

References

- [1] A. Argyriou, R. Hauser, C. Micchelli, and M. Pontil. A DC-programming algorithm for kernel selection. In *Proceedings of the 23rd Annual International Conference on Machine Learning*, 2006.
- [2] A. Argyriou, C. Micchelli, and M. Pontil. Learning convex combinations of continuously parameterized basic kernels. In *COLT*, 2005.
- [3] F. Bach. Exploring large feature spaces with hierarchical multiple kernel learning. *Advances in Neural Information Processing Systems*, 2008.
- [4] F. Bach, G. R. G. Lanckriet, and M. I. Jordan. Multiple kernel learning, conic duality, and the smo algorithm. In *Proceedings of the 21st Annual International Conference on Machine Learning*, 2004.
- [5] M.-F. Balcan and A. Blum. On a theory of learning with similarity functions. In *Proceedings of the 23rd International Conference on Machine Learning*, 2006.
- [6] A. Ben-Hur and W. S. Noble. Kernel methods for predicting protein-protein interactions. In *ISMB, Supplement of Bioinformatics*, 2005.
- [7] B. E. Boser, I. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *COLT*, volume 5, pages 144–152, 1992.
- [8] O. Bousquet and D. J. L. Herrmann. On the complexity of learning the kernel matrix. In *Advances in Neural Information Processing Systems*, 2002.
- [9] S. Bucak, R. Jin, and A. Jain. Multi-label Multiple Kernel Learning by Stochastic Approximation: Application to Visual Object Recognition. *Advances in Neural Information Processing Systems*, 2010.
- [10] C. Chang and C. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27, 2011.
- [11] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee. Choosing multiple parameters for support vector machines. In *Machine Learning*, 2002.
- [12] T.-J. Chin, H. Wang, and D. Suter. Multi-structure model selection via kernel optimisation. *IEEE Computer Vision and Pattern Recognition*, 2010.

- [13] C. Cortes, P. Haffner, and M. Mohri. Rational kernels: Theory and algorithms. *The Journal of Machine Learning Research*, 5:1035–1062, 2004.
- [14] C. Cortes, M. Mohri, and A. Rostamizadeh. Learning sequence kernels. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2008), (invited lecture)*, Cancún, Mexico, October 2008.
- [15] C. Cortes, M. Mohri, and A. Rostamizadeh. L_2 regularization for learning kernels. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 109–116, 2009.
- [16] C. Cortes, M. Mohri, and A. Rostamizadeh. Learning non-linear combinations of kernels. *Advances in Neural Information Processing Systems*, 22:396–404, 2009.
- [17] C. Cortes, M. Mohri, and A. Rostamizadeh. Generalization bounds for learning kernels. *International Conference on Machine Learning*, 2010.
- [18] C. Cortes, M. Mohri, and A. Rostamizadeh. Two-stage learning kernel algorithms. In *Proceedings of the 27th Conference on Machine Learning*, 2010.
- [19] C. Cortes, M. Mohri, and A. Rostamizadeh. Ensembles of kernel predictors. *Uncertainty in Artificial Intelligence*, 2011.
- [20] C. Cortes and V. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3), 1995.
- [21] N. Cristianini, J. Shawe-Taylor, A. Elisseeff, and J. S. Kandola. On kernel-target alignment. In *Advances in Neural Information Processing Systems*, 2001.
- [22] A. Demiriz, K. Bennett, and J. Shawe-Taylor. Linear programming boosting via column generation. *Machine Learning*, 46(1), 2002.
- [23] A. Demiriz, K. Bennett, and J. Shawe-Taylor. Linear programming boosting via column generation. *Machine Learning*, 46(1):225–254, 2002.
- [24] F. Dinuzzo, C. Ong, P. Gehler, and G. Pillonetto. Learning output kernels with block coordinate descent. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.

- [25] O. Emanuelsson, H. Nielsen, S. Brunak, and G. von Heijne. Predicting subcellular localization of proteins based on their n-terminal amino acid sequence. *Journal of molecular biology*, 300(4):1005–1016, 2000.
- [26] R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.
- [27] V. Franc and S. Sonnenburg. Optimized cutting plane algorithm for large-scale risk minimization. *The Journal of Machine Learning Research*, 10:2157–2192, 2009.
- [28] K. Gai, G. Chen, and C. Zhang. Learning Kernels with Radiuses of Minimum Enclosing Balls. *Advances in Neural Information Processing Systems*, 2010.
- [29] P. Gehler and S. Nowozin. On feature combination for multiclass object classification. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 221–228, 2009.
- [30] Y. Grandvalet and S. Canu. Adaptive scaling for feature selection in SVMs. In *Advances in Neural Information Processing Systems*, 2003.
- [31] E. Hu, B. Wang, and S. Chen. BCDNPKL: Scalable non-parametric kernel learning using block coordinate descent. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.
- [32] Z. Hussain and J. Shawe-Taylor. Improved Loss Bounds For Multiple Kernel Learning. *International Conference on Artificial Intelligence and Statistics*, 2011.
- [33] P. Jain, B. Kulis, and I. Dhillon. Inductive regularized learning of kernel functions. *Advances in Neural Information Processing Systems*, 2010.
- [34] P. Jawanpuria, J. Nath, and G. Ramakrishnan. Efficient rule ensemble learning using hierarchical kernels. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.
- [35] T. Jebara. Multi-task feature and kernel selection for SVMs. In *Proceedings of the 21st International Conference on Machine Learning*, 2004.
- [36] R. Jin, S. Hoi, and T. Yang. Online multiple kernel learning: algorithms and mistake bounds. In *Algorithmic Learning Theory*, pages 390–404, 2010.

