Compute the number of parameters in a LSTM cell (*)

Let nfeatures denote the dimensionality of the input vectors, such that $\underline{x}_t \in \mathbb{R}^{nfeatures \times 1}$. $(\forall t \in \{1, 2, \dots, n\})$

Let nunits denote the dimensionality of the hidden state, such that $\underline{h}_t \in \mathbb{R}^{nunits \times 1}$.

Then, we have to take into consideration the fact that each LSTM cell operates on the basis of 4 sets of weight parameters such that:

a)
$$\left\{ \begin{array}{l} \underline{\underline{W}}_i \in \mathbb{R}^{nunits \times nfeatures} \\ \underline{\underline{U}}_i \in \mathbb{R}^{nunits \times nunits} \\ \underline{b}_i \in \mathbb{R}^{nunits \times 1} \end{array} \right\} \text{ For the input gate}$$

$$\boxed{\underline{i}_t = \sigma_g(\underline{\underline{U}}_i \cdot \underline{h}_{t-1} + \underline{\underline{W}}_i \cdot \underline{x}_t + \underline{b}_i)} \quad \underline{i}_t \in \mathbb{R}^{nunits \times 1}$$

where $\sigma_g(\cdot)$ is the sigmoid function

B)
$$\left\{ \begin{array}{l} \underline{\underline{W}}_f \in \mathbb{R}^{nunits \times nfeatures} \\ \underline{\underline{U}}_f \in \mathbb{R}^{nunits \times nunits} \\ \underline{b}_f \in \mathbb{R}^{nunits \times 1} \end{array} \right\} \text{ For the forget gate}$$

$$\boxed{\underline{f}_t = \sigma_g(\underline{\underline{U}}_f \cdot \underline{h}_{t-1} + \underline{\underline{W}}_f \cdot \underline{x}_t + \underline{b}_f)} \quad \underline{f}_t \in \mathbb{R}^{nunits \times 1}$$

(*) Function $\sigma_g(\cdot)$ is applied element wise.

γ)
$$\left.\begin{cases} \underline{\underline{W}}_c \in \mathbb{R}^{n_{units} \times n_{features}} \\ \underline{\underline{U}}_c \in \mathbb{R}^{n_{units} \times n_{units}} \\ \underline{b}_c \in \mathbb{R}^{n_{units} \times 1} \end{cases}\right\}$$
For the candidate cell state

$$\boxed{\underline{\tilde{c}}_t = \sigma_h (\underline{\underline{U}}_c \cdot \underline{h}_{t-1} + \underline{\underline{W}}_c \cdot \underline{x}_t + \underline{b}_c)} \qquad \tilde{c}_t \in \mathbb{R}^{n_{units} \times 1}$$

where $\sigma_h(\cdot)$ is the $\tanh(\cdot)$ function which is also applied element-wise.

δ)
$$\left.\begin{cases} \underline{\underline{W}}_o \in \mathbb{R}^{n_{units} \times n_{features}} \\ \underline{\underline{U}}_o \in \mathbb{R}^{n_{units} \times n_{units}} \\ \underline{b}_o \in \mathbb{R}^{n_{units} \times 1} \end{cases}\right\}$$
For the output gate.

$$\boxed{\underline{o}_t = \sigma_g (\underline{\underline{U}}_o \cdot \underline{h}_{t-1} + \underline{\underline{W}}_o \cdot \underline{x}_t + \underline{b}_o)} \qquad \underline{o}_t \in \mathbb{R}^{n_{units} \times 1}$$

ε) Finally, the new cell and hidden states are computed according to the following equations:

$$\boxed{\underline{c}_t = \underline{f}_t \circledast \underline{c}_{t-1} + \underline{i}_t \circledast \underline{\tilde{c}}_t} \qquad \underline{c}_t \in \mathbb{R}^{n_{units} \times 1}$$

$$\boxed{\underline{h}_t = \underline{o}_t \circledast \sigma_h(\underline{c}_t)} \qquad \underline{h}_t \in \mathbb{R}^{n_{units} \times 1}$$

Where $\circledast$ denotes the element-wise multiplication.

In this context, the total number of parameters N_{LSTM} for the LSTM cell will be given as:

$$N_{LSTM} = 4 * (h_{units} * n_{features} + h_{units} * h_{units} + h_{units} * 1)$$

(★) For our example we have that $h_{units} = 2 * n = 2 * 10 = 20$, $n_{features} = 1$

Thus, we get $N_{LSTM} = 4 * (20 * 1 + 20 * 20 + 20 * 1) \Rightarrow$

$N_{LSTM} = 4 * (20 + 400 + 20) \Rightarrow$

$N_{LSTM} = 4 * 440 = 1760$ total LSTM PARAMETERS

(★) The final output of the DENSE layer depends upon the weight matrix $\underline{W}_0 \in \mathbb{R}^{h_{units} * n_{features}}$ and the corresponding bias-term vector $\underline{b}_0 \in \mathbb{R}^{n_{features} * 1}$.

It will be given by the equation

$$\hat{y}_t = \sigma_n (\underline{W}_0^T \cdot \underline{h}_t + \underline{b}_0)$$

$\underline{W}_0$: [h_{units} * $n_{features}$]

$\underline{h}_t$: [h_{units} x 1]

$\underline{b}_0$: [$n_{features}$ x 1]

[$n_{features}$ x h_{units}] * [h_{units} x 1] + [$n_{features}$ x 1]

[$n_{features}$ x 1] + [$n_{features}$ x 1]

[$n_{features}$]

(★) Total Number of parameters for the Dense layer

$N_{DENSE} = h_{units} * n_{features} + n_{features}$

$= 20 * 1 + 1 = 21$