Assignment - 4

Details:

- Assignment Date : 20 October 2022
- Student Name : Chenchu Jaswanth Jangam
- Student Roll Number : 111519106053
- Maximum Marks : 2 Marks

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout, Embedding
from tensorflow.keras.optimizers import RMSprop
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing import sequence
```

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

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already	NaN	NaN	NaN

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   v1      5572 non-null   object
 1   v2      5572 non-null   object
dtypes: object(2)
memory usage: 87.2+ KB
```

```
X = df.v2
Y = df.v1
encoder = LabelEncoder()
Y = encoder.fit_transform(Y)
```

```
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)


tokenizer = Tokenizer(num_words=2000, lower=True)
tokenizer.fit_on_texts(X_train)
sequences = tokenizer.texts_to_sequences(X_train)
X_train = sequence.pad_sequences(sequences, maxlen=200)


model = Sequential()


model.add(Embedding(2000, 50, input_length=200))
model.add(LSTM(64))
model.add(Dense(256, activation="relu"))
model.add(Dropout(0.5))
model.add(Dense(1,activation="sigmoid"))


model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 200, 50)           100000

 lstm (LSTM)                 (None, 64)                29440

 dense (Dense)               (None, 256)               16640

 dropout (Dropout)           (None, 256)               0

 dense_1 (Dense)             (None, 1)                 257

=================================================================
Total params: 146,337
Trainable params: 146,337
Non-trainable params: 0
_________________________________________________________________
```

```
model.compile(loss='binary_crossentropy', optimizer=RMSprop(), metrics=['accurac


model.fit(X_train, y_train, batch_size=128, epochs=10, validation_split=0.2)
```

```
Epoch 1/10
28/28 [==============================] - 4s 49ms/step - loss: 0.3426 - accu
Epoch 2/10
28/28 [==============================] - 1s 27ms/step - loss: 0.1026 - accu
Epoch 3/10
28/28 [==============================] - 1s 27ms/step - loss: 0.0417 - accu
Epoch 4/10
28/28 [==============================] - 1s 27ms/step - loss: 0.0253 - accu
Epoch 5/10
```

```
Epoch 5/10
28/28 [==============================] - 1s 26ms/step - loss: 0.0191 - accu
Epoch 6/10
28/28 [==============================] - 1s 28ms/step - loss: 0.0131 - accu
Epoch 7/10
28/28 [==============================] - 1s 26ms/step - loss: 0.0085 - accu
Epoch 8/10
28/28 [==============================] - 1s 26ms/step - loss: 0.0075 - accu
Epoch 9/10
28/28 [==============================] - 1s 26ms/step - loss: 0.0053 - accu
Epoch 10/10
28/28 [==============================] - 1s 26ms/step - loss: 0.0031 - accu
<keras.callbacks.History at 0x1ecbb7a4ee0>
```

```
model.save("model.h5")
```

```
test_sequences = tokenizer.texts_to_sequences(X_test)
X_test = sequence.pad_sequences(test_sequences, maxlen=200)
```

```
acc = model.evaluate(X_test, y_test)
```

```
35/35 [==============================] - 0s 12ms/step - loss: 0.0825 - accu
```

```
def predict(message):
    txt = tokenizer.texts_to_sequences(message)
    txt = sequence.pad_sequences(txt, maxlen=200)
    preds = model.predict(txt)
    if preds > 0.5:
        print("Spam")
    else:
        print("Not Spam")
```

```
predict(["Sorry, I'll call after the meeting."])
```

```
1/1 [==============================] - 0s 28ms/step
Not Spam
```

```
predict(["Congratulations!!! You won $50,000. Send message LUCKY100 to XXXXXXXXX
```

```
1/1 [==============================] - 0s 25ms/step
Spam
```

Colab paid products - Cancel contracts here