

REFERENCES

Abrams, B. 2003. "The Pit of Success." <https://blogs.msdn.microsoft.com/brada/2003/10/02/the-pit-of-success/>.

Baggerly, K, and K Coombes. 2009. "Deriving Chemosensitivity from Cell Lines: Forensic Bioinformatics and Reproducible Research in High-Throughput Biology." *The Annals of Applied Statistics* 3 (4): 1309–34.

Bartley, E AND Schliep, M . AND Hanks. 2019. "Identifying and Characterizing Extrapolation in Multivariate Response Data." *PLOS ONE* 14 (December): 1–20.

Biecek, Przemyslaw, and Tomasz Burzykowski. 2021. *Explanatory Model Analysis*. Chapman; Hall/CRC, New York. <https://ema.drwhy.ai/>.

Bohachevsky, I, M Johnson, and M Stein. 1986. "Generalized Simulated Annealing for Function Optimization." *Technometrics* 28 (3): 209–17.

Bolstad, B. 2004. *Low-Level Analysis of High-Density Oligonucleotide Array Data: Background, Normalization and Summarization*. University of California, Berkeley.

Box, GEP, W Hunter, and J Hunter. 2005. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. Wiley.

Bradley, R, and M Terry. 1952. "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons." *Biometrika* 39 (3/4): 324–45.

Breiman, L. 1996a. "Bagging Predictors." *Machine Learning* 24 (2): 123–40.

———. 1996b. "Stacked Regressions." *Machine Learning* 24 (1): 49–64.

———. 2001a. "Random Forests." *Machine Learning* 45 (1): 5–32.

- . 2001b. “Statistical Modeling: The Two Cultures.” *Statistical Science* 16 (3): 199–231.
- Carlson, B. 2012. “Putting Oncology Patients at Risk.” *Biotechnology Healthcare* 9 (3): 17–21.
- Chambers, J. 1998. *Programming with Data: A Guide to the S Language*. Berlin, Heidelberg: Springer-Verlag.
- Chambers, J, and T Hastie, eds. 1992. *Statistical Models in S*. Boca Raton, FL: CRC Press, Inc.
- Claeskens, G. 2016. “Statistical Model Choice.” *Annual Review of Statistics and Its Application* 3: 233–56.
- Cleveland, W. 1979. “Robust Locally Weighted Regression and Smoothing Scatterplots.” *Journal of the American Statistical Association* 74 (368): 829–36.
- Craig–Schapiro, R, M Kuhn, C Xiong, E Pickering, J Liu, T Misko, R Perrin, et al. 2011. “Multiplexed Immunoassay Panel Identifies Novel CSF Biomarkers for Alzheimer’s Disease Diagnosis and Prognosis.” *PLoS ONE* 6 (4): e18850.
- Cybenko, G. 1989. “Approximation by Superpositions of a Sigmoidal Function.” *Mathematics of Control, Signals and Systems* 2 (4): 303–14.
- Danowski, T, J Aarons, J Hydovitz, and J Wingert. 1970. “Utility of Equivocal Glucose Tolerances.” *Diabetes* 19 (7): 524–26.
- Davison, A, and D Hinkley. 1997. *Bootstrap Methods and Their Application*. Vol. 1. Cambridge university press.
- De Cock, D. 2011. “Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project.” *Journal of Statistics Education* 19 (3).
- Dobson, A. 1999. *An Introduction to Generalized Linear Models*. Chapman; Hall: Boca Raton.
- Durrleman, S, and R Simon. 1989. “Flexible Regression Models with Cubic Splines.” *Statistics in Medicine* 8 (5): 551–61.

