17

(a).

$h_1 = \delta(x w_1 + b_1)$ $\quad z_1 = x w_1 + b_1$

$h_2 = \delta(h_1 w + b)$ $\quad z_2 = h_1 w + b$

$h_3 = \delta(h_2 w + b)$ $\quad z_3 = h_2 w + b$

$h_4 = \delta(h_3 w + b)$ $\quad z_4 = h_3 w + b$

$y = \delta(h_4 w + b)$ $\quad z_5 = h_4 w + b$

$$\frac{\partial y}{\partial w_1} = \frac{\partial y}{\partial z_5}\cdot\frac{\partial z_5}{\partial h_4}\cdot\frac{\partial h_4}{\partial z_4}\cdot\frac{\partial z_4}{\partial h_3}\cdot\frac{\partial h_3}{\partial z_3}\cdot\frac{\partial z_3}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial w_1}$$

$$= \delta'(z_5)\cdot w\cdot\delta'(z_4)\cdot w\cdot\delta'(z_3)\cdot w\cdot\delta'(z_2)\cdot w\cdot\delta'(z_1)\cdot x$$

$$= w^4\cdot\delta'(z_5)\delta'(z_4)\cdot\delta'(z_3)\cdot\delta'(z_2)\cdot\delta'(z_1)\cdot x$$

$$\frac{\partial y}{\partial b_1} = \frac{\partial y}{\partial z_5}\cdot\frac{\partial z_5}{\partial h_4}\cdot\frac{\partial h_4}{\partial z_4}\cdot\frac{\partial z_4}{\partial h_3}\cdot\frac{\partial h_3}{\partial z_3}\cdot\frac{\partial z_3}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial b_1}$$

$$= w^4\,\delta'(z_5)\cdot\delta'(z_4)\cdot\delta'(z_3)\cdot\delta'(z_2)\cdot\delta'(z_1)$$

(b).

$h_1 = \delta(w_1 x + b_1)$ $\quad z_1 = w_1 x + b_1$

$h_2 = \delta(w h_1 + b)$ $\quad z_2 = w h_1 + b$

$h_3^* = \delta(w h_2 + h_1 + b)$ $\quad z_3^* = w h_2 + h_1 + b$

$h_4 = \delta(w h_3^* + b)$ $\quad z_4 = w h_3^* + b$

$y^* = \delta(w h_4 + h_3^* + b)$ $\quad z_5^* = w h_4 + h_3^* + b$

$$\frac{\partial y^*}{\partial w_1} = \frac{\partial y^*}{\partial z_5^*}\cdot\left[\frac{\partial z_5^*}{\partial h_4}\cdot\frac{\partial h_4}{\partial z_4}\cdot\frac{\partial z_4}{\partial h_3^*}\cdot\frac{\partial h_3^*}{\partial z_3^*}\left(\frac{\partial z_3^*}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial w_1}+\frac{\partial z_3^*}{\partial h_1}\cdot\frac{\partial h_1}{\partial w_1}\right)\right.$$

$$\left.+\frac{\partial z_5^*}{\partial h_3^*}\cdot\frac{\partial h_3^*}{\partial z_3^*}\left(\frac{\partial z_3^*}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial w_1}+\frac{\partial z_3^*}{\partial h_1}\cdot\frac{\partial h_1}{\partial w_1}\right)\right]$$

$$= \frac{\partial y^*}{\partial z_5^*}\cdot\left(\frac{\partial z_5^*}{\partial h_4}\cdot\frac{\partial h_4}{\partial z_4}\cdot\frac{\partial z_4}{\partial h_3^*}\cdot\frac{\partial h_3^*}{\partial z_3^*}+\frac{\partial z_5^*}{\partial h_3^*}\cdot\frac{\partial h_3^*}{\partial z_3^*}\right)\left(\frac{\partial z_3^*}{\partial h_2}\cdot\frac{\partial h_2}{\partial z_2}\cdot\frac{\partial z_2}{\partial h_1}\cdot\frac{\partial h_1}{\partial w_1}+\frac{\partial z_3^*}{\partial h_1}\cdot\frac{\partial h_1}{\partial w_1}\right)$$

$$= \delta'(z_5^*)\cdot\left(w\cdot\delta'(z_4)\cdot w\cdot\delta'(z_3^*)+1\cdot\delta'(z_3^*)\right)\left(w\cdot\delta'(z_2)\cdot w\cdot\delta'(z_1)x+1\cdot\delta'(z_1)x\right)$$

$$= \left(w^2\cdot\delta'(z_4)\cdot\delta'(z_3^*)+\delta'(z_3^*)\right)\left(w^2\delta'(z_2)+1\right)\delta'(z_5^*)\cdot\delta'(z_1)x$$

