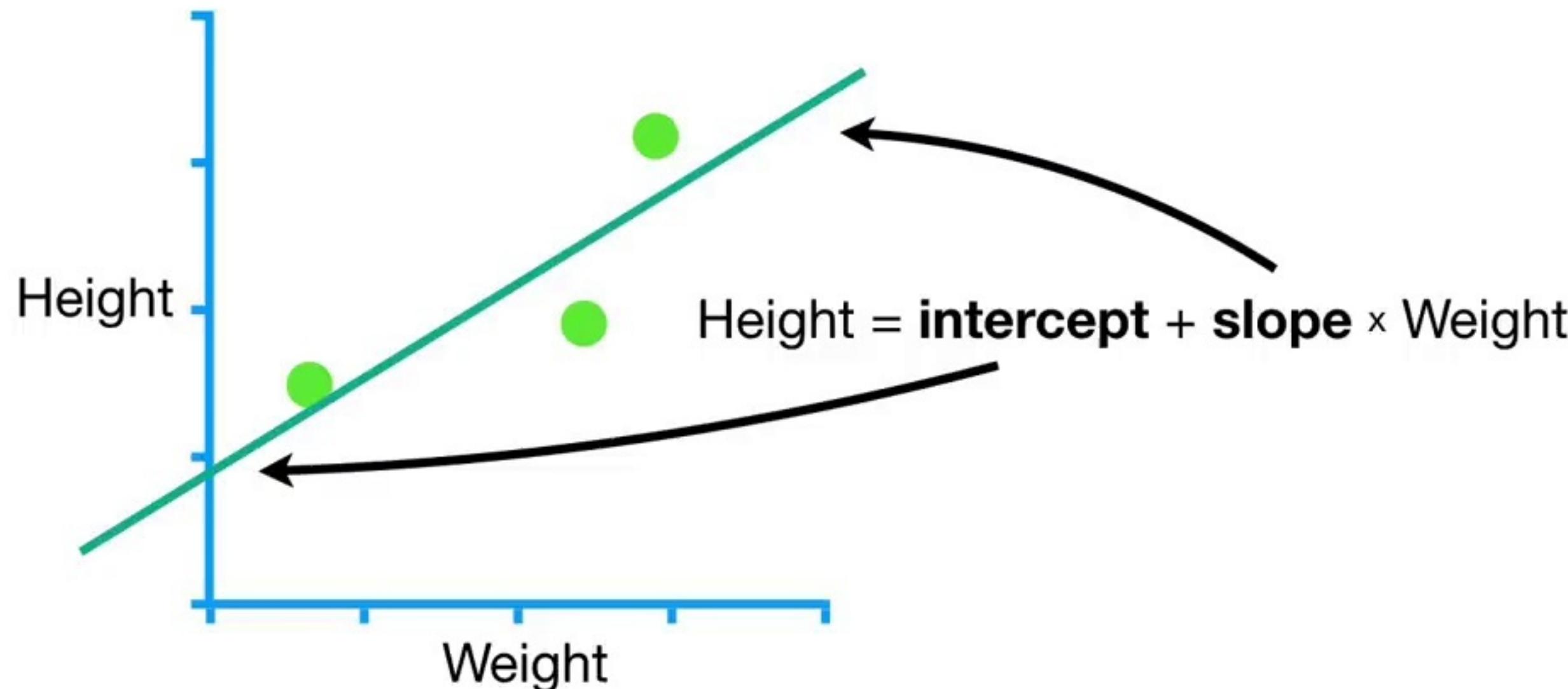


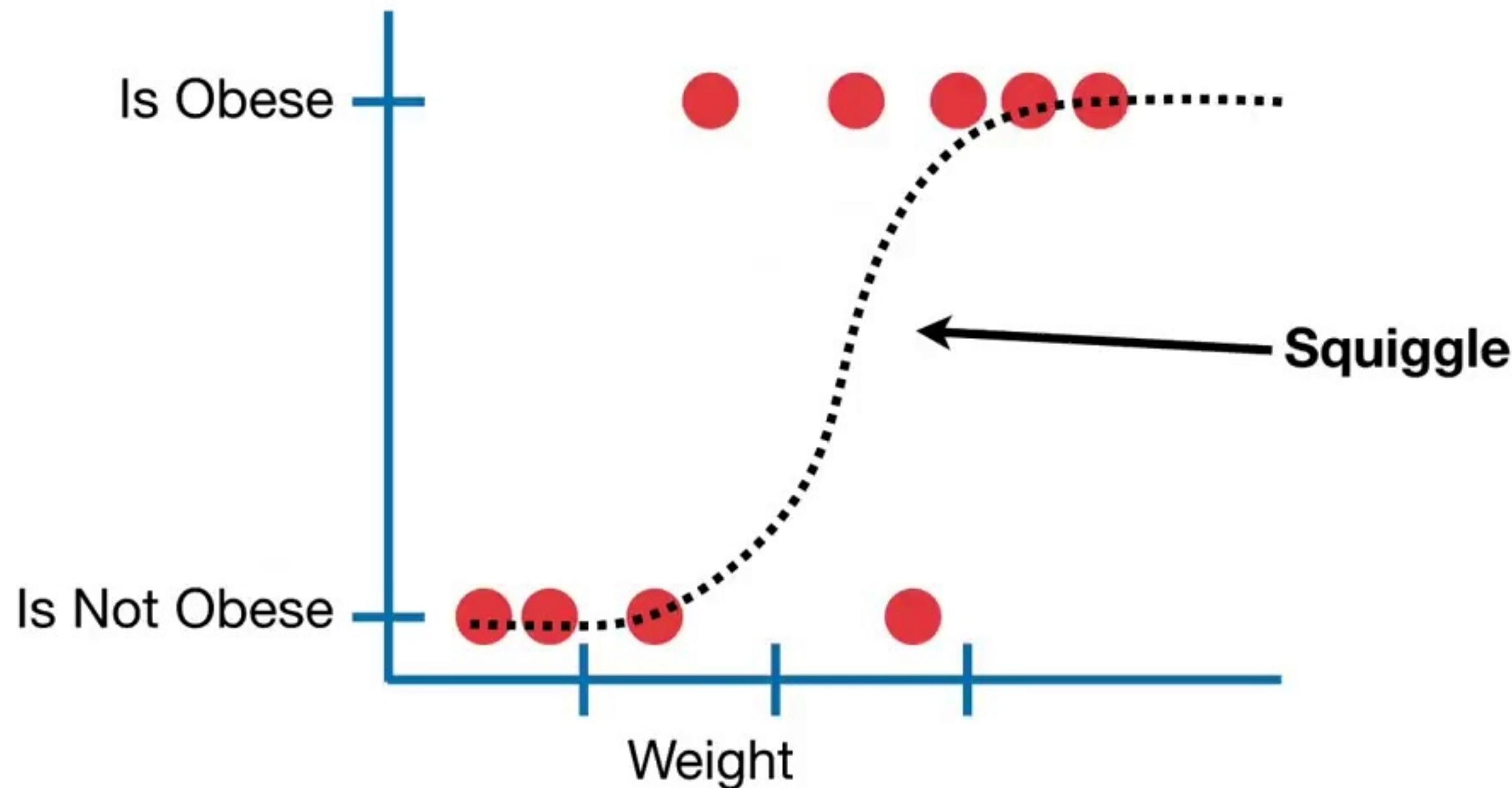
Gradient Descent Step-by-Step!!!

In Statistics, Machine Learning and other Data Science fields, we optimize a lot of stuff.

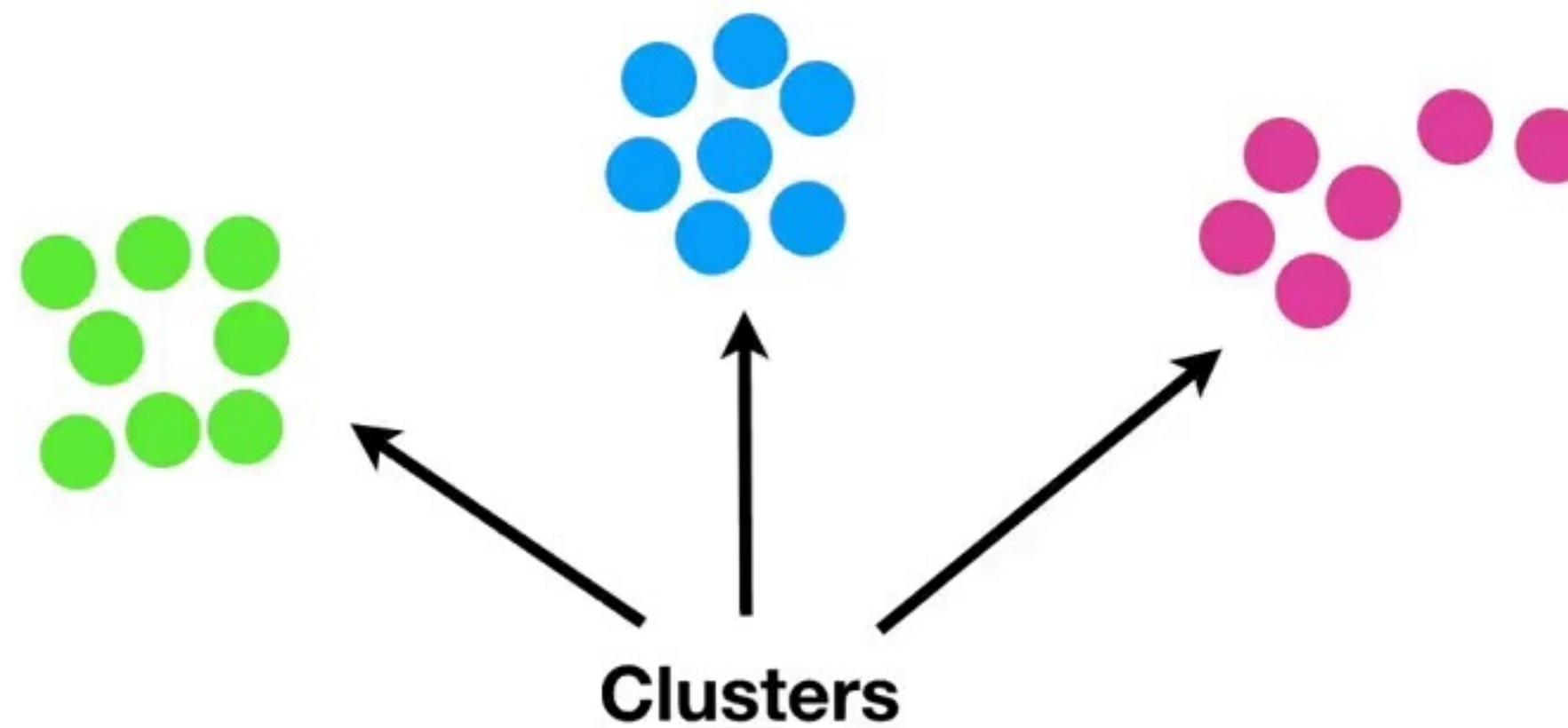
When we fit a line with **Linear Regression**, we optimize the **Intercept** and **Slope**.



When we use **Logistic Regression**,
we optimize a squiggle.



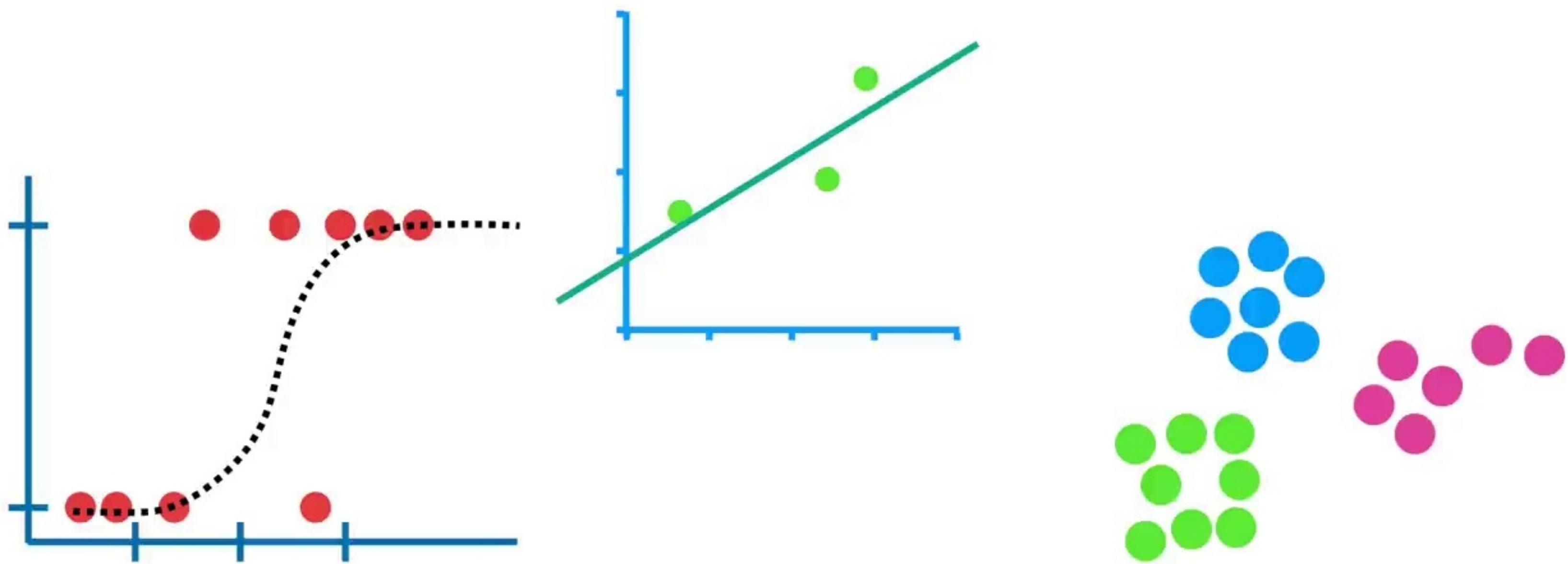
And when we use **t-SNE**, we optimize clusters.



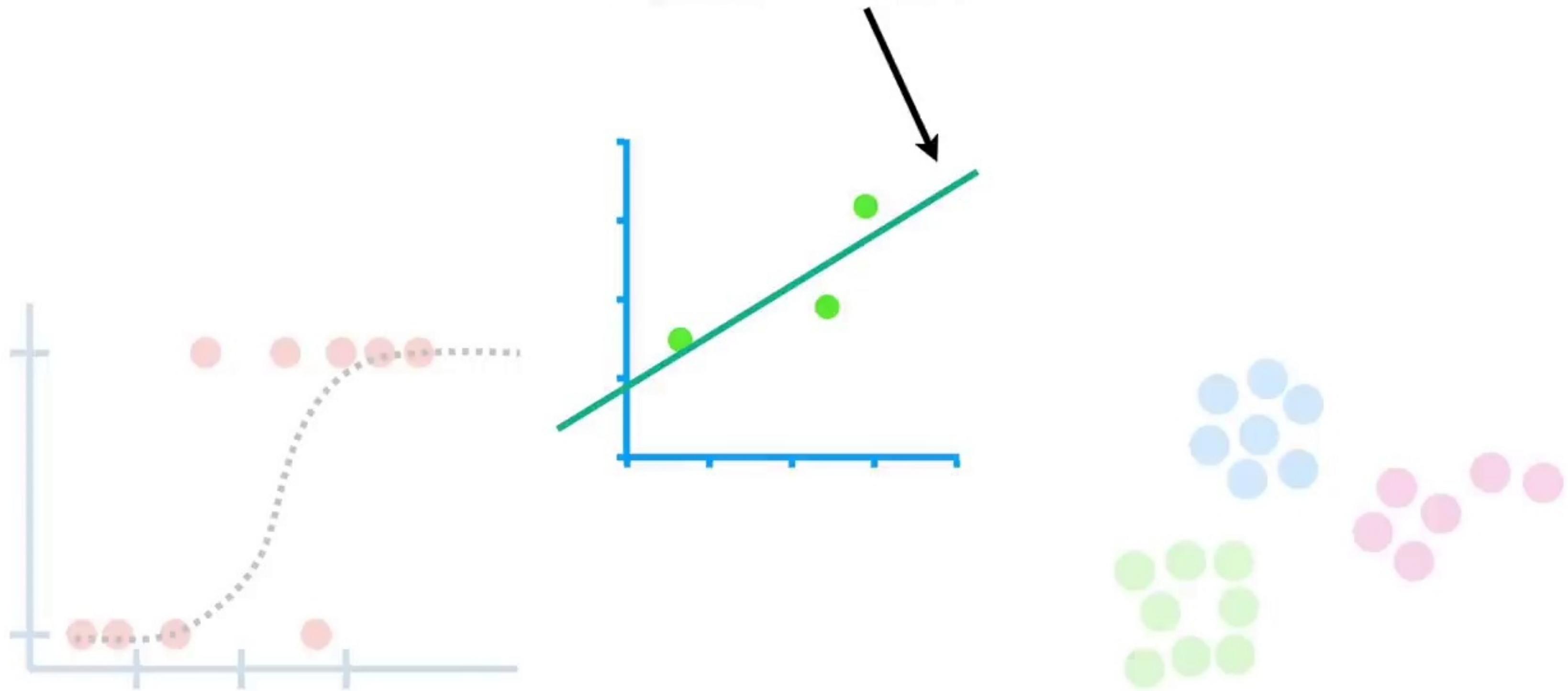
These are just a few examples of the stuff we optimize, there are tons more.



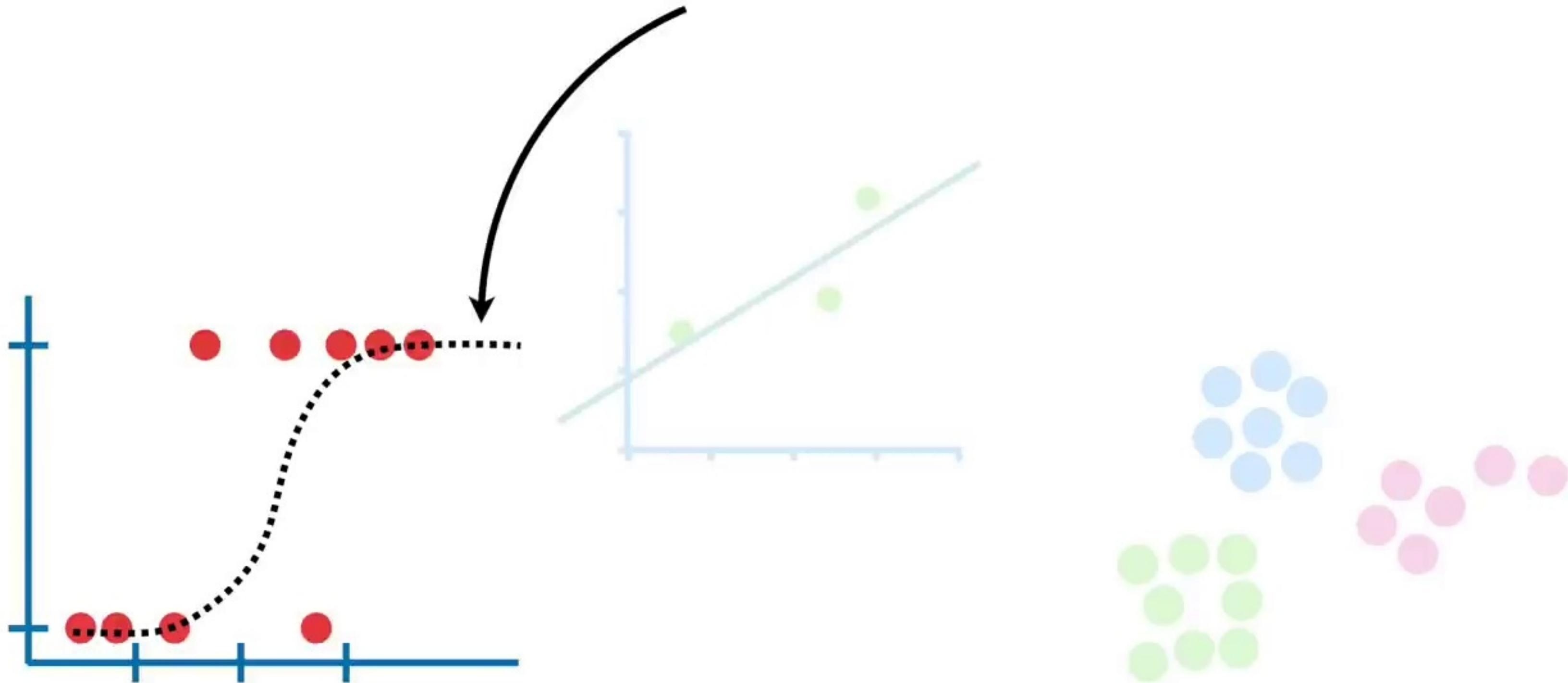
The cool thing is that **Gradient Descent** can optimize all these things and much more.



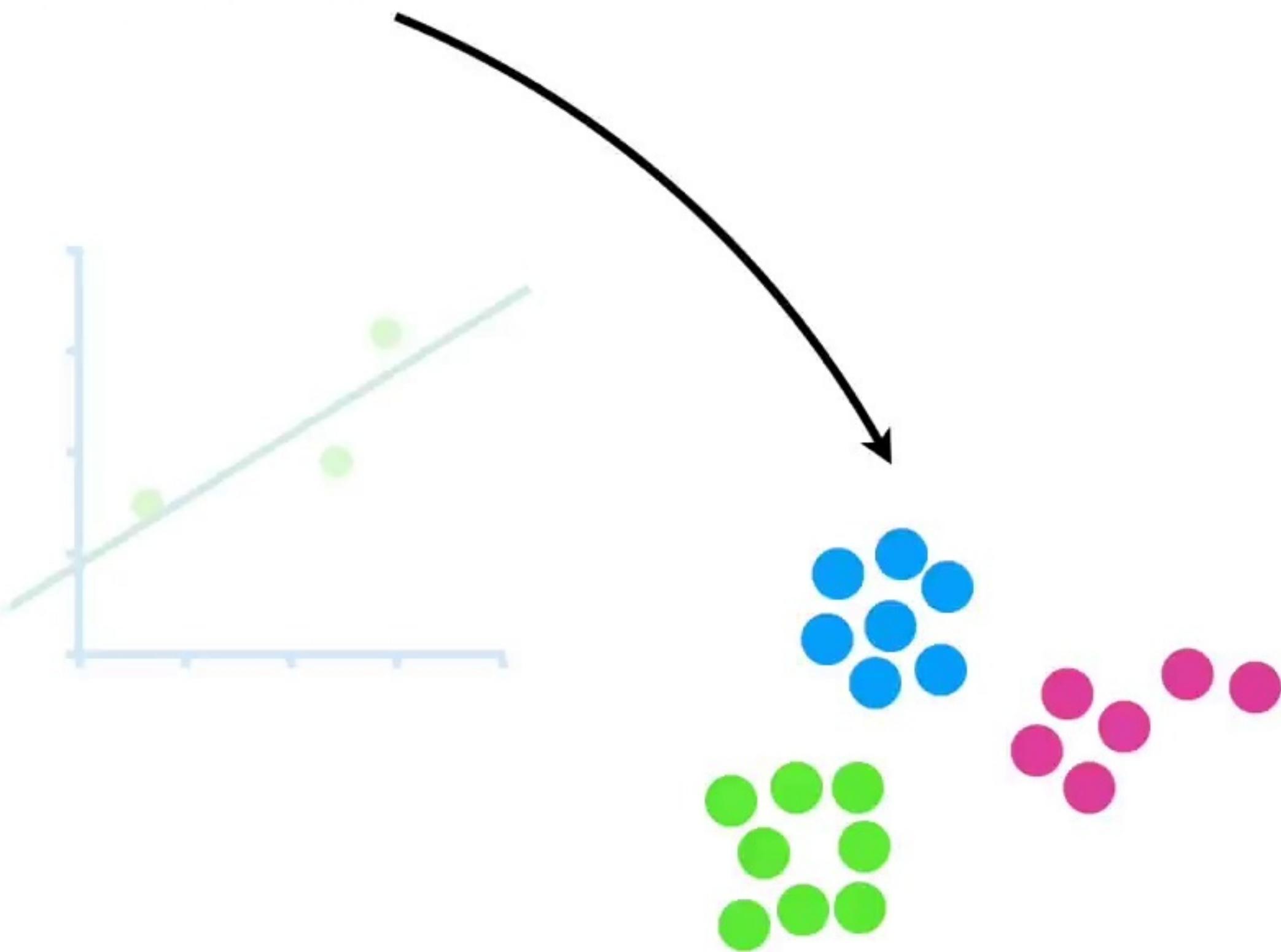
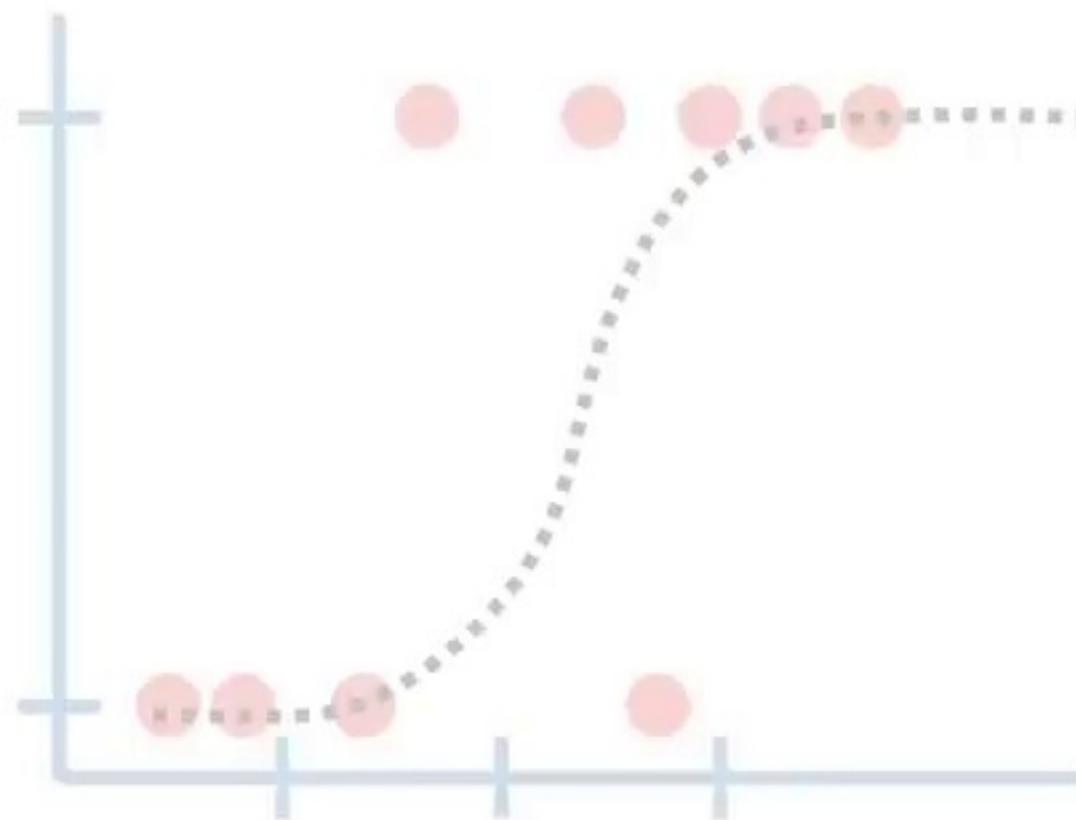
So if we learn how to optimize this line using
Gradient Descent...



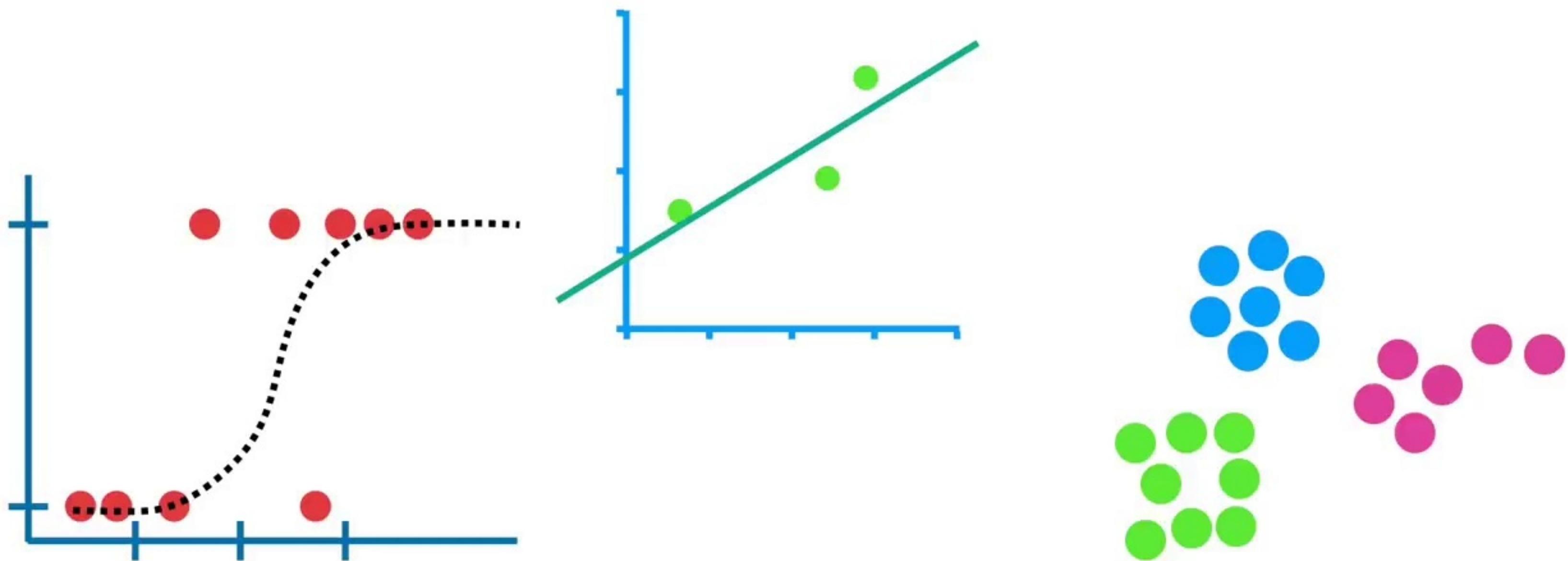
...then we will have learned the strategy that optimizes this squiggle...



