Methods 3: Multilevel Statistical Modeling and Machine Learning

true

Autumn 2023

Contents

Course repository: https://github.com/msaintemarie/methods3_a2023

Description

The purpose of the course is to further develop students' knowledge and skills in statistics and programming.

The course includes instruction on statistical techniques relying on the Generalised Linear Model, including multilevel modelling and basic machine learning procedures (e.g. predictive approaches and validation techniques). The course includes lectures on statistical and machine learning concepts as well as practical exercises in programming.

This course builds on the Methods 1 and 2 courses and provides skills and knowledge needed to understand and apply a wider variety of statistical and machine learning methods. The course prepares students for Bayesian computational modelling in the Methods 4 course, and provides techniques for data analysis relevant to exam projects in Social and Cultural Dynamics in Cognition, and in the Bachelor thesis project.

Objectives

Knowledge

In the evaluation of the student's performance, emphasis is placed on the extent to which the student is able to

- demonstrate understanding of statistical techniques relying on the Generalised Linear Model
- demonstrate understanding of hierarchical modeling methods
- demonstrate understanding of basic machine learning concepts.

Skills

In the evaluation of the student's performance, emphasis is placed on the extent to which the student is able to:

- build and evaluate models of hierarchically structured data
- integrate machine learning procedures in data analysis
- communicate analysis processes, results and interpretation.

Competences

In the evaluation of the student's performance, emphasis is placed on the extent to which the student is able to:

- independently decide on data analysis methods, given a data set and a research question
- justify decision making when pre-processing messy data for data analysis.

Format

Lectures are generally accompanied by readings and other resources. All key content is covered in the lecture. However, reading the accompanying materials is a very good idea as the contents are covered in more details, or from different angles, and in general the readings give you the chance to think through the topic and identify your doubts, which you can then bring up in class or on alternative channels.

Lectures should be as interactive as possible, and any question is welcome at any time during class.

> There are naïve questions, tedious questions, ill-phrased questions, questions put after inadequate self-criticism. But every question is a cry to understand the world. There is no such thing as a dumb question (Sagan C (2011). *The Demon-Haunted World: Science as a Candle in the, Dark*. Ballantine Books.)

Alternatively, you can also ask a question anonymously and in writing on a Google Document that will be made available to you for each class. You can write questions there before, during, and after lectures. I'll review it from time to time during and between classes to answer them as adequately as I can, in incremental and iterative fashion.

Evaluation

The course is aimed at giving you a conceptual and hands-on understanding of estimation and generalizability issues, together with data processing and coding skills. Accordingly, the exam consists in 4 portfolio assignments (indicated as A1 to A4 in the course schedule) challenging you to apply your understanding your understanding and programming skills.

RMarkdown files with instructions for each assignment will be made available online around the time of the relevant class. Both group and one-to-one support from the teacher and the instructors will only be provided during classroom hours. Each assignment will consist of a Github/Gitlab repository with all the code, and a document containing the answers to the questions assigned. You are strongly recommended to submit the assignments during the course, so you can get collective feedback. You must re-submit the whole portfolio - revised according to the feedback - for the final exam. You can submit the full portfolio even if you haven't managed to submit each assignment during the course.

Each assignment can be done alone or in groups. The latter is however strongly recommended, for various reasons. Most obviously, teamwork creates opportunities to establish connections and relationships that can be beneficial to you in the longer term, as it will offer you the opportunity to hone crucial collaboration skills such as communication, leadership, conflict resolution, time management, and resource allocation. But also, the pooling of perspectives, backgrounds, knowledge, skills, working styles, and resources made possible by teamwork will lead to a deeper understanding of the subject matter as well as more comprehensive, exhaustive, and creative work output for you and everyone else involved. Finally, shared responsibility and accountability can increase motivation and dedication, and getting constructive feedback from your teammates can both be a learning and confidence-boosting experience for you. Of course, problems specific to teamwork may arise during this course (communication issues, conflicting schedules, unequal contributions…). While we recommend to tackle these sooner than later, it is important to stress that learning to navigate and overcome these challenges is a valuable aspect of teamwork itself.

Schedule

Students are expected to read mandatory lectures before each session. **Important: class readings are subject to change, contingent on mitigating circumstances as well as the progress we make as a class**. Attending lectures and regularly checking the course website for updates is thus strongly recommended.

