Midterm - CS 687

Prithul Sarker (800138885б)

3/29/2021

1.

(a) True

(b) False

(c) False

(d) True

(e) True

3. (a) $\sigma(x) = \frac{1}{1+e^{-x}}$

$$\sigma'(x) = \lim_{h\to 0} \frac{\sigma(x+h)-\sigma(x)}{h}$$

$$= \lim_{h\to 0} \frac{\frac{1}{1+e^{-x-h}} - \frac{1}{1+e^{-x}}}{h}$$

$$= \lim_{h\to 0} \frac{1}{h} \cdot \frac{e^{-x} - e^{-x-h}}{(1+e^{-x})(1+e^{-x-h})}$$

$$= \lim_{h\to 0} \frac{1}{h} \cdot \frac{e^{-x} - e^{-x} \cdot e^{-h}}{(1+e^{-x})(1+e^{-x-h})}$$

$$= \lim_{h\to 0} \frac{1}{h} \cdot \frac{e^{-x}(1-e^{-h})}{(1+e^{-x})(1+e^{-x-h})}$$

$$= \lim_{h\to 0} \frac{1-e^{-h}}{h} \cdot \lim_{h \to 0} \frac{e^{-x}}{(1+e^{-x})(1+e^{-x-h})}$$

$$= 1 \cdot \frac{1+e^{-x}-1}{(1+e^{-x})^2}$$

$$= \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2}$$

$$= \frac{1}{1+e^{-x}}\left[1 - \frac{1}{1+e^{-x}}\right]$$

$$= \sigma(x)\,(1-\sigma(x))$$

So, $\sigma'(x) = \sigma(x)(1-\sigma(x))$

(d) $\tanh(x) = \dfrac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$

$$\tanh'(x) = \frac{d}{dx}\left[\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}\right]$$

$$= \frac{(e^{x}+e^{-x})\frac{d}{dx}(e^{x}-e^{-x}) - (e^{x}-e^{-x})\frac{d}{dx}(e^{x}+e^{-x})}{(e^{x}+e^{-x})^2}$$

$$= \frac{(e^x+e^{-x})(e^x+e^{-x}) - (e^x-e^{-x})(e^x-e^{-x})}{(e^x+e^{-x})^2}$$

$$= \frac{(e^x+e^{-x})^2 - (e^x-e^{-x})^2}{(e^x+e^{-x})^2}$$

$$= 1 - \left(\frac{e^x-e^{-x}}{e^x+e^{-x}}\right)^2$$

$$= 1 - \tanh^2(x)$$

So, $\tanh'(x) = 1 - \tanh^2(x)$

4.

(a) $y_1 = \mathrm{ReLU}(w_1 \cdot x + b_1) = \mathrm{ReLU}(\hat{w}_1 \cdot \hat{x})$

$y = \mathrm{ReLU}(w_2 \cdot y_1 + b_2) = \mathrm{ReLU}(\hat{w}_2 \cdot \hat{y}_1)$

[where $\hat{y}_1 = \begin{bmatrix} y_1 \\ 1 \end{bmatrix}$]

$$\hat{w}_1 = \left[w_1 \mid b_1 \right]$$

$$= \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & -2 & -1 & 2 \end{bmatrix}$$

size of $\hat{w}_1 = (2 \times 4)$

$$\hat{x} = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ 1 \end{bmatrix}$$

