References

Affenseller, M., Winkler, S., Wagner, S., & Beham, A. (2009). *Genetic algorithms and genetic programming: Modern concepts and practical applications*. New York: Chapman & Hall.

Akaike, H. (1973). Information theory and an extension to the maximum likelihood principle. In B. N. Petrov & F. Casaki (Eds.), *International Symposium on Information Theory* (pp. 267–281). Budapest: Akademia Kiado.

Angrist, J. D., & Pischke, J. (2009). *Mostly harmless econometrics*. Princeton: Princeton University Press.

Baca-García, E., Perez-Rodriguez, M. M., Basurte-Villamor, I., Saiz-Ruiz, J., Leiva-Murillo, J. M., de Prado-Cumplido, M., et al. (2006). Using data mining to explore complex clinical decisions: A study of hospitalization after a suicide attempt. *Jounal of Clinical Psychiatry*, *67*(7), 1124–1132.

Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge: Cambridge University Press.

Bartlett, P. L., & Traskin, M. (2007). Adaboost is Consistent. *Journal of Machine Learning Research*, *8*(2347–2368), 2007.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*(6), 561–571.

Berk, R. A. (2003). *Regression analysis: A constructive critique*. Newbury Park: Sage.

Berk, R. A. (2005). New claims about executions and general deterrence: Déjà vu all over again? *Journal of Empirical Legal Studies*, *2*(2), 303–330.

Berk, R. A. (2012). *Criminal justice forecasts of risk: A machine learning approach*. New York: Springer.

Berk, R. A., & Freedman, D. A. (2003). Statistical assumptions as empirical commitments. In T. Blomberg & S. Cohen (Eds.), *Law, punishment, and social control: Essays in honor of Sheldon Messinger, Part V* (pp. 235–254). Berlin: Aldine de Gruyter (November 1995, revised in second edition).

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

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

Berk, R. A., Brown, L., & Zhao, L. (2010). Statistical inference after model selection. *Journal of Quantitative Criminology*, *26*, 217–236.

Berk., R. A., Brown, L., Buja, A., Zhang, K., & Zhao, L. (2014). Valid post-selection inference. *Annals of Statistics*, *41*(2), 802–837

R.A. Berk, *Statistical Learning from a Regression Perspective*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-44048-4

Berk, R. A., Brown, L., Buja, A., George, E., Pitkin, E., Zhang, K., et al. (2014). Misspecified mean function regression: Making good use of regression models that are wrong. *Sociological Methods and Research*, *43*, 422–451.

Berk, R. A., & Bleich, J. (2013). Statistical procedures for forecasting criminal behavior: A comparative assessment. *Journal of Criminology and Public Policy*, *12*(3), 515–544.

Berk, R. A., & Bleich, J. (2014). Forecast violence to inform sentencing decisions. *Journal of Quantitative Criminology*, *30*, 79–96.

Berk, R. A., & Hyatt, J. (2015). Machine learning forecasts of risk to inform sentencing decisions. *Federal Sentencing Reporter*, *27*(4), 222–228.

Bhat, H. S., Kumer, N., & Vaz, G. (2011). Quantile regression trees. Working Paper, School of Natural Sciences, University of California, Merced.

Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*, *13*, 1063–1095.

Biau, G., Devroye, L., & Lugosi, G. (2008). Consistency of random forests and other averaging classifiers. *Journal of Machine Learning Research*, *9*, 2015–2033.

Biau, G., & Devroye, L. (2010). On the layered nearest neighbor estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification. *Journal Multivariate Analysis*, *101*, 2499–2518.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, *90*(430), 443–450.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *26*, 123–140.

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., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey: Wadsworth Press.

Breiman, L., Meisel, W., & Purcell, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics*, *19*, 135–144.

Bring, J. (1994). How to standardize regression coefficients. *The American Statistician*, *48*(3), 209–213.

Bühlmann, P. (2006). Boosting for high dimensional linear models. *The Annals of Statistics*, *34*(2), 559–583.

Bühlmann, P., & Yu, B. (2002). Analyzing bagging. *The Annals of Statistics*, *30*, 927–961.

