CS6957 NLP with Neural Networks
Mini Project 2

R. Rishanth (u1419542)

1.
Done

2.

Embedding	Mean		Concatenate	
	UAS	LAS	UAS	LAS
Glove 6B 50d	0.29	0.03	0.26	0.06
Glove 6B 300d	0.29	0.03	0.27	0.05
Glove 42B 300d	0.29	0.04	0.25	0.06
Glove 840B 300d	0.29	0.03	0.26	0.05

3.
Concatenating vectors works better which makes intuitive sense as it avoids the loss of information on averaging.
Different variants and sizes of Glove embeddings do not seem to make much difference at least in this setup with the given data. This could be because the model does not learn much during training due to the problem of data imbalance (SHIFT occurs 10x the second most occurring action). Different gradient descent variants and different weighting of loss functions were explored but to no avail.

4.
(a) Mary had a little lamb . (POS tags: PROPN AUX DET ADJ NOUN PUNCT)
Actions:
SHIFT SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det

(b) I ate the fish raw . (POS tags: PRON VERB DET NOUN ADJ PUNCT)
Actions:
SHIFT SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det

(c) With neural networks , I love solving problems . (POS tags: ADP ADJ NOUN PUNCT PRON VERB VERB NOUN PUNCT)
Actions:
SHIFT SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det SHIFT REDUCE_L_det

---

5.

Chen and Manning include the labels too in the input (as in the Extra credit setup) besides words in the stack and buffer and their POS tags.
The words they consider include not just top-k words on the stack and buffer but also the first and second leftmost and rightmost children of the top-2 words on the stack and the leftmost and rightmost children of the rightmost children of the top 2 words of the stack.
One of the main points they make is that such higher-order features are very important for the task of dependency parsing.
Not just the parse state but their non-linearity is also different. They use the cube function as the activation function instead of ReLU as they hypothesize that the cube function would help capture relationship between every pair of input features (words, POS, labels) and also of the triple as a whole.

Extra Credit
All results are on test set

Embedding	Mean		Concatenate	
	UAS	LAS	UAS	LAS
Glove 6B 50d	0.3	0.04	0.29	0.05
Glove 6B 300d	0.3	0.05	0.29	0.05
Glove 42B 300d	0.25	0.05	0.3	0.05
Glove 840B 300d	0.3	0.05	0.3	0.07
