Name: Asadollah Hill Galib

CSE 891: Deep learning

Homework 1

Q-1

$$\omega^{(1)} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad b^{(1)} = [0 \quad 0 \quad 0]$$

$$\omega^{(2)} = [1 \quad 1 \quad 1], \quad b^{(2)} = [-2.0]$$

Intuition: $\omega^{(1)}$ is dealing with comparing ~~distinct~~ successive pair in each row:

$$\omega^{(1)} \begin{cases} \text{1st row: } & -x_1 + x_2 \\ \text{2nd row: } & -x_2 + x_3 \\ \text{3rd row: } & -x_3 + x_4 \end{cases}$$

$\Rightarrow$ We want all of them to be greater than 0: which will evaluate to $[1 \quad 1 \quad 1]$ after activation.

$\omega^{(2)}$ is taking sum of the output from the previous layer. We want this sum to be 3, which ~~in~~ indicates satifying all pairs $(x_1 < x_2,\ x_2 < x_3,\ x_3 < x_4)$.

$b^{(2)}$ is for checking whether the sum is 3 or lesser. If, the sum is 3, it will evaluate to $3 - 2 = 1 \Rightarrow \phi(1) = 1$. Otherwise, it will evaluate to 0 or lesser $\Rightarrow \phi(0) = 0$.

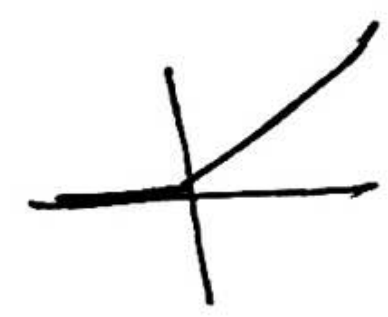
Q-2

$\frac{dL}{d\omega_1}$: Yes

$$\frac{dL}{d\omega_1} = y' \frac{dy}{d\omega_1} = y' \phi'(z)\, h_1 = 0 \quad [\text{as } h_1 = 0]$$

$\frac{dL}{d\omega_2}$: Yes

$$\frac{dL}{d\omega_2} = h_1' \frac{dh_1}{d\omega_2} = h_1' \phi'(z)\, h_3$$



$$= h_1' \phi'(-1)\, h_3$$

$$= 0 \quad [\text{as } Relu'(-1) = 0]$$

$\frac{dL}{d\omega_3}$: No

$$\frac{dL}{d\omega_3} = h_3' \frac{dh^3}{d\omega_3}$$

$$= (\omega_2' + \omega_4') \frac{dh^3}{d\omega_3}$$

It's not ~~gurantee~~ guaranteed to be 0, as it has two parts (ω_2' & ω_4') and we do not know anything about ω_4'.

So, $\frac{dL}{d\omega_3}$ can be 0 or other.

Q-3

1

$$\hat{f}_{\tau}(x) = \omega_1 a(\omega_0 x + b_0) + b_1$$

$$a(y) = \mathbb{I}(y \geq 0)$$

$$g(h,a,b,x) = h.\mathbb{I}(a \leq x \leq b)$$

Let, $n = 2$

$W_0 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $b_0 = \begin{bmatrix} -a \\ b \end{bmatrix}$ } to check ~~compare~~ < $x \geq a$ and $x \leq b$

$\omega_1 = \begin{bmatrix} h \\ h \end{bmatrix}$, $b_1 = -h$ } to get h when $a \leq x \leq b$ and 0 when ~~a~~ $x < a$ or $x > b$

Then, $\omega_0 x + b_0 = \begin{bmatrix} x - a \\ b - x \end{bmatrix}$

$\Rightarrow a(\omega_0 x + b_0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, if $\begin{cases} x \geq a \text{ and} \\ x \leq b \end{cases}$

$\Rightarrow \omega_1 a(\omega_0 x + b_0) + b_1 = h + h - h$
$= h = g(h,a,b,x)$ [when, $a \leq x \leq b$]

In any other cases, we will get
$a(\omega_0 x + b_0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ which will evaluate to $h - h = 0 = g(h,a,b,x)$ [when $x > b$ or $x < a$]

(*) Here, both $x < a$ and $x > b$ case is not possible, as $a \leq b$.

