

Journal Papers

Conference Papers

Tutorials

Open Problems

Book Chapters

Technical Reports

Theses

Journal Papers

- J73 Amid, E., Anil, R., Fifty, C., and Warmuth, M. K. (2023), “Layerwise Bregman Representation Learning with Applications to Knowledge Distillation”, *Transactions on Machine Learning Research* 02/2023, pp. 1–19, February 2023. [pdf](#)
- J72 Dereziński, M., Warmuth, M. K. and Daniel Hsu (2022), “Unbiased estimators for random design regression”, *Journal of Machine Learning Research*, vol. 22, pp. 1-46, August 2022. [pdf](#)
- J71 Chalopin, J., Chepoi, V., Moran, S., and Warmuth, M. K. (2021), “Unlabeled sample compression schemes and corner peelings for ample and maximum classes”, *Journal of Computer and System Sciences*, vol. 127, pp. 1-28, August 2022. [pdf](#)
- J70 Nakamura, A., Helmbold, D. P., and Warmuth, M. K. (2019), “Mistake bounds on a noise-free multi-armed bandit game”, *Journal of Information and Computation*, vol. 269, pp. 1-20, December 2019. [pdf](#)
- J69 Dereziński, M., and Warmuth, M. K. (2018), “Reverse iterative volume sampling for linear regression”, *Journal of Machine Learning Research*, vol. 19, pp. 1-39, August 2018. [pdf](#)
Long talk about volume sampling [pdf](#)
-
- J68 Nie, J., Kotłowski, W., and Warmuth, M. K. (2016), “Online PCA with optimal regret”, *Journal of Machine Learning Research*, vol. 17, pp. 1-49, September 2016. [pdf](#)
Talk [pdf](#)
- J67 Hazan, E., Kale, S., and Warmuth, M. K. (2016) “Learning rotations with little regret”, *Journal of Machine Learning*, vol. 104(1), pp. 129-148, March 2016. [pdf](#)
- J66 Warmuth, M. K., Kotłowski, W., and Zhou, S. (2014), “Kernelization of matrix updates”, *Journal of Theoretical Computer Science*, special issue for the 23rd International Conference on Algorithmic Learning Theory (ALT 12), vol. 558, pp. 159–178, September 2014. [pdf](#)
Talk [pdf](#)
- J65 Warmuth, M. K., Koolen, W. M., and Helmbold, D. P. (2012) “Combining initial segments of lists”, *Journal of Theoretical Computer Science*, special issue for the 22nd International Conference on Algorithmic Learning Theory (ALT 11), vol. 519, pp. 29–45, January 2014. [pdf](#)
- J64 Warmuth, M. K., and Kuzmin, D. (2011) “Online Variance Minimization”. *Journal of Machine Learning*, Vol. 87(1), pp. 1–32, November 2011. [pdf](#)
Talk [pdf](#)
- J63 Warmuth, M. K., and Kuzmin, D. (2009), “Bayesian Generalized Probability Calculus for Density Matrices”, *Journal of Machine Learning*, Vol. 78(1-2), pp. 63–101, January 2010. [pdf](#)
Talk [pdf](#)
Video of talk [@Simon's](#)
- J62 Helmbold, D. P., and Warmuth, M. K. (2009) “Learning Permutations with Exponential Weights” *Journal of Machine Learning Research*, vol 10, pp. 1705–1736, July 2009. [pdf](#)
Talk [pdf](#)
- J61 Warmuth, M. K., and Kuzmin, D. (2008), “Randomized Online PCA Algorithms with Regret Bounds that are Logarithmic in the Dimension”, *Journal of Machine Learning Research*, vol. 9, pp. 2217-2250, October 2008. [pdf](#)
Talk [pdf](#)
- J60 Kuzmin, D., and Warmuth, M. K. (2007) “Unlabeled Compression Schemes for Maximum Classes”, *Journal of Machine Learning Research*, vol. 8, pp. 2047-2081, September 2007. [pdf](#)
Talk [pdf](#)

- J59 Jun Liao, Manfred K. Warmuth, Sridar Govindarajan, Jon E. Ness, Rebecca P. Wang, Claes Gustafsson, Jeremy Minshull (2007) “Engineering proteinase K using machine learning and synthetic genes”, *BMC Biotechnology*, vol 7:16, 19 pages, March 2007. [pdf](#)
Talk [ppt](#)
- J58 Kivinen, J., Warmuth, M. K., and Hassibi, B. (2006) “The p-norm Generalization of the LMS Algorithm for Adaptive Filtering”, *IEEE Transactions on Signal Processing*, pp. 1782–1793, vol. 54(5), May 2006. [pdf](#)
Talk [pdf](#)
- J57 Rätsch, G. and Warmuth, M. K. (2005) “Efficient Margin Maximization with Boosting”, *Journal of Machine Learning Research*, Vol. 6, pp. 2131–2152, December 2005. [pdf](#)
- J56 Tsuda, K., Räetsch, G., and Warmuth, M. K. (2005) “Matrix Exponentiated Gradient Updates of On-line Learning and Bregman Projection”, *Journal of Machine Learning Research*, Vol. 6, pp. 995–1018, June 2005. [pdf](#)
- J55 Takimoto, E. and Warmuth, M. K. (2003) “Path Kernels and Multiplicative Update”, special issue of *Journal of Machine Learning Research: “Best of learning theory 2003”* (edited by Thore Graepel and Ralf Herbrich), Vol. 4, pp. 773-818, October 2003. [pdf](#)
Talk [pdf](#)
- J54 Warmuth, M. K., Liao, J., Rätsch, G., Mathieson, M., Putta, S. and Lemmen, C. (2003) “Active Learning with Support Vector Machines in the Drug Discovery Process”. *Journal of Chemical Information Sciences* vol. 43(2), pp. 667-673, 2003. [pdf](#)
- J53 Forster, J., and Warmuth, M. K. (2003) “Relative Loss Bounds for Temporal Difference Learning”, *Journal of Machine Learning*, Vol. 51-1, pp. 23-50, April 2003. [pdf](#)
- J52 Bousquet, O. and Warmuth, M.K. (2002) “Tracking a Small Set of Experts by Mixing Past Posteriors”, in a special issue of *Journal of Machine Learning Research* for COLT 2001 (edited by Phil Long), vol. 3(Nov):363-396, 2002. [pdf](#)
Talk [pdf](#)
- J51 Forster, J., and Warmuth, M. K. (2002) “Relative Expected Instantaneous Loss Bounds”, *Journal of Computer and System Sciences*, Vol. 64-1, pp. 76-102. Special issue of for the *13th Annual Conference on Computational Learning Theory (COLT 2000)*, February 2002. [pdf](#)
- J50 Takimoto, E., and Warmuth, M. K. (2002) “Predicting Nearly as Well as the Best Pruning of a Planar Decision Graph”, Special issue of *Theoretical Computer Science* for the *10th International Conference on Algorithmic Learning Theory (ALT 99)*, Vol. 288(2), pp. 217-235, 2002. [pdf](#)
- J49 Herbster, M. and Warmuth, M.K. (2001) “Tracking the Best Linear Predictor”, *Journal of Machine Learning Research*, Vol. 1, pp. 281-309, September 2001. [pdf](#)
- J48 Abe, N., Takeuchi, J., and Warmuth, M. K. (2001) “Polynomial Learnability of Stochastic Rules with respect to the KL-divergence and Quadratic Distance”, *IEICE Transactions on Information and Systems*, Vol. E84-D, No.3, pp. 299–315, 2001. [pdf](#)
- J47 Azoury, K., and Warmuth, M. K. (2001) “Relative Loss Bounds for On-line Density Estimation with the Exponential Family of Distributions”, in a special issue of the *Journal of Machine Learning on Theoretical Advances in On-line Learning, Game Theory and Boosting*, edited by Yoram Singer, Vol. 43(3), pp. 211-246, June 2001. [pdf](#)
- J46 Kivinen, J., and Warmuth, M. K. (2001) “Relative Loss Bounds for Multidimensional Regression Problems”. *Journal of Machine Learning*, Vol. 45(3), pp. 301-329. [pdf](#)
- J45 Helmbold, D. P., Panizza, S., and Warmuth, M. K. (2000) “Direct and Indirect Algorithms for On-line Learning of Disjunctions”, In special issue of *Theoretical Computer Science* for EUROCOLT 99, Vol. 284(1), pp. 109-142, July 2002. [pdf](#)

