Question 1.

Mark the following statements as true or false.

1. N-gram models do not properly represent long range dependency. T
2. LSTMs improve credit assignment compared to standard RNNs.   
   T (See 04\_LMvsClassification.pptx slide 30.)
3. Better language models exhibit higher perplexity. F
4. A high accuracy can be seen when we evaluate a classifier whose output is constant. T
5. Euclidean distance captures semantic similarity between documents well. F
6. Exact match queries in web search often produce either too few or too many results.   
   T (See sparse-vector-1.pptx slide 19.)

Question 2.

Given the five sentences below, calculate the following bigram probabilities. Assume that <s> is prepended to the beginning of each sentence, and </s> is appended to the end of each sentence.

|  |
| --- |
| “<s> you do not like them </s>”  “<s> so you say </s>”  “<s> try them </s>”  “<s> try them </s>”  “<s> and you may </s>” |

1. 1/3
2. 0.2
3. 1.0
4. 0.4
5. 1.0

Question 3.

Below are tables of bigram and unigram counts.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  |  | I | love | Italian | food | my | favorite | hobby |
|  | I | 41 | 2452 | 0 | 0 | 0 | 0 | 0 |
|  | love | 4 | 9 | 40 | 21 | 80 | 0 | 1 |
|  | Italian | 0 | 0 | 0 | 55 | 0 | 0 | 0 |
|  | food | 25 | 0 | 0 | 0 | 2 | 0 | 0 |
|  | my | 1 | 12 | 0 | 7 | 0 | 839 | 12 |
|  | favorite | 7 | 0 | 0 | 132 | 4 | 0 | 5 |
|  | hobby | 5 | 0 | 0 | 0 | 0 | 0 | 0 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| I | love | Italian | food | my | favorite | hobby |
| 32498 | 2943 | 73 | 486 | 5117 | 1310 | 105 |

1. Fill in the bigram probability table below. (Round to 4 decimal places.)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  |  | I | love | Italian | food | my | favorite | hobby |
|  | I | 0.0013 | 0.0755 | 0 | 0 | 0 | 0 | 0 |
|  | love | 0.0014 | 0.0031 | 0.0136 | 0.0071 | 0.0272 | 0 | 0.0003 |
|  | Italian | 0 | 0 | 0 | 0.7534 | 0 | 0 | 0 |
|  | food | 0.0514 | 0 | 0 | 0 | 0.0041 | 0 | 0 |
|  | my | 0.0002 | 0.0023 | 0 | 0.0014 | 0 | 0.1640 | 0.0023 |
|  | favorite | 0.0053 | 0 | 0 | 0.1008 | 0.0031 | 0 | 0.0038 |
|  | hobby | 0.0476 | 0 | 0 | 0 | 0 | 0 | 0 |

1. Fill in the bigram probability table below with add-1 smoothing. Assume that the size of the vocabulary is 8453. (Round to 4 decimal places.)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  |  | I | love | Italian | food | my | favorite | hobby |
|  | I | 0.0010 | 0.0599 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
|  | love | 0.0004 | 0.0009 | 0.0036 | 0.0019 | 0.0071 | 0.0001 | 0.0002 |
|  | Italian | 0.0001 | 0.0001 | 0.0001 | 0.0066 | 0.0001 | 0.0001 | 0.0001 |
|  | food | 0.0029 | 0.0001 | 0.0001 | 0.0001 | 0.0003 | 0.0001 | 0.0001 |
|  | my | 0.0001 | 0.0010 | 0.0001 | 0.0006 | 0.0001 | 0.0619 | 0.0010 |
|  | favorite | 0.0008 | 0.0001 | 0.0001 | 0.0136 | 0.0005 | 0.0001 | 0.0006 |
|  | hobby | 0.0007 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |

Question 4.

Answer the following questions.

1. Briefly explain the training objectives (tasks) of CBOW and skip-gram, respectively.  
     
   CBOW predicts the masked word based on the sum of the embeddings of the context words.  
   Skip-gram predicts each word in the context of the given word.
2. What is the role of softmax in these methods?  
     
   In both CBOW and skip-gram, the model produces scores over the vocabulary, and softmax is used to convert these scores into probabilities (by taking the exponential of the scores and normalizing).

Question 5.

We have a dataset with three class labels, namely Urgent, Normal, and Spam. The following is the confusion matrix for these classes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | True Class |  |  |
|  |  | Urgent | Normal | Spam |
| Predicted Class | Urgent | 7 | 8 | 9 |
|  | Normal | 1 | 2 | 3 |
|  | Spam | 3 | 2 | 1 |

1. Complete the tables below.

|  |  |  |
| --- | --- | --- |
|  | Class: Urgent |  |
|  | True  Urgent | True  not |
| System  Urgent | 7 | 17 |
| System  not | 4 | 8 |

|  |  |  |
| --- | --- | --- |
|  | Class: Normal |  |
|  | True  Normal | True  not |
| System  Normal | 2 | 4 |
| System  not | 10 | 20 |

|  |  |  |
| --- | --- | --- |
|  | Class: Spam |  |
|  | True  Spam | True  not |
| System  Spam | 1 | 5 |
| System  not | 12 | 18 |

1. Compute the Precision, Recall and F1-score for Each Class and complete the table below.

|  |  |  |  |
| --- | --- | --- | --- |
| Class | Precision | Recall | F1-score |
| Urgent | 0.29 | 0.64 | 0.4 |
| Normal | 0.33 | 0.17 | 0.22 |
| Spam | 0.17 | 0.08 | 0.11 |

7/(7+17) = 7/24 = ~0.291

7/(7+4) = 7/11 = ~0.64

F1: 2\*(precision\*recall)/(precision+recall)