

Saharon ROSSET, Ph.D.

CURRICULUM VITAE

Work Address: Raymond and Beverly Sackler School of
Mathematical Sciences,
Tel Aviv University,
Tel Aviv 69978, Israel
Work Tel No.: +972-3-640-8820
E-mail: saharon@post.tau.ac.il
Date and place of birth: 26/6/1969 in Evanston, IL
Marital status: Married+2
Citizenship: Israel and USA

EDUCATION

1992–1994 B.Sc., Mathematics
Tel Aviv University
Date of award: 1995 (Summa Cum Laude)

1995–1999 MSc., Statistics
Tel Aviv University
Date of award: 2000 (Summa Cum Laude)
Title of Masters Thesis: *Ranking: Methods for Flexible Evaluation
and Efficient Comparison of Classification Performance*
Name of Supervisor: Professor David Steinberg

1999–2003 Ph.D., Statistics
Stanford University, Stanford, CA
Date of award: 2003
Title of Doctoral Dissertation: *Topics in Regularization and Boosting*
Name of Supervisors: Professors Jerome Friedman and Trevor Hastie

ACADEMIC AND PROFESSIONAL EXPERIENCE

- 2007–present Senior Lecturer, Dept. of Statistics and Operations Research, Tel Aviv University, Israel.
- 2003–2007 Research Staff Member, Mathematical Sciences Dept., IBM T.J. Watson Research Center, NY.
Developing algorithmic and statistical solutions to problems in business and scientific domains.
- 2001–2003 Teaching assistant, Statistics Dept., Stanford University, CA.
- 1999–2001 Research assistant, Statistics Dept., Stanford University, CA.
Data analysis project funded by Blue Martini Ltd.
- 1997–2003 Data Mining researcher/consultant, Amdocs Ltd (USA & Israel).
Developing data mining solutions for the telecommunications industry.
- 1993–1997 Mathematical programmer, Ron Soferman Image Processing Ltd., Israel
Developing image processing and fax encryption software.

OTHER ACTIVITIES

- 2005–2006 Participated in Genographic project's media campaign: interviewed for dozens of media outlets worldwide.
- 2006–2007 Chair of IBM Research's Knowledge Discovery PIC (Professional Interest Community).
- From Jun. 2006 Action Editor for Journal of Machine Learning Research.
- From Dec. 2006 Member of the Editorial Board of Machine Learning Journal.

LIST OF PUBLICATIONS BY AREA**Statistical and Machine Learning**

1. Saharon Rosset, Eran Segal
Boosting Density Estimation
Neural Information Processing Systems 15, 2002.
2. Ji Zhu, Saharon Rosset, Trevor Hastie, Rob Tibshirani
1-norm Support Vector Machines
Neural Information Processing Systems 16, 2003.
3. Saharon Rosset, Ji Zhu, Trevor Hastie
Margin Maximizing Loss Functions
Neural Information Processing Systems 16, 2003.
4. Saharon Rosset, Ji Zhu, Trevor Hastie
Boosting as a Regularized Path to a Maximum Margin Classifier
Journal of Machine Learning Research, 5:Aug. (2004), 941–973.
5. Saharon Rosset, Ji Zhu
Discussion of “Least Angle Regression” by Efron et al.
Annals of Statistics, Apr. 2004.
6. Saharon Rosset, Ji Zhu
A Corrected Proof of the Result of “A Prediction Error Property of the Lasso” by Huang(2003)
Australia and New Zealand Journal of Statistics, 46:3 (2004), 505–510.
7. Jerome Friedman, Trevor Hastie, Saharon Rosset, Rob Tibshirani, Ji Zhu
Discussion of Three Boosting Papers
Annals of Statistics, Feb. 2004.
8. Trevor Hastie, Saharon Rosset, Rob Tibshirani, Ji Zhu
The Complete Regularization Path of the Support Vector Machine
Journal of Machine Learning Research, 5:Oct. (2004), 1391–1415.
9. Saharon Rosset
Model Selection via the AUC
Proceedings of the 21st International Conference on Machine Learning, ICML-04.
10. Trevor Hastie, Saharon Rosset, Rob Tibshirani, Ji Zhu
The Entire Regularization Path of the Support Vector Machine
Neural Information Processing Systems 17, 2004.
(short version of paper 8)

