Deep Learning : AI2100

Page No.:
Date. / /

Assignment 4

RAJ PATIL : CS18BTECH11039

1. a)

$$\frac{\partial E}{\partial W} = \sum_{i=1}^{3} \frac{\partial E_i}{\partial W}$$

$$\frac{\partial E}{\partial U} = \sum_{i=1}^{3} \frac{\partial E_i}{\partial U}$$

$$\frac{\partial E}{\partial V} = \sum_{i=1}^{3} \frac{\partial E_i}{\partial V}$$

b)

$$\frac{\partial E_2}{\partial W} = \frac{\partial E_2}{\partial O_2}\frac{\partial O_2}{\partial H_2}\left(\frac{\partial H_2}{\partial W} + \frac{\partial H_2}{\partial H_1}\frac{\partial H_1}{\partial W}\right)$$

$$\frac{\partial E_2}{\partial U} = \frac{\partial E_2}{\partial O_2}\frac{\partial O_2}{\partial H_2}\left(\frac{\partial H_2}{\partial U} + \frac{\partial H_2}{\partial H_1}\frac{\partial H_1}{\partial U}\right)$$

$$\frac{\partial E_2}{\partial V} = \frac{\partial E_2}{\partial O_2}\frac{\partial O_2}{\partial V}$$

c)

$$\frac{\partial E_3}{\partial W} = \frac{\partial E_3}{\partial O_3}\frac{\partial O_3}{\partial H_3}\left(\frac{\partial H_3}{\partial W} + \frac{\partial H_3}{\partial H_2}\left(\frac{\partial H_2}{\partial W} + \frac{\partial H_2}{\partial H_1}\frac{\partial H_1}{\partial W}\right)\right)$$

$$\frac{\partial E_3}{\partial U} = \frac{\partial E_3}{\partial O_3}\frac{\partial O_3}{\partial H_3}\left(\frac{\partial H_3}{\partial U} + \frac{\partial H_3}{\partial H_2}\left(\frac{\partial H_2}{\partial U} + \frac{\partial H_2}{\partial H_1}\frac{\partial H_1}{\partial U}\right)\right)$$

$$\frac{\partial E_3}{\partial V} = \frac{\partial E_3}{\partial O_3}\frac{\partial O_3}{\partial V}$$

Page No.:
Date. / /

2) a) If you notice ~~Eq~~ the gradients in the initial stages of ~~E~~ the unrolled RNN during backpropogation, it can be observed that there is a chance for vanishing / exploding gradients due to a weight matrix being repititively applied backword along the chain.

Depending on the Weight matrix's magnitude range, repetitive multiplication can lead to vanishing gradients.

Some solutions to the same ~~con~~ are:-

- Use a sensible initialization strategy to oppose the same
- Allow for a potential by-pass for the gradients to flow back. LSTM's & GRU's use this & learn the prominence of this bypass to selectively alter it.

b) i) due to discrete interpolation of the words, important words such as "The", "Which", "How" which significantly affect the semantics of the data point are pushed further away from the object of the statement due to the extra inner words.
ie. instead of
~~The~~ How ~~the~~ green color flag,
we have
How... How, the the the the, green color flag

Those extra 3-"the's" might lead to slow learning of the importance of the word "How" in this case.

ii) ~~Ans~~ # Using a learnable cell-state & parametrized skip connections as in LSTMs or GRUs can help in dealing with the problem.

3.) Computing precision vs recall values.

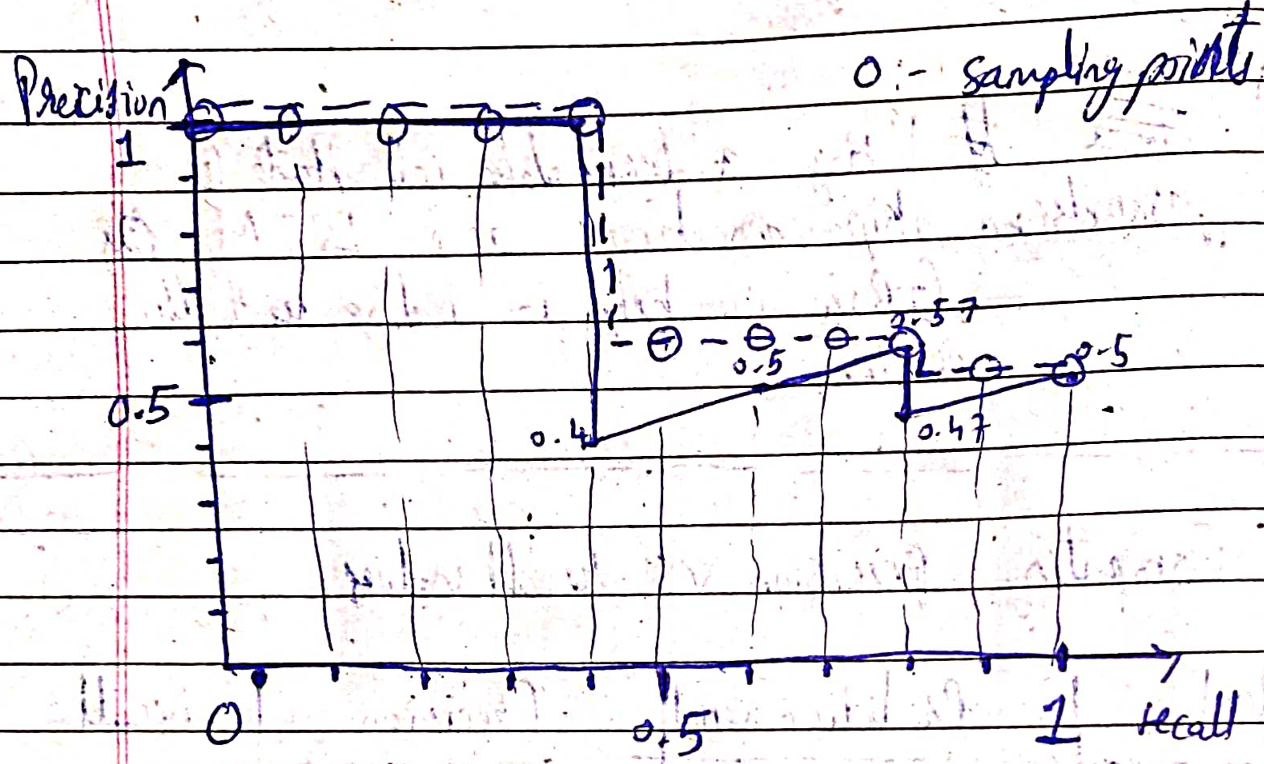
Rank	Prediction correct	Precision	Recall
1	1	1.0	0.2
2	1	1.0	0.4
3	0	0.67	0.4
4	0	0.5	0.4
5	0	0.4	0.4
6	1	0.5	0.6
7	1	0.57	0.8
8	0	0.5	0.8
9	0	0.44	0.8
10	1	0.5	1.0

