

BIBLIOGRAPHY

- [M2012] “Machine Learning: A Probabilistic Perspective” Murphy, K. P. - chapter 14.4.3, pp. 492-493, The MIT Press, 2012
- [RW2006] Carl Eduard Rasmussen and Christopher K.I. Williams, “Gaussian Processes for Machine Learning”, MIT Press 2006, Link to an official complete PDF version of the book [here](#) .
- [B1999] L. Breiman, “Pasting small votes for classification in large databases and on-line”, Machine Learning, 36(1), 85-103, 1999.
- [B1996] L. Breiman, “Bagging predictors”, Machine Learning, 24(2), 123-140, 1996.
- [H1998] T. Ho, “The random subspace method for constructing decision forests”, Pattern Analysis and Machine Intelligence, 20(8), 832-844, 1998.
- [LG2012] G. Louppe and P. Geurts, “Ensembles on Random Patches”, Machine Learning and Knowledge Discovery in Databases, 346-361, 2012.
- [B2001] 12. Breiman, “Random Forests”, Machine Learning, 45(1), 5-32, 2001.
- [B1998] 12. Breiman, “Arcing Classifiers”, Annals of Statistics 1998.
- [L2014] G. Louppe, “Understanding Random Forests: From Theory to Practice”, PhD Thesis, U. of Liege, 2014.
- [FS1995] Y. Freund, and R. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”, 1997.
- [ZZRH2009] J. Zhu, H. Zou, S. Rosset, T. Hastie. “Multi-class AdaBoost”, 2009.
- [D1997] 8. Drucker. “Improving Regressors using Boosting Techniques”, 1997.
- [HTF] T. Hastie, R. Tibshirani and J. Friedman, “Elements of Statistical Learning Ed. 2”, Springer, 2009.
- [VEB2009] Vinh, Epps, and Bailey, (2009). “Information theoretic measures for clusterings comparison”. Proceedings of the 26th Annual International Conference on Machine Learning - ICML ‘09. doi:10.1145/1553374.1553511. ISBN 9781605585161.
- [VEB2010] Vinh, Epps, and Bailey, (2010). “Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance”. JMLR <<http://jmlr.csail.mit.edu/papers/volume11/vinh10a/vinh10a.pdf>>
- [YAT2016] Yang, Algesheimer, and Tessone, (2016). “A comparative analysis of community detection algorithms on artificial networks”. Scientific Reports 6: 30750. doi:10.1038/srep30750.
- [B2011] Identification and Characterization of Events in Social Media, Hila Becker, PhD Thesis.
- [Mrl09] “Online Dictionary Learning for Sparse Coding” J. Mairal, F. Bach, J. Ponce, G. Sapiro, 2009
- [Jen09] “Structured Sparse Principal Component Analysis” R. Jenatton, G. Obozinski, F. Bach, 2009

- [R45f14345c000-1] 12. Breiman, “Random Forests”, Machine Learning, 45(1), 5-32, 2001.
- [Rf91cab2dc427-1] 12. Breiman, “Random Forests”, Machine Learning, 45(1), 5-32, 2001.
- [Rf91cab2dc427-2] P. Geurts, D. Ernst., and L. Wehenkel, “Extremely randomized trees”, Machine Learning, 63(1), 3-42, 2006.
- [Rc8f28bfad63f-1] P. Geurts, D. Ernst., and L. Wehenkel, “Extremely randomized trees”, Machine Learning, 63(1), 3-42, 2006.
- [Ra7d0c8995fbc-1] P. Geurts, D. Ernst., and L. Wehenkel, “Extremely randomized trees”, Machine Learning, 63(1), 3-42, 2006.
- [Guyon2015] I. Guyon, K. Bennett, G. Cawley, H.J. Escalante, S. Escalera, T.K. Ho, N. Macià, B. Ray, M. Saeed, A.R. Statnikov, E. Viegas, *Design of the 2015 ChaLearn AutoML Challenge*, IJCNN 2015.
