7. Workshop 2-3 : IMDB - NLP - RNN

> **[Reference] :**
>
> François Chollet, **Deep Learning with Python**, Chapter 6, Section 2, Manning, 2018. [http://www.deeplearningitalia.com/wp-content/uploads/2017/12/Dropbox_Chollet.pdf#](http://www.deeplearningitalia.com/wp-content/uploads/2017/12/Dropbox_Chollet.pdf) (http://www.deeplearningitalia.com/wp-content/uploads/2017/12/Dropbox_Chollet.pdf)

A first recurrent layer in Keras

The process we just naively implemented in Numpy corresponds to an actual Keras layer: the `SimpleRNN` layer:

In [1]:

```python
import keras
keras.__version__
```

```
/Users/macmini1/anaconda3/lib/python3.6/site-packages/h5p
y/__init__.py:36: FutureWarning: Conversion of the second
argument of issubdtype from `float` to `np.floating` is de
precated. In future, it will be treated as `np.float64 ==
np.dtype(float).type`.
  from ._conv import register_converters as _register_conv
erters
Using TensorFlow backend.
```

Out[1]:

```
'2.2.4'
```

In [2]:

```python
from keras.layers import SimpleRNN
```

There is just one minor difference: `SimpleRNN` processes batches of sequences, like all other Keras layers, not just a single sequence like in our Numpy example. This means that it takes inputs of shape `(batch_size, timesteps, input_features)`, rather than `(timesteps, input_features)`.

Like all recurrent layers in Keras, `SimpleRNN` can be run in two different modes: it can return either the full sequences of successive outputs for each timestep (a 3D tensor of shape `(batch_size, timesteps, output_features)`), or it can return only the last output for each input sequence (a 2D tensor of shape `(batch_size, output_features)`). These two modes are controlled by the `return_sequences` constructor argument. Let's take a look at an example:

In [3]:

```
from keras.models import Sequential
from keras.layers import Embedding, SimpleRNN

model = Sequential()
model.add(Embedding(10000, 32))
model.add(SimpleRNN(32))
model.summary()
```

```
__________________________________________________________________
_______
Layer (type)                 Output Shape              Par
am #
==================================================================
=======
embedding_1 (Embedding)      (None, None, 32)          320
000
__________________________________________________________________
_______
simple_rnn_1 (SimpleRNN)     (None, 32)                208
0
==================================================================
=======
Total params: 322,080
Trainable params: 322,080
Non-trainable params: 0
__________________________________________________________________
_______
```

In [4]:

```
model = Sequential()
model.add(Embedding(10000, 32))
model.add(SimpleRNN(32, return_sequences=True))
model.summary()
```

```
__________________________________________________________________
_______
Layer (type)                 Output Shape              Par
am #
==================================================================
=======
embedding_2 (Embedding)      (None, None, 32)          320
000
__________________________________________________________________
_______
simple_rnn_2 (SimpleRNN)     (None, None, 32)          208
0
==================================================================
=======
Total params: 322,080
Trainable params: 322,080
Non-trainable params: 0
__________________________________________________________________
_______
```

It is sometimes useful to stack several recurrent layers one after the other in order to increase the representational power of a network. In such a setup, you have to get all intermediate layers to return full sequences:

In [5]:

```
model = Sequential()
model.add(Embedding(10000, 32))
model.add(SimpleRNN(32, return_sequences=True))
model.add(SimpleRNN(32, return_sequences=True))
model.add(SimpleRNN(32, return_sequences=True))
model.add(SimpleRNN(32))  # This last layer only returns the last o
model.summary()
```

```
_________________________________________________________________
________
Layer (type)                 Output Shape              Par
am #
=================================================================
=======
embedding_3 (Embedding)      (None, None, 32)          320
000

_________________________________________________________________
________
simple_rnn_3 (SimpleRNN)     (None, None, 32)          208
0

_________________________________________________________________
________
simple_rnn_4 (SimpleRNN)     (None, None, 32)          208
0

_________________________________________________________________
________
simple_rnn_5 (SimpleRNN)     (None, None, 32)          208
0

_________________________________________________________________
________
simple_rnn_6 (SimpleRNN)     (None, 32)                208
0
=================================================================
=======
Total params: 328,320
Trainable params: 328,320
Non-trainable params: 0

_________________________________________________________________
________
```

Now let's try to use such a model on the IMDB movie review classification problem. First, let's preprocess the data:

In [6]:

```
from keras.datasets import imdb
from keras.preprocessing import sequence

max_features = 10000  # number of words to consider as features
maxlen = 500  # cut texts after this number of words (among top max
batch_size = 32

print('Loading data...')
(input_train, y_train), (input_test, y_test) = imdb.load_data(num_w
print(len(input_train), 'train sequences')
print(len(input_test), 'test sequences')

print('Pad sequences (samples x time)')
input_train = sequence.pad_sequences(input_train, maxlen=maxlen)
input_test = sequence.pad_sequences(input_test, maxlen=maxlen)
print('input_train shape:', input_train.shape)
print('input_test shape:', input_test.shape)
```

```
Loading data...
25000 train sequences
25000 test sequences
Pad sequences (samples x time)
input_train shape: (25000, 500)
input_test shape: (25000, 500)
```

