

References

- Abu-Mostafa, Y. (1995). Hints, *Neural Computation* **7**: 639–671.
- Ackley, D. H., Hinton, G. and Sejnowski, T. (1985). A learning algorithm for Boltzmann machines, *Trends in Cognitive Sciences* **9**: 147–169.
- Adam, B.-L., Qu, Y., Davis, J. W., Ward, M. D., Clements, M. A., Cazares, L. H., Semmes, O. J., Schellhammer, P. F., Yasui, Y., Feng, Z. and Wright, G. (2003). Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy mean, *Cancer Research* **63**(10): 3609–3614.
- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. and Verkamo, A. I. (1995). Fast discovery of association rules, *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Cambridge, MA.
- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*, Wiley, New York.
- Agresti, A. (2002). *Categorical Data Analysis (2nd Ed.)*, Wiley, New York.
- Ahn, J. and Marron, J. (2005). The direction of maximal data piling in high dimensional space, *Technical report*, Statistics Department, University of North Carolina, Chapel Hill.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *Second International Symposium on Information Theory*, pp. 267–281.

- Allen, D. (1974). The relationship between variable selection and data augmentation and a method of prediction, *Technometrics* **16**: 125–7.
- Ambrose, C. and McLachlan, G. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data, *Proceedings of the National Academy of Sciences* **99**: 6562–6566.
- Amit, Y. and Geman, D. (1997). Shape quantization and recognition with randomized trees, *Neural Computation* **9**: 1545–1588.
- Anderson, J. and Rosenfeld, E. (eds) (1988). *Neurocomputing: Foundations of Research*, MIT Press, Cambridge, MA.
- Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis, 3rd ed.*, Wiley, New York.
- Bach, F. and Jordan, M. (2002). Kernel independent component analysis, *Journal of Machine Learning Research* **3**: 1–48.
- Bair, E. and Tibshirani, R. (2004). Semi-supervised methods to predict patient survival from gene expression data, *PLOS Biology* **2**: 511–522.
- Bair, E., Hastie, T., Paul, D. and Tibshirani, R. (2006). Prediction by supervised principal components, *Journal of the American Statistical Association* **101**: 119–137.
- Bakin, S. (1999). Adaptive regression and model selection in data mining problems, *Technical report*, PhD. thesis, Australian National University, Canberra.
- Banerjee, O., Ghaoui, L. E. and d’Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data, *Journal of Machine Learning Research* **9**: 485–516.
- Barron, A. (1993). Universal approximation bounds for superpositions of a sigmoid function, *IEEE Transactions on Information Theory* **39**: 930–945.
- Bartlett, P. and Traskin, M. (2007). Adaboost is consistent, in B. Schölkopf, J. Platt and T. Hoffman (eds), *Advances in Neural Information Processing Systems 19*, MIT Press, Cambridge, MA, pp. 105–112.
- Becker, R., Cleveland, W. and Shyu, M. (1996). The visual design and control of trellis display, *Journal of Computational and Graphical Statistics* **5**: 123–155.
- Bell, A. and Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution, *Neural Computation* **7**: 1129–1159.

- Bellman, R. E. (1961). *Adaptive Control Processes*, Princeton University Press.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society Series B* **85**: 289–300.
- Benjamini, Y. and Yekutieli, Y. (2005). False discovery rate controlling confidence intervals for selected parameters, *Journal of the American Statistical Association* **100**: 71–80.
- Bickel, P. and Levina, E. (2004). Some theory for Fisher’s linear discriminant function, “Naive Bayes”, and some alternatives when there are many more variables than observations, *Bernoulli* **10**: 989–1010.
- Bickel, P. J., Ritov, Y. and Tsybakov, A. (2008). Simultaneous analysis of lasso and Dantzig selector, *Annals of Statistics*. to appear.
- Bishop, C. (1995). *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*, Springer, New York.
- Bishop, Y., Fienberg, S. and Holland, P. (1975). *Discrete Multivariate Analysis*, MIT Press, Cambridge, MA.
- Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*, Cambridge University Press.
- Breiman, L. (1992). The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error, *Journal of the American Statistical Association* **87**: 738–754.
- Breiman, L. (1996a). Bagging predictors, *Machine Learning* **26**: 123–140.
- Breiman, L. (1996b). Stacked regressions, *Machine Learning* **24**: 51–64.
- Breiman, L. (1998). Arcing classifiers (with discussion), *Annals of Statistics* **26**: 801–849.
- Breiman, L. (1999). Prediction games and arcing algorithms, *Neural Computation* **11**(7): 1493–1517.
- Breiman, L. (2001). Random forests, *Machine Learning* **45**: 5–32.
- Breiman, L. and Friedman, J. (1997). Predicting multivariate responses in multiple linear regression (with discussion), *Journal of the Royal Statistical Society Series B* **59**: 3–37.

- Breiman, L. and Ihaka, R. (1984). Nonlinear discriminant analysis via scaling and ACE, *Technical report*, University of California, Berkeley.
- Breiman, L. and Spector, P. (1992). Submodel selection and evaluation in regression: the X-random case, *International Statistical Review* **60**: 291–319.
- Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). *Classification and Regression Trees*, Wadsworth, New York.
- Bremaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer, New York.
- Brown, P., Spiegelman, C. and Denham, M. (1991). Chemometrics and spectral frequency selection, *Transactions of the Royal Society of London Series A*. **337**: 311–322.
- Bruce, A. and Gao, H. (1996). *Applied Wavelet Analysis with S-PLUS*, Springer, New York.
- Bühlmann, P. and Hothorn, T. (2007). Boosting algorithms: regularization, prediction and model fitting (with discussion), *Statistical Science* **22**(4): 477–505.
- Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models (with discussion), *Annals of Statistics* **17**: 453–555.
- Buja, A., Swayne, D., Littman, M., Hofmann, H. and Chen, L. (2008). Data visualization with multidimensional scaling, *Journal of Computational and Graphical Statistics*. to appear.
- Bunea, F., Tsybakov, A. and Wegkamp, M. (2007). Sparsity oracle inequalities for the lasso, *Electronic Journal of Statistics* **1**: 169–194.
- Burges, C. (1998). A tutorial on support vector machines for pattern recognition, *Knowledge Discovery and Data Mining* **2**(2): 121–167.
- Butte, A., Tamayo, P., Slonim, D., Golub, T. and Kohane, I. (2000). Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks, *Proceedings of the National Academy of Sciences* pp. 12182–12186.
- Candes, E. (2006). Compressive sampling, *Proceedings of the International Congress of Mathematicians*, European Mathematical Society, Madrid, Spain.
- Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n , *Annals of Statistics* **35**(6): 2313–2351.

- Chambers, J. and Hastie, T. (1991). *Statistical Models in S*, Wadsworth/Brooks Cole, Pacific Grove, CA.
- Chaudhuri, S., Drton, M. and Richardson, T. S. (2007). Estimation of a covariance matrix with zeros, *Biometrika* **94**(1): 1–18.
- Chen, L. and Buja, A. (2008). Local multidimensional scaling for nonlinear dimension reduction, graph drawing and proximity analysis, *Journal of the American Statistical Association*.
- Chen, S. S., Donoho, D. and Saunders, M. (1998). Atomic decomposition by basis pursuit, *SIAM Journal on Scientific Computing* **20**(1): 33–61.
- Cherkassky, V. and Ma, Y. (2003). Comparison of model selection for regression, *Neural computation* **15**(7): 1691–1714.
- Cherkassky, V. and Mulier, F. (2007). *Learning from Data (2nd Edition)*, Wiley, New York.
- Chui, C. (1992). *An Introduction to Wavelets*, Academic Press, London.
- Clifford, P. (1990). Markov random fields in statistics, in G. R. Grimmett and D. J. A. Welsh (eds), *Disorder in Physical Systems. A Volume in Honour of John M. Hammersley*, Clarendon Press, Oxford, pp. 19–32.
- Comon, P. (1994). Independent component analysis—a new concept?, *Signal Processing* **36**: 287–314.
- Cook, D. and Swayne, D. (2007). *Interactive and Dynamic Graphics for Data Analysis; with R and GGobi*, Springer, New York. With contributions from A. Buja, D. Temple Lang, H. Hofmann, H. Wickham and M. Lawrence.
- Cook, N. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction, *Circulation* **116**(6): 928–35.
- Copas, J. B. (1983). Regression, prediction and shrinkage (with discussion), *Journal of the Royal Statistical Society, Series B, Methodological* **45**: 311–354.
- Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification, *IEEE Transactions on Information Theory* **IT-11**: 21–27.
- Cover, T. and Thomas, J. (1991). *Elements of Information Theory*, Wiley, New York.
- Cox, D. and Hinkley, D. (1974). *Theoretical Statistics*, Chapman and Hall, London.

- Cox, D. and Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*, Chapman and Hall, London.
- Cressie, N. (1993). *Statistics for Spatial Data (Revised Edition)*, Wiley-Interscience, New York.
- Csiszar, I. and Tusnády, G. (1984). Information geometry and alternating minimization procedures, *Statistics & Decisions Supplement Issue 1*: 205–237.
- Cutler, A. and Breiman, L. (1994). Archetypal analysis, *Technometrics* **36**(4): 338–347.
- Dasarathy, B. (1991). *Nearest Neighbor Pattern Classification Techniques*, IEEE Computer Society Press, Los Alamitos, CA.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Daubechies, I., Defrise, M. and De Mol, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Communications on Pure and Applied Mathematics* **57**: 1413–1457.
- de Boor, C. (1978). *A Practical Guide to Splines*, Springer, New York.
- Dempster, A. (1972). Covariance selection, *Biometrics* **28**: 157–175.
- Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society Series B* **39**: 1–38.
- Devijver, P. and Kittler, J. (1982). *Pattern Recognition: A Statistical Approach*, Prentice-Hall, Englewood Cliffs, N.J.
- Dietterich, T. (2000a). Ensemble methods in machine learning, *Lecture Notes in Computer Science* **1857**: 1–15.
- Dietterich, T. (2000b). An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization, *Machine Learning* **40**(2): 139–157.
- Dietterich, T. and Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes, *Journal of Artificial Intelligence Research* **2**: 263–286.
- Donath, W. E. and Hoffman, A. J. (1973). Lower bounds for the partitioning of graphs, *IBM Journal of Research and Development* pp. 420–425.
- Donoho, D. (2006a). Compressed sensing, *IEEE Transactions on Information Theory* **52**(4): 1289–1306.

- Donoho, D. (2006b). For most large underdetermined systems of equations, the minimal ℓ^1 -norm solution is the sparsest solution, *Communications on Pure and Applied Mathematics* **59**: 797–829.
- Donoho, D. and Elad, M. (2003). Optimally sparse representation from overcomplete dictionaries via ℓ^1 -norm minimization, *Proceedings of the National Academy of Sciences* **100**: 2197–2202.
- Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**: 425–455.
- Donoho, D. and Stodden, V. (2004). When does non-negative matrix factorization give a correct decomposition into parts?, in S. Thrun, L. Saul and B. Schölkopf (eds), *Advances in Neural Information Processing Systems 16*, MIT Press, Cambridge, MA.
- Duan, N. and Li, K.-C. (1991). Slicing regression: a link-free regression method, *Annals of Statistics* **19**: 505–530.
- Duchamp, T. and Stuetzle, W. (1996). Extremal properties of principal curves in the plane, *Annals of Statistics* **24**: 1511–1520.
- Duda, R., Hart, P. and Stork, D. (2000). *Pattern Classification (2nd Edition)*, Wiley, New York.
- Dudoit, S., Fridlyand, J. and Speed, T. (2002a). Comparison of discrimination methods for the classification of tumors using gene expression data, *Journal of the American Statistical Association* **97**(457): 77–87.
- Dudoit, S., Yang, Y., Callow, M. and Speed, T. (2002b). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments, *Statistica Sinica* pp. 111–139.
- Edwards, D. (2000). *Introduction to Graphical Modelling, 2nd Edition*, Springer, New York.
- Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis, *Journal of the American Statistical Association* **70**: 892–898.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Annals of Statistics* **7**: 1–26.
- Efron, B. (1983). Estimating the error rate of a prediction rule: some improvements on cross-validation, *Journal of the American Statistical Association* **78**: 316–331.
- Efron, B. (1986). How biased is the apparent error rate of a prediction rule?, *Journal of the American Statistical Association* **81**: 461–70.