- [Mosley2013] L. Mosley, *A balanced approach to the multi-class imbalance problem*, IJCV 2010.
- [Kelleher2015] John. D. Kelleher, Brian Mac Namee, Aoife D’Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, 2015.
- [Urbanowicz2015] Urbanowicz R.J., Moore, J.H. *ExSTraCS 2.0: description and evaluation of a scalable learning classifier system*, *Evol. Intel.* (2015) 8: 89.
- [Manning2008] C.D. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, 2008.
- [Everingham2010] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, *The Pascal Visual Object Classes (VOC) Challenge*, IJCV 2010.
- [Davis2006] J. Davis, M. Goadrich, *The Relationship Between Precision-Recall and ROC Curves*, ICML 2006.
- [Flach2015] P.A. Flach, M. Kull, *Precision-Recall-Gain Curves: PR Analysis Done Right*, NIPS 2015.
- [HTF2009] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Second Edition, Section 10.13.2, Springer, 2009.
- [Mol2019] C. Molnar, *Interpretable Machine Learning*, Section 5.1, 2019.
- [NQY18] J. Nothman, H. Qin and R. Yurchak (2018). “Stop Word Lists in Free Open-source Software Packages”. In *Proc. Workshop for NLP Open Source Software*.
- [RR2007] “Random features for large-scale kernel machines” Rahimi, A. and Recht, B. - *Advances in neural information processing* 2007,
- [LS2010] “Random Fourier approximations for skewed multiplicative histogram kernels” Random Fourier approximations for skewed multiplicative histogram kernels - *Lecture Notes for Computer Science (DAGM)*
- [VZ2010] “Efficient additive kernels via explicit feature maps” Vedaldi, A. and Zisserman, A. - *Computer Vision and Pattern Recognition* 2010
- [VVZ2010] “Generalized RBF feature maps for Efficient Detection” Vempati, S. and Vedaldi, A. and Zisserman, A. and Jawahar, CV - 2010
- [R57cf438d7060-1] Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers, B. Zadrozny & C. Elkan, ICML 2001
- [R57cf438d7060-2] Transforming Classifier Scores into Accurate Multiclass Probability Estimates, B. Zadrozny & C. Elkan, (KDD 2002)
- [R57cf438d7060-3] Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods, J. Platt, (1999)
- [R57cf438d7060-4] Predicting Good Probabilities with Supervised Learning, A. Niculescu-Mizil & R. Caruana, ICML 2005

- [R2c55e37003fe-1] Ankerst, Mihael, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. "OPTICS: ordering points to identify the clustering structure." *ACM SIGMOD Record* 28, no. 2 (1999): 49-60.
- [R2c55e37003fe-2] Schubert, Erich, Michael Gertz. "Improving the Cluster Structure Extracted from OPTICS Plots." *Proc. of the Conference "Lernen, Wissen, Daten, Analysen" (LWDA)* (2018): 318-329.
- [1] Ankerst, Mihael, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. "OPTICS: ordering points to identify the clustering structure." *ACM SIGMOD Record* 28, no. 2 (1999): 49-60.
- [R68ae096da0e4-1] Rousseeuw, P.J., Van Driessen, K. "A fast algorithm for the minimum covariance determinant estimator" *Technometrics* 41(3), 212 (1999)
- [RVD] A Fast Algorithm for the Minimum Covariance Determinant Estimator, 1999, American Statistical Association and the American Society for Quality, *TECHNOMETRICS*
- [RVDriessen] A Fast Algorithm for the Minimum Covariance Determinant Estimator, 1999, American Statistical Association and the American Society for Quality, *TECHNOMETRICS*
- [R9f63e655f7bd-Rousseeuw1984] P. J. Rousseeuw. Least median of squares regression. *J. Am Stat Ass*, 79:871, 1984.