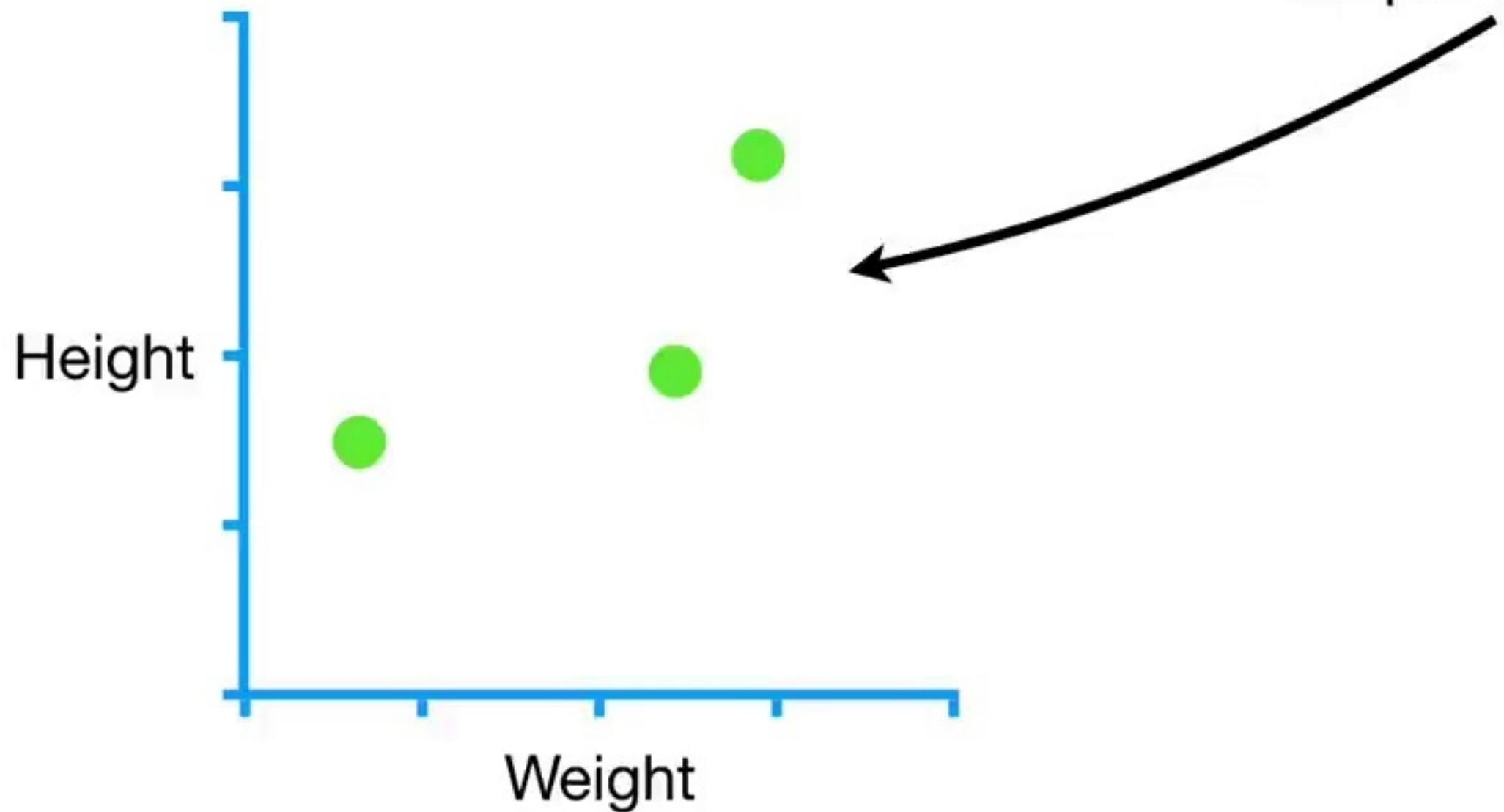
...and these clusters...

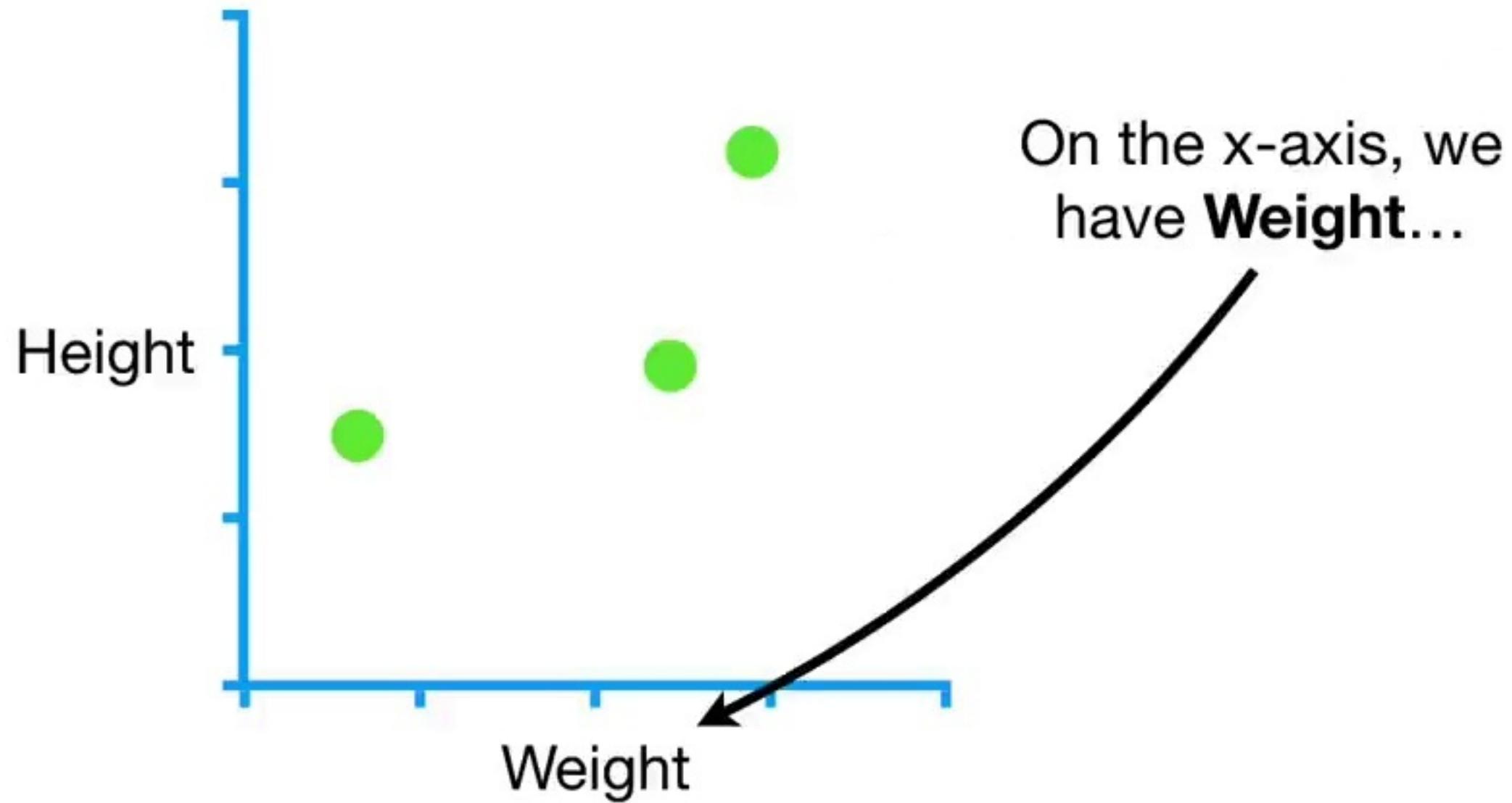


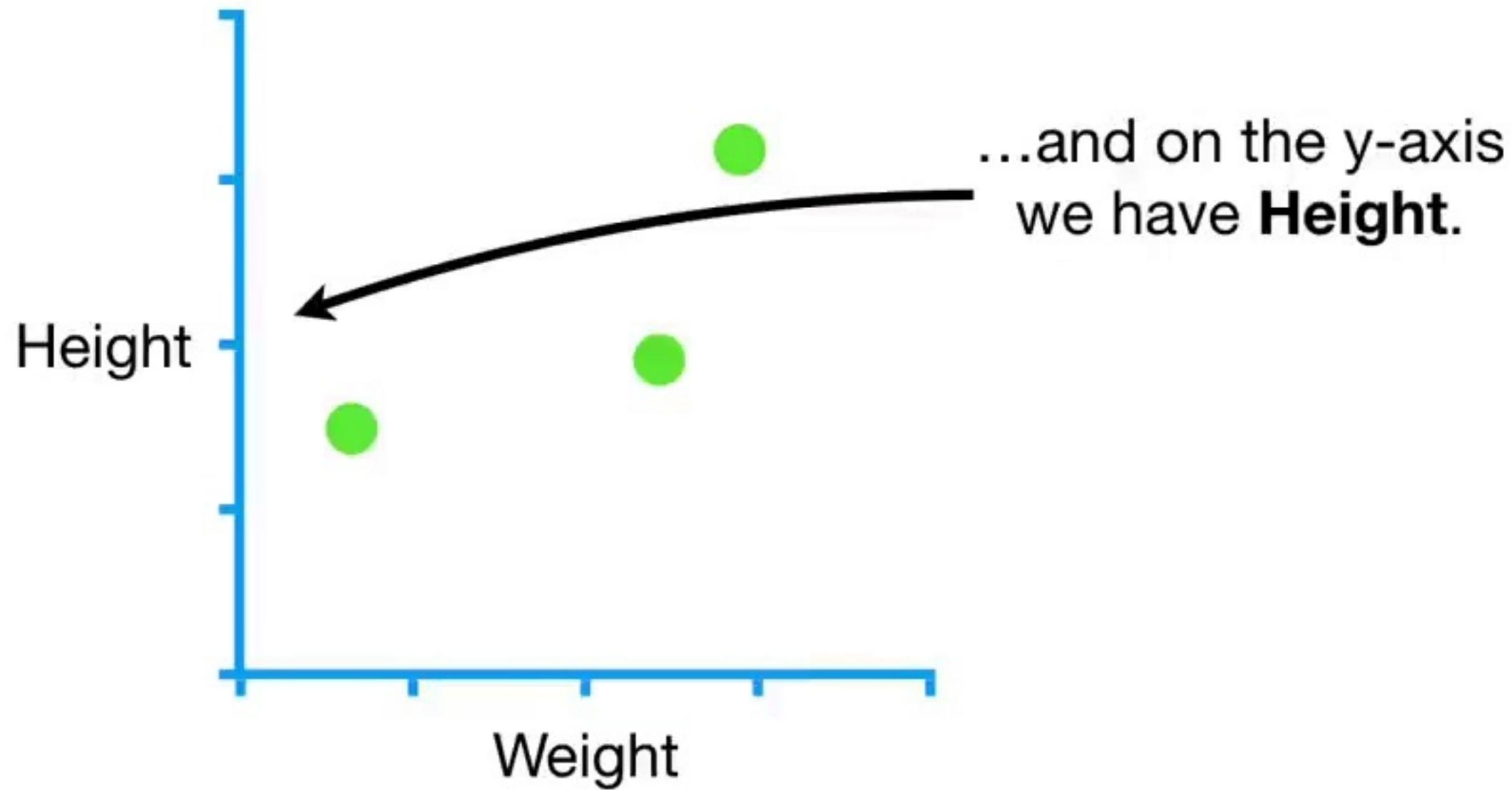
...and many more of the optimization problems
we have in Statistics, Machine Learning and Data
Science.



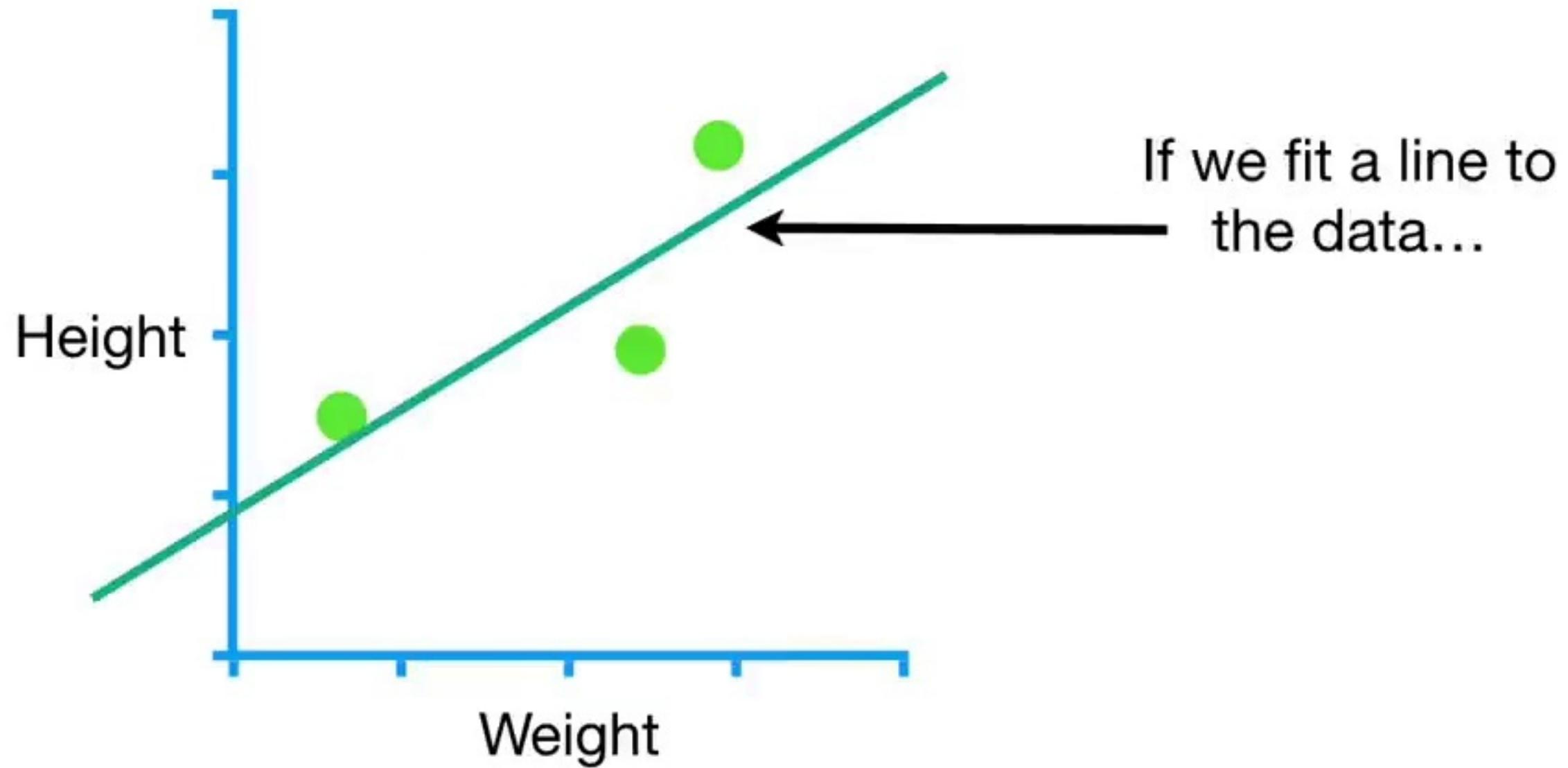
So let's start with a simple data set.

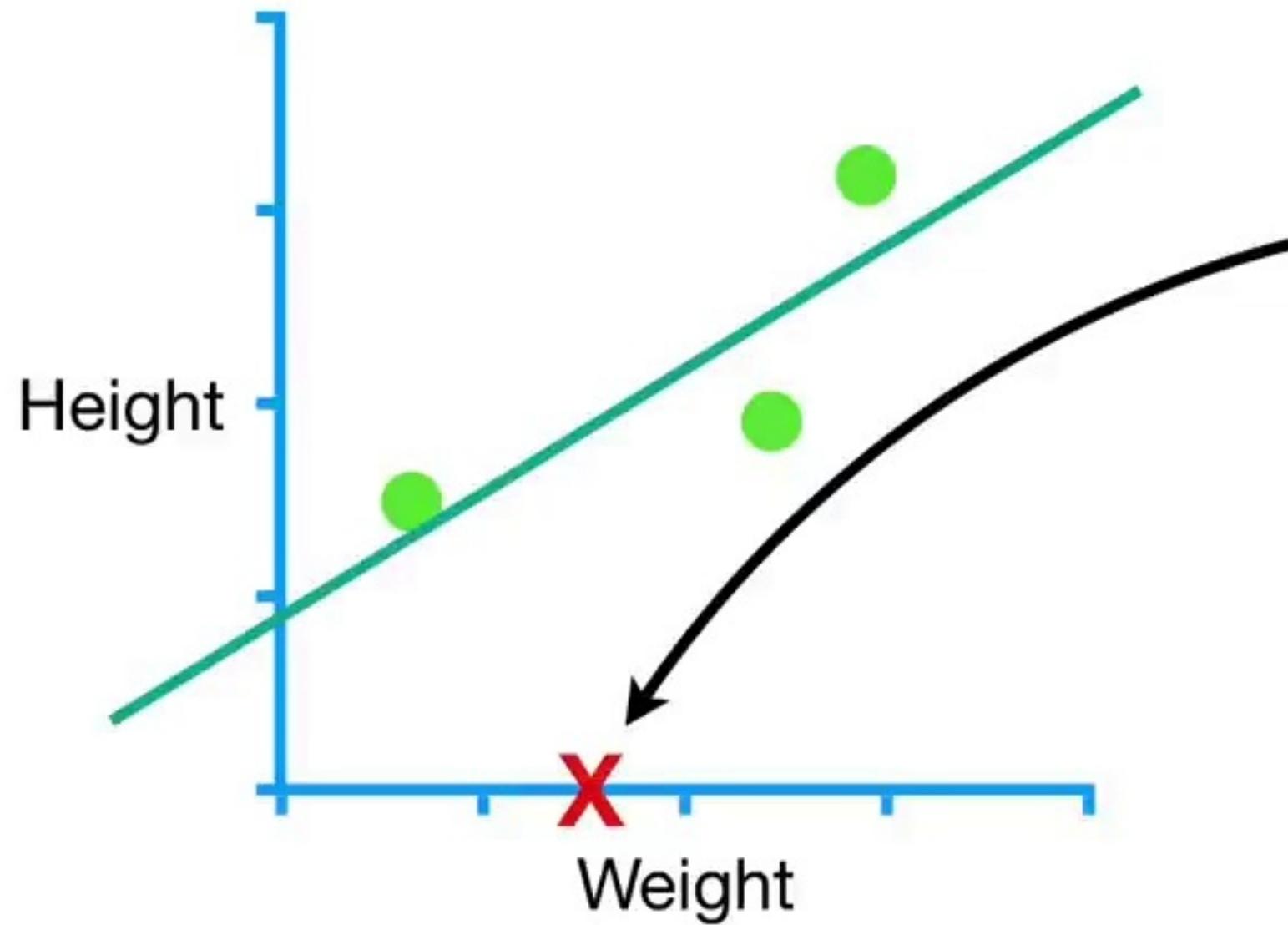




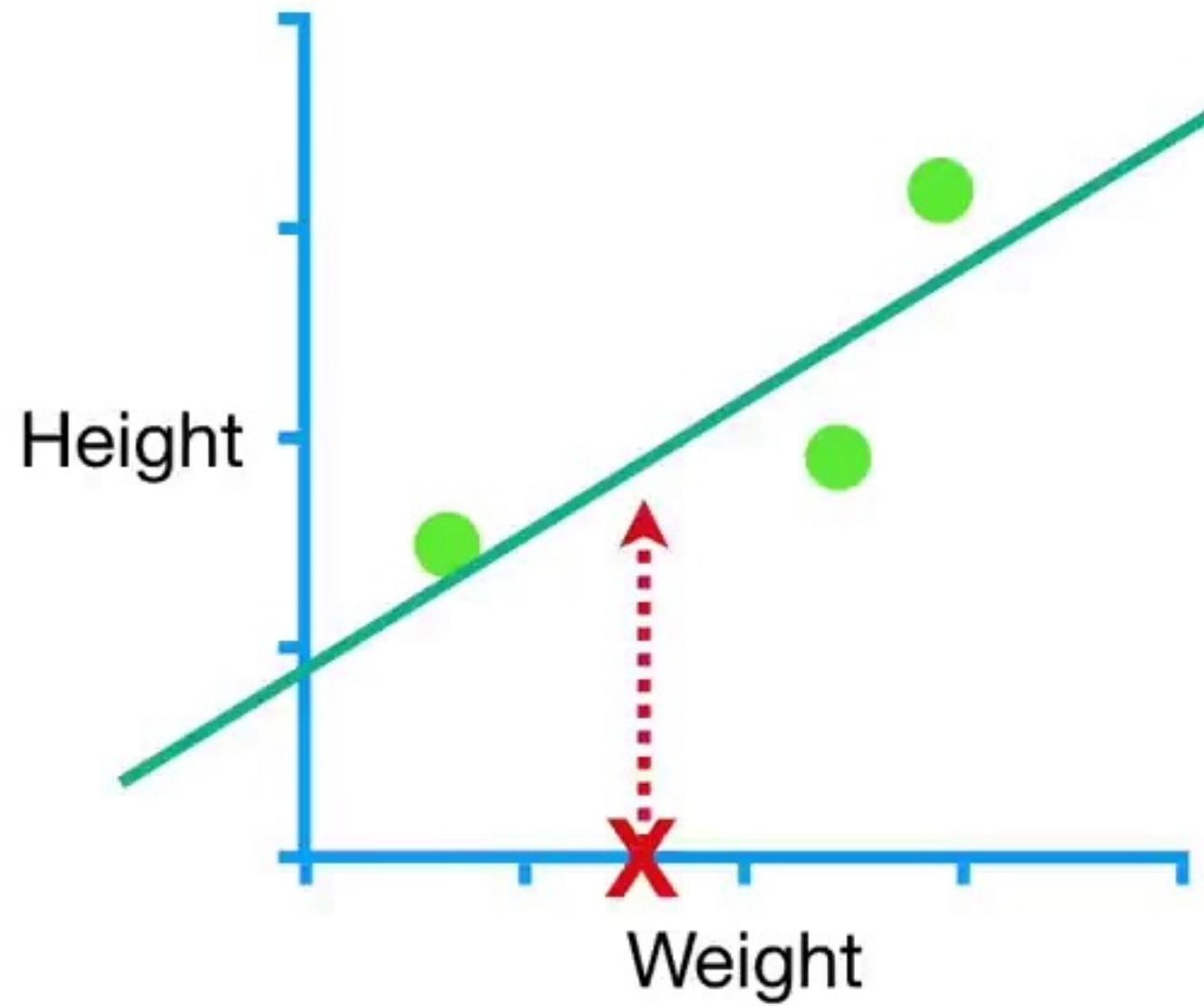


...and on the y-axis
we have **Height**.

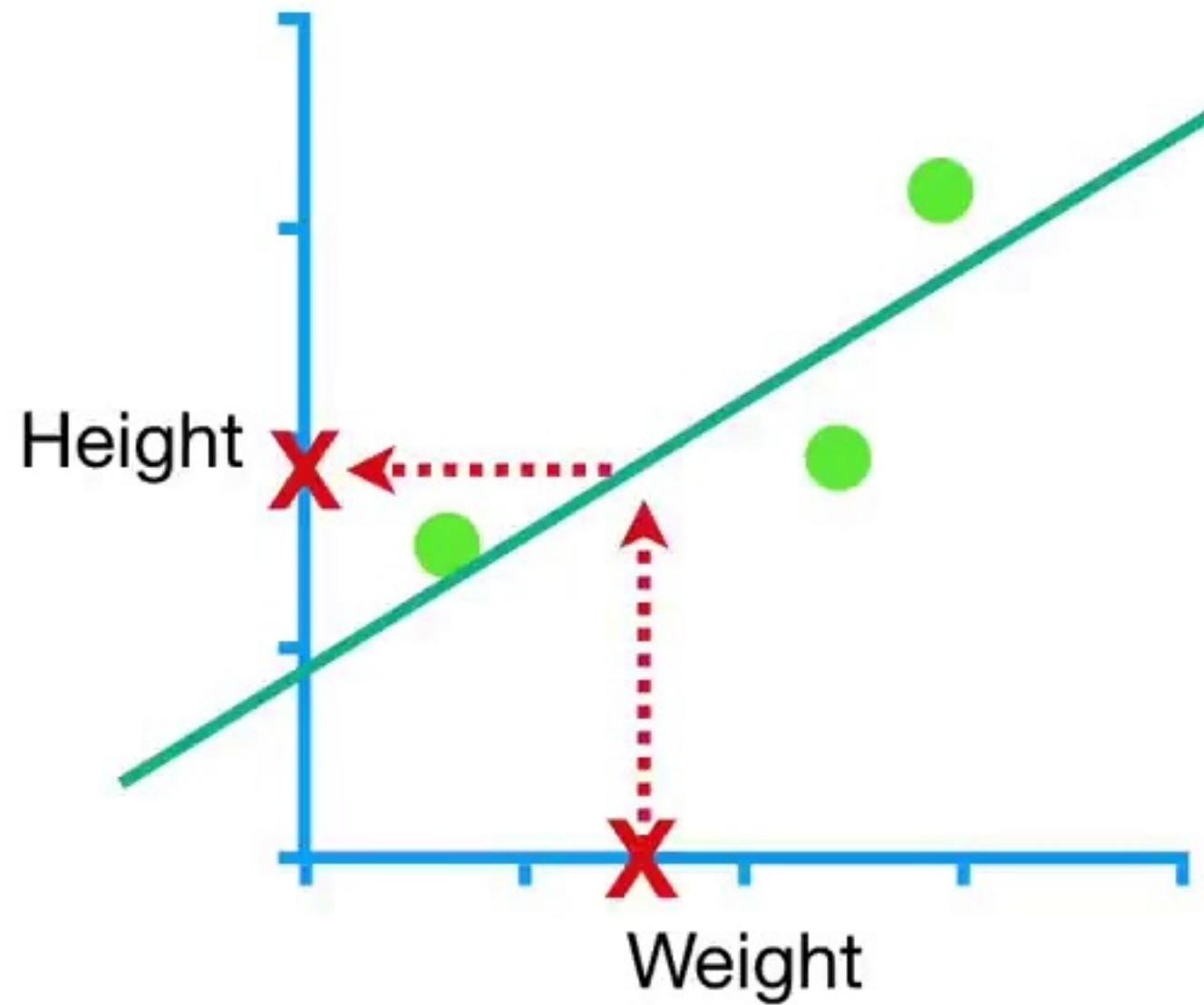




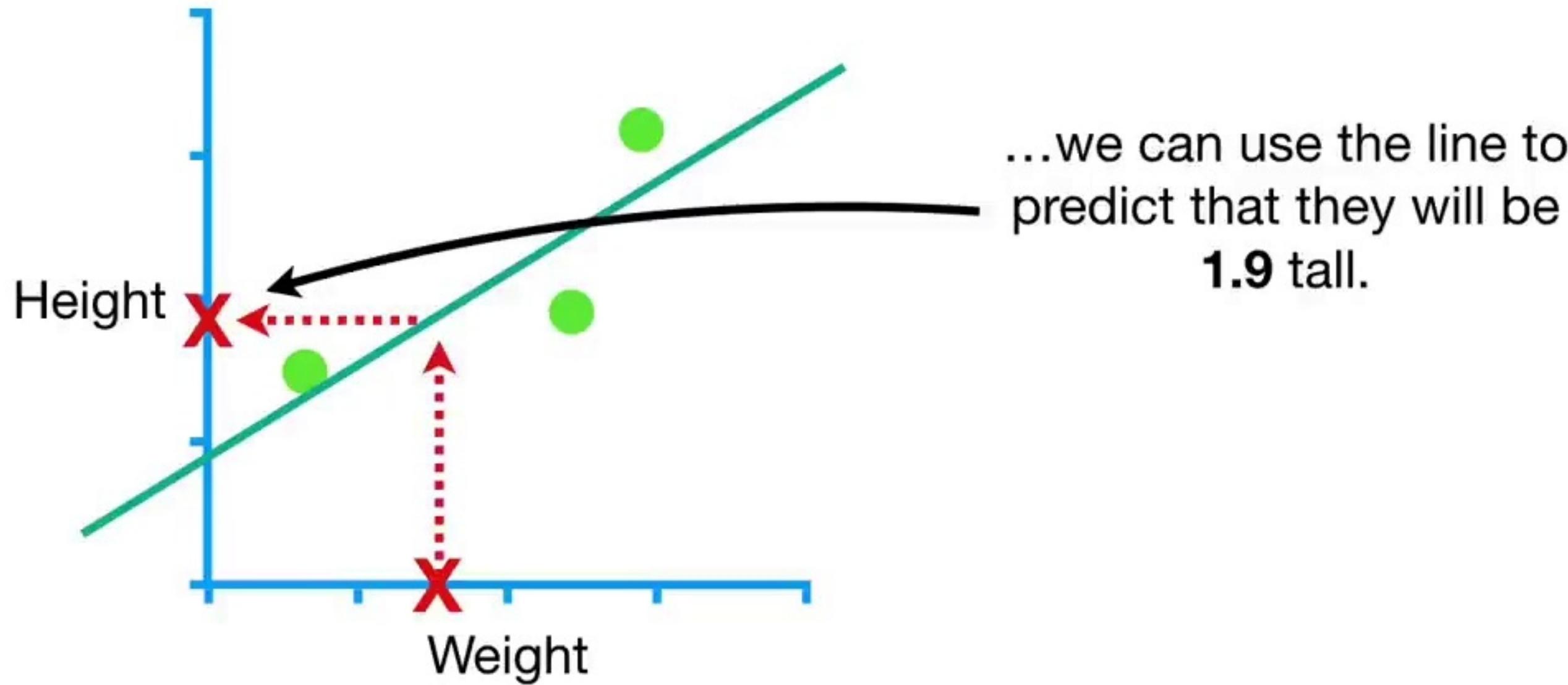
...and someone tells us
that they weigh **1.5**...



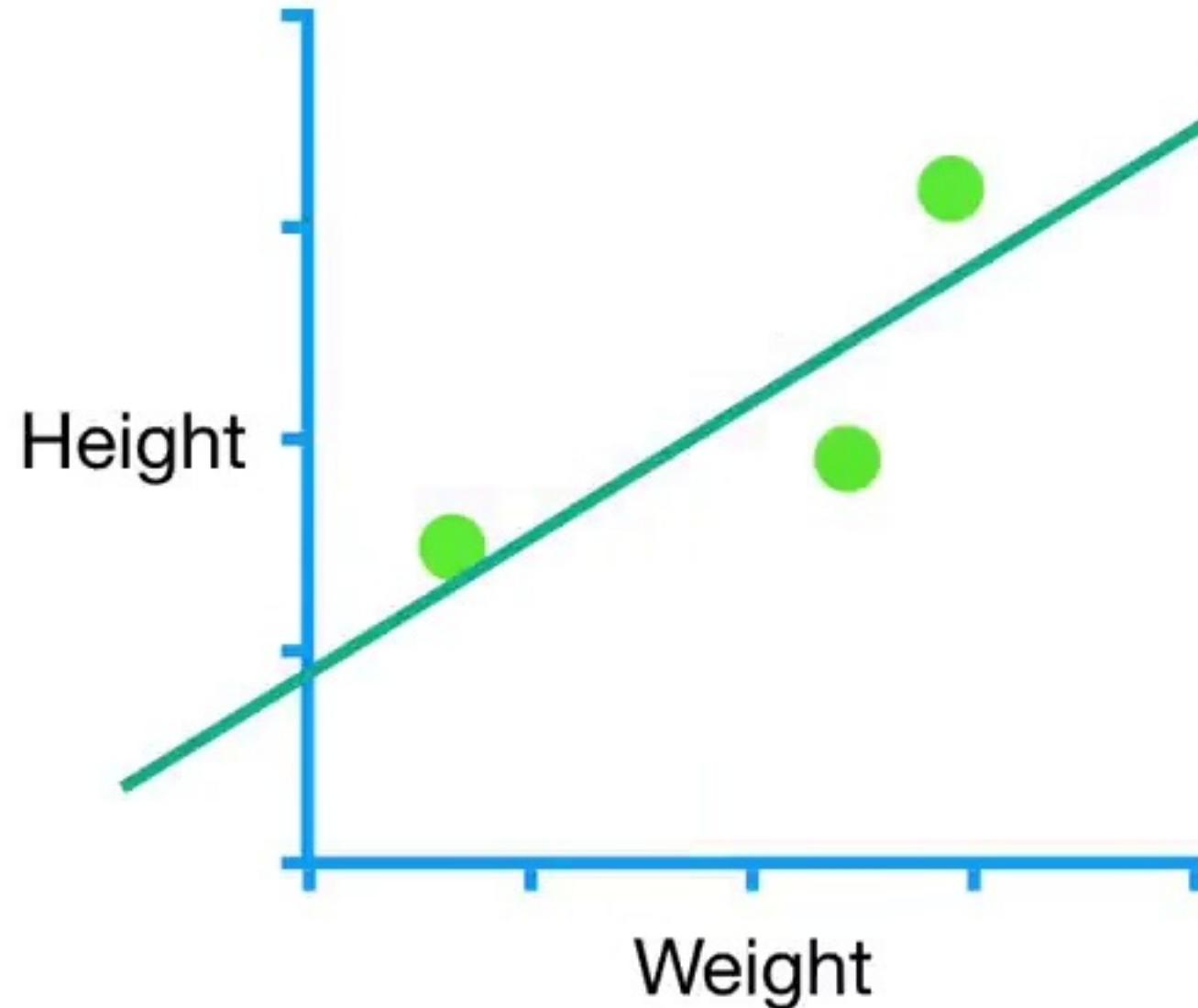
...we can use the line to
predict that they will be
1.9 tall.



