Introduction

AU STAT-427/627

Emil Hvitfeldt

2021-08-30

Welcome

- Introductions

- Syllabus

- Material

- Questions

Questions

You are encouraged to ask questions when you have them rather than wait for me to ask for questions. If you have a question, chances are that something else has a question.

Attendence and Camera

Both highly encouraged

Due to COVID-19 both should be practiced according to what makes you safe

About Me

- Data Analyst at Teladoc Health

- R package developer, about 10 packages on CRAN (textrecipes, themis, paletteer, prismatic, textdata)

- Co-author of "Supervised Machine Learning for Text Analysis in R"

Statistical Machine Learning

American University 427/627

This website contains most of the information and material that will be used for one of the sections of the course Statistical Machine Learning at American University.

The navigation bar contains information about the syllabus, schedule, readings, labs and assignments.

License

This online work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International. Visit here for more information about the license.

Colophon

This book was written in RStudio using distill. The complete source is available on GitHub. All packages versions are being handled with renv.

Syllabus

An Introduction to Statistical Learning with Applications in R

Will be our main textbook

We will cover most of the material, and follow it mostly chronologically

Tidymodels labs for ISLR

https://emilhvitfeldt.github.io/ISLR-tidymodels-labs/index.html

The Elements of Statistical Learning: Data Mining, Inference, and Prediction

Supplementary textbook

I Will sometimes refer to this book when more detail is needed

Tidy Modeling with R

I Will sometimes refer to this book when more detail is needed

Syllabus

Come to me before it is too late

I'm here to help, my main goal for this course is to make you succeed

Late assignment

There are some (limited) late penalties

It is more important for me that you turn something in than that you give up. You will always get points (sometimes reduced) for late assignments

Contact me if you are having a hard time or need to turn in late

Lecture

Mondays

Focused on intuition, concepts, and statistics

Labs

Thursdays

A hands-on section where we work together on the implementation side in R

These should be turned in WITH explanatory text.

Assignments

There will be 10 assignments

It Will contain a mix of conceptual questions and practical coding exercises about the weekly topic

Midterm

Be similar to an assignment

Final Project
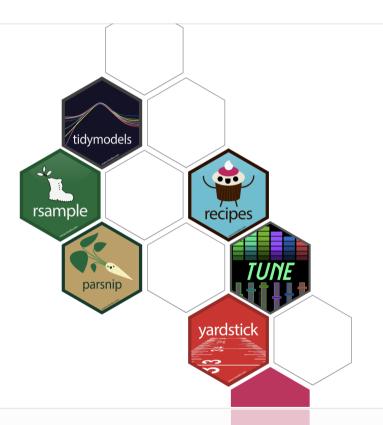
We end the class with a final machine learning project

You will find a data set and analyze it with the tools you have learned in the class

The project will be a document(25%) and a presentation given to the class(5%)

Data will have to be approved by me

TIDYMODELS

The tidymodels framework is a collection of packages for modeling and machine learning using **tidyverse** principles.

Install tidymodels with:

```
install.packages("tidymodels")
```

tidymodels feedback

Any and all feedback regarding the use of {tidymodels} is appreciated

Both how I am teaching it and how it is to use

Material 1/2

- Introduction, motivation, and examples. Understanding large and complex data sets. Statistical learning. First steps in R. [Chap. 1-2].

- Review of regression modeling and analysis; implementation in R. [Chap. 3].

- Classification problems and classification tools. Logistic regression and review of linear discriminant analysis. [Chap. 4]

- Resampling methods; bootstrap. [Chap. 5 and lecture notes].

Material 2/2

- High-dimensional data and shrinkage. Ridge regression. LASSO. Model selection methods and dimension reduction. Principal components. Partial least squares. [Chap. 6]

- Nonlinear trends and splines. [Chap. 7; 7.4-7.5]

- Regression trees and decision trees [Chap. 8]

- Introduction to support vector machines [Chap. 9]

- Clustering methods [Chap. 12]

What you will learn

- How the foundational Machine learning models work

- The intuition behind them

- How to use them

- Using {tidymodels}

What you won't

- How they are coded

- State of the art (SOTA) methods

- Deep learning

- Survival Analysis

Questions?