



IN PARTNERSHIP WITH:
CNRS

**Ecole normale supérieure de
Paris**

Activity Report 2012

Project-Team SIERRA

Statistical Machine Learning and Parsimony

IN COLLABORATION WITH: Département d'Informatique de l'Ecole Normale Supérieure

RESEARCH CENTER
Paris - Rocquencourt

THEME
**Optimization, Learning and Statistical
Methods**

Table of contents

1. Members	1
2. Overall Objectives	1
2.1. Statement	1
2.2. Highlights of the Year	2
3. Scientific Foundations	2
3.1. Supervised Learning	2
3.2. Unsupervised Learning	2
3.3. Parsimony	2
3.4. Optimization	2
4. Application Domains	3
5. Software	3
5.1. SPAMS (SPArse Modeling Software)	3
5.2. SiGMa - Simple Greedy Matching: a tool for aligning large knowledge-bases	3
5.3. minFunc (2012 version)	4
5.4. prettyPlot	4
5.5. SegAnnot	4
6. New Results	4
6.1. A Stochastic Gradient Method with an Exponential Convergence Rate for Strongly-Convex Optimization with Finite Training Sets	4
6.2. Convex Relaxation for Combinatorial Penalties	5
6.3. Kernel change-point detection	5
6.4. On the Equivalence between Herding and Conditional Gradient Algorithms	5
6.5. V -fold cross-validation and V -fold penalization in least-squares density estimation	5
6.6. Machine learning for Neuro-imaging	6
6.7. SiGMa: Simple Greedy Matching for Aligning Large Knowledge Bases	6
6.8. Block-Coordinate Frank-Wolfe Optimization for Structural SVMs	6
6.9. A convex relaxation for weakly supervised classifiers	7
6.10. Multi-Class Cosegmentation	7
6.11. A latent factor model for highly multi-relational data	7
6.12. Semi-supervised NMF with time-frequency annotations for single-channel source separation	8
7. Bilateral Contracts and Grants with Industry	8
8. Partnerships and Cooperations	8
8.1. National Initiatives	8
8.1.1.1. Calibration	8
8.1.1.2. Detect	9
8.2. European Initiatives	9
8.3. International Initiatives	10
8.4. International Research Visitors	10
9. Dissemination	10
9.1. Scientific Animation	10
9.1.1. Editorial boards	10
9.1.2. Area chairs	10
9.1.3. Reviewing	10
9.1.4. Other	11
9.1.5. Workshop and conference organization	11
9.2. Teaching - Supervision - Juries	11
9.2.1. Teaching	11
9.2.2. Supervision	11
9.3. Invited presentations	11

9.4. Prizes and awards	13
9.5. Popularization	13
10. Bibliography	13

Project-Team SIERRA

Keywords: Machine Learning, Statistics, Convex Optimization, Data Mining

Creation of the Project-Team: January 01, 2011 .

1. Members

Research Scientists

Sylvain Arlot [Junior Researcher CNRS]

Francis Bach [Team leader, Senior researcher “détaché” at Inria from Corps des Mines, HdR]

Guillaume Obozinski [Ingénieur Expert de Recherche]

Engineer

Jean-Paul Chieze [Ingénieur SED]

PhD Students

Louise Benoît

Florent Couzinié-Devy

Fajwel Fogel

Edouard Grave

Toby Hocking [graduated on November 20, 2012]

Armand Joulin [graduated on December 17, 2012]

Rémi Lajugie

Loic Landrieu

Augustin Lefèvre [graduated on October 3, 2012]

Alex Mesnil

Anil Nelakanti [Cifre Ph.D. with Xerox]

Fabian Pedregosa

Thomas Schatz

Matthieu Solnon

Post-Doctoral Fellows

Simon Lacoste-Julien [financed by Mairie de Paris and ERC grant]

Nicolas Le Roux [financed by ERC grant, until September 20, 2012]

Ronny Luss [financed by associated team STATWEB, until September 1, 2012]

Mark Schmidt [financed by ERC grant]

Nino Shervashidze [since December 1, 2012]

Visiting Scientist

Michael Jordan [financed by Fondation de Sciences Mathématiques de Paris]

Administrative Assistant

Lindsay Polienor

2. Overall Objectives

2.1. Statement

Machine learning is a recent scientific domain, positioned between applied mathematics, statistics and computer science. Its goals are the optimization, control, and modelisation of complex systems from examples. It applies to data from numerous engineering and scientific fields (e.g., vision, bioinformatics, neuroscience, audio processing, text processing, economy, finance, etc.), the ultimate goal being to derive general theories and algorithms allowing advances in each of these domains. Machine learning is characterized by the high quality and quantity of the exchanges between theory, algorithms and applications: interesting theoretical problems almost always emerge from applications, while theoretical analysis allows the understanding of why and when popular or successful algorithms do or do not work, and leads to proposing significant improvements.

Our academic positioning is exactly at the intersection between these three aspects—algorithms, theory and applications—and our main research goal is to make the link between theory and algorithms, and between algorithms and high-impact applications in various engineering and scientific fields, in particular computer vision, bioinformatics, audio processing, text processing and neuro-imaging.

