NIT Rourkela

Lab-4, CS6379

NLP Lab, Spring'25

Topics: RNN model for Text classification

Problem Statement:

1. Build an elementary RNN Model for Text Classification Task.
   Follow the below steps to execute your program.

   a) Load the dataset using pandas 'read_json()' method as the dataset is in json file format

   b) Pre-process the dataset to combine the 'headline' and 'short_description' of the dataset.

   c) Clean the text data to move forward with tokenization and vectorization of text inputs before we feed vectorized text data to the RNN model.

   d) Tokenization and vectorization of text data to create a word index of the sentences and split the dataset into train and test datasets.

   e) Now as you have 'train' and 'test' data prepared, you can build an RNN model using the 'Embedding()' and 'SimpleRNN()' layers of Kera's library.

   f) Compile the model with the 'rmsprop' optimizer and 'accuracy' as validation metrics followed by fitting the model to the 'X_train' and 'y_train' data. you can evaluate the model using the 'model.evaluate()' method on test data. Congrats! you have just built your first model using word embedding and RNN layers.