BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

I SEMESTER 2023-2024 CS F429 –Natural Language Processing, Mid Semester Exam (Closed Book)

Time: 90 mins **Weightage: 30%** **Date: 8th Oct 2023**

1. Compare
   (a) language model with n-gram vs sequence modeling with RNN
   (b) sigmoid vs tanh
   (c) PoS vs NER
   (d) CBoW vs Word2Vec [5]

2. Consider period (.) disambiguation problem. Design a good set of five features and a two-layer NN to classify that each occurrence of a period is end-of-the-sentence (EoS), Shortening of the word, or other. [4]

3. (a) Consider the following models of logistic regression for a binary classification with a sigmoid function *g*:
   M1: $P(Y=1|X,w_1,w_2)=g(w_1X_1+w_2X_2)$
   M2: $P(Y=1|X,w_1,w_2)=g(w_0+w_1X_1+w_2X_2)$
   We have three training examples: $x^{(1)} = [1, 1]^T$ $x^{(2)} = [1, 0]^T$ $x^{(3)} = [0, 0]^T$ $y^{(1)} = 1$ $y^{(2)} = -1$ $y^{(3)} = 1$.
   Does it matter how the third example is labelled in model M1 i.e., would the learned value of $w = (w_1, w_2)$ be different if we change the label of the third example to -1? Does it matter in model M2? Briefly explain your answer.
   (b) What are the three updated weights' values (w0, w1, w2) after applying one iteration of mini-batch gradient decent on M2 with initial weights (0.1, 0.2, 0.2), learning rate 0.1, batch size 3? [3+3]

4. (a) Suppose a classifier predicts each possible class with equal probability. If there are 10 classes, what will the cross-entropy error be on a single example and why?
   (b) Suppose we have a loss function $f(x; y; \theta)$, defined in terms of parameters, inputs x and labels y. Suppose that, for some special input and label pair (x0, y0), the loss equals zero. Justify whether the "it follows that the gradient of the loss with respect to θ is equal to zero" is true or false.
   (c) What is the effect of window size on word2vec embeddings? [4]

5. (a) Give 2 examples of how we can evaluate word vectors as intrinsic and extrinsic evaluation.
   (b) An RNN model predicts a positive sentiment for the following sentence:
   Yesterday turned out to be a terrible day. I overslept my alarm clock, and to make matters worse, my dog ate my homework. At least my dog seems happy...
   Why might the model misclassify the appropriate sentiment of the sentence? [4]

6. (a) Consider the following sentences with their class:

Language independent system data driven dependency parsing	*-dependency parsing*
Algorithm deterministic incremental dependency parsing	*-dependency parsing*
Transition based techniques projective dependency parsing	*-dependency parsing*
Structured models sentiment analysis	*-sentiment analysis*

   Use Naïve Bayes to classify "*dependency sentiment analysis data*"

   (b) Suppose you run your classifier on 10 documents and recorded the predicted classes (DP or SA) with actual class given in table below. Construct a confusion matrix and compute the macro-averaged and micro-averaged precision.

Doc.	Actual	Predicted	Doc.	Actual	Predicted
1	DP	SA	6	SA	SA
2	DP	DP	7	DP	SA
3	SA	DP	8	SA	SA
4	DP	DP	9	DP	DP
5	SA	SA	10	SA	DP

   We changed the classifier slightly and found that only prediction for 6th doc. changed from SA to DP. What will be the effect of the change on the averages? [7]