Neural Networks and Deep Learning with Business Applications

SP Jain School of Global Management

Assignment L9

Sentiment Analysis using BERT for Embeddings

In this assignment, you will develop a sentiment analysis model leveraging the BERT model for word embeddings, combined with your choice of classification techniques. You will explore the power of BERT embeddings in understanding the context of words in text data and apply advanced classification methods for sentiment analysis.
The dataset consists of text reviews, each labeled as either "positive" or "negative".
Tasks for Submission:

1. Implement a sentiment analysis model using BERT for embeddings.
    - Utilize the BERT model to convert text data into embeddings.
    - You are free to choose any classification technique (like LSTM, CNN, or others) to classify the sentiments as positive or negative based on the BERT embeddings.
2. Train and test your model.
    - Split the dataset into training and testing sets.
    - Use an appropriate loss function and optimizer for binary classification.
    - Train your model and report its accuracy on the testing set.
3. Analyze and discuss your results.
    - Provide insights into the performance of your model and the effectiveness of BERT embeddings in sentiment analysis.
    - Include plots of training loss and accuracy over time.
    - Present a confusion matrix of your model's predictions on the testing set.

Submission Format: Your submission should include the following components:

- Source code of your sentiment analysis model utilizing BERT for embeddings.
- A comprehensive report describing your choice of classification technique, model architecture, training process, and a detailed analysis of the results.
- Relevant visualizations or plots that support your analysis.

Additional Instructions:

- Ensure that your code is well-commented and adheres to best practices for readability and maintainability.
- The report should be clear and succinct, highlighting the critical aspects and findings of your project.