

Bibliography

- [1] J Aach and GM Church. Aligning gene expression time series with time warping algorithms. *Bioinformatics*, 17(6):495–508, 2001.
- [2] YS Abu-Mostafa, M Magdon-Ismail, and HT Lin. *Learning from data*. AMLBook, 2012.
- [3] G Al-Naymat, S Chawla, and J Taheri. Sparsedtw: a novel approach to speed up dynamic time warping. In *Proceedings of the Eighth Australasian Data Mining Conference-Volume 101*, pages 117–127. Australian Computer Society, Inc., 2009.
- [4] DJ Anderson and R Adolphs. A framework for studying emotions across species. *Cell*, 157(1):187–200, 2014.
- [5] DJ Anderson and P Perona. Toward a science of computational ethology. *Neuron*, 84(1):18–31, 2014.
- [6] K Asahina, K Watanabe, BJ Duistermars, E Hoopfer, CR González, EA Eyjólfssdóttir, P Perona, and DJ Anderson. Tachykinin-expressing neurons control male-specific aggressive arousal in drosophila. *Cell*, 156(1):221–235, 2014.
- [7] K Bache and M Lichman. UCI machine learning repository, 2013.
- [8] Z Bar-Joseph, G Gerber, DK Gifford, TS Jaakkola, and I Simon. A new approach to analyzing gene expression time series data. In *Proceedings of the Sixth Annual International Conference on Computational Biology*, pages 39–48. ACM, 2002.
- [9] C Bargmann, W Newsome, A Anderson, E Brown, K Deisseroth, J Donoghue, P MacLeish, E Marder, R Normann, J Sanes, et al. Brain 2025: a scientific vision. *Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Working Group Report to the Advisory Committee to the Director, NIH [http://www.nih.gov/science/brain/2025/\(US National Institutes of Health, 2014\)](http://www.nih.gov/science/brain/2025/(US%20National%20Institutes%20of%20Health,%202014))*, 2014.
- [10] S Ben-David, J Blitzer, K Crammer, A Kulesza, F Pereira, and JW Vaughan. A theory of learning from different domains. *Machine learning*, 79(1-2):151–175, 2010.

- [11] S Ben-David, J Blitzer, K Crammer, and F Pereira. Analysis of representations for domain adaptation. *Advances in Neural Information Processing Systems*, 19:137, 2007.
- [12] DJ Berndt and J Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.
- [13] S Bickel, M Brückner, and T Scheffer. Discriminative learning for differing training and test distributions. In *Proceedings of the 24th International Conference on Machine Learning*, pages 81–88. ACM, 2007.
- [14] S Bickel, M Brückner, and T Scheffer. Discriminative learning under covariate shift. *The Journal of Machine Learning Research*, 10:2137–2155, 2009.
- [15] J Blitzer, M Dredze, and F Pereira. Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Annual Meeting-Association For Computational Linguistics*, volume 45, page 440, 2007.
- [16] J Blitzer, R McDonald, and F Pereira. Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 120–128. Association for Computational Linguistics, 2006.
- [17] A Blum. Learning from labeled and unlabeled data using graph mincuts. In *Proceedings of the 18th International Conference on Machine Learning*, 2001.
- [18] A Blum and T Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, pages 92–100. ACM, 1998.
- [19] SP Boyd and L Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [20] K Branson, AA Robie, J Bender, P Perona, and MH Dickinson. High-throughput ethomics in large groups of drosophila. *Nature Methods*, 6(6):451–457, 2009.
- [21] XP Burgos-Artizzu, P Dollár, D Lin, DJ Anderson, and P Perona. Social behavior recognition in continuous video. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1322–1329. IEEE, 2012.
- [22] EG Caiani, A Porta, G Baselli, M Turiel, S Muzzupappa, F Pieruzzi, C Crema, A Malliani, and S Cerutti. Warped-average template technique to track on a cycle-by-cycle basis the cardiac filling phases on left ventricular volume. In *Computers in Cardiology 1998*, pages 73–76. IEEE, 1998.

