# References

[Aka01] Shotaro Akaho, *A kernel method for canonical correlation analysis*, Proceedings of International Meeting on Psychometric Society (IMPS2001), 2001.

[AN01] Shun-Ichi Amari and Hiroshi Nagaoka, *Methods of information geometry*, AMS vol. 191, 2001.

[Aro50] N. Aronszajn, *Theory of reproducing kernels*, Transactions of the American Mathematical Society **68** (1950), 337 – 404.

[Bac08a] Francis Bach, *Exploring large feature spaces with hierarchical multiple kernel learning*, Advances in Neural Information Processing Systems 21, 2008, to appear.

[Bac08b] Francis R. Bach, *Consistency of the group lasso and multiple kernel learning*, J. Mach. Learn. Res. **9** (2008), 1179–1225.

[Bac08c] ———, *Graph kernels between point clouds.*, ICML (William W. Cohen, Andrew McCallum, and Sam T. Roweis, eds.), ACM International Conference Proceeding Series, vol. 307, ACM, 2008, pp. 25–32.

[BCR84] Christian Berg, Jens Peter Reus Christensen, and Paul Ressel, *Harmonic analysis on semigroups*, Graduate Texts in Mathematics, no. 100, Springer Verlag, 1984.

[Ber05] Dennis S. Bernstein, *Matrix mathematics: Theory, facts, and formulas with application to linear systems theory*, Princeton University Press, 2005.

[BH03] O. Bousquet and D.J.L. Herrmann, *On the complexity of learning the kernel matrix*, Advances in neural information processing systems (2003), 415–422.

[BJ02] Francis Bach and Michael Jordan, *Kernel independent component analysis*, Journal of Machine Learning Research **3** (2002), 1–48.

[BLJ04] Francis R. Bach, Gert R. G. Lanckriet, and Michael I. Jordan, *Multiple kernel learning, conic duality, and the smo algorithm*, ICML '04: Proceedings of the twenty-first international conference on Machine learning (New York, NY, USA), ACM, 2004, p. 6.

[Boc33] S. Bochner, *A theorem on Fourier-Stieltjes integrals*, Mathematische Annalen **108** (1933), 378–410.

[Boc48] ———, *Vorlesungen "uber Fouriersche Integrale*, New York (1948).

[BTA03] Alain Berlinet and Christine Thomas-Agnan, *Reproducing kernel hilbert spaces in probability and statistics*, Kluwer Academic Publishers, 2003.

[BV04] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[Cat04] Olivier Catoni, *Statistical learning theory and stochastic optimization*, LNM, no. 1851, Springer Verlag, 2004.

[CBK09] Varun Chandola, Arindam Banerjee, and Vipin Kumar, *Anomaly detection: A survey*, ACM Computing Surveys (2009).

[CF07] Marco Cuturi and Kenji Fukumizu, *Kernels on structured objects through nested histograms*, Advances in Neural Information Processing Systems 19 (B. Schölkopf, J. Platt, and T. Hoffman, eds.), MIT Press, Cambridge, MA, 2007.

[CFV05] Marco Cuturi, Kenji Fukumizu, and Jean-Philippe Vert, *Semigroup kernels on measures*, Journal of Machine Learning Research **6** (2005), 1169–1198.

[Cha07] Olivier Chapelle, *Training a support vector machine in the primal*, Neural Comput. **19** (2007), no. 5, 1155–1178.

[CHV99] O. Chapelle, P. Haffner, and V. Vapnik, *SVMs for histogram based image classification*, IEEE Transactions on Neural Networks **10** (1999), no. 5, 1055.

[CO02] Lehel Csató and Manfred Opper, *Sparse on-line gaussian processes*, Neural Computation **14** (2002), no. 3, 641–668.

[CS02] Felipe Cucker and Stephen Smale, *On the mathematical foundations of learning*, BAMS: Bulletin of the American Mathematical Society **39** (2002).

[Cut07] Marco Cuturi, *Permanents, transportation polytopes and positive-definite kernels on histograms*, Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI) 2007, 2007, pp. 732 – 737.

[CV95] Corinna Cortes and Vladimir Vapnik, *Support-vector networks*, Machine Learning **20** (1995), 273.

