CS-C3240 - Machine Learning D, Lecture, 5.9.2022-14.10.2022

Report Outline

Introduction:

- Explain the background (real-life scenario) of your ML application.
- Briefly outline the structure of this report

Problem Formulation:

- Formalise the application an ML problem.
- Clearly explain the data points, features and labels of this ML problem
- Explain the source of the dataset

Methods:

- State the number of datapoints, briefly describe the dataset and/or any data preprocessing needed.
- Explain your feature selection process (no theoretical justification needed).
- Describe and explain (why?) your choice of ML model(s)/hypothesis space(s)*, e.g., linear predictors, etc.
- Describe and explain (why?) your choice of loss function(s)*, e.g., logistic loss
- Explain the process of model validation - how did you split the data into training, validation and test sets. What are the sizes of each set and why did you make such design choice.

Stage 1
23 Sep, 23:59
You are only required to discuss 1 method

Results:

- Compare and discuss the training and validation errors obtained for all ML methods considered.
- Which is the final chosen method and why?
- What is the test error of the final chosen method?

Conclusion:

- Summarise the report and your findings.
- Are the results suggesting that the problem is solved satisfactorily, or might there be room for improvement?
- Explain the limitation of the methods and how it can be further improved.

Bibliography/References

Appendices

- Your code with which you preprocessed the data, trained and evaluated the models, etc. (for stage 1 you only need to include code for the progress you have made on the project so far at that stage)

Stage 2
7 Oct, 23:59
You are required to discuss at least 2 methods

*Choose from the ones covered in the course.