

Project Descriptions

Description:

Welcome to the R Machine Learning Projects repository, a comprehensive collection of projects showcasing the application of various machine learning algorithms using the R programming language. This repository serves as a valuable resource for both beginners and experienced practitioners looking to delve into the world of statistical modeling and predictive analytics.

Project Topics:

k-Nearest Neighbors (KNN): In this project on k-Nearest Neighbors (KNN), the model is fitted with varying values of K. The training and test error rates are plotted against K, revealing insights into model performance. The analysis aims to identify the optimal K value, assessing its associated training and test error rates. Additionally, a visual representation of the training data is presented, including the decision boundary for the optimal K, prompting observations and reflections on the model's overall interpretability and performance.

Linear Regression: A comprehensive approach to constructing a reliable linear regression model is the primary focus of this project. It begins with an exploratory analysis of the dataset, followed by the identification of necessary transformations. Subset selection techniques are applied to determine the optimal model, and model assumptions are evaluated to ensure reliability. Finally, the model is utilized for predictive purposes, resulting in a systematic and comprehensive approach to linear regression modeling.

Logistic Regression: In this project, a logistic regression model is constructed, and a summary of estimates of the regression coefficients, the standard errors of the estimates, and 95% confidence intervals of the coefficients is provided. Moreover, the estimated coefficients of at least two predictors are interpreted and Provided the training error rate for the model.

Discriminant Analysis (LDA, QDA): The GitHub project involves implementing LDA and QDA models using training data to visualize the decision boundaries they create. It also includes computing confusion matrices and overall misclassification rates for both training and test datasets. Finally, the performance of LDA and QDA is compared to assess their strengths and weaknesses in terms of classification accuracy and generalization capabilities.

Resampling Methods (LOOCV, Bootstrap): Understand the significance of resampling methods like Leave-One-Out Cross-Validation (LOOCV) and Bootstrap through projects that emphasize model assessment and validation.

Model Selection and Regularization: Explore the nuances of model selection and regularization techniques, including projects that highlight the importance of balancing model complexity and performance.

Clustering: Dive into projects showcasing clustering algorithms, where unsupervised learning techniques are applied to identify patterns and group similar data points.

Dimensionality Reduction Our project explores the use of dimensionality reduction techniques, including Principal Component Analysis (PCA), Principal Component Regression (PCR), and Partial Least Squares Regression (PLS). We identify critical variables, reduce data dimensionality, and model relationships between variables.

Tree-Based Methods: Learn about decision trees, random forests, and gradient boosting through projects that harness the power of tree-based methods for classification and regression tasks.

Support Vector Machines (SVM): Discover the capabilities of Support Vector Machines in projects that cover both linear and non-linear classification problems.