- J44 Helmbold, D.P., Kivinen, J. and Warmuth, M.K. (1999) “Relative Loss Bounds for Single Neurons”, *IEEE Transactions on Neural Networks*, Vol. 10(6), pp. 1291–1304, November 1999. [pdf](#)
- J43 Helmbold, D.P., Schapire, R.E., Singer, Y. and Warmuth, M.K. (1998) “On-line Portfolio Selection using Multiplicative Updates”, *Mathematical Finance*, Vol. 8(4), pp. 325–347, 1998. [pdf](#)
- J42 Haussler, D., Kivinen, J. and Warmuth, M. K. (1998) “Sequential Prediction of Individual Sequences under General Loss Functions”, *IEEE Transactions on Information Theory*, Vol. 44(5), pp. 1906–1925, September 1998. [pdf](#)
- J41 Maass, M. and Warmuth, M.K. (1998) “Efficient Learning with Virtual Threshold Gates”, in the Journal of *Information and Computation*, 141(1), pp. 66–83, February 1998. [pdf](#)
- J40 Auer, P. and Warmuth, M. K. (1998) “Tracking the Best Disjunction”, in the special issue on context sensitivity and concept drift of the *Journal of Machine Learning*, Vol. 32(2), pp. 127-150, August 1998. [pdf](#)
- J39 Herbster, M. and Warmuth, M.K. (1998) “Tracking the Best Expert”, in the special issue on context sensitivity and concept drift of the *Journal of Machine Learning*, Vol. 32(2), pp. 151-178, August 1998. [pdf](#)
- J38 Kivinen, J., Warmuth, M.K., and Auer, P. (1997) “The Perceptron Algorithm vs. Winnow: Linear vs. Logarithmic Mistake Bounds When Few Input Variables are Relevant”, in the special issue of *Artificial Intelligence*, on *Relevance*, Vol. 97(1–2), pp. 325–343, December 1997. [pdf](#)
- J37 Cesa-Bianchi, N., Freund, Y., Helmbold, D. P., Haussler, D., Schapire, R. E., and Warmuth, M. K. (1997) “How to Use Expert Advice”, *Journal of the ACM*, Vol. 44(3), pp. 427–485, May 1997.. [pdf](#)
- J36 Kivinen, J., and Warmuth, M.K. (1997) “Exponentiated Gradient versus Gradient Gradient Descent for Linear Predictors”, in the Journal of *Information and Computation*, Vol. 132, No. 1, pp. 1–64, January 1997. [pdf](#)
Correction of the EG+- update [pdf](#)
- J35 Helmbold, D.R., Schapire, R.E., Singer, Y. and Warmuth, M.K. (1997) “A Comparison of New and Old Algorithms for a Mixture Estimation Problem”, *Journal of Machine Learning*, Vol. 27(1), pp. 97–119. [pdf](#)
- J34 Schapire, R.E, and M.K. Warmuth (1996) “On the Worst-case Analysis of Temporal-difference Learning Algorithms”, in the special issue of *Machine Learning*, on Reinforcement Learning, Vol. 22, (1,2,3), pp. 95–121. [pdf](#)
- J33 Cesa-Bianchi, N., Long, P.M., and Warmuth, M.K. (1996) “Worst-case Quadratic Loss Bounds for Prediction Using Linear Functions and Gradient Descent”, in *IEEE Transactions on Neural Networks*, Vol. 7, No. 3, pp. 604–619, May 1996. [pdf](#)
- J32 Cesa-Bianchi, N., Y. Freund, D.P. Helmbold, and M.K. Warmuth (1996) “On-line Prediction and Conversion Strategies”, *Machine Learning*, Vol. 25 (1), pp. 71–110. [pdf](#)
- J31 Goldman, S. A. and M.K. Warmuth (1995) “Learning Binary Relations Using Weighted Majority Voting”, in *Machine Learning*, Vol. 20 (3), pp. 245–271, September 1995. [pdf](#)
- J30 Floyd, S., and M.K. Warmuth (1995) “Sample Compression, Learnability, and the Vapnik-Chervonenkis Dimension”, *Machine Learning*, Vol. 21 (3), pp. 269–304, December 1995. [pdf](#)
[Earlier technical report that first introduced compression schemes and gave a sample size bound for them.](#)
- J29 Littlestone, N., P.M. Long, and M.K. Warmuth (1995) “On-Line Learning of Linear Functions”, *Journal of Computational Complexity*, Vol. 5, pp. 1–23, 1995. [pdf](#)
[Earlier technical report containing an on-line algorithm for solving a system of linear equations.](#)

- J28 Helmbold, D., and M.K. Warmuth (1995) “On Weak Learning”, in the special issue of *Journal of Computer and System Sciences* for the *Fifth Annual Conference on Computational Learning Theory*, (Pittsburg, Pennsylvania, August 1992), Vol. 50, No. 3, pp. 551–573 (June 1995). [pdf](#)
- J27 Haussler, D., N. Littlestone, and M.K. Warmuth (1994) “Predicting 0,1 Functions on Randomly Drawn Points”, *Information and Computation*, 115(2), pp. 248–292 (December 1994). [pdf](#)
- J26 Long, P. M., and M.K. Warmuth (1994) “Composite Geometric Concepts and Polynomial Predictability”, *Information and Computation*, 113(2), pp. 230–252 (September 1994). [pdf](#)
- J25 Cesa-Bianchi, N., A. Krogh, and M.K. Warmuth (1994) “Bounds on Approximate Steepest Descent for Likelihood Maximization in Exponential Families”, *IEEE Transaction on Information Theory*, Vol. 40, No. 4, pp. 1215–1220 (July 1994). [pdf](#)
- J24 Littlestone, N, and M.K. Warmuth (1994) “The Weighted Majority Algorithm”, *Information and Computation*, 108(2), pp. 212–261 (February 1, 1994). [pdf](#)
- J23 Bodlaender, H. L., S. Moran, and M.K. Warmuth (1994) “The Distributed Bit Complexity of the Ring: from the Anonymous to the Non-anonymous Case”, *Information and Computation*, 108(1), 34–50 (January 1994). [pdf](#)
- J22 Pitt, L., and M.K. Warmuth (1993) “The Minimum Consistent DFA Problem Cannot be Approximated Within Any Polynomial”, *Journal of the Association for Computing Machinery*, Vol. 40, No. 1, January 1993, pp. 95–142. [pdf](#)
- J21 Helmbold, D., R. Sloan, and M.K. Warmuth (1992) “Learning Integer Lattices”, *Siam Journal on Computing*, vol. 21(2), 240–266 (April 1992). [pdf](#)
- J20 Abe, N. and M.K. Warmuth (1992) “On the Computational Complexity of Approximating Distributions by Probabilistic Automata”, in the special issue of *Machine Learning* for the *Third Annual Conference on Computational Learning Theory*, (Rochester, NY, August 1990), vol. 9(2/3), 1992. [pdf](#)
- J19 Moran, S., and Warmuth, M. K. (1992) “Gap Theorems for Distributed Computation”, *SIAM Journal on Computing*, 22(2). [pdf](#)
- J18 Haussler, D., Kearns, M., Littlestone N. and Warmuth, M. K. (1991) “Equivalence of Models for Polynomial Learnability”, *Information and Computation*, 95(2), 129–161, (December 91). [pdf](#)
- J17 Pitt, L. and Warmuth, M. K. (1990) “Prediction Preserving Reducibility”, in special issue of the *Journal of Computer and System Sciences*, for the *Third Annual Conference of Structure in Complexity Theory* (Washington, DC., June 88), vol. 41(3), 430–467, December 1990. [pdf](#)
- J16 Helmbold, D., Sloan, R., and Warmuth, M. K. (1990) “Learning Nested Differences of Intersection-closed Concept Classes”, in the special issue of the *Machine Learning*, for the *Second Workshop on Computational Learning Theory*, (Santa Cruz, California, July 1989), Vol. 5(2), pp. 165–196 June 1990. [pdf](#)
- J15 Ratner, Daniel and Warmuth, M. K. (1990) “The $[n^2 - 1]$ -Puzzle and Related Relocation Problem”, *Journal of Symbolic Computation*, 10 (2), 111–137 (August 1990). [pdf](#)
- J14 Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M. K. (1989) “Learnability and the Vapnik-Chervonenkis dimension”, *Journal of the Association for Computing Machinery*, 36, pp. 929–965 (1989). [pdf](#)
- J13 Anderson, R., Mayr, E., and Warmuth, M. K. (1989) “Parallel Approximation Algorithms for Bin Packing”, *Information and Computation*, 82(3), 262–277 (September 1989). [pdf](#)
- J12 Simons, B., and Warmuth, M. K. (1989) “A Fast Algorithm for Multiprocessor Scheduling of Unit-length Jobs”, *SIAM Journal on Computing*, 18 (4), 690–710 (August 1989). [pdf](#)

