

The MSPA Modeling Sequence

Introduction

- The modeling sequence provides the backbone of the MSPA program and your predictive analytics education.
- Our modeling sequence is rigorous. We are a *Predictive Analytics* program, not a *Business Analytics* program. If you compare our curriculum to other analytics programs, then you will see that we are more rigorous.
- The modeling courses will have larger work loads, and hence larger time demands than the non-modeling courses. Many students should take the modeling courses one course per quarter.

Modeling Courses

- Predict 410: Regression and Multivariate Analysis (Core - SAS)
- Predict 411: Generalized Linear Models (Core - SAS)
- Predict 413: Time Series and Forecasting (Core - R)
- Predict 422: Practical Machine Learning (Core - R)
- Predict 454: Advanced Modeling Techniques (Elective - R)

Predict 410: Regression and Multivariate Analysis (Core - SAS)

This course develops the foundations of predictive modeling by: introducing the conceptual foundations of regression and multivariate analysis; developing statistical modeling as a process that includes exploratory data analysis, model identification, and model validation; and discussing the difference between the uses of statistical models for statistical inference versus predictive modeling. The high level topics covered in the course include: exploratory data analysis, statistical graphics, linear regression, automated variable selection, principal components analysis, exploratory factor analysis, and cluster analysis. In addition students will be introduced to the SAS statistical software, and its use in data management and statistical modeling.

Predict 411: Generalized Linear Models (Core - SAS)

This course extends linear 'OLS' regression by introducing the concept of Generalized Linear Model 'GLM' regression. The course reviews traditional linear regression as a special case of GLM's, and then continues with logistic regression, poisson regression, and survival analysis. The course is heavily weighted towards practical application with large data sets containing missing values and outliers. It addresses issues of data preparation, model development, model validation, and model deployment.

Predict 413: Time Series and Forecasting (Core - R)

The objective of this course is to cover key analytical techniques used in the analysis and forecasting of time series data. Specific topics include the role of forecasting in organizations, exponential smoothing methods, stationary and non-stationary time series, autocorrelation and partial correlation functions, univariate ARIMA models, seasonal models, Box-Jenkins methodology, regression models with ARIMA errors, transfer function modeling, intervention analysis, and multivariate time series analysis.

Predict 422: Practical Machine Learning (Core - R)

The rapid advancement of computational methods from machine/statistical learning, data mining, and pattern recognition provides unprecedented opportunities for understanding large, complex datasets. This course takes a practical approach to introduce several machine learning methods with business applications in marketing, finance, and other areas. The course aims to provide a practical survey of modern machine learning techniques that can be applied to make informed business decisions: regression and classification methods, resampling methods and model selection, regularization, perceptron and artificial neural networks, tree-based methods, support vector machines and kernel methods, principal components analysis, and clustering methods. At the end of the course students will have a basic understanding of how each of these methods learn from data to find underlying patterns useful for prediction, classification, and exploratory data analysis. Further, each student will learn how to implement machine learning methods in the R statistical programming language for improved decision-making in real business situations. The course format is a combination of textbook readings and lecture slides, R Lab video sessions, and group discussions. Weekly quizzes and programming assignments using R will be used to reinforce both machine learning concepts and practice. The final project will involve students applying multiple machine learning methods to solve a practical business problem in marketing.

Predict 454: Advanced Modeling Techniques (Elective - R)

This course acts as a 'modeling capstone' through the application of the topics covered in the core modeling sequence and their extension to the more advanced topics from modern statistical and machine learning. The high level objectives of this course are to teach students how to build statistical models on larger data sets, and how to build statistical models on a modeling team. To meet these objectives each student will complete a series of individual assignments on the standard predictive modeling problems of regression, binary classification, and multiclass classification; and a larger unstructured team project on a large data set. High level topics typically included in this course are: exploratory data analysis with advanced statistical graphics, classification and regression trees, Random Forest, model complexity and regularization techniques such as the LASSO, boosting, model ensembles, neural networks, and genetic algorithms. The exact set of topics covered could vary from course to course or instructor to instructor, but the topics covered should be clearly interpreted by the student from the assigned readings.

Understanding Courses By Their Books

Check the course description against the course books. The course books are more meaningful than the course description.

<http://www.stat.rice.edu/~bhatticr/CourseBooks.shtml>