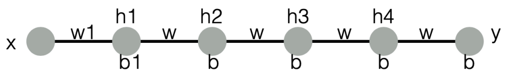
BS6207 Assignment 1 Jiao Yan

Question 1:



$h_1 = \sigma(z_1) = \sigma(w_1 x + b_1)$
$h_2 = \sigma(z_2) = \sigma(wh_1 + \mathrm{b})$
$h_3 = \sigma(z_3) = \sigma(wh_2 + \mathrm{b})$
$h_4 = \sigma(z_4) = \sigma(wh_3 + \mathrm{b})$

$\mathrm{y} = \sigma(z_y) = \sigma(wh_4 + \mathrm{b})$

$$\frac{\partial \mathrm{y}}{\partial w_1} = \frac{\partial \mathrm{y}}{\partial z_y}\cdot\frac{\partial z_y}{\partial h_4}\cdot\frac{\partial h_4}{\partial z_4}\cdot\frac{\partial z_4}{\partial h_3}\cdot\frac{\partial h_3}{\partial z_3}\cdot\frac{\partial z_3}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial w_1}$$

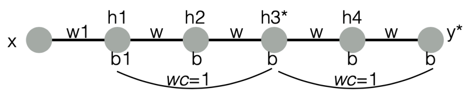
$$= \sigma'(z_y)\cdot w\cdot\sigma'(z_4)\cdot w\cdot\sigma'(z_3)\cdot w\cdot\sigma'(z_2)\cdot w\cdot\sigma'(z_1)\cdot x$$

$$= w^4\cdot\sigma'(z_y)\cdot\sigma'(z_4)\cdot\sigma'(z_3)\cdot\sigma'(z_2)\cdot\sigma'(z_1)\cdot x$$

$$\frac{\partial \mathrm{y}}{\partial b_1} = \frac{\partial \mathrm{y}}{\partial z_y}\cdot\frac{\partial z_y}{\partial h_4}\cdot\frac{\partial h_4}{\partial z_4}\cdot\frac{\partial z_4}{\partial h_3}\cdot\frac{\partial h_3}{\partial z_3}\cdot\frac{\partial z_3}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial b_1}$$

$$= \sigma'(z_y)\cdot w\cdot\sigma'(z_4)\cdot w\cdot\sigma'(z_3)\cdot w\cdot\sigma'(z_2)\cdot w\cdot\sigma'(z_1)\cdot 1$$

$$= w^4\cdot\sigma'(z_y)\cdot\sigma'(z_4)\cdot\sigma'(z_3)\cdot\sigma'(z_2)\cdot\sigma'(z_1)$$



$h_1 = \sigma(z_1) = \sigma(w_1 x + b_1)$ $\quad z_1 = w_1 x + b_1$
$h_2 = \sigma(z_2) = \sigma(wh_1 + \mathrm{b})$ $\quad z_2 = wh_1 + \mathrm{b}$
$h_3{}^* = \sigma(z_3{}^*) = \sigma(wh_2 + h_1 + \mathrm{b})$ $\quad z_3{}^* = wh_2 + h_1 + \mathrm{b}$
$h_4 = \sigma(z_4) = \sigma(wh_3{}^* + \mathrm{b})$ $\quad z_4 = wh_3{}^* + \mathrm{b}$
$\mathrm{y}^* = \sigma(z_{y^*}) = \sigma(wh_4 + h_3{}^* + \mathrm{b})$ $\quad z_{y^*} = wh_4 + h_3{}^* + \mathrm{b}$

$$\frac{\partial \mathrm{y}^*}{\partial w_1} = \frac{\partial \mathrm{y}^*}{\partial z_{y^*}}\cdot[\frac{\partial z_{y^*}}{\partial h_4}\cdot\frac{\partial h_4}{\partial z_4}\cdot\frac{\partial z_4}{\partial h_3{}^*}\cdot\frac{\partial h_3{}^*}{\partial z_3{}^*}\cdot(\frac{\partial z_3{}^*}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial w_1} + \frac{\partial z_3{}^*}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial w_1})$$

