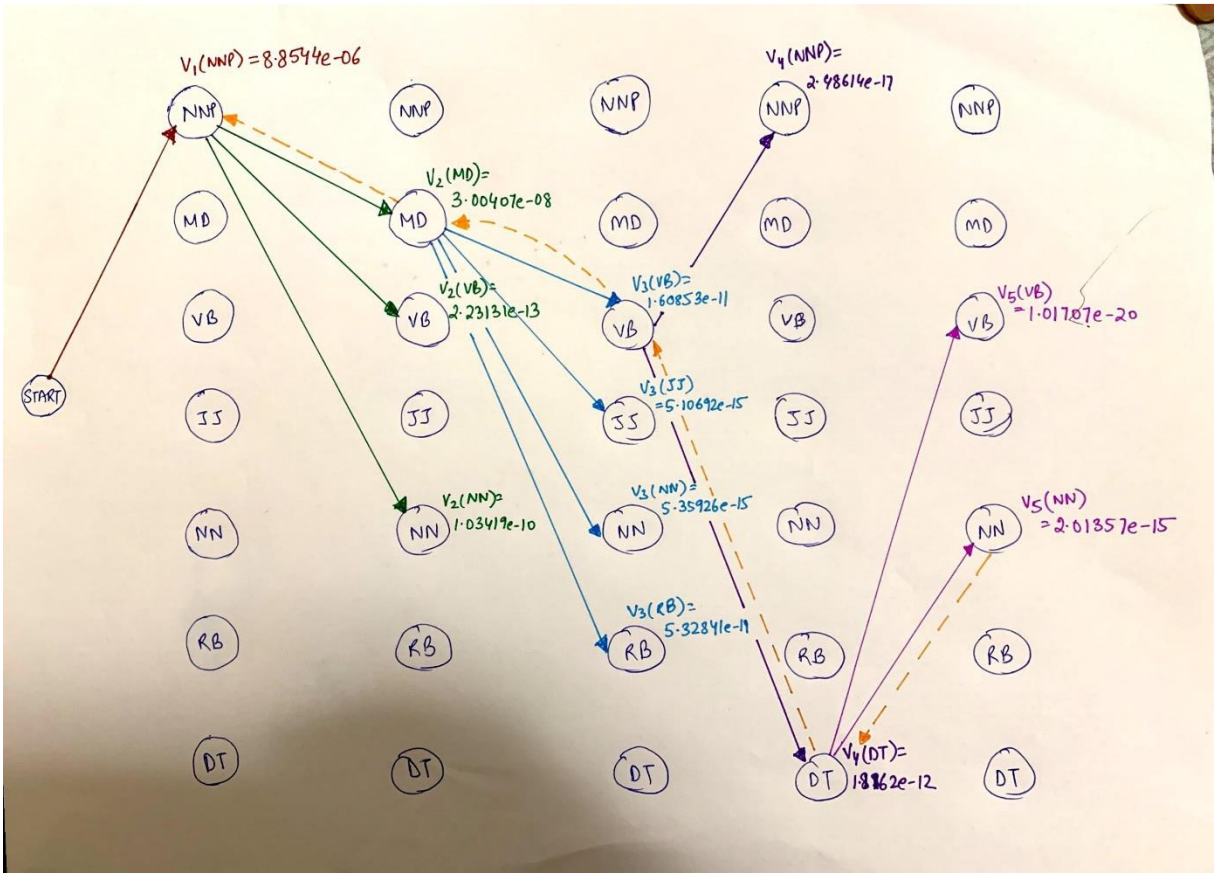
Homework 3

Shruti Agrawal(sxa178830)

1. Regular Expressions (50 points)

a) Trellis is added below. Calculations for each state follow.



The calculations shown are for the max obtained at each state for every observation state. B_ij indicates the backtracking state for the state j at observation i

Obs_State (START) → State (Janet)

$P(NNP \mid START) = 0.2767$
$P(Janet \mid NNP) = 0.000032$

$V_1^{(NNP)} = 0.2767 \times 0.000032 = 8.8544e-06$
$B_1^{(NNP)} = 0$

Obs_State (Janet) → State (will)

$P(MD \mid NNP) = 0.011$
$P(will \mid MD) = 0.308431$

$P(VB \mid NNP) = 0.0009$
$P(will \mid VB) = 0.000028$

$P(NN \mid NNP) = 0.0584$
$P(will \mid NN) = 0.0002$

~~$V_1(NNP) \times P(MD \mid NNP) \times P(will \mid MD) = 9.73984e-08$~~
~~$V_1(NNP) \times P(VB \mid NNP) \times P(will \mid VB) = 7.96896e-09$~~
~~$V_1(NNP) \times P(NN \mid NNP) \times P(will \mid NN) = 5.17097e-07$~~

