

Model selection and comparison: Bayesian models

Miao Cai

2019-09-12

Contents

1 p_{loo} and Pareto k values

1

1 p_{loo} and Pareto k values

This is a quick note on different cases why Pareto k values can be large. I'll try to extend this to a longer explanation and a case study. Any feedback on this draft is appreciated

- If all Pareto k small, model is likely to be ok (although there can be better models)
- If high Pareto k values
 - If $p_{loo} \ll$ the number of parameters p , then the model is likely to be misspecified. PPC is likely to detect the problem, too. Try using overdispersed model, or add more structural information (nonlinearity, mixture model, etc.).
 - If $p_{loo} >$ the number of parameters p , then the model is likely to be badly misspecified. If the number of parameters $p \ll n$, then PPC is likely to detect the problem, too. Case example https://rawgit.com/avehtari/modelselection_tutorial/master/roaches.html 41
 - If $p_{loo} >$ the number of parameters p , then the model is likely to be badly misspecified. If the number of parameters p is relatively large compared to the number of observations $p > n/5$ (more accurately we should count number of observations influencing each parameter as in hierarchical models some groups may have small n and some groups large n), it is possible that PPC doesn't detect the problem. Case example Recommendations for what to do when k exceeds 0.5 in the loo package? 58
 - If $p_{loo} <$ the number of parameters p and the number of parameters p is relatively large compared to the number of observations $p > n/5$, it is likely that model is so flexible or population prior is so weak that it's difficult to predict for left out observation even if the model is true one. Case example is the simulated 8 schools in <https://arxiv.org/abs/1507.04544> 24 and Gaussian processes and spatial models with short correlation lengths.