Tasks

1. Fine-tune the BERT model on the training dataset. You can use any network architecture on top of the BERT, the final layer must be softmax for classification. Choose a suitable number of epochs and a suitable batch size. Test the fine-tuned model on the test dataset. Report test accuracy.

2. Look at the actual classification predictions of the fine-tuned model on the test dataset. Take at least 10 examples the model predicts correctly and at least 10 examples which the model predicts incorrectly, from these make at least 3 general observations (an observation should not be specific to any one example).

3. Come up with at least 5 examples demonstrating BERT model's contextual embeddings analogous to the "dog" and "cat" example towards the end of the slides. Include your examples in the report along with the cosine similarity scores.