NLP
Assignment 3

Alireza Moradi

April 23, 2021

1 Written

a

i. This method dampens the gradients slopes and by using the right hyperparameters, recent changes will be more important than the latest gradient. So if there's a sudden high change in gradient, it will be applied during the course of several steps instead of being applied at a single step, So we are absorbing the shock with this low variance and smoothing the curve, This will cause in severely less oscillations and faster, easier convergence.

ii. According to the formula, places with smaller gradients will get larger updates and vice-versa. This method makes the update amplitudes in all directions closer, which can reduce the shock.

b

i. Because during training we are dropping neurons with the probability of p_{drop} and keeping neurons with the probability of $p_{keep} = 1 - p_{drop}$. We want expected value of $h_{drop} = h$ so we have:

$$h_{drop} = \gamma d \circ h$$

We know expected value of $d = 1 - p_{drop}$, so the above equation leads to:

$$\gamma = \frac{1}{1 - p_{drop}} = \frac{1}{p_{keep}}$$

ii. Because dropout is a technique to prevent overfitting help the network to generalize better, and overfitting happens only during training time (when we are updating the weights). During test time we need everything that the network has learned in order to have the best performance and shutting down neurons during test time only hurts the performance as we are not learning anything.

2 Coding

a

Stack	Buffer	New dependency	Transition
[ROOT]	[I, parsed, this, sentence, correctly]		Initial Configuration
[ROOT,I]	[parsed, this, sentence, correctly]		`Shift`
[ROOT,I,parsed]	[this, sentence, correctly]		`Shift`
[ROOT,parsed]	[this, sentence, correctly]	parsed → I	`Left-Arc`
[ROOT,parsed,this]	[sentence, correctly]		`Shift`
[ROOT,parsed,this,sentence]	[correctly]		`Shift`
[ROOT,parsed,sentence]	[correctly]	sentence → this	`Left-Arc`
[ROOT,parsed]	[correctly]	parsed → sentence	`Right-Arc`
[ROOT,parsed,correctly]	[]		`Shift`
[ROOT,parsed]	[]	parsed → correctly	`Right-Arc`
[ROOT]	[]	ROOT → l	`Right-Arc`

Table 1

b

It takes n steps to move the sentence from buffer to stack and another n steps to remove the sentence from stack, so $2n$ steps in total.

e

dev	88.69
test	89.01

Table 2: UAS scores

f

	Error type	Incorrect dependency	Correct dependency
i	Verb Phrase Attachment Error	wedding → fearing	heading → fearing
ii	Coordination Attachment Error	makes → rescue	rush → rescue
iii	Prepositional Phrase Attachment Error	named → midland	guy → midland
iv	Modifier Attachment Error	elements → most	crucial → most

Table 3