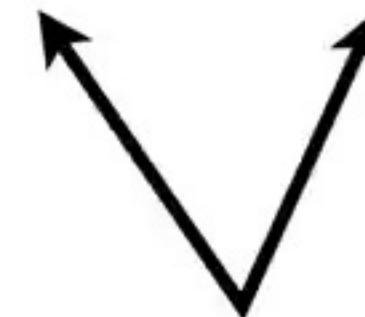
...we can use the line to
predict that they will be
1.9 tall.



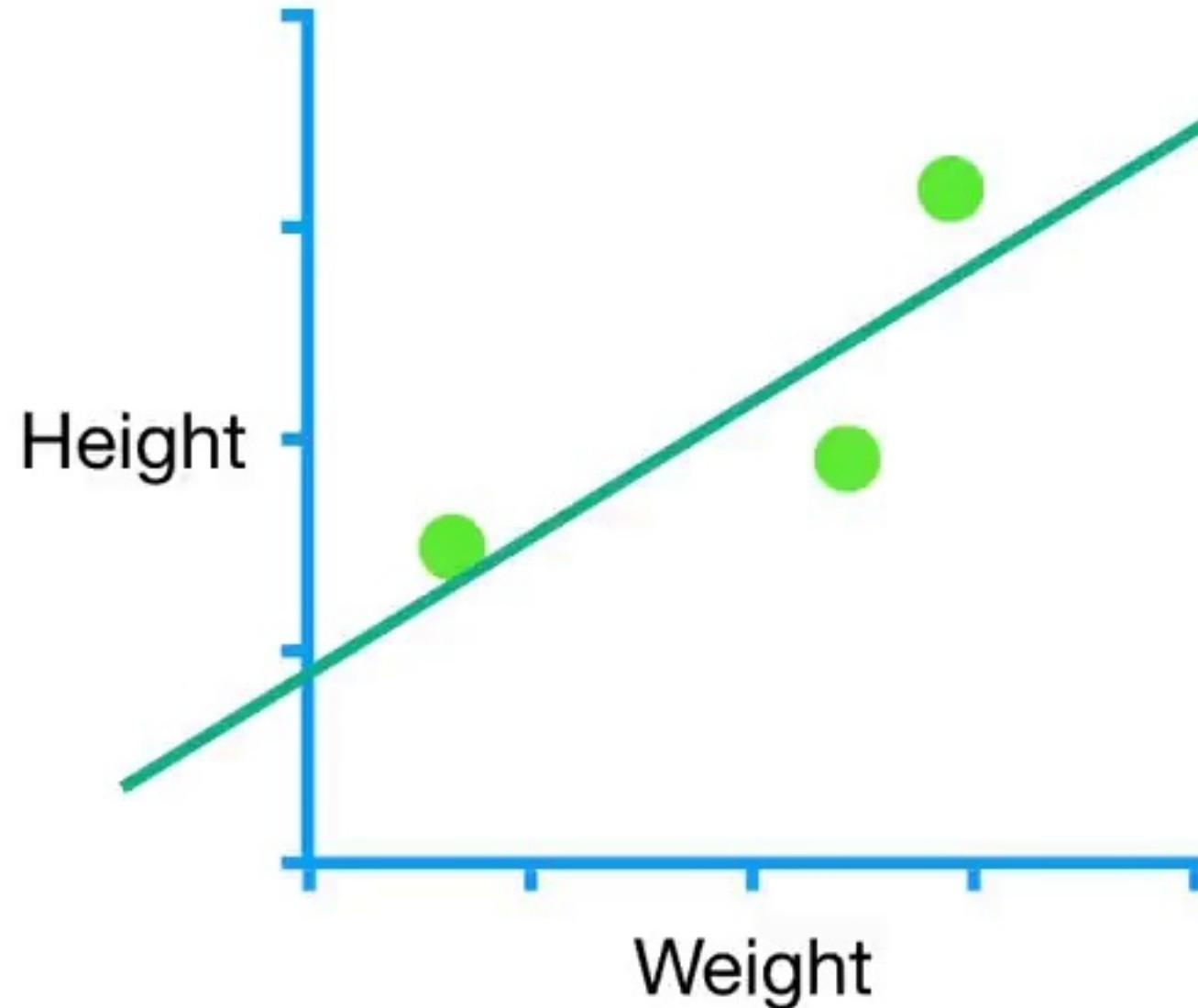
Predicted Height = intercept + slope × **Weight**



So let's learn how **Gradient Descent** can fit a line to data by finding the optimal values for the **Intercept** and the **Slope**.

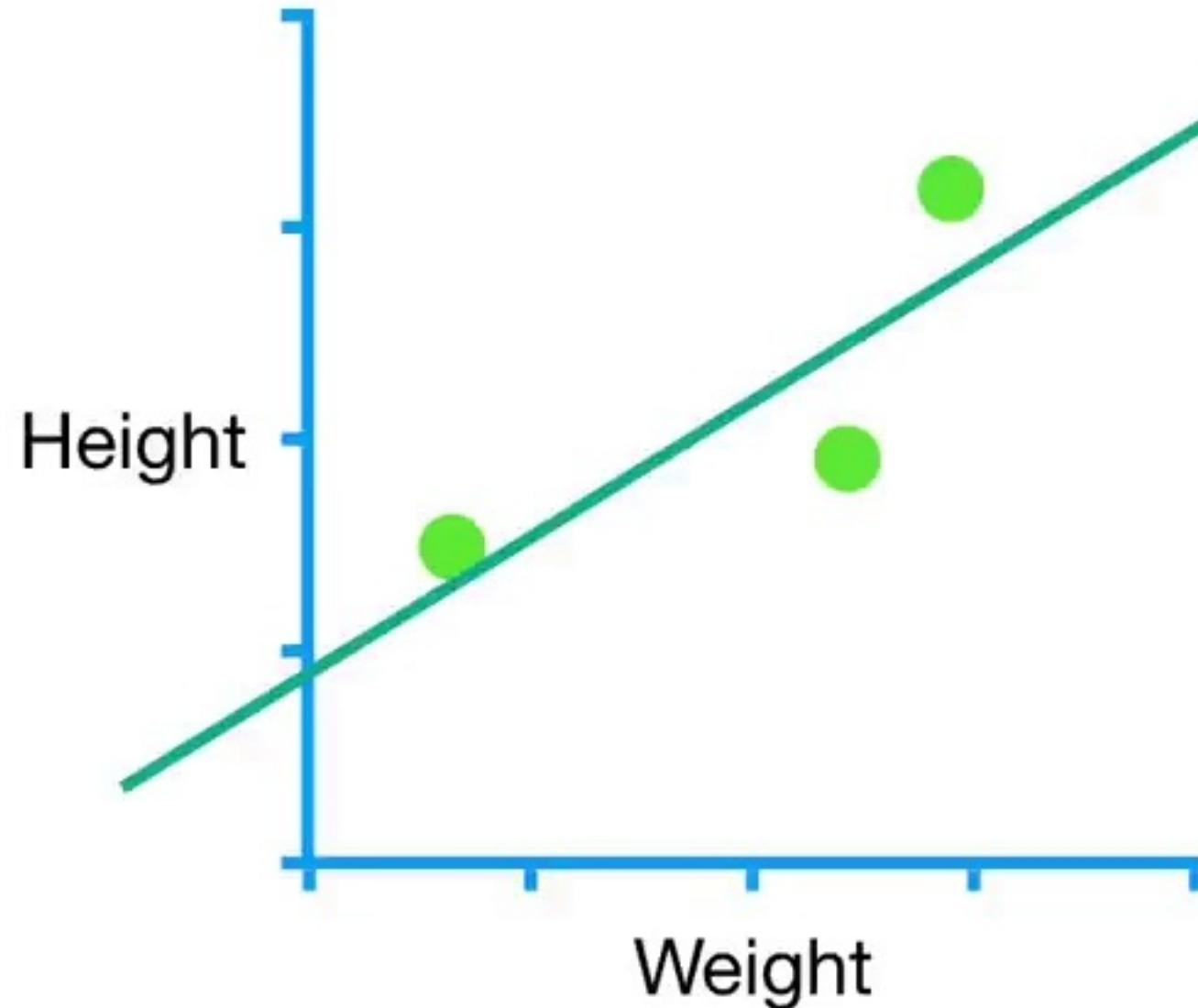


Predicted Height = intercept + slope × Weight



Actually, we'll start by using
Gradient Descent to find the
Intercept.

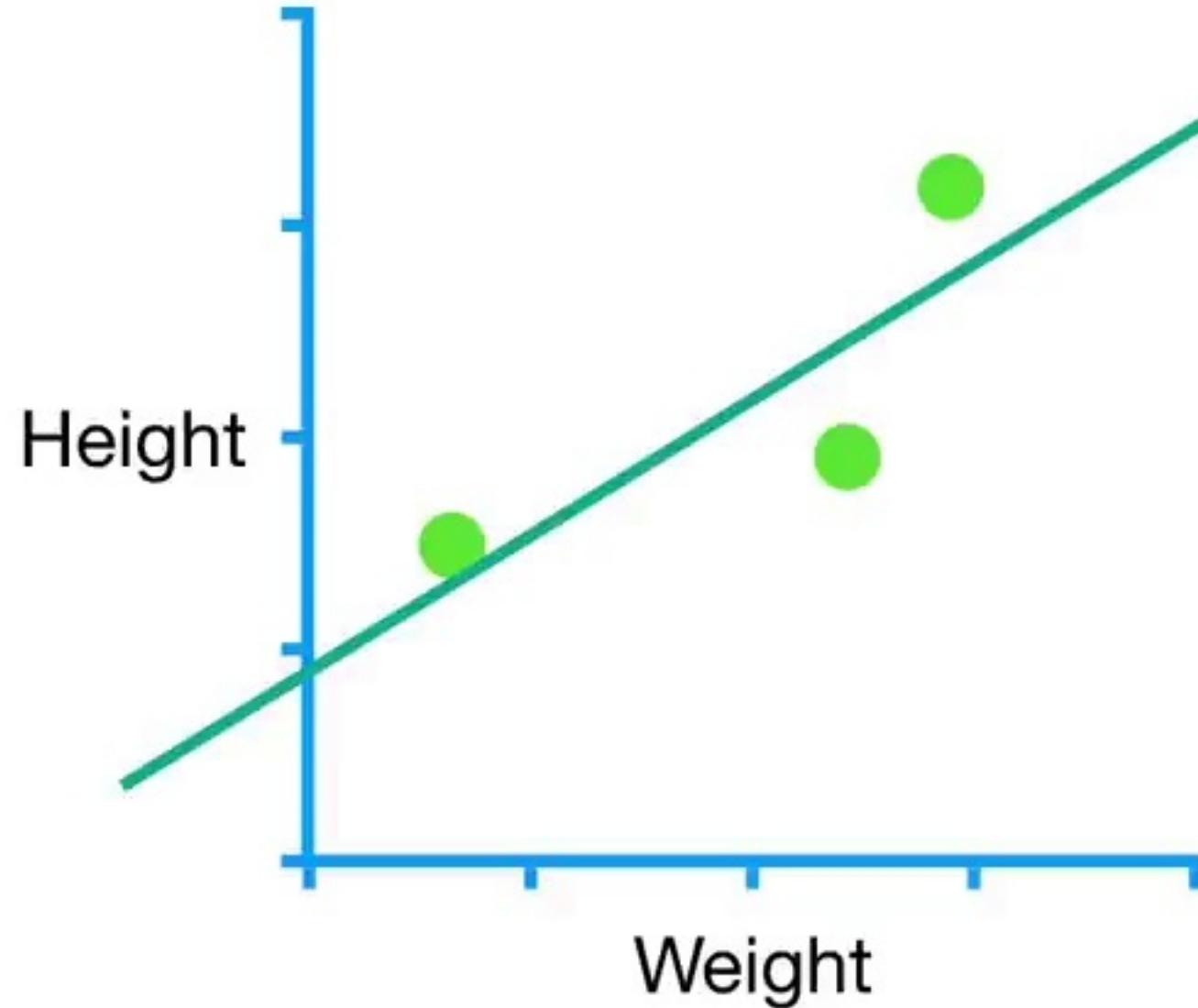
Predicted Height = intercept + slope \times Weight



Then, once we understand how
Gradient Descent works, we'll
use it to solve for the **Intercept**
and the **Slope**.

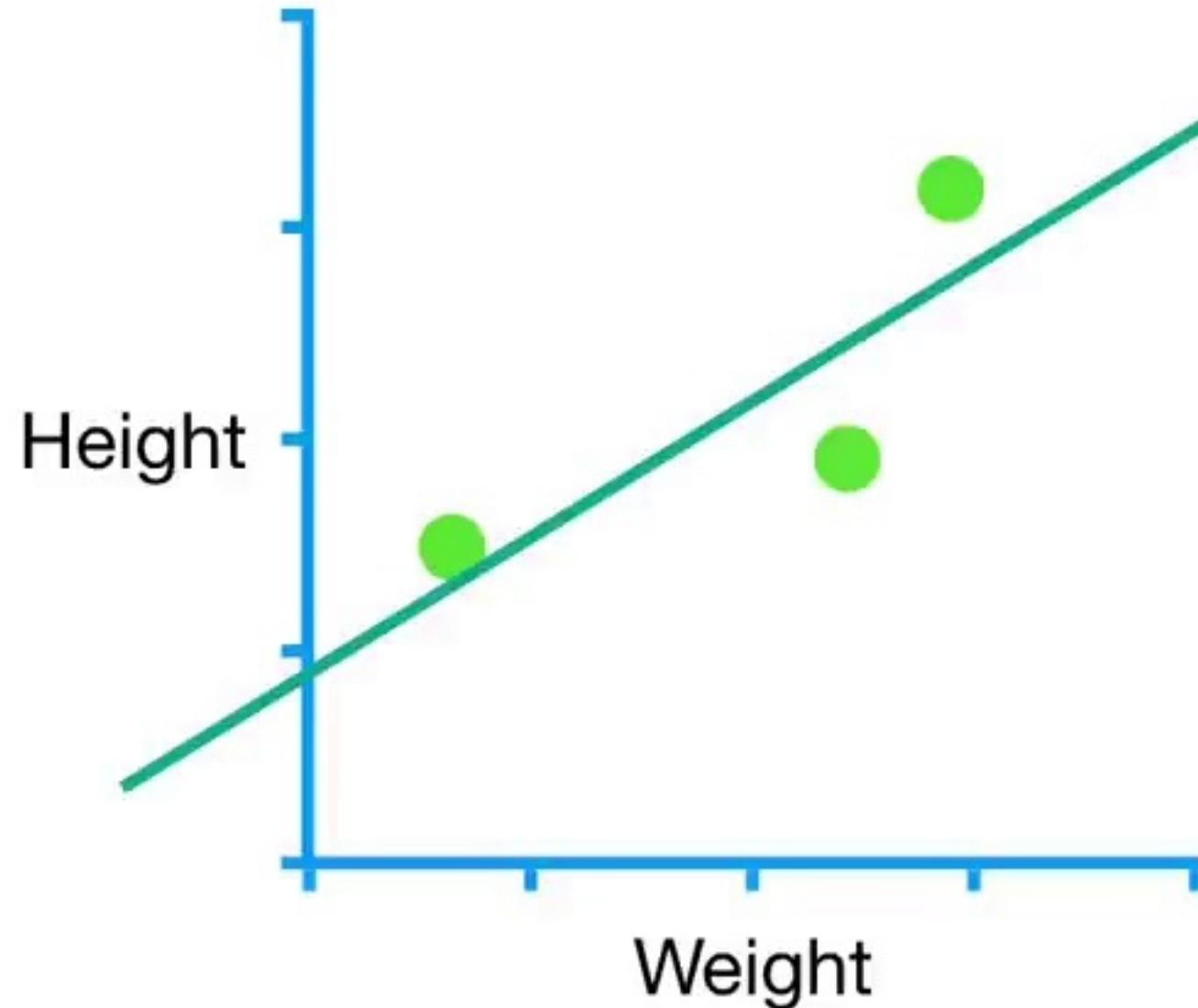


Predicted Height = intercept + slope × **Weight**



So for now, let's just plug in
the **Least Squares** estimate
for the **Slope**, 0.64.

Predicted Height = intercept + 0.64 × **Weight**

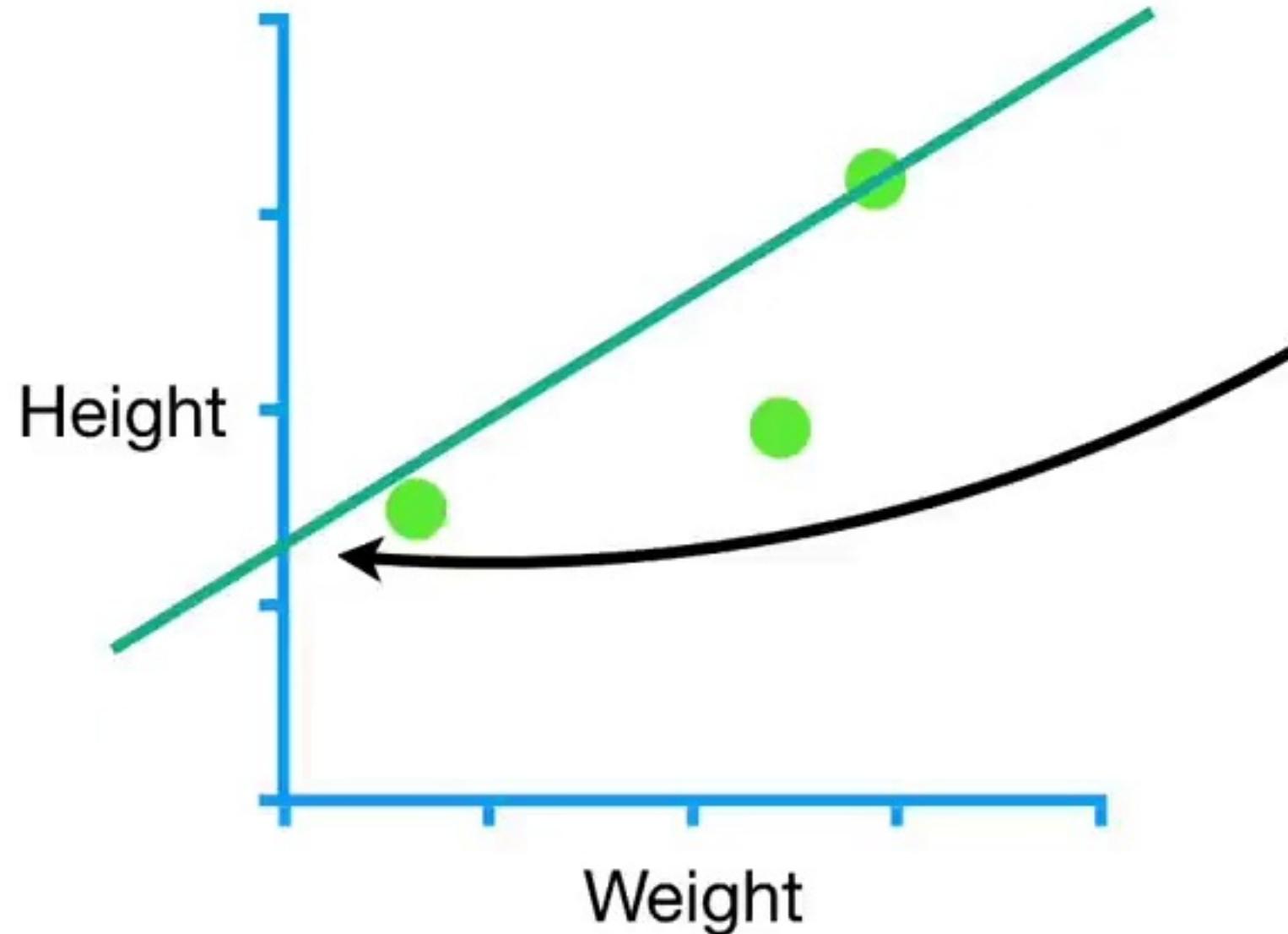


So for now, let's just plug in
the **Least Squares** estimate
for the **Slope**, 0.64.

0.64

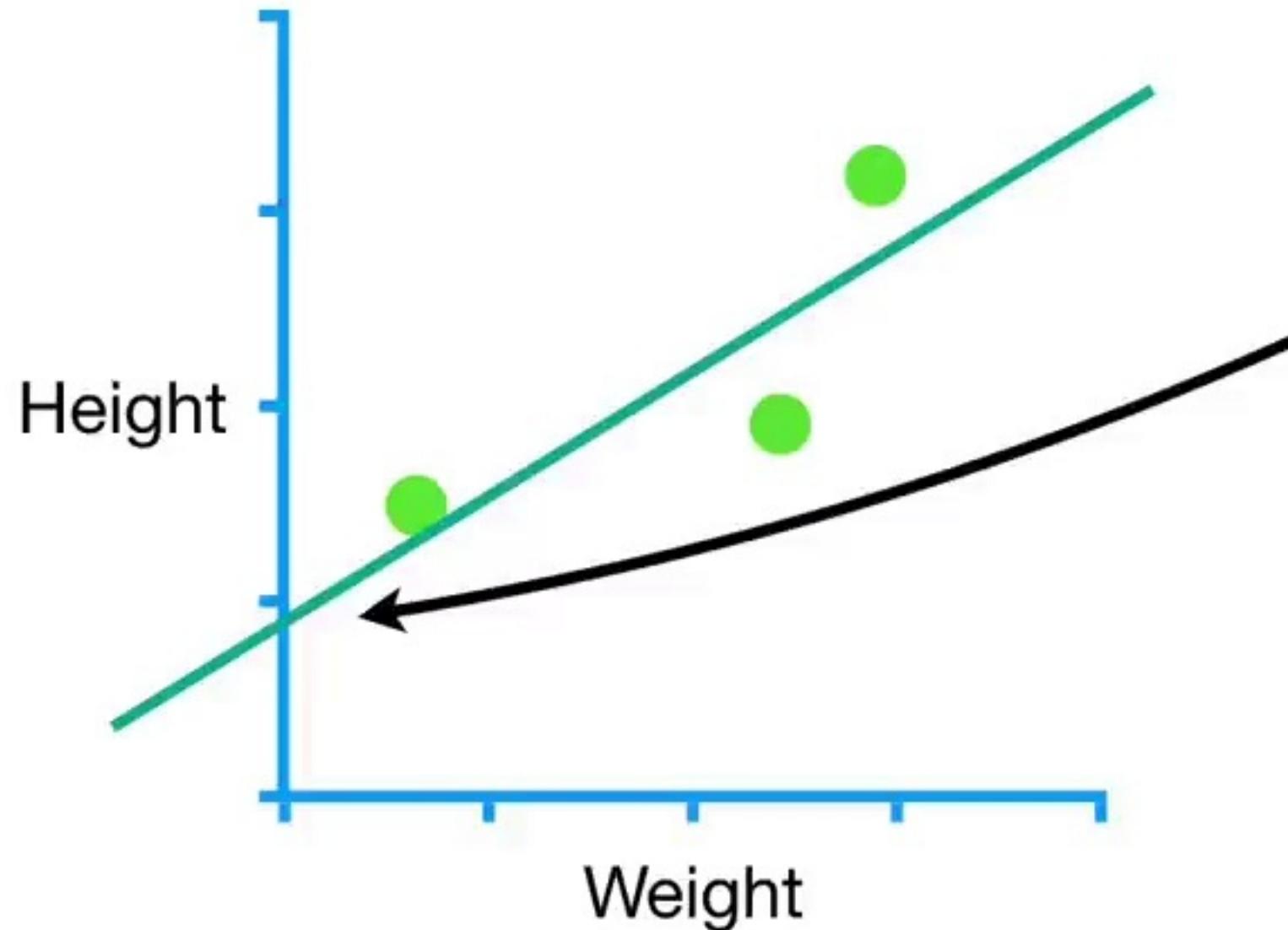


Predicted Height = intercept + $0.64 \times \text{Weight}$



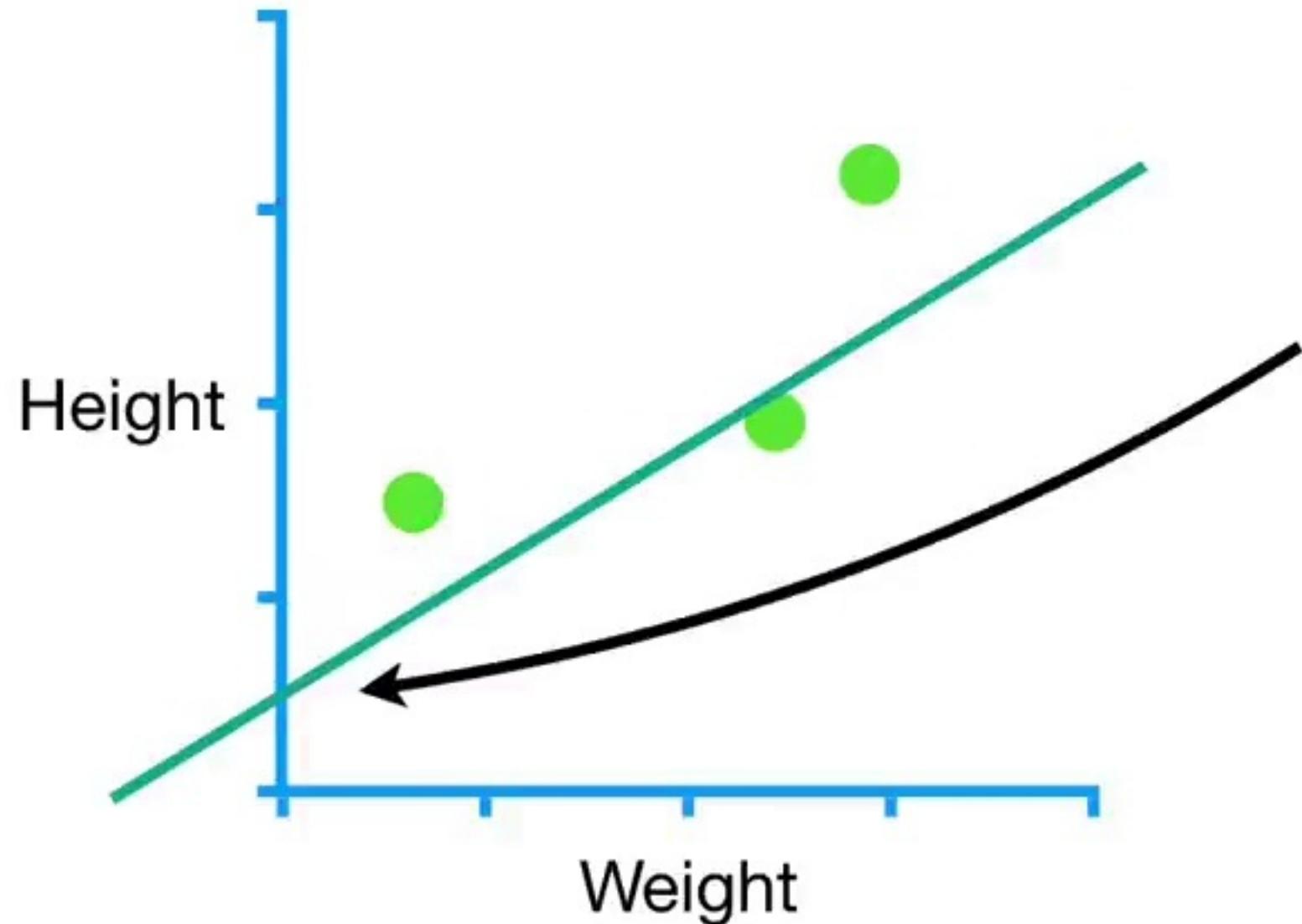
...and we'll use **Gradient Descent** to find the optimal value for the Intercept.

Predicted Height = intercept + 0.64 × **Weight**



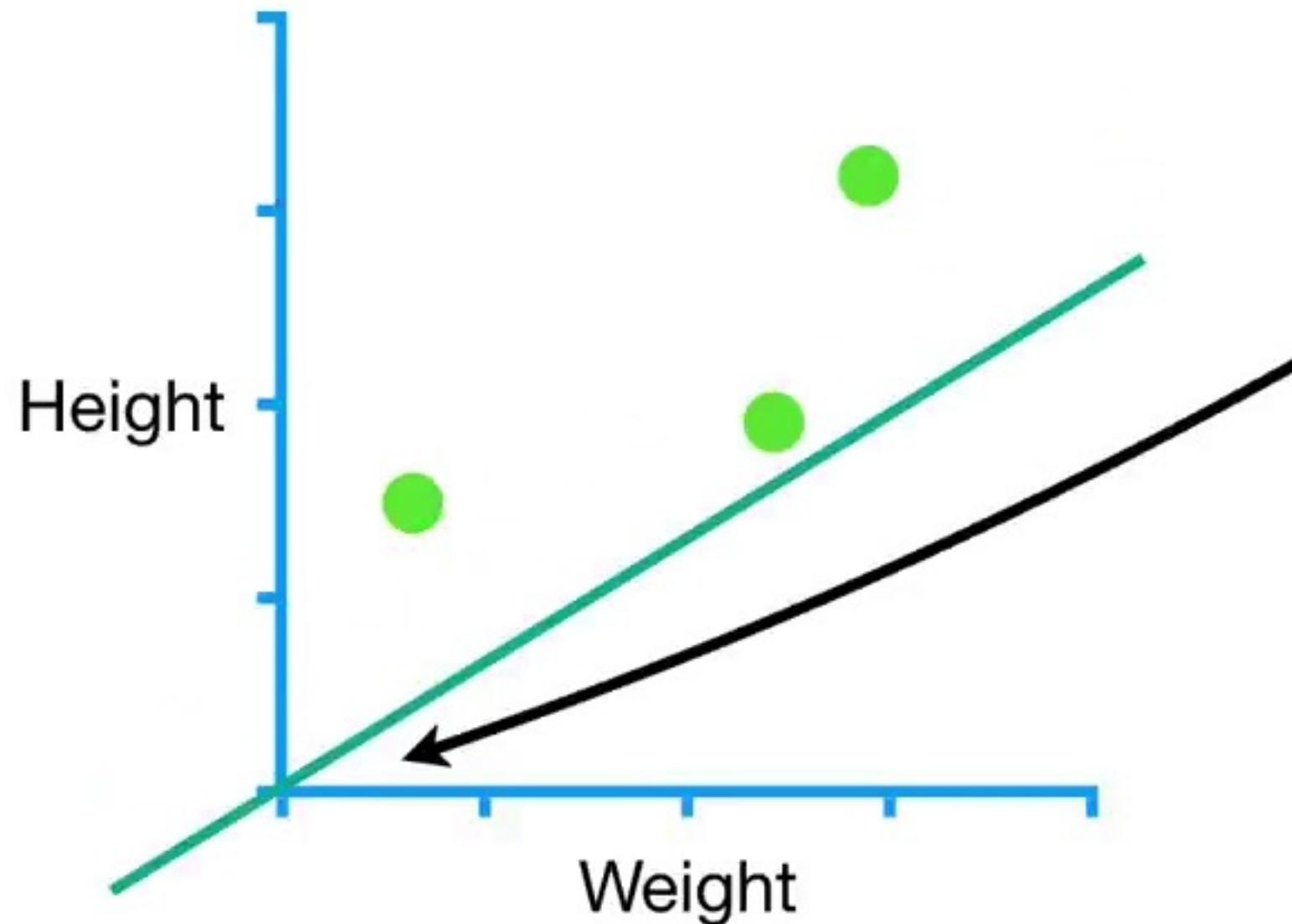
...and we'll use **Gradient Descent** to find the optimal value for the Intercept.

