Q1

$$z = \begin{bmatrix} 0.4 \\ 0.8 \end{bmatrix}.$$

$$h_0 = \mathrm{ReLU}(w_0^T z + b_0)$$

$$= \mathrm{ReLU}\left(\begin{bmatrix} 0.38 \\ 1.4 \\ 0.96 \\ 1.06 \\ 0.7 \\ 1.54 \\ 1.62 \\ 1.06 \end{bmatrix}\right) = \begin{bmatrix} 0.38 \\ 1.4 \\ 0.96 \\ 1.06 \\ 0.7 \\ 1.54 \\ 1.62 \\ 1.06 \end{bmatrix}$$

Since $w_1^{3\times3\times2}$, our h_1 should also have 2. channels.

$$h_1^1 = \begin{bmatrix} 0.38 & 1.4 \\ 0.96 & 1.06 \end{bmatrix} \quad h_2^2 = \begin{bmatrix} 0.7 & 1.54 \\ 1.62 & 1.06 \end{bmatrix}$$

Because the filter is $3\times3\times2$, We need to pad 2 pixel of 0 at each edge.

$$h_1^1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.38 & 1.4 & 0 & 0 \\ 0 & 0 & 0.96 & 1.06 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad h_2^2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 1.54 & 0 & 0 \\ 0 & 0 & 1.62 & 1.06 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$C_1 = \mathrm{conv2D}(h_1^1, w_1)$

$C_2 = \mathrm{conv2D}(h_1^2, w_2)$

$$\begin{bmatrix} 0.114 & 0.572 & 0.636 & 0.28 \\ 0.134 & 0.528 & 1.634 & 0.352 \\ 0.286 & 1.63 & 1.436 & 0.246 \\ 0.48 & 0.914 & 0.52 & 0.106 \end{bmatrix} \quad \begin{bmatrix} 0.21 & 0.602 & 0.658 & 0.77 \\ 0.626 & 1.58 & 2.62 & 0.992 \\ 0.744 & 2.664 & 1.734 & 0.626 \\ 0.972 & 0.798 & 0.143 & 0.212 \end{bmatrix}$$

After adding bias,

$$\begin{array}{c|cccc|}
 & h_1[0] & h_2[1] & h_2[2] & h_2[3] \\
h_2 & 0.424 & 1.674 & 1.894 & 1.35 \\
h_2[4] & 0.96 & 2.508 & 5.054 & 1.444 \\
h_2[8] & 1.13 & 4.594 & 3.87 & 1.072 \\
h_2[12] & 1.752 & 1.812 & 1.15 & 0.418
\end{array} \; h_2[15].$$

Given TA's policy, except I index from 0.

$$H_2 = \begin{bmatrix} 0.424 \\ 1.674 \\ 1.894 \\ 1.35 \\ 0.96 \\ 2.508 \\ 5.054 \\ 1.444 \\ 1.13 \\ 4.594 \\ 3.87 \\ 1.072 \\ 1.752 \\ 1.812 \\ 1.15 \\ 0.418 \end{bmatrix}$$

$$m = w_2^T H_2 + b_2 = 11.9786.$$

$$p(y=0) = \hat{y} = \sigma(m) = 0.999.$$

Q2.

$p(y=0)$ means the input H_2 offered by the generation network. To maximiz the discriminator. the loss function is:

$$\mathcal{L}(\theta^D) - \log p(y=0 \mid \theta^D)$$

Negative log likelihood because we want to use gradient descent.

