Assignment 4

CS5370: Deep Learning for Vision/AI5100: Deep Learning/AI2100: Deep Learning
IIT-Hyderabad
Jan-Apr 2021

Max Marks: 50
Due: 30th Mar 2021 11:59 pm

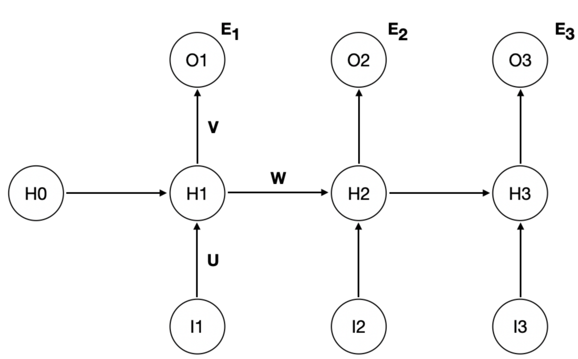
Instructions

- Please use Google Classroom to upload your submission by the deadline mentioned above. Your submission should comprise of a single ZIP file, named `<Your_Roll_No>_Assign4`, with all your solutions, including code.
- 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 10 grace days for late submission of assignments, of which upto 4 grace days can be used for a single assignment. Late submissions will automatically use your grace days balance, if you have any left. You can see your balance on the Marks and Grace Days document, soon to be shared under the course Google drive.
- You have to 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.

1 Theory (15 marks)

You can submit your response as a PDF document, which can be typed out in LaTeX/Word, or handwritten and scanned. If handwritten, please ensure legibility of answers.

1. *(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) $\frac{\partial E_3}{\partial W}, \frac{\partial E_3}{\partial U}, \frac{\partial E_3}{\partial V}$

2. *(1+2=3 marks)* In the context of the vanishing gradient problem, answer the following questions:

   (a) Why do recurrent models suffer from the vanishing gradient problem, and how to solve it?

   (b) Consider you are solving a text series prediction task where you need to generate next word of a given sentence. The dataset that you are using has the following patterns (assume that you don't have an option to preprocess/modify the dataset):

   - The the the the use use use black color ball.
   - Which which which which the the the blue color boat.
   - Is is is is is is is red color car.
   - How how how the the the the green color flag.

   Assuming that the time steps are fixed to 10, answer the following questions:

   i. The above data surely has some problem. Can you relate the problem to the vanishing gradient problem?

   ii. If the problem can be associated to the vanishing gradients, then what modification to the recurrent architecture/training design would be helpful in this case? If not, please justify.

3. *(3 marks)* You have designed a novel Object Detector (IIT-Det) and have trained it on a dataset of flowers that you collected on campus. The test split of the dataset contains 5 instances of 'Roses'. The following table contains all the predictions of 'Roses' class across all test images ranked in the descending order of the predicted confidence. Compute the interpolated AP (11 points) for the

Rank	1	2	3	4	5	6	7	8	9	10
Is a correct prediction?	True	True	False	False	False	True	True	False	False	True

'Roses' class for IIT-Det. *Hint:* You can refer to the Pascal VOC 2007 evaluation protocol (see Equation 1 in Pascal VOC paper.)

4. *(2 marks)* Consider the Focal Loss $FL(p) = -(1-p)^{\gamma}(\log p)$, where γ is a hyperparameter and p is the prediction probability. If we set the value of γ=0, what is the relationship between the Focal loss and standard Cross Entropy Loss? In general, what happens to the focal loss for correctly and incorrectly classified points?

5. *(2 marks)* For bounding box localization in object detection networks, typically the L_2 norm between predicted and ground truth bounding boxes(bbox) is used as the loss function. However, the evaluation metric is IoU (Intersection over Union) between the predicted bbox and the ground-truth bbox. Show that you can have the same L_2 norm between two bounding-boxes with differing IoU values. Give an intuitive explanation of why this happens.

6. *(2 marks)*

   (a) *(1 mark)* Consider the transpose convolution of a 3×3 input with all 1s (all entries of matrix are 1) with a 7×7 filter of all 1s using stride 1 and no padding. What is the size of the result of such a transpose convolution?

   (b) *(1 mark)* Write 2D transposed convolution in matrix multiplication form. You can assume a 2×2 input and a 2×2 kernel.

2 Programming (35 marks)

- The programming questions are shared in "Assignment_4.zip". Please follow the instructions in the notebook. Turn-in the notebook via Google Classroom once you finish your work.
- Marks breakdown is as follows:

  Part-1 *(17 marks)*

  Please Note:- 15 marks for code completion and 2 marks for reporting train and test loss, final accuracies and the plots as requested.

  - Question 1: 2 marks
  - Question 2: 6 marks
  - Question 3: 9 marks

  Part-2 *(18 marks)*

  - Code : 9 marks
  - Question 4: 1 marks
  - Question 5: 3 marks
  - Question 6: 1 mark
  - Question 7: 1 mark
  - Question 8: 3 marks