Assignment 2 (Li Yonghao)

Question 1:

(1)
$$z_5 = wh_4 + b \qquad y = \sigma(z_5)$$
$$z_4 = wh_3 + b \qquad h_4 = \sigma(z_4)$$
$$z_3 = wh_2 + b \qquad h_3 = \sigma(z_3)$$
$$z_2 = wh_1 + b \qquad h_2 = \sigma(z_2)$$
$$z_1 = w_1 x + b_1 \qquad h_1 = \sigma(z_1)$$

$$\frac{\partial y}{\partial z_5} = \sigma'(z_5)$$
$$\frac{\partial y}{\partial z_4} = \frac{\partial y}{\partial z_5} \cdot \frac{\partial z_5}{\partial h_4} \cdot \frac{\partial h_4}{\partial z_4} = w\sigma'(z_5)\sigma'(z_4)$$
$$\frac{\partial y}{\partial z_3} = \frac{\partial y}{\partial z_4} \cdot \frac{\partial z_4}{\partial h_3} \cdot \frac{\partial h_3}{\partial z_3} = w^2\sigma'(z_5)\sigma'(z_4)\sigma'(z_3)$$
$$\frac{\partial y}{\partial z_2} = \frac{\partial y}{\partial z_3} \cdot \frac{\partial z_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} = w^3\sigma'(z_5)\sigma'(z_4)\sigma'(z_3)\sigma'(z_2)$$
$$\frac{\partial y}{\partial z_1} = \frac{\partial y}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} = w^4\sigma'(z_5)\sigma'(z_4)\sigma'(z_3)\sigma'(z_2)\sigma'(z_1)$$
$$\therefore \frac{\partial y}{\partial w_1} = \frac{\partial y}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1} = x w^4\sigma'(z_5)\sigma'(z_4)\sigma'(z_3)\sigma'(z_2)\sigma'(z_1)$$
$$\frac{\partial y}{\partial b_1} = \frac{\partial y}{\partial z_1} \cdot \frac{\partial z_1}{\partial b_1} = w^4\sigma'(z_5)\sigma'(z_4)\sigma'(z_3)\sigma'(z_2)\sigma'(z_1)$$

(2)
$$z_5' = wh_4 + h_3^* + b \qquad y^* = \sigma(z_5)$$
$$z_4' = wh_3 + b \qquad h_4 = \sigma(z_4)$$
$$z_3' = wh_2 + h_1 + b \qquad h_3^* = \sigma(z_3)$$
$$z_2' = wh_1 + b \qquad h_2 = \sigma(z_2)$$
$$z_1' = w_1 x + b_1 \qquad h_1 = \sigma(z_1)$$

$$\frac{\partial y^*}{\partial z_5'} = \sigma'(z_5')$$
$$\frac{\partial y^*}{\partial z_4'} = \frac{\partial y^*}{\partial z_5'} \cdot \frac{\partial z_5'}{\partial h_4} \cdot \frac{\partial h_4}{\partial z_4'} = w \cdot \sigma'(z_5') \cdot \sigma'(z_4')$$
$$\frac{\partial y^*}{\partial z_3'} = \frac{\partial y^*}{\partial z_4'} \cdot \frac{\partial z_4'}{\partial h_3^*} \cdot \frac{\partial h_3^*}{\partial z_3'} + \frac{\partial y^*}{\partial z_5'} \cdot \frac{\partial z_5'}{\partial h_3^*} \cdot \frac{\partial h_3^*}{\partial z_3'} = w^2\sigma'(z_5')\sigma'(z_4')\sigma'(z_3') + \sigma'(z_5')\sigma'(z_3')$$
$$\frac{\partial y^*}{\partial z_2'} = \frac{\partial y^*}{\partial z_3'} \cdot \frac{\partial z_3'}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2'} = w^3\sigma'(z_5')\sigma'(z_4')\sigma'(z_3')\sigma'(z_2') + w\sigma'(z_5')\sigma'(z_3')\sigma'(z_2')$$
$$\frac{\partial y^*}{\partial z_1'} = \frac{\partial y^*}{\partial z_2'} \cdot \frac{\partial z_2'}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1'} + \frac{\partial y^*}{\partial z_3'} \cdot \frac{\partial z_3'}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1'} = w^4\sigma'(z_5')\sigma'(z_4')\sigma'(z_3')\sigma'(z_2')\sigma'(z_1') + w^2\sigma'(z_5')\sigma'(z_3')\sigma'(z_2')\sigma'(z_1')$$
$$+ w^2\sigma'(z_5')\sigma'(z_4')\sigma'(z_3')\sigma'(z_1') + \sigma'(z_5')\sigma'(z_3')\sigma'(z_1')$$
$$\therefore \frac{\partial y^*}{\partial w_1} = \frac{\partial y^*}{\partial z_1'} \cdot \frac{\partial z_1'}{\partial w_1} = x w^4\sigma'(z_5')\sigma'(z_4')\sigma'(z_3')\sigma'(z_2')\sigma'(z_1') + x w^2\sigma'(z_5')\sigma'(z_3')\sigma'(z_2')\sigma'(z_1')$$
$$+ x w^2\sigma'(z_5')\sigma'(z_4')\sigma'(z_3')\sigma'(z_1') + x\sigma'(z_5')\sigma'(z_3')\sigma'(z_1')$$
$$\frac{\partial y^*}{\partial b_1} = \frac{\partial y^*}{\partial z_1'} \cdot \frac{\partial z_1'}{\partial b_1} = w^4\sigma'(z_5')\sigma'(z_4')\sigma'(z_3')\sigma'(z_2')\sigma'(z_1') + w^2\sigma'(z_5')\sigma'(z_3')\sigma'(z_2')\sigma'(z_1')$$
$$+ w^2\sigma'(z_5')\sigma'(z_4')\sigma'(z_3')\sigma'(z_1') + \sigma'(z_5')\sigma'(z_3')\sigma'(z_1')$$

