### Introduction to Natural Language Processing

# Assignment 4

# ELMO

Hardik Sharma

#### Introduction

In this detailed report, I meticulously document my journey implementing the ELMO model, a cutting-edge technique in natural language processing. Exploring its complex architecture and contextual word embeddings, I navigate through the training process, sharing the methodologies used and challenges faced. Through evaluating its performance across diverse NLP tasks, I highlight both strengths and limitations.

We also train the model on a downstream classification task, combining the embeddings in several possible ways.

For the learnable function, I use a single hidden layer with relu sigmoid activation.

#### Results

Below is the graph for the model accuracy on the test data, plotted vs the time taken for training the model. From the raw data itself.

Model Accuracies:

Frozen Lambdas :  $\lambda_0 = 0.5203, \lambda_1 = 0.0315, \lambda_2 = 0.8535$  Learned Lambdas :  $\lambda_0 = 1.6256, \lambda_1 = 0.0126, \lambda_2 0.6851$ 

Aggregation	Accuracy
Trainable	88.88%
Frozen	87.24%
General Function	89.65%

### Conclusion

Overall, the learable parameter performs better than when the model has fixed parameter, a learnable function also performs well, but isn't as good as the learnable parameter one, and also takes some more time to train due to a larget increase in the number of parameters.

ELMO architecture, therefore, does outperform previous architectures with similar setting, however, it takes a longer time to train. Therefore, the choice of the underlying model depends on the tasks aat hand.