

## Assignment-4

Assignment Date	17 October 2022
Student Name	M.Hema Preethi
Student Roll Number	962719106009
Maximum Marks	2Marks

Question 1:

Download the dataset

Link:

[https://drive.google.com/file/d/1Siqx5H5R86tRp2YZKzzd4\\_iEfjChZ3ob/view?usp=sharing](https://drive.google.com/file/d/1Siqx5H5R86tRp2YZKzzd4_iEfjChZ3ob/view?usp=sharing)

Question 2:

Import required library

Solution:

```
import pandas as pd
import numpy as np
from keras import utils
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from keras.models import Model
from keras.layers import LSTM, Activation, Dense, Dropout, Input, Embedding
from keras.optimizers import RMSprop
from keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
from keras.utils import to_categorical
```

### Question 3:

Read dataset and do pre-processing

Solution:

Read dataset

```
!unzip "/content/archive.zip"
```

```
df = pd.read_csv('spam.csv', delimiter=',', encoding='latin-1')  
df
```

Pre processing

```
df.drop(['Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], axis=1, inplace=True)  
df
```

```
sns.countplot(df.v1, palette='Set3')  
plt.xlabel('Label')  
plt.title('Number of ham and spam messages')
```

```
X = df.v2  
Y = df.v1  
le = LabelEncoder()  
Y = le.fit_transform(Y)  
Y = Y.reshape(-1,1)
```

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.15)
```

```
max_words = 1000  
max_len = 150  
tok = Tokenizer(num_words=max_words)  
tok.fit_on_texts(X_train)  
sequences = tok.texts_to_sequences(X_train)  
sequences_matrix = utils.pad_sequences(sequences, maxlen=max_len)
```

```
sequences_matrix.shape
```

```
sequences_matrix.ndim
```

```
sequences_matrix = np.reshape(sequences_matrix,(4736,150,1))
```

```
sequences_matrix.ndim
```





Question 4:

Create model

Solution:

```
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.layers import Embedding
```

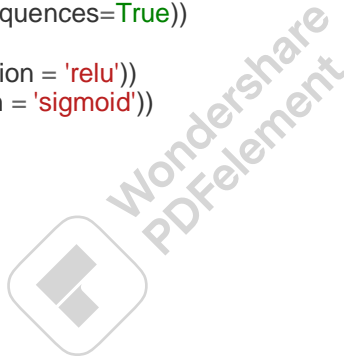
```
model = Sequential()
```

### Question 5:

Add layers(LSTM,Dense-(Hidden layers),output)

### Solution:

```
model.add(Embedding(max_words,50,input_length=max_len))
model.add(LSTM(units=64,input_shape = (sequences_matrix.shape[1],1),return_sequences=True))
model.add(LSTM(units=64,return_sequences=True))
model.add(LSTM(units=64,return_sequences=True))
model.add(LSTM(units=64))
model.add(Dense(units = 256,activation = 'relu'))
model.add(Dense(units = 1,activation = 'sigmoid'))
```



### Question 6:

Compile the model

### Solution:

```
model.summary()
model.compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy'])
```

Question 7:

Fit the model

Solution:

```
X = model.fit(sequences_matrix,Y_train,batch_size=128,epochs=5,validation_split=0.2)
```

Question 8:

Save the model

Solution:

```
model.save
```

## Question 9:

### Test the model

### Solution:

```
test_sequences = tok.texts_to_sequences(X_test)
test_sequences_matrix = utils.pad_sequences(test_sequences,maxlen=max_len)

accr = model.evaluate(test_sequences_matrix,Y_test)

l = accr[0]
a =accr[1]
print('Test set\n Loss: {:.3f}\n Accuracy: {:.3f}'.format(l,a))
```

