UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

StatisticAlps 2024

2324-StatisticAlps

Title

REGRESSION MODELING STRATEGIES
Strategies for good statistical practice, developing accurate predictive models that validate, choosing between statistical models and machine learning, introduction to Bayesian regression modeling, and complete R examples

Teacher(s)

- Frank E Harrell Jr, Department of Biostatistics School of Medicine Vanderbilt University
- Drew Levy, GoodScience, Inc.

Language

English

Short description

The course provides methods for estimating the shape of the relationship between predictors and response checking in a suitable way the assumptions and avoiding overfitting. Methods for data reduction will be introduced to deal with the common case where the number of potential predictors is large in comparison with the number of observations. Methods of model validation (bootstrap and cross–validation) will be covered, as will auxiliary topics such as modeling interaction surfaces, efficiently utilizing partial covariable data by using multiple imputation,

variable selection, overly influential observations, collinearity, and shrinkage. The methods covered will apply to almost any regression model, including ordinary least squares, longitudinal models, logistic regression models, ordinal regression, quantile regression, longitudinal data analysis, and survival models. Statistical models will be contrasted with machine learning.

The course mainly refers to the book “Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis” Springer Series in Statistics, Frank E. Harrell, Jr., Springer International Publishing, 2015

CFU / Hours

Teaching period

1-6 September 2024

Sustainable Development Goals

QUALITY EDUCATION