**CS5560 Knowledge Discovery and Management**

Problem Set 4 June 26 (T), 2017

**Name**: Sadanand Kallakuri

**Class ID**: 9

**I.N-Gram Consider a mini-corpus of three sentences**

**<s> I am Sam </s>**

**<s> Sam I am </s>**

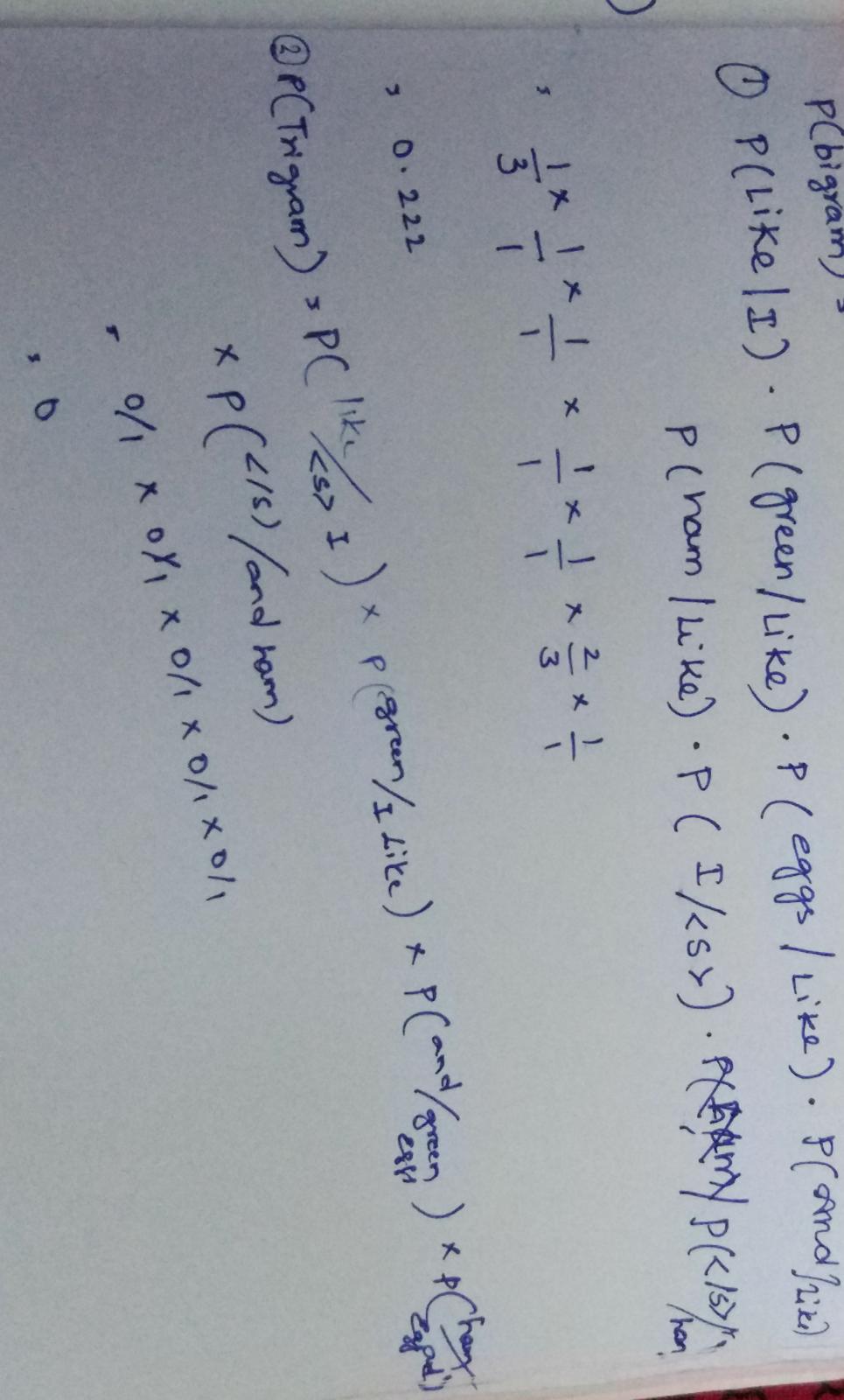
**<s> I like green eggs and ham </s>**

**1)Compute the probability of sentence “I like green eggs and ham” using the appropriate bigram probabilities.**

**2)Compute the probability of sentence “I like green eggs and ham” using the appropriate trigram probabilities. Question 1**

P( wi | wi-1 ) = count ( wi-1, wi ) / count ( wi-1 )

Probability that wordi-1 is followed by wordi = [Num times we saw wordi-1 followed by wordi] / [Num times we saw wordi-1]



**Describe the word2vec model**

Word2vec is a group of related models that are used to produce word embeddings. These models are shallow, two-layer neural networks that are trained to reconstruct linguistic contexts of words. Word2vec takes as its input a large corpus of text and produces a vector space, typically of several hundred dimensions, with each unique word in the corpus being assigned a corresponding vector in the space. Word vectors are positioned in the vector space such that words that share common contexts in the corpus are located in close proximity to one another in the space.

**Describe the differences of the following approaches   
Continuous Bag-of-Words model, Continuous Skip-gram model For**

Cbow

The context words form the input layer. Each word is encoded in one-hot form, so if the vocabulary size is V these will be V-dimensional vectors with just one of the elements set to one, and the rest all zeros. There is a single hidden layer and an output layer.  
  
The training objective is to maximize the conditional probability of observing the actual output word (the focus word) given the input context words, with regard to the weights.

Skip gram :

It is constructed with the focus word as the single input vector, and the target context words are now at the output layer.

The activation function for the hidden layer simply amounts to copying the corresponding row from the weights matrix W1 (linear) as we saw before. At the output layer, we now output C multinomial distributions instead of just one. The training objective is to mimimize the summed prediction error across all context words in the output layer.

**For the sentence “morning fog, afternoon light rain,”Place the words on the skip-gram Word2Vec model below. Draw a CBOW model using the same words.**