- [R9f63e655f7bd-Rousseeuw] A Fast Algorithm for the Minimum Covariance Determinant Estimator, 1999, American Statistical Association and the American Society for Quality, *TECHNOMETRICS*
- [R9f63e655f7bd-ButlerDavies] R. W. Butler, P. L. Davies and M. Jhun, Asymptotics For The Minimum Covariance Determinant Estimator, *The Annals of Statistics*, 1993, Vol. 21, No. 3, 1385-1400
- [RVD] A Fast Algorithm for the Minimum Covariance Determinant Estimator, 1999, American Statistical Association and the American Society for Quality, *TECHNOMETRICS*
- [RVDriessen] A Fast Algorithm for the Minimum Covariance Determinant Estimator, 1999, American Statistical Association and the American Society for Quality, *TECHNOMETRICS*
- [1] Dhillon, I. S. (2001, August). Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 269-274). ACM.
- [1] Kluger, Y., Basri, R., Chang, J. T., & Gerstein, M. (2003). Spectral biclustering of microarray data: coclustering genes and conditions. *Genome research*, 13(4), 703-716.
- [1] I. Guyon, "Design of experiments for the NIPS 2003 variable selection benchmark", 2003.
- [1] J. Friedman, "Multivariate adaptive regression splines", *The Annals of Statistics* 19 (1), pages 1-67, 1991.
- [2] L. Breiman, "Bagging predictors", *Machine Learning* 24, pages 123-140, 1996.
- [1] J. Friedman, "Multivariate adaptive regression splines", *The Annals of Statistics* 19 (1), pages 1-67, 1991.
- [2] L. Breiman, "Bagging predictors", *Machine Learning* 24, pages 123-140, 1996.
- [1] J. Friedman, "Multivariate adaptive regression splines", *The Annals of Statistics* 19 (1), pages 1-67, 1991.
- [2] L. Breiman, "Bagging predictors", *Machine Learning* 24, pages 123-140, 1996.
- [1] 10. Zhu, H. Zou, S. Rosset, T. Hastie, "Multi-class AdaBoost", 2009.
- [1] T. Hastie, R. Tibshirani and J. Friedman, "Elements of Statistical Learning Ed. 2", Springer, 2009.
- [1] G. Celeux, M. El Anbari, J.-M. Marin, C. P. Robert, "Regularization in regression: comparing Bayesian and frequentist methods in a poorly informative situation", 2009.
- [1] S. Marsland, "Machine Learning: An Algorithmic Perspective", Chapter 10, 2009. <http://seat.massey.ac.nz/personal/s.r.marsland/Code/10/ile.py>
- [R33e4ec8c4ad5-1] Y. Freund, R. Schapire, "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting", 1995.

- [R33e4ec8c4ad5-2] 10. Zhu, H. Zou, S. Rosset, T. Hastie, “Multi-class AdaBoost”, 2009.
- [R0c261b7dee9d-1] Y. Freund, R. Schapire, “A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting”, 1995.
- [R0c261b7dee9d-2] 8. Drucker, “Improving Regressors using Boosting Techniques”, 1997.
- [Rb1846455d0e5-1] L. Breiman, “Pasting small votes for classification in large databases and on-line”, *Machine Learning*, 36(1), 85-103, 1999.
- [Rb1846455d0e5-2] L. Breiman, “Bagging predictors”, *Machine Learning*, 24(2), 123-140, 1996.
- [Rb1846455d0e5-3] T. Ho, “The random subspace method for constructing decision forests”, *Pattern Analysis and Machine Intelligence*, 20(8), 832-844, 1998.
- [Rb1846455d0e5-4] G. Louppe and P. Geurts, “Ensembles on Random Patches”, *Machine Learning and Knowledge Discovery in Databases*, 346-361, 2012.
- [R4d113ba76fc0-1] L. Breiman, “Pasting small votes for classification in large databases and on-line”, *Machine Learning*, 36(1), 85-103, 1999.
- [R4d113ba76fc0-2] L. Breiman, “Bagging predictors”, *Machine Learning*, 24(2), 123-140, 1996.