Machine learning is now a vast field of research and the team focuses on the following aspects: supervised learning (kernel methods, calibration), unsupervised learning (matrix factorization, statistical tests), parsimony (structured sparsity, theory and algorithms), and optimization (convex optimization, bandit learning). These four research axes are strongly interdependent, and the interplay between them is key to successful practical applications.

2.2. Highlights of the Year

- Rodolphe Jenatton (former PhD student, graduated in 2011) received two thesis prizes (Fondation Hadamard and AFIA).
- Francis Bach received the Inria young researcher prize.
- Monograph published in the collection *Foundations and Trends in Machine Learning*: “Optimization with sparsity-inducing penalties”.

3. Scientific Foundations

3.1. Supervised Learning

This part of our research focuses on methods where, given a set of examples of input/output pairs, the goal is to predict the output for a new input, with research on kernel methods, calibration methods, and multi-task learning.

3.2. Unsupervised Learning

We focus here on methods where no output is given and the goal is to find structure of certain known types (e.g., discrete or low-dimensional) in the data, with a focus on matrix factorization, statistical tests, dimension reduction, and semi-supervised learning.

3.3. Parsimony

The concept of parsimony is central to many areas of science. In the context of statistical machine learning, this takes the form of variable or feature selection. The team focuses primarily on structured sparsity, with theoretical and algorithmic contributions (this is the main topic of the ERC starting investigator grant awarded to F. Bach).

3.4. Optimization

Optimization in all its forms is central to machine learning, as many of its theoretical frameworks are based at least in part on empirical risk minimization. The team focuses primarily on convex and bandit optimization, with a particular focus on large-scale optimization.

4. Application Domains

4.1. Application Domains

Machine learning research can be conducted from two main perspectives: the first one, which has been dominant in the last 30 years, is to design learning algorithms and theories which are as generic as possible, the goal being to make as few assumptions as possible regarding the problems to be solved and to let data speak for themselves. This has led to many interesting methodological developments and successful applications. However, we believe that this strategy has reached its limit for many application domains, such as computer vision, bioinformatics, neuro-imaging, text and audio processing, which leads to the second perspective our team is built on: Research in machine learning theory and algorithms should be driven by interdisciplinary collaborations, so that specific prior knowledge may be properly introduced into the learning process, in particular with the following fields:

- Computer vision: object recognition, object detection, image segmentation, image/video processing, computational photography. In collaboration with the Willow project-team.
- Bioinformatics: cancer diagnosis, protein function prediction, virtual screening. In collaboration with Institut Curie.
- Text processing: document collection modeling, language models.
- Audio processing: source separation, speech/music processing. In collaboration with Telecom Paris-tech.
- Neuro-imaging: brain-computer interface (fMRI, EEG, MEG). In collaboration with the Parietal project-team.

5. Software

5.1. SPAMS (SPArse Modeling Software)

Participants: Jean-Paul Chieze [correspondant], Guillaume Obozinski [correspondant].

SPAMS (SPArse Modeling Software) is an optimization toolbox for solving various sparse estimation problems: dictionary learning and matrix factorization, solving sparse decomposition problems, solving structured sparse decomposition problems. It is developed by Julien Mairal (former Willow PhD student, co-advised by F. Bach and J. Ponce), with the collaboration of Francis Bach (Inria), Jean Ponce (Ecole Normale Supérieure), Guillermo Sapiro (University of Minnesota), Rodolphe Jenatton (Inria) and Guillaume Obozinski (Inria). It is coded in C++ with a Matlab interface. Recently, interfaces for R and Python have been developed by Jean-Paul Chieze (Inria). Currently 650 downloads and between 1500 and 2000 page visits per month. See <http://spams-devel.gforge.inria.fr/>.

5.2. SiGMa - Simple Greedy Matching: a tool for aligning large knowledge-bases

Participant: Simon Lacoste-Julien [correspondant].

SiGMa - Simple Greedy Matching: a tool for aligning large knowledge-bases

Version 1. Webpage: <http://mlg.eng.cam.ac.uk/slacoste/sigma/>.

The tool SiGMa (Simple Greedy Matching) is a knowledge base alignment tool implemented in Python. It takes as input two knowledge bases, each represented as a list of triples of (entity, relationship, entity), in addition to a partial alignment between the relationships from one knowledge base to the other, and gives as output an ordered list of proposed entity matches between the two knowledge base (where the order corresponds heuristically to a notion of certainty about these matches). The matching decisions are made in a greedy fashion, combining information about the relationship graph as well as a pairwise similarity scores defined between the entities. The code handles various sources of information to be used for this score, such as a similarity defined on strings, dates, and other entity properties – and gives a few options to the user.

We also provide two large-scale knowledge base alignment benchmark datasets with tens of thousands of ground truth pairs: YAGO aligned to IMDb as well as Freebase aligned to IMDb.

Participants outside of Sierra: Konstantina Palla, Alex Davies, Zoubin Ghahramani (Machine Learning Group, Department of Engineering, University of Cambridge); Gjergji Kasneci, Thore Graepel (Microsoft Research Cambridge)

See <http://mlg.eng.cam.ac.uk/slacoste/sigma/>.