$V_1(NNP) \times P(MD \mid NNP) \times P(will \mid MD) = 3.00407e-08$
$V_1(NNP) \times P(VB \mid NNP) \times P(will \mid VB) = 2.23131e-13$
$V_1(NNP) \times P(NN \mid NNP) \times P(will \mid NN) = 1.03419e-10$

$V_2(MD) = 3.00407e-08$ $B_2(MD) = NNP$
$V_2(VB) = 2.23131e-13$ $B_2(VB) = NNP$
$V_2(NN) = 1.03419e-10$ $B_2(NN) = NNP$

Obs_State (will) → State (back)

$P(VB|MD) = 0.7968$
$P(back|VB) = 0.000672$

$P(JJ|MD) = 0.0005$
$P(back|JJ) = 0.00034$

$P(NN|MD) = 0.0008$
$P(back|NN) = 0.000923$

$P(RB|MD) = 0.1698$
$P(back|RB) = 0.010446$

$V_2(MD) \times P(VB|MD) \times P(back|VB) = 1.60853e-11$
$V_2(MD) \times P(JJ|MD) \times P(back|JJ) = 5.10692e-15$
$V_2(MD) \times P(NN|MD) \times P(back|NN) = 5.35926e-15$
$V_2(MD) \times P(RB|MD) \times P(back|RB) = 5.32841e-11$

$V_3(VB) = 1.60853e-11$ $B_3(VB) = MD$
$V_3(JJ) = 5.10692e-15$ $B_3(JJ) = MD$
$V_3(NN) = 5.35926e-15$ $B_3(NN) = MD$
$V_3(RB) = 5.32841e-11$ $B_3(RB) = MD$

Obs_State (back) → State (the)

$P(NNP|VB) = 0.0322$
$P(the|NNP) = 0.000048$

$P(DT|VB) = 0.2231$
$P(the|VB) = 0.506099$

$V_3(VB) \times P(NNP|VB) \times P(the|NNP) = 2.48614e-17$
$V_3(VB) \times P(DT|VB) \times P(the|\cancel{VB}\ DT) = 1.8162e-12$

$V_4(NNP) = 2.48614e-17 \quad B_4(NNP) = VB$
$V_4(DT) = 1.8162e-12 \quad B_4(DT) = VB$

Obs. state (the) → state (bill)

$P(VB|DT) = 0.0002$
$P(\cancel{VB}\ bill|VB) = 0.000028$

$P(NN|DT) = 0.4744$
$P(bill|NN) = 0.002337$

$V_4(DT) \times P(VB|DT) \times P(bill|VB) = 1.01707e-20$
$V_4(DT) \times P(NN|DT) \times P(bill|NN) = 2.01357e-15$

$V_5(VB) = 1.01707e-20 \quad B_5(VB) = DT$
$V_5(NN) = 2.01357e-15 \quad B_5(NN) = DT$

max value at final state (bill) = 2.01357e-15
Tag for 'bill' = NN

backtracking :
$B_5(NN) = DT \quad \therefore Tag(the) = DT$
$B_4(DT) = VB \quad \therefore Tag(back) = VB$
$B_3(VB) = MD \quad \therefore Tag(will) = MD$
$B_2(MD) = NNP \quad \therefore Tag(Janet) = NNP$

Ans: Janet will back the bill
NNP MD VB DT NN

b) The code along with instructions and output file is attached.

2. Telephone Number (50 points)

a) Secretariat/NNP is/VBZ expected/VBN to/TO race/?? tomorrow/NN
For NN|race, features satisfied are f1, f6
For VB|race, features satisfied are f2, f4, f5

$$P(NN|race) = \frac{e^{0.3}e^{-0.2}}{(e^{0.3}e^{-0.2} + e^{0.75}e^{0.10}e^{0.15})}$$

$$P(NN|race) = 0.2891$$

Also,

$$P(VB|race) = \frac{e^{0.75}e^{0.10}e^{0.15}}{(e^{0.3}e^{-0.2} + e^{0.75}e^{0.10}e^{0.15})}$$

$$P(VB|race) = 0.711$$

Thus, we choose VB as the POS Tag for race.

b) the/DT race/?? for/IN outer/JJ space/NN
For NN|race, features satisfied are f1, f3
For VB|race, features satisfied are f4, f5

$$P(NN|race) = \frac{e^{0.3}e^{0.9}}{(e^{0.3}e^{0.9} + e^{0.1}e^{0.15})}$$

$$P(NN|race) = 0.7211$$

Also,

$$P(VB|race) = \frac{e^{0.1}e^{0.15}}{(e^{0.3}e^{0.9} + e^{0.1}e^{0.15})}$$

$$P(VB|race) = 0.2789$$

Thus, we choose NN as the POS Tag for race.