THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Machine Learning I
DATS 6202
Due: *June-28-2017*

Final Project

The objective of the final project is to apply what you have learned in this course to a real world problem – a problem in your area of interest. You can apply any machine learning network or algorithms that we cover in the course. Below is a list of websites that contain data sets that might be suitable for deep learning.

Websites with datasets

Data Sets For BI/Analytics/Visulization Projects
https://sqlbelle.com/2015/01/16/data-sets-for-bianalyticsvisualization-projects/

Kaggle competitions and data sets
https://www.kaggle.com/

Flowing Data
http://flowingdata.com/

Fivethirtyeight
https://github.com/fivethirtyeight/data

R DataSet
https://vincentarelbundock.github.io/Rdatasets/datasets.html

Proposal

After you have selected a topic, a network and a data set, submit a proposal of what you plan to do for the project. The proposal should be a few hundred words, and should address the following items.

- What problem did you select, and why did you select it?
- What database will you use? Is it large enough to train a machine learning network or different algorithms?
- What neural network will you use? Will it be a standard form of the network, or will you have to customize it? What algorithms you will be used?.
- What software will you use to implement the neural network or different algorithms? Why?

- What reference materials will you use to obtain sufficient background on applying the chosen network or algorithm to the specific problem that you selected?
- How will you judge the performance of the network? What metrics will you use?
- Provide a rough schedule for completing the project.

Final report

The final report should be submitted as a PDF file to the BlackBoard by Wednesday, June 28 at 11:59 pm and should contain the following sections.

- Put the proposal in the very first sections of your report.
- Introduction. An overview of the project and an outline of the report.
- Description of the data set.
- Description of the machine learning network and training algorithm or other algorithms that you used. Provide some background information on the development of the algorithm and include necessary equations and figures.
- Experimental setup. Describe how you are going to use the data to train and test the network. Explain how you will implement the network in the chosen framework and how you will judge the performance. Write a complete report with theoretical description and verify this mathematical concepts with applying it with actual data. Provide enough information about the codes tat you have written. Write your codes in sperate subroutines and call the functions if needed?. Explain each subroutine.
- Results. Describe the results of your experiments, using figures and tables wherever possible. Include all results (including all figures and tables) in the main body of the report, not in appendices. Provide an explanation of each figure and table that you include. Your discussions in this section will be the most important part of the report.
- **Summary and conclusions.** Summarize the results you obtained, explain what you have learned, and suggest improvements that could be made in the future.
- References.
- A separate appendix should contain documented computer listings.