$$+ \frac{\partial z_{y^*}}{\partial h_3{}^*}\cdot\frac{\partial h_3{}^*}{\partial z_3{}^*}\cdot(\frac{\partial z_3{}^*}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial w_1} + \frac{\partial z_3{}^*}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial w_1})]$$

$$= \frac{\partial \mathrm{y}^*}{\partial z_{y^*}}\cdot(\frac{\partial z_{y^*}}{\partial h_4}\cdot\frac{\partial h_4}{\partial z_4}\cdot\frac{\partial z_4}{\partial h_3{}^*}\cdot\frac{\partial h_3{}^*}{\partial z_3{}^*} + \frac{\partial z_{y^*}}{\partial h_3{}^*}\cdot\frac{\partial h_3{}^*}{\partial z_3{}^*})\cdot(\frac{\partial z_3{}^*}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial w_1} + \frac{\partial z_3{}^*}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial w_1})$$

$$= \sigma'(z_{y^*})(w\cdot\sigma'(z_4)\cdot w\cdot\sigma'(z_3{}^*) + 1\cdot\sigma'(z_3{}^*))\cdot(w\cdot\sigma'(z_2)\cdot w\cdot\sigma'(z_1)\cdot x + 1\cdot\sigma'(z_1)\cdot x)$$

$$= w^4\cdot\sigma'(z_{y^*})\cdot\sigma'(z_4)\cdot\sigma'(z_3{}^*)\cdot\sigma'(z_2)\cdot\sigma'(z_1)\cdot x + w^2\cdot\sigma'(z_{y^*})\cdot\sigma'(z_4)\cdot\sigma'(z_3{}^*)\cdot\sigma'(z_1)\cdot x +$$

$$w^2\cdot\sigma'(z_{y^*})\cdot\sigma'(z_3{}^*)\cdot\sigma'(z_2)\cdot\sigma'(z_1)\cdot x + \sigma'(z_{y^*})\cdot\sigma'(z_3{}^*)\cdot\sigma'(z_1)\cdot x$$

$$\frac{\partial \mathrm{y}^*}{\partial b_1} = \frac{\partial \mathrm{y}^*}{\partial z_{y^*}}\cdot[\frac{\partial z_{y^*}}{\partial h_4}\cdot\frac{\partial h_4}{\partial z_4}\cdot\frac{\partial z_4}{\partial h_3{}^*}\cdot\frac{\partial h_3{}^*}{\partial z_3{}^*}\cdot(\frac{\partial z_3{}^*}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial b_1} + \frac{\partial z_3{}^*}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial b_1})$$

$$+ \frac{\partial z_{y^*}}{\partial h_3{}^*}\cdot\frac{\partial h_3{}^*}{\partial z_3{}^*}\cdot(\frac{\partial z_3{}^*}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial b_1} + \frac{\partial z_3{}^*}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial b_1})]$$

$$= \frac{\partial \mathrm{y}^*}{\partial z_{y^*}}\cdot(\frac{\partial z_{y^*}}{\partial h_4}\cdot\frac{\partial h_4}{\partial z_4}\cdot\frac{\partial z_4}{\partial h_3{}^*}\cdot\frac{\partial h_3{}^*}{\partial z_3{}^*} + \frac{\partial z_{y^*}}{\partial h_3{}^*}\cdot\frac{\partial h_3{}^*}{\partial z_3{}^*})\cdot(\frac{\partial z_3{}^*}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial b_1} + \frac{\partial z_3{}^*}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial b_1})$$

$$= \sigma'(z_{y^*})(w\cdot\sigma'(z_4)\cdot w\cdot\sigma'(z_3{}^*) + 1\cdot\sigma'(z_3{}^*))\cdot(w\cdot\sigma'(z_2)\cdot w\cdot\sigma'(z_1)\cdot 1 + 1\cdot\sigma'(z_1)\cdot 1)$$

