Carnegie Mellon University


# PhD in Machine Learning

**Course Requirements**
The curriculum for the Machine Learning Ph.D. is built on a foundation of **five core** courses and **three electives** (plus the **Data Analysis Project** requirement). These five courses also comprise the required courses for the MS degree. Together with the Data Analysis Project requirement, these should be completed for MS degree within 3 years but many students do it in 2 or 2.5 years.

Note: If a student has taken some of the MLD core courses before joining the MLD PhD program, and has not counted these courses toward any other PhD-level degree, the student may count these courses toward the MLD PhD. In this situation the student will need to take fewer than 5 new core courses to graduate. A student must always take at least three elective courses while registered in the MLD PhD program, irrespective of any courses taken before joining the PhD program. Students who took 10-701 in Spring 2014 or earlier can use it as a core course, even if they weren't part of the MLD PhD program at the time they took 10-701.

A typical full-time, graduate course load during the first two years consists each term of two classes (at 12 graduate units per class) plus 24 units of advanced research. Thus, during the first two years, a student has the opportunity to take several elective classes in addition to the five required courses.

The ML curriculum joins courses with a Computer Science main theme and those with a Probability and Statistics main theme. These may be grouped, as follows:

In CS, relevant sub-fields include: Databases; Machine Learning, Data Mining and algorithms applications in areas such as Robotics, Information Retrieval and AI.

In Statistics (including Philosophy), the sub-fields include: Statistical modeling (e.g., hierarchical and times series); Bayes' Nets, Causation, and experimental design. The curriculum is based on core academic courses on Intermediate Statistics, Machine Learning, Statistical Machine Learning, Multimedia Databases, and Algorithms.

These core courses together provide a foundation in machine learning, statistics, probability, and algorithms.

- 10-715 Advanced Introduction to Machine Learning
- 10-702 Statistical Machine Learning
- 10-705 Intermediate Statistics

Plus any two of the following courses:

- 10-708 Probabilistic Graphical Models
- 10-725 Convex Optimization
- 15-826 Multimedia Databases and Data Mining
- 15-750 Algorithms or 15-853 Algorithms in the Real World

Possible electives

Data Analysis Project Requirement, in the second year, which serves in lieu of an MS thesis.

Here is a typical schedule for the first two years of study.

| **FALL - 1st Year** | **SPRING - 1st Year** |
|---|---|
| 10-715 Advanced Introduction to Machine Learning | 10-702 Statistical Machine Learning |
| 10-705 Intermediate Statistics | Core course or elective |
| 10-920 Research | 10-920 Research |
| **FALL - 2nd Year** | **SPRING - 2nd Year** |

| | |
|---|---|
| Core course or Elective | Research for Data Analysis Project |
| Elective | Elective |
| 10-920 Research | 10-920 Research |
| | |

**The Data Analysis Project requirement:**
During the second year a Ph.D. student is required to demonstrate data analysis and machine learning skills in the context of a focused project. The Data Analysis Project may be carried out either at Carnegie Mellon or at a sponsoring corporate institution under the joint supervision of the sponsor and a ML faculty. It will be concluded by a written report (in lieu of a Masters Thesis) in which the student demonstrates an ability to approach data mining problems in a way that cuts across existing disciplinary boundaries. The requirement includes a presentation in the ML Journal Club and also the submission of a DAP Paper. Passing this requirement will be the judgment of the DAP committee.

Student must form an official "DAP committee" of three faculty to evaluate the document. The committee will consist of the advisor, the Journal club instructor(s), and one other faculty member selected by the student. The third member is typically someone with an interest in the analysis of the data set, and does not have to be an expert in ML or part of the student's thesis committee.The student should form the committee as early as possible during the DAP research process, and inform Diane of who the members are. Two faculty from the committee are required to attend the presentation.