[CV05a] Marco Cuturi and Jean-Philippe Vert, *The context-tree kernel for strings*, Neural Networks **18** (2005), no. 8.

[CV05b] Marco Cuturi and Jean-Philippe Vert, *Semigroup kernels on finite sets*, Advances in Neural Information Processing Systems 17 (Lawrence K. Saul, Yair Weiss, and Léon Bottou, eds.), MIT Press, Cambridge, MA, 2005, pp. 329–336.

[CVBM] Marco Cuturi, Jean-Philippe Vert, Øystein Birkenes, and Tomoko Matsui, *A kernel for time series based on global alignments*, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, vol. II.

[CVBM02] Olivier Chapelle, Vladimir Vapnik, Olivier Bousquet, and Sayan Mukherjee, *Choosing multiple parameters for support vector machines*, Machine Learning **46** (2002), no. 1/3, 131.

[DBK⁺97] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, and V. Vapnik, *Support vector regression machines*, Advances in neural information processing systems (1997), 155–161.

[Dev55] A. Devinatz, *The representation of functions as Laplace-Stieltjes integrals*, Duke Mathematical Journal **24** (1955), 481–498.

[EGG03] W. Ehm, M.G. Genton, and T. Gneiting, *Stationary covariances associated with exponentially convex functions*, Bernoulli (2003), 607–615.

[EPE04] T. Evgeniou, M. Pontil, and A. Elisseeff, *Leave one out error, stability, and generalization of voting combinations of classifiers*, Machine Learning **55** (2004), no. 1, 71–97.

[FBG07] K. Fukumizu, F.R. Bach, and A. Gretton, *Statistical consistency of kernel canonical correlation analysis*, The Journal of Machine Learning Research **8** (2007), 361–383.

[FBJ04] Kenji Fukumizu, Francis Bach, and Michael Jordan, *Dimensionality reduction for supervised learning with reproducing kernel hilbert spaces*, Journal of Machine Learning Research **5** (2004), 73–99.

[FCS04] H. Frölich, O. Chapelle, and B. Schökopf, *Feature selection for support vector machines using genetic algorithms*, International Journal on Artificial Intelligence Tools **13** (2004), no. 4, 791–800.

[Fis36] Ronald Fisher, *The use of multiple measurements in taxonomic problems.*, Annals of Eugenics **7** (1936), 179–188.

[FS03] ois Fleuret Franc and Hichem Sahbi, *Scale-invariance of support vector machines based on the triangular kernel*, In 3rd International Workshop on Statistical and Computational Theories of Vision, 2003.

[GD05] Kristen Grauman and Trevor Darrell, *The pyramid match kernel: Discriminative classification with sets of image features*, ICCV, IEEE Computer Society, 2005, pp. 1458–1465.

[GJP95] Federico Girosi, Michael Jones, and Tomaso Poggio, *Regularization theory and neural networks architectures*, Neural Computation **7** (1995), no. 2, 219–269.

[Haa05]  Bernard Haasdonk, *Feature space interpretation of SVMs with indefinite kernels*, IEEE Trans. Pattern Anal. Mach. Intell **27** (2005), no. 4, 482–492.

[Hau99]  David Haussler, *Convolution kernels on discrete structures*, Tech. report, UC Santa Cruz, 1999, USCS-CRL-99-10.

[HB05]  M. Hein and O. Bousquet, *Hilbertian metrics and positive definite kernels on probability measures*, Proceedings of AISTATS 2005 (Z. Ghahramani and R. Cowell, eds.), January 2005.

[HB07]  Zaid Harchaoui and Francis Bach, *Image classification with segmentation graph kernels*, CVPR, 2007.

[HK02]  Bernard Haasdonk and Daniel Keysers, *Tangent distance kernels for support vector machines*, Proceedings of the International Conference on Pattern Recognition (2), 2002, pp. 864–868.

[Hoe62]  A.E. Hoerl, *Application of ridge analysis to regression problems*, Chemical Engineering Progress **58** (1962), no. 3, 54–59.

[Hof07]  H. Hoffmann, *Kernel PCA for novelty detection*, Pattern Recognition **40** (2007), no. 3, 863–874.

[HSS08]  T. Hofmann, B. Scholkopf, and A.J. Smola, *Kernel methods in machine learning*, Annals of Statistics **36** (2008), no. 3, 1171.