Predicted Height = intercept + 0.64 × **Weight**



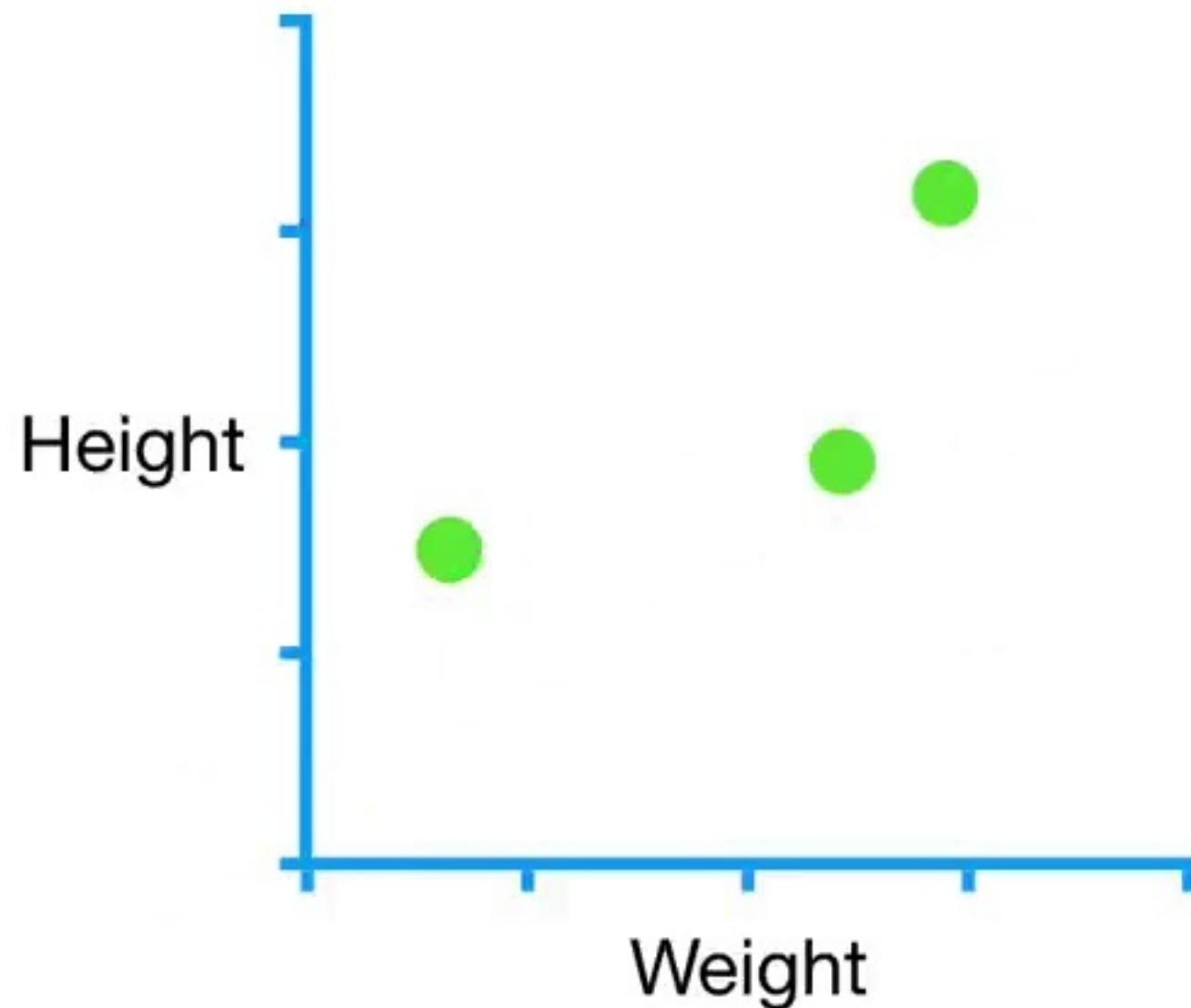
...and we'll use **Gradient Descent** to find the optimal value for the Intercept.

Predicted Height = intercept + $0.64 \times \text{Weight}$



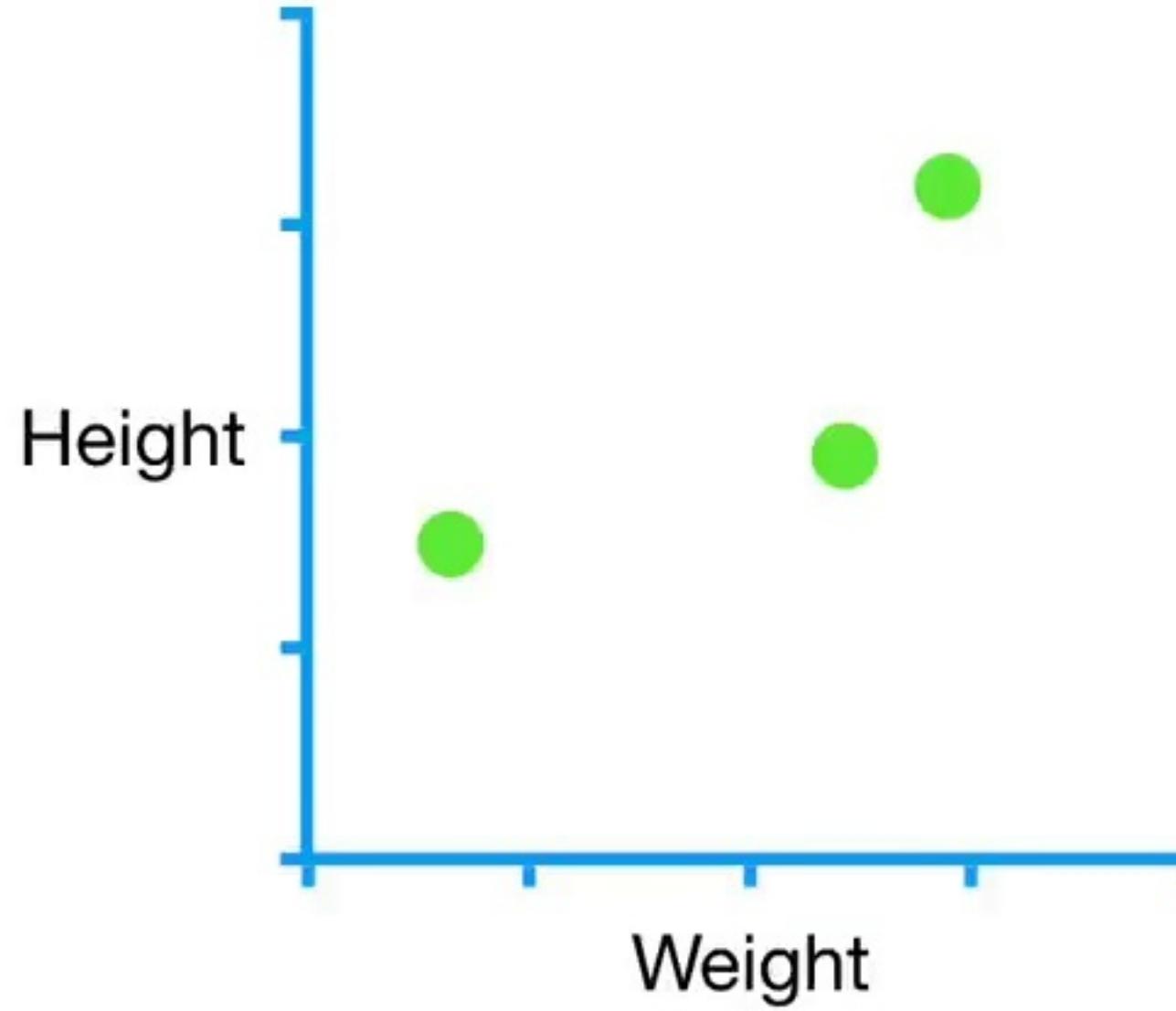
...and we'll use **Gradient Descent** to find the optimal value for the Intercept.

Predicted Height = intercept + 0.64 × **Weight**



The first thing we do is
pick a random value for
the **Intercept**.

Predicted Height = intercept + $0.64 \times \text{Weight}$



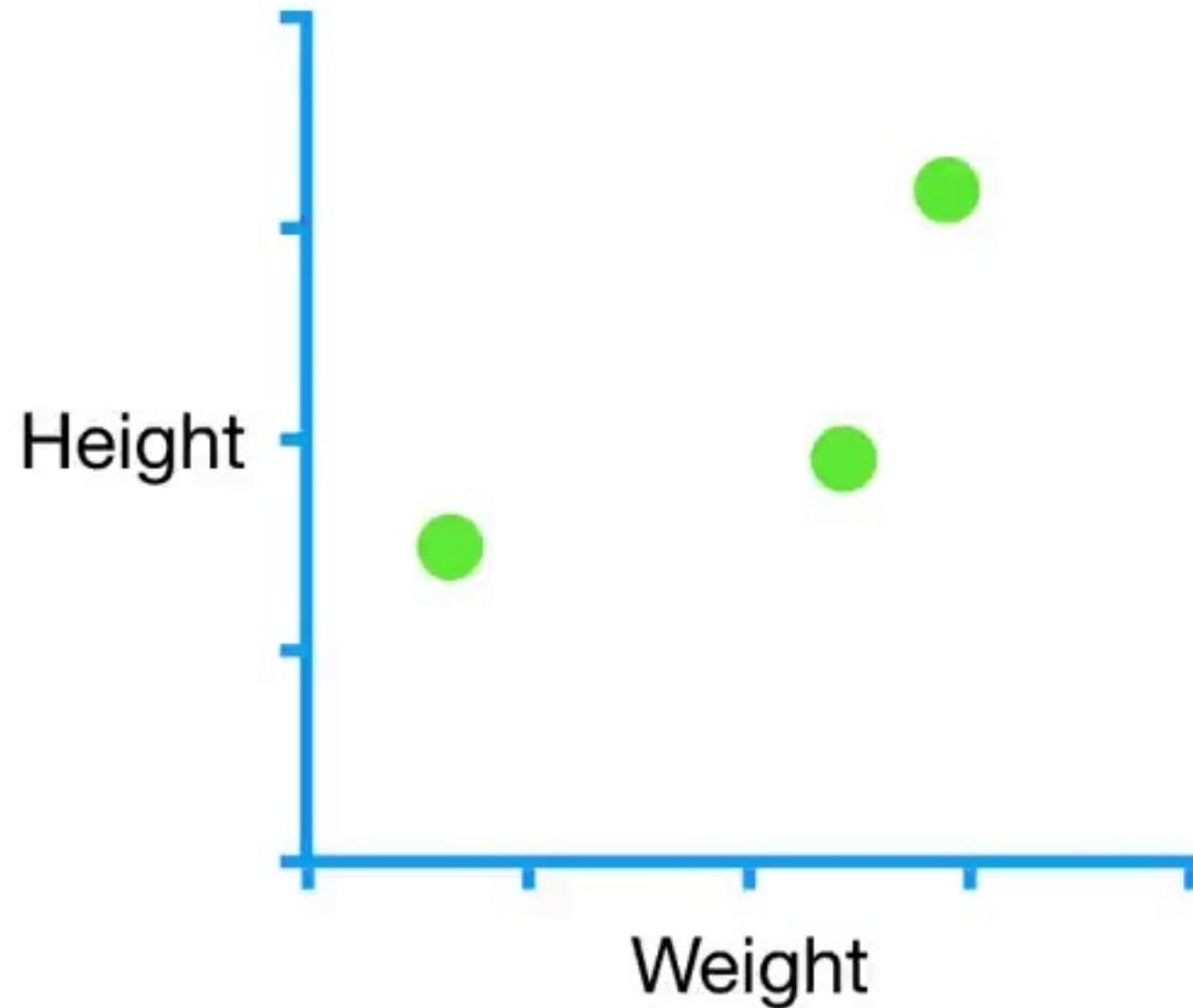
The first thing we do is pick a random value for the **Intercept**.

This is just an initial guess that gives **Gradient Descent** something to improve upon.

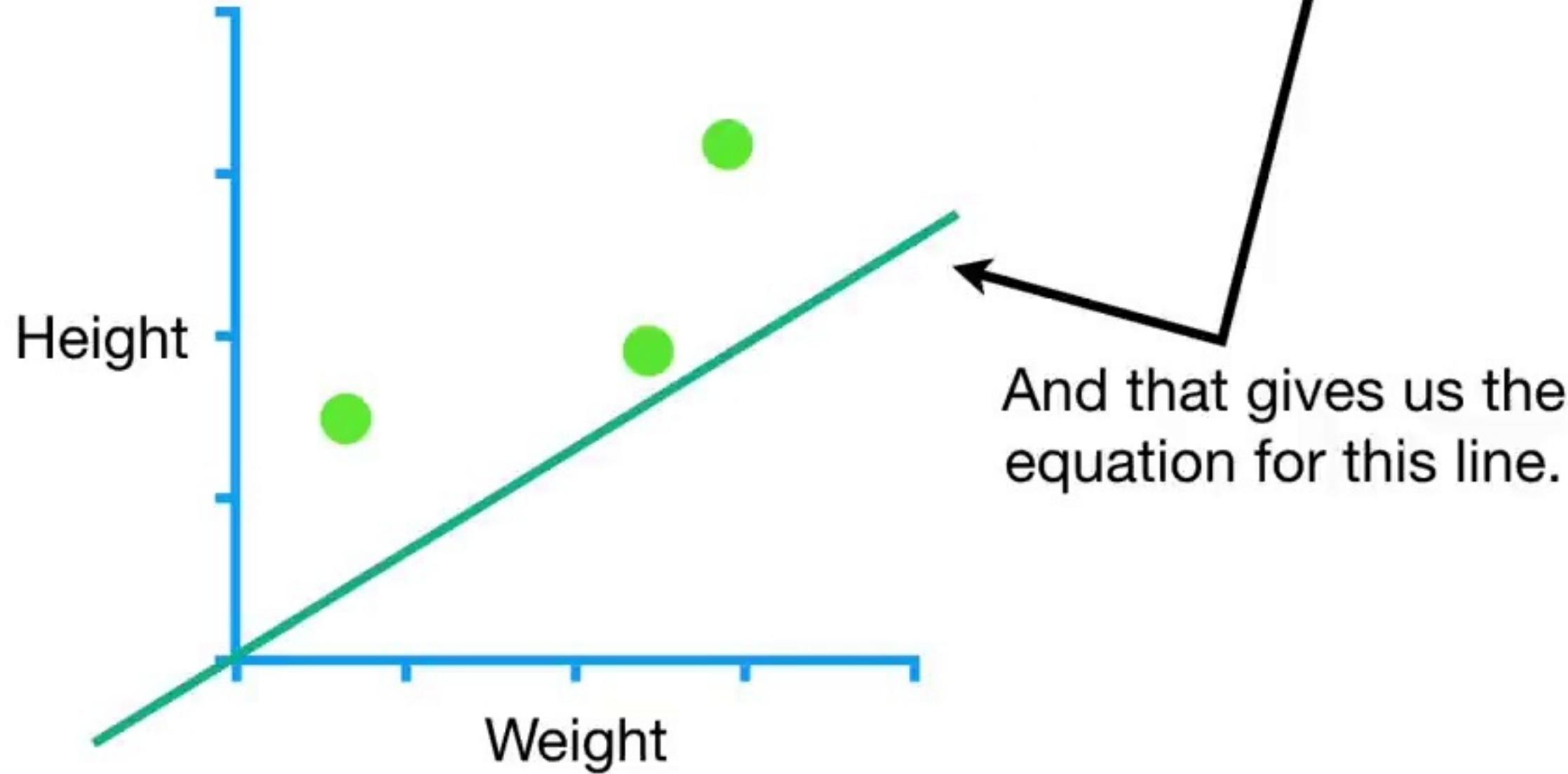
Predicted Height = 0 + 0.64 × Weight

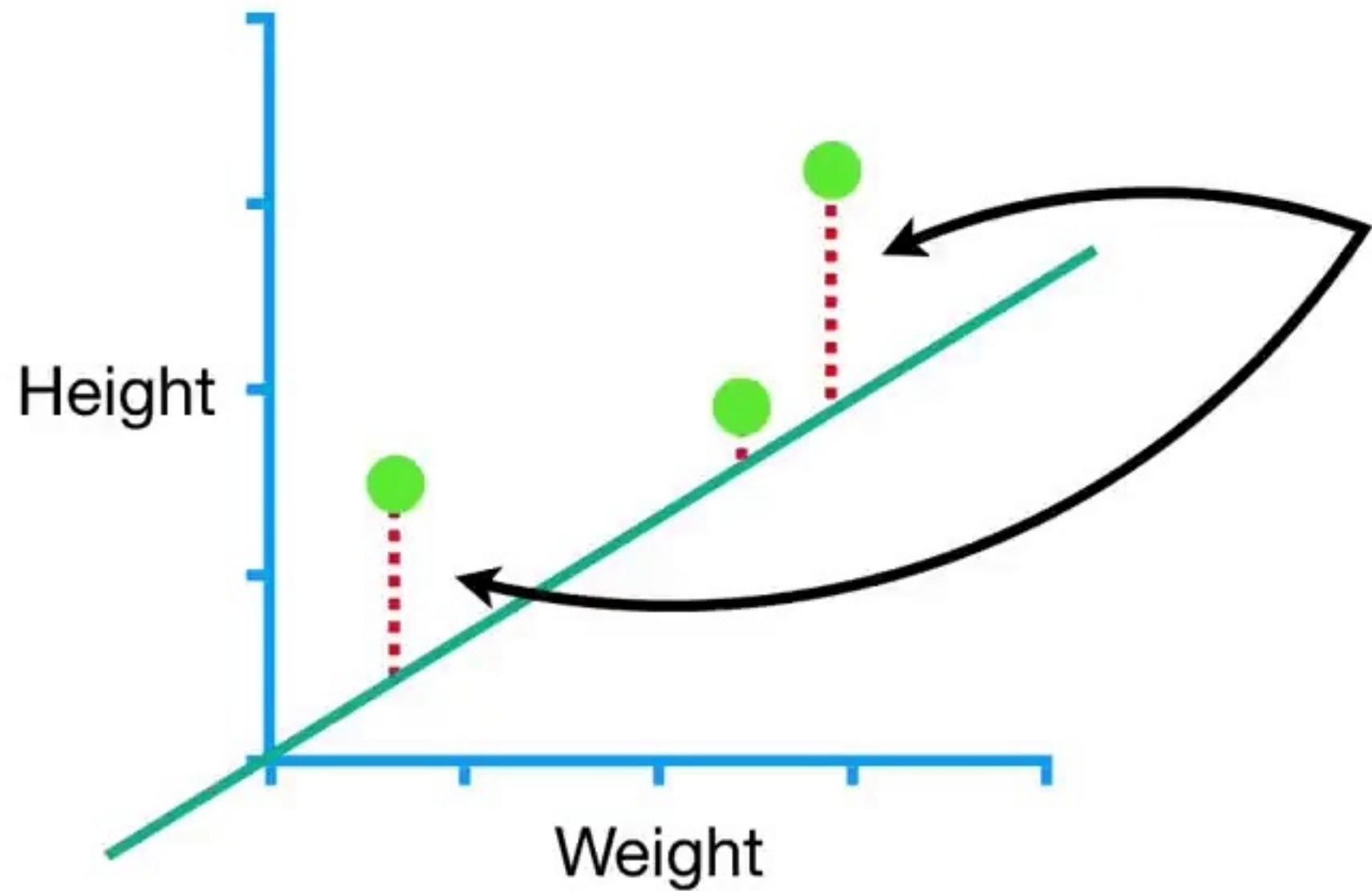


In this case, we'll use **0**,
but any number will do.

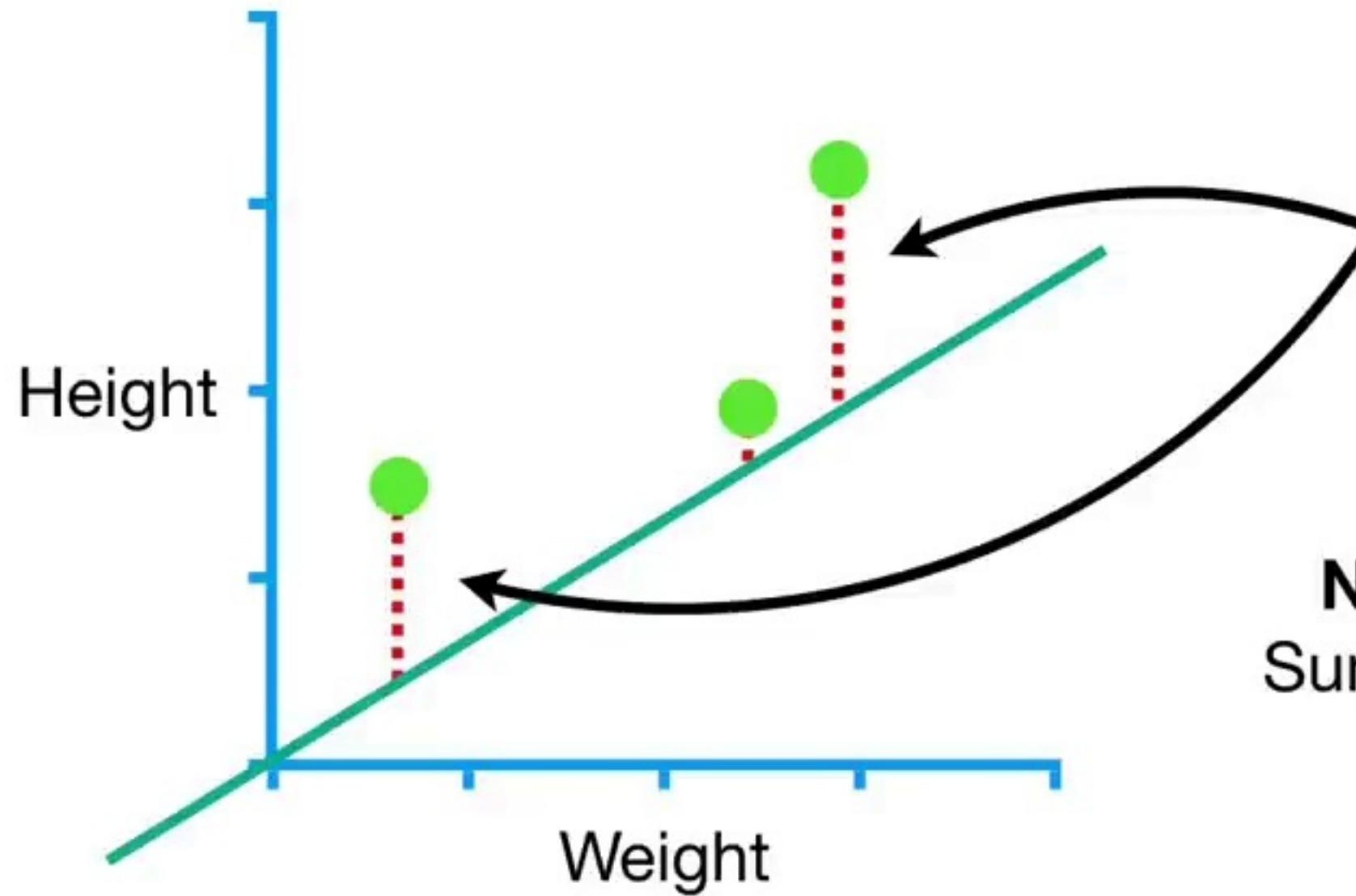


Predicted Height = 0 + 0.64 × Weight



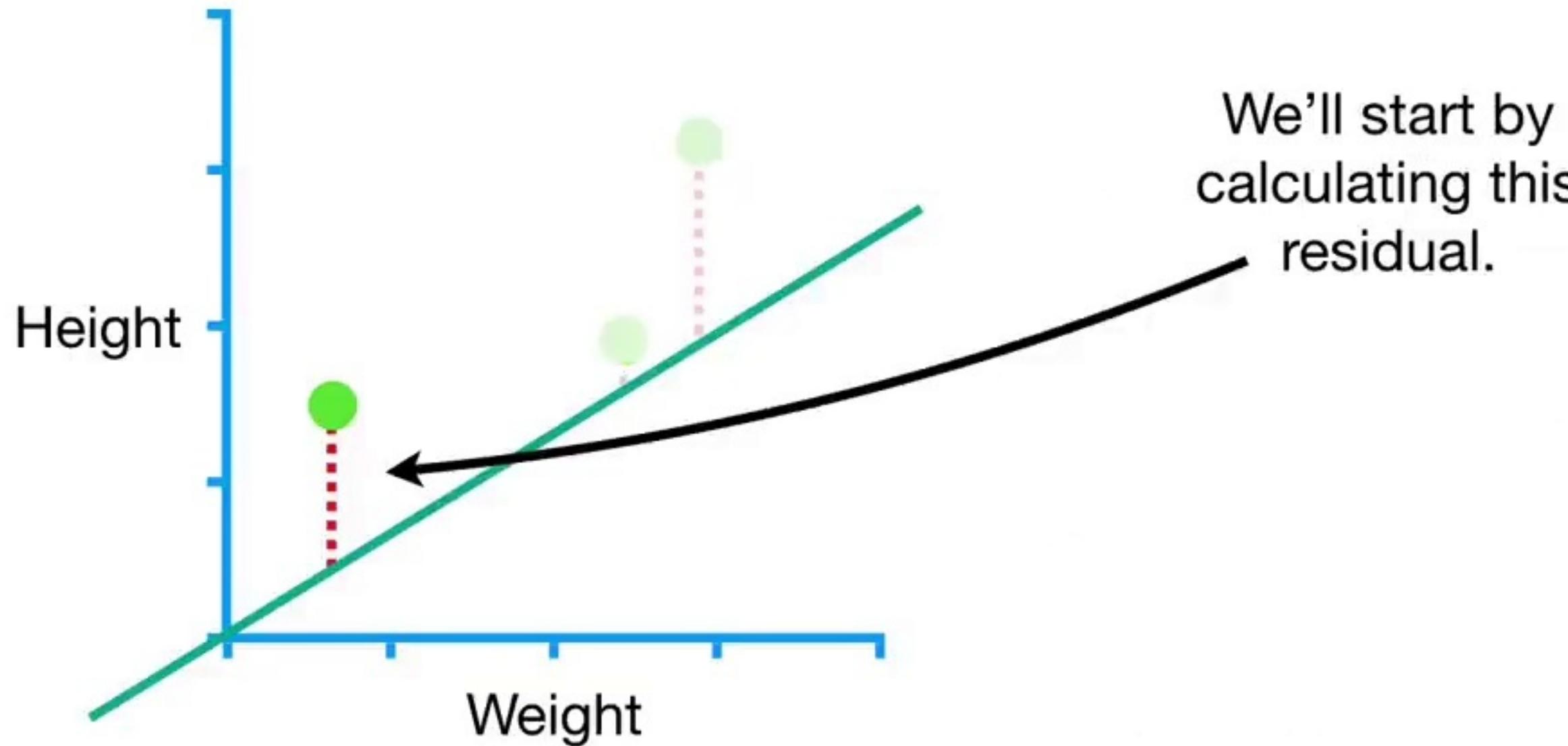


In this example, we will evaluate how well this line fits the data with the **Sum of the Squared Residuals.**

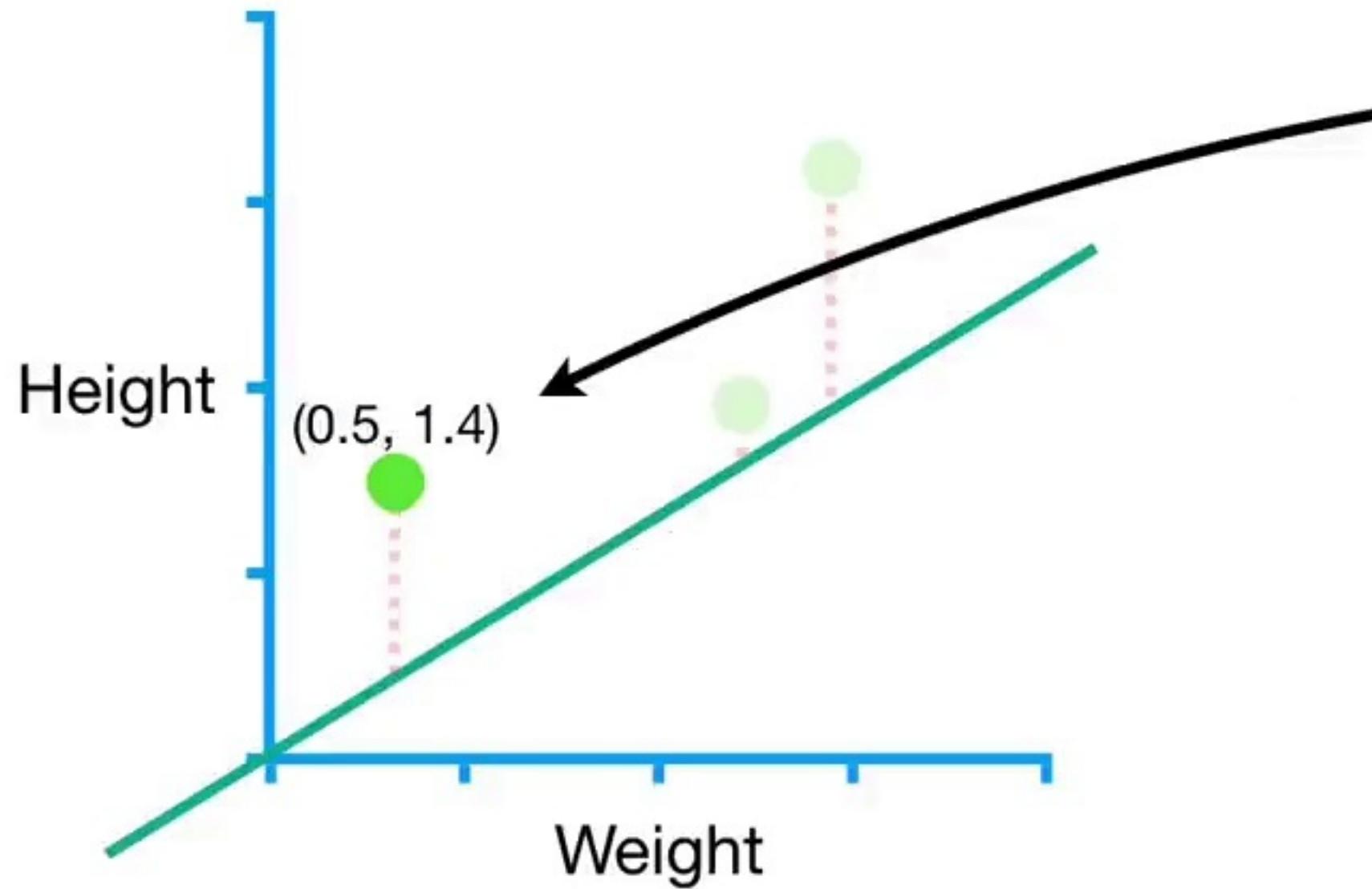


In this example, we will evaluate how well this line fits the data with the **Sum of the Squared Residuals.**