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

Bühlmann, P., & Yu, B. (2006). Sparse boosting. *Journal of Machine Learning Research*, *7*, 1001–1024.

Bühlmann, P., & van de Geer, S. (2011). *Statistics for high dimensional data*. New York: Springer.

Buja, A., & Rolke, W. (2007). Calibration for simultaneity: (Re) sampling methods for simultaneous inference with application to function estimation and functional data. Working Paper. https://www-stat.wharton.upenn.edu/~buja/.

Buja, A., & Stuetzle, W. (2000). Smoothing effects of bagging. Working Paper. http://www-stat.wharton.upenn.edu/~buja/.

Buja, A., & Stuetzle, W. (2006). Observations on bagging. *Statistica Sinica*, *16*(2), 323–352.

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

Buja, A., Stuetzle, W., & Shen, Y. (2005). Loss functions for binary class probability estimation and classification: Structure and applications. Unpublished manuscript, Department of Statistics, The Wharton School, University of Pennsylvania.

Buja, A., Berk, R. A., Brown, L., George, E., Pitkin, E., Traskin, M. et al. (2015). Models as approximations — a conspiracy of random regressors and model violations against classical inference in regression. $imsart - stsver.2015/07/30 : Buja_$et_al_Conspiracy-v2.tex date: July 23, 2015.

Camacho, R., King, R., & Srinivasan, A. (2006). 14th International conference on inductive logic programming. *Machine Learning*, *64*, 145–287.

Candel, A., Parmar, V., LeDell, E., & Arora, A. (2016). Deep learning with H_2O. Mountain View: H_2O.ai Inc.

Candes, E., & Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n (with discussion). *Annals of Statistics*, *35*(6), 2313–2351.

Chaudhuri, P., Lo, W.-D., Loh, W.-Y., & Yang, C.-C. (1995). Generalized regression trees. *Statistic Sinica*, *5*, 641–666.

Chaudhuri, P., & Loh, W.-Y. (2002). Nonparametric estimation of conditional quantiles using quantile regression trees. *Bernoulli*, *8*(5), 561–576.

Chen, P., Lin, C., & Schölkopf, B. (2004). A tutorial on ν-support vector machines. Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan.

Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. arXiv:1603.02754v1 [cs.LG].

Chipman, H. A., George, E. I., & McCulloch, R. E. (1998). Bayesian CART model search. *Journal of the American Statistical Association*, *93*(443), 935–948.

Chipman, H. A., George, E. I., & McCulloch, R. E. (1999). Hierarchical priors for Bayesian CART shrinkage. *Statistics and Computing*, *10*(1), 17–24.

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *Annals for Applied Statistics*, *4*(1), 266–298.

Christianini, N., & Shawe-Taylor, J. (2000). *Support vector machines* (Vol. 93(443), pp. 935–948). Cambridge, UK: Cambridge University Press.

Choi, Y., Ahn, H., & Chen, J. J. (2005). Regression trees for analysis of count data with extra Poisson variation. *Computational Statistics & Data Analysis*, *49*, 893–915.

Clarke, B., Fokoué, E, & Zhang, H. H. (2009). *Principles and theory of data mining and machine learning* New York: Springer.

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.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: Wiley.

Cook, D. R., & Weisberg, S. (1999). *Applied regression including computing and graphics*. New York: Wiley.

Crawley, M. J. (2007). *The R book*. New York: Wiley.

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

Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning*. New York: Wiley.

de Boors, C. (2001). *A practical guide to splines* (revised ed.). New York: Springer.

Deng, L., & Yu, D. (2014). *Deep learning: Methods and applications*. Boston: Now Publishers Inc.

Dijkstra, T. K. (2011). Ridge regression and its degrees of freedom. Working Paper, Department Economics & Business, University of Gronigen, The Netherlands.