2.1 Gradient of prediction : $\nabla \hat{y}$.

$$\nabla \hat{y} = \frac{\partial -\log \hat{y}}{\partial \hat{y}} = -\frac{1}{\hat{y}} =$$

2.2 Gradient of H_2 : ∇H_2.

$$\nabla_{H_2} = \frac{\partial m}{\partial H_2} \frac{\partial \hat{y}}{\partial m} \nabla \hat{y}$$

$$= w_2 \hat{y}(1-\hat{y}) \nabla \hat{y}.$$

2.3 Gradient of h_2: ∇h_2

$$\nabla h_2 = \begin{bmatrix} \nabla H_2[0] & \cdots & \nabla H_2[3] \\ \nabla H_2[4] & & \vdots \\ \vdots & & \vdots \\ \nabla H_2[12] & \cdots & \nabla H_2[15] \end{bmatrix}$$

2.4 Gradient of w_1: ∇w_1

$$H_2 = c_1^T h_1^1 + c_2^T h_1^2 + b_1 \Rightarrow h_1 = C(H_2 - b_1).$$

where C_1 is a 4×16 matrix, which is based on $w_1^1 = \begin{bmatrix} 0,0 & 0,1 & 0,2 \\ 1,0 & 1,1 & 1,2 \\ 2,0 & 2,1 & 2,2 \end{bmatrix}$

$$C_1 = \left[\begin{array}{cccccccccccccccc}
w_1^1[0,0] & w_1^1[0,1] & w_1^1[0,2] & 0 & w_1^1[1,0] & w_1^1[1,1] & w_1^1[1,2] & 0 & w_1^1[2,0] & w_1^1[2,1] & w_1^1[2,2] & 0 & 0 & 0 & 0 & 0 \\
0 & w_1^1[0,0] & w_1^1[0,1] & w_1^1[0,2] & 0 & w_1^1[1,0] & w_1^1[1,1] & w_1^1[1,2] & 0 & w_1^1[2,0] & w_1^1[2,1] & w_1^1[2,2] & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & w_1^1[0,0] & w_1^1[0,1] & w_1^1[0,2] & 0 & w_1^1[1,0] & w_1^1[1,1] & w_1^1[1,2] & 0 & w_1^1[2,0] & w_1^1[2,1] & w_1^1[2,2] & 0 \\
0 & 0 & 0 & 0 & 0 & w_1^1[0,0] & w_1^1[0,1] & w_1^1[0,2] & 0 & w_1^1[1,0] & w_1^1[1,1] & w_1^1[1,2] & 0 & w_1^1[2,0] & w_1^1[2,1] & w_1^1[2,2]
\end{array}\right]$$

Image H_2

fliter $C[1][\,]$

$$\begin{array}{|c|c|c|c|} \hline H_0 & H_1 & H_2 & H_3 \\ \hline H_4 & H_5 & H_6 & H_7 \\ \hline H_8 & H_9 & H_{10} & H_{11} \\ \hline H_{12} & H_{13} & H_{14} & H_{15} \\ \hline \end{array} \otimes \begin{array}{|c|c|c|c|} \hline W_{00} & W_{01} & W_{02} & 0 \\ \hline W_{10} & W_{11} & W_{12} & 0 \\ \hline W_{20} & W_{21} & W_{22} & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline \end{array}$$

We can got C_2 In a similar way.

$$W_{00}\times H_0 + W_{01}\times H_1 + W_{02}H_2 + 0H_3$$

$$+ W_{10}H_4 + W_{11}H_5 + W_{12}H_6 + 0\cdot H_7$$

$$+ W_{20}H_8 + W_{21}H_9 + H_{22}H_{10} + 0\cdot H_{11}$$

$$+ 0\cdot H_{12} + 0\cdot H_{13} + 0\cdot H_{14} + 0\cdot H_{15}$$

$$\nabla C_1 = \frac{\partial H_2}{\partial C_1} \nabla H_2 .$$

$$= \sum_{i=1}^{16} \frac{\partial H_2[i]}{\partial C_1} \nabla H_2[i] .$$

$$= \sum_{i=1}^{16} \left[\frac{\partial H_2[i]}{\partial C_1[1]}, \ldots, \frac{\partial H_2[i]}{\partial C_1[16]} \right] \cdot \nabla H_2[i] .$$

