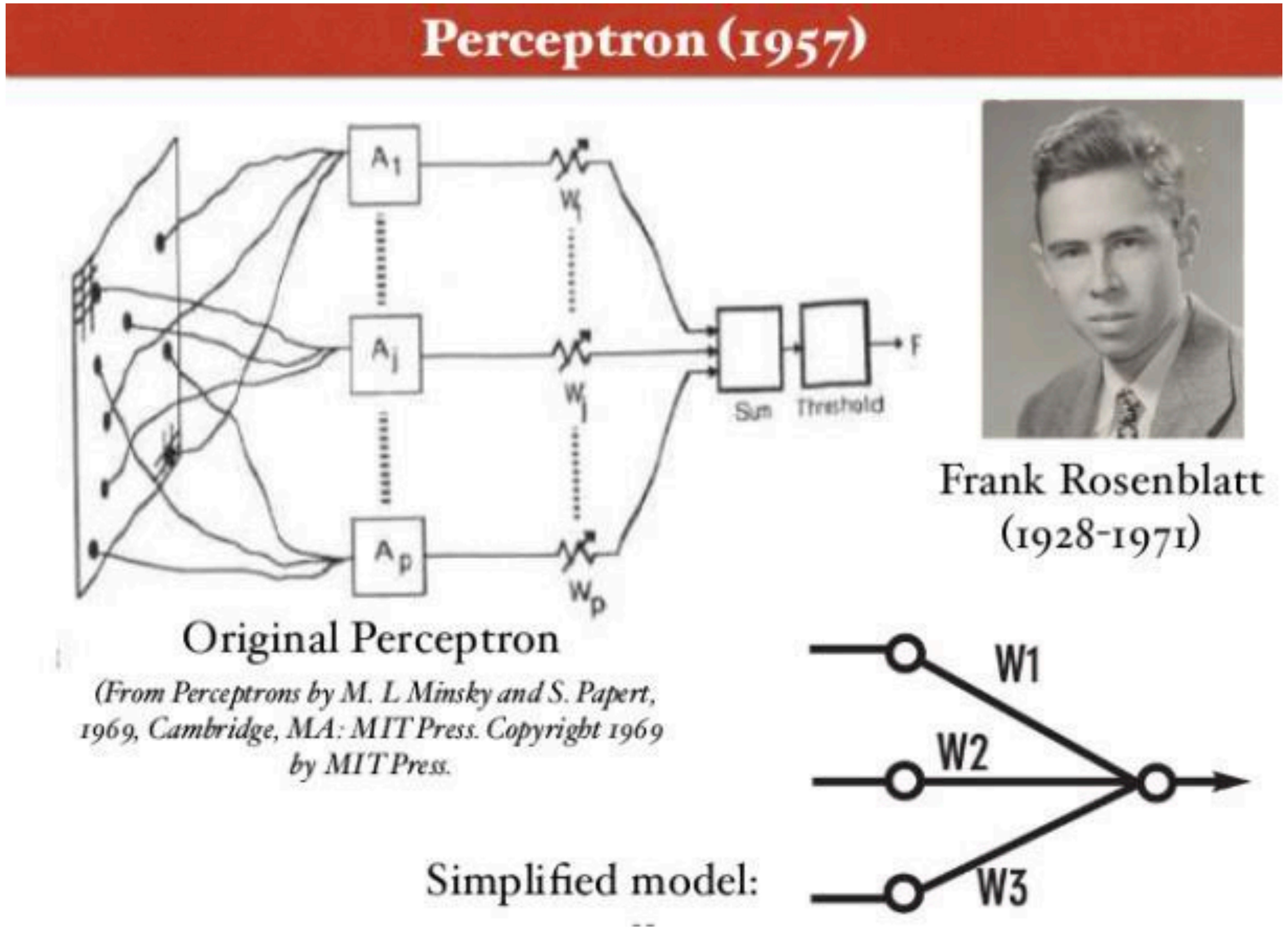
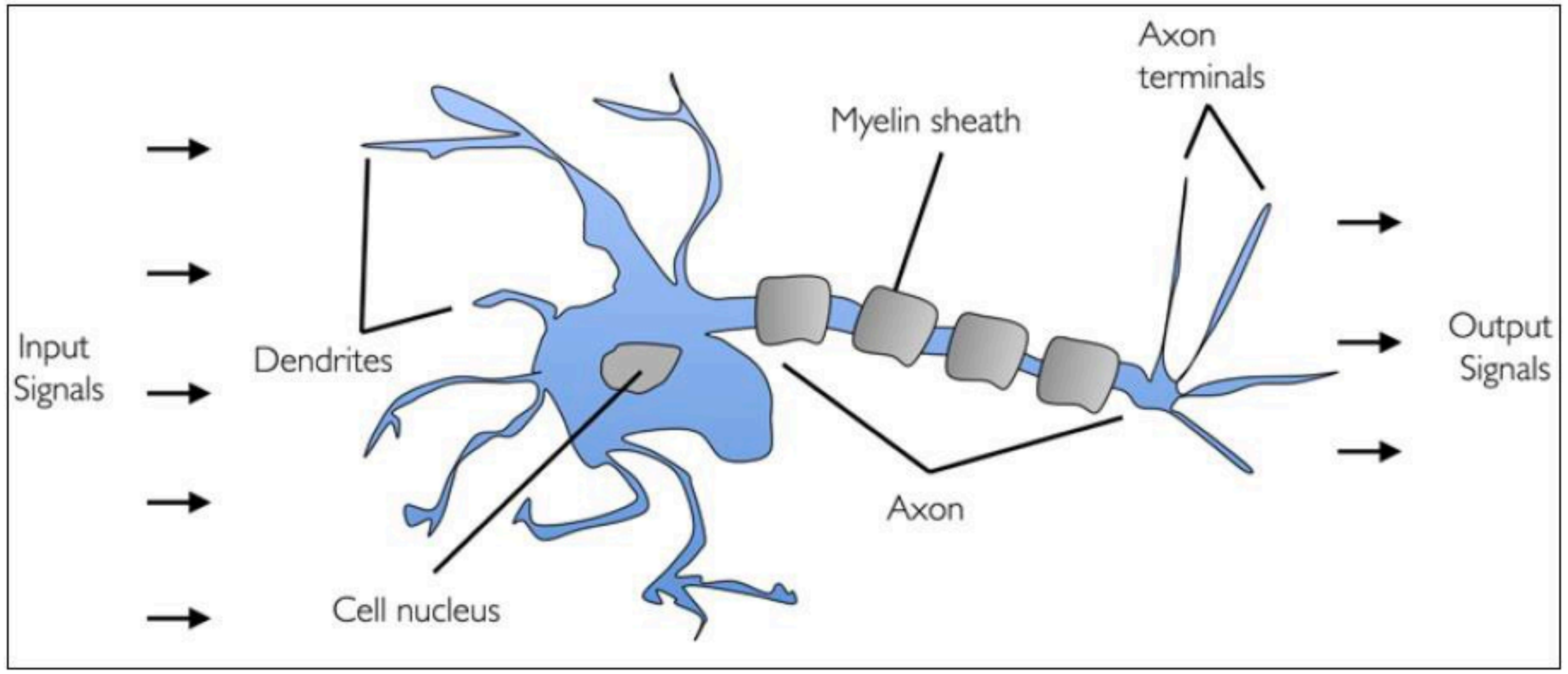
01.PERCEPTRON

Perceptron

미국 정신분석 학자 프랭크 로잔블라트에 의해 창안됨



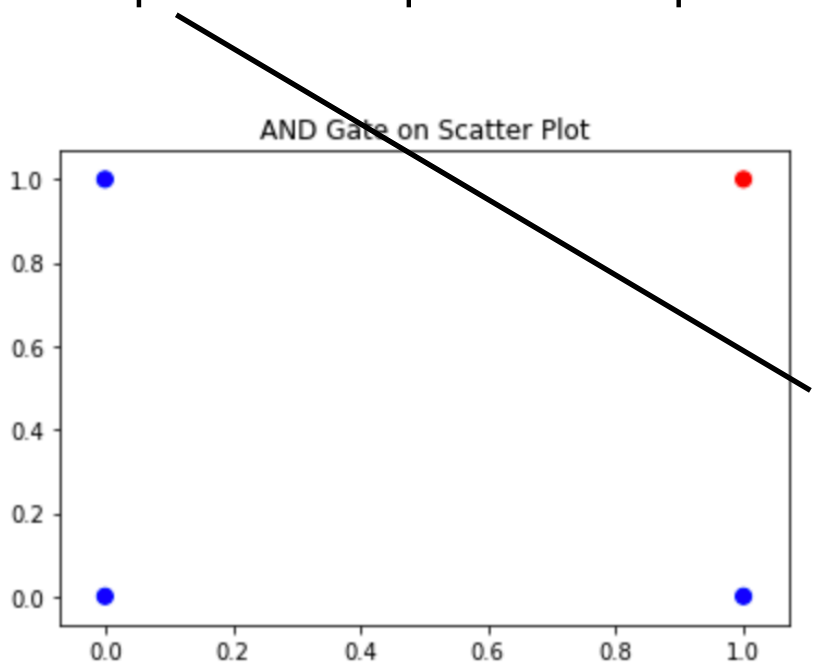


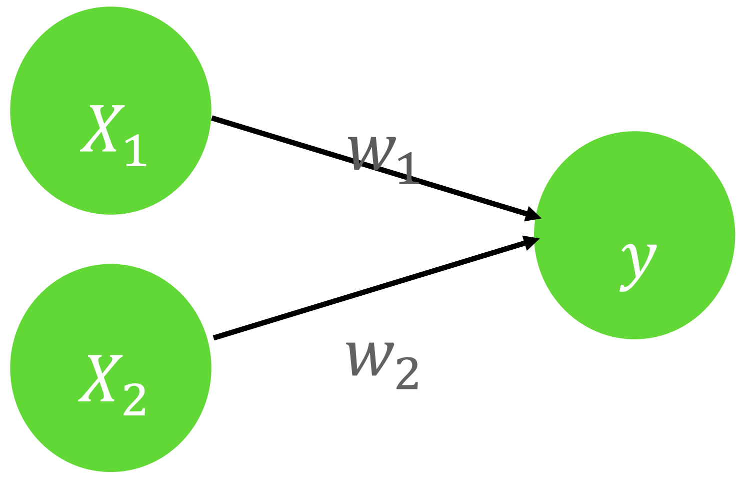
$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

퍼셉트론 으로 논리 회로 구현

AND

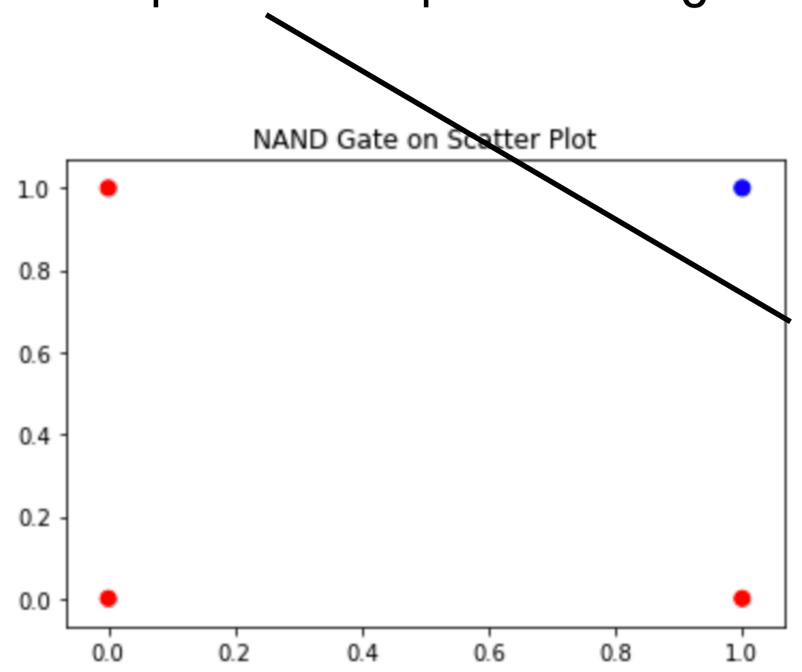
X1	X2	Y
0	0	0
1	0	0
0	1	0
1	1	1

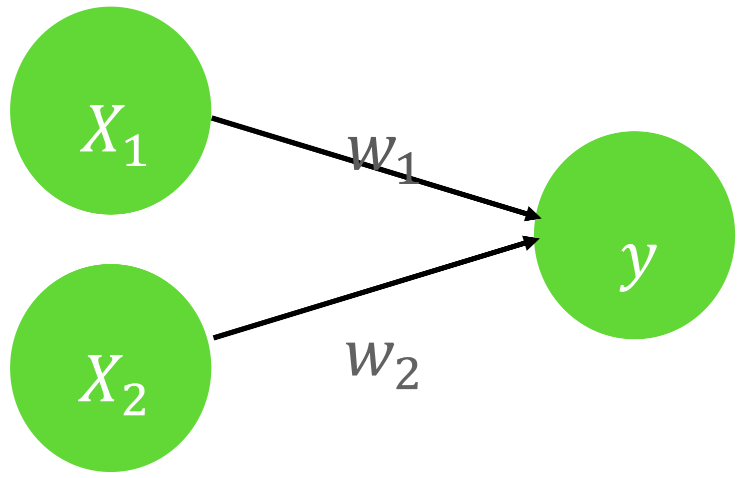




NDND

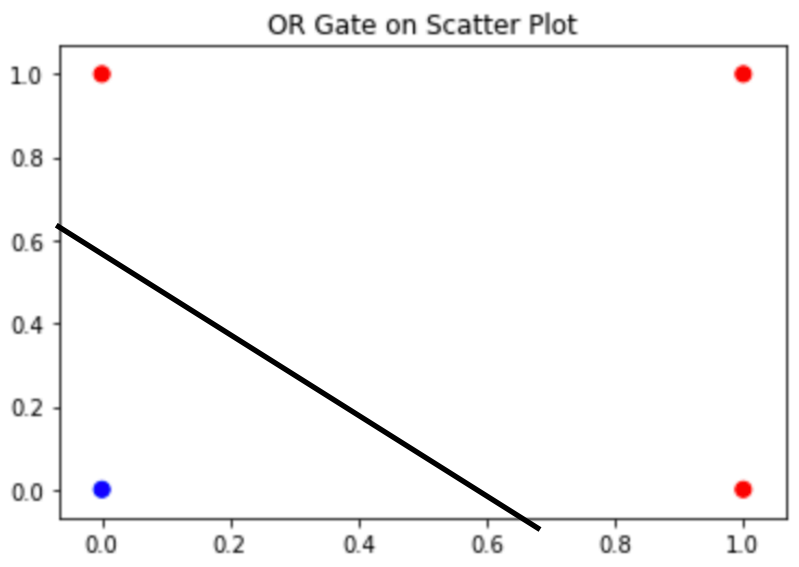
X1	X2	Y
0	0	1
1	0	1
0	1	1
1	1	0

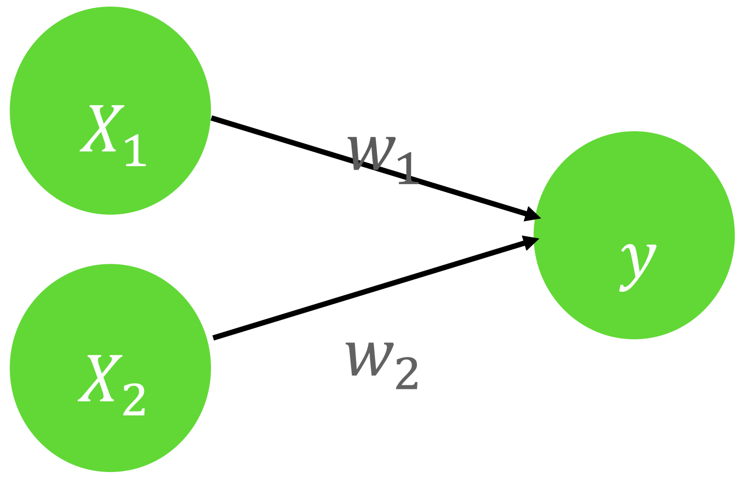




OR

X1	X2	Y
0	0	0
1	0	1
0	1	1
1	1	1

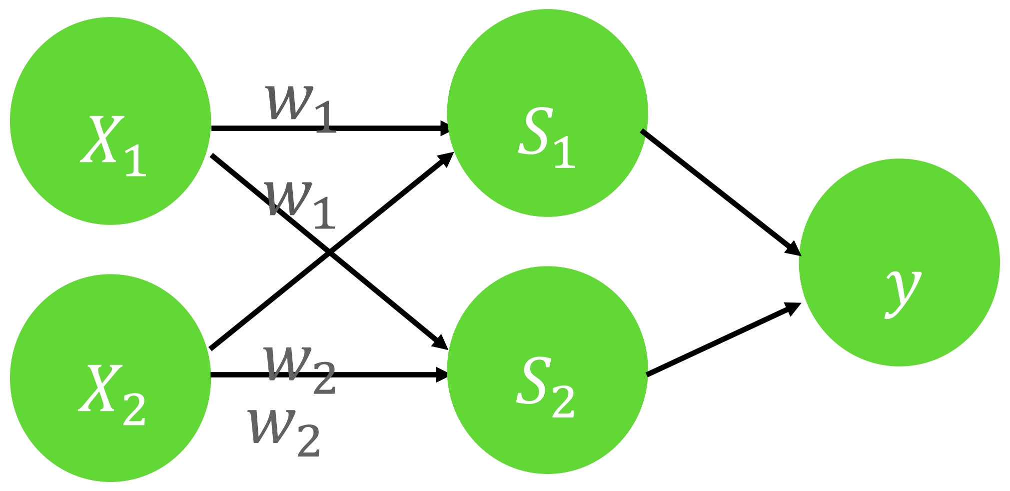




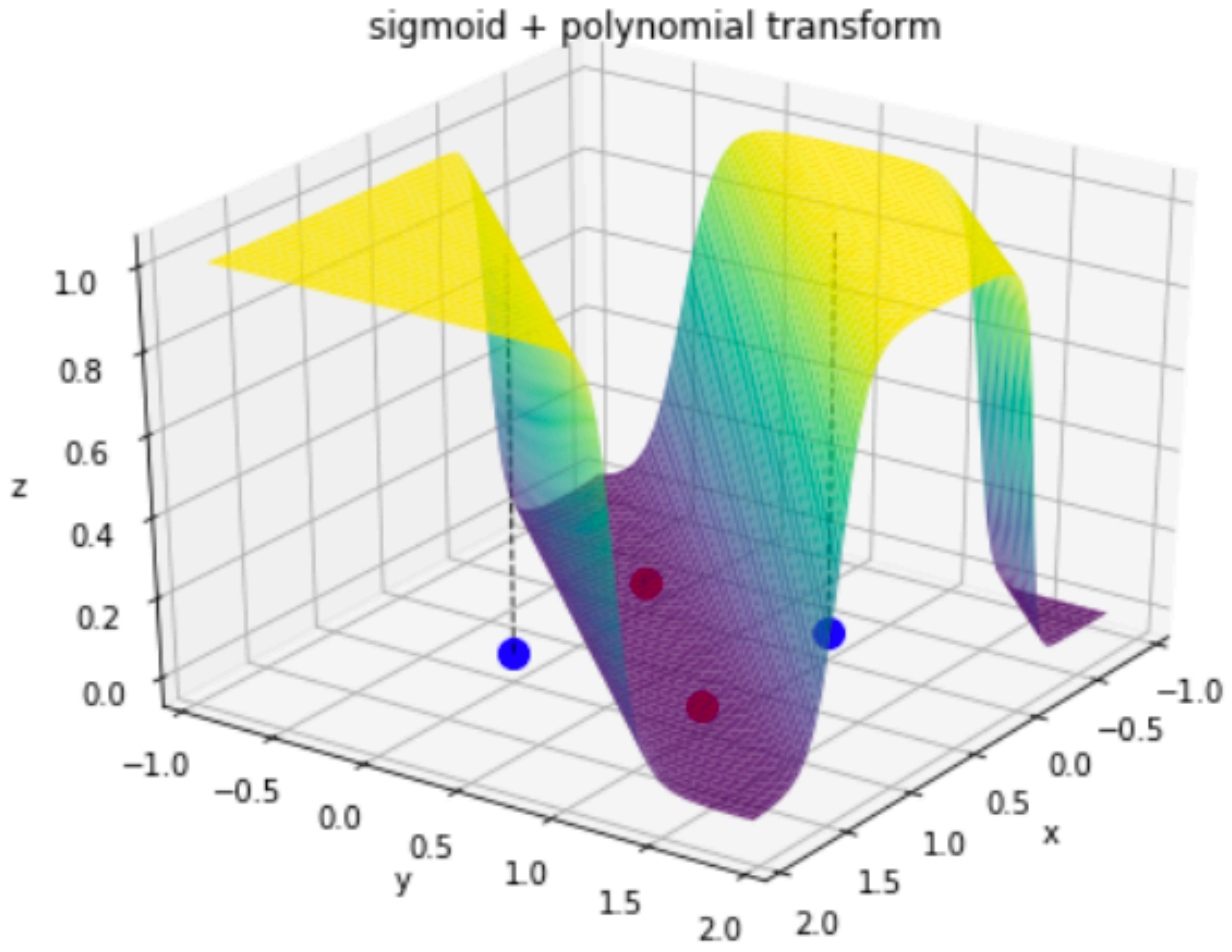
XOR

X1	X2	Y
0	0	0
1	0	1
0	1	1
1	1	1

?



		NAND	OR	AND
X1	X2	S1	S2	Y
0	0	1	0	0
1	0	1	1	1
0	1	1	1	1
1	1	0	1	0



01. 신경망

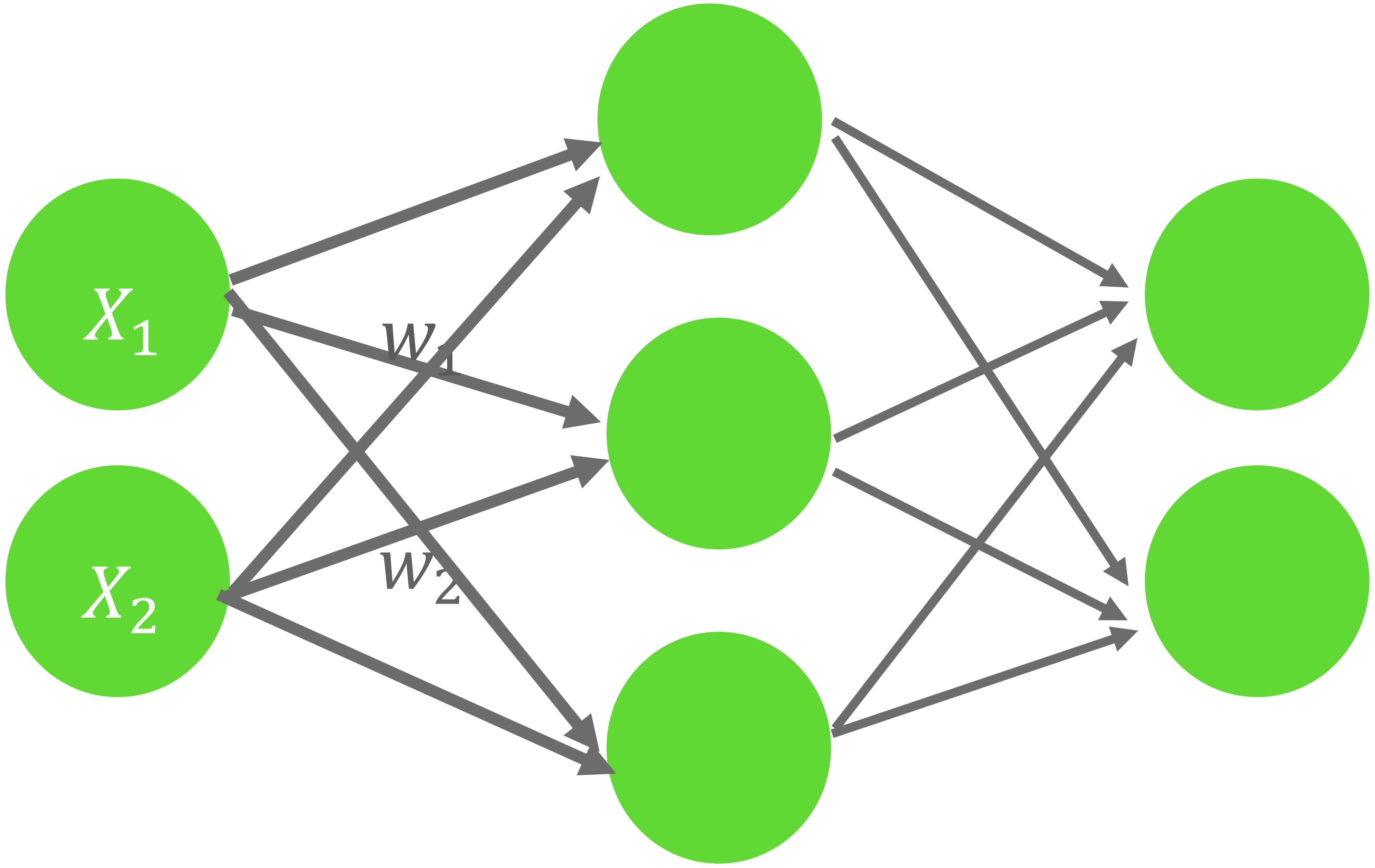
어떤 작업 **T**에 대한 컴퓨터 프로그램의 성능을 **P**로 측정했을 때 경험 **E**로 인해 성능이 향상 됐다면,
이 컴퓨터 프로그램은 작업 **T**와 성능 측정 **P**에 대해 경험 **E**로 학습한 것이다

Tom Mitchell, 1997

Type of Machine Learning

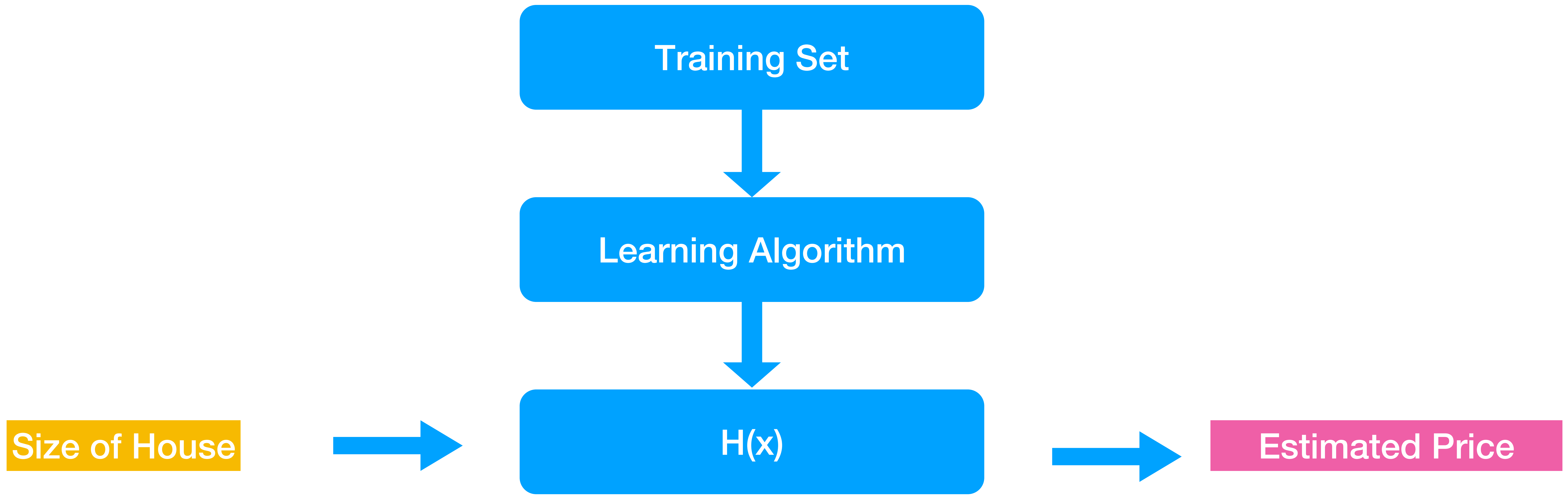
지도 / 비지도 / 준지도 학습
온라인 학습 / 배치 학습
사례 기반 학습 / 모델 기반 학습