Q-3

(2)

$$f(x) = -x^2 + 1$$

$$\hat{f}_0(x) = 0$$

New function, $\hat{f}_1(x) = \hat{f}_0(x) + g(h_1, a_1, b_1, x)$

where, $a_1 = -1$, $b_1 = 1$ and

$h_1 = \frac{1}{2}\max(f(x))$ [$\max(f(x))$ can be found by setting $f'(x) = 0$:

$= \frac{1}{2} \cdot 1 = \frac{1}{2}$

$\Rightarrow -2x = 0 \Rightarrow x = 0$

$\Rightarrow f(0) = 1$]

We need to show,

$$\|f - \hat{f}_1\| \leq \|f - \hat{f}_0\|$$

$\|f - \hat{f}_0\|$

$= \int_{-1}^{1} |f(x) - 0|\, dx$

$= \int_{-1}^{1} |-x^2 + 1|\, dx$

$= \left[-\frac{x^3}{3} + x\right]_{-1}^{1}$

$= -\frac{2}{3} + 2$

$= \frac{4}{3}$

$\|f - \hat{f}_1\|$

$= \int_{-1}^{1} |f(x) - \hat{f}_1(x)|\, dx$

$= \int_{-1}^{1} |-x^2 + 1 - h_1|\, dx$ [$\hat{f}_1(x) = g(h_1, -1, 1, x) = h_1 \cdot \mathbb{I}(-1 \leq x \leq 1)$

$= \int_{-1}^{1} |-x^2 + 1 - \frac{1}{2}|\, dx$

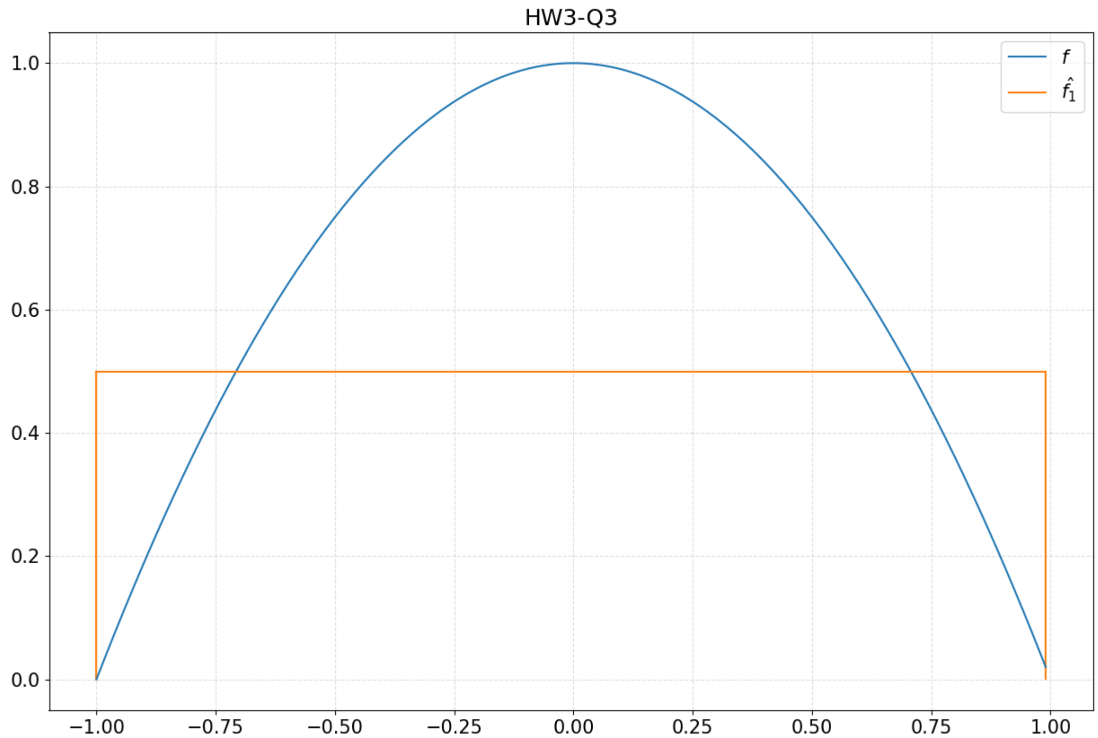
$= \left[-\frac{x^3}{3} + \frac{1}{2}x\right]_{-1}^{1}$

$= -\frac{2}{3} + 1$

$= \frac{1}{3} < \frac{4}{3} = \|f - \hat{f}_0\|$

So, $\|f - \hat{f}_1\| < \|f - \hat{f}_0\|$

Question 3(2): Plot



Q-3

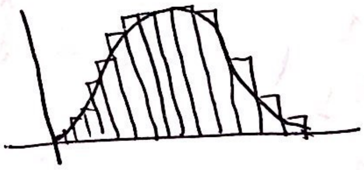
3

$\hat{f}_0(x) = 0$

$\hat{f}_{i+1}(x) = \hat{f}_i(x) + g(h_{i+1}, a_{i+1}, b_{i+1}, x)$ .. (1)

We have to construct a series of (1) with a fixed N, which satisfies ~~H fo~~ $\|f - \hat{f}_{i+1}\| \le \|f - \hat{f}_i\|$.

We can achieve this by incrementally adding N equal area parts' of the interval. The intuition is from the notion of integration, like we can get the area under a curve by splitting it into infinitesimal rectangles ~~pieces~~ and adding them up.



