Yufei Lin
Mar 23^{th} 2019
Computer Linguistic

Homework 2

Question 1

a) Total Bigram Count: 126

b) Count(.) = 8, Count(. DT): 1

$$\therefore P(DT|.) = \frac{Count(.DT)}{Count(.)} = \frac{1}{8} = 0.125$$

c) Count(DT) = 22, Count(DT JJ): 3

$$\therefore P(JJ|DT) = \frac{Count(DTJJ)}{Count(DT)} = \frac{3}{22} = 0.136$$

d) Count(DT) = 22, Count(DT NN): 17

$$\therefore P(NN|DT) = \frac{Count(DTNN)}{Count(DT)} = \frac{17}{22} = 0.773$$

Question 2

a)

$$P(t_n|t_{n-2}\ t_{n-1}) = \frac{Count(t_{n-2}\ t_{n-1}\ t_n)}{Count(t_{n-2}\ t_{n-1})}$$

b) Count(IN DT) = 14, Count(NN): 12

$$\therefore P(NN|DT) = \frac{Count(DTNN)}{Count(DT)} = \frac{12}{14} = 0.857$$

Question 3

Total Word Count: 127

a) Count(JJ): 7, Count(DT): 22, Count(DT JJ): 3, Count(VBD DT) = 2, Count(VBD DT JJ): 1

$$\therefore P(JJ) = \frac{7}{127},\ P(JJ|DT) = \frac{3}{22},\ P(JJ|VBD\ DT) = \frac{1}{2}$$

$$\therefore \hat{P}(JJ|VBD\ DT) = 0.1 \cdot \frac{7}{127} + 0.7 \cdot \frac{3}{22} + 0.2 \cdot \frac{1}{2} = 0.201$$

b) Count(DT): 22, Count(VBD): 7, Count(VBD DT): 2, Count(PRP VBD) = 2,

Count(PRP VBD DT): 0

$$\therefore P(DT) = \frac{22}{127},\ P(DT|VBD) = \frac{2}{22},\ P(JJ|PRP\ VBD) = \frac{0}{2}$$

$$\therefore \hat{P}(DT|PRP\ VBD) = 0.1 \cdot \frac{22}{127} + 0.7 \cdot \frac{2}{22} + 0.2 \cdot \frac{0}{2} = 0.081$$