1.

$$b_h = \begin{bmatrix} -0.5 \\ -1.5 \\ -2.5 \end{bmatrix} \qquad b_y = -0.5 \qquad V = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$W = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{matrix} h_1^{t-1} \\ h_2^{t-1} \\ h_3^{t-1} \end{matrix}$$ \# This means we only take h_2^{t-1} into account for next timestep and ignore h_1^{t-1} and h_3^{t-1}

$$U = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

activation: $h(z) = \begin{cases} 1 & , \text{ if } z \geq 0 \\ 0 & , \text{ if } z < 0 \end{cases}$

2. (a) $\overline{h^{(t)}} = \overline{g^{(t+1)}}(1 - \tanh^2(W_{gx}x^{(t+1)} + W_{gh}h^{(t)})) \cdot W_{gh} +$
$\overline{i^{(t+1)}} \sigma'(W_{ix}x^{(t+1)} + W_{ih}h^{(t)}) \cdot W_{ih} +$
$\overline{f^{(t+1)}} \sigma'(W_{fx}x^{(t+1)} + W_{fh}h^{(t)}) \cdot W_{fh} +$
$\overline{o^{(t+1)}} \sigma'(W_{ox}x^{(t+1)} + W_{oh}h^{(t)}) \cdot W_{oh}$

$\overline{c^{(t)}} = \overline{c^{(t+1)}} f^{(t+1)} + \overline{h^{(t)}} o^{(t)}(1 - \tanh^2(c^{(t)}))$

$\overline{g^{(t)}} = \overline{c^{(t)}} i^{(t)}$

$\overline{o^{(t)}} = \overline{h^{(t)}} \tanh(c^{(t)})$

$\overline{f^{(t)}} = \overline{c^{(t)}} c^{(t-1)}$

$\overline{i^{(t)}} = \overline{c^{(t)}} g^{(t)}$

(b) $\overline{W_{ix}} = \sum_t \overline{i^{(t)}} \sigma'(W_{ix}x^{(t)} + W_{ih}h^{(t-1)}) x^{(t)}$