reconsider whether we should drop something out and study some issues more in detail.

Another source of criticism was that many of the students did not have that much experience in using machine learning software, for some, even importing a data set into analysis software was a challenge. The course did not teach the use of data analysis software, and it was designed language or software independent, although the example codes were given in GNU R.

All course materials, including the LaTeX source code of the slides, are available from the course web site at `http://www.cis.hut.fi/Opinnot/T-61.3050/`.

3.2. Second Course "Machine Learning and Neural Networks"

As mentioned before, our laboratory had up to the teaching year 2006-2007 two courses on neural networks, "Principles of Neural Computing" and "Advanced Course in Neural Computing", both worth of 5 ECTS credit points. They were installed in the late 1990's by Profs. Juha Karhunen and Samuel Kaski. The textbook used in both courses was the well-known Prof. Haykin's book (Haykin, 1998).

As the importance of neural networks has decreased in the field of machine learning in favor of probabilistic graphical modeling and other methods during the recent years, we decided to compress our two earlier neural networks courses mentioned above into a single one in the new Macadamia master's programme. The new course, "Machine Learning and Neural Networks", was lectured for the first time from November to mid-December in 2007, with 4 hours lectures and exercises in a week. The course was designed and lectured by Prof. Juha Karhunen, who had lectured and developed on previous years both of our earlier neural networks courses.

Even though Haykin's book (Haykin, 1998) covers probably best the topics discussed in the new course and a new edition of it is under finalization, we chose as the main textbook for the new course Ham's and Kostanic's book (Ham & Kostanic, 2001). The main reason for this choice was that matters are discussed throughout too extensively and thoroughly in Haykin's about 850 pages long book (Haykin, 1998) for the needs of a single course with 5 ECTS credit points.

In Ham's and Kostanic's book, the theory of neural networks is covered much more concisely in the first five chapters covering about 240 pages. Some of this theory is on matters that are not so relevant any more, such as Adaline and Hopfield networks, but we skipped almost completely such out-of-date topics in our course. The remainder in this book deals with various applications of neural networks, but they were not considered either in our course. This was because several of these applications chapters are on linear problems, while a major justification for using neural networks is their ability to tackle difficult problems using distributed nonlinear processing. Linear problems based on second-order statistics can usually be handled much more efficiently using standard numerical and signal processing methods than using slowly converging and inaccurate stochastic gradient type neural algorithms.

The new course "Machine Learning and Neural Networks" consists of the following 12 lectures, each covering one topic area discussed in our course:

1. Introduction to neural networks, examples of their applications.
2. Models of neuron, activation functions, network architectures.
3. Single neuron models and learning rules: least-mean squares (LMS) algorithm, basic perceptron, their weaknesses.
4. Hebbian learning and principal component analysis (PCA), preprocessing of data.
5. Feedforward multilayer perceptron (MLP) networks, backpropagation learning algorithms, their properties and some improvements.
6. Advanced optimization algorithms for multilayer perceptron networks: conjugate gradient algorithm, Levenberg-Marquardt algorithm.
7. Model assessment and selection: generalization, overlearning, regularization, bias-variance decomposition, validation and cross-validation.
8. Radial-basis function (RBF) neural networks and their learning algorithms.
9. Support vector machines for classification and nonlinear regression.
10. Independent component analysis (ICA): basic principles, criteria, learning algorithms, and some applications.
11. Self-organizing maps (SOM) and learning vector quantization (LVQ).
12. Processing of temporal information in feedforward networks, simple recurrent network.

Each of these "lectures" covers one subject entity. The number of slides in them varies greatly, and presenting one of them took in practice often less ore more time than one oral lecture of 2×45 minutes. Most of the lectures were based on Ham's and Kostanic's book (Ham & Kostanic, 2001), but it does not cover all the lecture topics or covers them poorly. The lecture on model assessment and selection was compiled from several sources, in particular from Bishop's book (Bishop, 2006), because these important matters are discussed very little in (Ham & Kostanic, 2001). The lecture on support vector machines was based on Chapter 6 in Haykin's book (Haykin, 1998), and similarly the last lecture on processing of temporal information. Finally, independent component analysis was presented following the treatment in Hyvärinen's and Oja's tutorial article (Hyvärinen & Oja, 2000), which is still a good and highly readable introduction to the basic independent component analysis.