Duvenaud, D., Lloyd, J. R., Grosse, R., Tenenbaum, J. B., & Ghahramani, Z. (2013). Structure discovery in nonparametric regression through compositional kernel search. *Journal of Machine Learning Research W&CP*, *28*(3), 1166–1174.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2011). Fairness through awareness. Retrieved November 29, 2011, from arXiv:1104.3913v2 [cs.CC]

Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O., & Roth, A. (2015). The reusable holdout: Preserving validity in adaptive data analysis. *Science*, *349*(6248), 636–638.

Edgington, E. S., & Ongehena, P. (2007). *Randomization tests* (4th ed.). New York: Chapman & Hall.

Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Annals of Mathematical Statistics*, *34*, 447–456.
Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 59–82.
Efron, B. (1986). How biased is the apparent error rate of prediction rule? *Journal of the American Statistical Association*, *81*(394), 461–470.
Efron, B., & Tibshirani, R. (1993). *Introduction to the bootstrap*. New York: Chapman & Hall.
Ericksen, E. P., Kadane, J. B., & Tukey, J. W. (1989). Adjusting the 1980 census of population and housing. *Journal of the American Statistical Association*, *84*, 927–944.
Exterkate, P., Groenen, P. J. K., Heij, C., & Van Dijk, D. J. C. (2011). Nonlinear forecasting with many predictors using kernel ridge regression. Tinbergen Institute Discussion Paper 11-007/4.
Fan, J., & Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, *20*(4), 2008–2036.
Fan, J., & Gijbels, I. (1996). *Local polynomial modeling and its applications*. New York: Chapman & Hall.
Fan, G., & Gray, B. (2005). Regression tree analysis using TARGET. *Journal of Computational and Graphical Statistics*, *14*, 206–218.
Fan, J., & 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, pp. 595–622).
Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *Journal of the Royal Statistical Society*, *B70*, 849–911.
Fan, J., & Gijbels, I. (1996). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, *20*(4), 2008–2036.
Faraway, J. (2004). Human animation using nonparametric regression. *Journal of Computational and Graphical Statistics*, *13*, 537–553.
Faraway, J. J. (2014). Does data splitting improve prediction? *Statistics and computing*. Berlin: Springer
Finch, P. D. (1976). The poverty of statisticism. *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, *6b*, 1–46.
Freedman, D. A. (1981). Bootstrapping regression models. *Annals of Statistics*, *9*(6), 1218–1228.
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. (2009a). *Statistical models cambridge*. UK: Cambridge University Press.
Freedman, D. A. (2009b). Diagnostics cannot have much power against general alternatives. *International Journal of Forecasting*, *25*, 833–839.
Freund, Y., & Schapire, R. (1996). Experiments with a new boosting algorithm. In *Macine Learning: Proceedings for the Thirteenth International Conference* (pp. 148–156). San Francisco: Morgan Kaufmann.
Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of online learning and an application to boosting. *Journal of Computer and System Sciences*, *55*, 119–139.
Freund, Y., & Schapire, R. E. (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). Computational statistics and data analysis. *Stochastic Gradient Boosting*, *38*, 367–378.
Friedman, J. H., & Hall, P. (2000). On bagging and nonlinear estimation. Technical Report. Department of Statistics, Stanford University.

