Syllabus

Methods 3: Multilevel Statistical Modeling and Machine Learning

The syllabus contains:

1. The lesson plan
2. Exam details
3. The reading list for each lesson and practical exercise

Lesson plan

Classes: Tuesdays 11.00-13.00

Instructors: Emil Trenckner Jessen and Lau Møller Andersen

Instructor sessions: Wednesdays 10.00-12.00 and 16.00-18.00

Languages: *R* (linear mixed effects modelling) and *Python* (machine learning)

Suggested Integrated Development Environments (IDEs): *RStudio;* and *Spyder* installed through *Conda* (a package and environment management system)

- Week 1: *Introduction and Why are we here?*, September 14 & 15
    - Instructor sessions: *Setting up* R *and* Python *and recollection of the general linear model*
- Week 2: *Linear Mixed Effects Models*, September 21 & 22
    - Instructor sessions: *Modelling random effects – and how do they differ from fixed effects?*
- Week 3: *Generalized Linear Mixed Effects Models*, September 28 & 29
    - Instructor sessions: *What to do when the response variable is not continuous?*
- Week 4: *Explanation and Prediction*, October 5 & 6
    - Instructor sessions: *Why are good explanations sometimes bad?*
- Week 5: *Evaluating and comparing models,* October 12 & 13
    - Instructor sessions: *How do we assess how models compare to one another?*
- Week 6: *Mid-way evaluation and Machine Learning Intro,* November 2 & 3
    - Instructor sessions: *Moving the goal away from explanation towards prediction and getting* Python *running*
- Week 7: *Linear regression revisited (machine learning),* November 9 & 10

    - Instructor sessions: *How to constrain our models to make them more predictive*
- Week 8: *Logistic regression (machine learning)*, November 16 & 17
    - Instructor sessions: *Categorizing responses based on informed guesses*
- Week 9: *Dimensionality Reduction, Principled Component Analysis (PCA)*, November 23 & 24
    - Instructor sessions: *What to do with very rich data?*
- Week 10: *Organising and preprocessing messy data* November 30 and December 1
    - Instructor sessions: *How to clean up?*
- Week 11: *Final evaluation and wrap-up of course*, December 7 & 8
    - Instructor sessions: *Ask anything!*

Exam details

(Underlined text below has been copied from the Academic Regulations)

Aid:

Not defined

Assessment:

Passed /failed

Grading:

Internal co-examination

Notes

Ordinary examination:
The exam consists of a portfolio containing a number of assignments. The total length of the portfolio is: 3-7 assignments.

Their form and length will be announced on Brightspace by the teacher at the start of the semester. The portfolio may include products. Depending on their length, and subject to the teacher's approval, these products can replace some of the standard pages in the portfolio.

It must be possible to carry out an individual assessment. So if some parts of the portfolio have been produced by a group, it must be stated clearly which parts each student is responsible for, and which parts the group as a whole is responsible for.

The complete portfolio must be submitted for assessment in the Digital Exam system by the date specified in the exam plan. Each student submits a portfolio.

Final portfolio:

Revised assignments, handed in as short reports with reproducible code on GitHub

Assignment 1: Using mixed effects modelling to model hierarchical data (Winter & Grawunder, 2012)

Assignment 2: Mixed effects modelling of response times, response counts, and accuracy (Andersen et al., 2019)

Assignment 3: Using logistic regression to classify subjective experience from brain data

Assignment 4: Dimensionality reduction; finding the signal among the noise

Re-examination:
The exam consists of a portfolio that contains the same type and number of assignments as the ordinary exam. The total length of the portfolio is: 3-7 assignments.

Their form and length will be announced on Brightspace by the teacher at the start of the semester. The portfolio may include products. Depending on their length, and subject to the teacher's approval, these products can replace some of the standard pages in the portfolio.

It must be possible to carry out an individual assessment. So if some parts of the portfolio have been produced by a group, it must be stated clearly which parts each student is responsible for, and which parts the group as a whole is responsible for.