- J11 Gonczarowski, J., and Warmuth, M. K. (1989) “Scattered Versus Context-sensitive Rewriting”, *Acta Informatica*, 27, 81–95 (1989). [pdf](#)
- J10 Attiya, C., Snir, M., and Warmuth, M. K. (1988) “Computing on an Anonymous Ring”, *Journal of the Association for Computing Machinery*, 35(4), 845–675. [pdf](#)
- J9 Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M. K. (1987) “Occam’s Razor”, *Journal of Information Processing Letters*, 24, 377–380 (April 87). [pdf](#)
- J8 Dahlhaus, E., and Warmuth, M. K. (1986) “Membership for Growing Context Sensitive Grammars is Polynomial”, *Journal of Computer and System Sciences*, 33(3), 456–472. [pdf](#)
- J7 Dolev, D., Upfal, E., and Warmuth, M. K. (1986) “The Parallel Complexity of Scheduling with Precedence Constraints”, *Journal of Parallel and Distributed Computing*, 3, 553–576. [pdf](#)
- J6 Gonczarowski, J., and Warmuth, M. K. (1986) “Manipulating Derivation Forests by Scheduling Techniques”, *Journal of Theoretical Computer Science*, 45, 87–119. [pdf](#)
- J5 Dolev, D., and Warmuth, M. K. (1985) “Scheduling Flat Graphs”, *SIAM Journal on Computing*, 14(3), 638–657. [pdf](#)
- J4 Dolev, D., and Warmuth, M. K. (1985) “Profile Scheduling of Opposing forest and level orders”, *SIAM Journal on Algebraic and Discrete Methods*, 6(4), 665–687. [pdf](#)
- J3 Gonczarowski, J., and Warmuth, M. K. (1985) “Applications of Scheduling Theory to Formal Language Theory”, *Journal of Theoretical Computer Science*, 37(2), 217–243. [pdf](#)
- J2 Dolev, D., and Warmuth, M. K. (1984) “Scheduling Precedence Graphs of Bounded Height”, *Journal of Algorithms*, 5, 48–59. [pdf](#)
- J1 Warmuth, M. K., and Haussler, D. (1984) “On the Complexity of Iterated Shuffle”, *Journal of Computer and Systems Sciences*, 28, 345–358. [pdf](#)

Book Chapters

- BC1 Warmuth, M. K. (2016), “The Blessing and the Curse of the Multiplicative Updates”, pp. 265–289, Chapter 10 in *Evolutionary Games in Natural, Social, and Virtual Worlds* by Daniel Friedman and Barry Sinervo, Oxford University Press, 2016. [pdf](#)
Talk [pdf](#) [\[video\]](#)

Relevant Technical Reports and Unpublished Manuscripts

- T4 Smith, A. M., and Warmuth, M. K. (2010), “Learning Rotations Online”, Technical Report UCSC-SOE-10-08, Department of Computer Science, University of California, Santa Cruz, February 16, 2010. [pdf](#)
- T3 Krakowski, K. A., Mahony, R. E., Williamson, R. C. and Warmuth, M. K., (2007) “A geometric View of Non-Linear On-Line Gradient Descent”, unpublished manuscript, October 2007. [pdf](#)
- T2 Littlestone, N., P.M. Long, and Warmuth, M. K. (1995), “On-Line Learning of Linear Functions”, Technical Report UCSC-CRL-91-29, Department of Computer, University of California, Santa Cruz, October 1, 1991 (Revised October 30, 1991). [pdf](#)
- T1 Littlestone, N. and Warmuth, M. K. (1986), “Relating Data Compression and Learnability”, Technical Report, Department of Computer and Information Sciences, Santa Cruz, CA, June 10, 1986. [pdf](#)

Theses

- 1 Warmuth, M. K. (1981) “Scheduling on profiles of constant breadth,” Doctoral dissertation, Department of Computer Science, University of Colorado, Boulder, Colorado. University of Boulder library link: <http://libraries.colorado.edu/search/o?SEARCH=8864597>
- 2 Warmuth, M. K. (1980) “M processor unit-execution-time scheduling reduces to M-1 weakly connected components,” Master’s thesis, Department of Computer Science, University of Colorado, Boulder, Colorado. University of Boulder library link: <http://libraries.colorado.edu/search/o?SEARCH=7651548>

Conference Papers

- C120 Abernethy, J., Agarwal, A., Marinov, T. V., and Warmuth, M. K. (2024), “A Mechanism for Sample-Efficient In-Context Learning for Sparse Retrieval Tasks” Proceedings of The 35th International Conference on Algorithmic Learning Theory, PMLR 237:3-46, San Diego, February 2024. [pdf](#)
- C119 Amid, E., Nielsen, F., Nock, R. and Warmuth, M. K. (2024), “Optimal Transport with Tempered Exponential Measures”, Proceedings of 38th Annual AAAI Conference on Artificial Intelligence (AAAI 24), 38(10), pp. 10838-10846, Vancouver Canada, February 2024. [pdf](#)
- C118 Nock, R., Amid, E., and Warmuth, M. K. (2023), “Boosting with Tempered Exponential Measures”, *Advances in Neural Information Processing Systems 36 (NeurIPS 23)*, New Orleans, December 2023. [pdf](#)
- Poster [pdf](#)
- C117 Amid, E., Nock, R. and Warmuth, M. K. (2023), “Clustering above Exponential Families with Tempered Exponential Measures”, Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS23), PMLR 206:2994-3017, Valencia, Spain, April 2023. [pdf](#)
- Software [zip](#)
- Poster [pdf](#)
- C116 Amid, E., Anil, R. and Warmuth, M. K. (2022), “LocoProp: Enhancing BackProp via Local Loss Optimization”, Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS 22), virtual conference, PMLR: Volume 151:9626-9642, virtual conference, March 2022. [pdf](#)
- C115 Amid, E., Kotłowski, W., and Warmuth, M. K. (2021), “A case where a spindly two-layer linear network decisively outperforms any neural network with a fully connected input layer”, 32nd International Conference on Algorithmic Learning Theory (ALT 21), Paris, France, March 2021. [pdf](#)
- ALT talk [pdf](#)
- Long talk [video](#)
- [pdf](#)
- C114 Amid, E., and Warmuth, M. K. (2020) “Reparameterizing Mirror Descent as Gradient Descent”, *Advances in Neural Information Processing Systems 33 (NeurIPS 20)*, pp. 8430–8439, virtual conference, December 2020. [pdf](#)
- Video [link](#)
- Talk [pdf](#)
- Poster [pdf](#)
- C113 Talebi, T., Amid, E., Milanfar, P. , and Warmuth, M. K. (2020), “Rank-smoothed Pairwise Learning in Perceptual Quality Assessment”, IEEE International Conference on Image Processing (ICIP 20), virtual conference, October 2020. [pdf](#)
- Video [mp4](#)
- Talk [pdf](#)
- C112 Amid, E., and Warmuth, M. K. (2020), “Divergence-Based Motivation for Online EM and Combining Hidden Variable Models”, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI 20)*, in Proceedings of Machine Learning Research, vol 124, August 2020, virtual conference. [pdf](#)
- Paper w. appendix [link](#)
- Video [link](#)
- Talk [pdf](#)
- C111 Amid, E., and Warmuth, M. K. (2020), “Winnowing with Gradient Descent”, Proceedings of the 33rd Conference on Learning Theory (COLT 20), in Proceedings of Machine Learning Research, vol 125:1–20, virtual conference, July 2020. [pdf](#)