$$= \left(w^2\delta'(z_4)+1\right)\left(w^2\delta'(z_2)+1\right)\delta'(z_5^*)\,\delta'(z_3^*)\cdot\delta'(z_1)x$$

$$= w^4 \Delta'(z_8^*)\Delta'(z_4)\cdot\Delta'(z_5^*)\Delta'(z_2)\cdot\Delta'(z_1)x$$
$$+ w^2\left(\Delta'(z_4)+\Delta'(z_2)\right)\Delta'(z_5^*)\Delta'(z_3^*)\cdot\Delta'(z_1)x$$
$$+ \Delta'(z_5^*)\Delta'(z_3^*)\Delta'(z_1)x$$

$$\frac{\partial y^*}{\partial b_1} = w^4\Delta'(z_8^*)\cdot\Delta'(z_4)\cdot\Delta'(z_5^*)\cdot\Delta'(z_2)\cdot\Delta'(z_1)$$
$$+ w^2\left(\Delta'(z_4)+\Delta'(z_2)\right)\Delta'(z_5^*)\Delta'(z_3^*)\Delta'(z_1)$$
$$+ \Delta'(z_5^*)\cdot\Delta'(z_3^*)\cdot\Delta'(z_1)$$

only consider $\left|\frac{\partial y^*}{\partial b_1}\right|$ and $\left|\frac{\partial y}{\partial b_1}\right|$ because $\frac{\partial y^*}{\partial w_1}$ and $\frac{\partial y}{\partial w_1}$ are same

$$\frac{\partial y^*}{\partial b_1} - \frac{\partial y}{\partial b_1} = w^4\Delta'(z_1)\cdot\Delta'(z_2)\underbrace{\left[\Delta'(z_5^*)\Delta'(z_4^*)\Delta'(z_3^*) - \Delta'(z_3)\Delta'(z_4)\Delta'(z_5)\right]}_{(A)}$$
$$+ w^2\left[\Delta'(z_4)+\Delta'(z_2)\right]\Delta'(z_5^*)\cdot\Delta'(z_3^*)\cdot\Delta'(z_1)$$
$$+ \Delta'(z_5^*)\cdot\Delta'(z_3^*)\cdot\Delta'(z_1) \Big\} (B).$$

B is always >0 but there are 3 situation about A ($A>0$, $A=0$ and $A<0$)

if $A\geq 0$ then $\frac{\partial y^*}{\partial b_1} \geq \frac{\partial y}{\partial b_1}$ and $\left|\frac{\partial y^*}{\partial w_1}\right| \geq \left|\frac{\partial y}{\partial w_1}\right|$

BUT: if $A<0$, which is $\Delta'(z_5^*)\cdot\Delta'(z_4^*)\Delta'(z_3^*) - \Delta'(z_3)\cdot\Delta'(z_4)\cdot\Delta'(z_5) < 0$

the only condition I have is $\Delta'(x)>0$, which can not solve the problem. (because the monotonicity of $\Delta'(x)$ is not known.

in this situation:

It seems very hard to find out whether $\frac{\partial y^*}{\partial b_1} \geq \frac{\partial y}{\partial b_1}$ or not because I do not know what exactly the monotonicity of $\Delta'(x)$

<2>

① the slope in $0x = -1$ and slope between x_0 and the peek of the bump is 1

Set point x_0 as $(0,0)$ then point "0" is $(-1,1)$, peek of the bump is (h,h)

because α is learning rate and is 0.3 then the step is :

$(-1,1) \Rightarrow (-0.7, 0.7) \Rightarrow (-0.4, 0.4) \Rightarrow (-0.1, 0.1) \Rightarrow (0.2, 0.2)$

when comes to point $(0.2, 0.2)$, because the slop becomes 1, so next it will come to $(-0.1, 0.1)$, and the slop change to -1 again, the next step is $(0.2, 0.2)$, so gradient descent will stuck between $(-0.1, 0.1)$ and $(0.2, 0.2)$ around point x_0.

② as long as $h > x_0 = 0.41018429512996585$, the adam will be stuck at x_0.

I choose step as 0.001 then max height of h I get is 0.410

<3>

the label of the parms shows at right side. ⇒

the label of node in graph shows below

```
a = Variable(torch.tensor([1.0]), requires_grad=True)
b = Variable(torch.tensor([1.0]), requires_grad=True)
c = Variable(torch.tensor([1.0]), requires_grad=True)
d = Variable(torch.tensor([1.0]), requires_grad=True)
e = Variable(torch.tensor([1.0]), requires_grad=True)
f = Variable(torch.tensor([1.0]), requires_grad=True)
x0 = Variable(torch.tensor([.5]), requires_grad=True)

# updater step - first step
x1 = b*x0 + c
x2 = x0 + x1*f # one step
x3 = (x0-e*x1)**a + torch.sin(d*x2)
x4 = x3*x2 + d*x1
```

ode

s tree
abel