Friedman, J. H., Hastie, T., & Tibshirani, R. (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., & Zhu, J. (2004). Discussion of boosting papers. *Annals of Statistics*, *32*, 102–107.
Gareth, M., & Radchenko, P. (2007). Sparse generalized linear models. Working Paper, Department of Statistics, Marshall School of Business, University of California.
Gareth, M., & Zhu, J. (2007). Functional linear regression that's interpretable. Working Paper, Department of Statistics, Marshall School of Business, University of California.
Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42.
Ghosh, M., Reid, N., & Fraser, D. A. S. (2010). Ancillary statistics: A review. *Statistica Sinica*, *20*, 1309–1332.
Gifi, A. (1990). *Nonlinear multivariate analysis*. New York: Wiley.
Good, P. I. (2004). *Permutation, parametric and bootstrap tests of hypotheses*. New York: Springer.
Grandvalet, Y. (2004). Bagging equalizes influence. *Machine Learning*, *55*, 251–270.
Granger, C. W. J., & Newbold, P. (1986). *Forecasting economic time series*. New York: Academic Press.
Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*. New York: Chapman & Hall.
Grubinger, T., Zeileis, A., & Pfeiffer, K.-P. (2014). Evtree: Evolutionary learning of globally optimal classification and regression trees in R. *Journal of Statistical Software*, *61*(1). http://www.jstatsoft.org/.
Hall, P. (1997). *The bootstrap and Edgeworth expansion* New York: Springer.
Hand, D., Manilla, H., & Smyth, P. (2001). *Principles of data mining*. Cambridge, MA: MIT Press.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). New York: Springer.
Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. New York: Chapman & Hall.
Hastie, T. J., & Tibshirani, R. J. (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. Ph.D. dissertation for the Department of Statistics, UCLA.
Hoeting, J., Madigan, D., Raftery, A., & Volinsky, C. (1999). Bayesian model averaging: A practical tutorial. *Statistical Science*, *14*, 382–401.
Horváth, T., & Yamamoto, A. (2006). International conference on inductive logic programming. *Journal of Machine Learning*, *64*, 3–144.
Hothorn, T., & Lausen, B. (2003). Double-bagging: Combining classifiers by bootstrap aggregation. *Pattern Recognition*, *36*, 1303–1309.
Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674.
Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Symposium on Mathematical Statistics and Probability*, *I*, 221–233.
Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307.
Hsu, C., Chung, C., & Lin, C. (2010). A practical guide to support vector classification. Department of Computer Science and Information Engineering National Taiwan University, Taipei, Taiwan. http://www.csie.ntu.edu.tw/~cjlin/libsvm.
Ishwaran, H. (2015). The effect of splitting on random forests. *Machine Learning*, *99*, 75–118.
Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, T. S. (2008). Random survival forests. *The Annals of Applied Statistics*, *2*(3), 841–860.
Ishwaran, H., Gerds, T. A., Kogalur, U. B., Moore, R. D., Gange, S. J., & Lau, B. M. (2014). Random survival forests for competing risks. *Biostatistics*, *15*(4), 757–773.
Janson, L., Fithian, W., & Hastie, T. (2015). Effective degrees of freedom: A flawed metaphor. *Biometrika*, *102*(2), 479–485.

