Stat 847 - Reading Assignment 2: Model Selection and Averaging

DUE: Friday February 16, 2024 by 11:59pm Eastern

NOTES

Your assignment must be submitted by the due date listed at the top of this document, and it must be submitted electronically in .pdf format via Crowdmark.

Organization and comprehensibility is part of a full solution. Consequently, points will be deducted for solutions that are not organized and incomprehensible. Furthermore, if you submit your assignment to Crowdmark, but you do so incorrectly in any way (e.g., you upload your Question 2 solution in the Question 1 box), you will receive a 5% deduction (i.e., 5% of the assignment's point total will be deducted from your point total).

Reading: Model Selection and Model Averaging [12 marks]

Open the UWaterloo Library website, lib.uwaterloo.ca , and use your WatIAM account to search for an open the book **Model Selection and Model Averaging**. By Gerda Claeskens and Nils Lid Hjort. The following questions can be answered by reading the "1 - Model Selection: data examples and introduction" chapter.

Please put your answers to questions 1-3, 4-6, and 7-8 in three separate pages, this can be done in Word with Crtl + Enter, or in Markdown with `\newpage`.

In Section 1.1 "Introduction", answer the following with quotes or your own words (your choice)

1. (1 mark) What is George Box's maxim about models?
2. (1 mark) Give an example of the principle of parsimony in action in statistical modelling.
3. (1 mark) In one sentence, why would you employ model averaging?

In Section 1.3 “Who wrote ‘The Quiet Don’?”, answer the following in your own words.

4. (1 mark) What are the three data sets used in this model selection problem?
5. (2 marks) What are the three competing models that were suggested?
6. (3 marks) How were the models compared to each other? (A vague description in your own words is good. Details about the methodology aren’t necessary, but do mention what data specifically was important)

In Section 1.6 “Football match prediction”, answer the following in your own words.

7. (2 marks) Describe the dataset being used in this problem.
8. (1 mark) Figure 1.5 shows the distribution of football scores, which are whole numbers. Why are there clouds of data points instead of values only at the whole numbers?