Mandatory readings are written in **bold** characters

#Week	Date	Subject	Readings	Slides	Assignment	Tutor
1	Sept 5 & Sept 7	Introduction to the course Introduction to Git	**1**,2,3	W1	A1	LWL
2	Sept 12 & Sept 14	Set Theory Probabilities		W2	A1	MHSM
3	Sept 19	Bayesian Analysis	**4**,**5**,6,7	W3	A1	
4	Sept 26 & Sept 28	Multilevel Linear Models		W4	A2	LWL
5	Oct 3 & Oct 5	Explanation vs. Prediction	**8**,9,10	W5	A2	MHSM
6	Oct 10 & Oct 12	Physiological Data & Advanced MLMs	**11**,**12**,13	W6	A3	LWL
7	Oct 24 & Oct 26	Advanced MLMs		W7	A3	MHSM
8	Oct 21 & Nov 2	Prediction Feature Selection	**14**	W8	A3	LWL
9	Nov 7 & Nov 9	Cumulative Approaches Midway Evaluation	**15**,16,17,18	W9	A4	MHSM
10	Nov 14 & Nov 16	Classification problems	19	W10	A4	LWL
11	Nov 21 & Nov 23	Generalized Multilevel Models & Classification		W11	A4	MHSM
12	Nov 28 & Nov 30	Machine Learning Pipelines I		W12	A4	LWL
13	Déc 5 & Déc 7	Machine Learning Pipelines II		W13	A4	MHSM

References

1. Dienes Z. Chapter 3: Neyman, pearson and hypothesis testing. In: Understanding psychology as a science: An introduction to scientific and statistical inference. Springer; 2008.
2. Vasishth S, Nicenboim B. Statistical methods for linguistic research: Foundational ideas–part I. Language and Linguistics Compass. 2016;108:349–69.
3. Wickham H, Çetinkaya-Rundel M, Grolemund G. R for data science. O'Reilly; 2023.
4. DeBruine LM, Barr DJ. Understanding mixed-effects models through data simulation. Advances in Methods and Practices in Psychological Science. 2021;41:2515245920965119.
5. Fusaroli R, Weed E, Fein D, Naigles L. Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. Cognition. 2019;183:1–18.
6. Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CE, et al. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ. 2018;6:e4794.
7. Winter B. Statistics for linguists: An introduction using r. In Routledge; 2019.
8. Rocca R, Yarkoni T. Putting psychology to the test: Rethinking model evaluation through benchmarking and prediction. Advances in Methods and Practices in Psychological Science. 2021;43:25152459211026864.

9. Chekroud A. Why validation matters: A demonstration predicting antipsychotic response using 5 rcts. Schizophrenia Bulletin. 2018;44Suppl 1:S157.
10. Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science. 2001;163:199–231.
11. Fusaroli R, Bjørndahl JS, Roepstorff A, Tylén K. A heart for interaction: Shared physiological dynamics and behavioral coordination in a collective, creative construction task. Journal of Experimental Psychology: Human Perception and Performance. 2016;429:1297.
12. Ferrer E, Helm JL. Dynamical systems modeling of physiological coregulation in dyadic interactions. International journal of psychophysiology. 2013;883:296–308.
13. Konvalinka I, Sebanz N, Knoblich G. The role of reciprocity in dynamic interpersonal coordination of physiological rhythms. Cognition. 2023;230:105307.
14. Kuhn M, Johnson K. Feature engineering and selection: A practical approach for predictive models. In: Chapter 10. Boca Raton, FL: Chapman; Hall/CRC; 2019.
15. Parola A, Simonsen A, Bliksted V, Fusaroli R. Voice patterns in schizophrenia: A systematic review and bayesian meta-analysis. Schizophrenia research. 2020;216:24–40.
16. Parola A, Simonsen A, Lin JM, Zhou Y, Wang H, Ubukata S, et al. Voice patterns as markers of schizophrenia: Building a cumulative generalizable approach via a cross-linguistic and meta-analysis based investigation. Schizophrenia Bulletin. 2023;49Supplement_2:S125–41.
17. Grames EM, Schwartz D, Elphick CS. A systematic method for hypothesis synthesis and conceptual model development. Methods in Ecology and Evolution. 2022;139:2078–87.
18. Harrer M, Cuijpers P, Furukawa T, Ebert D. Doing meta-analysis with r: A hands-on guide. New York: Chapman; Hall/CRC; 2021.
19. Kuhn M, Johnson K. Chapter 11: Measuring performance in classification problems. In: Feature engineering and selection: A practical approach for predictive models. Boca Raton, FL: Chapman; Hall/CRC; 2019.