HW1_msalem

My First R Markdown

Problem 2 - Part A

What I hope to learn in this class includes, among other things, the ability to effectively use Git to the extent of observing a positive impact on my ability to create reproducible projects; developing a degree of proficiency in R Markdown that would allow me to migrate analyses in conducted in R to a readable and neat document in a smooth and flowing manner; and finally, I hope to be able to build a better understanfing of Deep Learning.

Problem 2 - Part B

$$Poisson(\lambda) : P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!};\;\; x = 0,\, 1,\, ...;\;\; 0 \leq \lambda < \infty \tag{1}$$

$$Beta(\alpha, \beta) : f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1};\;\; 0 \leq x \leq 1;\;\; \alpha > 0;\;\; \beta > 0 \tag{2}$$

$$Normal(\mu, \sigma^2) : f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{(2\sigma^2)}} \;\; ; \; -\infty < x < \infty \;\; ; \; -\infty < \mu < \infty \;\; ; \; \sigma > 0 \tag{3}$$

Problem 3 - Guidelines and Challenges for Reproducible Research

Guidelines

1. Establish detailed tracking of how you reached each result, including tracking of pre and post processing steps. A tracking method that allows for direct execution is the most suited for ease of replication.
2. Minimize manual manipulation of data by using prepackaged, popular tools to homogenize data manipulations.
3. Version tracking and archiving of software used.
4. Version tracking of code and scripts
5. Recording and archiving intermediate results and making sure they are accessible and visible to others.
6. Use seed numbers, document them, and point out where they were used.
7. Storing the underlying data of plots.
8. Generate a hierarchical tracking system that makes it easier for would-be replicators to access desired detailed intermediate results.
9. Always connect results to their textual interpretations especially when communicating textual interpretations without providing a full version of the output.
10. Providing public access to all project resources.

Challenges

1. Challenges to achieving the above includes the amount of effort that needs to be invested in order to track the minutiae of the workflow process and maintaining those in a directly executable form which may be considerable.
2. Challenges include tool accessibility and ubiquity.Not all researchers have access to the same tools. It is also often the case that there exist multiple tools providing similar functionality, researchers may have varying preferences on which to use. It also may be the case in some contexts that data manipulations can be more efficiently accomplished using personalized methods than prepackaged ones.
3. Challenges include the future non-compatibility of the archived software with newer platforms. Software archiving also requires clear and detailed descriptions of the changes implemented between versions whose provision responsibility lies with the original creator of the software.
4. Effort investment in maintaining the necessary logs, as well as the storage resources required by such archiving.
5. Effort investment, increased storage requirements, over-flooding work with intermediate results that overshadow the final conclusion and could detract from the intended main focus of the work.
6. Some researchers may not be familiar we the concept of seed numbers.
7. Plots may be extracted directly from other work with no data provided by the original author. Storing additional data entails additional storage and logging requirements.
8. This may require knowledge in an additional platform such as hypertext.
9. It may be difficult to locate and readily retrieve results associated with a particular textual interpretation if not appropriately planned beforehand.
10. Requires a repository to host all the project's elements that is accessible to researchers at large.

Plots for Problem 4