Week 4

11-777

Due: Thursday September 21 by 4:30 pm

Instructions

A small yet significant part of the course is answering a few questions based on a paper you just read. By design, some of these questions are open-ended and the purpose is to start a thinking process. Some of those questions may already have been answered in recent literature (or in the given paper itself). We encourage you to, both, come up with your own ideas and survey recent papers before answering these questions.

A few key points to take care of:

- Answer **two out of the three questions** (you are welcome to answer all 3), unless stated otherwise.
- A lot of questions ask you to suggest changes or make alterations. It would be nice to support it with logical / mathematical arguments. Cite all sources used in the process of coming up with your answer. Figures and equations, if they support your arguments will be appreciated.

Visualizing and Understanding Recurrent Networks

1. The datasets used in this paper include War and Peace and Linux Kernel (a codebase). If you were to build your own dataset to test out properties of RNNs (synthetic or from the "real world") how would you do that?
2. In the paper it is shown that a recurrent model works better than a 20-gram language model. Provide some empirical results from the paper to support this claim along with your thoughts on why should such an improvement be expected?
3. Section 4.4 of the paper shows an error analysis of the model. Specifically, an n-gram model helps the model eliminate 18% of errors. This shows that gated RNNs do not fully utilize the previous context even for the previous 9 characters. Why do you think that is the case? How would you address this problem?