- Video [link](#)
 Talk [pdf](#)
 Long talk that also covers C114 [pdf](#)
- C110 Amid, E., and Warmuth, M. K. (2019), “An Implicit Form of Krasulina’s k-PCA Update without the Orthonormality Constraint”. Proceedings of the *Thirtyfourth AAAI Conference on Artificial Intelligence (AAAI 20)*. New York, NY, February 2020. [pdf](#)
 Poster [pdf](#)
- C109 Amid, E., Warmuth, M. K., Anil, R., and Koren, T. (2019), ”Robust Bi-Tempered Logistic Loss Based on Bregman Divergences”, *Advances in Neural Information Processing Systems 32 (NeurIPS 19)*, pp. 14987–14996, Vancouver, December 2019. [pdf](#)
 Poster [pdf](#)
 Talk [pdf](#)
[\[Updated arXiv version with appendix\]](#)
- C108 Chalopin, J., Chepoi, V., Moran, S., and Warmuth, M. K. (2019), “Unlabeled sample compression schemes and corner peelings for ample and maximum classes”, 46th International Colloquium on Automata, Languages and Programming (ICALP 2019), proceeding published by Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, vol. 132, pp. 34:1–34:15, location Patras, Greece, July 2019. [pdf](#)
- C107 Dereziński, M., Clarkson, K. L., Mahoney, M. W., and Warmuth, M. K. (2019), “Minimax experimental design: Bridging the gap between statistical and worst-case approaches to least squares regression”, 32nd Conference on Learning Theory (COLT 19), Phoenix AZ, USA, June 2019, in *Proceedings of Machine Learning Research* vol. 99:1050-1069, 2019. [pdf](#)
 Poster [pdf](#)
- C106 Kempka, M., Kotłowski, W., and Warmuth, M. K. (2019), “Adaptive scale-invariant online algorithms for learning linear models”, 36th International Conference for Machine Learning (ICML 19), Los Angeles CA, USA, June 2019, in *Proceedings of Machine Learning Research* vol. 97:3321-3330. [pdf](#)
 Supplement [pdf](#)
- C105 Amid, E., Warmuth, M. K., and Srinivasan, S. (2019), “Two-temperature logistic regression based on the Tsallis divergence”, 22nd International Conference on Artificial Intelligence and Statistics (AISTATS 19), April 2019, Naha, Okinawa, Japan. In *Proceedings of Machine Learning Research* vol 89:2388-2396, 2019. [pdf](#)
 Supplement [pdf](#)
 Poster [pdf](#)
- C104 Dereziński, M., Warmuth, M. K., and Hsu, D. (2019), “Correcting the bias in least squares regression with volume-rescaled sampling”, 22nd International Conference on Artificial Intelligence and Statistics (AISTATS 19), April 2019, Naha, Okinawa, Japan. In *Proceedings of Machine Learning Research* vol 89:944-953, 2019. [pdf](#)
 Supplement [pdf](#)
 Poster [pdf](#)
- C103 Cortes, C., Kuznetsov, V., Mohri, M., Rahmanian, H., and Warmuth, M. K. (2019), “Online Non-Additive Path Learning under Full and Partial Information”, 30th International Conference on Algorithmic Learning Theory (ALT 19), Chicago, USA, March 2019. In *Proceedings of Machine Learning Research* vol 98:274-299, 2019. [pdf](#)
 Talk [pdf](#)
- C102 Dereziński, M., Warmuth, M. K., and Hsu, D. (2018), “Leveraged volume sampling for linear regression”, *Advances in Neural Information Processing Systems 31 (NeurIPS 18)*, Montreal, December 2018. [pdf](#)
 Supplement [pdf](#)

- C101 Dereziński, M., and Warmuth, M. K. (2018), “Subsampling for Ridge Regression via Regularized Volume Sampling”, International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, Vol. 84:716–725. Playa Blanca, Lanzarote, Canary Islands, April 2018. [pdf](#)
Talk [pdf](#)
Poster [pdf](#)
- C100 Dereziński, M., and Warmuth, M. K. (2017), “Unbiased estimates for linear regression via volume sampling”, *Advances in Neural Information Processing Systems 30 (NeurIPS 17)*, Los Angeles, December 2017. [pdf](#)
Poster [pdf](#)
- C99 Rahmanian, H., and Warmuth, M. K. (2017), “Online dynamic programming”, *Advances in Neural Information Processing Systems 30 (NeurIPS 17)*, Los Angeles, December 2017. [pdf](#)
Poster [pdf](#)
- C98 Moran, S., and Warmuth, M. K. (2016), “Labeled compression schemes for extremal classes”, *Proceeding of the 27th International Conference on Algorithmic Learning Theory (ALT 16)*, Lecture Notes in Computer Science, vol 9925, Springer Verlag, pp. 34–49, Bari Italy, October 2016. [pdf](#)
- C97 Nakamura, A., Helmbold, D. P., and Warmuth, M. K. (2016), “Noise Free Multi-armed Bandit Game”, Proceedings of 10th International Conference on Language and Automata Theory and Applications, In Lecture Notes in Computer Science 9618, Springer Verlag, pp. 412–423, Prague, March 2016. [pdf](#)
Talk [pdf](#)
- C96 Bartlett, P. L., Koolen, W. M., Malek, A., Takimoto, E., and Warmuth, M. K. (2015), “Minimax Fixed-Design Linear Regression”, *JMLR: Workshop and Conference Proceedings vol 40:1–24, Proceedings of 28th Conference on Learning Theory (COLT 15)*, Paris, July 2015. [pdf](#)
- C95 Cortes, C., Kuznetsov, V., Mohri, M., and Warmuth, M. K. (2015), “On-Line Learning Algorithms for Path Experts with Non-Additive Losses”, *JMLR: Workshop and Conference Proceedings vol 40:1–24, Proceedings of 28th Conference on Learning Theory (COLT 15)*, Paris, July 2015. [pdf](#)
- C94 Dereziński, M., and Warmuth, M. K. (2014) “The limits of squared Euclidean distance regularization”, *Advances in Neural Information Processing Systems 27 (NeurIPS 14)*, Montreal, December 2014. [pdf](#)
NeurIPS spotlight [pdf](#)
NeurIPS poster [pdf](#)
- C93 Erven, T. V., Kotłowski, W., and Warmuth, M. K. (2014), “Follow the Leader with Dropout Perturbations”, *JMLR Workshop and Conference Proceedings 35, Proceedings of 27th Conference on Learning Theory (COLT 14)*, pp. 949-974, Barcelona, June 14. [pdf](#)
Colt talk [pdf](#)
More recent talk [pdf](#)
- C92 Nie, J., Kotłowski, W., and Warmuth, M. K. (2013), “On-line PCA with Optimal Regrets”, *Proceeding of the 24th International Conference on Algorithmic Learning Theory (ALT 13)*, Lecture Notes in Artificial Intelligence 8139, Springer-Verlag, pp. 98-112, Singapore, October 2013. [pdf](#)
Talk [pdf](#)
- C91 Koolen, W. M., Nie, J., and Warmuth, M. K., “Learning a set of directions”, 26th Annual Conference on Learning Theory (COLT 13), in *Proceedings of Machine Learning Research* vol. 30:851-866, Princeton, June 2013. [pdf](#)
Poster [pdf](#)
Talk [pdf](#)
- C90 Koolen, W. M., Adamskiy, D., and Warmuth, M. K., (2012) “Putting Bayes to sleep”, *Advances in Neural Information Processing Systems 25 (NeurIPS 12)*, Lake Tahoe, December 2012. [pdf](#)