The complete portfolio must be handed in for assessment in the Digital Exam system by the date specified in the exam plan. Each student submits a portfolio.

Literature

Week 1: *Introduction and Why are we here?*, September 14 & 15

- Readings:
  - *Mandatory*: an introduction to hypothesis testing
    - (Chapter 3) Dienes, Z., 2008. Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference. Palgrave Macmillan, Basingstoke.
  - *Recommended:* these two articles are part of the heated argument (not a good model for disagreement in science) regarding how scientific hypotheses are to be evaluated using statistics
    - Fisher, R., 1955. Statistical Methods and Scientific Induction. Journal of the Royal Statistical Society: Series B (Methodological) 17, 69–78. https://doi.org/10.1111/j.2517-6161.1955.tb00180.x
    - Neyman, J., 1956. Note on an Article by Sir Ronald Fisher. Journal of the Royal Statistical Society: Series B (Methodological) 18, 288–294. https://doi.org/10.1111/j.2517-6161.1956.tb00236.x
  - *Optional:* collaboration highlight the reproducibility problem in science.
    - Collaboration, O.S., 2015. Estimating the reproducibility of psychological science. Science 349. https://doi.org/10.1126/science.aac4716
- Practical exercise (P1): *Setting up* R *and* Python *and recollection of the general linear model*

- *Mandatory*:
  - setup GitHub/GitLab in RStudio: https://happygitwithr.com/ (especially Chapter 12)

Week 2: *Linear Mixed Effects Models*, September 21 & 22

- Readings:
  - *Mandatory:*
    - Winter, B., 2013. Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499 [cs]. https://arxiv.org/abs/1308.5499
  - *Recommended:*
    - Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E.D., Robinson, B.S., Hodgson, D.J., Inger, R., 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ 6, e4794. https://doi.org/10.7717/peerj.4794
  - *Optional:*
    - *(*Reference book, especially chapter 12) Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.
- Practical exercise (P2): *Modelling random effects – and how do they differ from fixed effects?*
  - **Assignment 1**: ***Using mixed effects modelling to model hierarchical data***
    - Background literature: Winter, B., Grawunder, S., 2012. The phonetic profile of Korean formal and informal speech registers. Journal of Phonetics 40, 808–815. https://doi.org/10.1016/j.wocn.2012.08.006

Week 3: *Generalized Linear Mixed Effects Models*, September 28 & 29

- Readings:
  - *Mandatory:*
    - (Chapters 6 & 13) Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.
  - *Recommended:*
  - *Optional:*
- Practical exercise (P3): *What to do when the response variable is not continuous?*
  - **Assignment 2; Part 1 – fitting the models: *Mixed effects modelling of response times, response counts, and accuracy***
    - Background literature: Andersen, L.M., Overgaard, M., Tong, F., 2019. Visual expectations change subjective experience without changing performance. Consciousness and Cognition 71, 59–69. https://doi.org/10.1016/j.concog.2019.03.007

Week 4: *Explanation and Prediction*, October 5 & 6

- Readings:
  - *Mandatory:*
    - Yarkoni, T., Westfall, J., 2017. Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspect Psychol Sci 12, 1100–1122. https://doi.org/10.1177/1745691617693393
  - *Recommended:*
    - Breiman, L., 2001. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). Statistical Science 16, 199–231. https://doi.org/10.1214/ss/1009213726
  - *Optional:*
- Practical exercise (P4): *Why are good explanations sometimes bad?*

Week 5: *Evaluating and comparing models*, October 12 & 13

- Readings:
  - *Mandatory:*
    - Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24, 127–135. https://doi.org/10.1016/j.tree.2008.10.008
  - *Recommended:*
  - *Optional:*
    - (background): Andersen, L.M., Overgaard, M., Tong, F., 2019. Visual expectations change subjective experience without changing performance. Consciousness and Cognition 71, 59–69. https://doi.org/10.1016/j.concog.2019.03.007
- Practical exercise (P5): *How do we assess how models compare to one another?*
  - **Assignment 2; Part 2 – evaluating and comparing the models: *Mixed effects modelling of response times, response counts, and accuracy***

