Machine Learning Summer Schools

Gunnar Rätsch
Bernhard Schölkopf

PASCAL meeting, Bled
History
MLSS Tübingen, Germany, August 20 - August 31, 2007
MLSS Taipei, Taiwan, July 24 - August 2, 2006
MLSS Canberra, Australia, February 6-17, 2006
MLSS Chicago, USA, May 16-27, 2005
MLSS Canberra, Australia, January 23 - February 5, 2005
MLSS Berder, France, September 12-25, 2004 (Wiki)
MLSS Tübingen, Germany, August 4-16, 2003
MLSS Canberra, Australia, February 2-14, 2003
MLSS Canberra, Australia, February 11-22, 2002
Sponsors

- PASCAL
- Nokia
- Google
- Microsoft Research
- Advanced Unibyte

--------------------------------------------------------

- Max Planck Society
- NICTA
Application Statistics

- **in total**
  - 333 regular applications
  - Many more informal inquiries

- **107 accepted participants**

- **Acceptance rate**
  - 32% (70% for PASCAL students)

- **25 countries, average 4 per country (median 3)**

<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>4</td>
</tr>
<tr>
<td>Canada</td>
<td>6</td>
</tr>
<tr>
<td>Chile</td>
<td>1</td>
</tr>
<tr>
<td>Denmark</td>
<td>2</td>
</tr>
<tr>
<td>Finland</td>
<td>3</td>
</tr>
<tr>
<td>France</td>
<td>10</td>
</tr>
<tr>
<td>Germany</td>
<td>21</td>
</tr>
<tr>
<td>Greece</td>
<td>1</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>1</td>
</tr>
<tr>
<td>Hungary</td>
<td>2</td>
</tr>
<tr>
<td>Israel</td>
<td>4</td>
</tr>
<tr>
<td>Italy</td>
<td>4</td>
</tr>
<tr>
<td>Japan</td>
<td>2</td>
</tr>
<tr>
<td>Malaysia</td>
<td>1</td>
</tr>
<tr>
<td>Norway</td>
<td>1</td>
</tr>
<tr>
<td>Portugal</td>
<td>2</td>
</tr>
<tr>
<td>Slovenia</td>
<td>1</td>
</tr>
<tr>
<td>South Africa</td>
<td>3</td>
</tr>
<tr>
<td>Spain</td>
<td>3</td>
</tr>
<tr>
<td>Switzerland</td>
<td>4</td>
</tr>
<tr>
<td>Thailand</td>
<td>2</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>3</td>
</tr>
<tr>
<td>Turkey</td>
<td>2</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>10</td>
</tr>
<tr>
<td>USA</td>
<td>9</td>
</tr>
</tbody>
</table>
Application Statistics II

- 10 715€ for grants

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate</td>
<td>10</td>
</tr>
<tr>
<td>PhD Student</td>
<td>71</td>
</tr>
<tr>
<td>Postdoc</td>
<td>9</td>
</tr>
<tr>
<td>Professor</td>
<td>6</td>
</tr>
<tr>
<td>Industrial</td>
<td>6</td>
</tr>
</tbody>
</table>
Lecture Courses (4-8h)

- Andrew Blake: Topics in Image and Video Processing
- Olivier Bousquet: Statistical Learning Theory
- Nicolò Cesa-Bianchi: Online Learning
- Arnaud Doucet: Sequential Monte Carlo Methods
- Zoubin Ghahramani: Graphical models
- László Györfi: Machine Learning and Finance
- Kenji Fukumizu: Kernel Methods for Dependence and Causality
- Carl E. Rasmussen: Bayesian Inference and Gaussian Processes
- Gunnar Rätsch: Introduction to Bioinformatics
- Bernhard Schölkopf & Alex Smola: Introduction to Kernel Methods
- Lieven Vandenberghe: Convex Optimisation
Practical Courses (2h)

- Joaquin Quiñonero Candela: Gaussian Processes
- Manuel Davy: Practical Sampling
- Matthias Hein & Ulrike von Luxburg: Spectral Clustering and Other Graph Based Algorithms
- Matthias Seeger: Variational Bayesian Inference
- Yee Whye Teh: Dirichlet Processes
Evening Speakers (1h)

- Andreas Dengel: Learning Mental Associations as a Means to Build Organisational Memories
- Uwe Hanebeck: Stochastic Information Processing in Sensory Networks
- Oliver Kohlbacher: Lost in Translation -- Solving biological problems with Machine Learning
- Joachim Weickert: Regularization in Image Analysis

