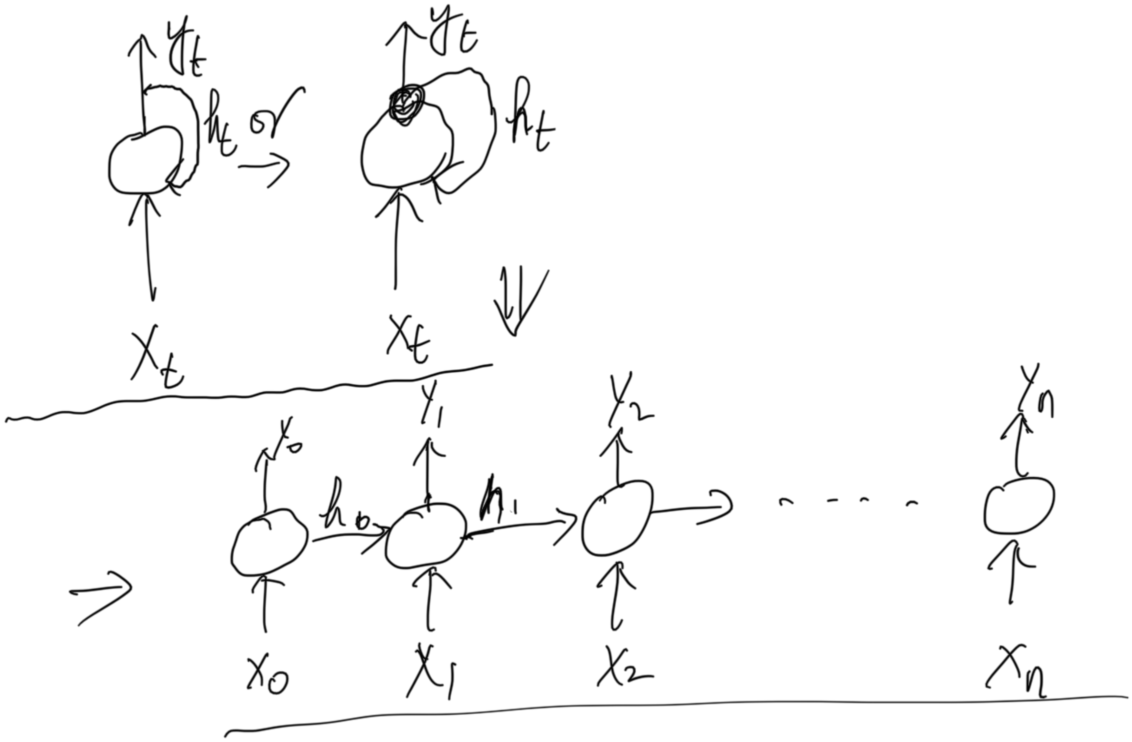
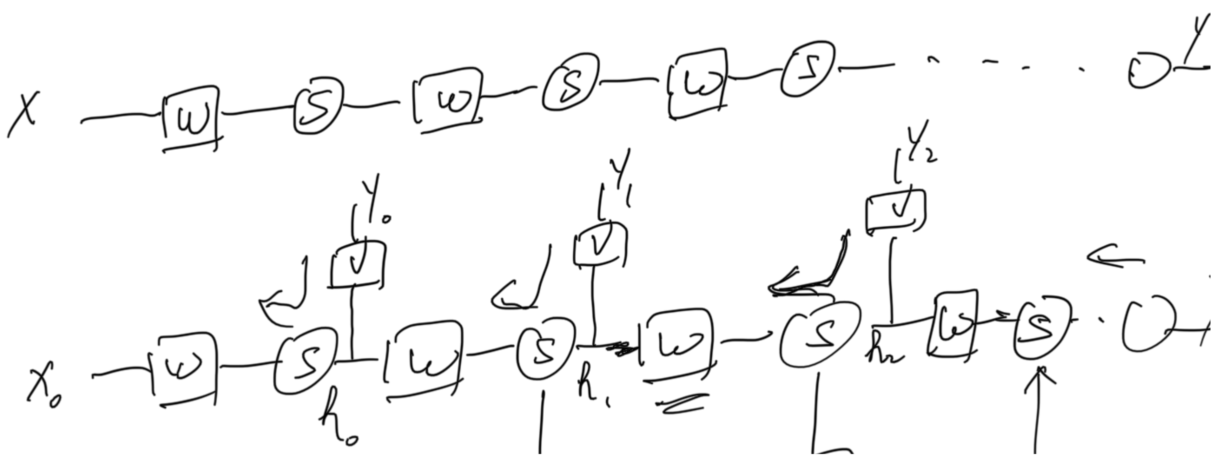
Lecture 7