for precision at ith rank, $P_i = \sum_{j=1}^{i} cp_j / i$

for recall at ith rank, $r_i = \sum_{j=1}^{i} cp_j / \sum_{j=1}^{10} cp_j$

5 in this case

plotting graph & interpolation while observing max



The dotted line represents the max that'll be sampled ~~interpolate~~ at 11 points (0 through 1.0).

$$\therefore AP = \frac{(5\times 1.0 + 4\times 0.57 + 2\times 0.5)}{11}$$

$$= \frac{5 + 1 + 2.28}{11} \rightarrow \frac{8.28}{11}$$

$$= 0.752$$
$$\cong 0.75$$

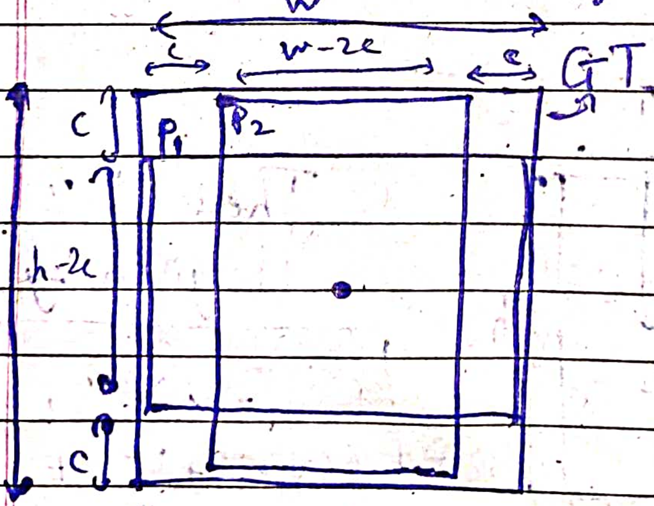
$$\boxed{\text{Interpolated AP} = 0.75}$$

5) Consider the following example :-

(representing bounding boxes by a 4-tuple)

(x, y, w, h)

center   width & height.



GT :- ground truth

P_1 :- prediction 1

P_2 :- prediction 2

fixing center for all for simplicity.

now $L_2(GT, P_1) = (w-w)^2 + (2c)^2 + 0^2 + 0^2$

$L_2(GT, P_2) = (2c)^2 + (H-H)^2 + 0^2 + 0^2$

$$IOU(GT, P_1) = \frac{(H-2c)\cancel{W}}{H\cancel{W}} = \frac{H-2c}{H}$$

$$IOU(GT, P_2) = \frac{(\cancel{H})(W-2c)}{\cancel{H}W} = \frac{W-2c}{W}$$

∴ these need not be the same.

even though they have the same L2 norm w.r.t. GT.

The L2 norm varies equally with differing entities like position & length.
But IOU varies non-uniformly with respect to them in different ranges of the variables
(for instance, when one box is inside the other or when they are completely outside)

There is no one-one correspondance between the two metrics and that is why the same L2 norm can correspond to differing IOUs.

6.) a) final output size :-

$$(3-1)(1) + (7-1) + 1 \rightarrow 9 \times 9$$

this results in a 9×9 output

b) for 2×2 input & 2×2 filter,
output size should be

3×3

which is a 9-vector when flattened

to go from a 4 vector to a 9 vector we need a

9×4 matrix

we ~~will~~ can represent the transposed convolution by

1.) Premultiplying the flattened input with a 9×4 matrix

2.) reshaping the obtained 9 vector into a 3×3 output.