Quiz 6: Introduction to Sequence Models

Introduction to Supervised Learning

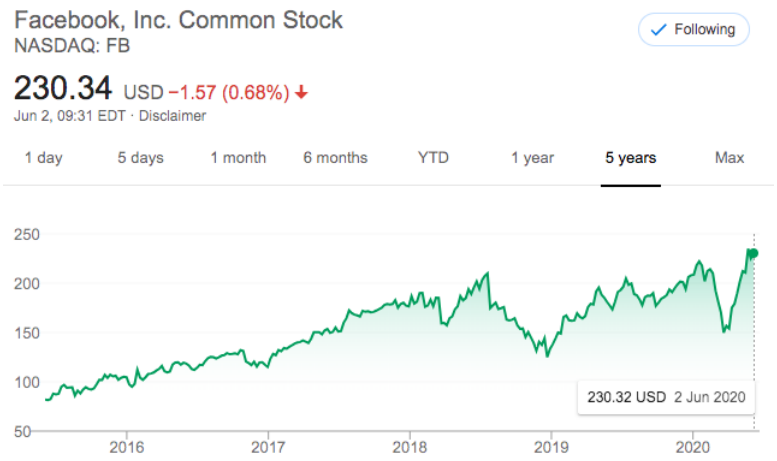
Email address *

mikesymmonds@gmail.com

Please enter your name: *

Michael Symmonds

Based on some information of the past T data points, we want to predict one of the three following categories for the next return of FB: category 0 if the return is < -1% , category 1 if the return is between -1% and +1% and category 2 if the return is > 1%



Here is the description of the training data:

- At each time step t, we have a feature vector x_t of size D representing the information we have gathered about the FB stock at time t.
- The whole sequence of feature vectors is: $x_1, \ldots, x_F$
- The corresponding sequence of targets is: $y_1, \ldots, y_F$ (where each $y_i \in \{0, 1, 2\}$)
- We have the following sequences of features and the corresponding targets:

Sequences	Targets
$x_1, \ldots, x_T$	y_{T+1}
$x_2, \ldots, x_{T+1}$	y_{T+2}
$\vdots$	$\vdots$
$x_{F-T}, \ldots, x_{F-1}$	y_F

Preprocessing

How many sequences do we have in our training data? 1 point

- ○ F
- ◉ F - T
- ○ F - T - 1

Let N be the number of sequences. What is the shape of our training tensor data? 1 point

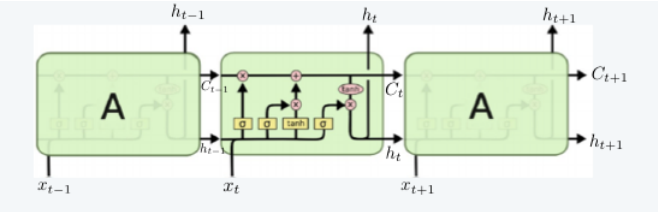
- ○ (N, D)
- ◉ (N, T, D)
- ○ (N, T)

What is the shape of our training target data after the one-hot encoding of the targets? 1 point

- ◉ (N, 3)
- ○ (N,)
- ○ (N, T, 3)

The LSTM layer

We want to use an LSTM layer to process the sequences. Let d be the output vector size at each time step t.



Why choosing an LSTM layer over a standard RNN layer? 1 point

Due to the vanishing gradient problem associated with RNN layer. Due to the long dependencies of the input structures, it is difficult for the output hT to depend on the earliest inputs. An LSTM layer, and the use of the cell states, allow for these long dependencies to be reflected in the output hT.

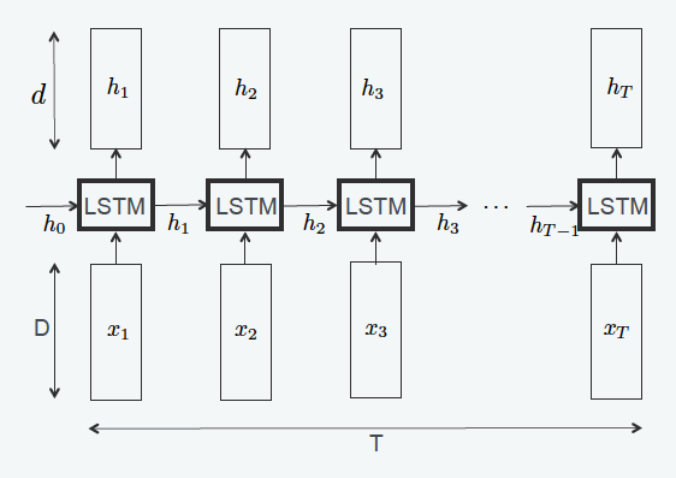
How does the sigmoid activation function protect the cell state? 1 point

