Exercise 3

Most frequent tag baseline

(b) Accuracy on the development set is – 0.923

HMM tagger

Our pruning policy was to set

The optimal are:

Dev: Accuracy of Viterbi hmm: 0.95627788718

Train and dev evaluation elapsed: 7.20700001717 seconds

MEMM

(c) For the Viterbi algorithm we used the following optimisations for the evaluation phase:

(1) If one of the inspected words (‘t’ ‘u’ or ‘v’) was never seen with the evalutated tag we skipped that tag evaluation(In the dynamic programming table per ‘u’, ‘v’.

(2) If the entry in the dynamic programming table is zero for the previous tags  
 (i.e in code - if DP[idx][keyt][keyu] == 0), we skipped the prediction phase for the given v with those tags (evaluating on those tags will be a waste of time we are multiplying that prediction probability with zero)

(3) we have added a parallel implementation for the evaluation phase.   
Because the system is not online, and the hyposesys doesn’t improve from the dev set, we can evaluate on several sentences at once.

(d)

0.9614 – Viterbi

0.9593 – Greedy

(e) Some error examples:

Sentence:

['The', 'bill', ',', 'whose', 'backers', 'include', 'Chairman', 'Dan', 'Rostenkowski', '-LRB-', 'D.', ',', 'Ill.', '-RRB-', ',', 'would', 'prevent', 'the', 'Resolution', 'Trust', 'Corp.', 'from', 'raising', 'temporary', 'working', 'capital', 'by', 'having', 'an', 'UNK', 'bank', 'or', 'thrift', 'issue', 'debt', 'that', 'would', "n't", 'be', 'counted', 'on', 'the', 'federal', 'budget', '.']

Errors:  
greedy is wrong on working VBG

viterbi is wrong on working VBG

Sentence:

['Another', '$', '20', 'billion', 'would', 'be', 'raised', 'through', 'Treasury', 'bonds', ',', 'which', 'pay', 'lower', 'interest', 'rates', '.']

['But', 'the', 'RTC', 'also', 'requires', '``', 'working', "''", 'capital', 'to', 'maintain', 'the', 'bad', 'assets', 'of', 'thrifts', 'that', 'are', 'sold', ',', 'until', 'the', 'assets', 'can', 'be', 'sold', 'separately', '.']

Errors:  
greedy is wrong on working NN

viterbi is wrong on working NN

Sentence:

['firstWord', 'other', 'working', 'capital', ',', 'he', 'said', ',', 'the', 'RTC', 'would', 'be', 'forced', 'to', 'delay', 'other', 'thrift', 'resolutions', 'until', 'cash', 'could', 'be', 'raised', 'by', 'selling', 'the', 'bad', 'assets', '.']

Errors:

greedy is wrong on firstWord VB

viterbi is wrong on firstWord VB

greedy is wrong on working NN

viterbi is wrong on working NN

From examining the errors we can see two major aspects

1. Most of the errors are due to Unknown words, even after replacing the “UNK” with a basic definition this is not enough information for the hypothesis to guess the right tag
2. For words with several possible tags (like working -NN/VBG) the evaluator is mostly mistaken, and probably has more weight on frequency then sentence context