Advanced probabilistic methods

Lecture 10: Concluding remarks

Pekka Marttinen

Aalto University

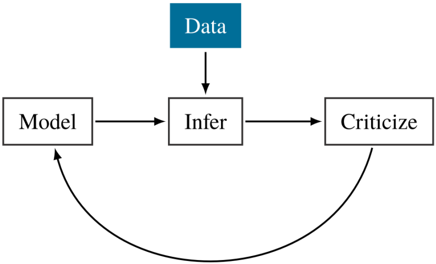
March, 2021

Lecture 10 overview

- A brief summary of the course
- Some suggestions

Course contents

- Ingredients of probabilistic modeling
  - **Models**: Bayesian networks, Sparse Bayesian linear regression, Gaussian mixture models, latent linear models
  - **Methods for inference**: maximum likelihood, maximum a posteriori (MAP), Laplace approximation, expectation maximization (EM), Variational Bayes (VB), Stochastic variational inference (SVI)
  - **Ways to select between models:** Bayesian model selection, AIC, BIC, ...



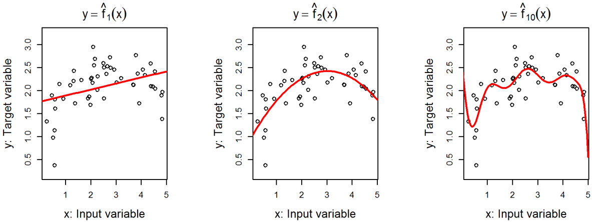
Box's loop (Blei, 2014)

Which model to use?

- Incremental approach
  - start with a simple model
  - assess fit (inspect residuals, compare model prediction with real data,...)
  - improve model if needed
  - See *Philosophy and practice of Bayesian statistics* (2012) by Gelman and Shalizi
- "Let the data speak"
  - start with a flexible model that can adapt to different kinds of data
  - often used in machine learning
- Construct alternative models based on understanding of the problem
  - Continue with model comparison (checking the fit still needed)

Which inference to use? (1/2)

- If the model is good, the results are often ok with any sensible model fitting technique.
- Bayesian approaches avoids some pitfalls of maximum likelihood (e.g. singularities in GMM) and yield uncertainty estimates and regularization
- Ready-made software, e.g. *Stan* for MCMC, *Pyro* (*PyTorch*) for VB, may be considered the first thing to try, if analytical integration is not possible.
    - Easy and fast if they work
    - Difficult to debug if they don't



Which inference to use? (2/2)

- **MCMC**:
  - pros: asymptotically you get correct probabilities
  - cons: assessing convergence, computation time
  - Recommeneded usage[1]: any 'conventional' statistical analysis where it's important to get the probabilities correct and computation time is not an issue
- **Variational inference**:
  - pros: faster, may be more stable than MCMC
  - cons: underestimates uncertainty
  - Recommeneded usage: 'machine learning' type models, where the goal is more on prediction and exact probabilities are not needed (e.g. deep learning models)

[1] Based on the present understanding of the lecturer. Other views may also be justified, therefore: think what is suitable in your case and try different options.

Reminders

- Details about the exam will be posted in MyCourses
  - You can ask more details in Slack.
  - Take advantage of the remaining exercise sessions to get clarifications.
- Please give feedback, this will be used when developing the course further next year.
  - An extra exercise point granted for providing feedback.

Thank you!

- Thanks and good luck!
- If you have any questions, you can post those to Slack.