Jiang, W. (2004). Process consistency for adaboost. *Annals of Statistics*, *32*, 13–29.
Jiu, J., Zhang, J., Jiang, X., & Liu, J. (2010). The group dantzig selector. *Journal of Machine Learning Research*, *9*, 461–468.
Joachims, T. (1998). Making large-scale SVM learning practical. In B. Schölkopf, C. J. C. Burges, & A. J. Smola (Eds.), *Advances in kernel methods - support vector learning*. Cambridge, MA: MIT Press.
Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6234), 255–260.
Karatzoglou, A., Smola, A., & Hornik, K. (2015). kernlab – An S4 Package for Kernel Methods in R. https://cran.r-project.org/web/packages/kernlab/vignettes/kernlab.pdf.
Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, *29*(2), 119–127.
Katatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). Kernlab – An S4 package for Kernel methods in R. *Journal of Statistical Software*, *11*(9). http://www.jstatsoft.org.
Kaufman, S., & Rosset, S. (2014). When does more regularization imply fewer degrees of freedom? Sufficient conditions and counter examples from the lasso and ridge regression. *Biometrica*, *101*(4), 771–784.
Kapelner, A., & Bleich, J. (2014). BartMachine: Machine learning for bayesian additive regression trees. arXiv:1312.2171v3 [stat.ML].
Kessler, R. C., Warner, C. H., & Ursine, R. J. (2015). Predicting suicides after psychiatric hospitalization in US army soldiers: The army study to assess risk and resilience in service members (Army STARRS). *JAMA Psychiatry*, *72*(1), 49–57.
Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York: Springer.
Krieger, A., Long, C., & Wyner, A. (2001). Boosting noisy data. In *Proceedings 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., & Wasserman, L. (2008). Rodeo: sparse greedy nonparametric regression. *Annals of Statistics*, *36*(1), 28–63.
Lamiell, J. T. (2013). Statisticism in personality psychologists' use of trait constructs: What is it? How was it contracted? Is there a cure? *New Ideas in Psychology*, *31*(1), 65–71.
Leamer, E. E. (1978). *Specification searches: Ad hoc inference with non-experimental data*. New York: Wiley.
LeBlanc, M., & Tibshirani, R. (1996). Combining estimates on regression and classification. *Journal of the American Statistical Association*, *91*, 1641–1650.
Lee, S. K. (2005). On generalized multivariate decision tree by using GEE. *Computational Statistics & Data Analysis*, *49*, 1105–1119.
Lee, S. K., & Jin, S. (2006). Decision tree approaches for zero-inflated cont data. *Journal of Applied Statistics*, *33*, 853–865.
Leeb, H., & Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, *21*, 21–59.
Leeb, H., & Pötscher, B. M. (2006). Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics*, *34*(5), 2554–2591.
Leeb, H., & Pötscher, B. M. (2008). Model selection. In T. G. Anderson, R. A. Davis, J.-P. Kreib & T. Mikosch (Eds.), *The handbook of financial time series* (pp. 785–821). New York: Springer.
Lin, Y., & Jeon, Y. (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., & Rubin, D. (2015). *Statistical analysis with missing data* (3rd ed.). New York: Wiley.
Liu, J., Wonka, P., & Ye, J. (2012). Multi-stage Dantzig selector. *Journal of Machine Learning Research*, *13*, 1189–1219.
Loh, W.-L. (2014). Fifty years of classification and regression trees (with discussion). *International Statistical Review*, *82*(3), 329–348.

