

Text Generation

Data Processing

Extracted all the data from the link provided, then removed all the special characters and punctuations from the text. After this, split the data using spaces and created a dictionary that has all the unique words and a corpus with all the words, then converted the list of words into a list of sentences by taking 30+1 words in each sentence. 30 words as features and the 31st word as the target variable. Then tokenized the data(converted the words into numbers) and converted the one-dimensional list of numbers to NumPy array and selected 1st 30 columns/elements of each row as features/inputs and the 31st element as the target variable, also converted the Y(target variable) to categorical.

Modelling

Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) are used to generate the text and both are working relatively the same. In both the networks embedding layer is used as an input layer which takes in the tokenized data then 2 LSTM and GRU layers are used with both layers having 64 units respectively, 1 dense layer having 128 nodes and an output layer.

Hyperparameters:

Optimizer: Adam optimizer was used with a learning rate of 0.001 because it is very versatile, it works well for both deep and shallow networks, it is good for sparse tensors as well.

Activation Function: Rectified Linear Unit (ReLU) was used because it is very less computation hungry and is able to add non-linearity to the network.

A **batch size** of 32 was used so that sufficient features can be learned and also be able to fit them in the memory. (In April 2018, Yann Lecun even tweeted "Friends don't let friends use mini-batches larger than 32"). And was run for 500 epochs.

Sample Input: *"My Name is Roshan Pandey, Generate me some text from Poirot Investigates by Agatha Christie of length 50 words."*

Output: *"found at lord willard nor rapidly him i turned the leaves ah on her apron and appearance we shall be obliged to introduce my life monsieur poirot he could i smiled forward in a low voice a man who made no italian before he had guests the mode upon me"*