$$\begin{cases} \sigma'(z) > 0 \\ w^4 \geq 0 \\ w^2 \geq 0 \end{cases} \Rightarrow \begin{array}{l} w^4\sigma'(z_5')\sigma'(z_4')\sigma'(z_3')\sigma'(z_2')\sigma'(z_1') + w^2\sigma'(z_5')\sigma'(z_3')\sigma'(z_2')\sigma'(z_1') + \\ + w^2\sigma'(z_5')\sigma'(z_4')\sigma'(z_3')\sigma'(z_1') + \sigma'(z_5')\sigma'(z_3')\sigma'(z_1') > 0 \end{array}$$

So, we can just compare $w^4\sigma'(z_5)\sigma'(z_4)\sigma'(z_3)\sigma'(z_2)\sigma'(z_1)$
and $w^4\sigma'(z_5')\sigma'(z_4)\sigma'(z_3')\sigma'(z_2)\sigma'(z_1)$

the only difference is $\begin{cases} \sigma'(z_5)\sigma'(z_3) & z_5 = wh_4 + b,\ z_3 = wh_2 + b \\ \sigma'(z_5')\sigma'(z_3') & z_5' = wh_4 + h_3^* + b,\ z_3' = wh_2 + h_1 + b \end{cases}$

We don't know the monotonicity of the derivate $\sigma'(z)$, so we can't get the result $\left|\frac{\partial y}{\partial w_1}\right| \leq \left|\frac{\partial y^*}{\partial w_1}\right|$ and $\left|\frac{\partial y}{\partial b_1}\right| \leq \left|\frac{\partial y^*}{\partial b_1}\right|$

Question 2:

1. Apply standard gradient descend:

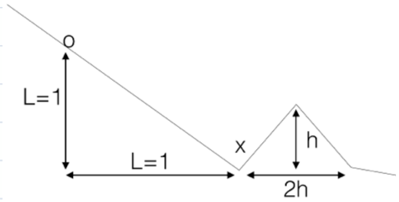
$x_0 = 0,\ y_0 = 1,\ a = 0.3,\ g \begin{cases} 1, & 1 < x < 1+h \\ -1, & x < 1 \text{ or } x > 1+h \end{cases}$

$x_n = x_{n-1} - ag$

$(x_1, y_1) = (0.3, 0.7),\ (x_2, y_2) = (0.6, 0.4)$

$(x_3, y_3) = (0.9, 0.1),\ (x_4, y_4) = (1.2, 0.2)$

$(x_5, y_5) = (0.9, 0.1),\ (x_6, y_6) = (1.2, 0.2)$



It stuck around point ‘x’, my model will oscillate back and forth near x and fall into a local optimal solution.

2. Adam optimization:

```
m,v,m_,v_,theta = [0],[0],[0],[0],[0]
b1,b2,a,t = 0.9,0.999,0.3,0
for i in range(8):
    t = t + 1
    if(theta[-1]<1):
        g = -1
    else: g = 1
    m.append(b1 * m[-1] + (1 - b1) * g)
    v.append(b2 * v[-1] + (1 - b2) * (g**2))
    m_.append(m[-1] / (1 - b1**t))
    v_.append(v[-1] / (1 - b2**t))
    theta.append(theta[-1] - a * m_[-1] * (v_[-1]**0.5))
theta
```

```
Out[5]: [0,
 0.3,
 0.6000000000000021,
 0.9000000000000021,
 1.2000000000000035,
 1.35348343141804,
 1.4101842951299737,
 1.3985127716140122,
 1.3362158088702167]
```

I use the adam potimisation with parameters given by the question, I run it 8 times to see where it will converge, finally I find that $\theta = 1.41$ is the critical value. So, the max height ‘h’ of the bump in which the adam optimiser will escape the local min at ‘x’ is:

h(max) = 1.41 – 1 = 0.41

Question 3:

You are given the code

comp_tree.py

Which generates this tree given on the right. Label all the nodes