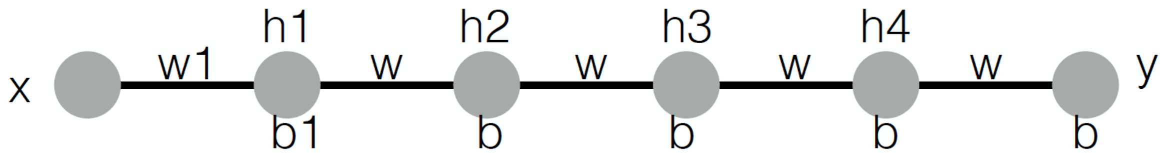
BS6207 Homework 2

Isaac Lin

1. Gradients of weights of 2 different connections



$z_1 = w_1 \cdot x + b_1$
$h_1 = \sigma(w_1 \cdot x + b_1)$

$z_2 = w \cdot h_1 + b$
$h_2 = \sigma(w \cdot h_1 + b)$

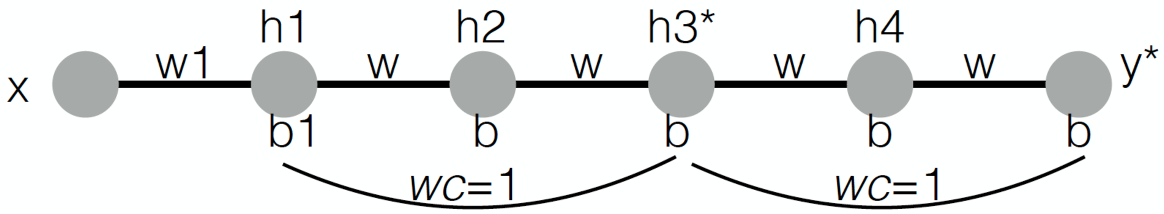
$z_3 = w \cdot h_2 + b$
$h_3 = \sigma(w \cdot h_2 + b)$

$z_4 = w \cdot h_3 + b$
$h_4 = \sigma(w \cdot h_3 + b)$

$z_5 = w \cdot h_4 + b$
$y = \sigma(w \cdot h_4 + b)$

$$\frac{dy}{dw_1} = \sigma'(z_5)w \cdot \sigma'(z_4)w \cdot \sigma'(z_3)w \cdot \sigma'(z_2)w \cdot \sigma'(z_1)x$$
$$= w^4 x \cdot \sigma'(z_5)\sigma'(z_4)\sigma'(z_3)\sigma'(z_2)\sigma'(z_1)$$

$$\frac{dy}{db_1} = \sigma'(z_5)w \cdot \sigma'(z_4)w \cdot \sigma'(z_3)w \cdot \sigma'(z_2)w \cdot \sigma'(z_1)$$
$$= w^4 \cdot \sigma'(z_5)\sigma'(z_4)\sigma'(z_3)\sigma'(z_2)\sigma'(z_1)$$



$$z_1 = w_1 \cdot x + b_1$$
$$h_1 = \sigma(w_1 \cdot x + b_1)$$

$$z_2 = w \cdot h_1 + b$$
$$h_2 = \sigma(w \cdot h_1 + b)$$

$$z_3 = w \cdot h_2 + h_1 + b$$
$${h_3}^* = \sigma(w \cdot h_2 + h_1 + b)$$

$$z_4 = w \cdot {h_3}^* + b$$
$$h_4 = \sigma(w \cdot {h_3}^* + b)$$

$$z_5 = w \cdot h_4 + {h_3}^* + b$$
$$y^* = \sigma(w \cdot h_4 + {h_3}^* + b)$$

$$\frac{dy^*}{dw_1} = \sigma'(z_5)w_c \cdot \sigma'(z_3)w_c \cdot \sigma'(z_1)x$$
$$= x \cdot \sigma'(z_5)\sigma'(z_3)\sigma'(z_1)$$

