

1 Logistic Regression Model

For this task, we have been provided with the IMDB Sentiment Dataset. This dataset has been split into training, validation, and test sets in the ratio **7:3:10** =>

Training Set: 17500, **Validation Set:** 7500, and **Test Set:** 25000

The dataset is loaded into a *DataLoader* for training, validation, and testing with a custom batching function specified using the `collate fn` parameter. Before loading the dataset into the *DataLoader*, the input vectors have been padded with `< pad >` token, which maps to dictionary index '1' to ensure all input vectors in a batch are of the same length.

The model used is a *Logistic Regression* with an embedding layer that takes in text and outputs a sentiment score, processed by a fully connected layer for the final output. Loss Function used is *BCEntropyLossWithLogits* which also internally applies *sigmoid* activation function on the output layer to fit model predictions between values of 0 and 1.

The model is trained using *Adam* optimizer with a learning rate of 1e-2 which controls the step size of the optimization algorithm.

Logistic Regression Model:

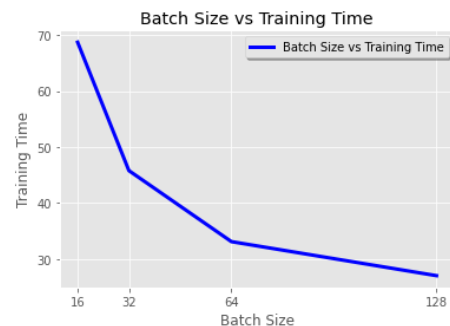
LR((embedding): Embedding(25002, 100)

(fc): Linear(in_features=100, out_features=1, bias=True))

1.1 Batch Training Time Data

S.No	Batch Size	Training Time (sec)
1.	16	68.71
2.	32	45.78
3.	64	33.09
4.	128	27.01

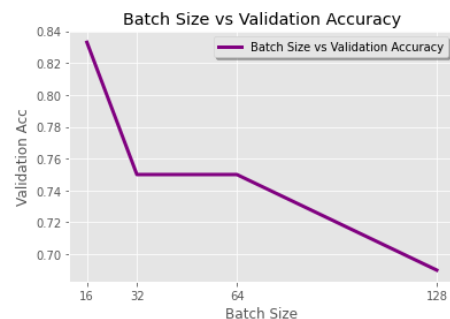
Table 1: Batch Size vs Training Time



1.2 Validation Accuracy Data

S.No	Batch Size	Validation Accuracy
1.	16	0.833
2.	32	0.75
3.	64	0.75
4.	128	0.69

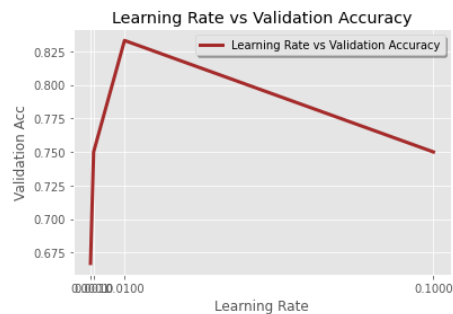
Table 2: Batch Size vs Validation Accuracy



1.3 Learning Rate vs Validation Accuracy

S.No	Learning Rate	Validation Accuracy
1.	1e-1	0.75
2.	1e-2	0.833
3.	1e-3	0.75
4.	1e-4	0.667

Table 3: Learning Rate vs Validation Accuracy



1.4 Best Model

Fixed Parameters:

Input Dimension: Vocab Size = 25002

Output Dimension: 1

Loss Function: BCEWithLogitsLoss (Binary Cross Entropy)

Hyper-Parameters:

S.No	Hyper-Paramter	Value
1.	Epochs	7
2.	Optimizer	Adam
3.	Learning Rate	1e-2
4.	Embedding Dimension	100
5.	Batch Size	16

