Mini project 2

1. Uploaded hidden.txt
2.

Embedding	Mean		Concat	
	UAS	LAS	UAS	LAS
GloVe 6B 50d	0.32	0.26	0.68	0.62
GloVe 6B 300d	0.40	0.35	0.70	0.63
GloVe 42B 300d	0.41	0.37	0.71	0.66
GloVe 840B 300d	0.43	0.38	0.74	0.67

3. As the Glove embedding gets bigger the performance gets improved across all two types of input representation.
   The concatanation gives better result. This happens because information gets lost in the mean representation while in concatanation there is no loss of information.
4. (a) SHIFT SHIFT SHIFT SHIFT SHIFT REDUCE_L_det REDUCE_L_det REDUCE_L_aux REDUCE_R_nmod SHIFT REDUCE_R_punct
   (b) SHIFT SHIFT REDUCE_L_nsubj SHIFT SHIFT REDUCE_L_conj SHIFT REDUCE_R_conj REDUCE_R_obj SHIFT REDUCE_R_punct
   (c) SHIFT SHIFT SHIFT REDUCE_L_amod REDUCE_L_case SHIFT REDUCE_R_case SHIFT SHIFT REDUCE_L_nsubj REDUCE_L_case SHIFT SHIFT REDUCE_R_obj REDUCE_R_advcl SHIFT REDUCE_R_punct
5. One big differece between our representation and the representation in the paper is that besides using stack, buffer, pos they included the labels too. They also included higher order features, a lot more complicaated than our top-2 words from stack and buffer. Instead of using the ReLU function, they use the cube function as an activation function. They think this helps capture the relationships between all pairs of input features. These are some of the key differences I found between the two representations.