$$\frac{dy^*}{db_1} = \sigma'(z_5)w_c \cdot \sigma'(z_3)w_c \cdot \sigma'(z_1)$$
$$= \sigma'(z_5)\sigma'(z_3)\sigma'(z_1)$$

$$\frac{\left|\frac{dy}{dw_1}\right|}{\left|\frac{dy^*}{dw_1}\right|} = \left|\frac{w^4 x \cdot \sigma'(z_5)\sigma'(z_4)\sigma'(z_3)\sigma'(z_2)\sigma'(z_1)}{x \cdot \sigma'(z_5)\sigma'(z_3)\sigma'(z_1)}\right|$$
$$= |w^4 \cdot \sigma'(z_4)\sigma'(z_2)|$$

Since $w < 1, w^4 \ll 1$ and $0 < \sigma' < 0.25$. Thus $\frac{\left|\frac{dy}{dw_1}\right|}{\left|\frac{dy^*}{dw_1}\right|} = |w^4 \cdot \sigma'(z_4)\sigma'(z_2)| < 1.$

$$\frac{\left|\frac{dy}{db_1}\right|}{\left|\frac{dy^*}{db_1}\right|} = \left|\frac{w^4 \cdot \sigma'(z_5)\sigma'(z_4)\sigma'(z_3)\sigma'(z_2)\sigma'(z_1)}{\sigma'(z_5)\sigma'(z_3)\sigma'(z_1)}\right|$$
$$= |w^4 \cdot \sigma'(z_4)\sigma'(z_2)|$$

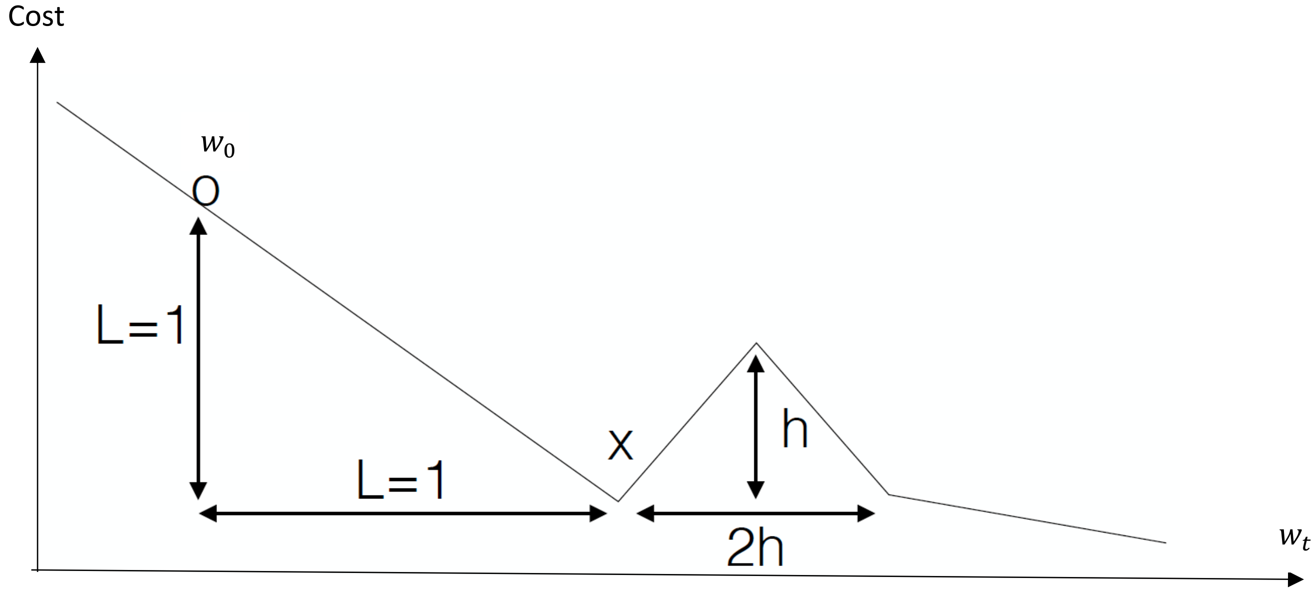
Since $w < 1, w^4 \ll 1$ and $0 < \sigma' < 0.25$. Thus $\frac{\left|\frac{dy}{db_1}\right|}{\left|\frac{dy^*}{db_1}\right|} = |w^4 \cdot \sigma'(z_4)\sigma'(z_2)| < 1.$

2. Adam Optimizer

a. What happens if you apply standard gradient descent?