Table 4: Hyper-Parameters for Logistic Regression

Training Accuracy: 0.917

Validation Accuracy: 0.833

Test Accuracy: 0.75



Figure 1: Plot of Training vs Validation Loss

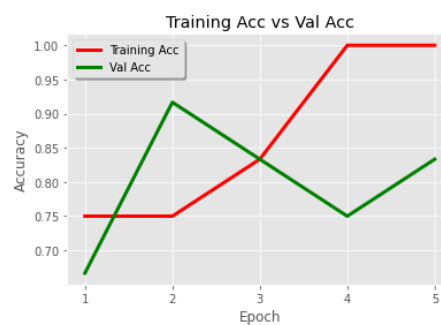


Figure 2: Plot of Training vs Validation Accuracy

1.5 Observation for the errors that Logistic Regression makes

To analyze errors, we examine the incorrect predictions made by the best-performing model in the first 10 sentences in the dev set. In our case, this model was trained for 5 epochs.

```

Input Sentence: Abu , THE THIEF OF BAGDAD , helps King Ahmed regain his kingdom from a wicked sorcerer.<br /><br />As Europe was going to war and significant sections of the world was going
Gold Label: pos
Pred Label: 0
Input Sentence: Nahhh ! Leila ( Grace Mills ) is a teenager turned on to Satan ( and LSD ) by her archaeologist fiancé Richard . There 's a neighborhood hippie demon cult hanging out at the
Gold Label: neg
Pred Label: 0
Input Sentence: While the original First Blood had its far - fetched moments , it was at least exciting in parts . In Rambo : First Blood , Part II the emphasis is shifted very much onto com
Gold Label: neg
Pred Label: 1
Input Sentence: It 's curious that the two stars of Meet The People were a pair of movie stars who went into the new medium of television and became even bigger successes and who both went :
Gold Label: neg
Pred Label: 0
Input Sentence: Brokedown Palace is not the kind of movie I would ever like to see . I also did not like the movie when some Aussie man smuggled drugs in Thailand and accused Claire Danes ar
Gold Label: neg
Pred Label: 0
Input Sentence: First of all let me say that I had to think a lot about writing a comment for this movie . The best review for this kind of Cinema can be just the silence . Movie addressed t
Gold Label: neg
Pred Label: 1
Input Sentence: Yes , Be My Love was Mario Lanza 's skyrocket to fame and still is popular today . His voice was strong and steady , so powerful in fact that MGM decided to use him in The G
Gold Label: pos
Pred Label: 0
Input Sentence: This film is self indulgent rubbish . Watch this film if you merely want to hear spoken Gaelic or enjoy the pleasant soundtrack . Watch for any other reason and you will be c
Gold Label: neg
Pred Label: 1
Input Sentence: The first half hour of the movie had a steady pace and introduced the characters . however all of a sudden everything was happening too quick , a lame reason for Akshey Kumar
Gold Label: neg
Pred Label: 1
Input Sentence: This was a less than exciting short film I saw between features on Turner Classic Movies recently . While the film popped out due to its very intense Technicolor , the film :
Gold Label: neg
Pred Label: 0

```

Figure 3: Dev Set: Gold Labels vs Pred Labels

As you can clearly see, the **Logistic Regression** Model made 4 incorrect predictions. The model appears to make simple decisions based on the positive or negative sentiment of the word embeddings in a sentence, without taking into account how the tone may have changed as the sentence progresses. This could be due to the limited capabilities of Logistic Regression as it cannot remember any context from the previous inputs.

2 LSTM Model

Before loading the dataset into the DataLoader, we need to pad the input sequences the same way we did for our Logistic Regression model, but there is a catch when we do padding for **LSTM**. Since we are building an LSTM model, we need it to process and backpropagate loss for only the non-padded elements to reduce the computational overhead. So we also pass a length tensor for each batch that contains the length of each input vector in a batch that helps the LSTM cell recognize the non-padded elements for each input sequence. This is done by packing the padded input sequence (using *pack_padded_sequence*) before feeding it into the LSTM cell and unpacking the output (using *pad_packed_sequence*) of the LSTM cell.