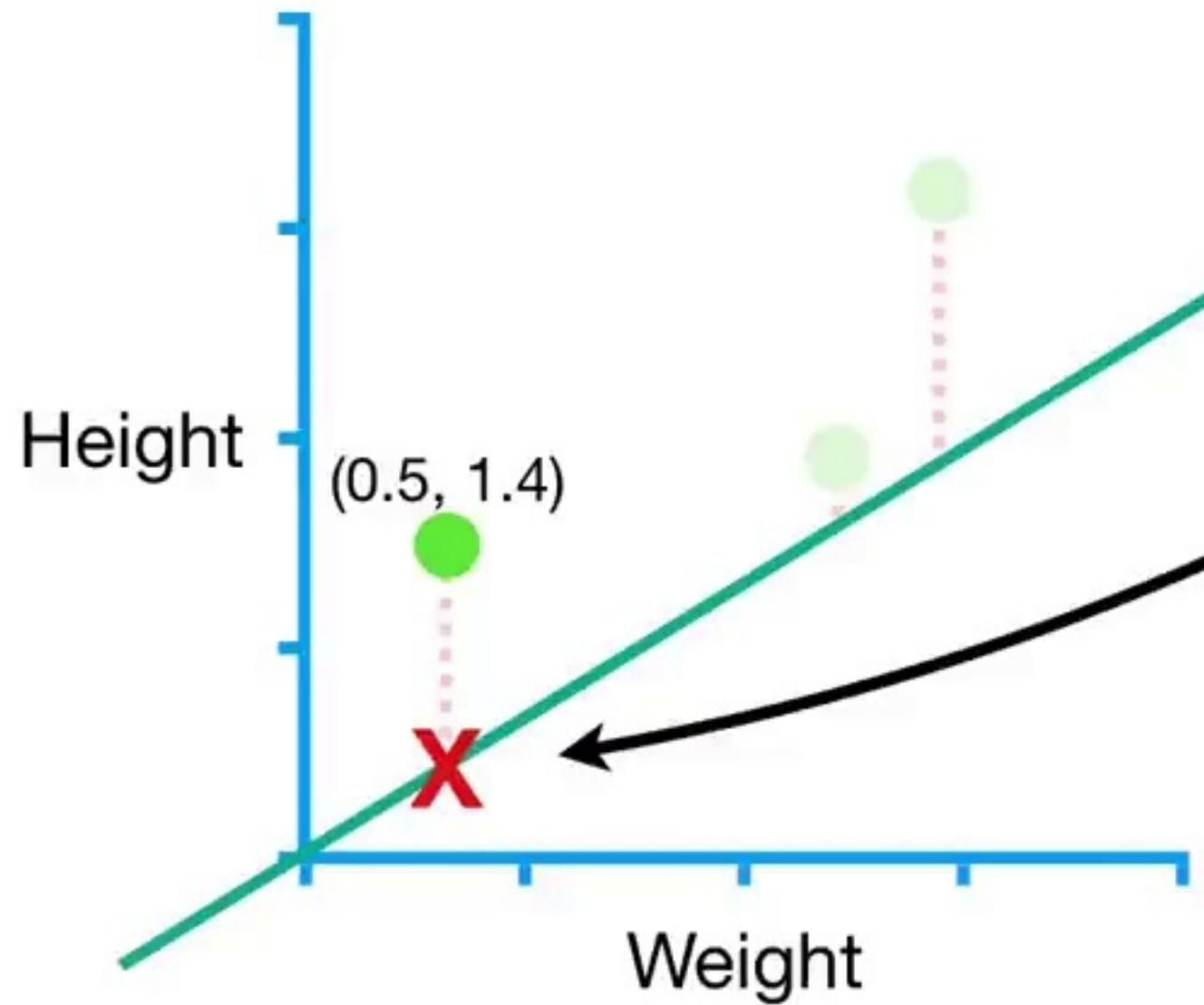
NOTE: In Machine Learning lingo, The Sum of the Squared Residuals is a type of **Loss Function.**



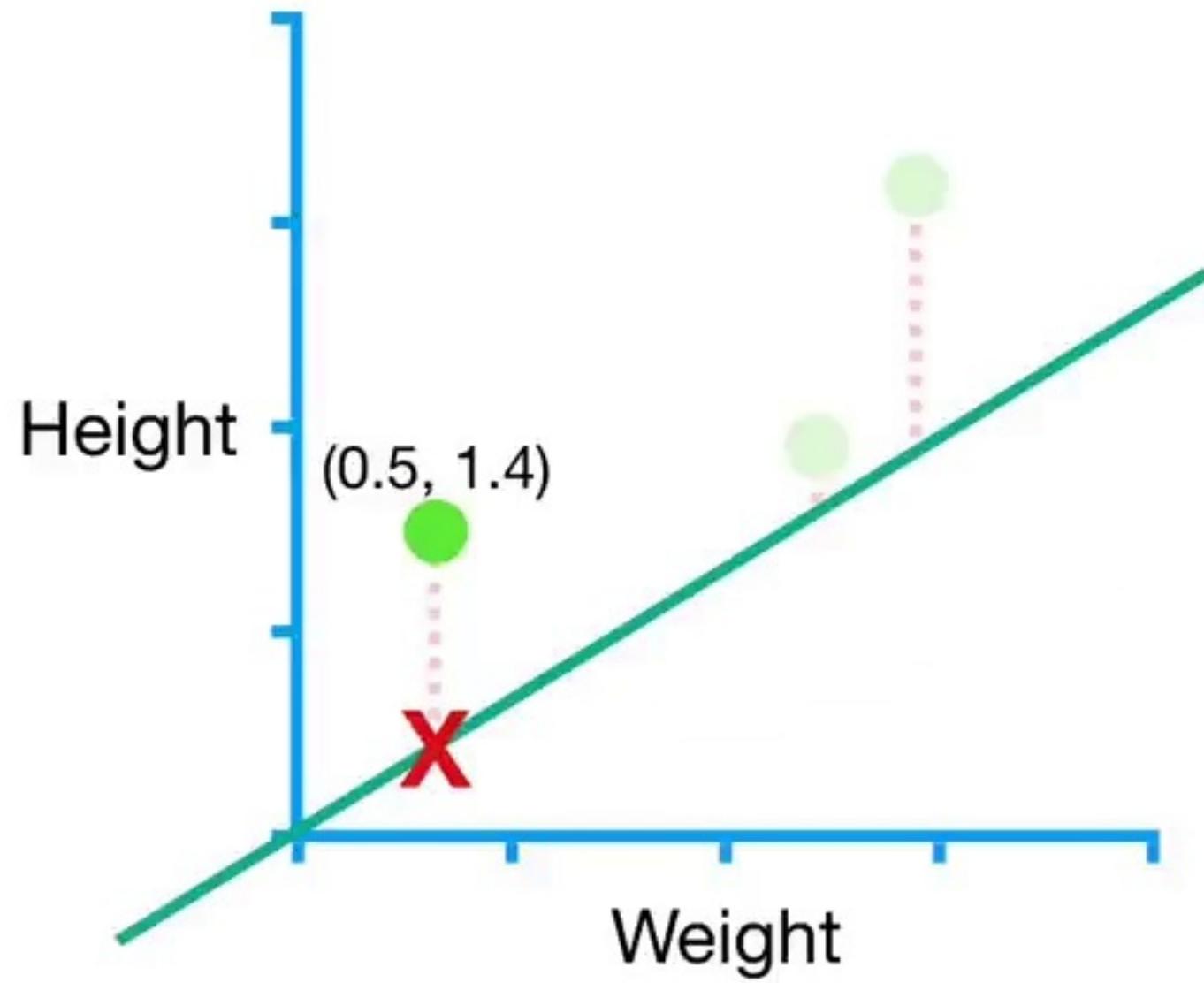
We'll start by calculating this residual.



This datapoint
represents a
person with
Weight 0.5 and
Height 1.4.



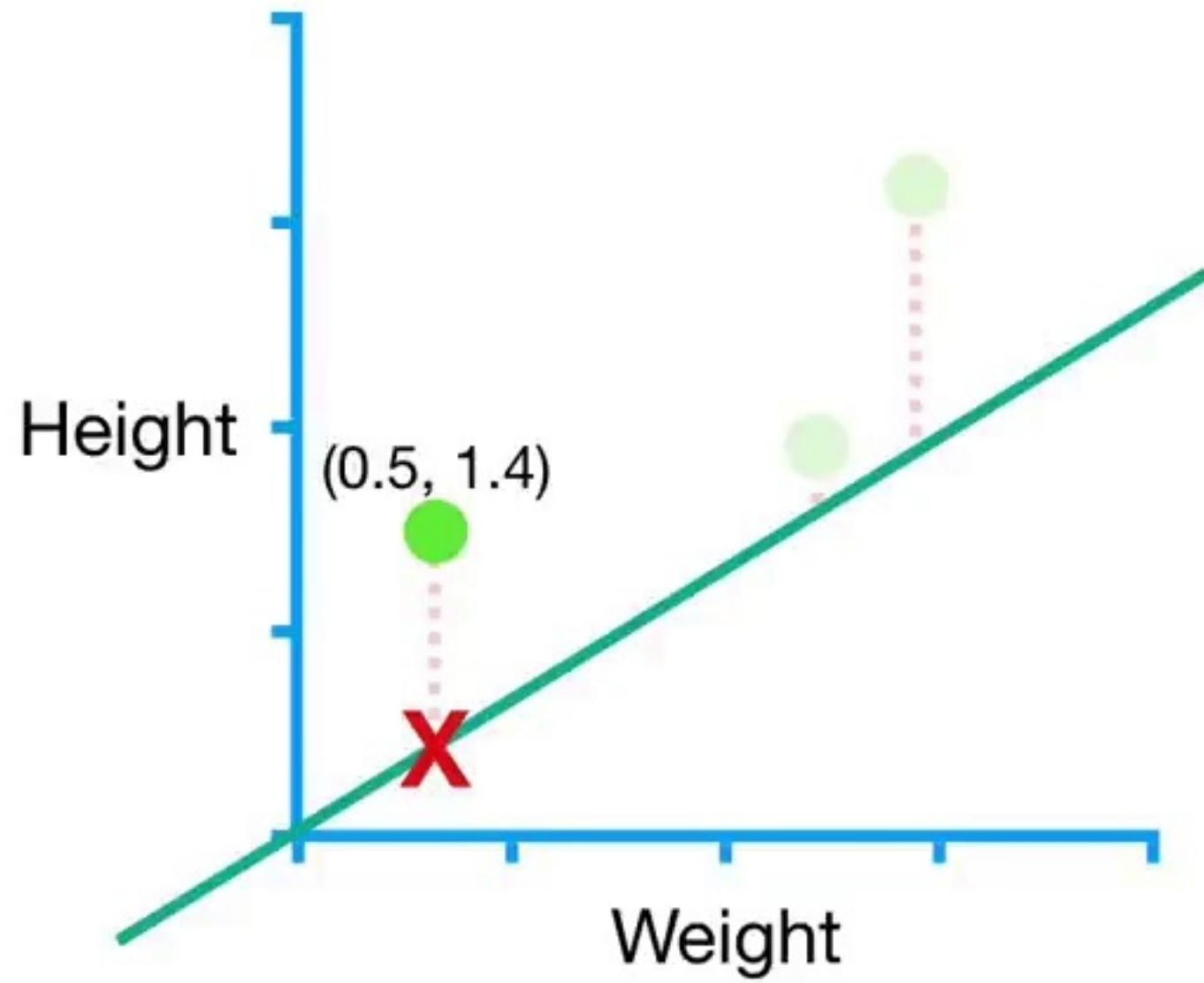
We get the **Predicted Height**, the point on the line...



We get the **Predicted Height**, the point on the line...

...by plugging
Weight = 0.5 into the equation for the line...

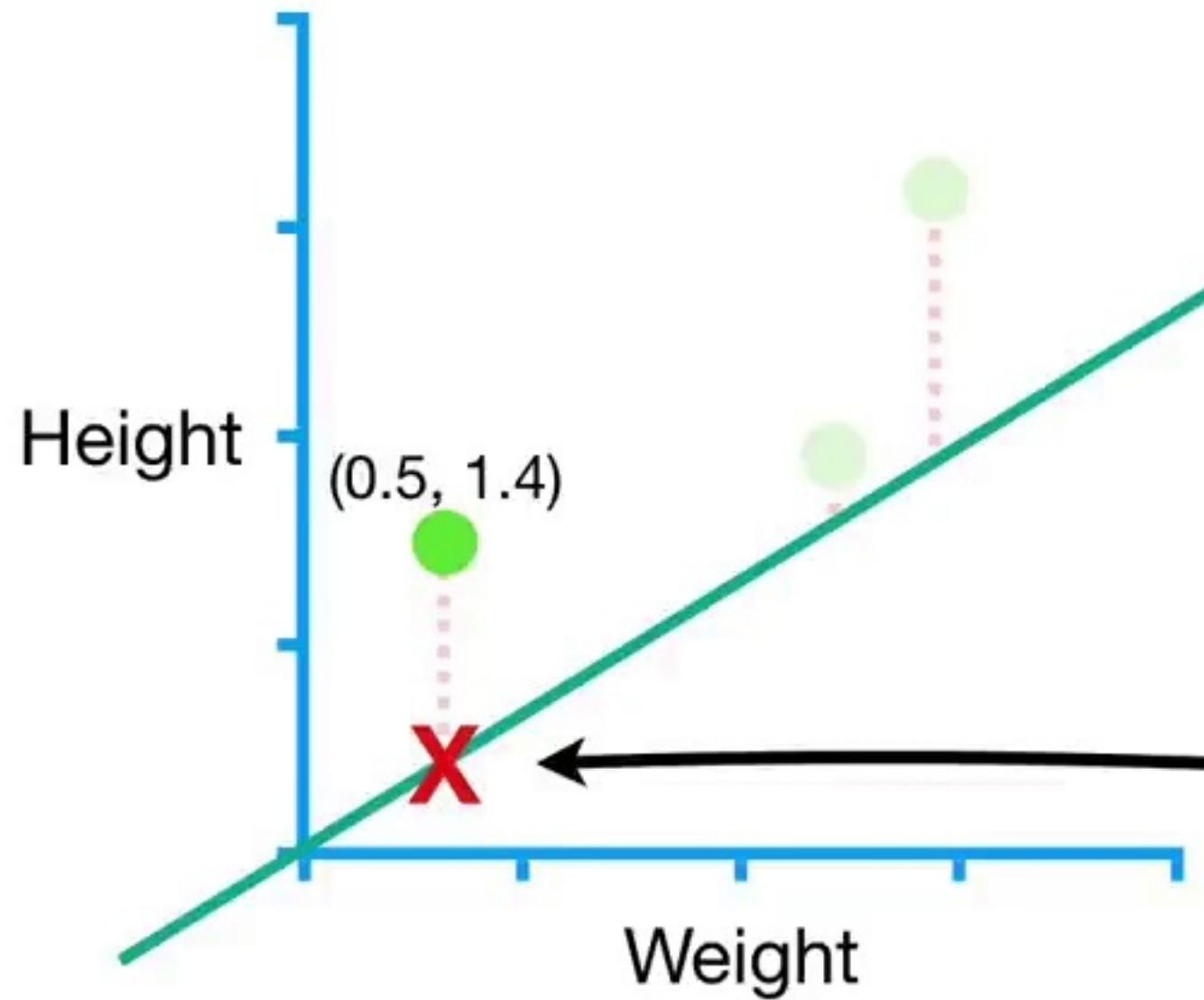
$$\text{Predicted Height} = 0 + 0.64 \times \text{Weight}$$



We get the **Predicted Height**, the point on the line...

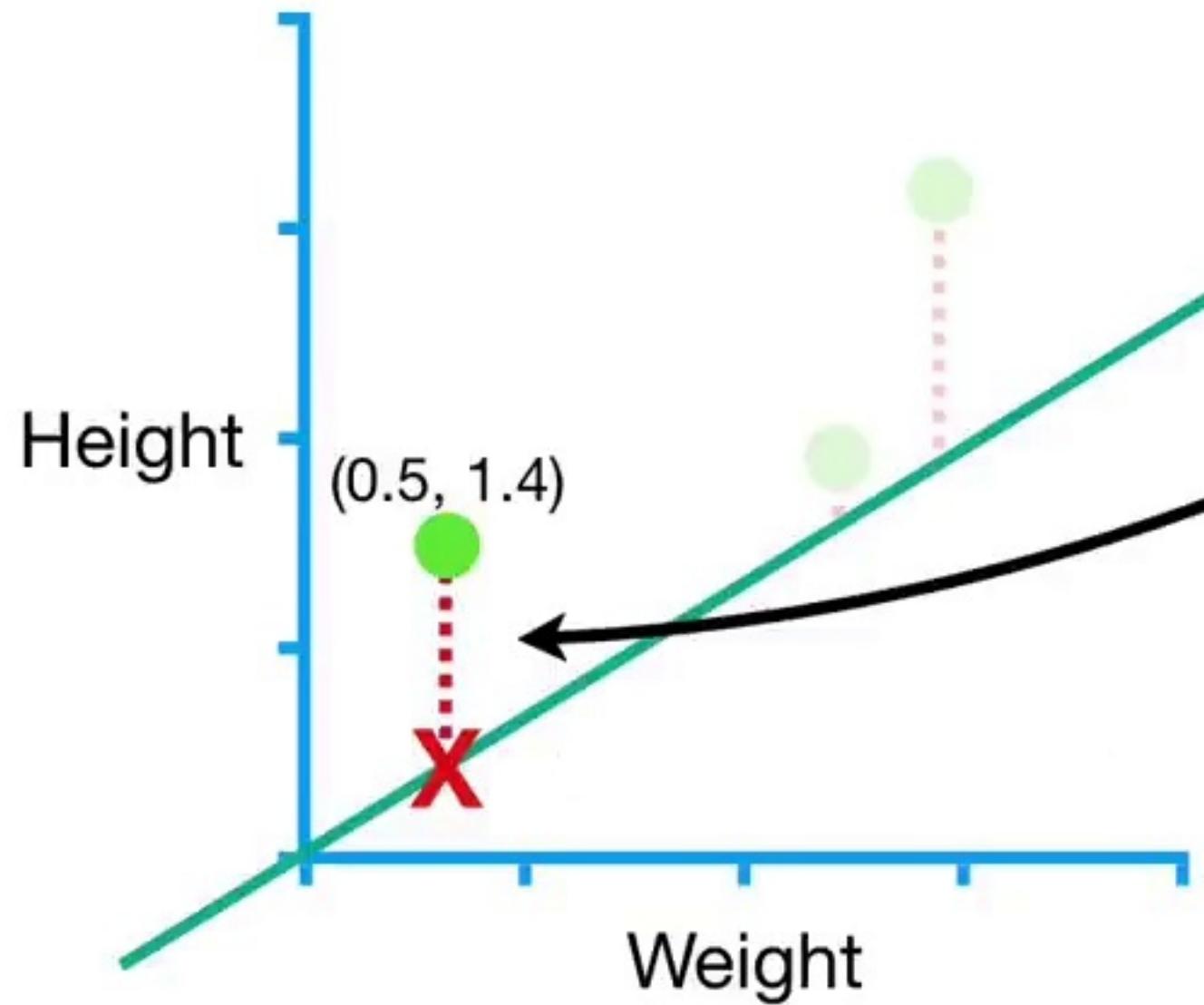
...by plugging
Weight = 0.5 into the equation for the line...

$$\text{Predicted Height} = 0 + 0.64 \times 0.5$$



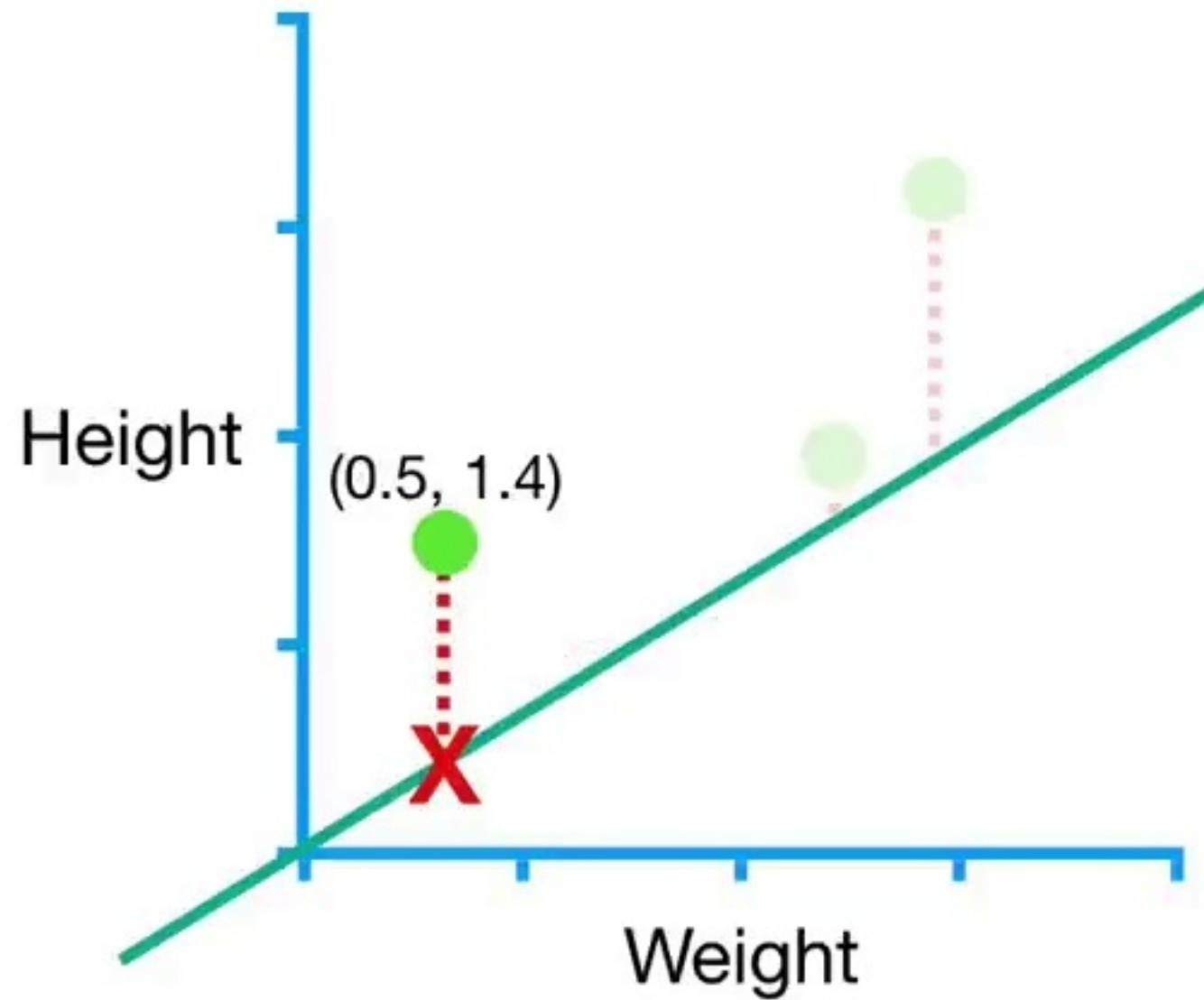
...and the **Predicted Height**
is **0.32**.

$$\text{Predicted Height} = 0 + 0.64 \times 0.5 = 0.32$$



The residual is the difference between the **Observed Height**, and the **Predicted Height**...

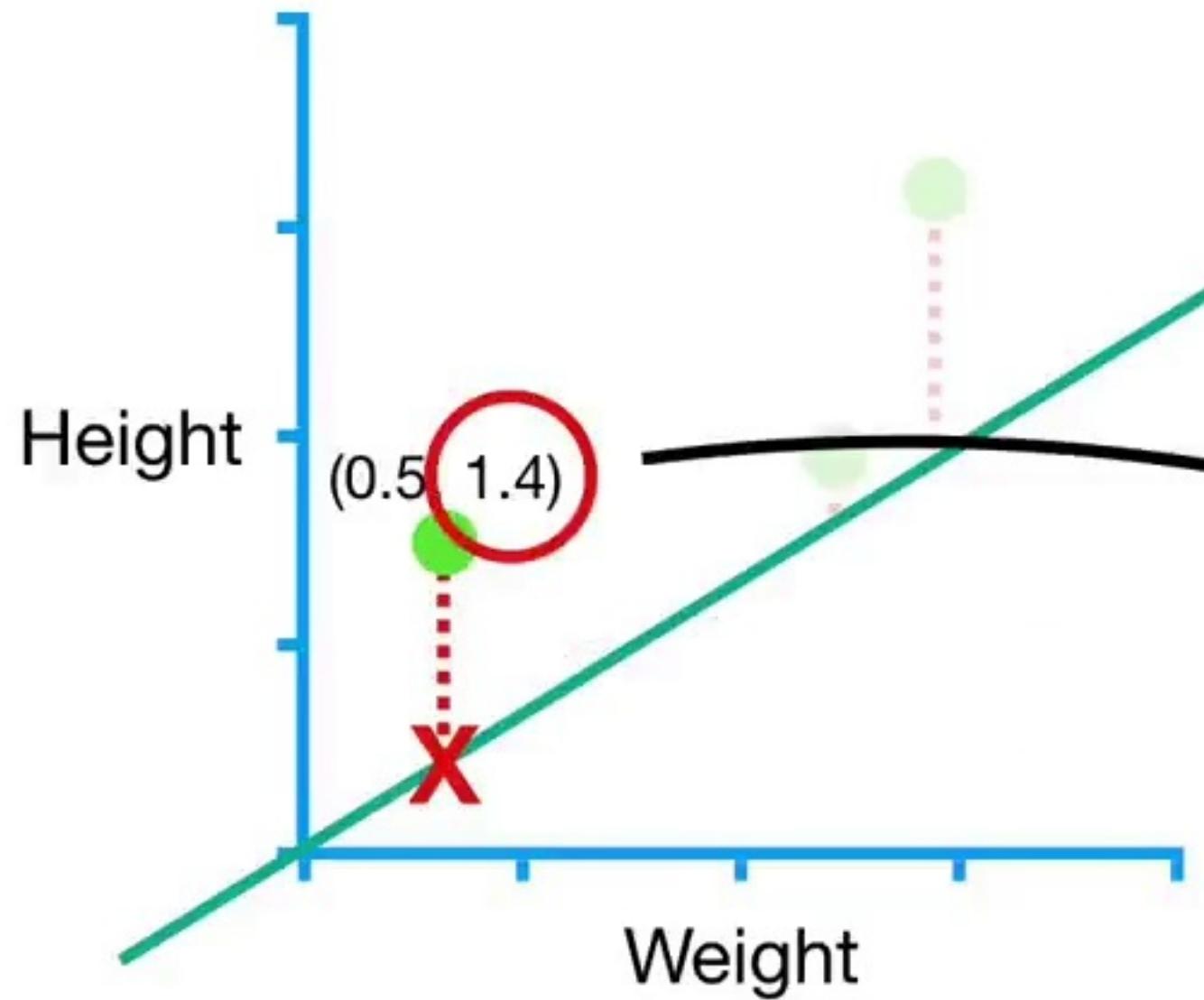
$$\text{Predicted Height} = 0 + 0.64 \times 0.5 = 0.32$$



The residual is the difference
between the **Observed Height**,
and the **Predicted Height**...

$$\text{Residual} = \text{Observed Height} - \text{Predicted Height}$$

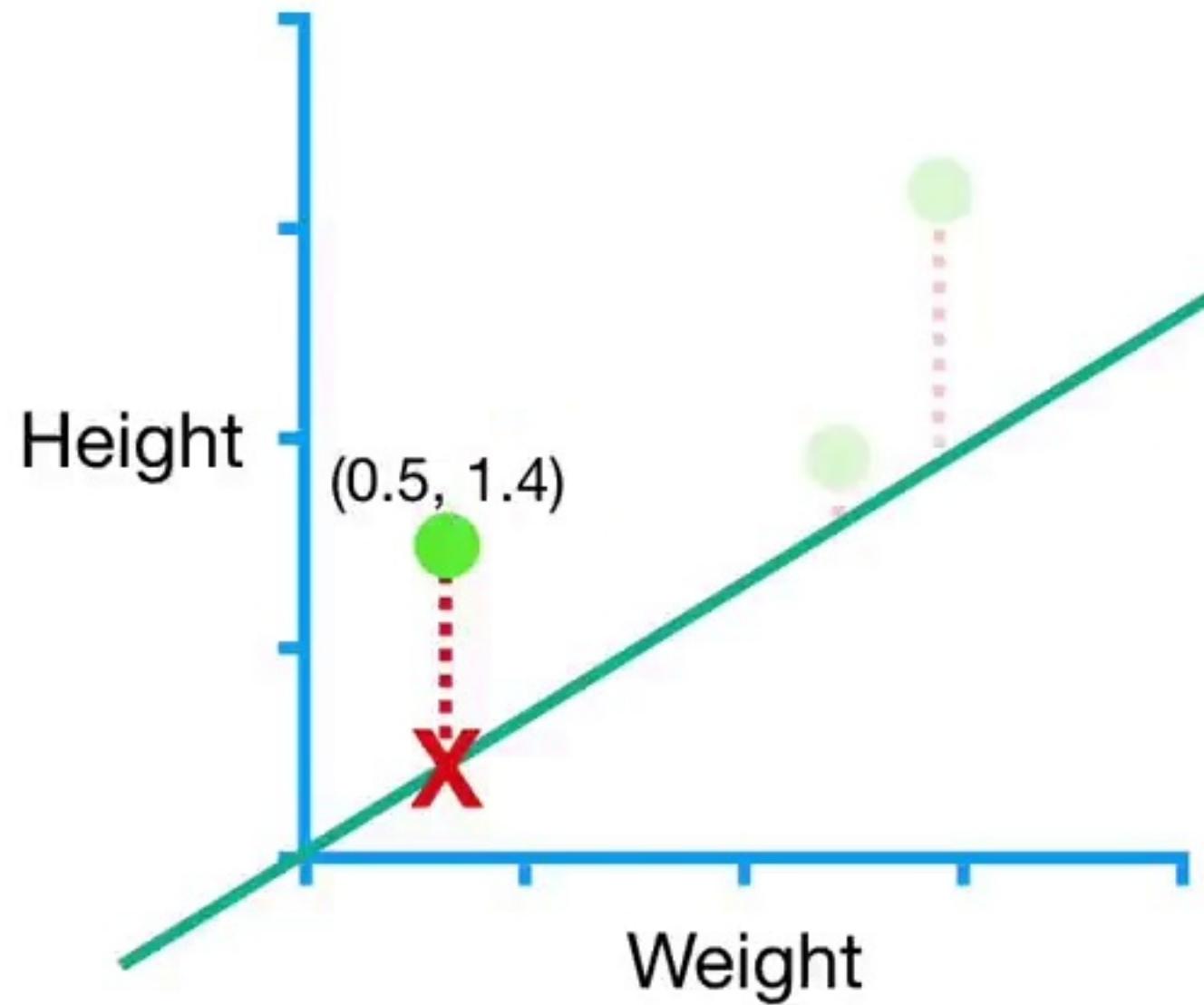
$$\text{Predicted Height} = 0 + 0.64 \times 0.5 = 0.32$$



...so we calculate the difference
between **1.4** (the **Observed Height**)...

