ASSIGNMENT 3

1. MACHINE LEARNING & NN

a) Adam Optimizer

SGD: $\theta \leftarrow \theta - \alpha \nabla_\theta J_{minibatch}(\theta)$

Adam: $m \leftarrow \beta_1 m + (1-\beta_1)\nabla_\theta J_{minibatch}(\theta)$

$v \leftarrow \beta_2 v + (1-\beta_2)(\nabla_\theta J_{minibatch}(\theta) \odot \nabla_\theta J_{minibatch}(\theta))$

$\theta \leftarrow \theta - \alpha m / \sqrt{v}$

i) Momentum adjusts the learning rate throughout the training, giving to a new gradient update a very small rate, but a massive one to expectation of the previous updates. It stabilize the direction and speed up convergence

ii) Since v is an uncentered variance, the division by $\sqrt{v}$ normalizes the parameters update. It reduces the impact of the high values and increase the contribution of the smaller ones.

b) Dropout

i) $\mathbb{E}[d] = 0 \cdot p_{drop} + 1 \cdot (1 - p_{drop}) = 1 - p_{drop}$

Hence $\gamma = \frac{1}{1-p_{drop}}$ since we need to rescale the neurons left to maintain $\mathbb{E}_{p_{drop}}[h_{drop}]_i = h_i$

ii) The dropout suppose to prevent overfitting, that can happen only during training.

2. DEPENDENCY PARSING

a)

stack	buffer	new dependency	transition	
[ROOT]	[I, parsed, this, sentence, correctly]		Initial Configuration	0
[ROOT, I]	[parsed, this, sentence, correctly]		SHIFT	1
[ROOT, I, parsed]	[this, sentence, correctly]		SHIFT	2
[ROOT, parsed]	[this, sentence, correctly]	parsed → I	LEFT-ARC	3
[ROOT, parsed, this]	[sentence, correctly]		SHIFT	4
[ROOT, parsed, this, sentence]	[correctly]		SHIFT	5
[ROOT, parsed, sentence]	[correctly]	sentence → this	LEFT-ARC	6
[ROOT, parsed]	[correctly]	parsed → sentence	RIGHT-ARC	7
[ROOT, parsed, correctly]	[]		SHIFT	8
[ROOT, parsed]	[]	parsed → correctly	RIGHT-ARC	9
[ROOT]	[]	ROOT → parsed	RIGHT-ARC	10

b) 2n: n to move from buffer to stack & n to pop everything out of it.

e)

```
Epoch 10 out of 10
100%|
Average Train Loss: 0.06719671878802441
Evaluating on dev set
1445850it [00:00, 47540289.73it/s]
- dev UAS: 88.60
New best dev UAS! Saving model.

================================================================================
TESTING
================================================================================
Restoring the best model weights found on the dev set
Final evaluation on test set
2919736it [00:00, 70159039.72it/s]
- test UAS: 88.91
Done!
```

f)

i) Error type: Verb Phrase Attachment Error
Incorrect dependency: wedding → fearing
Correct dependency: heading → fearing

ii) Error type: Coordination Attachment Error
Incorrect dependency: makes → rescue
Correct dependency: rush → rescue

iii) Error type: Prepositional Phrase Attachment Error
Incorrect dependency: named → midland
Correct dependency: guy → midland

iv) Error type: Modifier Attachment Error
Incorrect dependency: elements → most
Correct dependency: crucial → most