Final Project

COM SCI X 450.4 - Machine Learning

Thursday 1st October, 2020

1 Introduction

The final project will be your chance to use and demonstrate everything you have learned throughout the course. You will pick a problem you care about using a real dataset and will propose a solution using machine learning. You can work individually or in groups of up to 3 students. The problem can be any of the topics discussed in our class (Classification, Regression, Clustering, Anomaly Detection, Reinforcement Learning, Recommender Systems, Bias in machine learning models, ...) as long as it is possible for you to measure performance and compare the trade-offs of using multiple models.

2 Project Proposal (5% of final grade for the course)

Your first deliverable will be a proposal. The proposal should contain the following sections:

1. **Problem** - Define the problem being studied: What are you trying to solve, what are the goals and how can you measure success in the end.
2. **Dataset** - Explain which dataset you will be using, which method you used to create it (if applicable), and summarize statistics with at least two plots of the data.
3. **Project Plan** - Attach the tasks you believe are necessary to complete your project and a rough estimate of the required time.

The proposal should have between 2-3 pages, including images. The sections can be used for the final report, consider the proposal as a checkpoint to make sure you are on track. It should be delivered at the end of Week 4.

3 Final Presentation (10% of final grade for the course)

You will record a 10 minutes presentation. Your final presentation should be a summary of the work done. It would be best if you assumed that the audience is composed of the clients that hired you to solve the problem. Make it enjoyable, focus on the macro details, and try to explain the methods in a way that non-technical audiences can understand. Being

able to communicate your work is one of the essential parts of data science and machine learning. Your presentation should include all the points from your final report. Focus on the trade-offs of each model you selected and how you decided which one was the best fit to solve your problem.

If working in groups, both students are expected to participate in the presentation equally.

4 Final Report (25% of final grade for the course)

Your final report will be the last deliverable of the class and represent a big part of your grade. It should contain the following sections:

1. **Problem** - Define the problem being studied: What are you trying to solve, what are the goals, and how can you measure success in the end.

2. **Dataset** - Explain which dataset you will be using, which method you used to create it (if applicable), and summarize statistics with at least two plots of the data.

3. **Proposed Solution** - Explain how you plan to solve this problem. How are you splitting your data? What kind of problem is this? What are metrics are you planning to use to measure success, and why? Those are some of the questions you might want to answer here. This section is probably the most critical part of your report, as the reasoning usually is more important than results.

4. **Modeling** - Describe the models you have decided to use to solve your problems. Explain why you have selected them (at least **three** models), how they work, and if there is any scenario that might limit the use of the model.

5. **Results** - Compare the models using the metrics you proposed in the previous sections. Use images and plots if necessary, and explain the trade-offs of each model.

6. **Conclusion** - Write your conclusion. Which model should be used to solve the problem? Is this the same model you would use if inference time were a constraint? What about the size of the model? Think of scenarios for when to use each of the paths you explored and propose the best solution for each of them.

Your final report should have between **3-6 pages** and have at least (feel free to add more sections if necessary) the sections mentioned above. If you have any questions or concerns, do not hesitate to contact the instructor. This project should take time to finish, but it will be a great addition to your portfolio.