CS 224n Assignment 3: Dependency Parsing

1 Machine learning & Neural Network (8 points)

(a) Adam Optimizer

i. Adam uses a trick called momentum by keeping track of m, a rolling average of the gradients:

$$m \leftarrow \beta_1 m + (1-\beta_1)\nabla_\theta J_{miniatch}(\theta)$$
$$\theta \leftarrow \theta - \alpha m$$

Briefly explain how using m stops the updates from varying as much and why this low variance may be helpful to learning, overall.

Solution:
Using smoothed gradient estimator, each update will be mostly like the previous one (especially when β_1 is set as 0.9), so it removes some of the noise and oscillations that gradient descent has, so is likely to coverage faster. It is also like computing average gradient over a larger batch, so that it is closer to the actual gradient.

ii. Adam also uses adaptive learning rates by keeping track of v, a rolling average of the magnitudes of the gradients:

$$m \leftarrow \beta_1 m + (1-\beta_1)\nabla_\theta J_{miniatch}(\theta)$$
$$v \leftarrow \beta_2 v + (1-\beta_2)(\nabla_\theta J_{minibatch}(\theta) \bigodot J_{minibatch}(\theta))$$
$$\theta \leftarrow \theta - \alpha \bigodot m/\sqrt{v}$$

Since Adam divides the update by $\sqrt{v}$, which of the model parameters will get larger updates? Why might this help with learning?

Solution:
The parameter with higher $m/\sqrt{v}$ will gets larger updates. $\sqrt{v}$ estimates the uncentered variance of the gradient. Higher $m/\sqrt{v}$ means there is less uncertainty about whether the direction of m corresponds to the direction of the true. So when it is certain about the update direction, it updates faster.

(b) Dropout

i. What must equal in terms of p_{drop}? Briefly justify your answer.

Solution: $1/(1-p_{drop})$
To make $E_{p_{drop}}[\gamma d \circ h] = h$, where d is 0 with probability p_{drop}, we have:
$\gamma[p_{drop} * 0 + (1-p_{drop}) * h] = h \Rightarrow \gamma = 1/(1-p_{drop})$

ii. Why should we apply dropout during training but not during evaluation?

Solution:
Complex relationships in the training set due to sampling noise may not exist in the evaluation set. During training, we want to reduce overfitting by applying the dropout, but during evaluation, we want to use the full knowledge in the network (not dropping any hidden units). To make sure the expected output for any hidden units during training is the same as the actual output during evaluation, we need to apply the weight γ above.

2 Neural Transition-Based Dependency Parsing (42 points)

(a) **Solution**:

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 → 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) A sentence containing n words will be parsed in how many steps (in terms of n)? Briefly explain why

Solution:
A sentence containing n words will be parsed in 2n steps: one SHIFT step and one reduce (LEFT-ARC or RIGHT-ARC) step.

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

Result:
The best UAS on the dev set: 88.24
The UAS on the test set: 89.02

(f) In this question are four sentences with dependency parses obtained from a parser. Each sentence has one error, and there is one example of each of the four types above. For each sentence, state the type of error, the incorrect dependency, and the correct dependency.

Solution:

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

ii. Error type: Coordination Attachment Error
Incorrect dependency: rescue → and
Correct dependency: rescue → rush

ii. Error type: Prepositional Phrase Attachment Error
Incorrect dependency: named → Midland
Correct dependency: guy → Midland

ii. Error type: Modifier Attachment Error
Incorrect dependency: element → most
Correct dependency: crucial → most