A Summary of the Methods

This is just my opinion!

Show 25 entries

Search:

Method	Class	Outcome Data	Method of Inference	Numeric Predictions	Classifications	Special Features	Limitations
Linear Regression	Classical Model	Continuous Measures	Coefficients	Yes	Only by applying thresholds	Inferences, confidence intervals for the coefficients, interpretability	Linearity assumption, requires more rows than columns, categorical inputs create additional columns
Logistic Regression	Classical Model	Binary Data	Coefficients, Odds Ratios	Yes	Yes	Inferences, confidence intervals for the coefficients, interpretability	Logistic assumption, requires more rows than columns, categorical inputs create additional columns, Effective sample size based on smallest category
Cox Proportional Hazards Regression	Classical Model	Censored Survival Data	Coefficients, Hazard Ratios	Yes	Only by applying thresholds	Inferences, confidence intervals for the coefficients, interpretability	Linearity assumption, requires more rows than columns, categorical inputs create additional columns, Effective sample size based on smallest category

Method	Class	Outcome Data	Method of Inference	Numeric Predictions	Classifications	Special Features	Limitations
K Nearest Neighbors	Simple Machine Learning	Numeric or Categorical	None	Yes	Yes	Very simple to implement, adapts to any distribution, K parameter can be tuned	High variability, poor predictive performance relative to other methods
Ridge Regression	Regularized Regression	Continuous, Binary, or Censored Survival Data	Coefficients / Odds Ratios / Hazard Ratios	Yes	Only by applying thresholds	Good for inference, Regularization avoids overfitting by shrinking the coefficient estimates, Shrinkage parameter can be tuned	Poor predictive performance relative to other methods
Lasso Regression	Regularized Regression	Continuous, Binary, or Censored Survival Data	Coefficients / Odds Ratios / Hazard Ratios	Yes	Only by applying thresholds	Good for inference, Regularization avoids overfitting by shrinking the coefficient estimates, Shrinkage parameter can be tuned, Lasso can perform variable selection, can work with a large number of predictors relative to the sample size	Poor predictive performance relative to other methods

Method	Class	Outcome Data	Method of Inference	Numeric Predictions	Classifications	Special Features	Limitations
Decision Trees	Tree-based Models	Numeric or Categorical	Tree Diagrams	Yes	Yes	Creates interpretable models for decisions	Results can change drastically with small changes in the data, poor predictive performance
Random Forest	Tree-based Models	Numeric or Categorical	Limited: Variable Importance Measures	Yes	Yes	Excellent predictive performance, highly adaptable, Parameters can be tuned	Slow running time due to bootstrapping. Considered a black box algorithm
Gradient Boosting	Tree-based Models	Numeric or Categorical	Limited	Yes	Yes	Excellent predictive performance, highly adaptable, Parameters can be tuned	The results depend a lot on finding the right combination of the tuning parameters. Considered a black box algorithm
Support Vector Machines	Support Vector Machines	Numeric or Categorical	None	Yes	Yes	Good predictive performance, kernals allow for numerous kinds of classifying boundaries (e.g. radial)	Selection of kernals is improved by understanding the trends, but the data are highly complex. Considered a black box algorithm

Method	Class	Outcome Data	Method of Inference	Numeric Predictions	Classifications	Special Features	Limitations
Neural Networks	Neural Networks	Numeric or Categorical	None	Yes	Yes	Excellent predictive performance, particularly with deep learning technology.	Less direction on the best ways to tune the algorithm. Considered a black box algorithm, even among many machine learning practitioners!

Showing 1 to 11 of 11 entries

Previous 1 Next