1. Design an LSTM that calculates the parity of an input bit-string (each $x_i$ is one of 0 or 1). The parity of the input so far ($x_1, ..., x_t$) should be stored in $c_t$. Provide the values (determined by hand) for the weight matrices ($W_{xf}, W_{xi}, W_{xo}, W_{xc}, W_{hf}, W_{hi}, W_{ho}, W_{hc}$), the biases ($b_f, b_i, b_o, b_c$), and $h_0$. Provide an explanation of how the weight matrices you provide calculate the parity correctly.

For the purposes of this assignment (to make for easy calculations), assume that:

\[
\sigma(x) = \begin{cases} 
1 & \text{if } x > 0 \\
0 & \text{otherwise}
\end{cases}
\]

\[
\tanh(x) = \begin{cases} 
1 & \text{if } x > 0 \\
0 & \text{if } x = 0 \\
-1 & \text{otherwise}
\end{cases}
\]

Hint: XOR($a, b$) is equivalent to $(a \land \neg b) \lor (\neg a \land b)$. Can you use that equivalency when updating $c_t$ (equation 4 of Notes on Gated Units).

**Solution:** Let’s use induction on $t$.

Assume that $c_{t-1}$ correctly contains the parity of $x_1, ..., x_{t-1}$. Then, the parity of $x_1, ..., x_t$ is just XOR($c_{t-1}, x_t$). The hint points us to

\[
c_t = f_t \odot c_{t-1} + i_t \odot \tanh(x_t W_{xc} + h_{t-1} W_{hc} + b_c)
\]

If we could ensure that:

\[
f_t = \neg x_t \quad (1)
\]

\[
i_t = x_t \quad (2)
\]

\[
\tanh(x_t W_{xc} + h_{t-1} W_{hc} + b_c) = \neg c_t \quad (3)
\]

then we’d know that $c_t$ contains the parity of $x_1, ..., x_t$.

If we define:

\[
W_{xf} = [-1]
\]

\[
W_{hf} = [0]
\]

\[
b_f = [1]
\]

then $f_t = 1$ when $x_t$ is 0 and $f_t = 0$ when $x_t$ is 1. Thus, equation (1) is satisfied.

Let’s ensure that $i_t = x_t$ by defining:

\[
W_{xi} = [1]
\]

\[
W_{hi} = [0]
\]

\[
b_i = [0]
\]
So, $i_t = 1$ iff $x_t = 1$ and equation (2) is satisfied.

Let’s ensure that $o_t = 1$ by defining:

$$W_{xi} = [0]$$

$$W_{hi} = [0]$$

$$b_i = [1]$$

Now, assuming that $c_t$ is either 0 or 1, $h_t$ has the same value.

If we define:

$$W_{xc} = [0]$$

$$W_{hc} = [-1]$$

$$b_c = [1]$$

then $\tanh(x_t W_{xc} + h_{t-1} W_{hc} + b_c) = 1$ when $h_{t-1}(= c_{t-1})$ is 0 and = 0 when $h_{t-1}(= c_{t-1})$ is 1. So, equation (3) is satisfied.

$h_0$ should be set to $[0]$ so that the base case is met (parity of no inputs is 0).

Thus, we’ve shown that for all $t > 0$, $c_t$ contain the parity of the binary inputs $x_1, ..., x_t$.

Note that there are many different solutions!