Consider a scalar time series
$$\{x^{(t)} : t = 1, 2, \cdots \}.$$

Output time series:
$$\{y^{(t)} : t = 1, 2, \cdots \}.$$

$$y^{(t)} = x^{(t)} + x^{(t-1)} + x^{(t-2)} \qquad \text{for} \quad t \geq 1.$$

$\{x^{(t)}\}$:	1	2	3	4	5	
$\{y^{(t)}\}$:	1	3	6	9	12	$\cdots$

Goal :   Build a model s.t $\hat{y}(t) = y(t) \qquad \forall t$

① $\quad \hat{y}(t) = FCNN(x^{(t)}) \quad \times$ no memory.

② $\quad \hat{y}(t) = $ General linear function on $\{x^{(1)}, \cdots, x^{(t)}\}$
$$= \sum_{s=1}^{t} a^{(s)} x^{(s)}$$
$\hookrightarrow$ fit to data $\quad \times$

③ $\quad h_1^{(t)} = x^{(t)}$
$h_2^{(t)} = \cancel{\#} h_1^{(t-1)} = x^{(t-1)}$
$h_3^{(t)} = h_2^{(t-1)} = x^{(t-2)}$

Model : $\hat{y}(t) = h_1^{(t)} + h_2^{(t)} + h_3^{(t)}$
$$= x^{(t)} + x^{(t-1)} + x^{(t-2)} = y^{(t)}$$

Vectorize above :
$$\begin{pmatrix} h_1^{(t)} \\ h_2^{(t)} \\ h_3^{(t)} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} h_1^{(t-1)} \\ h_2^{(t-1)} \\ h_3^{(t-1)} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} x^{(t)}$$
$\underbrace{\qquad}_{h(t)}$ $\underbrace{\qquad}_{W}$ $\underbrace{\qquad}_{h(t-1)}$ $\underbrace{\qquad}_{U}$

$$\underbrace{(h_?)}_{h(t)} \quad \underbrace{(\qquad)}_{W} \quad (\underbrace{h_3}_{h(t-1)}) \quad \begin{pmatrix} \smile \\ U \end{pmatrix}$$

$$\hat{y}(t) = (1 \qquad 1 \qquad 1) \begin{pmatrix} h_1^{(t)} \\ h_2^{(t)} \\ h_3^{(t)} \end{pmatrix}$$

$$\underbrace{\qquad}_{V} \qquad \underbrace{\qquad}_{h(t)}$$

$$h(t) = \sigma_r \left(W h^{(t-1)} + U x^{(t)} + b \right)$$
$$\hat{y}(t) = \sigma_o \left(V \cdot h^{(t)} + c \right)$$