

CS-E407520 - Special Course in Machine Learning and Data Science: Bayesian Workflows

Session 4: Model Checking

May 13, 2024



Schedule for today's session

Time	Activity
45 min	Discussion of workflow diaries
10 min	Break
15 min	Primer for next workflow steps
15 min	Example implementation

Reminder

Don't forget to submit your "work in progress" workflow diaries on MyCourses if you haven't already!

Workflow Diary Discussion

Discussion of workflow diaries

When it is your turn:

1. present your workflow diary via projector or screen
2. briefly summarise your data and research question
3. teaching staff will guide discussion of your implementation of the workflow steps

Your goals were:

Parameter estimate and model fit checks

- check model fitting diagnostics
- print and/or plot the marginal (and bivariate) posteriors for important parameters
- comment on the parameter estimates; do they make sense? are they surprising?

Posterior predictive checks

- create appropriate plots to check how model predictions compare to observations
- comment on any potential deficiencies in your model

Influence and sensitivity checks

- use cross validation to identify influential data points (justify if not applicable)
- perform a prior/likelihood sensitivity analysis and comment on results

Let's take a break! (10 min)

Some suggestions for recharging during breaks :

- move your body
- open a window or go outside
- drink some water
- try to avoid checking e-mails, messengers, or social media

Primer: Model Expansion and Selection

Model Expansion

- Repeatedly emphasised to 'start simple and then build complexity'
- Key part of Bayesian workflow is to *iteratively* expand our model (add complexity)
 - Allows for faster identification of fitting issues (and syntax mis-specification!)
- But how should you decide what to add/change in a model? Consider:
 - Research question or modelling goal
 - Diagnostic issues encountered so far

How to Expand a Model

- Adding more complex model structures
 - Hierarchies or clusters in the data
 - Time structure or dependence
- Changing existing model parameterisations
 - Different functional forms or distributional families
 - Alternative priors
 - More computationally efficient expression of current model
- Combining previously separate models
 - Multiple univariate models to single multivariate model
 - Joint modelling of groups/clusters

How to Expand a Model

- You expand and fit alternative models, now what?

Model Selection

- There is no 'one-size-fits-all' approach
 - How you choose a model is dependent on what you want the model to do
- Prediction
 - Cross-validation (LOO/K-Fold)
 - Held-out test data
- Explanation
 - R^2
 - Interpretability of model parameters
- Performance/Efficiency
 - Achieving the above as cheaply or quickly as possible

Model Selection

- It is also recommended to start with the general diagnostics/checks (R-hat, PPC, etc.)
 - Quickly identify models that are clearly mis-specified or inappropriate
 - Avoid unnecessary effort of more complex selection/comparison methods
- Is selection always needed?

Alternatives to Model Selection

- Your goal might not be to select the single “best” model
 - What if you want to use information from all candidate models?
- Bayesian Model Averaging or Stacking
 - Combine models for inference or prediction
- Multiverse analysis
 - Fit all possible modelling choices and investigate differences in inferences/results

Key Takeaways

- Your approach to model expansion and selection (or not!) is guided by your modelling goal
- Be explicit about what you are trying to model and what you want your model to do

Your Goals for the Next Week: Model Expansion

- Given your current model, propose at least two expanded or alternative models adjusting:
 - Prior choices
 - Hierarchical structure
 - Outcome family or model parameterisation
- Make sure to justify your choices/proposals, considering:
 - Your research goal/question
 - Previous results/diagnostics in the workflow so far
- Estimate/fit your alternative model(s)

Your Goals for the Next Week: Model Selection

- Choose, and justify, your criteria for selection:
 - Prediction
 - Explanation
- Perform the selection using one of the below methods for your chosen criteria:
 - Prediction
 - Cross-validation (LOO/K-Fold)
 - Held-out test data
 - Explanation
 - R^2
 - Interpretability of model parameters
- Summarise the selection results in a single table

Resources for the Week

Readings

- Workflow Book:
 - Ch.6: Comparing and improving models
 - Ch.19: Model building and expansion: Golf putting

Resources

- Aki's Cross-Validation FAQ
- nabiximols case study

Relevant Papers

- Riha et al. (2024); Bayesian Multiverse
- Piironen & Vehtari (2017); Bayesian model selection