Hands-on session 5

Transformers

Instructor: Dr. Souvik Chakraborty

APL 745: Deep Learning for Mechanics

10 April, 2024

Submission due on 23 April, 2024

Instructions:

(i) You are allowed to discuss in a group; however, you must submit your own handwritten homework copy (no computer-typed submission will be accepted). Further, copying homework from your friends is forbidden. If found copied, the homework submission will be considered invalid for all the students involved and will be graded zero.

(ii) Write all the steps, including your reasoning and the formulae you referred to, if any. If sufficient reasoning is not provided and steps are unclear, step marks will not be given.

(iii) For practical submissions, the codes are accepted in *.ipynb* or *.py* format.

(iv) Unless mentioned otherwise, only *numpy*, *pytorch* and *matplotlib* libraries may be used. Direct commands for algorithms to be implemented are not allowed.

(v) The *.rar* file containing all submission related files shall be named in the format, **Name_Entrynumber.rar**

Question 1. Building transformer from scratch for Natural Language Processing

Task: Develop transformer architectures with three encoders and three decoders from scratch for language translation. Below are the detailed instructions:

1) **Dataset:** Download the data from this link (eng_frenh.csv). The data contains 2 columns, one column has English words/sentences and the other one has French words/sentences. The unique values are different because the same English word has a different French representation.
2) **Model Implementation:** Implement a transformer architecture with three encoders and three decoders from scratch for language translation from English to French using Pytorch and/or NumPy. **Ensure that your implementations do not use any pre-built transformers model found in *PyTorch* or *TensorFlow*. However, you can use the basic *PyTorch* functionalities, including automatic differentiation (*autograd*) and backpropagation.**
3) **Data Preprocessing:(30 marks)**
   - Load the dataset and examine its contents to understand the available features (English and French sentences). Split the data into training and testing data sets such that training data contains 70% and testing data contains the rest of the 30%.
   - Encode the words into the word embedding vectors using an appropriate tokenizer and embedding space. Please follow the link for the details
   - Write a code for a suitable positional encoding function that generates positional vectors representing the positions of words in a sentence and can be added to the corresponding word embedding vectors.
4) **Model and Training configurations:(40 marks)**
   - A basic transformer model with configurations for translating English sentences to French with
     - Configuration 1: A single attention heads
     - Configuration 2: Multi attention heads of 8
     - Configuration 3: Multi attention heads of 32
     - Configuration 4: Multi-attention heads of 8 and increased number of encoders and recorders to 5.
   - Train each model configuration using the training data. Monitor the training loss and validation accuracy.

5) **Loss Function:**
   - Choose categorical cross entropy as the loss function for training the models.
6) **Evaluation: (10 marks)**
   - After training each model configuration, evaluate the model's performance on the test dataset.
7) **Plotting:(10 marks)**
   - Plot the epoch-wise loss for each training configuration (1, 2, 3 and 4). Create separate plots for each model configuration.
8) **Comparison and Analysis:(10 marks)**
   - Compare the performance of the first three models and comment on how the usage of multiple attention heads improves the results. Also, comment on the accuracy of the model configurations 2 and 4 models based on the test accuracy. Provide an analysis of your findings on the strengths and weaknesses of each model.

Use ADAM optimizer and 200 epochs, while training the models.