

## Week 6 Homework

Lv Xinpeng

### 1) The forward propagation of LSTM:

In the forget-gate:

$$f^{(t)} = \sigma(W_f h^{(t-1)} + U_f x^{(t)} + b_f)$$

Then , for the input-gate:

$$i^{(t)} = \sigma(W_i h^{(t-1)} + U_i x^{(t)} + b_i)$$

$$a^{(t)} = \tanh(W_a h^{(t-1)} + U_a x^{(t)} + b_a)$$

Next is the update of the cell-state;

$$C^{(t)} = C^{(t-1)} \odot f^{(t)} + i^{(t)} \odot a^{(t)}$$

Final is the output-gate;

$$o^{(t)} = \sigma(W_o h^{(t-1)} + U_o x^{(t)} + b_o)$$

$$h(t) = o^{(t)} \odot \tanh(C^{(t)})$$

So, here is the predicted y:

$$\hat{y}^{(t)} = \sigma(Vh^{(t)} + c)$$

( $W$  is the weight ,  $\odot$  is Hadamard product and  $\sigma$  is the activate function. )

### 2) The forward propagation of GRU:

In the update-gate  $z_t$  :

$$z_t = \sigma(W_z \cdot x_t + U_z \cdot h_{t-1})$$

and in the reset-gate  $r_t$  :

$$r_t = \sigma(W_r \cdot x_t + U_r \cdot h_{t-1})$$

The update of  $x$  from reset-gate :

$$\acute{h}_t = \tanh(W_{\acute{h}} \cdot x_t + r_t \odot U_{\acute{h}} h_{t-1})$$

So, the final memory of the current time:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \acute{h}_t$$

( $W, U$  is the weight and  $\odot$  is Hadamard product) .