References

Akaike, H. (1973) "Information Theory and an Extension to the Maximum Likelihood Principle." In B.N. Petrov and F. Casaki (Eds.), *International Symposim on Information Theory* Budapest: Akademia Kiado: 267–281.

Bartlett, P.L., Jordon, M.I., and J. D. McAuliffe (2006) "Comment." *Statistical Science* 21(3): 341–346.

Berk, R.A. (2003) *Regression Analysis: A Constructive Critique*. Newbury Park, CA.: Sage.

Berk, R.A. (2005a) "New Claims About Executions and General Deterrence: Déjà Vu All Over Again?." *Journal of Empirical Legal Studies* 2(2)L 303–330 .

Berk, R.A. (2005b) "Randomized Experiments as the Bronze Standard." *Journal of Experimental Criminology*, 1(4):417–433.

Berk, R.A. and S. Rothenberg (2004) "Water Resource Dynamics in Asian Pacific Cities." Statistics Department Preprint Series, UCLA.

Berk, R.A., He, Y., and S. Sorenson (2005b) "Developing a Practical Forecasting Screener for Domestic Violence Incidents." *Evaluation Review* 29(4): 358–382.

Berk, R.A., Kriegler, B., and D. Ylvisaker (2008) "Counting the Homeless in Los Angeles County." In D. Nolan and S. Speed (Eds.), *Probability and Statistics: Essays in Honor of David A. Freedman*, Monograph Series for the Institute of Mathematical Statistics, forthcoming.

Berk, R.A., Kriegler B. and J. Baek (2006) "Forecasting Dangerous Inmate Misconduct: An Application of Ensemble Statistical Procedures. *Journal of Quantitative Criminology* 22(2): 131-145. forthcoming.

Berk, R.A., Li, A. and L. Hickman (2005a) "Statistical Difficulties in Determining the Role of Race in Capital Cases: A Re-analysis of Data from the State of Maryland." *Journal of Quantitative Criminology*, 21(4): 365–390.

Berk, R.A., Ladd, H., Graziano, H., and J. Baek (2003) "A Randomized Experiment Testing Inmate Classification Systems." *Journal of Criminology and Public Policy*, 2, No. 2: 215–242.

Bickel, P.J., Ritov, Y., and A Zakai (2006) "Some Theory for Generalized Boosting Algorithms." *Journal of Machine Learning Research* 7: 705–769.

Bishop, C.M. (2006) *Pattern Recognition and Machine Learning*. New York: Springer.

Bousquet, O. and B. Scholköph (2006) "Comment." *Statistical Science* 21(3): 337–340.

Box, G.E.P. (1976) "Science and Statistics." *Journal of the American Statistical Association* 71: 791–799.

Breiman, L. (1996) "Bagging Predictors." *Machine Learning* 26:123–140.

Breiman, L. (1999) "Prediction Games and Arcing Algorithms." *Neural Computation* 11: 1493–1517.

Breiman, L. (2000) "Some Infinity Theory for Predictor Ensembles." *Technical Report 522*, Department of Statistics, University of California, Berkeley.

Breiman, L. (2001a) "Random Forests." *Machine Learning* 45: 5–32.

Breiman, L. (2001b) "Statistical Modeling: Two Cultures" (with discussion). *Statistical Science* 16: 199–231.

Breiman, L, (2004) "The 2002 Wald Memorial Lectures: Population Theory for Boosting Ensembles." *The Annals of Statistics* 32 (1): 1–11.

Breiman, L., Friedman, J.H., Olshen, R.A., and C.J. Stone, (1984) *Classification and Regression Trees*. Monterey, CA: Wadsworth Press.

Brown, L.D., Zhao, L., Mao, V., and J. Xu. (2005) "A Pseudo-Bayesian Procedure for Nonparametric Regression Estimation and Pointwise Confidence Bands." Invited presentation at Bernoulli/IMS meeting, Barcelona.

Bühlmann, P. (2006) "Boosting for High Dimensional Linear Models." *The Annals of Statistics* 34 (2): 559–583.

Bühlmann, P. and B. Yu (2002), "Analyzing Bagging." *The Annals of Statistics* 30: 927–961.

Bühlmann, P. and B. Yu (2004), "Discussion." *The Annals of Statistics* 32: 96–107.

Bühlmann, P. and B. Yu (2006) "Sparse Boosting." *Journal of Machine Learning Research* 7: 1001–1024.

Buja, A. and W. Rolke (2007) "Calibration for Simultaneity: (Re) Sampling Methods for Simultaneous Inference with Application to Function Estimation and Functional Data." Working paper at www-stat. wharton.upenn.edu/ buja/.

Buja, A. and W. Stuetzle (2000) "Smoothing Effects of Bagging." Working paper at www-stat. wharton. upenn.edu/ buja/.

Buja, A. and W. Stuetzle (2006) "Observations on Bagging." *Statistica Sinica* 16(2) 323–352.

Buja, A., Mease, D., and A.J. Wyner (2008) "Discussion of Bühlmann and Hothorn." *Statistical Science*, forthcoming.

Buja, A., Stuetzle, W. and Y. Shen (2005) "Loss Functions for Binary Class Probability Estimation and Classification: Structure and Applications."

Unpublished manuscript, Department of Statistics, The Wharton School, University of Pennsylvania.