- [23] CC Chang and CJ Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.
- [24] T Chen, Z Zheng, Q Lu, W Zhang, and Y Yu. Feature-based matrix factorization. *arXiv preprint arXiv:1109.2271*, 2011.
- [25] C Cortes, Y Mansour, and M Mohri. Learning bounds for importance weighting. *Advances in Neural Information Processing Systems*, 23:442–450, 2010.
- [26] C Cortes and M Mohri. Domain adaptation and sample bias correction theory and algorithm for regression. *Theoretical Computer Science*, 519:103–126, 2014.
- [27] C Cortes, M Mohri, M Riley, and A Rostamizadeh. Sample selection bias correction theory. In *Algorithmic Learning Theory*, pages 38–53. Springer, 2008.
- [28] R Courant and D Hilbert. *Methods of mathematical physics*, volume 1. CUP Archive, 1966.
- [29] I Csisz et al. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.
- [30] B Dacorogna. *Introduction to the Calculus of Variations*, volume 4. World Scientific, 2004.
- [31] H Dankert, L Wang, ED Hoopfer, DJ Anderson, and P Perona. Automated monitoring and analysis of social behavior in drosophila. *Nature Methods*, 6(4):297–303, 2009.
- [32] V Digalakis, D Rtischev, and L Neumeyer. Speaker adaptation using constrained estimation of Gaussian mixtures. *Speech and Audio Processing, IEEE Transactions on*, 3(5):357–366, 1995.
- [33] E Eyjolfsdottir, S Branson, XP Burgos-Artizzu, ED Hoopfer, J Schor, DJ Anderson, and P Perona. Detecting social actions of fruit flies. In *Computer Vision–ECCV 2014*, pages 772–787. Springer, 2014.
- [34] J Gao, W Fan, J Jiang, and J Han. Knowledge transfer via multiple model local structure mapping. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 283–291. ACM, 2008.
- [35] WT Gibson, CR. Gonzalez, CM Fernandez, L Ramasamy, T Tabachnik, RR Du, PD Felsen, MM Maire, P Perona, and DJ Anderson. Behavioral response to a repetitive visual threat stimulus express a persistent state of defensive arousal in *Drosophila*. *Current Biology*, May 2015.
- [36] CR González and YS Abu-Mostafa. Mismatched training and test distributions can outperform matched ones. *Neural Computation*, 27(2):365–387, February 2015.

- [37] T Groves and T Rothenberg. A note on the expected value of an inverse matrix. *Biometrika*, 56(3):690–691, 1969.
- [38] J Huang, AJ Smola, A Gretton, K Borgwardt, and B Scholkopf. Correcting sample selection bias by unlabeled data. *Advances in Neural Information Processing Systems*, 19:601, 2007.
- [39] H Jhuang, E Garrote, X Yu, V Khilnani, T Poggio, AD Steele, and T Serre. Automated home-cage behavioural phenotyping of mice. *Nature Communications*, 1:68, 2010.
- [40] J Jiang and C Zhai. Instance weighting for domain adaptation in nlp. In *Annual Meeting-Association For Computational Linguistics*, volume 45, page 264, 2007.
- [41] M Kabra, AA Robie, M Rivera-Alba, S Branson, and K Branson. Jaaba: interactive machine learning for automatic annotation of animal behavior. *Nature Methods*, 10(1):64–67, 2013.
- [42] T Kanamori, S Hido, and M Sugiyama. A least-squares approach to direct importance estimation. *The Journal of Machine Learning Research*, 10:1391–1445, 2009.
- [43] T Kanamori and H Shimodaira. Active learning algorithm using the maximum weighted log-likelihood estimator. *Journal of Statistical Planning and Inference*, 116(1):149–162, 2003.
- [44] Y Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 426–434. ACM, 2008.
- [45] JB Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.
- [46] C Leggetter and P Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech and Language*, 9(2):171, 1995.
- [47] Y Mansour, M Mohri, and A Rostamizadeh. Domain adaptation: Learning bounds and algorithms. *arXiv preprint arXiv:0902.3430*, 2009.
- [48] A Margolis. A literature review of domain adaptation with unlabeled data. *University of Washington*, http://ssli.ee.washington.edu/~amargoli/review_Mar23.pdf, 2011.
- [49] L Mason, J Baxter, PL Bartlett, and MR Frean. Boosting algorithms as gradient descent. In S.A. Solla, T.K. Leen, and K. Müller, editors, *Advances in Neural Information Processing Systems 12*, pages 512–518. MIT Press, 2000.
- [50] K Murphy and M Dunham. Probabilistic model toolkit for MATLAB and Octave (pmtk3).