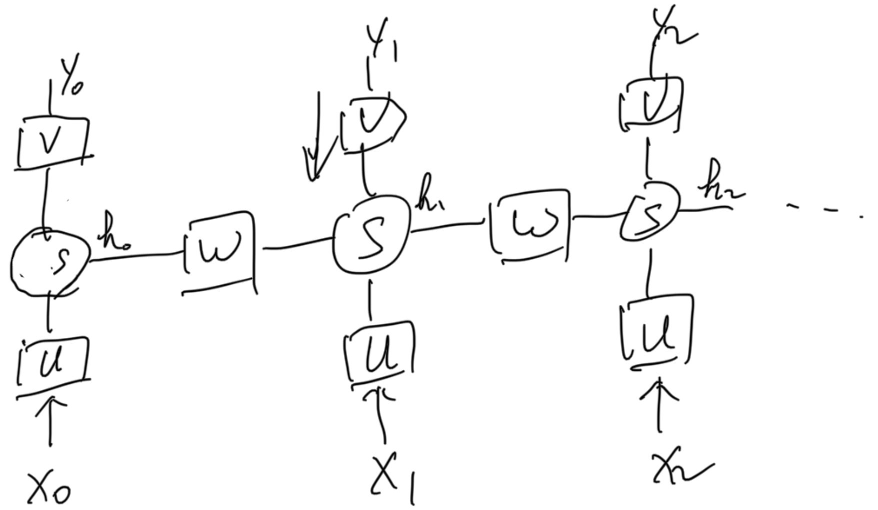


$$h_t = f(x_t, h_{t-1}) \qquad h_1 = f(x_1, h_0)$$

$$y_t = g(x_t, h_{t-1})$$

$$= g(V h_t)$$

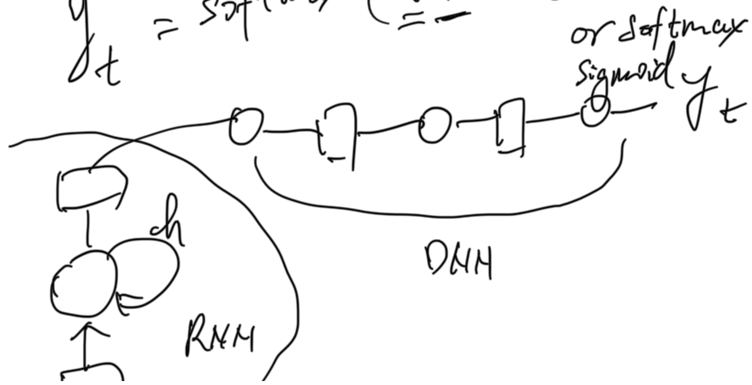




w, u, v

$$h_t = \tanh(w h_{t-1} + u x_t + b_h)$$

$$y_t = \mathrm{softmax}(\underline{V}\underline{h} + b_y)$$



$$1 : (y_0^1, \dots, y_n^1) \text{ vs. } (\hat{y}_0^1, \dots, \hat{y}_n^1)$$

$$\vdots$$

$$m : (y_0^m, \dots, y_n^m) \text{ vs } (\hat{y}_0^m, \dots, \hat{y}_n^m)$$

$$L = \sum_{j=1}^{m} \sum_{i=1}^{n} L(y_i^j, \hat{y}_i^j)$$

$$\frac{\partial L}{\partial V} = \frac{\partial L}{\partial y} \frac{\partial y}{\partial O} \frac{\partial O}{\partial V}$$