size of $\hat{x} = (4 \times 1)$

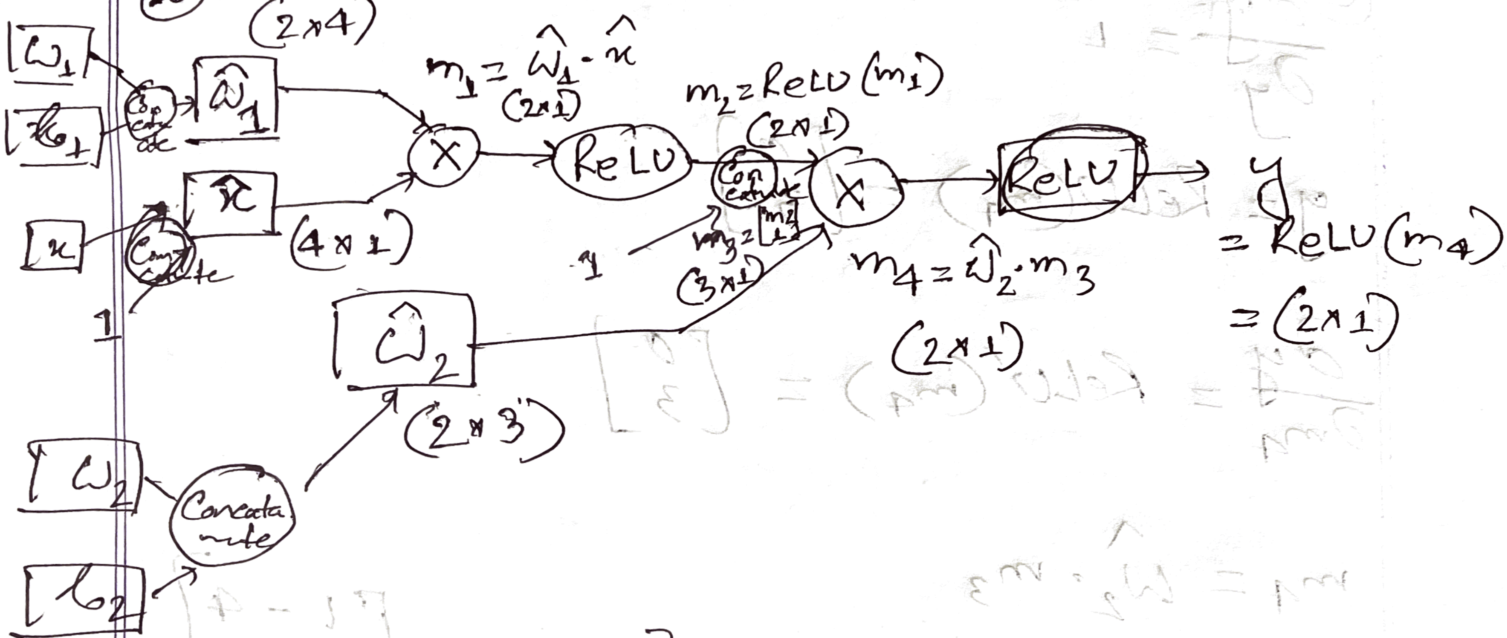
$$\hat{w}_2 = \left[w_2 \mid b_2 \right]$$

