Machine Learning Homework 2
Brandon Peck
Bjp9pq

1) Hidden Markov Models
1.1 A Toy Problem
1.

START	G	C	A	C	T	G	END
H	.3+.6 $V_1 = .9$ $B_1 = START$	.9+.4+.3 $V_2 = 1.6$ $B_2 = H$	1.6+.2+.4 $V_3 = 2.2$ $B_3 = H$	2.3+.3+.2 $V_4 = 2.8$ $B_4 = L$	3+.2+.2 $V_5 = 3.4$ $B_5 = L$	3.8+.3+.2 $V_6 = 4.3$ $B_6 = L$	4.5+.3 $V_7 = 4.8$ $B_7 = L$
L	.2+.4 $V_1 = .6$ $B_1 = START$	.9+.2+.4 $V_2 = 1.5$ $B_2 = H$	1.6+.3+.4 $V_3 = 2.3$ $B_3 = H$	2.3+.2+.5 $V_4 = 3.0$ $B_4 = L$	3+.3+.5 $V_5 = 3.8$ $B_5 = L$	3.8+.2+.5 $V_6 = 4.5$ $B_6 = L$	4.5+.3 $V_7 = 4.8$ $B_7 = L$

2.
H – H – L – L – L – L

1.2 POS Tagging
1.
K = 1
V = 24509

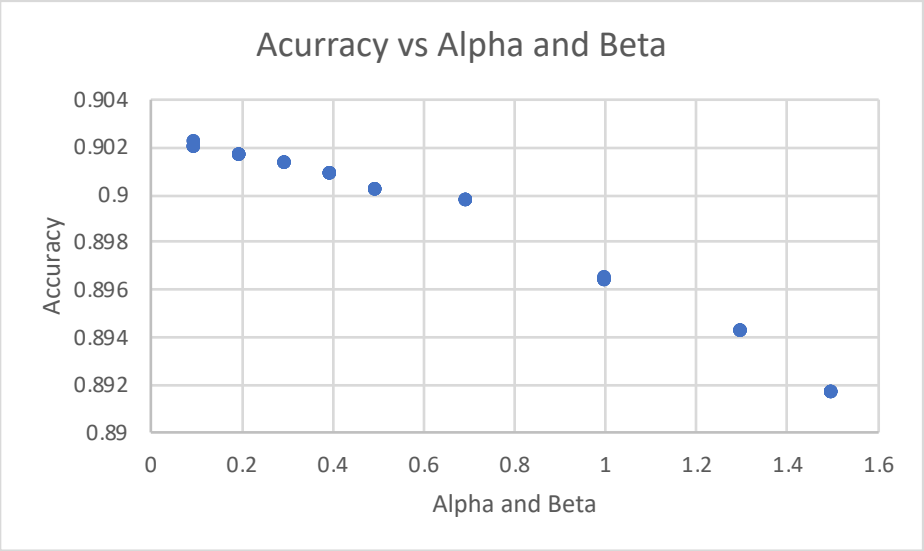
5.
In implementation, I used this equation from the Eisenstein textbook to obtain the Viterbi variable:

$$= p_{w|y}(w_m \mid y_m) \times \max_{y_{m-1}} p_y(y_m \mid y_{m-1}) \times v_{m-1}(y_{m-1}). \quad [7.53]$$

Each Viterbi variable is computed by *maximizing* over a set of *products*.

Dev Data Accuracy: 89.52%

7.



Model accuracy was maximized with the lowest values of Alpha and Beta tested, where Alpha=.1, Beta=.1 Accuracy=90.2%. This could be because when calculating the conditional probabilities lower Alpha, Beta values decrease the denominator and so reward transitions and emissions where higher counts of the variables exist.

2) Conditional Random Fields
1.
Dev Data Accuracy: 99.48%

2.
I experimented with the following features:
First letter, last letter, first two letters, last two letters, previous word, next word, word as lower case, word is upper case, word is digit, word is a title.
All of these combined resulted in 99.455% on the dev set.

The combination of features that I found that resulted in the highest accuracy was:
First letter, last letter, previous word, next word, word as lower case
Using these features resulted in 99.58% on the dev set.

3.
Dev set accuracy with averaged perceptron: 99.987%