whe. $\partial C_1[j]$ is the column of C_1

$$\nabla W_1'[0,0] = \nabla C_1[0,0] + \nabla C_1[1,1] + \nabla C_1[2,4] + \nabla C_1[3,5]$$

$$\nabla W_1'[0,1] = \nabla C_1[0,1] + \nabla C_1[1,2] + \nabla C_1[2,5] + \nabla C_1[3,6]$$

$$\nabla W_1'[0,2] = \nabla C_1[0,2] + \nabla C_1[1,3] + \nabla C_1[2,6] + \nabla C_1[3,7] .$$

$$\nabla W_1'[1,0] = \nabla C_1[0,4] + \nabla C_1[1,5] + \nabla C_1[2,8] + \nabla C_1[3,9]$$

$$\nabla W_1'[1,1] = \nabla C_1[0,5] + \nabla C_1[1,6] + \nabla C_1[2,9] + \nabla C_1[3,10]$$

$$\nabla W_1'[1,2] = \nabla C_1[0,6] + \nabla C_1[1,7] + \nabla C_1[2,10] + \nabla C_1[3,11]$$

$$\nabla W_1'[2,0] = \nabla C_1[0,8] + \nabla C_1[1,9] + \nabla C_1[2,12] + \nabla C_1[3,13]$$

$$\nabla W_1'[2,1] = \nabla C_1[0,9] + \nabla C_1[1,10] + \nabla C_1[2,13] + \nabla C_1[3,14]$$

$$\nabla W_1'[2,2] = \nabla C_1[0,10] + \nabla C_1[1,11] + \nabla C_1[2,14] + \nabla C_1[3,15]$$

This is the gradient matrix of. w_1^1 : ∇w_1^1

we can also use the same method to get: ∇w_1^2

so $\nabla w_1 = [\nabla w_1^1, \nabla w_1^2]$.

2.5 Gradient of h_1

$$H_2 = C_1^T h_1^2 + C_2^T h_1^1 + b_1$$

$$\nabla h_1^1 = \frac{\partial H_2}{\partial h_1^1} \nabla H_2.$$

$$\nabla h_1^2 = \frac{\partial H_2}{\partial h_1^2} \nabla H_2$$

2.6 Gradient of h_0.

$$h_1 = \text{reshape}(h_0) :$$

$$\nabla h_0 = \begin{bmatrix} \nabla h_1^1 \\ \nabla h_1^2 \end{bmatrix}$$

2.7. Gradient of w_0

$$h_0 = \text{Relu}(w_0 z_1 + b_0)$$

$$\nabla w_0 = \underset{8\times 2.}{\overset{8\times 1}{\frac{\partial h_0}{\partial w_0}}} \nabla h_0 .$$

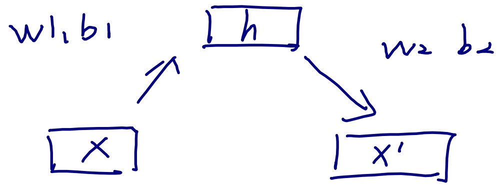
$$= \left[\frac{\partial h_0[1]}{\partial w_0}, \ldots, \frac{\partial h_0[8]}{\partial w_0} \right]^{1\times 8.} \nabla h_0 .$$

Each $\frac{\partial h_0[i]}{\partial w_0}$ is a 8×2 matrix.

$$= \begin{bmatrix} \boxed{\frac{\partial h_0[i]}{\partial w_0[1][1]}, \frac{\partial h_0[i]}{\partial w_0[1][2]}} \\ \vdots \\ \frac{\partial h_0[i]}{\partial w_0[8][1]} \quad \frac{\partial h_0[i]}{\partial w_0[8][2]} \end{bmatrix} 8\times 2 .$$

For each. $\frac{\partial h_0[i]}{\partial w_0[j][\,]} = \begin{cases} z^T, & i = j \ \& \ h_0[i] > 0. \\ 0, & \text{otherwise.} \end{cases}$

Q2. Auto Encoder.