5.3. minFunc (2012 version)

Participant: Mark Schmidt [correspondant].

minFunc is a Matlab function for unconstrained optimization of differentiable real-valued multivariate functions using line-search methods. It uses an interface very similar to the Matlab Optimization Toolbox function `fminunc`, and can be called as a replacement for this function. On many problems, minFunc requires fewer function evaluations to converge than `fminunc` (or `minimize.m`). Further it can optimize problems with a much larger number of variables (`fminunc` is restricted to several thousand variables), and uses a line search that is robust to several common function pathologies.

The default parameters of minFunc call a quasi-Newton strategy, where limited-memory BFGS updates with Shanno-Phua scaling are used in computing the step direction, and a bracketing line-search for a point satisfying the strong Wolfe conditions is used to compute the step direction. In the line search, (safeguarded) cubic interpolation is used to generate trial values, and the method switches to an Armijo back-tracking line search on iterations where the objective function enters a region where the parameters do not produce a real valued output (i.e. complex, NaN, or Inf). See <http://www.di.ens.fr/~mschmidt/Software/minFunc.html>.

5.4. prettyPlot

Participant: Mark Schmidt [correspondant].

The prettyPlot function is a simple wrapper to Matlab's plot function for quickly making nicer-looking plots. Here are the features: Made the default line styles bigger, and the default fonts nicer. Options are passed as a structure, instead of through plot's large number of different functions. You can pass in cell arrays to have lines of different lengths. You can pass an $n \times 3$ matrix of colors, and cell arrays of line-styles and/or markers. It will cycle through the given choices. All markers are placed on top of (all) lines, you do not have to put a marker on every data point, and you can use different spacing between markers for different lines. You can change only the upper or lower x-limit (y-limit), rather than having to specify both. There is some support for making nicer-looking error lines. See <http://www.di.ens.fr/~mschmidt/Software/prettyPlot.html>.

5.5. SegAnnot

Participant: Toby Hocking [correspondant].

SegAnnot: an R package for fast segmentation of annotated piecewise constant signals. Tech report and R package. Standard segmentation models for piecewise constant signals do not always agree with an expert's visual interpretation of the signal, as encoded using a set of annotations. This R package implements a dynamic programming algorithm which can be used to quickly find a segmentation model in agreement with expert annotations. Collaboration with Guillem Rigau (Inria - AgroParisTech). See <http://hal.inria.fr/hal-00759129> and <http://segannot.r-forge.r-project.org/>.

6. New Results

6.1. A Stochastic Gradient Method with an Exponential Convergence Rate for Strongly-Convex Optimization with Finite Training Sets

Participants: Francis Bach, Mark Schmidt, Nicolas Le Roux [correspondant].

In [21], we propose a new stochastic gradient method for optimizing the sum of a finite set of smooth functions, where the sum is strongly convex. While standard stochastic gradient methods converge at sublinear rates for this problem, the proposed method incorporates a memory of previous gradient values in order to achieve a linear convergence rate. In a machine learning context, numerical experiments indicate that the new algorithm can dramatically outperform standard algorithms, both in terms of optimizing the training objective and reducing the testing objective quickly.

6.2. Convex Relaxation for Combinatorial Penalties

Participants: Francis Bach, Guillaume Obozinski [correspondant].

In [15], we propose an unifying view of several recently proposed structured sparsity-inducing norms. We consider the situation of a model simultaneously (a) penalized by a set-function defined on the support of the unknown parameter vector which represents prior knowledge on supports, and (b) regularized in L_p -norm. We show that the natural combinatorial optimization problems obtained may be relaxed into convex optimization problems and introduce a notion, the lower combinatorial envelope of a set-function, that characterizes the tightness of our relaxations. We moreover establish links with norms based on latent representations including the latent group Lasso and block-coding, and with norms obtained from submodular functions.

6.3. Kernel change-point detection

Participant: Sylvain Arlot [correspondant].

In [16], we tackle the change-point problem with data belonging to a general set. We propose a penalty for choosing the number of change-points in the kernel-based method of Harchaoui and Cappé (2007). This penalty generalizes the one proposed for one dimensional signals by Lebarbier (2005). We prove it satisfies a non-asymptotic oracle inequality by showing a new concentration result in Hilbert spaces. Experiments on synthetic and real data illustrate the accuracy of our method, showing it can detect changes in the whole distribution of data, even when the mean and variance are constant. Our algorithm can also deal with data of complex nature, such as the GIST descriptors which are commonly used for video temporal segmentation.

Collaboration with Alain Celisse (University Lille 1; Inria Lille, MODAL team) and Zaïd Harchaoui (Inria Grenoble, LEAR team).

6.4. On the Equivalence between Herding and Conditional Gradient

Algorithms

Participants: Francis Bach [correspondant], Simon Lacoste-Julien, Guillaume Obozinski.

In [5], we show that the herding procedure of Welling (2009) takes exactly the form of a standard convex optimization algorithm—namely a conditional gradient algorithm minimizing a quadratic moment discrepancy. This link enables us to invoke convergence results from convex optimization and to consider faster alternatives for the task of approximating integrals in a reproducing kernel Hilbert space. We study the behavior of the different variants through numerical simulations. The experiments indicate that while we can improve over herding on the task of approximating integrals, the original herding algorithm tends to approach more often the maximum entropy distribution, shedding more light on the learning bias behind herding.

6.5. V -fold cross-validation and V -fold penalization in least-squares density estimation

Participant: Sylvain Arlot [correspondant].

