

1 (a) i) since momentum uses the gradient from previous steps, it has a $(1-\beta_1)$ multiplied to it, then the updates are moving in a direction as previous gradients. also if the learning rate is small then we will have converge. in the other hand if the learning rate is larger then we will miss the optimized solution.

ii) if the gradient has a big value, thus \sqrt{v} gets smaller so our parameters won't change a lot. on the other hand if gradient has a small value then \sqrt{v} gets larger & makes the updates normalized.

$$\begin{aligned} \text{b) i) } E_{p_{\text{drop}}} [h_{\text{drop}}]_i &= E_{p_{\text{drop}}} [\delta d_i \times h_i] \\ &= \delta E_{p_{\text{drop}}} [d_i \times h_i] \\ &= \delta [p_{\text{drop}} + (1-p_{\text{drop}}) \cdot h_i] \\ &= \delta (1-p_{\text{drop}}) h_i \quad \text{zhi} \\ \Rightarrow \delta &= \frac{1}{1-p_{\text{drop}}} \end{aligned}$$

ii) dropout is used to prevent overfitting in training & it shouldn't be used in evaluation because applying dropout would result in different set of prediction each time, making it hard to compare the performance of network.

2. a)

Stack	Buffer	new dependency	transition
[Root]	[I, Parsed, this, sentence, correctly]		initial
[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

stack	buffer	new depend	transition
[root, parsed, correctly]	[]		shift
[root, parsed]	[]	parsed → correctly	right-arc
[root]	[]	root → parsed	right-arc

b) 2N: because we move word from buffer to stack & remove it.

e) dev set: 88.44
test set: 89.74

f) i) error type: verb phrase error

incorrect dependency: citing → ,

correct dependency: block → citing

ii) error type: modifier attachment error

incorrect dependency: left → early

correct dependency: Friday → early

iii) error type: prepositional phrase attachment error

incorrect dependency: declined → decision

correct " : comment → decision

iv) error type: coordination attachment error

incorrect dependency: affects → one

correct " : car → one