

CS224n: Assignment 3

Amirali Abdullah

Winter 2020

1 a (i)

Momentum keeps variance in the updates low, because a large proportion of the weights update is allotted to be the same as the previous step.

1 b (i)

We must have that $\gamma = \frac{1}{1-p_i}$. (By the Linearity of Expectations, so that $E(\gamma d \odot h_i) = E(\gamma) * E(d) * E(h_i) = \frac{1}{1-p_i} * (1 - p_i) * h_i = h_i$).

1 b (ii)

We do not apply dropout at runtime (usually) since this will not help the model learn further robustness at evaluation, and when the signal is sparse we need all the inputs available.

2 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 \rightarrow I	LEFT-ARC
[ROOT, parsed, this]	[sentence, correctly]		SHIFT
[ROOT, parsed, this, sentence]	[correctly]		SHIFT
[ROOT, parsed, sentence]	[correctly]	sentence \rightarrow this	LEFT-ARC
[ROOT, parsed]	[correctly]	parsed \rightarrow sentence	RIGHT-ARC
[ROOT, parsed, correctly]	[]		SHIFT
[ROOT, parsed]	[]	parsed \rightarrow correctly	RIGHT-ARC
[ROOT]	[]	ROOT \rightarrow parsed	RIGHT ARC