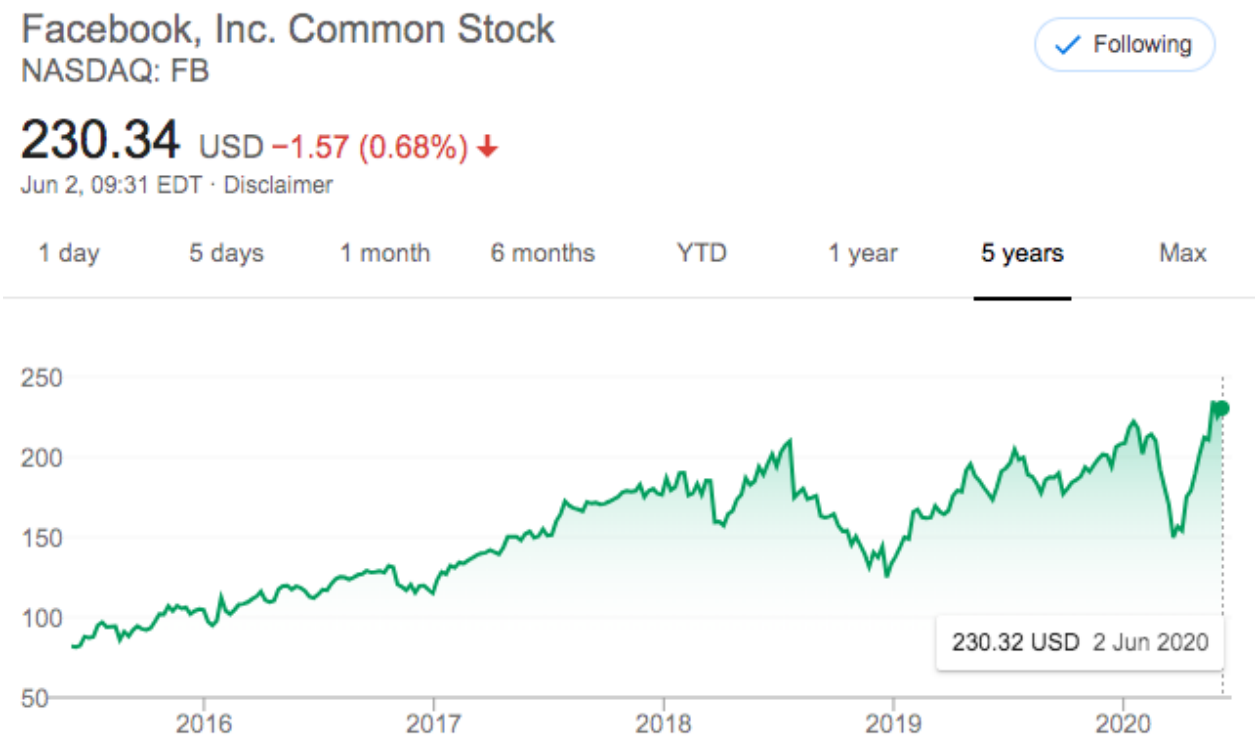
Quiz 6: Introduction to Sequence Models

Introduction to Supervised Learning

* Indicates required question

1. Please enter your name: *

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

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

Mark only one oval.

- F
- F - T
- F - T - 1

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

Mark only one oval.

- (N, D)
- (N, T, D)
- (N, T)

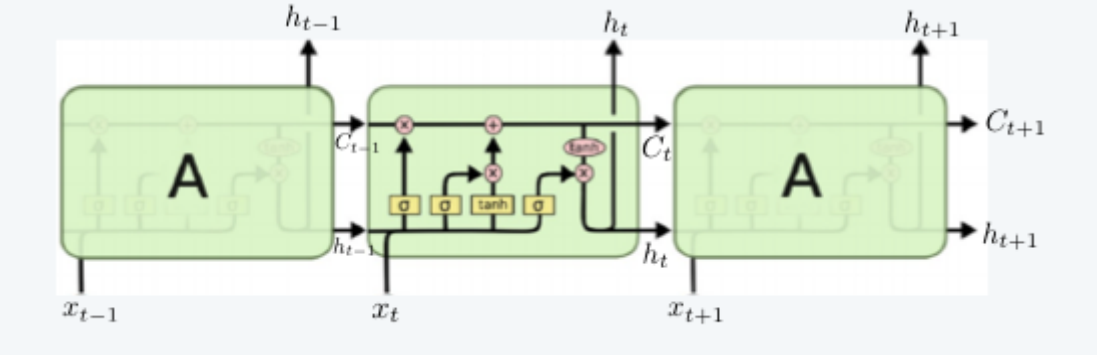
4. What is the shape of our training target data after the one-hot encoding of the targets? 1 point

Mark only one oval.

- (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.