$$= w^4\cdot\sigma'(z_{y^*})\cdot\sigma'(z_4)\cdot\sigma'(z_3{}^*)\cdot\sigma'(z_2)\cdot\sigma'(z_1) + w^2\cdot\sigma'(z_{y^*})\cdot\sigma'(z_4)\cdot\sigma'(z_3{}^*)\cdot\sigma'(z_1) +$$

$$w^2\cdot\sigma'(z_{y^*})\cdot\sigma'(z_3{}^*)\cdot\sigma'(z_2)\cdot\sigma'(z_1) + \sigma'(z_{y^*})\cdot\sigma'(z_3{}^*)\cdot\sigma'(z_1)$$

So far the derivation is as shown above.

Question 2:

1) Standard gradient descent:

$$Loss(x) = \begin{cases} -x+1 \\ x-1-h \\ -x+1+h \\ -0.3x+\frac{(1+2h)}{0.3} \end{cases} \quad \frac{\partial \text{Loss}}{\partial \text{x}} = \begin{cases} -1 & (x<1) \\ 1 & (1<x<1+h) \\ -1 & (1+h<x<1+2h) \\ -0.3 & (x>1+2h) \end{cases}$$

$$x_0 = 0$$

$$x_1 = x_0 - a \times \frac{\partial \text{Loss}}{\partial x_0} = 0 - 0.3 \times (-1) = 0.3$$

$$x_2 = x_1 - a \times \frac{\partial \text{Loss}}{\partial x_1} = 0.3 - 0.3 \times (-1) = 0.6$$

$$\left.\begin{aligned} x_3 &= x_2 - a \times \frac{\partial \text{Loss}}{\partial x_2} = 0.6 - 0.3 \times (-1) = 0.9 \\ x_4 &= x_3 - a \times \frac{\partial \text{Loss}}{\partial x_3} = 0.9 - 0.3 \times (-1) = 1.2 \end{aligned}\right\} \text{repeating}$$

$$x_5 = x_4 - a \times \frac{\partial \text{Loss}}{\partial x_4} = 1.2 - 0.3 \times (1) = 0.9$$

$$x_6 = x_5 - a \times \frac{\partial \text{Loss}}{\partial x_5} = 0.9 - 0.3 \times (-1) = 1.2$$

....

2) Adam Optimizer:

Gradient with Adam Optimizer:

Within 9999 iterations, if the x can escape the bump, which means x>(1+h), the height h is not large enough

After 9999 iterations, if the x cannot escape the bump, we believe the height h is too large to escape