- Poster [pdf](#)
- Latest talk [pdf](#)
- C89 Warmuth, M. K., Kotłowski, W., and Zhou, S. (2012) “Kernelization of matrix updates, when and how?” *Proceeding of the 23rd International Conference on Algorithmic Learning Theory (ALT 12)*, Lecture Notes in Artificial Intelligence 7568, Springer-Verlag, Lyon, France, October 2012. [pdf](#)
Talk [pdf](#)
- C88 Warmuth, M. K., Kotłowski, W., and Koolen, W. M., (2011) “Learning Eigenvectors for free”, *Advances in Neural Information Processing Systems 24 (NeurIPS 11)*, pp. 945–953, Granada Spain, December 2011. [pdf](#)
- C87 Warmuth, M. K., Koolen, W. M., and Helmbold, D. P. (2011) “Combining initial segments of lists”, *Proceeding of the 22nd International Conference on Algorithmic Learning Theory (ALT 11)*, Springer-Verlag, pp. 219–233, Helsinki, October 2011. [pdf](#)
Talk [pdf](#)
- C86 Abernethy, J., and Warmuth, M. K. (2011), “Repeated games against budgeted adversaries”, *Advances in Neural Information Processing Systems 23 (NeurIPS 10)*, pp. 1–9, Vancouver, Canada, December 2010. [pdf](#)
- C85 Wouter M. Koolen, Warmuth, M. K., and Kivinen, J. (2010) “Hedging structured concepts”, *Proceedings of the 23rd Annual Conference on Learning Theory (COLT 10)*, pp. 93-104, Haifa, June 2010. [pdf](#)
Talk [pdf](#)
- C84 Hazan, E., Kale, S., and Warmuth, M. K. (2010) “Learning rotations with little regret”, *Proceedings of the 24rd Annual Conference on Learning Theory (COLT 10)*, pp. 144-154, Haifa, June 2010. [pdf](#)
Talk [pdf](#)
Corrigendum [pdf](#)
- C83 Warmuth, M. K., Glocer, K. A., and Vishwanathan, S.V.N. (2008) “Entropy Regularized LPBoost”, in *Proceeding of the 19th International Conference on Algorithmic Learning Theory (ALT 08)*, Springer-Verlag, pp. 256-271, Budapest, October 2008. [pdf](#)
Talk [pdf](#)
Talk w. connection to zero-sum games [pdf](#)
Thesis w. much more yet unpublished related work [pdf](#)
- C82 Abernethy, J., Warmuth, M. K., and Yellin, J. (2008) “Optimal Strategies from Random Walks” *Proceedings of the 21st Annual Conference on Learning Theory (COLT 08)*, pp. 437–445, Helsinki, July 2008. [pdf](#)
Talk [pdf](#)
- C81 Warmuth, M. K., Glocer, K., and Rätsch, G. (2007) “Boosting Algorithms for Maximizing the Soft Margin”, *Advances in Neural Information Processing Systems 20 (NeurIPS 07)*, MIT Press, pp. 1585–1592, December 2007. [pdf](#)
Poster [pdf](#)
- C80 Warmuth, M. K., (2007) “Winnowing Subspaces”, *Proceedings of the 24th International Conference for Machine Learning (ICML 07)*, ACM International Conference Proceedings Series, Corvallis, Oregon, pp. 999-1006, June 2007. [pdf](#)
Talk [pdf](#)
- C79 Kuzmin, D., and Warmuth, M. K. (2007) “Online Kernel PCA with Entropic Matrix Updates”, *Proceedings of the 24th International Conference for Machine Learning (ICML 07)*, ACM International Conference Proceedings Series, Corvallis, Oregon, pp. 465-472, June 2007. [pdf](#)
Talk [pdf](#)

- C78 Helmbold, D. P., and Warmuth, M. K. (2007) “Learning Permutations with Exponential Weights”, *Proceedings of the 20th Annual Conference on Learning Theory (COLT 07)*, pp. 469-483, San Diego, June 2007. [pdf](#)
Talk [pdf](#)
- C77 Warmuth, M. K., and Kuzmin, D. (2006) “Randomized PCA Algorithms with Regret Bounds that are Logarithmic in the Dimension”, *Advances in Neural Information Processing Systems 19 (NeurIPS 06)*, MIT Press, pp. 1481-1498, Cambridge, MA., December 2006. [pdf](#)
Talk [pdf](#)
- C76 Warmuth, M. K., and Kuzmin, D. (2006) “A Bayesian Probability Calculus for Density Matrices”, in *Proceedings of the Twentysecond Conference on Uncertainty in Artificial Intelligence (UAI 06)*. Morgan Kaufmann, San Francisco, pp. 503-511, August 2006. [pdf](#)
Talk [pdf](#) [video](#)
- C75 Warmuth, M. K., Liao, J., and Rätsch, G. (2006) “Totally Corrective Boosting Algorithms that Maximize the Margin”, *Proceedings of the 23th International Conference for Machine Learning (ICML 06)*, ACM International Conference Proceedings Series, vol. 148, pp. 1001-1008 Carnegie Mellon, Pittsburgh PA., June 2006. [pdf](#)
Talk [pdf](#)
- C74 Warmuth, M. K., and Kuzmin, D. (2006) “Online Variance Minimization” *Proceedings of the 19th Annual Conference on Learning Theory (COLT 06)*, pp. 514-528, Carnegie Mellon, Pittsburgh PA, June 2006. [pdf](#)
Talk [pdf](#)
- C73 Abernethy, J., Langford, J., and Warmuth, M. K. (2006) “Continuous Experts and the Binning Algorithm”, *Proceedings of the 19th Annual Conference on Learning Theory (COLT 06)*, Carnegie Mellon, pp. 544-558, Pittsburgh PA, June 2006. [pdf](#)
Abernethy’s talk [pdf](#)
- C72 Warmuth, M. K. (2005) “A Bayes Rule for Density Matrices”, *Advances in Neural Information Processing Systems 18 (NeurIPS 05)*, MIT Press, Cambridge, MA., December 2005. [pdf](#)
Talk [pdf](#)
- C71 Warmuth, M. K., and Vishwanathan, S.V.N. (2005) “Leaving the Span”, in *Proceedings of the 18th Annual Conference on Learning Theory (COLT 05)*, Bertinoro, Italy, pp. 366-381, June 2005. [pdf](#)
Talk [pdf](#)
- C70 Kuzmin, D., and Warmuth, M. K. (2005) “Unlabeled Compression Schemes for Maximum Classes”, in *Proceedings of the 18th Annual Conference on Computational Learning Theory (COLT 05)*, Bertinoro, Italy, pp. 591-605, June 2005. [pdf](#)
Talk [pdf](#)
- C69 Tsuda, K., Rätsch, G., and Warmuth, M. K. (2004) “Matrix Exponentiated Gradient Updates for On-line Learning and Bregman Projections”, *Advances in Neural Information Processing Systems 17 (NeurIPS 04)*, MIT Press, Cambridge, MA., December 2004. [pdf](#)
- C68 Hatano, K., and Warmuth, M. K. (2003) “Boosting Versus Covering”, *Advances in Neural Information Processing Systems 16 (NeurIPS 03)*, MIT Press, Cambridge, MA., December 2003. [pdf](#)
Talk [pdf](#)
- C67 Garg, A., and Warmuth, M. K. (2003) “Inline Updates for HMMs”, invited submission to special session on “Advanced Machine Learning Algorithms for Speech and Language Processing” at *EuroSpeech 2003* (organised by Mazin Rahim and Rob Schapire), pp. 1005-1008, Geneve, Switzerland, September 2003. [pdf](#)