- Faraway, J. 2016. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. CRC press.
- Fox, J. 2008. *Applied Regression Analysis and Generalized Linear Models*. Second. Thousand Oaks, CA: Sage.
- Frazier, R. 2018. "A Tutorial on Bayesian Optimization." <http://arxiv.org/abs/1807.02811>.
- Freund, Y, and R Schapire. 1997. "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting." *Journal of Computer and System Sciences* 55 (1): 119–39.
- Friedman, J. 1991. "Multivariate Adaptive Regression Splines." *The Annals of Statistics* 19 (1): 1–141.
- . 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics* 29 (5): 1189–1232.
- Friedman, J, T Hastie, and R Tibshirani. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33 (1): 1.
- Geladi, P., and B Kowalski. 1986. "Partial Least-Squares Regression: A Tutorial." *Analytica Chimica Acta* 185: 1–17.
- Gentleman, R, V Carey, W Huber, R Irizarry, and S Dudoit. 2005. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Berlin, Heidelberg: Springer-Verlag.
- Goodfellow, I, Y Bengio, and A Courville. 2016. *Deep Learning*. MIT Press.
- Hand, D, and R Till. 2001. "A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems." *Machine Learning* 45 (August): 171–86.
- Hill, A, P LaPan, Y Li, and S Haney. 2007. "Impact of Image Segmentation on High-Content Screening Data Quality for SK-BR-3 Cells." *BMC Bioinformatics* 8 (1): 340.
- Ho, T. 1995. "Random Decision Forests." In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1:278–82. IEEE.
- Hosmer, D, and Sy Lemeshow. 2000. *Applied Logistic Regression*. New York: John Wiley; Sons.

Hyndman, R, and G Athanasopoulos. 2018. *Forecasting: Principles and Practice*. OTexts.

Ismay, C, and A Kim. 2021. *Statistical Inference via Data Science: A Modern Dive into R and the Tidyverse*. Chapman; Hall/CRC. <https://moderndive.com/>.

Jaworska, J, N Nikolova-Jeliazkova, and T Aldenberg. 2005. "QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review." *Alternatives to Laboratory Animals* 33 (5): 445–59.

Johnson, D, P Eckart, N Alsamadisi, H Noble, C Martin, and R Spicer. 2018. "Polar Auxin Transport Is Implicated in Vessel Differentiation and Spatial Patterning During Secondary Growth in Populus." *American Journal of Botany* 105 (2): 186–96.

Joseph, V, E Gul, and S Ba. 2015. "Maximum Projection Designs for Computer Experiments." *Biometrika* 102 (2): 371–80.

Jungsu, K, D Basak, and D Holtzman. 2009. "The Role of Apolipoprotein E in Alzheimer's Disease." *Neuron* 63 (3): 287–303.

Kerleguer, A., J.-L. Koeck, M. Fabre, P. G  r  me, R. Teyssou, and V. Herv  . 2003. "Use of Equivocal Zone in Interpretation of Results of the Amplified Mycobacterium Tuberculosis Direct Test for Diagnosis of Tuberculosis." *Journal of Clinical Microbiology* 41 (4): 1783–4.

Kirkpatrick, S, D Gelatt, and M Vecchi. 1983. "Optimization by Simulated Annealing." *Science* 220 (4598): 671–80.

Koklu, M, and IA Ozkan. 2020. "Multiclass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques." *Computers and Electronics in Agriculture* 174: 105507.

Krueger, T, D Panknin, and M Braun. 2015. "Fast Cross-Validation via Sequential Testing." *Journal of Machine Learning Research* 16 (33): 1103–55.

Kruschke, J, and T Liddell. 2018. "The Bayesian New Statistics: Hypothesis Testing, Estimation, Meta-Analysis, and Power Analysis from a Bayesian Perspective." *Psychonomic Bulletin and Review* 25 (1): 178–206.

Kuhn, Max. 2014. "Futility Analysis in the Cross-Validation of Machine Learning Models."

<http://arxiv.org/abs/1405.6974>.

Kuhn, M, and K Johnson. 2013. *Applied Predictive Modeling*. Springer.

———. 2020. *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press.

Lambert, D. 1992. "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing." *Technometrics* 34 (1): 1–14.

Littell, R, J Pendergast, and R Natarajan. 2000. "Modelling Covariance Structure in the Analysis of Repeated Measures Data." *Statistics in Medicine* 19 (13): 1793–1819.

Long, J. 1992. "Measures of Sex Differences in Scientific Productivity*." *Social Forces* 71 (1): 159–78.

Lundberg, Scott M., and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–77. NIPS'17. Red Hook, NY, USA: Curran Associates Inc.

Mangiafico, S. 2015. "An R Companion for the Handbook of Biological Statistics."

<https://rcompanion.org/handbook/>.

Maron, O, and A Moore. 1994. "Hoeffding Races: Accelerating Model Selection Search for Classification and Function Approximation." In *Advances in Neural Information Processing Systems*, 59–66.

McCullagh, P, and J Nelder. 1989. *Generalized Linear Models*. London: Chapman; Hall.

McDonald, J. 2009. *Handbook of Biological Statistics*. Sparky House Publishing.

McElreath, R. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC press.