Standard gradient descent only looks for a local minimum, not a global minimum. Therefore, after several epochs, the cost function at the point 'o' would converge to 'x'. Since gradients at both ends of 'x' is increasing, it will be stuck at point 'x'.

b. Applying Adam optimization, what is the max height 'h' in which the Adam optimizer will escape the local minimum at 'x'?



$\alpha = 0.3$
$\beta_1 = 0.9$
$\beta_2 = 0.999$
$\varepsilon = 0$

$$dw_t = \begin{cases} -1 & \text{if } w_t - w_0 < 1 \\ 1 & \text{if } 1 < w_t - w_0 < 1 + h \\ -1 & \text{if } w_t - w_0 > 1 + h \\ 0 & \text{if } w_t - w_0 = 1 \text{ or } 1 + h \end{cases}$$

Adam Optimization Equations

$$g_t = dw_{t-1}$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\, g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)\, g_t^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$w_t = w_{t-1} - \frac{\alpha \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

For example, we define $w_0 = 0$

For the first instance:
$t = 1$
$\mathrm{g_t} = -1$
$\mathrm{m_t} = 0 + (1 - \ 0.9) \cdot -1 = -0.1$

$$\mathrm{v_t} = 0 + (1 - \ 0.999) \cdot (-1)^2 = 0.001$$

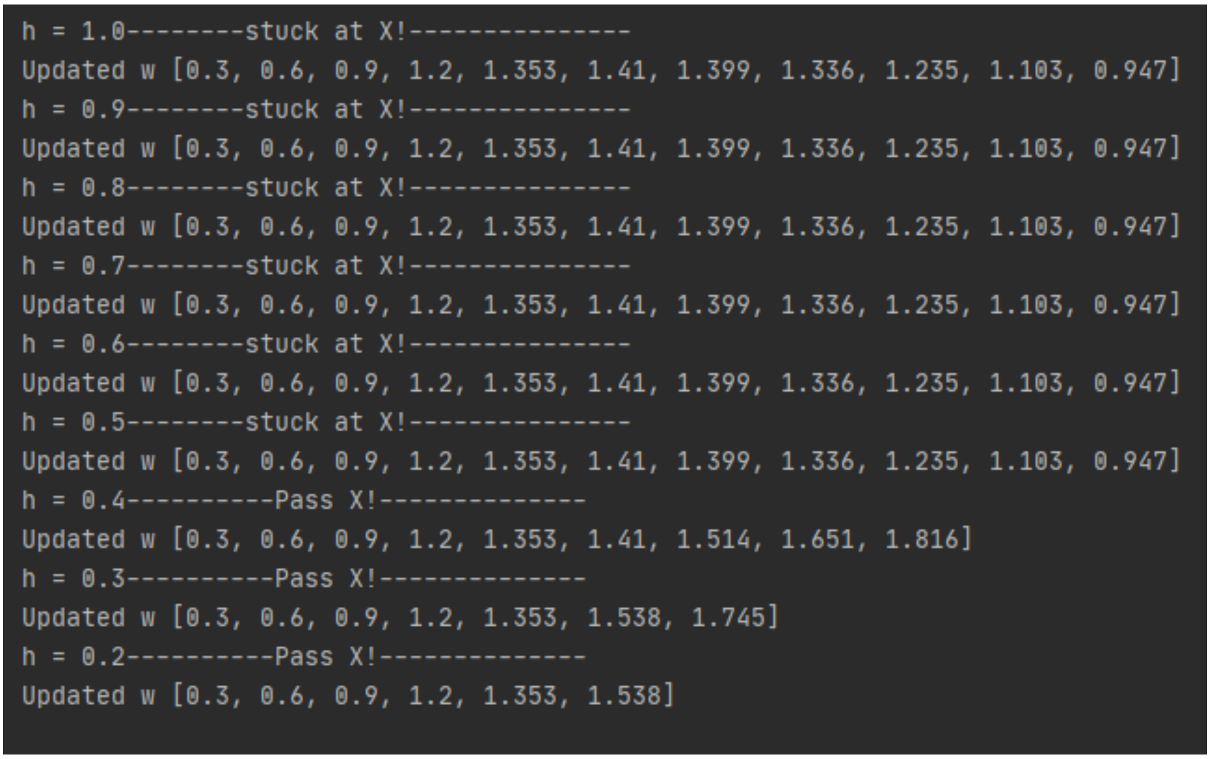
$$\widehat{\mathrm{m}}_\mathrm{t} = \ \frac{\mathrm{m_t}}{1 - \beta_1^\mathrm{t}} = -\frac{0.1}{1 - 0.9} = -1$$

