Sudha Amarnath

SAN JOSÉ STATE
UNIVERSITY

Graduate and Extended Studies

FA19: CMPE-297 Sec 01 - Special Topics

Prof. Chandrasekar Vuppalapati

Learning Objective: size the tensor object.

1. A dataset of text documents where each document by the counts of how many times each word appears is 20,000 common words. A set of 500 documents need to be represented into a Data Frame. Can you write the tensor shape?

- In this dataset, each single data point can be encoded as a vector, and thus a batch of data will be encoded as a 2D tensor (that is, an array of vectors), where the first axis is the samples axis and the second axis is the features axis.

- The vector data is encoded as:
  Vector data—2D tensors of shape (samples, features)

- A dataset of text documents, where we represent each document by the counts of how many times each word appears in it (out of a dictionary of 20,000 common words). Each document can be encoded as a vector of 20,000 values (one count per word in the dictionary), and thus an entire dataset of 500 documents can be stored in a tensor of shape (500, 20000).

- Python execution to print the shape of the tensor for the given example. The dataframe is initialized with value 0 since only shape is needed.

```
In [1]:  import tensorflow as tf
         import numpy as np

         x = 500
         y = 20000

         df=[ [0] * y ] * x

         nparr = np.array(df)
         sess = tf.Session()
         tensor = tf.convert_to_tensor(nparr, dtype=tf.int8)
         sess.run(tensor)
         tensor_shape = tensor.get_shape()
         tensor_shape
         print('Array Dimension = %sD' % nparr.ndim)
         print('Tensor Shape =', tensor_shape)

         Array Dimension = 2D
         Tensor Shape = (500, 20000)
```

2. You joined a public traded company with stock ticker symbol: MYTICKER. a fictitious stock. The Data Frame shall contain stock price for every minute, the **current price** of the stock, the **highest price** in **the past minute**, and the **lowest price** in the past minute. You need to hold one year's worth of the stock. Can you write Tensor & Shape of the tensor? (assume that there are 390 minutes in a trading day, and 250 days of active trading in a year).

- We can store the given dataset in a 3D tensor with an explicit time axis (since there is a time notion). Each sample can be encoded as a sequence of vectors (a 2D tensor), and thus a batch of data will be encoded as a 3D tensor.

- The time series data is encoded as:
  Timeseries data or sequence data—3D tensors of shape (samples, timesteps, features)

- A dataset of stock prices for MYTICKER where every minute, we store the current price of the stock, the highest price in the past minute, and the lowest price in the past minute. Thus every minute is encoded as a 3D vector, an entire day of trading is encoded as a 2D tensor of shape (390, 3) (there are 390 minutes in a trading day), and 250 days' worth of data can be stored in a 3D tensor of shape (250, 390, 3). Here, each sample would be one day's worth of data.

- Python execution to print the shape of the tensor for the given example. The dataframe is initialized with value 0 since only shape is needed.

In [2]:
```python
import tensorflow as tf
import numpy as np

x = 250
y = 390
z = 3

df=[ [ [0] * z ] * y ] * x

nparr = np.array(df)
sess = tf.Session()
tensor = tf.convert_to_tensor(nparr, dtype=tf.int8)
sess.run(tensor)
tensor_shape = tensor.get_shape()
tensor_shape
print('Array Dimension = %sD' % nparr.ndim)
print('Tensor Shape =', tensor_shape)
```

```
Array Dimension = 3D
Tensor Shape = (250, 390, 3)
```