**The Third Year**
During the third year, a Ph.D. student completes the elective course requirements. One of these three electives is taken from the offerings in Statistics. The other two advanced electives, chosen in consultation with the students advisor, form a concentration in one of the allied disciplines with SCS, Biology, Philosophy, or GSIA. For those candidates seeking an academic position after completing the ML Ph.D. degree, the thoughtful selection of these three elective courses is particularly important. As in the each of the first two years, coursework is supplemented by 24 units/term of research.

**The Fourth Year and Beyond**
A Ph.D. student typically presents a thesis proposal no later than the start of the fourth year, and then spends the fourth and sometimes fifth year working on their thesis research.

**Research**
Responsible Conduct of Research Training
**Students must complete training for NSF and NIH grants. A copy of the certificate must be given to the MLD Business Manager, Colleen Everett.**

It is expected that all Ph.D. students engage in active research from their first semester. Moreover, advisor selection occurs in the first month of entering the Ph.D. program, with the option to change at a later time. Roughly half of a student's time should be allocated to research and lab work, and half to courses until these are completed.

This **Research page** is a list of some of the projects for which ML faculty may be interested in recruiting students. Within each project there can be lines of research which range in size from a semester's work to an entire thesis (or beyond). So, this page is intended as a resource for students looking for a thesis advisor, for a Data Analysis project, or to collaborate for any other reason.

Carnegie Mellon University

MACHINE LEARNING
DEPARTMENT

# Relevant Courses offered for Spring 2015

## CORE COURSES:

10-701 Machine Learning
10-702 Statistical Machine Learning
10-708 Probabilistic Graphical Models
10-725 Convex Optimization
15-750 Graduate Algorithms

## ML JOURNAL CLUB

10-915, ML PhD students must register if you plan to satisfy either the Speaking Skills or Data Analysis Project (DAP) requirements.

Incoming PhD students should register in Spring.

Research Courses:
10-920 Grad Reading & Research
10-930 Dissertation Research

Students should register for 10-920 R & R until they propose.
After you propose, register for Dissertation Research.

## Suggested Research Depth Electives:

***For ML PhD students, two advanced electives, chosen in consultation with the student's advisor, form a research depth concentration. Approved Research Depth electives are listed below.***
***Full list of*** *Approved Electives*

**Research Depth in AI:**
10-708 Probabilistic Graphical Models
10-725 Convex Optimization
15-780 Graduate Artificial Intelligence

**Research Depth in CNBC Track:**
03-762 Systems Neuroscience
36-759 Statistical Models of the Brain
**Applicable Courses from the University of Pittsburgh**
*http://www.cmu.edu/hub/registration/undergraduates/cross/outgoing.html*

**Research Depth in Algorithms & Theory**
10-725 Convex Optimization
15-859 (E) Special Topics in Theory: Advanced Algorithms

**Research Depth in NLP or Text Analysis:**
10-708 Probabilistic Graphical Models
11-741 Machine Learning for Text Mining
11-745 Adv. Statistical Learning Seminar (6)
11-761 Language and Statistics

**Research Depth in Computational Biology:**
02-710 Computational Genomics
10-708 Probabilistic Graphical Models

**Research Depth in Computer Vision:**
10-725 Convex Optimization
16-720 Computer Vision
16-822 Geometry-Based Methods in Vision
16-824 Learning Based Methods in Vision

## Suggested Electives from Statistics

(For ML PhD Students, one elective or courses combined for a total of 12 units must be chosen from Statistics)
36-728 Time Series
36-752 Adv. Probability Overview

## Other electives from SCS approved but don't have a category:

10-704 Information Processing & Learning
11-755 Machine Learning for Signal Processing
18-755 Networks in the Real World

Carnegie Mellon University


---

# Electives

## ML PhD students:

One elective or courses combined for a total of 12 units must be chosen from Statistics.
You must also have a research depth of 24 units.

