Project Report 3

The mid-term project represents an integration of statistical learning techniques implemented in the first half of the course. You are encouraged to focus on the same dataset and reuse code and results from the bi-weekly Reports and to incorporate improvements from the peer code review and instructor comments. The mid-term project report should:

1. Compare *TWO* models based on techniques discussed in Chapters 2-7, and
2. Interpret the results of the best model in the context of a compelling research question. Include the following five sections:

1. Introduction

- Describe dataset and motivating scientific question or hypothesis.
- Credit data source and cite relevant literature including peer-reviewed publications to help the reader understand the dataset and research question.

2. Methods

- Briefly summarize the methods used for analysis.
- Justify the two models you have chosen to compare for your dataset. You may choose to compare different modeling frameworks (e.g. LDA vs logistic regression) or models that include different sets of predictors (e.g. multiple regression model vs. a regularized model).
- Use one of the resampling approaches from Chp. 5 to compare models.
- Your report should include code used to perform the analysis, but the rendered report should be easily understandable by someone who is not a programmer.

3. Results

- Describe and briefly interpret the results.
- One or more visualizations (including formatted tables) that support the conclusions made. Do not include extraneous visualizations.
- Indicate which model gives best performance on your dataset. Given your knowledge of the dataset and methods implemented, why do you expect this to be the case?
- If there are any problems applying the technique to your dataset, describe them.

4. Discussion

- Compare and contrast method performance in the context of trade-offs discussed in ISL Chp. 2. Given these considerations, which model do you recommend?
- Describe as a future direction at least one other method that might provide better performance (see ESL or next ISL chapters)? Why?
- How do the results bear on the question/hypothesis set up in the Introduction?

5. References

- See https://bookdown.org/yihui/rmarkdown-cookbook/bibliography.html#ref-R-bookdown for information on formatting a bibliography section.