

# CSE 5525 Speech and Language Processing (Spring 2020)

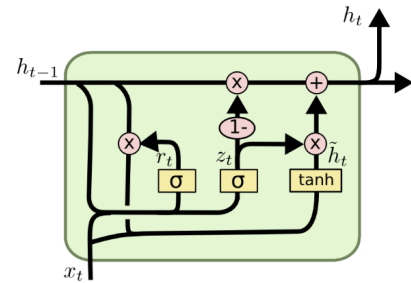
## In-class Exercise: Recurrent Neural Networks

Wei Xu, The Ohio State University

March 4, 2020

Here are the defining equations for a Gated Recurrent Unit (GRU) cell:

$$\begin{aligned} z_t &= \sigma(W^{(z)}x_t + U^{(z)}h_{t-1}) \\ r_t &= \sigma(W^{(r)}x_t + U^{(r)}h_{t-1}) \\ \tilde{h}_t &= \tanh(W^{(h)}x_t + r_t \circ U^{(h)}h_{t-1}) \\ h_t &= (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t \end{aligned}$$



(1) Recall that  $\circ$  denotes element-wise multiplication and that  $\sigma$  denote the sigmoid function  $\sigma(a) = 1/(1 + e^{-a})$ .  $h_t$  and  $x_t$  are column vectors. Assume that the  $h_t$  are of dimension  $d_h$  and that the  $x_t$  are of dimensions  $d_x$ . What are the dimensions of  $W^{(z)}$ ,  $U^{(z)}$ ,  $W^{(r)}$ ,  $U^{(r)}$ ,  $W^{(h)}$ , and  $U^{(h)}$ ? Define clearly which numbers are rows and columns.

$W^{(z)}$ :

$U^{(z)}$ :

$W^{(r)}$ :

$U^{(r)}$ :

$W^{(h)}$ :

$U^{(h)}$ :

(2) True or False.  $z_t$  and  $r_t$  can be viewed as probability distributions (i.e., their entries are non-negative and their entries sum to 1). JUSTIFY YOUR ANSWER.