[HTF09]  T. Hastie, R. Tibshirani, and J. Friedman, *Elements of statistical learning: Data mining, inference, and prediction (2nd edition)*, Springer Verlag, 2009.

[Jag75]  P. Jagers, *Branching processes with biological applications*, John Wiley & Sons, 1975.

[JDH99]  T. Jaakkola, M. Diekhaus, and D. Haussler, *Using the fisher kernel method to detect remote protein homologies.*, 7th Intell. Sys. Mol. Biol. (1999), 149–158.

[JDH00]  Tommi Jaakkola, Mark Diekhans, and David Haussler, *A discriminative framework for detecting remote protein homologies*, Journal of Computational Biology **7** (2000), no. 1,2, 95–114.

[JH99]  T. S. Jaakkola and D. Haussler, *Exploiting Generative Models in Discriminative Classifiers*, Advances in Neural Information Processing Systems 11 (M. S. Kearns, S. A. Solla, and D. A. Cohn, eds.), MIT Press, 1999.

[JKH04]  Tony Jebara, Risi Kondor, and Andrew Howard, *Probability product kernels*, Journal of Machine Learning Research **5** (2004), 819–844.

[Joa02] Thorsten Joachims, *Learning to classify text using support vector machines: Methods, theory, and algorithms*, Kluwer Academic Publishers, 2002.

[Jol02] I. T. Jolliffe, *Principal component analysis*, Springer, New York, NY, USA, 2002.

[KL02] Risi Kondor and John Lafferty, *Diffusion kernels on graphs and other discrete input spaces.*, Proceedings of the Nineteenth International Conference on Machine Learning, 2002, pp. 315–322.

[KSB09] Risi Imre Kondor, Nino Shervashidze, and Karsten M. Borgwardt, *The graphlet spectrum*, ICML, 2009, p. 67.

[KSC07] S.S. Keerthi, V. Sindhwani, and O. Chapelle, *An efficient method for gradient-based adaptation of hyperparameters in SVM models*, Advances in Neural Information Processing Systems **19** (2007), 673.

[KTI03] H. Kashima, K. Tsuda, and A. Inokuchi, *Marginalized kernels between labeled graphs*, Proceedings of the Twentieth International Conference on Machine Learning (T. Faucett and N. Mishra, eds.), AAAI Press, 2003, pp. 321–328.

[KW71] G. S. Kimeldorf and G. Wahba, *Some results on tchebycheffian spline functions*, Journal of Mathematical Analysis and Applications **33** (1971), 82–95.

[LBC$^+$04] Gert R. G. Lanckriet, Tijl De Bie, Nello Cristianini, Michael I. Jordan, and William Stafford Noble, *A statistical framework for genomic data fusion*, Bioinformatics **20** (2004), no. 16, 2626–2635.

[LD08] Ronny Luss and Alexandre D'Aspremont, *Support vector machine classification with indefinite kernels*, Advances in Neural Information Processing Systems 20 (J.C. Platt, D. Koller, Y. Singer, and S. Roweis, eds.), MIT Press, Cambridge, MA, 2008, pp. 953–960.

[LEN02] Christina Leslie, Eleazar Eskin, and William Stafford Noble, *The spectrum kernel: a string kernel for svm protein classific ation*, Proc. of PSB 2002, 2002, pp. 564–575.

[LEWN03] Christina Leslie, Eleazar Eskin, Jason Weston, and William Stafford Noble, *Mismatch string kernels for svm protein classification*, NIPS 15 (Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, eds.), MIT Press, 2003.

[LL05] John Lafferty and Guy Lebanon, *Diffusion kernels on statistical manifolds*, JMLR **6** (2005), 129–163.

[Mer09] T. Mercer, *Functions of positive and negative type and their connection with the theory of integral equations*, Philosophical Transactions of the Royal Society of London, Series A **209** (1909), 415–446.

[MMR+01] K.R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, *An introduction to kernel-based learning algorithms*, IEEE transactions on neural networks **12** (2001), no. 2, 181–201.

[MP06] C.A. Micchelli and M. Pontil, *Learning the kernel function via regularization*, Journal of Machine Learning Research **6** (2006), no. 2, 1099.