$$\frac{\partial L}{\partial W} = \frac{\partial \sum_{i=1}^{n} L(y_i^j, \hat{y}_i^j)}{\partial W}$$

$$= \sum_{i=1}^{n} \frac{\partial L(y_i^j, \hat{y}_i^j)}{\partial W}$$

$$\frac{\partial L(y_i^j, \hat{y}_i^j)}{\partial W} = \frac{\partial L}{\partial y_i} \frac{\partial y_i}{\partial h_i} \frac{\partial h_i}{\partial W}$$

$$\frac{\partial h_i}{\;} = \frac{\partial h_{i-1}}{\;} \cdots \frac{\partial h_2}{\;}$$

$$\partial h_{i-1} \quad \partial h_{i-2} \quad \_ \quad \partial h_1$$

if all of $\frac{\partial h_i}{\partial h_{i-1}} < 1$, ≈ 0, vanishing

if all of $\frac{\partial h_i}{\partial h_{i-1}} > 1$, $\approx \infty$, Exploding

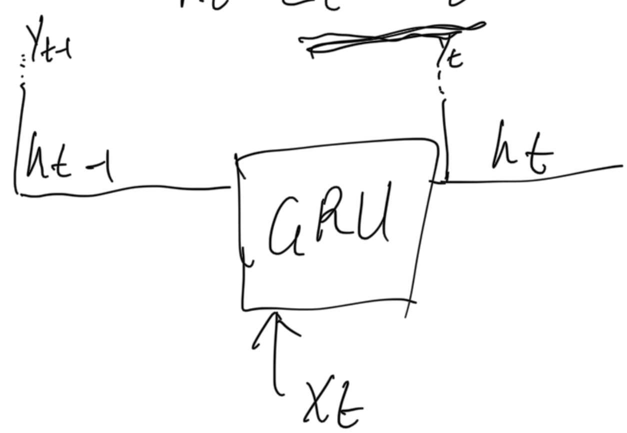
GRU. $\overset{x}{10\times1}$ $\overset{h}{20\times1}$

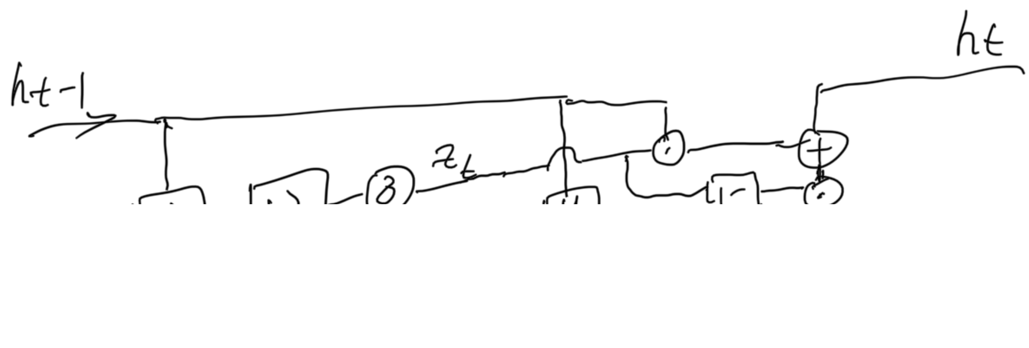
update gate $\overset{20\times1}{z_t} = \sigma\left(\underline{\overset{20\times10}{W_z}\overset{10\times1}{x_t} + \overset{20\times20}{U_z}\overset{20\times1}{h_{t-1}}} + \overset{20\times1}{b_z}\right) \Leftarrow$ $3\times(200+40$ $\underline{186=}$

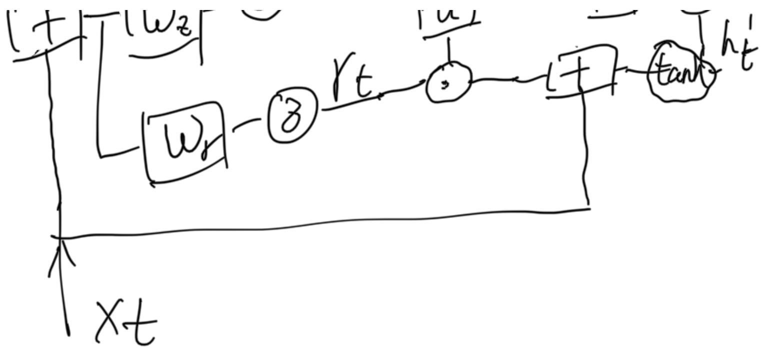
Reset gate $\overset{20\times1}{r_t} = \sigma\left(\overset{20\times10}{W_r}x_t + \overset{20\times20}{U_r}h_{t-1} + \overset{20\times1}{b_r}\right) \Leftarrow$

