Ian Fayorsey

Comprehensive Project Round 1 Response

1. **We are concerned about reproducibility due to the last minute issues with the template. It is acceptable to not use the template, so long as there are clear steps presented to allow us to replicate the analysis. The submission does not show the code used in the analysis (it can be in an appendix).** Response : I acknowledge that my first version did not have an appendix attached, and have included one in the most recent version.
2. **We felt that it was clear you had taken some time to investigate the methods and present appropriate background in chapter 2. However, each section felt underdeveloped in terms of notation and related explanation. Your descriptions may benefit from more specifics. You may have been borrowing notation from your sources, but it’s not clear here. For example, you use both K and k for the number of classes, G is undefined in 2.3 and it’s unclear if the X in 2.3 is a vector or matrix, etc. We recommend establishing notation that you can use throughout the chapter (as best you can), and introducing that to start, then doing method specific notation to start each method’s section**. Response: Yes, the notation in my first round was not always succinct. I have defined all the my symbols and ensured they are consistent and clearly defined and remain consistent throughout the analysis.
3. **Toy example – Consider including a toy example to demonstrate each method from section 2 as you present each method. Something as simple as classifying the iris data set by species can show what the method output looks like, and what you might expect from your results.** Response: I was under the impression applying these algorithms to a new data type and discussing their predictions was how I demonstrated my understanding of the methods. Time permitting, I may incorporate a few toy examples.
4. **Analysis – When reading Chapter 3, it was a bit of work for the reader to go back to Chapter 1 and find the details about the data set involved in the analysis. A quick recap of what the data is, how many observations are involved in training/test, and what the variables are would be a way to strengthen the section, before jumping into the results. (Again, specifics/details are helpful).** Response: I have included a recap of the data and how I broke down the test/train sets.
5. **Runtimes – We found inconsistencies presented in terms of the runtimes in the sentence description versus the table**. Response: I have addressed inconsistencies between table runtimes and how they are described in sentences.
6. Formatting/Proofreading – There are numerous spelling and other errors that could be **addressed by running spell check and proofreading the submission. At least one Figure is cut off (which you had pointed out previously). Items like this should be addressed. We suggest examining Figures/tables, etc. for appropriate numbering, titles, and captions.** Response : I have addressed the spelling errors, updated the formatting and corrected the captions for the tables and figures.
7. **We found several acronyms that were used before the name of the item was presented. Please check that you state the name before using the acronym.** Response : The acronyms have been spelt out before using them.
8. **In 3.0.1, the term "significantly faster" is used. Was a hypothesis test used? Is this a colloquial use of “significantly”?** Response: The use of “significantly” was colloquial, and was not supposed to imply a hypothesis test.
9. **In 3.1, 2nd paragraph of the conclusion, you suggest that some techniques were not suitable for the data. Can you explain (briefly) why they weren't suitable and what characteristics would have been necessary for the techniques?** Response: : I have addressed why some techniques were not suitable for the data.