- [37] J. S. Kandola, J. Shawe-Taylor, and N. Cristianini. On the extensions of kernel alignment. technical report 120, Department of Computer Science, Univ. of London, UK, 2002.
- [38] S.-J. Kim, A. Zymnis, A. Magnani, K. Koh, and S. Boyd. Learning the kernel via convex optimization. In *ICASSP '08*, 2008.
- [39] M. Kloft, U. Brefeld, S. Sonnenburg, P. Laskov, K. Müller, and A. Zien. Efficient and accurate lp-norm multiple kernel learning. *Advances in Neural Information Processing Systems*, 22(22):997–1005, 2009.
- [40] M. Kloft, U. Brefeld, S. Sonnenburg, and A. Zien. Lp-norm multiple kernel learning. *The Journal of Machine Learning Research*, 2011.
- [41] V. Koltchinskii and D. Panchenko. Empirical margin distributions and bounding the generalization error of combined classifiers. *Annals of Statistics*, 30, 2002.
- [42] V. Koltchinskii and M. Yuan. Sparse recovery in large ensembles of kernel machines on-line learning and bandits. In *Conference on Learning Theory*, 2008.
- [43] A. Kundu, V. Tankasali, C. Bhattacharyya, and A. Ben-Tal. Efficient algorithms for learning kernels from multiple similarity matrices with general convex loss functions. *Advances in Neural Information Processing Systems*, 2010.
- [44] G. Lanckriet, N. Cristianini, P. Bartlett, L. Ghaoui, and M. Jordan. Learning the kernel matrix with semidefinite programming. *The Journal of Machine Learning Research*, 5:27–72, 2004.
- [45] G. Lanckriet, T. De Bie, N. Cristianini, M. Jordan, and W. Noble. A statistical framework for genomic data fusion. *Bioinformatics*, 20(16):2626, 2004.
- [46] D. P. Lewis, T. Jebara, and W. S. Noble. Nonstationary kernel combination. In *Proceedings of the 23rd International Conference on Machine Learning*, 2006.
- [47] D. Luenberger. *Linear and Nonlinear Programming*. Addison Wesley, 1984.
- [48] P. Machart, T. Peel, S. Anthoine, L. Ralaivola, and H. Glotin. Stochastic low-rank kernel learning for regression. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.

- [49] Q. Mao and I. Tsang. Parameter-Free Spectral Kernel Learning. *Uncertainty in Artificial Intelligence*, 2010.
- [50] A. Martins, N. Smith, E. Xing, P. Aguiar, M. Figueiredo, and P. Lisboa. Online Learning of Structured Predictors with Multiple Kernels. *International Conference on Artificial Intelligence and Statistics*, 2011.
- [51] M. Meila. Data centering in feature space. In *AISTATS*, 2003.
- [52] H. Mewes, D. Frishman, U. Güldener, G. Mannhaupt, K. Mayer, M. Mokrejs, B. Morgenstern, M. Münsterkötter, S. Rudd, and B. Weil. Mips: a database for genomes and protein sequences. *Nucleic acids research*, 30(1):31, 2002.
- [53] C. Micchelli and M. Pontil. Learning the kernel function via regularization. *JMLR*, 6, 2005.
- [54] L. Mukherjee, V. Singh, J. Peng, and C. Hinrichs. Learning kernels of normalized cuts: Convex relaxations and applications. *IEEE Computer Vision and Pattern Recognition*, 2010.
- [55] J. S. Nath, G. Dinesh, S. Raman, C. Bhattacharyya, and A. B.-T. K. R. Ramakrishnan. On the algorithmics and applications of a mixed-norm based kernel learning formulation. In *Advances in Neural Information Processing Systems*, 2009.
- [56] C. Ong and A. Zien. An automated combination of kernels for predicting protein subcellular localization. *Algorithms in Bioinformatics*, pages 186–197, 2008.
- [57] C. S. Ong, A. Smola, and R. Williamson. Learning the kernel with hyperkernels. *The Journal of Machine Learning Research*, 6, 2005.
- [58] F. Orabona and L. Jie. Ultra-fast optimization algorithm for sparse multi kernel learning. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.
- [59] F. Orabona, L. Jie, and B. Caputo. Online-batch strongly convex multi kernel learning. *IEEE Computer Vision and Pattern Recognition*, 2010.
- [60] J.-B. Pothin and C. Richard. Optimizing kernel alignment by data translation in feature space. In *ICASSP*, 2008.
- [61] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet. Simplemkl. *Journal of Machine Learning Research*, 9:2491–2521, 2008.