*Both Statistics and Tepper offer "mini" half-term courses. Two such "mini" courses are equivalent to one (12 unit) graduate course.*

## Suggested Electives from Statistics

36-703 Intermediate Probability
36-707 Regression Analysis
36-708 Experimental Design, 6 units, A4 mini
36-709 Linear Models, 6 units, A3 mini
36-720 Discrete Multivariate Analysis, 6 units, A2 mini
36-722 Applied Continuous Multivariate Analysis, 6 units, A4 mini
36-724 Applied Bayesian Methods, 6 units, A3 mini
36-728 Time Series
36-737 Applied Multilevel & Hierarchical Models, 6 units, A2 mini
36-752 Adv. Probability Overview
36-754 Adv. Probability
36-755 Advanced Statistical Theory I
36-759 Statistical Models of the Brain
36-781 Advanced Statistical Methods I mini, 6 units, A1 mini
36-782 Advanced Statistical Methods II mini, 6 units, A2 mini
36-786 Bayesian Theoretical Statistics 1, 6 units, A1 mini
36-787 Bayesian Theoretical Statistics 2, 6 units, A2 mini
36-825 Statistics Journal Club
36-835 Foundations of Statistics Seminar
36-900 Selected Topics of the Contemporary Frontiers of High Dimensional Inference
36-905 Seminar on Latent Variable Models, 6 units

## Suggested Depth Requirement Electives from SCS

**AI:**
10-725 Convex Optimization
10-708 Probablistic Graphical Models
15-780 Graduate Artificial Intelligence
15-857 Analytical Performance Modeling & Design of Computer Systems
15-859 (M) Randomized Algorithms
15-887 Planning, Execution, and Learning

**Algorithms & Theory:**
10-725 Convex Optimization
15-855 Computational Complexity Theory
15-857 Analytical Performance Modeling & Design of Computer Systems
15-859 Special Topics in Theory - check for appropriate topics
15-859 (B) Machine Learning Theory
15-859 (E) Special Topics in Theory: Advanced Algorithms
16-811 Mathematical Fundamentals for Robotics
21-801 Adv. Topics Discrete Math (Random Graphs)

**Computational Biology:**
02-750 Automation of Biological Research: Robotics & Machine Learning
02-710 Computational Genomics
10-708 Probablistic Graphical Models

**Computer Vision:**
10-725 Convex Optimization
16-720 Computer Vision
16-822 Geometry-Based Methods in Vision

16-824 Learning Based Methods in Vision

**Databases:**
15-823 Advanced Database Topics

**NLP or Text Analysis:**
10-708 Probablistic Graphical Models
10-710/11-763 Structured Prediction for Language & Other Discrete Data
10-802/11-772 Analysis of Social Media
11-711 Algorithms for NLP
11-741 Machine Learning for Text Mining
11-744 Experimental Information Retrieval
11-745 Advanced Statistical Learning Seminar (6 units)
11-761 Language and Statistics
11-762 Language and Statistics II
11-773 Text-Driven Forecasting

**Robotics:**
02-750 Automation of Biological Research: Robotics & Machine Learning
15-887 Planning, Execution, and Learning
16-811 Mathematical Fundamentals for Robotics
16-831 Statistical Techniques in Robotics
16-899C Adaptive Control and Reinforcement Learning

**Other electives from SCS approved but don't have a Depth Requirement category:**
10-704 Information Processing & Learning
11-745 Adv. Statistical Learning Seminar (6 units)
11-755 Machine Learning for Signal Processing
15-830 Computational Methods in Sustainable Energy
18-755 Networks in the Real World

## Suggested Depth Requirement Electives for CNBC Track

03-762 Advanced Cellular Neuroscience
03-763 Systems Neuroscience
15-883 Computational Models of Neural Systems
36-759 Statistical Models of the Brain
85-719 Introduction to Parallel Distributed Processing
85-765 Cognitive Neuroscience (12 units)

*Applicable Courses from the University of Pittsburgh (Please see http://www.cmu.edu/hub/registration/undergraduates/cross/outgoing.html)*
NROSCI 2100 Cellular & Molecular Neurobiology
NROSCI 2102/2103 Systems Neurobiology
MATH 3375 Computational Neuroscience

