SHARIF UNIVERSITY OF TECHNOLOGY

# Deep Learning Assignment 3

Saeedreza Zouashkiani

February 1, 2023

## 1

### 1.1

We can see from the relations that if $z_t$ is zero, then the previous state $h_t$ is ignored and the current state is only determined by new information. Meaning that it is the update gate. On the other hand, if $r_t$ is zero, then the previous state's effect is ignored and the new state is only determined by the previous state. Meaning that it is the reset gate.

### 1.2

$$\frac{\partial J}{\partial W} = \sum_{k=1}^{t} \frac{\partial J^{(t)}}{\partial y_t} \frac{\partial y_t}{\partial W} = \sum_{k=1}^{t} \frac{\partial J^{(t)}}{\partial \hat{y}_k} \frac{\partial \hat{y}_k}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W} \tag{1}$$

$$\frac{\partial h_t}{\partial h_k} = \prod_{j=k+1}^{t} \frac{\partial h_j}{\partial h_{j-1}} \tag{2}$$

$$\frac{\partial h_j}{\partial h_{j-1}} = z_j + (1 - z_j) \frac{\partial \hat{h}_j}{\partial h_{j-1}} \tag{3}$$

We can see that if $z_j = 1$ then $\frac{\partial h_j}{\partial h_{j-1}} = 1$, therefore vanishing gradient problem is solved.