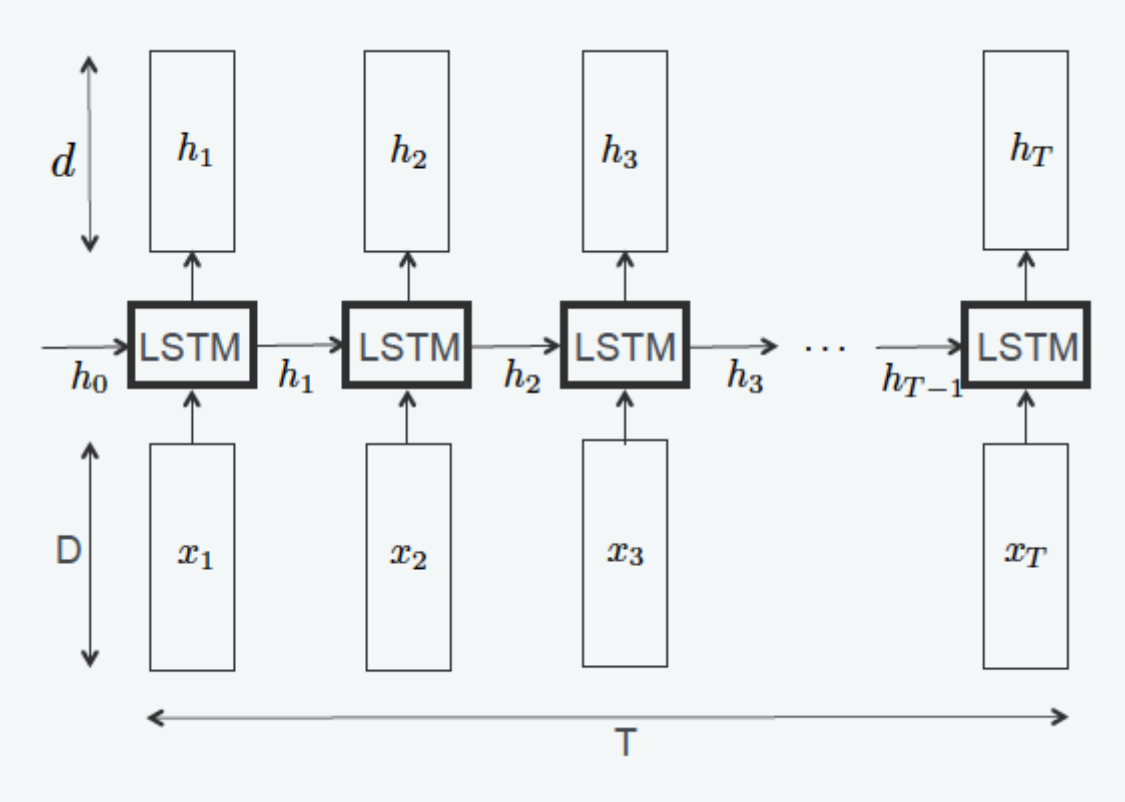


5. Why choosing an LSTM layer over a standard RNN layer? 1 point

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

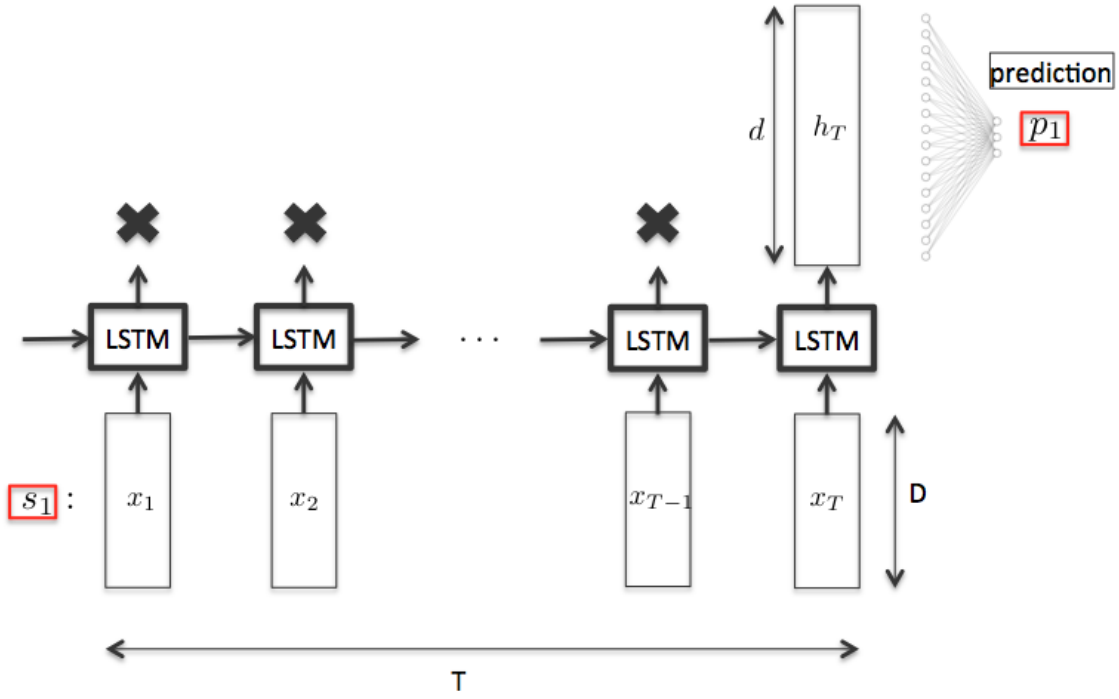
7. List all the parameters of the LSTM layer that should be learned using Gradient Descent. 1 point

8. 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



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.



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

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

11. What loss function should be used? 1 point

Programming Session

12. Did you understand the problem?

Mark only one oval.

- Yes
- No

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

13. Any comment?