## Suggested Concentration Electives from School of Public Policy & Management:

10-830/90-904 Research Seminar in Machine Learning & Policy, 6 units, A3 mini
10-831/90-921 Special topics in Machine Learning & Policy, 6 units, A4 mini

## Suggested Concentration Electives from Tepper (Must follow Tepper special registration rules)

**Finance Track:**
45-814 Options
46-926 Linear Models/Equity Portfolio Management
46-929 Financial Time Series Analysis
46-944 Stochastc Calc Fin 1

**Marketing Track:**
15-892 Foundations of Electronic Marketplaces (CS course)
47-800 Intermediate Microeconomic Analysis
47-741 Seminar in Marketing I
47-742 Seminar in Marketing II
47-743 Seminar in Marketing III
47-744 Analytical and Structural Marketing Models
45-821 Marketing with Electronic & Social Media
45-824 Database Marketing

**Information Systems Track:**
47-800 Intermediate Microeconomic Analysis
45-870 Management of Information Systems
45-871 Information Strategy, Systems and Economics
47-951 Seminar in Information Systems I
47-952 Seminar in Information Systems II
47-953 Seminar in Information Systems III
47-954 Seminar in Information Systems IV

**NOTE: Tepper courses are on the mini-system.45-* and 46-* are Master level courses and the 47-**

*** are PhD level courses Suggested**

## Concentration Electives from Philosophy

80-605 Rational Choice
80-614 Logic in Artificial Intelligence
80-616 Probability and Artificial Intelligence
80-621 Causality in the Social Sciences

Carnegie Mellon University


---

# PhD Dissertations

## [All are .pdf files]

Learning Statistical Features of Scene Images
**Wooyoung Lee, 2014**

Towards Scalable Analysis of Images and Videos
**Bin Zhao, 2014**

Statistical Text Analysis for Social Science
**Brendan T. O'Connor, 2014**

Modeling Large Social Networks in Context
**Qirong Ho, 2014**

Semi-Cooperative Learning in Smart Grid Agents
**Prashant P. Reddy, 2013**

On Learning from Collective Data
**Liang Xiong, 2013**

Exploiting Non-sequence Data in Dynamic Model Learning
**Tzu-Kuo Huang, 2013**

Mathematical Theories of Interaction with Oracles
**Liu Yang, 2013**

Cortical spatiotemporal plasticity in visual category learning
**Yang Xu, 2013**

Short-Sighted Probabilistic Planning
**Felipe W. Trevizan, 2013**

Statistical Models and Algorithms for Studying Hand and Finger Kinematics and their Neural Mechanisms
**Lucia Castellanos, 2013**

Approximation Algorithms and New Models for Clustering and Learning
**Pranjal Awasthi, 2013**

Uncovering Structure in High-Dimensions: Networks and Multi-task Learning Problems
**Mladen Kolar, 2013**

Learning with Sparsity: Structures, Optimization and Applications
**Xi Chen, 2013**

GraphLab: A Distributed Abstraction for Large Scale Machine Learning
**Yucheng Low, 2013**

Graph Structured Normal Means Inference
**James Sharpnack, 2013 (Joint Statistics & ML PhD)**

Probabilistic Models for Collecting, Analyzing, and Modeling Expression Data
**Hai-Son Phuoc Le, 2013**

Learning Large-Scale Conditional Random Fields
**Joseph K. Bradley, 2013**

New Statistical Applications for Differential Privacy
**Rob Hall, 2013 (Joint Statistics & ML PhD)**

Parallel and Distributed Systems for Probabilistic Reasoning
**Joseph Gonzalez, 2012**

Spectral Approaches to Learning Predictive Representations
**Byron Boots, 2012**

Attribute Learning using Joint Human and Machine Computation
**Edith L. M. Law, 2012**

Statistical Methods for Studying Genetic Variation in Populations
**Suyash Shringarpure, 2012**