Let's train a simple recurrent network using an `Embedding` layer and a `SimpleRNN` layer:

In [7]:

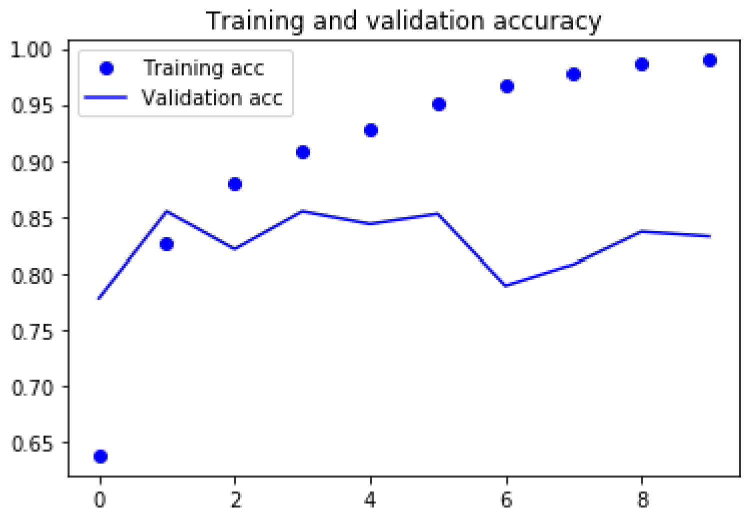
```
from keras.layers import Dense

model = Sequential()
model.add(Embedding(max_features, 32))
model.add(SimpleRNN(32))
model.add(Dense(1, activation='sigmoid'))

model.compile(optimizer='rmsprop', loss='binary_crossentropy', metr
history = model.fit(input_train, y_train,
                    epochs=10,
                    batch_size=128,
                    validation_split=0.2)
```

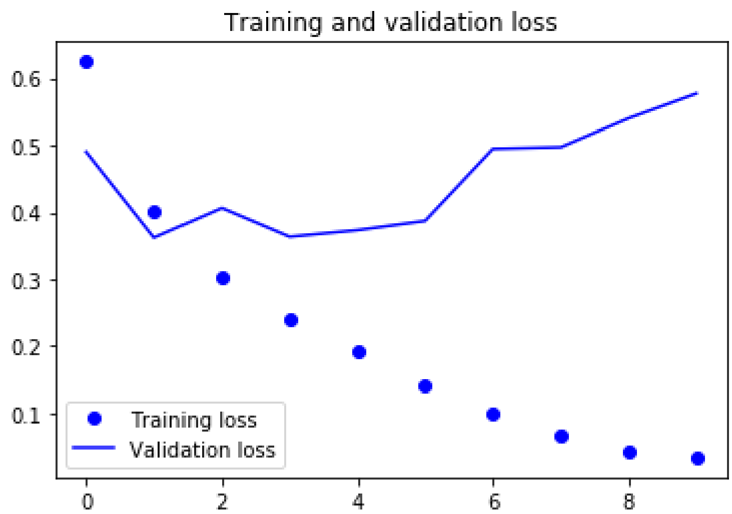
```
Train on 20000 samples, validate on 5000 samples
Epoch 1/10
20000/20000 [==============================] - 29s 1ms/ste
p - loss: 0.6249 - acc: 0.6379 - val_loss: 0.4902 - val_ac
c: 0.7780
Epoch 2/10
20000/20000 [==============================] - 27s 1ms/ste
p - loss: 0.4022 - acc: 0.8273 - val_loss: 0.3624 - val_ac
c: 0.8548
Epoch 3/10
20000/20000 [==============================] - 28s 1ms/ste
p - loss: 0.3016 - acc: 0.8807 - val_loss: 0.4064 - val_ac
c: 0.8212
Epoch 4/10
20000/20000 [==============================] - 28s 1ms/ste
p - loss: 0.2406 - acc: 0.9076 - val_loss: 0.3637 - val_ac
c: 0.8548
Epoch 5/10
20000/20000 [==============================] - 28s 1ms/ste
p - loss: 0.1939 - acc: 0.9274 - val_loss: 0.3735 - val_ac
c: 0.8436
Epoch 6/10
20000/20000 [==============================] - 28s 1ms/ste
p - loss: 0.1406 - acc: 0.9508 - val_loss: 0.3872 - val_ac
c: 0.8526
Epoch 7/10
20000/20000 [==============================] - 28s 1ms/ste
p - loss: 0.0988 - acc: 0.9668 - val_loss: 0.4946 - val_ac
c: 0.7890
Epoch 8/10
20000/20000 [==============================] - 28s 1ms/ste
p - loss: 0.0663 - acc: 0.9774 - val_loss: 0.4971 - val_ac
c: 0.8080
Epoch 9/10
20000/20000 [==============================] - 28s 1ms/ste
p - loss: 0.0433 - acc: 0.9871 - val_loss: 0.5404 - val_ac
c: 0.8368
Epoch 10/10
20000/20000 [==============================] - 28s 1ms/ste
p - loss: 0.0344 - acc: 0.9899 - val_loss: 0.5771 - val_ac
c: 0.8326
```

