

Predicting survival time of XX disease

(STATS/CSE 780 course project)

Pratheepa Jeganathan
Department of Mathematics and Statistics
McMaster University

03/04/2022

Outline

Motivation (1 slide)

- ▶ State the problem?
- ▶ Why does this problem important?
- ▶ Any solution already exists for the problem?
- ▶ Goal of this project?
- ▶ etc.?

Data (1 slide)

- ▶ Source?
- ▶ Is it a data frame?
 - ▶ What is in rows?
 - ▶ What is in columns?
- ▶ Results of exploratory analysis?
 - ▶ Data types, type of response if any?
 - ▶ Correlation analysis?
 - ▶ Outliers?
 - ▶ Missingness?
 - ▶ Data transformation, etc. ?
- ▶ $n > p$ or $n < p$ or $n \approx p$?
- ▶ etc. ?

Data (Example)

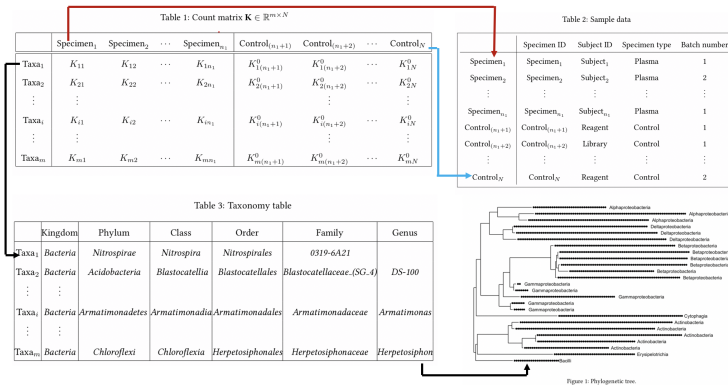


Figure 1: Source:xxx.

K_{ij} abundance of i -taxon in j -th sample.

Methods (1 slide)

- ▶ What are the two methods you compared?
- ▶ Why those two methods?
- ▶ Algorithms of the methods?
- ▶ Any statistical transformation used?
- ▶ Any other pre-processing (feature engineering) used?
- ▶ Any feature selection (filter, or wrapper, or embedded) used?
- ▶ etc.?

Methods (Example)¹

- ▶ KNN and DT for classification.
- ▶ Decision trees - partition the predictor space into simple regions.
 - ▶ Predict y_0 of a new data point x_0 using the response of training observations in the region to which x_0 belongs.
 - ▶ How to find the partitions?
- ▶ KNN -

¹An introduction to statistical learning (James et al. 2013)

Results (1 slide)

- ▶ What are the results of applying the methods?
 - ▶ Visualize the results?
 - ▶ Compare the methods using graphs?
 - ▶ Interpret the model/results?
 - ▶ etc.?

Discussion (1 slide)

- ▶ Discuss problems related to the methods and data -
 - ▶ Curse of dimensionality?
 - ▶ Multiple data types?
 - ▶ Interpretability?
 - ▶ Reproducibility?
 - ▶ Stability?
 - ▶ etc. ?

Thank You!

References

- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. Springer.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. "O'Reilly Media, Inc."
- Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. CRC Press.