Data Mining Meets HCI: Making Sense of Large Graphs
**Duen Horng (Polo) Chau, 2012**

Learning with Limited Supervision by Input and Output Coding
**Yi Zhang, 2012**

Target Sequence Clustering
**Benjamin Shih, 2011**

Nonparametric Learning in High Dimensions
**Han Liu, 2010 (Joint Statistics & ML PhD)**

Structural Analysis of Large Networks: Observations and Applications
**Mary McGlohon, 2010**

Modeling Purposeful Adaptive Behavior with the Principle of Maximum Causal Entropy
**Brian D. Ziebart, 2010**

Tractable Algorithms for Proximity Search on Large Graphs
**Purnamrita Sarkar, 2010**

Rare Category Analysis
**Jingrui He, 2010**

Coupled Semi-Supervised Learning
**Andrew Carlson, 2010**

Fast Algorithms for Querying and Mining Large Graphs
**Hanghang Tong, 2009**

Efficient Matrix Models for Relational Learning
**Ajit Paul Singh, 2009**

Exploiting Domain and Task Regularities for Robust Named Entity Recognition
**Andrew O. Arnold, 2009**

Theoretical Foundations of Active Learning
**Steve Hanneke, 2009**

Generalized Learning Factors Analysis: Improving Cognitive Models with Machine Learning
**Hao Cen, 2009**

Detecting Patterns of Anomalies
**Kaustav Das, 2009**

Dynamics of Large Networks
**Jurij Leskovec, 2008**

Computational Methods for Analyzing and Modeling Gene Regulation Dynamics
**Jason Ernst, 2008**

Stacked Graphical Learning
**Zhenzhen Kou, 2007**

Actively Learning Specific Function Properties with Applications to Statistical Inference
**Brent Bryan, 2007**

Approximate Inference, Structure Learning and Feature Estimation in Markov Random Fields
**Pradeep Ravikumar, 2007**

Scalable Graphical Models for Social Networks
**Anna Goldenberg, 2007**

Measure Concentration of Strongly Mixing Processes with Applications
**Leonid Kontorovich, 2007**

Tools for Graph Mining
**Deepayan Chakrabarti, 2005**

Automatic Discovery of Latent Variable Models
**Ricardo Silva, 2005**

Carnegie Mellon University

MACHINE LEARNING
DEPARTMENT

---

# Student Data Analysis Projects

Students are required to demonstrate their grasp of fundamental data analysis and machine learning concepts and techniques in the context of a focused project. The project should focus on a substantive problem involving the analysis of one or more data sets and the application of state-of-the art machine learning and data mining methods, or on suitable simulations where this is deemed appropriate. Or, the project may focus on machine learning methodology and demonstrate its applicability to substantial examples from the relevant literature. The project may involve the development of new methodology or extensions to existing methodology.

A General Approach to Prediction and Forecasting Crime Rates with Gaussian Processes [.pdf] -Seth Flaxman, 5/14

Time-varying Linear Regression with Total Variation Regularization [.pdf] - Matthew Wytock, 5/14

An integrated approach to validating ChIP-Seq using A* Lasso for Sparse Bayesian Network Learning [.pdf] - Jing Xiang, 4/14

Robust Data-Driven State Estimation for Smart Grid [.pdf] - Yang Weng, 3/14

Prioritizing Malware Analysis [.pdf] - Zhen Tang, 2/14

Automated Discovery of Novel Anomalous Patterns [.pdf] - Edward McFowland III, 12/13

Dynamic Pattern Detection with Temporal Consistency and Connectivity Constraints [.pdf] - Skyler Speakman, 11/13

Sigma-Optimality for Active Learning on Gaussian Random Fields [.pdf] -Yifei Ma, 11/13

Beyond Poisson: Modeling Inter-Arrival Times of Requests in a Datacenter [.pdf] - Da-Cheng Juan, 11/13

Experimental Evaluation of Feature Selection Methods for Clustering [.pdf] -Martin Azizyan, 10/13

