Fine-Tuning RoBERTa

By: Shruti Wagle

Code Description

Data loading & pre-processing:
We start by loading the AG News dataset using PyTorch. The data is tokenized using a RoBertTokenizer. The dataset is finally split into the train, test, and validation datasets with batch size 256.

Creating & training the model:
This program uses a pre-trained RoBERT Classification model with an Adam optimizer and Cross-Entropy loss function. We freeze the pre-trained RoBERTa parameters and reduce the total trainable parameters to 593668. Finally, we train the model for 3 epochs and evaluate it against the validation dataset. The model with the highest validation accuracy is saved and used to measure the accuracy of the test dataset.

Evaluating the results:
This model gives the best validation accuracy of 90.61% in epoch 3 with a training accuracy of 87.31%, train loss of 0.370, and validation loss of 0.267. These results match the expected values. As expected, the final test accuracy is 90.59%, and the test loss is 0.277. The model is then checked against other metrics giving the following results.

Precision: 0.906
Recall: 0.762
F1 Score: 0.828

Scope for Improvement

We can reduce loss and improve performance metrics by implementing the following:

1. Use different learning rates, optimizers, & smaller batch sizes.
2. Run more epochs.
3. Pre-process the data and fine-tune it. Apply data augmentation.
4. Focus on improving recall by adding more layers to the model.
5. Implement a different model