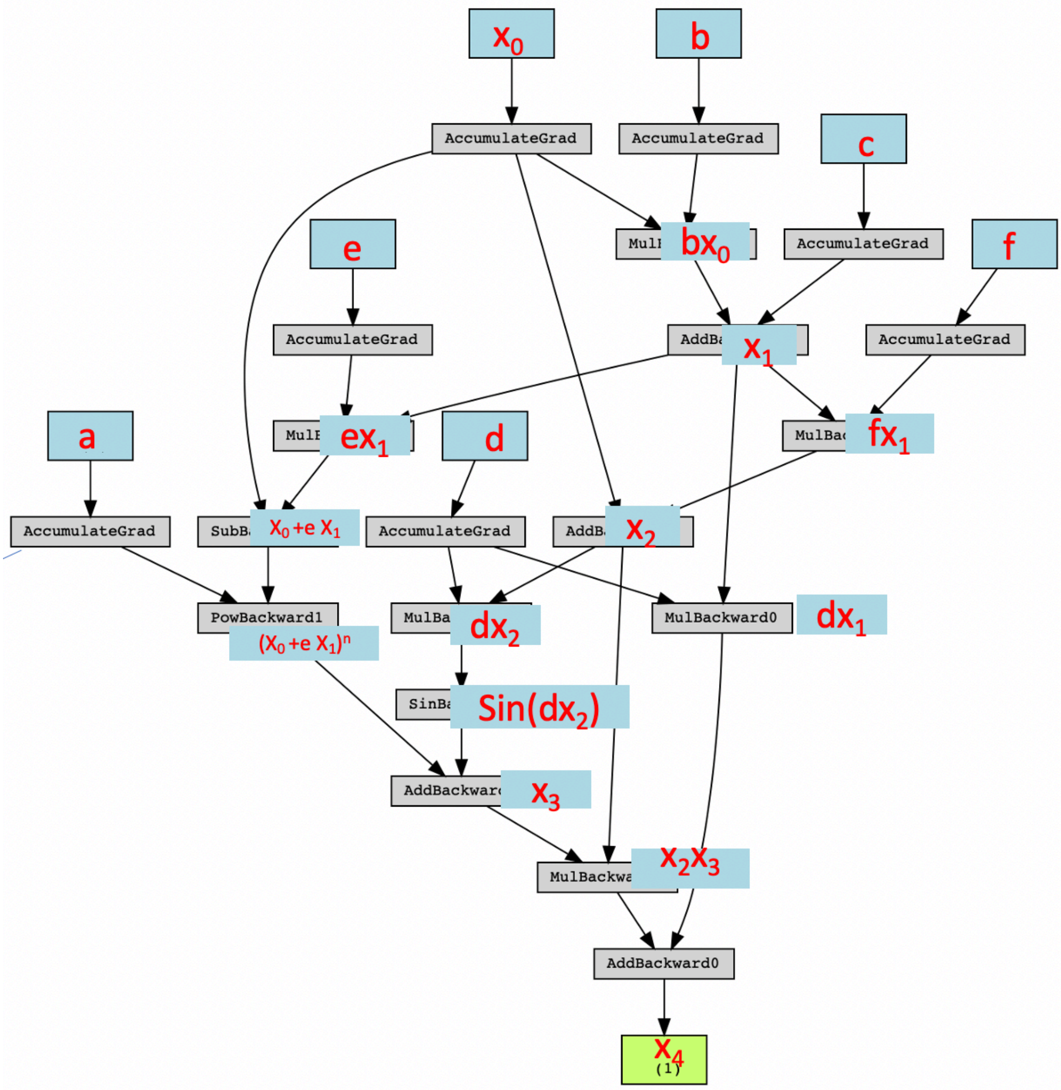
```python
def de_function(x, h):
    if x<1:
        return -1
    elif 1<x< (1+h):
        return 1
    elif (1+h)< x<(1+2*h):
        return -1
    else:
        return -0.3

def find_h( derivative, n_iter, alpha, beta1, beta2, eps=0):
    for h in np.arange(0.3,1,0.0001):
        x = 0
        m = 0.0
        v = 0.0
        for t in range(1,n_iter):
            g = de_function(x,h)
            m = beta1 * m + (1 - beta1) * g
            v = beta2 * v + (1 - beta2) * g**2
            mhat = m / (1.0 - beta1**t)
            vhat = v / (1.0 - beta2**t)
            x = x - alpha * mhat / (vhat**0.5 )
            if x>(1+h):
                break
        if x<(1+h):
            print("The maximum height of the bump h=", h)
            break

n_iter = 10000
alpha = 0.3
beta1 = 0.9
beta2 = 0.999
find_h( de_function, n_iter, alpha, beta1, beta2)
```

```
The maximum height of the bump h= 0.41019999999998785
```

Question 3:



x0
b
AccumulateGrad
AccumulateGrad
c
e
bx0
AccumulateGrad
f
AccumulateGrad
x1
AccumulateGrad
a
ex1
d
fx1
AccumulateGrad
x0+e x1
AccumulateGrad
x2
PowBackward1
(x0+e x1)n
dx2
MulBackward0
dx1
Sin(dx2)
x3
x2x3
AddBackward0
x4
(1)