Latent Session Model for Web User Clustering: A case study on modeling users of an online real estate website [.pdf] - Haijie Gu, 8/13

Semi-supervised Data Clustering with Coupled Non-negative Matrix Factorization: Sub-category Discovery of Noun Phrases in NELL's Knowledge Base [.pdf] - Chunlei Liu, 8/13

Analysis of Crime in Pittsburgh [.pdf] - Aaditya Ramdas, 5/13

Fast and Effective Similarity Searching in Large MIDI databases [.pdf] - Guangyu Xia, 5/13

What Makes Paris Look Like Paris? [.pdf] - Carl Doersch, 5/13

Semi-supervised context-aware discovery of unknown audio concepts [.pdf] - Antonio Juarez, 3/13

Learning Frames from Text with an Unsupervised Latent Variable Model [.pdf] - Brendan O'Connor, 10/12

TREEGL: Reverse Engineering Tree-Evolving Gene Networks Underlying Developing Biological Lineages [.pdf] - Ankur Parikh, 10/12

Automated Learning of Subcellular Location Patterns in Confocal Fluorescence Images from Human Protein Atlas [.pdf] - Jieyue Li, 10/12

Conditional Sparse Coding and Multiple Regression for Grouped Data [.pdf] - Min Xu, 5/12

Active, semi-supervised learning to utilize human oracles [.pdf] - Robert Fisher, 5/12

Decoding Word Semantics from Magnetocencephalography Time Series Transformations [.pdf] - Alona Fyshe, 5/12

Layered Timeseries Analysis for Smart Grid Agents [.pdf] - Prashant Reddy, 5/12

Learning Global Properties of Scene Images Based on Their Correlational Structures [.pdf] - Wooyoung Lee, 5/12

Understanding the Interaction between Interests, Conversations and Friendships in Facebook [.pdf] - Qirong Ho, 4/12

Trade-offs in Explanatory Model Learning [.pdf] - Madalina Fiterau, 3/12

Dynamics of Visual Category Learning with Magnetoencephalography [.pdf] - Yang Xu, 12/11

Online Detection of Unusual Events in Videos via Dynamic Sparse Coding [.pdf] - Bin Zhao, 11/11

Valid Statistical Inference on Automatically Matched Files [.pdf] - Rob Hall, 11/11

Modeling Correlated Purchase Behavior in Large-Scale Networks:A Markov Random Field (MRF) Approach [.pdf] - Liye Ma, 4/11

Multi-factor Analysis for Classifying fMRI Brain Images [.pdf] - Sung Won Park, 4/11

Learning the Sparsity Parameter in a Generalized Fast Subset Sums Framework for Bayesian Event Detection [.pdf] - Kan Shao, 4/11

CANTINA+: A Feature-rich Machine Learning Framework for Detecting Phishing Web Sites [.pdf] - Guang Xiang, 4/11

Comparing Data Sources in High Dimensions [.pdf] - Di Liu, 3/11

Cross-Species Queries of Large Gene Expression Databases [.pdf] - Hai-Son Le, 3/11

Clustering Under Natural Stability Assumptions [.pdf] - Pranjal Awasthi, 3/11

Automated Unmixing of Complex Protein Subcellular Location Patterns [.pdf] - Tao Peng, 2/11

Extracting Subpopulations from Large Social Networks [.pdf] - Bin Zhang, 2/11

Inferring Rates of Domain Shuffling Using a Birth-Death and Gain Model - Maureen Stolzer, 1/11

Anomaly Detection for Astronomical Data [.pdf] - Liang Xiong, 12/10

An Efficient Proximal Gradient Method for General Structured Sparse Learning [.pdf] - Xi Chen, 11/10

Learning Dynamic Models from Non-sequenced Data [.pdf] - Tzu-Kuo Huang, 11/10

Multiple Domain User Personalization [.pdf] - Yucheng Low, 11/10

Learning Opponent's Strategies In the RoboCup Small Size League [.pdf] - Felipe Trevizan, 10/10

