



Iran University of Science & Technology
School of Computer Engineering

Assignment #4

Natural language processing

BY:

DR. Behrouz Minaei, Fall 2024

Teaching Assistants:

Iman Barati

Mahdi Khoursha

Due: 1403/09/26

Contents

Notes	3
Problem 1	4
Problem 2	4
Problem 3	4
Problem 4	5

Notes

1. Submit the answers in a complete PDF file and the code for the questions in the .ipynb format (including the notebook cell outputs) in a compressed file named HW4_StudentID.zip by the specified deadline.
2. A total of 72 + 48 hours of delay in submitting the answers is allowed across all projects. After that, for each additional day of delay, 10% of the score will be deducted.
3. If a student submits the project earlier than the deadline and achieves 75% of the score, up to 24 hours will be added to their allowable delay time.
4. The maximum delay for submitting each assignment is 5 days, and after 5 days, submission will not be accepted.
5. It is important to note that the explanation of the code and the obtained results must be included in the PDF file. Code without a report will result in a score deduction.
6. The evaluation of the assignment will be based on the correctness of the solution and the completeness and accuracy of the report.
7. Assignments must be completed individually, and group work on assignments is not allowed.
8. Please allocate sufficient time for the assignment and avoid leaving it until the last days.
9. You can ask your questions in the relevant group.

good luck.

Problem 1

In the field of NLP, Named Entity Recognition (NER) is a crucial task often approached as a sequence labeling problem. Recently, a novel approach called **GPT-NER** has been proposed, which leverages large language models (LLMs) to address the challenges of NER by reframing it as a text generation task. Read the paper [GPT-NER: Named Entity Recognition via Large Language Models](#) and answer the following questions based on the methodology and findings presented in the article: (The first four sections of the article are sufficient to answer these questions.) **(30 points)**

- Explain the core difference between traditional NER approaches and GPT-NER.
- How does GPT-NER transform the sequence labeling task into a text generation task? Provide an example.
- Describe the role of the "self-verification" strategy in GPT-NER. Why is it important for addressing hallucination issues?

What are the three main components of the prompt construction in GPT-NER? Provide a brief explanation of each.

Problem 2

Maximum Entropy Classifiers **(20 points)**

- Define generative and discriminative models and explain their key differences.
- Why are Maximum Entropy models more suitable for classification tasks compared to Naive Bayes?
- Provide a real-world example of using a Maximum Entropy model in natural language processing and explain how to construct suitable features for this model.
- What is the importance of using appropriate feature weights in Maximum Entropy models? Why does this improve model accuracy?

Problem 3

Named Entity Recognition **(20 points)**

- Explain the concept of Named Entity Recognition (NER) and provide examples of its applications.
- Using the IOB tagging scheme, label the following sentence:
- Sentence: "Ali studies at IUST in Tehran"
- What are the differences between Maximum Entropy Markov Models (MEMMs) and Conditional Random Fields (CRFs)?
- How are Named Entity Recognition (NER) models used in an Information Retrieval system?

Problem 4

Refer to the **RNN_NLP_Task** notebook, read the problem thoroughly, and complete the specified steps. **(40 points)**