Loader, C. (2004). Smoothing: Local regression techniques. In J. Gentle, W. Hardle, & Y. Mori (Eds.), *Handbook of computational statistics*. New York: Springer.
Lockhart, R., Taylor, J., Tibshirani, R. J., & Tibshirani, R. (2014). A significance test for the lasso (with discussion). *Annals of Statistics*, *42*(2), 413–468.
Loh, W.-Y. (2002). Regression trees with unbiased variable selection and interaction detection. *Statistica Sinica*, *12*, 361–386.
Ma, Y., & Gao, G. (2014). *Support vector machines applications*. New York: Springer.
Maindonald, J., & Braun, J. (2007). *Data analysis and graphics using R* (2nd ed.). Cambridge, UK: Cambridge University Press.
Madigan, D., Raftery, A. E., Volinsky, C., & Hoeting, J. (1996). Bayesian model averaging. In *AAA Workshop on Integrating Multiple Learned Models* (pp. 77–83). Portland: AAAI Press.
Mallows, C. L. (1973). Some comments on CP. *Technometrics*, *15*(4), 661–675.
Manly, B. F. J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology*. New York: Chapman & Hall.
Mammen, E., & van de Geer, S. (1997). Locally adaptive regression splines. *The Annals of Statistics*, *25*(1), 387–413.
Mannor, S., Meir, R., & Zhang, T. (2002). The consistency of greedy algorithms for classification. In J. Kivensen & R. H. Sloan (Eds.), *COLT 2002*. LNAI (Vol. 2375, pp. 319–333).
Maronna, R., Martin, D., & Yohai, V. (2006). *Robust statistics: Theory and methods*. New York: Wiley.
Marsland, S. (2014). *Machine learning: An algorithmic perspective* (2nd ed.). New York: Chapman & Hall.
Mathlourthi, W., Fredette, M. & Larocque, D. (2015). Regression trees and forests for non-homogeneous poisson processes. *Statistics and Probability Letters*, *96*, 204–211.
McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). New York: Chapman & Hall.
McGonagle, K. A., Schoeni, R. F., Sastry, N., & Freedman, V. A. (2012). The panel study of income dynamics: Overview, recent innovations, and potential for life course research. *Longitudinal and Life Course Studies*, *3*(2), 268–284.
Mease, D., & Wyner, A. J. (2008). Evidence contrary to the statistical view of boosting (with discussion). *Journal of Machine Learning*, *9*, 1–26.
Mease, D., Wyner, A. J., & Buja, A. (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., & Bühlmann, P. (2006). High dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, *34*(3), 1436–1462.
Mentch, L., & Hooker, G. (2015). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. Cornell University Library. arXiv:1404.6473v2 [stat.ML].
Meyer, D., Zeileis, A., & Hornik, K. (2007). The strucplot framework: Visualizing multiway contingency tables with vcd. *Journal of Statistical Software*, *17*(3), 1–48.
Michelucci, P., & Dickinson, J. L. (2016). The power of crowds: Combining human and machines to help tackle increasingly hard problems. *Science*, *351*(6268), 32–33.
Milborrow, S. (2001). rpart.plot: Plot rpart models. An enhanced version of plot.rpart. R Package.
Mitchell, M. (1998). *An introduction to genetic algorithms*. Cambridge: MIT Press.
Moguerza, J. M., & Munõz, A. (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.
Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, *55*, 765–799.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge: MIT Press.
Murrell, P. (2006). *R graphics*. New York: Chapman & Hall/CRC.
Nagin, D. S., & Pepper, J. V. (2012). *Deterrence and the death penalty*. Washington, DC: National Research Council.
Neal, R., & Zhang, J. (2006). High dimensional classification with bayesian neural networks and dirichlet diffusion trees). In I. Guyon, S. Gunn, M. Nikravesh & L. Zadeh (Eds.), *Feature extraction, foundations and applications*. New York: Springer.
Peña, D. (2005). A new statistic for influence in linear regression. *Technometrics*, *47*, 1–12.
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.
Ridgeway, G. (1999). The state of boosting. *Computing Science and Statistics*, *31*, 172–181.
Ridgeway, G. (2012). Generalized boosted models: A guide to the gbm package. Available at from gbm() documentation in R.
Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge, UK: Cambridge University Press.
Rosset, S., & Zhu, J. (2007). Piecewise linear regularized solution paths. *The Annals of Statistics*, *35*(3), 1012–1030.
Rozeboom, W. W. (1960). The fallacy of null-hypothesis significance tests. *Psychological Bulletin*, *57*(5), 416–428.
Rubin, D. B. (1986). Which ifs have causal answers. *Journal of the American Statistical Association*, *81*, 961–962.
Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics*, *2*(3), 808–840.
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., & Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, *22*, 1346–1370.
Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*. Cambridge, UK: Cambridge University Press.
Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
Shakhnarovich, G. (Ed.). (2006). *Nearest-neighbor methods in learning and vision: Theory and practice*. Cambridge, MA: MIT Press.
Schapire, R. E., Freund, Y., Bartlett, P., & Lee, W.-S. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. *The Annals of Statistics*, *26*(5), 1651–1686.
Schapire, R. E. (1999). A brief introduction to boosting. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*.