$$\hat{\mathrm{v}}_\mathrm{t} = \frac{\mathrm{v_t}}{1 - \beta_2^\mathrm{t}} = \frac{0.001}{1 - 0.999} = 1$$

$$\mathrm{w_t} = \mathrm{w_{t-1}} \ - \frac{\alpha \widehat{\mathrm{m}}_\mathrm{t}}{\sqrt{\hat{\mathrm{v}}_\mathrm{t}} + \epsilon} = 0 - \frac{0.3*(-1)}{\sqrt{1}+0} = 0.3$$

Weight of $w_0 = 0$ is updated to $w_1 = 0.3$ in the next instance.

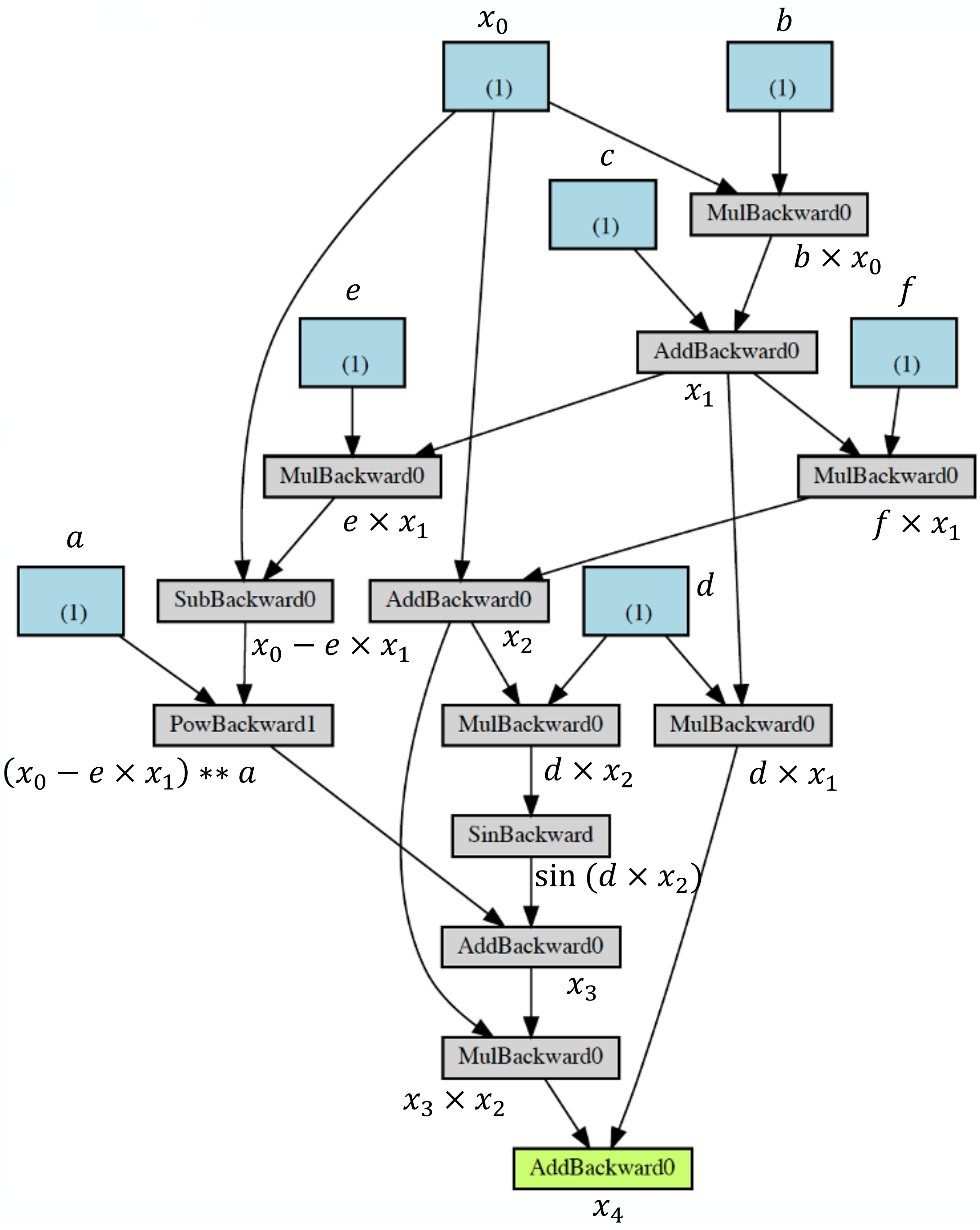
Below shows different combinations of h.



```
h = 1.0--------stuck at X!---------------
Updated w [0.3, 0.6, 0.9, 1.2, 1.353, 1.41, 1.399, 1.336, 1.235, 1.103, 0.947]
h = 0.9--------stuck at X!---------------
Updated w [0.3, 0.6, 0.9, 1.2, 1.353, 1.41, 1.399, 1.336, 1.235, 1.103, 0.947]
h = 0.8--------stuck at X!---------------
Updated w [0.3, 0.6, 0.9, 1.2, 1.353, 1.41, 1.399, 1.336, 1.235, 1.103, 0.947]
h = 0.7--------stuck at X!---------------
Updated w [0.3, 0.6, 0.9, 1.2, 1.353, 1.41, 1.399, 1.336, 1.235, 1.103, 0.947]
h = 0.6--------stuck at X!---------------