- [62] C. Saunders, A. Gammerman, and V. Vovk. Ridge regression learning algorithm in dual variables. In *Proceedings of the 15th International Conference on Machine Learning*, volume 521, 1998.
- [63] S. Sonnenburg and G. Rätsch. Accurate splice site prediction for *caenorhabditis elegans*. In *MIT Press series on Computational Molecular Biology*, pages 277–298, 2004.
- [64] S. Sonnenburg, G. Rätsch, S. Henschel, C. Widmer, J. Behr, A. Zien, F. Bona, A. Binder, C. Gehl, and V. Franc. The SHOGUN machine learning toolbox. *The Journal of Machine Learning Research*, 99:1799–1802, 2010.
- [65] S. Sonnenburg, G. Rätsch, and C. Schäfer. Learning interpretable svms for biological sequence classification. In *RECOMB*, 2005.
- [66] S. Sonnenburg, G. Ratsch, and C. Schafer. A general and efficient multiple kernel learning algorithm. In *Advances in Neural Information Processing Systems*, 2006.
- [67] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf. Large scale multiple kernel learning. *The Journal of Machine Learning Research*, 7:1531–1565, 2006.
- [68] S. Sonnenburg, A. Zien, and G. Rätsch. Arts: accurate recognition of transcription starts in human. *Bioinformatics*, 22(14), 2006.
- [69] N. Srebro and S. Ben-David. Learning bounds for support vector machines with learned kernels. In *COLT*, 2006.
- [70] M. Sugiyama, M. Yamada, M. Kimura, and H. Hachiya. On information-maximization clustering: Tuning parameter selection and analytic solution. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.
- [71] O. Tamuz, C. Liu, S. Belongie, O. Shamir, and A. Kalai. Adaptively learning the crowd kernel. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.
- [72] M. Tan, C. Wang, and E. Tsang. Learning sparse svm for feature selection on very high dimensional datasets. In *International Conference on Machine Learning*, 2010.

- [73] L. Tang, J. Chen, and J. Ye. On multiple kernel learning with multiple labels. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, 2009.
- [74] L. van der Maaten. Learning discriminative fisher kernels. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.
- [75] V. Vapnik and A. Y. Chervonenkis. A note on one class of perceptrons. *Automation and Remote Control*, 25, 1964.
- [76] M. Varma and B. Babu. More generality in efficient multiple kernel learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1065–1072, 2009.
- [77] M. Varma and D. Ray. Learning the discriminative power-invariance trade-off. In *Computer Vision, 2007. IEEE 11th International Conference on*, pages 1–8, 2007.
- [78] S. Vishwanathan, Z. Sun, N. Theera-Ampornpunt, and M. Varma. Multiple kernel learning and the smo algorithm. *Advances in Neural Information Processing Systems, Vancouver, BC, Canada*, 2010.
- [79] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik. Feature selection for svms. In *Advances in Neural Information Processing Systems*, 2000.
- [80] Z. Xu, R. Jin, H. Yang, I. King, and M. Lyu. Simple and efficient multiple kernel learning by group lasso. *International Conference on Machine Learning*, 2010.
- [81] F. Yan, K. Mikolajczyk, M. Barnard, H. Cai, and J. Kittler. ℓ -p Norm Multiple Kernel Fisher Discriminant Analysis for Object and Image Categorisation. *IEEE Computer Vision and Pattern Recognition*, 2010.
- [82] J. Yang, Y. Li, Y. Tian, L. Duan, and W. Gao. Group-sensitive multiple kernel learning for object categorization. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 436–443, 2009.
- [83] Y. Ying and C. Campbell. Generalization bounds for learning the kernel problem. In *Conference on Learning Theory*, 2009.
- [84] J. Zhuang, I. Tsang, and S. Hoi. Two-Layer Multiple Kernel Learning. *International Conference on Artificial Intelligence and Statistics*, 2011.

- [85] A. Zien and C. Ong. Multiclass multiple kernel learning. In *Proceedings of the 24th International Conference on Machine Learning*, pages 1191–1198, 2007.