The sigmoid function is the activation function applied at each gate. The gate is a way of protecting the cell state (memory). We use the sigmoid transformation which has the highest gradient around 0 and moves quickly to either zero or one. Since most values are close to zero or one, when we perform the sigmoid activation, inputs with values close to zero are lost (forgotten) but those close to 1 are remembered and incorporated into the memory process. Thus the sigmoid acts as a filter for important information.

List all the parameters of the LSTM layer that should be learned using Gradient Descent. 1 point

The four weight matrices, Wf, Wi, Wc and Wo and the three associated bias terms, bf, bi, bc and bo.

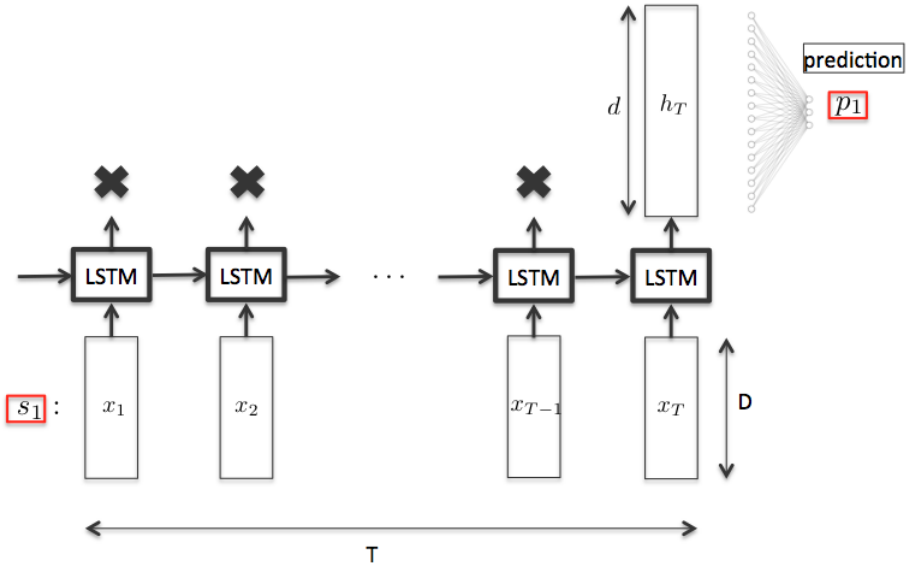
For each sequence x_1, ..., x_T, let h_1, ..., h_T represent the output vectors. What information is represented by the vector h_t for each t in {1, ..., T}? 1 point



ht is the value of the memory cell for each t in the range specified. It represents the information gained from the current input, xt, as well as memory stored in the cell state at step t-1 (ct-1). Hence it is a compound of these two pieces of information.

The Supervised Model

Let's describe the forward propagation for the first sequence s_1 = x_1, ..., x_T. The sequence is fed into an LSTM layer. We only keep the last output vector h_T of size d. The vector h_T is then fed into a Dense layer to output a vector of size 3.



Describe the evolution of the shape of data after each layer transformation: The LSTM layer and the Dense layer. 1 point

The input data that is inputted into the LSTM framework is (N,T,D). After the LSTM layer, the outputted data is of shape (N,d) where N is the number of samples and d is the dimension of the LSTM layer (the vector characterising the data). This data is then fed into the dense layer which outputs a probability distribution of dimension 3. Hence the output would be (N,3), with a probability distribution for each sample

What activation function should be used in the Dense layer? 1 point

Softmax

What loss function should be used? 1 point

Categorical Cross Entropy

Programming Session

Did you understand the problem?

- (●) Yes
- ( ) No

Do you have any questions about the Coursework?

Feel free to send us an email if you need more support.