$$
\begin{aligned}
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
c'_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= f_t \circ c_{t-1} + i_t \circ c'_t \\
h_t &= o_t \circ \tanh(c_t)
\end{aligned} \tag{1}
$$

Variables:
M: input vector dimension
N: output vector dimension
(H_i: hidden layer i vector dimension)
x_t $[M \times 1]$: input vector
h_t $[N \times 1]$: output vector
$h_0 = 0$
c_t $[N \times 1]$: cell state vector
$c_0 = 0$
W $[N \times M]$, U $[N \times N]$ and b $[N \times 1]$: parameter matrices and vector (W is for weight, U is for update?, and b is for bias?)
f_t, i_t and o_t: gate vectors
f_t $[N \times 1]$: Forget gate vector. Weight of remembering old information.
i_t $[N \times 1]$: Input gate vector. Weight of acquiring new information.
o_t $[N \times 1]$: Output gate vector. Output candidate.

$$
\begin{aligned}
\delta h_t + &= target_t - h_t \\
\delta o_t &= \delta h_t \circ \tanh(c_t) \\
\delta c_t + &= \delta h_t \circ o_t \circ \tanh'(c_t) \\
\delta i_t &= \delta c_t \circ c'_t \\
\delta f_t &= \delta c_t \circ c_{t-1} \\
\delta c'_t &= \delta c_t \circ i_t \\
\delta c_{t-1} &= \delta c_t \circ f_t \\
\delta \hat{i}_t &= \delta c_t \circ c'_t \\
\delta \hat{f}_t &= \delta c_t \circ c_{t-1} \\
\delta \hat{c'}_t &= \delta c'_t \circ (1 - c'^2_t) \\
\delta \hat{i}_t &= \delta i_t \circ i_t \circ (1 - i_t) \\
\delta \hat{f}_t &= \delta f_t \circ f_t \circ (1 - f_t) \\
\delta \hat{o}_t &= \delta o_t \circ o_t \circ (1 - o_t) \\
\delta W_{i,f,o,c} &= \delta \hat{i}, \hat{f}, \hat{o}, \hat{c'}_t x_t^T \\
\delta U_{i,f,o,c} &= \delta \hat{i}, \hat{f}, \hat{o}, \hat{c'}_t h_{t-1}^T \\
\delta b_{i,f,o,c} &= \delta \hat{i}, \hat{f}, \hat{o}, \hat{c'}_t \\
\delta h_{t-1} &= \sum U_{i,f,o,c}^T \delta \hat{i}, \hat{f}, \hat{o}, \hat{c'}_t
\end{aligned}
\tag{2}
$$

$\delta h_T = 0$
$\delta c_T = 0$