$$\eta$$



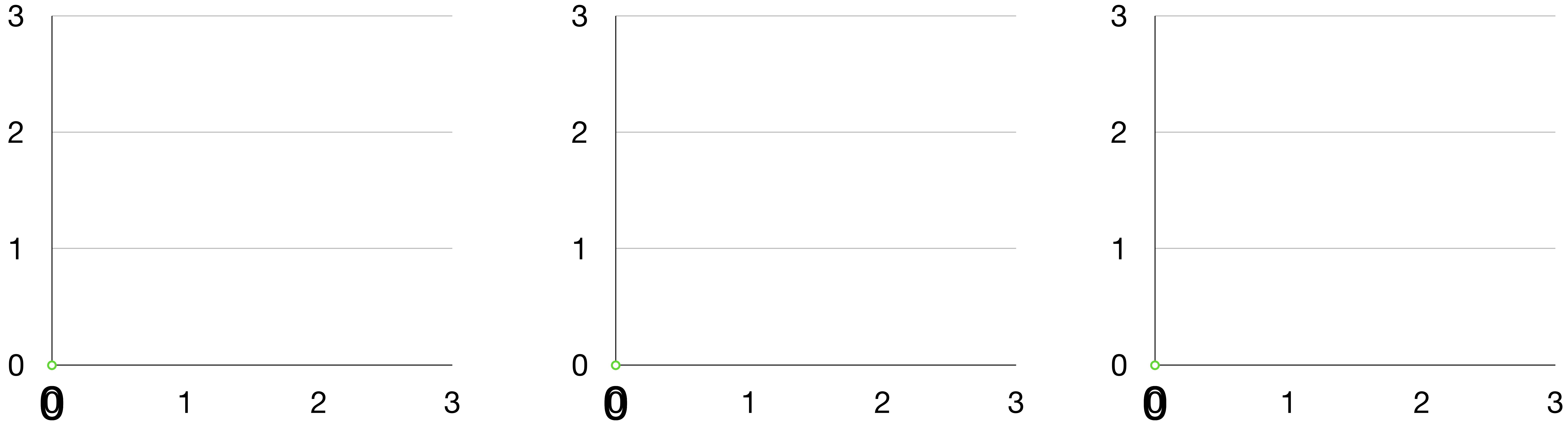
X_1
X_2
w_1
w_2

02. Linear Regression

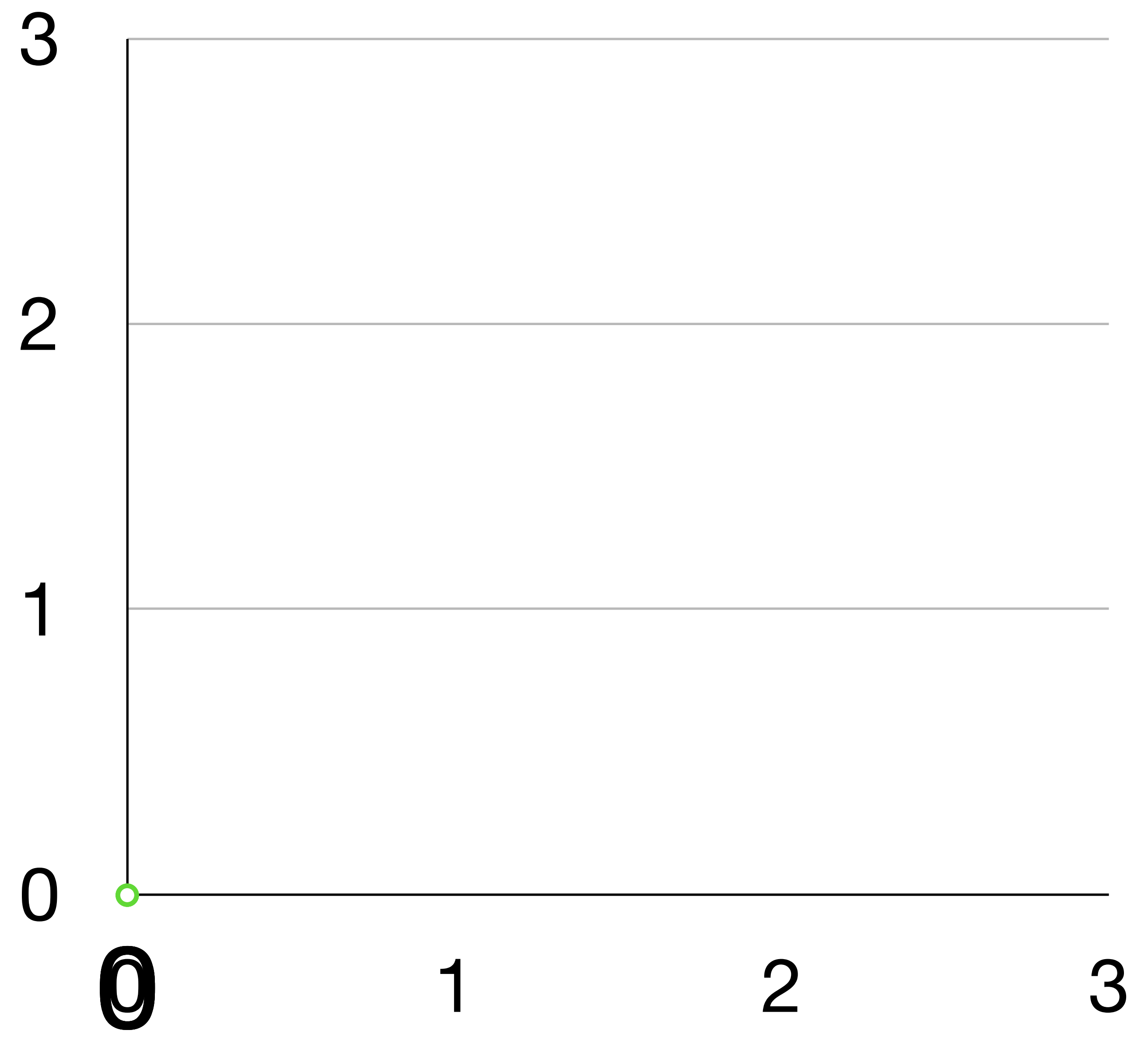


$$h(x) = \theta_0 + \theta_1 x$$

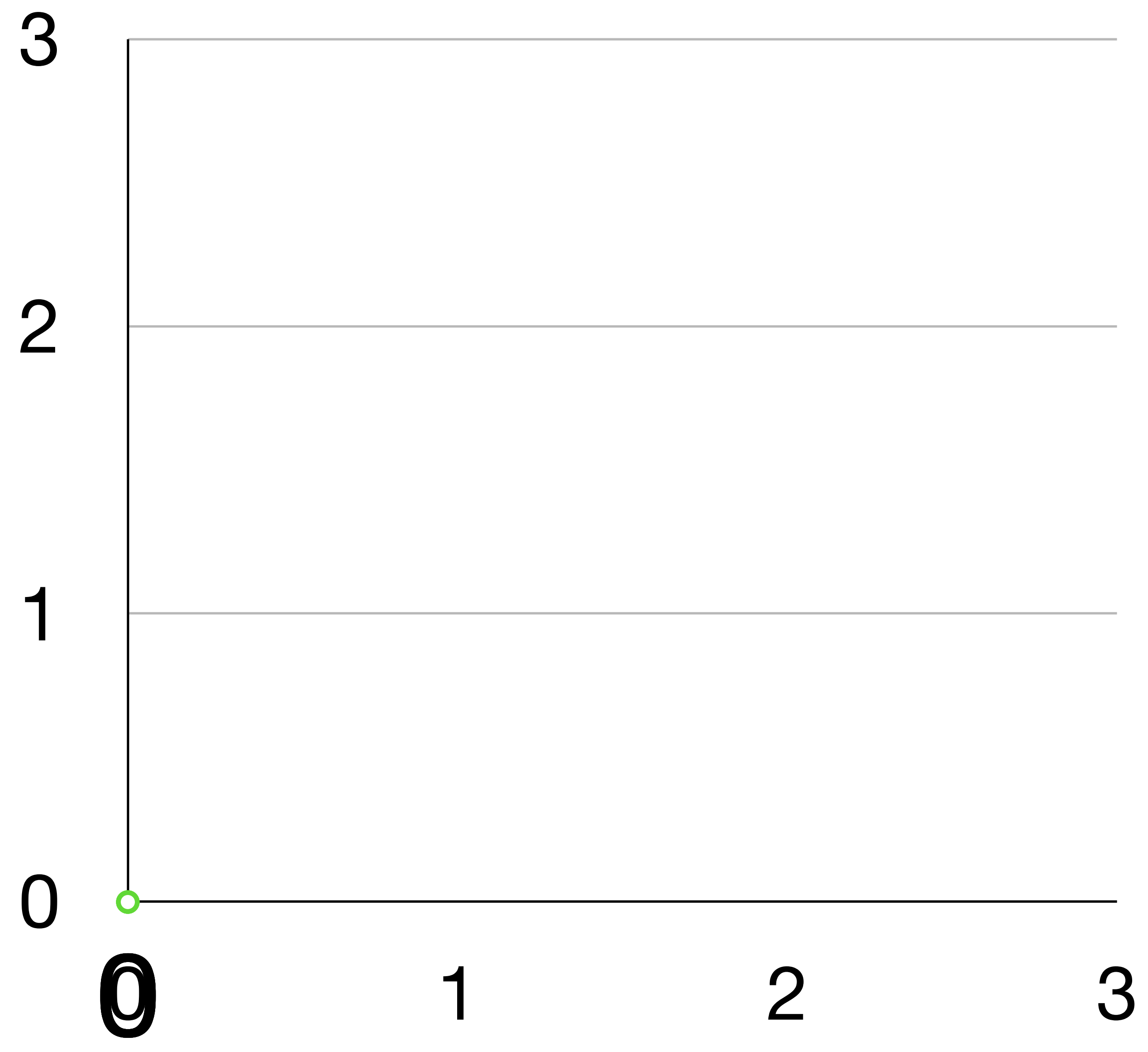
$$h(x) = \theta_0 + \theta_1 x$$



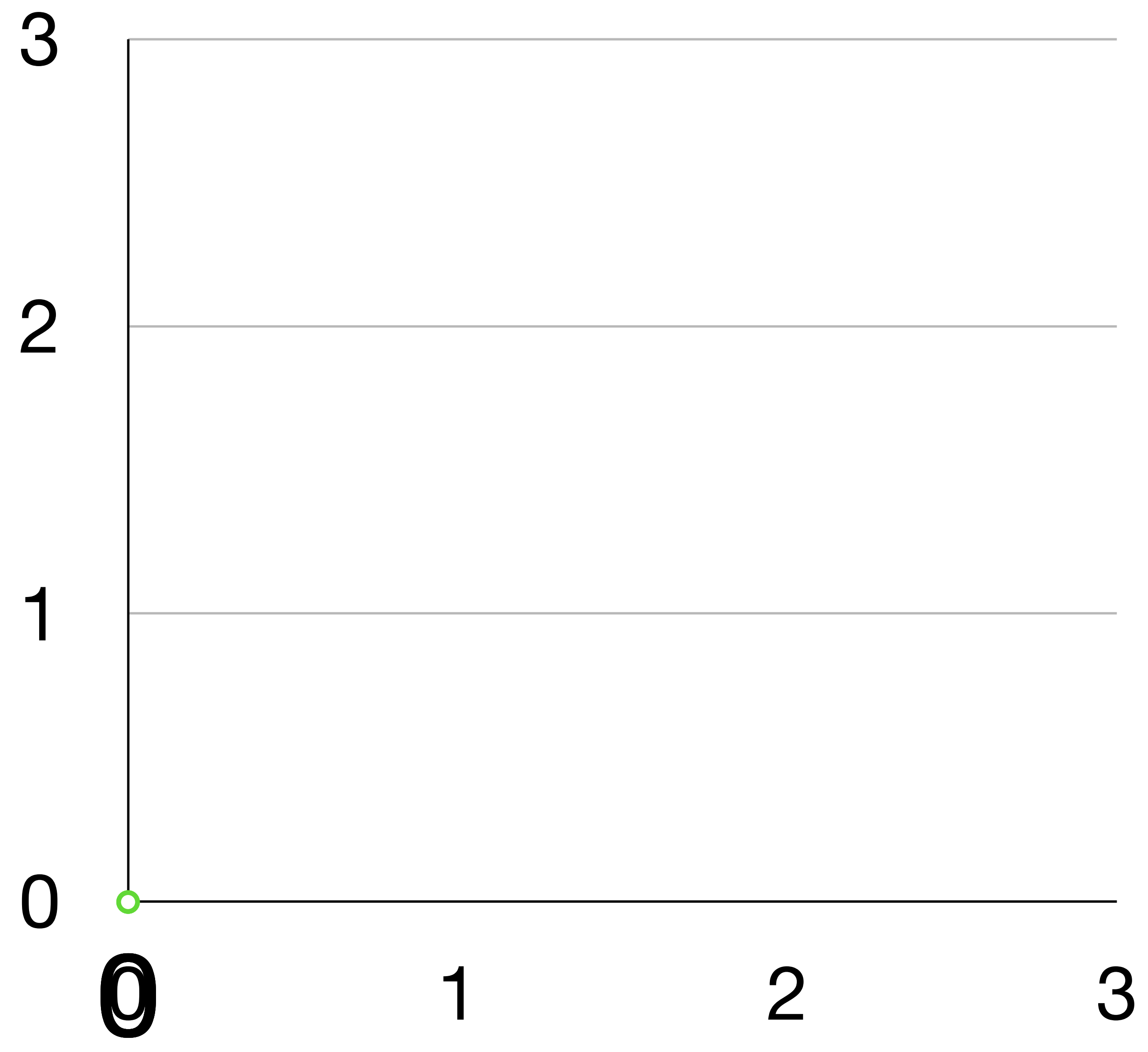
$$h(x) = \theta_0 + \theta_1 x$$



$\theta_0 = 1.5$
$\theta_1 = 0$

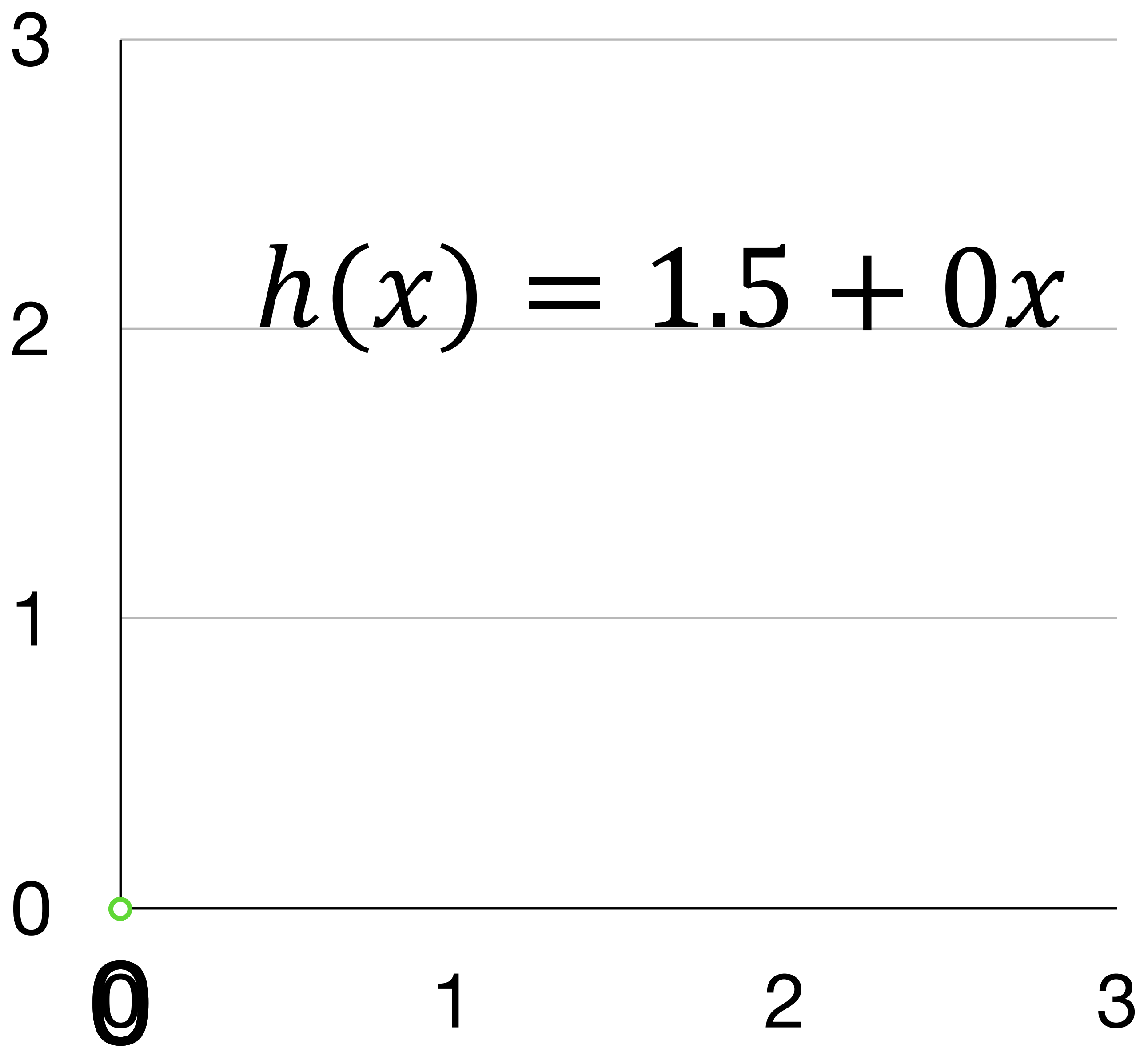


$\theta_0 = 0$
$\theta_1 = 0.5$

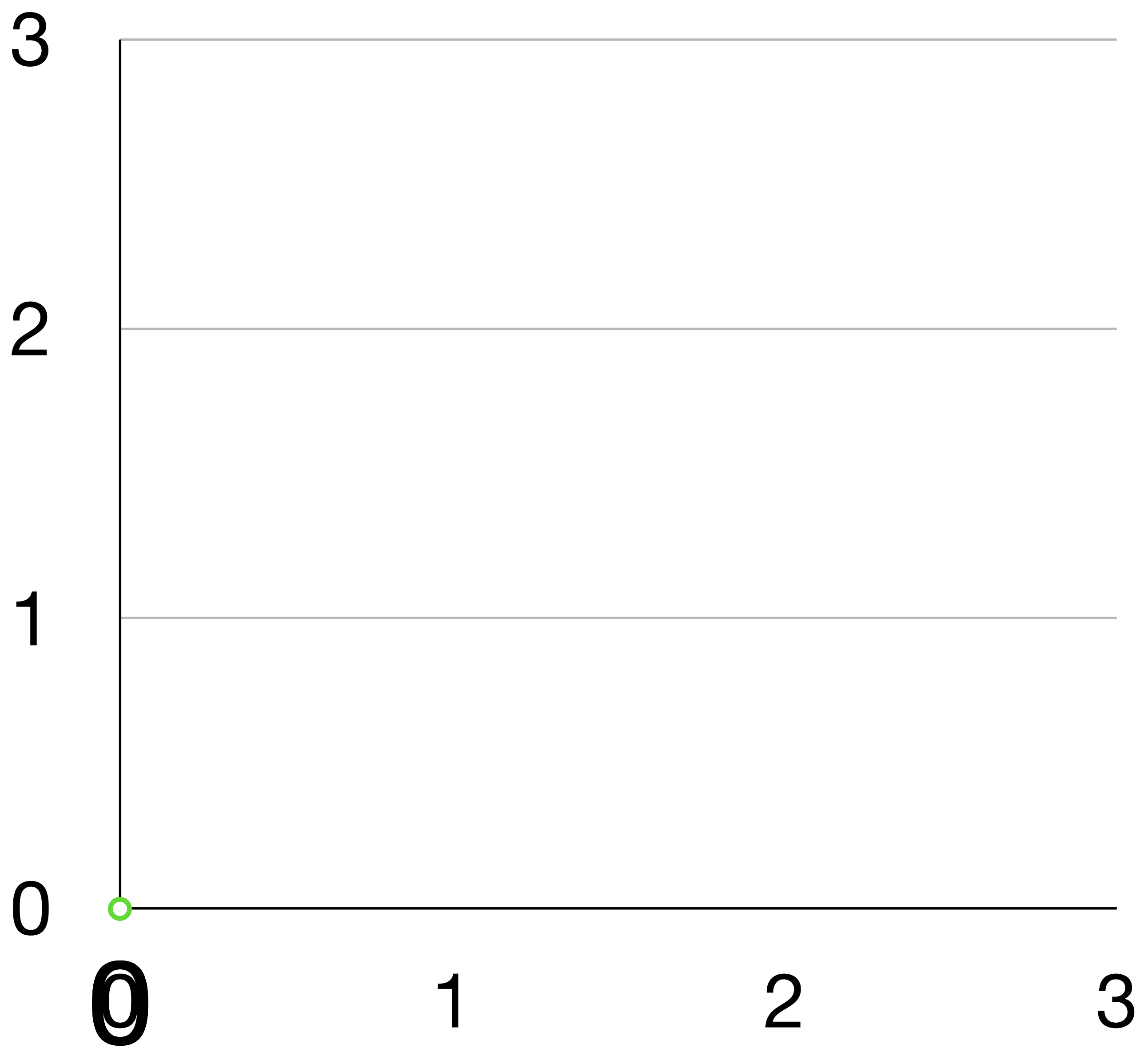


$\theta_0 = 1$
$\theta_1 = 0.5$

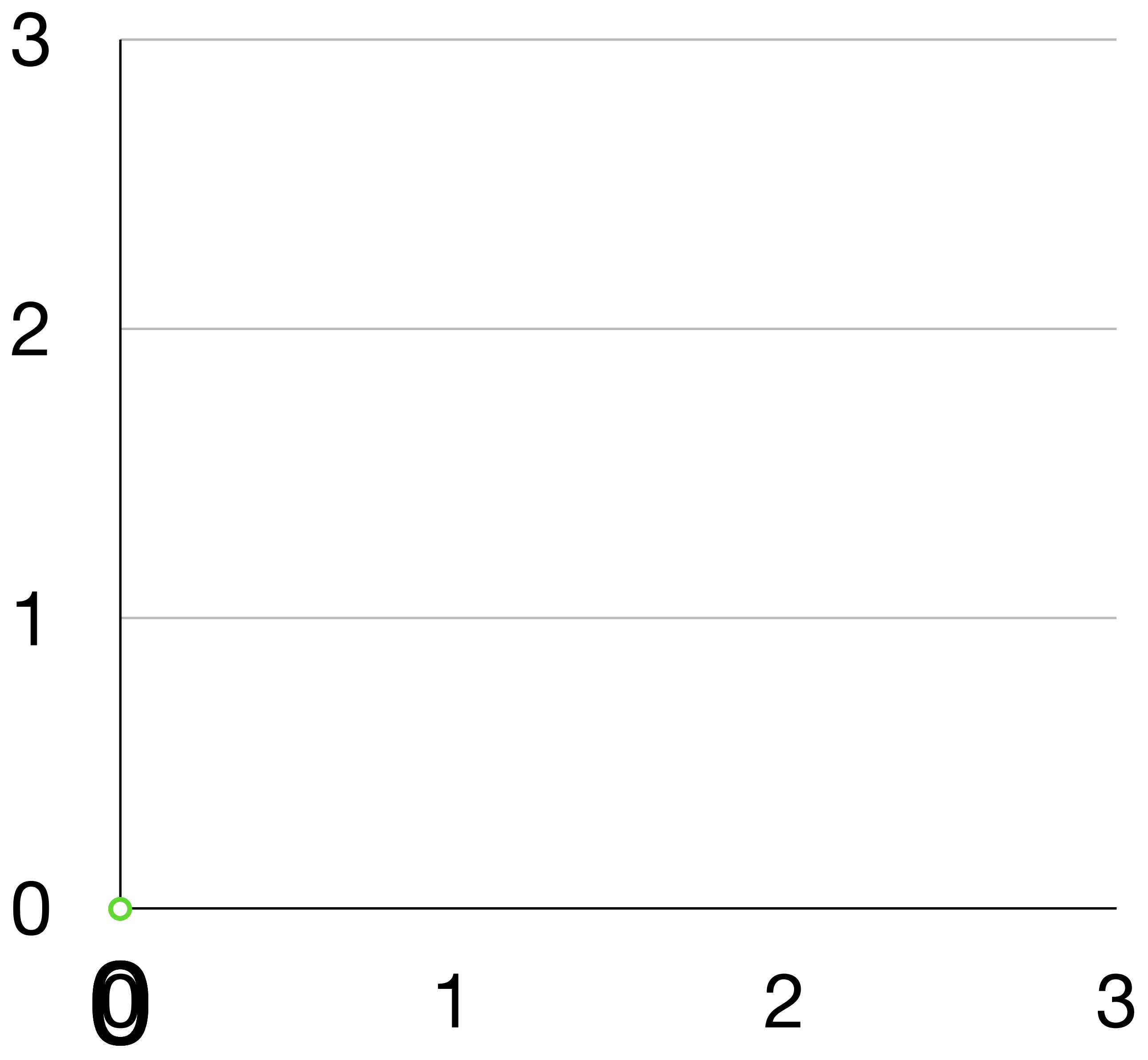
$$h(x) = \theta_0 + \theta_1 x$$



$\theta_0 = 1.5$
$\theta_1 = 0$

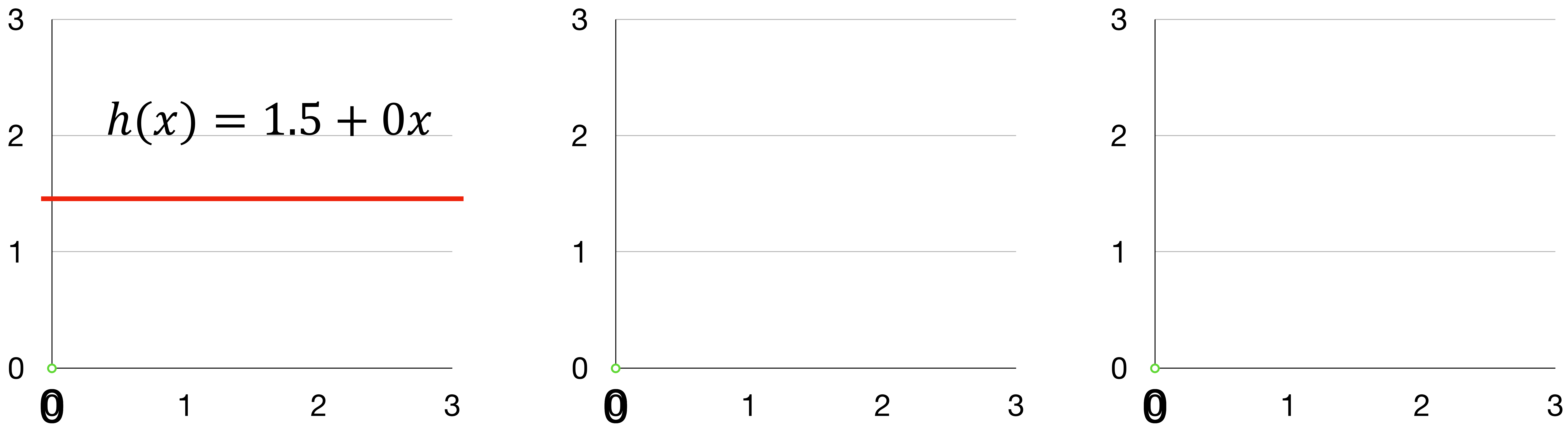


$\theta_0 = 0$
$\theta_1 = 0.5$



$\theta_0 = 1$
$\theta_1 = 0.5$

$$h(x) = \theta_0 + \theta_1 x$$

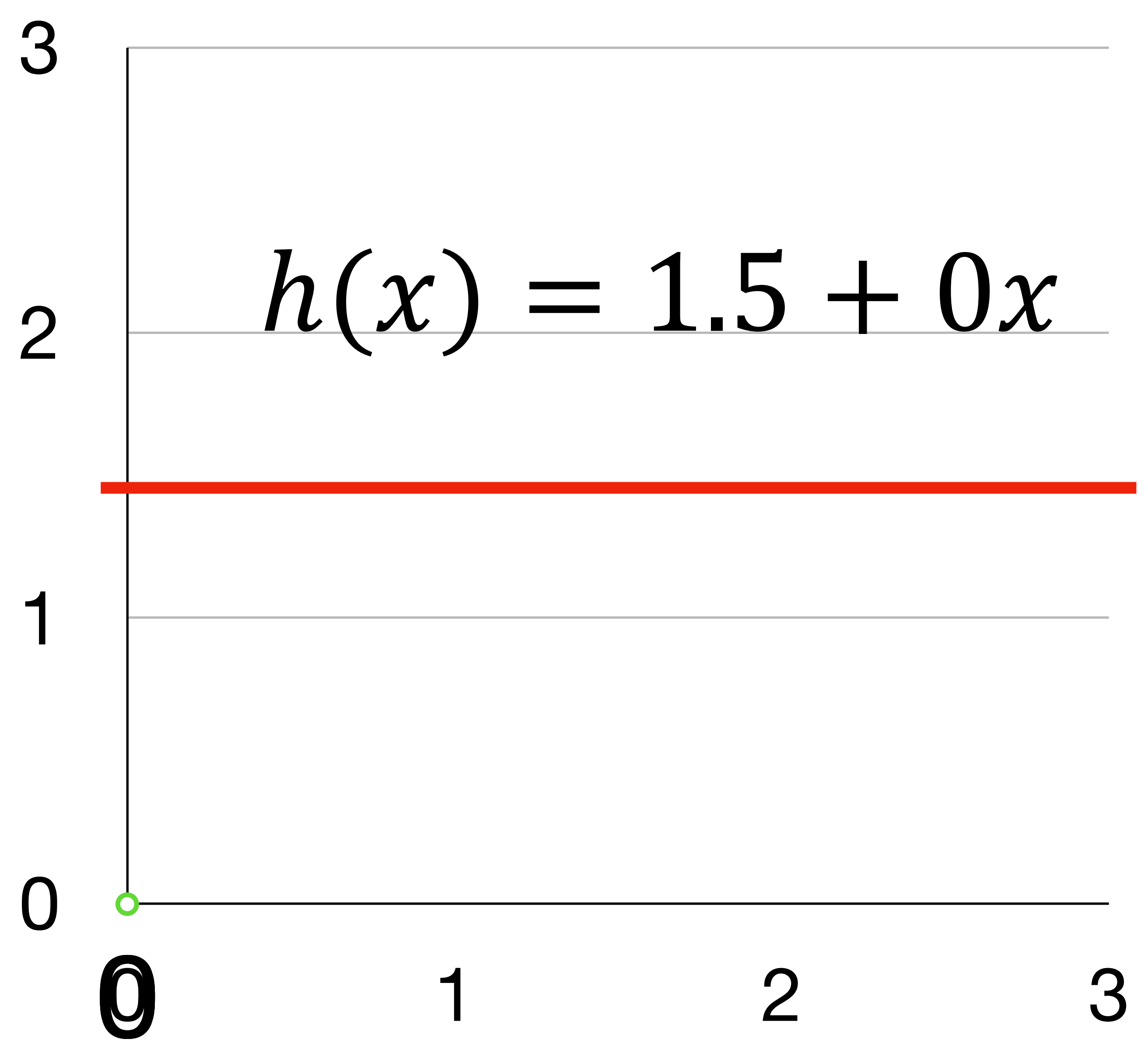


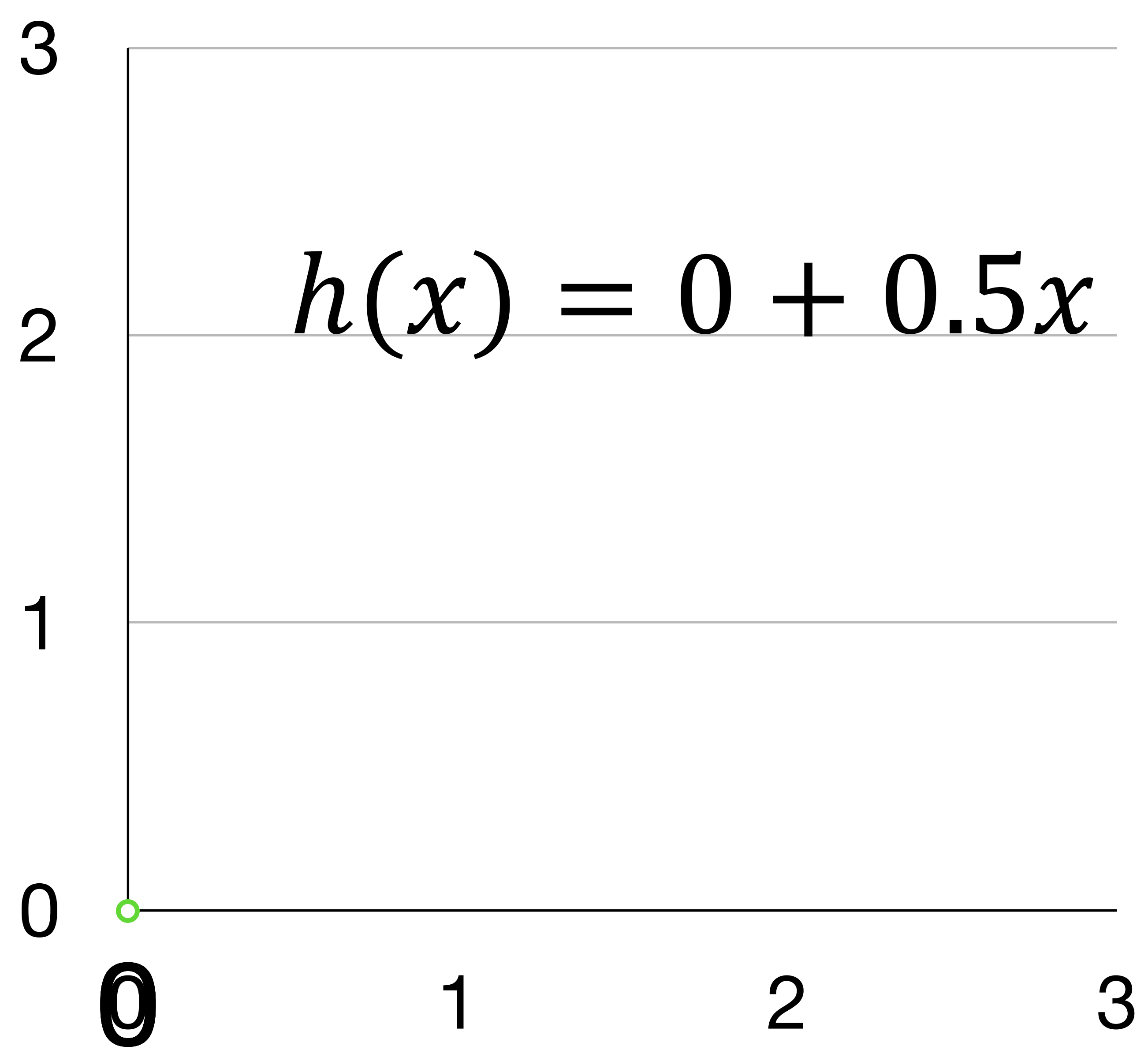
$\theta_0 = 1.5$
$\theta_1 = 0$

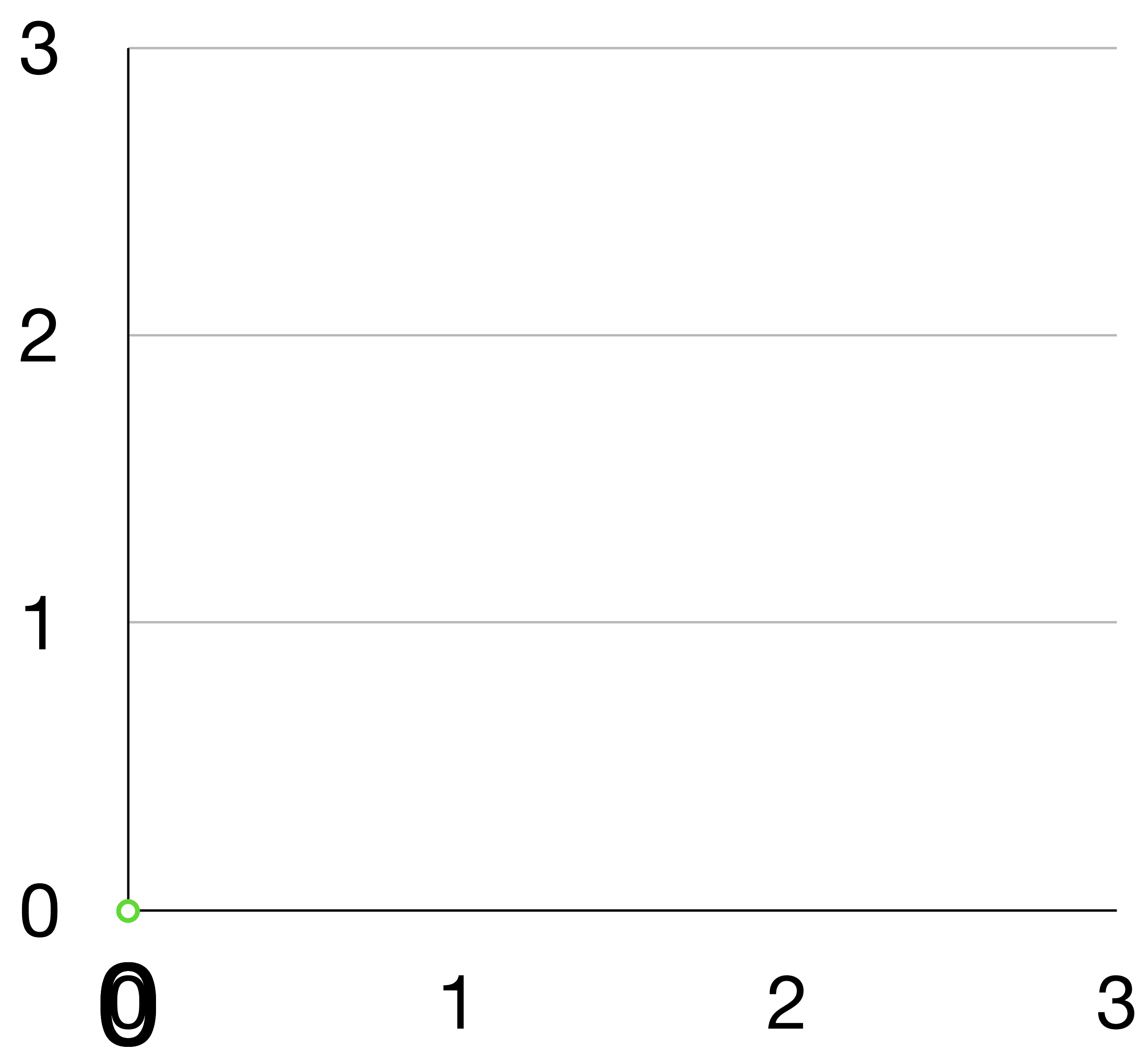
$\theta_0 = 0$
$\theta_1 = 0.5$

$\theta_0 = 1$
$\theta_1 = 0.5$

$$h(x) = \theta_0 + \theta_1 x$$





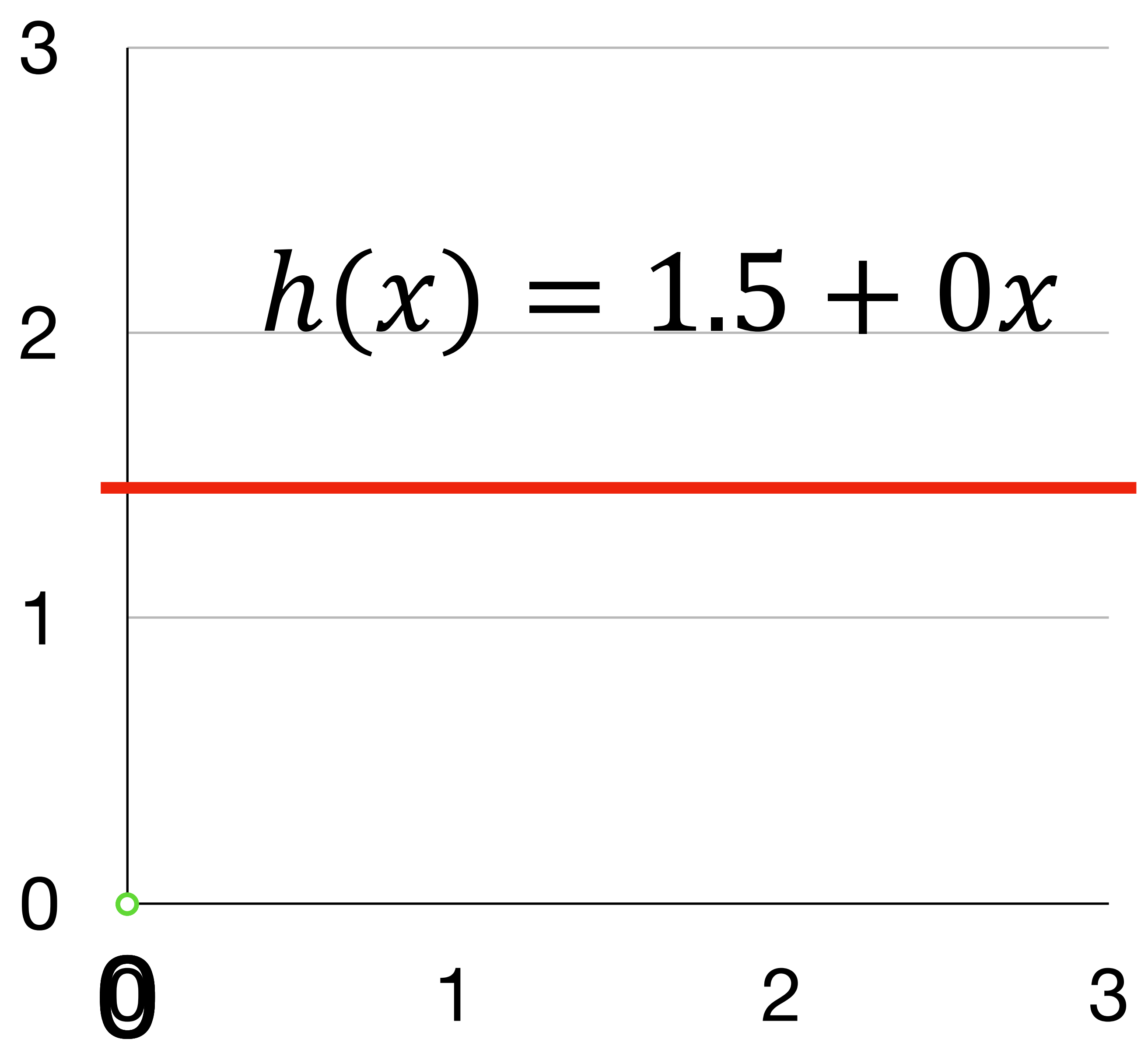


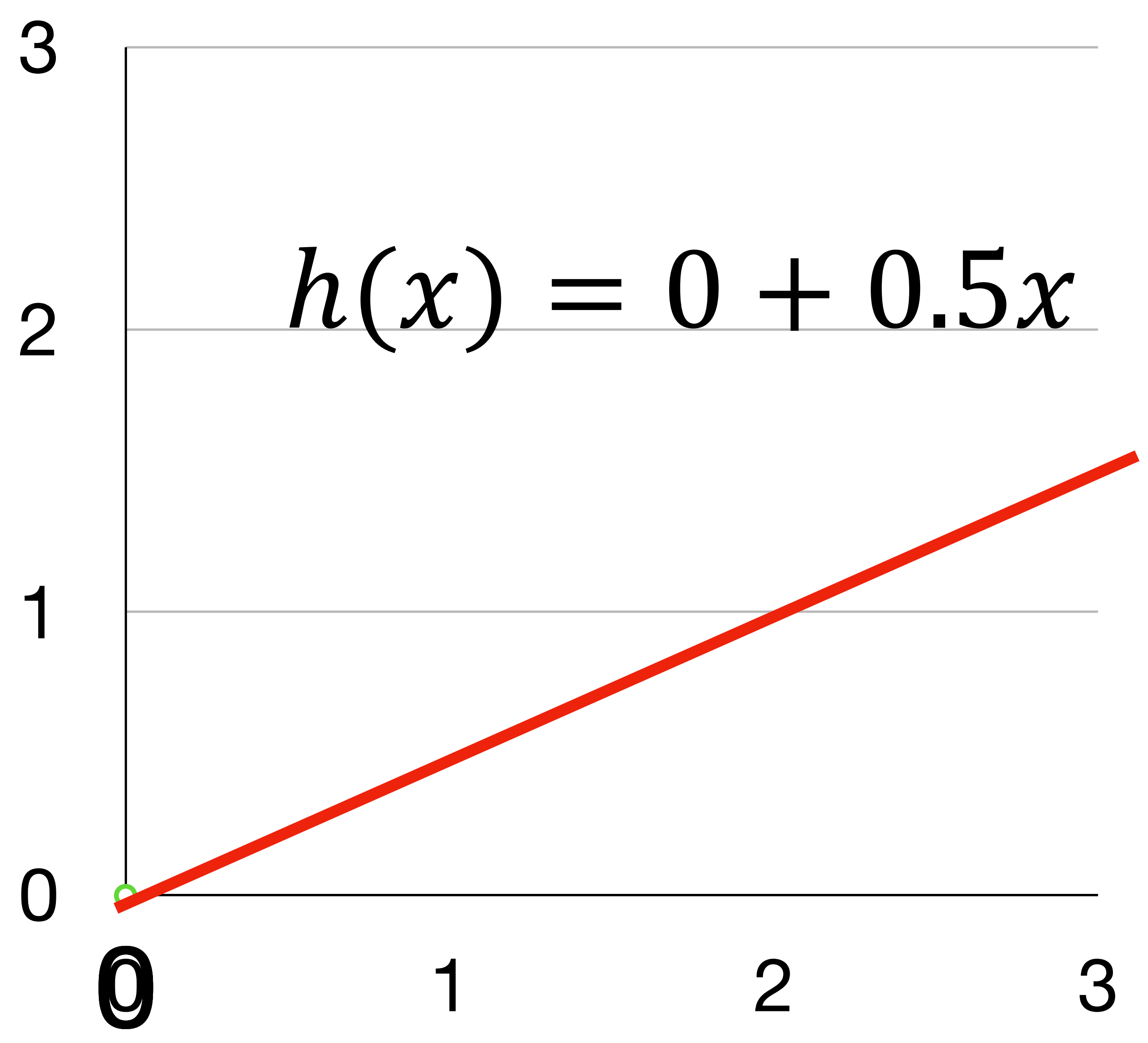
$\theta_0 = 1.5$
$\theta_1 = 0$

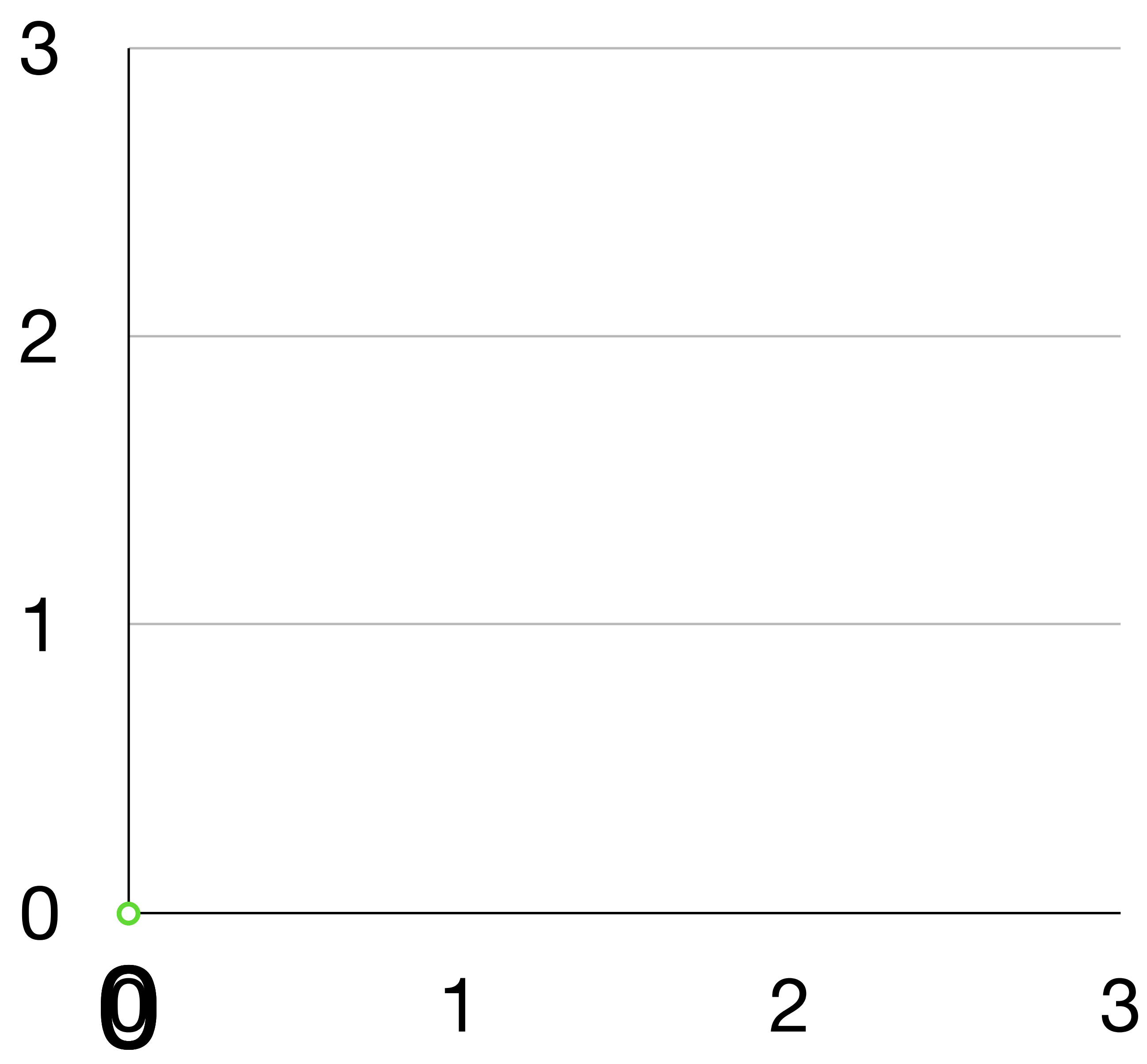
$\theta_0 = 0$
$\theta_1 = 0.5$

$\theta_0 = 1$
$\theta_1 = 0.5$

$$h(x) = \theta_0 + \theta_1 x$$





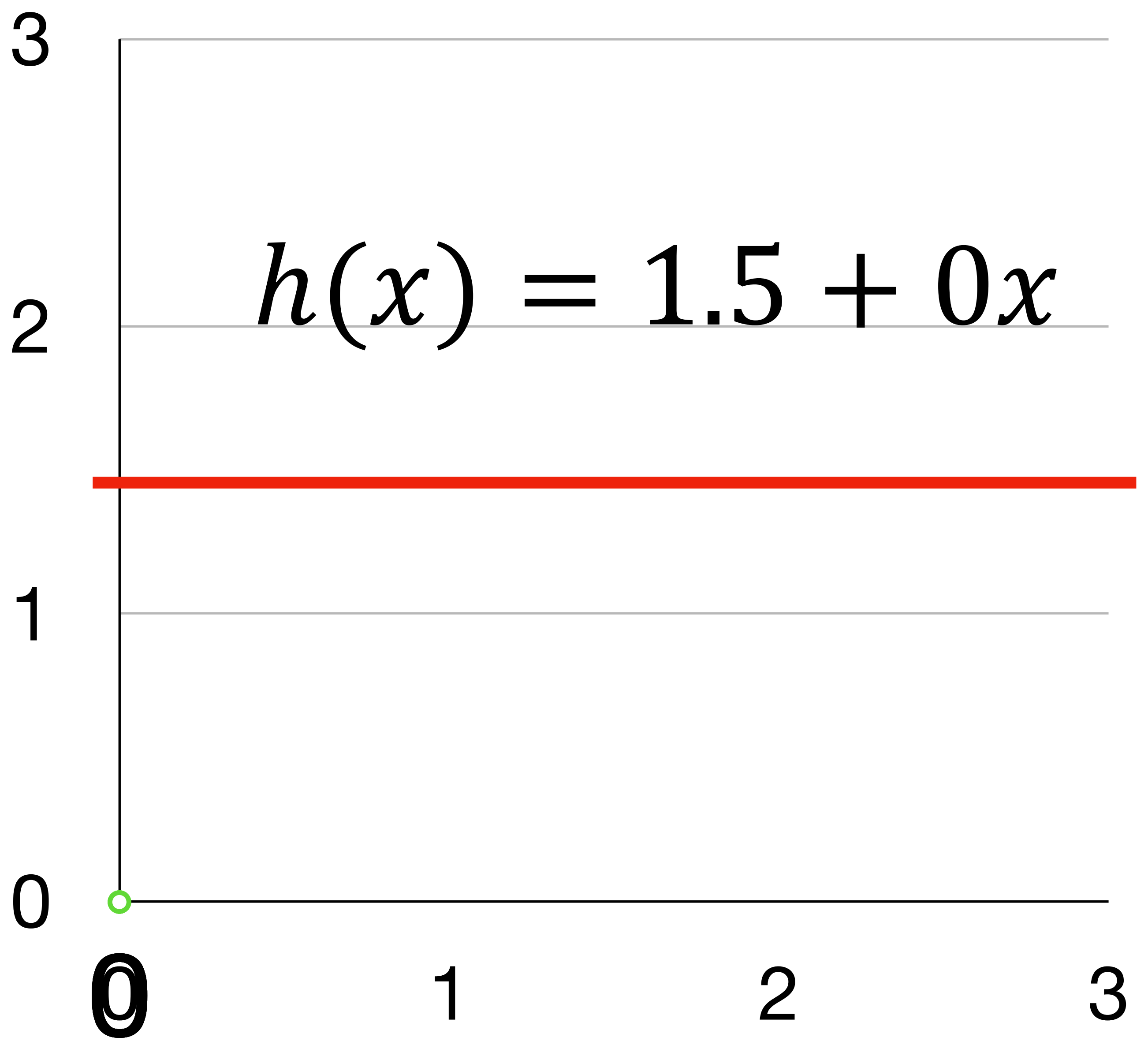


$\theta_0 = 1.5$
$\theta_1 = 0$

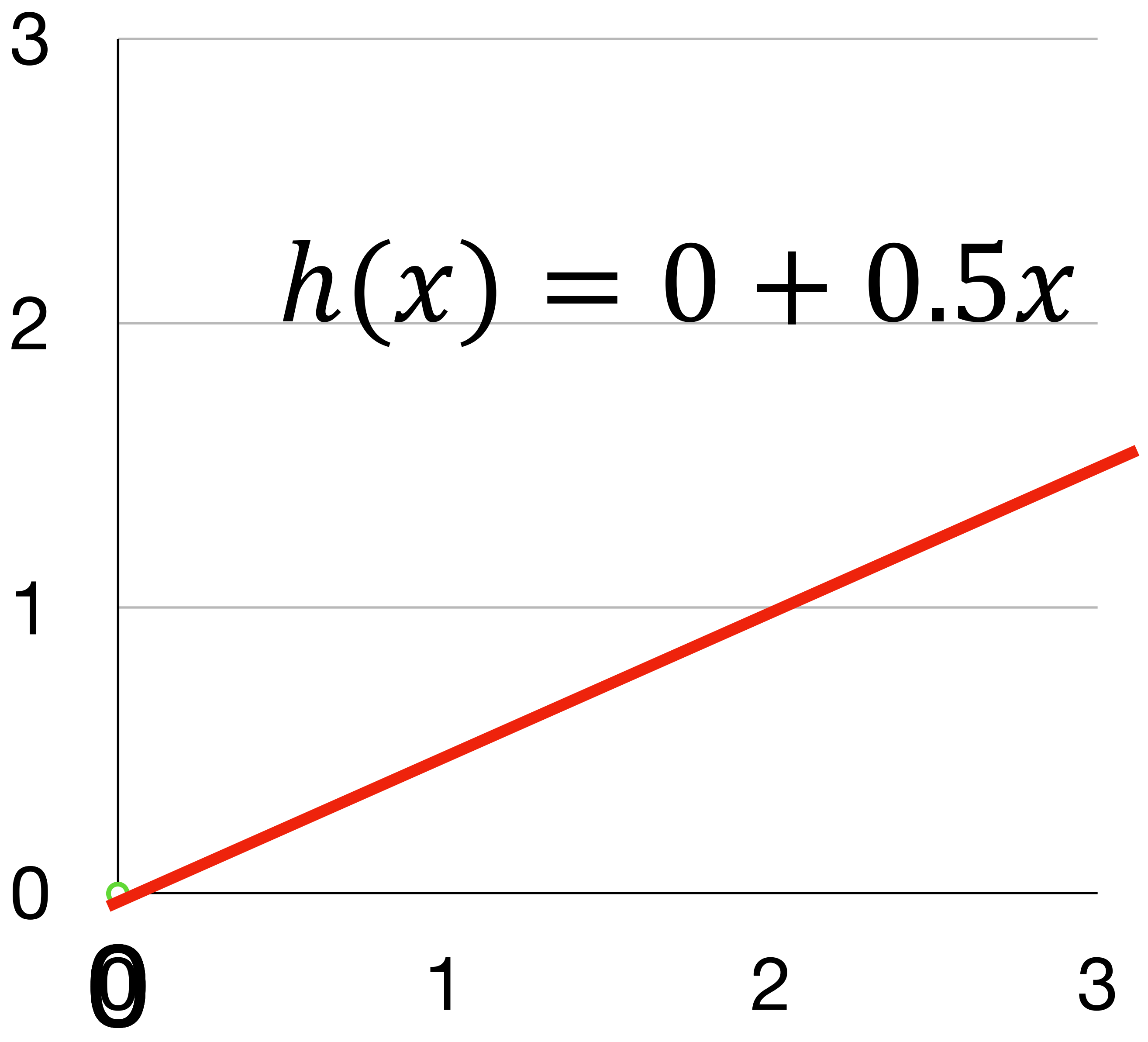
$\theta_0 = 0$
$\theta_1 = 0.5$

$\theta_0 = 1$
$\theta_1 = 0.5$

