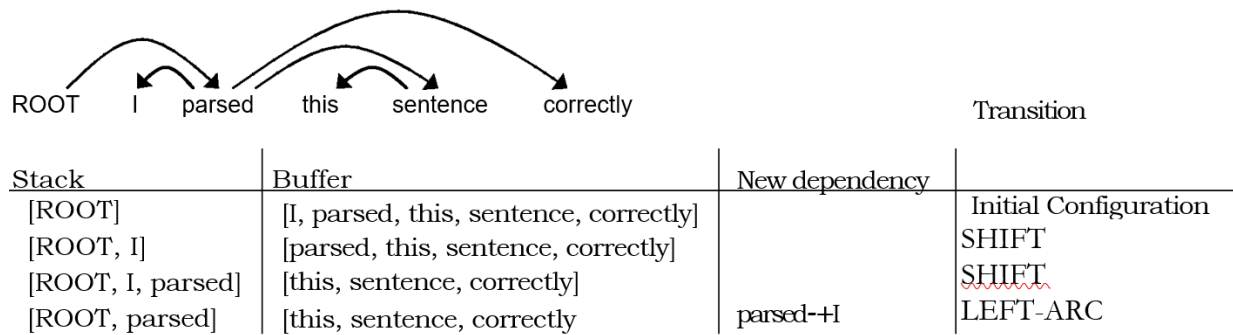


1. Machine Learning & Neural Networks (8 points)

- a. Adam uses a combination of AdaGrad and RMSProp optimization algorithms that can handle sparse gradients on noisy problems. Momentum accumulates on the updates over a time period using an exponentially weighted moving average to update the weights (parameters β_1 , β_2 , both control the decay rate over the moving averages). Adam's second beta parameter uses second moments of the gradients as gradients become sparser and/or variance increases over the trajectory of SGD.
 - ii. Adam contains a combination of two parts: a. The update direction for the i th coordinate is given by the sign of $m_{t,i}$ and b. the update magnitude for the i th coordinate is determined by the global learning rate and square root of v_t . Taking the square-root of v_t allow for this weighting method to apply more effective learning rates to all recently updated and normalized weights. Both beta parameters take advantage of the square-root of v_t . Learning becomes more effective because the application of square-root of v_t shortens the update in directions of high variance, adapting for varying reliability of the gradient in different coordinates.
- b. Gamma must equal to $1/p_{drop}$, regardless of the vector output's magnitude, the active dropout should remain the same.
 - ii. Dropout should be applied during training but not during evaluation as its purpose as a hyperparameter is to be used to avoid overfitting. After assessing the final model's parameters / hyperparameters on the evaluation set, its not best practice to tune the model any further. Using dropout on a model during evaluation will bias it's error rate estimate compared to the results of the training set.

2. Neural Transition-Based Dependency Parsing (44 points)



a.	[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. Sentences (n words) are parsed in 2(n) steps. Individual words are added or removed to the stack (n steps) and creates dependencies.

```
(base) PS C:\Users\orion.darley> python C:\Users\orion.darley\Desktop\Stanford_AI\CS224\
a3\student\parser_transitions.py part_c
SHIFT test passed!
LEFT-ARC test passed!
RIGHT-ARC test passed!
parse test passed!
```

c.

```
(base) PS C:\Users\orion.darley> python C:\Users\orion.darley\Desktop\Stanford_AI\CS224\
a3\student\parser_transitions.py part_d
minibatch_parse test passed!
(base) PS C:\Users\orion.darley>
```

d.

e. Report the best UAS your model achieves on the dev set and the UAS it achieves on the test set.

```
=====
=====
TRAINING
=====
=====
Epoch 1 out of 10
100%|██████████| 1848/1848 [02:04<00:00, 17.46it/s]
Average Train Loss: 0.186638488583428
Evaluating on dev set
1445850it [00:00, 30825045.18it/s]
- dev UAS: 83.91
New best dev UAS! Saving model.
```

```

Epoch 10 out of 10
100%| 1848/1848 [01:59<00:00, 17.57it/s]
Average Train Loss: 0.06640204614899181
Evaluating on dev set
1445850it [00:00, 46267552.99it/s]
- dev UAS: 88.68
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, 31151297.52it/s]
- test UAS: 89.20
Done!

```

- f. i. Verb Phrase Attachment Error
 - Incorrect dependency: wedding → fearing
 - Correct dependency: heading → fearing
- ii. Conditional Attachment Attachment Error
 - Incorrect dependency: makes → rescue
 - Correct dependency: rush → rescue
- iii. Prepositional Phrase Attachment Attachment Error
 - Incorrect dependency: named → midland
 - Correct dependency: guy → midland
- iv. Modified Attachment Attachment Error
 - Incorrect dependency: most → elements
 - Correct dependency: most → crucial