In [22], we study V -fold cross-validation for model selection in least-squares density estimation. The goal is to provide theoretical grounds for choosing V in order to minimize the least-squares risk of the selected estimator. We first prove a non asymptotic oracle inequality for V -fold cross-validation and its bias-corrected version (V -fold penalization), with an upper bound decreasing as a function of V . In particular, this result implies V -fold penalization is asymptotically optimal. Then, we compute the variance of V -fold cross-validation and related criteria, as well as the variance of key quantities for model selection performances. We show these variances depend on V like $1 + 1/(V - 1)$ (at least in some particular cases), suggesting the performances increase much from $V = 2$ to $V = 5$ or 10 , and then is almost constant. Overall, this explains the common advice to take $V = 10$ —at least in our setting and when the computational power is limited—, as confirmed by some simulation experiments.

Collaboration with Matthieu Lerasle (CNRS, University Nice Sophia Antipolis).

6.6. Machine learning for Neuro-imaging

Participants: Fabian Pedregosa [correspondant], Francis Bach, Guillaume Obozinski.

In the course of the year 2011-2012 two articles were submitted and accepted in international workshops. The first published article, **Improved brain pattern recovery through ranking approaches** ([12]) was presented at the 2nd International Workshop on Pattern Recognition in NeuroImaging in London, July 2012 and proposes a new approach for the problem of estimating the coefficients of a generalized linear model with monotonicity constraint. For this, we explore the use of ranking techniques, which are popular in the context of information retrieval but novel for medical imaging applications.

The second published article, **Learning to rank from medical imaging data** ([11]) uses the same techniques as the previous article to solve a more fundamental problem, that is, to predict a quantitative (and potentially non-linear) variable from a set of noisy measurements. We show on simulations and two fMRI datasets that this approach is able to predict the correct ordering on pairs of images, yielding higher prediction accuracy than standard regression and multiclass classification techniques.

Collaboration with the Parietal project-team (A. Gramfort, B. Thirion, G. Varoquaux)

6.7. SiGMa: Simple Greedy Matching for Aligning Large Knowledge Bases

Participant: Simon Lacoste-Julien [correspondant].

The Internet has enabled the creation of a growing number of large-scale knowledge bases in a variety of domains containing complementary information. Tools for automatically aligning these knowledge bases would make it possible to unify many sources of structured knowledge and answer complex queries. However, the efficient alignment of large-scale knowledge bases still poses a considerable challenge. In [20], we present Simple Greedy Matching (SiGMa), a simple algorithm for aligning knowledge bases with millions of entities and facts. SiGMa is an iterative propagation algorithm which leverages both the structural information from the relationship graph as well as flexible similarity measures between entity properties in a greedy local search, thus making it scalable. Despite its greedy nature, our experiments indicate that SiGMa can efficiently match some of the world's largest knowledge bases with high precision. We provide additional experiments on benchmark datasets which demonstrate that SiGMa can outperform state-of-the-art approaches both in accuracy and efficiency.

Collaboration with Konstantina Palla, Alex Davies, Zoubin Ghahramani (Machine Learning Group, Department of Engineering, University of Cambridge); Gjergji Kasneci (Max Planck Institut fur Informatik); Thore Graepel (Microsoft Research Cambridge).

6.8. Block-Coordinate Frank-Wolfe Optimization for Structural SVMs

Participants: Simon Lacoste-Julien [correspondant], Mark Schmidt.

In [19], we propose a randomized block-coordinate variant of the classic Frank-Wolfe algorithm for convex optimization with block-separable constraints. Despite its lower iteration cost, we show that it achieves the same convergence rate in duality gap as the full Frank-Wolfe algorithm. We also show that, when applied to the dual structural support vector machine (SVM) objective, this yields an online algorithm that has the same low iteration complexity as primal stochastic subgradient methods. However, unlike stochastic subgradient methods, the stochastic Frank-Wolfe algorithm allows us to compute the optimal step-size and yields a computable duality gap guarantee. Our experiments indicate that this simple algorithm outperforms competing structural SVM solvers.

Collaboration with Martin Jaggi (Centre de Mathématiques Appliquées, Ecole Polytechnique); Patrick Pletscher (Machine Learning Laboratory, ETH Zurich).

6.9. A convex relaxation for weakly supervised classifiers

Participants: Armand Joulin [correspondant], Francis Bach.

In [8], we introduce a general multi-class approach to weakly supervised classification. Inferring the labels and learning the parameters of the model is usually done jointly through a block-coordinate descent algorithm such as expectation-maximization (EM), which may lead to local minima. To avoid this problem, we propose a cost function based on a convex relaxation of the soft-max loss. We then propose an algorithm specifically designed to efficiently solve the corresponding semidefinite program (SDP). Empirically, our method compares favorably to standard ones on different datasets for multiple instance learning and semi-supervised learning, as well as on clustering tasks.

6.10. Multi-Class Cosegmentation

Participants: Armand Joulin [correspondant], Francis Bach.

