Assignment 4

Deep Learning
IIT-Hyderabad
Jan-May 2022

Max Marks: 25
Due: 01st May 2022 11:59 pm

This homework is intended to cover exercises in the following topics:

- Recurrent Neural Networks, LSTM/GRU

Instructions

- Please upload your submission on Piazza by the deadline mentioned above. Your submission should comprise of a single file (PDF/ZIP), named `<Your_Roll_No>_Assign4`, with all your solutions.
- For late submissions, 10% is deducted for each day (including weekend) late after an assignment is due. Note that each student begins the course with 12 grace days for late submissionof assignments (of which atmost 7 can be used for a given submission). Late submissions will automatically use your grace days balance, if you have any left. You can see your balance on the CS5480 Marks and Grace Days document (soon to be shared).
- Please use PYTHON for the programming questions.
- Please read the department plagiarism policy. Do not engage in any form of cheating - strict penalties will be imposed for both givers and takers. Please talk to instructor or TA if you have concerns.

Questions: Theory

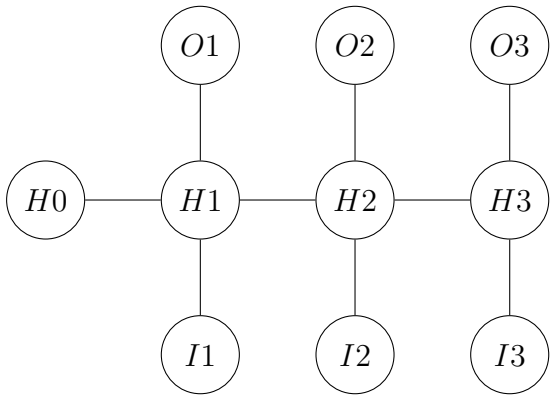
1. *(2+2+2=6 marks)* We learned in class that RNNs are versatile, and can be used in several tasks related to data with a temporal component. For each of the following tasks, state how you would design an RNN to do that task. In particular, specify: (i) how many outputs you have in your RNN; (ii) what probability distribution would the outputs be over (e.g. distributed over all species of cats); (iii) which inputs are fed at each time step; and (iv) if there is any change in architecture between train and test times (and if so, what change).

   (a) **Named-Entity Recognition:** For each word in a sentence, classify that word as either a person, organization, location, or none. *Inputs:* A sentence containing n words.

(b) **Sentiment Analysis:** Classify the sentiment of a sentence rangingfrom negative to positive (integer values from 0 to 4). *Inputs:* A sentence containing n words.

(c) **Language models:** generating text from a chatbot that was trained to speak like you by predicting the next word in the sequence. *Input:* A single start word or token that is fed into the first time step of the RNN.

2. *(3 marks)* Consider the following time-unrolled RNN:



where H_i , O_i , I_i are the hidden states, output states and input states respectively. U, V, W are weights of the RNN connecting different modules as shown in the figure. If errors corresponding to each output O_i are E_i , and E is the total error of the network, compute the following gradients:

(a) $\frac{\partial E}{\partial W}$, $\frac{\partial E}{\partial U}$, $\frac{\partial E}{\partial V}$

(b) $\frac{\partial E_2}{\partial W}$, $\frac{\partial E_2}{\partial U}$, $\frac{\partial E_2}{\partial V}$

(c) $\frac{\partial E_3}{\partial W}$, $\frac{\partial E_3}{\partial U}$, $\frac{\partial E_3}{\partial V}$

Questions: Programming

1. *(6 marks)* Go through the `Assignment_4_1.ipynb` provided with this assignment. The notebook has the necessary questions and skeleton code to implement this assignment. In particular, find out and report the caption generated by the model when an input image is being fed to the encoder. You may have to do the following to complete this question:

   (a) Implementation of Encoder-CNN module

   (b) Implementation of Decoder-LSTM module

   (c) Input pre-processing, loading vocabulary and parameters of encoder-decoder modules in order to perform inference with the given input image.

2. *(10 marks)* In this question, you will implement a sequence-to-sequence GRU for translation. Please refer to the provided `Assignment_4_2.ipynb` notebook. (If you'd like to practice working with RNNs, please also see the practice question provided.) You may have to do the following to complete this question:

   (a) *(1 mark)* Prepare the data by converting sentences to word lists and pairing appropriate words.

(b) *(3 marks)* Build the encoder and decoder architectures.

(c) *(1 mark)* Prepare the training data and convert to tensors.

(d) *(3 marks)* Train the model with occasional teacher forcing.

(e) *(2 marks)* Evaluate your trained network.

1 Practice Exercises

NO SUBMISSION REQUIRED; PLEASE USE THIS FOR PRACTICE AND LEARNING.

1. In this question, we will perform time series prediction using an LSTM model on the provided `airline-passengers.csv` file. You can find the skeleton code for this question in the `Assignment_4_3.ipynb` notebook, provided herewith. Your tasks will be to:

   (a) Create the model

   (b) Train the model

   (c) Test the model

   (d) Report the mean of all values predicted by the LSTM network on test data

2. The attention mechanism in deep neural network models talks about q (query), v (value) and k (key) vectors. Let us look at the case of single-head attention. Let us say q, k, v vectors are from $\mathbb{R}^d$ and there are m value vectors and m key vectors. Then attention vector and attention weights can be defined as:

$$z = \sum_{i=1}^{m} (v_i \alpha_i)$$

$$\alpha_i = \frac{\exp(k_i^T) q}{\sum_{i=1}^{m} \exp(k_i^T q)}$$

   (a) Let us say z is evaluated to v_j for some j. What does this mean? Explain using query, key and value vectors.

   (b) Now, take an orthogonal set of key vectors $\{k_1, \ldots, k_m\}$ where all key vectors are orthogonal, that is $k_i \perp k_j$ for all $i \neq j$. Let $\|k_i\| = 1$ for all i. Let $\mathrm{v}_a, \mathrm{v}_b \in \{\mathrm{v}_1, ..., \mathrm{v}_m\}$ be two of the value vectors from a set of m arbitrary vectors. Express query vector q such that the output z is roughly equal to the average of v_a and v_b, that is, $1/2^*(\mathrm{v}_a+\mathrm{v}_b)$.

3. Below is the modified objective function for a Generative Adversarial Network (GAN) where a and b are the class labels for the generated images and real images. Given a fixed generator G, derive the optimum discriminator $D^*(x)$:

$$min_D[\frac{1}{2}\mathbb{E}_{x\sim P_{data}(x)}[(D(x)-b)^2] + \frac{1}{2}\mathbb{E}_{z\sim p_z(z)}[(D(G(z))-a)^2]$$

4. Why does Inception Score (IS) fail in evaluating a GAN? What can be done to overcome that failure?