Support Vector Machines and Kernel Methods

T-61.183 Special Course in Computer and Information Science III, Spring 2003 (4 cr)

Prof. Juha Karhunen
Helsinki University of Technology

General information

- The course is based on selected parts the book

  B. Schölkopf and A. Smola: Learning with Kernels

- Subtitle: Support Vector Machines, Regularization, Optimization, and Beyond.
- Publisher: The MIT Press. Price: 60 US dollars.
- About 640 pages $\Rightarrow$ We shall skip some parts of the book in this course.
- Comments and reviews on the book have been quite positive.
- The book is mathematically fairly demanding and theoretical.
- We shall skip advanced special topics and overly theoretical portions of the book in our course.

- Home page of the book:
  http://www.learning-with-kernels.org/
- There is plenty of additional useful material available.
- A tutorial review article "An introduction to kernel-based learning algorithms" by Prof. K.-R. Müller et al., IEEE Trans. on Neural Networks, March 2001.
- This article is a kind of summary of the seminar book.
- It is delivered to you, and you should read the article; it makes easier for you to follow the seminar later on.
- Before reading this article, please read Chapter 1: A Tutorial Introduction of the book.
- It tries to explain the central ideas in a simple manner.
- Chapter 1 is partly overlapping with the review article, which covers

using rather condensed style more material.

- Prof. Müller's slides based on these tutorial materials are available on the web page of the course.
- Another book: N. Christiani and J. Shave-Taylor, "An Introduction to Support Vector Machines and other kernel-based learning methods", Cambridge Univ. Press 2000, about 200 pages.
- Concise but clear presentation.
- A third book has been ordered: R. Herbrich (Ed.), "Learning Kernel Classifiers", MIT Press 2002, about 340 pages.
- Focuses on classification learning, good for beginners.
- Useful web sites containing articles, MATLAB or C^{++} software, and other information:

http://www.kernel-machines.org/

http://www.isis.ecs.soton.ac.uk/resources/svminfo/

http://www.torch.ch/

- Support vector machines (SVM's) and related kernel methods are currently a hot topic in learning.
- Basic idea: an implicitly computed feature mapping into a higher dimensional space.
- After this, 'nonlinear' classifiers and representations can be implemented using linear techniques.
- Reasons for the popularity of SVM's and other kernel methods:
  - Theoretically more tractable than for example neural networks.
  - Usually very good statistical performance.
  - A large number of efficient algorithms and implementations are

available.

  - Computationally feasible in high-dimensional problems.

- SVM's are especially useful in pattern classification.
- Kernel PCA can be used for nonlinear representation.
- The method performs often well, but is somewhat heuristic.
- Neural network and Bayesian methods are still quite useful in many situations.
- Kernel methods cannot be applied to all problems.
- Basic restriction: the algorithm to be kernelized should be representable using inner products.

Organization of the seminar

- Four (4) credit points as usual in our seminar courses.

- Weekly in the lecture room T4 on Mondays 14:15-16.
- No seminar meeting on 31st March due to conference travels.
- For a more detailed programme, see a separate file on the web page of the seminar.
- Responsible teacher: Prof. Juha Karhunen
- Email: `Juha.Karhunen@hut.fi`, room TB327, tel. 451 3270.
- Course assistant: MSc. Karthikesh Raju
- His email: `karthik@james.hut.fi`; room TB330.
- The language of the course is English due to foreign participants.
- The course is intended mainly for graduate students.
- Can be taken by undergraduate students who are mature enough.
- Sufficient mathematical background, most of Dipl.Eng. studies done.

- In Appendix B of the book, some mathematical prerequisites are presented.
- You should already know probability theory and linear algebra.
- You can read necessary background on functional analysis from there if necessary (about 5 pages).
- Knowlegde on some of our laboratory's courses (Principles of Pattern Recognition, Neural Networks) is helpful but not necessary.
- There will be no examination; this would be difficult for graduate students working full-time in corporations etc.
- You should write your name, study book number, email address, and department to the enrollment list circulating.
- The last two columns: Insert a cross (x) in you are a graduate student, or wish to buy the book.

- The price of the book is not too high, 60 USD.
- The delivery time is typically some 2-4 weeks.
- Alternatively, you can content with a copy of the book.
- A master copy made by Karthik is already available.
- You can take your own copies from it yourself.
- Please reserve your own talk from the list circulating in the seminar meetings as soon as possible.

Requirements for passing the course

- Sufficient participation (about 70%) in the seminar meetings.
  - Put a cross to the attendance list whenever you attend.
- You must prepare and present your own talk.
  - Usually one talk per meeting, about 1 hour long.

- If there are two talks, they could be about 45 min each.
- Skip proofs, excessive theory and details, and try to explain clearly the most important matters of the scope of your talk.
- All the figures of the book are available on its home page.
- If possible or reasonable, prepare .pdf slides of your talk.
- These will be inserted to the web page of the seminar.

- Solving sufficient amount of the given problems.
  - Solving 50% of the total number of problems suffices for the mark "accepted" (hyväksytty).
  - Solving 90-95% of problems is required for the mark "accepted with distinction (kiittäen hyväksytty)
  - Typically there will be $2-3$ problems per each talk.
  - You should select yourself $2-3$ problems on the portion of the book covered in your talk.

- You must be able to solve these problems yourself!
- This guarantees that the problems are not too difficult.
- Give the correct solutions of the problems chosen by you to the course organizer Prof. Juha Karhunen
- Problems are useful because they force people to read the corresponding parts of the book.
- It is preferable but not necessary to return your solutions to the problems given within 2 weeks.

• Performing given computer assigment(s).

- Easy assigments giving some hands-on experience using available data sets and MATLAB or C$^+$+ software.