- [R4d113ba76fc0-3] T. Ho, “The random subspace method for constructing decision forests”, *Pattern Analysis and Machine Intelligence*, 20(8), 832-844, 1998.
- [R4d113ba76fc0-4] G. Louppe and P. Geurts, “Ensembles on Random Patches”, *Machine Learning and Knowledge Discovery in Databases*, 346-361, 2012.
- [Rd7ae0a2ae688-1] Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua. “Isolation forest.” *Data Mining*, 2008. ICDM’08. Eighth IEEE International Conference on.
- [Rd7ae0a2ae688-2] Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua. “Isolation-based anomaly detection.” *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6.1 (2012): 3.
- [R6e47e53bacbd-1] P. Geurts, D. Ernst., and L. Wehenkel, “Extremely randomized trees”, *Machine Learning*, 63(1), 3-42, 2006.
- [R6e47e53bacbd-2] Moosmann, F. and Triggs, B. and Jurie, F. “Fast discriminative visual codebooks using randomized clustering forests” *NIPS* 2007
- [R1b90ac3ca370-Yates2011] R. Baeza-Yates and B. Ribeiro-Neto (2011). *Modern Information Retrieval*. Addison Wesley, pp. 68-74.
- [R1b90ac3ca370-MRS2008] C.D. Manning, P. Raghavan and H. Schütze (2008). *Introduction to Information Retrieval*. Cambridge University Press, pp. 118-120.
- [Re310f679c81e-1] Guyon, I., Weston, J., Barnhill, S., & Vapnik, V., “Gene selection for cancer classification using support vector machines”, *Mach. Learn.*, 46(1-3), 389–422, 2002.
- [R6f4d61ceb411-1] Guyon, I., Weston, J., Barnhill, S., & Vapnik, V., “Gene selection for cancer classification using support vector machines”, *Mach. Learn.*, 46(1-3), 389–422, 2002.
- [1] [Mutual Information](#) on Wikipedia.
- [2] A. Kraskov, H. Stogbauer and P. Grassberger, “Estimating mutual information”. *Phys. Rev. E* 69, 2004.
- [3] B. C. Ross “Mutual Information between Discrete and Continuous Data Sets”. *PLoS ONE* 9(2), 2014.
- [4] L. F. Kozachenko, N. N. Leonenko, “Sample Estimate of the Entropy of a Random Vector:”, *Probl. Peredachi Inf.*, 23:2 (1987), 9-16
- [1] [Mutual Information](#) on Wikipedia.
- [2] A. Kraskov, H. Stogbauer and P. Grassberger, “Estimating mutual information”. *Phys. Rev. E* 69, 2004.

- [3] B. C. Ross “Mutual Information between Discrete and Continuous Data Sets”. PLoS ONE 9(2), 2014.
- [4] L. F. Kozachenko, N. N. Leonenko, “Sample Estimate of the Entropy of a Random Vector”, Probl. Peredachi Inf., 23:2 (1987), 9-16
- [Rcd31b817a31e-1] Stef van Buuren, Karin Groothuis-Oudshoorn (2011). “mice: Multivariate Imputation by Chained Equations in R”. Journal of Statistical Software 45: 1-67.
- [Rcd31b817a31e-2] S. F. Buck, (1960). “A Method of Estimation of Missing Values in Multivariate Data Suitable for use with an Electronic Computer”. Journal of the Royal Statistical Society 22(2): 302-306.
