Text and Sequences Assignment Report

Nikhil Kumar Sampath

IMDB:

Objective:

To build an embedding model that can train on the IMDb reviews dataset and learn the features in the reviews and finally make a binary prediction of whether a review is negative or positive.

Varied Sample Sizes: (Scratch Models)

#No.	Training	Validation	Test	Performance on Test Set (Loss, Accuracy)
Model 2	100	10000	5000	(0.693, 0.517)
Model 3	1000	10000	5000	(0.524, 0.744)
Model 4	25000	10000	5000	(0.290, 0.908)
Model 5	**35000**	**10000**	**5000**	**(0.257, 0.926)**

Findings:

- The initial 2 models i.e., Model 2 and Model 3 were primarily trained on a small amount of data and the network architecture used to run this model was very simple just with an embedding layer. The model's performance was not that bad but we can't even say that it is exceptionally good as well.

- When it comes to Model 4 and Model 5, we started to use Convolution 1D along with Embedding Layers and Dense Layers to see if the model's performance changed, we did observe a few things:

  a. An increase in training samples provides the model to learn more and more about the data, in this case providing more samples helped the model to learn the positive and negative elements in a review, which then helped to generalize well on the test set.

b. The use of Conv1D along with the Embedding Layer provided more robustness in the model's performance, Conv1D works through the filters and as the filters move along the review the model tends to learn more and more about the review which helps to better predict when applied on the test set.

c. Change in Hyper-Parameters is a key thing during the model-building phase, by observing the prior model's performance as a modeller it is important to know which knob has to be turned to increase the model's performance.

d. Embedding Vector Dimensions, Learning Rate, Conv1D Layers, Dense Layers, Dropout Rate and Nodes are the key hyperparameters that have been fine-tuned along the way to increase the model's generalization ability.

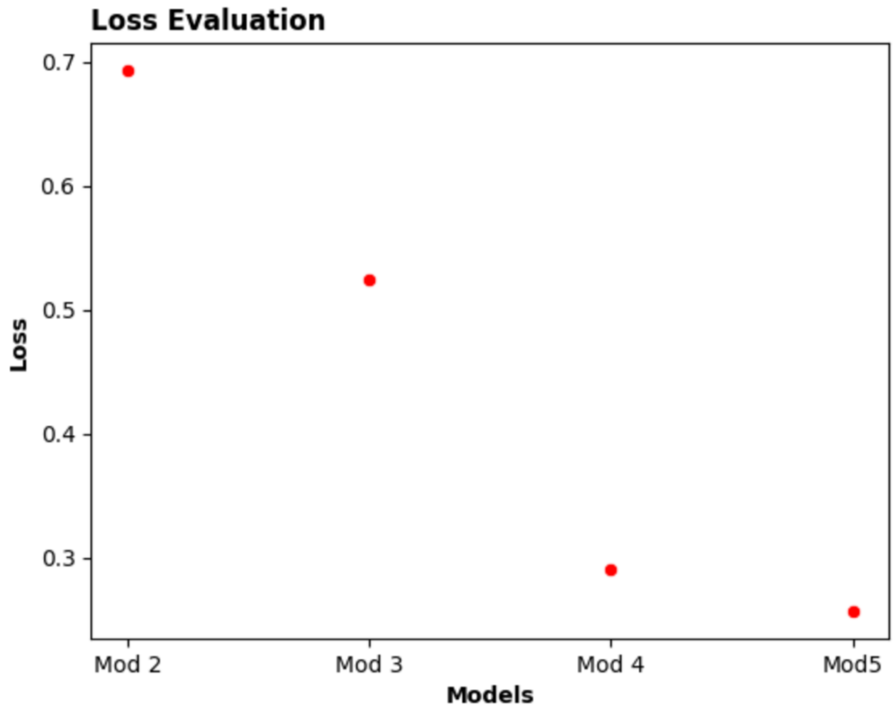
- Plotting the training and validation loss/accuracy is helpful to pick the sweet spot where the model starts to overfit.

