Hadi Sheikhi 97521369

1.

a.

i. This formula is exponential average between $\frac{1}{1-\beta_1}$last items. So if we consider $\beta_1 = 0.9$ then we have average between last 10 gradients and if we have a jump in gradients, it should be done in consecutive steps to show up in updates.

ii. Term $\frac{m}{\sqrt{v}}$ called signal-to-noise ratio. when the signal to noise ratio goes high updates are larger. It depends on β_1 and β_2 to choose how to decrease or increase SNR. When this term is closer to 0, then update steps goes slower and stepsize is closer to 0. SNR goes close to 0 towards the optimum. An important property of Adam's update rule is its careful choice of step sizes and causes the learning process choose the true way and goes toward the optima.

b. G

i.

$$E_{p_{drop}}\left[h_{drop}\right]_i = p_{drop} \times h_{drop} + \left(1 - p_{drop}\right) \times h_{drop} = p_{drop} \times \gamma \times 1 \times h + 0$$
$$= h \rightarrow p_{drop} \times \gamma = 1 \rightarrow \gamma = \frac{1}{p_{drop}}$$

ii. In the training phase, some weights are not involving in the prediction. So we use dropout to make other weights affect the prediction. Now in the testing phase, it’s better to use all weights and it's useless to drop some neurons!

Hadi Sheikhi 97521369

2.

a.

Stack	Buffer	New dependency	Transition
[ROOT]	[I, parsed, this, sentence, correctly]		Initialize
[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→parsed	RIGHT-ARC

b. We need (n) shifts and because each two of them has relation with each other, so we need (n – 1) arcs and one arc from the root. So it should be done in "2n" steps.

e. The best dev UAS was 87.63, and the test UAS was 88.02.

f. .

i. .
Error Type: Modifier Attachment Error
Incorrect dependency: $wedding \rightarrow fearing$
Correct dependency: $heading \rightarrow fearing$

ii. .
Error Type: Coordination Attachment Error
Incorrect dependency: $makes \rightarrow rescue$
Correct dependency: $rush \rightarrow rescue$

iii. .
Error Type: Propositional Phrase Attachment Error
Incorrect dependency: $named \rightarrow Midland$
Correct dependency: $loan \rightarrow Midland$

iv.
Error Type: Modifier Attachment Error
Incorrect dependency: $elements \rightarrow most$
Correct dependency: $crucial \rightarrow most$