WRITE-UP

1. Unknown words
   In hmm (by mle-estimators) we handled unknown words by the signatures method. when our signatures were suffixes with length of up to 4 letters. The word's suffix had to be in one of the saved formats.
2. Pruning strategy:
   we had 2 strategies:
   - if the emission value of (word, POS) was too low we omitted the possibility without calculation (in MEMM – $if \;\; w\emptyset(x, y[:i-1], i, y_i) < 0.2$)
   - we only calculated the probabilities for the top k source cells in the matrix
3. Results
   In the hmm-greedy the test recall in detecting the labels were around 75% recall. In the HMM-Viterbi the results were around 90% recall.
   MEMM greedy gives an accuracy of 95%
   MEMM gives recall of 97%
4. Differences MEMM-Max-Ent HMM
   the are differences: HMM is relaying only on probabilities, and its difficult to handle new words or new POS orders. On the other hand the Max-Ent model has can prosses the representation of a word more accurately hence the better results
5. Differenced between Datasets
   The NER dataset contains a – lot of 'O' there for its easy to get high recall (by just label all words as 'O') while in the POS dataset the labels are more uniformly distributed.
6. Changes HMM for NER
   I would make the prediction more locally based. I think the importance of previous labels are much less significant in that case.
7. Changes MEMM
   I would try to change the loss function. And give higher loss for making an error-prediction for non-'O' labels