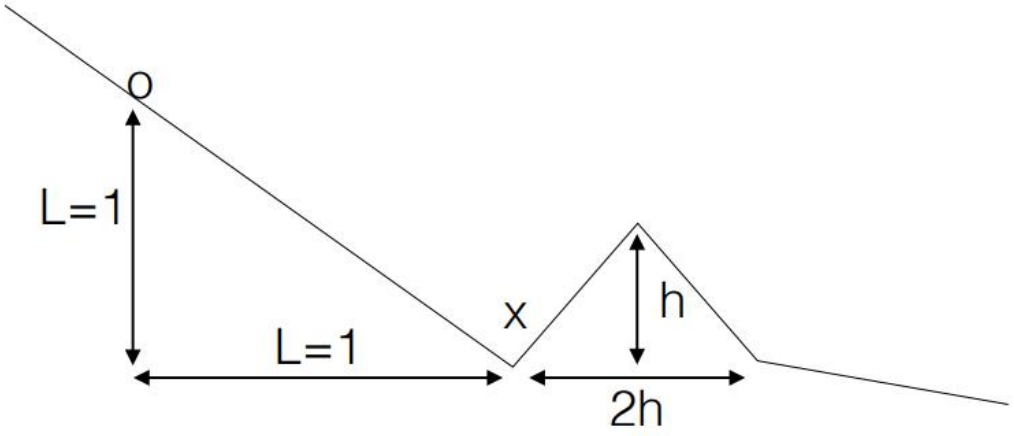
Assignment 2

1.
2.



1. SGD

I set the x axis for O is 0.

For the first optimize, the gradient = -1 and the a = 0.3. so the O will move from 0 to 0.3.

For the second optimize, the gradient = -1 and the a = 0.3. so the O will move from 0 to 0.6.

For the third optimize, the gradient = -1 and the a = 0.3. so the O will move from 0 to 0.9.

For the fourth optimize, the gradient = -1 and the a = 0.3. so the O will move from 0 to 1.2.

For the fifth optimize, the gradient = 1 and the a = 0.3. so the O will move from 0 to 0.9.

For the sixth optimize, the gradient = -1 and the a = 0.3. so the O will move from 0 to 1.2.

And converge.

2. Adam

If we want to escape from the local min x(x axis=1), we need the point to move to the position that larger than the peek(x axis > 1 + h). since h is known, I assume that h is positive infinite, and to see with given parameter and adam, what is the max length that the point can move. And the h should be smaller than that value.

The following is the code for adam.

```
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))
```

