**Summary of the history of methods discussed in “An Introduction to Statistical Learning with Applications in R”**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Methods | Year of development | Citation of the original paper | Note | Reference |
| Linear Regression | 1886 | Galton, Francis (1886). "Regression Towards Mediocrity in Hereditary Stature". *The Journal of the Anthropological Institute of Great Britain and Ireland.* 15: 246–263. | It is believed that Francis Galton first used the term “regression”, though using least squares to fit lines was performed by by Legendre (1805) and Gauss (1809) | <https://en.wikipedia.org/wiki/Linear_regression> |
| Generalized Linear Models (Logistic Regression) | 1972 | Nelder, John; Wedderburn, Robert (1972). "Generalized Linear Models". *Journal of the Royal Statistical Society. Series A*. 135 (3): 370–384.  The first and classical textbook on GLM is: McCullagh, Peter; Nelder, John (1989). Generalized Linear Models (2nd ed.). Chapman and Hall/CRC. | Nelder and Wedderburn first formulated GLM that unifies this class of models, though logistic regression was used by others before them. | <https://en.wikipedia.org/wiki/Generalized_linear_model> |
| KNN | 1951 | Fix, Evelyn; Hodges, Joseph L. (1951). "*Discriminatory Analysis*”*.* Nonparametric Discrimination: Consistency Properties (Report). USAF School of Aviation Medicine, Randolph Field, Texas. | | <https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm> |
| Linear Discriminant Analysis | 1936 | Fisher, R. A. (1936). "The Use of Multiple Measurements in Taxonomic Problems". *Annals of Eugenics*. 7 (2): 179–188. | R. A. Fisher first proposed LDA for two class problem, but extensions of DA continues even nowadays. | <https://en.wikipedia.org/wiki/Linear_discriminant_analysis> |
| Naïve Bayes | 1700s | Bayes' theorem was named after Thomas Bayes (1702–1761), and it is believed that naive Bayes classifiers was used at his time, though extensions continues even nowadays. | | <https://stats.stackexchange.com/questions/18212/origin-of-the-na%C3%AFve-bayes-classifier> |
| Bootstrap | 1979 | Efron, Bradley (1979). "Bootstrap methods: Another look at the jackknife". *The Annals of Statistics*. 7: 1–26. | | <https://en.wikipedia.org/wiki/Bootstrapping_(statistics)> |
| Ridge Regression | 1970 | Hoerl, Arthur E.; Kennard, Robert W. (1970). "Ridge Regression: Applications to Nonorthogonal Problems". *Technometrics*. 12 (1): 69–82. | The theory was first introduced by Hoerl and Kennard in 1970, though the idea of regularization was introduced by Andrey Tikhonov in 1940s. | <https://en.wikipedia.org/wiki/Ridge_regression> |
| Lasso | 1996 | Tibshirani, Robert (1996). "Regression Shrinkage and Selection via the lasso". *Journal of the Royal Statistical Society. Series B.* 58 (1): 267–88. | The formal framework of Lasso in statistics was due to Tibshirani in 1996, though it was developed independently in geophysics literature in 1986. | <https://en.wikipedia.org/wiki/Lasso_(statistics)> |
| Principle Component Analysis | 1901 | Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space". *Philosophical Magazine*. 2 (11): 559–572. | | <https://en.wikipedia.org/wiki/Principal_component_analysis> |
| Partial Least Square | 1960s | Wold, Herman (1966). "Estimation of principal components and related models by iterative least squares". In *Multivariate Analysis*. New York: Academic Press. pp. 391–420. | | <https://en.wikipedia.org/wiki/Partial_least_squares_regression> |
| Regression Spline | Before 1960s | Ahlberg & Nilson (1967) *The theory of splines and their applications*, Academic Press. | The Ahlberg & Nilson reference is the earliest I found as a systematic treatment of this method. | <https://en.wikipedia.org/wiki/Curve_fitting> |
| Smoothing Spline | Before 1960s | Reinsch, Christian H (1967). "Smoothing by Spline Functions". *Numerische Mathematik*. 10 (3): 177–183. | The Reinsch reference is the earliest I found for this method. | <https://en.wikipedia.org/wiki/Smoothing_spline> |
| Local Regression | 1860s |  | Despite the long history, local regression methods received little attention until the late 1970s. | Clive Loader (1999). “Local Regression and Likelihood”, Springer |
| Generalized Additive Models | 1986 | Hastie, Trevor and Tibshirani, Robert. (1986), Generalized Additive Models, *Statistical Science*, Vol. 1, No 3, 297-318. | |  |
| Decision Trees | 1963 | Morgan, J.N. & Sonquist, J.A. (1963). “Problems in the analysis of survey data, and a proposal”. *J. Amer. Statist. Assoc.*, 58, 415–434. | It is believed that Morgan & Sonquist published the first regression tree algorithm, though there are many extensions later. | Wei-Yin Loh, “Fifty Years of Classification and Regression Trees”, *International Statistical Review* (2014), 82, 3, 329–348. |
| Bagging | 1990s | Breiman, Leo (1996). "Bagging predictors". *Machine Learning*. 24 (2): 123–140. | The same paper was published in 1994 as a Technical Report in Department of Statistics, University of California Berkeley |  |
| Random Forest | 2001 | Breiman L (2001). “Random Forests”. Machine Learning. 45 (1): 5–32. | |  |
| Boosting | 1997 | Yoav Freund and Robert E. Schapire (1997). “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”. *Journal of Computer and System Sciences*, 55(1):119-139. | |  |
| BART | 2010 | Chipman et al (2010). “BART: BAYESIAN ADDITIVE REGRESSION TREES”. *The Annals of Applied Statistics* 4 (1), 266–298. | |  |
| SVM | 1990s | Cortes, Corinna; Vapnik, Vladimir (1995). "Support-vector networks". *Machine Learning*. 20 (3): 273–297. | |  |
| Neural Networks | 1980s | Rumelhart, David; Hinton, Geoffrey; Williams, Ronald. (1986). "Learning representations by back-propagating errors". *Nature*. 323 (6088): 533–536. | Neural networks rose to fame in late 1980s, though the idea can de dated back to 1940s. The Rumelhart reference was for the back-propagation algorithm. | <https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/History/history1.html> |