The model used is an *LSTM* with an embedding layer that takes in the text to give an embedded output. Dropout is applied to this embedded output, and it is fed to an LSTM cell. Since we are using a Bi-LSTM in this case, we concatenate the final *forward* (*hidden[-2, :, :]*) and *backward* (*hidden[-1, :, :]*) hidden layers and apply dropout, processed by a fully connected layer for the final output.

The model is trained using Adam optimizer with a learning rate of 1e-2 and using Binary Cross Entropy Loss Function.

LSTM Model:

```

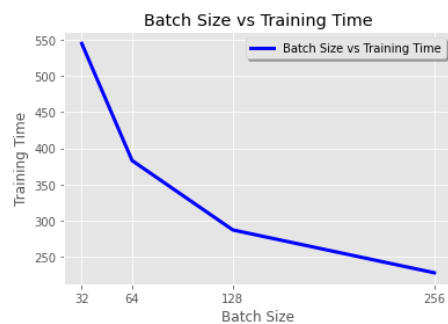
LSTM( (embedding): Embedding(25002, 100)
(lstm): LSTM(100, 256, num_layers=2, dropout=0.2, bidirectional=True)
(fc): Linear(in_features=512, out_features=1, bias=True)
(dropout): Dropout(p=0.2, inplace=False) )

```

2.1 Batch Training Time Data

S.No	Batch Size	Training Time (sec)
1.	32	545.12
2.	64	382.94
3.	128	286.81
4.	256	227.46

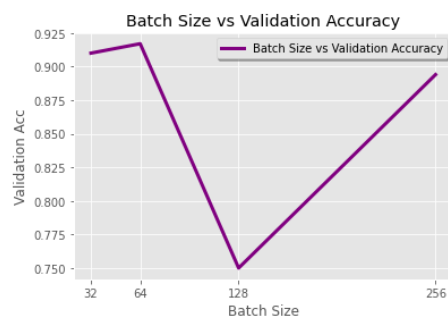
Table 5: Batch Size vs Training Time



2.2 Validation Accuracy Data

S.No	Batch Size	Validation Accuracy
1.	32	0.91
2.	64	0.917
3.	128	0.75
4.	256	0.894

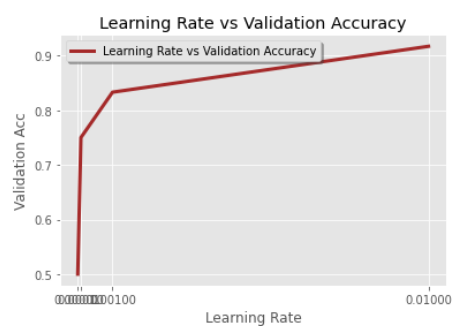
Table 6: Batch Size vs Validation Accuracy



2.3 Learning Rate vs Validation Accuracy

S.No	Learning Rate	Validation Accuracy
1.	1e-2	0.917
2.	1e-3	0.833
3.	1e-4	0.75
4.	1e-5	0.50

Table 7: Learning Rate vs Validation Accuracy



2.4 Best Model

Fixed Parameters:

Input Dimension: Vocab Size = 25002

Hidden Dimension: 256

Output Dimension: 1

Number of Layers: 2

Type of LSTM Cell: Bi-Directional

Dropout: 0.2

Loss Function: BCEWithLogitsLoss (Binary Cross Entropy)

S.No	Hyper-Paramter	Value
1.	Epochs	5
2.	Optimizer	Adam
3.	Learning Rate	1e-2
4.	Embedding Dimension	100
5.	Batch Size	64

Table 8: Hyper-Parameters for LSTM

Hyper-Parameters:

