

Assignment 6

Fudi(Fred) Wang

November 2019

Part 1

Forward process of LSTM

Let $x(t)$ be the input in time step t , $h(t)$ be the hidden state of $x(t)$ and $\hat{y}(t)$ be the output for $x(t)$.

Forget gate:

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f)$$

Input gate:

$$\begin{aligned} i_t &= \sigma(W_i h_{t-1} + U_i x_t + b_i) \\ a_t &= \tanh(W_a h_{t-1} + U_a x_t + b_a) \end{aligned}$$

Update of cell state

$$C_t = C_{t-1} \odot f_t + i_t \odot a_t$$

where \odot is the Hadamard product.

Output gate:

$$\begin{aligned} o_t &= \sigma(W_o h_{t-1} + U_o x_t + b_o) \\ h(t) &= o_t \odot \tanh(C_t) \end{aligned}$$

Eventually, we have

$$\hat{y}_t = \sigma(V h_t + c)$$

Forward process of GRU

Update gate:

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

Reset gate:

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

Candidate activation:

$$\tilde{h}_t = \tanh(W x_t + U(r_t \odot h_{t-1}))$$

So the activation value h_t becomes:

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