

Name: _____

Delay lines

7. (10 points) Recall the specification of a standard recurrent neural network (RNN): input x_t of dimension $\ell \times 1$, state s_t of dimension $m \times 1$, and output y_t of dimension $v \times 1$. The weights in the network, then, are

$$W^{sx} : m \times \ell$$

$$W^{ss} : m \times m$$

$$W^O : v \times m$$

with activation functions f_1 and f_2 . **Throughout this problem, for simplicity, we will treat all offsets as equal to 0.** Finally, the operation of the RNN is described by

$$s_t = f_1(W^{sx}x_t + W^{ss}s_{t-1})$$

$$y_t = f_2(W^O s_t) \quad .$$

- (a) Consider an RNN defined by $\ell = 1$, $m = 2$, $v = 1$, $f_1 = f_2 =$ the identity function, and

$$W^{sx} = \begin{bmatrix} 5 \\ 6 \end{bmatrix} \quad W^{ss} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad W^O = [-3 \quad -2]$$

Assuming the initial state is all 0, and the input sequence is $[[1], [-1]]$, what is the output sequence?

Name: _____

(b) We can think of the RNN as mapping input sequences to output sequences. Jody thinks that if we remove f_1 and f_2 then the mapping from input sequence to output sequence can be achieved by a hypothesis of the form $Y = WX$. In the case of a length 3 sequence, assuming inputs and outputs are 1-dimensional, $s_0 = [0]$, $X = [x_1, x_2, x_3]^T$, $Y = [y_1, y_2, y_3]^T$, and W is 3×3 .

i. Is Jody right? Yes No

ii. If Jody is right, provide a definition for W in Jody's model in terms of W^{sx} , W^{ss} , and W^O of the original RNN that makes them equivalent. If Jody is wrong, explain why.

Name: _____

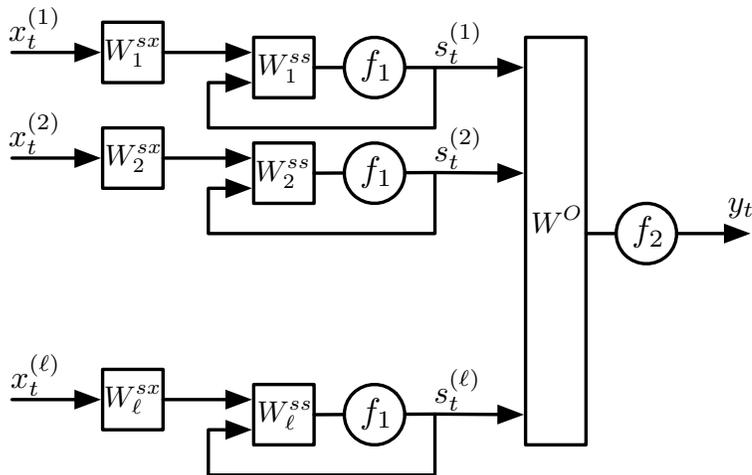
(c) Pat thinks a different RNN model would be good. Its operation is defined by

$$s_t^{(i)} = f_1 \left(W_i^{sx} x_t^{(i)} + W_i^{ss} s_{t-1}^{(i)} \right)$$

$$y_t = f_2 \left(W^O s_t \right) .$$

where the dimension of the state, $m = k \cdot \ell$, so there are k state dimensions for each input dimension, $s^{(i)}$ is the i th group of k dimensions in the state vector, $x^{(i)}$ is the i th entry in the input vector, W_i^{sx} is $k \times 1$ and W_i^{ss} is $k \times k$.

Here is a diagram.



- i. Can this model represent the same set of state machines as a regular RNN?
 - Yes No
- ii. If yes, explain how to convert the weights of a regular RNN into weights for Pat's model.
 If no, describe a concrete input/output relationship (for example, the output y_t is the sum of all the inputs $x_t^{(1)}, \dots, x_t^{(\ell)}$) that **can** be represented by a regular RNN but cannot be represented by Pat's model, for any value of k .