11. Saharon Rosset, Ji Zhu, Hui Zou, Trevor Hastie
A Method for Inferring Label Sampling Mechanisms in Semi-Supervised Learning
Neural Information Processing Systems 17, 2004.
12. Saharon Rosset
Tracking Curved Regularized Optimization Solution Paths
Neural Information Processing Systems 17, 2004.
13. Saharon Rosset
Robust Boosting and Its Relation to Bagging
Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD-05.
14. Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, Keith Knight
Sparsity and Smoothness via the Fused Lasso
Journal of the Royal Statistical Society, Series B, Vol. 67 (2005), No. 1.
15. Saharon Rosset, Ji Zhu
Sparse, Flexible and Efficient Modeling using L1 Regularization
In: I. Guyon and S. Gunn (eds.), Feature Extraction, Foundations and Applications.
Springer, 2006.
16. Saharon Rosset, Ji Zhu
Piecewise Linear Regularized Solution Paths
Annals of Statistics, Vol. 35:3. (2007), 1012–1030.
17. Saharon Rosset, Grzegorz Swirszcz, Ji Zhu, Nathan Srebro
 l_1 Regularization in Infinite Dimensional Feature Spaces
Twentieth Conference on Learning Theory (COLT-07), 2007.
18. Ji Zhu, Saharon Rosset, Hui Zou, Trevor Hastie
Multi-class AdaBoost
Under review by the Journal of the Royal Statistical Society, Series B.
19. Giles Hooker, Saharon Rosset
Prediction-Based Regularization Using Data Augmented Regression
Under review by the Journal of the American Statistical Association.

Practical Data Mining

1. Saharon Rosset
Ranking: Methods for Flexible Evaluation and Efficient Comparison of Classification Performance
Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, KDD-98.
2. Saharon Rosset, Uzi Murad, Einat Neumann, Yizhak Idan, Gadi Pinkas
Discovery of Fraud Rules for Telecommunications: Challenges and Solutions
Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD-99.
3. Saharon Rosset, Einat Neumann, Uri Eick, Nurit Vatnik, Shuki Idan
Evaluation of Prediction Models for Campaign Planning
Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD-01.
4. Saharon Rosset, Einat Neumann, Uri Eick, Nurit Vatnik, Shuki Idan
Lifetime Value Modeling and Its Use for Customer Retention Planning
Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD-02.
Winner of best application paper award
(short version of paper 5)
5. Saharon Rosset, Einat Neumann, Uri Eick, Nurit Vatnik
Lifetime Value Models for Decision Support
Data Mining and Knowledge Discovery Journal, Vol. 7 (2003), 321–339.
6. Saharon Rosset, Einat Neumann
Integrating Customer Value Considerations into Predictive Modeling
Proceedings of the Third IEEE Conference on Data Mining, ICDM-03.
7. Saharon Rosset, Claudia Perlich, Bianca Zadrozny, Srujana Merugu, Sholom Weiss, Rick Lawrence
Customer Wallet Estimation
1st NYU workshop on CRM and Data Mining, 2005.
8. Saharon Rosset, Claudia Perlich, Bianca Zadrozny
Ranking-Based Evaluation of Regression Models
Proceedings of the Fifth IEEE Conference on Data Mining, ICDM-05.
(short version of paper 12)

9. Sofus Macskassy, Foster Provost, Saharon Rosset
ROC Confidence Bands: An Empirical Evaluation
Proceedings of the 22nd International Conference on Machine Learning, ICML-05.
10. Saharon Rosset, Richard Lawrence
Data Enhanced Predictive Modeling for Sales Targeting
2006 SIAM Conference on Data Mining.
11. Srujana Merugu, Saharon Rosset, Claudia Perlich
A New Multi-View Regression Method with an Application to Customer Wallet Estimation
Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD-06.
12. Saharon Rosset, Claudia Perlich, Bianca Zadrozny
Ranking-Based Evaluation of Regression Models
Knowledge and Information Systems, 12:3 (2007), 331-353.
13. Claudia Perlich, Saharon Rosset
Identifying Bundles of Product Options using Mutual Information Clustering
2007 SIAM Conference on Data Mining, 2007.
14. Claudia Perlich, Saharon Rosset, Bianca Zadrozny, Richard Lawrence
High-Quantile Modeling for Customer Wallet Estimation and Other Applications
Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD-07.

Computational Biology

1. Saharon Rosset
Efficient Inference on Known Phylogenetic Trees Using Poisson Regression
Proc. of the 5th European Conference on Computational Biology (ECCB-2006),
Bioinformatics 22, Suppl 2.
2. Doron Behar, Saharon Rosset, et al. (ten authors)
The Genographic Project Public Participation Mitochondrial DNA Database
PLoS Genetics, 3:6 (2007), e104.
3. Doron Behar, Ene Metsaplu, Toomas Kivisild, Saharon Rosset, et al. (15 authors)
Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora
Under revision for PLoS One (tentatively accepted 8/07).
4. Doron Behar, Richard Villems, – 11 other authors –, Saharon Rosset and the Genographic Consortium
The dawn of human matrilineal diversity: Insights from complete mitochondrial genomes
Under review by Science (sent out for reviews 10/07).
5. Saharon Rosset, Spencer Wells, David Soria, Chris Tyler-Smith, Ajay Royyuru, Doron Behar and the Genographic Consortium
Maximum Likelihood Estimation of Site-Specific Mutation Probabilities in the Hyper-Variable Region of Human mtDNA
Pending submission to Molecular Biology and Evolution.

