REPORT ON ASSIGNMENT 4 (MACHINE LEARNING LAB)

❖ PROCEDURE

The *Amazon Product Review* dataset was downloaded that included

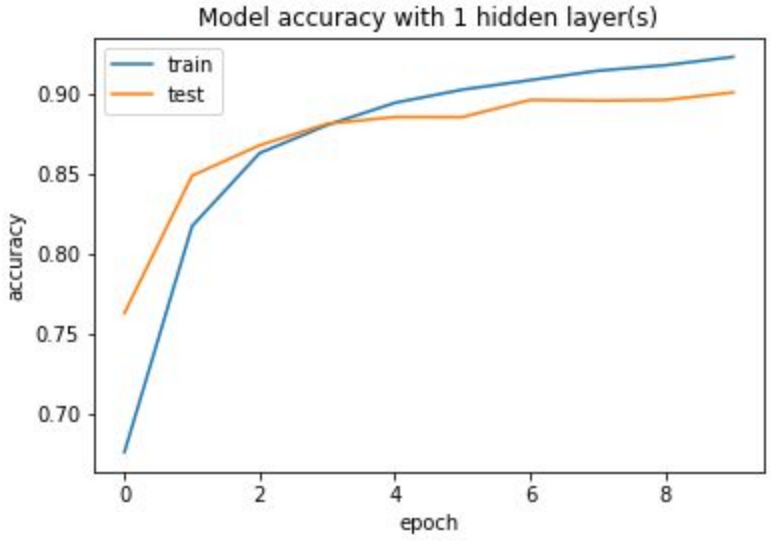
400,000 reviews from different customers having positive (label 1) and negative (label 2) reviews.

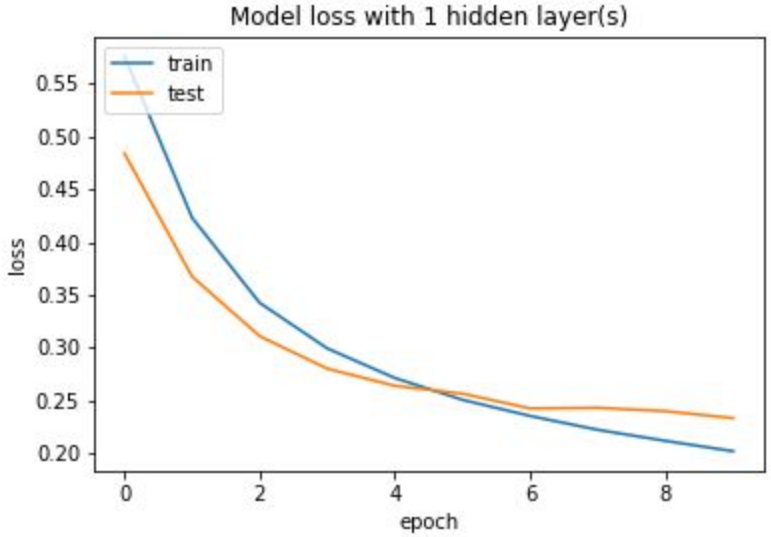
★ PRE-PROCESSING OF THE DATASET

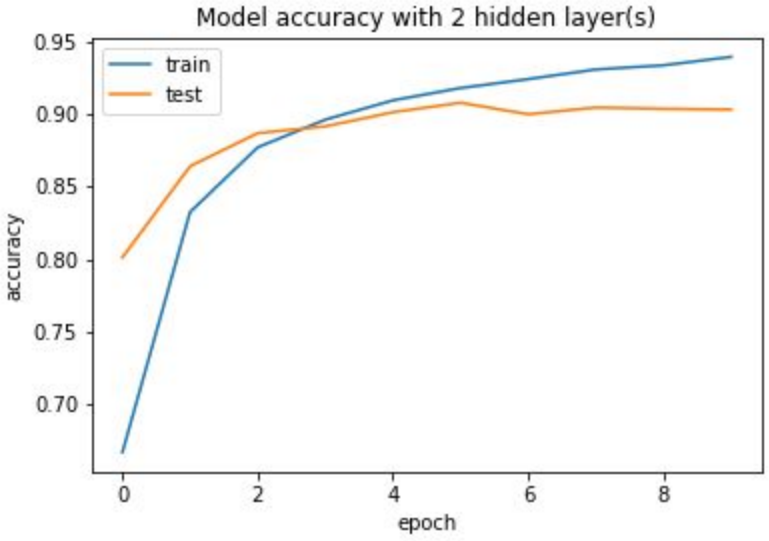
- The dataset was converted into *dataframe* object (using pandas Data-frame) with two columns:
  ['Sentiment_class_label', 'Review_Text'] where the first column represents positive(1) or negative(0) review and 'Text' represents the corresponding text sentence.
- Punctuations were removed from the sentences.
- For each sentence, the number of word tokens is counted and for those in which token count is less than 25, further steps are done.
- Two lists were created:
  - *DWORDS*: It stores all the unique words present.
  - *WCOUNT*: It stores the corresponding frequencies of those unique words.
- A list of words is created as a vocabulary, where word count is more than 5 for that particular word over all available text review.
- Based on this vocabulary, a one-hot representation of the sentences is formed.
- Finally, the dataset is divided into training and test sets in the ratio of 9:1.

