assignment 11

May 28, 2023

0.1 Assignment 11

[ ]:

```
import numpy as np
from keras.models import Sequential
from keras.layers import LSTM, Dense
from keras.callbacks import ModelCheckpoint
from keras.optimizers import RMSprop
```

[ ]:

```
text = open("01 - The Fellowship Of The Ring.txt", 'r', encoding='latin-1').
 ↪read().lower()
print('Corpus length:', len(text))
```

```
Corpus length: 1021113
```

[ ]:

```
maxlen = 60
step = 3
sentences = []
next_chars = []

for i in range(0, len(text) - maxlen, step):
    sentences.append(text[i: i + maxlen])
    next_chars.append(text[i + maxlen])
print('Number of sequences:', len(sentences))


chars = sorted(list(set(text)))
print('Unique characters:', len(chars))
char_indices = dict((char, chars.index(char)) for char in chars)


print('Vectorization...')
x = np.zeros((len(sentences), maxlen, len(chars)), dtype=bool)
y = np.zeros((len(sentences), len(chars)), dtype=bool)
for i, sentence in enumerate(sentences):
    for t, char in enumerate(sentence):
        x[i, t, char_indices[char]] = 1
    y[i, char_indices[next_chars[i]]] = 1
```

```
Number of sequences: 340351
```

```
Unique characters: 63
Vectorization…
```

[ ]:

```python
model = Sequential()
model.add(LSTM(128, input_shape=(maxlen, len(chars))))
model.add(Dense(len(chars), activation='softmax'))

optimizer = RMSprop(learning_rate=0.01)
model.compile(loss='categorical_crossentropy', optimizer=optimizer)
```

[ ]:

```python
checkpoint = ModelCheckpoint('assignment11_model.h5', monitor='loss',␣
 ↪save_best_only=True, mode='min')
```

[ ]:

```python
def sample(preds, temperature=1.0):
    preds = np.asarray(preds).astype('float64')
    preds = np.log(preds) / temperature
    exp_preds = np.exp(preds)
    preds = exp_preds / np.sum(exp_preds)
    probas = np.random.multinomial(1, preds, 1)
    return np.argmax(probas)
```

[ ]:

```python
import random, os

if not os.path.exists('results'):
    os.makedirs('results')

for epoch in range(1, 20):
    print('epoch', epoch)
    epoch_results = (f'results/epoch_{epoch}_results.md')
    # Fit the model for 1 epoch on the available training data
    model.fit(x, y,
              batch_size=128,
              epochs=1)

    with open(epoch_results, 'w') as f:
        start_index = random.randint(0, len(text) - maxlen - 1)
        generated_text = text[start_index: start_index + maxlen]
        f.write('--- Generating with seed: "' + generated_text + '"\n')

        for temperature in [0.2, 0.5, 1.0, 1.2]:
            f.write('------ temperature:' + str(temperature) + "\n")
            f.write(generated_text)

            for i in range(400):
                sampled = np.zeros((1, maxlen, len(chars)))
                for t, char in enumerate(generated_text):
                    sampled[0, t, char_indices[char]] = 1.
                preds = model.predict(sampled, verbose=0)[0]
                next_index = sample(preds, temperature)
                next_char = chars[next_index]

                generated_text += next_char
                generated_text = generated_text[1:]

                f.write(next_char)
            f.write('\n')
```

```
epoch 1
2659/2659 [==============================] - 100s 37ms/step - loss: 1.7032
epoch 2
2659/2659 [==============================] - 105s 39ms/step - loss: 1.4473
epoch 3
2659/2659 [==============================] - 106s 40ms/step - loss: 1.3906
epoch 4
2659/2659 [==============================] - 104s 39ms/step - loss: 1.3588
epoch 5
2659/2659 [==============================] - 99s 37ms/step - loss: 1.3381
epoch 6
2659/2659 [==============================] - 101s 38ms/step - loss: 1.3206
epoch 7
2659/2659 [==============================] - 95s 36ms/step - loss: 1.3100
epoch 8
2659/2659 [==============================] - 100s 38ms/step - loss: 1.3013
epoch 9
2659/2659 [==============================] - 101s 38ms/step - loss: 1.2940
epoch 10
2659/2659 [==============================] - 96s 36ms/step - loss: 1.2872
epoch 11
2659/2659 [==============================] - 95s 36ms/step - loss: 1.2798
epoch 12
2659/2659 [==============================] - 96s 36ms/step - loss: 1.2732
epoch 13
2659/2659 [==============================] - 97s 37ms/step - loss: 1.2691
epoch 14
2659/2659 [==============================] - 93s 35ms/step - loss: 1.2650
epoch 15
2659/2659 [==============================] - 92s 35ms/step - loss: 1.2611
epoch 16
2659/2659 [==============================] - 97s 37ms/step - loss: 1.2568
epoch 17
2659/2659 [==============================] - 93s 35ms/step - loss: 1.2543
epoch 18
2659/2659 [==============================] - 94s 35ms/step - loss: 1.2506
epoch 19
2659/2659 [==============================] - 123s 46ms/step - loss: 1.2483
```

```
C:\Users\Gabe\AppData\Local\Temp\ipykernel_23540\2163503365.py:3:
RuntimeWarning: divide by zero encountered in log
  preds = np.log(preds) / temperature
```