- Efron, B. and Tibshirani, R. (1991). Statistical analysis in the computer age, *Science* **253**: 390–395.
- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, London.
- Efron, B. and Tibshirani, R. (1996). Using specially designed exponential families for density estimation, *Annals of Statistics* **24**(6): 2431–2461.
- Efron, B. and Tibshirani, R. (1997). Improvements on cross-validation: the 632+ bootstrap: method, *Journal of the American Statistical Association* **92**: 548–560.
- Efron, B. and Tibshirani, R. (2002). Microarrays, empirical Bayes methods, and false discovery rates, *Genetic Epidemiology* **1**: 70–86.
- Efron, B., Hastie, T. and Tibshirani, R. (2007). Discussion of “Dantzig selector” by Candès and Tao, *Annals of Statistics* **35**(6): 2358–2364.
- Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression (with discussion), *Annals of Statistics* **32**(2): 407–499.
- Efron, B., Tibshirani, R., Storey, J. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment, *Journal of the American Statistical Association* **96**: 1151–1160.
- Evgeniou, T., Pontil, M. and Poggio, T. (2000). Regularization networks and support vector machines, *Advances in Computational Mathematics* **13**(1): 1–50.
- Fan, J. and Fan, Y. (2008). High dimensional classification using features annealed independence rules, *Annals of Statistics*. to appear.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.
- Fan, J. and Li, R. (2005). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association* **96**: 1348–1360.
- Fiedler, M. (1973). Algebraic connectivity of graphs, *Czechoslovak Mathematics Journal* **23**(98): 298–305.
- Fienberg, S. (1977). *The Analysis of Cross-Classified Categorical Data*, MIT Press, Cambridge.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems, *Eugen.* **7**: 179–188.

- Fisher, W. (1958). On grouping for maximum homogeneity, *Journal of the American Statistical Association* **53**(284): 789–798.
- Fix, E. and Hodges, J. (1951). Discriminatory analysis—nonparametric discrimination: Consistency properties, *Technical Report 21-49-004,4*, U.S. Air Force, School of Aviation Medicine, Randolph Field, TX.
- Flury, B. (1990). Principal points, *Biometrika* **77**: 33–41.
- Forgy, E. (1965). Cluster analysis of multivariate data: efficiency vs. interpretability of classifications, *Biometrics* **21**: 768–769.
- Frank, I. and Friedman, J. (1993). A statistical view of some chemometrics regression tools (with discussion), *Technometrics* **35**(2): 109–148.
- Freund, Y. (1995). Boosting a weak learning algorithm by majority, *Information and Computation* **121**(2): 256–285.
- Freund, Y. and Schapire, R. (1996a). Experiments with a new boosting algorithm, *Machine Learning: Proceedings of the Thirteenth International Conference*, Morgan Kaufman, San Francisco, pp. 148–156.
- Freund, Y. and Schapire, R. (1996b). Game theory, on-line prediction and boosting, *Proceedings of the Ninth Annual Conference on Computational Learning Theory*, Desenzano del Garda, Italy, pp. 325–332.
- Freund, Y. and Schapire, R. (1997). A decision-theoretic generalization of online learning and an application to boosting, *Journal of Computer and System Sciences* **55**: 119–139.
- Friedman, J. (1987). Exploratory projection pursuit, *Journal of the American Statistical Association* **82**: 249–266.
- Friedman, J. (1989). Regularized discriminant analysis, *Journal of the American Statistical Association* **84**: 165–175.
- Friedman, J. (1991). Multivariate adaptive regression splines (with discussion), *Annals of Statistics* **19**(1): 1–141.
- Friedman, J. (1994a). Flexible metric nearest-neighbor classification, *Technical report*, Stanford University.
- Friedman, J. (1994b). An overview of predictive learning and function approximation, in V. Cherkassky, J. Friedman and H. Wechsler (eds), *From Statistics to Neural Networks*, Vol. 136 of *NATO ISI Series F*, Springer, New York.
- Friedman, J. (1996). Another approach to polychotomous classification, *Technical report*, Stanford University.

- Friedman, J. (1997). On bias, variance, 0-1 loss and the curse of dimensionality, *Journal of Data Mining and Knowledge Discovery* **1**: 55–77.
- Friedman, J. (1999). Stochastic gradient boosting, *Technical report*, Stanford University.
- Friedman, J. (2001). Greedy function approximation: A gradient boosting machine, *Annals of Statistics* **29**(5): 1189–1232.
- Friedman, J. and Fisher, N. (1999). Bump hunting in high dimensional data, *Statistics and Computing* **9**: 123–143.
- Friedman, J. and Hall, P. (2007). On bagging and nonlinear estimation, *Journal of Statistical Planning and Inference* **137**: 669–683.
- Friedman, J. and Popescu, B. (2003). Importance sampled learning ensembles, *Technical report*, Stanford University, Department of Statistics.
- Friedman, J. and Popescu, B. (2008). Predictive learning via rule ensembles, *Annals of Applied Statistics*, to appear.
- Friedman, J. and Silverman, B. (1989). Flexible parsimonious smoothing and additive modelling (with discussion), *Technometrics* **31**: 3–39.
- Friedman, J. and Stuetzle, W. (1981). Projection pursuit regression, *Journal of the American Statistical Association* **76**: 817–823.
- Friedman, J. and Tukey, J. (1974). A projection pursuit algorithm for exploratory data analysis, *IEEE Transactions on Computers, Series C* **23**: 881–889.
- Friedman, J., Baskett, F. and Shustek, L. (1975). An algorithm for finding nearest neighbors, *IEEE Transactions on Computers* **24**: 1000–1006.
- Friedman, J., Bentley, J. and Finkel, R. (1977). An algorithm for finding best matches in logarithmic expected time, *ACM Transactions on Mathematical Software* **3**: 209–226.
- Friedman, J., Hastie, T. and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion), *Annals of Statistics* **28**: 337–307.
- Friedman, J., Hastie, T. and Tibshirani, R. (2008a). Response to “Mease and Wyner: Evidence contrary to the statistical view of boosting”, *Journal of Machine Learning Research* **9**: 175–180.
- Friedman, J., Hastie, T. and Tibshirani, R. (2008b). Sparse inverse covariance estimation with the graphical lasso, *Biostatistics* **9**: 432–441.

- Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software* **33**(1): 1–22.
- Friedman, J., Hastie, T., Hoefling, H. and Tibshirani, R. (2007). Pathwise coordinate optimization, *Annals of Applied Statistics* **2**(1): 302–332.
- Friedman, J., Hastie, T., Rosset, S., Tibshirani, R. and Zhu, J. (2004). Discussion of three boosting papers by Jiang, Lugosi and Vayatis, and Zhang, *Annals of Statistics* **32**: 102–107.
- Friedman, J., Stuetzle, W. and Schroeder, A. (1984). Projection pursuit density estimation, *Journal of the American Statistical Association* **79**: 599–608.
- Fu, W. (1998). Penalized regressions: the bridge vs. the lasso, *Journal of Computational and Graphical Statistics* **7**(3): 397–416.
- Furnival, G. and Wilson, R. (1974). Regression by leaps and bounds, *Technometrics* **16**: 499–511.
- Gelfand, A. and Smith, A. (1990). Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association* **85**: 398–409.
- Gelman, A., Carlin, J., Stern, H. and Rubin, D. (1995). *Bayesian Data Analysis*, CRC Press, Boca Raton, FL.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.
- Genkin, A., Lewis, D. and Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization, *Technometrics* **49**(3): 291–304.
- Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery rates, *Annals of Statistics* **32**(3): 1035–1061.
- Gersho, A. and Gray, R. (1992). *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, MA.
- Girosi, F., Jones, M. and Poggio, T. (1995). Regularization theory and neural network architectures, *Neural Computation* **7**: 219–269.
- Golub, G. and Van Loan, C. (1983). *Matrix Computations*, Johns Hopkins University Press, Baltimore.
- Golub, G., Heath, M. and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics* **21**: 215–224.

- Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri, M., Bloomfield, C. and Lander, E. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, *Science* **286**: 531–536.
- Goodall, C. (1991). Procrustes methods in the statistical analysis of shape, *Journal of the Royal Statistical Society, Series B* **53**: 285–321.
- Gordon, A. (1999). *Classification (2nd edition)*, Chapman and Hall/CRC Press, London.
- Green, P. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall, London.
- Greenacre, M. (1984). *Theory and Applications of Correspondence Analysis*, Academic Press, New York.
- Greenshtein, E. and Ritov, Y. (2004). Persistence in high-dimensional linear predictor selection and the virtue of overparametrization, *Bernoulli* **10**: 971–988.
- Guo, Y., Hastie, T. and Tibshirani, R. (2006). Regularized linear discriminant analysis and its application in microarrays, *Biostatistics* **8**: 86–100.
- Guyon, I., Gunn, S., Nikravesh, M. and Zadeh, L. (eds) (2006). *Feature Extraction, Foundations and Applications*, Springer, New York.
- Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines, *Machine Learning* **46**: 389–422.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.
- Hammersley, J. M. and Clifford, P. (1971). Markov field on finite graphs and lattices, unpublished.
- Hand, D. (1981). *Discrimination and Classification*, Wiley, Chichester.
- Hanley, J. and McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve, *Radiology* **143**: 29–36.
- Hart, P. (1968). The condensed nearest-neighbor rule, *IEEE Transactions on Information Theory* **14**: 515–516.
- Hartigan, J. A. (1975). *Clustering Algorithms*, Wiley, New York.

- Hartigan, J. A. and Wong, M. A. (1979). [(Algorithm AS 136] A k -means clustering algorithm (AS R39: 81v30 p355-356), *Applied Statistics* **28**: 100–108.
- Hastie, T. (1984). *Principal Curves and Surfaces*, PhD thesis, Stanford University.
- Hastie, T. and Herman, A. (1990). An analysis of gestational age, neonatal size and neonatal death using nonparametric logistic regression, *Journal of Clinical Epidemiology* **43**: 1179–90.
- Hastie, T. and Simard, P. (1998). Models and metrics for handwritten digit recognition, *Statistical Science* **13**: 54–65.
- Hastie, T. and Stuetzle, W. (1989). Principal curves, *Journal of the American Statistical Association* **84**(406): 502–516.
- Hastie, T. and Tibshirani, R. (1987). Nonparametric logistic and proportional odds regression, *Applied Statistics* **36**: 260–276.
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall, London.
- Hastie, T. and Tibshirani, R. (1996a). Discriminant adaptive nearest-neighbor classification, *IEEE Pattern Recognition and Machine Intelligence* **18**: 607–616.
- Hastie, T. and Tibshirani, R. (1996b). Discriminant analysis by Gaussian mixtures, *Journal of the Royal Statistical Society Series B.* **58**: 155–176.
- Hastie, T. and Tibshirani, R. (1998). Classification by pairwise coupling, *Annals of Statistics* **26**(2): 451–471.
- Hastie, T. and Tibshirani, R. (2003). Independent components analysis through product density estimation, in S. T. S. Becker and K. Obermayer (eds), *Advances in Neural Information Processing Systems 15*, MIT Press, Cambridge, MA, pp. 649–656.
- Hastie, T. and Tibshirani, R. (2004). Efficient quadratic regularization for expression arrays, *Biostatistics* **5**(3): 329–340.
- Hastie, T. and Zhu, J. (2006). Discussion of “Support vector machines with applications” by Javier Moguerza and Alberto Munoz, *Statistical Science* **21**(3): 352–357.
- Hastie, T., Botha, J. and Schnitzler, C. (1989). Regression with an ordered categorical response, *Statistics in Medicine* **43**: 884–889.

