1. First, we separated the corpus to 2 halves, and we regarded each word that was on the second half but wasn’t on the first half as unknown word. Then we checked the frequency of each of the tags for the unknown word. In order to approximate the probability of an unknown word to belong to one of the possible tags. Second, we added the following signatures:
2. We added the following suffixes - "ing", "ed", "s", "er", "est", "dom", "ism", "ist", "al", "ity", "ment", "ness", "tion", "ship", "ate", "en", "ify", "fy", "sion", "ize", "able", "ful", "ish", "less", "ive"
3. We added the following prefixes - "un", "de", "re", "in", "anti", "auto", "Auto", "Anti", "Un", "De", "Re", "In", "im", "Im", "Pre", "pre", "extra", "Extra", "over", "Over"

And we took the weighted sum of those two components.

1. Based on the training set, we created a dictionary containing all possible labels encountered for each word. In Viterbi algorithm, we looped, for each word, only on the possible labels listed in the dictionary for that word.
2. We received the following results:

POS Dataset:

* Greedy - 0.927686516938
* Viterbi - 0.953436199118
* GreedyMaxEnt - 0.953710397089
* MEMM - 0.956626866416

NER Dataset:

* Greedy - 0.925413936174
* Viterbi - 0.948117414401
* GreedyMaxEnt - 0.965586102602
* MEMM - 0.973690333088

#TODO: The results are just cause

1. Greedy Tagging is based on local information and that implies the correct tag of a given word, in the greedy hmm we estimated the probability of a given tag using a limited local information namely the identity of the last three tags and the current word given a tag, on the other hand in the greedy memm we extracted much more extensive set of local features for example, suffixes, prefixes, identity of previous and next words, etc… This allows more reliable local cues which will explain the better accuracy of the Greed Max Entropy. Also because Viterbi memm uses global information the accuracy does improve by much by using those extensive features.
2. #TODO: Not exactly for now.