- [51] A Ng, M Jordan, Y Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 2:849–856, 2002.
- [52] S Pan, J Kwok, and Q Yang. Transfer learning via dimensionality reduction. In *Proceedings of the 23rd National Conference on Artificial Intelligence*, volume 2, pages 677–682, 2008.
- [53] E Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, pages 1065–1076, 1962.
- [54] J Quiñonero-Candela, M Sugiyama, A Schwaighofer, and ND Lawrence. *Dataset shift in machine learning*. Cambridge, MA: MIT Press, 2009.
- [55] J Ren, X Shi, W Fan, and P Yu. Type-independent correction of sample selection bias via structural discovery and re-balancing. In *Proceedings of the Eighth SIAM International Conference on Data Mining, SDM*, pages 565–576, 2008.
- [56] M Rosenblatt et al. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.
- [57] S Rosset, J Zhu, H Zou, and T Hastie. A method for inferring label sampling mechanisms in semi-supervised learning. *Ann Arbor*, 1001:48109, 2005.
- [58] S Sabato and R Munos. Active regression by stratification. In Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 469–477. Curran Associates, Inc., 2014.
- [59] S Salvador and P Chan. Fastdtw: Toward accurate dynamic time warping in linear time and space. *KDD Workshop on Mining Temporal and Sequential Data*, pages 70–80, 2004.
- [60] J Sherman and WJ Morrison. Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix. *The Annals of Mathematical Statistics*, 20(4):620–624, 1949.
- [61] H Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227–244, 2000.
- [62] Y Sogawa, T Ueno, Y Kawahara, and T Washio. Active learning for noisy oracle via density power divergence. *Neural Networks*, 46:133–143, 2013.
- [63] M Sugiyama. Active learning in approximately linear regression based on conditional expectation of generalization error. *The Journal of Machine Learning Research*, 7:141–166, 2006.

- [64] M Sugiyama and M Kawanabe. *Machine learning in non-stationary environments: Introduction to covariate shift adaptation*. MIT Press, 2012.
- [65] M Sugiyama, M Krauledat, and K Müller. Covariate shift adaptation by importance weighted cross validation. *The Journal of Machine Learning Research*, 8:985–1005, 2007.
- [66] M Sugiyama and S Nakajima. Pool-based active learning in approximate linear regression. *Machine Learning*, 75(3):249–274, 2009.
- [67] M Sugiyama, S Nakajima, H Kashima, P Von Buenau, and M Kawanabe. Direct importance estimation with model selection and its application to covariate shift adaptation. *Advances in Neural Information Processing Systems*, 20:1433–1440, 2008.
- [68] L Torgo. LIACC regression datasets.
- [69] JT Vogelstein, Y Park, T Ohyama, Rex A Kerr, JW Truman, CE Priebe, and M Zlatic. Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science*, 344(6182):386–392, 2014.
- [70] DP Wiens. Robust weights and designs for biased regression models: Least squares and generalized m-estimation. *Journal of Statistical Planning and Inference*, 83(2):395–412, 2000.
- [71] B Zadrozny. Learning and evaluating classifiers under sample selection bias. In *Proceedings of the 21st International Conference on Machine Learning*, page 114. ACM, 2004.
- [72] C Zhang, L Zhang, and J Ye. Generalization bounds for domain adaptation. *Advances in Neural Information Processing Systems*, 2012.