- [Re4616ef910fb-1] Peter J. Huber, Elvezio M. Ronchetti, Robust Statistics Concomitant scale estimates, pg 172
- [Re4616ef910fb-2] Art B. Owen (2006), A robust hybrid of lasso and ridge regression. <https://statweb.stanford.edu/~owen/reports/hhu.pdf>
- [R80ce5b25cf9d-1] <https://en.wikipedia.org/wiki/RANSAC>
- [R80ce5b25cf9d-2] <https://www.sri.com/sites/default/files/publications/ransac-publication.pdf>
- [R80ce5b25cf9d-3] <http://www.bmva.org/bmvc/2009/Papers/Paper355/Paper355.pdf>
- [1] “Least Angle Regression”, Efron et al. <http://statweb.stanford.edu/~tibs/ftp/lars.pdf>
- [2] Wikipedia entry on the Least-angle regression
- [3] Wikipedia entry on the Lasso
- [1] “Least Angle Regression”, Efron et al. <http://statweb.stanford.edu/~tibs/ftp/lars.pdf>
- [2] Wikipedia entry on the Least-angle regression
- [3] Wikipedia entry on the Lasso
- [R7f4d308f5054-1] Tenenbaum, J.B.; De Silva, V.; & Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. Science 290 (5500)
- [R62e36dd1b056-1] Roweis, S. & Saul, L. Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323 (2000).
- [R62e36dd1b056-2] Donoho, D. & Grimes, C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. Proc Natl Acad Sci U S A. 100:5591 (2003).
- [R62e36dd1b056-3] Zhang, Z. & Wang, J. MLLE: Modified Locally Linear Embedding Using Multiple Weights. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.70.382>
- [R62e36dd1b056-4] Zhang, Z. & Zha, H. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. Journal of Shanghai Univ. 8:406 (2004)
- [1] Roweis, S. & Saul, L. Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323 (2000).
- [2] Donoho, D. & Grimes, C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. Proc Natl Acad Sci U S A. 100:5591 (2003).
- [3] Zhang, Z. & Wang, J. MLLE: Modified Locally Linear Embedding Using Multiple Weights. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.70.382>
- [4] Zhang, Z. & Zha, H. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. Journal of Shanghai Univ. 8:406 (2004)
- [1] Wikipedia entry for the Average precision
- [1] Brodersen, K.H.; Ong, C.S.; Stephan, K.E.; Buhmann, J.M. (2010). The balanced accuracy and its posterior distribution. Proceedings of the 20th International Conference on Pattern Recognition, 3121-24.
- [2] John. D. Kelleher, Brian Mac Namee, Aoife D’Arcy, (2015). Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies.

- [1] [Wikipedia entry for the Brier score.](#)
- [1] J. Cohen (1960). “A coefficient of agreement for nominal scales”. *Educational and Psychological Measurement* 20(1):37-46. doi:10.1177/001316446002000104.
- [2] R. Artstein and M. Poesio (2008). “Inter-coder agreement for computational linguistics”. *Computational Linguistics* 34(4):555-596.
- [3] [Wikipedia entry for the Cohen’s kappa.](#)
- [1] [Wikipedia entry for the Confusion matrix](#) (Wikipedia and other references may use a different convention for axes)
- [1] [Wikipedia entry for the F1-score](#)
- [1] R. Baeza-Yates and B. Ribeiro-Neto (2011). *Modern Information Retrieval*. Addison Wesley, pp. 327-328.
- [2] [Wikipedia entry for the F1-score](#)
- [1] Grigorios Tsoumakas, Ioannis Katakis. Multi-Label Classification: An Overview. *International Journal of Data Warehousing & Mining*, 3(3), 1-13, July-September 2007.
- [2] [Wikipedia entry on the Hamming distance](#)
- [1] [Wikipedia entry on the Hinge loss](#)
- [2] Koby Crammer, Yoram Singer. On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines. *Journal of Machine Learning Research* 2, (2001), 265-292
- [3] [L1 AND L2 Regularization for Multiclass Hinge Loss Models](#) by Robert C. Moore, John DeNero.