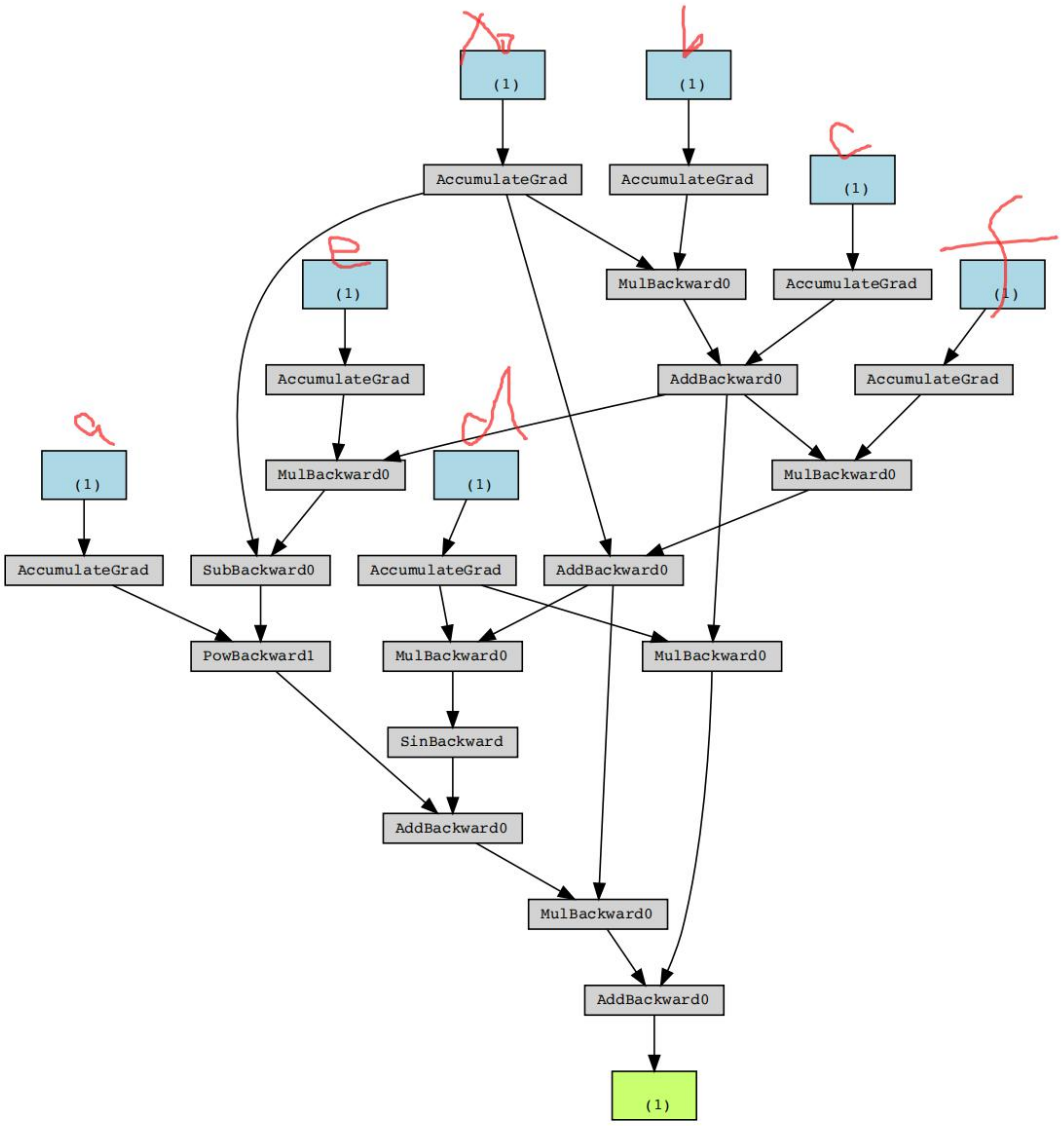
And result is as follow

```
theta
```

```
[0,
 0.3,
 0.6000000000000021,
 0.9000000000000021,
 1.2000000000000035,
 1.35348343141804,
 1.4101842951299737,
 1.3985127716140122,
 1.3362158088702167,
 1.2351152242311723,
 1.1034768180909487,
 0.9473026798481269,
 0.8547005754938964,
 0.8147392651127069,
 0.818858241483421]
```

We can see that at the beginning and the gradient = -1, the Theta is optimized just like the SGD, but when theta = 1.2, the gradient change to 1. But unlike the SGD, the next theta is not 0.9 but with the momentum, the next theta = 1.35. With this momentum, the max value that O can reach is 1.41. And after that the O start to go back to the local min x. So if we want to escape form the local min, the highest point(1.41) should have the gradient < 0.
So the max value of h is 1.41 – 1 = 0.41.

3.


(1)
(1)
AccumulateGrad
AccumulateGrad
(1)
(1)
MulBackward0
AccumulateGrad
(1)
AccumulateGrad
AddBackward0
AccumulateGrad
(1)
MulBackward0
(1)
MulBackward0
AccumulateGrad
SubBackward0
AccumulateGrad
AddBackward0
PowBackward1
MulBackward0
MulBackward0
SinBackward
AddBackward0
MulBackward0
AddBackward0
(1)