

UNITED STATES MILITARY ACADEMY

HOMEWORK 4

CS386X: APPLIED NEURAL NETWORKS

SECTION I1

MAJ SARAH SMITH

BY

CDT VICTOR OLAWALE-APANPA '24, CO C3

WEST POINT, NEW YORK

7 MAY 2024

U6A

I CERTIFY THAT I HAVE COMPLETELY DOCUMENTED ALL SOURCES THAT I USED TO COMPLETE THIS ASSIGNMENT AND THAT I ACKNOWLEDGED ALL ASSISTANCE I RECEIVED IN THE COMPLETION OF THIS ASSIGNMENT.

_____ I CERTIFY THAT I DID NOT USE ANY SOURCES OR RECEIVE ANY ASSISTANCE REQUIRING DOCUMENTATION WHILE COMPLETING THIS ASSIGNMENT.

SIGNATURES: _____

A handwritten signature in black ink, appearing to be 'V. Olawale-Apanpa', written over a horizontal line.

The primary purpose of this assignment is to conduct sentiment analysis and develop a model that accurately predicts the label for a given tweet about the weather (Positive, Negative, Neutral)

Question 4:

Accuracy is an appropriate metric, however, while it provides a quick snapshot of overall model performance, it is not the most informative. Precision and recall help understand the effectiveness of a model in identifying each class. Precision measures the accuracy of positive predictions, which is beneficial for high cost false positives. Recall measures the ability of a model to find all the relevant cases within a dataset important when there is a high missing positive cost. These metrics better convey the model's performance, highlighting overall correctness but also how well the model manages the trade-offs between different types of errors. For example, if the dataset has more neutral sentiments but the cost of misclassifying negative sentiments as positive is very high, precision and recall for the negative class become more important.

Question 3:

Parameters are the components of the model that are learned from the training data. Parameters include weights and biases that are adjusted through a learning algorithm (ex, backpropagation). The goal of training a model is to find the optimal set of parameters that minimize the loss function, which measures how well the model's predictions match the actual data.

Alternatively, Hyperparameters, govern the entire training process. These values are set before training begins and are used to control the behavior of the learning algorithm. Hyperparameters are crucial because they directly influence the architecture of the model and how the model is trained.

1. Batch Size – Increasing the batch size can lead to faster computational speed by leveraging parallel processing capabilities, but a smaller batch size leads to more noise in the gradient estimates, which might increase training time.
2. Learning Rate – A higher learning rate can accelerate training but risks overshooting while a low learning rate may result in slow convergence
3. Number of Training Epochs – Increasing the number of epochs allows the model more iterations to learn from the data, potentially improving accuracy at the risk of overfitting if not monitored
4. Optimizer – AdamW, for example, provides faster convergence in the initial training phase
5. Scheduler – can enhance training by adapting the learning rate based on training progress

In this assignment a pre-trained BERT model ("google-bert/bert-base-cased") and compared its performance before and after fine-tuning. Initially, the base model showed low accuracy, around

35.61%, indicating it struggled with sentiment analysis. However, after fine-tuning on the given dataset, accuracy improved to 97.51%, showing a substantial improvement. This highlighted the importance of fine-tuning pre-trained models for task-specific datasets, allowing them to better capture nuances and improve predictive accuracy. This comparison highlighted the importance of fine-tuning for specialized tasks like sentiment analysis, enhancing the model's effectiveness.

Question 1:

The distribution of text lengths shows significant variation, primarily ranging between 20 to 140 characters with notable peaks around 40-50 and 110 characters. Regarding class distribution, the dataset appears relatively balanced across Negative, Neutral, and Positive sentiments, especially in the training data. This balance is beneficial for training as it minimizes the risk of bias. However, I noticed a slight imbalance in the validation dataset, where there are fewer instances in the Positive category compared to Negative and Neutral. This could potentially affect the model's accuracy on unseen data if similar imbalances exist. Lastly, the text length distributions are consistent across different classes in both the training and validation datasets. This uniformity is beneficial as it ensures that the input features concerning text length are consistent across different sentiments, preventing the learning process from being biased by varying text lengths among classes.

Question 2:

1. `finiteautomata/bertweet-base-sentiment-analysis`
Bertweet is designed specifically for Twitter data, utilizing a RoBERTa-based architecture that excels in handling social media text. Its pre-training on a corpus of Tweets means it is well-adapted to informal language, like slang and abbreviations. The Bertweet model has been directly fine-tuned for sentiment analysis, enhancing its ability to discern sentiment effectively in such environments. Additionally, despite its robust capabilities, Bertweet is relatively efficient computationally, making it a better choice.
2. `google-bert/bert-large-cased`
BERT Large cased has 3 times more parameters than the base model. More parameters allow a model to learn more complex functions or relationships within the data which allows for the model to capture finer details and nuances.

WORKS CITED

Ali, Saleem CDT D3 '24. Assistance given to author, Electronic Communication. CDT Ali and I communicated over iMessage as well as met in person. I asked him for clarification for this assignment. I was initially confused about the purpose as not many changes needed to be made in order to get a working notebook. However, he corrected me the assignment was primary analysis. West Point, NY. 7 May 2024

Gemini1.0Pro. Assistance given to the author, AI. I used the following prompt in Gemini1.0Pro: I repeatedly pasted in error messages that I was receiving from using specifically Dataloaders and others errors in both version of my code. Additionally, I asked in Question 2 about the benefits of uniformity. Google, (<https://gemini.google.com>). West Point, NY, 6MAY2024.

Training a Classifier — PyTorch Tutorials 2.2.1+cu121 Documentation.
pytorch.org/tutorials/beginner/blitz/cifar10_tutorial.html.