There is a little overlap with our first machine learning course "Machine Learning: Basic Principles" on model assessment and selection as well as on principal component analysis, but this was not considered harmful, because these matters are important and discussed from different viewpoints in these courses. There was a lot of work in designing the course and writing the lecture slides and solutions to the exercise problems. During the forthcoming years, the course will most probably remain largely as it is now, but the new promising paradigm Extreme learning machine (Huang et al., 2006) will be included in the discussion of multilayer perceptron networks.

We used the exercise problems of our two earlier neural networks courses, but also added new problems. Both Haykin's book (Haykin, 1998) and Ham's and Kostanic's book (Ham & Kostanic, 2001) have accompanying solutions manuals, making it easier to select instructive problems of suitable difficulty level. Computer demonstrations were also presented in context with the exercises to give the students an idea of how the methods discussed perform in practice. Another reason for presenting demos is that it is difficult to design suitable exercise problems for example on highly nonlinear multilayer perceptron networks.

The course also contains a computer assignment, which is selected randomly for each student from five possible assignments. Two of them are on multilayer perceptron networks, two on self-organizing maps, and one on independent component analysis. All the lecture slides, exercise problems and their solutions, prepared using Latex, are available in .pdf form on the home page of the course `http://www.cis.hut.fi/Opinnot/T-61.5130/`. The original Latex source files can be requested from the lecturer of the course, Prof. Juha Karhunen (email: `Juha.Karhunen@tkk.fi`).

3.3. Third Course "Machine Learning: Advanced Probabilistic Methods"

This course is the most advanced course in the Macadamia program and builds on the previous courses covered in earlier sections. This gives the opportunity to rely on a certain background knowledge taught on the prerequisite courses and concentrate on probabilistic methods without the need to recapitulate on the underlying basics. The textbook used on the course is the Bishop's recent book (Bishop, 2006) on pattern recognition and machine learning. During the course, a subset of five chapters are covered, the textbook is complemented with additional material when needed.

The main topic of the course is probabilistic inference and learning in the context of machine learning, with special emphasis on the framework of graphical models. The course is started with the presentation of Bayesian networks and the Bayes's theorem as the solution to "answer questions", that is, to perform inference using the model and the available evidence. After the rather general presentation on Bayesian networks and the algorithmic possibilities of performing exact inference, the course covers mixture models and the EM algorithm. The order of the material has been designed in the hope that the students realize that mixture models are "just simple Bayesian networks", so the earlier material applies here as well. Naturally, we enforce this view in the lectures and the exercises actively. The EM algorithm is first presented in the context of mixture models, which makes the presentation easier than in the most general settings. First, the EM algorithm is presented as is, and on the later lectures the principles of derivation are presented. The course continues on the topic of models for sequential data, again with the emphasis on the models being "just a little bit more complex Bayesian networks". Models building on Markov-chains, such as Hidden Markov models and the possibilities for extensions are presented. Up to this point, exact inference and the EM algorithm have been the the main tools. Towards the end of the course, approximate inference with sampling and variational algorithms are presented. This is backed up earlier material in the Bayesian networks framework, where the possible reasons for the difficulty or even infeasibility of inference were reviewed.