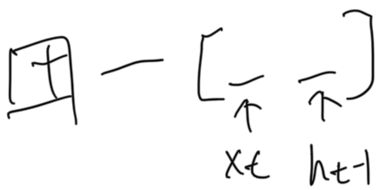
$$h_t' = \tanh\left(\overset{20\times10}{W_h}x_t + r_t \odot \overset{20\times20}{U_h}h_{t-1} + \overset{20\times1}{b_h}\right) \Leftarrow$$

$$h_t = z_t \odot h_{t-1} + (1-z_t) \odot h_t'$$

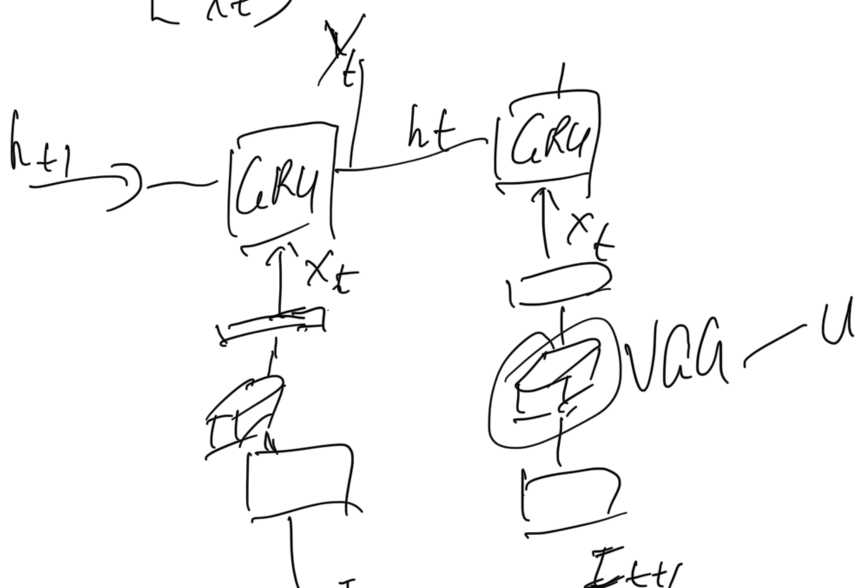