So, with a fixed n, we can generate new a_{i+1}, b_{i+1}, and h_{i+1}:

$$\left.\begin{aligned} a_{i+1} &= a_i + (\delta + \underbrace{0.00001}_{\text{to make sure that the splits are non-overlapping}}) \\ b_{i+1} &= a_i + \delta \end{aligned}\right\} \text{ where, } \delta = \frac{b-a}{N}$$

$$h_{i+1} = (f(a_{i+1}) + f(b_{i+1}))/2$$

Let's show that, $\|f - \hat{f}_{i+1}\| < \|f - \hat{f}_i\|$

$\|f - \hat{f}_i\|$

$\Rightarrow \int_{-1}^{1} |-x^2+1 - \hat{f}_{i-1} - g(h_i, a_i, b_i, x)\, dx$

$\Rightarrow \int_{-1}^{1} |-x^2+1 - \hat{f}_{i-1} - h_i|\, dx$

$\Rightarrow \left[-\frac{x^3}{3} + x - \hat{f}_{i-1} - h_i\right]_{-1}^{+1}$

$\Rightarrow \frac{4}{3} - \left[\hat{f}_{i-1} - h_i\right]_{-1}^{1} \ldots (2)$

$\|f - \hat{f}_{i+1}\|$

$= \int_{-1}^{1} |-x^2+1 - \hat{f}_i - g(h_{i+1}, a_{i+1}, b_{i+1}, x)|\, dx$

$= \int_{-1}^{1} |-x^2+1 - \hat{f}_{i-1} - h_i - h_{i+1}|\, dx$

$= \left[-\frac{x^3}{3} + x - \hat{f}_{i-1} - h_i - h_{i+1}\right]_{-1}^{1}$

$= \frac{4}{3} - \left[\hat{f}_{i-1} - h_i\right]_{-1}^{1} - \left[h_{i+1}\right]_{-1}^{1} \ldots (3)$

$< \frac{4}{3} - \left[\hat{f}_{i-1} - h_i\right]_{-1}^{1}$ [because, $\left[h_{i+1}\right]_{-1}^{1}$ evaluates to

$= \|f - \hat{f}_i\|$

positive : $h_{i+1} = \frac{(f(a_{i+1}) + f(b_{i+1}))}{2}$

~~both~~ both $f(a_{i+1})$ & $f(b_{i+1})$

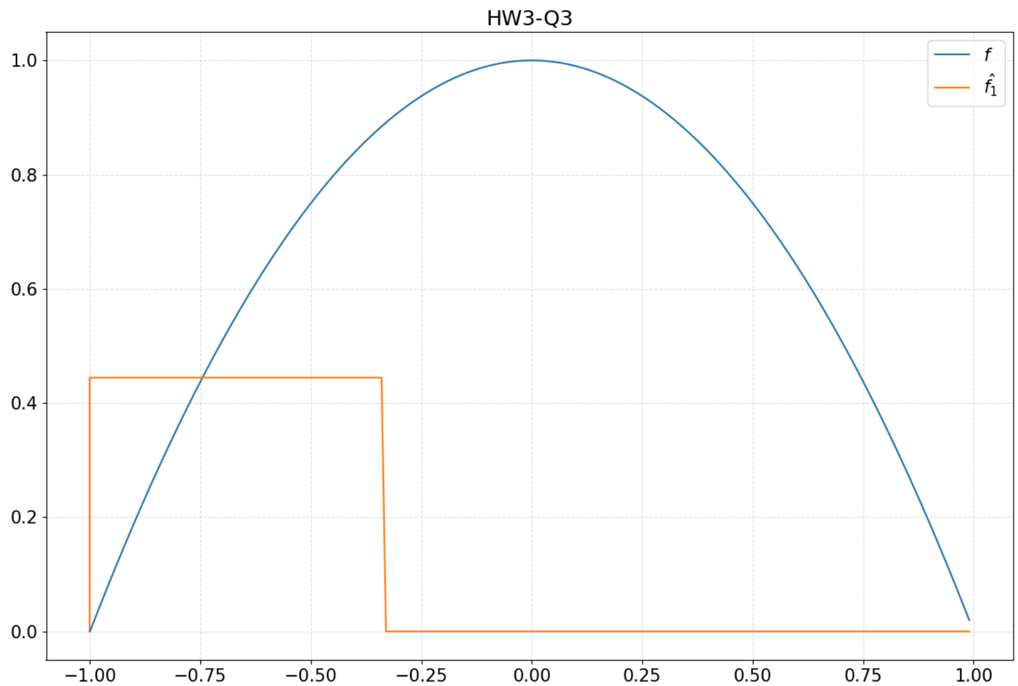
are ~~positive~~ greater or equal to 0.

