Project Breakdown

Christian Pascual

4/27/2019

(Limit 6 pages, limit 8 figures/tables)

Project Breakdown

Introduction (Assigned to Christian)

Describe your data set. Provide proper motivation for your work.

- What questions are you trying to answer?
- How did you prepare and clean the data?

EDA/Unsupervised Analysis (Assigned to Christian)

- Is there any interesting structure present in the data?
- What were your findings
- Visualize -compare boxplots of control vs case for each protein
- Correlation Plot
- visualize what the correlation is between variables
- PCA
- visualize scree plot
- visualize first vs second principal component
- If you cannot find anything interesting, then describe what you tried and show that there isn't much visible structure.

Models

- What predictor variables did you include?

Models to use:

- Logistic-LASSO (Assigned to Christian)
- KNN (Assigned to)
- Random Forest/Bagging (Assigned to)
- Boosting (AdaBoost?) (Assigned to)
- SVM (Assigned to)

For each of the models:

- What techniques did you use
- why did you choose it?
- What assumptions, if any, are being made by using this technique?
- If there were tuning parameters, how did you pick their values?
- What are the limitations of the models you used (if there are any)?

After all of the models have been tuned: (Assigned to Christian)

- How did you conduct model selection?

- Explain/visualize the final model you select.
- Which variables play important roles in predicting the response?
- Discuss the training/test performance if you have a test data set.

Conclusion (Assigned to Christian)

- What were your findings?
- Are they what you expect?
- What insights into the data can you make?

Appendix

- model tuning