- [1] [Wikipedia entry for the Jaccard index](#)
- [1] Baldi, Brunak, Chauvin, Andersen and Nielsen, (2000). Assessing the accuracy of prediction algorithms for classification: an overview
- [2] [Wikipedia entry for the Matthews Correlation Coefficient](#)
- [3] Gorodkin, (2004). Comparing two K-category assignments by a K-category correlation coefficient
- [4] Jurman, Riccadonna, Furlanello, (2012). A Comparison of MCC and CEN Error Measures in MultiClass Prediction
- [1] [Wikipedia entry for the Precision and recall](#)
- [2] [Wikipedia entry for the F1-score](#)
- [3] [Discriminative Methods for Multi-labeled Classification](#) *Advances in Knowledge Discovery and Data Mining* (2004), pp. 22-30 by Shantanu Godbole, Sunita Sarawagi
- [1] [Wikipedia entry for the Receiver operating characteristic](#)
- [2] Fawcett T. An introduction to ROC analysis[J]. *Pattern Recognition Letters*, 2006, 27(8):861-874.
- [3] [Analyzing a portion of the ROC curve.](#) McClish, 1989
- [1] [Wikipedia entry for the Receiver operating characteristic](#)
- [2] Fawcett T. An introduction to ROC analysis[J]. *Pattern Recognition Letters*, 2006, 27(8):861-874.
- [1] [Wikipedia entry on the Coefficient of determination](#)
- [1] Tsoumakas, G., Katakis, I., & Vlahavas, I. (2010). Mining multi-label data. In *Data mining and knowledge discovery handbook* (pp. 667-685). Springer US.
- [1] Tsoumakas, G., Katakis, I., & Vlahavas, I. (2010). Mining multi-label data. In *Data mining and knowledge discovery handbook* (pp. 667-685). Springer US.

- [1] Vinh, Epps, and Bailey, (2010). Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance, JMLR
- [2] Wikipedia entry for the Adjusted Mutual Information
- [Hubert1985] L. Hubert and P. Arabie, Comparing Partitions, Journal of Classification 1985 <https://link.springer.com/article/10.1007%2FBF01908075>
- [wk] https://en.wikipedia.org/wiki/Rand_index#Adjusted_Rand_index
- [1] T. Calinski and J. Harabasz, 1974. "A dendrite method for cluster analysis". Communications in Statistics
- [1] Davies, David L.; Bouldin, Donald W. (1979). "A Cluster Separation Measure". IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-1 (2): 224-227
- [1] Andrew Rosenberg and Julia Hirschberg, 2007. V-Measure: A conditional entropy-based external cluster evaluation measure
- [1] E. B. Fowlkes and C. L. Mallows, 1983. "A method for comparing two hierarchical clusterings". Journal of the American Statistical Association
- [2] Wikipedia entry for the Fowlkes-Mallows Index
- [1] Andrew Rosenberg and Julia Hirschberg, 2007. V-Measure: A conditional entropy-based external cluster evaluation measure
- [1] Peter J. Rousseeuw (1987). "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". Computational and Applied Mathematics 20: 53-65.
- [2] Wikipedia entry on the Silhouette Coefficient
- [1] Peter J. Rousseeuw (1987). "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". Computational and Applied Mathematics 20: 53-65.
- [2] Wikipedia entry on the Silhouette Coefficient
- [1] Andrew Rosenberg and Julia Hirschberg, 2007. V-Measure: A conditional entropy-based external cluster evaluation measure
- [R16529824bff2-1] Bishop, Christopher M. (2006). "Pattern recognition and machine learning". Vol. 4 No. 4. New York: Springer.
- [R16529824bff2-2] Hagai Attias. (2000). "A Variational Bayesian Framework for Graphical Models". In Advances in Neural Information Processing Systems 12.
- [R16529824bff2-3] Blei, David M. and Michael I. Jordan. (2006). "Variational inference for Dirichlet process mixtures". Bayesian analysis 1.1
- [R2eddaeecc0849-1] "Solving multiclass learning problems via error-correcting output codes", Dietterich T., Bakiri G., Journal of Artificial Intelligence Research 2, 1995.
- [R2eddaeecc0849-2] "The error coding method and PICTs", James G., Hastie T., Journal of Computational and Graphical statistics 7, 1998.