[MRW+99] S. Mika, G. Ratsch, J. Weston, B. Schökopf, and K.R. Muller, *Fisher discriminant analysis with kernels*, Neural networks for signal processing IX (1999), 41–48.

[MUA+04] P. Mahé, N. Ueda, T. Akutsu, J.-L. Perret, and J.-P. Vert, *Extensions of marginalized graph kernels*, Proceedings of the Twenty-First International Conference on Machine Learning (ICML 2004) (R. Greiner and D. Schuurmans, eds.), ACM Press, 2004, pp. 552–559.

[MUA+05] P. Mahe, N. Ueda, Akutsu, J.-L. T., Perret, and J.-P. Vert, *Graph kernels for molecular structure-activity relationship analysis with support vector machines*, Journal of Chemical Information and Modeling **45** (2005), no. 4, 939–951.

[NS07] Assaf Naor and Gideon Schechtman, *Planar earthmover is not in $l_1$*, SIAM J. Comput. **37** (2007), no. 3, 804–826.

[Par62] Emanuel Parzen, *Extraction and detection problems and reproducing kernel Hilbert spaces*, Journal of the Society for Industrial and Applied Mathematics. Series A, On control **1** (1962), 35–62.

[RBCG07] Alain Rakotomamonjy, Francis Bach, Stéphane Canu, and Yves Grandvalet, *More efficiency in multiple kernel learning*, International Conference on Machine Learning (ICML), 2007, IDIAP-RR 07-18.

[SC08] I. Steinwart and A. Christmann, *Support Vector Machines*, Springer Verlag, 2008.

[SCDV98] Patrice Y. Simard, Yann A. Le Cun, John S. Denker, and Bernard Victorri, *Transformation invariance in pattern recognition – tangent distance and tangent propagation*, Lecture Notes in Computer Science, Springer, 1998, pp. 239–274.

[Sch42] I.J. Schoenberg, *Positive definite functions on spheres*, Duke Math. J. **9** (1942), 96–108.

[See02] Matthias Seeger, *Covariance kernels from bayesian generative models*, Advances in Neural Information Processing Systems 14 (T. G. Dietterich, S. Becker, and Z. Ghahramani, eds.), MIT Press, 2002, pp. 905–912.

[SG02] N. Smith and M. Gales, *Speech recognition using svms*, Advances in Neural Information Processing Systems 14 (Cambridge, MA) (T. G. Dietterich, S. Becker, and Z. Ghahramani, eds.), MIT Press, 2002.

[SNB05] Vikas Sindhwani, Partha Niyogi, and Mikhail Belkin, *Beyond the point cloud: from transductive to semi-supervised learning*, Proceedings of ICML '05: Twenty-second international conference on Machine learning, ACM Press, 2005.

[SNNS02a] H. Shimodaira, K.-I. Noma, M. Nakai, and S. Sagayama, *Dynamic time-alignment kernel in support vector machine*, Advances in Neural Information Processing Systems 14 (Cambridge, MA) (T. G. Dietterich, S. Becker, and Z. Ghahramani, eds.), MIT Press, 2002.

[SNNS02b] _____, *Dynamic time-alignment kernel in support vector machine*, Advances in Neural Information Processing Systems 14 (Cambridge, MA) (T. G. Dietterich, S. Becker, and Z. Ghahramani, eds.), MIT Press, 2002.

[SPSt+99] Bernhard Schökopf, John C. Platt, John Shawe-taylor, Alex J. Smola, and Robert C. Williamson, *Estimating the support of a high-dimensional distribution*, Neural Computation **13** (1999), 2001.

[SRSS06] Sören Sonnenburg, Gunnar Rätsch, Christin Schäfer, and Bernhard Schölkopf, *Large scale multiple kernel learning*, Journal of Machine Learning Research **7** (2006), 1531–1565.

[SS02] Bernhard Schölkopf and Alexander J. Smola, *Learning with kernels: Support vector machines, regularization , optimization, and beyond*, MIT Press, 2002.

[SSM98] B. Schölkopf, A. Smola, and K. Müller, *Nonlinear component analysis as a kernel eigenvalue problem*, Neural Computation **10** (1998), no. 5, 1299–1319.

[STC04] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*, Cambridge Univ Pr, 2004.

[STV04] Bernhard Schölkopf, Koji Tsuda, and Jean-Philippe Vert, *Kernel methods in computational biology*, MIT Press, 2004.