Cai, T.T. and J. Lv (2007) "Discussion: The Dantzig Selector: Statistical Estimation When p is much larger than n." *Annals of Statistics* 35(6): 2365–2369.

Camacho, R., King, R., and A. Srinivasan (2006) "14th International Conference on Inductive Logic Programming." *Machine Learning* 64: 145–287.

Candes, E. and T. Tao. (2007) "The Dantzig Selector: Statistical Estimation When p is Much Larger than n" (with discussion). *Annals of Statistics* 35(6): 2313-2351.

Carlin, B.P. and T.A. Lewis (1996) *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman & Hall.

Chen, P., Lin, C., and B. Schölkopf (2004) " A Tutorial on ν-Support Vector Machines." Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan.

Chipman, H.A., George, E.I., and McCulloch, R.E. (1998) "Bayesian CART Model Search. *Journal of the American Statistical Association*

Christianini, N, and J. Shawe-Taylor. (2000) *Support Vector Machines*. Cambridge, UK: Cambridge University Press. 93(443): 935-948.

Cleveland, W. (1979) "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 78: 829–836.

Cleveland, (W.) 1993) *Visualizing Data*. Summit, New Jersey: Hobart Press.

Committee on Reducing Porpoise Mortality from Tuna Fishing (1992) *Dolphins and the Tuna Industry*. Washington, DC, National Academy Press.

Cook, D.R. and S. Weisberg (1999) *Applied Regression Including Computing and Graphics*. New York: John Wiley and Sons.

Crawley, M.J. (2007) *The R Book*. New York: John Wiley and Sons.

Dalgaard, P. (2002) *Introductory Statistics with R*. New York: Springer.

Damšar, J. (2006) "Statistical Comparisons of Classifiers over Multiple Data Sets." *Journal of Machine Learning Research* 7: 1–30.

Danilov, D.L. and J.R. Magnus (2004) "On the Harm that Ignoring Pre-Testing Can Cause." *Journal of Econometrics* 122: 27–46.

Dasu, T. and T. Johnson (2003) *Exploratory Data Mining and Data Cleaning*. New York: John Wiley and Sons.

de Boors, C. (2001) *A Practical Guide to Splines*, revised edition. New York: Springer.

Deloncle, A., Berk, R.A., D'Andrea, F., and M. Ghil (2006) "Weather Regmine Prediction Using Statistical Learning." UCLA Department of Statistics Preprint # 435.

Drummond, C. and R.C. Holte (2006) "Cost Curves: An Improved Method for Visualizing Classifier Performance." *Machine Learning* 65 (1): 95–130.

Efron, B. (1981) "Nonparametric Standard Errors and Confidence Intervals." *Canadian Journal of Statistics* 9: 139–172.

Efron, B. (1983) "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation." *Journal of the American Statistical Association* 78: 892-898.

Efron, B. (1987) "Better Bootstrap Confidence Intervals." *Journal of the American Statistical Association* 82: 171–185.

Efron, B. (2004) "The Estimation of Prediction Error: Covariance Penalties and Cross-Validation" (with discussion). *Journal of the American Statistical Association* 99: 619–642.

Efron, B. and R. Tibshirani (1993) *Introduction to the Bootstrap*. New York: Chapman & Hall.

Efron, B. and R. Tibshirani (1997) "Improvements on Cross-Validation: the 632+ Bootstrap Method," *Journal of the American Statistical Association* 92: 548–560.

Efron, B., Hastie, T., Johnstone, I., and R. Tibshirani (2004) "Least Angle Regression" (with discussion). *The Annals of Statistics* 32(2): 407–499.

Efron, B., Hastie, T., and R. Tibshirani (2007) "Discussion: The Dantzig Selector: Statistical Estimation When p is much larger than n." *Annals of Statistics* 35(6): 2358–2364.

Eibl, G. and K. Pfeiffer (2005) "Multiclass Boosting for Weak Classifiers." *Journal of Machine Learning Research* 6: 189–210.

Fan, J. and I. Gijbels. (1996) *Local Polynomial Modeling and its Applications*. New York: Chapman & Hall.

Fan, J. and Li. R. (2006) "Statistical Challenges with Dimensionality: Feature Selection in Knowledge Discovery." In M. Sanz-Sole, J. Soria, J.L. Varona, J. Verdera (eds.), Proceedings of the International Congress of Mathematicians Vol. III, 595–622.

Fan G., and B. Gray (2005) "Regression Tree Analysis Using TARGET." *Journal of Computational and Graphical Statistics* 14: 206–218.

Faraway, J. (2004) "Human Animation Using Nonparametric Regression." *Journal of Computational and Graphical Statistics* 13: 537–553.

Freedman, D.A. (1987) "As Others See Us: A Case Study in Path Analysis" (with discussion). *Journal of Educational Statistics* 12: 101–223.

Freedman, D.A. (2004) "Graphical Models for Causation and the Identification Problem." *Evaluation Review* 28: 267–293.

Freedman, D.A. (2005) *Statistical Models* Cambridge, UK: Cambridge University Press.

Freedman, D.A. and R. Berk (2008) "Weighting Regressions by Propensity Scores." *Evaluation Review*, forthcoming.

