CSC485 A1

Wilson Hsu

May 2020

Q1

(a)

Stack	Buffer	New Dependency	Transition
[ROOT]	[Nadia, rode, the, old, donkey, with, stripes]		Initial Config
[ROOT, Nadia]	[rode, the, old, donkey, with, stripes]		SHIFT
[ROOT, Nadia, rode]	[the, old, donkey, with, stripes]		SHIFT
[ROOT, rode]	[the, old, donkey, with, stripes]	rode $\xrightarrow{nsubj}$ Nadia	LEFT-ARC
[ROOT, rode, the]	[old, donkey, with, stripes]		SHIFT
[ROOT, rode, the, old]	[donkey, with, stripes]		SHIFT
[ROOT, rode, the, old, donkey]	[with, stripes]		SHIFT
[ROOT, rode, the, donkey]	[with, stripes]	donkey $\xrightarrow{amod}$ old	LEFT-ARC
[ROOT, rode, donkey]	[with, stripes]	donkey $\xrightarrow{det}$ the	LEFT-ARC
[ROOT, rode, donkey, with]	[stripes]		SHIFT
[ROOT, rode, donkey, with, stripes]	[]		SHIFT
[ROOT, rode, donkey, with]	[]	with $\xrightarrow{pobj}$ stripes	RIGHT-ARC
[ROOT, rode, donkey]	[]	donkey $\xrightarrow{prep}$ with	RIGHT-ARC
[ROOT, rode]	[]	rode $\xrightarrow{dobj}$ donkey	RIGHT-ARC

(b)

Assuming that ROOT arc is not taken into consideration. It will take us exactly $(n + (n-1) +$ Initial Config$) = 2n$ steps to parse a sentence with n words. Because there are n words that we will have to shift from buffer into the stack, which takes n steps, then there are $n-1$ arcs that we will have to create in order for the graph to be connected and acyclic. Finally, adding in the initial config step will give us $2n$ steps.

(c)

The gap degree of the non-projective tree is 1. It can be found by going through the words in the sentence one by one and counting the gap degree of each word, then take the maximum of that. With the non-projective tree given, we find that only the word "a dog" has a gap degree of 1, because it takes 2 contiguous sub-string to piece together "a dog" and its descendants. All other words have gap degree of 0.

Q2

Code file submitted

Q3

(a)

We know the mean μ and variance σ^2 is calculated by $\mu = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}$ for a $unif(a,b)$.
Therefore in order for $\mu = 0, \sigma^2 = \frac{2}{m}$.

We know $a+b=0 \sim (1)$, $\frac{(b-a)^2}{12} = \frac{2}{m} \sim (2)$

$$
\begin{aligned}
a+b &= 0 && \sim (1) \\
a &= -b \\
\frac{(b-a)^2}{12} &= \frac{2}{m} && \sim (2) \\
\frac{(-2a)^2}{12} &= \frac{2}{m} && \text{from } a=-b \\
\frac{a^2}{3} &= \frac{2}{m} \\
a &= \pm\sqrt{\frac{6}{m}} \\
a &= -\sqrt{\frac{6}{m}} && \text{Stated in A1 that } a \text{ is lower bound} \\
b &= \sqrt{\frac{6}{m}}
\end{aligned}
$$

Therefore our $a = -\sqrt{\frac{6}{m}}, b = \sqrt{\frac{6}{m}}$ for our uniform distribution $unif(a,b)$.

(b)

Without tweaking hyper-parameters:

```
Best weights will be saved to: weights.pt
Epoch  1 training loss: 0.333
         validation LAS: 0.820 (BEST!) UAS: 0.841
Epoch  2 training loss: 0.18
         validation LAS: 0.844 (BEST!) UAS: 0.864
Epoch  3 training loss: 0.158
         validation LAS: 0.857 (BEST!) UAS: 0.875
Epoch  4 training loss: 0.147
         validation LAS: 0.859 (BEST!) UAS: 0.876
Epoch  5 training loss: 0.138
         validation LAS: 0.867 (BEST!) UAS: 0.883
Epoch  6 training loss: 0.131
         validation LAS: 0.872 (BEST!) UAS: 0.887
Epoch  7 training loss: 0.126
         validation LAS: 0.873 (BEST!) UAS: 0.888
Epoch  8 training loss: 0.121
         validation LAS: 0.877 (BEST!) UAS: 0.893
Epoch  9 training loss: 0.117
         validation LAS: 0.878 (BEST!) UAS: 0.893
Epoch 10 training loss: 0.113
         validation LAS: 0.884 (BEST!) UAS: 0.898


==============================================================
==============
TESTING
==============================================================
==============
Restoring the best model weights found on the dev set.
Test LAS: 0.886       UAS: 0.901
Done.
```

After changing n epochs to 20 and hidden size to 450:

```
Epoch 19 training loss: 0.068
         validation LAS: 0.894 (BEST!) UAS: 0.909
Epoch 20 training loss: 0.0662
         validation LAS: 0.895 (BEST!) UAS: 0.908

==============================================================
==============
TESTING
==============================================================
==============
Restoring the best model weights found on the dev set.
Test LAS: 0.898       UAS: 0.911
Done.
```

CSC485, Summer 2020: Assignment 1

Family name: Hsu
First name: Wilson wei-sheng
Student Number: 1004146970
Date: May 20, 2020

I declare that this assignment, both my paper and electronic submissions, is my own work, and is in accordance with the University of Toronto Code of Behaviour on Academic Matters and the Code of Student Conduct.

Signature: Wilson Hsu