- C66 Singh, R., Warmuth, M. K., Raj, B. and Lamere, P. (2003) “Classification with Free Energy at Raised Temperatures”, *EuroSpeech* 2003, pp. 1773-1776, Geneva, Switzerland, September 2003. [pdf](#)
Poster [pdf](#)
- C65 Kivinen, J., Warmuth, M. K., and Hassibi, B. (2003) “The p-norm Generalization of the LMS Algorithm for Adaptive Filtering”, in the *13th IFAC Symposium on System Identification (SYSID-2003)*, Rotterdam, Netherlands, Elsevier Publishing Company, pp. 1755-1760, August 2003. [pdf](#)
- C64 Gramacy, R. B., Warmuth, M. K., Brandt, S. A., and Ari, I. (2002) “Adaptive Caching by Refetching”, *Advances in Neural Information Processing Systems 15 (NeurIPS 02)*, MIT Press, Cambridge, MA., December 2002. [pdf](#)
Talk [pdf](#)
MS thesis with better algorithms and experiments [pdf](#)
- C63 Takimoto, E., and Warmuth, M. K. (2002) “Path Kernels and Multiplicative Updates”, in *Proceedings of the 15th Annual Conference on Computational Learning Theory (COLT 02)*, Sidney, Australia, pp. 74–89, July 2002. [pdf](#)
- C62 Rätsch, G., and Warmuth, M. K. (2002) “Maximizing the Margin with Boosting”, in *Proceedings of the 15th Annual Conference on Computational Learning Theory (COLT 02)*, Sidney, Australia, pp. 334–350, July 2002. [pdf](#)
- C61 Rätsch, R., Mika, S., and Warmuth, M. K., “On the Convergence of Leveraging”, *Advances in Neural Information Processing Systems 14 (NeurIPS 01)*, pp. 487–494, MIT Press, Cambridge, MA., December 2001. [pdf](#)
- C60 Warmuth, M. K., Rätsch, R., Mathieson, M., Liao, J., and Lemmen, C. “Active Learning in the Drug Discovery Process”, *Advances in Neural Information Processing Systems 14 (NeurIPS 01)*, pp. 1449-1456, MIT Press, Cambridge, MA, December 2001. [pdf](#)
- C59 Bousquet, O., and Warmuth, M. K. (2001) “Tracking a Small Set of Experts by Mixing Past Posteriors”, in *Proceedings of the 14th Annual Conference on Computational Learning Theory (COLT 01)*, pp. 31-47, Amsterdam, Fulcrum award for best student paper, July 2001.
- C58 Takimoto, E., and Warmuth, M. K. (2000) “The Last-step Minimax Algorithm”, in *Proceeding of the 11th International Conference on Algorithmic Learning Theory (ALT 00)*, Lecture Notes in Artificial Intelligence, volume 1968, Springer Verlag. pp. 279–290, Sydney, Australia, December 2000. [pdf](#)
- C57 Räetsch, G., Warmuth, M. K., Mika, S., Onoda, T., Lemm, S., and Mueller, K. (2000). “Barrier Boosting”, in *Proceedings of the 13th Annual Conference on Computational Learning Theory (COLT 00)*, Stanford, CA, Morgan Kaufmann, San Francisco, CA, pp. 170-179, June 2000. [pdf](#)
- C56 Takimoto, E., and Warmuth, M. K. (2000) “The Minimax Strategy for Gaussian Density Estimation”, in *Proceedings of the 13th Annual Conference on Computational Learning Theory (COLT 00)*, Stanford, CA, Morgan Kaufmann, San Francisco, CA, pp. 100-106, June 2000. [pdf](#)
- C55 Forster, J., and Warmuth, M. K. (2000) “Relative Expected Instantaneous Loss Bounds”, in *Proceedings of the 13th Annual Conference on Computational Learning Theory (COLT 00)*, Stanford, CA, Morgan Kaufmann, San Francisco, CA, pp. 90-99, June 2000. [pdf](#)
- C54 Forster, J., and Warmuth, M. K. (2000) “Relative Loss Bounds for Temporal Difference Learning”, *Proceedings of the 17th International Conference (ICML 00)*, Stanford, CA, Morgan Kaufmann, San Francisco, CA, pp. 295-302, June 2000.
- C53 Takimoto, E., and Warmuth, M. K. (1999) “Predicting Nearly as Well as the Best Pruning of a Planar Decision Graph”, in *Proceeding of the 10th International Conference on Algorithmic Learning Theory (ALT 99)*, Tokyo, Springer Verlag, Berlin, pp. 335–346, December 1999. [pdf](#)

- C52 Azoury, K. S., and Warmuth, M. K. (1999) “Relative Loss Bounds for On-line Density Estimation with the Exponential Family of Distributions”, in *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI 99)*, Stockholm, Morgan Kaufmann, San Francisco, pp. 31–40, July 1999. [pdf](#)
- C51 Kivinen, J., and Warmuth, M. K. (1999) “Boosting as Entropy Projection”, in *Proceedings of the 12th Annual Conference on Computational Learning Theory (COLT 99)*, Santa Cruz, CA, ACM Press, New York, NY, pp. 134–144, July 1999. [pdf](#)
- C50 Kivinen, J., and Warmuth, M. K. (1999) “Averaging Expert Predictions”, *Proceedings of EUROCOLT 1999*, Springer-Verlag, pp. 153–167, March 1999. [pdf](#)
- C49 Helmbold, D. P., Panizza, S., and Warmuth, M. K. (1999) “Direct and Indirect Algorithms for On-line Learning of Disjunctions”, *Proceedings of EUROCOLT 1999*, Springer-Verlag, pp. 138–152, March 1999. [pdf](#)
- C48 Singer, Y. and Warmuth, M. K. (1998) “Batch and On-line Parameter Estimation of Gaussian Mixtures Based on the Joint Entropy”, *Advances in Neural Information Processing Systems 11 (NeurIPS 98)*, MIT Press, Cambridge, MA, pp. 578–584, December 1998. [pdf](#)
- C47 Gentile, C. and Warmuth, M. K. (1998) “Linear Hinge Loss and Average Margin”, *Advances in Neural Information Processing Systems 11 (NeurIPS 98)*, MIT Press, Cambridge, MA, pp. 225–231, December 1998. [pdf](#)
- C46 Herbster, M. and Warmuth, M.K. (1998) “Tracking the Best Regressor”, in *Proceedings 11th Annual Conference on Computational Learning Theory (COLT 98)*, pp. 24–31, July 1998. [pdf](#)
- C45 Warmuth, M. K., and Jagota, A. (1998) “Continuous And Discrete-Time Nonlinear Gradient Descent: Relative Loss Bounds and Convergence,” in the *Fifth International Symposium on Artificial Intelligence and Mathematics*, Florida, January 1998. [pdf](#)
- C44 Kivinen, J., and M.K. Warmuth (1997) “Relative Loss Bounds for Multidimensional Regression Problems”, *Advances in Neural Information Processing Systems 10 (NeurIPS 97)*, MIT Press, Cambridge, MA”, pp. 287–293, December 1997. [pdf](#)
- C43 Freund, Y., R.E. Schapire, Y. Singer, and M.K. Warmuth (1997) “Theory and Applications of Predictors that Specialize”, *Twenty Ninth Annual ACM Symposium on Theory of Computing (STOC 97)*, pp. 334–343, May 1997. [pdf](#)
- C42 Singer, Y., and M.K. Warmuth (1996) “Training Algorithms for Hidden Markov Models Using Entropy Based Distance Function”, *Advances in Neural Information Processing Systems 9 (NeurIPS 96)*, MIT Press, Cambridge, MA, pp. 641–647, Denver, CO., December 1996. [pdf](#)
- C41 Helmbold, D.R., R.E. Schapire, Y. Singer, and M.K. Warmuth (1996) “On-line Portfolio Selection Using Multiplicative Updates,” *Machine Learning: Proceedings of the Thirteenth International Conference (ICML 96)*, Bari, Italy, Morgan Kaufmann Publishers, pp. 243–251, San Francisco, CA, July 1996. [pdf](#)
- C40 Auer, P., S. Kwek, M. Maass, and M.K. Warmuth (1996) “Learning of Depth Two Neural Networks with Constant Fan-in at the Hidden Nodes,” in *Proceedings 9th Annual Conference on Computational Learning Theory*, pp. 333–343. Bari, Italy, ACM Press, New York, NY, July 1996. [pdf](#)
- C39 Helmbold, D. P., J. Kivinen and M.K. Warmuth (1995) “Worst-case Loss Bounds for Single Neurons,” *Advances in Neural Information Processing Systems (NeurIPS ‘95)*, Morgan Kaufmann Publishers, pp. 309–315, December 1995. [pdf](#)
- C38 Auer, P., M. Herbster, and M.K. Warmuth (1995) “Exponentially Many Local Minima for Single Neurons”, *Advances in Neural Information Processing Systems 8 (NeurIPS 95)*, Morgan Kaufmann Publishers, pp. 315–322, December 1995. [pdf](#)