Freedman, D.A., Navidi, W., and S.C. Peters (1988) "On the Impact of Variable Selection in Fitting Regression Equations." In T.K. Dijkstra (ed.), *On Model Uncertainty and Its Statistical Implications*, 1–16, Springer, Berlin

Freund, Y., and R. Schapire. (1996) "Experiments with a New Boosting Algorithm. *Macine Learning: Proceedings for the Thirteenth International Conference* 148–156. Morgan Kaufmann, San Francisco.

Freund, Y., and R.E. Schapire (1999) "A Short Introduction to Boosting." *Journal of the Japanese Society for Artificial Intelligence* 14: 771–780.

Friedman, J.H. (1991) "Multivariate Adaptive Regression Splines (with discussion)," *The Annals of Statistics* 19: 1–82.

Friedman, J. H. (2001) "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics* 29: 1189–1232

Friedman, J.H. (2002) "Stochastic Gradient Boosting," *Computational Statistics and Data Analysis* 38: 367–378.

Friedman, J.H. and P. Hall (2000) "On Bagging and Non-Linear Estimation." *Technical Report*, Department of Statistics, Stanford University.

Friedman, J.H., Hastie, T., and R. Tibshirani (2000). "Additive Logistic Regression: A Statistical View of Boosting" (with discussion). *Annals of Statistics* 28: 337–407.

Friedman, J.H., Hastie, T., Rosset S., Tibshirani, R., and J. Zhu (2004) "Discussion of Boosting Papers," *Annals of Statistics* 32: 102–107.

Gareth, M. and P. Radchenko (2007) "Sparse Generalized Linear Models." Working Paper, Department of Statistics, Marshall School of Business, University of California.

Gareth, M. and J. Zhu (2007) "Funtional Linear Regression That's Interpretable." Working Paper, Department of Statistics, Marshall School of Business, University of California.

Geurts, P., Ernst, D., and L. Wehenkel (2006) "Extemely Randomized Trees." *Machine Learning* 63 (1): 3–42.

Gifi, A. (1990) *Nonlinear Multivariate Analysis*. New York: John Wiley and Sons.

Goldenshluger, A. and A. Tsybakov (2001) "Adaptive Prediction and Estimation in Linear Regression with Infinitely Many Parameters." *Annals of Statistics* 29(6): 1601–1619.

Grandvalet, Y. (2004) "Bagging Equalizes Influence." *Machine Learning* 55: 251–270.

Granger, C.W.J. and P. Newbold, (1986) *Forecasting Economic Time Series*. New York: Academic Press.

Green, P.J. and B.W. Silverman (1994) *Nonparametric regression and Generalized Linear Models*. New York: Chapman & Hall.

Green, W.H. (2003) *Econometric Analysis*, fifth edition. New York: Prentice Hall.

Hand, D., Manilla, H., and P. Smyth (2001) *Principles of Data Mining*. Cambridge, MA: MIT Press.

Hastie, T., and J. Zhu (2006) "Comment." *Statistical Science* 21(3): 352–356.

Hastie, T., Tibshirani, R. and J. Friedman (2001) *The Elements of Statistical Learning*. New York: Springer-Verlag.

Hastie, T.J. and R.J. Tibshirani. (1990) *Generalized Additive Models*. New York: Chapman & Hall.

Hastie, T.J. and R.J. Tibshirani. (1996) "Discriminant Adaptive Nearest Neighbor Classification." *IEEE Pattern Recognition and Machine Intelligence* 18: 607–616.

He, Y. (2006) "Missing Data Imputation For Tree-Based Models." PhD dissertation for the Department of Statistics, UCLA.

Heckman, J. (1999) "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective." *The Quarterly Journal of Economics*, February: 45–97.

Heckman, J, Ichimura, H., and P. Todd (1998) "Matching as an Economic Evaluation Estimator," *Review of Economic Studies* 65: 261–294.

Hoeting, J., Madigan, D., Raftery, A., and C. Volinsky (1999) "Bayesian Model Averaging: A Practical tutorial." *Statistical Science* 14: 382–401.

Horváth, T. and A. Yamamoto (2006) "International Conference on Inductive Logic Programming." *Journal of Machine Learning* 64:3–144.

Hothorn, T., Berthold, L., Brenner, A., and M. Radespiel-Tröger. (2002) "Bagging Survival Trees." Department of Medical Informatics, Biometry and Epidemiology, Freidrich-Alexander University Erlangen-Nuremberg, preprint.

Hothorn, T., and B. Lausen (2003) "Double-Bagging: Combining Classifiers by Bootstrap Aggregation." *Pattern Recognition* 36: 1303–1309.

Hothorn, T., Hornik, K., and A. Zeileis (2006) "Unbiased Recursive Partitioning: A Conditional Inference Framework." *Journal of Computational and Graphical Statistics* 15(3): 651–674.

Hurvich, C.M., and C. Tsai (1989) "Regression and Time Series Model Selection in Small Samples." *Biometrika* 76(2): 297–307.

Hsu, C., Chung, C., and C. Lin (2007) "A Practical Guide to Support Vector Classification," Department of Computer Science and Information Engineering, National Taiwan University, Taipsi, Taiwan. http://www.csie.ntu.edu.tw/ cjlin/libsvm.

Imbens, G. (2004) "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review." *The Review of Economics and Statistics*: 86: 4–29.

Jiang, W. (2004) "Process Consistency for Adaboost." *Annals of Statistics* 32: 13–29.

