CS 395

Deep Learning

Assignment 2

Due on Thursday, 4/8/2021, 11:59pm.

Instructions:

- Submit a detailed report about your experiments, You are free to choose the template you want.
- Choose whatever deep learning toolkit, any language you see it fit …
- Late policy as explained in the syllabus.
- Don't copy any code from internet for this assignment, you can use build functions in deep learning toolkits to perform high level architectures such as RNN, LSTM, etc, but you are not allowed to copy any code form internet that does similar task.

Note: For this assignment, submit your local copy of code and answers. Submit the file to Blackboard under Assignment1 using this format:
Yourfirstname_lastname_Assignment2.zip, inside this zip file, you should have a file name **report.pdf** to report your results**.**

Deadline: 4/8/2021, 11:59pm.

Assignment:

In this assignment, you will design and implement a character-level Recurrent Neural Network (RNN).

Part 1: Data and Representation

- Select you training dataset. Keep in mind, first of all, that your dataset has to be big in order to learn the structure, typically RNNs have been trained on highly diverse texts such as novels, Eminem lyrics, programming code, etc. (so be creative here!) Easy option, Gutenberg Books is a source of free books where you may download full novels in a .txt format.
- You need to use a character-level representation for this model. Convert the chosen training set to the extended ASCII of 256 characters, read characters one at a time and convert it into a one-hot-encoding. Each character will map to a vector of ones and zeros, where the one indicates which of the characters is present. Your RNN will read in binary vector of length-256.

Dataset choices:

- Your own choice, it is fun to work with your choice of data. Examples:
  - http://www.ffts.com/recipes.htm
  - A Midsummer Night's Dream", Data2 under /gpfs1/home/s/w/swshah/Assignment2_Datasets
  - Federalist papers prepared by project glutenberg. Data3 under /gpfs1/home/s/w/swshah/Assignment2_Datasets
  - Shakespeare's Sonnets: Sonnet 1, Data4, under /gpfs1/home/s/w/swshah/Assignment2_Datasets

Part2: Training

Train your recurrent neural network using the dataset you created in Part1. You are free to choose learning parameters (sequence length, learning rate, stopping criteria, etc.), to make it easy for you, you can use build-in functions from deep learning toolkits to implement and train RNN architectures. Complete the following tasks:

- Report your training procedure. Plot the training loss vs. # of training epochs.
- During training, choose 5 breaking points (for example, you train the network for 100 epochs and you choose the end of epoch 20,40,60,80,100) and show how well your network learns through time. You can do it by feeding in the network a chunk of your training text and show what is the output of the network. Report your result.

Part3: Experiments

Experiment with different Network Structures. Discuss your findings if you change these parameters:

- Number of hidden units. Try doubling and halving your number of hidden units. plot the training loss vs. the # of training epochs and show your text sampling results.
- Sequence length. Try doubling and halving your length of sequence that feeds into the network. plot the training loss vs. the #of training epochs and show your text sampling results.
- Replace your RNN with LSTM and GRU. plot the training loss vs. the #of training epochs and show your text sampling results.