

Applied Supervised Learning with R

Course Summary

Description

Applied Supervised Learning with R perfectly balances theory and exercises. Each module is designed to build on the learnings of the previous module. The course contains multiple activities that use real-life business scenarios for you to practice and apply your new skills in a highly relevant context. Applied Supervised Learning with R will make you a pro at identifying your business problem, selecting the best-supervised machine learning algorithm to solve it, and fine-tuning your model to exactly deliver your needs without overfitting itself.

Objectives

By the end of the course, students will be able to:

- Develop analytical thinking to precisely identify a business problem
- Wrangle data with dplyr, tidyr, and reshape2
- Visualize data with ggplot2
- Validate your supervised machine learning model using the k-fold algorithm
- Optimize hyperparameters with grid and random search and Bayesian optimization
- Deploy your model on AWS Lambda with Plumber
- Improve a model's performance with feature selection and dimensionality reduction

Topics

- R for Advanced Analytics
- Exploratory Analysis of Data
- Introduction to Supervised Learning
- Regression
- Classification
- Feature Selection and Dimensionality Reduction
- Model Improvements
- Model Deployment
- Capstone Project - Based on Research Papers

Audience

This course is specially designed for novice and intermediate data analysts, data scientists, and data engineers who want to explore various methods of supervised machine learning and its various use cases. Some background in statistics, probability, calculus, linear algebra, and programming will help you thoroughly understand and follow the content of this course.

Duration

Three Days

Applied Supervised Learning with R

Course Outline

I. *R for Advanced Analytics*

- A. Working with Real-World Datasets
- B. Reading Data from Various Formats of Data
- C. Data Structures in R
- D. Data Processing and Transformation
- E. The Apply Family of Functions
- F. Data Visualization

II. *Exploratory Analysis of Data*

- A. Univariate Analysis
- B. Bivariate Analysis
- C. Multivariate Analysis
- D. Categorical Dependent and Numeric/Continuous Independent Variables
- E. Categorical Dependent and Categorical Independent Variable

III. *Introduction to Supervised Learning*

- A. Regression and Classification Problems
- B. Machine Learning Workflow
- C. Regression
- D. Classification
- E. Evaluation Metrics

IV. *Regression*

- A. Linear Regression
- B. Model Diagnostics
- C. Quantile Regression
- D. Polynomial Regression
- E. Ridge Regression
- F. Lasso Regression
- G. Elastic Net Regression
- H. Poisson Regression
- I. Cox Proportional-Hazards Regression Model

V. *Classification*

- A. Classification
- B. Techniques for Supervised Learning
- C. Logistic Regression
- D. Evaluating Classification Models
- E. Evaluating Logistic Regression
- F. Decision Trees
- G. XGBoost
- H. Deep Neural Networks

VI. *Feature Selection and Dimensionality Reduction*

- A. Feature Engineering
- B. One-Hot Encoding
- C. Feature Selection
- D. Feature Reduction
- E. Variable Clustering
- F. Linear Discriminant Analysis for Feature Reduction

VII. *Model Improvements*

- A. Bias-Variance Trade-off
- B. Underfitting and Overfitting
- C. Cross-Validation
- D. K-Fold Cross-Validation
- E. Hold-One-Out Validation
- F. Hyperparameter Optimization
- G. Grid Search Optimization
- H. Random Search Optimization
- I. Bayesian Optimization

VIII. *Model Deployment*

- A. Introduction to plumber
- B. Docker
- C. Amazon Web Services
- D. Introducing AWS SageMaker
- E. What is Amazon Lambda?
- F. What is Amazon API Gateway?
- G. Building Serverless ML Applications

IX. *Capstone Project - Based on Research Papers*

- A. The mlr Package
- B. Implementing Multilabel Classifier using the mlr and OpenML Packages
- C. Constructing a Learner
- D. Predictions