Bottom-up, fully unsupervised segmentation remains a daunting challenge for computer vision. In the cosegmentation context, on the other hand, the availability of multiple images assumed to contain instances of the same object classes provides a weak form of supervision that can be exploited by discriminative approaches. Unfortunately, most existing algorithms are limited to a very small number of images and/or object classes (typically two of each). In [9], we propose a novel energy-minimization approach to cosegmentation that can handle multiple classes and a significantly larger number of images. The proposed cost function combines spectral- and discriminative-clustering terms, and it admits a probabilistic interpretation. It is optimized using an efficient EM method, initialized using a convex quadratic approximation of the energy. Comparative experiments show that the proposed approach matches or improves the state of the art on several standard datasets.

Collaboration with the Willow project-team (J. Ponce).

6.11. A latent factor model for highly multi-relational data

Participants: Nicolas Le Roux, Guillaume Obozinski [correspondant].

Many data such as social networks, movie preferences or knowledge bases are multi-relational, in that they describe multiple relations between entities. While there is a large body of work focused on modeling these data, modeling these multiple types of relations jointly remains challenging. Further, existing approaches tend to breakdown when the number of these types grows. In [7], we propose a method for modeling large multi relational datasets, with possibly thousands of relations. Our model is based on a bilinear structure, which captures various orders of interaction of the data, and also shares sparse latent factors across different relations. We illustrate the performance of our approach on standard tensor-factorization datasets where we attain, or outperform, state-of-the-art results. Finally, a NLP application demonstrates our scalability and the ability of our model to learn efficient and semantically meaningful verb representations.

Collaboration with R. Jenatton (CMAP, Ecole Polytechnique) and Antoine Bordes (CNRS, Université de Technologie de Compiègne).

6.12. Semi-supervised NMF with time-frequency annotations for single-channel source separation

Participants: Francis Bach, Augustin Lefèvre [correspondant].

In [10], we formulate a novel extension of nonnegative matrix factorization (NMF) to take into account partial information on source-specific activity in the spectrogram. Results on single-channel source separation show that time-frequency annotations allow to disambiguate the source separation problem, and learned annotations open the way for a completely unsupervised learning procedure for source separation with no human intervention.

Collaboration with C. Févotte (Laboratoire traitement et communication de l'information (LTCI), CNRS: UMR5141 - Institut Télécom - Télécom ParisTech).

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Grants with Industry

Participant: Francis Bach.

Google Research Award: "Large scale adaptive machine learning with finite data sets".

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

8.1.1.1. Calibration

Participant: Sylvain Arlot.

S. Arlot, Membre du projet ANR Calibration

Titre: Statistical calibration

Coordinator: University Paris Dauphine

Leader: Vincent Rivoirard

Other members: 34 members, mostly among CEREMADE (Paris Dauphine), Laboratoire Jean-Alexandre Dieudonné (Université de Nice) and Laboratoire de Mathématiques de l'Université Paris Sud

Instrument: ANR Blanc

Duration: Jan 2012 - Dec 2015

Total funding: 240 000 euros

Webpage: <https://sites.google.com/site/anrcalibration/>

8.1.1.2. Detect

Participants: Sylvain Arlot, Francis Bach, Rémi Lajugie.

Title: New statistical approaches to computer vision and bioinformatics

Coordinator: Ecole Normale Supérieure (Paris)

Leader of the project: Sylvain Arlot

Other members: J. Sivic (Willow project-team, ENS), A. Celisse (University Lille 1), T. Mary-Huard (AgroParisTech), E. Roquain and F. Villers (University Paris 6).

Instrument: ANR, Young researchers Program

Duration: Sep 2009 - Aug 2012

Total funding: 70000 Euros

See also: <http://www.di.ens.fr/~arlot/ANR-DETECT.htm>

Abstract: The Detect project aims at providing new statistical approaches for detection problems in computer vision (in particular, detecting and recognizing human actions in videos) and bioinformatics (e.g., simultaneously segmenting CGH profiles). These problems are mainly of two different statistical nature: multiple change-point detection (i.e., partitioning a sequence of observations into homogeneous contiguous segments) and multiple tests (i.e., controlling a priori the number of false positives among a large number of tests run simultaneously).

8.2. European Initiatives

8.2.1. FP7 Projects

8.2.1.1. SIERRA

Participants: Francis Bach [correspondant], Simon Lacoste-Julien, Augustin Lefèvre, Nicolas Le Roux, Mark Schmidt.

Title: SIERRA – Sparse structured methods for machine learning

Type: IDEAS

Instrument: ERC Starting Grant (Starting)

Duration: December 2009 - November 2014

Coordinator: Inria (France)

See also: <http://www.di.ens.fr/~fbach/sierra>

Abstract: Machine learning is now a core part of many research domains, where the abundance of data has forced researchers to rely on automated processing of information. The main current paradigm of application of machine learning techniques consists in two sequential stages: in the representation phase, practitioners first build a large set of features and potential responses for model building or prediction. Then, in the learning phase, off-the-shelf algorithms are used to solve the appropriate data processing tasks. While this has led to significant advances in many domains, the potential of machine learning techniques is far from being reached: the tenet of this proposal is that to achieve the expected breakthroughs, this two-stage paradigm should be replaced by an integrated process where the

8.3. International Initiatives

8.3.1. Inria Associate Teams

8.3.1.1. STATWEB

Participants: Francis Bach [correspondant], Ronny Luss.