- Hastie, T., Buja, A. and Tibshirani, R. (1995). Penalized discriminant analysis, *Annals of Statistics* **23**: 73–102.
- Hastie, T., Kishon, E., Clark, M. and Fan, J. (1992). A model for signature verification, *Technical report*, AT&T Bell Laboratories. <http://www-stat.stanford.edu/~hastie/Papers/signature.pdf>.
- Hastie, T., Rosset, S., Tibshirani, R. and Zhu, J. (2004). The entire regularization path for the support vector machine, *Journal of Machine Learning Research* **5**: 1391–1415.
- Hastie, T., Taylor, J., Tibshirani, R. and Walther, G. (2007). Forward stagewise regression and the monotone lasso, *Electronic Journal of Statistics* **1**: 1–29.
- Hastie, T., Tibshirani, R. and Buja, A. (1994). Flexible discriminant analysis by optimal scoring, *Journal of the American Statistical Association* **89**: 1255–1270.
- Hastie, T., Tibshirani, R. and Buja, A. (2000). Flexible discriminant and mixture models, in J. Kay and M. Titterton (eds), *Statistics and Artificial Neural Networks*, Oxford University Press.
- Hastie, T., Tibshirani, R. and Friedman, J. (2003). A note on “Comparison of model selection for regression” by Cherkassky and Ma, *Neural computation* **15**(7): 1477–1480.
- Hathaway, R. J. (1986). Another interpretation of the EM algorithm for mixture distributions, *Statistics & Probability Letters* **4**: 53–56.
- Hebb, D. (1949). *The Organization of Behavior*, Wiley, New York.
- Hertz, J., Krogh, A. and Palmer, R. (1991). *Introduction to the Theory of Neural Computation*, Addison Wesley, Redwood City, CA.
- Hinton, G. (1989). Connectionist learning procedures, *Artificial Intelligence* **40**: 185–234.
- Hinton, G. (2002). Training products of experts by minimizing contrastive divergence, *Neural Computation* **14**: 1771–1800.
- Hinton, G., Osindero, S. and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets, *Neural Computation* **18**: 1527–1554.
- Ho, T. K. (1995). Random decision forests, in M. Kavavaugh and P. Storms (eds), *Proc. Third International Conference on Document Analysis and Recognition*, Vol. 1, IEEE Computer Society Press, New York, pp. 278–282.

- Hoefling, H. and Tibshirani, R. (2008). Estimation of sparse Markov networks using modified logistic regression and the lasso, submitted.
- Hoerl, A. E. and Kennard, R. (1970). Ridge regression: biased estimation for nonorthogonal problems, *Technometrics* **12**: 55–67.
- Hothorn, T. and Bühlmann, P. (2006). Model-based boosting in high dimensions, *Bioinformatics* **22**(22): 2828–2829.
- Huber, P. (1964). Robust estimation of a location parameter, *Annals of Mathematical Statistics* **53**: 73–101.
- Huber, P. (1985). Projection pursuit, *Annals of Statistics* **13**: 435–475.
- Hunter, D. and Lange, K. (2004). A tutorial on MM algorithms, *The American Statistician* **58**(1): 30–37.
- Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications, *Neural Networks* **13**: 411–430.
- Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*, Wiley, New York.
- Izenman, A. (1975). Reduced-rank regression for the multivariate linear model, *Journal of Multivariate Analysis* **5**: 248–264.
- Jacobs, R., Jordan, M., Nowlan, S. and Hinton, G. (1991). Adaptive mixtures of local experts, *Neural computation* **3**: 79–87.
- Jain, A. and Dubes, R. (1988). *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, N.J.
- James, G. and Hastie, T. (1998). The error coding method and PICTs, *Journal of Computational and Graphical Statistics* **7**(3): 377–387.
- Jancey, R. (1966). Multidimensional group analysis, *Australian Journal of Botany* **14**: 127–130.
- Jensen, F. V., Lauritzen, S. and Olesen, K. G. (1990). Bayesian updating in recursive graphical models by local computation, *Computational Statistics Quarterly* **4**: 269–282.
- Jiang, W. (2004). Process consistency for Adaboost, *Annals of Statistics* **32**(1): 13–29.
- Jiroušek, R. and Přeučil, S. (1995). On the effective implementation of the iterative proportional fitting procedure, *Computational Statistics and Data Analysis* **19**: 177–189.
- Johnson, N. (2008). A study of the NIPS feature selection challenge, Submitted.

- Joliffe, I. T., Trendafilov, N. T. and Uddin, M. (2003). A modified principal component technique based on the lasso, *Journal of Computational and Graphical Statistics* **12**: 531–547.
- Jones, L. (1992). A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training, *Annals of Statistics* **20**: 608–613.
- Jordan, M. (2004). Graphical models, *Statistical Science (Special Issue on Bayesian Statistics)* **19**: 140–155.
- Jordan, M. and Jacobs, R. (1994). Hierarchical mixtures of experts and the EM algorithm, *Neural Computation* **6**: 181–214.
- Kalbfleisch, J. and Prentice, R. (1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.
- Kearns, M. and Vazirani, U. (1994). *An Introduction to Computational Learning Theory*, MIT Press, Cambridge, MA.
- Kittler, J., Hatef, M., Duin, R. and Matas, J. (1998). On combining classifiers, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **20**(3): 226–239.
- Kleinberg, E. M. (1990). Stochastic discrimination, *Annals of Mathematical Artificial Intelligence* **1**: 207–239.
- Kleinberg, E. M. (1996). An overtraining-resistant stochastic modeling method for pattern recognition, *Annals of Statistics* **24**: 2319–2349.
- Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators, *Annals of Statistics* **28**(5): 1356–1378.
- Koh, K., Kim, S.-J. and Boyd, S. (2007). An interior-point method for large-scale L1-regularized logistic regression, *Journal of Machine Learning Research* **8**: 1519–1555.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection, *International Joint Conference on Artificial Intelligence (IJCAI)*, Morgan Kaufmann, pp. 1137–1143.
- Kohonen, T. (1989). *Self-Organization and Associative Memory (3rd edition)*, Springer, Berlin.
- Kohonen, T. (1990). The self-organizing map, *Proceedings of the IEEE* **78**: 1464–1479.

- Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Paatero, A. and Saarela, A. (2000). Self-organization of a massive document collection, *IEEE Transactions on Neural Networks* **11**(3): 574–585. Special Issue on Neural Networks for Data Mining and Knowledge Discovery.
- Koller, D. and Friedman, N. (2007). *Structured Probabilistic Models*, Stanford Bookstore Custom Publishing. (Unpublished Draft).
- Kressel, U. (1999). Pairwise classification and support vector machines, in B. Schölkopf, C. Burges and A. Smola (eds), *Advances in Kernel Methods - Support Vector Learning*, MIT Press, Cambridge, MA., pp. 255–268.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing, *Technometrics* **34**(1): 1–14.
- Lange, K. (2004). *Optimization*, Springer, New York.
- Lauritzen, S. (1996). *Graphical Models*, Oxford University Press.
- Lauritzen, S. and Spiegelhalter, D. (1988). Local computations with probabilities on graphical structures and their application to expert systems, *J. Royal Statistical Society B.* **50**: 157–224.
- Lawson, C. and Hansen, R. (1974). *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ.
- Le Cun, Y. (1989). Generalization and network design strategies, *Technical Report CRG-TR-89-4*, Department of Computer Science, Univ. of Toronto.
- Le Cun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W. and Jackel, L. (1990). Handwritten digit recognition with a back-propagation network, in D. Touretzky (ed.), *Advances in Neural Information Processing Systems*, Vol. 2, Morgan Kaufman, Denver, CO, pp. 386–404.
- Le Cun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE* **86**(11): 2278–2324.
- Leathwick, J., Elith, J., Francis, M., Hastie, T. and Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding new zealand: an analysis using boosted regression trees, *Marine Ecology Progress Series* **77**: 802–813.
- Leathwick, J., Rowe, D., Richardson, J., Elith, J. and Hastie, T. (2005). Using multivariate adaptive regression splines to predict the distributions of New Zealand’s freshwater diadromous fish, *Freshwater Biology* **50**: 2034–2051.

- Leblanc, M. and Tibshirani, R. (1996). Combining estimates in regression and classification, *Journal of the American Statistical Association* **91**: 1641–1650.
- LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE* **86**(11): 2278–2324.
- Lee, D. and Seung, H. (1999). Learning the parts of objects by non-negative matrix factorization, *Nature* **401**: 788.
- Lee, D. and Seung, H. (2001). Algorithms for non-negative matrix factorization, *Advances in Neural Information Processing Systems, (NIPS 2001)*, Vol. 13, Morgan Kaufman, Denver., pp. 556–562.
- Lee, M.-L. (2004). *Analysis of Microarray Gene Expression Data*, Kluwer Academic Publishers.
- Lee, S.-I., Ganapathi, V. and Koller, D. (2007). Efficient structure learning of markov networks using l_1 -regularization, in B. Schölkopf, J. Platt and T. Hoffman (eds), *Advances in Neural Information Processing Systems 19*, MIT Press, Cambridge, MA, pp. 817–824.
- Leslie, C., Eskin, E., Cohen, A., Weston, J. and Noble, W. S. (2004). Mismatch string kernels for discriminative protein classification, *Bioinformatics* **20**(4): 467–476.
- Levina, E. (2002). *Statistical issues in texture analysis*, PhD thesis, Department. of Statistics, University of California, Berkeley.
- Lin, H., McCulloch, C., Turnbull, B., Slate, E. and Clark, L. (2000). A latent class mixed model for analyzing biomarker trajectories in longitudinal data with irregularly scheduled observations, *Statistics in Medicine* **19**: 1303–1318.
- Lin, Y. and Zhang, H. (2006). Component selection and smoothing in smoothing spline analysis of variance models, *Annals of Statistics* **34**: 2272–2297.
- Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data (2nd Edition)*, Wiley, New York.
- Lloyd, S. (1957). Least squares quantization in PCM., *Technical report*, Bell Laboratories. Published in 1982 in *IEEE Transactions on Information Theory* **28** 128–137.
- Loader, C. (1999). *Local Regression and Likelihood*, Springer, New York.

- Loh, W. and Vanichsetakul, N. (1988). Tree structured classification via generalized discriminant analysis, *Journal of the American Statistical Association* **83**: 715–728.
- Lugosi, G. and Vayatis, N. (2004). On the bayes-risk consistency of regularized boosting methods, *Annals of Statistics* **32**(1): 30–55.
- Macnaughton Smith, P., Williams, W., Dale, M. and Mockett, L. (1965). Dissimilarity analysis: a new technique of hierarchical subdivision, *Nature* **202**: 1034–1035.
- MacKay, D. (1992). A practical Bayesian framework for backpropagation neural networks, *Neural Computation* **4**: 448–472.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, eds. L.M. LeCam and J. Neyman, University of California Press, pp. 281–297.
- Madigan, D. and Raftery, A. (1994). Model selection and accounting for model uncertainty using Occam’s window, *Journal of the American Statistical Association* **89**: 1535–46.
- Mardia, K., Kent, J. and Bibby, J. (1979). *Multivariate Analysis*, Academic Press.
- Mason, L., Baxter, J., Bartlett, P. and Frean, M. (2000). Boosting algorithms as gradient descent, **12**: 512–518.
- Massart, D., Plastria, F. and Kaufman, L. (1983). Non-hierarchical clustering with MASLOC, *The Journal of the Pattern Recognition Society* **16**: 507–516.
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, Chapman and Hall, London.
- McCulloch, W. and Pitts, W. (1943). A logical calculus of the ideas imminent in nervous activity, *Bulletin of Mathematical Biophysics* **5**: 115–133. Reprinted in Anderson and Rosenfeld (1988), pp 96–104.
- McLachlan, G. (1992). *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York.
- Mease, D. and Wyner, A. (2008). Evidence contrary to the statistical view of boosting (with discussion), *Journal of Machine Learning Research* **9**: 131–156.
- Meinshausen, N. (2007). Relaxed lasso, *Computational Statistics and Data Analysis* **52**(1): 374–393.

- Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso, *Annals of Statistics* **34**: 1436–1462.
- Meir, R. and Rätsch, G. (2003). An introduction to boosting and leveraging, in S. Mendelson and A. Smola (eds), *Lecture notes in Computer Science*, Advanced Lectures in Machine Learning, Springer, New York.
- Michie, D., Spiegelhalter, D. and Taylor, C. (eds) (1994). *Machine Learning, Neural and Statistical Classification*, Ellis Horwood Series in Artificial Intelligence, Ellis Horwood.
- Morgan, J. N. and Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal, *Journal of the American Statistical Association* **58**: 415–434.
- Murray, W., Gill, P. and Wright, M. (1981). *Practical Optimization*, Academic Press.
- Myles, J. and Hand, D. (1990). The multiclass metric problem in nearest neighbor classification, *Pattern Recognition* **23**: 1291–1297.
- Nadler, B. and Coifman, R. R. (2005). An exact asymptotic formula for the error in CLS and in PLS: The importance of dimensional reduction in multivariate calibration, *Journal of Chemometrics* **102**: 107–118.
- Neal, R. (1996). *Bayesian Learning for Neural Networks*, Springer, New York.
- Neal, R. and Hinton, G. (1998). *A view of the EM algorithm that justifies incremental, sparse, and other variants; in Learning in Graphical Models, M. Jordan (ed.)*, Dordrecht: Kluwer Academic Publishers, Boston, MA., pp. 355–368.
- Neal, R. and Zhang, J. (2006). High dimensional classification with bayesian neural networks and dirichlet diffusion trees, in I. Guyon, S. Gunn, M. Nikravesh and L. Zadeh (eds), *Feature Extraction, Foundations and Applications*, Springer, New York, pp. 265–296.
- Onton, J. and Makeig, S. (2006). Information-based modeling of event-related brain dynamics, in Neuper and Klimesch (eds), *Progress in Brain Research*, Vol. 159, Elsevier, pp. 99–120.
- Osborne, M., Presnell, B. and Turlach, B. (2000a). A new approach to variable selection in least squares problems, *IMA Journal of Numerical Analysis* **20**: 389–404.
- Osborne, M., Presnell, B. and Turlach, B. (2000b). On the lasso and its dual, *Journal of Computational and Graphical Statistics* **9**: 319–337.

- Pace, R. K. and Barry, R. (1997). Sparse spatial autoregressions, *Statistics and Probability Letters* **33**: 291–297.
- Page, L., Brin, S., Motwani, R. and Winograd, T. (1998). The pagerank citation ranking: bringing order to the web, *Technical report*, Stanford Digital Library Technologies Project. <http://citeseer.ist.psu.edu/page98pagerank.html>.
- Park, M. Y. and Hastie, T. (2007). l_1 -regularization path algorithm for generalized linear models, *Journal of the Royal Statistical Society Series B* **69**: 659–677.
- Parker, D. (1985). Learning logic, *Technical Report TR-87*, Cambridge MA: MIT Center for Research in Computational Economics and Management Science.
- Parmigiani, G., Garrett, E. S., Irizarry, R. A. and Zeger, S. L. (eds) (2003). *The Analysis of Gene Expression Data*, Springer, New York.
- Paul, D., Bair, E., Hastie, T. and Tibshirani, R. (2008). “Pre-conditioning” for feature selection and regression in high-dimensional problems, *Annals of Statistics* **36**(4): 1595–1618.
- Pearl, J. (1986). On evidential reasoning in a hierarchy of hypotheses, *Artificial Intelligence* **28**: 9–15.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann, San Francisco, CA.
- Pearl, J. (2000). *Causality: Models, Reasoning and Inference*, Cambridge University Press.
- Peterson and Anderson, J. R. (1987). A mean field theory learning algorithm for neural networks, *Complex Systems* **1**: 995–1019.
- Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V., Steinberg, S. M., Mills, G. B., Simone, C., Fishman, D. A., Kohn, E. and Liotta, L. A. (2002). Use of proteomic patterns in serum to identify ovarian cancer, *Lancet* **359**: 572–577.
- Platt, J. (1999). *Fast Training of Support Vector Machines using Sequential Minimal Optimization; in Advances in Kernel Methods—Support Vector Learning*, B. Schölkopf and C. J. C. Burges and A. J. Smola (eds), MIT Press, Cambridge, MA., pp. 185–208.
- Quinlan, R. (1993). *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo.
- Quinlan, R. (2004). C5.0, www.rulequest.com.

- Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J., Poggio, T., Gerald, W., Loda, M., Lander, E. and Golub, T. (2001). Multiclass cancer diagnosis using tumor gene expression signature, *PNAS* **98**: 15149–15154.
- Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*, Springer, New York.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.
- Rätsch, G. and Warmuth, M. (2002). Maximizing the margin with boosting, *Proceedings of the 15th Annual Conference on Computational Learning Theory*, pp. 334–350.
- Ravikumar, P., Liu, H., Lafferty, J. and Wasserman, L. (2008). Spam: Sparse additive models, in J. Platt, D. Koller, Y. Singer and S. Roweis (eds), *Advances in Neural Information Processing Systems 20*, MIT Press, Cambridge, MA, pp. 1201–1208.
- Ridgeway, G. (1999). The state of boosting, *Computing Science and Statistics* **31**: 172–181.
- Rieger, K., Hong, W., Tusher, V., Tang, J., Tibshirani, R. and Chu, G. (2004). Toxicity from radiation therapy associated with abnormal transcriptional responses to DNA damage, *Proceedings of the National Academy of Sciences* **101**: 6634–6640.
- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*, Cambridge University Press.
- Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length, *Annals of Statistics* **11**: 416–431.
- Robbins, H. and Munro, S. (1951). A stochastic approximation method, *Annals of Mathematical Statistics* **22**: 400–407.
- Roosen, C. and Hastie, T. (1994). Automatic smoothing spline projection pursuit, *Journal of Computational and Graphical Statistics* **3**: 235–248.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain, *Psychological Review* **65**: 386–408.
- Rosenblatt, F. (1962). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Spartan, Washington, D.C.

- Rosenwald, A., Wright, G., Chan, W. C., Connors, J. M., Campo, E., Fisher, R. I., Gascoyne, R. D., Muller-Hermelink, H. K., Smeland, E. B. and Staudt, L. M. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large b-cell lymphoma, *The New England Journal of Medicine* **346**: 1937–1947.
- Rosset, S. and Zhu, J. (2007). Piecewise linear regularized solution paths, *Annals of Statistics* **35**(3): 1012–1030.
- Rosset, S., Zhu, J. and Hastie, T. (2004a). Boosting as a regularized path to a maximum margin classifier, *Journal of Machine Learning Research* **5**: 941–973.
- Rosset, S., Zhu, J. and Hastie, T. (2004b). Margin maximizing loss functions, in S. Thrun, L. Saul and B. Schölkopf (eds), *Advances in Neural Information Processing Systems 16*, MIT Press, Cambridge, MA.
- Rousseauw, J., du Plessis, J., Benade, A., Jordaan, P., Kotze, J., Jooste, P. and Ferreira, J. (1983). Coronary risk factor screening in three rural communities, *South African Medical Journal* **64**: 430–436.
- Roweis, S. T. and Saul, L. K. (2000). Locally linear embedding, *Science* **290**: 2323–2326.
- Rumelhart, D., Hinton, G. and Williams, R. (1986). Learning internal representations by error propagation, in D. Rumelhart and J. McClelland (eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, The MIT Press, Cambridge, MA., pp. 318–362.
- Sachs, K., Perez, O., Pe’er, D., Lauffenburger, D. and Nolan, G. (2005). Causal protein-signaling networks derived from multiparameter single-cell data, *Science* **308**: 523–529.
- Schapire, R. (1990). The strength of weak learnability, *Machine Learning* **5**(2): 197–227.
- Schapire, R. (2002). The boosting approach to machine learning: an overview, in D. Denison, M. Hansen, C. Holmes, B. Mallick and B. Yu (eds), *MSRI workshop on Nonlinear Estimation and Classification*, Springer, New York.
- Schapire, R. and Singer, Y. (1999). Improved boosting algorithms using confidence-rated predictions, *Machine Learning* **37**(3): 297–336.
- Schapire, R., Freund, Y., Bartlett, P. and Lee, W. (1998). Boosting the margin: a new explanation for the effectiveness of voting methods, *Annals of Statistics* **26**(5): 1651–1686.

- Schölkopf, B., Smola, A. and Müller, K.-R. (1999). Kernel principal component analysis, in B. Schölkopf, C. Burges and A. Smola (eds), *Advances in Kernel Methods—Support Vector Learning*, MIT Press, Cambridge, MA, USA, pp. 327–352.
- Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* **6**(2): 461–464.
- Scott, D. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley, New York.
- Seber, G. (1984). *Multivariate Observations*, Wiley, New York.
- Segal, M. (2004). Machine learning benchmarks and random forest regression, *Technical report*, eScholarship Repository, University of California. http://repositories.edlib.org/cbmb/bench_rf_regn.
- Shao, J. (1996). Bootstrap model selection, *Journal of the American Statistical Association* **91**: 655–665.
- Shenoy, P. and Shafer, G. (1988). An axiomatic framework for Bayesian and belief-function propagation, *AAAI Workshop on Uncertainty in AI*, North-Holland, pp. 307–314.
- Short, R. and Fukunaga, K. (1981). The optimal distance measure for nearest neighbor classification, *IEEE Transactions on Information Theory* **27**: 622–627.
- Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- Silvey, S. (1975). *Statistical Inference*, Chapman and Hall, London.
- Simard, P., Cun, Y. L. and Denker, J. (1993). Efficient pattern recognition using a new transformation distance, *Advances in Neural Information Processing Systems*, Morgan Kaufman, San Mateo, CA, pp. 50–58.
- Simon, R. M., Korn, E. L., McShane, L. M., Radmacher, M. D., Wright, G. and Zhao, Y. (2004). *Design and Analysis of DNA Microarray Investigations*, Springer, New York.
- Sjöstrand, K., Rostrup, E., Ryberg, C., Larsen, R., Studholme, C., Baezner, H., Ferro, J., Fazekas, F., Pantoni, L., Inzitari, D. and Waldemar, G. (2007). Sparse decomposition and modeling of anatomical shape variation, *IEEE Transactions on Medical Imaging* **26**(12): 1625–1635.
- Speed, T. and Kiiveri, H. T. (1986). Gaussian Markov distributions over finite graphs, *Annals of Statistics* **14**: 138–150.

- Speed, T. (ed.) (2003). *Statistical Analysis of Gene Expression Microarray Data*, Chapman and Hall, London.
- Spiegelhalter, D., Best, N., Gilks, W. and Inskip, H. (1996). Hepatitis B: a case study in MCMC methods, in W. Gilks, S. Richardson and D. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice*, Interdisciplinary Statistics, Chapman and Hall, London, pp. 21–43.
- Spielman, D. A. and Teng, S.-H. (1996). Spectral partitioning works: Planar graphs and finite element meshes, *IEEE Symposium on Foundations of Computer Science*, pp. 96–105.
- Stamey, T., Kabalin, J., McNeal, J., Johnstone, I., Freiha, F., Redwine, E. and Yang, N. (1989). Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate II radical prostatectomy treated patients, *Journal of Urology* **16**: 1076–1083.
- Stone, C., Hansen, M., Kooperberg, C. and Truong, Y. (1997). Polynomial splines and their tensor products (with discussion), *Annals of Statistics* **25**(4): 1371–1470.
- Stone, M. (1974). Cross-validated choice and assessment of statistical predictions, *Journal of the Royal Statistical Society Series B* **36**: 111–147.
- Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, *Journal of the Royal Statistical Society Series B*. **39**: 44–7.
- Stone, M. and Brooks, R. J. (1990). Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression (Corr: V54 p906-907), *Journal of the Royal Statistical Society, Series B* **52**: 237–269.
- Storey, J. (2002). A direct approach to false discovery rates, *Journal of the Royal Statistical Society B*. **64**(3): 479–498.
- Storey, J. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value, *Annals of Statistics* **31**: 2013–2025.
- Storey, J. and Tibshirani, R. (2003). Statistical significance for genomewide studies, *Proceedings of the National Academy of Sciences* **100**–: 9440–9445.
- Storey, J., Taylor, J. and Siegmund, D. (2004). Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach., *Journal of the Royal Statistical Society, Series B* **66**: 187–205.

- Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economics, Societies and Nations.*, Little, Brown.
- Swayne, D., Cook, D. and Buja, A. (1991). Xgobi: Interactive dynamic graphics in the X window system with a link to S, *ASA Proceedings of Section on Statistical Graphics*, pp. 1–8.
- Tanner, M. and Wong, W. (1987). The calculation of posterior distributions by data augmentation (with discussion), *Journal of the American Statistical Association* **82**: 528–550.
- Tarpey, T. and Flury, B. (1996). Self-consistency: A fundamental concept in statistics, *Statistical Science* **11**: 229–243.
- Tenenbaum, J. B., de Silva, V. and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction, *Science* **290**: 2319–2323.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B* **58**: 267–288.
- Tibshirani, R. and Hastie, T. (2007). Margin trees for high-dimensional classification, *Journal of Machine Learning Research* **8**: 637–652.
- Tibshirani, R. and Knight, K. (1999). Model search and inference by bootstrap “bumping”, *Journal of Computational and Graphical Statistics* **8**: 671–686.
- Tibshirani, R. and Wang, P. (2007). Spatial smoothing and hot spot detection for CGH data using the fused lasso, *Biostatistics* **9**: 18–29.
- Tibshirani, R., Hastie, T., Narasimhan, B. and Chu, G. (2001a). Diagnosis of multiple cancer types by shrunken centroids of gene expression, *Proceedings of the National Academy of Sciences* **99**: 6567–6572.
- Tibshirani, R., Hastie, T., Narasimhan, B. and Chu, G. (2003). Class prediction by nearest shrunken centroids, with applications to DNA microarrays, *Statistical Science* **18**(1): 104–117.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso, *Journal of the Royal Statistical Society, Series B* **67**: 91–108.
- Tibshirani, R., Walther, G. and Hastie, T. (2001b). Estimating the number of clusters in a dataset via the gap statistic, *Journal of the Royal Statistical Society, Series B.* **32**(2): 411–423.
- Tropp, J. (2004). Greed is good: algorithmic results for sparse approximation, *IEEE Transactions on Information Theory* **50**: 2231– 2242.

- Tropp, J. (2006). Just relax: convex programming methods for identifying sparse signals in noise, *IEEE Transactions on Information Theory* **52**: 1030–1051.
- Valiant, L. G. (1984). A theory of the learnable, *Communications of the ACM* **27**: 1134–1142.
- van der Merwe, A. and Zidek, J. (1980). Multivariate regression analysis and canonical variates, *The Canadian Journal of Statistics* **8**: 27–39.
- Vapnik, V. (1996). *The Nature of Statistical Learning Theory*, Springer, New York.
- Vapnik, V. (1998). *Statistical Learning Theory*, Wiley, New York.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, Wiley, New York.
- von Luxburg, U. (2007). A tutorial on spectral clustering, *Statistics and Computing* **17**(4): 395–416.
- Wahba, G. (1980). Spline bases, regularization, and generalized cross-validation for solving approximation problems with large quantities of noisy data, *Proceedings of the International Conference on Approximation theory in Honour of George Lorenz*, Academic Press, Austin, Texas, pp. 905–912.
- Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia.
- Wahba, G., Lin, Y. and Zhang, H. (2000). GACV for support vector machines, in A. Smola, P. Bartlett, B. Schölkopf and D. Schuurmans (eds), *Advances in Large Margin Classifiers*, MIT Press, Cambridge, MA., pp. 297–311.
- Wainwright, M. (2006). Sharp thresholds for noisy and high-dimensional recovery of sparsity using ℓ_1 -constrained quadratic programming, *Technical report*, Department of Statistics, University of California, Berkeley.
- Wainwright, M. J., Ravikumar, P. and Lafferty, J. D. (2007). High-dimensional graphical model selection using ℓ_1 -regularized logistic regression, in B. Schölkopf, J. Platt and T. Hoffman (eds), *Advances in Neural Information Processing Systems 19*, MIT Press, Cambridge, MA, pp. 1465–1472.
- Wasserman, L. (2004). *All of Statistics: a Concise Course in Statistical Inference*, Springer, New York.
- Weisberg, S. (1980). *Applied Linear Regression*, Wiley, New York.

- Werbos, P. (1974). *Beyond Regression*, PhD thesis, Harvard University.
- Weston, J. and Watkins, C. (1999). Multiclass support vector machines, in M. Verleysen (ed.), *Proceedings of ESANN99*, D. Facto Press, Brussels.
- Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, Wiley, Chichester.
- Wickerhauser, M. (1994). *Adapted Wavelet Analysis from Theory to Software*, A.K. Peters Ltd, Natick, MA.
- Widrow, B. and Hoff, M. (1960). Adaptive switching circuits, *IRE WESCON Convention record*, Vol. 4. pp 96-104; Reprinted in Andersen and Rosenfeld (1988).
- Wold, H. (1975). Soft modelling by latent variables: the nonlinear iterative partial least squares (NIPALS) approach, *Perspectives in Probability and Statistics, In Honor of M. S. Bartlett*, pp. 117–144.
- Wolpert, D. (1992). Stacked generalization, *Neural Networks* **5**: 241–259.
- Wu, T. and Lange, K. (2007). The MM alternative to EM, unpublished.
- Wu, T. and Lange, K. (2008). Coordinate descent procedures for lasso penalized regression, *Annals of Applied Statistics* **2**(1): 224–244.
- Yee, T. and Wild, C. (1996). Vector generalized additive models, *Journal of the Royal Statistical Society, Series B.* **58**: 481–493.
- Yuan, M. and Lin, Y. (2007). Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society, Series B* **68**(1): 49–67.
- Zhang, P. (1993). Model selection via multifold cross-validation, *Annals of Statistics* **21**: 299–311.
- Zhang, T. and Yu, B. (2005). Boosting with early stopping: convergence and consistency, *Annals of Statistics* **33**: 1538–1579.
- Zhao, P. and Yu, B. (2006). On model selection consistency of lasso, *Journal of Machine Learning Research* **7**: 2541–2563.
- Zhao, P., Rocha, G. and Yu, B. (2008). The composite absolute penalties for grouped and hierarchical variable selection, *Annals of Statistics*. (to appear).
- Zhu, J. and Hastie, T. (2004). Classification of gene microarrays by penalized logistic regression, *Biostatistics* **5**(2): 427–443.
- Zhu, J., Zou, H., Rosset, S. and Hastie, T. (2005). Multiclass adaboost, Unpublished.

- Zou, H. (2006). The adaptive lasso and its oracle properties, *Journal of the American Statistical Association* **101**: 1418–1429.
- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society Series B.* **67**(2): 301–320.
- Zou, H., Hastie, T. and Tibshirani, R. (2006). Sparse principal component analysis, *Journal of Computational and Graphical Statistics* **15**(2): 265–28.
- Zou, H., Hastie, T. and Tibshirani, R. (2007). On the degrees of freedom of the lasso, *Annals of Statistics* **35**(5): 2173–2192.

Author Index

- Abu-Mostafa, Y. 95, 474
Ackley, D. H. 645
Adam, B.-L. 664
Agrawal, R. 489–491, 578
Agresti, A. 385, 638, 640
Ahn, J. 695
Akaike, H. 257
Allen, D. 257
Ambroise, C. 247
Amit, Y. 602
Anderson, J. R. 641
Anderson, T. 645
Angelo, M. 654, 658
Ardekani, A. M. 664
- Bach, F. 569
Baezner, H. 551
Bair, E. 676, 679–683, 693
Bakin, S. 90
Bakiri, G. 605, 606
Banerjee, O. 636
Barnhill, S. 658
Barron, A. 415
Barry, R. 371
Bartlett, P. 384, 615
- Baskett, F. 480
Baxter, J. 384
Becker, R. 369
Bell, A. 578
Bellman, R. E. 22
Benade, A. 122
Bengio, Y. 404, 407, 408, 414, 644
Benjamini, Y. 687, 689, 693
Bentley, J. 480
Best, N. 292
Bibby, J. 94, 135, 441, 539, 559,
578, 630, 679
Bickel, P. 652
Bickel, P. J. 89
Bishop, C. 38, 233, 414, 623, 645
Bishop, Y. 629, 638
Bloomfield, C. 663
Boser, B. 404, 414
Botha, J. 334
Bottou, L. 404, 407, 408, 414, 644
Boyd, S. 125, 632
Breiman, L. 85, 243, 251, 257, 292,
308, 310, 334, 339, 367,
384, 451, 453, 455, 554,
587, 602

- Bremaud, P. 577
 Brin, S. 577
 Brooks, R. J. 81
 Brown, P. 679
 Bruce, A. 181
 Bühlmann, P. 87, 361, 384
 Buja, A. 110, 297, 441, 446, 451, 455, 565, 574, 576, 578
 Bunea, F. 91
 Burges, C. 455
 Butte, A. 631

 Caligiuri, M. 663
 Callow, M. 686, 693
 Campo, E. 674
 Candes, E. 86, 89, 613
 Carlin, J. 292
 Cazares, L. H. 664
 Chambers, J. 334
 Chan, W. C. 674
 Chaudhuri, S. 631, 633
 Chen, L. 574, 576, 578
 Chen, S. S. 68, 94
 Cherkassky, V. 38, 239, 257
 Chu, G. 684, 693
 Chui, C. 181
 Clark, L. 331
 Clark, M. 539
 Clements, M. A. 664
 Cleveland, W. 369
 Clifford, P. 629
 Cohen, A. 668, 669
 Coifman, R. R. 679
 Coller, H. 663
 Comon, P. 578
 Connors, J. M. 674
 Cook, D. 565, 578
 Cook, N. 317
 Copas, J. B. 94, 610
 Cover, T. 257, 465, 481
 Cox, D. 292, 645
 Cressie, N. 171
 Csiszar, I. 292
 Cun, Y. L. 407, 471, 481
 Cutler, A. 554

 Dale, M. 526
 Dasarathy, B. 480, 481
 d'Aspremont, A. 636
 Daubechies, I. 92, 181
 Davis, J. W. 664
 de Boer, C. 181
 De Mol, C. 92
 de Silva, V. 573
 Defrise, M. 92
 Dempster, A. 292, 449, 633
 Denham, M. 679
 Denker, J. 404, 407, 414, 471, 481
 Devijver, P. 480
 Dietterich, T. 286, 602, 605, 606, 623
 Donath, W. E. 578
 Donoho, D. 68, 86, 91, 94, 179, 181, 554, 613
 Downing, J. 663
 Drton, M. 631, 633
 du Plessis, J. 122
 Duan, N. 480
 Dubes, R. 508, 522
 Duchamp, T. 541
 Duda, R. 38, 135
 Dudoit, S. 686, 693
 Duin, R. 624

 Edwards, D. 645
 Efron, B. 73, 86, 90, 94, 97, 98, 128, 231, 254, 257, 292, 334, 568, 609, 692, 693
 Elad, M. 613
 Elith, J. 375, 376, 378
 Eskin, E. 668, 669
 Evgeniou, T. 168, 181, 455

 Fan, J. 92, 216, 539, 654
 Fan, Y. 654
 Fazekas, F. 551
 Feng, Z. 664
 Ferreira, J. 122
 Ferro, J. 551
 Fiedler, M. 578
 Fienberg, S. 585, 629, 638