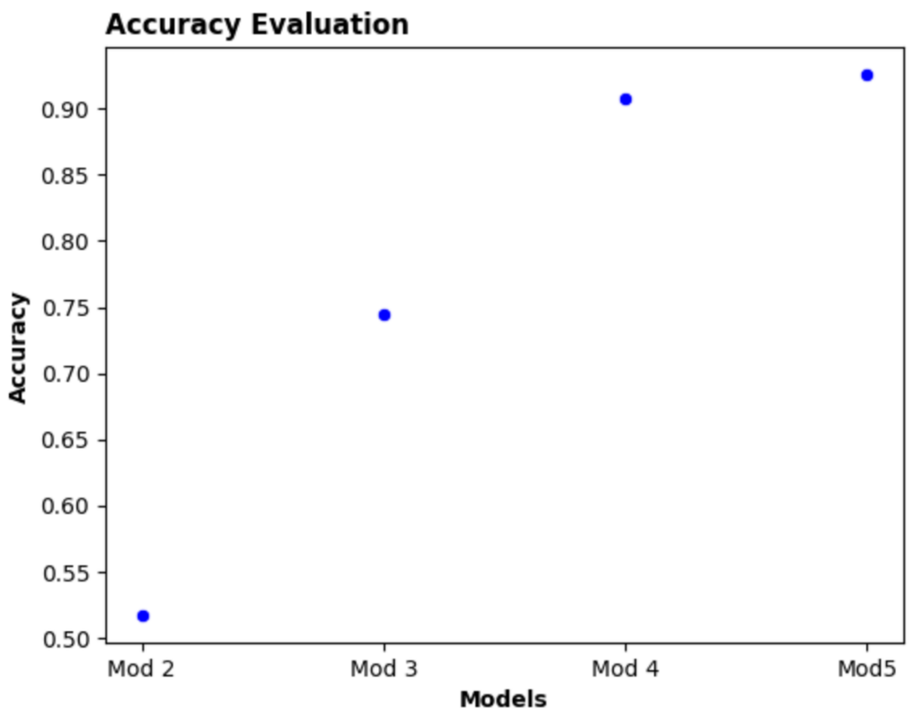
- Building a complex network architecture to let the model overfit and then fine-tuning the parameters is proven to be the best way to work where the end goal is to better generalize on the test set.

- Last but not least whenever the model tends to not perform well always provide more data for the model to learn.

Final Evaluation: (Base Models)

As we can see below the final model that was deemed to be the best performer across all the 4 models built can be counted as the 5th Model, this model was built with 35000 Training Samples, 1000 Validation Samples and 5000 Test Samples.





Pre-Trained Network:

The vector dimension of the GloVe network was used to build these models and the network constitutes LSTM layers.

Varied Sample Sizes: (Pre-Trained Models)

#No.	Training	Validation	Test	Performance on Test Set (Loss, Accuracy)
Pre_Model1	100	10000	5000	(0.699, 0.452)
Pre_Model2	1000	10000	5000	(0.440, 0.798)
Pre_Model3	10000	10000	5000	(1.013, 0.524)
Pre_Model4	15000	10000	5000	(0.696, 0.374)
Pre_Model5	**15000**	**10000**	**5000**	**(0.332, 0.851)**

Findings:

- Underfitting is a serious concern than Overfitting is the major learning through these models that have been built.

- If a model tends to underfit it implies that the performance on the training set itself is bad and poor which in turn indicates that the performance on the unseen data is going to be even worse. Model 3 and Model 4 can be considered as underfit models, these models didn't achieve greater performance on the training set and when tested out performance on the test set was even worse.

- Sometimes complex architectures are not always the best option to build a better model, simple architectures with fewer input layers and nodes can learn better features. Model 5 was built with a simple network architecture and as a result, the model didn't underfit and the performance on the training set was quite good. Eventually, test set performance also stood out better than others.

Final Comments:

As we can see below the final model that was deemed to be the best performer across all the 5 models built can be counted as the 5th Model, this model was built with 15000 Training Samples, 1000 Validation Samples and 5000 Test Samples.