Title: Fast Statistical Analysis of Web Data via Sparse Learning

Inria principal investigator: Francis Bach

International Partner (Institution - Laboratory - Researcher):

University of California Berkeley (United States) - EECS and IEOR Departments - Laurent El Ghaoui

Duration: 2011 - 2013

See also: <http://www.di.ens.fr/~fbach/statweb.html>

The goal of the proposed research is to provide web-based tools for the analysis and visualization of large corpora of text documents, with a focus on databases of news articles. We intend to use advanced algorithms, drawing from recent progresses in machine learning and statistics, to allow a user to quickly produce a short summary and associated timeline showing how a certain topic is described in news media. We are also interested in unsupervised learning techniques that allow a user to understand the difference between several different news sources, topics or documents.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

Michael Jordan (U.C. Berkeley, <http://www.cs.berkeley.edu/~jordan>), is spending one year in our team, starting September 2012, financed by the Fondation de Sciences Mathématiques de Paris and Inria.

9. Dissemination

9.1. Scientific Animation

9.1.1. Editorial boards

F. Bach: Journal of Machine Learning Research, Action Editor.

F. Bach: IEEE Transactions on Pattern Analysis and Machine Intelligence, Associate Editor.

F. Bach: Information and Inference, Associate Editor.

F. Bach: SIAM Journal on Imaging Sciences, Associate Editor.

G. Obozinski: Journal of Machine Learning Research, Member of the Editorial Board.

9.1.2. Area chairs

G. Obozinski: International conference on Artificial Intelligence and Statistics (AISTATS) 2012.

F. Bach: International Conference on Machine Learning, 2012.

S. Lacoste-Julien, F. Bach: Conference on Uncertainty in Artificial Intelligence, 2012.

9.1.3. Reviewing

Journals: Annals of Statistics, Machine Learning, Journal of Machine Learning Research (JMLR), IEEE Transaction on Information Theory, Transaction in Pattern Recognition and Machine Intelligence (TPAMI), Information and Inference (IMAI), Scandinavian Journal of Statistics (SJS), Statistics and Computing (STO), Annales de l'IHP, Annals of Statistics

Conferences: UAI, ECML, NIPS, CVPR, ICML, COLT, AISTATS.

9.1.4. Other

S. Arlot is member of the board for the entrance exam in Ecole Normale Supérieure (mathematics, voie B/L).

9.1.5. Workshop and conference organization

M. Schmidt, Session Organizer at International Conference on Continuous Optimization (July 27 - August 1, 2013).

G. Obozinski, Co-organiser of the workshop "Sparsity, Dictionaries and Projections in Machine Learning and Signal Processing" at ICML 2012, Edinburgh, Scotland. <http://www.di.ens.fr/~obozinski/ICML2012workshop/>.

F. Bach: International Conference on Machine Learning, 2012, Workshop co-chair.

F. Bach: Co-organizer of the NIPS workshop on "Analysis Operator Learning vs. Dictionary Learning: Fraternal Twins in Sparse Modeling", <https://sites.google.com/site/dlaoplnips2012/>.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence : S. Arlot, F. Bach, G. Obozinski, "Apprentissage statistique", 35h, Ecole Normale Supérieure, Filière "Math-Info", première année.

Licence: G. Obozinski, Introduction aux modèles graphiques (4h) in Enseignement spécialisé "Apprentissage artificiel" for second year students at Ecole des Mines.

Mastère: S. Arlot and F. Bach, "Statistical learning", 24h, Mastère M2, Université Paris-Sud, France.

Mastère: G. Obozinski, N. Le Roux, Introduction à l'apprentissage machine appliqué aux neurosciences et à la cognition Co-teaching (6h) in the course of the Master Recherche en Sciences Cognitives co-habilitated by EHESS, ENS and Université Paris Descartes.

Mastère: F. Bach, G. Obozinski, Introduction aux modèles graphiques (30h), Master MVA (Ecole Normale Supérieure de Cachan).

Doctorat: S. Arlot, "Model selection via penalization, resampling and cross-validation, with application to change-point detection", 6h, Université de Cergy.

Doctorat: G. Obozinski, Probabilistic graphical models for Information Retrieval, in the Russian Summer School in Information Retrieval (RuSSIR 2012), Yaroslavl, Russia.

Doctorat: F. Bach, International Computer Vision Summer School 2012, 3h, Sicily.

Doctorat: F. Bach, Summer school on Visual Recognition and Machine Learning, 3h, Grenoble.

Doctorat: F. Bach: Machine learning summer school (MLSS), 3h, Kyoto, Japan.

9.2.2. Supervision

PhD : Toby Hocking, "Learning algorithms and statistical software, with applications to bioinformatics", ENS Cachan, November 20, 2012, Advisors: F. Bach, J.-P. Vert (Ecole des Mines de Paris - Institut Curie).

PhD: Augustin Lefèvre, "Dictionary learning methods for single-channel audio source separation", ENS Cachan, October 3, 2012, Advisors: F. Bach and C. Févotte (Telecom Paristech).

PhD: Armand Joulin, "Convex optimization for co-segmentation", ENS Cachan, December 17, 2012, Advisors: F. Bach and J. Ponce (Willow project-team).

9.3. Invited presentations

S. Arlot, "Optimal model selection with V-fold cross-validation: how should V be chosen?", World Congress in Probability and Statistics 2012, Istanbul.