[SWE+02] Bernhard Schölkopf, Jason Weston, Eleazar Eskin, Christina Leslie, and William Stafford Noble, *A kernel approach for learning from almost orthogonal patterns*, Proceedings of ECML 2002, 13th European Conference on Machine Learning, Helsinki, Finland, August 19-23, 2002 (Tapio Elomaa, Heikki Mannila, and Hannu Toivonen, eds.), Lecture Notes in Computer Science, vol. 2430, Springer, 2002, pp. 511–528.

[TAKM04] Koji Tsuda, Shotaro Akaho, Motoaki Kawanabe, and Klaus-Robert Müller, *Asymptotic properties of the fisher kernel*, Neural Computation **16** (2004), no. 1, 115–137.

[Tib96] R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society. Series B (Methodological) (1996), 267–288.

[TKA02] K. Tsuda, T. Kin, and K. Asai, *Marginalized kernels for biological sequences*, Bioinformatics **18** (2002), no. Suppl 1, 268–275.

[TKR$^+$02] Koji Tsuda, Motoaki Kawanabe, Gunnar Rätsch, Sören Sonnenburg, and Klaus-Robert Müller, *A new discriminative kernel from probabilistic models*, Neural Computation **14** (2002), no. 10, 2397–2414.

[TV06] Choon Hui Teo and S. V. N. Vishwanathan, *Fast and space efficient string kernels using suffix arrays*, In Proceedings, 23rd ICMP, ACM Press, 2006, pp. 929–936.

[Vap98] Vladimir N. Vapnik, *Statistical learning theory*, Wiley, 1998.

[VBKS08] SVN Vishwanathan, K.M. Borgwardt, I.R. Kondor, and N.N. Schraudolph, *Graph kernels*, Journal of Machine Learning Research **9** (2008), 1–37.

[Ver02] Jean-Philippe Vert, *A tree kernel to analyze phylogenetic profiles*, Bioinformatics **18** (2002), S276–S284.

[Ver06] J.-P. Vert, *Classification of biological sequences with kernel methods*, Lecture Notes in Computer Science **4201** (2006), 7.

[VK03] Jean-Philippe Vert and Minoru Kanehisa, *Graph-driven features extraction from microarray data using diffusion kernels and kernel cca*, Advances in Neural Information Processing Systems 15 (Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, eds.), MIT Press, 2003.

[VSA04] Jean-Philippe Vert, Hiroto Saigo, and Tatsuya Akutsu, *Local alignment kernels for protein sequences*, Kernel Methods in Computational Biology (Bernhard Schölkopf, Koji Tsuda, and Jean-Philippe Vert, eds.), MIT Press, 2004.

[VV05] Régis Vert and Jean-Philippe Vert, *Consistency and convergence rates of one-class svm and related algorithms*, Tech. Report 1414, LRI, Université Paris Sud, 2005.

[VY05] Jean-Philippe Vert and Yoshihiro Yamanishi, *Supervised graph inference*, Advances in Neural Information Processing Systems 17 (Lawrence K. Saul, Yair Weiss, and Léon Bottou, eds.), MIT Press, 2005.

[Wah90] G. Wahba, *Splines models for observational data*, Series in Applied Mathematics, Vol. 59, SIAM, 1990.

[Wat00] C. Watkins, *Dynamic alignment kernels*, Advances in Large Margin Classifiers (A.J. Smola, P.L. Bartlett, B. Schölkopf, and D. Schuu rmans, eds.), MIT Press, 2000, pp. 39–50.

[WST95] F. M. J. Willems, Y. M. Shtarkov, and Tj. J. Tjalkens, *The context-tree weighting method: basic properties*, IEEE Transactions on Information Theory (1995), 653–664.

[ZCL05] Dell Zhang, Xi Chen, and Wee Sun Lee, *Text classification with kernels on the multinomial manifold*, SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, ACM Press, 2005, pp. 266–273.

[ZH02] J. Zhu and T. Hastie, *Kernel logistic regression and the import vector machine*, Advances in Neural Information Processing Systems 14 (Cambridge, MA) (T. G. Dietterich, S. Becker, and Z. Ghahramani, eds.), MIT Press, 2002, pp. 1081–1088.