Join the AI revolution !

Deep Learning Course – Capstone Project

Spam Checker with LSTM

Phone : +91 9742800566

Offices : Bengaluru

Email : vivek@cellstrat.com

Web : www.cellstrat.com

Problem Statement

- You are given Messaging spam data with three possible labels Ham, Spam and Info
- You need to build an LSTM model that classifies the messages correctly

Pre-requisites

- You need to have the followings softwares installed
  - Python 3.5, 3.6 or 3.7 (as TensorFlow works with these Python versions only)
  - TensorFlow
  - Keras on TensorFlow
  - Jupyter notebook

Program & data

- Extract the ipynb file and the data in the same folder

Note on data size and runs

- There are 29000 training samples and 1000 test samples.
- The student is expected to run 1 Epoch (however in production one might run many more Epochs)

CELLSTRAT

Assignment overview

- You are GIVEN the following parts of the program already :-
  - Import modules
  - Read the data into Pandas dataframes
  - Pre-process messages to make them a certain length

- You are expected to build a Keras LSTM model that can learn to classify the messages

- The next slide has the details of the model that you are to produce.

Task to be completed:

The steps to be completed

- Building an LSTM model with Keras Sequential.
- Add an Embedding Layer with input_dim as length of tokenizer word index, input length as 50 and output dimensions as 100.
- Add just one layer of LSTM to Keras model with 10 units and one Dense Layer with 3 units with softmax activation.
- Do a Keras Compile
- Do a Keras fit with just 1 Epoch and validation_split=0.25
- Predict on Test Dataset and print the accuracy on the test dataset

Training/ Evaluation

- For training step, run 1 Epoch
- For prediction step, calculate the accuracy on test dataset

Reading material

For ideas look at RNN module tutorial

Thank you

Vivek Singhal
Co-Founder and AI Data Scientist, CellStrat
9742800566
vivek@cellstrat.com