- S. Arlot, "Resampling-based estimation of the accuracy of satellite ephemerides", Inaugural Conference of the Laboratory Fibonacci, Scuola Normale Superiore di Pisa, 2012.
- S. Arlot, "Choix de V pour la sélection de modèles par validation croisée V-fold en estimation de densité", Séminaire parisien de statistique, IHP, Paris, 2012.
- S. Arlot, "Calibration automatique d'estimateurs linéaires à l'aide de pénalités minimales, application à la régression multi-tâches.", Séminaire de Statistique de l'IMT, Toulouse, 2012.
- F. Bach, International Conference on Pattern Recognition Applications and Methods, Faro, Portugal, 2012 (keynote speaker).
- F. Bach, Rank Prize Symposium (invited talk), Lake District, England, 2012.
- F. Bach, University of Cambridge (two seminars), 2012.
- F. Bach, Schlumberger workshop on Mathematical Models of Sound Analysis, IHES (invited talk), 2012.
- F. Bach, Joint Pattern Recognition Symposium of the German Association for Pattern Recognition (DAGM) (invited talk), Graz, Austria, 2012.
- F. Bach, International Workshop on Machine Learning for Signal Processing (plenary lecture), Santander, Spain, 2012.
- F. Bach, Seminar Max-Planck Institute, Tübingen, October 2012.
- E. Grave, Laboratoire d'Informatique de Paris 6, Université Pierre et Marie Curie (Seminar), 2012.
- S. Lacoste-Julien, "Harnessing the structure of data for discriminative machine learning", Colloque du Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montréal, Canada, February 2012
- S. Lacoste-Julien, "Structured Alignment Methods in Machine Learning", SMT seminar at the LIMSI, Orsay, France, July 2012
- S. Lacoste-Julien, "Frank-Wolfe optimization insights in machine learning", machine learning seminar, University of Toronto, Toronto, Canada, August 2012.
- S. Lacoste-Julien, "Frank-Wolfe optimization insights in machine learning", Machine Learning Group seminar, University of Cambridge, Cambridge, UK, August 2012.
- S. Lacoste-Julien, "Harnessing the structure of data in machine learning", invited talk, Department of Engineering Science, University of Oxford, Oxford, UK, September 2012.
- S. Lacoste-Julien, "Frank-Wolfe optimization insights in machine learning", invited talk, Stanford AI Lab, Stanford University, Stanford, USA, December 2012.
- G. Obozinski, Swiss Statistical Seminar, Bern, Switzerland, April 2012.
- G. Obozinski, Séminaire de Statistiques, Université de Pennsylvanie, Philadelphia, PA, USA, May 2012.
- G. Obozinski, Séminaire de Statistiques, Université Paris 11, May 2012.
- G. Obozinski, Journées de Statistiques (Conférence annuelle de la Société Française de Statistiques), Université Libre de Bruxelles, Belgium, May 2012.
- G. Obozinski, Congrès mondial de Probabilités et Statistiques, Istanbul, Turkey, July 2012.
- G. Obozinski, séminaire du CEREMADE, Université Paris-Dauphine, December 2012.
- M. Schmidt, NAIS Workshop on Advances in Large-Scale Optimization, Edinburgh, May 24-25, 2012.
- M. Schmidt, International Symposium on Mathematical Programming, Berlin, August 19-24, 2012.
- M. Schmidt, University of British Columbia, "Linearly-Convergent Stochastic-Gradient Methods", Seminar, December 10, 2012.
- M. Schmidt, - Simon Fraser University, "Opening up the black box: Faster methods for non-smooth and big-data optimization", Seminar, December 11, 2012.

9.4. Prizes and awards

M. Schmidt, NSERC Postdoctoral Fellowship (January 2012 - December 2013).

S. Lacoste-Julien, Research in Paris fellowship 2011-2012.

F. Bach: Inria young researcher prize, 2012.

R. Jenatton: Thesis prize from Fondation Hadamard, 2012.

R. Jenatton: Thesis prize from AFIA, accessit, 2012.

9.5. Popularization

Participation to Inria-Rocquencourt "Fête de la Science", 2012.

10. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journals

- [1] F. BACH, R. JENATTON, J. MAIRAL, G. OBOZINSKI. *Structured sparsity through convex optimization*, in "Statistical Science", 2012, vol. 27, n^o 4, p. 450-468, <http://hal.inria.fr/hal-00621245>.
- [2] M. P. FRIEDLANDER, M. SCHMIDT. *Hybrid Deterministic-Stochastic Methods for Data Fitting*, in "SIAM Journal on Scientific Computing", 2012, vol. 34, n^o 3, p. A1380-A1405, 22 pages [DOI : 10.1137/110830629], <http://hal.inria.fr/inria-00626571>.
- [3] R. JENATTON, A. GRAMFORT, V. MICHEL, G. OBOZINSKI, E. EGER, F. BACH, B. THIRION. *Multi-scale Mining of fMRI data with Hierarchical Structured Sparsity*, in "SIAM Journal on Imaging Sciences", July 2012, vol. 5, n^o 3, p. 835-856 [DOI : 10.1137/110832380], <http://hal.inria.fr/inria-00589785>.
- [4] M. SOLNON, S. ARLOT, F. BACH. *Multi-task Regression using Minimal Penalties*, in "Journal of Machine Learning Research", September 2012, vol. 13, p. 2773-2812, <http://hal.inria.fr/hal-00610534>.

