Computational Statistics Final Project

U20 Math 575, Fall 2021

Objective

For our final project, you will be asked to apply computational statistics methods to a topic of your choice using an analysis-friendly dataset.

Analysis ideas

Below are some suggestions to help you get started:

- Using bootstrap algorithm to estimating model coefficients, or population statistics
- Implement a random variable simulation algorithm (such as Metropolis-Hastings) using different type of proposal functions
- Using non-parametric methods to estimate optimal curve for bi-variate data

Dataset idaes

and here are some ideas for datastes (click for link)

- *Kaggle*
- *UCI Machine Learning repository*
- *Pre-loaded datasets in R*

R pacakge references

Below are some example of R packages that could be applied in your final project. Check for examples using these packages for more ideas:

- R package for kernel density estimation: https://cran.r-project.org/web/packages/ks/vignettes/kde.pdf
- R package for Markov chain Monte Carlo simulation: https://cran.r-project.org/web/packages/mcmc/mcmc.pdf
- R package for bootstrap algorithm: https://cran.r-project.org/web/packages/boot/index.html

Detailed guideline

While you are not expected to build a computationally complex model, your work needs to show logical flow, and demonstrates the Bayesian analysis concepts discussed in the course. This includes the following:

1. *Description of the problem:* What is the problem you are trying to solve? What is the motivation and significance behind this? Why might your approach be useful here?

2. *Description of your data:* What are the variables of interest and their summary? What are some caveats of the data (such as data quality issues) that we need to be aware of, if any?

3. *Formulation of your analysis approach:* How is the model or estimation algorithm defined?

4. *Computational approach:* What methods are you using to analyze the data? You are encouraged to use existing R packages.

5. *Results and conclusion:* What is the takeaway from your analysis? What makes your approach advantageous (or challenging) in your problem? What are the next steps in your analysis?

6. *References:* Citation of references and data sources (if applicable)

Format

The project will be divided into 3 parts:

1. Outline (5pt): Please describe the problem of your choice and an initial outline of the steps you plan on taking to complete the project.

2. Write-up (20 pt): Please prepare a detailed write-up of your project that addresses the points listed under the `Detailed guideline` section.

Please submit all portions of your project on Canvas. Links will be available before due dates.

Timeline

The following are the key due dates for the project:

- **Tuesday, November 30, 2021** : Outline of final project

- **Sunday, December 19, 2021**: Write-up of final project

Writing and communication guidelines

Please keep your project write-up as concise as you can. Please write your paper in the style of a professional report, which means it is important to check the flow of your work to make sure it is coherent, include any exhibitions to facilitate your ideas, and cite all sources properly. For styling, you are highly encouraged to use LaTeX or Rmarkdown to type up the project, but Word is acceptable as well.