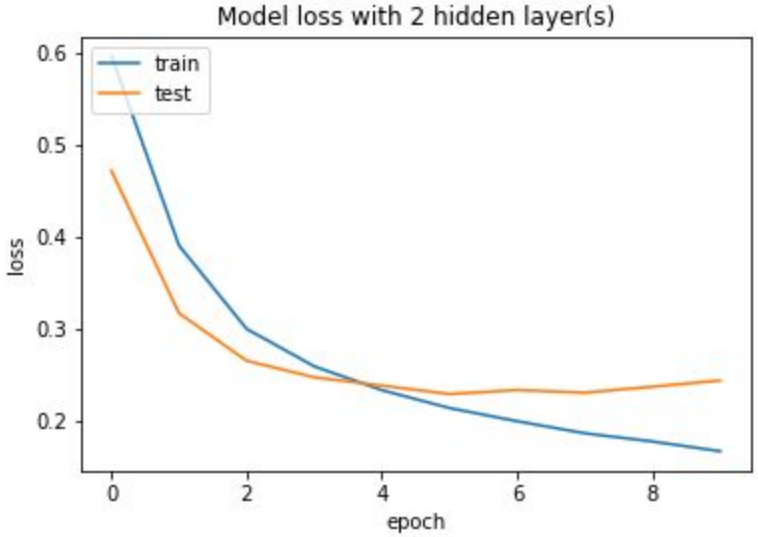
★TRAINING AND CLASSIFICATION (IN KERAS)

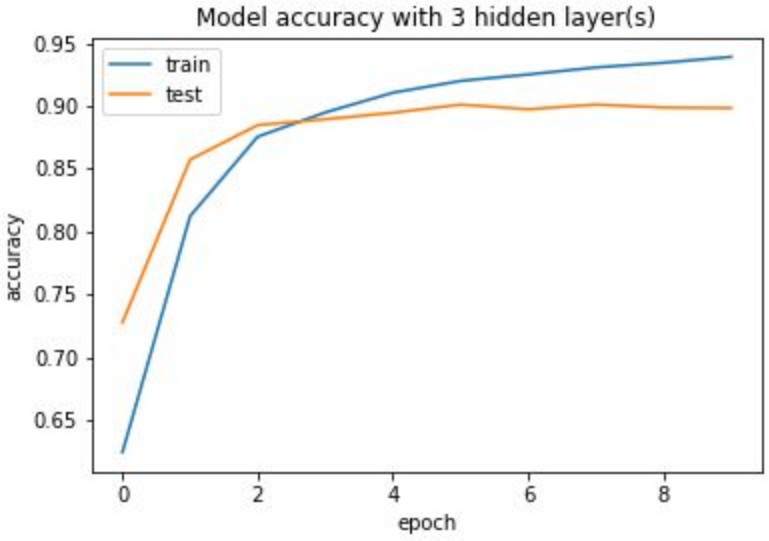
- With the training data, a fully connected neural network was built and the number of hidden layers was varied (1, 2, and 3). The results were obtained and the loss curves for training and validation losses, as well as training and validation accuracies, were plotted.
- For each architecture, the following remain constant:
    - Number of epochs: 10
    - Loss function: Sigmoid cross-entropy
    - Optimizer: Stochastic Gradient Descent (SGD)
    - Batch size: 30

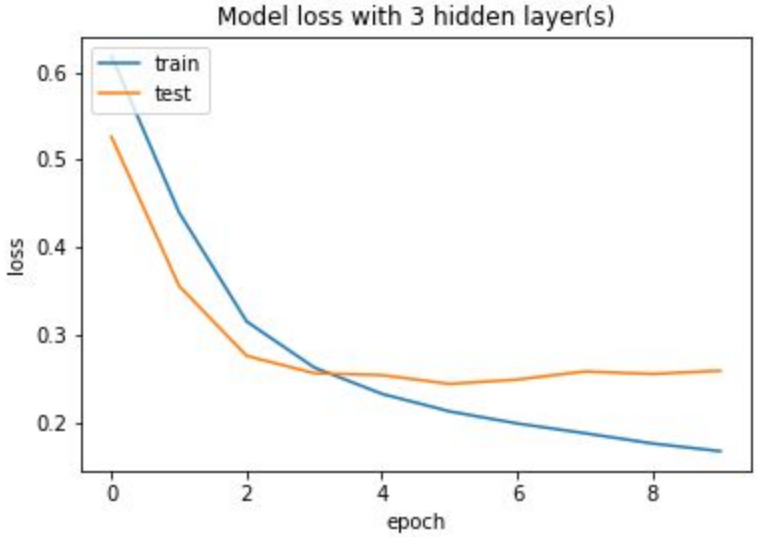

Model accuracy with 1 hidden layer(s)
train
test
accuracy
epoch
0.90
0.85
0.80
0.75
0.70
0
2
4
6
8


Model loss with 1 hidden layer(s)
train
test
loss
epoch
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0
2
4
6
8


Model accuracy with 2 hidden layer(s)
train
test
accuracy
epoch
0.95
0.90
0.85
0.80
0.75
0.70
0
2
4
6
8


Model loss with 2 hidden layer(s)
train
test
loss
epoch
0.6
0.5
0.4
0.3
0.2
0
2
4
6
8


Model accuracy with 3 hidden layer(s)
train
test
accuracy
epoch
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0
2
4
6
8


Model loss with 3 hidden layer(s)
train
test
loss
epoch
0.6
0.5
0.4
0.3
0.2
0
2
4
6
8