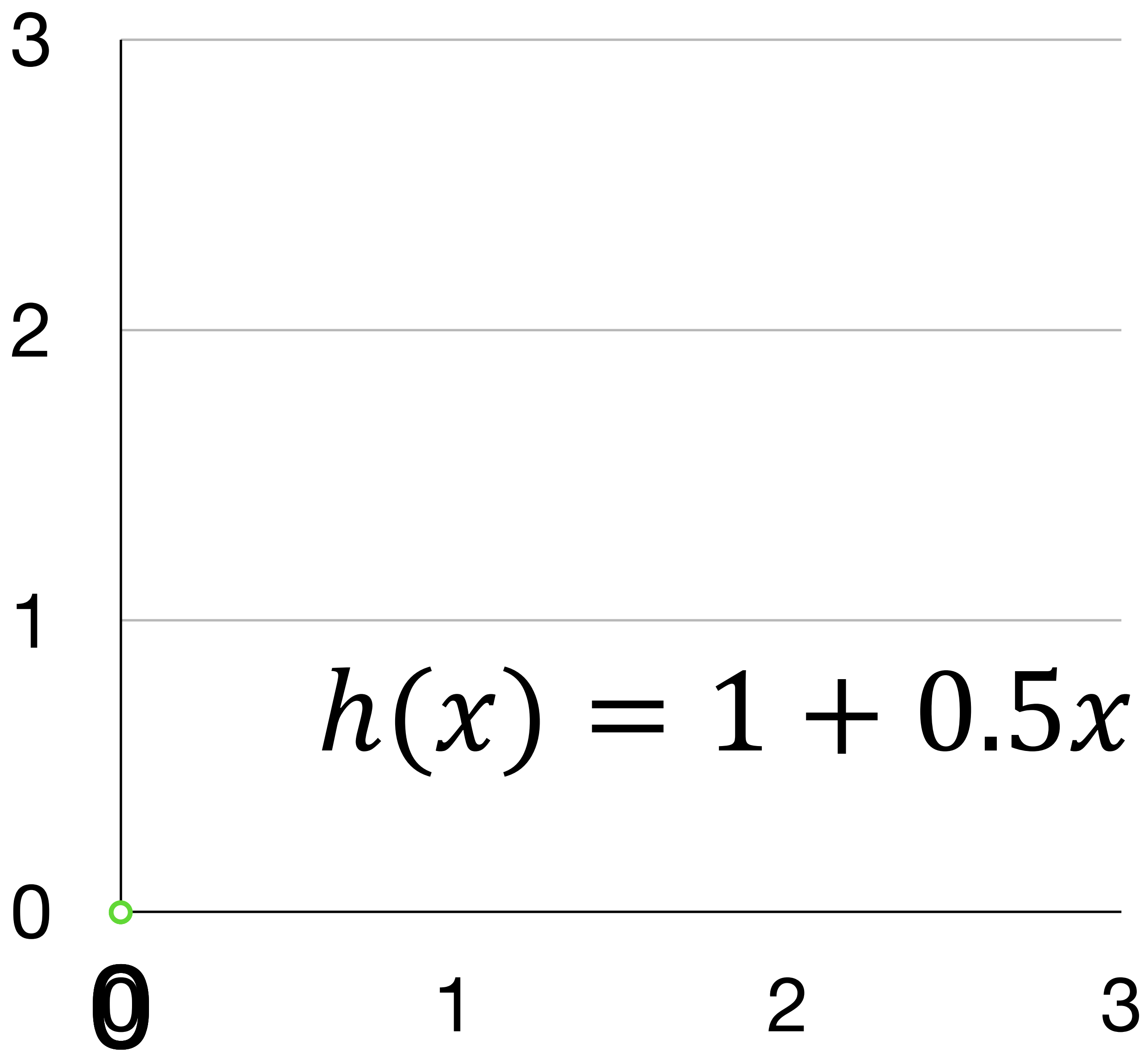
$$h(x) = \theta_0 + \theta_1 x$$



$\theta_0 = 1.5$
$\theta_1 = 0$

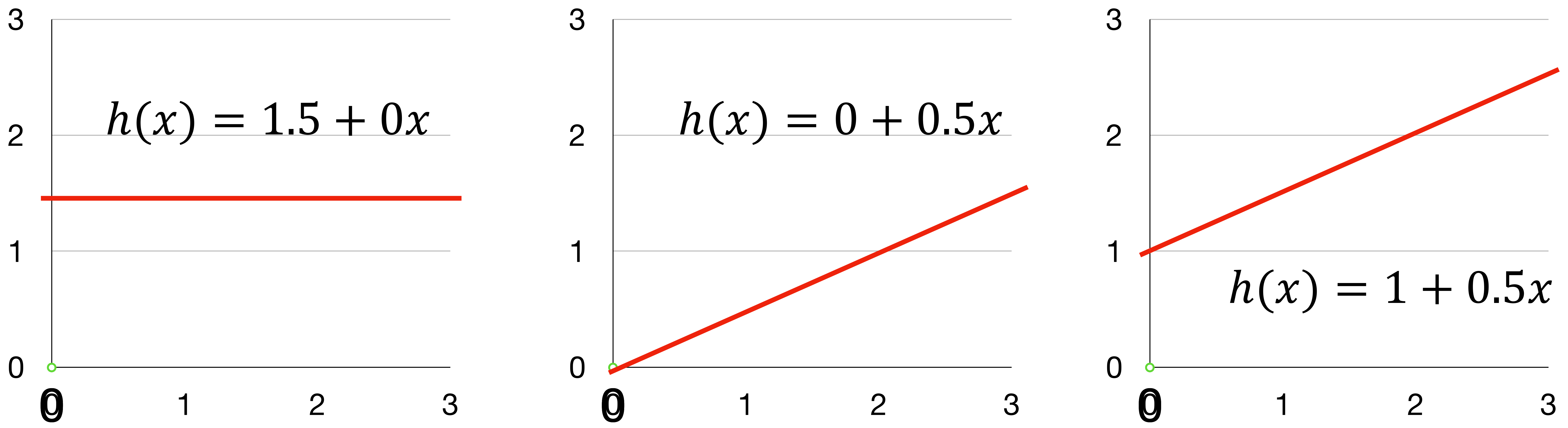


$\theta_0 = 0$
$\theta_1 = 0.5$



$\theta_0 = 1$
$\theta_1 = 0.5$

$$h(x) = \theta_0 + \theta_1 x$$

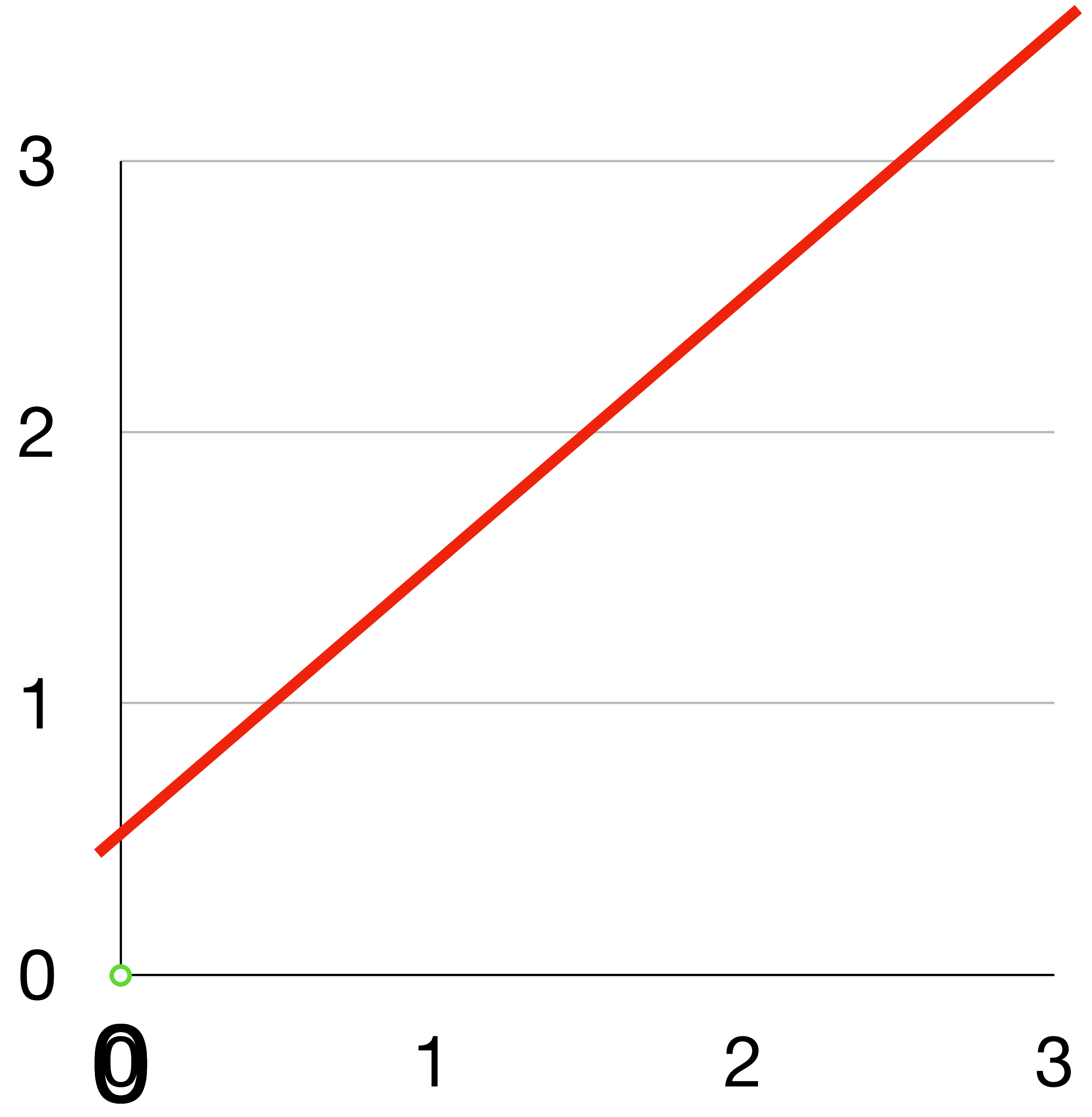


$\theta_0 = 1.5$
$\theta_1 = 0$

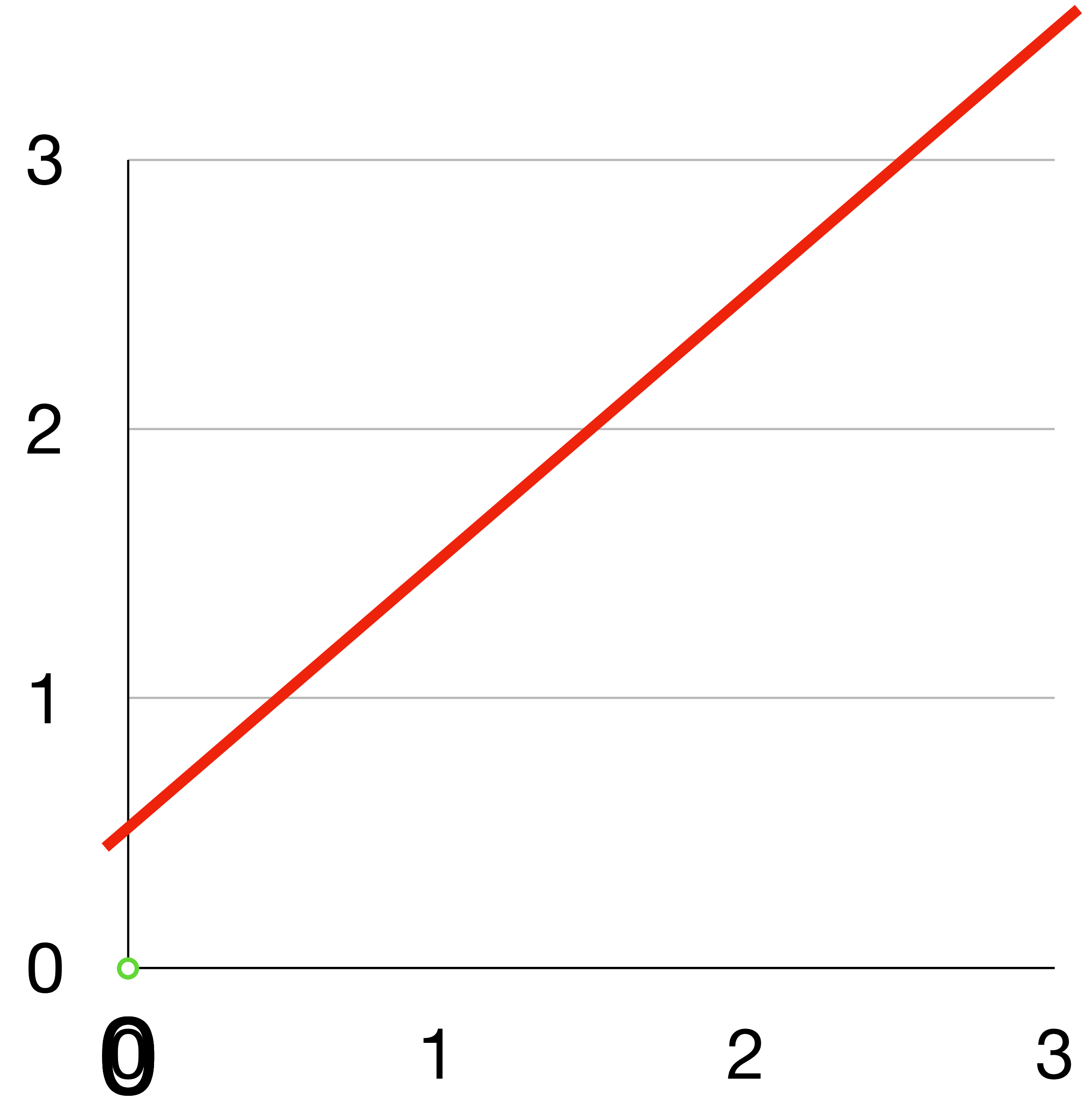
$\theta_0 = 0$
$\theta_1 = 0.5$

$\theta_0 = 1$
$\theta_1 = 0.5$

$$h(x) = \theta_0 + \theta_1 x$$

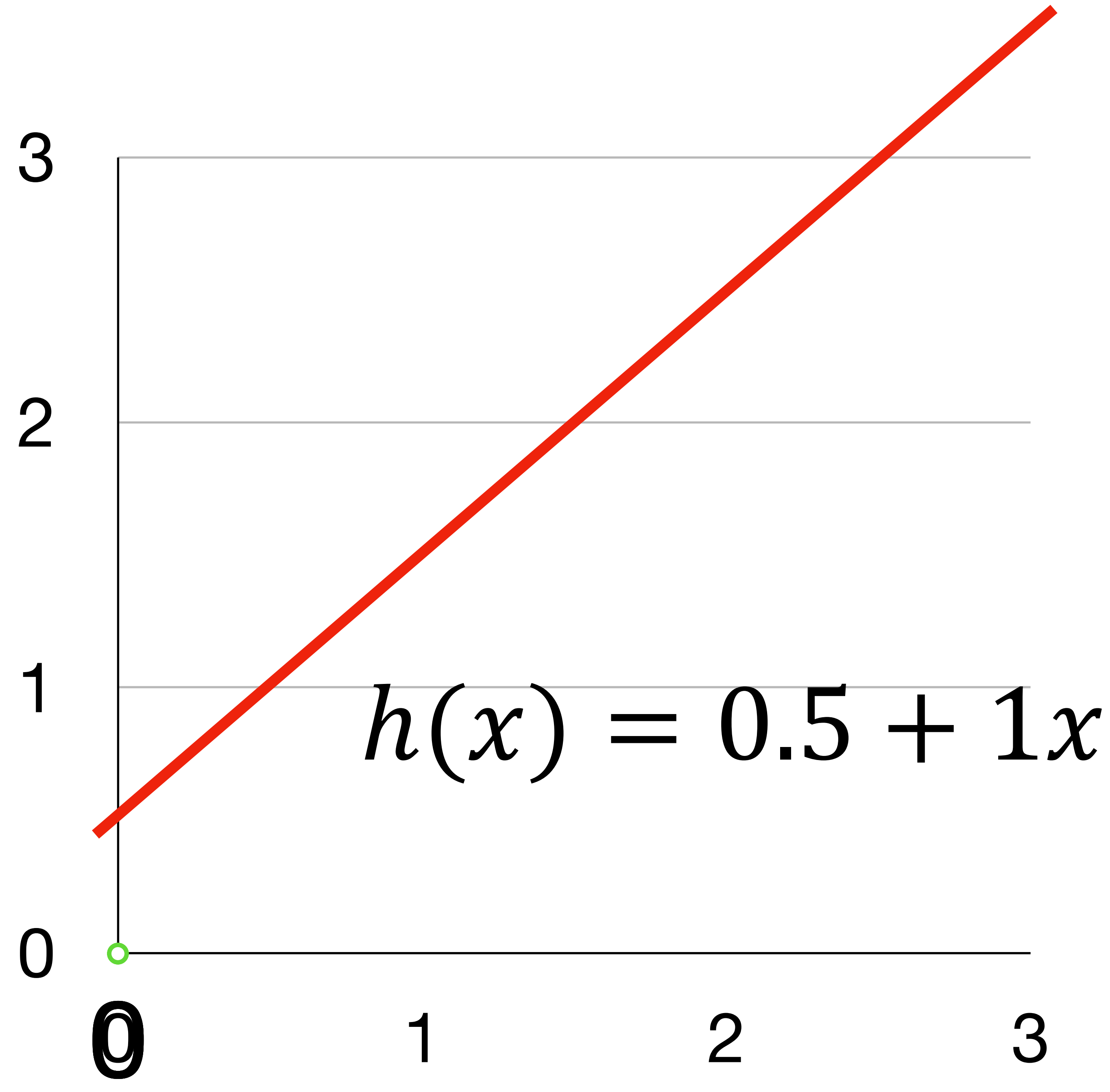


$$h(x) = \theta_0 + \theta_1 x$$

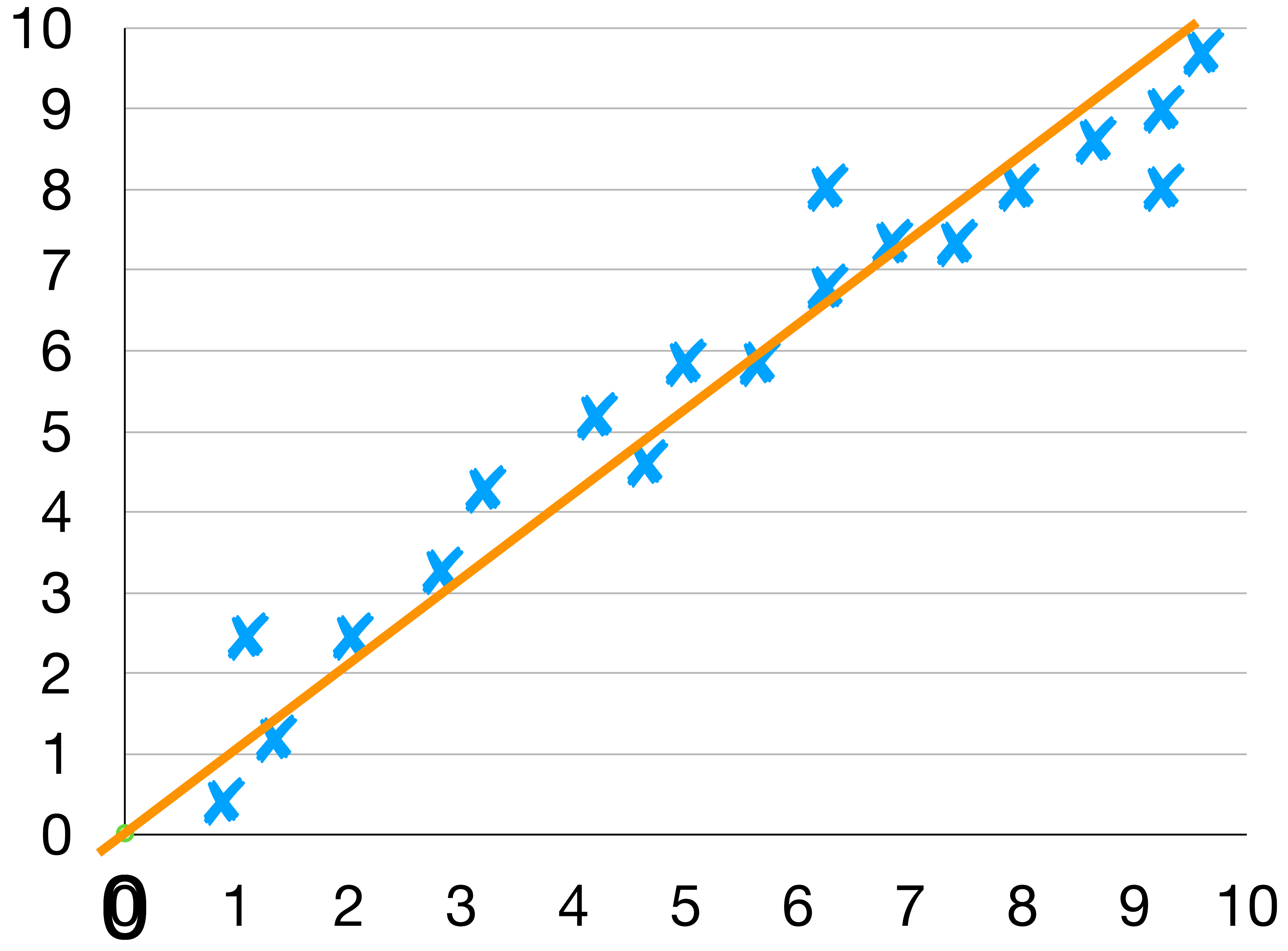


$$\theta_0 = 0.5$$
$$\theta_1 = 1$$

$$h(x) = \theta_0 + \theta_1 x$$



$$\theta_0 = 0.5$$
$$\theta_1 = 1$$

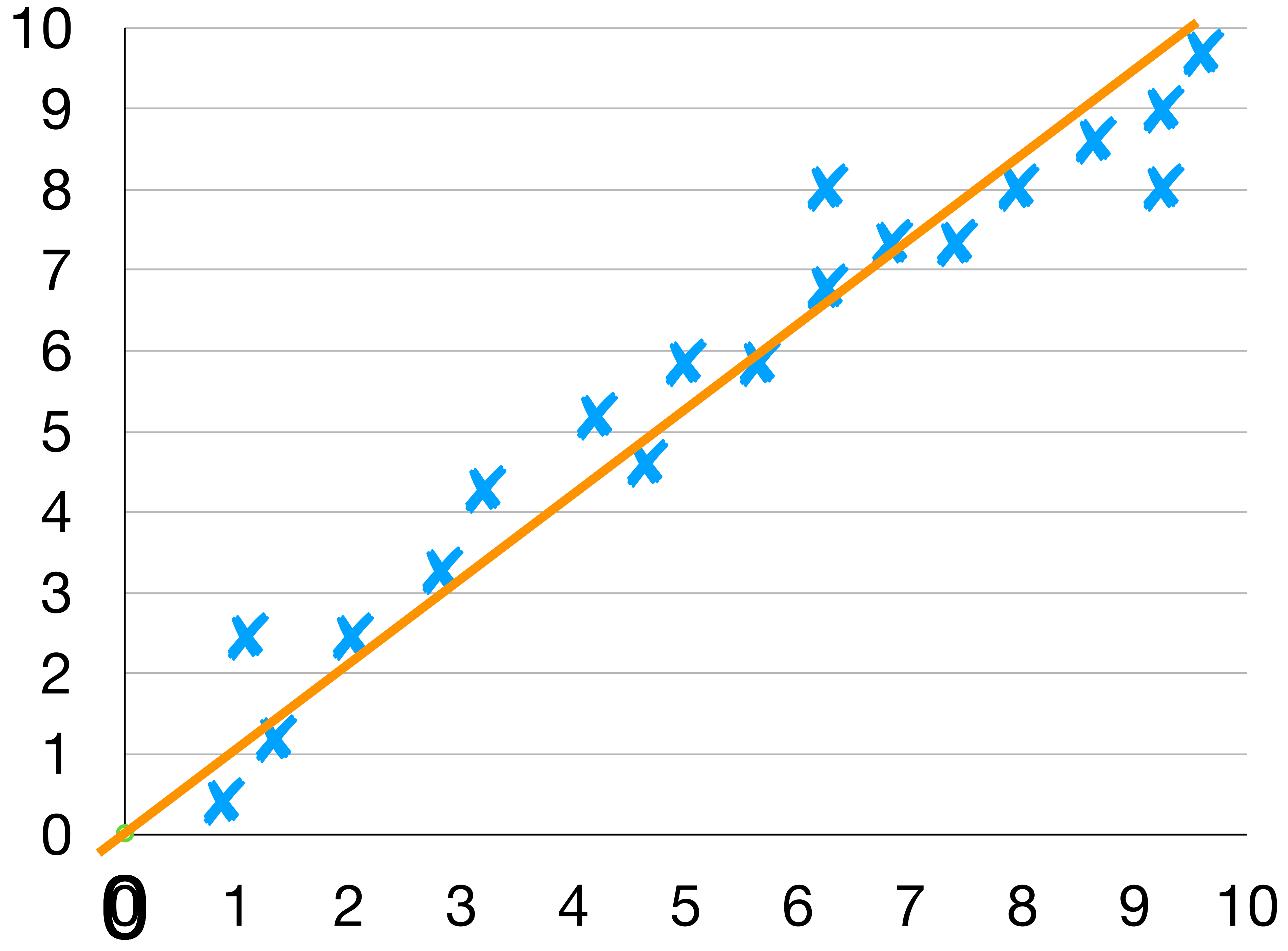


Hypothesis

$$h(x) = \theta_0 + \theta_1 x$$

Cost Function (Square error function)

$$\underbrace{J(\theta_0, \theta_1)}_{\text{Minimize}} = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$



Hypothesis

$$h(x) = \theta_0 + \theta_1 x$$

Cost Function(Square error function)

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}$$

Minimize

Hypothesis

$$h(x) = \theta_0 + \theta_1 x$$

Parameters

$$\theta_0, \theta_1$$

Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})$$

Goal

Minimize $J(\theta_0, \theta_1)$

Simplified

Hypothesis

$$h(x) = \theta_1 x$$

Parameters

$$\theta_0 = 0, \theta_1 = 1$$

Cost Function

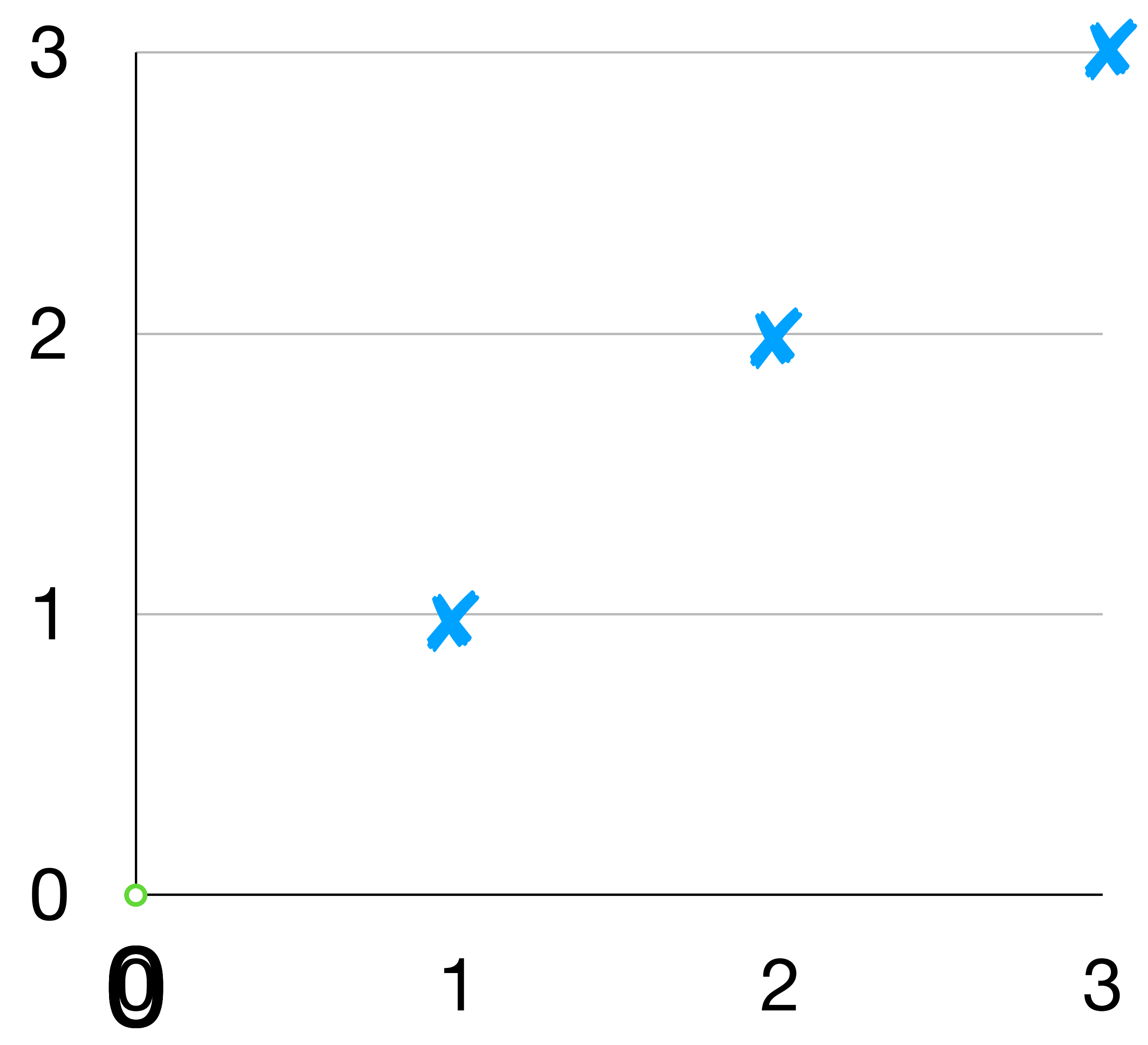
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (\theta_1 x^{(i)} - y^{(i)})^2$$