Joachims, T (1998) "Making Large-Scale SVM Learning Practical," in B. Schölkopf, C.J.C. Burges, and A.J. Smola (Eds.) *Advances in Kernel Methods — Support Vector Learning* Cambridge, MA: MIT Press.

Krieger, A., Long, C., and A. Wyner (2001) "Boosting Noisy Data." *Prroceedings of the International Conference on Machine Learning*, Amsterdam: Mogan Kauffman.

Kriegler, B. (2007) "Boosting the Quantile Distribution: A Cost-Sensitive Statistical Learning Procedure." Department of Statistics, UCLA, working paper.

Lafferty, J. and L. Wasserman (2008) "Rodeo: Sparse Greedy Nonparametric Regression." *Annals of Statistics* 36(1): 28-63.

Leamer, E.E. (1978) *Specification Searches: Ad Hoc Inference with Non-Experimental Data*. New York, John Wiley.

LeBlanc, M. and R. Tibshirani (1996) "Combining Estimates on Regression and Classification." *Journal of the American Statistical Association* 91: 1641–1650.

Leeb, H. and B.M. Pötscher 2006) "Can One Estimate the Conditional Distribution of Post-Model-Selection Estimators?" *The Annals of Statistics* 34(5): 2554–2591.

Lennert-Cody, C.E. and R.A. Berk (2005) "Statistical Learning Procedures for Monitoring Regulatory Compliance: An Application to Fisheries Data. UCLA Department of Statistics Preprint Series #426.

Li, Y., Lie, Y., and J. Zhu (2007) "Quantile Regression in Reproducing Kernel Hilbert Spaces." *Journal of the American Statistical Association* 102: 255–268.

Lin, Y. and Y. Jeon (2006) "Random Forests and Adaptive Nearest Neighbors." *Journal of the American Statistical Association* 101: 578–590.

Lipton, P. (2005) "Testing Hypotheses: Prediction and Prejudice." *Science* 307: 219–221.

Little, R. and D. Rubin (2002) *Statistical Analysis with Missing Data*, second edition. New York: John Wiley.

Loader, C. (1999) *Local Regression and Likelihood*. New York: Springer-Verlag.

Loader, C. (2004) "Smoothing: Local Regression Techniques," in J. Gentle, W. Hardle, and Y. Mori, *Handbook of Computational Statistics*. NewYork: Springer-Verlag.

Loh, W.-Y. (2002) "Regression Trees with Unbiased Variable Selection and Interaction Detection." *Statistica Sinica* 12: 361–386.

Lugosi, G. and N. Vayatis (2004) "On Bayes Risk Consistency of Regularized Boosting Methods." *Annals of Statistics* 32(1): 30–55.

Luo, X, Stefanski, L.A., and D.B. Boos (2006) "Tuning Variable Selection Procedures by Adding Noise." *Technometrics* 48(2): 165–175.

Lutz, R.W. and P. Büehlmann (2006) "Conjugate Direction Boosting." *Journal of Computational and Graphical Statistics* 15: 287–311.

Maindonald, J. and J. Braun (2007) *Data Analysis and Graphics Using R*, second edition. Cambridge, UK: Cambridge University Press.

Madigan, D., Raftery, A.E., Volinsky, C., and J. Hoeting (1996) "Bayesian Model Averaging." In *AAAI Workshop on Integrating Multiple Learned Models*: 77–83, AAAI Press.

Manly, B.F.J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*. New York: Chapman & Hall.

Mannor, S., Meir, R. and T. Zhang (2002) "The Consistency of Greedy Algorithms for Classification." In J. Kivensen and R.H. Sloan (Eds.), COLT 2002, LNAI 2375: 319–333.

Manski, C.F. (1990) "Nonparametric Bounds on Treatment Effects." *American Economic Review Papers and Proceedings* 80: 319–323.

McCaffrey, D., Ridgeway, G., and A. Morral (2004) "Propensity Score Estimation with Boosted Regression for Evaluating Adolescent Substance Abuse Treatment." *Psychological Methods* 9: 403–425.

McCullagh, P. and J.A. Nelder (1989) *Generalized Linear Models*, second edition. New York: Chapman and Hall.

Mease, D. and A.J. Wyner (2008) "Evidence Contrary to the Statistical View of Boosting." *Journal of Machine Learning* 9: 1-26

Mease, D., Wyner, A.J. and A. Buja (2007) "Boosted Classification Trees and Class Probability/Quantile Estimation." *Journal of Machine Learning* 8: 409–439.

Meinshausen, N. (2006) "Quantile Regression Forests." *Journal of Machine Learning Research* 7: 983–999.

Meinshausen, N. and P. Bühlmann (2006) "High Dimensional Graphs and Variable Selection with the Lasso." *The Annals of Statistics* 34(3) 1436–1462.

Meinshausen, N., Rocha, G., B. Yu. (2007) "Discussion: The Dantzig Selector: Statistical Estimation When p is much larger than n." *Annals of Statistics* 35(6): 2373–2384.

Mocan, H.N. and K. Gittings (2003) "Getting off Death Row: Commuted Sentences and the Deterrent Effect of Capital Punishment." *Journal of Law and Economics* 46: 453-478.

Moguerza, J.M. and A. Munõz (2006) "Support Vector Machines with Applications." *Statistical Science* 21(3): 322–336.