$$= \begin{bmatrix} 1 & -1 & 1 \\ -4 & 2 & -1 \end{bmatrix}$$

size of $\hat{w}_2 = (2 \times 3)$

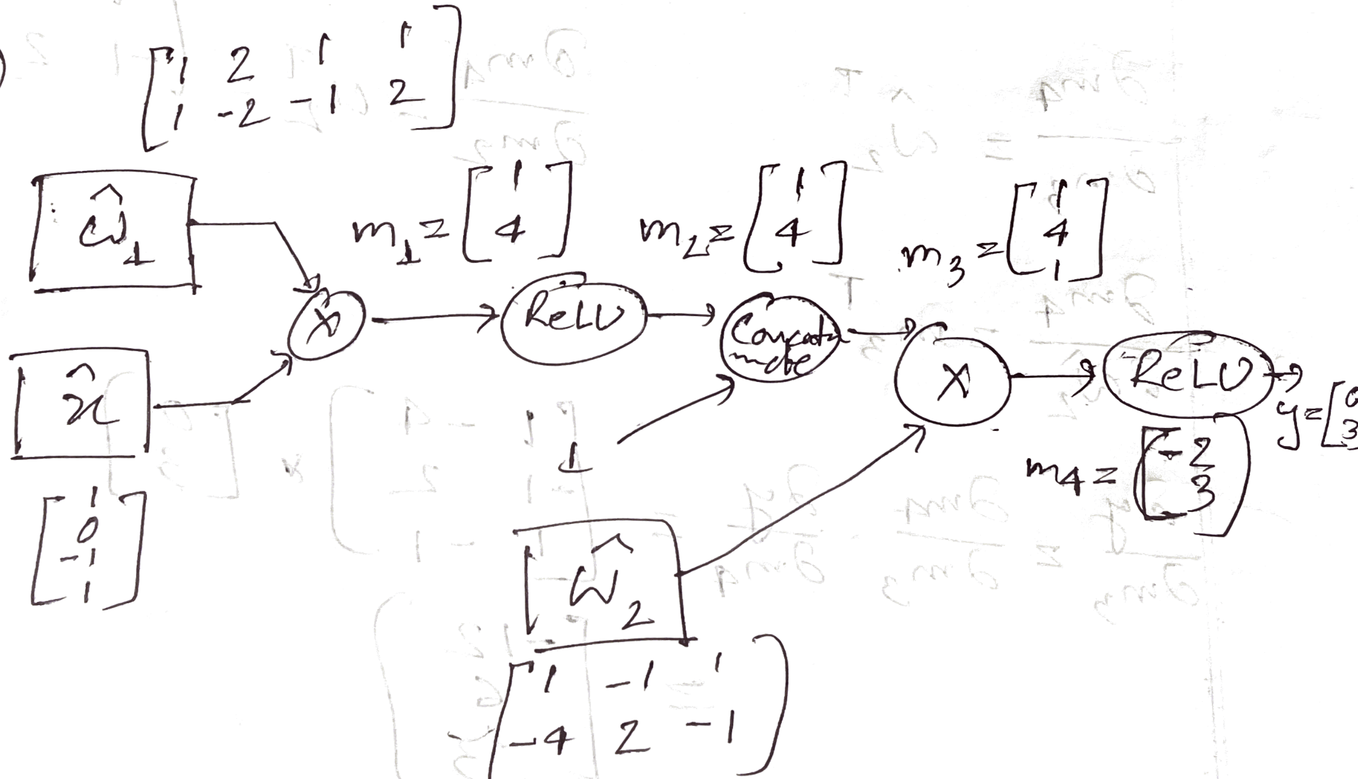
(c)

(b)



(2×4)
w_1
b_1
Concatenate
$\hat{w}_1$
x
Concatenate
$\hat{x}$
(4×1)
1
$m_1 = \hat{w}_1 \cdot \hat{x}$
(2×1)
×
ReLU
$m_2 = \mathrm{ReLU}(m_1)$
(2×1)
Concatenate
1
$m_3 = \begin{bmatrix} m_2 \\ 1 \end{bmatrix}$
(3×1)
×
$m_4 = \hat{w}_2 \cdot m_3$
(2×1)
ReLU
y
$= \mathrm{ReLU}(m_4)$
$= (2\times1)$
$\hat{w}_2$
(2×3)
w_2
b_2
Concatenate

(d)



$\begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & -2 & -1 & 2 \end{bmatrix}$
$\hat{w}_1$
$\hat{x}$
$\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$
×
$m_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$
ReLU
$m_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$
Concatenate
1
$m_3 = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}$
×
ReLU
$y = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$
$m_4 = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$
$\hat{w}_2$
$\begin{bmatrix} 1 & -1 & 1 \\ -4 & 2 & -1 \end{bmatrix}$

(e) $\frac{\partial y}{\partial y} = 1$

$y = ReLU(m_4) = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$

$\frac{\partial y}{\partial m_4} = ReLU'(m_4) = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$

$m_4 = \hat{\omega}_2 \cdot m_3$

$\frac{\partial m_4}{\partial m_3} = \hat{\omega}_2^T$ $\qquad \frac{\partial m_4}{\partial m_2} = \omega_2^T = \begin{bmatrix} 1 & -4 \\ -1 & 2 \end{bmatrix}$

