Homework 3

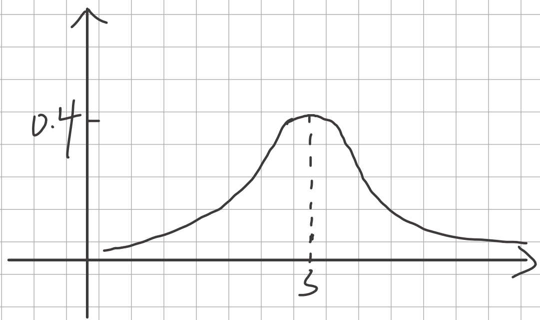
Yuanteng Chen

1. Normalization Layers.

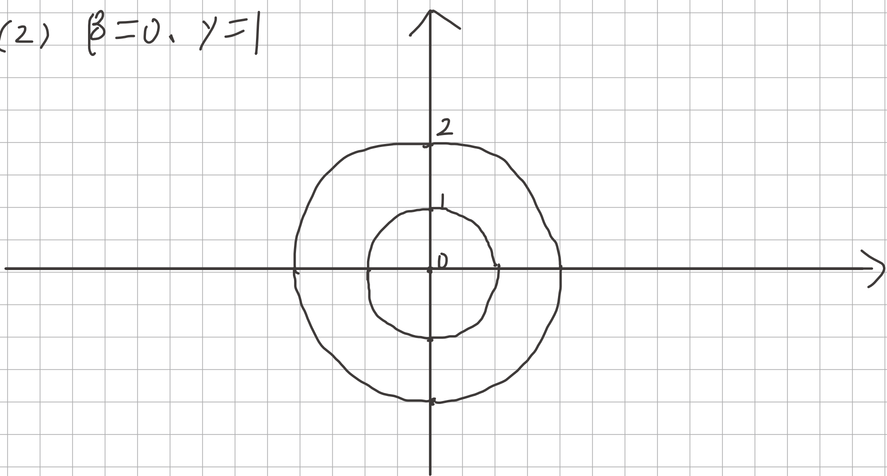
(1) $\beta = 3$, $\gamma = 1$

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$f(3) = \frac{1}{1 \cdot \sqrt{2\pi}} \simeq 0.4$$



(2) $\beta = 0$, $\gamma = 1$



2. Understanding Convolution as Finite Impulse Response Filter.

$$y(n) = x(n) * h(n) = \sum_{i=-\infty}^{\infty} x(n-i) \cdot h(i) = \sum_{i=-\infty}^{\infty} x(i) \cdot h(n-i)$$

$$x(n) = \begin{cases} 1, & n = 0, 1, 2, \cdots, L-1 \\ 0, & \text{otherwise} \end{cases}$$

(a). $$h(n) = (\tfrac{1}{2})^n u(n) = \begin{cases} (\tfrac{1}{2})^n, & n = 0, 1, 2 \\ 0, & \text{otherwise} \end{cases}$$

Compute and plot the convolution of $x(n)$ and $h(n)$.

Solution:

① $n<0$. $y(n) = \sum_{i=-\infty}^{\infty} x(i) \cdot h(n-i)$

as $x(n) = 1$, $n = 0, 1, 2, 3, \cdots L-1$

$\therefore$ $y(n) = \sum_{i=0}^{\infty} h(n-i) = \sum_{i=0}^{L-1} h(n-i)$

as $n<0$, $i = 0 \cdots L-1$, $\therefore$ $n-i<0$

$\therefore$ $h(n) = 0$. $\therefore$ $y(n) = \sum_{i=0}^{L-1} h(n-i) = 0$

② $0 \le n < L-1$

$y(n) = \sum_{i=-\infty}^{\infty} x(i) \cdot h(n-i)$

as $n-i<0$ $(n<i)$, $h(n-i) = 0$

$y(n) = \sum_{i=0}^{\infty} h(n-i) = \sum_{i=0}^{n} h(n-i) = \sum_{i=0}^{n} \frac{1}{2^i}$

③ $n \geq L-1$

$$y(n) = \sum_{i=-\infty}^{\infty} x(i) \cdot h(n-i)$$

$$= \sum_{i=0}^{\infty} h(n-i)$$

as $h(n) \,!= 0$ $(n \geq 0)$

$$\therefore y(n) = \sum_{i=0}^{L-1} h(n-i) = \sum_{i=0}^{L-1} \frac{1}{2^i}$$

$$\therefore y(n) = \begin{cases} 0, & n < 0 \\ \sum_{i=0}^{n} \frac{1}{2^i}, & 0 \leq n < L-1 \\ \sum_{i=0}^{L-1} \frac{1}{2^i}, & n \geq L-1 \end{cases}$$

(b) $x_2(n) = x(n-N)$ $N=5$.

$y_2(n) = h(n) * x_2(n)$

Sol:

$$y(n-5) = \begin{cases} 0, & n-5 < 0 \\ \sum_{i=0}^{n-5} \frac{1}{2^i}, & 0 \leq n-5 < L-1 \\ \sum_{i=0}^{L-1} \frac{1}{2^i}. & n \geq L-1 \end{cases} = y_2(n)$$

find: the root of shift invariance of convolution is time invariance of finite impulse response filter.