(local Professors which are experts in their field)
## Schedule Week 1

<table>
<thead>
<tr>
<th>Time</th>
<th>Monday 20.08</th>
<th>Tuesday 21.08</th>
<th>Wednesday 22.08</th>
<th>Thursday 23.08</th>
<th>Friday 24.08</th>
<th>Saturday 25.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00-10:45</td>
<td>Smola 1</td>
<td>Rasmussen 1</td>
<td>Rasmussen 3</td>
<td>Ghahramani 3</td>
<td>Ghahramani 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schölkopf 1</td>
<td>Rasmussen 2</td>
<td>Rasmussen 4</td>
<td>Ghahramani 4</td>
<td>Ghahramani 6</td>
<td>Vandenberghe 1</td>
</tr>
<tr>
<td>11:15-13:00</td>
<td>Schölkopf 2</td>
<td>Smola 2</td>
<td>Rätsch5</td>
<td>Fukumizu 1</td>
<td>Fukumizu 3</td>
<td>Vandenberghe 2</td>
</tr>
<tr>
<td></td>
<td>Schölkopf 3</td>
<td>Smola 3</td>
<td>Rätsch6</td>
<td>Fukumizu 2</td>
<td>Fukumizu 4</td>
<td>Vandenberghe 3</td>
</tr>
<tr>
<td>13:00-14:00</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>14:00-16:45</td>
<td>Hein, Luxburg</td>
<td>Hein, Luxburg</td>
<td>Quinonero Candela</td>
<td>Quinonero Candela</td>
<td>Seeger</td>
<td></td>
</tr>
<tr>
<td>16:45–17:15</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>17:15-19:00</td>
<td>Rätsch 1</td>
<td>Rätsch 3</td>
<td>Ghahramani 1</td>
<td>Rasmussen 5</td>
<td>Student Pres.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rätsch 2</td>
<td>Rätsch 4</td>
<td>Ghahramani 2</td>
<td>Rasmussen 6</td>
<td>Student Pres.</td>
<td></td>
</tr>
<tr>
<td>19:00-20:00</td>
<td>Welcome</td>
<td>Dinner</td>
<td>BBQ</td>
<td>Dinner</td>
<td>Dinner</td>
<td></td>
</tr>
<tr>
<td>20:30-21:30</td>
<td>Dinner</td>
<td>BBQ</td>
<td>Kohlbacher</td>
<td>Dinner</td>
<td>Dinner</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bar</td>
<td>Bar</td>
<td>Bar</td>
<td>Bar</td>
<td>Bar</td>
<td></td>
</tr>
</tbody>
</table>
## Schedule Week 2

<table>
<thead>
<tr>
<th>Time</th>
<th>Sun 26.08.</th>
<th>Mon 27.08.</th>
<th>Tue 28.08.</th>
<th>Wed 29.08.</th>
<th>Thrs 30.08.</th>
<th>Fri 31.08.</th>
<th>Sat 01.09.</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00-10:45</td>
<td>Bousquet 1</td>
<td>Bousquet 3</td>
<td>Bousquet 5</td>
<td>Cesa-Bianchi 3</td>
<td>Doucet 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bousquet 2</td>
<td>Bousquet 4</td>
<td>Bousquet 6</td>
<td>Cesa-Bianchi 4</td>
<td>Doucet 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:45-11:15</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td></td>
<td>Departure</td>
</tr>
<tr>
<td>11:15-13:00</td>
<td>Teh 1</td>
<td>Blake 3</td>
<td>Cesa-Bianchi 1</td>
<td>Doucet 1</td>
<td>Cesa-Bianchi 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teh 2</td>
<td>Blake 4</td>
<td>Cesa-Bianchi 2</td>
<td>Doucet 2</td>
<td>Cesa-Bianchi 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:00-14:00</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>14:00-16:45</td>
<td>Teh or break</td>
<td>Teh or break</td>
<td>Seeger or break</td>
<td>Davy or break</td>
<td>Davy or break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:45-17:15</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>17:15-19:00</td>
<td>Blake 1</td>
<td>Fukumizu 5</td>
<td>Györfi 1</td>
<td>Doucet 3</td>
<td>Györfi 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blake 2</td>
<td>Fukumizu 6</td>
<td>Györfi 2</td>
<td>Doucet 4</td>
<td>Györfi 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:00-20:00</td>
<td>Conference Dinner</td>
<td>Dinner</td>
<td>BBQ</td>
<td>Dinner</td>
<td>Dinner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20:30-21:30</td>
<td>Weickert</td>
<td>Hanebeck</td>
<td>Final Party</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>late</td>
<td>Bar</td>
<td>Bar</td>
<td>Bar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MLSS Courses Canberra 2002

- Reinforcement Learning (Peter Bartlett)
- Boosting (Ron Meir)
- Statistical Learning Theory and Empirical Processes (Shahar Mendelson)
- Online Learning and Bregman Divergences (Gunnar Rätsch and Manfred Warmuth)
- Support Vector Machines and Kernels (Bernhard Schölkopf)
- Bayesian Kernel Methods (Alex Smola)

plus the following short courses:
- Learning for Control: Adaptive Control Problems are different (Brian Anderson)
- Nonparametric Estimation of Component Distributions in a Multivariate Mixture (Peter Hall)
- Algorithms for Association Rules (Markus Hegland)
- Online Loss Bounds (Jyrki Kivinen)
- Learning from Structured Data (John Lloyd)
- A Unified Approach to Deduction and Induction (Arun Sharma)
- Inductive Principles (Bob Williamson)
MLSS Courses Canberra 2003