Mojirsheibani, M. (1997) "A Consistent Combined Classification Rule." *Statistics & Probability Letters* 36: 411-419.

Mojirsheibani, M. (1999) "Combining Classifiers vis Discretization." *Journal of the American Statistical Association* 94: 600-609.

Murrell, P. (2006) *R Graphics*. New York: Chapman & Hall/CRC.

Peña, D. (2005) "A New Statistic for Influence in Linear Regression." *Technometrics* 47: 1–12.

Polzehl, J. and S. Spokoiny (2000) "Adaptive Weights Smoothing with Applications to Image Restoration." *Journal of the Royal Statistical Association*, Soc. B, 62, Part 2: 335–354.

Polzehl, J. and S. Spokoiny (2002). "Varying Coefficient Regression Modeling by Adaptive Weights Smoothing." Manuscript.

Quinlan, R. (1993) *Programs in Machine Learning*. San Mateo, CA: Morgan Kaufman.

Raftery, A.D. (1995) "Bayesian Model Selection in Social Research." *Sociological Methodology* 25: 111–163.

Reunanen, J. (2003) "Overfitting in Making Comparisons between Variable Selection Methods." *Journal of Machine Learning Research* 3: 1371–1382.

Ridgeway, G. (1999) "The State of Boosting." *Computing Science and Statistics*, 31: 172–181.

Ridgeway, G. (2005) "Generalized Boosted Models: A Guide to the gbm Package." Available at http: //i-pensieri.com/gregr/papers/gbm-vignette.pdf

Ripley, B.D., (1996) *Pattern Recognition and Neural Networks*. Cambridge, UK: Cambridge University Press.

Rosenbaum, P.R. (2002) *Observational Studies*. second edition. New York: Springer-Verlag.

Rosset, S. and J. Zhu (2007) "Piecewise Linear Regularized Solution Paths." *The Annals of Statistics* 35 (3): 1012–1030.

Rubin, D. B. (1986) "Which Ifs Have Causal Answers." *Journal of the American Statistical Association* 81: 961–962.

Ruppert, D. (1997) "Empirical-Bias Bandwidths for Local Polynomial Nonparametric Regression and Density Estimation." *Journal of the American Statistical Association* 92: 1049–1062.

Ruppert, D. and M.P. Wand (1994) "Multivariate Locally Weighted Least Squares Regression." *Annals of Statistics* 22: 1346–1370.

Ruppert, D., Wand, M.P. and R.J Carroll (2003) *Semiparametric Regression*. Cambridge, UK: Cambridge University Press.

Schwartz, G. (1978) "Estimating the Dimension of a Model." *Annals of Statistics* 6: 461–464.

Segal, M.R. (2003) "Machine Learning Benchmarks and Random Forest Regression." Working paper, Department of Biostatistics University of California San Francisco (http://www.ics.uci.edu/ mlearn/MLRepository.html).

Shakhnarovich, G. (Ed.) (2006) *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice*. Cambridge, MA: MIT Press.

Shapire, R.E. (1999) "A Brief Introduction to Boosting." *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*.

Shapire, R.E. (2002) "The Boosting Approach to Machine Learning: An Overview." *MSRI Workshop on Non-Linear Estimation and Classification*.

Simonoff, J.S., and C. Tsai (1999) "Semiparametric and Additive Model Selection Using an Improved Akaike Information Criterion." *Journal of Computational and Graphical Statistics* 8(1): 22–40.

Stein, C. (1956) "Inadmissability of the Usual Estimator for the Mean of Multivariate Normal Distribution." *Proceedings of the 3rd Berkeley Symposium* 1: 197–206.

Strobl, C., Bouestreix, A., Zeileis, A., and T. Hothorn (2007) "Bias in Random Forest Variable Importance Measures: Illustrations, Sources, and a Solution." *Bioinformatics* 8(25).

Su, X., Wang, M., and J. Fan (2004) "Maximum Likelihood Regression Trees." *Journal of Computational and Graphical Statistics* 13: 586–598.

Sutton, R.S. and A.G. Barto. (1999) *Reinforcement Learning*. Cambridge, MA: MIT Press.

Svetnik, V., Liaw, A., and C.Tong. (2003) "Variable Selection in Random Forest with Application to Quantitative Structure-Activity Relationship." Working paper, Biometrics Research Group, Merck & Co., Inc.

Therneau, T.M. and E.J. Atkinson (1997) "An Introduction to Recursive Partitioning Using the RPART Routines." Technical Report # 61, Mayo Foundation.

Thompson, S.K. (2002) *Sampling*, second edition. New York: John Wiley.

Tibshirani, R.J. (1996) "Regression Shrinkage and Selection Via the LASSO." *Journal of the Royal Statistical Society*, Series B, 25: 267–288.

Traskin, M. (2008) "The Role of Bootstrap Sample Size in the Consistency of the Random Forest Algorithm." Technical Report, Department of Statistics, University of Pennsylvania.

Vapnick, V. (1996) *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.

Wahba, G. (2006) "Comment." *Statistical Science* 21(3): 347–351

Wang, H., Li, G., and F. Jiang (2007) "Robust Regression Shrinkage and Consistent Variable Selection Through the LAD-Lasso." *Journal of Business and Economic Statistics* 25(3): 347–355.

Witten, I.H. and E. Frank. (2000). *Data Mining*. New York: Morgan and Kaufmann.

