**M.E. Computer Science & Engineering, 1st Year, 2nd Semester Examination, 2025**

# Natural Language Processing

**Time – 3 hours** **Full Marks - 100**

**Answer any five questions**

1. a. Compare Needleman-Wunsch algorithm and the Levenshtein Edit Distance algorithm. *5*
   b. Discuss and write the Damerau-Levenshtein edit distance algorithm. *5*
   c. Find out the edit distance and alignment between the two strings "*processing*" and "*impression*", considering the following costs for the edit operations. Highlight the alignment in the edit distance matrix. *10*

```
insertion cost = deletion cost = substitution cost = 1
```

2. a. Briefly discuss about Word2Vec embedding. *5*
   b. Explain the intuition behind using noisy channel model for spelling correction.Discussthe noisy channel model for non-word spelling correction. *6*
   c. Formulate an HMM model for part of speech tagging. Just give the equations, no explanation required. *4*
   d. Write and explain the Forward algorithm for computing the probability of an observation sequence, given the model. *5*

3. a. Compare Sparse Vectors vs Dense Vectors in NLP. *5*
   b. Define hyponym and hypernym. Discuss the properties of hyponymy. Differentiate between hyponym and instance relations. *5*
   c. Given the following term-context matrix, compute which of the following word pairs - [data, information] and [lemon, orange], is more similar according to distributional similarity using PPMI and add-*2* smoothing. *10*

| term \ context | computer | digital | boil | result | fry |
|---|---|---|---|---|---|
| data | 2 | 2 | 0 | 1 | 0 |
| information | 1 | 6 | 0 | 2 | 0 |
| eggplant | 0 | 0 | 1 | 0 | 1 |
| potato | 0 | 0 | 1 | 0 | 2 |

4. a. Explain why precision and recall are not suitable for machine translation evaluation. *2*
   b. Briefly discuss about the BLEU MT evaluation metric, its strengths and weaknesses. *5*
   c. Discuss how hypothesis recombination can be used to reduce the search space in SMT decoding. *3*

d. Compute the alignment probabilities and the translation probabilities according to the EM algorithm assuming no NULL token and only 1-to-1 alignments for the following toy parallel training corpus. Show the first 3 iterations. *10*

| Translation pair id | Source Language | Target Language |
|---|---|---|
| 1 | big house | casa grande |
| 2 | the house | la casa |

5. (a) How to find the correct class c from a document d in Generative and Discriminative classifiers? *4*
(b) Consider the following 6 features for Sentiment Classification task in a Movie Review document:
count(positive Lexicon) ε doc,
count(negative Lexicon) ε doc,
1 if "no" ε doc 0 otherwise,
count(1st and 2nd pronouns) ε doc,
1 if "!" ε doc 0 otherwise,
log (word count of doc)
The feature weights are: 2.5, -5.0, -1.2, 0.5, 2.0, 0.7 and bias = 0.1
The first 5 feature weights are: 3,2,1,3,0 and word count of doc is 66.
Note: sigmoid(0.833) = 0.70 and ln(66) = 4.19
Compute whether the document belongs to positive or negative sentiment. *6*
(c) Derive the cross-entropy loss for a single observation x in Logistic Regression. Find the cross-entropy loss in the Question 5.(b) when the true output is positive (y=1) and the true output is negative (y=0). *6*
(d) What is Overfitting in a Machine Learning model? How overfitting can be avoided in Logistic Regression and in Neural Networks? *4*

6. (a) Two non-linear activation functions besides sigmoid are tanh and RelU. Please explain why sigmoid can be applied only to the final output layer of Neural Network and not tanh or ReLU. Which activation functions can be applied at the intermediate hidden layers of a Neural Network? *4*
(b) Describe a simplified feed-forward neural language model with a sliding window of length 3 (three) words. *6*
(c) Discuss the Byte Pair Encoding (BPE) token learner and segmenter algorithms with examples. *6*
(d) Stemming reduces terms to stems, chopping off affixes crudely. Explain why stemming is used in Information Retrieval Systems instead of Morphological Parsing. *4*

7. (a) How RNNs and their variants LSTMs introduce a different way of representing time? *4*
(b) Describe the four architectures for NLP tasks with RNNs. *6*
(c) What roles are played by Forget gate, Add gate and Output gate in LSTM architecture? *6*
(d) What is the problem with passing context vector from the final hidden state of the encoder? How is this problem solved with LSTM attention? *4*

8. (a) How Large Language models differ from N-gram based Language models? *3*
   (b) Discuss the perplexity Intrinsic Evaluation method for language models along with examples. *4*
   (c) What are the three architectures for Large Language Models? *3*
   (d) What is top-k sampling in LLM decoding? What are the problems with top-k sampling? What is nucleus sampling? What is temperature sampling? *6*
   (e) What is Low Rank Adaptation (LoRA)? *4*