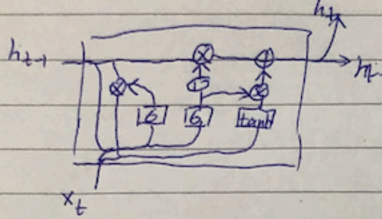
DL Week 9

Date No.

Task 1

1. 

2. $W^z: d_h \times d_x$, $W^r: d_h \times d_x$, $W: d_h \times d_x$

$U^z: d_h \times d_h$, $U^r: d_h \times d_h$, $U^h: d_h \times d_h$

3. LSTMs have a separate update gate and forget gate which are $G_u = \sigma(W_u[a_{t-1}, x_t] + b_u)$ and $G_f = \sigma(W_f[a_{t-1}, x_t] + b_f)$. As a result, it is more complex than GRU.

Task 2. To defend an adversarial attacks, one can add a temperature in the training phase. $\text{softmax}(x, T)[i] = \frac{e^{x_i/T}}{\sum_k e^{x_k/T}}$. This will result in a nearly 0 gradient which can defend attacks using gradient method.

To overcome the defendance, one can use logits layer to avoid dealing with softmax output, $d(x) = \max_{c \neq t} f_c(x) - f_t(x)$ $\min_x \|x - x_0\|^2 + c\max(d(x), 0)$

Task 3. There might not be optimized minimum value because g_w can be $-\infty$. With ubounded loss, the loss is still sufficient to satisfy the goal.

Task 4. $p(00) = \frac{1}{4} \times 0.8 \times 0.8 + \frac{1}{4} \times 0.8 \times 0.8 + \frac{1}{4} \times 0.8 \times 0.8 + \frac{1}{4} \times 0.8 \times 0.8 = 0.64$

$p(01) = - - - - = 0.64$

$p(10) = - - - - = 0.64$

$p(11) = - - - - = 0.64$