- Information Geometry (Shun-Ichi Amari)
- Concentration Inequalities (Gabor Lugosi)
- Unsupervised Learning (Zoubin Ghahramani)

plus short courses by
- Eleazar Eskin
- Peter Hall
- Markus Hegland
- John Lloyd
- Shahar Mendelson
- Mike Osborne
- Gunnar Rätsch
- Alex Smola
- S.V.N. Vishwanathan
- Bob Williamson
- Petra Philips
MLSS Courses Tübingen 2003

Long Courses
- Statistical Learning Theory O. Bousquet
- Independent Component Analysis J-F. Cardoso
- Probabilistic Models and Gaussian Processes C.E. Rasmussen
- Kernel Algorithms I B. Schölkopf
- Kernel Algorithms II A. Smola
- Pattern Classification E. Yom-Tov

Short Courses
- Monte Carlo Simulation Methods C. Andrieu
- Bioinformatics P. Baldi
- Stochastic Approximation L. Bottou
- Concentration Inequalities S. Boucheron
- Mathematical Tools for Machine Learning C. Burges
- Minimum Description Length P. Grünwald
- Information Retrieval and Language Technology T. Joachims
- Foundations of Learning S. Smale
- Principles and Practice of Bayesian Learning M. Tipping

Practical Sessions
- Simulation Methods M. Davy
- Support Vector Machines A. Elisseeff, A. Gretton, J. Weston
- Pattern Classification – From Data to Decision E. Yom-Tov

Evening Talks
- On Learning Vector-Valued Functions M. Pontil
- Empirical Inference V. Vapnik

Student Talks
- Half-day mini-conference, talks and posters
Some Quotations (Tübingen 2003 & 2007)

2003:
◆ “Excellent organization! [...] Also the price for students was very reasonable!-
   More practical sessions? They were really excellent and for PhD students
cannot be recommended highly enough.”
◆ “Most of the lecturers at our university could have learned quite a lot from the
   mostly non-everyday-lecturers at the summer school.”
◆ “Excellent, I really enjoyed the opportunity to see famous people and to have
   the opportunity to contact other students. [...] Please, try to keep the effort as
   many years as possible because it is worth it.”

2007:
◆ “Yes, I liked it a lot. It was a great mixture of study and social activities, being
   able to learn new things in an area you are interested in, and get to know
   many interesting people. [...] I liked the practical sessions very much [...]”
◆ “I think it would be useful to send out [...] some readings [before the summer
   school], so that it will be easier to attend some of the lectures [...]”
◆ “I very much enjoyed MLSS. The lectures were consistently of a high quality,
   and the environment was fantastic. I liked the fact that much of the content
   was accessible to people from outside the field like myself.”
Elements of a Machine Learning Syllabus

Block 1: **Mathematics**
- analysis *
- linear algebra *
- functional analysis *
- numerical mathematics and mathematical programming *
- probability theory *
- statistics *
- empirical process theory and concentration inequalities
- approximation theory
- multivariate analysis
- nonparametric statistics
- differential geometry
- convex analysis

Block 2: **Computer Science and Engineering**
- computer programming *
- signal processing *
- Software development
- computer vision
- robotics
- control theory

Block 3: **Human Sciences**
- experimental psychology
- neuroscience
- linguistics

Block 4: **Philosophy**
- philosophy of science
- epistemology

Block 5: **Specialized Courses**
- (cf. Machine Learning Summer Schools)

* undergraduate material
A “Parallel Distributed” European MSc/PhD in ML?

- group of 10 institutions offering graduate level credits
- credits = courses, lab rotations, etc.
- credits can also be earned in summer schools
- typically, students will spend periods of 6 months in each place, giving sufficient time to complete a course
- different sites offer different specialized courses
- thesis can be written at any participating lab and is reviewed by professors from different sites
- degree is awarded by a PASCAL member university, ideally a prestigious one
- model: "intercollegiate London MSc. in mathematics"

Benefits:
- might foster also scientific collaborations (often collaborations are driven by students)
- will help retain top European students, and attract strong overseas students
MLSS Courses Berder 2004

- Graphical Models and Variational Methods (C. Bishop)
- Computer Vision (A. Blake)
- Advanced Statistical Learning Theory (O. Bousquet)
- Regularization (S. Canu)
- Simulation Methods (M. Davy)
- Information Retrieval and Text Mining (T. Hofman)
- Control Systems (J. Moore)
- Boosting (G. Rätsch)
- Kernel Methods I (B. Schölkopf)
- Statistical Learning Theory (J. Shawe-Taylor)
- Kernel Methods II (A. Smola)
- Empirical Inference (V. Vapnik)
- Signal Processing (R. Williamson)
- Machine Learning in Bioinformatics (A. Zien)

Plus practical sessions:
- Semi-supervised Learning (O. Chapelle)
- Simulation Methods (M. Davy)
- Independent Component Analysis (A. Gretton)