

Assignment 11

author: Rachel Nelson

class: DSC650

Assignment 11

Using section 8.1 in Deep Learning with Python as a guide, implement an LSTM text generator. Train the model on the Enron corpus or a text source of your choice. Save the model and generate 20 examples to the results directory of dsc650/assignments/assignment11/.

```
1 # Listing 8.1 Reweighting a probability distribution to a different temperature
import numpy as np
def reweight_distribution(original_distribution, temperature=0.5):
    distribution = np.log(original_distribution) / temperature
    distribution = np.exp(distribution)
    return distribution / np.sum(distribution)

2 # Listing 8.2 Downloading and parsing the initial text file
import tensorflow.keras
import numpy as np
path = tensorflow.keras.utils.get_file(
    'nietzsche.txt',
    origin='https://s3.amazonaws.com/text-datasets/nietzsche.txt')
text = open(path).read().lower()
print('Corpus length:', len(text))

Corpus length: 600901

3 # Listing 8.3 Vectorizing sequences of characters
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=np.bool)
y = np.zeros((len(sentences), len(chars)), dtype=np.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: 200281
Unique characters: 59
Vectorization...

D:\College\venv\lib\site-packages\ipykernel_launcher.py:17: DeprecationWarning: `np.bool` is a deprecated
Deprecated in NumPy 1.20; for more details and guidance: https://numpy.org/devdocs/release/1.20.0-notes.h
D:\College\venv\lib\site-packages\ipykernel_launcher.py:18: DeprecationWarning: `np.bool` is a deprecated
Deprecated in NumPy 1.20; for more details and guidance: https://numpy.org/devdocs/release/1.20.0-notes.h

4 # Listing 8.4 Single-layer LSTM model for next-character prediction
from keras import layers
model = tensorflow.keras.models.Sequential()
model.add(layers.LSTM(128, input_shape=(maxlen, len(chars))))
model.add(layers.Dense(len(chars), activation='softmax'))
model.save_weights('model.h5')

5 optimizer = tensorflow.keras.optimizers.RMSprop(lr=0.01)
model.compile(loss='categorical_crossentropy', optimizer=optimizer)

D:\College\venv\lib\site-packages\keras\optimizer_v2\rmsprop.py:130: UserWarning: The `lr` argument is de
super(RMSprop, self).__init__(name, **kwargs)

6 # Listing 8.6 Function to sample the next character given the model's predictions
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)

```

```

10 # Listing 8.7 Text-generation loop
import random
import sys

# Using 5 epochs at 4 temperatures for 20 results
for epoch in range(1, 6):
    print('epoch', epoch)
    model.fit(x, y, batch_size=128, epochs=1)

    start_index = random.randint(0, len(text) - maxlen - 1)
    generated_text = text[start_index: start_index + maxlen]
    print('--- Generating with seed: "' + generated_text + '"')

    for temperature in [0.2, 0.5, 1.0, 1.2]:
        print('----- temperature:', temperature)
        sys.stdout.write(generated_text)
        # reduced range from 400 to 60 for quicker processing
        for i in range(60):
            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:]

sys.stdout.write(next_char)

print()
with open('text_{}_temperature_{}.txt'.format(epoch, temperature), 'w') as f:
    f.write('generated_text')

epoch 1
1565/1565 [=====] - 71s 45ms/step - loss: 1.4258████████████████████████████████████████████████████████████████
--- Generating with seed: "found its readers in a wide circle of lands and
peoples--it "
----- temperature: 0.2
found its readers in a wide circle of lands and
peoples--it is not the world of the supposing the still the persony of
----- temperature: 0.5
is not the world of the supposing the still the persony of the permitive and of the most the faith they
----- temperature: 1.0
the permitive and of the most the faith they are and as a constant just corvant of
all the partic superiors these suspect
----- temperature: 1.2
nstant just corvant of
all the partic superiors these suspecting, of rued"wer is ay thoseing to that understandness and t
epoch 2
1565/1565 [=====] - 79s 51ms/step - loss: 1.4087████████████████████████████████████████████████████████████
--- Generating with seed: "rom the poor man (for example, a prince who
deprives a plebe"
----- temperature: 0.2
rom the poor man (for example, a prince who
deprives a plebeian and subling of the soul and subling of the soul and subl
----- temperature: 0.5
ian and subling of the soul and subling of the soul and sublisted in the way to the sympathy of what to s
----- temperature: 1.0
isted in the way to the sympathy of what to see in the commands were to eace that it got side in this was
----- temperature: 1.2
nds were to eace that it got side in this was and considered to the other, from embeddols, commands

ahour love anot, an
epoch 3
1565/1565 [=====] - 72s 46ms/step - loss: 1.3954████████████████████████████████████████████████████████
--- Generating with seed: "e
right to define all active force unequivocally as will to "
----- temperature: 0.2
e
right to define all active force unequivocally as will to the stand the ways the perhaps the individual,
----- temperature: 0.5
the stand the ways the perhaps the individual, and the stand the same and his spiritual prililitur goes o
----- temperature: 1.0
the same and his spiritual prililitur goes of the command the strength of fundrenth, of the fisedvences,
----- temperature: 1.2
he strength of fundrenth, of the fisedvences, was a definitl, stand, has execthiog over-inequality, the
epoch 4
1565/1565 [=====] - 71s 46ms/step - loss: 1.3839████████████████████████████████████████████████████████
--- Generating with seed: "uestions of its origin and
beginning: must one not be almost"
----- temperature: 0.2
uestions of its origin and
beginning: must one not be almost the sense of the act and consideration of the superficial a
----- temperature: 0.5
the sense of the act and consideration of the superficial and part of the more a sense, with the connect
----- temperature: 1.0
nd part of the more a sense, with the connection of the considerable tporminus the philosophy as little b
----- temperature: 1.2
```