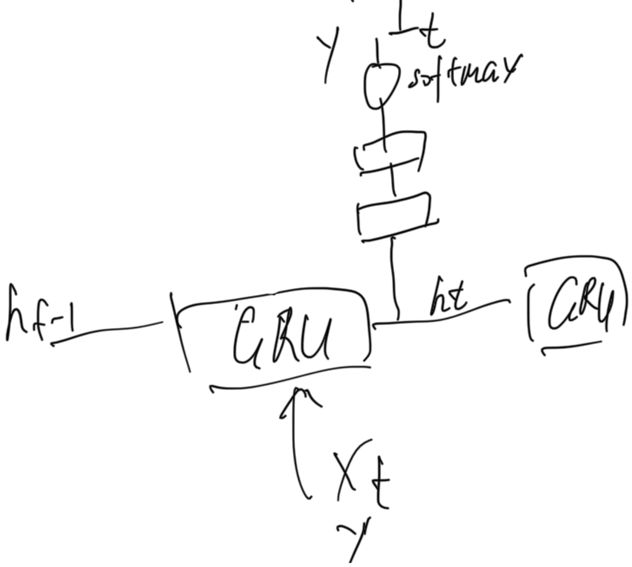


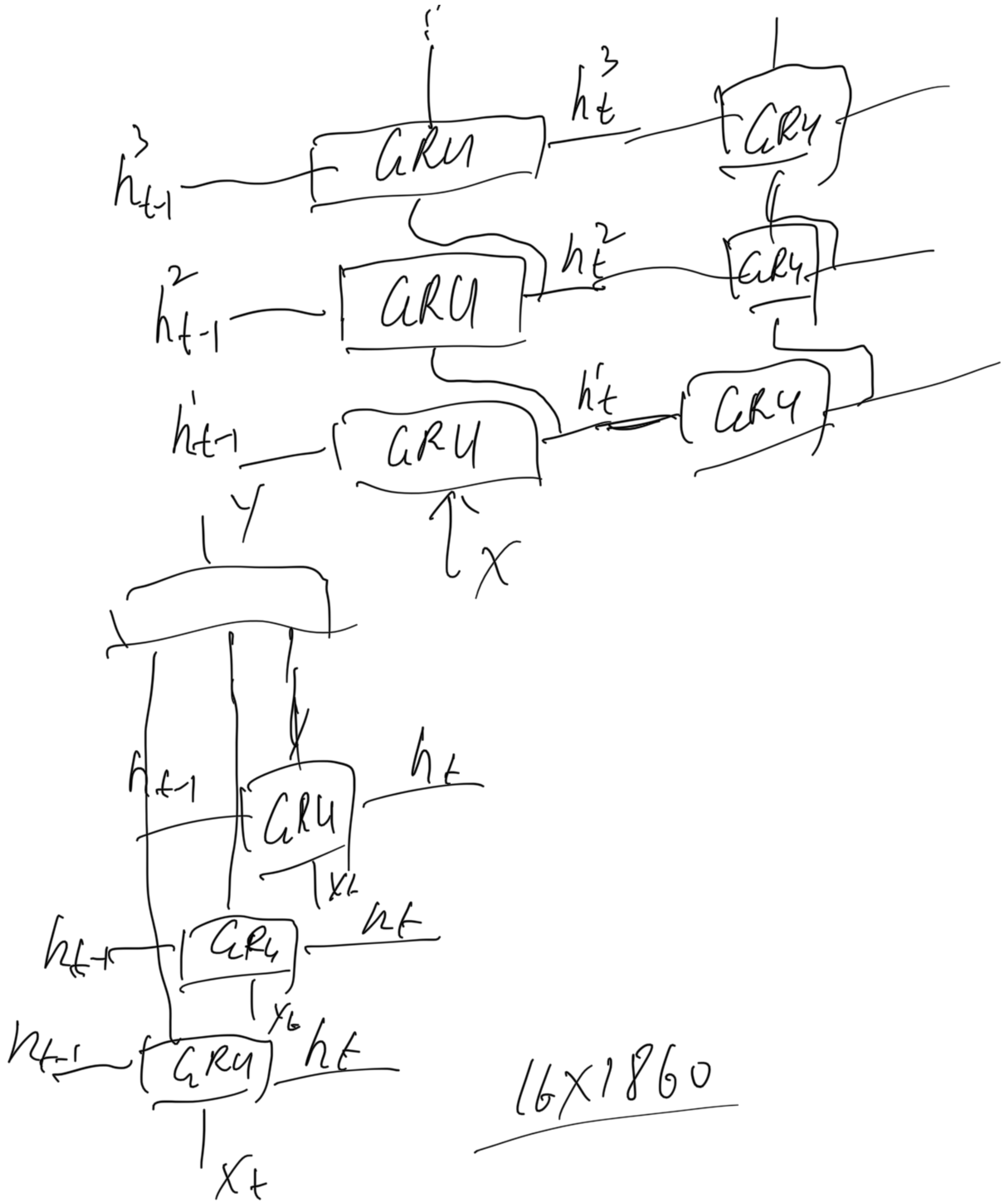




$$[w \quad u]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}$$







h_{t-1}^3
GRU
h_t^3
GRU
h_{t-1}^2
GRU
h_t^2
GRU
h_{t-1}^1
GRU
h_t^1
GRU
Y
X
h_{t-1}
GRU
h_t
x_t
h_{t-1}
GRU
h_t
x_t
h_{t-1}
GRU
h_t
x_t

16X1860