Let's display the training and validation loss and accuracy:

In [9]:

```
import matplotlib.pyplot as plt
%matplotlib inline

acc = history.history['acc']
val_acc = history.history['val_acc']
loss = history.history['loss']
val_loss = history.history['val_loss']

epochs = range(len(acc))

plt.plot(epochs, acc, 'bo', label='Training acc')
plt.plot(epochs, val_acc, 'b', label='Validation acc')
plt.title('Training and validation accuracy')
plt.legend()

plt.figure()

plt.plot(epochs, loss, 'bo', label='Training loss')
plt.plot(epochs, val_loss, 'b', label='Validation loss')
plt.title('Training and validation loss')
plt.legend()

plt.show()
```





As a reminder, in chapter 3, our very first naive approach to this very dataset got us to 88% test accuracy. Unfortunately, our small recurrent network doesn't perform very well at all compared to this baseline (only up to 85% validation accuracy). Part of the problem is that our inputs only consider the first 500 words rather the full sequences -- hence our RNN has access to less information than our earlier baseline model. The remainder of the problem is simply that `SimpleRNN` isn't very good at processing long sequences, like text. Other types of recurrent layers perform much better. Let's take a look at some more advanced layers.

A concrete LSTM example in Keras

Now let's switch to more practical concerns: we will set up a model using a LSTM layer and train it on the IMDB data. Here's the network, similar to the one with `SimpleRNN` that we just presented. We only specify the output dimensionality of the LSTM layer, and leave every other argument (there are lots) to the Keras defaults. Keras has good defaults, and things will almost always "just work" without you having to spend time tuning parameters by hand.

In [10]:

```
from keras.layers import LSTM

model = Sequential()
model.add(Embedding(max_features, 32))
model.add(LSTM(32))
model.add(Dense(1, activation='sigmoid'))

model.compile(optimizer='rmsprop',
              loss='binary_crossentropy',
              metrics=['acc'])
history = model.fit(input_train, y_train,
                    epochs=10,
                    batch_size=128,
                    validation_split=0.2)
```

```
Train on 20000 samples, validate on 5000 samples
Epoch 1/10
20000/20000 [==============================] - 95s 5ms/ste
p - loss: 0.5138 - acc: 0.7628 - val_loss: 0.3759 - val_ac
c: 0.8636
Epoch 2/10
20000/20000 [==============================] - 93s 5ms/ste
p - loss: 0.2944 - acc: 0.8833 - val_loss: 0.4668 - val_ac
c: 0.7766
Epoch 3/10
20000/20000 [==============================] - 93s 5ms/ste
p - loss: 0.2365 - acc: 0.9088 - val_loss: 0.3169 - val_ac
c: 0.8806
Epoch 4/10
20000/20000 [==============================] - 92s 5ms/ste
p - loss: 0.1974 - acc: 0.9269 - val_loss: 0.3397 - val_ac
c: 0.8676
Epoch 5/10
20000/20000 [==============================] - 93s 5ms/ste
p - loss: 0.1726 - acc: 0.9378 - val_loss: 0.3528 - val_ac
c: 0.8826
Epoch 6/10
20000/20000 [==============================] - 93s 5ms/ste
p - loss: 0.1547 - acc: 0.9427 - val_loss: 0.3533 - val_ac
c: 0.8836
Epoch 7/10
20000/20000 [==============================] - 92s 5ms/ste
p - loss: 0.1425 - acc: 0.9485 - val_loss: 0.3672 - val_ac
c: 0.8866
Epoch 8/10
20000/20000 [==============================] - 93s 5ms/ste
p - loss: 0.1294 - acc: 0.9559 - val_loss: 0.3297 - val_ac
c: 0.8834
Epoch 9/10
20000/20000 [==============================] - 92s 5ms/ste
p - loss: 0.1146 - acc: 0.9606 - val_loss: 0.3535 - val_ac
c: 0.8830
Epoch 10/10
20000/20000 [==============================] - 93s 5ms/ste
p - loss: 0.1077 - acc: 0.9612 - val_loss: 0.3439 - val_ac
c: 0.8826
```

In [11]:

```
acc = history.history['acc']
val_acc = history.history['val_acc']
loss = history.history['loss']
val_loss = history.history['val_loss']

epochs = range(len(acc))

plt.plot(epochs, acc, 'bo', label='Training acc')
plt.plot(epochs, val_acc, 'b', label='Validation acc')
plt.title('Training and validation accuracy')
plt.legend()

plt.figure()

plt.plot(epochs, loss, 'bo', label='Training loss')
plt.plot(epochs, val_loss, 'b', label='Validation loss')
plt.title('Training and validation loss')
plt.legend()

plt.show()
```