Wood, S.N. (2000) "Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties." *Journal of the Royal Statistical Society*, B 62(2):413–428.

Wood, S.N. (2003) "Thin Plate Regression Splines." *Journal of the Royal Statistical Society* B 65(1):95–114.

Wood, S.N. (2004) "Stable and Eficient Multiple Smoothing Parameter Estimation for Generalized Additive Models." *Journal of the American Statistical Association*, Vol. 99: 673–686.

Wood, S.N. (2006) *Generalized Additive Models* New York: Chapman & Hall.

Wu, Y., Tjelmeland, H., and M. West (2007) "Bayesian CART: Prior Specification and Posterior Simulation." *Journal of Computational and Graphical Statistics* 16(1): 44–66.

Wyner, A.J. (2003) "Boosting and Exponential Loss." In C.M. Bishop and B.J. Frey (Eds.) *Proceedings of the Ninth Annual Conference on AI and Statistics* Jan 3–6, Key West, Florida.

Xu, M. and M.W. Golay (2006) "Data-Guided Model Combination by Decomposition and Aggregation." *Machine Learning* 63 (1): 43–67.

Zellner, A., Keuzenkamp, H.A., and M. McAleer (Eds.) (2001) *Simplicity, Inference and Modelling.* Cambridge, UK: Cambridge University Press.

Zhang, C. (2005) "General Empirical Bayes Wavelet Methods and Exactly Adaptive Minimax Estimation." *The Annals of Statistics* 33(1): 54–100.

Zhang, H. and B. Singer (1999) *Recursive Partitioning in the Health Sciences*. New York: Springer-Verlag.

Zhang, T. and B. Yu (2005) "Boosting with Early Stopping: Convergence and Consistency." *Annals of Statistics*, 33(4): 1538–1579.

Zou, H. (2006) "The Adaptive Lasso and Its Oracle Properties." *The Journal of the American Statistical Association* 101 (467) 1418–1429.

Zou, H. and T. Hastie (2005) "Regularization and Variable Selection via Elastic Net." *Journal of the Royal Statistical Association*, series B, 67, Part 2: 301–320.

Zou, H. and Hastie, T., and R. Tibshirani (2005) "Space Principal Component Analysis." *Journal of Computational and Graphical Statistics* 15: 265–286.

Index

Adaboost, 258–266, 269, 270, 275, 276, 296
added variable plot, 85
added variable plots, 34
AIC, *see* The Akaike Information Criterion
Akaike Information Criterion, 30–37, 47, 136
Alaike Information Criterion, 339
algorithmic methods, 22, 186
asymptotic properties, 16–17, 97, 263

B-splines, 52, 57–60, 71, 86, 98
  degree one, 58
  degree three, 60
  degree two, 58
  degree zero, 58
backfitting, 85–86
bagging, 22, 193, 194, 198, 200, 204, 213, 271, 273, 303, 306, 333
  bias, 180–181, 183
  quantitative response, 187–188
  unbalanced data, 189
  variance, 179–181, 183–186
bandwidth, 76, 77, 80
basis functions, 37–38, 40, 41, 53, 57, 70, 94, 104, 118, 158, 159, 170, 171, 195, 261, 271, 272, 303
Bayes classifiers, 26
Bayes error, 114
Bayes error rate, 26
Bayes risk, 26
Bayesian Information Criterion, 30–37, 136, 339
Bayesian model averaging, 169
Baysian methods, 68
bias, 16–17, 24, 32, 33, 63, 66–68, 112, 135–137
bias–variance tradeoff, 17–20, 41, 55–56, 69, 70, 73, 74, 129, 170, 181, 190, 197, 307
BIC, *see* Bayesian Information Criterion
boosting, 21, 333, 336
bootstrap, 32–33, 68, 95, 179

CART, *see* classification and regression trees
causal model, 7, 12, 14, 79, 331–332
classification, 23, 41, 108, 117–119
classification and regression trees, XI, 40, 170, 171, 173, 174, 176, 180, 181, 183, 186, 187, 189, 193–196, 200, 204, 206, 207, 209–211, 222, 230, 233, 235, 261, 267, 268, 271–272, 276, 278, 304, 321, 332, 333, 335
  bias, 149–150, 158, 161, 174–178
  classification, 138–139
  classification errors, 122–128
  forecasting, 138–139
  missing data, 132–135
  statistical inference, 135–138
  surrogate variables, 134–135
  variance, 150–154, 162, 174–178
classifier, 23, 104, 199
complexity, 17
confidence interval, 14–16, 40, 63, 66, 68, 69, 73, 94–96, 136

confusion tables, 108–110
consistency, 16–340
coplot, 6, 7, 46
cost complexity, 129–130, 139–145
cross-validation, 31, 63, 154, 170–172, 325, 339
  bias, 171
curse of dimensionality, 38, 80, 84

Dantzig Selector, 66–67
data snooping, 7, 33, 44, 45, 95, 340, 341
data summary, 13
data-generation mechanism, *see* data-generation process
data-generation process, 9–11, 22, 23, 40, 43, 138, 329, 339
degrees of freedom, 28–29, 47, 136
description, 12, 15, 41
deviance, 100, 116, 135, 136, 158, 224
drop-one cross-validation, 31