Goal

Minimize $J(\theta_1)$

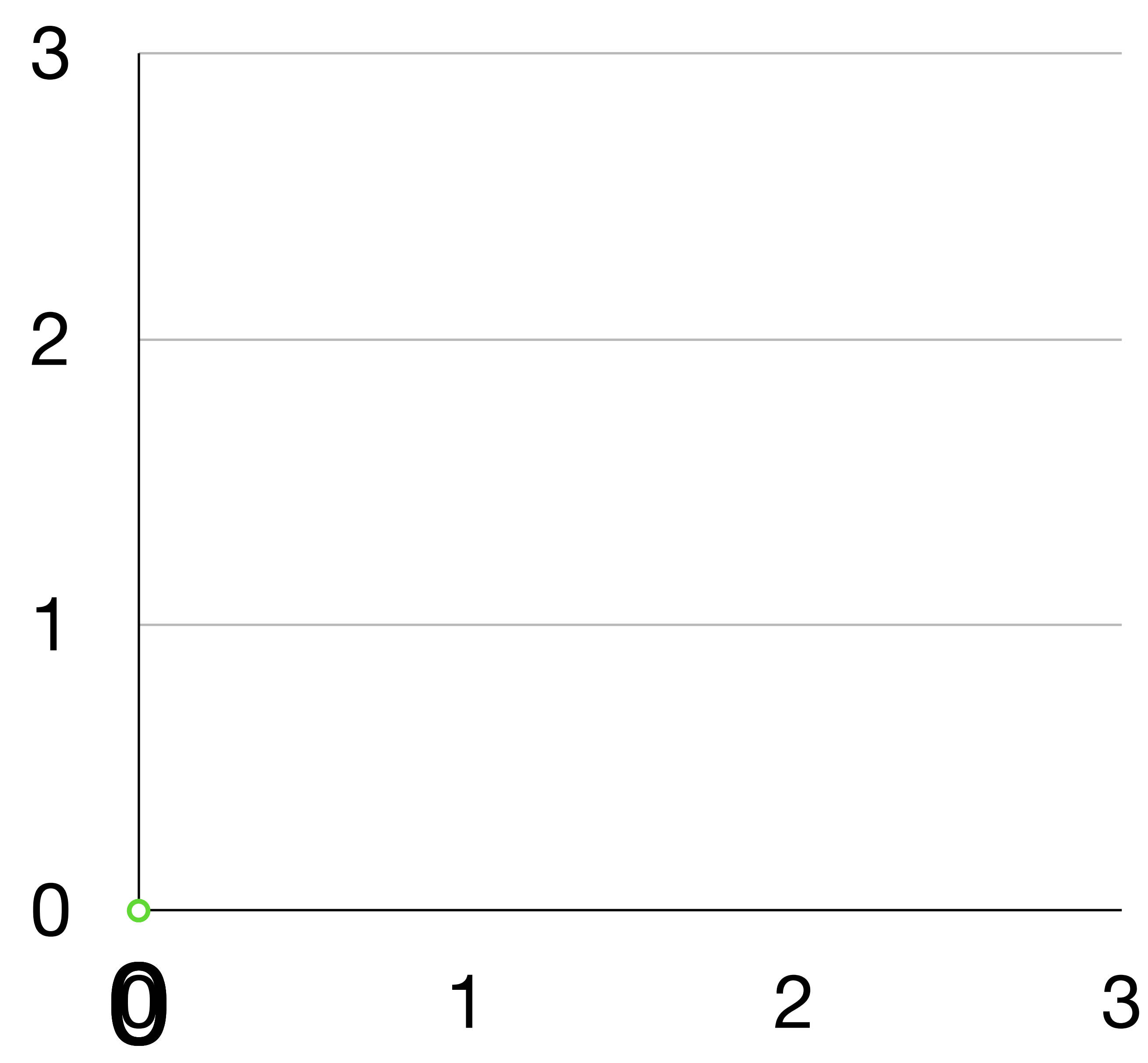
$$h_\theta(x)$$

for fixed θ_1, this is a function of x



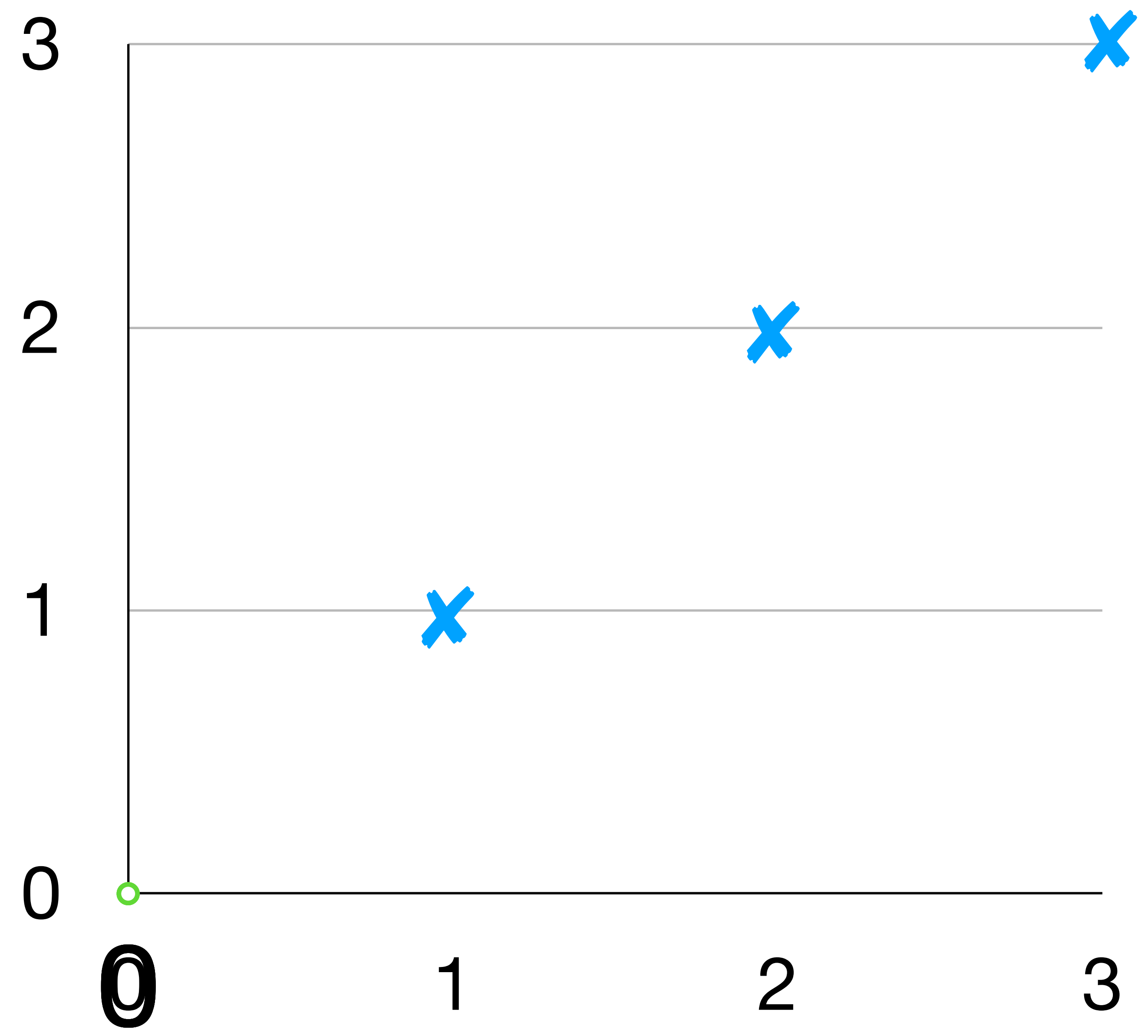
$$J(\theta_1)$$

function of the parameter θ_1



$$h_\theta(x)$$

for fixed θ_1, this is a function of x

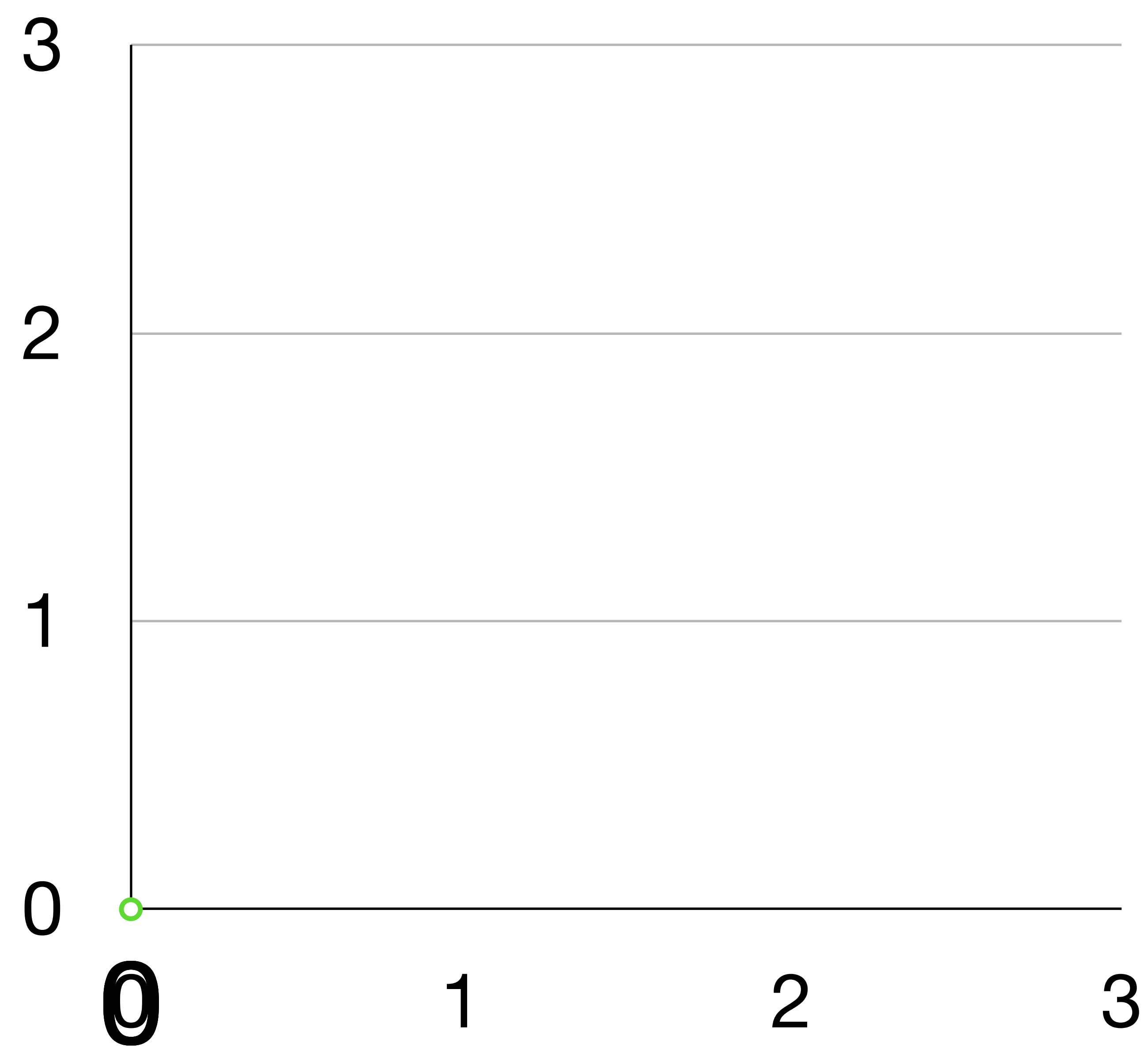


$h_\theta(x)$

$\theta_1 = 1$

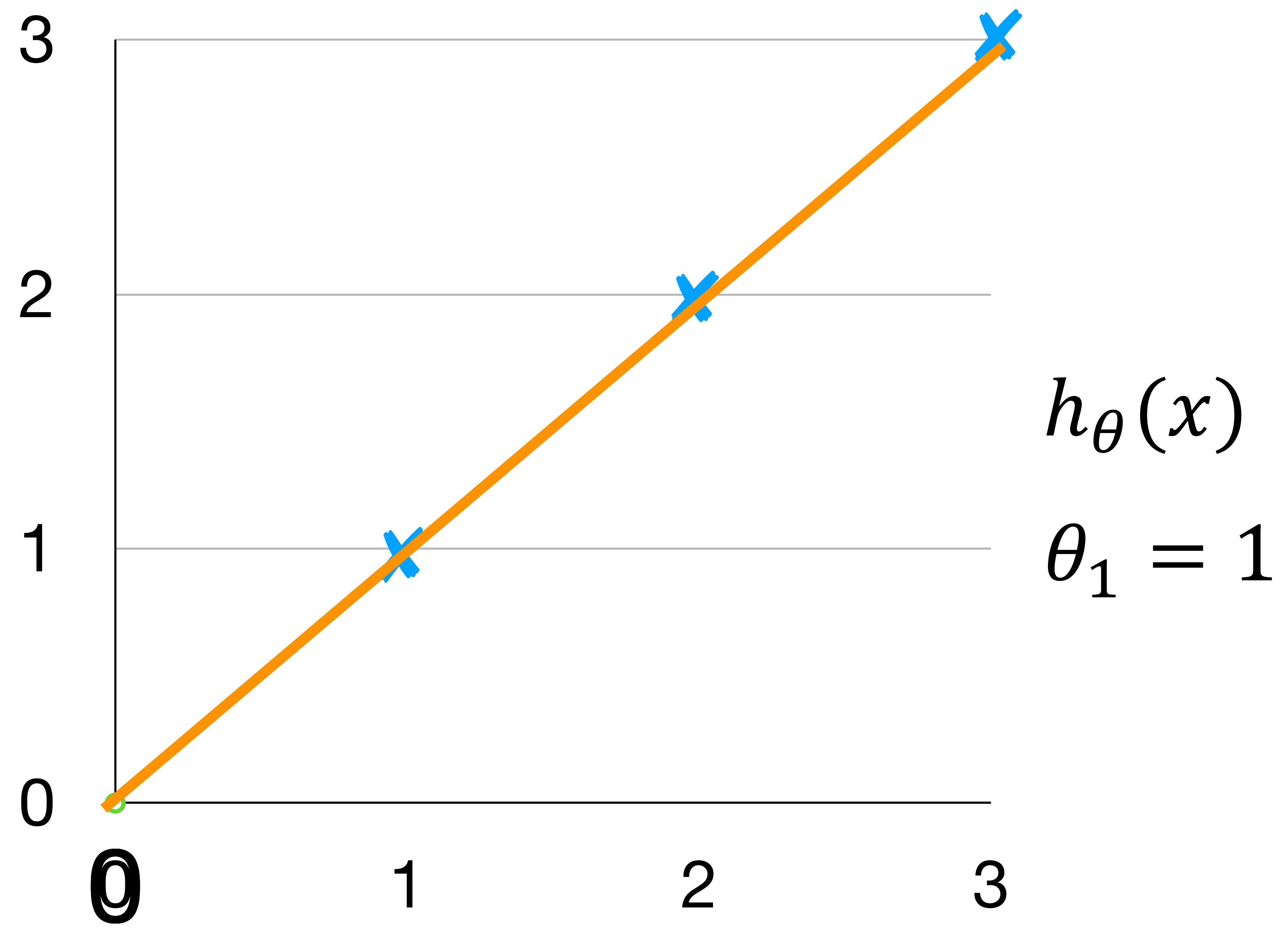
$$J(\theta_1)$$

function of the parameter θ_1



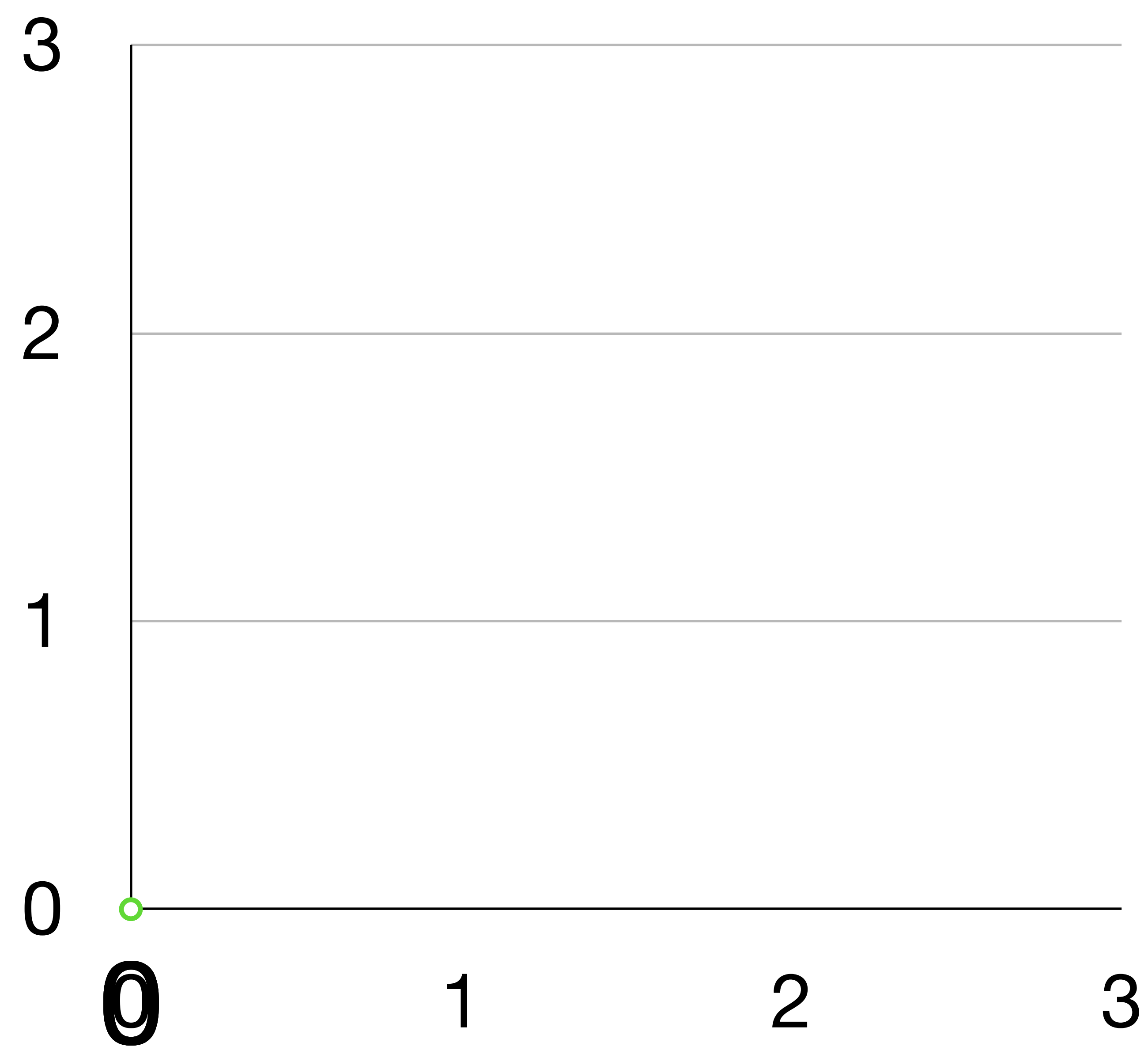
$$h_\theta(x)$$

for fixed θ_1, this is a function of x



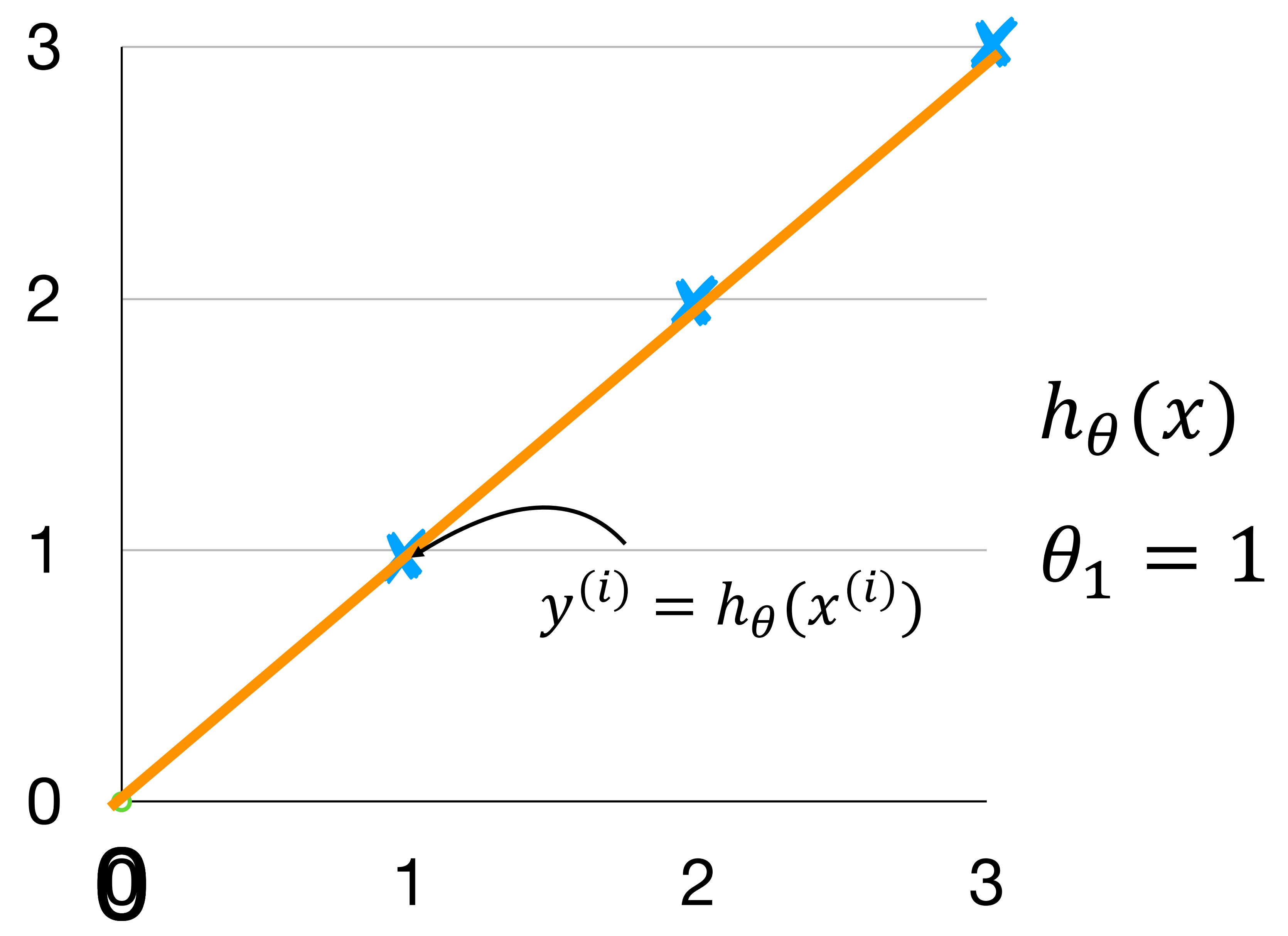
$$J(\theta_1)$$

function of the parameter θ_1



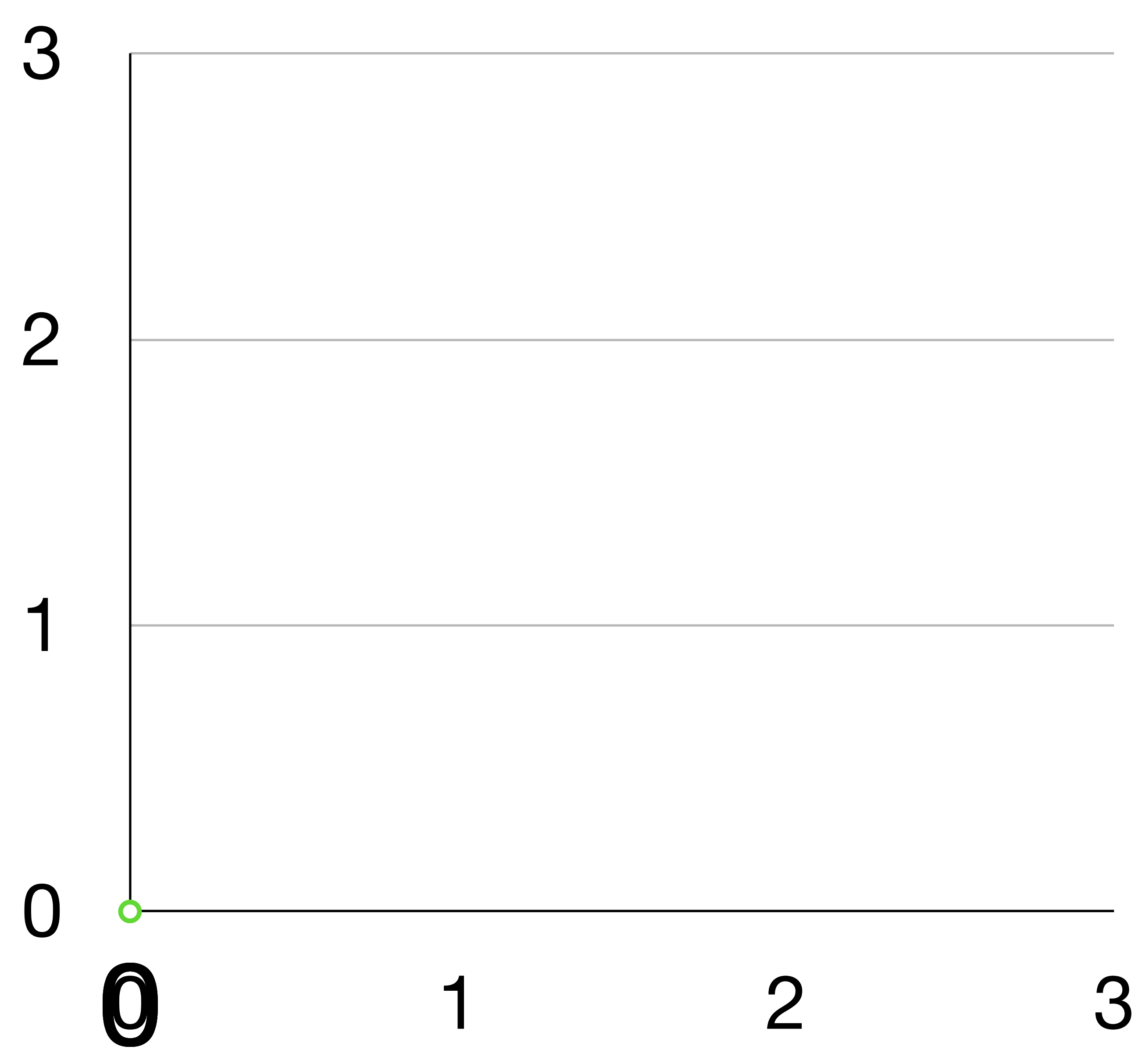
$$h_\theta(x)$$

for fixed θ_1, this is a function of x



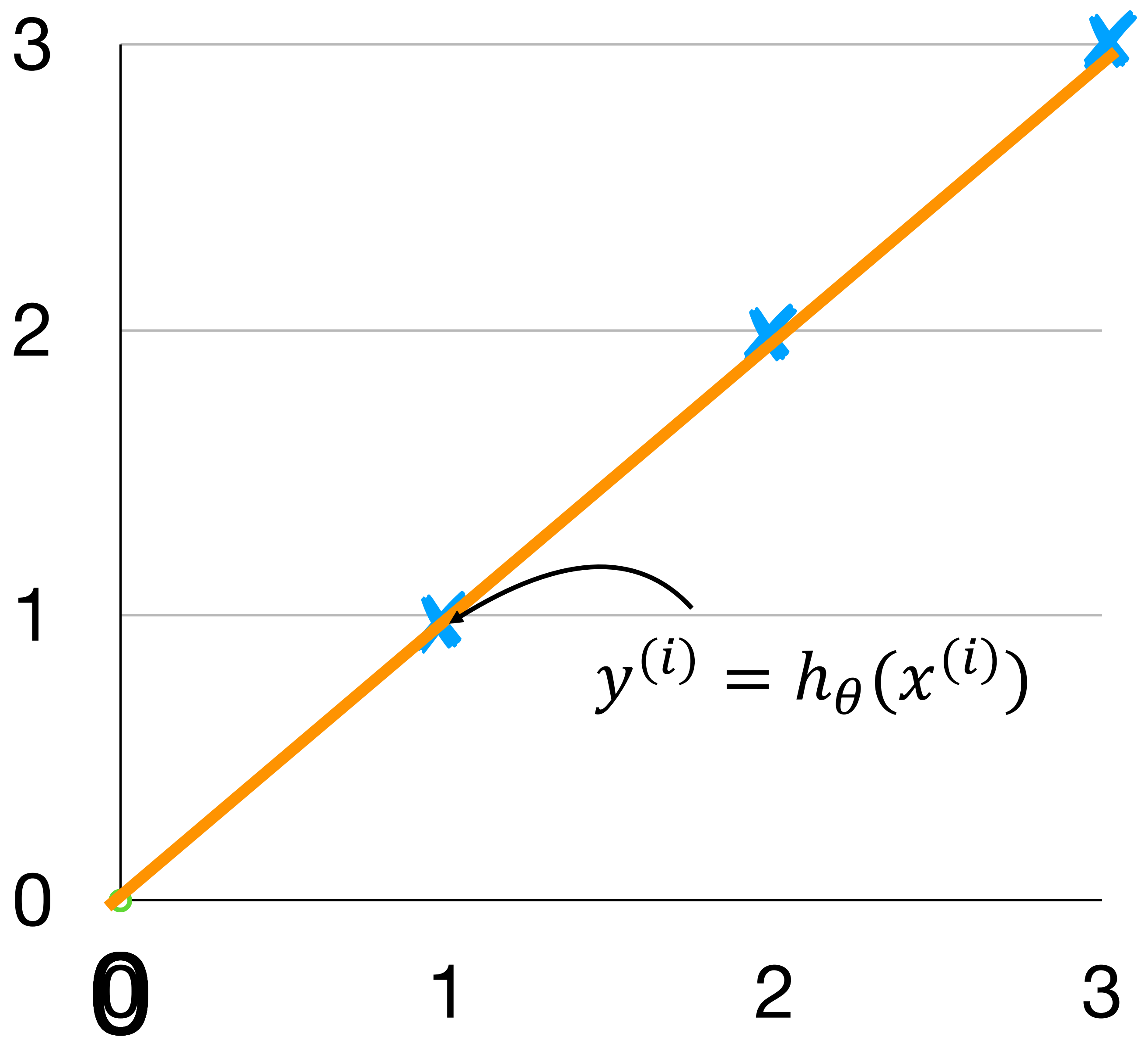
$$J(\theta_1)$$

function of the parameter θ_1



$h_\theta(x)$

for fixed θ_1, this is a function of x



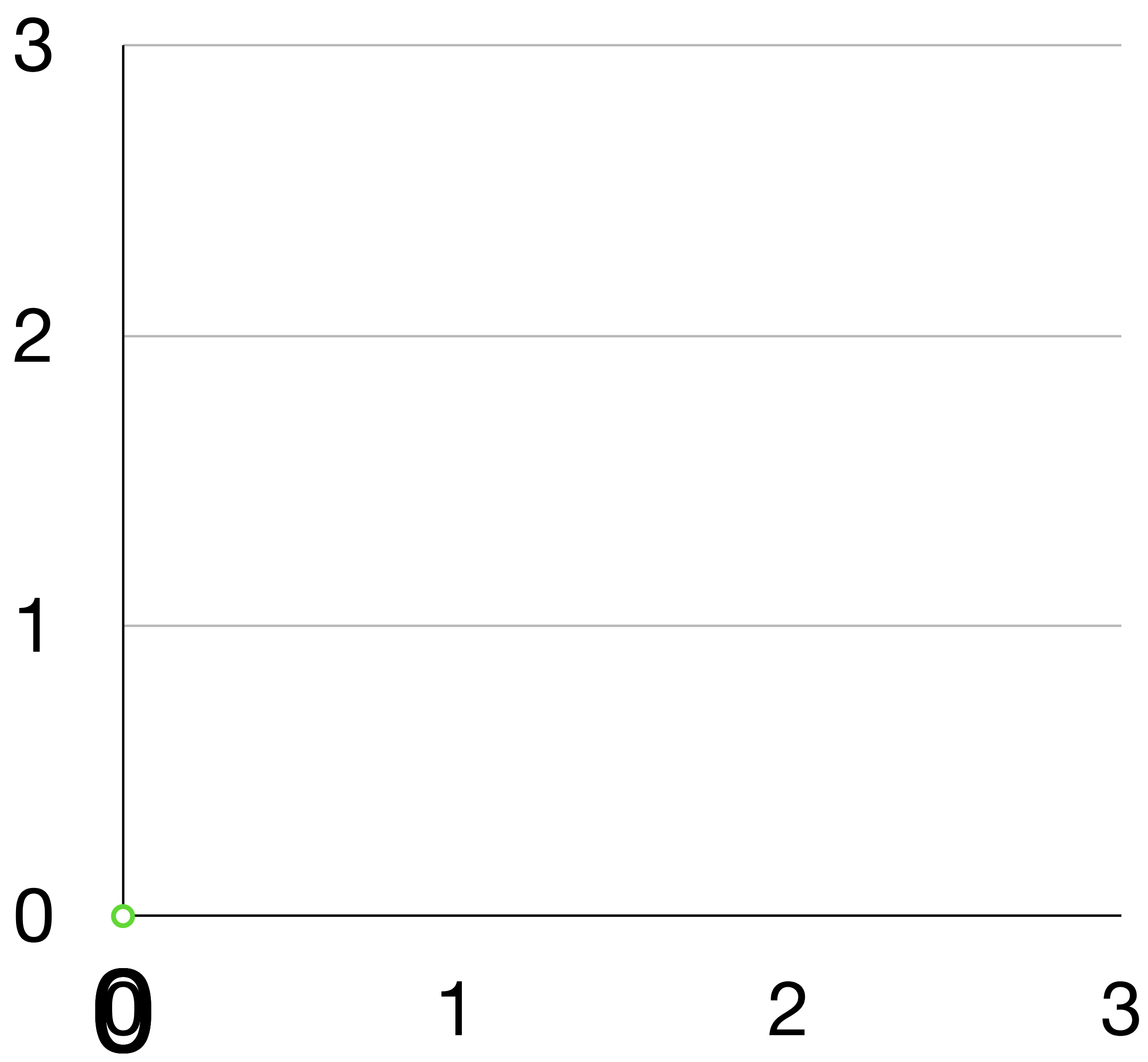
$h_\theta(x)$

$\theta_1 = 1$

$$J(1) = \frac{1}{2m}[(1-1)^2 + (2-2)^2 + (3-3)^2]$$

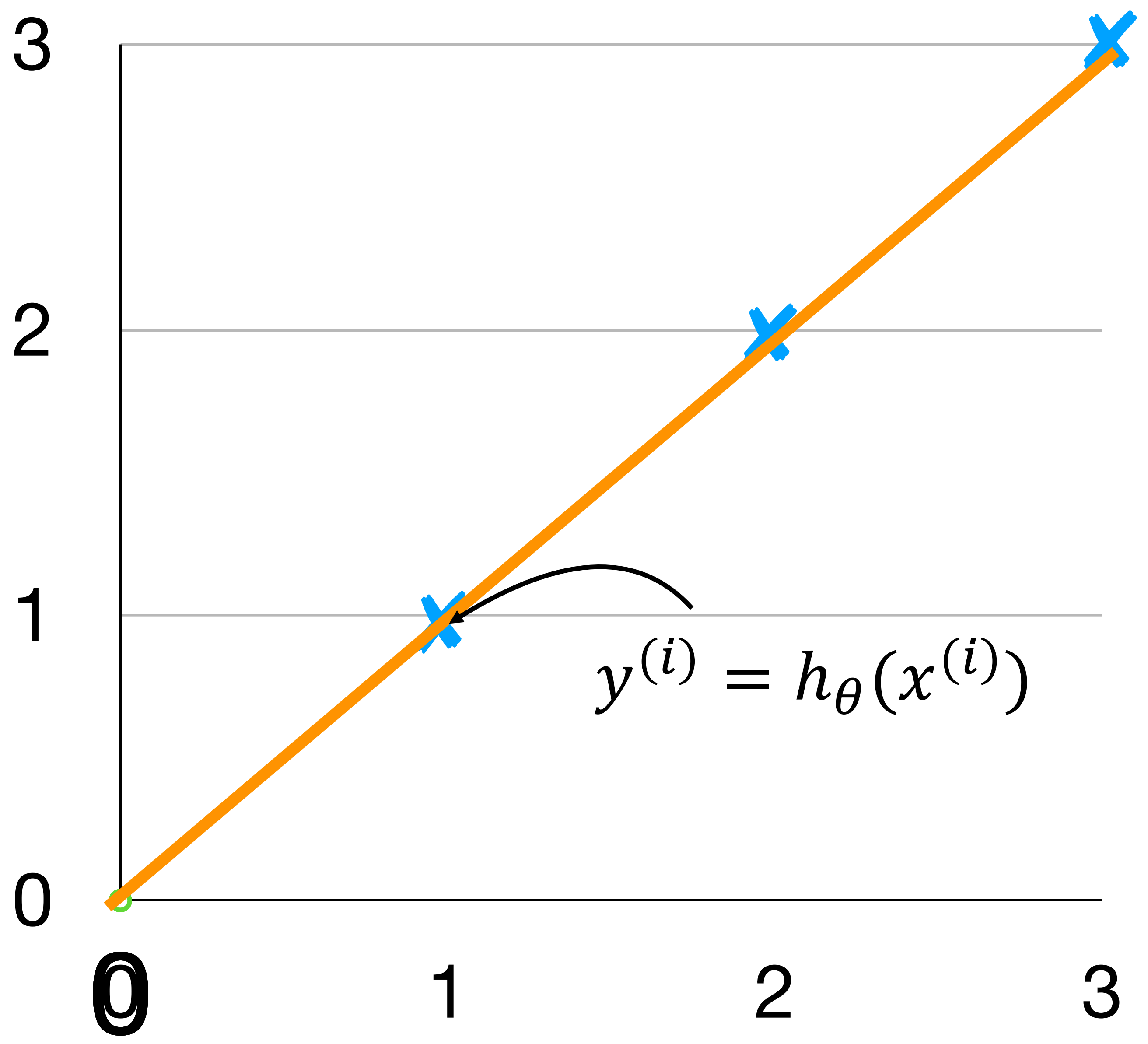
$J(\theta_1)$

function of the parameter θ_1



$$h_\theta(x)$$

for fixed θ_1, this is a function of x

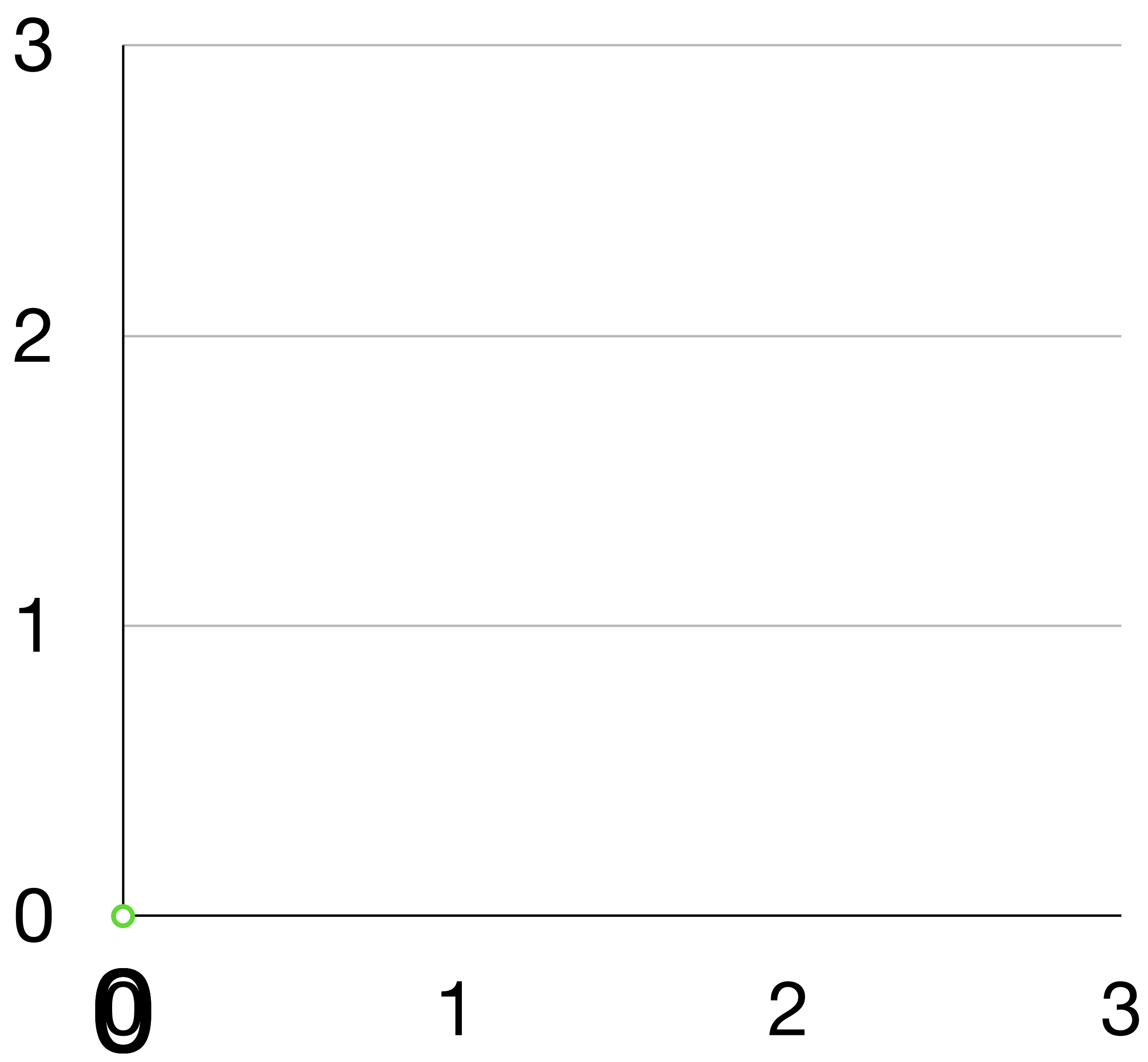


$h_\theta(x)$

$\theta_1 = 1$

$$J(\theta_1)$$

function of the parameter θ_1

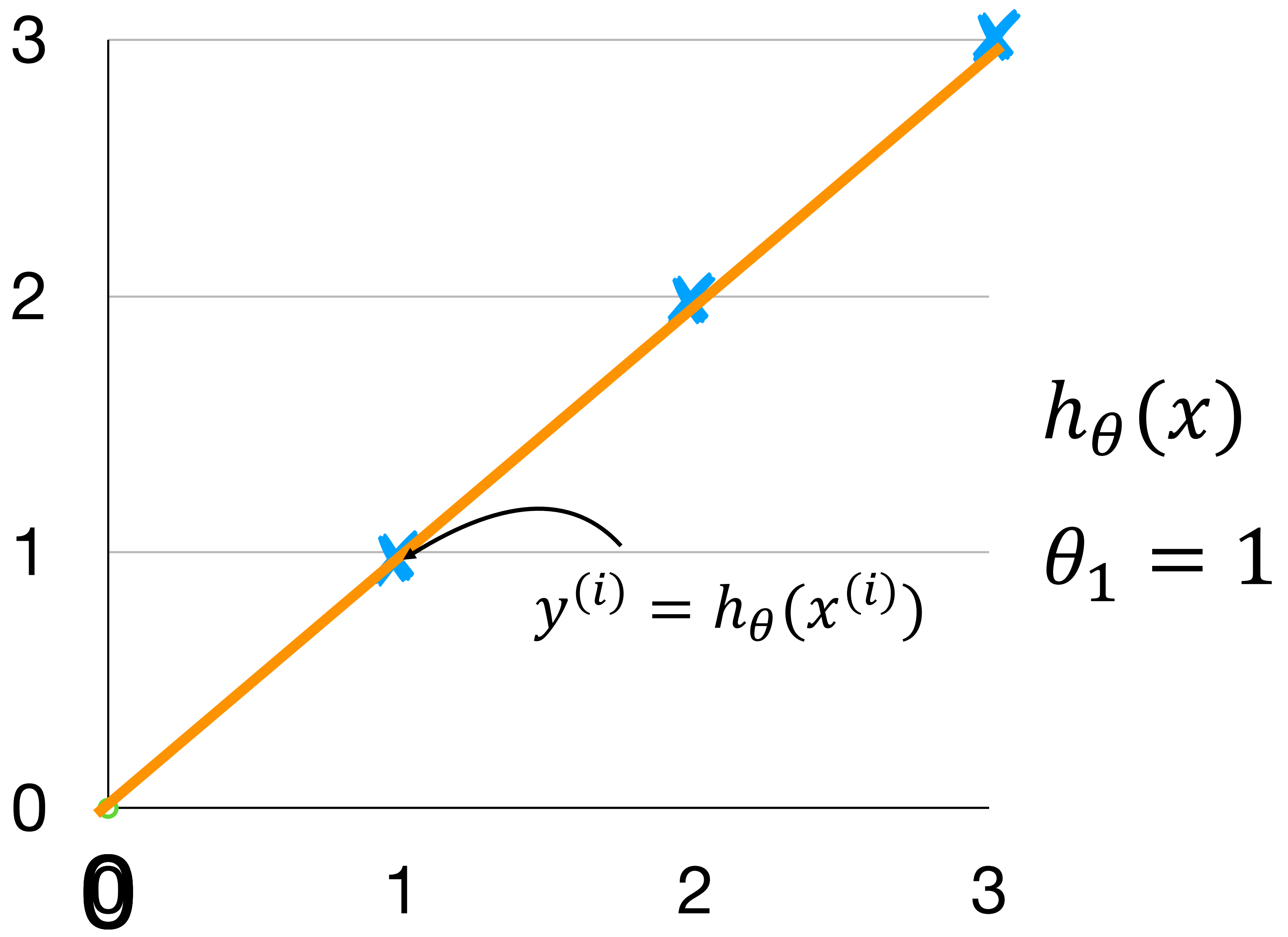


$$J(1) = \frac{1}{2m}[(1-1)^2 + (2-2)^2 + (3-3)^2]$$

$$J(1) = \frac{1}{2*3}[0^2 + 0^2 + 0^2]$$

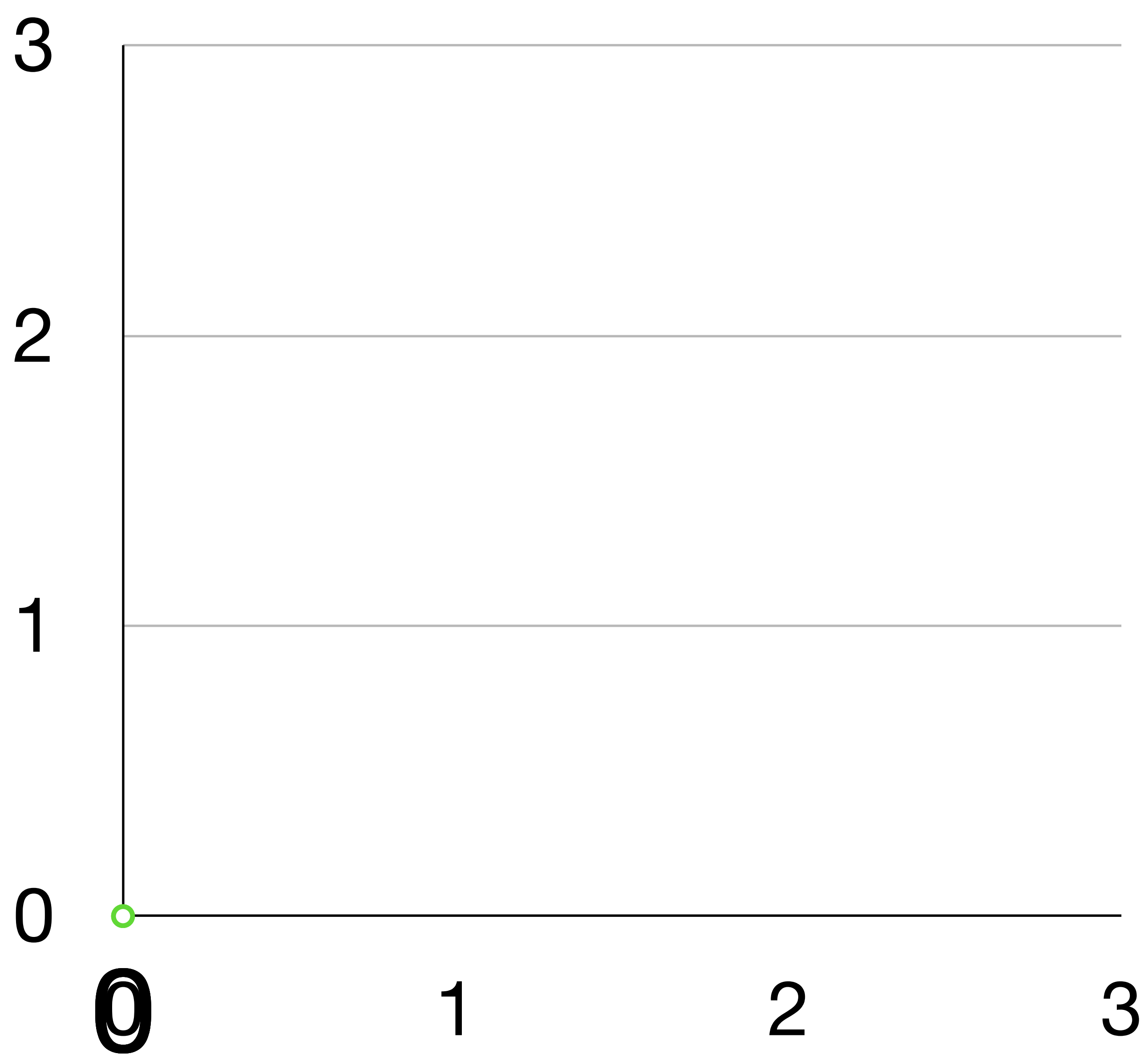
$$h_\theta(x)$$

for fixed θ_1, this is a function of x



$$J(\theta_1)$$

function of the parameter θ_1



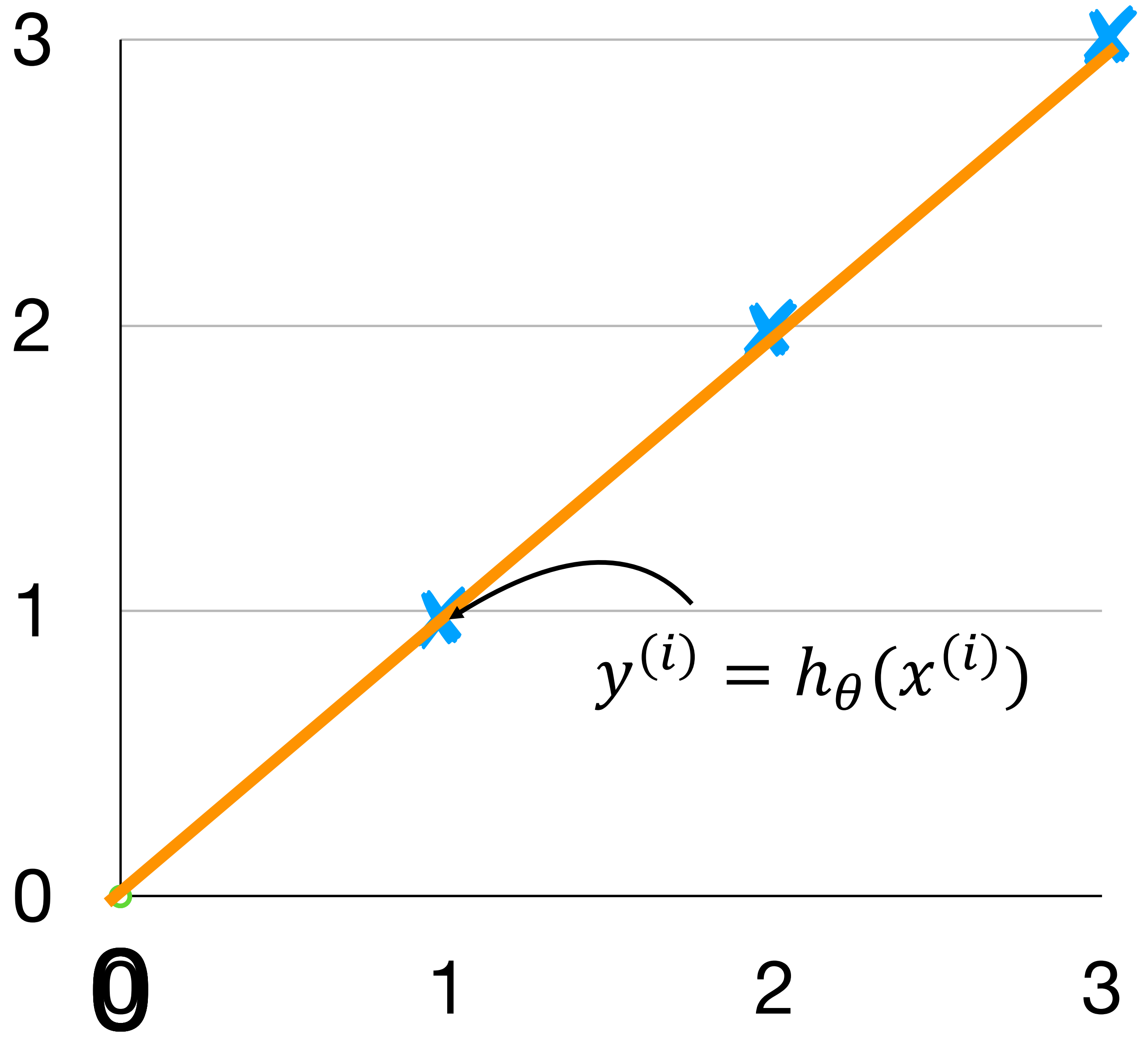
$$J(1) = \frac{1}{2m}[(1-1)^2 + (2-2)^2 + (3-3)^2]$$

$$J(1) = \frac{1}{2*3}[0^2 + 0^2 + 0^2]$$

$$J(1) = \frac{1}{2*3}[0] = 0$$

$h_\theta(x)$

for fixed θ_1, this is a function of x

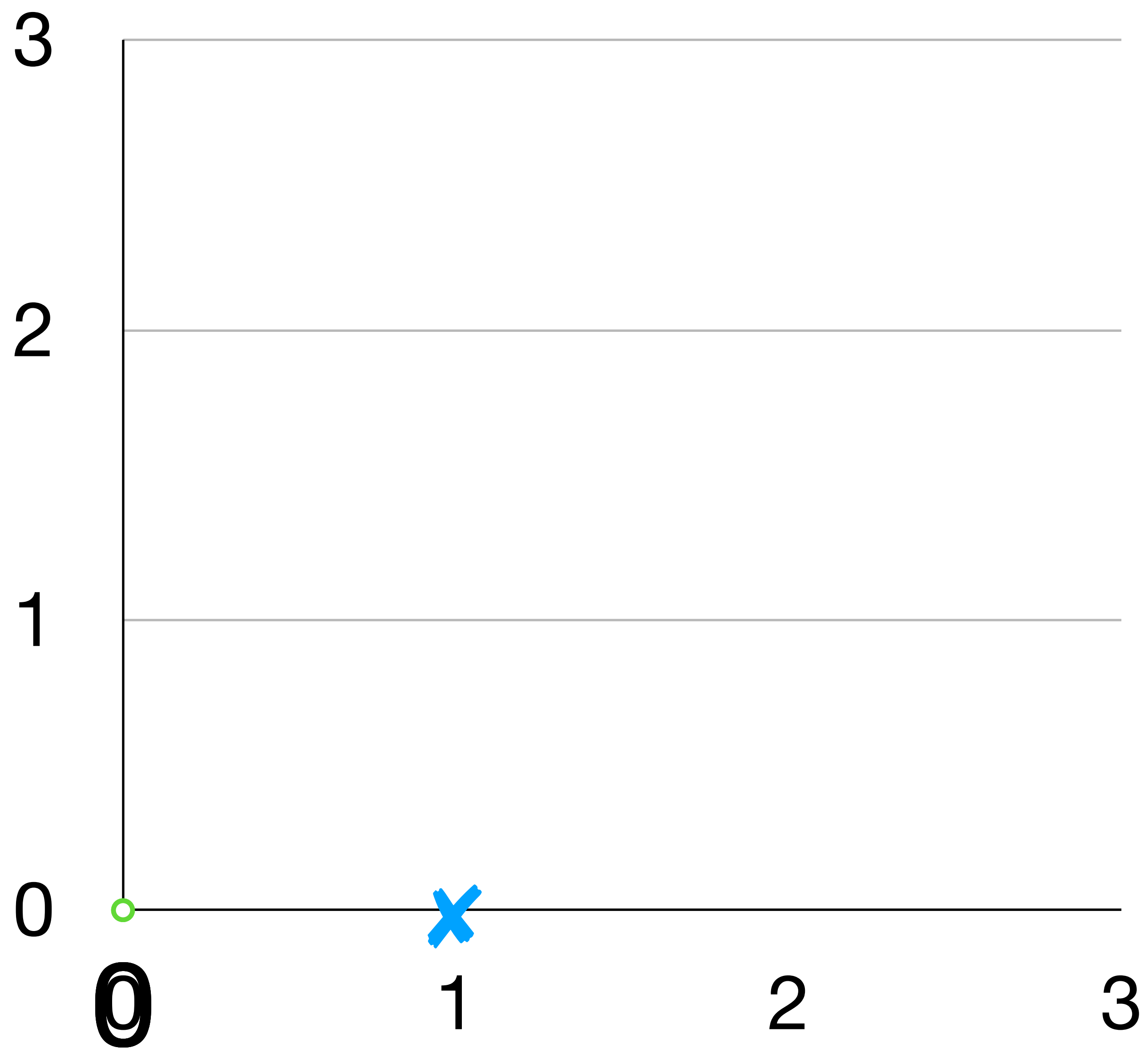


$h_\theta(x)$

$\theta_1 = 1$

$J(\theta_1)$

function of the parameter θ_1



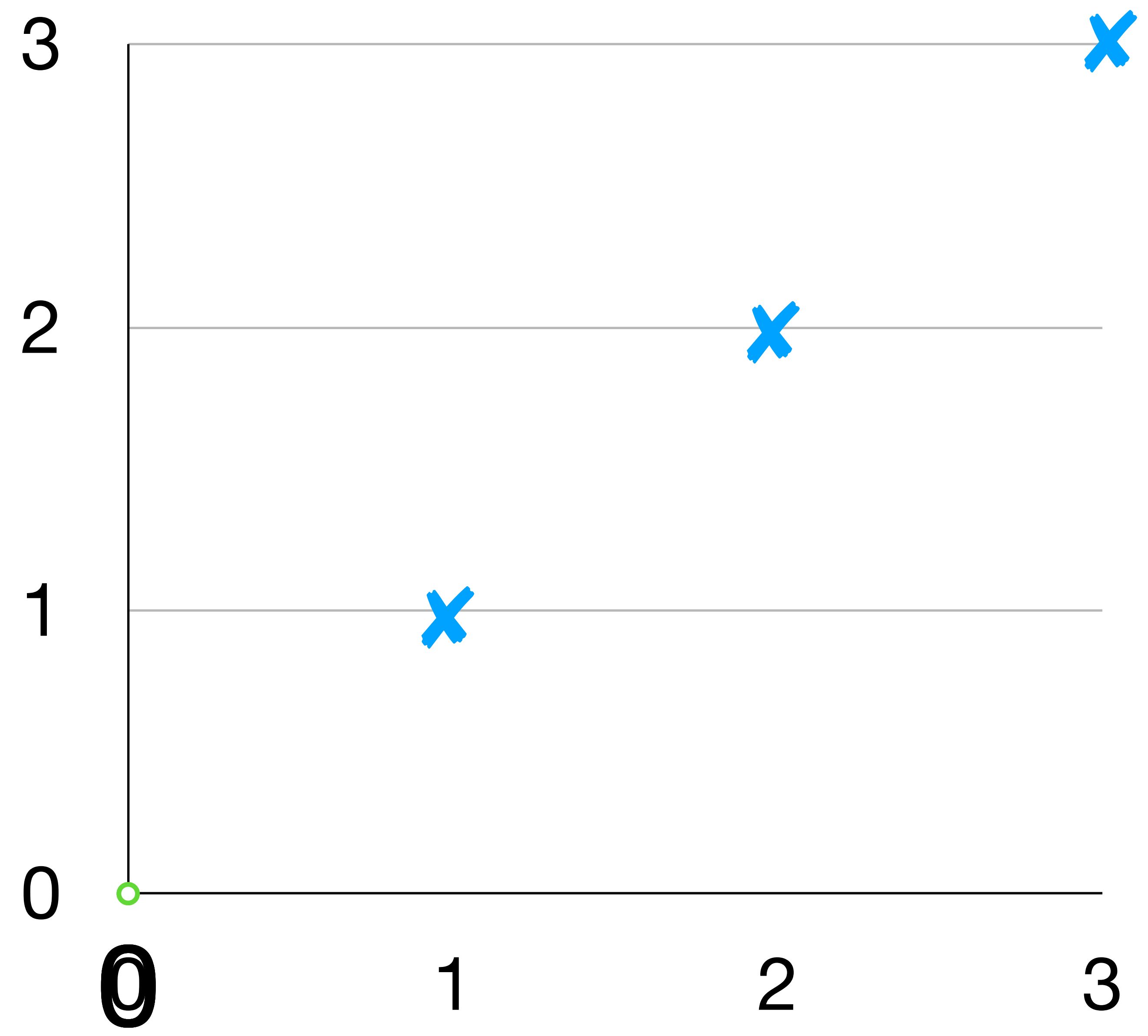
$$J(1) = \frac{1}{2m}[(1-1)^2 + (2-2)^2 + (3-3)^2]$$

$$J(1) = \frac{1}{2 * 3}[0^2 + 0^2 + 0^2]$$

$$J(1) = \frac{1}{2 * 3}[0] = 0$$

$$h_\theta(x)$$

for fixed θ_1, this is a function of x

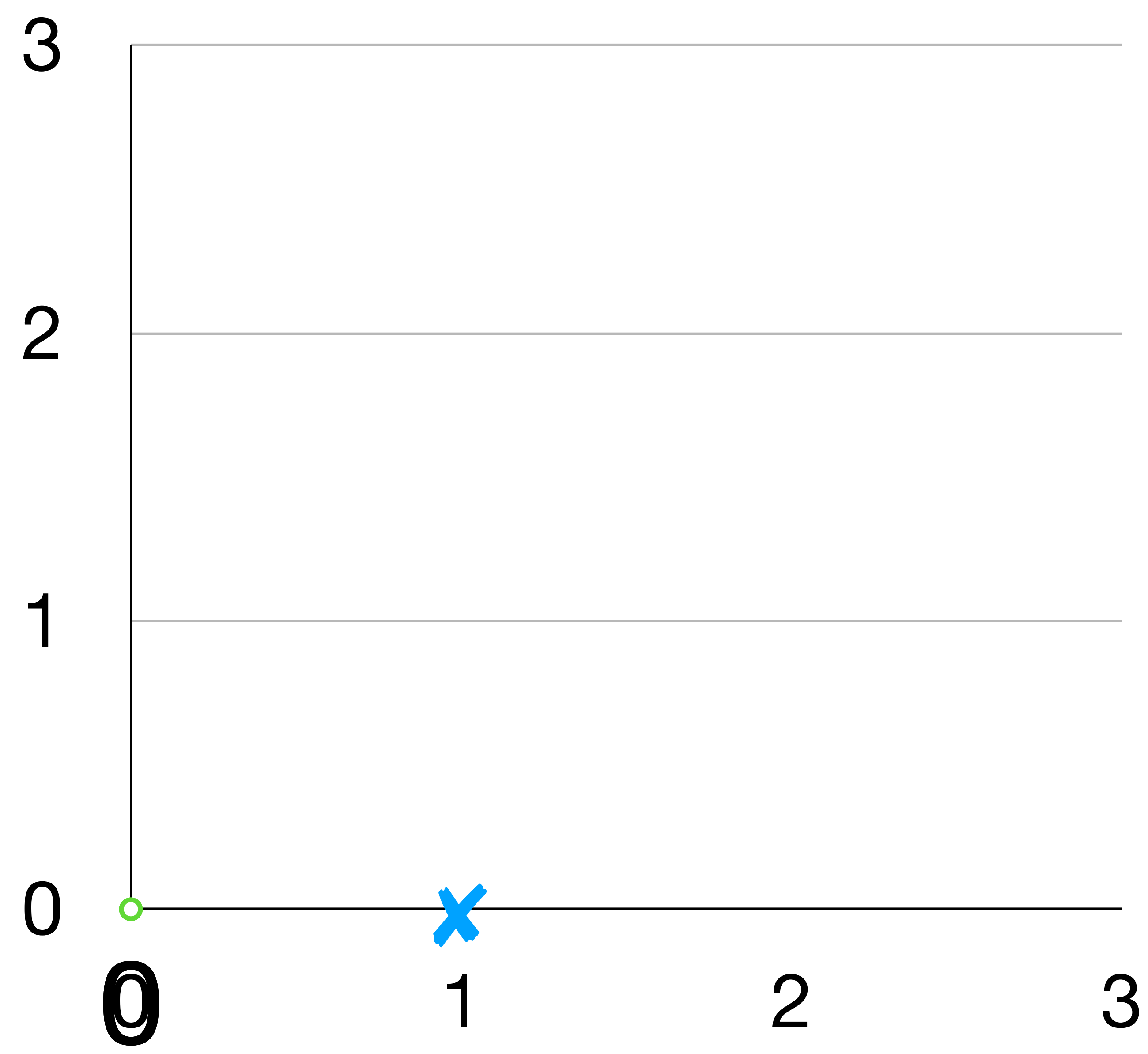


$h_\theta(x)$

$\theta_1 = 0.5$

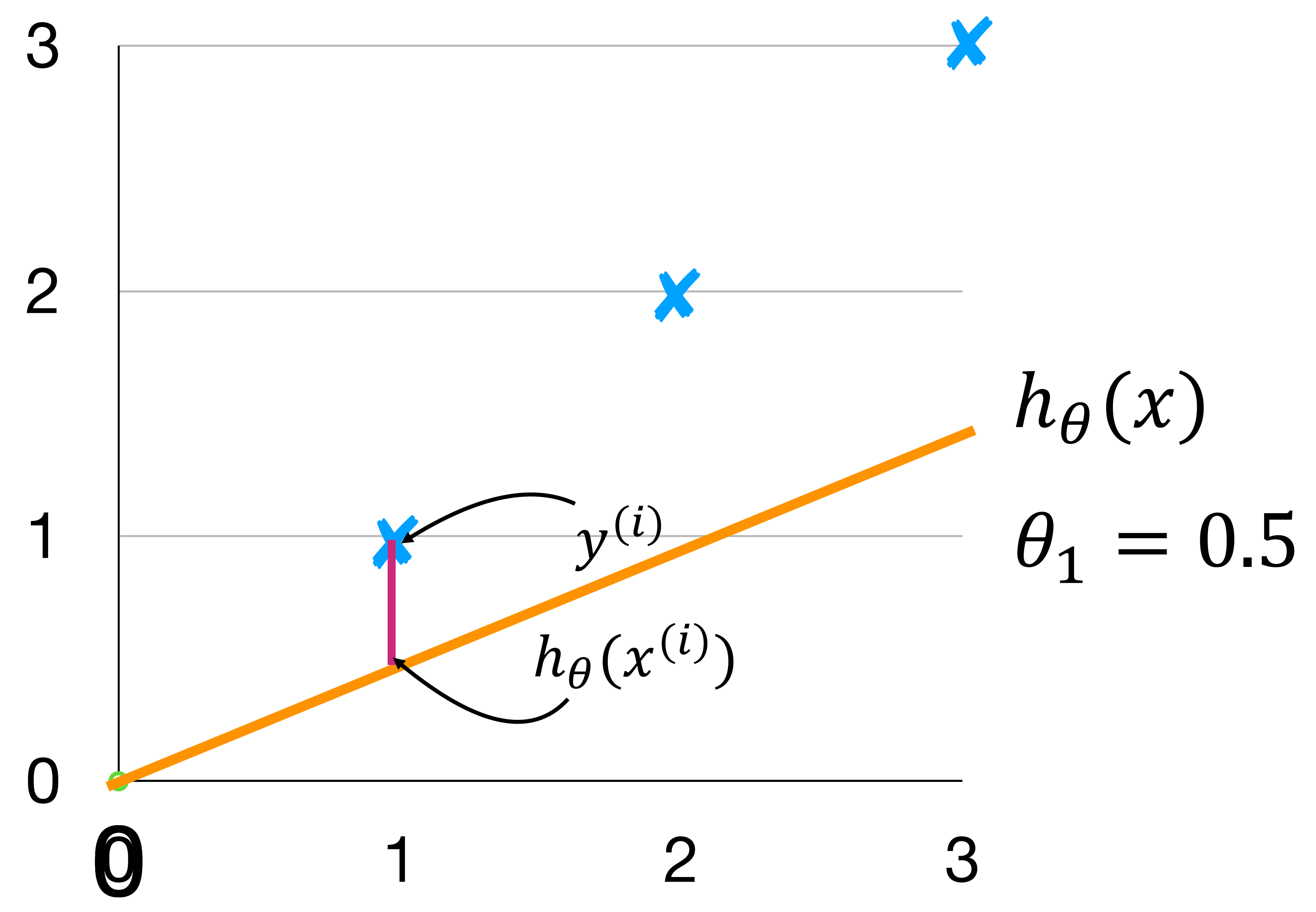
$$J(\theta_1)$$

function of the parameter θ_1



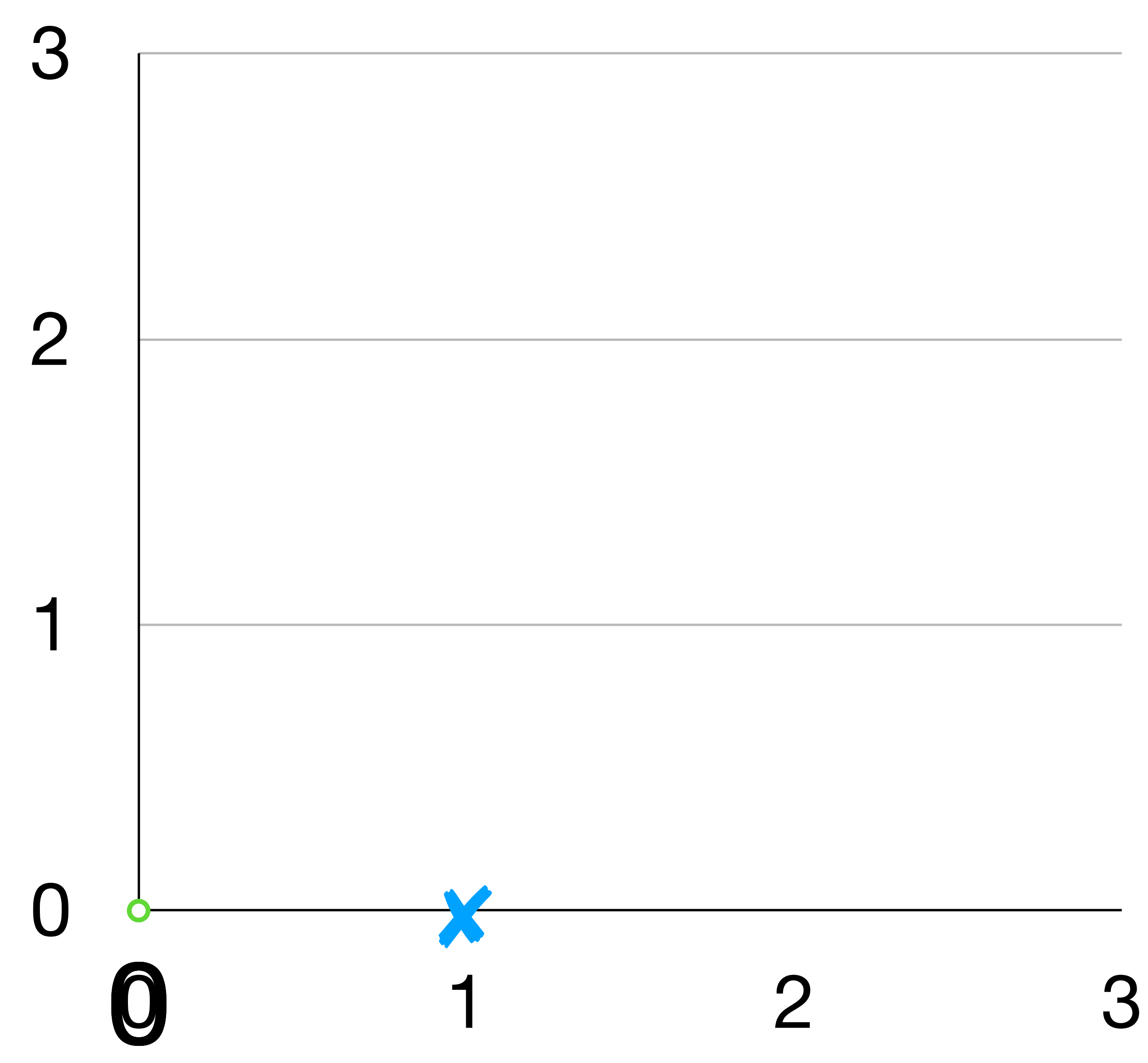
$$h_\theta(x)$$

for fixed θ_1, this is a function of x



$$J(\theta_1)$$

function of the parameter θ_1

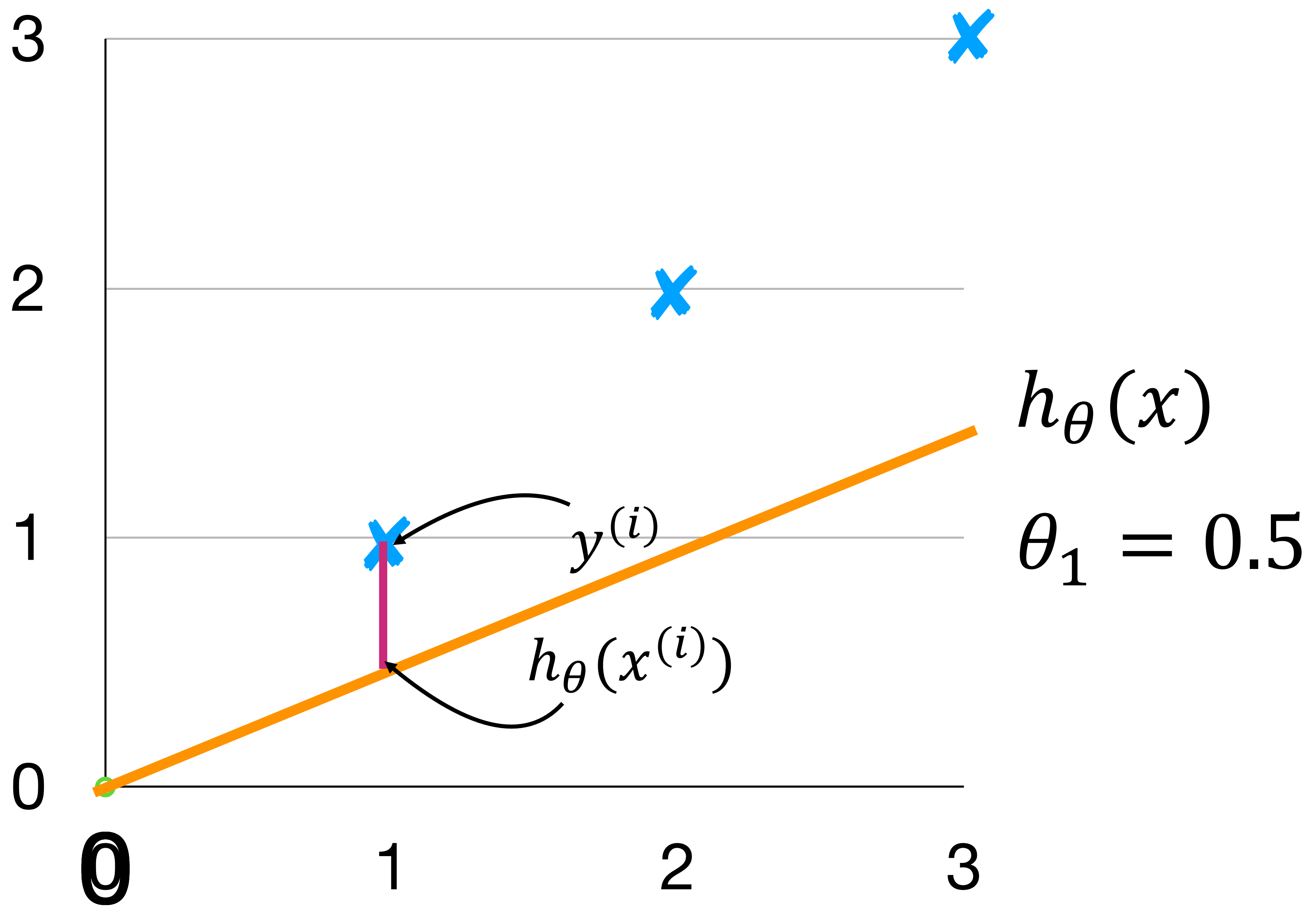


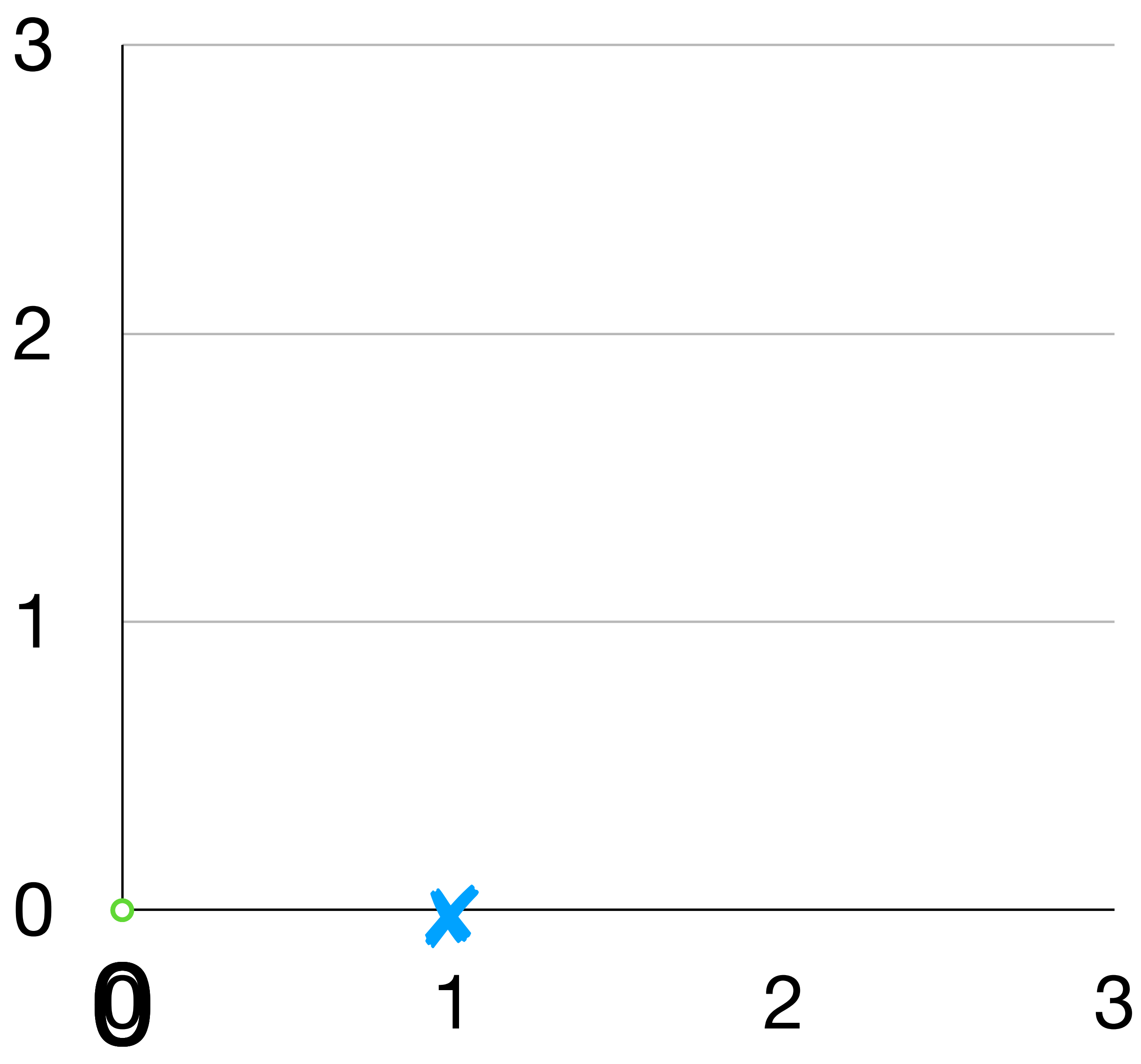
$h_\theta(x)$

for fixed θ_1, this is a function of x

$J(\theta_1)$

function of the parameter θ_1

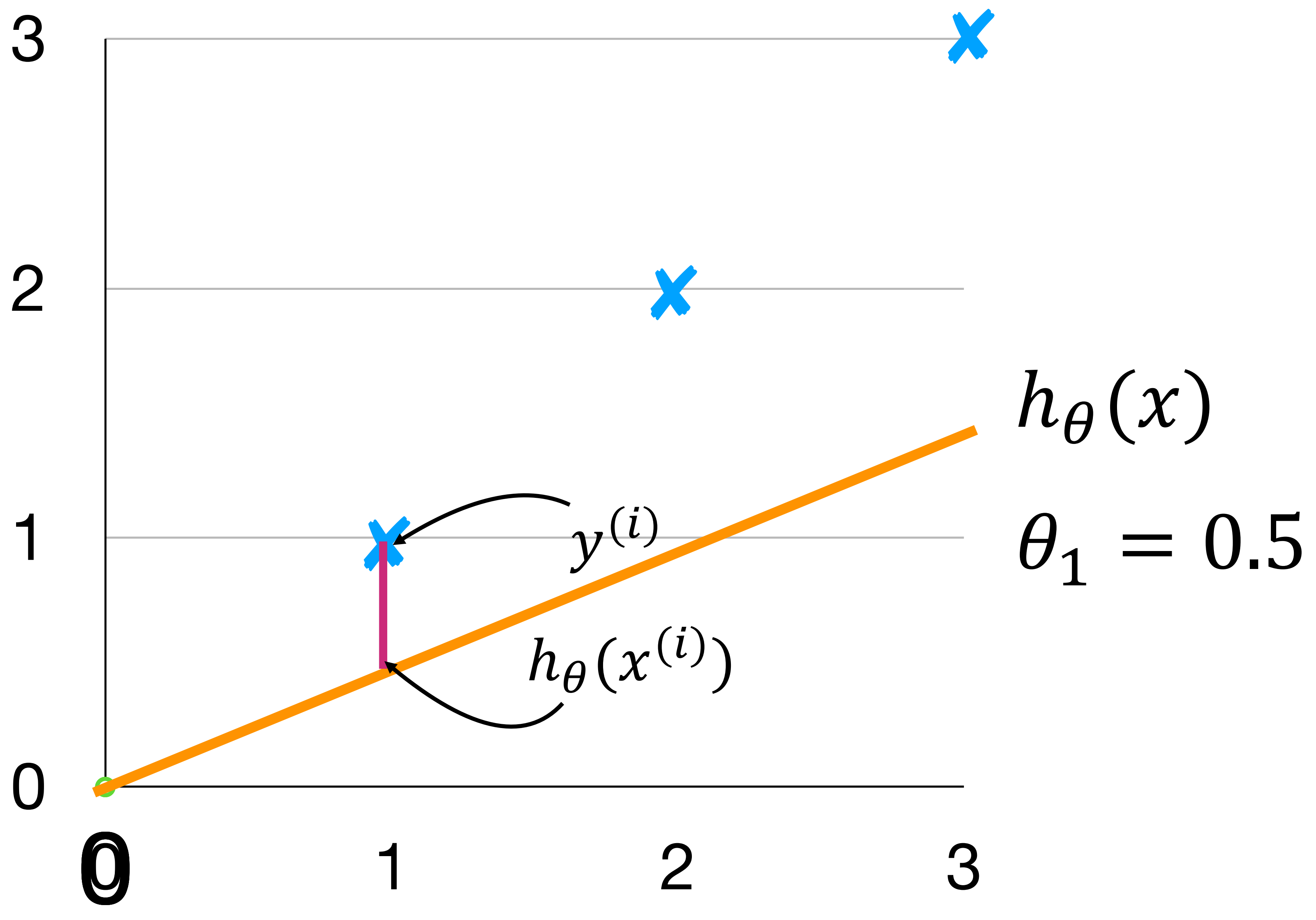




$$J(0.5) = \frac{1}{2m}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right]$$