- C37 Auer, P. and M.K. Warmuth (1995) “Tracking the Best Disjunction,” in *Proc. of the 36th Symposium on the Foundations of Comp. Sci. (FOCS 95)*, pp. 312–321, Milwaukee, WI, October 1995, [pdf](#)
- C36 Herbster, M. and Warmuth, M.K. (1995) “Tracking the Best Expert,” in *Machine Learning: Proceedings of the Twelfth International Conference (ICML 95)*, Morgan Kaufmann Publishers, San Francisco, CA, pp. 286–294, Tahoe City, CA, July 1995.
- C35 Maass, M. and Warmuth, M.K. (1995) “Efficient Learning with Virtual Threshold Gates,” in *Machine Learning: Proceedings of the Twelfth International Conference (ICML 95)*, Morgan Kaufmann Publishers, San Francisco, CA, pp. 378–386, Tahoe City, CA, July 1995.
- C34 Helmbold, D.R., Schapire, R.E., Singer, Y., and Warmuth M.K.(1995) “A Comparison of New and Old Algorithms for a Mixture Estimation Problem,” in *Proc. 8th Annual Conference on Computational Learning Theory (COLT 95)*, ACM Press, New York, NY, pp. 69–78, Santa Cruz, CA, July 1995. [pdf](#)
- C33 Kivinen, J. and Warmuth, M.K. (1995) “The Perceptron Algorithm vs. Winnow: linear vs. logarithmic mistake bounds when few input variables are relevant,” in *Proc. 8th Annual Conference on Computational Learning Theory (COLT 95)*, ACM Press, New York, NY, pp. 289–300, Santa Cruz, CA, July 1995. [pdf](#)
- C32 Kivinen, J., and Warmuth, M. K. (1995) “Additive Versus Exponentiated Gradient Updates for Linear Prediction,” *Proceedings of Twenty Seventh Annual ACM Symposium on Theory of Computing (STOC 95)*, pp. 209–218, Reno, May 1995. [pdf](#)
- C31 Haussler, D., Kivinen, J. and Warmuth, M. K. (1995) “Tight Worst-case Loss Bounds for Predicting with Expert Advice,” *Proceedings of 2nd EUROCOLT 1995*, pp. 69–83, Barcelona, Spain, March 1995. [pdf](#)
- C30 Schapire, R. E., and Warmuth, M. K. (1994) “On the Worst-case Analysis of Temporal-difference Learning Algorithms,” in the *Proceedings of Eleventh International Conference on Machine Learning (ICML 94)*, pp. 266–274, July 1994.
- C29 Cesa-Bianchi, N., Freund, Y., Helmbold, D. P., and Warmuth, M. K. (1993) “On-line Prediction and Conversion Strategies,” *Proceedings of EUROCOLT 1993*.
- C28 Kivinen, J., and Warmuth, M. K. (1993) “Using Experts for Predicting Continuous Outcomes,” In *Computational Learning Theory: EUROCOLT 93*, “New Series Number 53”, pages 109-120. Clarendon Press, Oxford, March 1994. [pdf](#)
- C27 Cesa-Bianchi, N., Long, P. M., and Warmuth, M. K. (1993) “Worst-case Quadratic Loss Bounds for a Generalization of the Widrow-Hoff Rule,” in the *Proceedings of the Sixth Annual Workshop on Computational Learning Theory (COLT 93)*, Santa Cruz, pp. 429–438, July 1993. [pdf](#)
- C26 Goldman, S. A. and Warmuth, M. K. (1993) “Learning Binary Relations Using Weighted Majority Voting,” in the *Proceedings of the Sixth Annual Workshop on Computational Learning Theory*, Santa Cruz, pp. 453–462, July 1993. [pdf](#)
- C25 Cesa-Bianchi, N., Freund, Y., Helmbold, D. P., Haussler, D., Schapire, R. E., and Warmuth, M. K. (1993) “How to Use Expert Advice” *Proceedings of the Twenty Fifth Annual ACM Symposium on Theory of Computing (STOC 93)*, San Diego, 382–391, May 1993. [pdf](#)
- C24 Helmbold, D., and Warmuth, M. K. (1992) “Some Weak Learning Results” *Proceedings of the Fifth Annual Workshop on Computational Learning Theory (COLT 92)*, Pittsburg, PA, pp. 399–415, July 1992. [pdf](#)
- C23 Helmbold, D., and Warmuth, M. K. (1992) “On Weak Learning” *Proceedings of the Second NEC Research Symposium on Computation and Cognition*, Princeton, New Jersey, pp 101–125, May 1992.

- C22 Haussler, D. and M. K. Warmuth (1992) “The Probably Approximately Correct (PAC) and Other Learning Models,” in *The Mathematics of Generalization, Proceedings of the SFI/CNLS Workshop on Formal Approaches to Supervised Learning*, pp. 17–36, June 1992.
- C21 Abe, N., Takeuchi, J., and Warmuth, M. K. (1991) “Polynomial Learnability of Probabilistic Concepts with respect to the Kullback-Leibler Divergence” *Proceedings of the Fourth Annual Workshop on Computational Learning Theory (COLT 91)*, Santa Cruz, California, August 1991.
- C20 Littlestone, N., Long, P. M., and Warmuth, M. K. (1991) “On-Line Learning of Linear Functions” *Proceedings of Twenty Third Annual ACM Symposium on Theory of Computing (STOC 91)*, New Orleans, pp. 465–475, May 1991. [pdf](#)
- C19 Long, P. M., and Warmuth, M. K. (1990) “Composite Geometric Concepts and Polynomial Predictability” *Proceedings of Second Workshop on Computational Learning Theory (COLT 90)*, pp. 273–287, Rochester, NY, August 1990.
- C18 Abe, N., and Warmuth, M. K. (1990) “On the Computational Complexity of Approximating Distributions by Probabilistic Automata,” *Proceedings of Second Workshop on Computational Learning Theory (COLT 90)*, pp. 52–66, Rochester, NY, August 1990.
- C17 Helmbold, D., Sloan, R., and Warmuth, M. K. (1990) “Learning Integer Lattices” *Proceedings of Second Workshop on Computational Learning Theory (COLT 90)*, pp. 288–300, Rochester, NY, August 1990.
- C16 Warmuth, M. K. (1989) “Towards Representation Independence in PAC Learning,” Invited contribution to *Proceedings of the International Workshop on Analogical and Inductive Inference (AII) 1989*. Subseries of *Lecture Notes in Computer Science 397*, pp. 78–103, edited by J. Siekmann, Springer Verlag. Reinhardtsbrunn Castle, GDR October 1989. [pdf](#)
- C15 Bodlaender, H. L., Moran, S., and Warmuth, M. K. (1989) “The Distributed Bit Complexity of the Ring: From the anonymous to the Non-anonymous Case”, *Proceedings of Fundamentals of Computation Theory 1989*, Szeged, Hungary, Springer Verlag *Lecture Notes in Computer Science 380*, pages 58–67, August 1989.
- C14 Littlestone, N. and Warmuth, M. K. (1989) “The Weighted Majority Algorithm,” *Proceedings of Thirtieth Annual IEEE Symposium on Foundations of Computer Science (FOCS 89)*, pp. 256–261, October 1989. Also presented at *Second Workshop on Computational Learning Theory (COLT 89)*, Santa Cruz, California, July 1989. **Thirty year test of time award for FOCS.** [pdf](#)
- C13 Helmbold, D., Sloan, R., and Warmuth, M. K. (1989) “Learning Nested Differences of Intersection-closed Concept Classes” *Proceedings of Second Annual Workshop on Computational Learning Theory (COLT 89)*, Morgan Kaufmann Publishers Inc., San Francisco, pp. 41–56. Santa Cruz, California, July 1989,
- C12 Pitt, L., and Warmuth, M. K. (1989) “The Minimum consistent DFA Problem Cannot be Approximated Within any Polynomial” *Proceedings of Twenty First Annual ACM Symposium on Theory of Computing (STOC 89)*, pp. 421–432, Seattle, Washington, May 1989. Also presented at *Forth Annual Conference of Structure in Complexity Theory*, Portland, Oregon, May 1989. [pdf](#)
- C11 Anderson, R., Mayr, E., and Warmuth, M. K. (1988) “Parallel Approximation Algorithms for Bin Packing”, *Methods of Operations Research 60, Proceedings of the XIII. Symposium on Operations Research*, Paderborn, September 1988.
- C10 Haussler, D., Littlestone, N., and Warmuth, M. K. (1988) “Predicting 0,1 Functions on Randomly Drawn Points” *Proceedings of Twenty Ninth Annual IEEE Symposium on Foundations of Computer Science (FOCS 88)*, pp. 100–109, October 88. Also in *Proceeding of First Workshop on Computational Learning Theory (COLT 88)*, pp. 280–296, Boston, Massachusetts, August 1988. [pdf](#)

