

**Nitin Panwar**

Technical Lead (Data Science), Naukri.com

[Blog](#) [About](#)

How to prepare Time Series Data for LSTM Networks

LSTM stands for Long short term memory, LSTMs came into picture to overcome the disadvantage of RNN. RNN has a disadvantage that it can't store long sequences.

It is not always straight forward to feed the data to LSTM model. LSTM expects three dimensional input in the Keras python deep learning library, And it doesn't like sequences of more than 200-400 time steps, so the data will need to be split into samples.

Python code to create data...

Python Code

```
# load data
data = list()
n = 5000
for i in range(n):
    data.append([i+1, (i+1)*10])
data = array(data)
print(data[:5, :])
print(data.shape)
```

Output

```
[[ 1 10] [ 2 20] [ 3 30] [ 4 40] [ 5 50]]  
(5000, 2)
```

If your time series data is uniform over time and there is no missing values, we can drop the time column.

Python code

```
# drop time  
data = data[:, 1]  
print(data.shape)
```

Output

```
(5000,)
```

In this case, 5,000 times steps are too long, therefore, we need to split it into multiple shorter sub-sequences. There are multiple ways to do that. e.g

1. Using overlapping sequences.
2. Using non-overlapping sequences.

Python code

```
# split into samples (e.g. 5000/200 = 25)
samples = list()
length = 200
# step over the 5,000 in jumps of 200
for i in range(0,n,length):
    sample = data[i:i+length]
    samples.append(sample)
print(len(samples))
```

Output

```
25
```

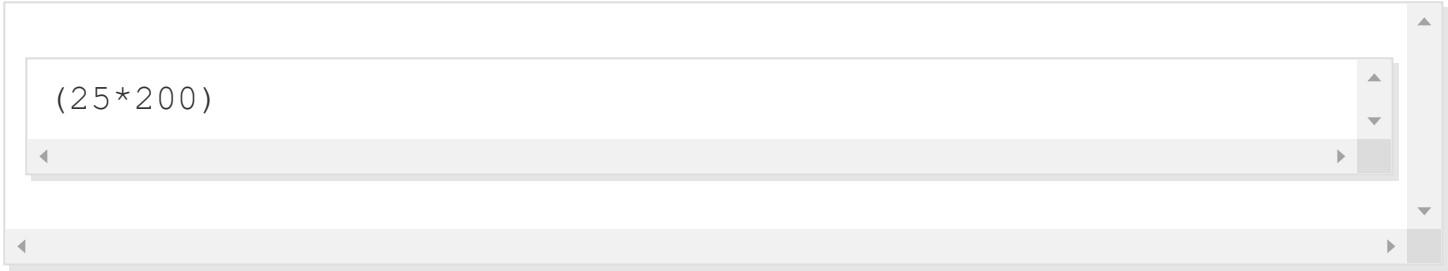
Reshape Subsequences

The LSTM needs data with the format of [samples, time steps, features] Here we have 25 samples, 200 time steps per sample and 1 feature. First we need to convert our list of arrays into a 2D numpy array of 25*20

Python code

```
data =array(samples)
print(data.shape)
```

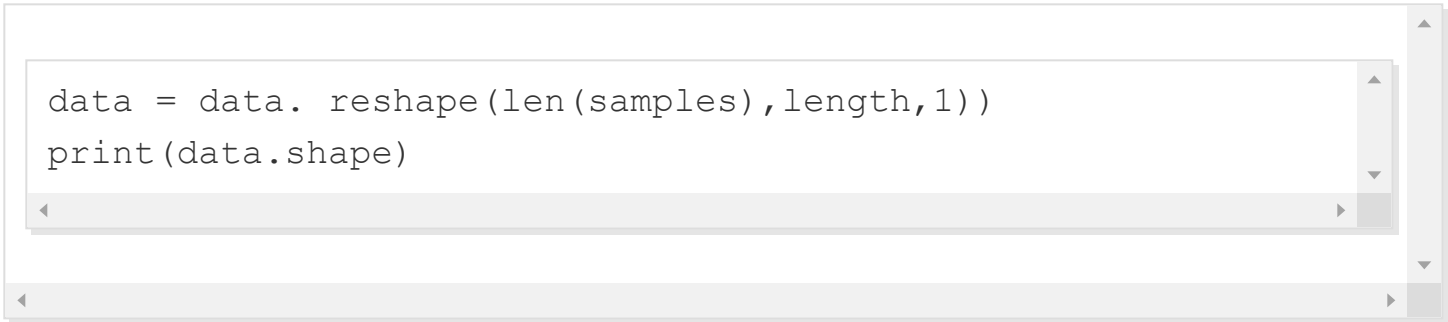
output



```
(25*200)
```

Next, we can use the `reshape()` function to add one additional dimension for our single feature.

Python code



```
data = data.reshape(len(samples), length, 1)
print(data.shape)
```

And this is it. The data can now be used as an input(X) to an LSTM model.

output



```
(25, 200, 1)
```

Written on November 24, 2017

0 Comments

Nitin Kumar Personal Pages

1

Login


Recommend

1

Tweet

Share

Sort by Newest



Start the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS ?

Name

Be the first to comment.

Subscribe

Add Disqus to your siteAdd DisqusAdd

Disqus' Privacy PolicyPrivacy PolicyPrivacy