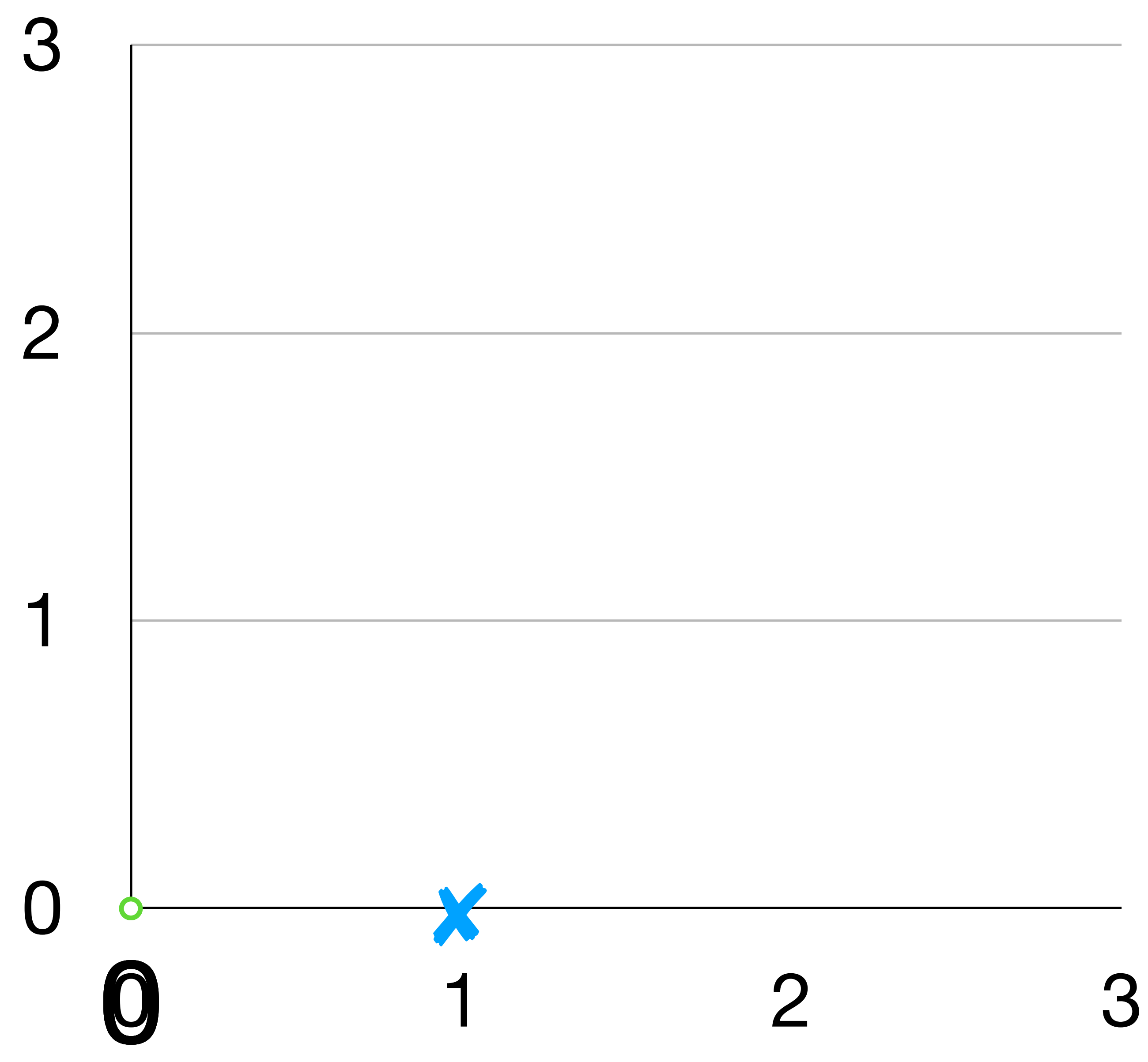
$$h_\theta(x)$$

for fixed θ_1, this is a function of x



$$J(\theta_1)$$

function of the parameter θ_1

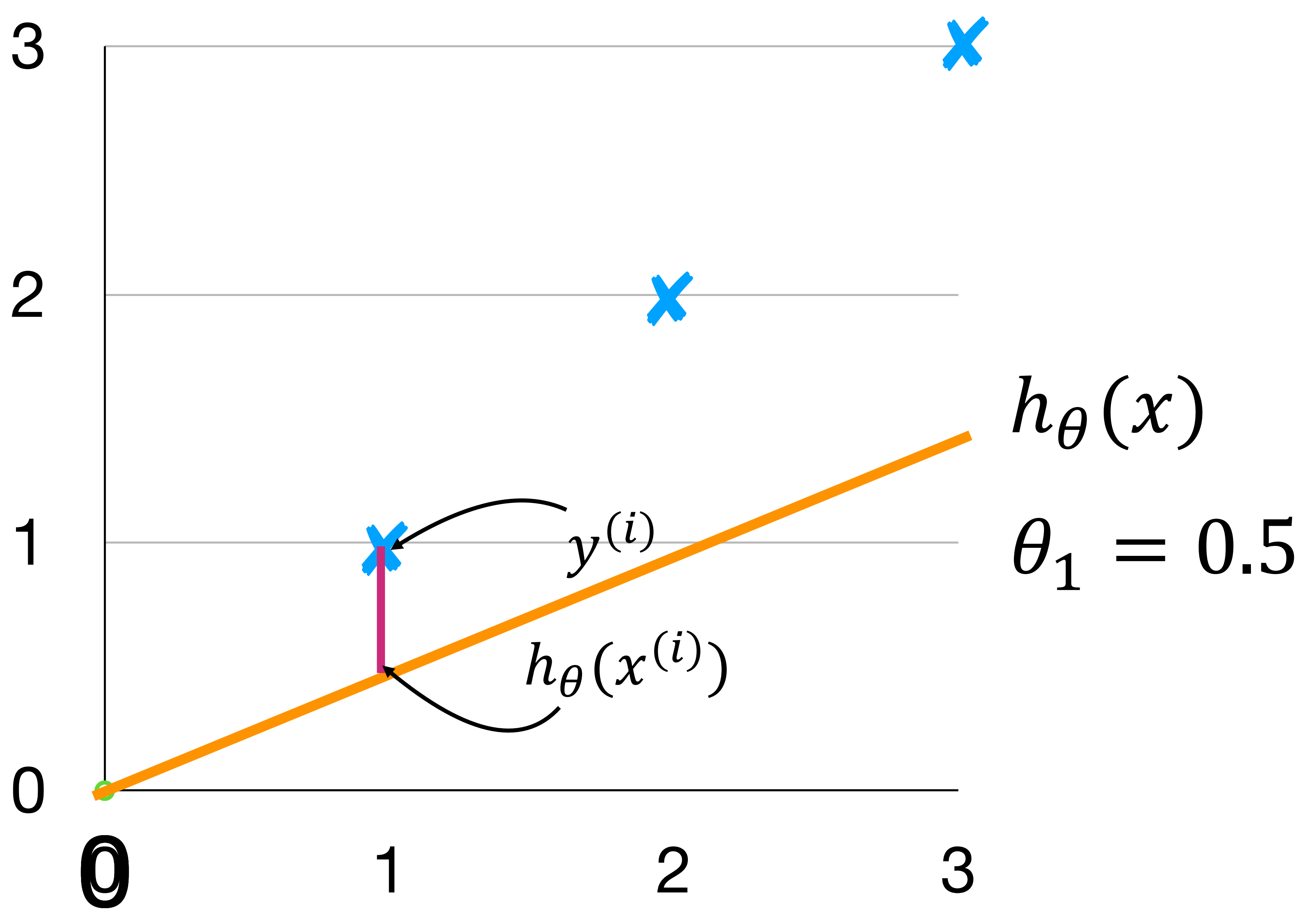


$$J(0.5) = \frac{1}{2m}[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2]$$

$$J(0.5) = \frac{1}{2*3}[(-0.5)^2 + (-1)^2 + (-1.5)^2]$$

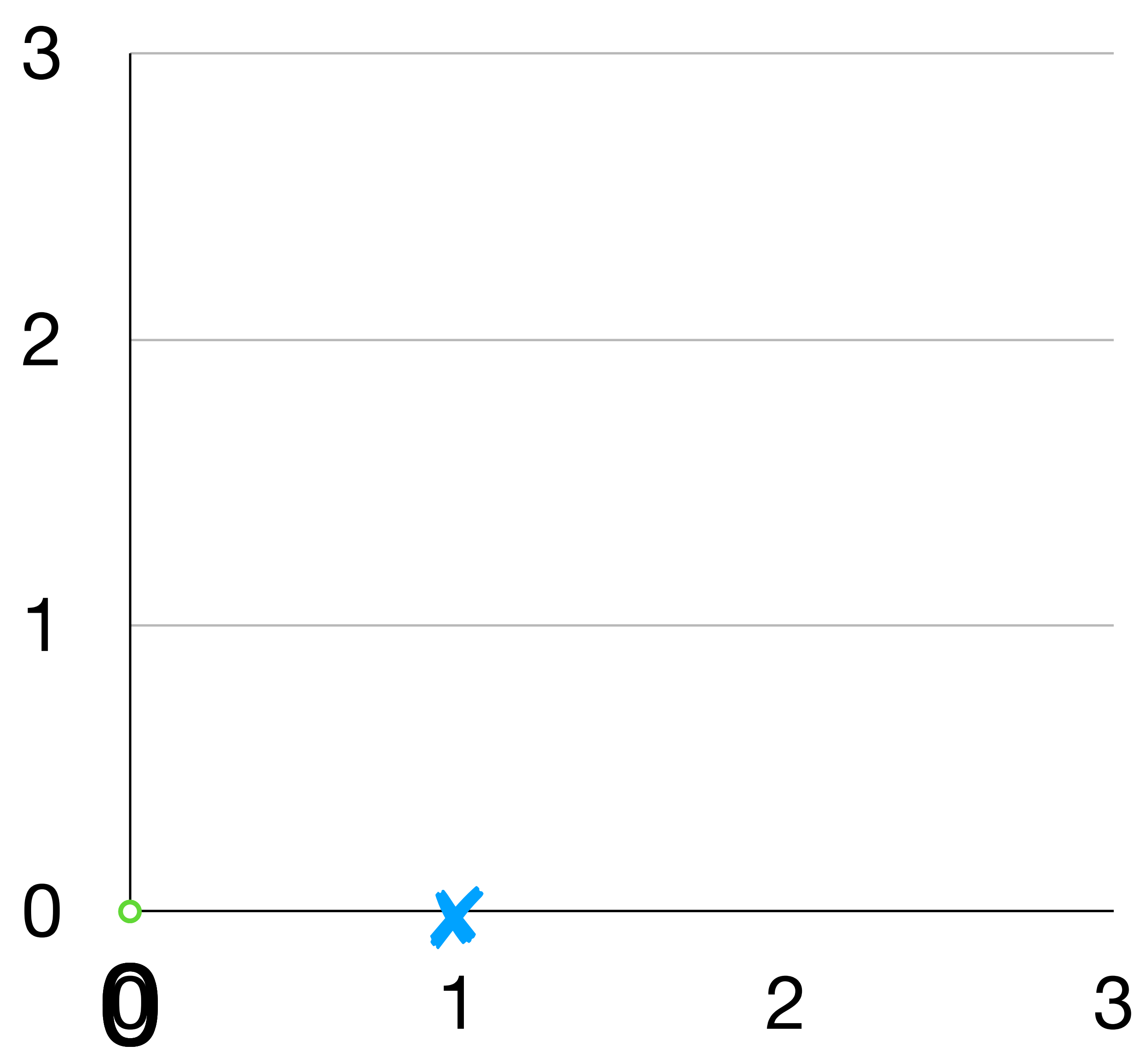
$h_\theta(x)$

for fixed θ_1, this is a function of x



$J(\theta_1)$

function of the parameter θ_1



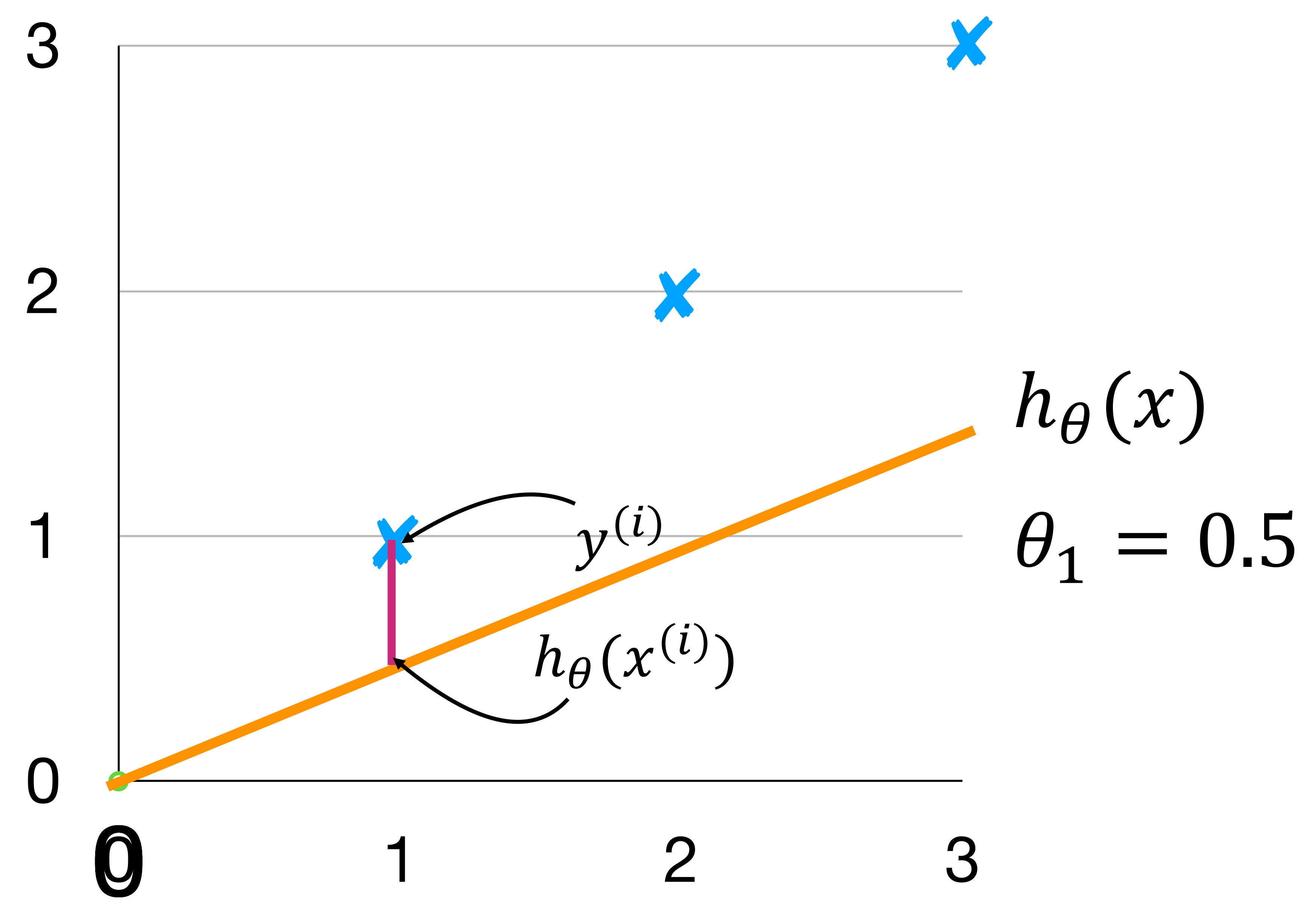
$$J(0.5) = \frac{1}{2m}[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2]$$

$$J(0.5) = \frac{1}{2*3}[(-0.5)^2 + (-1)^2 + (-1.5)^2]$$

$$J(0.5) = \frac{1}{2*3}[3.5] = 0.583$$

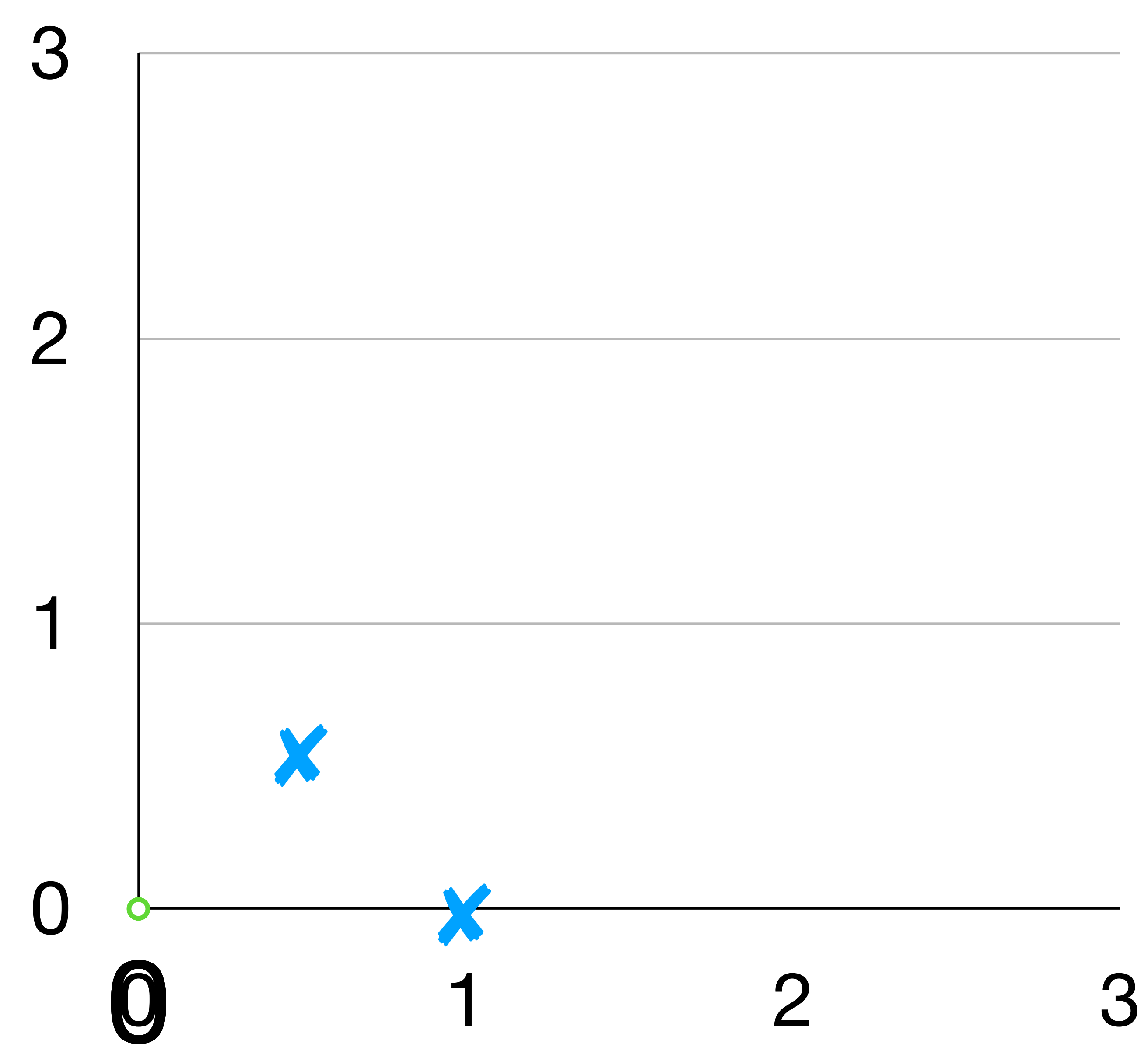
$h_\theta(x)$

for fixed θ_1, this is a function of x



$J(\theta_1)$

function of the parameter θ_1



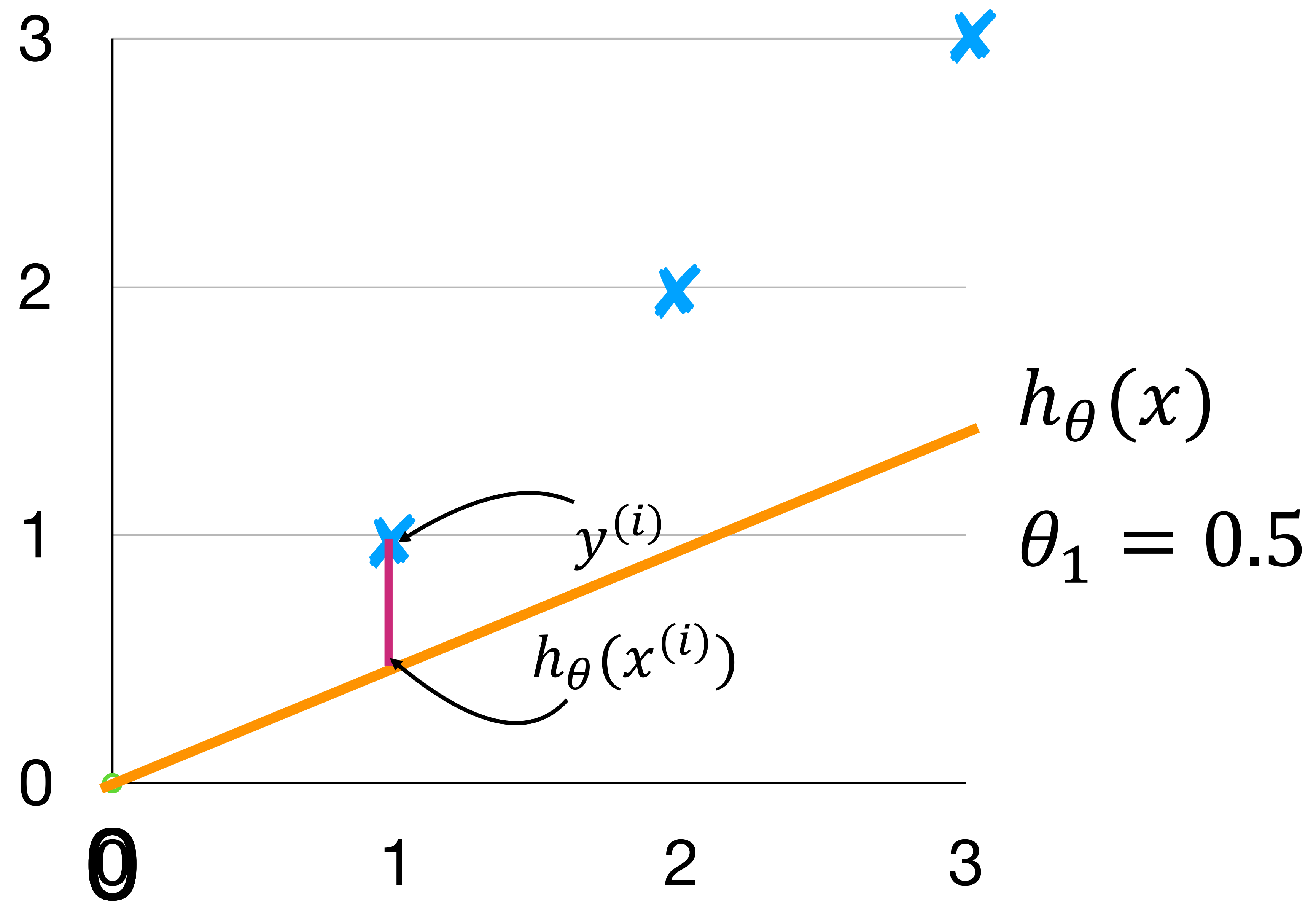
$$J(0.5) = \frac{1}{2m}[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2]$$

$$J(0.5) = \frac{1}{2*3}[(-0.5)^2 + (-1)^2 + (-1.5)^2]$$

$$J(0.5) = \frac{1}{2*3}[3.5] = 0.583$$

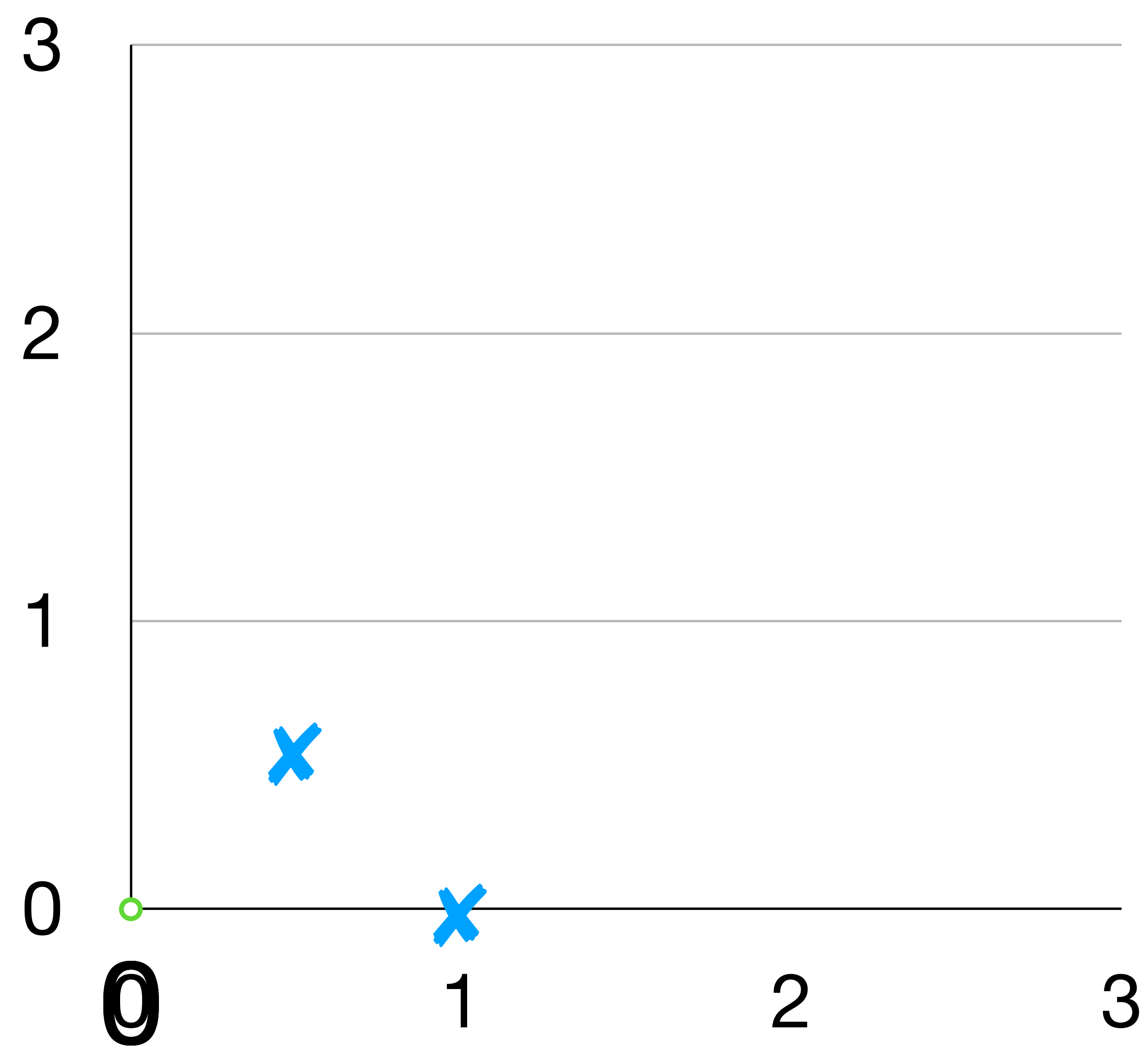
$$h_\theta(x)$$

for fixed θ_1, this is a function of x



$$J(\theta_1)$$

function of the parameter θ_1



$$J(0.5) = \frac{1}{2m}[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2]$$

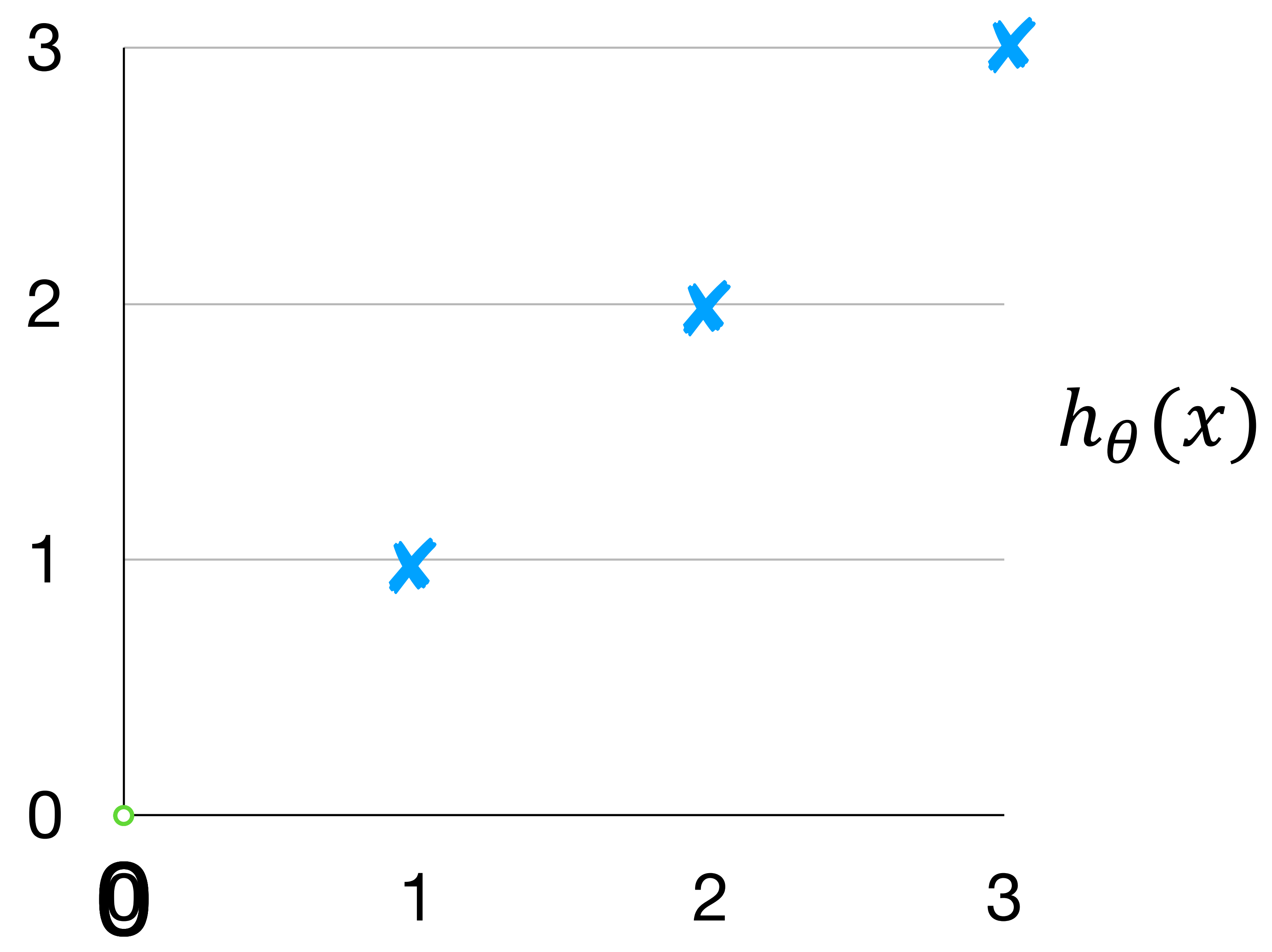
$$J(0.5) = \frac{1}{2*3}[(-0.5)^2 + (-1)^2 + (-1.5)^2]$$