$\frac{\partial m_4}{\partial \hat{\omega}_2} = m_3^T$

$\frac{\partial y}{\partial m_3} = \frac{\partial m_4}{\partial m_3} \cdot \frac{\partial y}{\partial m_4} = \begin{bmatrix} 1 & -4 \\ -1 & 2 \\ 1 & -1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 3 \end{bmatrix}$

$= \begin{bmatrix} -12 \\ 6 \\ -3 \end{bmatrix}$

$$\cancel{m_3 = \begin{bmatrix} m_2 \\ 1 \end{bmatrix}}$$

$$\cancel{\frac{\partial m_3}{\partial m_2} = \begin{bmatrix} -12 \\ 6 \end{bmatrix} \begin{bmatrix} 1 \\ \end{bmatrix}}$$

$$\frac{\partial y}{\partial m_2} = \frac{\partial y}{\partial m_4} \cdot \frac{\partial m_4}{\partial m_2} = \cancel{\begin{bmatrix} 0 \\ 3 \end{bmatrix}} \times \begin{bmatrix} 1 & -4 \\ -1 & 2 \end{bmatrix} \times \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} -12 \\ 6 \end{bmatrix}$$

$$m_2 = \mathrm{ReLU}(m_1) = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$$\frac{\partial m_2}{\partial m_1} = \mathrm{ReLU}'(m_1) = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

elementwise

$$\frac{\partial y}{\partial m_1} = \frac{\partial m_2}{\partial m_1} \times \frac{\partial y}{\partial m_2} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \times \begin{bmatrix} -12 \\ 6 \end{bmatrix} = \begin{bmatrix} -12 \\ 24 \end{bmatrix}$$

$$m_1 = \hat{w}_1 \cdot \hat{x}$$

$$\frac{\partial m_1}{\partial \hat{x}} = \hat{w}_1^T = \begin{bmatrix} 1 & 1 \\ 2 & -2 \\ 1 & -1 \\ 1 & 2 \end{bmatrix}$$

$$\frac{\partial m_1}{\partial \hat{w}_1} = \hat{x}^T = \begin{bmatrix} 1 & 0 & -1 & 1 \end{bmatrix}$$

$$\frac{\partial m_1}{\partial x} = w_1^T = \begin{bmatrix} 1 & 1 \\ 2 & -2 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial m_1}{\partial w_1} = x^T = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$$

$$\nabla_{w_1} y = \frac{\partial y}{\partial w_1} = \frac{\partial y}{\partial m_1} \cdot \frac{\partial m_1}{\partial w_1} = \begin{bmatrix} -12 \\ 24 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} -12 & 0 & 12 \\ 24 & 0 & -24 \end{bmatrix}$$

$$\nabla_{x} y = \frac{\partial y}{\partial x} = \frac{\partial m_1}{\partial x} \cdot \frac{\partial y}{\partial m_1} = \begin{bmatrix} 12 \\ -72 \\ -36 \end{bmatrix}$$

$$\nabla_{\hat{w}_1} y = \frac{\partial y}{\partial \hat{w}_1} = \frac{\partial y}{\partial m_1} \cdot \frac{\partial m_1}{\partial \hat{w}_1} = \begin{bmatrix} -12 & 0 & 12 & -12 \\ 24 & 0 & -24 & 24 \end{bmatrix}$$

$$\nabla_{\hat{x}} y = \frac{\partial y}{\partial \hat{x}} = \frac{\partial m_1}{\partial \hat{x}} \cdot \frac{\partial y}{\partial m_1} = \begin{bmatrix} 12 \\ -72 \\ -36 \\ 36 \end{bmatrix}$$

$$\nabla_{\hat{w}_2} y = \frac{\partial y}{\partial \hat{w}_2} = \frac{\partial y}{\partial m_4} \cdot m_3^T = \begin{bmatrix} 0 \\ 3 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 3 & 12 & 3 \end{bmatrix}$$

$$\nabla_{w_2} y = \frac{\partial y}{\partial w_2} = \begin{bmatrix} 0 & 0 \\ 3 & 12 \end{bmatrix}$$

5.

