7thSep 2018

**Attendance: 10%, Continuous evaluation: 70%, Viva-20%**

**Assignment No. 4**

The problem is to predict the sentiment of a given product review (sentence) whether it is positive review or negative review.  
  
Example:      
Sent 1:: The product is great I liked it. (Positive review)  Class Label 1  
Sent 2:: The item was not upto mark and price is high. (Negative review) Class Label 0.

* Download the sentiment analysis dataset from https://www.kaggle.com/bittlingmayer/amazonreviews/downloads/test.ft.txt.bz2/2on amazon reviews.
* Pre-process on the dataset by doing the following:
  1. Convert it into data-frame object (using pandas Data-frame) with two columns [‘Sentiment\_class\_label’, ‘Review\_Text’] where first column represent positive(1) or negative(0) review and ‘Text’ represent the corresponding text sentence.
  2. Remove the punctuation token from column ‘Review\_Text’.
  3. Count number of words in for each sentence (‘Review\_Text’) and add new column ‘Word\_Length’.
  4. Select only the rows which have ‘Word\_Length’ less than 25.
  5. Create a list of vocabulary(words) where word count is more than 5 for that particular word over all available text review.
* Create a “one hot vector” representation for each sentence based on created vocubualary set. Devide the dataest into training and test set with ratio of 90:10, respectively.
* Train a Fully Connected Layer on tensorflow or Keras package having hidden layers size of 1,2,3 using sigmoid cross entropy loss function.
* Plot the graph ofloss and compare the results on test set for each of the iteration.

Submit a report with result.