EDA, 1
effective degrees of freedom, 28
efficiency, 16–17
elastic net, 65–66, 69
entropy, 114, 115
equivalent degrees of freedom, 28
Euclidian distance, 80, 82
expected prediction error, 31–34
exploratory data analysis, 1

F-test, 48, 96, 97, 136
forecasting, 13, 15, 41, 108, 331
  costs of errors, 122–128, 139–145

GAM, *see* generalized additive model
generalized additive model, 84–88, 91
generalized cross-validation statistic, 32, 56, 80, 82, 86
generalized linear model, 66, 84
gentle Adaboost, 259
Gini index, 114
GLM, *see* generalized linear model
Granger causality, 214, 216
greedy algorithms, 264
GVC, *see* generalized cross-validation statistic

hat matrix, 27, 28

impurity, 113–117, 130, 131, 155, 194, 213, 243

jackknife, 31
jitter, 3, 40
Jous-Boost, 276, 294

knots, 50, 53–54, 56–57, 70–71, 76
Kullbeck–Leibler information, 30

L_0–penalty, 62
L_1–penalty, 61, 64–66
L_2–penalty, 61, 62
lasso, 64–67, 69, 266
  adaptive, 65
least angle regression(LARS), 264–266
leave-one-out cross-validation, 31
linear basis expansion, 37
linear estimators, 26–27
linear loss, 25
locally adaptive smoothers, 81, 91–93
locally weighted regression, 39, 73–80
logitboost, 262
loss function, 14, 23–26, 42, 44, 157
  asymmetric, 25
  symmetric, 25, 26
lowess, 75, 82–84, 86
  robust, 77

M-estimators, 77
margin, 154, 173, 261
Matlab, 43
mean squared error, 29, 33, 243
  decomposition, 180
missing data, 131
model evaluation, 29–34
model selection, 34–37, 44, 135
model-based sampling, 14, 95, 114
multidimensional scaling, 231
multivariate adaptive regression splines, 158–159
multivariate smoothers, 80–88

N-fold cross-validation, 31, 71
natural cubic splines, 54–57, 70–71
nearest neighbor methods, 24, 73–76, 110–112, 204–210, 302

omitted variable, 79
out-of-bag observations, 173–174, 186

overfitting, 17–20, 33, 36, 42, 44, 45, 60, 80, 130, 170, 187, 202
overtuning, 33, 42, 45, 339

pairwise deletion, 134
penalized smoothing, 60–69, 86
piecewise linear basis, 49–53
point-by-point confidence interval, 72, 87
polynomial regression splines, 53–54
principal components analysis, 69
prior, 139–145
prior probability, 125–128
pruning, 128–131, 155

quadratic loss, 24, 25
quantile classification, 276

R^2, 29, 47, 48
  adjusted, 29, 47, 48
random assignment, 15
random forests, 257, 258, 261, 263, 267, 273, 274, 276, 279, 283, 285, 290, 303, 308, 321, 326, 332, 333, 335, 336, 338, 339, 341
  clustering, 231
  costs, 210–212
  dependence, 203
  generalization error, 200–204
  margin, 200–201, 203–204
  mean square error, 233
  missing data, 231–232
  outliers, 232
  partial dependence plots, 222–226
  predictor importance, 213–216
  proximity matrix, 229–232
  quantile, 234–235, 245–248
  quantitative response, 233–234
  specialized basis functions, 195, 197
  strength, 202–203
  tuning, 197, 211, 234–236
  variance, 195
real Adaboost, 259
receiver operator characteristic curve, 251
recursive partitioning, 105
regression analysis, 2–8
  definition, 2, 23
regression splines, 49–60
regression trees, 154–157
regularization, 62, 64, 65
residual degrees of freedom, 28
resubstitution, 29
ridge regression, 62–65, 71
rodeo, 67–68
rug plot, 72

Salford Systems, 43, 249
shrinkage, 61–69, 95, 179
smoother, 52
smoother matrix, 26, 28, 72
smoothing, 8
smoothing splines, 70–73, 81–84
span, 76, 77, 81, 82
sparsity, 66, 67
squared loss, *see* quadratic loss
stagewise algorithms, 261, 268
stagewise regression, 36, 105, 264–266
statistical inference, 7, 14–16, 63, 65, 68–69, 93–97, 330
statistical learning
  definition, 23
statistical test, 14–16, 66, 68, 69, 96–97, 136
Stein estimator, 68
stepwise regression, 8, 35–36, 47, 104, 137, 264
stochastic gradient boosting, 266–274
  partial dependence plots, 273–274
  predictor importance, 274
  tuning, 271–273
support vector classifier, 303–307, 309–312
  decision boundary, 304
  hard boundary, 310
  margin boundary, 306
  separating hyperplane, 304
  soft boundary, 311
  target variable, 303
support vector machines, 307–309, 312–315, 336
  classification error costs, 321–322, 325
  hinge loss function, 315
  hockey stick loss function, 315
  kernel function, 312–314, 324
  linear kernel, 316
  overfitting, 325
  overtuning, 325

quantitative response, 315
radial kernel, 318
regularized sum of squares, 314
separating hyperplane, 315
tuning, 318–321, 325

test data, 34, 44, 339
thin plate splines, 81
training data, 44
tree diagrams, 105–108
truncated power series basis, 53
tuning, 42, 43, 62, 65, 70, 338

unbalanced data, 149
unbiased, 24

variable importance, 159

weak classifiers, 183
window, 76

(continued from page ii)