The course contents and the organization has been designed by Jaakko Hollmén, also he lectured the course. The practical pen-and-paper course exercises have been designed together with Tapani Raiko, who led the exercise sessions. The ultimate goal behind the design of the course is to teach the fundamental principles behind probabilistic inference, that is, the Bayes's theorem and the computational principles in its execution in different models, and to apply the principles in model construction and learning from data. Whereas the principles are of great importance, even more important is to tie together the world of principals with the pragmatic implementations and real-world data sets. Towards this aim, practical implementations in a concrete machine learning scenario are presented during the lectures and exercises. Moreover, a term project on the topic of mixture modeling is given. Contrary to our earlier standard of giving programming exercises to the students, a full software package BernoulliMix has been used. In the term project, the students concentrate on the modeling aspects, and thinking of results instead of "just getting the programs to work". The design of the exercise and the contents of the BernoulliMix program package is reviewed in an independent article in the current volume, see (Hollmén & Raiko, 2008). The BernoulliMix home page is at `http://www.cis.hut.fi/jhollmen/BernoulliMix`, where the manual including the exercises may be found.

The preliminary experiences about the course running for the first time during the Spring 2008 have been positive. During the last lecture, the exam requirements were presented, and lots of effort was put on relating all the taught material to each other and emphasizing important topics. This was especially appreciated among the students in their free-form feedback after the last lecture, they hoped for this kind of relating even more in the beginning parts on the course. Paradoxically, in the beginning phases of the course, the teacher may only related the material learned so far (but some careful forward references to coming material and topics may be done). This can be seen as a positive feedback on the used framework relying on graphical models, or more specifically Bayesian networks and them placing all the subsequent material in that framework.

The course material, including lecture notes, problems and solutions to the exercises may be found on the course home page at `http://www.cis.hut.fi/Opinnot/T-61.5140/`.

4. Dual Degrees

Macadamia has agreements for a dual degree with three other Master's programmes in the Cluster[1] network. The student would spend the first year in the home university and the second year in the partner university and get a dual degree from both universities. Hopefully the number of partners would increase in the future. The three pilot agreements are with Master's programme in Artificial Intelligence (MIA) at Universitat Politècnica de Catalunya (UPC), with Master's programme in Information Technology (MTI) — Specialisation on Information Management and Use at UPC, and with Master's programme in Information Systems and Computer Engineering — major in Intelligent Systems at Universidade Técnica de Lisboa, Instituto Superior Técnico (IST).

[1] Consortium Linking Universities of Science and Technology for Education and Research

References

Alpaydin, E. (2004). *Introduction to machine learning*. The MIT Press.

Bishop, C. (2006). *Pattern recognition and machince learning*. Springer-Verlag.

Ham, F., & Kostanic, I. (2001). *Principles of neurocomputing for science & engineering*. McGraw-Hill.

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Adaptive Computation and Machine Learning Series. MIT Press.

Haykin, S. (1998). *Neural networks: A comprehensive foundation, 2nd ed.* Prentice-Hall.

Hollmén, J., & Raiko, T. (2008). Learning mixture models - courseware for finite mixtures of Bernoulli distributions. *Teaching Machine Learning: Workshop on open problems and new directions*. Saint-Étienne, France.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, *70*, 489–501.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. John Wiley & Sons.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, *13*, 411–430.

Kohonen, T. (1995). *Self-organizing maps*. Springer-Verlag.

Mannila, H., & Toivonen, H. (1998). *Knowledge discovery in databases: the search for frequent patterns*. Available at http://www.cis.hut.fi/Opinnot/T-61.5060/t122103-2003.pdf.

Oja, E., Simula, O., Karhunen, J., Mannila, H., & Kaski, S. (2007–). Macadamia: Master's programme in machine learning and data mining. Academic coordinator: T. Raiko. http://www.cis.hut.fi/macadamia/.

Ware, C. (2004). *Information visualization: Perception for design, 2nd edition*. Morgan Kaufmann.

Table 1. Courses given in the programme. The size of the courses are given in credit points (ECTS). The total for the two-year programme is 120 ECTS. Note that the Special Courses have a varying topic (5–6 topics per year) so many of them may be included in the curriculum.