- [R2eddaeecc0849-3] "The Elements of Statistical Learning", Hastie T., Tibshirani R., Friedman J., page 606 (second-edition) 2008.
- [Rca479bb49841-1] Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000, May). LOF: identifying density-based local outliers. In ACM sigmod record.
- [Rf9b6baee8229-1] J. Goldberger, G. Hinton, S. Roweis, R. Salakhutdinov. "Neighbourhood Components Analysis". Advances in Neural Information Processing Systems. 17, 513-520, 2005. <http://www.cs.nyu.edu/~roweis/papers/ncanips.pdf>

- [Rf9b6baee8229-2] Wikipedia entry on Neighborhood Components Analysis https://en.wikipedia.org/wiki/Neighbourhood_components_analysis
- [Rf3e1504535de-1] I.K. Yeo and R.A. Johnson, “A new family of power transformations to improve normality or symmetry.” *Biometrika*, 87(4), pp.954-959, (2000).
- [Rf3e1504535de-2] G.E.P. Box and D.R. Cox, “An Analysis of Transformations”, *Journal of the Royal Statistical Society B*, 26, 211-252 (1964).
- [1] I.K. Yeo and R.A. Johnson, “A new family of power transformations to improve normality or symmetry.” *Biometrika*, 87(4), pp.954-959, (2000).
- [2] G.E.P. Box and D.R. Cox, “An Analysis of Transformations”, *Journal of the Royal Statistical Society B*, 26, 211-252 (1964).
- [R0fecf191e4b8-1] Ping Li, T. Hastie and K. W. Church, 2006, “Very Sparse Random Projections”. https://web.stanford.edu/~hastie/Papers/Ping/KDD06_rp.pdf
- [R0fecf191e4b8-2] D. Achlioptas, 2001, “Database-friendly random projections”, <https://users.soe.ucsc.edu/~optas/papers/jl.pdf>
- [1] https://en.wikipedia.org/wiki/Johnson%E2%80%93Lindenstrauss_lemma
- [2] Sanjoy Dasgupta and Anupam Gupta, 1999, “An elementary proof of the Johnson-Lindenstrauss Lemma.” <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.45.3654>
- [R20c70293ef72-1] LIBSVM: A Library for Support Vector Machines
- [R20c70293ef72-2] Platt, John (1999). “Probabilistic outputs for support vector machines and comparison to regularized likelihood methods.”
- [Rb1ec977cd307-1] https://en.wikipedia.org/wiki/Decision_tree_learning
- [Rb1ec977cd307-2] L. Breiman, J. Friedman, R. Olshen, and C. Stone, “Classification and Regression Trees”, Wadsworth, Belmont, CA, 1984.
- [Rb1ec977cd307-3] T. Hastie, R. Tibshirani and J. Friedman. “Elements of Statistical Learning”, Springer, 2009.
- [Rb1ec977cd307-4] L. Breiman, and A. Cutler, “Random Forests”, https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm
- [Ra37b7e3adb19-1] https://en.wikipedia.org/wiki/Decision_tree_learning
- [Ra37b7e3adb19-2] L. Breiman, J. Friedman, R. Olshen, and C. Stone, “Classification and Regression Trees”, Wadsworth, Belmont, CA, 1984.
- [Ra37b7e3adb19-3] T. Hastie, R. Tibshirani and J. Friedman. “Elements of Statistical Learning”, Springer, 2009.
- [Ra37b7e3adb19-4] L. Breiman, and A. Cutler, “Random Forests”, https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm
- [Rdd99a0224c6e-1] P. Geurts, D. Ernst., and L. Wehenkel, “Extremely randomized trees”, *Machine Learning*, 63(1), 3-42, 2006.
- [R4939d63d5a49-1] P. Geurts, D. Ernst., and L. Wehenkel, “Extremely randomized trees”, *Machine Learning*, 63(1), 3-42, 2006.
- [1] [Wikipedia entry for the Jaccard index](#)