Q1. (8 *points* in total)

(a) Yes. (1 *points*)

(b) (1,0), (3,1), and (3,-1) (3 *points*)

(c) We compute the decision boundary using the support vectors:

$$w_1 \cdot 1 + w_2 \cdot 0 + b = -1$$
$$w_1 \cdot 3 + w_2 \cdot 1 + b = 1$$
$$w_1 \cdot 3 + w_2 \cdot -1 + b = 1$$

By solving the equation, we can get $w_1 = 1$, $w_2 = 0$, and $b = -2$. The decision boundary is x_1-2=0. (3 *points*)

(d) (0,0) is negative (0.5 *points*), because $w_1 \cdot 0 + w_2 \cdot 0 + b = -2 < -1$. (3,3) is positive (0.5 *points*), because $w_1 \cdot 3 + w_2 \cdot 3 + b = 1 >= 1$

Q2. (12 *points* in total)

(a)

Consider timestamp t_1

$$
\begin{aligned}
f_1 &= \sigma(W_f[x_1, y_0] + b_f) \\
&= \sigma(\begin{pmatrix}0.7\\0.4\\0.1\end{pmatrix}\begin{pmatrix}0.3\\0.6\\0\end{pmatrix} + 0.1) \\
&= \sigma(0.7 \cdot 0.3 + 0.4 \cdot 0.6 + 0.1 \cdot 0 + 0.1) \\
&= \sigma(0.55) \\
&= 0.6341 \\
i_1 &= \sigma(W_i[x_1, y_0] + b_i) \\
&= \sigma(\begin{pmatrix}0.2\\0.6\\0.7\end{pmatrix}\begin{pmatrix}0.3\\0.6\\0\end{pmatrix} + 0.4) \\
&= \sigma(0.2 \cdot 0.3 + 0.6 \cdot 0.6 + 0.7 \cdot 0 + 0.4) \\
&= \sigma(0.82) \\
&= 0.6942 \\
a_1 &= tanh(W_a[x_1, y_0] + b_a) \\
&= tanh(\begin{pmatrix}0.3\\0.2\\0.1\end{pmatrix}\begin{pmatrix}0.3\\0.6\\0\end{pmatrix} + 0.3) \\
&= tanh(0.3 \cdot 0.3 + 0.2 \cdot 0.6 + 0.1 \cdot 0 + 0.3) \\
&= tanh(0.51) \\
&= 0.4699 \\
o_1 &= \sigma(W_o[x_1, y_0] + b_o) \\
&= \sigma(\begin{pmatrix}0.6\\0.3\\0.1\end{pmatrix}\begin{pmatrix}0.3\\0.6\\0\end{pmatrix} + 0.2) \\
&= \sigma(0.6 \cdot 0.3 + 0.3 \cdot 0.6 + 0.1 \cdot 0 + 0.2) \\
&= \sigma(0.56) \\
&= 0.6365 \\
s_1 &= f_1 \cdot s_0 + i_1 \cdot a_1 \\
&= 0.6341 \cdot 0 + 0.6942 \cdot 0.4699 \\
&= 0.3262 \\
y_1 &= o_1 \cdot tanh(s_1) \\
&= 0.6365 \cdot tanh(0.3262) \\
&= 0.2006
\end{aligned}
$$

Consider timestamp t_2

$$
\begin{aligned}
f_2 &= \sigma(W_f[x_2, y_1] + b_f) \\
&= \sigma(\begin{pmatrix}0.7\\0.4\\0.1\end{pmatrix}\begin{pmatrix}0.1\\1.0\\0.2006\end{pmatrix} + 0.1) \\
&= \sigma(0.7 \cdot 0.1 + 0.4 \cdot 1.0 + 0.1 \cdot 0.2006 + 0.1) \\
&= \sigma(0.5901) \\
&= 0.6434 \\
i_2 &= \sigma(W_i[x_2, y_1] + b_i) \\
&= \sigma(\begin{pmatrix}0.2\\0.6\\0.7\end{pmatrix}\begin{pmatrix}0.1\\1.0\\0.2006\end{pmatrix} + 0.4) \\
&= \sigma(0.2 \cdot 0.1 + 0.6 \cdot 1.0 + 0.7 \cdot 0.2006 + 0.4) \\
&= \sigma(1.1604) \\
&= 0.7614 \\
a_2 &= tanh(W_a[x_2, y_1] + b_a) \\
&= tanh(\begin{pmatrix}0.3\\0.2\\0.1\end{pmatrix}\begin{pmatrix}0.1\\1.0\\0.2006\end{pmatrix} + 0.3) \\
&= tanh(0.3 \cdot 0.1 + 0.2 \cdot 1.0 + 0.1 \cdot 0.2006 + 0.3) \\
&= tanh(0.5501) \\
&= 0.5006 \\
o_2 &= \sigma(W_o[x_2, y_1] + b_o) \\
&= \sigma(\begin{pmatrix}0.6\\0.3\\0.1\end{pmatrix}\begin{pmatrix}0.1\\1.0\\0.2006\end{pmatrix} + 0.2) \\
&= \sigma(0.6 \cdot 0.1 + 0.3 \cdot 1.0 + 0.1 \cdot 0.2006 + 0.2) \\
&= \sigma(0.5801) \\
&= 0.6411 \\
s_2 &= f_2 \cdot s_1 + i_2 \cdot a_2 \\
&= 0.6434 \cdot 0.3262 + 0.7614 \cdot 0.5006 \\
&= 0.5910 \\
y_2 &= o_2 \cdot tanh(s_2) \\
&= 0.6411 \cdot tanh(0.5910) \\
&= 0.3402
\end{aligned}
$$

Suggested Marking Scheme: deduct 0.5 point for each wrongly computed gate variable.

(b)

At t_1,

$$\begin{aligned} error &= y_1 - y \\ &= 0.2006 - 0.2 \\ &= 0.0006 \end{aligned}$$

At t_2,

$$\begin{aligned} error &= y_2 - y \\ &= 0.3402 - 0.4 \\ &= -0.0598 \end{aligned}$$

Suggested Marking Scheme: deduct 1 point for each wrongly computed error.