Parallel Splash Belief Propagation [.pdf] - Joseph Gonzalez, 5/10

Trends and Differences in Industrial Safety Perception survey results: Subset Selection and Minimax Bound based Hypothesis Testing - Liu Yang, 5/10

Learning to Tag using Noisy Labels [.pdf] - Edith Law, 4/10

Polonium: Tera-Scale Graph Mining for Malware Detection [.pdf] - Duen Horng Chau, 4/10

Data Mining with MapReduce: Graph and Tensor Algorithms with Applications [.pdf] - Charalampos Tsourakakis, 4/10

Learning Directed Graphical Models from Nonlinear and Non-Gaussian Data [.pdf] - Robert Tillman, 3/10

Semi-parametric Methods for Estimating Time-varying Graph Structure [.pdf] - Mladen Kolar, 2/10

Genetic Population Structure in Pacific Islanders [.pdf] - Suyash Shringarpure, 2/10

Discovery of Student Strategies using Hidden Markov Model Clustering [.pdf] - Benjamin Shih, 1/10

Grasping in Primates: Mechanics and Neural Basis [.pdf] - Lucia Castellanos, 12/09

Parallel WalkSAT with Clause Learning [.pdf] - Austin McDonald, 12/09

Semi-Supervised Discovery of Named Entities and Relations from the Web [.pdf] - Sophie Wang, 11/09

Learning Stable Linear Dynamical Systems [.pdf] - Byron Boots, 6/09

Structured Correspondence Topic Models for Mining Captioned Figures in Biological Literature [.pdf] - Amr Ahmed, 5/09

FilterBoost: Regression and Classification on Large Datasets [.pdf] - Joseph Bradley, 5/09

Learning Compressible Models [.pdf] - Yi Zhang, 5/09

Center-Piece Subgraphs: Problem Definition and Fast Solutions [.pdf] - Hanghang Tong, 8/08

Graph-Based Semi-Supervised Learning as a Generative Model [.pdf] - Jingrui He, 8/08

Information Propagation on the Web: Patterns and a Model [.pdf] - Mary McGlohon, 11/07

Maximum Likelihood Estimation in Latent Class Models for Contingency Table Data [.pdf] - Yi Zhou, 11/07

A Comparison of Methods for Transductive Transfer Learning [.pdf] - Andrew Arnold, 5/07

Learning Factors Analysis - A General Method for Cognitive Model Evaluation and Improvement [.pdf] - Hao Cen, 5/07

The Complexity of Interactive Machine Learning [.pdf] - Stephen Hanneke, 5/07

T-cube: Fast Extraction of Time Series from Large Datasets [.pdf] - Maheshkumar Sabhnani, 5/07

Learning Selectively conditioned Forest Structures with Applications to DBNs and Classification [.pdf] - Brian Ziebart, 5/07

Large-Scale Automated Analysis of Location Patterns in Randomly-Tagged 3T3 Cells [.pdf] - Juchang Hua,

4/07

Continuous Hidden Process Model for Time Series [.pdf] - Yanxin Shi, 4/07

Modeling Networks Using Kronecker Multiplication [.pdf] - Jurij Leskovec, 4/07

Gene Family Classification using a Semi-Supervised Learning Method [.pdf] - Nan Song, 1/07

Feature Reduction for Improved Recognition of Subcellular Location Patterns in Fluorescence Microspoe Images [.pdf] - Kai Huang, 11/06

Intelligent Light Control using Sensor Networks [.pdf] - Vipul Singhvi, 9/06

Incremental Hierarchical Clustering of Text Documents [.pdf] - Nachiketa Sahoo, 5/06

Using Customer's Reported Forecasts to Predict Future Sales [.pdf] - Nihat Altintas, 5/06

Data Mining in Macroeconomic Data Sets [.pdf] - Ping Chen, 4/06

Dynamic Social Network Analysis using Latent Space Models [.pdf] - Purnamrita Sarkar, 4/06

