Supply Chain Analytics I: Predictive Analytics

Updated on June 21, 2024

1 Course description

Today's supply chains are complex and create a vast amount of data. Companies face the challenge of extracting information out of their data to improve their operations and to gain a competitive advantage. Supply Chain Analytics I covers descriptive and predictive analytics. You will learn how to identify patterns and to make accurate predictions and inferences. The course covers core concepts of statistical or machine learning and its application using Python.

The course consists of lectures and a project. In the project, the concepts that are covered in the lectures are applied to a data set. It prepares students to identify improvement opportunities that exist in actual supply chains, to quantify possible improvements, and to control supply chain performance.

2 Administration

Class times	Monday, 14:00 - 15:30 (Hoersaal XVIII) Monday 16:00 - 17:30 (Hoersaal XVIII) Wednesday 14:00 - 15:30(Hoersaal XVIII)
KLIPS-Number	14271.0901 (14271.5009 for incomings)
ECTS	6
Short title	Predictive Analytics
Faculty	Nicolas Fugger Mail: fugger@wiso.uni-koeln.de
Teaching assistants	Dylan Gellert Mail: dylan.gellert@uni-koeln.de
Secretary	Stephanie Rauscher Mail: rauscher@wiso.uni-koeln.de
Assignments	There are 5 homework assignments covering the theoretical concepts. Students can earn 1 bonus point per assignment by uploading complete solutions to Ilias on time.
Grading	Project 40%, Exam 60%
Language	English
Prerequisites	Basic knowledge in statistics and Python. The necessary knowledge about Python can, e.g., be acquired in the free online course Python Basics for Data Science.
Project submission	...
Exam	... (cheat sheet, handwritten, two-pages)

3 Session overview

3.1 Self-learning phases

During the self-learning phases, you should read and understand the chapters of the book An Introduction to Statistical Learning. You will apply this knowledge in the Assignments. You do not need to read chapters in parantheses.

Statistical Learning

- Chapters: 2.1, 2.1.1, 2.1.2, (2.1.3), (2.1.4), 2.1.5, 2.2, 2.2.1, 2.2.2, 2.2.3
- Terms and concepts to know: input variables: predictors, independent variables, features, variables; output variable: response, dependent variable; prediction vs. inference; training data vs. test data; parametric vs. non-parametric; overfitting, noise, flexibility/degrees of free-

dom; regression vs. classification; quantitative, qualitative/categorical, and indicator variables; mean squared error; training error vs. test error; cross-validation; bias-variance trade-off*;Bayes classifier, Bayes decision boundary, conditional probability; K-nearest neighbors; error rate; ordinary least squares; reducible error vs. irreducible error;

Linear Regression

- Chapters: 3.1, 3.1.1, 3.1.2, 3.1.3, 3.2, 3.2.1, 3.2.2, 3.3, 3.3.1, 3.3.2. 3.3.3, (3.4), (3.5)

- Terms and concepts to know: simple linear regression; intercept, slope, coefficient, parameter; residual sum of squares (RSS), total sum of squares (TSS), and R^2; residual standard error $RSE = \sqrt{RSS/(n-p-1)}$; residual plot; multiple linear regression; interaction effect, main effect, hierarchical principle; prediction interval; dummy variable; baseline; confidence interval; hypothesis test, null hypothesis, alternative hypothesis; t-statistic; p-value; F-statistic $F = \frac{(TSS-RSS)/p}{RSS/(n-p-1)}$; forward, backward, and mixed selection; polynomial regression; correlation of error terms; non-constant variance of error terms or heteroscedasticity; outliers, studentized residual; high-leverage points, leverage statistic; collinearity, multi-collinearity, variance inflation factor; standard error $Var(\hat{\mu}) = SE(\hat{\mu})^2 = \sigma^2/n$

Classification

- Chapters: 4.1, 4.2, 4.3, 4.3.1, 4.3.2, 4.3.3, 4.3.4, (4.3.5), 4.4, 4.4.1, (4.4.2, only: confusion matrix, sensitivity, specificity, ROC curve), 4.4.3 (only difference to LDA for $p = 1$), 4.4.4, (4.5), (4.5.1), (4.5.2), (4.6), (4.6.1), (4.6.2), (4.6.3)