- Finkel, R. 480
 Fisher, N. 334
 Fisher, R. A. 136, 455
 Fisher, R. I. 674
 Fisher, W. 310
 Fishman, D. A. 664
 Fix, E. 481
 Flury, B. 578
 Forgy, E. 578
 Francis, M. 375, 376, 378
 Frank, I. 81, 82, 94
 Frean, M. 384
 Freiha, F. 3, 49
 Freund, Y. 337, 383, 384, 615
 Fridlyand, J. 693
 Friedman, J. 38, 81, 82, 85, 92–94,
 111, 121, 126, 251, 257,
 258, 308, 310, 334, 339,
 345, 365, 367, 384, 391,
 414, 437, 451, 453, 475,
 480, 565, 578, 602, 611,
 617–621, 623, 636, 657,
 661, 667
 Friedman, N. 629, 630, 645
 Fu, W. 91, 92
 Fukunaga, K. 475
 Furnival, G. 57
 Fusaro, V. 664

 Gaasenbeek, M. 663
 Ganapathi, V. 642
 Gao, H. 181
 Gascoyne, R. D. 674
 Gelfand, A. 292
 Gelman, A. 292
 Geman, D. 292, 602
 Geman, S. 292
 Genkin, A. 661
 Genovese, C. 693
 Gerald, W. 654, 658
 Gershon, A. 514, 515, 526, 578
 Ghaoui, L. E. 636
 Gijbels, I. 216
 Gilks, W. 292
 Gill, P. 96, 421

 Girosi, F. 168, 174, 181, 415
 Golub, G. 257, 335, 535
 Golub, T. 631, 654, 658, 663
 Goodall, C. 578
 Gordon, A. 578
 Gray, R. 514, 515, 526, 578
 Green, P. 181, 183, 334
 Greenacre, M. 455
 Greenshtein, E. 91
 Guo, Y. 657
 Guyon, I. 658

 Haffner, P. 404, 407, 408, 414, 644
 Hall, P. 292, 602, 619
 Hammersley, J. M. 629
 Hand, D. 135, 475
 Hanley, J. 317
 Hansen, M. 328
 Hansen, R. 93
 Hart, P. 38, 135, 465, 480, 481
 Hartigan, J. A. 510, 578
 Hastie, T. 72, 73, 78, 86, 88, 90,
 92–94, 97, 98, 110, 121,
 122, 126, 137, 174, 216,
 257, 297, 299, 304, 334,
 339, 345, 348, 349, 375,
 376, 378, 384, 385, 414,
 428, 431, 434, 437, 441,
 446, 451, 455, 475, 478,
 480, 481, 519, 539, 550,
 565, 568, 578, 606, 609–
 611, 614, 615, 636, 657,
 658, 660–662, 664, 667,
 676, 679–683, 693
 Hatef, M. 624
 Hathaway, R. J. 292
 Heath, M. 257
 Hebb, D. 414
 Henderson, D. 404, 414
 Herman, A. 334
 Hertz, J. 414
 Hinkley, D. 292
 Hinton, G. 292, 334, 408, 414, 644,
 645
 Hitt, B. A. 664

- Ho, T. K. 602
 Hochberg, Y. 687, 689, 693
 Hodges, J. 481
 Hoefling, H. 92, 93, 642, 667
 Hoerl, A. E. 64, 94
 Hoff, M. 396, 414
 Hoffman, A. J. 578
 Hofmann, H. 578
 Holland, P. 629, 638
 Hong, W. 684
 Hothorn, T. 87, 361, 384
 Howard, R. 404, 414
 Huard, C. 663
 Hubbard, W. 404, 414
 Huber, P. 349, 414, 435, 565, 578
 Hunter, D. 294
 Hyvärinen, A. 560, 562, 578, 583
- Ihaka, R. 455
 Inskip, H. 292
 Inzitari, D. 551
 Izenman, A. 84
- Jackel, L. 404, 414
 Jacobs, R. 334
 Jain, A. 508, 522
 James, G. 606
 Jancey, R. 578
 Jensen, F. V. 629
 Jiang, W. 384
 Jiroušek, R. 640
 Johnson, N. 412
 Johnstone, I. 3, 49, 73, 86, 94, 97,
 98, 179, 181, 609, 613
 Joliffe, I. T. 550
 Jones, L. 415
 Jones, M. 168, 174, 181, 415
 Jooste, P. 122
 Jordaan, P. 122
 Jordan, M. 334, 569, 645
- Kabalin, J. 3, 49
 Kalbfleisch, J. 674, 693
 Karhunen, J. 583
 Kaski, S. 531, 532, 578
- Kaufman, L. 517, 526, 578
 Kearns, M. 380
 Kennard, R. 64, 94
 Kent, J. 94, 135, 441, 539, 559,
 578, 630, 679
 Kiiveri, H. T. 632
 Kim, S.-J. 125
 Kishon, E. 539
 Kittler, J. 480, 624
 Kleinberg, E. M. 602
 Knight, K. 91, 292, 666, 693
 Koh, K. 125
 Kohane, I. 631
 Kohavi, R. 243, 257
 Kohn, E. 664
 Kohonen, T. 462, 481, 531, 532,
 578
 Koller, D. 629, 630, 642, 645
 Kooperberg, C. 328
 Korn, E. L. 693
 Kotze, J. 122
 Kressel, U. 437
 Krogh, A. 414
- Ladd, C. 654, 658
 Lafferty, J. 90, 304
 Lafferty, J. D. 642
 Lagus, K. 531, 532, 578
 Laird, N. 292, 449
 Lambert, D. 376
 Lander, E. 654, 658, 663
 Lange, K. 92, 294, 583, 584
 Langford, J. C. 573
 Larsen, R. 551
 Latulippe, E. 654, 658
 Lauffenburger, D. 625
 Lauritzen, S. 629, 632, 645
 Lawson, C. 93
 Le Cun, Y. 404, 406–408, 414
 Leathwick, J. 375, 376, 378
 Leblanc, M. 292
 LeCun, Y. 644
 Lee, D. 552, 553
 Lee, M.-L. 693
 Lee, S.-I. 642

- Lee, W. 384, 615
 Leslie, C. 668, 669
 Levina, E. 652, 693
 Levine, P. J. 664
 Lewis, D. 661
 Li, K.-C. 480
 Li, R. 92
 Lin, H. 331
 Lin, Y. 90, 304, 428, 455
 Liotta, L. A. 664
 Little, R. 332, 647
 Littman, M. 578
 Liu, H. 90, 304
 Lloyd, S. 481, 578
 Loader, C. 209, 216
 Loda, M. 654, 658
 Loh, M. 663
 Loh, W. 310
 Lugosi, G. 384
- Ma, Y. 257
 Macnaughton Smith, P. 526
 MacKay, D. 623
 MacQueen, J. 481, 578
 Madigan, D. 257, 292, 661
 Makeig, S. 564, 565
 Mannila, H. 489–491, 578
 Mardia, K. 94, 135, 441, 539, 559,
 578, 630, 679
 Marron, J. 695
 Mason, L. 384
 Massart, D. 517
 Matas, J. 624
 McCullagh, P. 638, 640
 McCulloch, C. 331
 McCulloch, W. 414
 McLachlan, G. 135, 247
 McNeal, J. 3, 49
 McNeil, B. 317
 McShane, L. M. 693
 Mease, D. 384, 603
 Meinshausen, N. 91, 635, 642
 Meir, R. 384
 Mesirov, J. 654, 658, 663
 Mills, G. B. 664
- Mockett, L. 526
 Morgan, J. N. 334
 Motwani, R. 577
 Mukherjee, S. 654, 658
 Mulier, F. 38, 239
 Muller-Hermelink, H. K. 674
 Müller, K.-R. 547, 548
 Munro, S. 397
 Murray, W. 96, 421
 Myles, J. 475
- Nadler, B. 679
 Narasimhan, B. 693
 Neal, R. 268, 292, 409–412, 414,
 605, 623
 Nelder, J. 638, 640
 Noble, W. S. 668, 669
 Nolan, G. 625
 Nowlan, S. 334
- Oja, E. 560, 562, 578, 583
 Olesen, K. G. 629
 Olshen, R. 251, 308, 310, 334, 367,
 451, 453
 Onton, J. 564, 565
 Osborne, M. 76, 94
 Osindero, S. 644
- Paatero, A. 531, 532, 578
 Pace, R. K. 371
 Page, L. 577
 Palmer, R. 414
 Pantoni, L. 551
 Park, M. Y. 94, 126, 661
 Parker, D. 414
 Paul, D. 676, 679–683, 693
 Pearl, J. 629, 645
 Pe'er, D. 625
 Perez, O. 625
 Peterson 641
 Petricoin, E. F. 664
 Pitts, W. 414
 Plastria, F. 517
 Platt, J. 453
 Poggio, T. 168, 174, 181, 415, 455,
 654, 658

- Pontil, M. 168, 181, 455
 Popescu, B. 617–619, 621, 623
 Prentice, R. 674, 693
 Presnell, B. 76, 94
 Přeučil, S. 640
- Qu, Y. 664
 Quinlan, R. 312, 334, 624
- Radmacher, M. D. 693
 Raftery, A. 257, 292
 Ramaswamy, S. 654, 658
 Ramsay, J. 181, 578
 Rao, C. R. 455
 Rättsch, G. 384, 615
 Ravikumar, P. 90, 304, 642
 Redwine, E. 3, 49
 Reich, M. 654, 658
 Richardson, J. 375
 Richardson, T. S. 631, 633
 Ridgeway, G. 361
 Rieger, K. 684
 Rifkin, R. 654, 658
 Ripley, B. D. 38, 131, 135, 136,
 234, 308, 310, 400, 414,
 415, 455, 468, 480, 481,
 641, 645
 Rissanen, J. 257
 Ritov, Y. 89, 91
 Robbins, H. 397
 Rocha, G. 90
 Roosen, C. 414
 Rosenblatt, F. 102, 129, 414
 Rosenwald, A. 674
 Rosset, S. 89, 98, 348, 349, 385,
 426, 428, 434, 610, 611,
 615, 657, 661, 664, 666,
 693
 Rostrup, E. 551
 Rousseauw, J. 122
 Rousseuw, P. 517, 526, 578
 Rowe, D. 375
 Roweis, S. T. 573
 Rubin, D. 292, 332, 449, 647
 Rumelhart, D. 414
- Ryberg, C. 551
- Saarela, A. 531, 532, 578
 Sachs, K. 625
 Salojärvi, J. 531, 532, 578
 Saul, L. K. 573
 Saunders, M. 68, 94, 666, 693
 Schapire, R. 337, 380, 383, 384,
 615
 Schellhammer, P. F. 664
 Schnitzler, C. 334
 Schölkopf, B. 547, 548
 Schroeder, A. 391
 Schwarz, G. 233, 257
 Scott, D. 216
 Seber, G. 94
 Segal, M. 596
 Sejnowski, T. 578, 645
 Semmes, O. J. 664
 Seung, H. 552, 553
 Shafer, G. 629
 Shao, J. 257
 Shenoy, P. 629
 Short, R. 475
 Shustek, L. 480
 Shyu, M. 369
 Siegmund, D. 689
 Silverman, B. 181, 183, 216, 334,
 486, 567, 578
 Silvey, S. 292
 Simard, P. 407, 471, 480, 481
 Simon, R. M. 693
 Simone, C. 664
 Singer, Y. 384
 Sjöstrand, K. 551
 Slate, E. 331
 Slonim, D. 631, 663
 Smeland, E. B. 674
 Smith, A. 292
 Smola, A. 547, 548
 Sonquist, J. A. 334
 Spector, P. 243, 257
 Speed, T. 632, 686, 693
 Spiegelhalter, D. 292, 629
 Spiegelman, C. 679