Active Learning for Identifying Function Threshold Boundaries [.pdf] - Brent Bryan, 4/06

Anomoly Detection in Multivariate Time Series [.pdf] - Kustav Das, 3/06

On the Number of Experiments Sufficient and in the Worst Case Necessary to Identify All Causal Relations Among N Variables [.pdf] - Frederick Eberhardt, 9/05

N-1 Experiments Suffice to Determine the Causal Relations Among N Variables [.pdf] - Frederick Eberhardt, 9/05

Conditional Density Estimation using Finite Mixture Models with an Application to Astrophysics [.pdf] - Alex Rojas-Pena, 7/05

Location proteomics - Building subcellular location trees from high resolution 3D fluorescence microsope images of randomly-tagged proteins [.pdf] - Xiang Chen, 5/05

Tabu Search Enhanced Markov Blanket Classifier for High Dimensional Data Sets [.pdf] - Xue Bai, 1/05

Clustering Short Time Series Gene Expression Data [.pdf] - Jason Ernst, 11/04

A Hierarchical Graphical Model for Record Linkage [.pdf] - Pradeep Ravikumar, 5/04

Learning Robust Rules from Data: The GenTree Algorithm [.pdf] - Yiheng Li, 4/04

Advances in Network Tomography [.pdf] - Edoardo Airoldi, 10/03

Improved Recognition of Protein Subcellular Location Patterns via Feature Selection and Classifier Ensembles [.pdf] - Kai Huang, 8/03

Fractal Dimension for Data Mining [.pdf] - Sree Krishna Kumaraswamy, 7/03

Tools for Graph Mining [.pdf] - Yiping Zhan, 6/03

Using Machine Learning to Detect Cognitive States across Multiple Subjects [.pdf] - Xuerui Wang, 5/03

People Tracking Using Many Simple Sensors [.pdf] - Daniel Wilson, 5/03

Simultaneous Localization and Mapping using Sparse Extended Iinformation Filters [.pdf] - Yufeng Liu, 4/03

Multi-agent Learning in Extensive Games with Complete Information [.pdf] - Pu Huang, 1/03

Compromising Privacy with Trail Re-Identification: The REIDIT Algorithms [.pdf] - Bradley Malin, 12/02

Mining Computer Tutor-Student Interaction Data to Assess Students Reading and Predict Future Behavior [.pdf] - Peng Jia, 10/02

A Method for Automatically Finding Interpretations of Reduced Dimension Representations [.pdf] - Marc Fasnacht, 9/02

A Method for Automatically Finding Structural Motifs in Proteins [.pdf] - Marc Fasnacht, 9/02

Learning Rich Neural Network Topologies [.pdf] - Matteo Matteucci, 7/02

The Structure of the Unobserved [.pdf] - Ricardo Silva, 6/02

Learning from Labeled and Unlabeled Data with Label Propagation [.pdf] - Xiaojin Zhu, 6/02

The "DGX" Distribution for Mining Massive, Skewed Data [.pdf] - Zhiqiang Bi, 5/02

Diffusion Kernels on Graphs and Other Discrete Input Spaces [.pdf] - Risi Imre Kondor, 4/02

Large-scale Automated Forecasting using Fractals [.pdf] - Deepayan Chakrabarti, 4/02

Planning for Single and Multiple Actors in Markov Decision Processes with Deterministic Hidden State [.pdf] - Jamie Schulte, 12/01

Boosting and Maximum Likelihood for Expontial Models [.pdf] - Guy Lebanon, 9/01

Framework for using grocery data for early detection of Bio-terrorism attacks [.pdf] - Anna Goldenberg, 9/01

Causal Inference [.pdf] and Additive Models[.pdf] - Tianjiao Chu, 5/01

Using Error-Correcting Codes for Efficient Text Categorization with a Large Number of Categories [.pdf] -

Rayid Ghani, 5/01

A Boosting Approach to Topic Spotting on Subdialogues [.pdf] - Kary Myers, 2000