International Conferences with Proceedings

- [5] F. BACH, S. LACOSTE-JULIEN, G. OBOZINSKI. *On the Equivalence between Herding and Conditional Gradient Algorithms*, in "ICML 2012 International Conference on Machine Learning", Edimburgh, United Kingdom, 2012, <http://hal.inria.fr/hal-00681128>.
- [6] D. BUCHMAN, M. SCHMIDT, S. MOHAMED, D. POOLE, N. DE FREITAS. *On Sparse, Spectral and Other Parameterizations of Binary Probabilistic Models*, in "AISTATS 2012 - 15th International Conference on Artificial Intelligence and Statistics", La Palma, Spain, 2012, <http://hal.inria.fr/hal-00717714>.
- [7] R. JENATTON, N. LE ROUX, A. BORDES, G. OBOZINSKI. *A latent factor model for highly multi-relational data*, in "Neural Information Processing Systems (NIPS)", Lake Tahoe, Nevada, États-Unis, 2012, <http://hal.inria.fr/hal-00776335>.
- [8] A. JOULIN, F. BACH. *A convex relaxation for weakly supervised classifiers*, in "ICML 2012 International Conference on Machine Learning", Edinburgh, United Kingdom, June 2012, 641, <http://hal.inria.fr/hal-00717450>.

- [9] A. JOULIN, F. BACH, J. PONCE. *Multi-Class Cosegmentation*, in "CVPR 2012 : 25th IEEE Conference on Computer Vision and Pattern Recognition", Providence, United States, June 2012, 0109, <http://hal.inria.fr/hal-00717448>.
- [10] A. LEFÈVRE, F. BACH, C. FÉVOTTE. *Semi-supervised NMF with time-frequency annotations for single-channel source separation*, in "ISMIR 2012 : 13th International Society for Music Information Retrieval Conference", Porto, Portugal, October 2012, <http://hal.inria.fr/hal-00717366>.
- [11] F. PEDREGOSA, A. GRAMFORT, G. VAROQUAUX, E. CAUVET, C. PALLIER, B. THIRION. *Learning to rank from medical imaging data*, in "MLMI 2012 - 3rd International Workshop on Machine Learning in Medical Imaging", Nice, France, Inria, July 2012, <http://hal.inria.fr/hal-00717990>.
- [12] F. PEDREGOSA, A. GRAMFORT, G. VAROQUAUX, B. THIRION, C. PALLIER, E. CAUVET. *Improved brain pattern recovery through ranking approaches*, in "PRNI 2012 : 2nd International Workshop on Pattern Recognition in NeuroImaging", London, United Kingdom, July 2012, <http://hal.inria.fr/hal-00717954>.

Research Reports

- [13] R. JENATTON, R. GRIBONVAL, F. BACH. *Local stability and robustness of sparse dictionary learning in the presence of noise*, HAL-Inria, October 2012, 41, <http://hal.inria.fr/hal-00737152>.
- [14] H. KADRI, A. RAKOTOMAMONJY, F. BACH, P. PREUX. *Multiple Operator-valued Kernel Learning*, Inria, March 2012, n^o RR-7900, <http://hal.inria.fr/hal-00677012>.
- [15] G. OBOZINSKI, F. BACH. *Convex Relaxation for Combinatorial Penalties*, HAL-Inria, May 2012, 35 page, <http://hal.inria.fr/hal-00694765>.

Other Publications

- [16] S. ARLOT, A. CELISSE, Z. HARCHAOUI. *Kernel change-point detection*, 2012, <http://hal.inria.fr/hal-00671174>.
- [17] F. BACH. *Sharp analysis of low-rank kernel matrix approximations*, 2012, <http://hal.inria.fr/hal-00723365>.
- [18] T. D. HOCKING, G. SCHLEIERMACHER, I. JANOUÉIX-LEROSEY, O. DELATTRE, F. BACH, J.-P. VERT. *Learning smoothing models of copy number profiles using breakpoint annotations*, 2012, <http://hal.inria.fr/hal-00663790>.
- [19] S. LACOSTE-JULIEN, M. JAGGI, M. SCHMIDT, P. PLETSCHER. *Stochastic Block-Coordinate Frank-Wolfe Optimization for Structural SVMs*, 2012, 13 pages main text + 13 pages appendix (short version). Under review, <http://hal.inria.fr/hal-00720158>.
- [20] S. LACOSTE-JULIEN, K. PALLA, A. DAVIES, G. KASNECI, T. GRAEPEL, Z. GHAHRAMANI. *SiGMA: Simple Greedy Matching for Aligning Large Knowledge Bases*, July 2012, 10 pages + 2 pages appendix; 5 figures – initial preprint, <http://hal.inria.fr/hal-00768180>.
- [21] N. LE ROUX, M. SCHMIDT, F. BACH. *A Stochastic Gradient Method with an Exponential Convergence Rate for Strongly-Convex Optimization with Finite Training Sets*, 2012, <http://hal.inria.fr/hal-00674995>.

- [22] M. LERASLE, S. ARLOT. *V-fold cross-validation and V-fold penalization in least-squares density estimation*, 2012, <http://hal.inria.fr/hal-00743931>.