Residual = 1.4

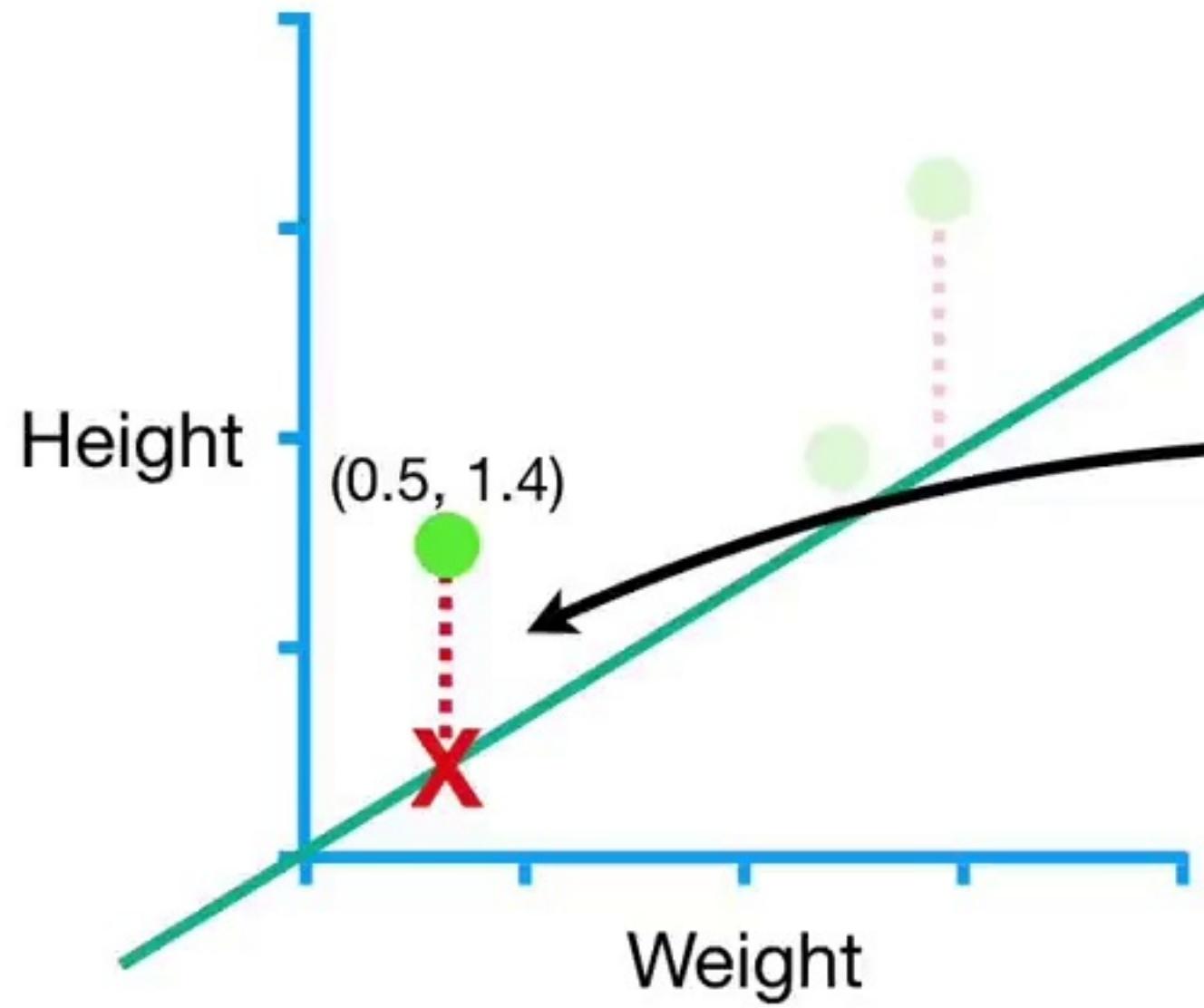
$$\text{Predicted Height} = 0 + 0.64 \times 0.5 = 0.32$$



...so we calculate the difference
between **1.4** (the **Observed Height**)...
...and **0.32** (the **Predicted Height**)...

$$\text{Residual} = 1.4 - 0.32$$

$$\text{Predicted Height} = 0 + 0.64 \times 0.5 = \boxed{0.32}$$



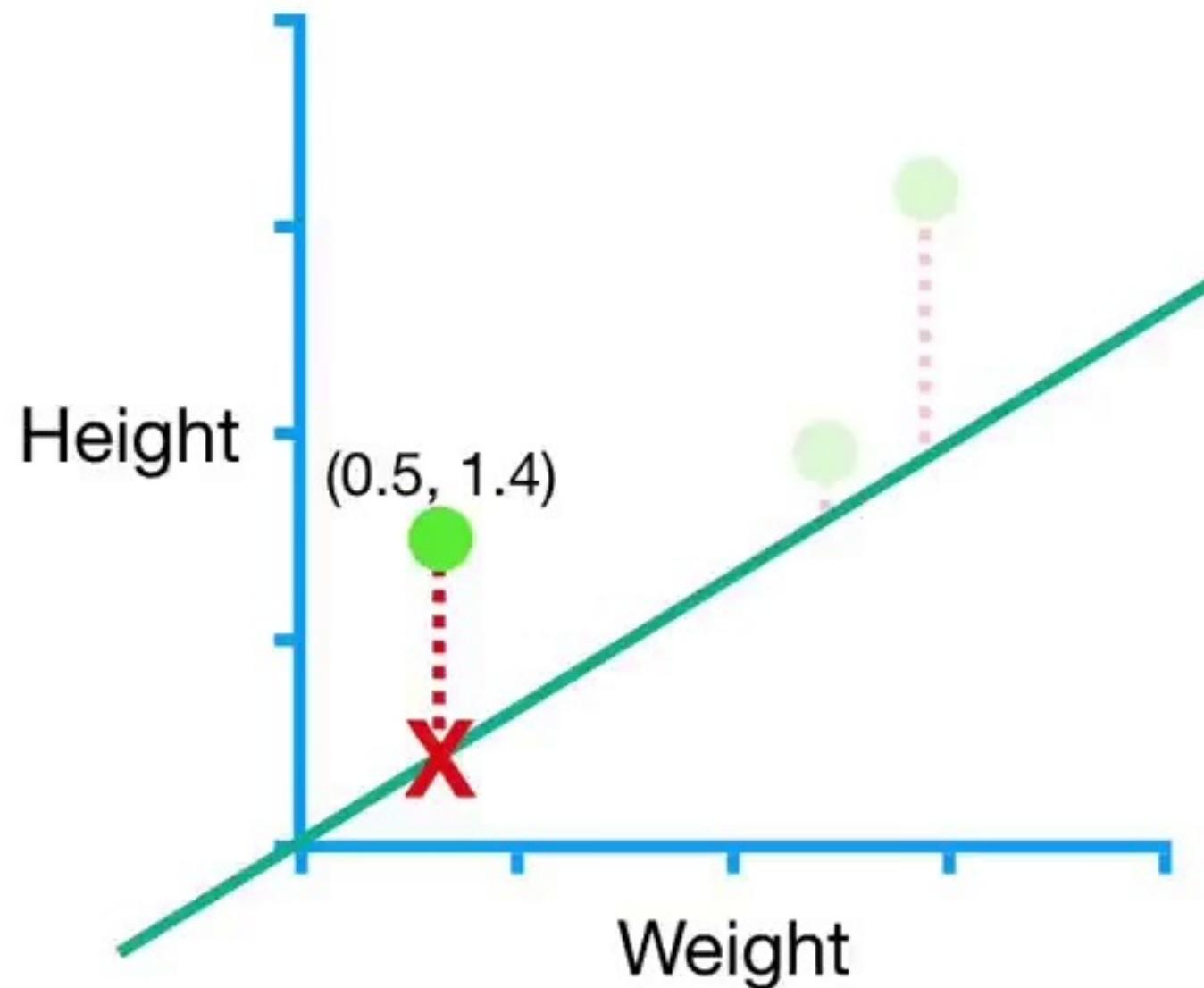
...and that gives us 1.1
for the residual.

$$\text{Residual} = 1.4 - 0.32 = 1.1$$

$$\text{Predicted Height} = 0 + 0.64 \times 0.5 = 0.32$$

Sum of squared residuals =

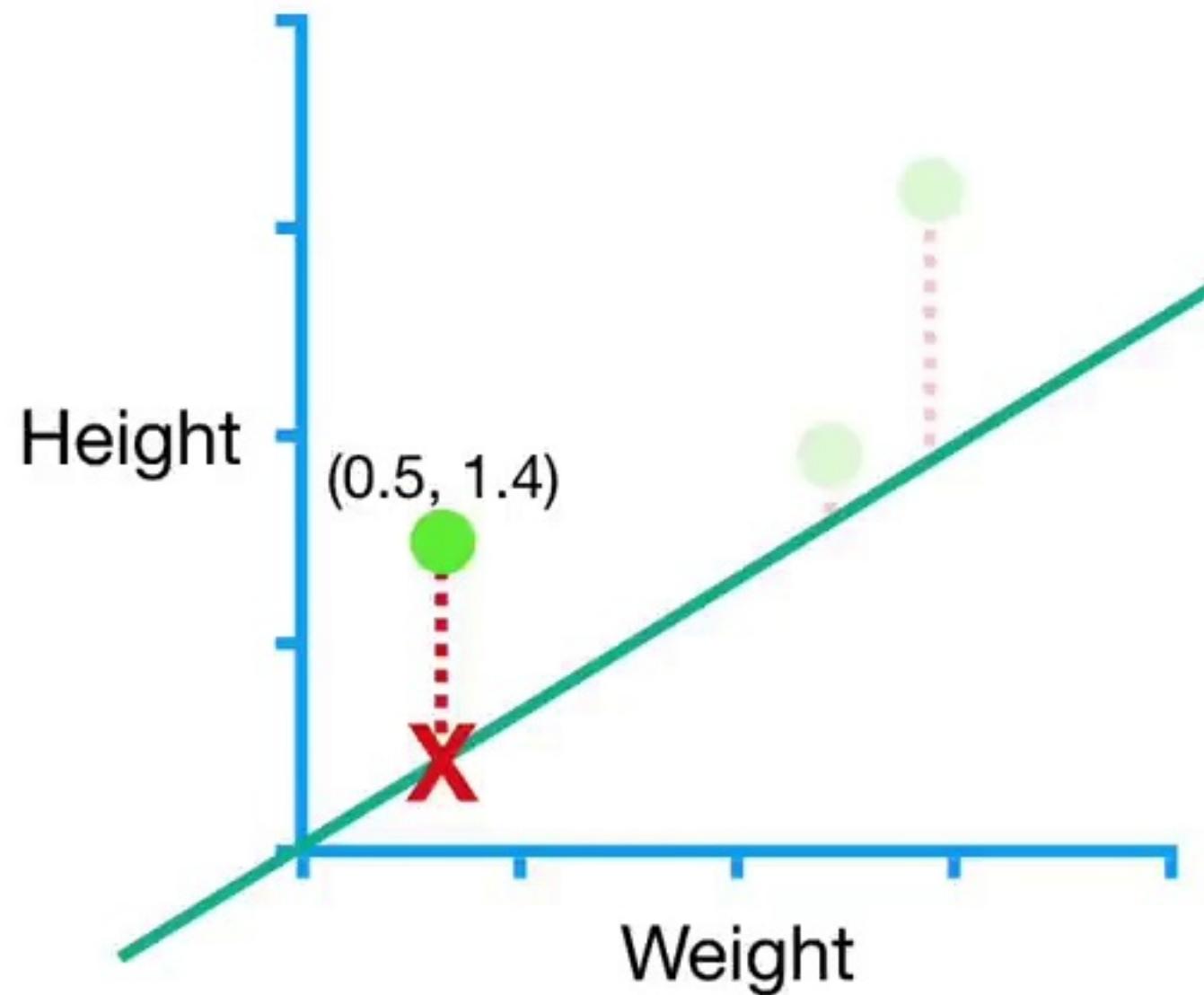
We'll keep track of the Sum of the Squared Residuals up here.



$$\text{Residual} = 1.4 - 0.32 = 1.1$$

$$\text{Predicted Height} = 0 + 0.64 \times 0.5 = 0.32$$

Sum of squared residuals = 1.1^2

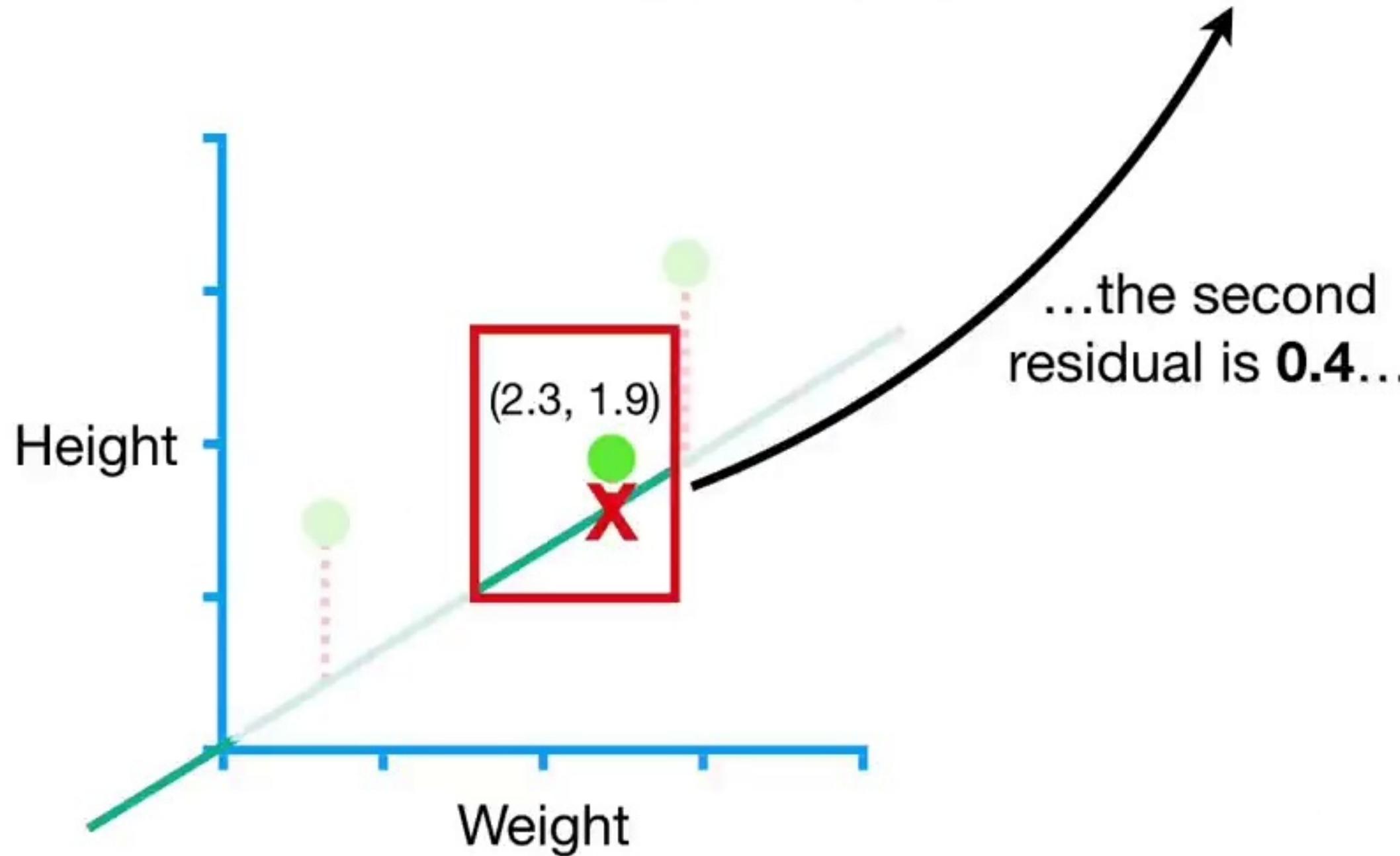


Residual = $1.4 - 0.32 = \boxed{1.1}$

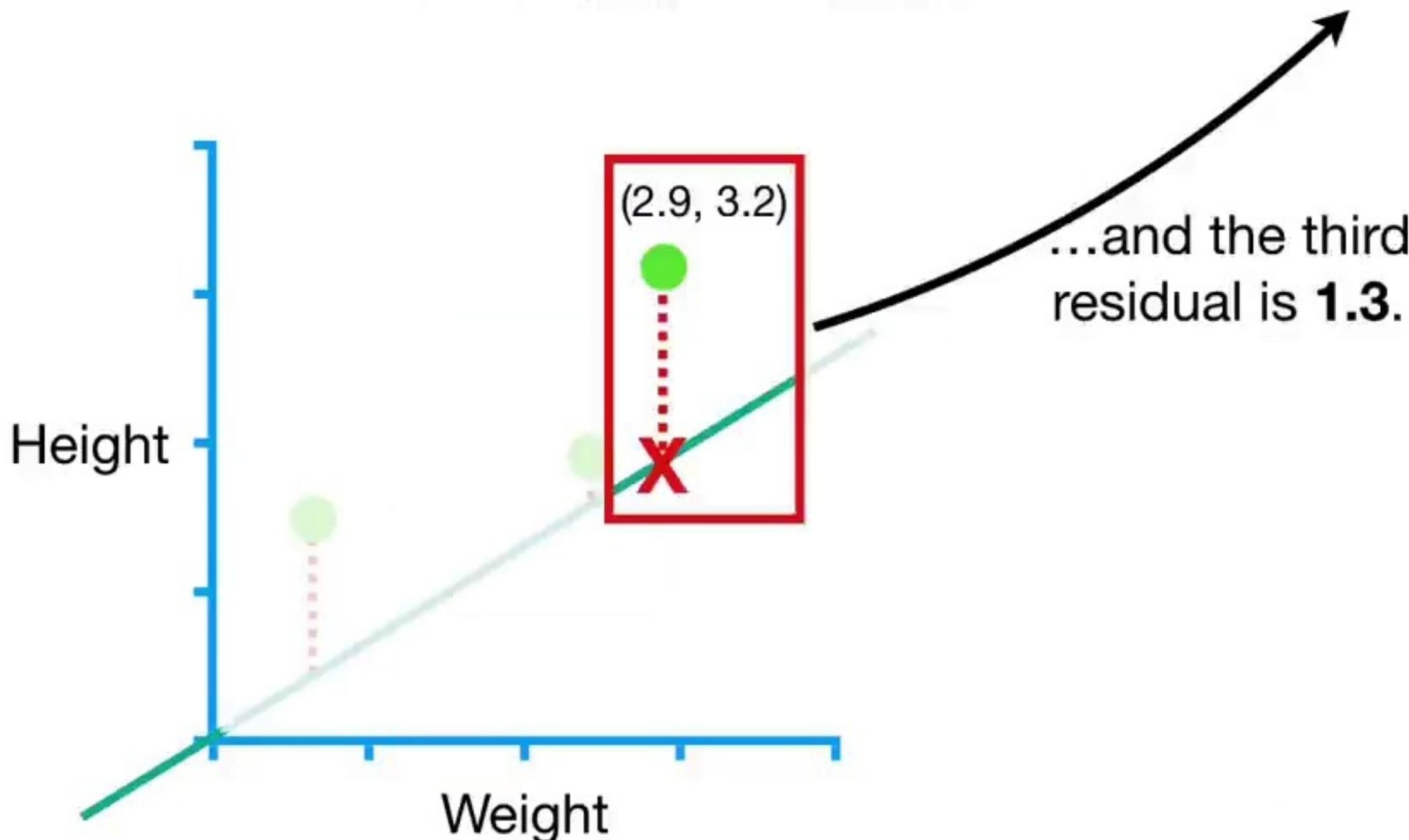
Here's the square of
the first residual...

Predicted Height = $0 + 0.64 \times 0.5 = 0.32$

$$\text{Sum of squared residuals} = 1.1^2 + 0.4^2$$

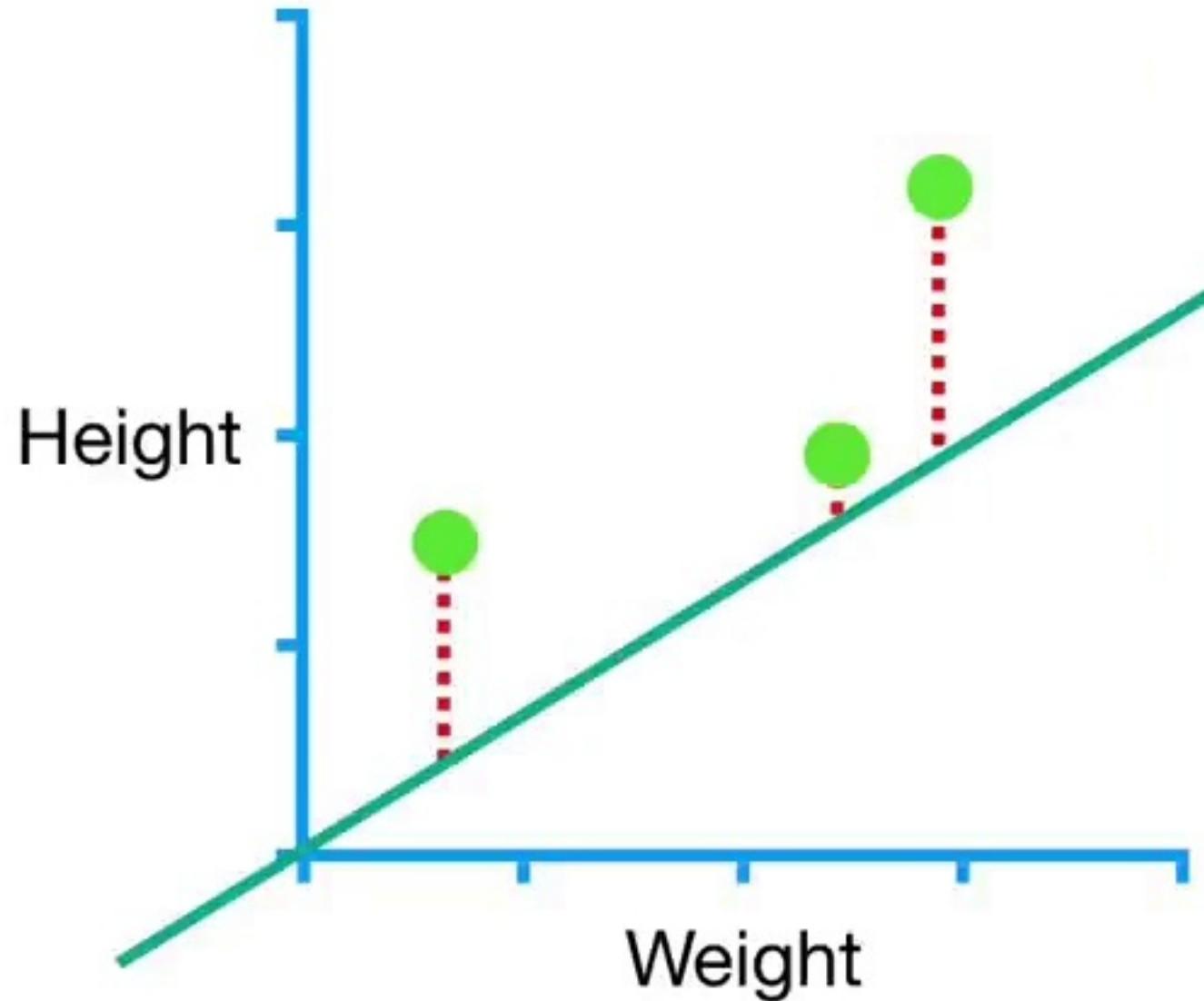


Sum of squared residuals = $1.1^2 + 0.4^2 + 1.3^2$



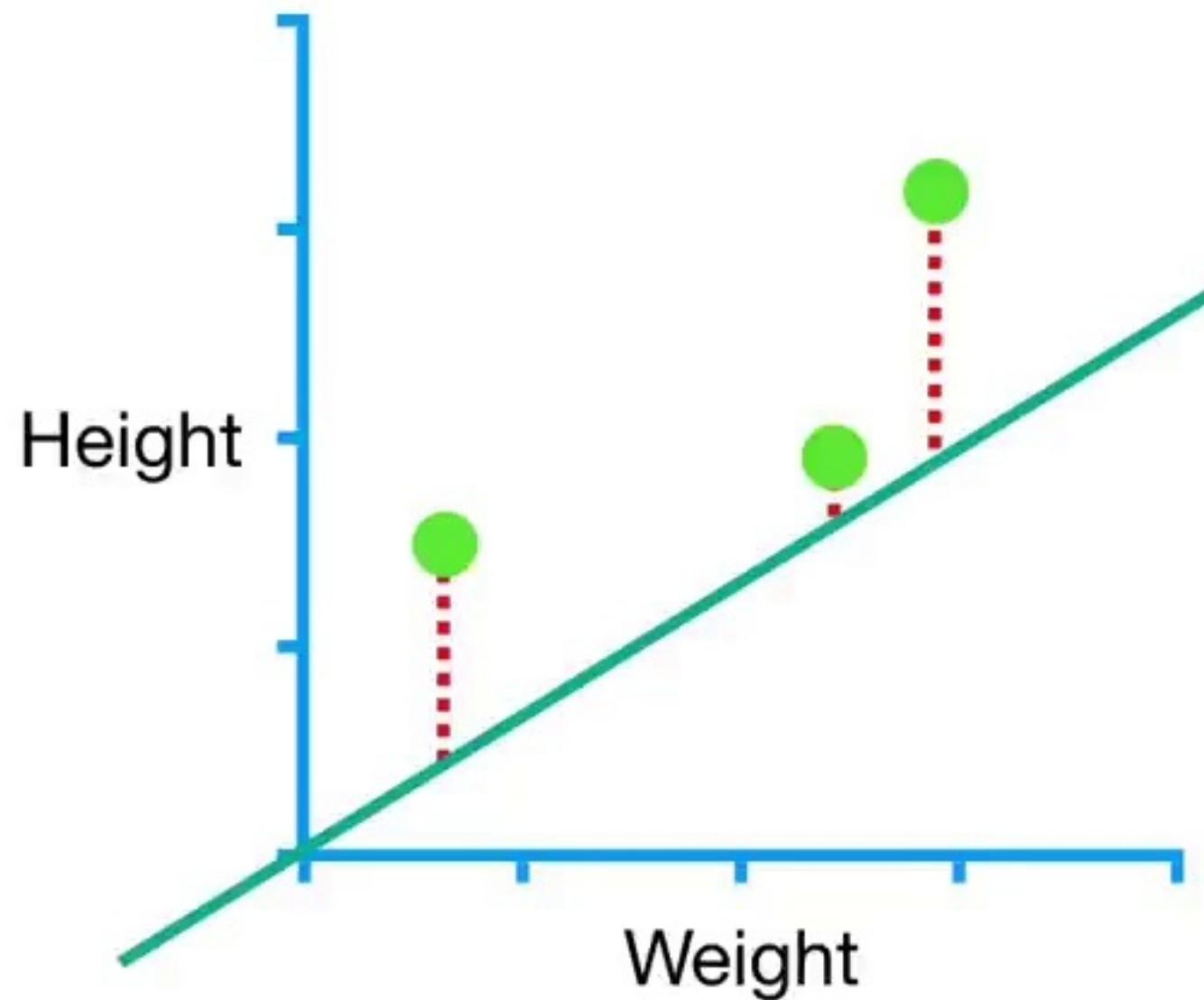
...and the third
residual is **1.3**.

$$\text{Sum of squared residuals} = 1.1^2 + 0.4^2 + 1.3^2 = \boxed{3.1}$$

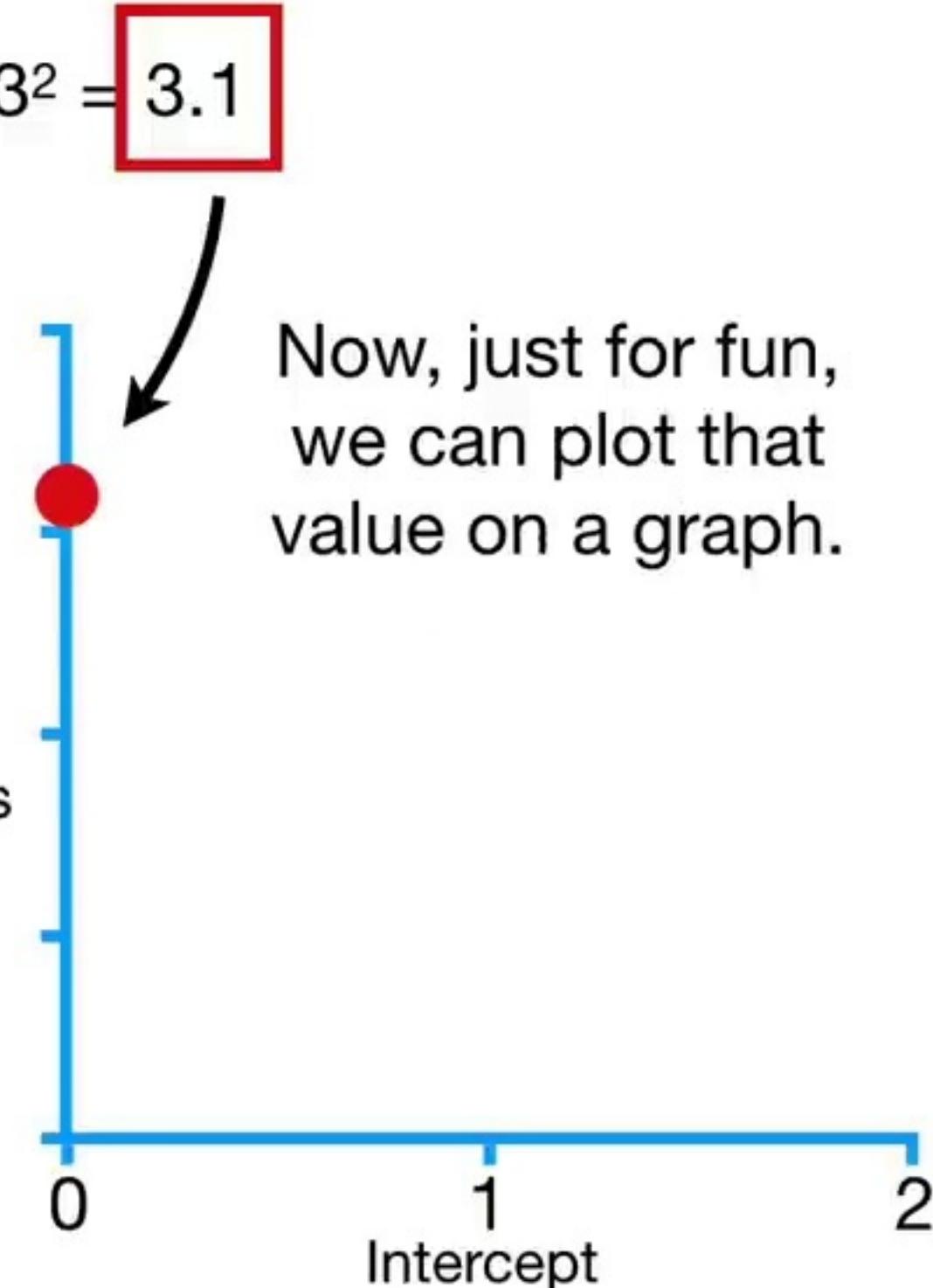


In the end, **3.1** is the Sum of the Squared Residuals.