$$J(0.5) = \frac{1}{2*3}[3.5] = 0.583$$

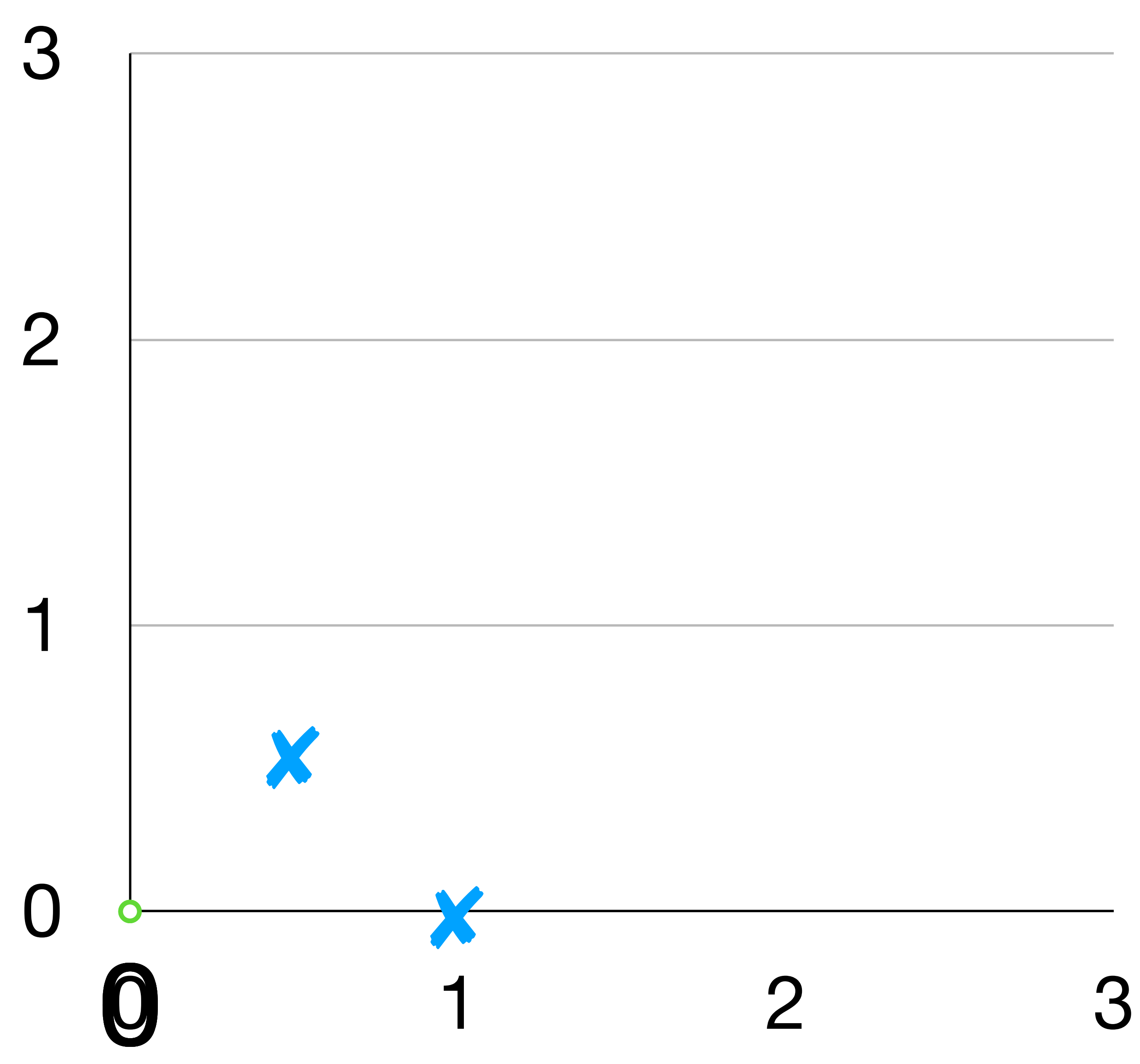
how about $\theta_1 = 0$

$J(\theta_1)$

for fixed θ_1, this is a function of x



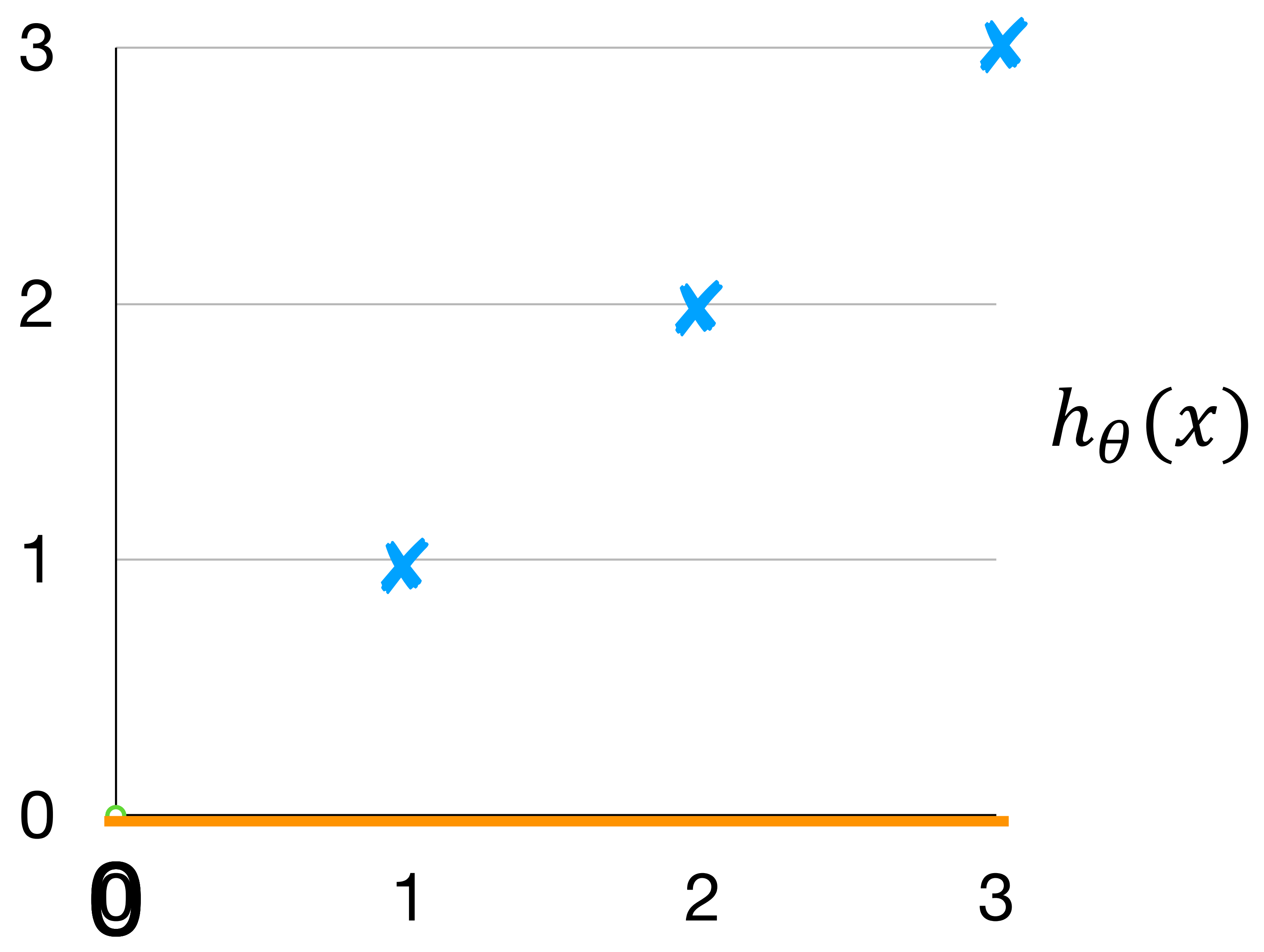
function of the parameter θ_1



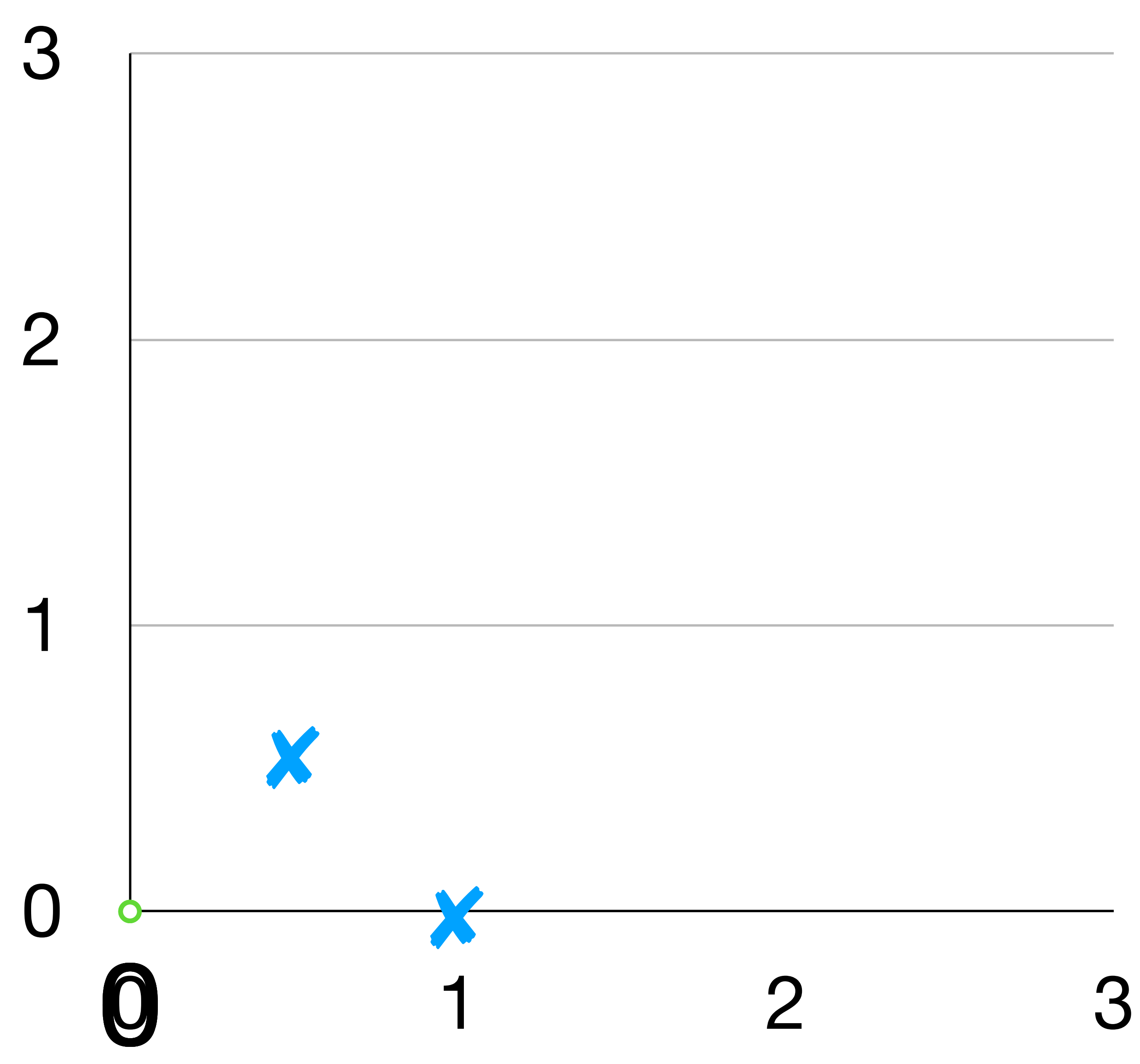
how about $\theta_1 = 0$

$J(\theta_1)$

for fixed θ_1, this is a function of x



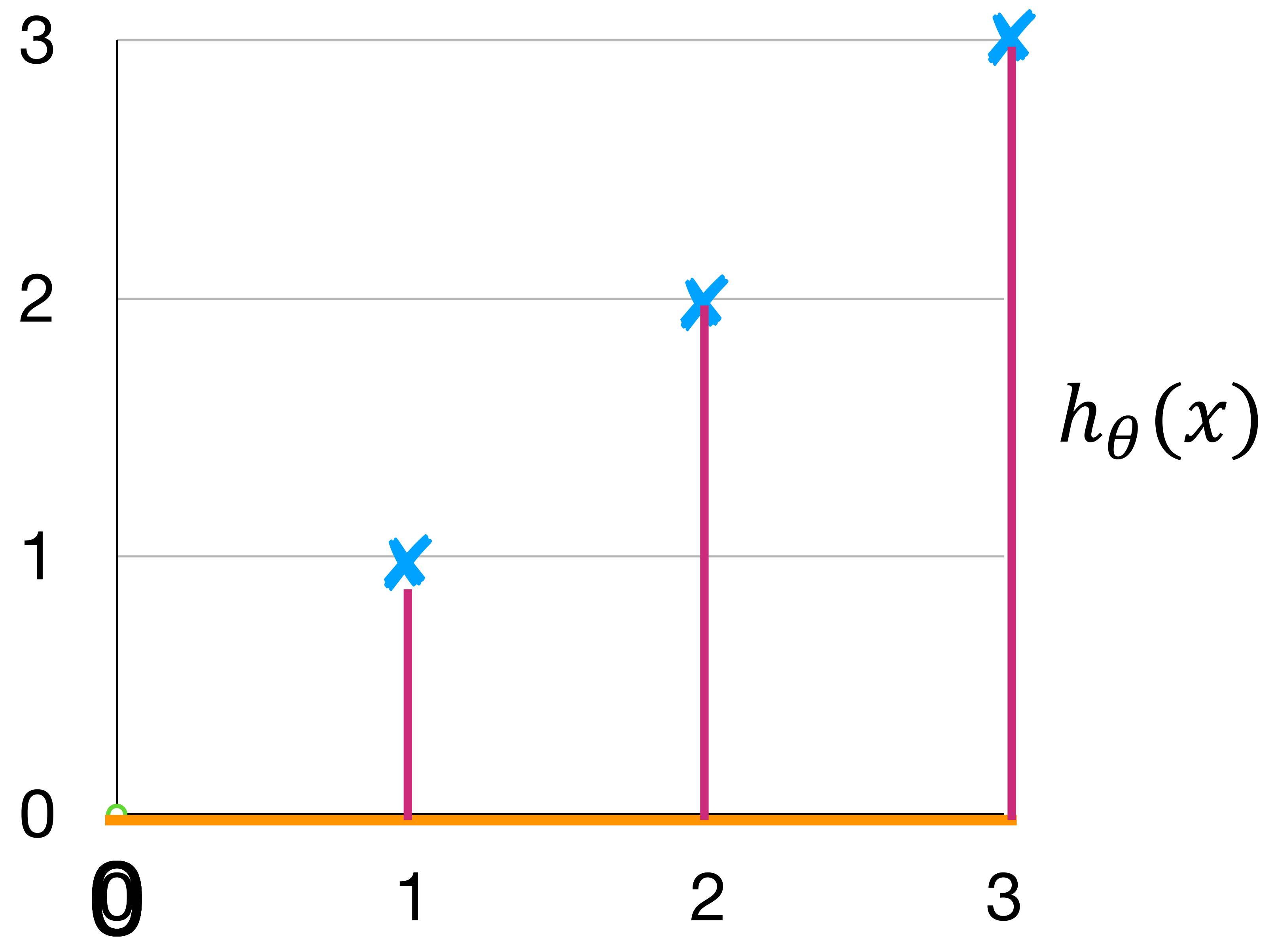
function of the parameter θ_1



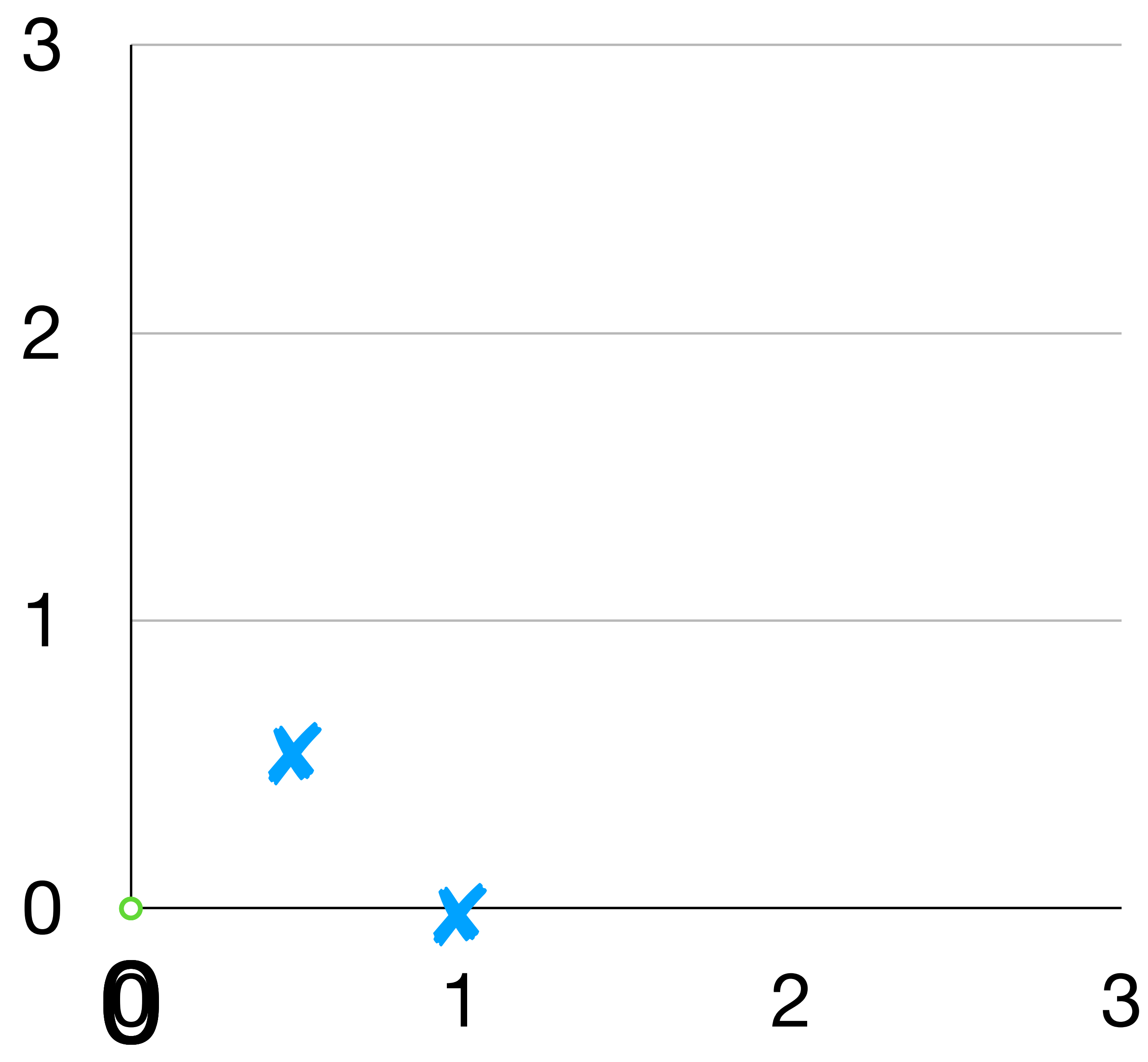
how about $\theta_1 = 0$

$J(\theta_1)$

for fixed θ_1, this is a function of x



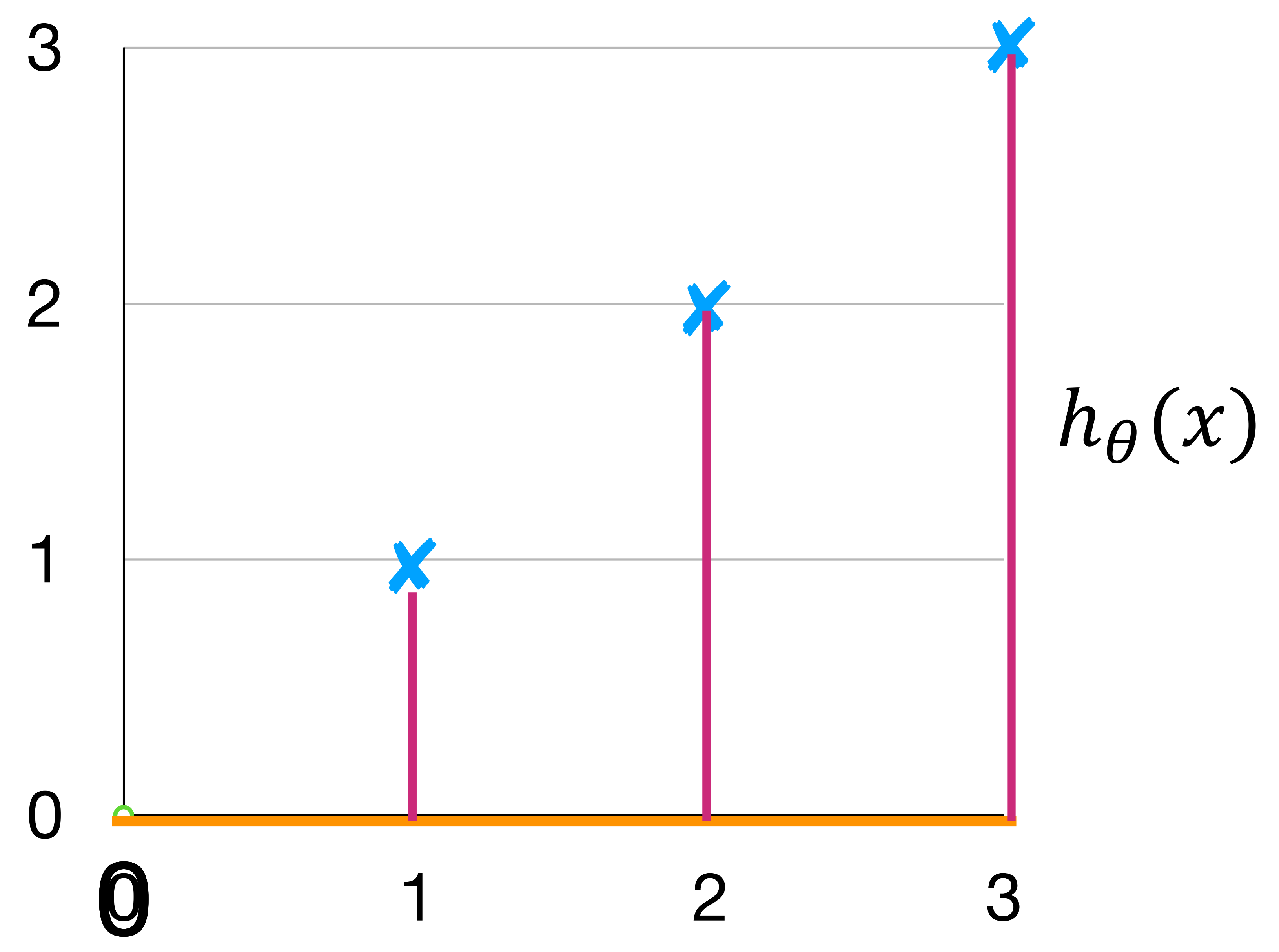
function of the parameter θ_1



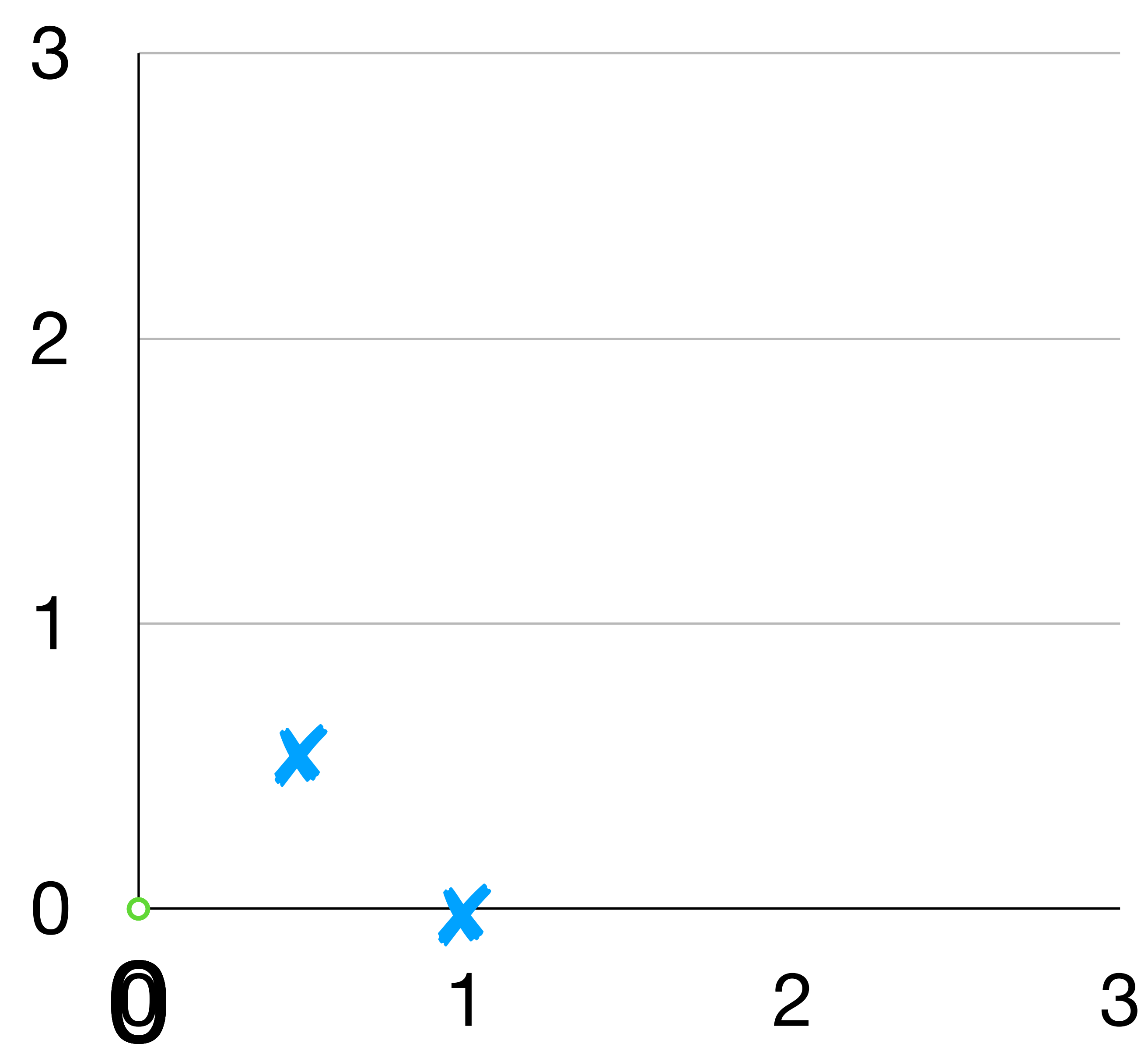
how about $\theta_1 = 0$

$J(\theta_1)$

for fixed θ_1, this is a function of x



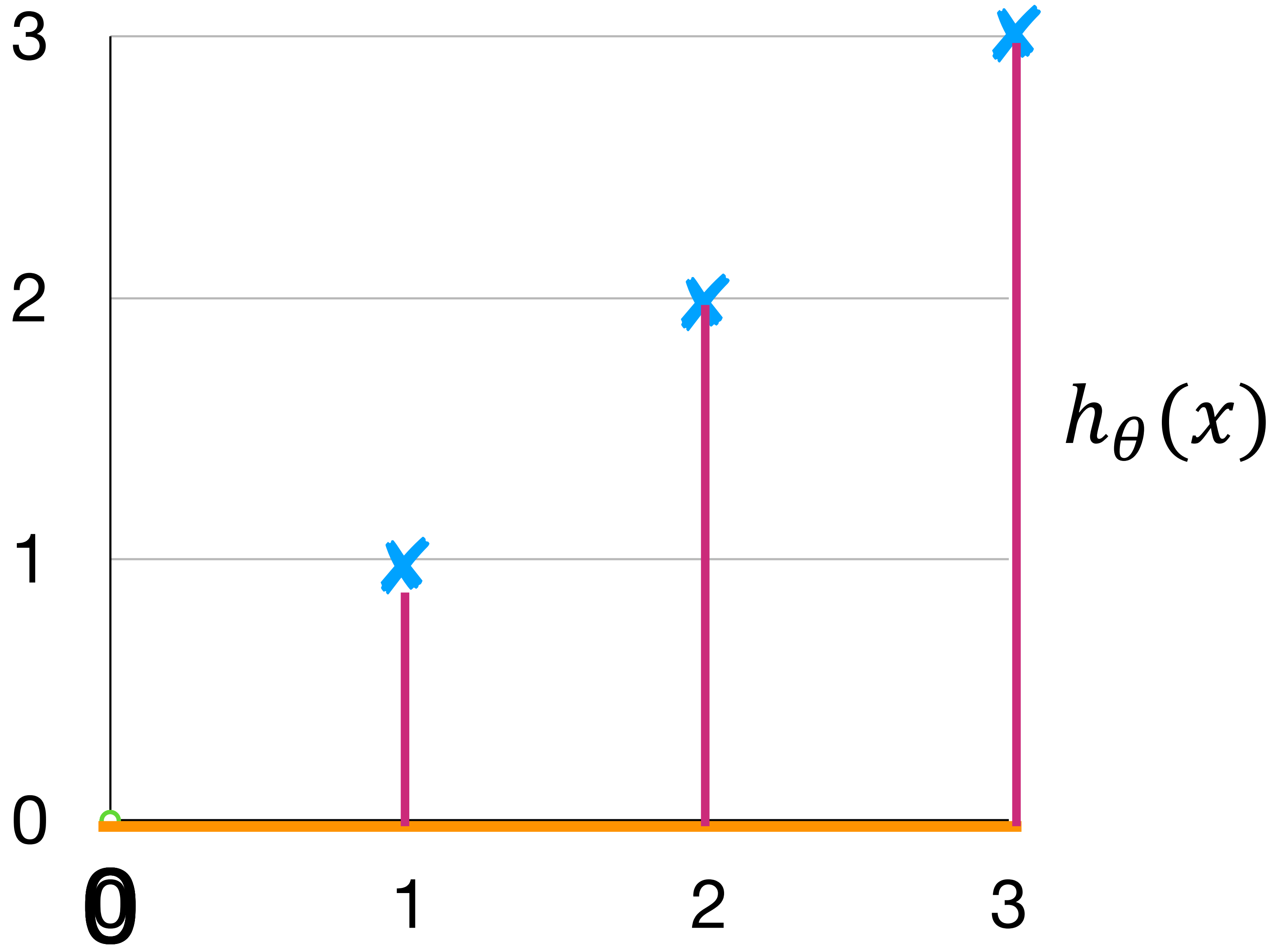
function of the parameter θ_1



$$J(0) = \frac{1}{2m}[(0-1)^2 + (0-2)^2 + (0-3)^2]$$

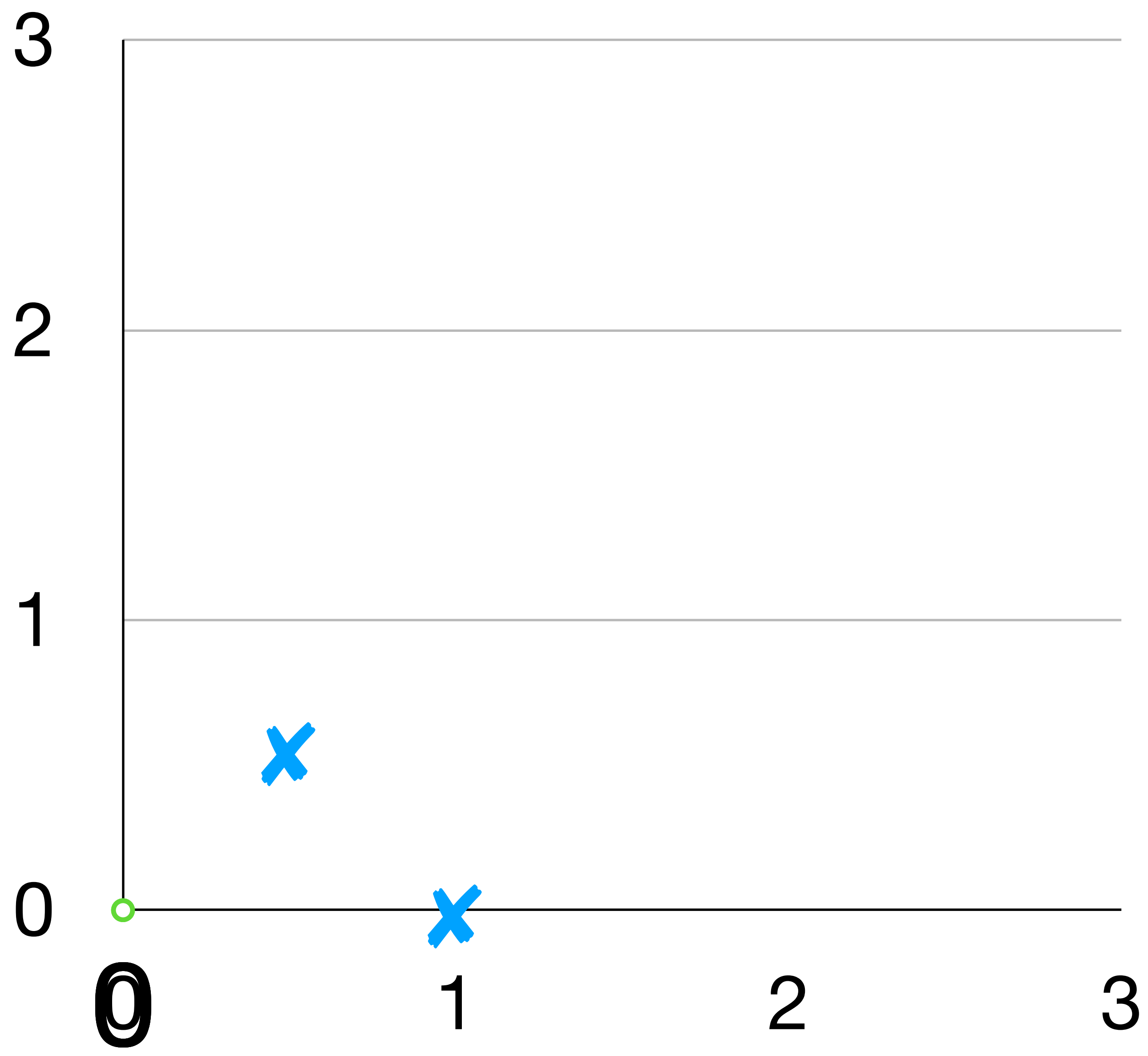
how about $\theta_1 = 0$

for fixed θ_1, this is a function of x



$J(\theta_1)$

function of the parameter θ_1



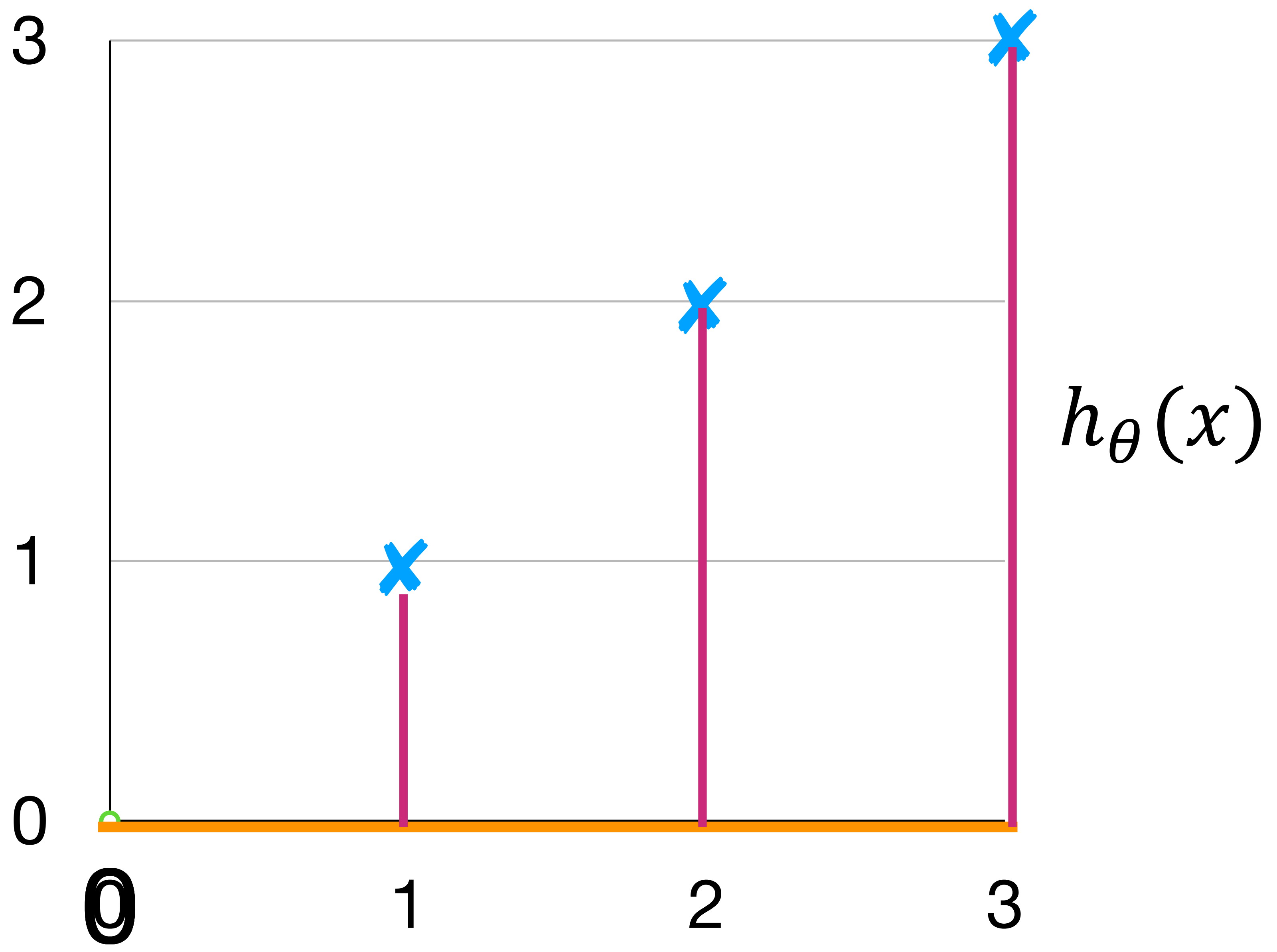
$$J(0) = \frac{1}{2m}[(0-1)^2 + (0-2)^2 + (0-3)^2]$$

$$J(0.5) = \frac{1}{2*3}[1+4+9]$$

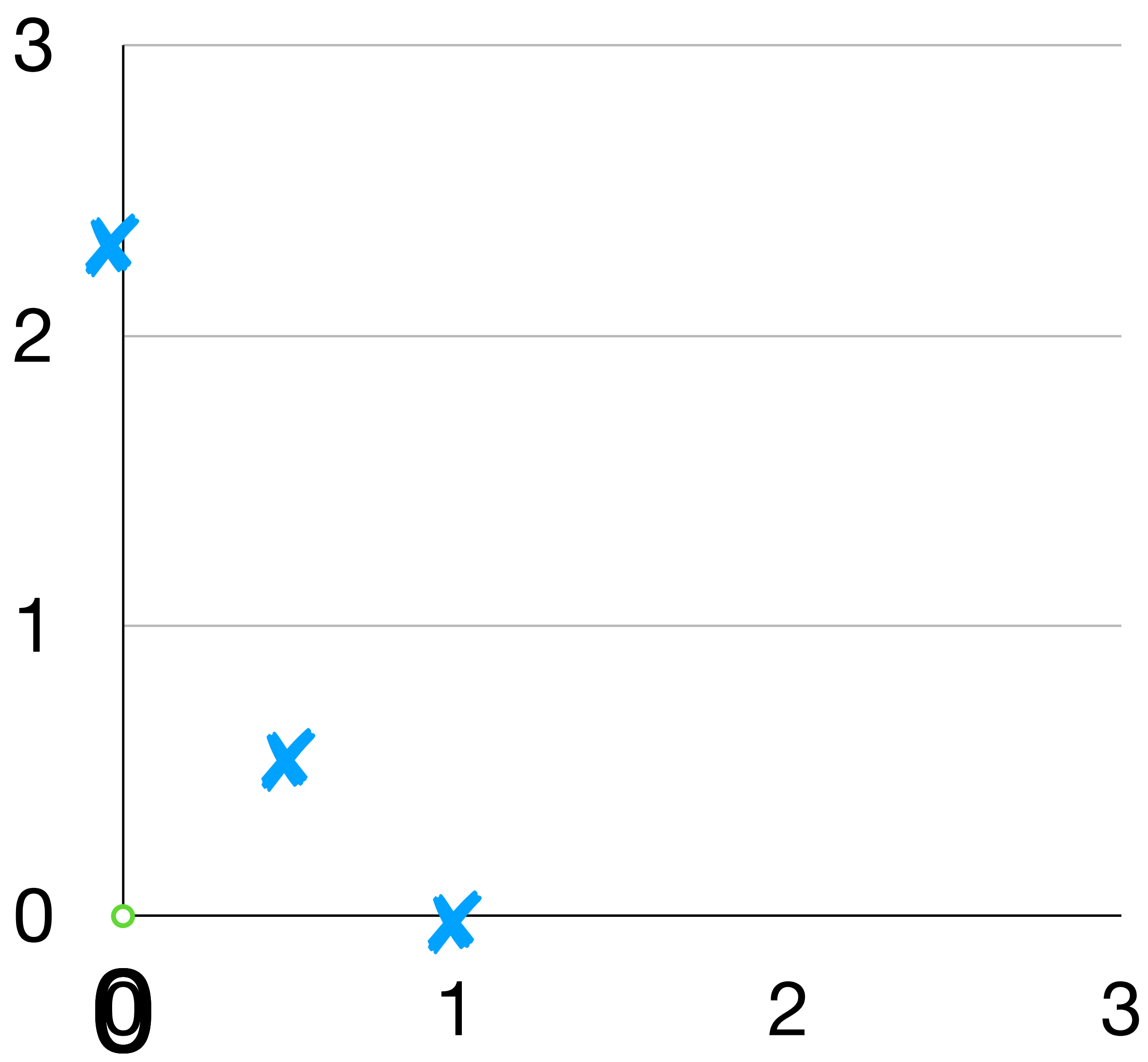
how about $\theta_1 = 0$

$J(\theta_1)$

for fixed θ_1, this is a function of x



function of the parameter θ_1



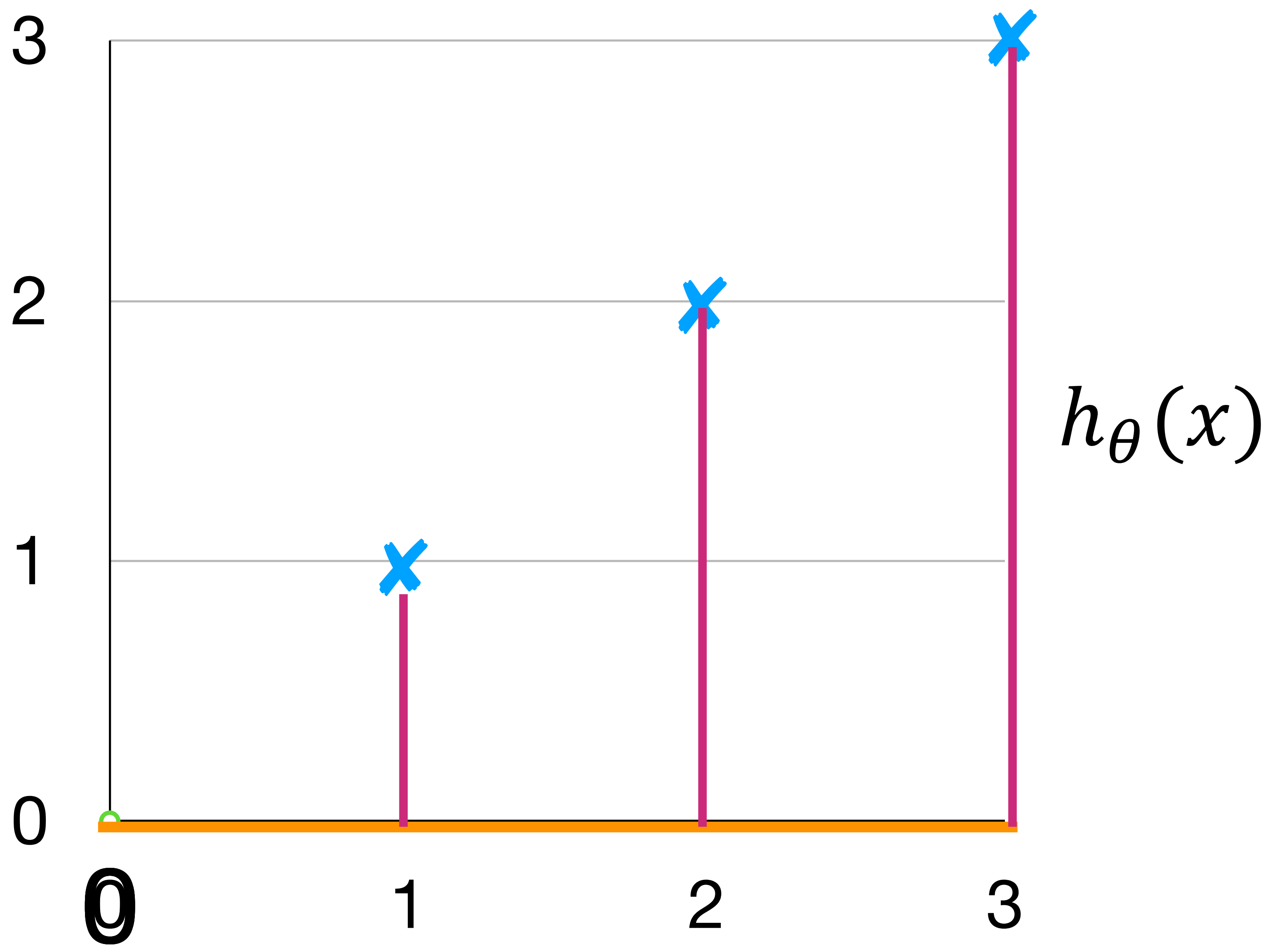
$$J(0) = \frac{1}{2m}[(0-1)^2 + (0-2)^2 + (0-3)^2]$$

$$J(0.5) = \frac{1}{2*3}[1+4+9]$$

$$J(0.5) = \frac{1}{2*3}[14] = \frac{14}{6} = 2.333$$

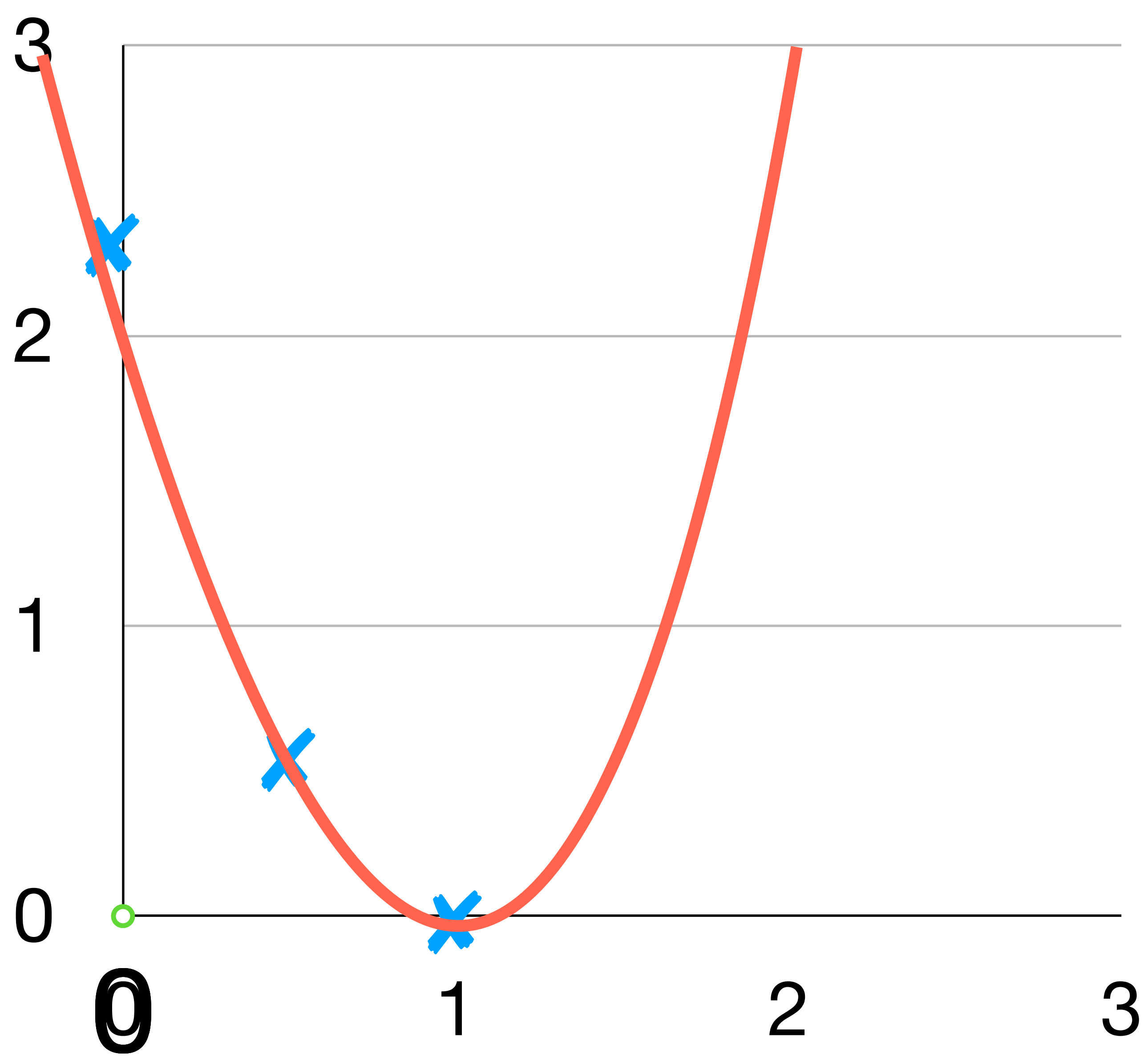
how about $\theta_1 = 0$

for fixed θ_1, this is a function of x



$J(\theta_1)$

function of the parameter θ_1



$$J(0) = \frac{1}{2m}[(0-1)^2 + (0-2)^2 + (0-3)^2]$$

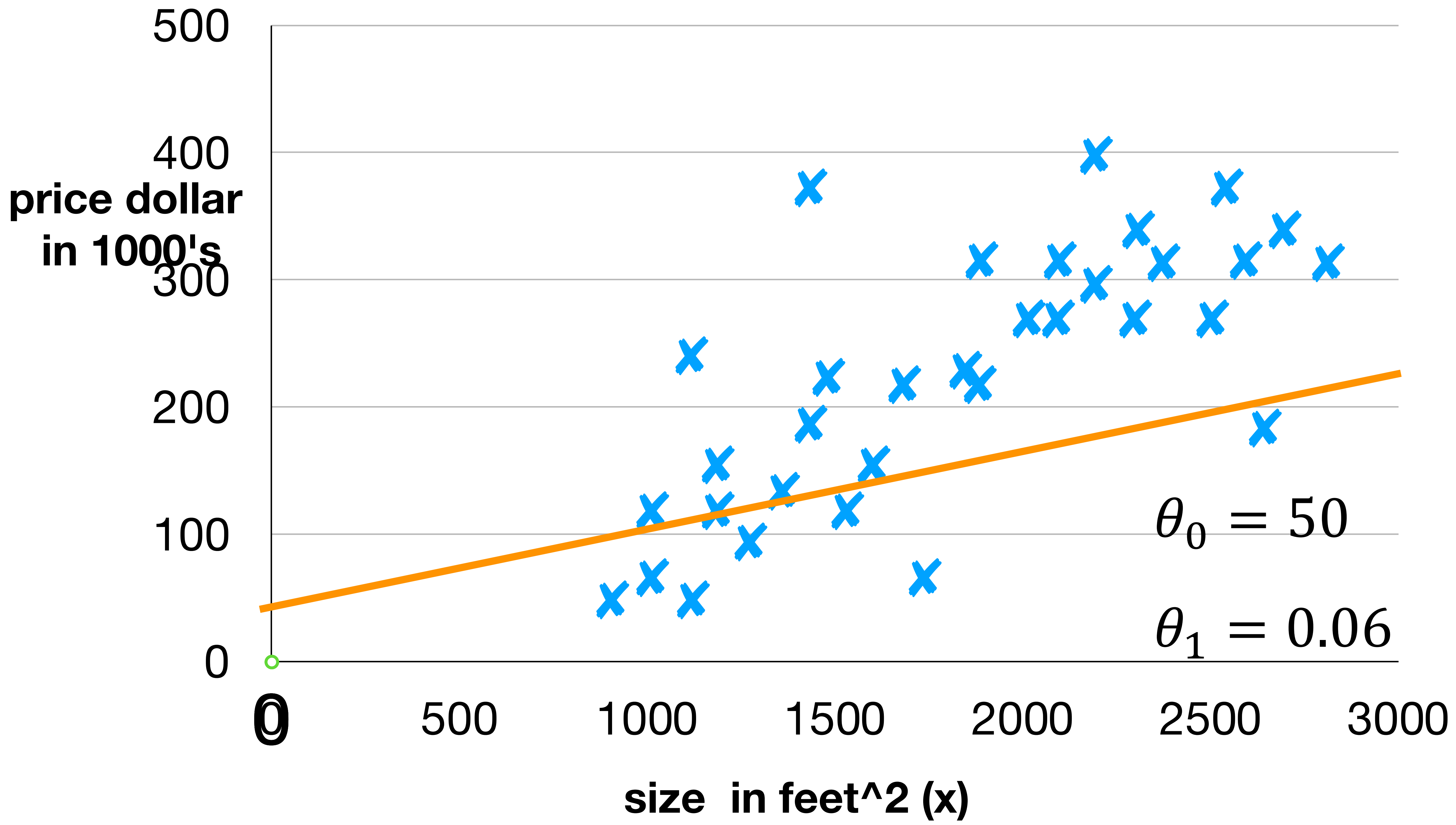
$$J(0.5) = \frac{1}{2*3}[1+4+9]$$

$$J(0.5) = \frac{1}{2*3}[14] = \frac{14}{6} = 2.333$$

Cost Function Intuition 2

$$h_\theta(x)$$

for fixed θ0, θ1 , this is a function of x

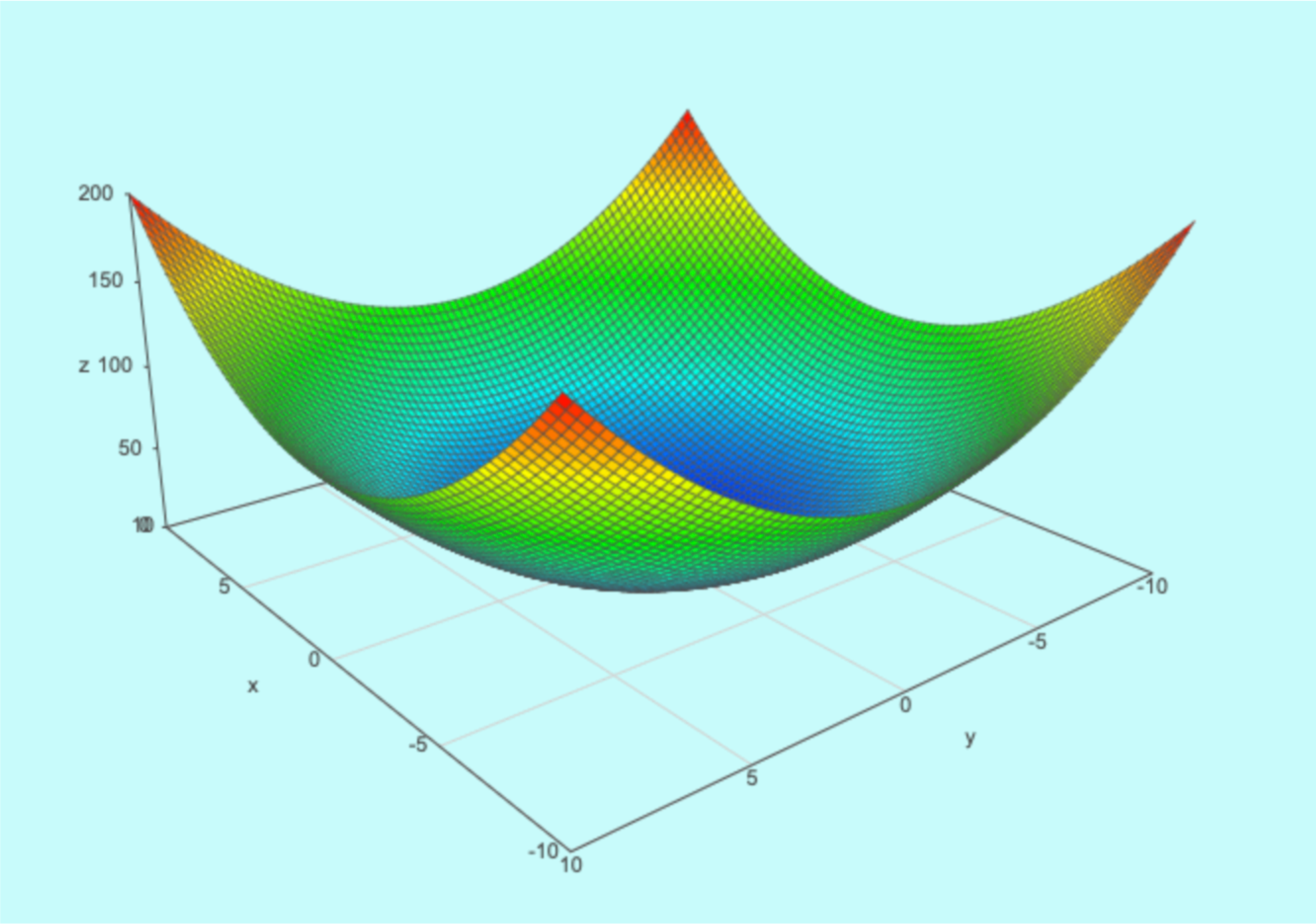


$$h_\theta(x) = 50 + 0.06x$$

$$J(\theta_o, \theta_1)$$

function of parameters θ0, θ1

?



http://al-roomi.org/3DPlot/index.html

02. Gradient Descent

Gradient Descent Algorithm

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) (for j = 0 and j = 1)$$

Correct

$$Temp_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$Temp_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_0 := Temp_0$$

$$\theta_1 := Temp_1$$

Simultaneous Update

Incorrect

$$Temp_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_0 := Temp_0$$

$$Temp_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_1 := Temp_1$$

Initial Value

$\theta_0 := 1 \quad \theta_1 := 2$

Update rule

$$\theta_j := \theta_j + \sqrt{\theta_0 + \theta_1} (for j = 0 and j = 1)$$

Update Result

Initial Value

$\theta_0 := 1 \quad \theta_1 := 2$

Update rule

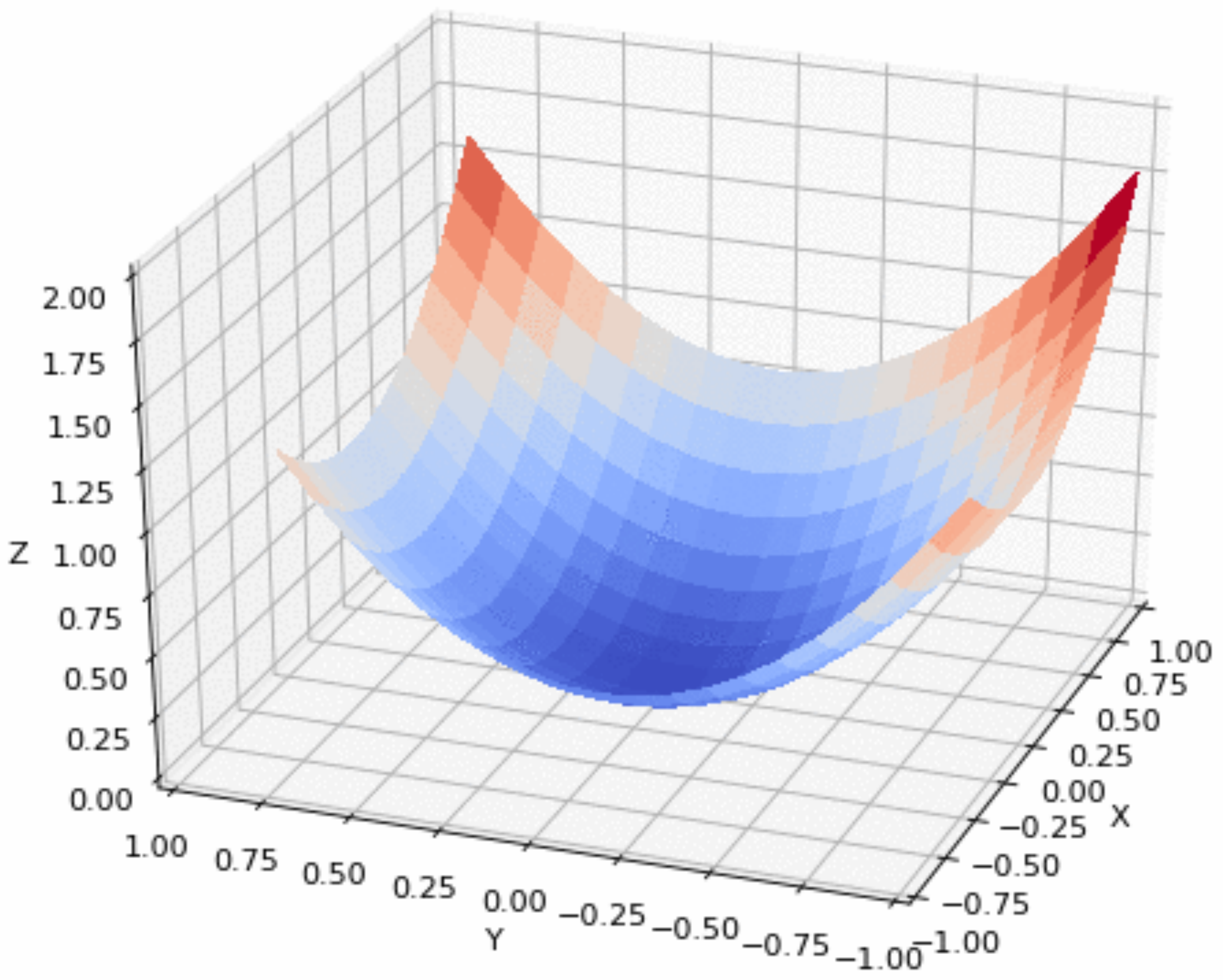
$$\theta_j := \theta_j + \sqrt{\theta_0 + \theta_1} (for j = 0 and j = 1)$$

Update Result

$\theta_0 = 1 + \sqrt{2} \qquad \theta_1 = 2 + \sqrt{2}$

Gradient Descent with Learning rate

Optimal Learning Rate



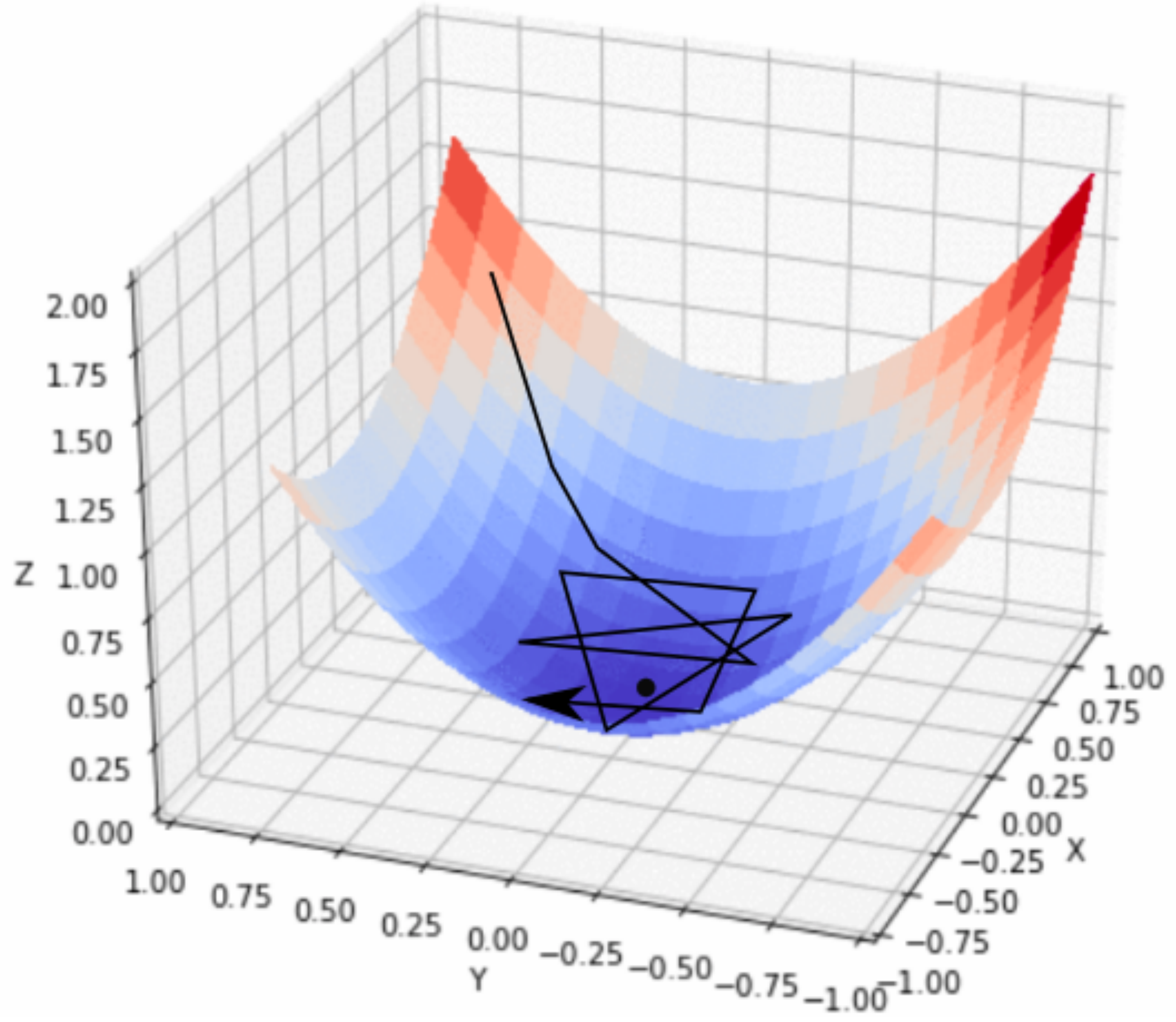
$$Temp_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$Temp_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_0 := Temp_0$$

$$\theta_1 := Temp_1$$

Too Large Learning Rate



Gradient Descent Algorithm

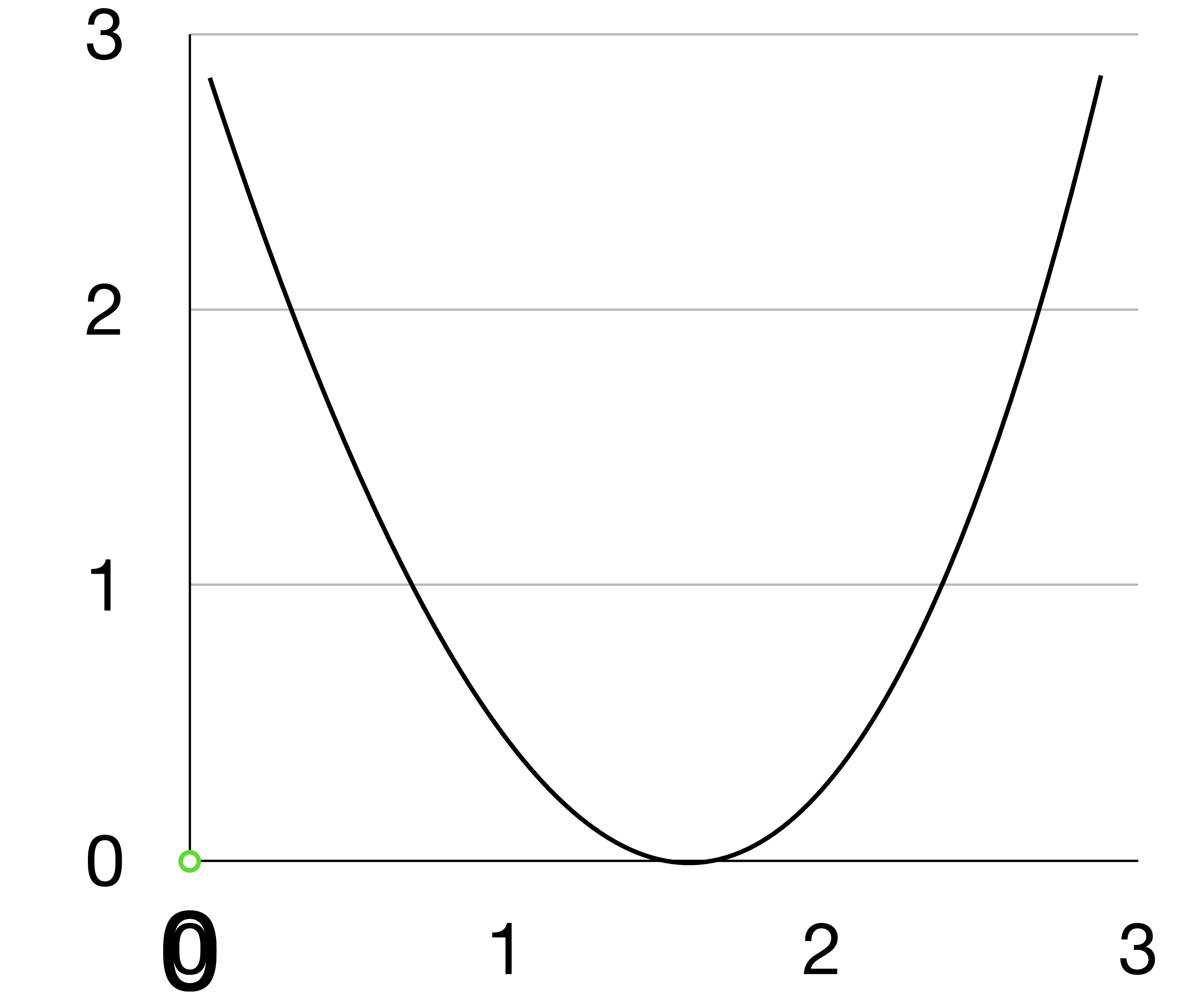
특정 값으로 수렴할때 까지 반복

$$\theta_j := \theta_j - \boxed{\alpha} \boxed{\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)} (forj = 0andj = 1)$$