McInnes, L, J Healy, and J Melville. 2020. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction."

McKay, M, R Beckman, and W Conover. 1979. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics* 21 (2): 239–45.

Mingqiang, Y, K Kidiyo, and R Joseph. 2008. "A Survey of Shape Feature Extraction Techniques." In *Pattern Recognition*, edited by PY Yin. Rijeka: IntechOpen. <https://doi.org/10.5772/6237>.

Molnar, Christopher. 2020. *Interpretable Machine Learning*.
<https://christophm.github.io/interpretable-ml-book/>.

Mullahy, J. 1986. "Specification and Testing of Some Modified Count Data Models." *Journal of Econometrics* 33 (3): 341–65.

Netzeva, T, A Worth, T Aldenberg, R Benigni, M Cronin, P Gramatica, J Jaworska, et al. 2005. "Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships: The Report and Recommendations of Ecvam Workshop 52." *Alternatives to Laboratory Animals* 33 (2): 155–73.

Olsson, D, and L Nelson. 1975. "The Nelder-Mead Simplex Procedure for Function Minimization." *Technometrics* 17 (1): 45–51.

Opitz, J, and S Burst. 2019. "Macro F1 and Macro F1." <http://arxiv.org/abs/1911.03347>.

Rasmussen, C, and C Williams. 2006. *Gaussian Processes for Machine Learning*. *Gaussian Processes for Machine Learning*. MIT Press.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.

Santner, T, B Williams, W Notz, and B Williams. 2003. *The Design and Analysis of Computer Experiments*. Springer.

Schmidberger, M, M Morgan, D Eddelbuettel, H Yu, L Tierney, and U Mansmann. 2009. "State of the Art in Parallel Computing with R." *Journal of Statistical Software* 31 (1): 1–27.
<https://www.jstatsoft.org/v031/i01>.

Schulz, E, M Speekenbrink, and A Krause. 2018. "A Tutorial on Gaussian Process Regression: Modelling, Exploring, and Exploiting Functions." *Journal of Mathematical Psychology* 85: 1–16.

Shahriari, B., K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas. 2016. "Taking the Human Out of the Loop: A Review of Bayesian Optimization." *Proceedings of the IEEE* 104 (1): 148–75.

- Shewry, M, and H Wynn. 1987. "Maximum Entropy Sampling." *Journal of Applied Statistics* 14 (2): 165–70.
- Shmueli, G. 2010. "To Explain or to Predict?" *Statistical Science* 25 (3): 289–310.
- Symons, S, and RG Fulcher. 1988. "Determination of Wheat Kernel Morphological Variation by Digital Image Analysis: I. Variation in Eastern Canadian Milling Quality Wheats." *Journal of Cereal Science* 8 (3): 211–18.
- Thomas, R, and D Uminsky. 2020. "The Problem with Metrics Is a Fundamental Problem for Ai." <http://arxiv.org/abs/2002.08512>.
- Van Laarhoven, P, and E Aarts. 1987. "Simulated Annealing." In *Simulated Annealing: Theory and Applications*, 7–15. Springer.
- Wasserstein, R, and N Lazar. 2016. "The ASA Statement on P-Values: Context, Process, and Purpose." *The American Statistician* 70 (2): 129–33.
- Wickham, H, M Averick, J Bryan, W Chang, L McGowan, R François, G Grolemund, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4 (43).
- Wickham, H, and G Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, Inc.
- Wolpert, D. 1992. "Stacked Generalization." *Neural Networks* 5 (2): 241–59.
- Wu, X, and Z Zhou. 2017. "A Unified View of Multi-Label Performance Measures." In *International Conference on Machine Learning*, 3780–8.
- Wundervald, B, A Parnell, and K Domijan. 2020. "Generalizing Gain Penalization for Feature Selection in Tree-Based Models." <http://arxiv.org/abs/2006.07515>.
- Xu, Q, and Y Liang. 2001. "Monte Carlo Cross Validation." *Chemometrics and Intelligent Laboratory Systems* 56 (1): 1–11.
- Yeo, I-K, and R Johnson. 2000. "A New Family of Power Transformations to Improve Normality or Symmetry." *Biometrika* 87 (4): 954–59.

Zeileis, A, C Kleiber, and S Jackman. 2008. "Regression Models for Count Data in R." *Journal of Statistical Software* 27 (8): 1–25. <https://www.jstatsoft.org/v027/i08>.