

Deep Learning Exam Questions

[Time: 5 hrs]

[Total Marks: 100]

Attempt any one case study.

Part I: LSTM

[Total Marks -50]

A company named 'XYZ' sells a variety of health products for people of all ages. Company management has decided to make key decisions on their logistics, supply and production for which they want to check the sales figures for a certain set of years which will help them to invest their bucks and time in taking decisions on key factors.

Goal: As a Deep Learning Engineer in the company, you are given historical data with respect to multiple regions of India and years (2005 to 2016). Your goal is to forecast sales for the years 2017 and 2018.

Constraints: You should use only LSTM to forecast sales (Should not use any other ML or DL models to generate results).

Data Description: Historical sales data of XYZ company from 2005 to 2016. Attribute Information:

1. **HQ** - This attribute represents HeadQuarters of the company.
2. **Country** - This attribute represents the country of the outlet.
3. **State_of_outlet** - This attribute represents the state of the outlet.
4. **City_of_outlet** - This attribute represents the city of the outlet.
5. **Month** - This attribute represents month(ranging from 1 to 12)
6. **Day** - This attribute represents day of the month(ranging from 1 to 31)
7. **Year** - This attribute represents year(ranging from 2005 to 2016)
8. **Total_Sales** - Target variable (total sales on particular day). This is basically the sale index value. For eg: if value is 60 then you can assume total sale happened is 60*10,000 USD.

Provided Files:

Train Data: This contains a csv file consisting of data from years 2005 to 2016. Use this data for training the model and forecast sales for the years 2017 and 2018.

Evaluation Criteria:

The evaluation metric for this problem statement is the Mean Absolute Error.

OR

Part II: RNN

[Total Marks - 50]

The Social Dilemma, a documentary-drama hybrid explores the dangerous human impact of social networking, with tech experts sounding the alarm on their own creations as the tech experts sound the alarm on the dangerous human impact of social networking. This dataset brings you the twitter responses made with the #TheSocialDilemma hashtag after watching the eye-opening documentary "The Social Dilemma" released in an OTT platform(Netflix) on September 9th, 2020. You can categorize the tweets into different groups to identify the sentiment of the users regarding the documentary.

Goal: You are hired as a deep learning engineer and you are asked to categorize the tweets into three categories.

Constraints: You should be using only RNN to generate results and should not be using any LSTM or ML classification models to generate results.

Data Description: The dataset was extracted using TwitterAPI, consisting of nearly 20000 tweets from twitter users all over the globe!

Attribute Information:

1. **user_name** - The name of the user, as they've defined it.
2. **user_location** - The user-defined location for this account's profile.
3. **user_description** - The user-defined UTF-8 string describing their account.
4. **user_created** - Time and date, when the account was created.
5. **user_followers** - The number of followers an account currently has.
6. **user_friends** - The number of friends an account currently has.
7. **user_favourites** - The number of favorites an account currently has.
8. **user_verified** - When true, indicates that the user has a verified account.
9. **date** - UTC time and date when the Tweet was created.
10. **hashtags** - All the other hashtags posted in the tweet along with #TheSocialDilemma
11. **source** - Utility used to post the Tweet, Tweets from the Twitter website have a source value - web
12. **is_retweet** - Indicates whether this Tweet has been Retweeted by the authenticating user.
13. **clean_text** - Cleaned text of the tweet.
14. **Sentiment** (target) - Indicates the sentiment of the tweet, consists of three categories: **Positive**, **neutral**, and **negative**.

Provided Files:

Train_data - Should be used to feed your model.

Test_data - Should be used only to generate predictions.

Evaluation Criteria:

The evaluation metric for this problem statement is the validation Accuracy Score.