Schapire, R. E., & Freund, Y. (2012). *Boosting* Cambridge: MIT Press.
Schmidhuber, J. (2014). Deep learning in neural networks: An overview. arXiv:1404.7828v4 [cs.NE].
Schwarz, D. F., König, I. R., & Ziegler, A. (2010). On safari to random jungle: A fast implementation of random forests for high-dimensional data. *Bioinformatics*, *26*(14), 1752–1758.
Scrucca, L. (2014). GA: A package for genetic algorithms in R. *Journal of Statistical Software*, *53*(4), 1–37.
Seligman, M. (2015). Rborist: Extensible, parallelizable implementation of the random forest algorithm. R package version 0.1-0. http://CRAN.R-project.org/package=Rborist.
Sill, M., Heilschher, T., Becker, N., & Zucknick, M. (2014). c060: Extended Inference with lasso and elastic-net regularized cox and generalized linear models. *Journal of Statistical Software*, *62*(5), 1–22
Sutton, R. S., & Barto, A. G. (2016). *Reinforcement learning* (2nd ed.). Cambridge, MA: MIT Press.
Therneau, T. M., & Atkinson, E. J. (2015). An introduction to recursive partitioning using the RPART routines. Technical Report, Mayo Foundation.
Thompson, S. K. (2002). *Sampling* (2nd ed.). New York: Wiley.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, *25*, 267–288.
Tibshirani, R. J. (2015). Adaptive piecewise polynomial estimation via trend filtering. *Annals of Statistics*, *42*(1), 285–323.
Vapnick, V. (1996). *The nature of statistical learning theory*. New York: Springer.
Wager, S. (2014). Asymptotic theory for random forests. Working Paper. arXiv:1405.0352v1.
Wager, E., Hastie, T., & Efron, B. (2014). Confidence intervals for random forests: The jackknife and infinitesimal jackknife. *Journal of Machine Learning Research*, *15*, 1625–1651.
Wager, S., & Walther, G. (2015). Uniform convergence of random forests via adaptive concentration. Working Paper. arXiv:1503.06388v1.
Wahba, G. (2006). Comment. *Statistical Science*, *21*(3), 347–351.
Wang, H., Li, G., & Jiang, F. (2007). Robust regression shrinkage and consistent variable selection through the LAD-lasso. *Journal of Business and Economic Statistics*, *25*(3), 347–355.
White, H. (1980a). Using least squares to approximate unknown regression functions. *International Economic Review*, *21*(1), 149–170.
White, H. (1980b). A heteroskedasticity-consistent covariance matix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838.
Weisberg, S. (2014). *Applied linear regression* (4th ed.). New York: Wiley.
Winham, S. J., Freimuth, R. R., & Beirnacka, J. M. (2103) A weighted random forests approach to improve predictive performance. *Statitical Analysis and Data Mining*, *6*(6), 496–505.
Witten, I. H., & Frank, E. (2000). *Data mining*. New York: Morgan and Kaufmann.
Wood, S. N. (2000). Modeling 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 efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*, 673–686.
Wood, S. N. (2006). *Generalized additive models* New York: Chapman & Hall.
Wright, M. N. & Ziegler, A. (2015). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. arXiv:1508.04409v1 [stat.ML].
Wu, Y., Tjelmeland, H., & West, M. (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 & B. J. Frey (Eds.), *Proceedings of the Ninth Annual Conference on AI and Statistics Jan* (pp. 3–6). Florida: Key West.
Wyner, A. J., Olson, M., Bleich, J., & Mease, D. (2015). Explaining the success of adaboost and random forests as interpolating classifiers. Working Paper. University of Pennsylvania, Department of Statistics.
Xu, B., Huang, J. Z., Williams, G., Wang, Q., & Ye, Y. (2012). Classifying very high dimensional data with random forests build from small subspaces. *International Journal of Data Warehousing and Mining*, *8*(2), 44–63.
Xu, M., & Golay, M. W. (2006). Data-guided model combination by decomposition and aggregation. *Machine Learning*, *63*(1), 43–67.
Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492–514.
Zelterman, D. (2014). A groaning demographic. *Significance*, *11*(5), 38–43.
Zemel, R., Wu, Y., Swersky, K., Pitassi, T., & Dwork, C. (2013). Learning fair representations. *Journal of Machine Learning Research, W & CP*, *28*(3), 325–333.
Zhang, C. (2005). General empirical bayes wavelet methods and exactly adaptive minimax estimation. *The Annals of Statistics*, *33*(1), 54–100.
Zhang, H., & Singer, B. (1999). *Recursive partitioning in the health sciences*. New York: Springer.
Zhang, H., Wang, M., & Chen, X. (2009). Willows: A memory efficient tree and forest construction package. *BMC Bioinformatics*, *10*(1), 130–136.

Ziegler, A., & König, I. R. (2014). Mining data with random forests: Current options for real world applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *4*(1), 55–63.

Zhang, T., & Yu, B. (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., & Hastie, T. (2005). Regularization and variable selection via elastic net. *Journal of the Royal Statistical Association, Series B*, *67*(2), 301–320.

Zou, H., Hastie, T., & Tibshirani, R. (2005). Space principal component analysis. *Journal of Computational and Graphical Statistics*, *15*, 265–286.

Index

R.A. Berk, *Statistical Learning from a Regression Perspective*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-44048-4