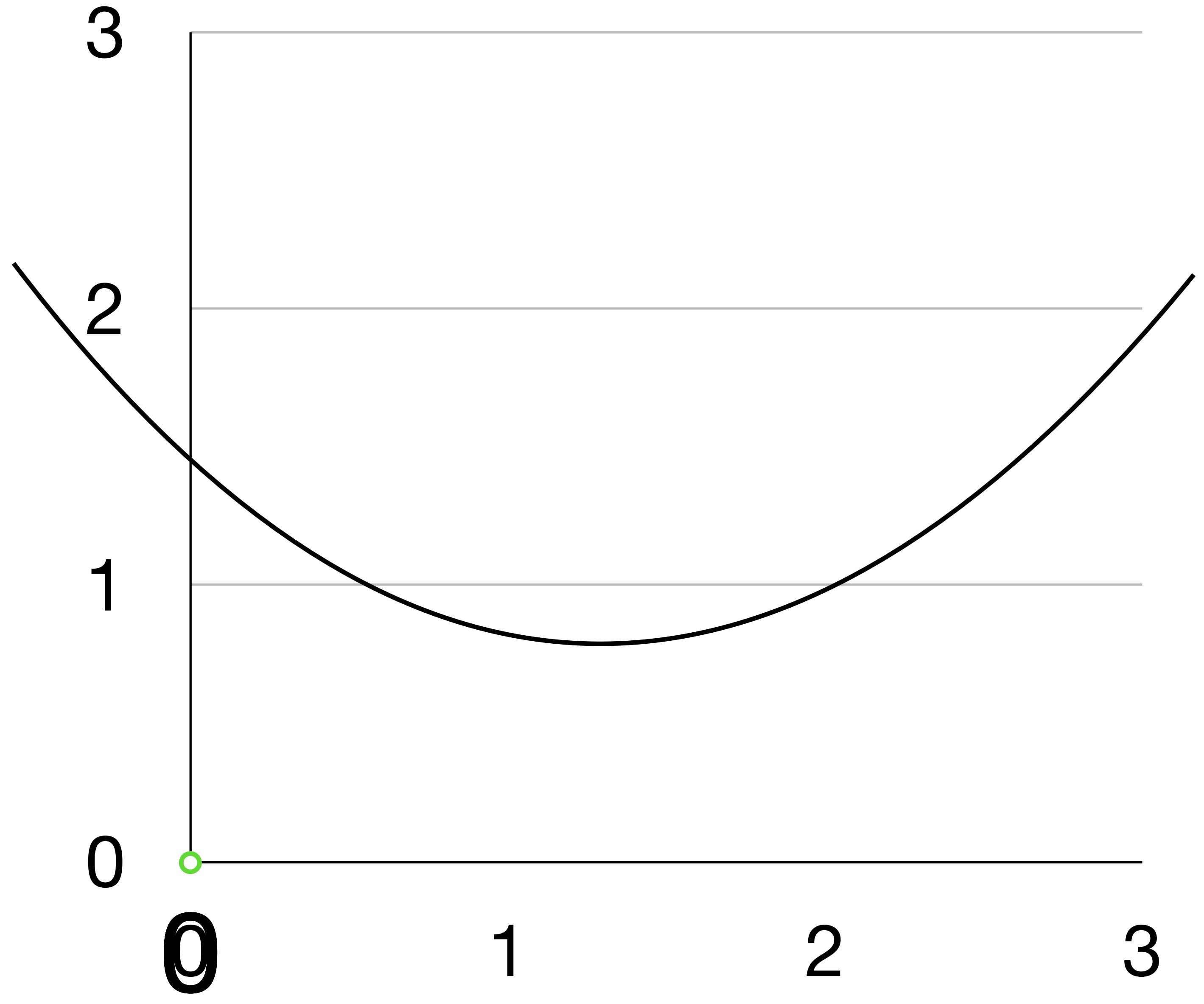
편미분계수

학습률(Learning Rate)

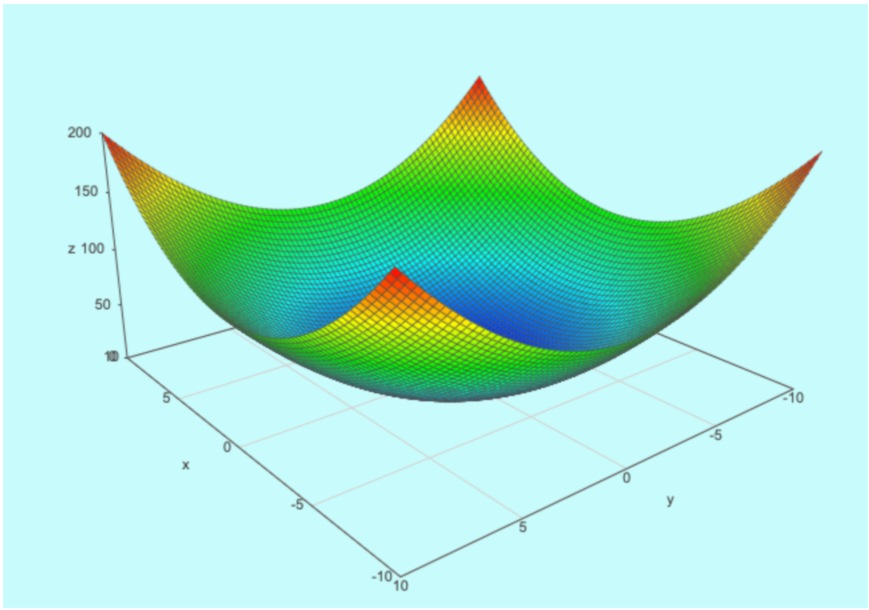
$J(\theta_0)$



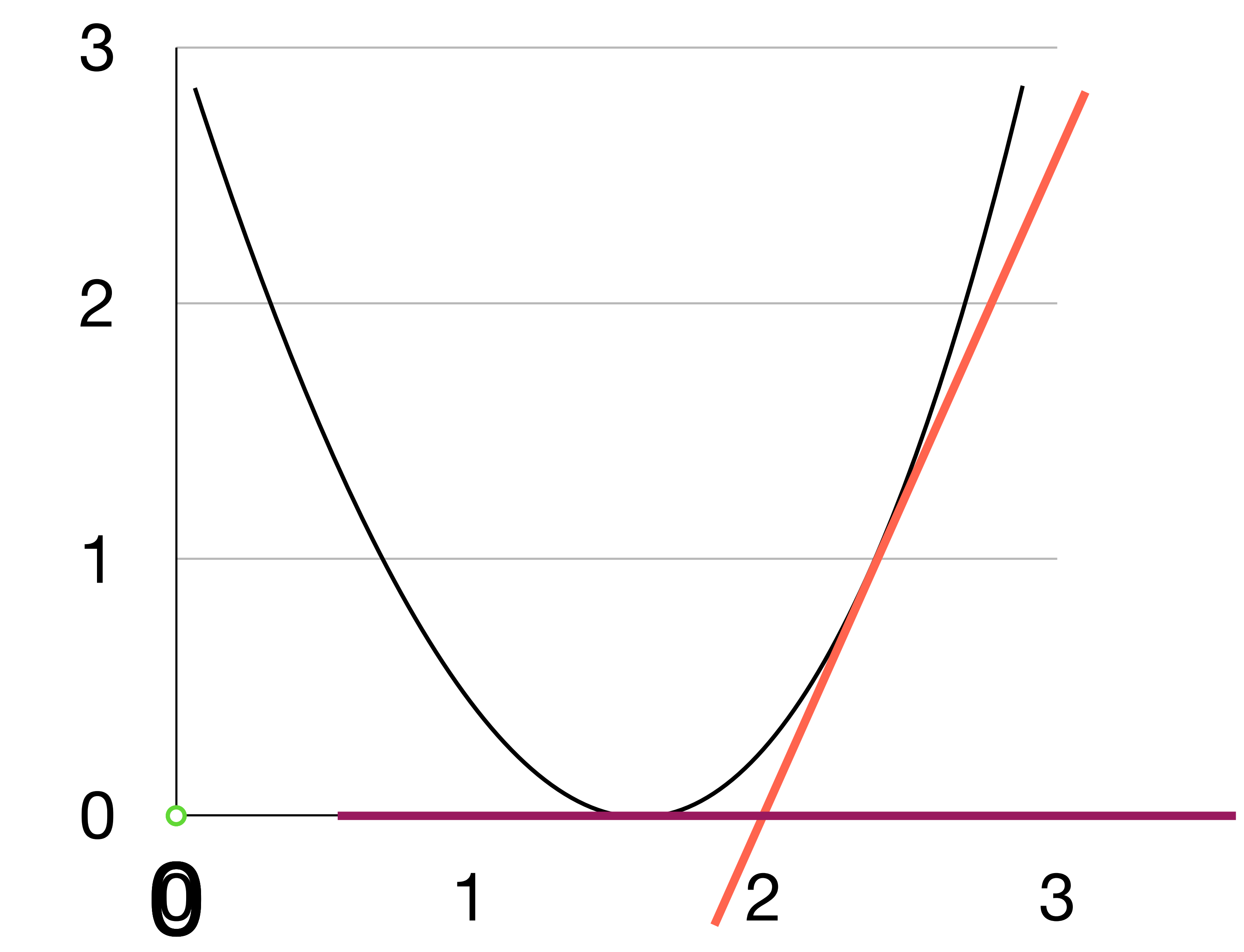
$J(\theta_1)$



$J(\theta_0, \theta_1)$

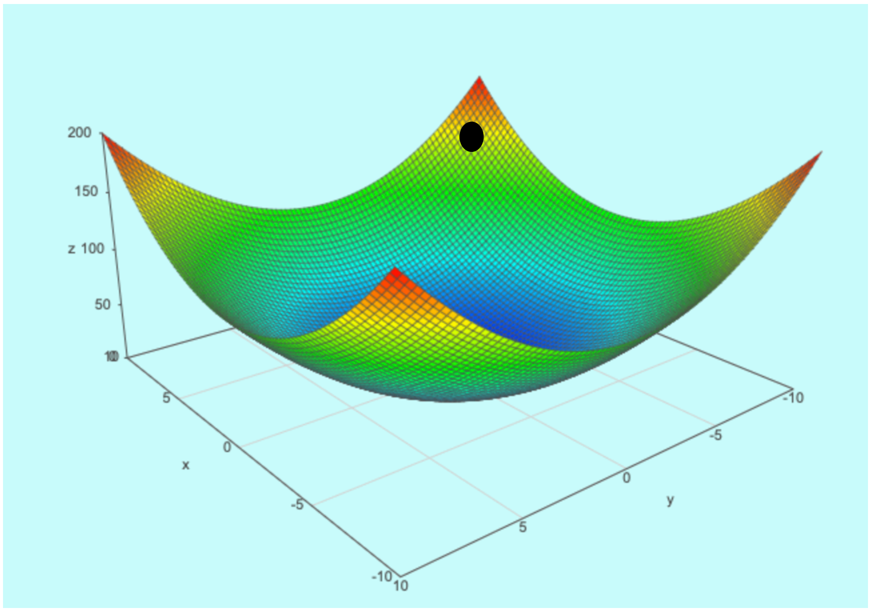


$J(\theta_0)$



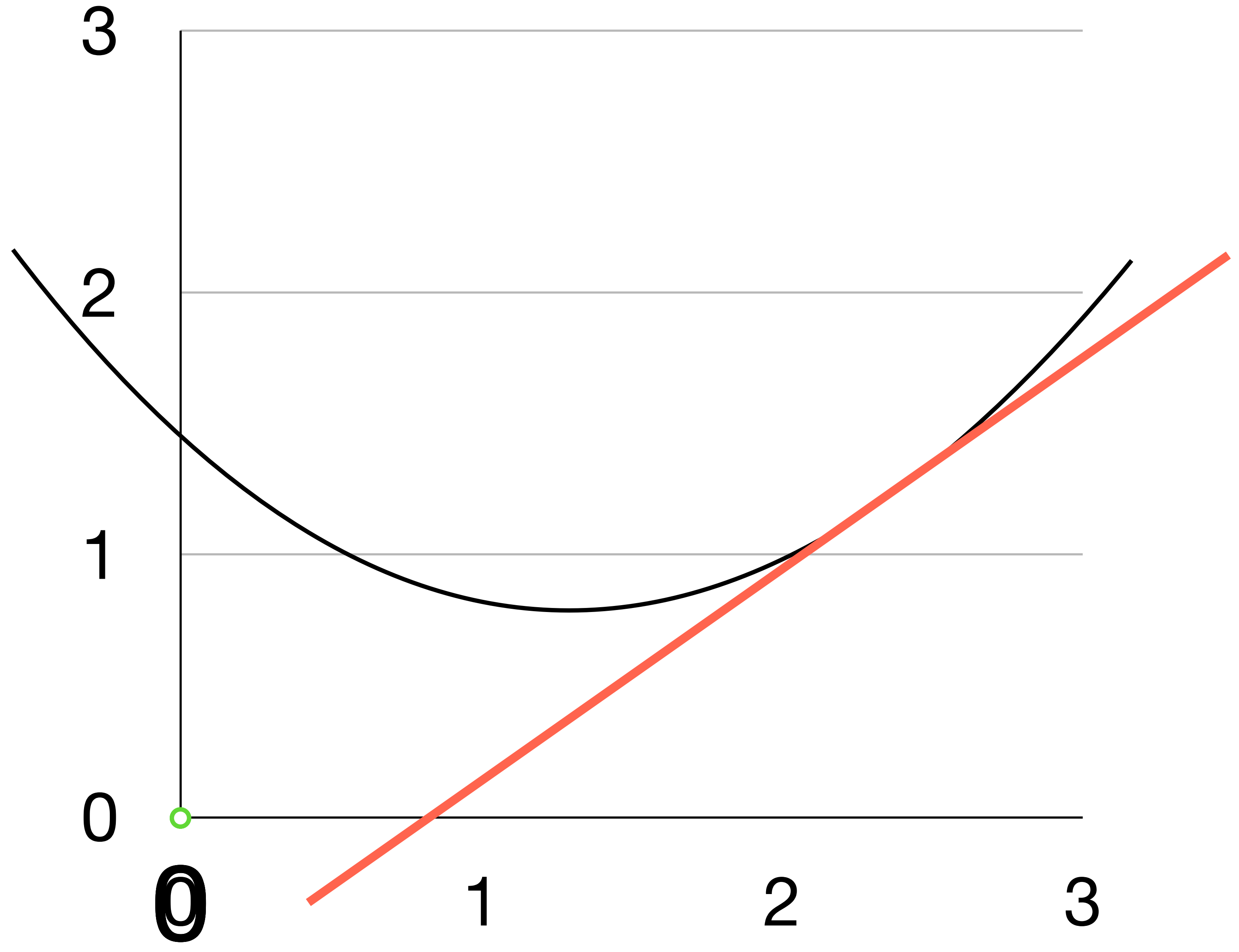
$$\theta_{0'} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$J(\theta_0, \theta_1)$

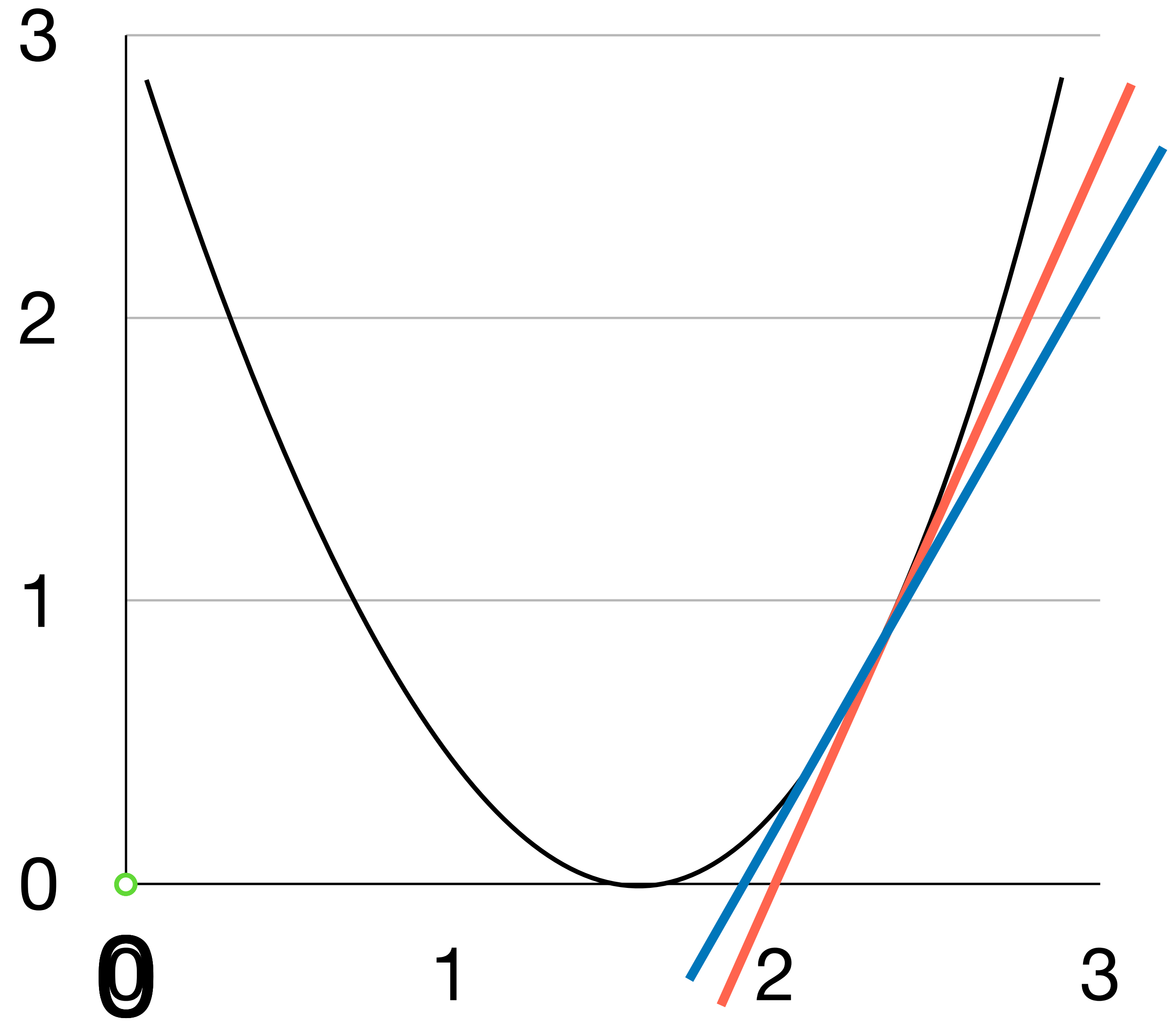


$$\theta_{1'} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$J(\theta_1)$



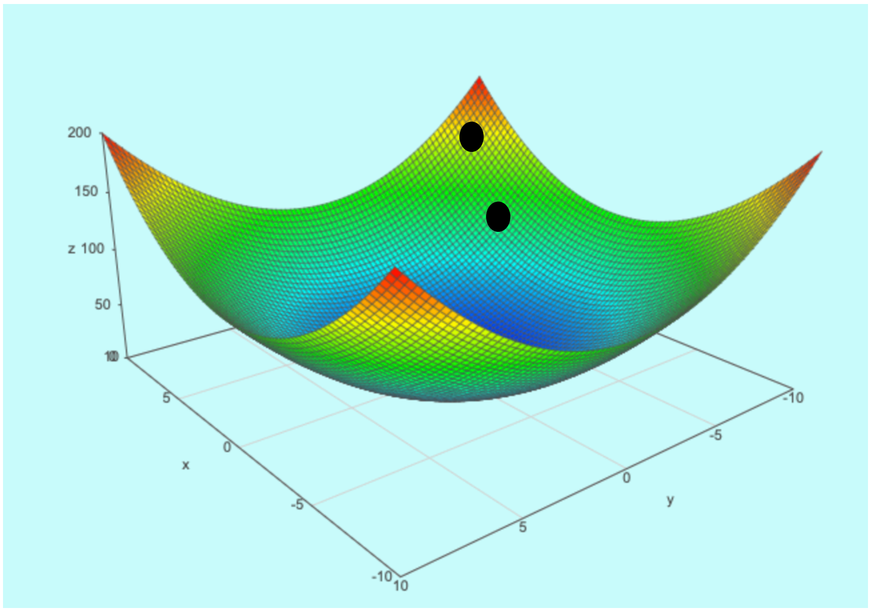
$J(\theta_0)$



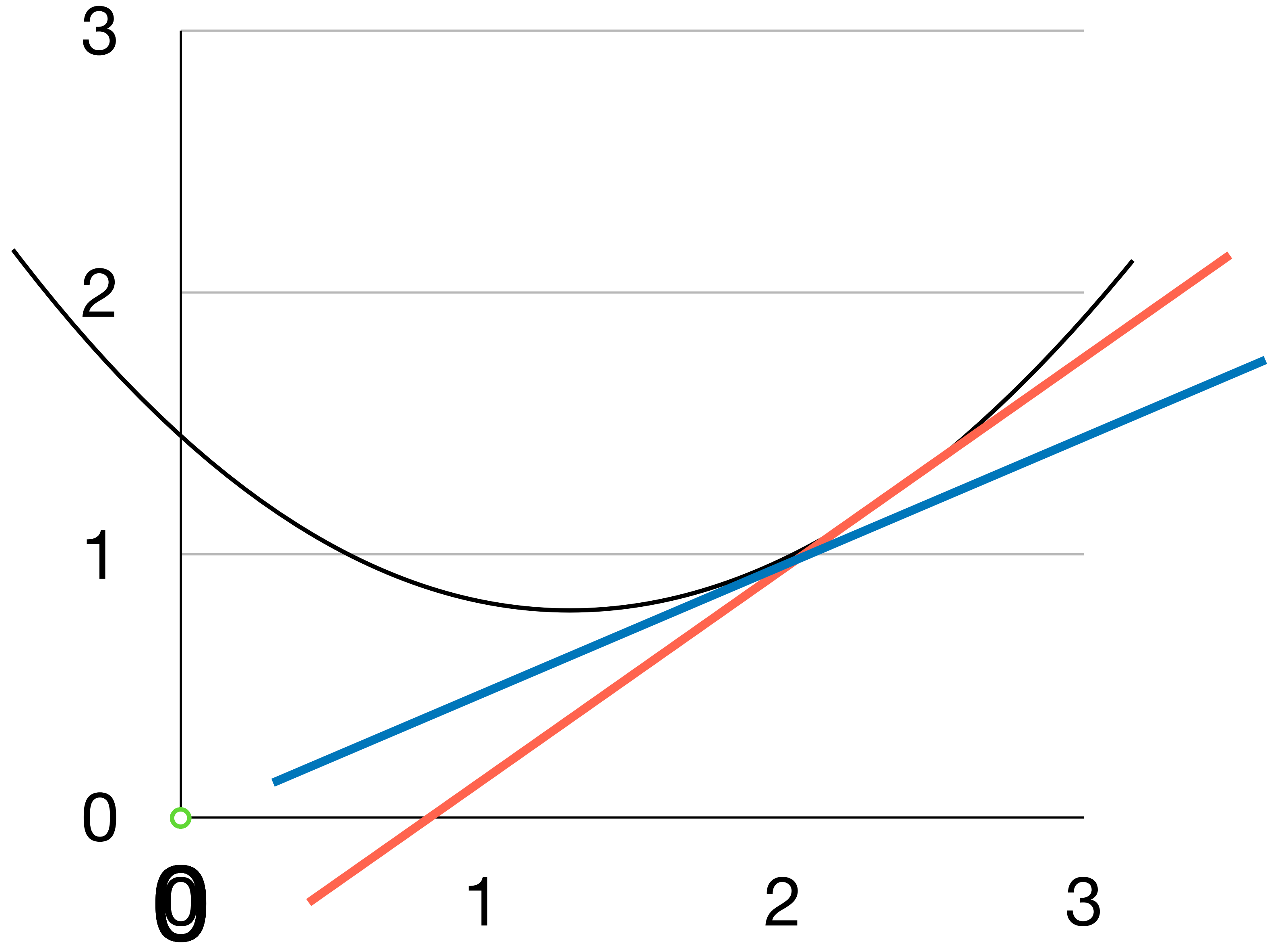
$$\theta_{0'} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_{0''} := \theta_{0'} - \alpha \frac{\partial}{\partial \theta_{0'}} J(\theta_0, \theta_1)$$

$J(\theta_0, \theta_1)$



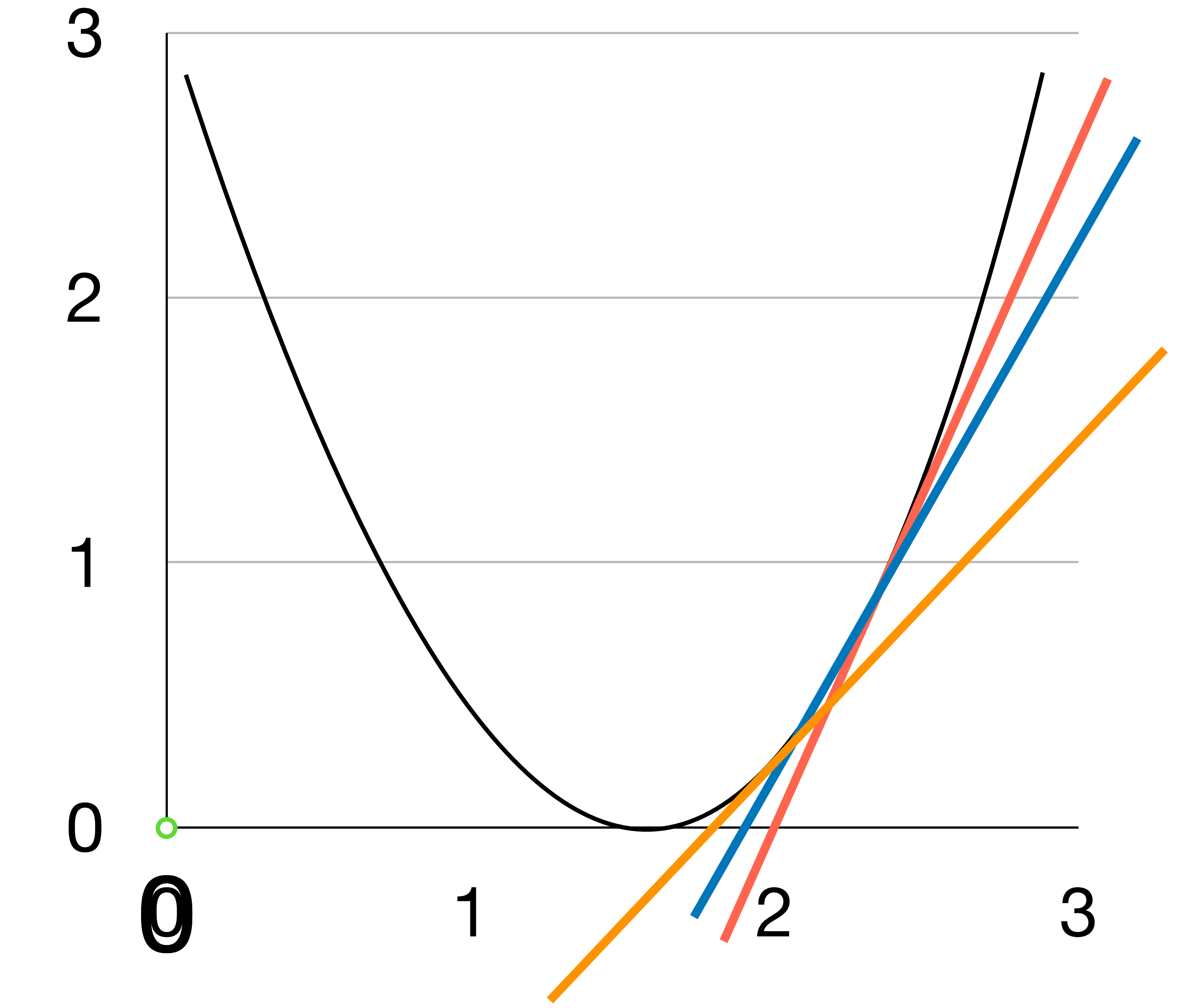
$J(\theta_1)$



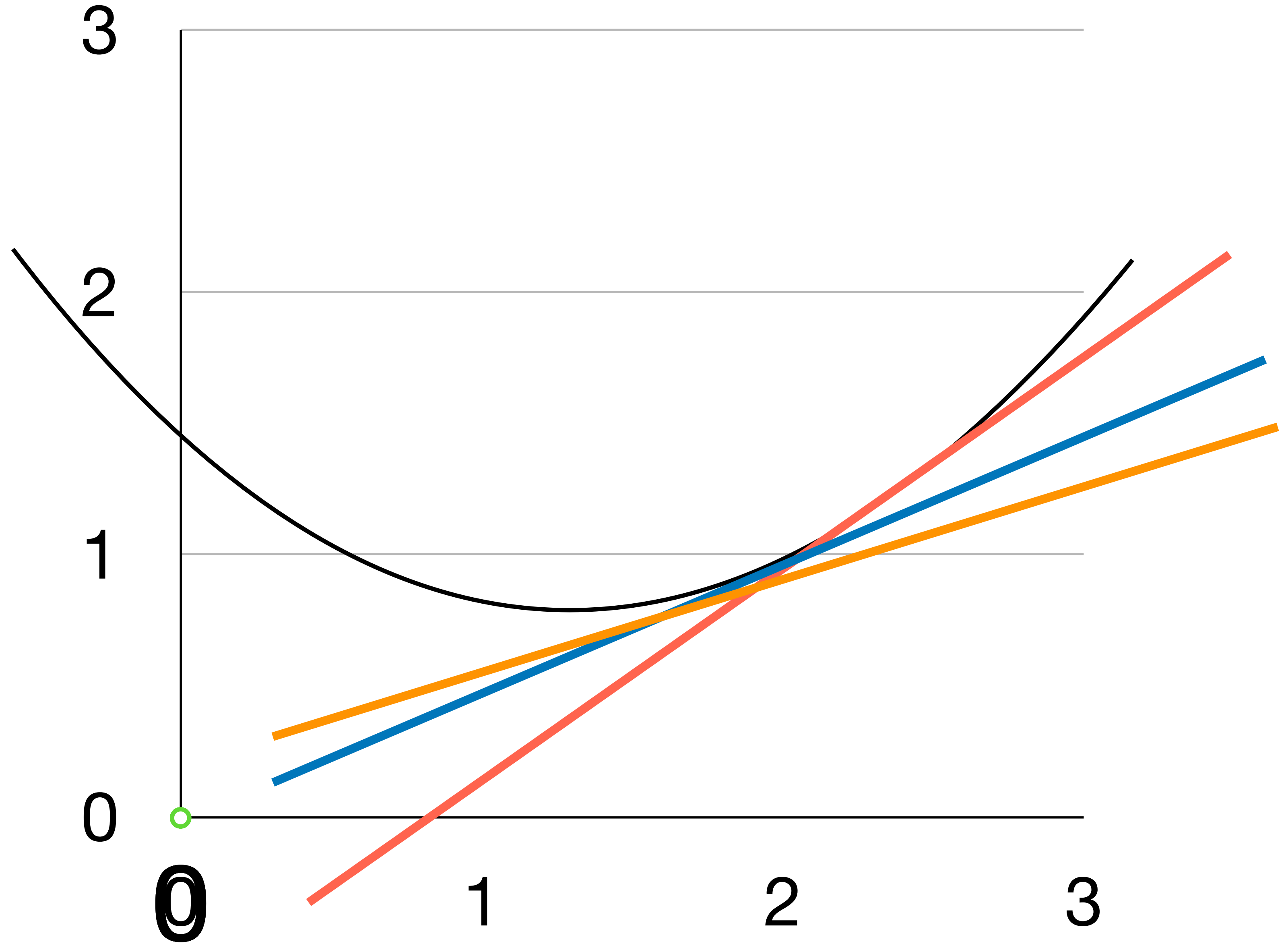
$$\theta_{1'} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_{1''} := \theta_{1'} - \alpha \frac{\partial}{\partial \theta_{1'}} J(\theta_0, \theta_1)$$

$J(\theta_0)$



$J(\theta_1)$

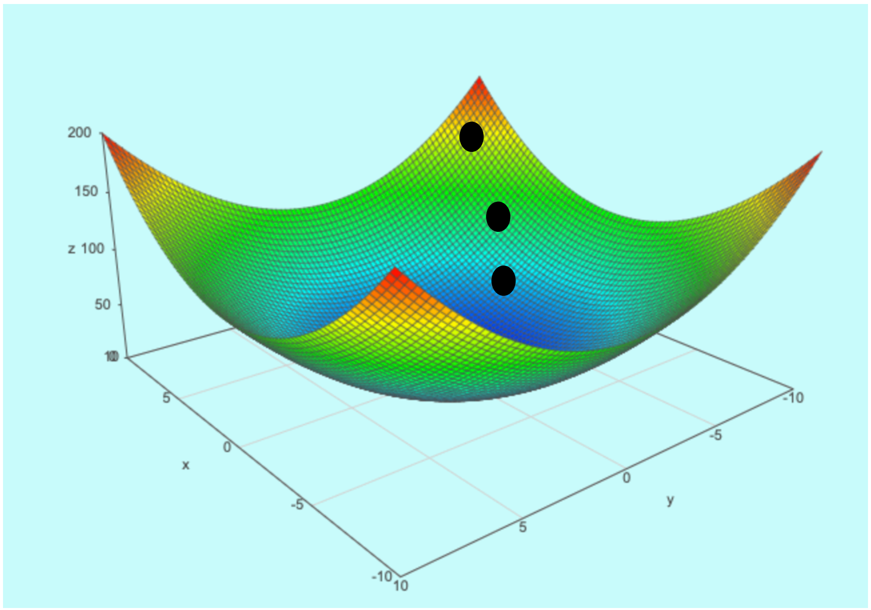


$$\theta_{0'} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_{0''} := \theta_{0'} - \alpha \frac{\partial}{\partial \theta_{0'}} J(\theta_0, \theta_1)$$

$$\theta_{0'''} := \theta_{0''} - \alpha \frac{\partial}{\partial \theta_{0''}} J(\theta_0, \theta_1)$$

$J(\theta_0, \theta_1)$

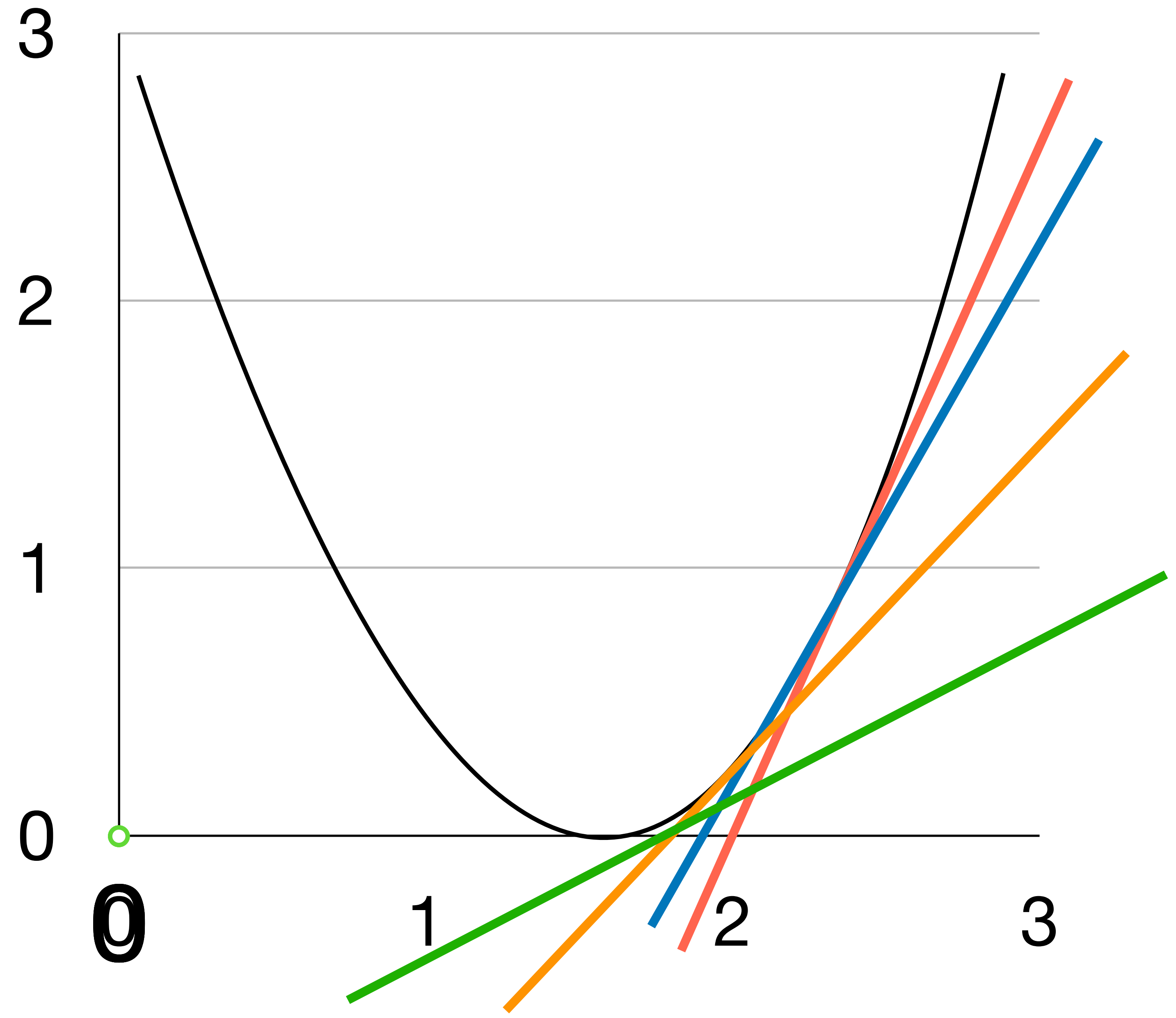


$$\theta_{1'} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

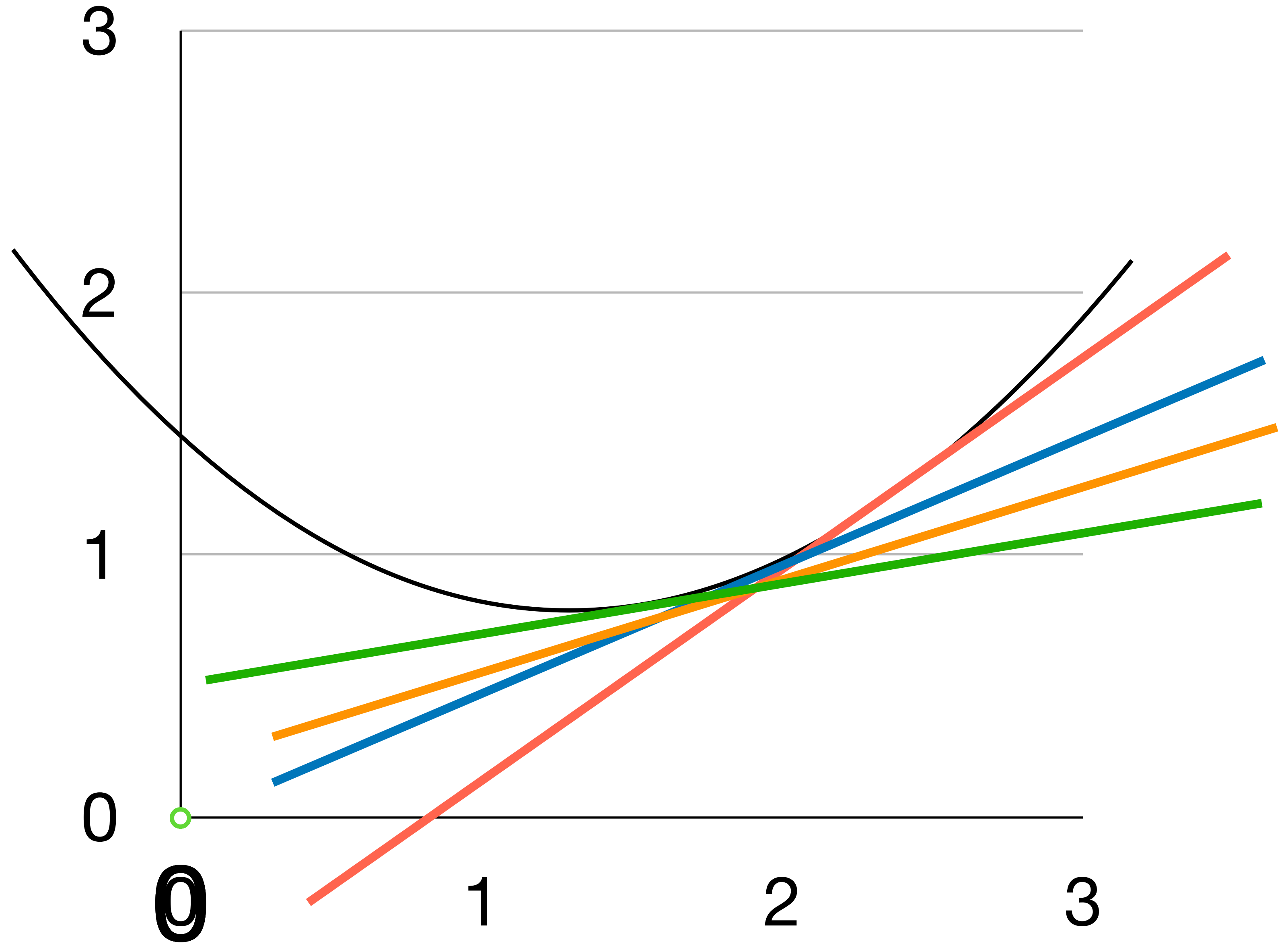
$$\theta_{1''} := \theta_{1'} - \alpha \frac{\partial}{\partial \theta_{1'}} J(\theta_0, \theta_1)$$

$$\theta_{1'''} := \theta_{1''} - \alpha \frac{\partial}{\partial \theta_{1''}} J(\theta_0, \theta_1)$$

$J(\theta_0)$



$J(\theta_1)$



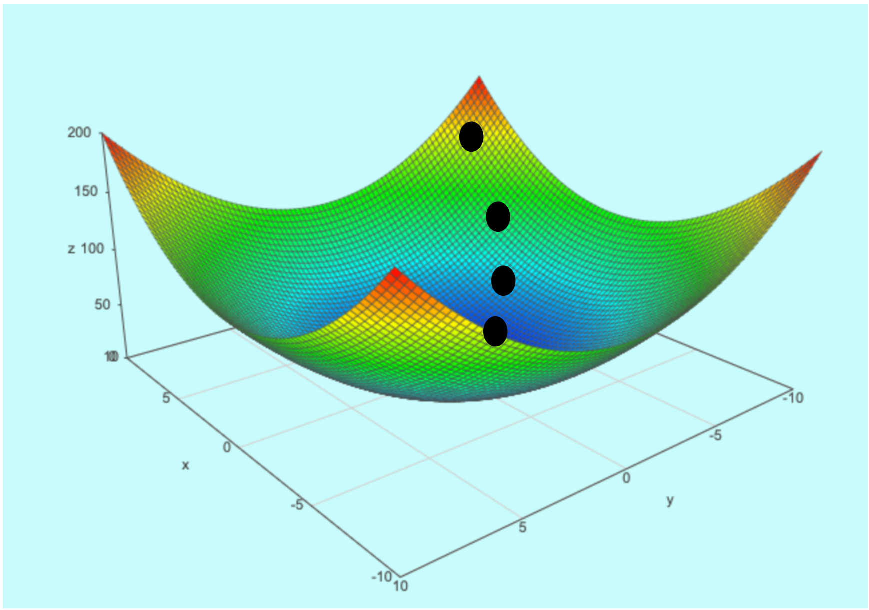
$$\theta_{0'} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_{0''} := \theta_{0'} - \alpha \frac{\partial}{\partial \theta_{0'}} J(\theta_0, \theta_1)$$

$$\theta_{0'''} := \theta_{0''} - \alpha \frac{\partial}{\partial \theta_{0''}} J(\theta_0, \theta_1)$$

$$\theta_{0''''} := \theta_{0'''} - \alpha \frac{\partial}{\partial \theta_{0'''}} J(\theta_0, \theta_1)$$

$J(\theta_0, \theta_1)$



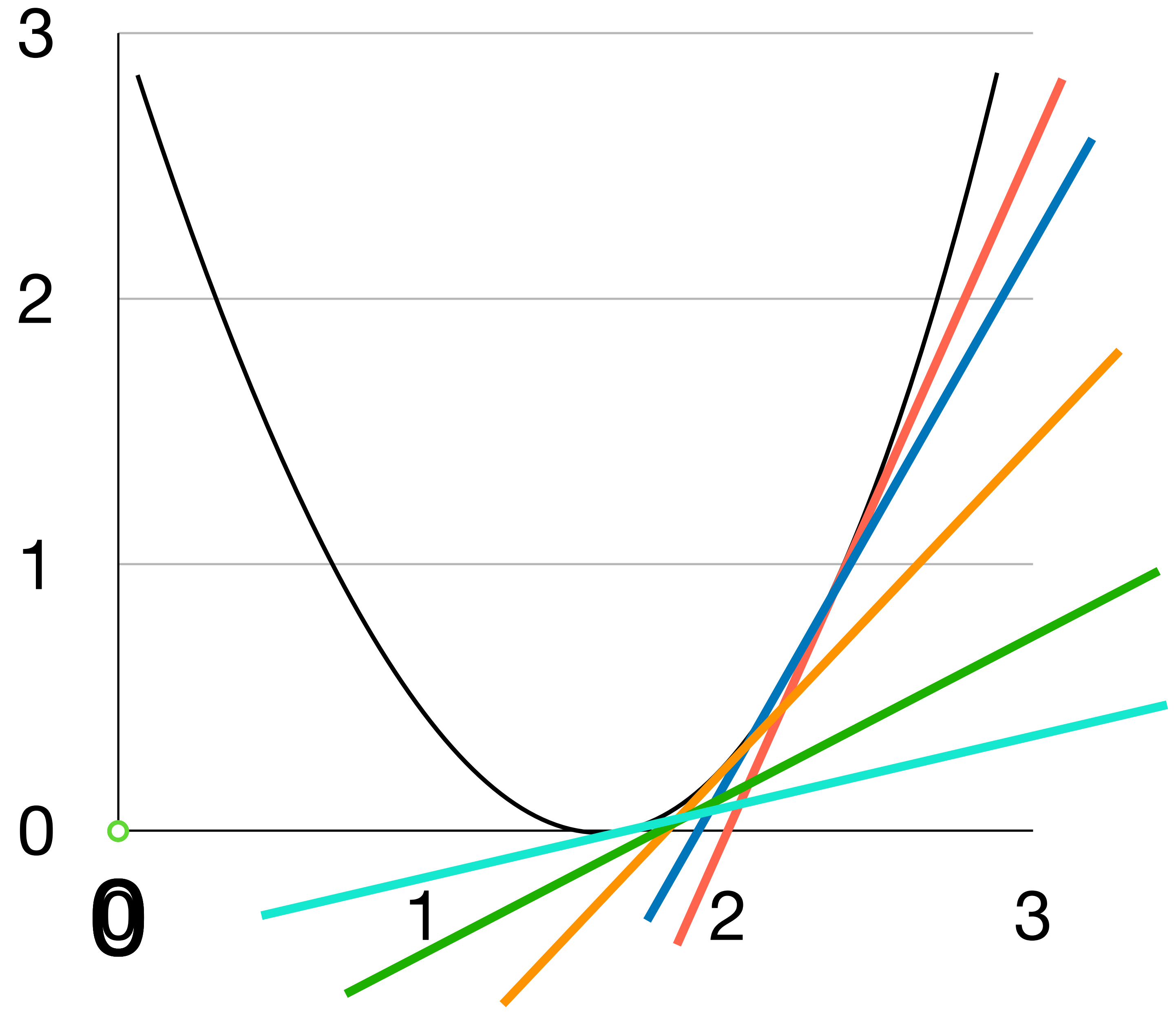
$$\theta_{1'} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_{1''} := \theta_{1'} - \alpha \frac{\partial}{\partial \theta_{1'}} J(\theta_0, \theta_1)$$

