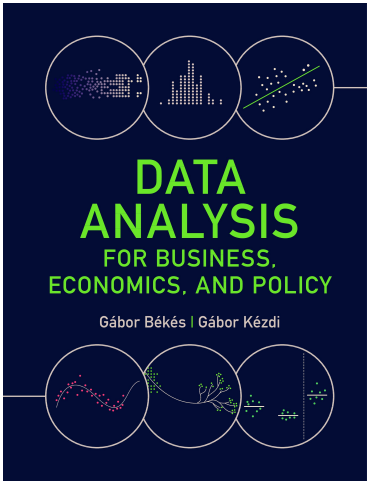
Intro to DA3

Gabor Bekes

Data Analysis 3: Prediction and Introduction to Machine Learning

2021

Slideshow for the Békés-Kézdi Data Analysis textbook



- Cambridge University Press, 2021 April
- Available in paperback, hardcover and e-book
- **gabors-data-analysis.com**
  - Download all data and code `https://gabors-data-analysis.com/data-and-code/`
- Data Analysis 3 covers **Chapter 13-18**

Data Analysis 3

1. Framework for prediction (prediction error, loss function, RMSE, prediction with regression, overfitting, cross-validation)
2. Model building and selection (process, feature and label engineering, LASSO)
3. Regression trees (CART, stopping rules, pruning, search algorithms, regression vs CART)
4. Random forest (boosting, decorrelating trees, regression vs random forest) and GBM
5. Probability prediction and classification (threshold selection, ROC/AUC, classification with logit vs. random forest)
6. Forecasting from time series data (serial correlation, cross-validation in time series, ARIMA, vector autoregression)

Course material

Material

- Four(!) weeks - 6 chapters
- Lots of material, data and code to work through.

Seminars and practice

- Great team:
- Seminars by (Jenő Pál, Emarsys / Phd CEU)
- TA: Máté Tóth (VP at Blackrock / Phd UPV)
- Office hours offered by us all

Assessment

- In class Quizzes (15p)
  - Beginning of class: Past lecture material
  - In class -
  - 17 quiz planned, 1p/quiz, max is 15p
- Assignments (85p)
  - 3 Assignments
  - 1 assignment may done in a pair of your choice
- Extra assignments for extra 5p

- No exam.
- To pass, you will need to get at least 50% of the overall grade.