Obligatory courses	ECTS
IT-Services at TKK	2
English language tests / course	3
Machine Learning: Basic Principles	5
Machine Learning and Neural Networks	5
Machine Learning: Advanced Probabilistic Methods	5
Algorithmic methods of data mining	5
Information Visualization	5
Research Project in Computer and Information Science	5–10
Master's thesis	30
Relevant courses	**ECTS**
Computer Vision	5
Statistical Natural Language Processing	5
High-Throughput Bioinformatics	5
Signal Processing in Neuroinformatics	5
Image Analysis in Neuroinformatics	5
Special Course in Computer and Information Science I–VI	3–7
Introduction to Bayesian Modelling	5
Combinatorial Models and Stochastic Algorithms	6
Search problems and algorithms	4
Parallel and distributed systems	4
Cryptography and data security	4
Computational Complexity Theory	5
Finnish 1A	2
Finnish 1B	2
Finnish 2A	2
Finnish 2B	2
Topics of Special Courses during 2006–2008	**ECTS**
Gaussian Processes for Machine Learning	6
Popular Algorithms in Data Mining and Machine Learning	5
Reinforcement Learning — Theory and Applications	6
Multimedia Retrieval	5
Introductory Elements of Functional Data Analysis	7
Independent Component Analysis	6
Information Networks	6
Variable Selection for Regression	6
Nonlinear Dimensionality Reduction	6
Modeling and Simulating Social Web	4
Decision support with data analysis	5
Data analysis and environmental informatics	5

Masterâ€™s in Machine Learning, Data Science and Artificial Intelligence â€“ Aalto University. Machine learning is one of the major strengths of Aalto University. This Masterâ€™s program will give you excellent opportunities for a career in research institutions or in the private sector in the rapidly developing fields of machine learning, data science, and artificial intelligence.Â The program covers areas such as large-scale data analytics, advanced machine learning and data-mining, information retrieval, natural language processing and web mining. It also includes foundational modules on topics such as statistics, probability and programming for data analytics. Main concepts of machine learning and data mining. Programming on Python language using the main scientific packages like Scikit-learn, Pandas, Numpy, etc. Manage real data and develop a desktop applications for machine learning and data mining. Course content. 7 sections â€¢ 83 lectures â€¢ 12h 24m total length.Â He has an Industrial Engineering degree from Pontificia Universidad Catolica del Peru (Lima-Peru) and Master in Business Administration (MBA) from ESAN Graduated School of Business (Lima-Peru). He is also an experience developer of machine learning and data science models in many fields of the industry and services like Marketing, Logistics, Finance, Manufacture, Quality Control, Computer Vision, NLP, Deep Learning apps and many others. An online Masterâ€™s degree in Machine Learning and Data Science from Imperial College London. This new Master's program is designed to help propel your engineering or data science career forward.Â Unlike other masterâ€™s in data science programmes that teach Machine Learning with a computer science focus, this degree prepares students with the mathematical and statistical theory needed to truly understand machine learning, as well as the practical skills to deal with real world applications that they need to be successful in their careers. The programme will train students in the mathematical, computational, and statistical foundations of machine learning, giving them the ability to critique data analysis and implement scalable machine learning solutions. The Masterâ€™s program in Business Analytics and Big Data (MBD) at IE Business School is a 10-month intensive program. Students will engage in three intensive team challenges to get practical, hands-on experience working directly with industry experts. The four major program areas are business transformation, data science, big data technologies, and professional skills.Â The methods of machine learning and data mining are applicable and needed in a wide variety of fields ranging from process industry to data science. Recent spearhead application areas include. bioinformatics.Â The major in Machine Learning, Data Science and Artificial Intelligence (Macadamia) covers a wide range of topics in modern computational data analysis and modeling methodologies. Master's Degree Programme in Computing Science â€“ Machine Learning. Featured help_outline. Read More. Tampere University. Tampere, Finland. Machine Learning is an engineering programme, where the particular emphasis is on implementing machine learning algorithms with an application-oriented focus, with application + Featured help_outline. MSc.Â We integrate statistical modeling and analysis with machine learning, data mining, and data manage + Featured help_outline. MSc.Â Machine Learning develops algorithms to find patterns or make predictions from empirical data and this masterâ€™s programme will teach you to master these skills. Machine Learni + Featured help_outline.