$x \in R^{N\times 1}$

$w_1 \in R^{N\times M}$.

$b_1 \in R^{M\times 1}$

$h \in R^{M\times 1}$

$w_2 \in R^{M\times N}$

$b_2 \in R^{N\times 1}$

$x' \in R^{N\times 1}$.

$$\begin{cases} h = \mathrm{Relu}(w_1^T x + b_1) \\ x' = \sigma(w_2^T h + b_2) \\ w^1, w^2 = \arg\min\limits_{w^1, w^2} \sum\limits_{j=1}^{N} (x_i - x_i')^2. \end{cases}$$

Derive the gradient for w_1 & w_2.

2.1 Gradient for x': $\nabla x'$

$$\nabla x' = \begin{bmatrix} 2(x[1] - x'[1]) \\ \vdots \\ 2(x[N] - x'[N]) \end{bmatrix}$$

2.2 Gradient for W_2: ∇W_2.

$$z_2 = w_2^T h + b_2$$

$$x' = \sigma(z_2)$$

$$\nabla W_2 = \frac{\partial z_2}{\partial W_2} \frac{\partial x'}{\partial z_2} \nabla x'$$

$$= \frac{\partial z_2}{\partial w_2} \begin{bmatrix} x'[1](1-x'[1]) \\ \vdots \\ x'[i](1-x'[i]) \\ \vdots \\ x'[N](1-x'[N]) \end{bmatrix} \nabla x'$$

where $\frac{\partial z_2}{\partial W_2} = \left[\frac{\partial z_2[1]}{\partial W_2}, \ldots, \frac{\partial z_2[N]}{\partial W_2} \right]^{1\times N}.$

Each. $\frac{\partial z_2[i]}{\partial W_2} = \begin{bmatrix} \frac{\partial z_2[i]}{\partial W_2[1][1]} & \cdots & \frac{\partial z_2[i]}{\partial W_2[1][N]} \\ \vdots & & \\ \frac{\partial z_2[i]}{\partial W_2[M][1]} & \cdots & \frac{\partial z_2[i]}{\partial W_2[M][N]} \end{bmatrix}$ $M\times N.$

Each. $\frac{\partial z_2[i]}{\partial W_2[\,][j]} = \begin{cases} h & \text{if } i=j \\ 0, & \text{otherwise.} \end{cases}$

2.3 Gradient of h: ∇h.

$$\nabla h = \frac{\partial z_2}{\partial h} \cdot \frac{\partial x'}{\partial z_2} \nabla x'$$

$$= W_2 \cdot x'(1-x') \nabla x'$$

2.4 Gradient of w_1:

$$h = \mathrm{ReLU}(w_1^T x + b_1).$$

$$\nabla W_1 = \frac{\overset{M\times 1}{\partial h}}{\underset{N\times M}{\partial W_1}} \nabla h$$

$$= \left[\frac{\partial h[1]}{\partial w_1}, \ldots, \frac{\partial h[M]}{\partial w_1} \right] \nabla h.$$

Each $\frac{\partial h[i]}{\partial w_1} = \begin{array}{c} \\ \begin{bmatrix} \boxed{\begin{matrix} \frac{\partial h[i]}{\partial w_1[1][1]} \\ \vdots \\ \frac{\partial h[i]}{\partial w_1[N][1]} \end{matrix}} & \begin{matrix} \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} \frac{\partial h[i]}{\partial w_1[1][M]} \\ \\ \frac{\partial h[i]}{\partial w_1[N][M]} \end{matrix} \end{bmatrix} \end{array}^{N\times M}$

each $\frac{\partial h[i]}{\partial w_1[\,][j]} = \begin{cases} x, & \text{if } i=j \ \& \ h[i]>0. \\ 0. & \text{otherwise.} \end{cases}$