$$\text{Sum of squared residuals} = 1.1^2 + 0.4^2 + 1.3^2 = 3.1$$

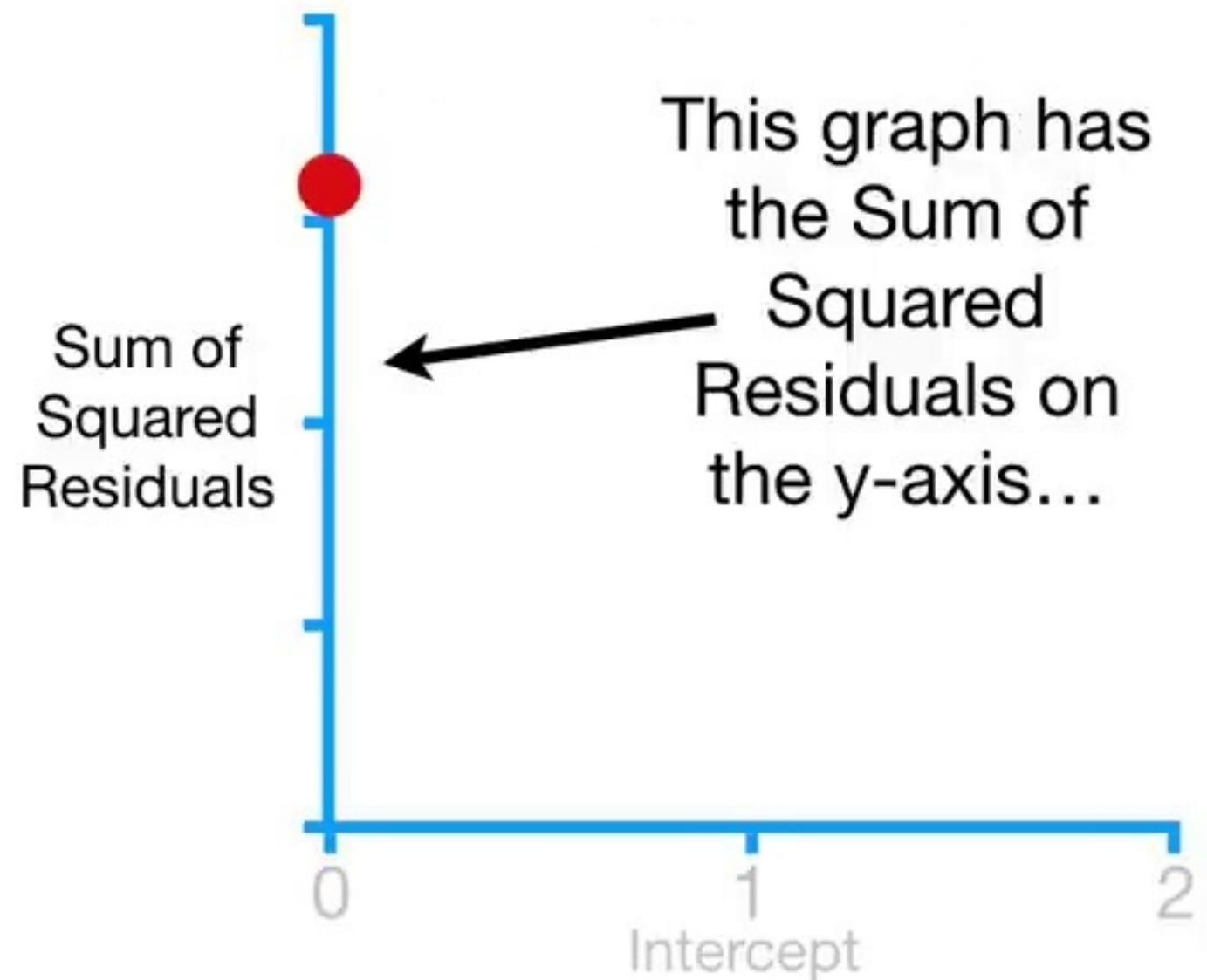
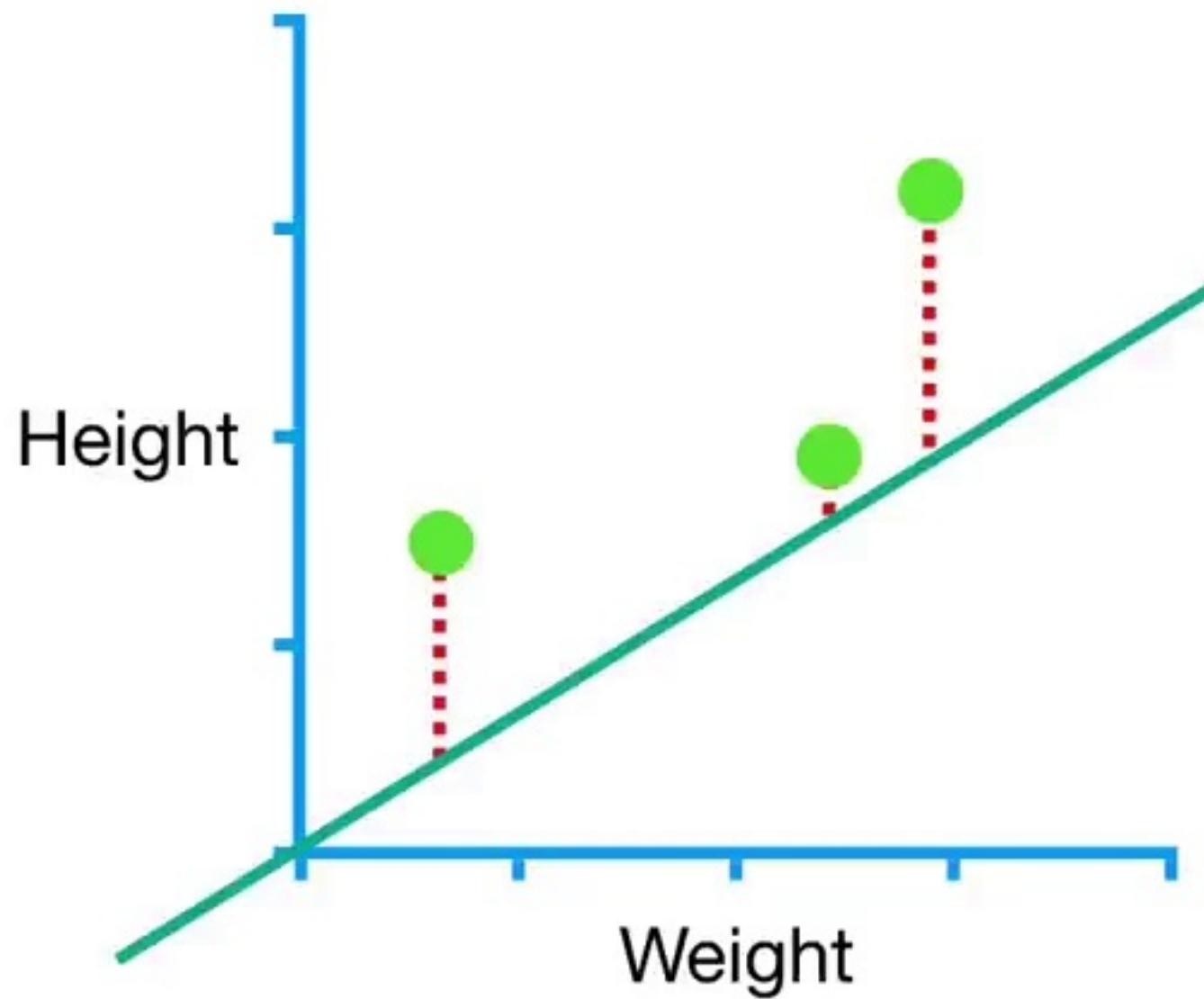


Sum of
Squared
Residuals

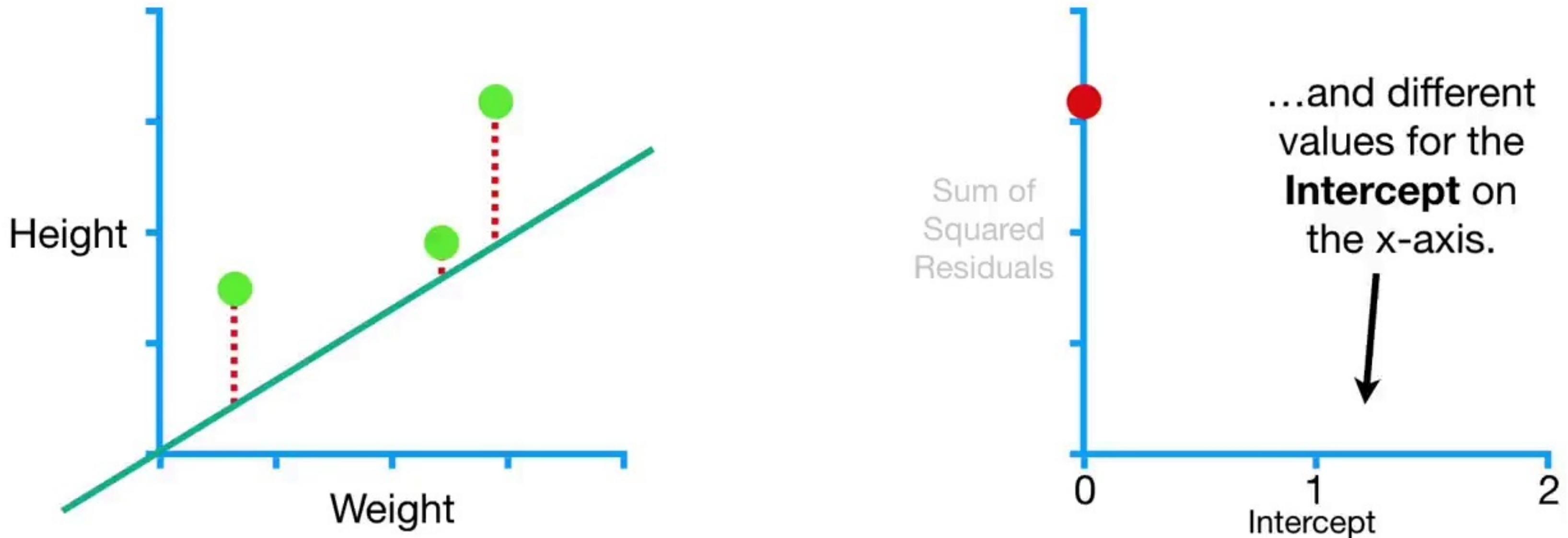


Now, just for fun,
we can plot that
value on a graph.

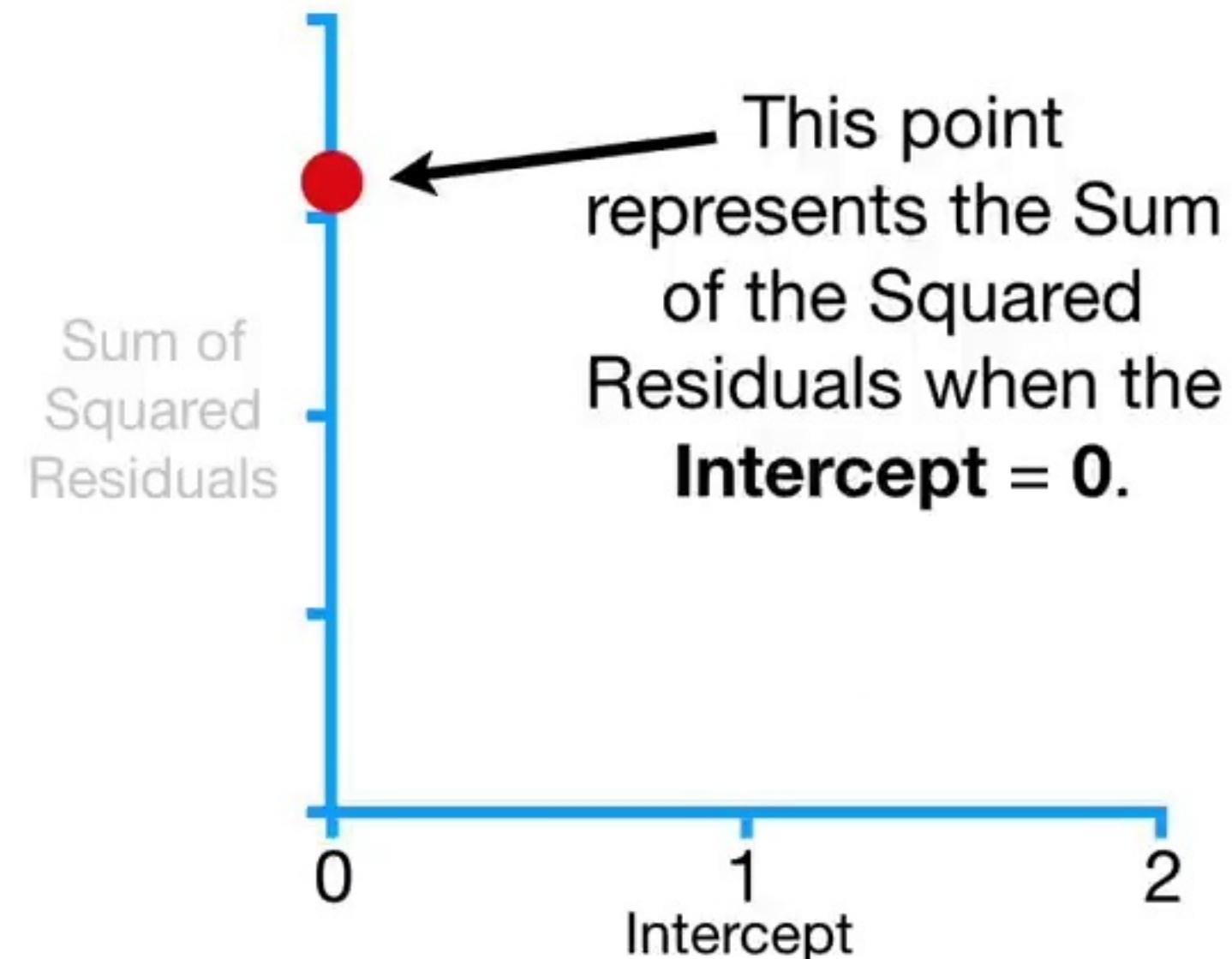
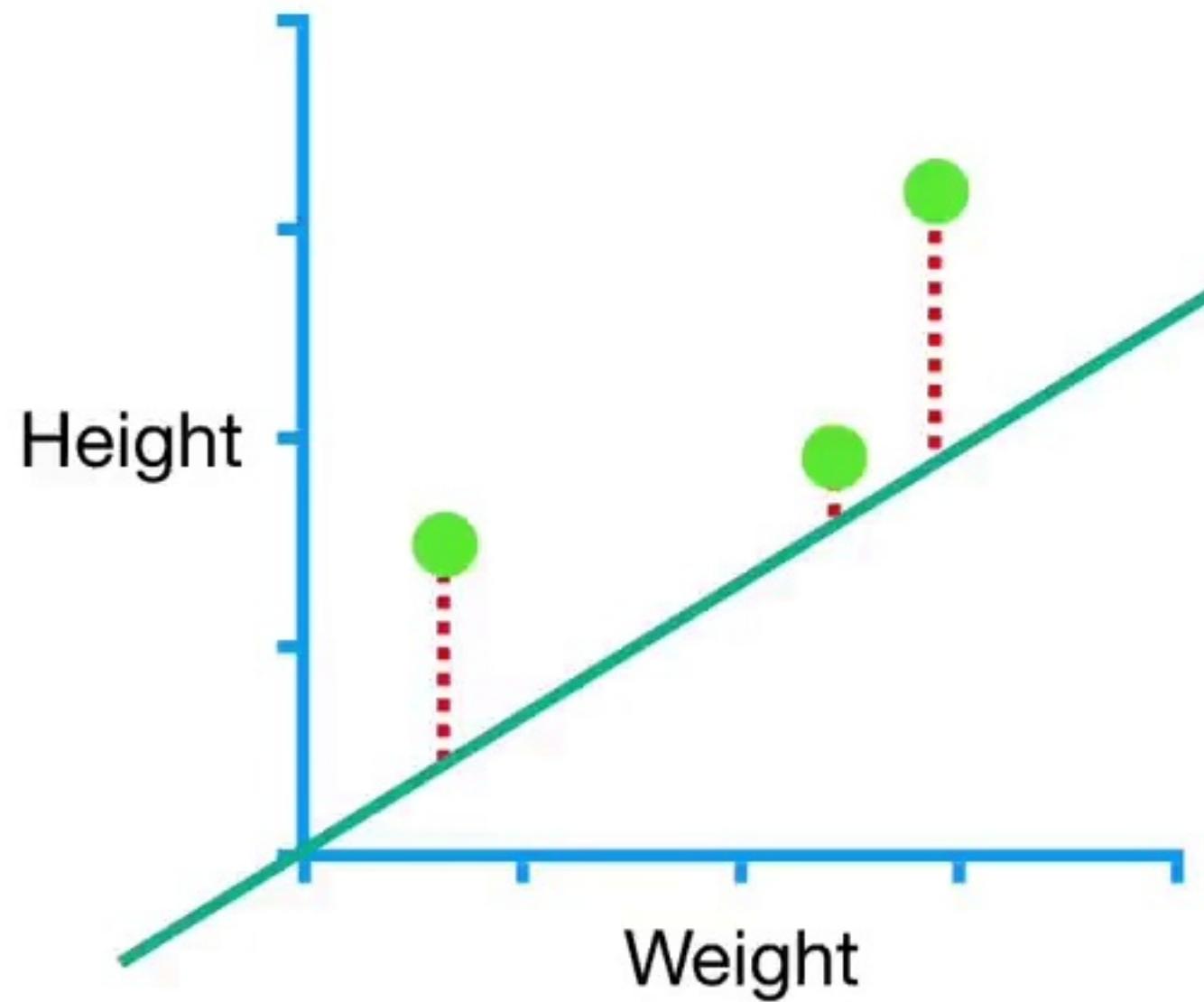
$$\text{Sum of squared residuals} = 1.1^2 + 0.4^2 + 1.3^2 = 3.1$$



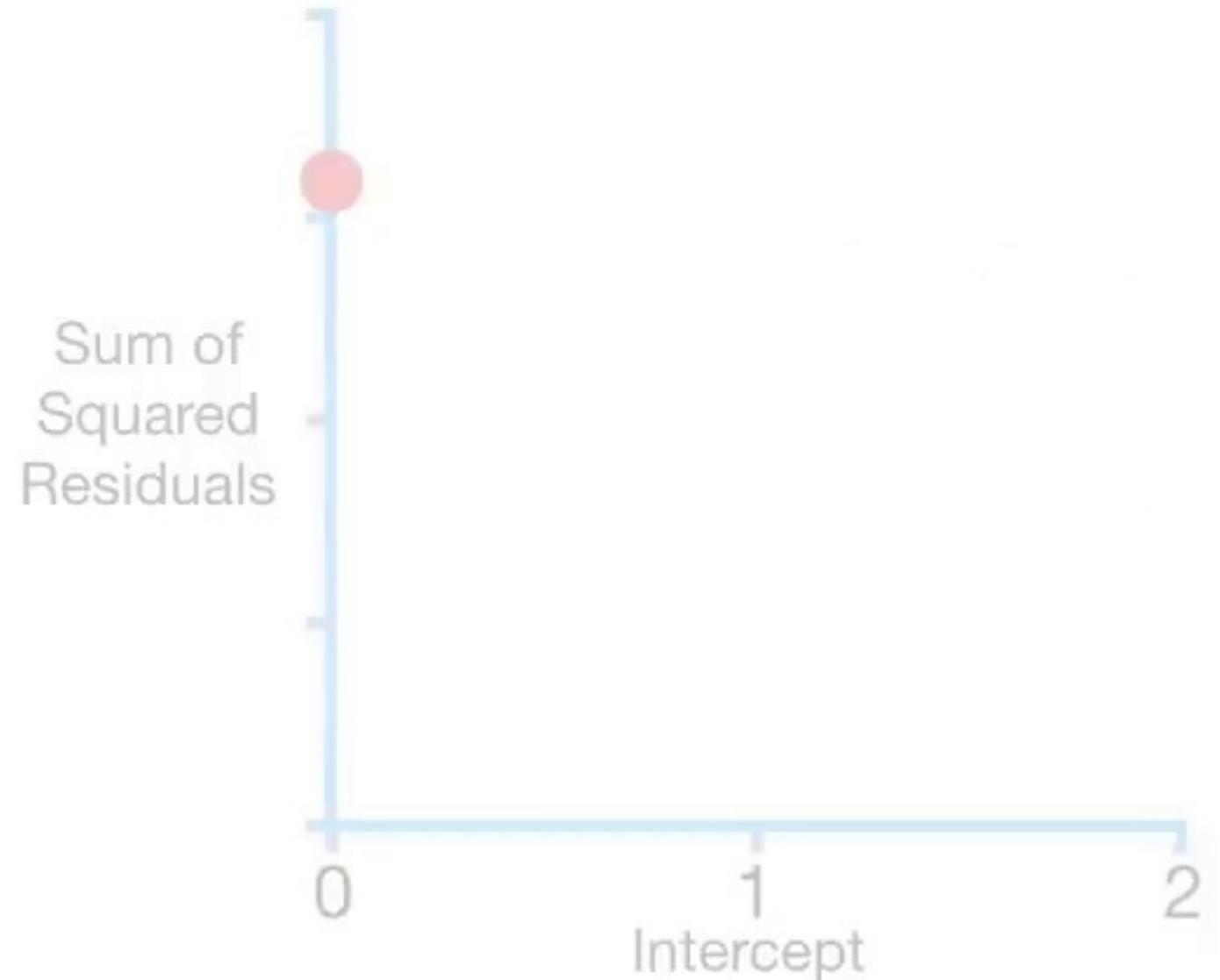
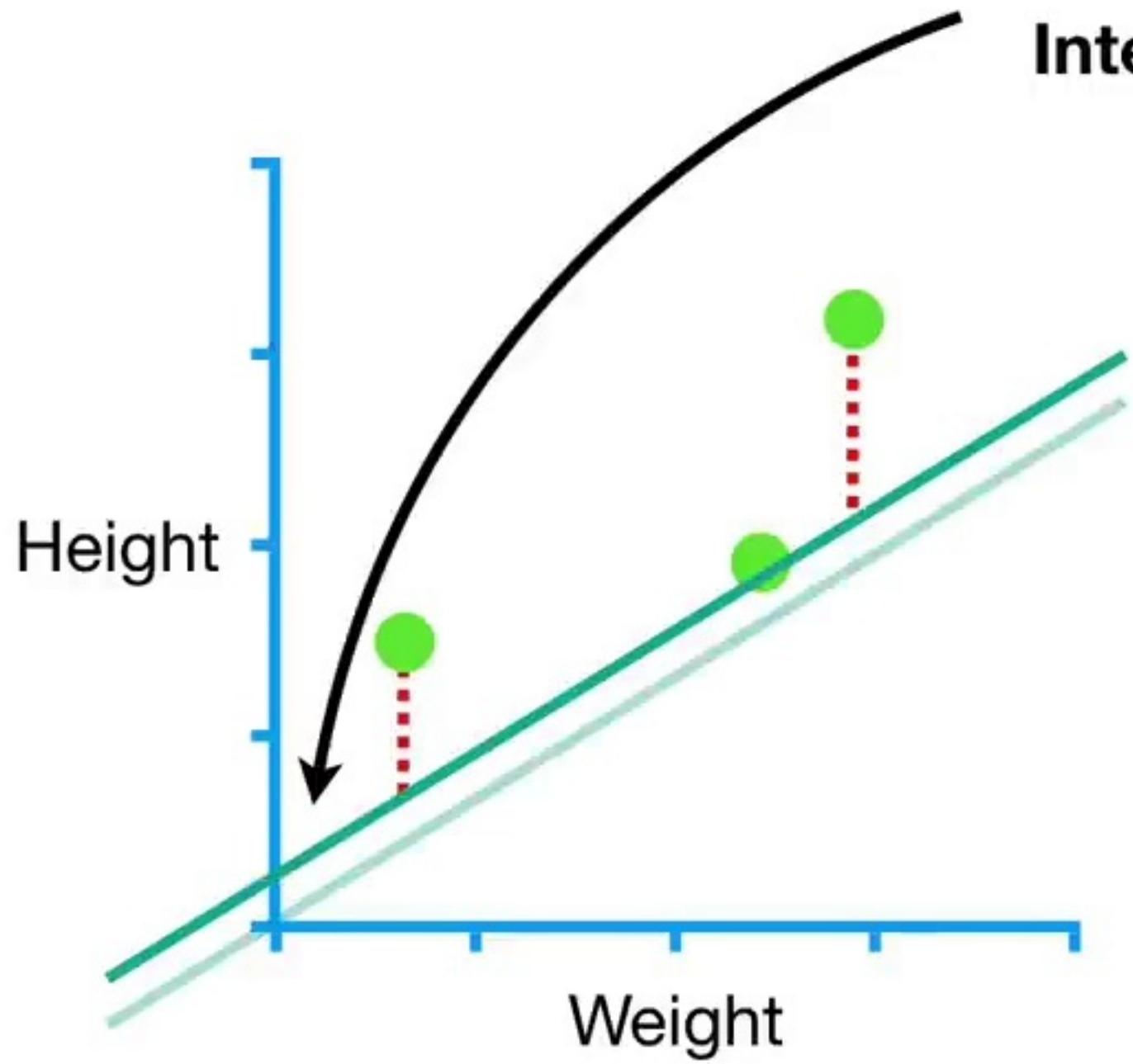
$$\text{Sum of squared residuals} = 1.1^2 + 0.4^2 + 1.3^2 = 3.1$$



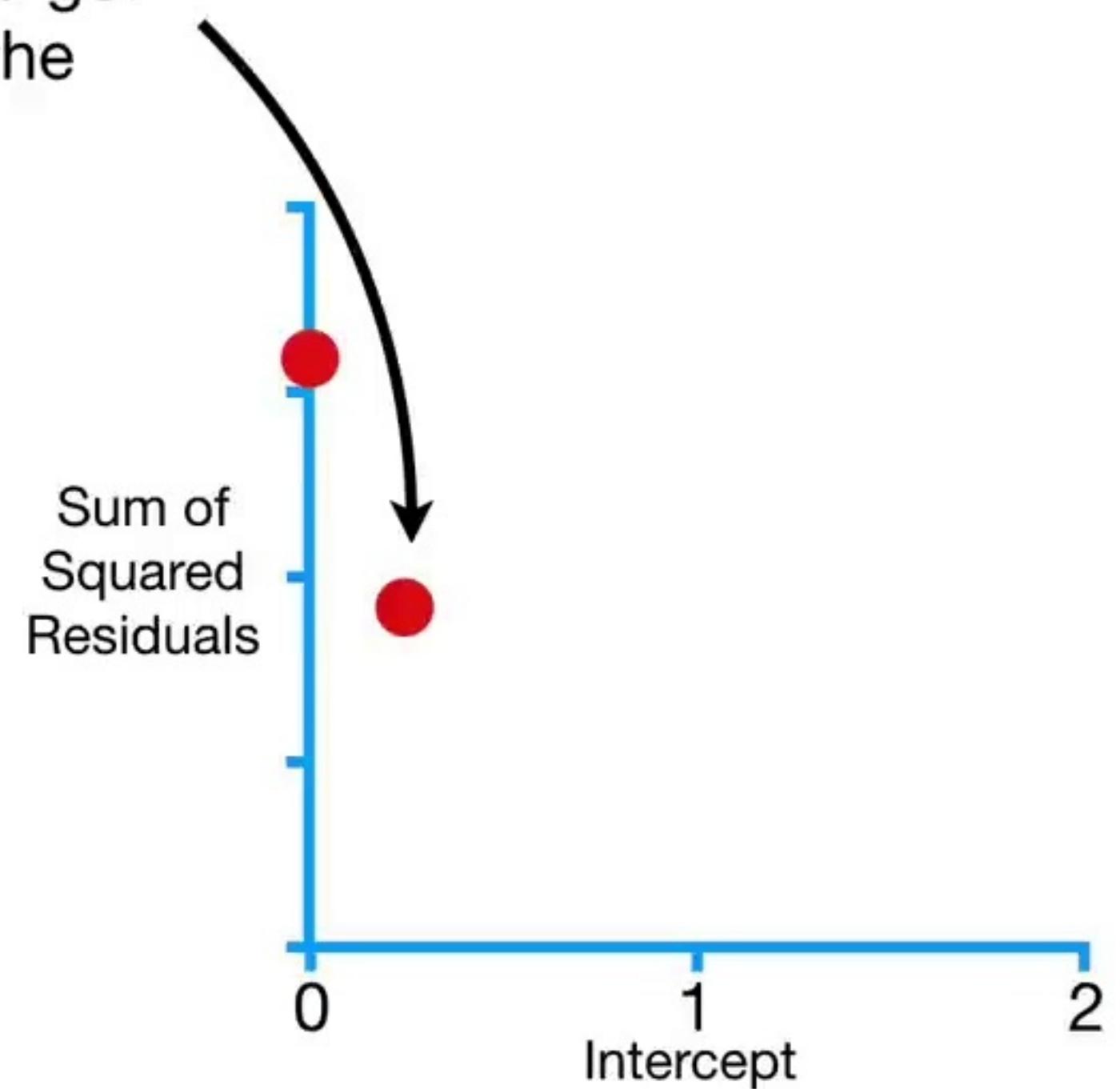
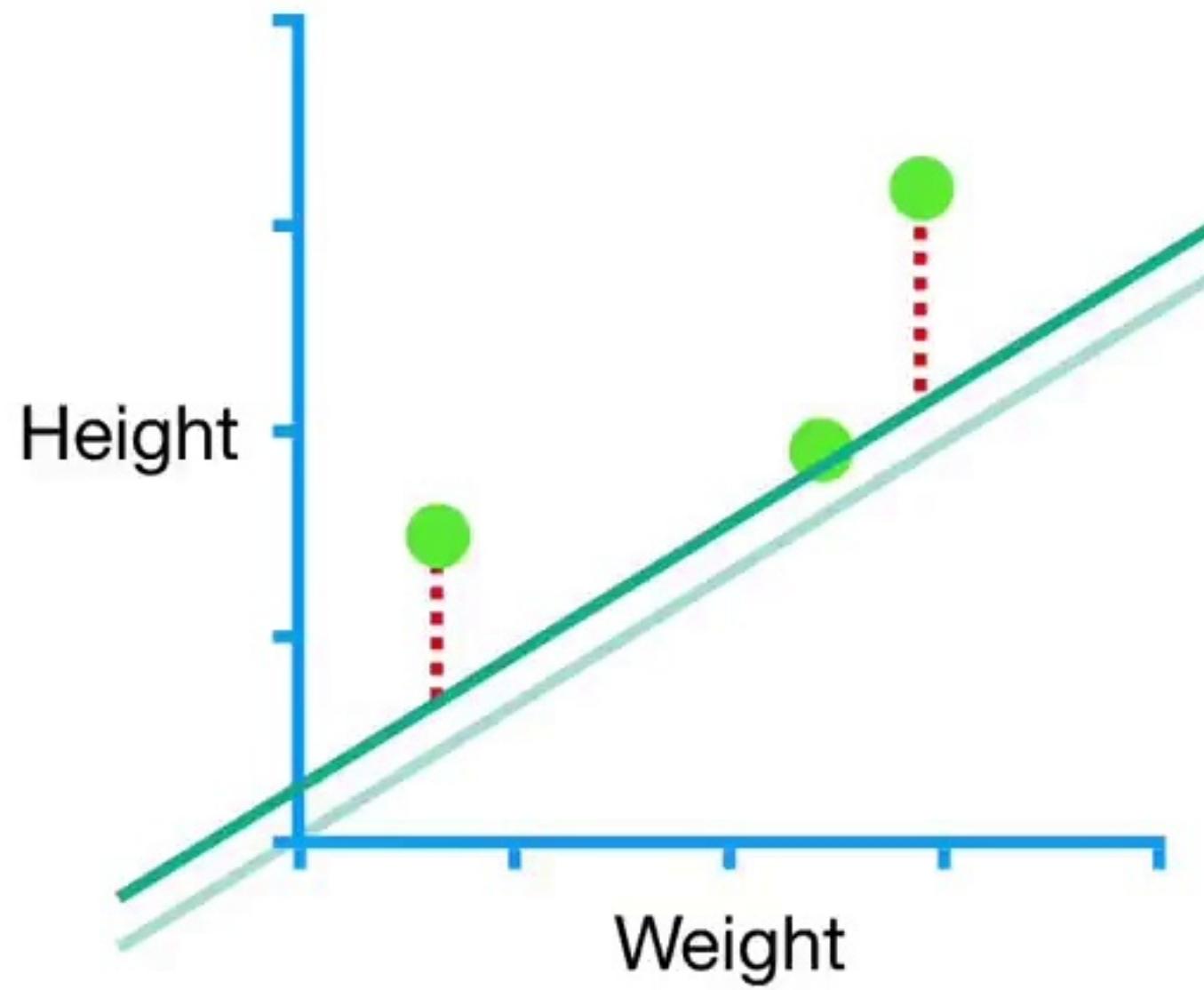
$$\text{Sum of squared residuals} = 1.1^2 + 0.4^2 + 1.3^2 = 3.1$$