- Spielman, D. A. 578
 Srikant, R. 489–491, 578
 Stamey, T. 3, 49
 Staudt, L. M. 674
 Steinberg, S. M. 664
 Stern, H. 292
 Stodden, V. 554
 Stone, C. 251, 308, 310, 328, 334,
 367, 451, 453
 Stone, M. 81, 257
 Storey, J. 689, 692, 693, 697, 698
 Stork, D. 38, 135
 Studholme, C. 551
 Stuetzle, W. 391, 414, 541, 578
 Surowiecki, J. 286
 Swayne, D. 565, 578

 Tamayo, P. 631, 654, 658, 663
 Tang, J. 684
 Tanner, M. 292
 Tao, T. 89, 613
 Tarpey, T. 578
 Taylor, J. 88, 94, 610, 614, 689
 Taylor, P. 375, 376, 378
 Teh, Y.-W. 644
 Tenenbaum, J. B. 573
 Teng, S.-H. 578
 Thomas, J. 257
 Tibshirani, R. 73, 78, 86, 88, 90,
 92–94, 97, 98, 110, 121,
 122, 126, 137, 216, 257,
 292, 297, 299, 304, 334,
 339, 345, 384, 428, 431,
 434, 437, 441, 446, 451,
 455, 475, 478, 480, 481,
 519, 550, 565, 568, 609–
 611, 614, 636, 642, 657,
 658, 660, 661, 666, 667,
 676, 679–684, 692, 693
 Toivonen, H. 489–491, 578
 Traskin, M. 384
 Trendafilov, N. T. 550
 Tropp, J. 91
 Truong, Y. 328
 Tsybakov, A. 89, 91

 Tukey, J. 414, 565, 578
 Turlach, B. 76, 94
 Turnbull, B. 331
 Tusher, V. 684, 692
 Tusnády, G. 292

 Uddin, M. 550

 Valiant, L. G. 380
 van der Merwe, A. 84
 Van Loan, C. 335, 535
 Vandenberghe, L. 632
 Vanichsetakul, N. 310
 Vapnik, V. 38, 102, 132, 135, 171,
 257, 438, 455, 658
 Vayatis, N. 384
 Vazirani, U. 380
 Verkamo, A. I. 489–491, 578
 Vidakovic, B. 181
 von Luxburg, U. 578

 Wahba, G. 168, 169, 181, 257, 268,
 428, 429, 455
 Wainwright, M. 91
 Wainwright, M. J. 642
 Waldemar, G. 551
 Walther, G. 88, 94, 519, 610, 614
 Wang, P. 667
 Ward, M. D. 664
 Warmuth, M. 615
 Wasserman, L. 90, 304, 626, 645,
 693
 Watkins, C. 658
 Wegkamp, M. 91
 Weisberg, S. 94
 Werbos, P. 414
 Wermuth, N. 645
 Weston, J. 658, 668, 669
 Whittaker, J. 632, 633, 641, 645
 Wickerhauser, M. 181
 Widrow, B. 396, 414
 Wild, C. 300
 Williams, R. 414
 Williams, W. 526
 Wilson, R. 57

- Winograd, T. 577
Wold, H. 94
Wolpert, D. 292
Wong, M. A. 510
Wong, W. 292
Wright, G. 664, 674, 693
Wright, M. 96, 421
Wu, T. 92, 294, 583
Wyner, A. 384, 603
- Yang, N. 3, 49
Yang, Y. 686, 693
Yasui, Y. 664
Yeang, C. 654, 658
Yee, T. 300
Yekutieli, Y. 693
Yu, B. 90, 91, 384
Yuan, M. 90
- Zhang, H. 90, 304, 428, 455
Zhang, J. 409–412, 605
Zhang, P. 257
Zhang, T. 384
Zhao, P. 90, 91
Zhao, Y. 693
Zhu, J. 89, 98, 174, 348, 349, 385,
426, 428, 434, 610, 611,
615, 657, 661, 664, 666,
693
Zidek, J. 84
Zou, H. 72, 78, 92, 349, 385, 550,
662, 693

Index

- L_1 regularization, *see* Lasso
- Activation function, 392–395
- AdaBoost, 337–346
- Adaptive lasso, 92
- Adaptive methods, 429
- Adaptive nearest neighbor methods, 475–478
- Adaptive wavelet filtering, 181
- Additive model, 295–304
- Adjusted response, 297
- Affine set, 130
- Affine-invariant average, 482, 540
- AIC, *see* Akaike information criterion
- Akaike information criterion (AIC), 230
- Analysis of deviance, 124
- Applications
 - abstracts, 672
 - aorta, 204
 - bone, 152
 - California housing, 371–372, 591
 - countries, 517
 - demographics, 379–380
 - document, 532
 - flow cytometry, 637
 - galaxy, 201
 - heart attack, 122, 146, 207
 - lymphoma, 674
 - marketing, 488
 - microarray, 5, 505, 532
 - nested spheres, 590
 - New Zealand fish, 375–379
 - nuclear magnetic resonance, 176
 - ozone, 201
 - prostate cancer, 3, 49, 61, 608
 - protein mass spectrometry, 664
 - satellite image, 470
 - skin of the orange, 429–432
 - spam, 2, 300–304, 313, 320, 328, 352, 593
 - vowel, 440, 464
 - waveform, 451
 - ZIP code, 4, 404, 536–539
- Archetypal analysis, 554–557
- Association rules, 492–495, 499–501

- Automatic relevance determination, 411
- Automatic selection of smoothing parameters , 156
- B-Spline, 186
- Back-propagation, 392–397, 408–409
- Backfitting, 297, 391
- Backward
 - selection, 58
 - stepwise selection, 59
- Backward pass, 396
- Bagging, 282–288, 409, 587
- Basis expansions and regularization, 139–189
- Basis functions, 141, 186, 189, 321, 328
- Batch learning, 397
- Baum–Welch algorithm, 272
- Bayes
 - classifier, 21
 - factor, 234
 - methods, 233–235, 267–272
 - rate, 21
- Bayesian, 409
- Bayesian information criterion (BIC), 233
- Benjamini–Hochberg method, 688
- Best-subset selection, 57, 610
- Between class covariance matrix, 114
- Bias, 16, 24, 37, 160, 219
- Bias-variance decomposition, 24, 37, 219
- Bias-variance tradeoff, 37, 219
- BIC, *see* Bayesian Information Criterion
- Boltzmann machines, 638–648
- Bonferroni method, 686
- Boosting, 337–386, 409
 - as lasso regression, 607–609
 - exponential loss and AdaBoost, 343
 - gradient boosting, 358
 - implementations, 360
 - margin maximization, 613
 - numerical optimization, 358
 - partial-dependence plots, 369
 - regularization path, 607
 - shrinkage, 364
 - stochastic gradient boosting, 365
 - tree size, 361
 - variable importance, 367
- Bootstrap, 249, 261–264, 267, 271–282, 587
 - relationship to Bayesian method, 271
 - relationship to maximum likelihood method, 267
- Bottom-up clustering, 520–528
- Bump hunting, *see* Patient rule induction method
- Bumping, 290–292
- C5.0, 624
- Canonical variates, 441
- CART, *see* Classification and regression trees
- Categorical predictors, 10, 310
- Censored data, 674
- Classical multidimensional scaling, 570
- Classification, 22, 101–137, 305–317, 417–429
- Classification and regression trees (CART), 305–317
- Clique, 628
- Clustering, 501–528
 - k*-means, 509–510
 - agglomerative, 523–528
 - hierarchical, 520–528
- Codebook, 515
- Combinatorial algorithms, 507
- Combining models, 288–290
- Committee, 289, 587, 605
- Comparison of learning methods, 350–352
- Complete data, 276

- Complexity parameter, 37
- Computational shortcuts
 - quadratic penalty, 659
- Condensing procedure, 480
- Conditional likelihood, 31
- Confusion matrix, 301
- Conjugate gradients, 396
- Consensus, 285–286
- Convolutional networks, 407
- Coordinate descent, 92, 636, 668
- COSSO, 304
- Cost complexity pruning, 308
- Covariance graph, 631
- C_p statistic, 230
- Cross-entropy, 308–310
- Cross-validation, 241–245
- Cubic smoothing spline, 151–153
- Cubic spline, 151–153
- Curse of dimensionality, 22–26

- Dantzig selector, 89
- Data augmentation, 276
- Daubechies symmlet-8 wavelets, 176
- De-correlation, 597
- Decision boundary, 13–15, 21
- Decision trees, 305–317
- Decoder, 515, *see* encoder
- Decomposable models, 641
- Degrees of freedom
 - in an additive model, 302
 - in ridge regression, 68
 - of a tree, 336
 - of smoother matrices, 153–154, 158
- Delta rule, 397
- Demmler-Reinsch basis for splines, 156
- Density estimation, 208–215
- Deviance, 124, 309
- Diagonal linear discriminant analysis, 651–654
- Dimension reduction, 658
 - for nearest neighbors, 479
- Discrete variables, 10, 310–311

- Discriminant
 - adaptive nearest neighbor classifier, 475–480
 - analysis, 106–119
 - coordinates, 108
 - functions, 109–110
- Dissimilarity measure, 503–504
- Dummy variables, 10

- Early stopping, 398
- Effective degrees of freedom, 17, 68, 153–154, 158, 232, 302, 336
- Effective number of parameters, 15, 68, 153–154, 158, 232, 302, 336
- Eigenvalues of a smoother matrix, 154
- Elastic net, 662
- EM algorithm, 272–279
 - as a maximization-maximization procedure, 277
 - for two component Gaussian mixture, 272
- Encoder, 514–515
- Ensemble, 616–623
- Ensemble learning, 605–624
- Entropy, 309
- Equivalent kernel, 156
- Error rate, 219–230
- Error-correcting codes, 606
- Estimates of in-sample prediction error, 230
- Expectation-maximization algorithm, *see* EM algorithm
- Extra-sample error, 228

- False discovery rate, 687–690, 692, 693
- Feature, 1
 - extraction, 150
 - selection, 409, 658, 681–683
- Feed-forward neural networks, 392–408

- Fisher's linear discriminant, 106–119, 438
- Flexible discriminant analysis, 440–445
- Forward
 - selection, 58
 - stagewise, 86, 608
 - stagewise additive modeling, 342
 - stepwise, 73
- Forward pass algorithm, 395
- Fourier transform, 168
- Frequentist methods, 267
- Function approximation, 28–36
- Fused lasso, 666

- Gap statistic, 519
- Gating networks, 329
- Gauss-Markov theorem, 51–52
- Gauss-Newton method, 391
- Gaussian (normal) distribution, 16
- Gaussian graphical model, 630
- Gaussian mixtures, 273, 463, 492, 509
- Gaussian radial basis functions, 212
- GBM, *see* Gradient boosting
- GBM package, *see* Gradient boosting
- GCV, *see* Generalized cross-validation
- GEM (generalized EM), 277
- Generalization
 - error, 220
 - performance, 220
- Generalized additive model, 295–304
- Generalized association rules, 497–499
- Generalized cross-validation, 244
- Generalized linear discriminant analysis, 438
- Generalized linear models, 125
- Gibbs sampler, 279–280, 641
 - for mixtures, 280
- Gini index, 309

- Global Markov property, 628
- Gradient Boosting, 359–361
- Gradient descent, 358, 395–397
- Graph Laplacian, 545
- Graphical lasso, 636
- Grouped lasso, 90

- Haar basis function, 176
- Hammersley-Clifford theorem, 629
- Hard-thresholding, 653
- Hat matrix, 46
- Helix, 582
- Hessian matrix, 121
- Hidden nodes, 641–642
- Hidden units, 393–394
- Hierarchical clustering, 520–528
- Hierarchical mixtures of experts, 329–332
- High-dimensional problems, 649
- Hints, 96
- Hyperplane, *see* Separating Hyperplane

- ICA, *see* Independent components analysis
- Importance sampling, 617
- In-sample prediction error, 230
- Incomplete data, 332
- Independent components analysis, 557–570
- Independent variables, 9
- Indicator response matrix, 103
- Inference, 261–294
- Information
 - Fisher, 266
 - observed, 274
- Information theory, 236, 561
- Inner product, 53, 668, 670
- Inputs, 10
- Instability of trees, 312
- Intercept, 11
- Invariance manifold, 471
- Invariant metric, 471
- Inverse wavelet transform, 179

