Name : Muhammad Taha Khan
Roll no. 21L-5238
Section : 7B

TASK 1

Input : 5 words
Hidden layer units : 7
Vocabulary : 10
Embedding : 15

$W_H = 7 \times 7$

$h_t = 7 \times 1$

$W_x = 7 \times 15$

$x_t = 15 \times 1$

$a_t = W_H h_{t-1} + W_x x_t$

$(7\times7)(7\times1) + (7\times15)(15\times1)$

$(7\times1) + (7\times1)$

$a_t : \boxed{7\times1}$

$h_t = \tanh(\underset{\boxed{7\times1}}{a_t})$

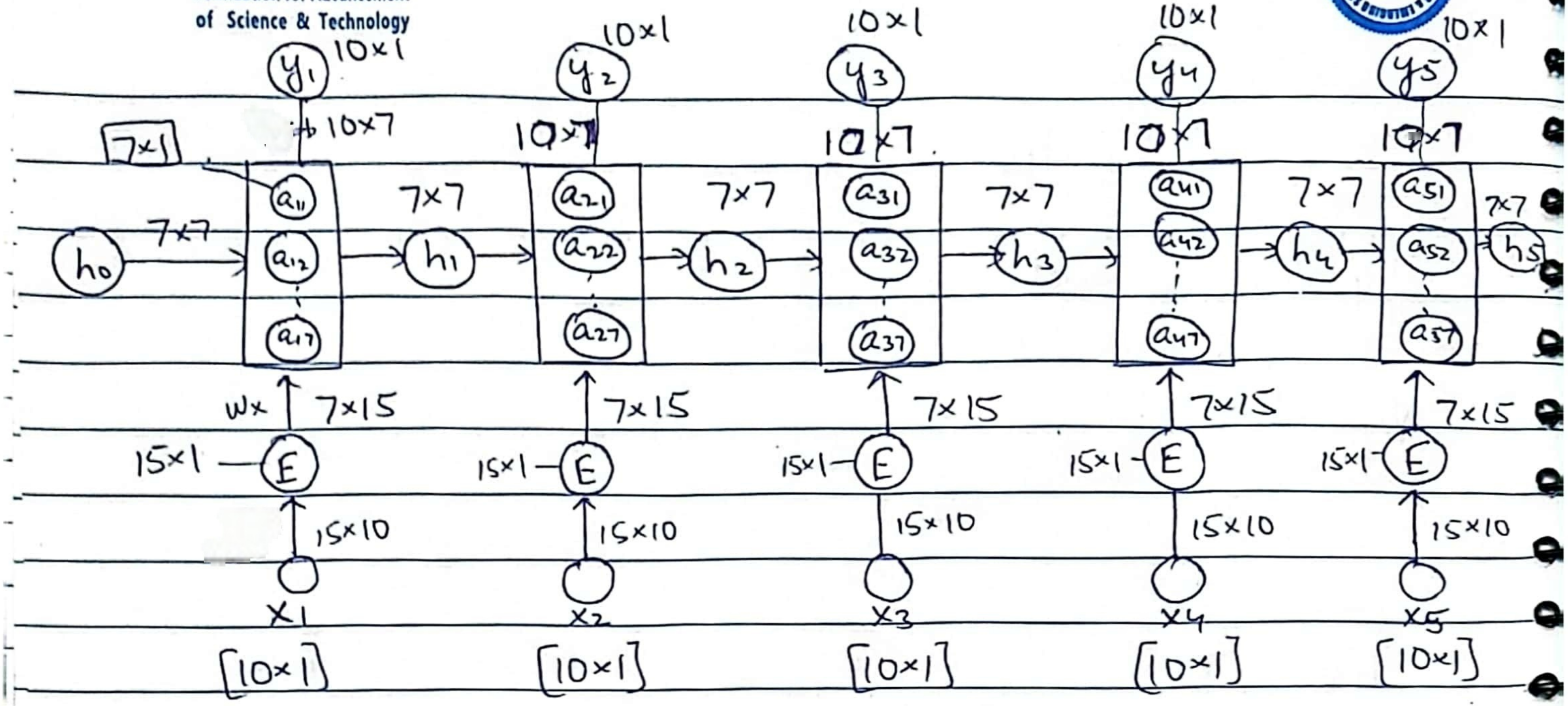
$h_t :$

$h_t : \boxed{7\times1}$

$y_t = \text{Softmax}(W_y h_t)$

$(10\times7)(7\times1)$

$y_t \downarrow (10\times1)$



TASK 2

Partial Derivative of W_H

Function Dependency with respect to W_H

$$L_t = -y_t \log(\hat{y}_t) \rightarrow \hat{y}_t = \text{softmax}(z_t) \rightarrow z_t = W_y h_t \rightarrow h_t = \tanh(W_H h_{t-1} + W_x x_t)$$

Chain Rule

$$\frac{\partial L_t}{\partial W_H} = \frac{\partial L_t}{\partial \hat{y}_t} \cdot \frac{\partial \hat{y}_t}{\partial z_t} \cdot \frac{\partial z_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial W_H}$$

As h_t and h_{t-1} also contain W_H, so chain rule goes recursively to h_0

$$\frac{\partial L_t}{\partial W_H} = \frac{\partial L_t}{\partial \hat{y}_t} \cdot \frac{\partial \hat{y}_t}{\partial z_t} \cdot \frac{\partial z_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial W_H} + \frac{\partial L_t}{\partial \hat{y}_t} \cdot \frac{\partial \hat{y}_t}{\partial z_t} \cdot \frac{\partial z_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial W_H}$$

$$+ \dots + \frac{\partial L_t}{\partial \hat{y}_t} \cdot \frac{\partial \hat{y}_t}{\partial z_t} \cdot \frac{\partial z_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_{t-n}} \cdot \frac{\partial h_{t-n}}{\partial W_H}$$

$$\frac{\partial L_t}{\partial W_H} = \sum_{k=0}^{n} \frac{\partial L_t}{\partial \hat{y}_t} \cdot \frac{\partial \hat{y}_t}{\partial z_t} \cdot \frac{\partial z_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_k} \cdot \frac{\partial h_k}{\partial W_H}$$

$$\frac{\partial L_{total}}{\partial W_H} = \sum_{t=1}^{n} \sum_{k=0}^{n} \frac{\partial L_t}{\partial \hat{y}_t} \cdot \frac{\partial \hat{y}_t}{\partial z_t} \cdot \frac{\partial z_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_k} \cdot \frac{\partial h_k}{\partial W_H}$$