(c)

$$y(m,n) = x(m,n) \cdot h(m,n) = \sum_{i,j=-\infty}^{\infty} x(m-i, n-j) \cdot h(i,j)$$

$$= \sum_{i,j=-\infty}^{\infty} x(i,j) \cdot h(m-i, n-j)$$

Solution:

$$h = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \qquad h' = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

$$\begin{bmatrix} n-6 & n-5 & n-4 \\ n-1 & n & n+1 \\ n+4 & n+5 & n+6 \end{bmatrix} * h' = -2(6+5\times2+4)$$

$$= -40$$

$$\therefore \begin{bmatrix} -40 & -40 & -40 \\ -40 & -40 & -40 \\ -40 & -40 & -40 \end{bmatrix}$$

d: i. stride, pad = 1

ii. stride, pad = 2, 1

(i)

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 4 & 5 & 0 \\ 0 & 6 & 7 & 8 & 9 & 10 & 0 \\ 0 & 11 & 12 & 13 & 14 & 15 & 0 \\ 0 & 16 & 17 & 18 & 19 & 20 & 0 \\ 0 & 21 & 22 & 23 & 24 & 25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} * h' = \begin{bmatrix} -19 & -28 & -32 & -36 & -29 \\ -30 & -40 & -40 & -40 & -30 \\ -30 & -40 & -40 & -40 & -30 \\ -30 & -40 & -40 & -40 & -30 \\ 49 & 68 & 72 & 76 & 59 \end{bmatrix}$$

(ii) stride = 2

$$\begin{bmatrix} -19 & -32 & -29 \\ -30 & -40 & -30 \\ 49 & 72 & 59 \end{bmatrix}$$

3. Normalization

(a) ① A

② B

(b) $[X_1, X_2, \cdots, X_n]$. $\hat{X_i} = X_i - \mu$, $\mu = \frac{1}{n}\sum_{j=1}^{n} X_j$

$[y_1, y_2, \cdots, y_n]$. $y_i = \gamma \hat{X_i} + \beta$

have a final loss L somewhere downstream

Calculate $\frac{\partial L}{\partial x_i}$ in terms of $\frac{\partial L}{\partial y_j}$ for $j = 1, \cdots, n$

Sol: $\frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial y_i} \cdot \frac{\partial y_i}{\partial x_i}$ $\quad (y_i = \gamma (X_i - \frac{1}{n}\sum_{j=1}^{n} X_j) + \beta)$

$= \frac{\partial L}{\partial y_i} \cdot \gamma \cdot (1 - \frac{1}{n})$

$= \frac{n-1}{n} \cdot \gamma \cdot \frac{\partial L}{\partial y_i}$

$$\begin{cases} \text{when } n = 1 \rightarrow \frac{\partial L}{X_1} = 0 \\ \text{when } n \rightarrow \infty \rightarrow \frac{n-1}{n} \rightarrow 1 \quad \frac{\partial L}{\partial x_i} = \gamma \frac{\partial L}{\partial y_i} \end{cases}$$

4. Optimizers

(a): 1. Given $\eta = 0.001$ $\beta_1 = 0.9$ $\beta_2 = 0.999$

2. Initialize time step $t \leftarrow 0$, parameter

$\theta_{t=0} \in R^n$, $m_{t=0} \leftarrow 0$; $V_{t=0} \leftarrow 0$

3: Repeat:

4: $t \leftarrow t+1$

5: $g_t \leftarrow \nabla f_t(\theta_{t-1})$

6: $m_t \leftarrow (1-\beta_1) m_{t-1} + \beta_1 g_t$ (A)

7: $V_t \leftarrow (1-\beta_2) V_{t-1} + \beta_2 g_t^2$ (B)

8: $\theta_t \leftarrow \theta_{t-1} - \eta \cdot \frac{m_t}{\sqrt{V_t}}$

9: Until the stopping condition is met

(b) L_2: $f_t^{reg} = f_t(\theta) + \frac{\lambda}{2} \|\theta\|_2^2$

weight decay: $\theta_{t+1} = (1-\gamma)\theta_t - \eta \nabla f(\theta_t)$

($\gamma = 0 \Leftrightarrow$ regular SGD)

Sol: in f_t^{reg}

$\nabla f_t^{reg} = \nabla f_t(\theta) + \lambda \theta_t$

plug in SGD-

$$\theta_{t+1}^{reg} = \theta_t^{reg} - \eta \nabla f_t^{reg}$$

$$= \theta_t^{reg} - \eta \nabla f(\theta_t) - \eta \lambda \theta_t$$

$$= (1-\eta\lambda)\theta_t^{reg} - \eta \nabla f(\theta_t)$$

when $\eta\lambda = \gamma$, SGD with weight decay using the original loss $f_t(\theta)$ is equivalent to regular SGD on L_2-reg loss $f_t^{reg}(\theta)$

5. Homework Process and Study Group.

(a). CSDN, past scribe

(c) 3 hours