Practical Course: „Machine Learning for Model Building in the Sciences“

Summer term 2016

Niels Landwehr, Tobias Scheffer

Organization: Overview

- Practical course worth 6 „Leistungspunkte“.
- Main contact: Niels Landwehr, Office 03.04.0.13, landwehr@cs.uni-potsdam.de.
- Webpage for course: http://www.cs.uni-potsdam.de/ml/teaching/ss16/pmlnm.html.
- Today: overview of format of course, brief sketch of possible topics.

Organization: Format

- Practical course covers topics in machine learning in the sciences.
- Significant prior knowledge in machine learning (at least two lectures) is required!

- Format: You get an individual project that you work on autonomously
  - Literature research, starting from a few scientific papers that we provide.
  - Familiarize yourself with the scientific application domain (data, literature).
  - Develop and implement a machine learning model.
  - Conduct experimental studies based on data from the scientific domain.
  - Detailed written report.
  - Defend your project in exam (usually with short presentation).

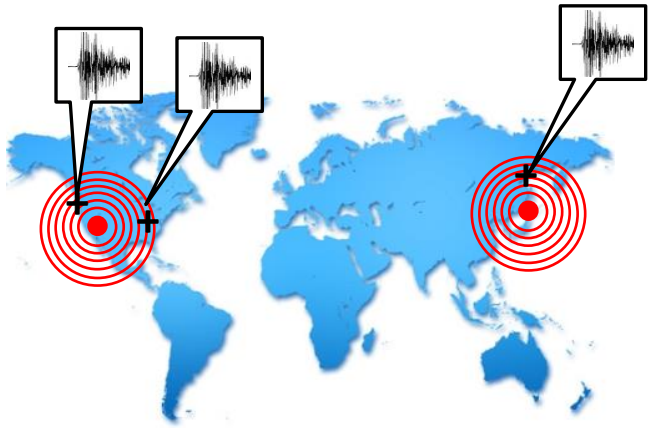
- 6 „Leistungspunkte“: Approximately 180 hours.

Signing Up for the Course

- If you would like to participate in the practical course, please write me an email and we organize a brief meeting.
- We will discuss possible topics with you.
- Next slides: brief overview over two possible application domains in the natural sciences, namely seismic risk analysis and eye movement models.
- You can also propose other topics, if you have an interesting domain to work on.

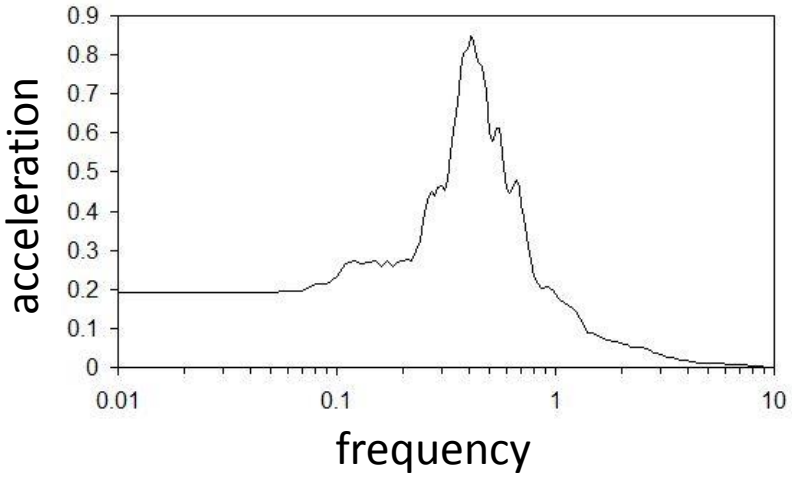
Application Area Seismic Risk Analysis

- *Ground motion models* predict the intensity of the ground motion in the case of an earthquake as a function of magnitude, distance to epicenter, soil properties etc.
- GMMs are used in statistical seismic risk analysis to infer the probability that a certain ground motion will be exceeded within a certain time span.
- Models are estimated from data recorded during previously observed earthquakes.
- Statistical / machine learning approaches.