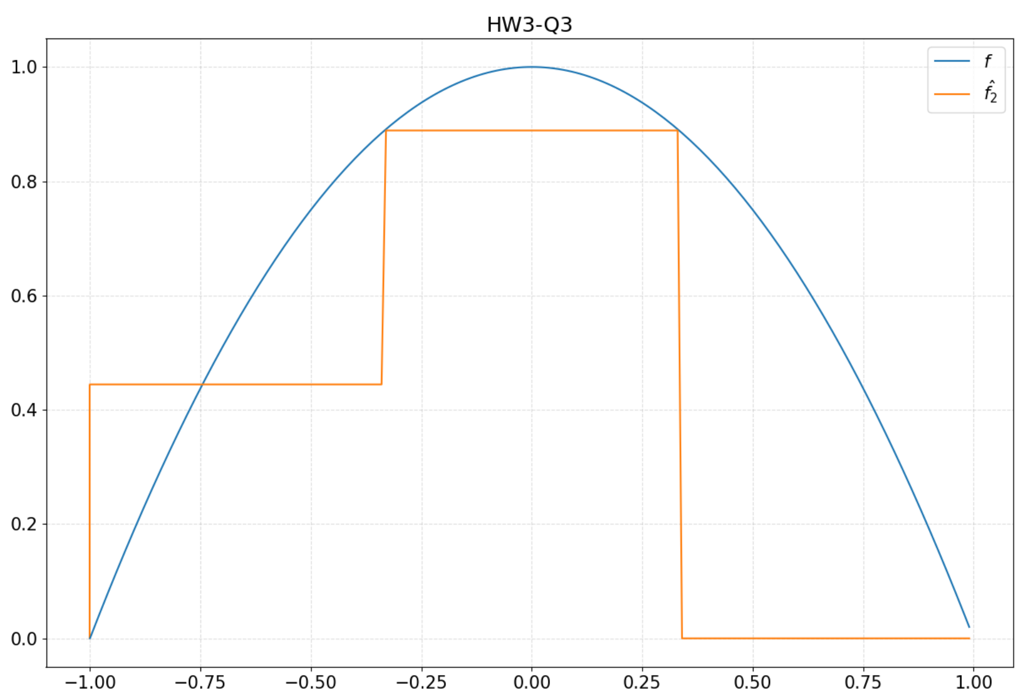
And, at least, one of them is greater than 0.

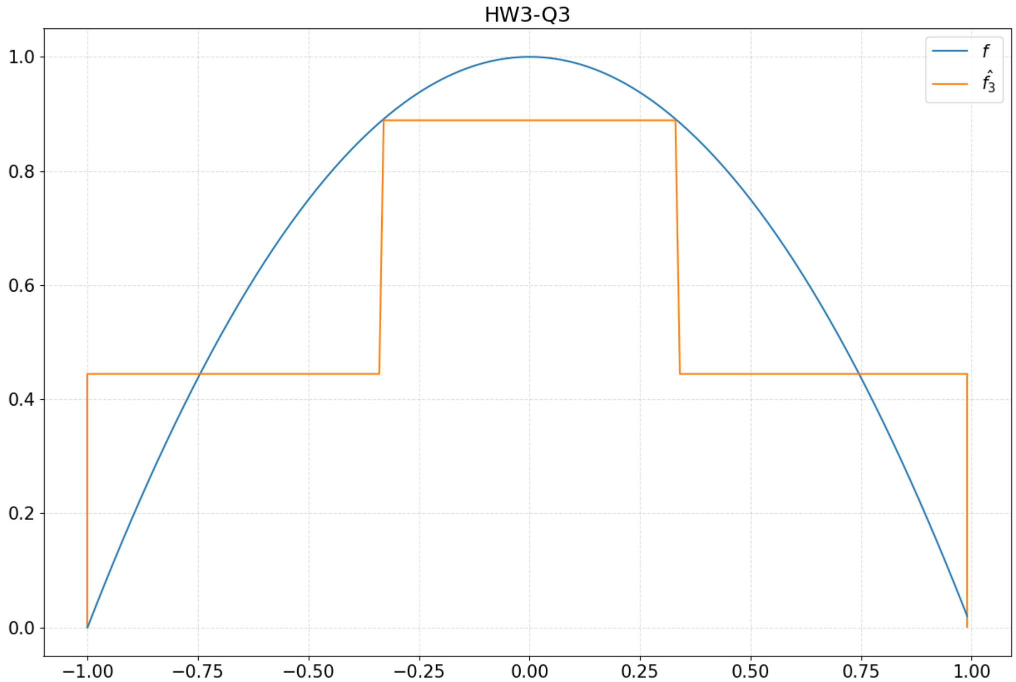
So, Equation 2 is greater than Equation 3

Scanned with CamScanner

Question 3(3): Plot







HW3-Q3
f
$\hat{f}_3$
1.0
0.8
0.6
0.4
0.2
0.0
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00