- C9 Haussler, D., Littlestone, N., Kearns, M., and Warmuth, M. K. (1988) “Equivalence of Models for Polynomial Learnability” *Proceedings of the first Workshop on Computational Learning Theory (COLT 88)*, pp. 42–55, Boston, Massachusetts, August 1988.
- C8 Pitt, L., and Warmuth, M. K. (1988) “Reductions Among Prediction Problems: On the Difficulty of Predicting Automata” *Proceeding of Third Annual Conference of Structure in Complexity Theory (STRUCTURES 88)*, Washington, DC, pp. 60–69, June 1988. [pdf](#)
- C7 Moran, Shlomo, and Warmuth, M. K. (1986) “Gap Theorems for Distributed Computation” *Proceeding of Fifth Annual ACM Symposium on Principles of Distributed Computing (PODC 86)*, pp. 131–140, Calgary, Alberta, Canada, August 1986. [pdf](#)
- C6 Ratner, Daniel, and Warmuth, M. K. (1986) “The Natural Extension of the 15-puzzle is Intractable” *Proceedings of AAAI Fifth National Conference on Artificial Intelligence*. Philadelphia, Pennsylvania, August 1986.
- C5 Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M. K. (1986) “Classifying Learnable Geometric Concepts with the Vapnik-Chervonenkis Dimension” *Proceedings of Eighteenth Annual ACM Symposium on Theory of Computing (STOC 86)*, pp. 273–282, Berkeley, CA, May 1986. [pdf](#)
- C4 Dahlhaus, E., and Warmuth, M. K. (1986) “Membership for Growing Context Sensitive Grammars is Polynomial” *Proceedings of Eleventh Colloquium on Trees in Algebra and Programming (CAAP)*, pp. 85–99. Nice, France, March 1986. [pdf](#)
- C3 Attiya, C., Snir, M., and Warmuth, M. K. (1985) “Computing on an Anonymous Ring” *Proceedings of Fourth Annual ACM Symposium on Principles of Distributed Computing (PODS 85)*, pp. 196–203. Minaki, Ontario, Canada, August 1985.
- C2 Haussler, D., and Warmuth, M. K. (1985) “On the Complexity of Iterated Shuffle” *Workshop on Computational Linguistics*. Santa Cruz, California.
- C1 Dolev, D., Upfal, E., and Warmuth, M. K. (1984) “Scheduling Trees in Parallel” *Proceedings of International Workshop on Parallel Computing and VLSI*, pp. 93–102. Amalfi, Italy, May 1984.

Open Problems

- P13 Manfred K. Warmuth and Ehsan Amid, “Learning sparse linear concepts by priming the features”, *Proceedings of the 36th Annual Conference on Learning Theory (COLT 23)*, in JMLR: Workshop and Conference Proceedings vol 195:1–6 (2023). [pdf](#)
Talk [pdf](#)
- P12 Wouter M. Koolen, Manfred K. Warmuth and Dmitry Adamskiy, “Online Sabotaged Shortest Path”, *Proceedings of the 28th Annual Conference on Learning Theory (COLT 15)*, in JMLR: Workshop and Conference Proceedings vol 40:1–24 (2015), pp. 1-3, Paris, July 2015. [pdf](#)
- P11 Manfred K. Warmuth and Wouter M. Koolen, “Shifting Experts on Easy Data”, *Proceedings of the 27th Annual Conference on Learning Theory (COLT 14)*, in JMLR: Workshop and Conference Proceedings 35 (2014), pp. 1295-1298, Barcelona, June 2014. [pdf](#)
Talk [pdf](#)
- P10 Nie, J., Warmuth, M.K., Vishwanathan, S.V.N. and Zhang, X. (2013) “Lower bounds for Boosting with Hadamard Matrices”, *Proceedings of the 26th Annual Conference on Learning Theory (COLT 13)*, in JMLR: Workshop and Conference Proceedings 30 (2013), pp. 1076-1079, Princeton, June 2013. [pdf](#)
Talk [pdf](#)
- P9 Kotłowski, W. and Warmuth, M.K. (2011) “Minimax algorithm for learning rotations”, *Proceedings of the 24th Annual Conference on Learning Theory (COLT 11)*, in JMLR: Workshop and Conference Proceedings 19 (2011), pp. 821-823, Budapest, June 2011. [pdf](#)
Talk [pdf](#)
- P8 Hazan, E., Kale, S., and Warmuth, M. K. (2010) “On-line variance minimization in $O(n^2)$ per trial?”, *Proceedings of the 23rd Annual Conference on Learning Theory (COLT 10)*, pp. 314-315, Haifa, June 2010. [pdf](#)
\$100 for solution! [pdf](#)
Talk [pdf](#)
- P7 Abernethy, J., and Warmuth, M. K. (2009) “Minimax Games with Bandits”, *Proceedings of the 22nd Annual Conference on Learning Theory (COLT 09)*, Toronto, June 2009. [pdf](#)
Talk [pdf](#)
- P6 Smith, A., and Warmuth, M. K. (2008) “Learning rotations”, *Proceedings of the 21th Annual Conference on Learning Theory (COLT 08)*, p. 517, Helsinki, July 2008. [pdf](#)
Talk [pdf](#)
- P5 Warmuth, M. K. (2007) “When is there a free matrix lunch,” *Proceedings of the 20th Annual Conference on Learning Theory (COLT 07)*, San Diego, June 2007. [pdf](#)
Talk [pdf](#)
- P4 Warmuth, M. K. (2006) “Can entropic regularization be replaced by squared Euclidean distance plus additional linear constraints,” *Proceedings of the 19th Annual Conference on Learning Theory (COLT 06)*, Carnegie Mellon, pp. 653-654, June 2006. [pdf](#)
- P3 Kuzmin, D., and Warmuth, M. K. (2005) “Optimum Follow the Leader Algorithm,” *Proceedings of the 18th Annual Conference on Computational Learning Theory (COLT 05)*, Bertinoro, Italy. [pdf](#)
Poster [pdf](#)
- P2 Warmuth, M. K. (2004) “The Optimal PAC Algorithm,” *Proceedings of the 17th Annual Conference on Computational Learning Theory (COLT 04)*, Banff, Canada, pp. 641–642. [pdf](#)
- P1 Warmuth, M. K. (2003) “Compressing to VC Dimension Many Points,” *Proceedings of the 16th Annual Conference on Computational Learning Theory (COLT 03)*, Washington, DC, pp. 743–744. [pdf](#)
\$600 for solution! [pdf](#)

Tutorials

- 2011 Machine Learning Summer School, “Survey of Boosting from an Optimization Perspective”, Purdue University, West Lafayette, Indiana, June 13 - June 24, 2011 [pdf](#) [\[video\]](#)
- 2010 Canberra “Recent Advances in Boosting Algorithms”, Machine Learning Summer School, Australian National University, Canberra, Australia, Sept. 27 - Oct. 6, 2010 [pdf](#)
- 2009 “Survey of Boosting from an Optimization Perspective” ICML 2009, Toronto, Canada, June 14, 2009 (with S.V. N. Vishwanathan) [pdf](#) [\[video\]](#)
- 2006 “On-line Learning and Bregman Divergences” Machine Learning Summer School, National Taiwan University of Science and Technology, Taipei, Taiwan, July 24 - Aug. 4, 2006 [pdf](#) [\[video\]](#)
- 2002 “On-line Learning”, DIMACS workshop on “Mining Massive Data Sets and Streams: Mathematical Methods and Algorithms for Homeland Defense”, Center for Communications Research, Princeton, New Jersey, June 18, 2001
- 2002 “On-line Learning and Bregmann Divergences” in Machine Learning Summer School, Australian National University, Canberra, Australia, Feb. 11 - Feb. 22, 2001 (with Gunnar Rätsch). (Lectures 1 [pdf](#), 2 [pdf](#), 3 [pdf](#)).
- 2001 “On-line Learning Algorithms - Methods and Open Problems”, Dagstuhl seminar on “Inference Principles and Model Selection”, organized by Joachim Buhmann and Bernhard Schoelkopf, July 26, 2001. [pdf](#)
- 2000 “Proving Relative Loss Bounds for On-line Learning Algorithms Using Bregman Divergences”, tutorial at the *13th Annual Conference on Computational Learning Theory*, June 2000 (with Claudio Gentile). [pdf](#)
- 1999 “On-line Learning and Relative Loss Bounds”, tutorial at *Advances in Neural Information Processing Systems 12 (NeurIPS 99)*, December 1999.
- 1999 “Relative Loss Bounds for On-line Learning Using Divergence Measures”, DIMACS Workshop: On-line Decision Making, organized by Yoav Freund and Rakesh Vohra, Rutgers University, July 12-15, 1999.
- 1997 “Multiplicative update algorithms - how to derive, apply and analyze this new family of learning algorithm”, tutorial at the *Fourteenth International Conference on Machine Learning (ICML 97)*, Nashville, TN, July 1997.
- 1995 “Amortized Analysis for On-Line Learning Algorithms”, tutorial at the *NeuroCOLT Workshop on Agnostic On-line Prediction*, Helsinki, September 1995.
- 1995 “Amortized Analysis for On-Line Learning Algorithms”, tutorial at the *Twelfth International Conference on Machine Learning (ICML 95)*, Tahoe, July 1995.
- 1994 “A common framework for deriving on-line learning algorithms and for proving worst-case loss bounds”, tutorial at Dagstuhl, Germany, June 27, 1994.