

1)

a) False. RNN can only access information from 1 previous time.

b) False. Does not consider any future input for the current state. However, the calculations does take into account historical information.

c) True. RNN are feedforward networks with a feedback loop and unrolled across time. Parameter sharing is necessary for efficiently processing sequential data.

d) False. Model size does not increase with size of input. The input is a vector and output is a attribute.

e) True. The input vector can be arbitrary in length. The input can't be so large that it causes computers to crash or something.

2)

 T_1

$$z_1 = \text{sig}([0.23 \ 0.12] \begin{bmatrix} 0.35 \\ 0.6 \end{bmatrix} + 0)$$

$$= 0.5381$$

$$r_1 = \text{sig}([0.23 \ 0.12] \begin{bmatrix} 0.45 \\ 0.12 \end{bmatrix} + 0)$$

$$= 0.5294$$

$$h_1 = \tanh([0.23 \ 0.12] \begin{bmatrix} 0.56 & 0.1 & 0.3 \\ 0.2 & 0.5 & 0.21 \end{bmatrix} + 0)$$

$$= [0.15084 \ 0.08281 \ 0.09372]$$

$$s_1 = (1 - 0.5381) [0.15084 \ 0.08281 \ 0.09372]$$

$$= [0.07005 \ 0.03825 \ 0.04332]$$

$$= [0.06952 \ 0.038167 \ 0.043195]$$

$$z_2 = \text{sig}([0.48 \ 0.98] \begin{bmatrix} 0.35 \\ 0.6 \end{bmatrix} + [0.06952 \ 0.038167 \ 0.043195] \begin{bmatrix} 0.12 \\ 0.67 \\ 0.34 \end{bmatrix})$$

$$= \text{sig}(6.756 + 0.04867) = 0.69097$$

$$r_2 = \text{sig}([0.48 \ 0.98] \begin{bmatrix} 0.45 \\ 0.12 \end{bmatrix} + s_1 \cdot \begin{bmatrix} 0.21 \\ 0.22 \\ 0.45 \end{bmatrix})$$

$$\text{sig}(0.3336 + 0.0425) = 0.5929$$

$$h_2 = \tanh([0.48 \ 0.98] \begin{bmatrix} 0.56 & 0.1 & 0.3 \\ 0.2 & 0.5 & 0.21 \end{bmatrix} + (s_1 \cdot 0.5929) \cdot \begin{bmatrix} 0.54 & 0.1 & 0.2 \\ 0.2 & 0.66 & 0.3 \\ 0.9 & 0.87 & 0.2 \end{bmatrix})$$

$$\nearrow s_t = (1 - 0.69097) \cdot h_2 + 0.69097 \cdot [0.06952 \ 0.038167 \ 0.043195]$$