Week 6: *Mid-way evaluation and Machine Learning Intro*, November 2 & 3

- Readings:
  - *Mandatory:*
    - Chapters 1 & 2: Raschka, S., 2015. Python Machine Learning. Packt Publishing Ltd.
  - *Recommended:*
    - Chapter 3 (detailed reading not necessary): Raschka, S., 2015. Python Machine Learning. Packt Publishing Ltd.
  - *Optional:*
- Practical exercise (P6): *Moving the goal away from explanation towards prediction and getting* Python *running*

Week 7: *Linear regression revisited (machine learning)*, November 9 & 10

- Readings:
  - *Mandatory:*
    - Chapter 10: Raschka, S., 2015. Python Machine Learning. Packt Publishing Ltd.
  - *Recommended:*
    - Have a look at the examples in section 1.1 https://scikit-learn.org/stable/supervised_learning.html#supervised-learning
  - *Optional*
- Practical exercise (P7): *How to constrain our models to make them more predictive*

Week 8: *Logistic regression (machine learning)*, November 16 & 17

- Readings:
  - *Mandatory:*
    - Chapter 3: Raschka, S., 2015. Python Machine Learning. Packt Publishing Ltd.
  - *Recommended:*
    - (Facilitating the time dimension): King, J.-R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends in Cognitive Sciences 18, 203–210. https://doi.org/10.1016/j.tics.2014.01.002
  - *Optional:*
    - (an argument for why machine learning may be useful for studies of awareness) Sandberg, K., Andersen, L.M., Overgaard, M., 2014. Using multivariate decoding to go beyond contrastive analyses in consciousness research. Front. Psychol. 5, 1250. https://doi.org/10.3389/fpsyg.2014.01250
- Practical exercise (P8): *Categorizing responses based on informed guesses*
  - **Assignment 3: Using logistic regression to classify subjective experience from brain data**

Week 9: *Dimensionality Reduction, Principled Component Analysis (PCA)*, November 23 & 24

- Readings:
  - *Mandatory:*
    - Chapter 5: Raschka, S., 2015. Python Machine Learning. Packt Publishing Ltd.
  - *Recommended:*
    - Shlens, J., 2014. A Tutorial on Principal Component Analysis. arXiv:1404.1100 [cs, stat]. https://arxiv.org/abs/1404.1100
  - *Optional:*
    - (Focus on algorithm, i.e. *Introduction* and *Materials and Methods*) Veraart, J., Novikov, D.S., Christiaens, D., Ades-aron, B., Sijbers, J., Fieremans, E., 2016.

Denoising of diffusion MRI using random matrix theory. NeuroImage 142, 394–406. https://doi.org/10.1016/j.neuroimage.2016.08.016

- Practical exercise (P9): *What to do with very rich data?*
  - **Assignment 4: Dimensionality reduction; finding the signal among the noise**

Week 10: *Organising and preprocessing messy data,* November 30 and December 1

- Readings:
  - *Mandatory:*
    - Chapter 4: Raschka, S., 2015. Python Machine Learning. Packt Publishing Ltd.
  - *Recommended:*
  - *Optional:*
    - Reference book: R for Data Science http://r4ds.had.co.nz/
- Practical exercise (P10): *How to clean up?*

Week 11: *Final evaluation and wrap-up of course*, December 7 & 8

- Readings:
  - *Mandatory:* None
  - *Recommended:*
    - Getting ready for Methods 4: Wagenmakers, E.-J., 2007. A practical solution to the pervasive problems of *p* values. Psychonomic Bulletin & Review 14, 779–804. https://doi.org/10.3758/BF03194105
  - *Optional:*
    - have a look at Andrew Gelman's blog: https://statmodeling.stat.columbia.edu/2016/03/07/29212/
- Practical exercise (P11): *Ask anything!*