- Terms and concepts to know: linear probability model; logistic model, logit; maximum likelihood, likelihood function; odds; z-statistic; prior, posterior, Bayes' theorem; linear discriminant analysis (LDA), quadratic discriminant analysis (QDA); density function, normal distribution; naive bayes, marginal distribution, joint distribution

Resampling Methods, Linear Model Selection, and Regularization

- Chapters: 5.1, 5.1.1, 5.1.2 (computational aspects not relevant), 5.1.3, 5.1.4, 5.1.5, 5.2, 6.1, 6.1.1, 6.1.2, 6.1.3, 6.2, 6.2.1, 6.2.2 (except Bayesian Interpreation pp. 248), 6.2.3, (6.3), (6.3.1), (6.3.2), (6.4), (6.4.1), (6.4.2), (6.4.3), (6.4.4)

- Terms and concepts to know: model assessment; cross-validation; validation set, hold-out set; leave-one-out cross-validation; k-fold cross-validation; bootstrap; best subset selection, forward stepwise selection, backward stepwise selection; deviance; C_p, AIC, BIC (basic ideas), adjusted R^2; one-standard error rule; shrinkage methods, tuning parameter, shrinkage penalty; ridge regression; scale equivariant, standardizing the predictors; lasso

Tree-Based Methods

- Chapters: 8.1, 8.1.1, 8.1.2, 8.1.3, 8.1.4, 8.2, 8.2.1, 8.2.2, 8.2.3, (8.2.4), 8.2.5 (except BART)
- Terms and concepts to know: decision tree, regression tree, classification tree; terminal node, leaf, internal node, branch; recursive binary splitting; cost complexity or weakest link pruning; error rate, Gini index, entropy; ensemble method; bagging; majority vote; variable importance; random forest; boosting

3.2 Concept sessions

The concepts sessions will take place on Mondays. In these sessions, students will presents their solutions to the assignments and we will discuss related topics. *Note: For a better learning experience, we will split the concept sessions into two groups from week 2 onward. The assignment to one of the two groups will be made in the first week.*

3.3 Python sessions

The Python sessions will take place on Wednesdays. In the first part of these sessions, we will jointly solve Python exercises. In the second part of these sessions, you will work on exercises in groups. In these sessions, you will learn how to apply the concepts using Python and you will learn how to master the project.

3.4 Group sessions

We recommend to prepare your assignments and your project in groups of up to 3 members.

4 Python support

If you have difficulties getting Python to work properly, please contact us after the sessions. You can also post your questions in the discussion forum on Ilias.

5 Submissions

5.1 Assignments

To earn bonus points you have to upload your complete solutions to Ilias on time. Assignments on the theoretical concepts are due Fridays at 18:00. By uploading your solutions you agree to present the solutions in the plenum. If you are (randomly) selected, you will only get bonus points if you present your solutions. The presentation of the solutions will not be graded. For each assignment you can get 1 bonus point. You can prepare your solutions in groups of up to 3 members but each student has to upload her/his own solution individually. Prepare slides with your solutions and upload them as a pdf-file. You can write down your answers in note form. Templates for Latex and PowerPoint are provided at Ilias.

Note: If you are unable to attend a session, but still want to receive bonus points for the assignment, you need to additionally upload a video of yourself presenting your solutions by Monday, 11:00 to Ilias. We recommend recording a Zoom meeting in which you share your solution slides.

5.2 Project

During the semester you will work on a project. You can find the data set and a Jupyter Notebook including a detailed description of your task at Ilias. You can work on and hand in the project in groups of up to 3 members. Save your solution and upload it to Ilias. *Note: Most of the tools and commands needed to complete the project will be discussed in the Python sessions. However, programming always implies browsing documentations and searching via Google.*

6 Resources

6.1 Concepts

6.1.1 Core

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning. New York: Springer

6.1.2 Additional

- In addition to reading the book you can watch: Online lectures based on the first edition of the book on edX
- If you want to check your understanding of the book and practice, you can find sample questions with solutions here: JakobGM

- If you are interested in the mathematical background of the concepts, chekc out: Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. New York: Springer

6.2 Programming

- We recommend downloading and installing Anaconda for running Python.
- We provide notebooks for the Python sessions and the project. You can find out more about it at Jupyter.
- You can find additional information about the Python libraries used in this course at: pandas, statsmodels, NumPy, scikit-learn, graphviz, seaborn, matplotlib