**CS 5316 – Natural Language Processing**

**Quiz 3**

**(Time limit: 12 minutes)**

Instructions: (1) Please write legibly. Unreadable answers will NOT be graded; (2) Write in the spaces provided for the questions only.

1. (5 points Estimate or )using (a) absolute discounting with aand (b) Kneser-Ney smoothing from a corpus with the following stats:

,, count of distinct words that follow t, count of distinct words that precede s, count of distinct bigrams = 20,000.

***See language modelling slides for formulas***

(a)

P\_absolute\_discounting = ( ( 20 - 0.75 ) /50 )=0.385

(b)

P\_kneser\_ney=max[((20 - 0.75)/50),0]/50 +(40/20000)+(0.75\*5/50)

1. (5 points)Given the following training dataset for a binary document classification problem, build a multinomial naïve Bayes classifier to predict the label for a new document with words:great, great, surprise . Show your working, and use add-one smoothing.

C1 =+, d1: great, information, excited

C2 = +, d2: excellent, news, surprise

C3 = -, d3: bad, news

C4 = -, d4: surprise

Solution:

P(+)=1/2

P(-)=1/2

P(great|+) = (1+1)/6+7

P(surprise|+) = (1+1)/6+7

P(great|-) = (0+1)/6+7

P(surprise |-) = (1+1)/6+7

Let d be the document.

P(d|+)=1/2\*(2/13)^3

P(d|-)=1/2\*(2/13)^2\*2/10