- IRLS, *see* Iteratively reweighted least squares
- Irreducible error, 224
- Ising model, 638
- ISOMAP, 572
- Isometric feature mapping, 572
- Iterative proportional scaling, 585
- Iteratively reweighted least squares (IRLS), 121

- Jensen's inequality, 293
- Join tree, 629
- Junction tree, 629

- K-means clustering, 460, 509–514
- K-medoid clustering, 515–520
- K-nearest neighbor classifiers, 463
- Karhunen-Loeve transformation (principal components), 66–67, 79, 534–539
- Karush-Kuhn-Tucker conditions, 133, 420
- Kernel
 - classification, 670
 - density classification, 210
 - density estimation, 208–215
 - function, 209
 - logistic regression, 654
 - principal component, 547–550
 - string, 668–669
 - trick, 660
- Kernel methods, 167–176, 208–215, 423–438, 659
- Knot, 141, 322
- Kriging, 171
- Kruskal-Shephard scaling, 570
- Kullback-Leibler distance, 561

- Lagrange multipliers, 293
- Landmark, 539
- Laplacian, 545
- Laplacian distribution, 72
- LAR, *see* Least angle regression
- Lasso, 68–69, 86–90, 609, 635, 636, 661
 - fused, 666
- Latent
 - factor, 674
 - variable, 678
- Learning, 1
- Learning rate, 396
- Learning vector quantization, 462
- Least angle regression, 73–79, 86, 610
- Least squares, 11, 32
- Leave-one-out cross-validation, 243
- LeNet, 406
- Likelihood function, 265, 273
- Linear basis expansion, 139–148
- Linear combination splits, 312
- Linear discriminant function, 106–119
- Linear methods
 - for classification, 101–137
 - for regression, 43–99
- Linear models and least squares, 11
- Linear regression of an indicator matrix, 103
- Linear separability, 129
- Linear smoother, 153
- Link function, 296
- LLE, *see* Local linear embedding
- Local false discovery rate, 693
- Local likelihood, 205
- Local linear embedding, 572
- Local methods in high dimensions, 22–27
- Local minima, 400
- Local polynomial regression, 197
- Local regression, 194, 200
- Localization in time/frequency, 175
- Loess (local regression), 194, 200
- Log-linear model, 639
- Log-odds ratio (logit), 119
- Logistic (sigmoid) function, 393
- Logistic regression, 119–128, 299
- Logit (log-odds ratio), 119
- Loss function, 18, 21, 219–223, 346
- Loss matrix, 310

- Lossless compression, 515
- Lossy compression, 515
- LVQ, *see* Learning Vector Quantization
- Mahalanobis distance, 441
- Majority vote, 337
- Majorization, 294, 553
- Majorize-Minimize algorithm, 294, 584
- MAP (maximum a posteriori) estimate, 270
- Margin, 134, 418
- Market basket analysis, 488, 499
- Markov chain Monte Carlo (MCMC) methods, 279
- Markov graph, 627
- Markov networks, 638–648
- MARS, *see* Multivariate adaptive regression splines
- MART, *see* Multiple additive regression trees
- Maximum likelihood estimation, 31, 261, 265
- MCMC, *see* Markov Chain Monte Carlo Methods
- MDL, *see* Minimum description length
- Mean field approximation, 641
- Mean squared error, 24, 285
- Memory-based method, 463
- Metropolis-Hastings algorithm, 282
- Minimum description length (MDL), 235
- Minorization, 294, 553
- Minorize-Maximize algorithm, 294, 584
- Misclassification error, 17, 309
- Missing data, 276, 332–333
- Missing predictor values, 332–333
- Mixing proportions, 214
- Mixture discriminant analysis, 449–455
- Mixture modeling, 214–215, 272–275, 449–455, 692
- Mixture of experts, 329–332
- Mixtures and the EM algorithm, 272–275
- MM algorithm, 294, 584
- Mode seekers, 507
- Model averaging and stacking, 288
- Model combination, 289
- Model complexity, 221–222
- Model selection, 57, 222–223, 230–231
- Modified regression, 634
- Monte Carlo method, 250, 495
- Mother wavelet, 178
- Multidimensional scaling, 570–572
- Multidimensional splines, 162
- Multiedit algorithm, 480
- Multilayer perceptron, 400, 401
- Multinomial distribution, 120
- Multiple additive regression trees (MART), 361
- Multiple hypothesis testing, 683–693
- Multiple minima, 291, 400
- Multiple outcome shrinkage and selection, 84
- Multiple outputs, 56, 84, 103–106
- Multiple regression from simple univariate regression, 52
- Multiresolution analysis, 178
- Multivariate adaptive regression splines (MARS), 321–327
- Multivariate nonparametric regression, 445
- Nadaraya-Watson estimate, 193
- Naive Bayes classifier, 108, 210–211, 694
- Natural cubic splines, 144–146
- Nearest centroids, 670
- Nearest neighbor methods, 463–483
- Nearest shrunken centroids, 651–654, 694
- Network diagram, 392
- Neural networks, 389–416

- Newton's method (Newton-Raphson procedure), 120–122
- Non-negative matrix factorization, 553–554
- Nonparametric logistic regression, 299–304
- Normal (Gaussian) distribution, 16, 31
- Normal equations, 12
- Numerical optimization, 395–396

- Object dissimilarity, 505–507
- Online algorithm, 397
- Optimal scoring, 445, 450–451
- Optimal separating hyperplane, 132–135
- Optimism of the training error rate, 228–230
- Ordered categorical (ordinal) predictor, 10, 504
- Ordered features, 666
- Orthogonal predictors, 53
- Overfitting, 220, 228–230, 364

- PageRank, 576
- Pairwise distance, 668
- Pairwise Markov property, 628
- Parametric bootstrap, 264
- Partial dependence plots, 369–370
- Partial least squares, 80–82, 680
- Partition function, 638
- Parzen window, 208
- Pasting, 318
- Path algorithm, 73–79, 86–89, 432
- Patient rule induction method (PRIM), 317–321, 499–501
- Peeling, 318
- Penalization, 607, *see* regularization
- Penalized discriminant analysis, 446–449
- Penalized polynomial regression, 171
- Penalized regression, 34, 61–69, 171
- Penalty matrix, 152, 189

- Perceptron, 392–416
- Piecewise polynomials and splines, 36, 143
- Posterior
 - distribution, 268
 - probability, 233–235, 268
- Power method, 577
- Pre-conditioning, 681–683
- Prediction accuracy, 329
- Prediction error, 18
- Predictive distribution, 268
- PRIM, *see* Patient rule induction method
- Principal components, 66–67, 79–80, 534–539, 547
 - regression, 79–80
 - sparse, 550
 - supervised, 674
- Principal curves and surfaces, 541–544
- Principal points, 541
- Prior distribution, 268–272
- Procrustes
 - average, 540
 - distance, 539
- Projection pursuit, 389–392, 565
 - regression, 389–392
- Prototype classifier, 459–463
- Prototype methods, 459–463
- Proximity matrices, 503
- Pruning, 308

- QR decomposition, 55
- Quadratic approximations and inference, 124
- Quadratic discriminant function, 108, 110

- Radial basis function (RBF) network, 392
- Radial basis functions, 212–214, 275, 393
- Radial kernel, 548
- Random forest, 409, 587–604
 - algorithm, 588

- bias, 596–601
- comparison to boosting, 589
- example, 589
- out-of-bag (OOB), 592
- overfit, 596
- proximity plot, 595
- variable importance, 593
- variance, 597–601
- Rao score test, 125
- Rayleigh quotient, 116
- Receiver operating characteristic (ROC) curve, 317
- Reduced-rank linear discriminant analysis, 113
- Regression, 11–14, 43–99, 200–204
- Regression spline, 144
- Regularization, 34, 167–176
- Regularized discriminant analysis, 112–113, 654
- Relevance network, 631
- Representer of evaluation, 169
- Reproducing kernel Hilbert space, 167–176, 428–429
- Reproducing property, 169
- Responsibilities, 274–275
- Ridge regression, 61–68, 650, 659
- Risk factor, 122
- Robust fitting, 346–350
- Rosenblatt’s perceptron learning algorithm, 130
- Rug plot, 303
- Rulefit, 623
- SAM, 690–693, *see* Significance Analysis of Microarrays
- Sammon mapping, 571
- SCAD, 92
- Scaling of the inputs, 398
- Schwarz’s criterion, 230–235
- Score equations, 120, 265
- Self-consistency property, 541–543
- Self-organizing map (SOM), 528–534
- Sensitivity of a test, 314–317
- Separating hyperplane, 132–135
- Separating hyperplanes, 136, 417–419
- Separator, 628
- Shape average, 482, 540
- Shrinkage methods, 61–69, 652
- Sigmoid, 393
- Significance Analysis of Microarrays, 690–693
- Similarity measure, *see* Dissimilarity measure
- Single index model, 390
- Singular value decomposition, 64, 535–536, 659
 - singular values, 535
 - singular vectors, 535
- Sliced inverse regression, 480
- Smoother, 139–156, 192–199
 - matrix, 153
- Smoothing parameter, 37, 156–161, 198–199
- Smoothing spline, 151–156
- Soft clustering, 512
- Soft-thresholding, 653
- Softmax function, 393
- SOM, *see* Self-organizing map
- Sparse, 175, 304, 610–613, 636
 - additive model, 91
 - graph, 625, 635
- Specificity of a test, 314–317
- Spectral clustering, 544–547
- Spline, 186
 - additive, 297–299
 - cubic, 151–153
 - cubic smoothing, 151–153
 - interaction, 428
 - regression, 144
 - smoothing, 151–156
 - thin plate, 165
- Squared error loss, 18, 24, 37, 219
- SRM, *see* Structural risk minimization
- Stacking (stacked generalization), 290
- Starting values, 397
- Statistical decision theory, 18–22

- Statistical model, 28–29
- Steepest descent, 358, 395–397
- Stepwise selection, 60
- Stochastic approximation, 397
- Stochastic search (bumping), 290–292
- Stress function, 570–572
- Structural risk minimization (SRM), 239–241
- Subset selection, 57–60
- Supervised learning, 2
- Supervised principal components, 674–681
- Support vector classifier, 417–421, 654
 - multiclass, 657
- Support vector machine, 423–437
- SURE shrinkage method, 179
- Survival analysis, 674
- Survival curve, 674
- SVD, *see* Singular value decomposition
- Symmlet basis, 176
- Tangent distance, 471–475
- Tanh activation function, 424
- Target variables, 10
- Tensor product basis, 162
- Test error, 220–223
- Test set, 220
- Thin plate spline, 165
- Thinning strategy, 189
- Trace of a matrix, 153
- Training epoch, 397
- Training error, 220–223
- Training set, 219–223
- Tree for regression, 307–308
- Tree-based methods, 305–317
- Trees for classification, 308–310
- Trellis display, 202
- Undirected graph, 625–648
- Universal approximator, 390
- Unsupervised learning, 2, 485–585
- Unsupervised learning as supervised learning, 495–497
- Validation set, 222
- Vapnik-Chervonenkis (VC) dimension, 237–239
- Variable importance plot, 594
- Variable types and terminology, 9
- Variance, 16, 25, 37, 158–161, 219
 - between, 114
 - within, 114, 446
- Variance reduction, 588
- Varying coefficient models, 203–204
- VC dimension, *see* Vapnik-Chervonenkis dimension
- Vector quantization, 514–515
- Voronoi regions, 510
- Wald test, 125
- Wavelet
 - basis functions, 176–179
 - smoothing, 174
 - transform, 176–179
- Weak learner, 383, 605
- Weakest link pruning, 308
- Webpages, 576
- Website for book, 8
- Weight decay, 398
- Weight elimination, 398
- Weights in a neural network, 395
- Within class covariance matrix, 114, 446