Updated w [0.3, 0.6, 0.9, 1.2, 1.353, 1.41, 1.399, 1.336, 1.235, 1.103, 0.947]
h = 0.5--------stuck at X!---------------
Updated w [0.3, 0.6, 0.9, 1.2, 1.353, 1.41, 1.399, 1.336, 1.235, 1.103, 0.947]
h = 0.4----------Pass X!-------------
Updated w [0.3, 0.6, 0.9, 1.2, 1.353, 1.41, 1.514, 1.651, 1.816]
h = 0.3----------Pass X!-------------
Updated w [0.3, 0.6, 0.9, 1.2, 1.353, 1.538, 1.745]
h = 0.2----------Pass X!-------------
Updated w [0.3, 0.6, 0.9, 1.2, 1.353, 1.538]
```

Therefore, h = 0.4 is the maximum height. More specifically, h = 0.41.

3. Label nodes


x_0
(1)
b
(1)
c
(1)
MulBackward0
$b \times x_0$
e
(1)
AddBackward0
f
(1)
x_1
MulBackward0
MulBackward0
$e \times x_1$
$f \times x_1$
a
(1)
SubBackward0
AddBackward0
(1)
d
$x_0 - e \times x_1$
x_2
PowBackward1
MulBackward0
MulBackward0
$(x_0 - e \times x_1) ** a$
$d \times x_2$
$d \times x_1$
SinBackward
$\sin(d \times x_2)$
AddBackward0
x_3
MulBackward0
$x_3 \times x_2$
AddBackward0
x_4