MAJOR INVITED TALKS (last three years)

- Aug. 2004 Invited talk at Joint Statistical Meetings, Toronto, Canada.
- Apr. 2005 Invited talk at Genographic project launch, Washington D.C.
- Jan. 2006 Invited talk at Shiluvim entrepreneurship forum, Tel Aviv University Engineering Dept. alumni club, Israel.
- May. 2006 Invited talk at Genographic project yearly meeting, Capetown, South Africa.
- Aug. 2006 Invited tutorial titled *Data Analytics for Marketing Decision Support* given at Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD-06, Philadelphia, PA.
(with Naoki Abe)
(also given at SIAM Data Mining 2007, SDM07)
- Aug. 2007 Invited talk at Joint Statistical Meetings, Salt Lake City, UT. (did not attend, talk given by colleague)
- Aug. 2007 Invited talk at Genographic project yearly meeting, Shanghai, China.
- Aug. 2007 Winner's talk at KDD-07 workshop on the KDD/Netflix Cup, San Jose, CA.

UNIVERSITY SEMINARS (last three years)

- Jan. 2004 Statistics seminar at University of Toronto, Canada.
- Feb. 2004 Statistics seminar at Rutgers University, NJ.
- Mar. 2004 Applied Math seminar at Yale University, CT.
- Mar. 2004 Machine Learning seminar at Hebrew University, Israel.
- Mar. 2004 School of Math Sciences colloquium, Tel Aviv University, Israel.
- Apr. 2004 Statistics seminar at University of Michigan, MI.
- Dec. 2004 Statistics seminar at Tel Aviv University, Israel.
- Apr. 2005 Statistics seminar at University of Michigan, MI.
- Nov. 2005 Statistics seminar at Stanford University, CA.
- Nov. 2005 Statistics seminar at University of California, Berkeley, CA.
- Jan. 2006 Statistics seminar at Tel Aviv University, Israel.

- Jan. 2007 Short course titled *Regularization and Boosting* at 9th School of Probability and Statistics, CIMAT, Guanajuato, Mexico.
- Feb. 2007 Computational biology seminar at University of Arizona, AZ.
- Feb. 2007 Statistics seminar at Stanford University, CA.
- Mar. 2007 Statistics seminar at Rutgers University, NJ.
- Oct. 2007 Statistics seminar at Tel Aviv University, Israel.

GRANTS, ACADEMIC AND PROFESSIONAL AWARDS

- 1999 Nimrod Lapid award for best master's thesis in Statistics at Tel Aviv university.
- 1999 Arik Peretz award for best masters thesis with applications in Biostatistics.
- 1999 Headed team that won the knowledge discovery task of KDD-Cup 99, the yearly international data mining competition.
- 2000 Advised team that won two of five sub-competitions at KDD-Cup 00.
- 2002 IBM Graduate Fellow, Frontiers of Engineering meeting, September 02, Irvine, CA.
- 2002 Primary author of best application paper award winner at KDD-2002.
- 2005–2008 NSF grant DMS-0505432 for \$115,597 (co-PI with Ji Zhu of Univ. of Michigan).
- 2006 IBM Research Outstanding Technical Achievement Award.
- 2007 IBM Research Outstanding Technical Achievement Award.
- 2007–2011 EU FP7 International Reintegration Grant for 100,000 Euros.
- 2007 Headed team that won one of two sub-competitions at KDD-Cup 07.

REFEREEING AND PROGRAM COMMITTEE MEMBERSHIP

Journals	Journal of the Royal Statistical Society; Journal of the American Statistical Society; Journal of Computational and Graphical Statistics; Statistica Sinica; Journal of Machine Learning Research; Machine Learning Journal; Data Mining and Knowledge Discovery Journal; AI in Medicine; Journal of Statistical Planning and Inference; Technometrics; Annals of Applied Statistics; Journal of Interactive Marketing
Conferences	Neural Information Processing Systems (NIPS); Intl. Conf. on Knowledge Discovery in Databases (KDD); Intl. Conf. on Machine Learning (ICML); IEEE Conf. on Data Mining (ICDM); Conf. of the American Assoc. of Artificial Intelligence (AAAI)

STUDENTS SUPERVISED

Summer 2004	Co-supervised summer student Arindam Bannerjee (now assistant professor at Univ. of Minnesota).
Summer 2005	Co-supervised summer student Srujana Merugu.

RESEARCH INTERESTS

My interests are in combining theory and applications in the areas of statistical modeling, machine learning and data mining. My current research stretches across the spectrum from developing modeling solutions for specific problems in business and scientific domains to methodological statistical learning research. Some of my current projects are:

1. *Customer wallet estimation*: solving a difficult and important business problem of modeling a customer's potential spending (an unobserved amount, as opposed to actual spending).
2. *Modeling genetic data in the Genographic project*: building statistical models for estimating mutation rates, classifying DNA samples and more.
3. *Efficient exact and approximate methods for regularized optimization*: combining statistical and computational considerations in designing new data modeling methodologies.

Date: November 2007