Training Accuracy: 0.93

Validation Accuracy: 0.917

Test Accuracy: 0.95

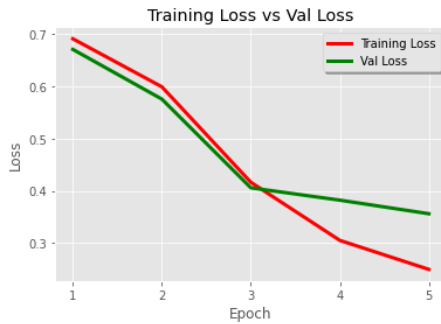


Figure 4: Plot of Training vs Validation Loss

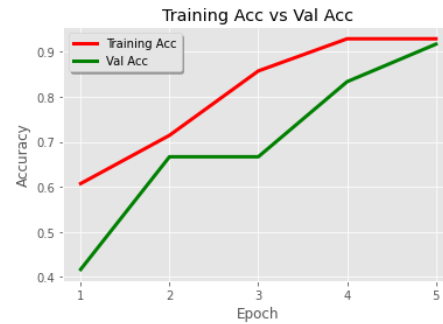


Figure 5: Plot of Training vs Validation Accuracy

2.5 Observation for the errors that LSTM makes

To analyze errors, we examine the incorrect predictions made by the best-performing model in the first 50 sentences in the dev set as our model didn't make any incorrect predictions in the first 10 sentences so we had to increase our observation pool. In our case, this model was trained for 5 epochs.

```

Input Sentence: I 'm a pretty old dude , old enough to remember the taste of Oreos and Coke as they were 50 - 55 years ago , when every taste for a kid was fresh . I wish I have somehow set
Gold Label: 0
Pred Label: 1
Input Sentence: Somehow , this movie manages to be invigorating , bittersweet , and heartwarming at the same time . Stars like Tony Shalhoub ( from Providence ) bring the tale to life . The
Gold Label: 0
Pred Label: 1
Input Sentence: Being a big fan of Corman 's horror movies I expected from his western a bit more than I got . Well , I was entertained all right . I had almost as many laughs as watching M
Gold Label: 1
Pred Label: 0
Input Sentence: Jason Alexander is a wonderful actor , but it 's ridiculous to cast him as a cuddly romantic lead . The fact that he dances so well , croons so effectively , and throws hims
Gold Label: 1
Pred Label: 0
Input Sentence: That movie was awesome ! I ca n't get over it 's songs . I think I 'm a little too old for musicals , but that movie deserves some credit here , guys ! My especial favorite
Gold Label: 0
Pred Label: 1
Input Sentence: Even if you know absolutely nothing about Ireland , you have to love " My Left Foot " ( and especially Daniel Day - Lewis ' performance in it ) . He plays cerebral palsy - a
Gold Label: 0
Pred Label: 1
Input Sentence: Its not sophisticated , and nobody in the credits had a great career , but taken as a whole , because there are no famous personalities ; the film seems more realistic than
Gold Label: 0
Pred Label: 1
Input Sentence: People do n't seem to be giving Lensman enough credit where its due . A few issues have been overlooked which are key to understanding the Lensman experience.<br /><br />The
Gold Label: 0
Pred Label: 1
Input Sentence: I have to say as being a fan of the man who created Halloween / The Fog / Christine / The Thing - probably his best films.<br /><br />Then you got this POS . I ca n't logica
Gold Label: 1
Pred Label: 0

```

Figure 6: Dev Set: Gold Labels vs Pred Labels

As you can clearly see, the LSTM Model made 9 incorrect predictions in the first 50 sentences. The LSTM model appears to have a better ability to grasp the context and nuances of sentiment variations in long text sequences. However, it still struggles to accurately classify sentiment when more subtle shifts occur within a long paragraph. We probably can get better results using global attention mechanism.

3 Model Correctness Testing

The scores produced by a mini-batch model on GPU differ from those produced by a single instance model. This is due to two main factors:

- 1) The mini-batch score vector has the same size as the batch, while a single instance model produces a score vector with only one element.
- 2) The scores for the same input text are significantly different between the models, as seen with the *Logistic Regression* and *LSTM* output scores. This could be because different network architectures (Logistic Regression, LSTM) learn the weights differently for the same instance, resulting in different predictions. Logistic Regression is a sort of linear approximation of the input text. At the same time, LSTM considers more non-linearity as it applies multiple 'tanh' activation functions internally to maintain short-term and long-term memory.