Topics in Seismic Risk Analysis

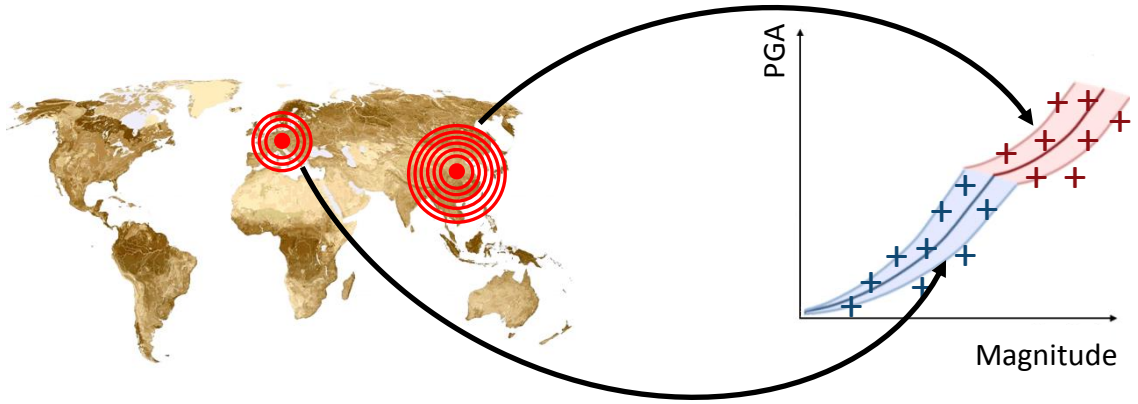
- Predicting so-called answer spectrum: ground motion at *k* different frequencies.



- Naive: solve *k* independent learning problems for the *k* frequencies.
- Alternative: solve joint learning problem, exploiting that neighboring frequencies are correlated.

Topics in Seismic Risk Analysis

- The relationship between inputs (distance, magnitude, soil) and output (ground acceleration) changes from region to region, e.g. due to tectonical features of region.
- However, (strong) earthquakes are rare: when training model for Region X, often need to additionally use data from Region Y.

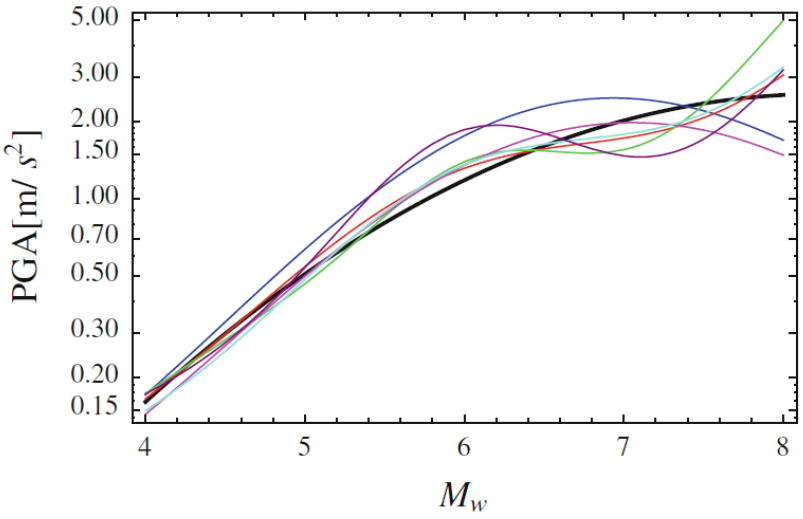


- How to exploit data from other regions consistently?

Topics in Seismic Risk Analysis

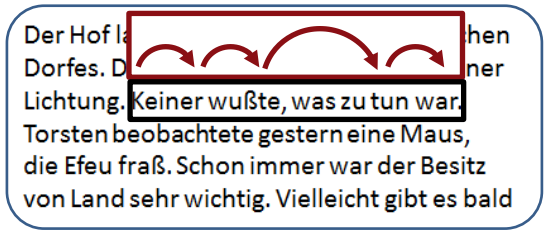
- Physically plausible extrapolation: predictions should be monotone e.g. in distance and magnitude.
- Nonlinear models do not always extrapolate nicely when they are applied to inputs that are outside of the range of values seen in the training data.

Training data magnitudes ≤ 6



Application Area Eye Movement Modelling

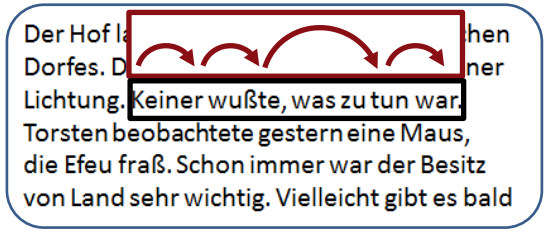
- Human eye movements reflect the interplay between vision, cognition, and motor control.
- Eye movements can be observed with high temporal and spatial precision in the lab.
- We specifically focus on eye movements during reading, where a reader fixates different words in a text while understanding what is written.



Topics in Eye Movement Modelling

- Can we predict eye movement patterns from the text being read, to better understand the interplay between text semantics and gaze control?
- What can we infer about a person based on observed eye movement patterns?

- Different aspects can be studied:
  - Predict whether a word will be fixated or how long it will be fixated
  - How do eye movements correlate for example with age, gender, native language, competency, or IQ?



Questions?

- Please watch our video lecture about how to do scientific projects: http://www.cs.uni-potsdam.de/ml/teaching/ws10/face/wa/Flash/wa.html

- Questions?