Ian Matheson and Hannah Haines
Group: “RIP Night King”
Martin

Homework #4

For our solution we used seqlearn. Since load_conll() needs to have three columns in the data in order to read it properly, we created a dummy column that consisted of all B’s. Our features function considers the two previous words as well as the the two future words in the sentence as well as if the word starts with a capital or not. Since the line consisted of the sentence ID counter as well as the word, we had to split sentence[i] and only consider the word of the line. Our structured perceptron model that then considers these features when predicting a tag. Once tags are predicted, we wrote our results to an output file. We split our training data by placing 20% of the data into a test file to test our predicted tags and once there was an adequate precision and recall percentage we added the 20% back into our training data.