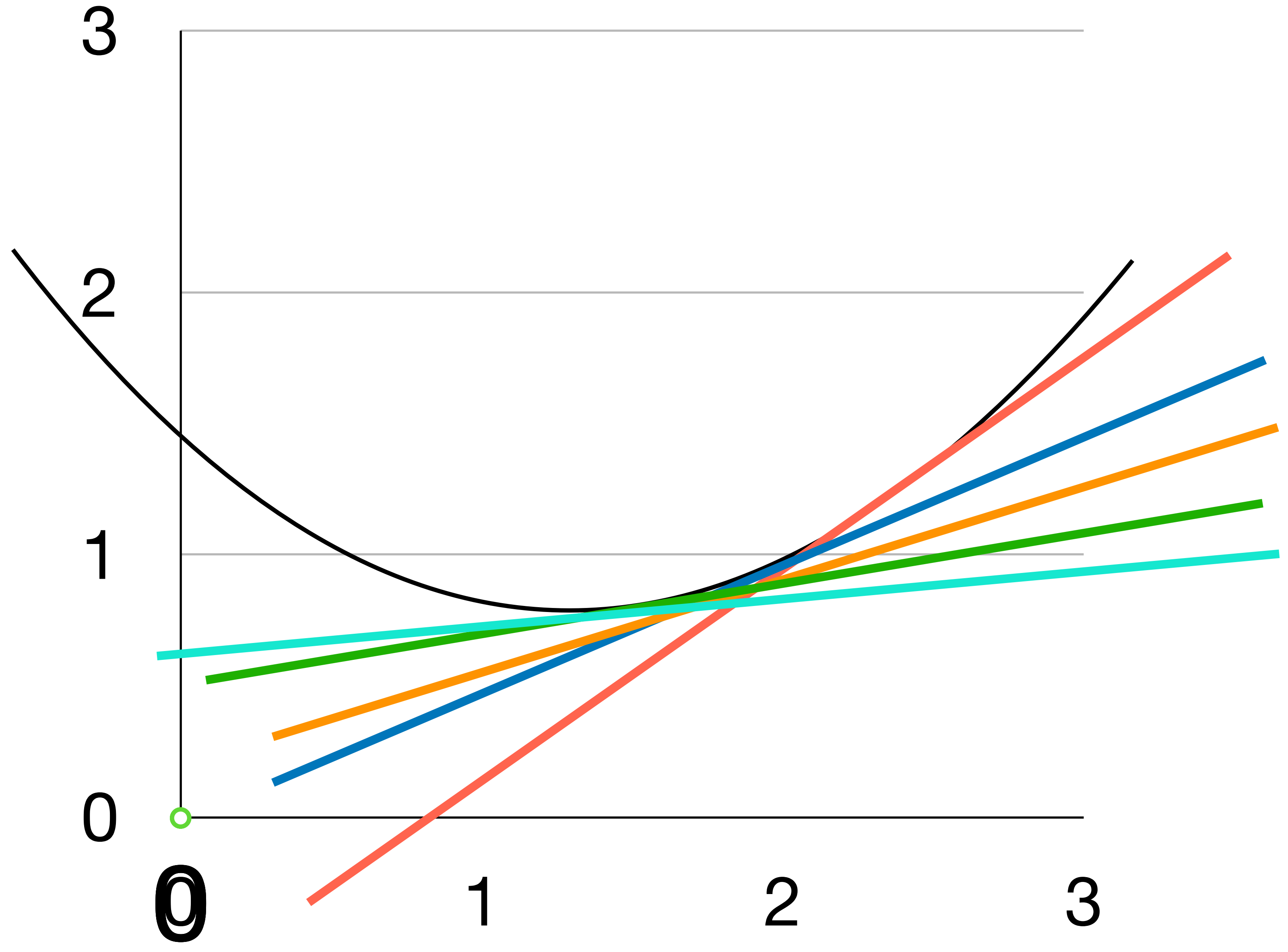
$$\theta_{1'''} := \theta_{1''} - \alpha \frac{\partial}{\partial \theta_{1''}} J(\theta_0, \theta_1)$$

$$\theta_{1''''} := \theta_{1'''} - \alpha \frac{\partial}{\partial \theta_{1'''}} J(\theta_0, \theta_1)$$

$J(\theta_0)$



$J(\theta_1)$



$$\theta_{0'} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

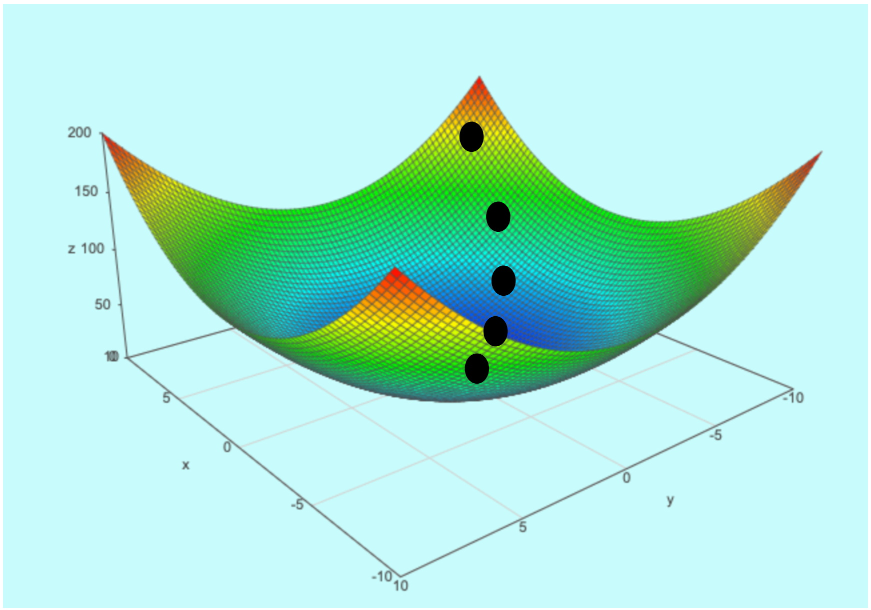
$$\theta_{0''} := \theta_{0'} - \alpha \frac{\partial}{\partial \theta_{0'}} J(\theta_0, \theta_1)$$

$$\theta_{0'''} := \theta_{0''} - \alpha \frac{\partial}{\partial \theta_{0''}} J(\theta_0, \theta_1)$$

$$\theta_{0''''} := \theta_{0'''} - \alpha \frac{\partial}{\partial \theta_{0'''}} J(\theta_0, \theta_1)$$

$$\theta_{0'''''} := \theta_{0''''} - \alpha \frac{\partial}{\partial \theta_{0''''}} J(\theta_0, \theta_1)$$

$J(\theta_0, \theta_1)$



$$\theta_{1'} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

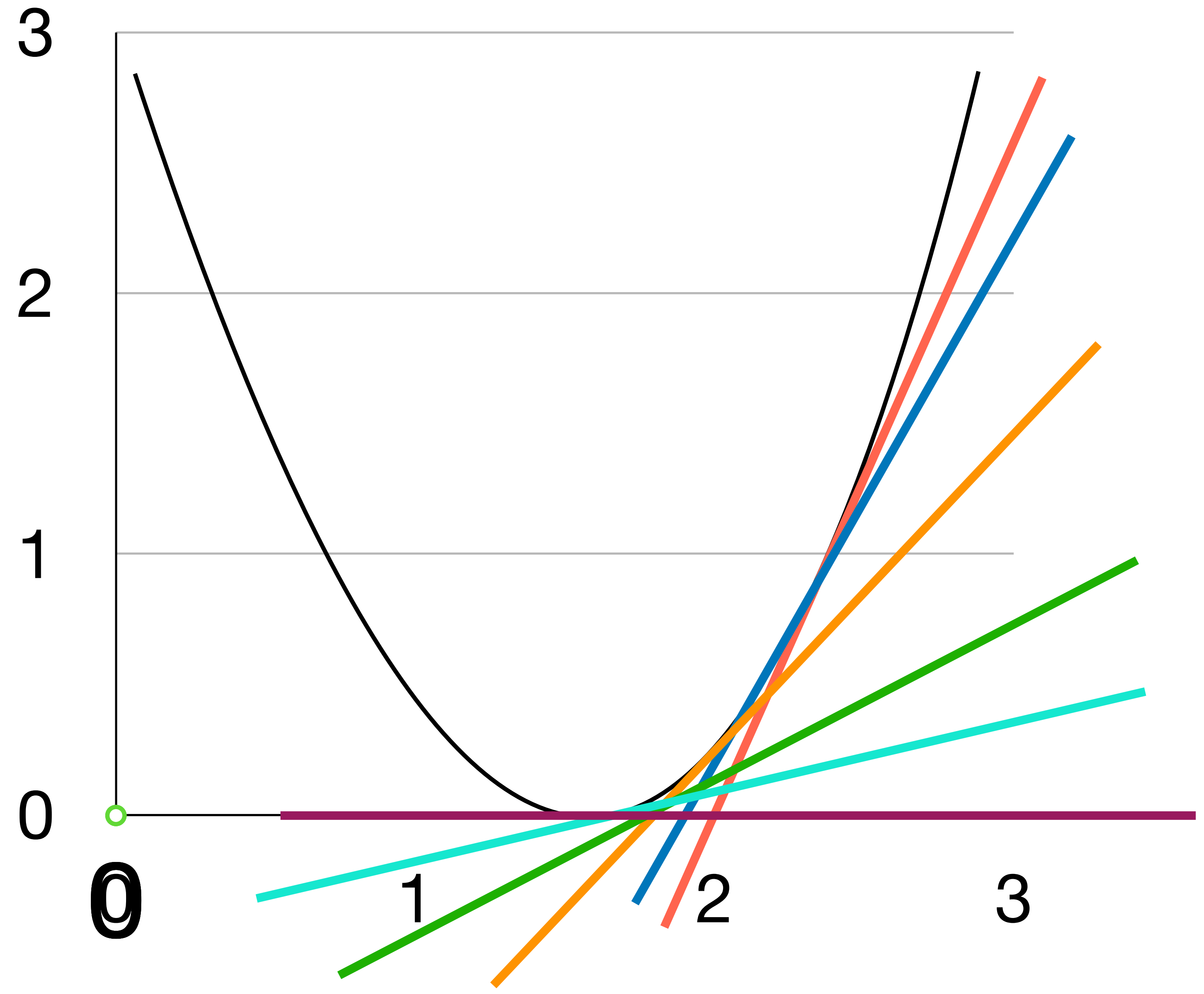
$$\theta_{1''} := \theta_{1'} - \alpha \frac{\partial}{\partial \theta_{1'}} J(\theta_0, \theta_1)$$

$$\theta_{1'''} := \theta_{1''} - \alpha \frac{\partial}{\partial \theta_{1''}} J(\theta_0, \theta_1)$$

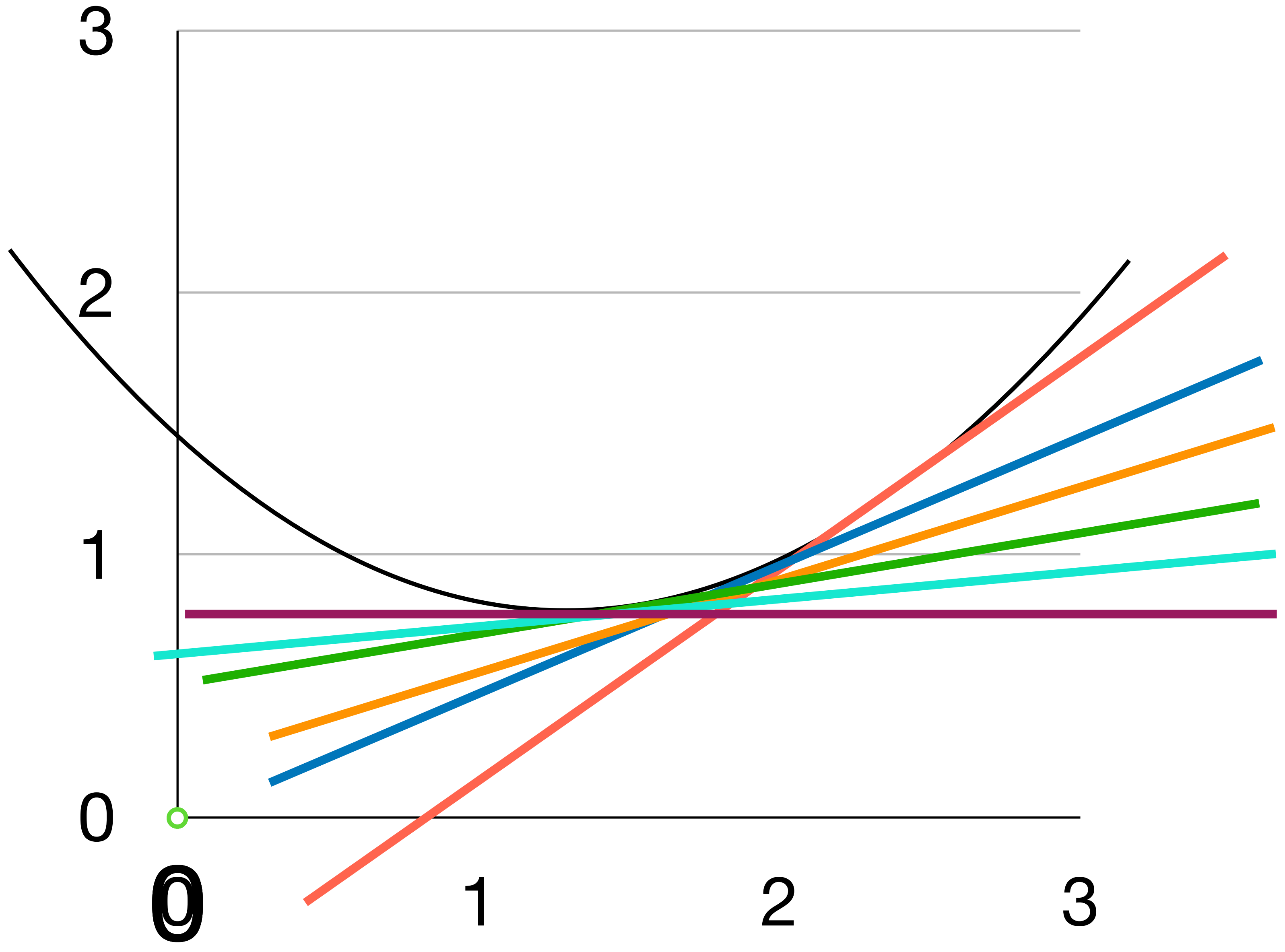
$$\theta_{1''''} := \theta_{1'''} - \alpha \frac{\partial}{\partial \theta_{1'''}} J(\theta_0, \theta_1)$$

$$\theta_{1'''''} := \theta_{1''''} - \alpha \frac{\partial}{\partial \theta_{1''''}} J(\theta_0, \theta_1)$$

$J(\theta_0)$



$J(\theta_1)$

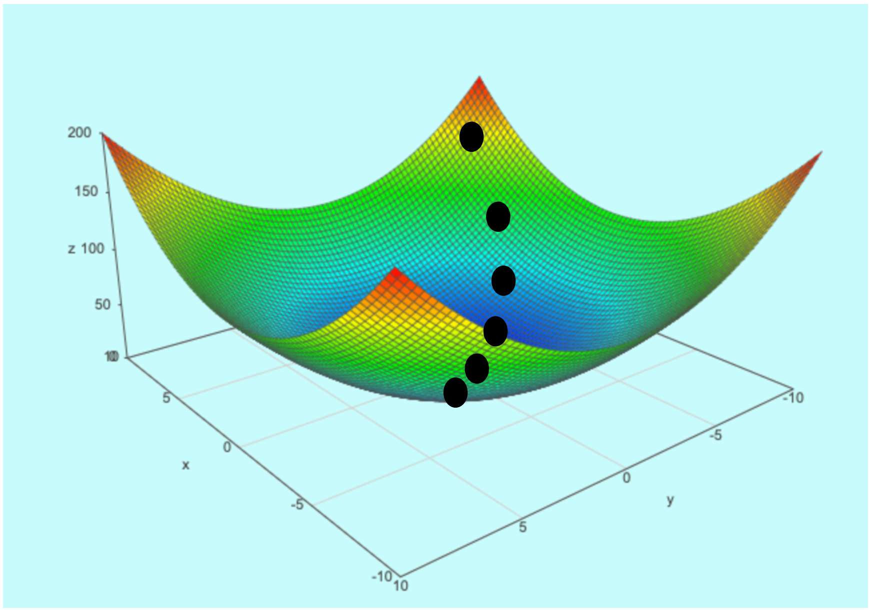


$$\theta_{0'} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \quad \theta_{0''''} := \theta_{0'''} - \alpha \frac{\partial}{\partial \theta_{0'''}} J(\theta_0, \theta_1)$$

$$\theta_{0''} := \theta_{0'} - \alpha \frac{\partial}{\partial \theta_{0'}} J(\theta_0, \theta_1) \quad \theta_{0'''''} := \theta_{0''''} - \alpha \frac{\partial}{\partial \theta_{0''''}} J(\theta_0, \theta_1)$$

$$\theta_{0'''} := \theta_{0''} - \alpha \frac{\partial}{\partial \theta_{0''}} J(\theta_0, \theta_1) \quad \theta_{0''''''} := \theta_{0'''''} - \alpha \frac{\partial}{\partial \theta_{0'''''}} J(\theta_0, \theta_1)$$

$J(\theta_0, \theta_1)$

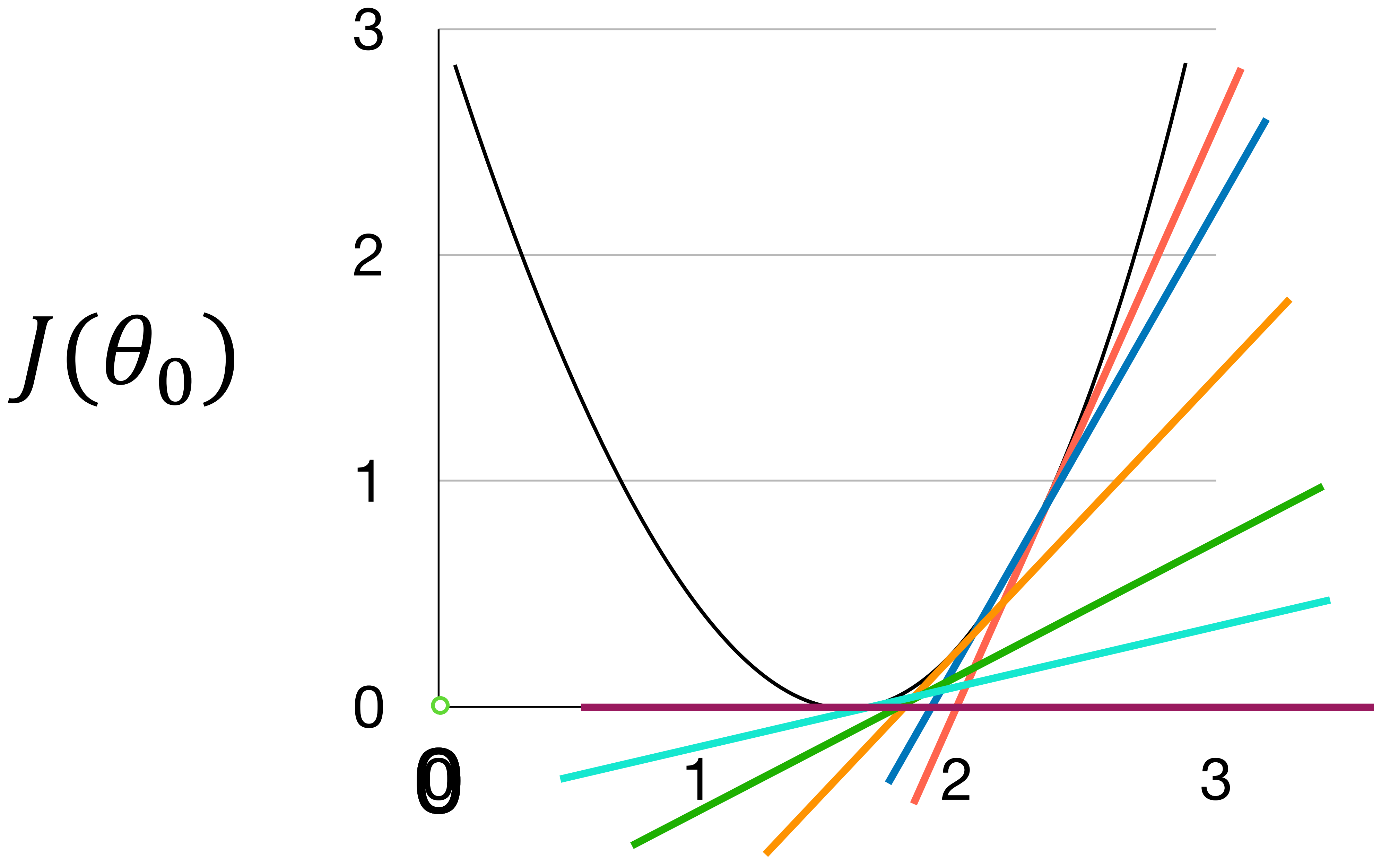


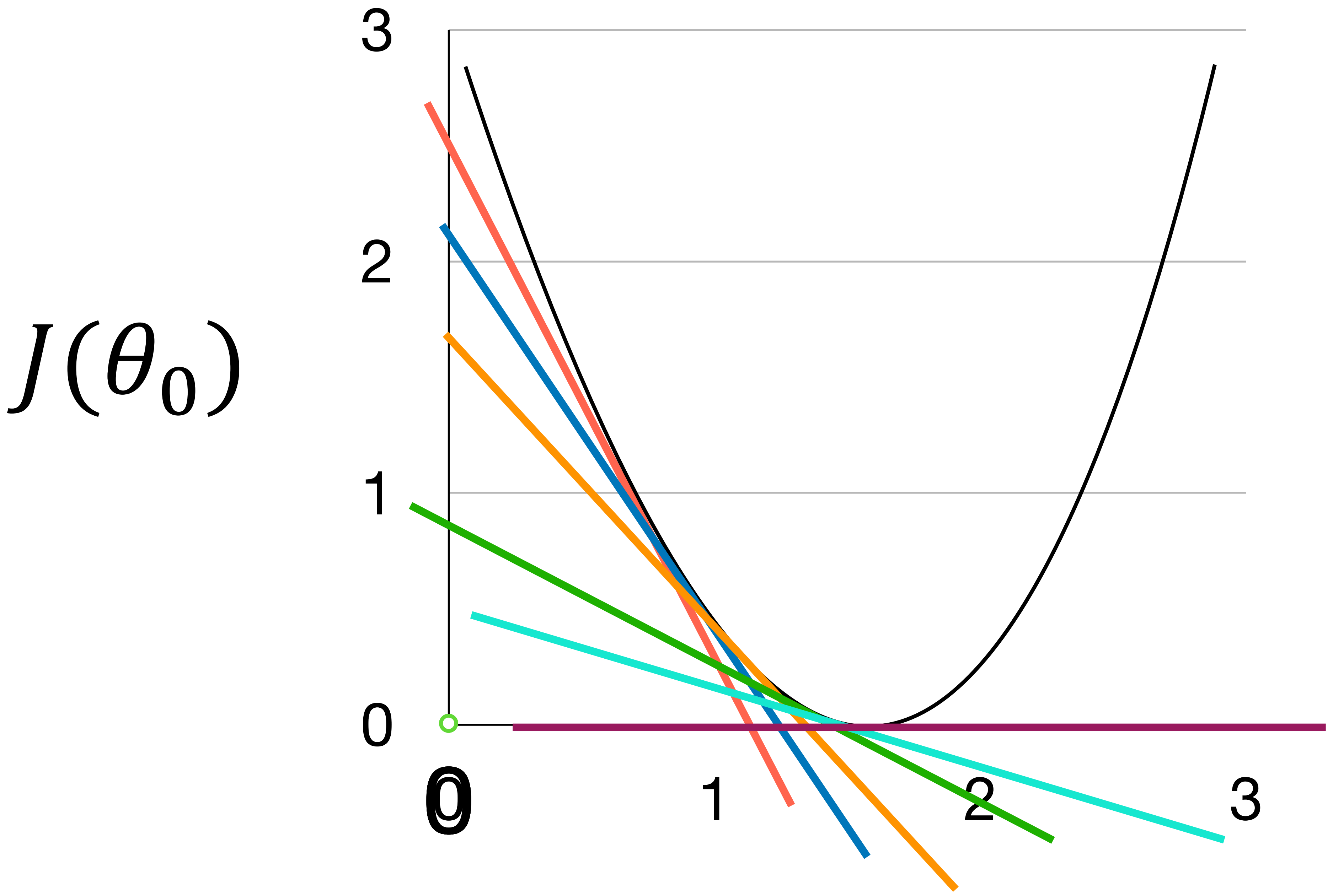
$$\theta_{1'} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \quad \theta_{1''''} := \theta_{1'''} - \alpha \frac{\partial}{\partial \theta_{1'''}} J(\theta_0, \theta_1)$$

$$\theta_{1''} := \theta_{1'} - \alpha \frac{\partial}{\partial \theta_{1'}} J(\theta_0, \theta_1) \quad \theta_{1'''''} := \theta_{1''''} - \alpha \frac{\partial}{\partial \theta_{1''''}} J(\theta_0, \theta_1)$$

$$\theta_{1'''} := \theta_{1''} - \alpha \frac{\partial}{\partial \theta_{1''}} J(\theta_0, \theta_1) \quad \theta_{1''''''} := \theta_{1'''''} - \alpha \frac{\partial}{\partial \theta_{1'''''}} J(\theta_0, \theta_1)$$

positive / negative Gradient





Positive

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

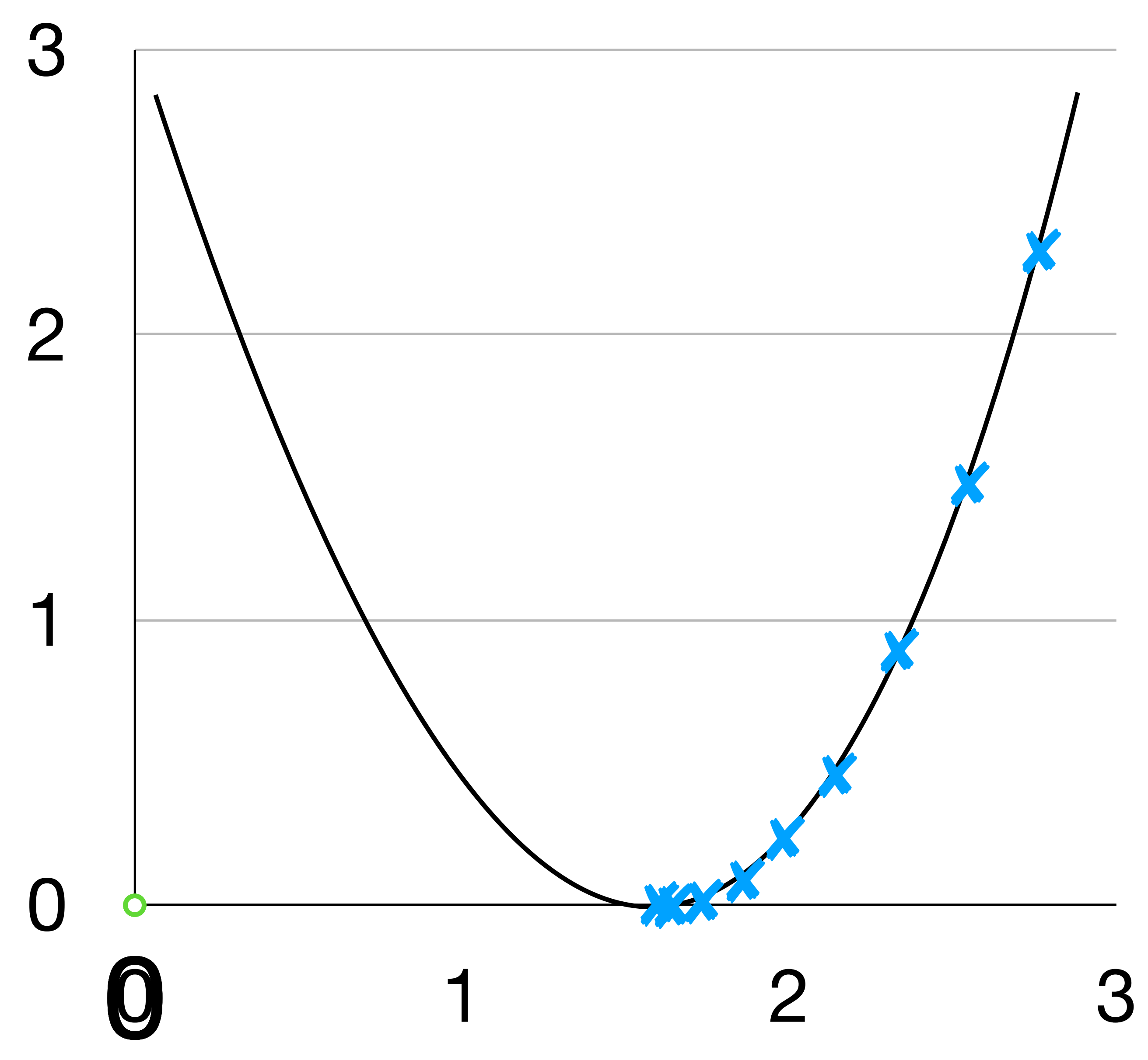
값이 감소

Negative

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

값이 증가

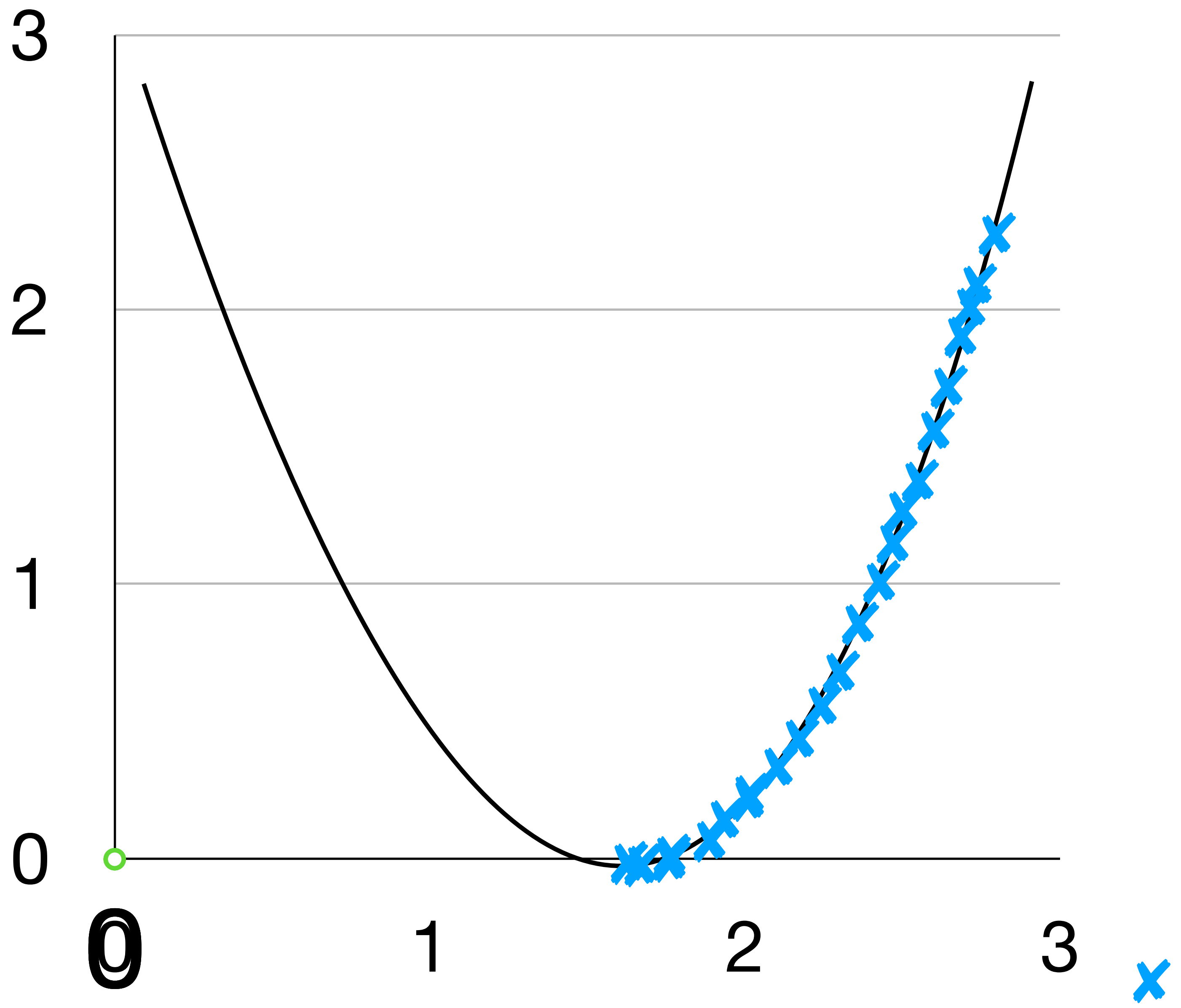
There is no need to decrease α over time



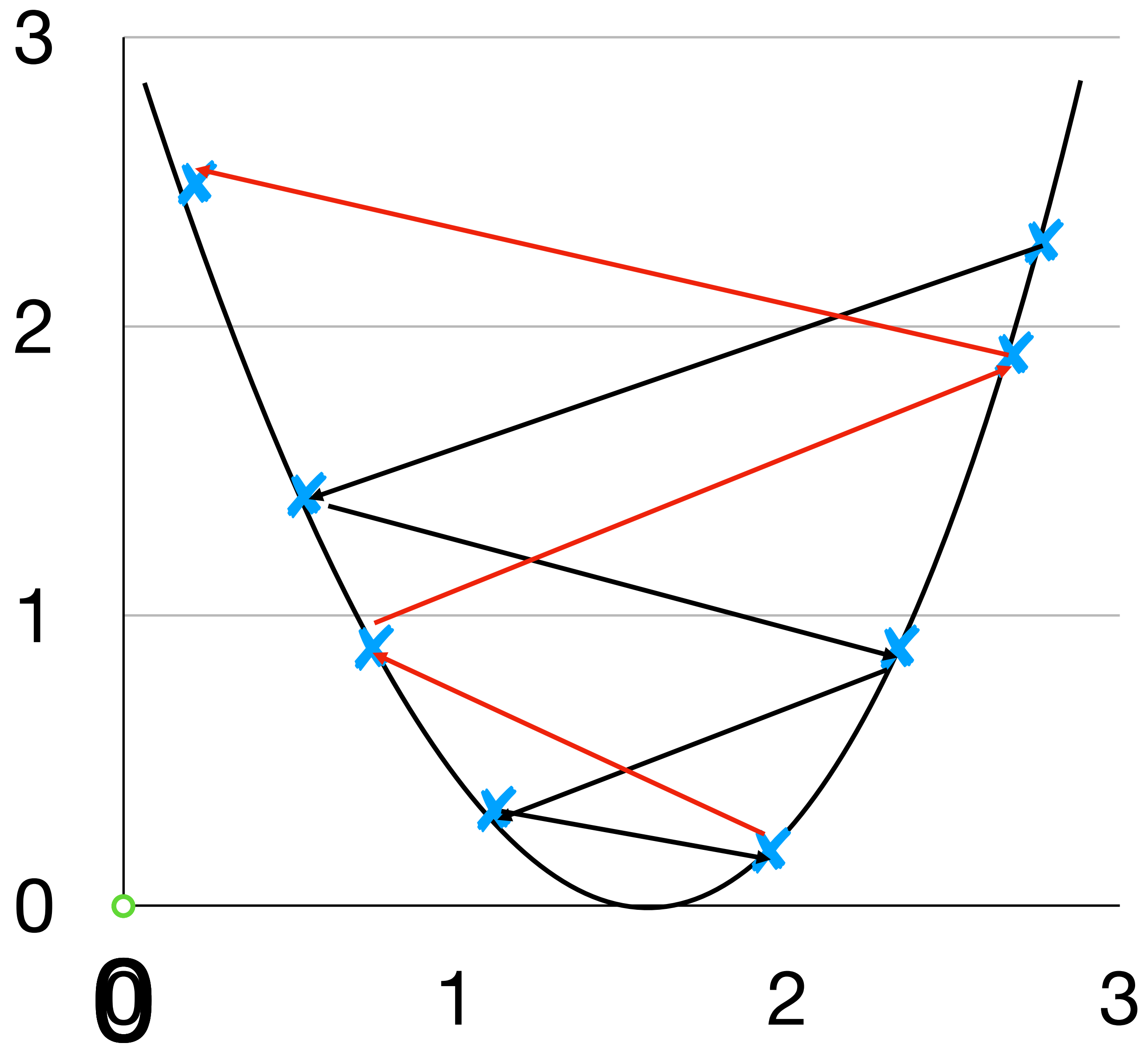
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

α 를 점점 작게 할 필요 없음.

Too small α, Too big α



너무 느림



수렴 불가

$$\frac{\partial}{\partial\theta_j}J(\theta_0,\theta_1) := \frac{\partial}{\partial\theta_j}\frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$:= \frac{\partial}{\partial\theta_j}\frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

$$j = 0\theta_0 := \frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1) = \frac{\partial}{\partial\theta_0}\frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2 = \frac{\partial}{\partial\theta_0}\frac{1}{2m}\sum_{i=1}^{m}\theta_0^2 + \theta_1^2 x^{(i)2} - y^{(i)2} + 2(\theta_0\theta_1 x^{(i)} - \theta_1 x^{(i)} y^{(i)} - \theta_0 y^{(i)})$$

$$= \frac{1}{2m}\sum_{i=1}^{m}2\theta_0 + 2\theta_1 x^{(i)} - 2y^{(i)} = \frac{1}{2m}\sum_{i=1}^{m}2(\theta_0 + \theta_1 x^{(i)} - y^{(i)}) \quad = \frac{1}{m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)}) \quad \boxed{= \frac{1}{m}\sum_{i=1}^{m}(h_\theta x^{(i)} - y^{(i)})}$$

$$j = 1\theta_1 := \frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1) = \frac{\partial}{\partial\theta_1}\frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2 = \frac{\partial}{\partial\theta_1}\frac{1}{2m}\sum_{i=1}^{m}\theta_0^2 + \theta_1^2 x^{(i)2} - y^{(i)2} + 2(\theta_0\theta_1 x^{(i)} - \theta_1 x^{(i)} y^{(i)} - \theta_0 y^{(i)})$$

$$= \frac{1}{2m}\sum_{i=1}^{m}2\theta_1 x^{(i)2} + 2\theta_0 x^{(i)} - 2x^{(i)} y^{(i)} = \frac{1}{2m}\sum_{i=1}^{m}2(\theta_1 x^{(i)2} + \theta_0 x^{(i)} - x^{(i)} y^{(i)}) = \frac{1}{m}\sum_{i=1}^{m}\theta_1 x^{(i)2} + \theta_0 x^{(i)} - x^{(i)} y^{(i)}$$

$$= \frac{1}{m}\sum_{i=1}^{m}x(\theta_1 x + \theta_0 - y^{(i)}) \boxed{= \frac{1}{m}\sum_{i=1}^{m}(h_\theta x^{(i)} - y^{(i)})x}$$

Gradient Descent

수렴 할때까지 무한 반복

$$\left\{ \begin{aligned} \theta_0 &:= \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta x^{(i)} - y^{(i)}) \\ \theta_1 &:= \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta x^{(i)} - y^{(i)}) x \end{aligned} \right\}$$

$$\theta_i := \theta_i - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta x^{(i)} - y^{(i)}) x_i (i = 0,1,2,3\ldots.n)$$

Batch Gradient Descent

$$\frac{\partial}{\partial\theta_j}J(\theta_0,\theta_1) := \frac{\partial}{\partial\theta_j}\frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

모든 값에 대해서 계산하지 때문에 Batch 라고 함

Hypothesis and matrix

$$h_\theta x = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_n x_n$$

Hypothesis and matrix

$$h_\theta x = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_n x_n$$

편의상 $X_0 = 1$

Gradient descent for multiple features

Hypothesis : $h_\theta(x) = \theta^T x = \theta_0 x_o + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$

Parameters : $\theta_0, \theta_1, \theta_2, \theta_3 \ldots$

Cost function : $J(\theta_0, \theta_1, \theta_2 \ldots \theta_n) = \theta_j - \frac{1}{2m} \alpha \sum_{i=1}^{m} (h_\theta x^{(i)} - y^{(i)})^2$

Linear Regression with Multiple Features

Size(feet)	Price($1000)
x	y
2104	460
1416	232
1534	315
852	178

Size (feet)	Number of bedrooms	Number of floors	Age of home (years)	Price ($1000)
x1	x2	x3	x4	y
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178

$$h_\theta(x) = \theta_0 + \theta_x$$

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \ldots + \theta_n x_n$$

n = number of features   m = number of examples

$x^{(i)}$ = input (features) of i^{th} training example

$x_j^{(i)}$ = value of feature j in i^{th} training example

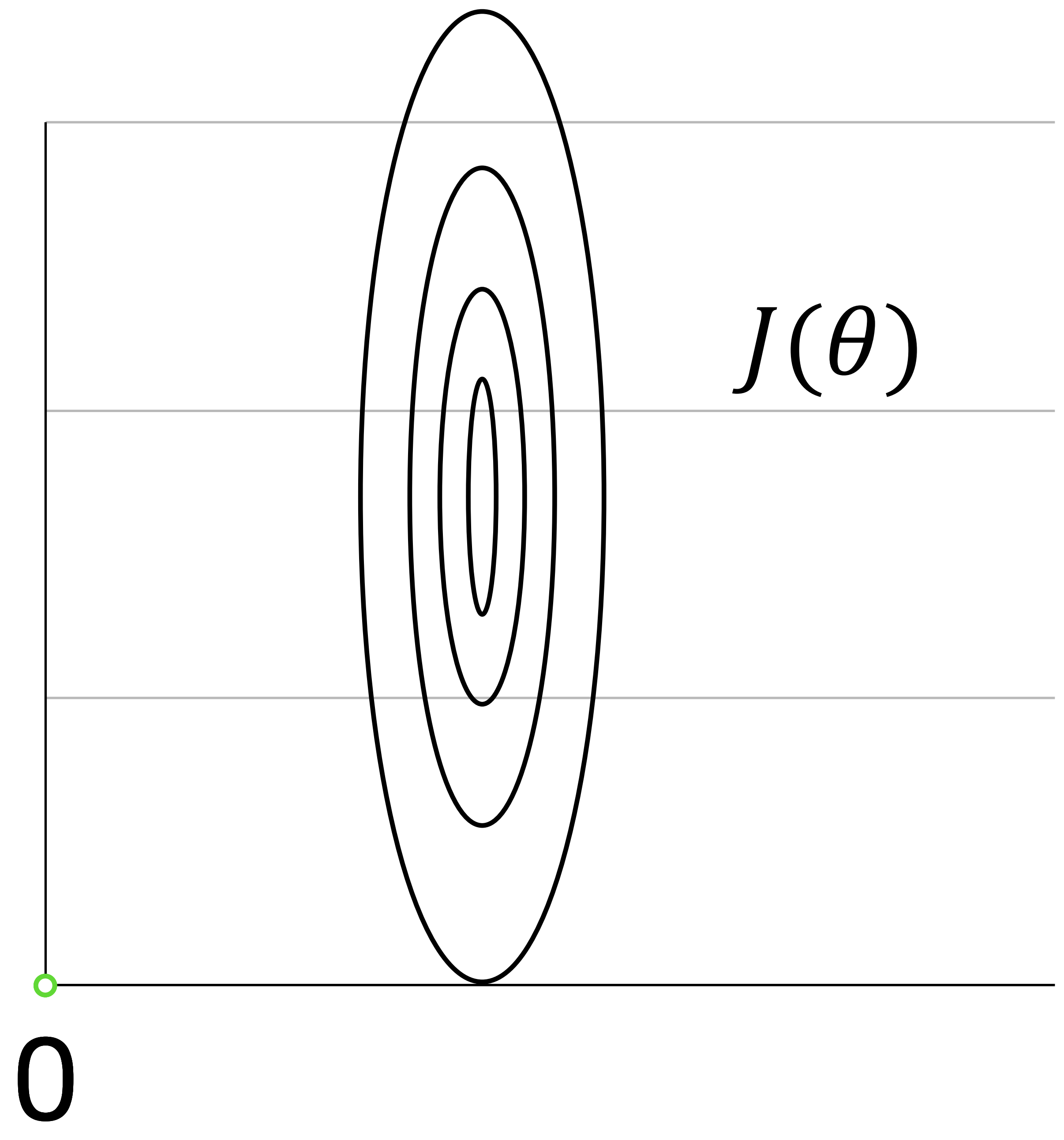
$$x^{(2)} = \begin{bmatrix} 1416 \\ 3 \\ 2 \\ 40 \end{bmatrix} \qquad x_3^{(2)} = 2 \qquad x_0^{(i)} = 1$$

Feature Scaling

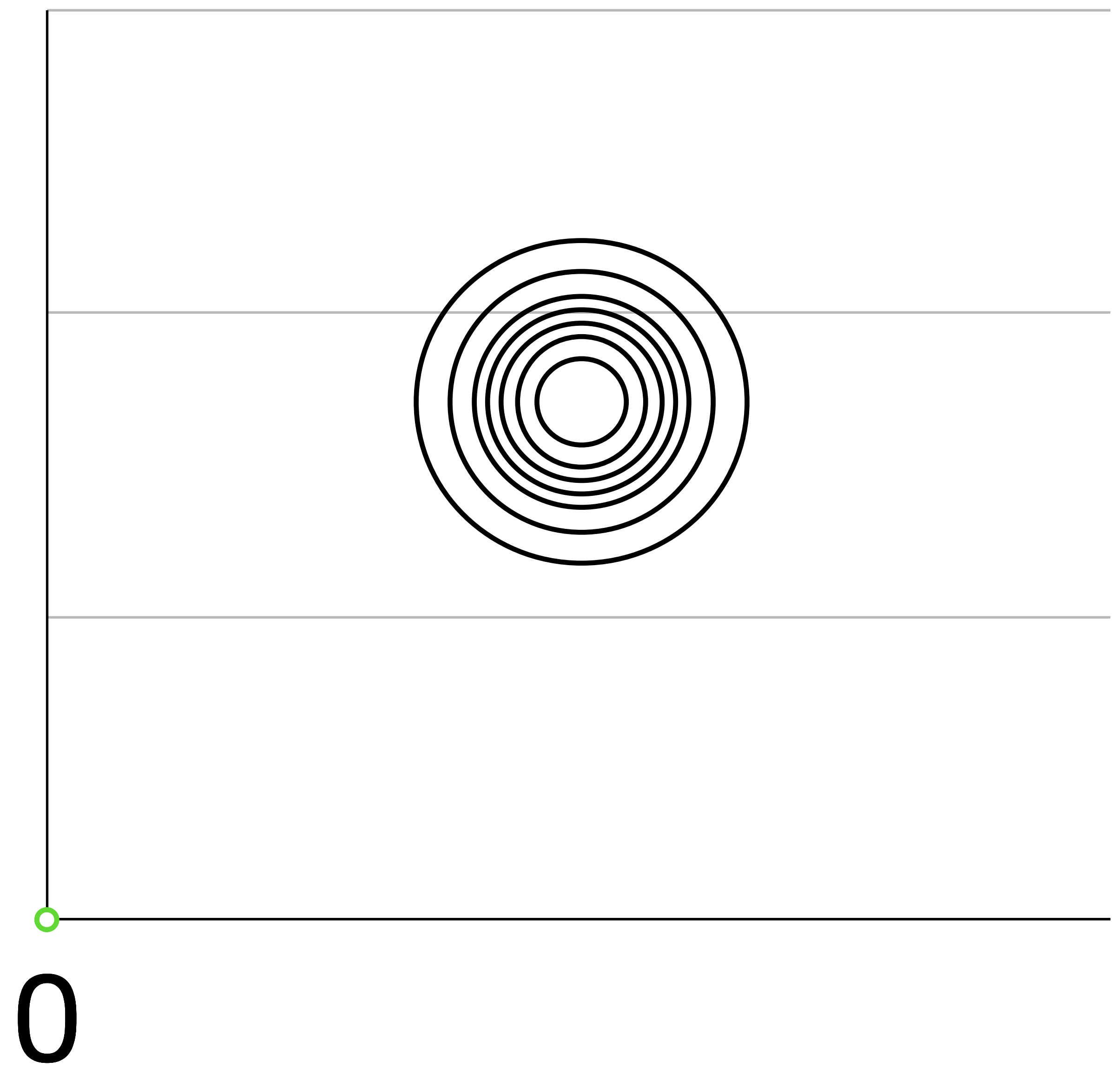
Concept : Feature 들 간의 값을 유사한 범위를 가지도록 하는것

x1 = size (0 ~ 2000)
x2 = number of bedrooms (1 ~ 5)

$$x_1 = \frac{size}{2000} \qquad x_2 = \frac{number of bedrooms}{5}$$



SLOW Gradient Descent



FAST Gradient Descent

모든 Feature 들 간의 값을 대략 적으로 -1 < X < 1 범위로 축소

$$0 \leq x \leq 3$$

$$-2 \leq x \leq 0.5$$

$$-100 \leq x \leq 100$$

$$-0.0001 \leq x \leq 0.0001$$

Mean Normalization

Concept : 평균을 빼고 최대값으로 나누기

$$x_1 = \frac{x_1 - \mu_1}{s_1} = \frac{size - 1000}{2000}$$

$$x_2 = \frac{x_2 - \mu_2}{s_2} = \frac{\#bedrooms - 2}{5}$$

$$-0.5 \le x_1 \le 0.5$$

$$-0.5 \le x_2 \le 0.5$$