Iacus: Simulation and Inference for Stochastic Differential Equations: With R Examples
Ibrahim/Chen/Sinha: Bayesian Survival Analysis
Jiang: Linear and Generalized Linear Mixed Models and Their Applications
Jolliffe: Principal Component Analysis, 2nd edition
Konishi/Kitagawa: Information Criteria and Statistical Modeling
Knottnerus: Sample Survey Theory: Some Pythagorean Perspectives
Kosorok: Introduction to Empirical Processes and Semiparametric Inference
Küchler/Sørensen: Exponential Families of Stochastic Processes
Kutoyants: Statistical Inference for Ergodic Diffusion Processes
Lahiri: Resampling Methods for Dependent Data
Lavallée: Indirect Sampling
Le Cam: Asymptotic Methods in Statistical Decision Theory
Le Cam/Yang: Asymptotics in Statistics: Some Basic Concepts, 2nd edition
Le/Zidek: Statistical Analysis of Environmental Space-Time Processes
Liese/Miescke: Statistical Decision Theory: Estimation, Testing, Selection
Liu: Monte Carlo Strategies in Scientific Computing
Manski: Partial Identification of Probability Distributions
Mielke/Berry: Permutation Methods: A Distance Function Approach, 2nd edition
Molenberghs/Verbeke: Models for Discrete Longitudinal Data
Morris/Tibshirani: The Science of Bradley Efron, Selected Papers
Mukerjee/Wu: A Modern Theory of Factorial Designs
Nelsen: An Introduction to Copulas, 2nd edition
Pan/Fang: Growth Curve Models and Statistical Diagnostics
Politis/Romano/Wolf: Subsampling
Ramsay/Silverman: Applied Functional Data Analysis: Methods and Case Studies
Ramsay/Silverman: Functional Data Analysis, 2nd edition
Reinsel: Elements of Multivariate Time Series Analysis, 2nd edition
Rosenbaum: Observational Studies, 2nd edition
Rosenblatt: Gaussian and Non-Gaussian Linear Time Series and Random Fields
Särndal/Swensson/Wretman: Model Assisted Survey Sampling
Santner/Williams/Notz: The Design and Analysis of Computer Experiments
Schervish: Theory of Statistics
Shaked/Shanthikumar: Stochastic Orders
Shao/Tu: The Jackknife and Bootstrap
Simonoff: Smoothing Methods in Statistics
Song: Correlated Data Analysis: Modeling, Analytics, and Applications
Sprott: Statistical Inference in Science
Stein: Interpolation of Spatial Data: Some Theory for Kriging
Taniguchi/Kakizawa: Asymptotic Theory for Statistical Inference for Time Series
Tanner: Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions, 3rd edition
Tillé: Sampling Algorithms
Tsaitis: Semiparametric Theory and Missing Data
van der Laan/Robins: Unified Methods for Censored Longitudinal Data and Causality
van der Vaart/Wellner: Weak Convergence and Empirical Processes: With Applications to Statistics
Verbeke/Molenberghs: Linear Mixed Models for Longitudinal Data
Weerahandi: Exact Statistical Methods for Data Analysis

Springer
the language of science

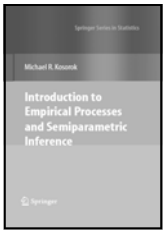
springer.com

Time Series Analysis with Applications in R

Johnathon D. Cryer and Kung-Sik Chan

Time Series Analysis With Applications in R, Second Edition, presents an accessible approach to understanding time series models and their applications. Although the emphasis is on time domain ARIMA models and their analysis, the new edition devotes two chapters to the frequency domain and three to time series regression models, models for heteroscedasticty, and threshold models. All of the ideas and methods are illustrated with both real and simulated data sets. A unique feature of this edition is its integration with the R computing environment.

2008. 2nd Ed., 494p. (Springer Texts in Statistics) Hardcover
ISBN 0-387-75958-6



Introduction to Empirical Processes and Semiparametric Inference

Michael R. Kosorok

This book provides a self-contained, linear, and unified introduction to empirical processes and semiparametric inference. These powerful research techniques are surprisingly useful for developing methods of statistical inference for complex models and in understanding the properties of such methods. The targeted audience includes those who have an interest in learning about and doing research in empirical processes and semiparametric inference but who would like to have a friendly and gradual introduction to the area.

2008. 476 pp. (Springer Series in Statistics) Hardcover
ISBN 978-0-387-74977-8

Introduction to Applied Bayesian Statistics and Estimation for Social Scientists

Scott M. Lynch

This book covers the complete process of Bayesian statistical analysis in great detail from the development of a model through the process of making statistical inference. The key feature of this book is that it covers models that are most commonly used in social science research, including the linear regression model, generalized linear models, hierarchical models, and multivariate regression models, and it thoroughly develops each real-data example in painstaking detail.

2007. 357 pp. (Statistics for Social and Behavioral Sciences) Hardcover
ISBN 978-0-387-71264-2

Easy Ways to Order► Call: Toll-Free 1-800-SPRINGER ▪ E-mail: orders-ny@springer.com ▪ Write: Springer, Dept. S8113, PO Box 2485, Secaucus, NJ 07096-2485 ▪ Visit: Your local scientific bookstore or urge your librarian to order.