Advan

5. (b)

Advantages:

i) CNN would need less computation w.r.to fully connected layers

ii) CNN has less trainable parameters than fully connected layers

iii) CNN value only depends on only its receptive field. where as in fully connected layers, it depends on all neurons of previous layers.

8. (a)

~~As~~ When a filter is used, the dimension reduces by its size but you also skip a single value. So, output = Input - Kernel size + 1

when ~~a~~ padding is used, same value is applied to its ~~sig~~ all sides. For row, it gets both side padding. So twice the value of P is used. So, output = input + 2×padding

when a stride is used, you skip the input by stride value. So it is divided.

output = $\frac{\text{Input}}{\text{stride}}$.

and finally,

combining all of

$$\text{output} = \frac{\text{Input} + 2\times \text{padding} - \text{Kernelsize}}{\text{stride}} + 1$$

$$M = \frac{N+2P-F}{S} + 1$$

8. b

So, input and output has to be same size.

$$N = \frac{N+2P-F}{S} + 1$$

$$\Rightarrow (N-1)S = N + 2P - F$$

$$\Rightarrow 2P = (N-1)S + F - N$$

$$= NS - S + F - N$$

$$= N(S-1) + F - S$$

$$\text{So, } P = \frac{(S-1)N + F - S}{2}$$

2.

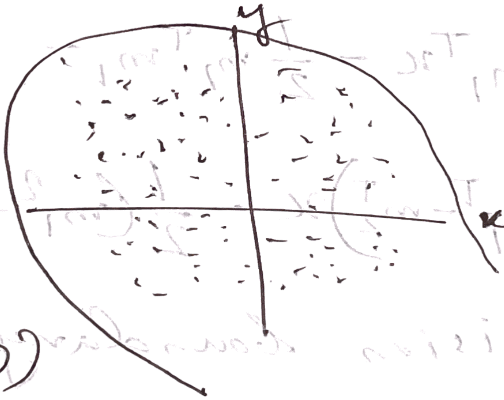
If y is the output and $x = \{x_1, x_2 \ldots x_n\}$ is the input

$$y = f(w_1 x_1 + w_2 x_2 + \ldots \ldots + w_n x_n + b)$$

where $w_1, w_2, \ldots w_n$ are the weights and b is the bias. f is the activation function.

Now, since the output is 1, the whole feature space is the decision boundary.

For example.



Decision boundary ~~it~~ can't be a circle because you need two layers for that.

In the figure, all input (2D) values are featured in xy plaice. So the non-linear line is the decision boundary

7. Since ~~is~~ the classes are linearly separable, the eqn might be

$$y = w \cdot x + b$$

~~y is~~ size of y is $(2,1)$ since it has ~~two~~ 2 outputs.

size of w is $(2, N)$

x is $(N, 1)$

b is $(2, 1)$

$$d_{12}(x) = m_1^T x - \frac{1}{2} m_1^T m_1 - m_2^T x + \frac{1}{2} m_2^T$$

$$= (m_1^T - m_2^T) x - \frac{1}{2}(m_1^2 - m_2^2)$$

The ~~d~~ecision boundary is the divider between two classes.

$w = m_1^T - m_2^T$

$b = -\frac{1}{2}(m_1^2 - m_2^2)$

and $y = w \cdot x + b$

no non-linearity is needed

6. (a)

$$504 = \frac{512 - F}{1} + 1$$

$$503 = 512 - F$$

$$F = 9$$

So, answer is $9 \times 9 \times 12$

(b) $25\cancel{0}2 \times 25\cancel{0}2 \times \cancel{3}12$ [since $\frac{504 - 2}{2} + 1$
$= 252$]

(c) Depth is same as the input. in this case 12 for first layer.

(d) $$\frac{252 - 3}{1} + 1 = 250$$

$250 \times 250 \times 6$

(e) $125 \times 125 \times 6$

Unwrapped $= 93750 \times 1$