CS5560 Knowledge Discovery and Management

Problem Set 4

June 26 (T), 2017

Name:

Class ID:

1. **N-Gram**

Consider a mini-corpus of three sentences

<s> I am Sam </s>

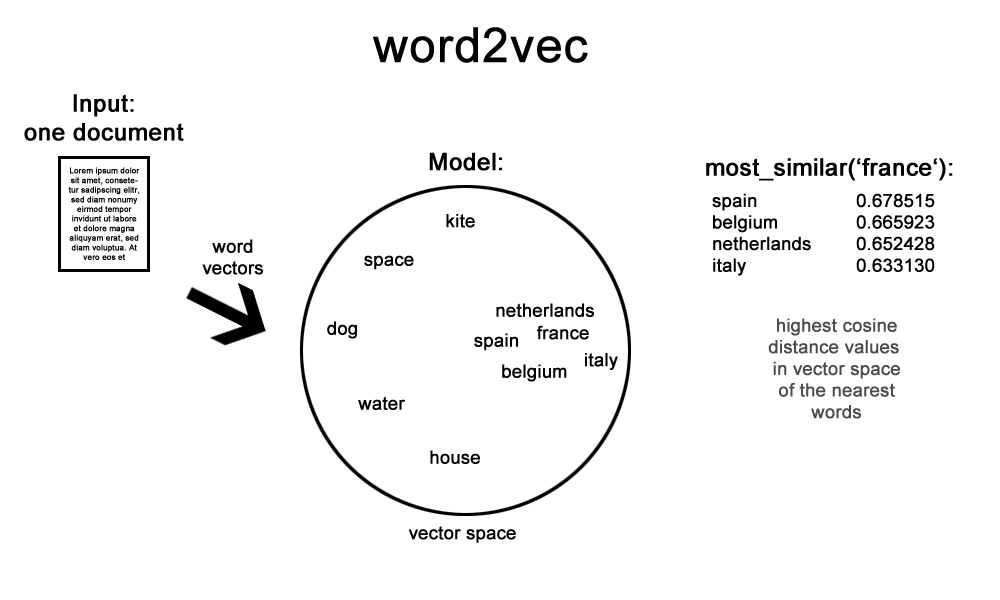
<s> Sam I am </s>

<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.
3. **Word2Vec**

Word2Vec reference: <https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>

Consider the following figure showing the Word2Vec model.



1. Describe the word2vec model
2. Describe How to extend this model for multiple documents. Also draw a similar diagram for the extended model.

Describe the differences of the following approaches

* Continuous Bag-of-Words model,
* Continuous Skip-gram model

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.