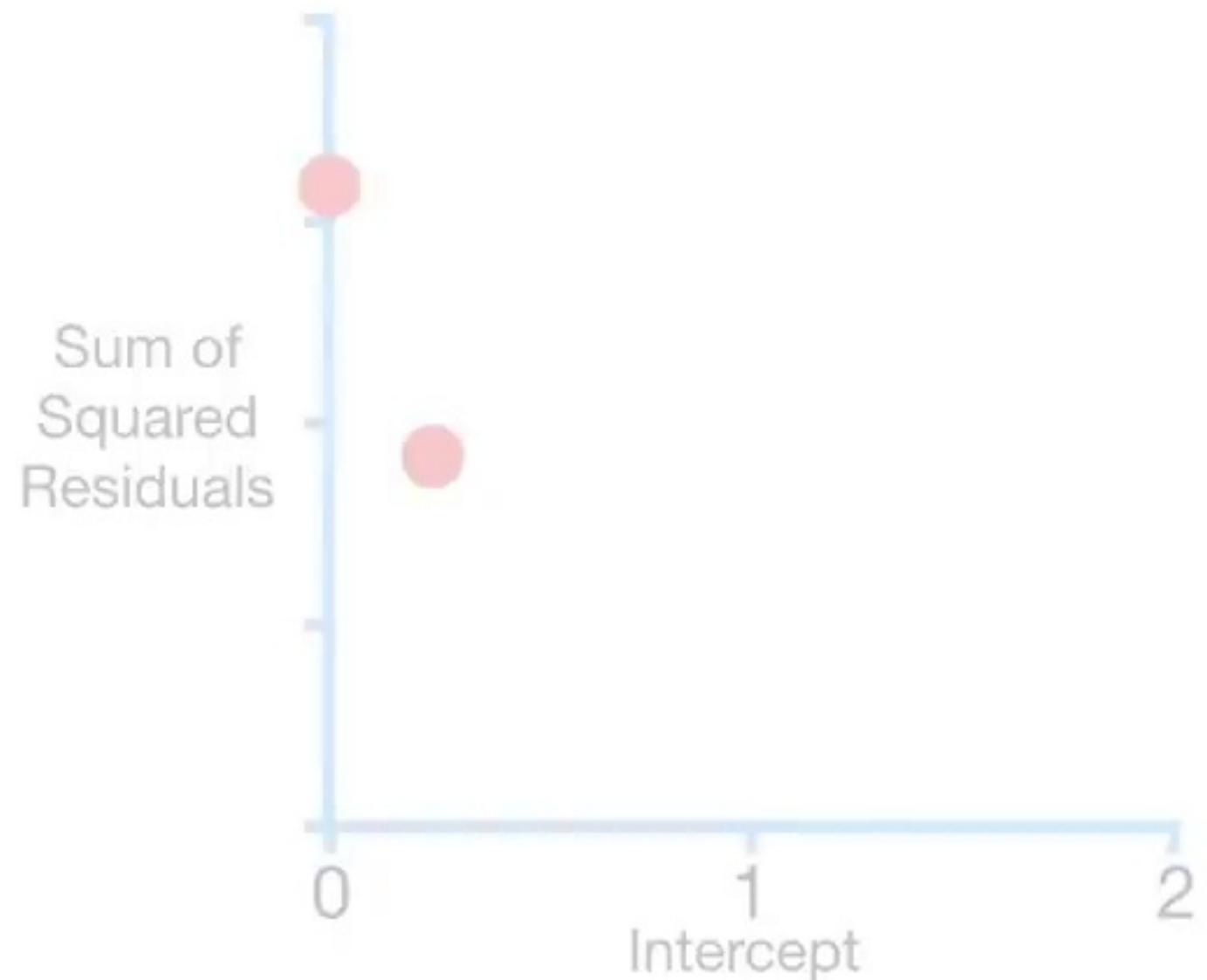
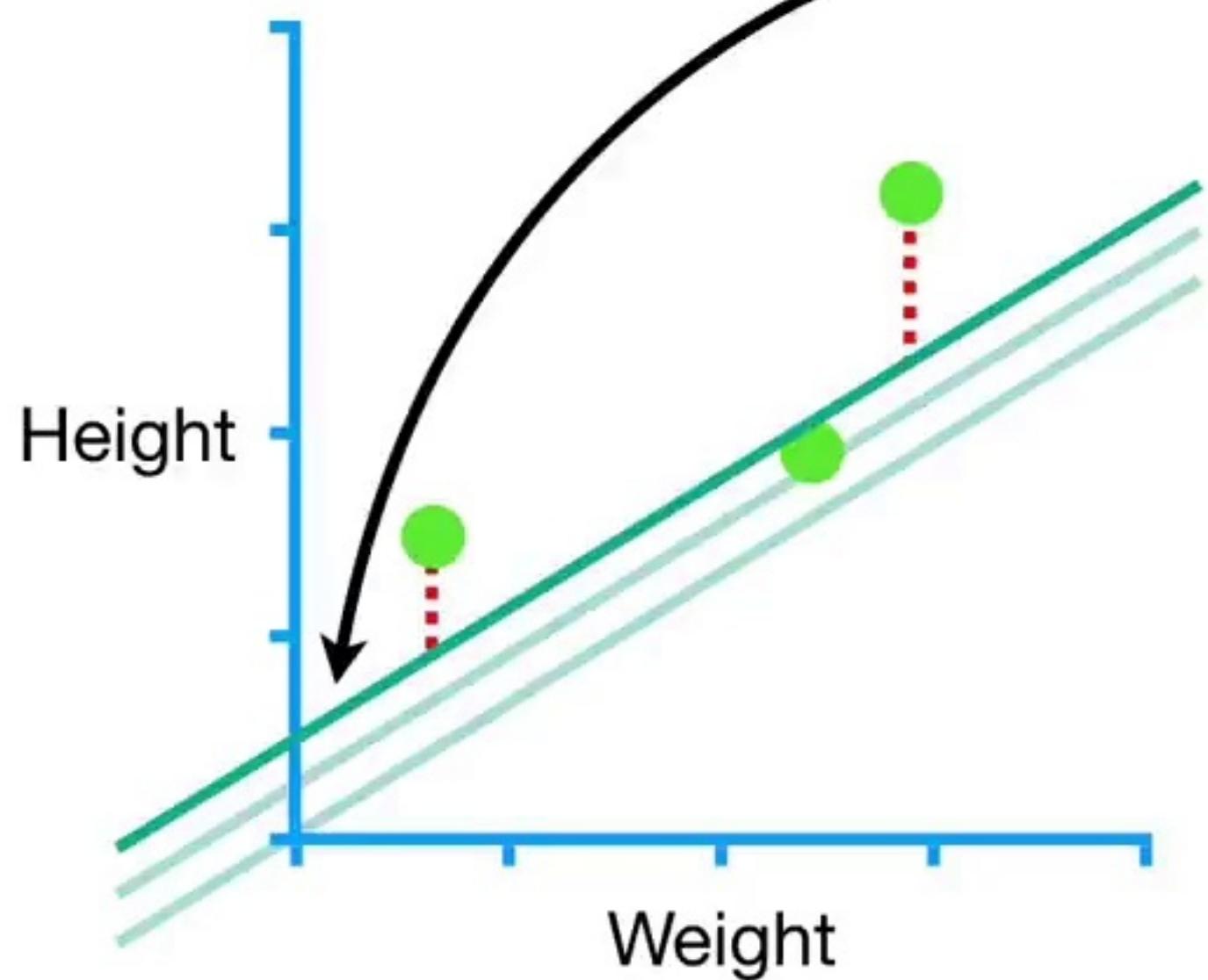
However, if the
Intercept = 0.25...



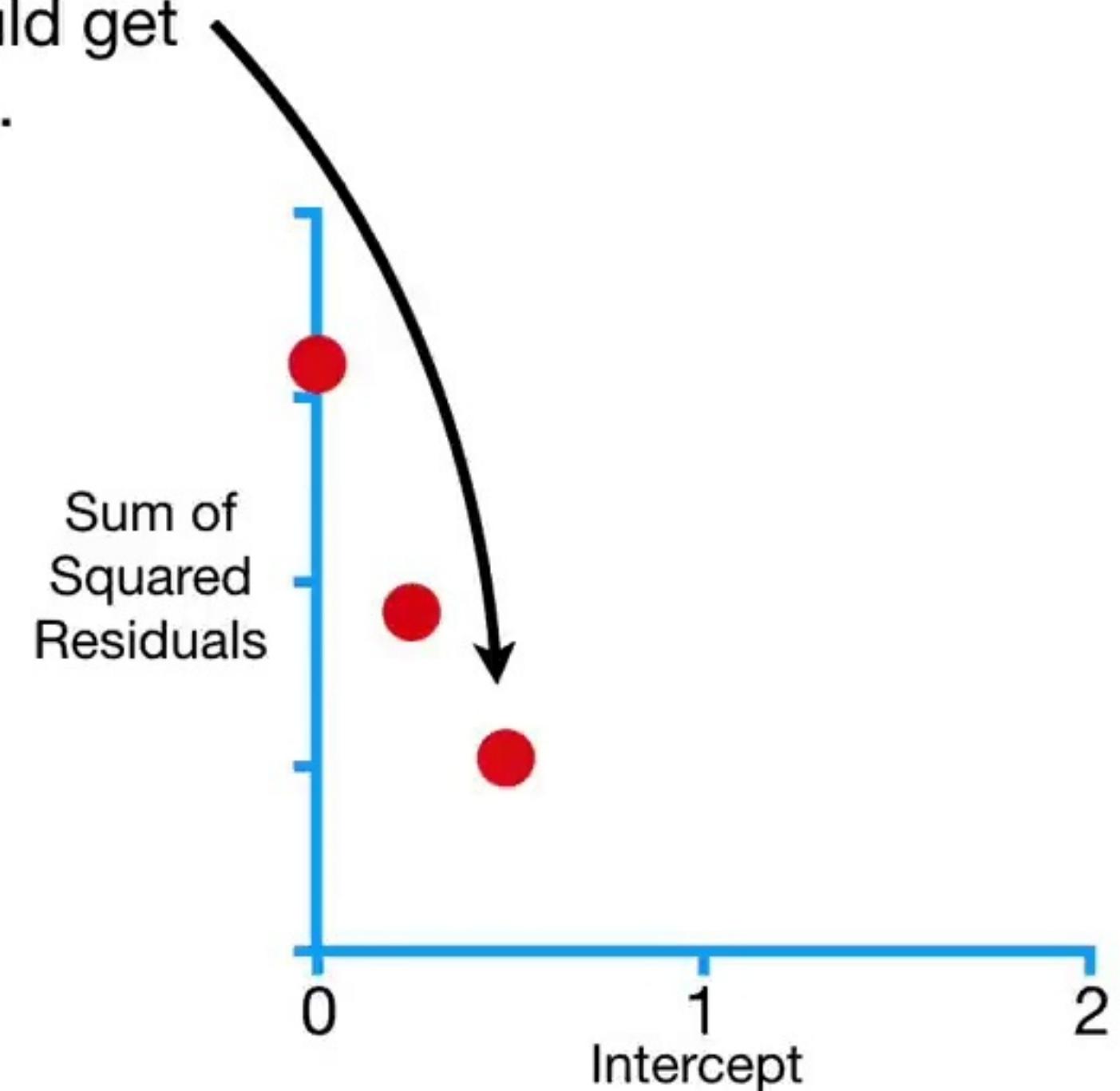
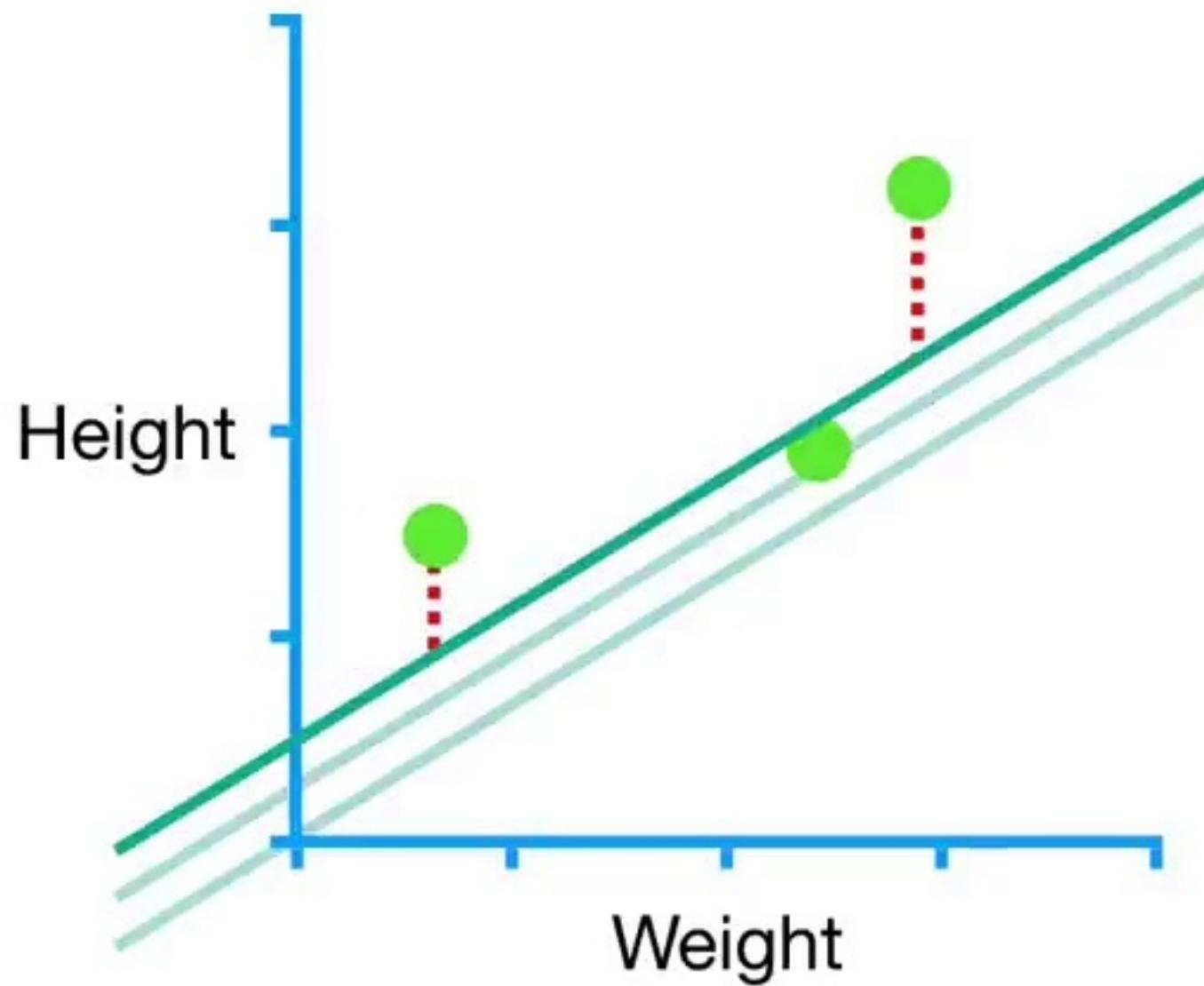
...then we would get
this point on the
graph.



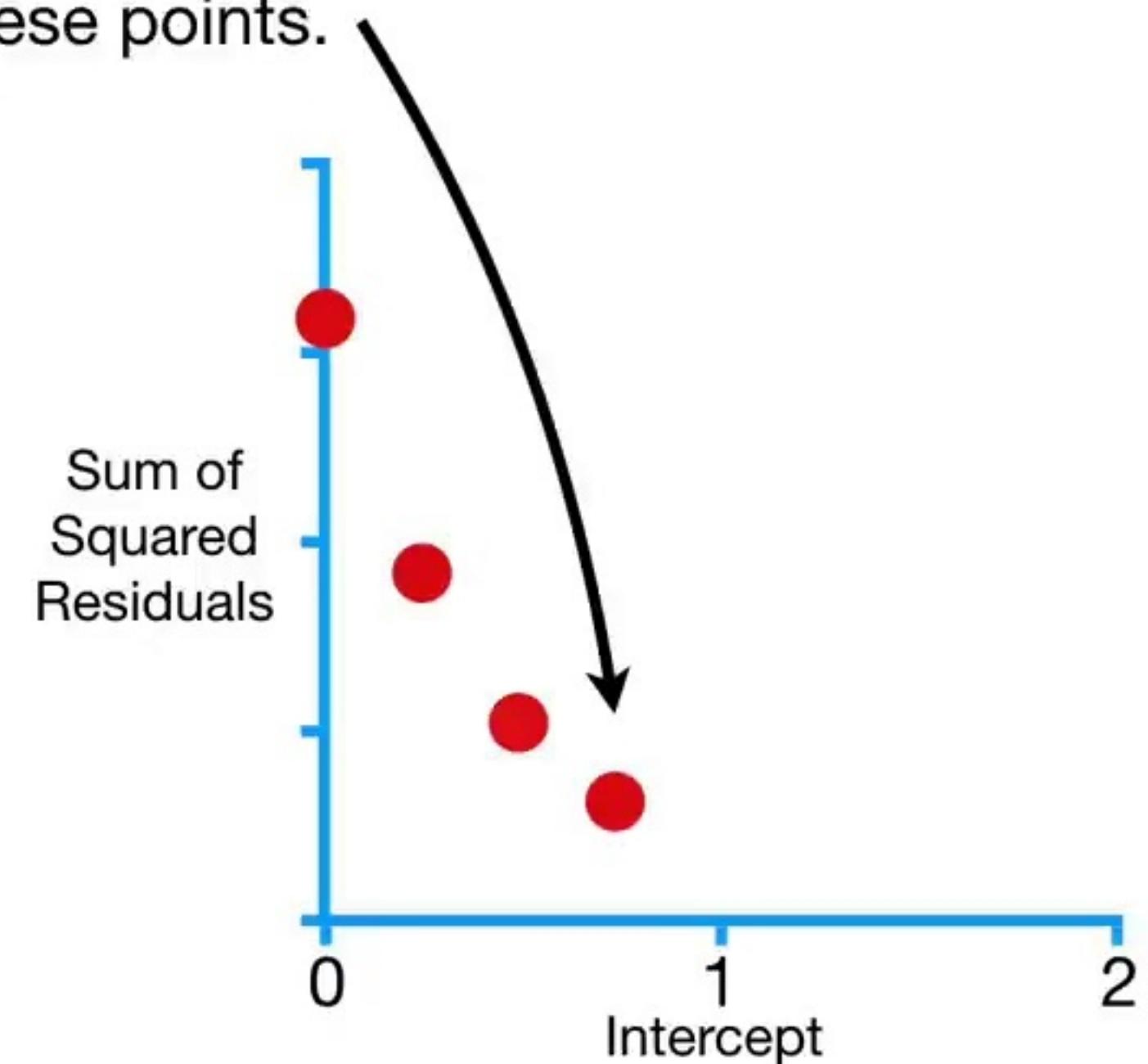
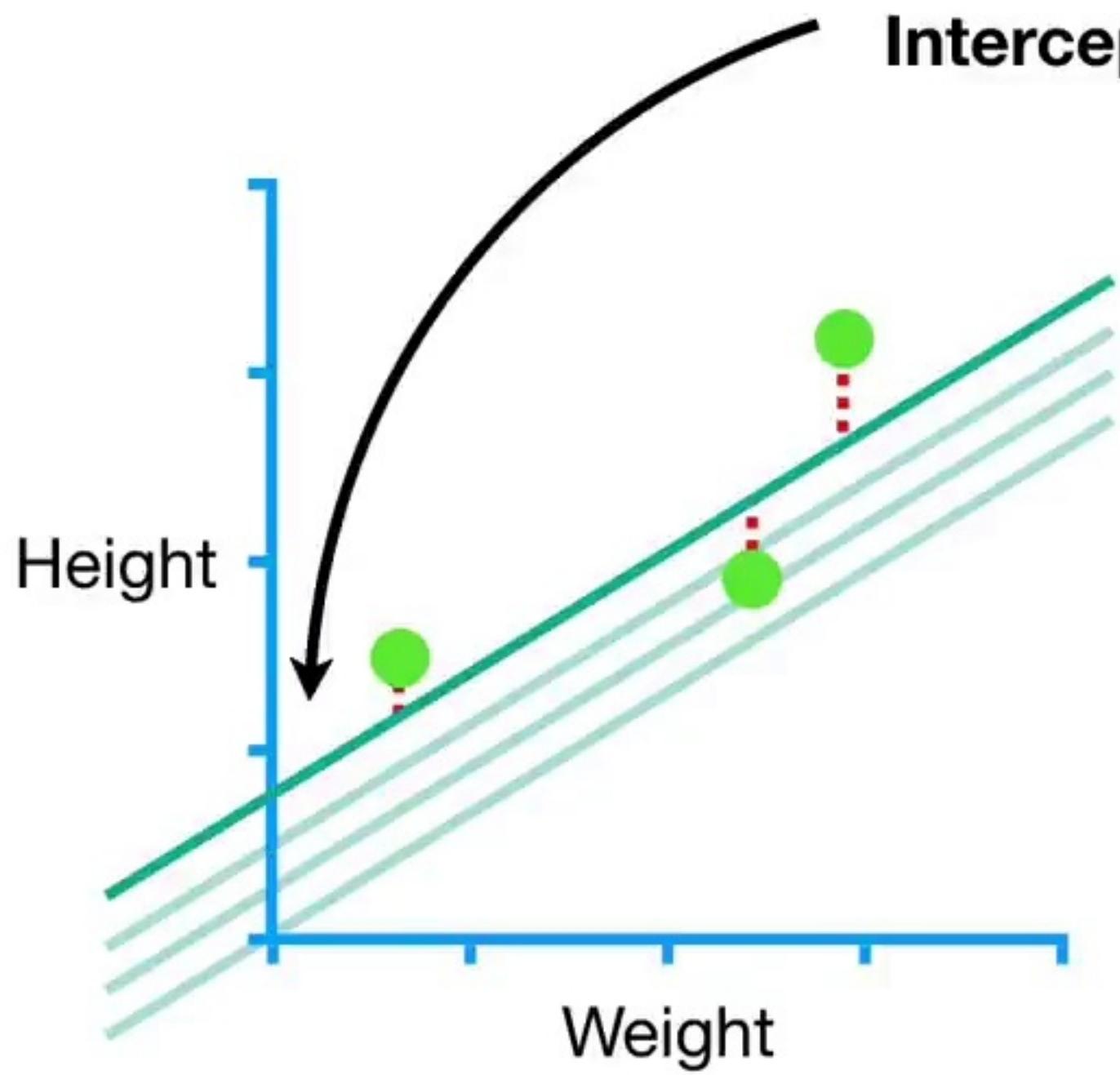
And if the
Intercept = 0.5...



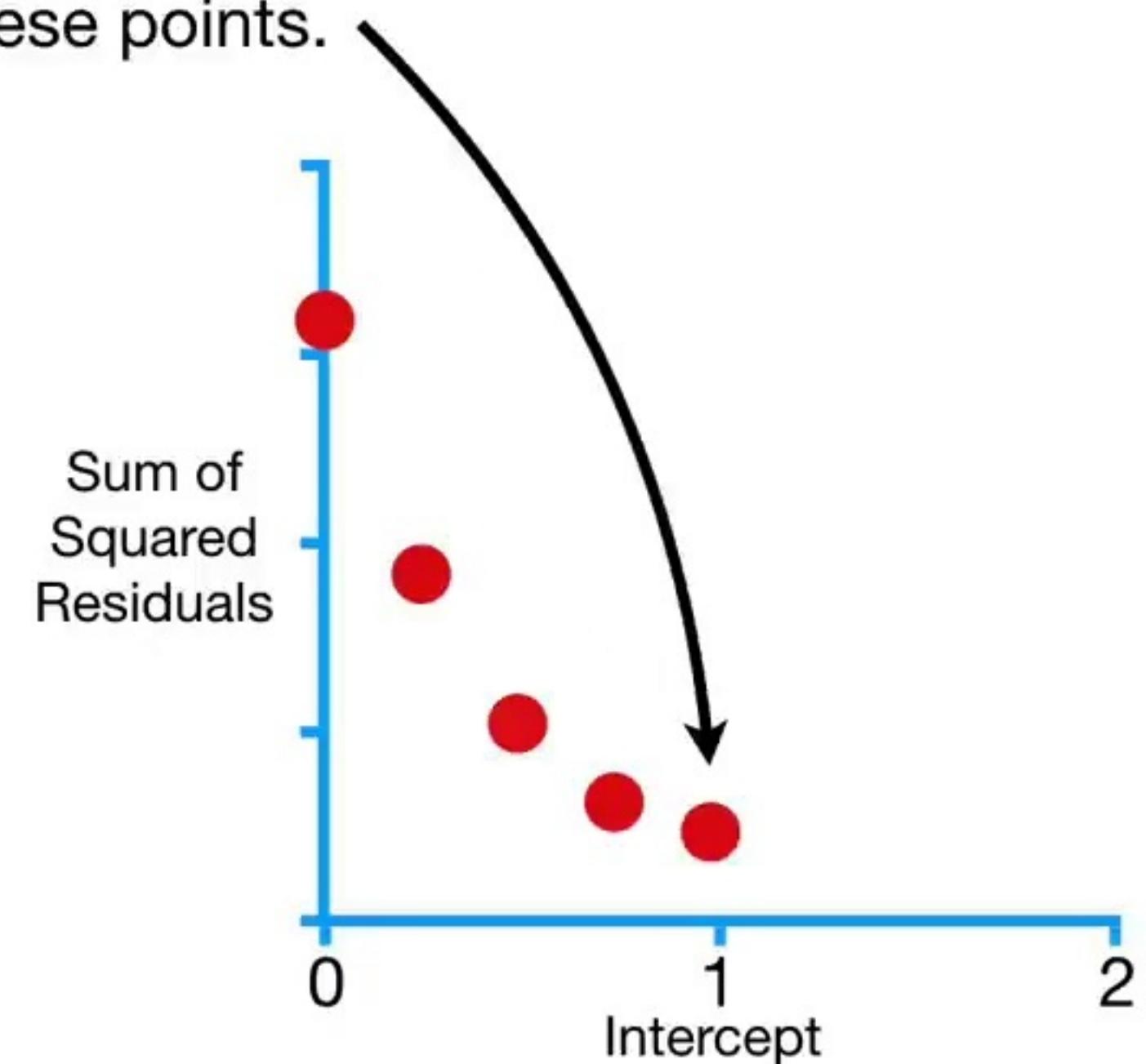
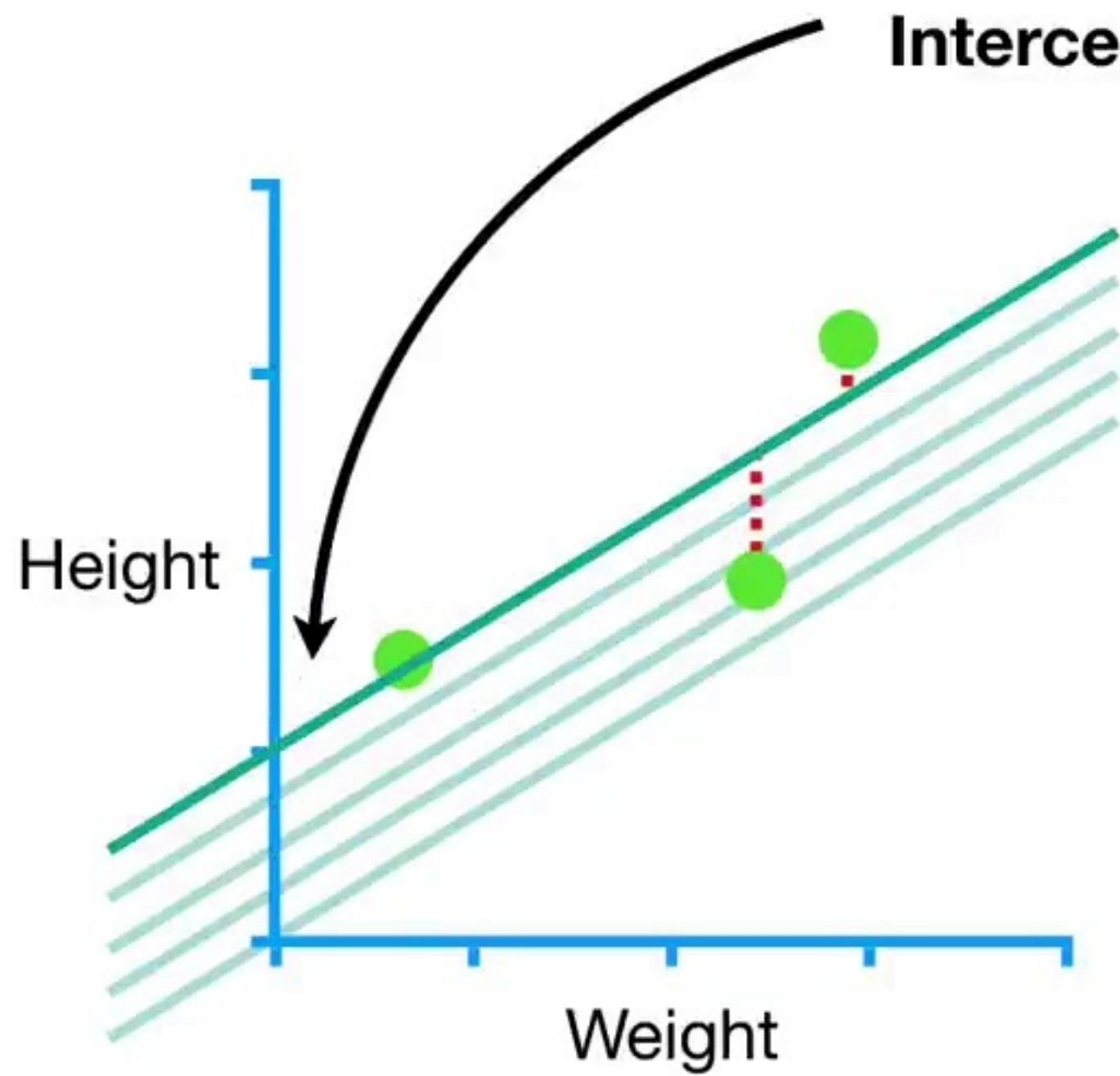
...then we would get
this point.



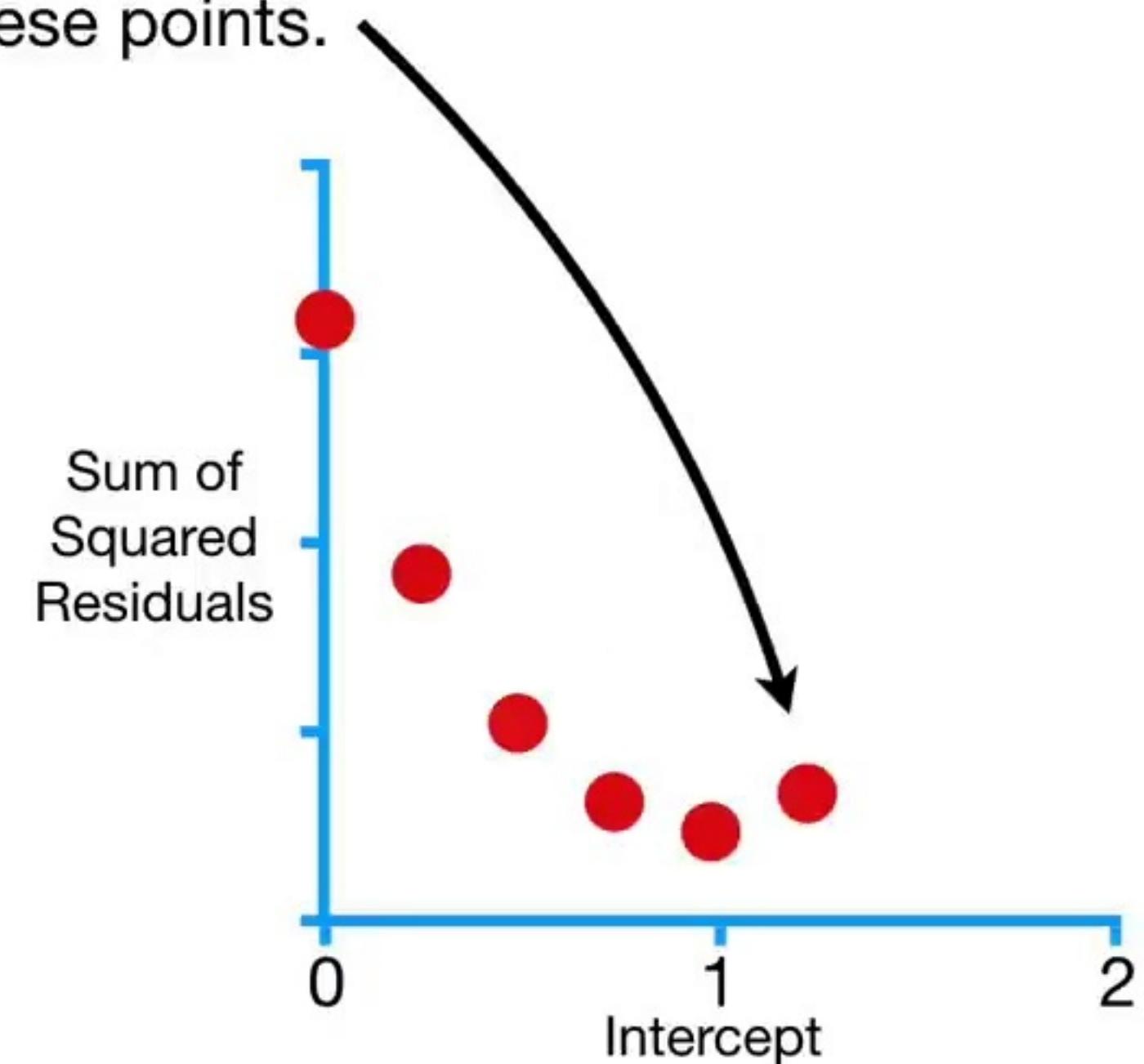
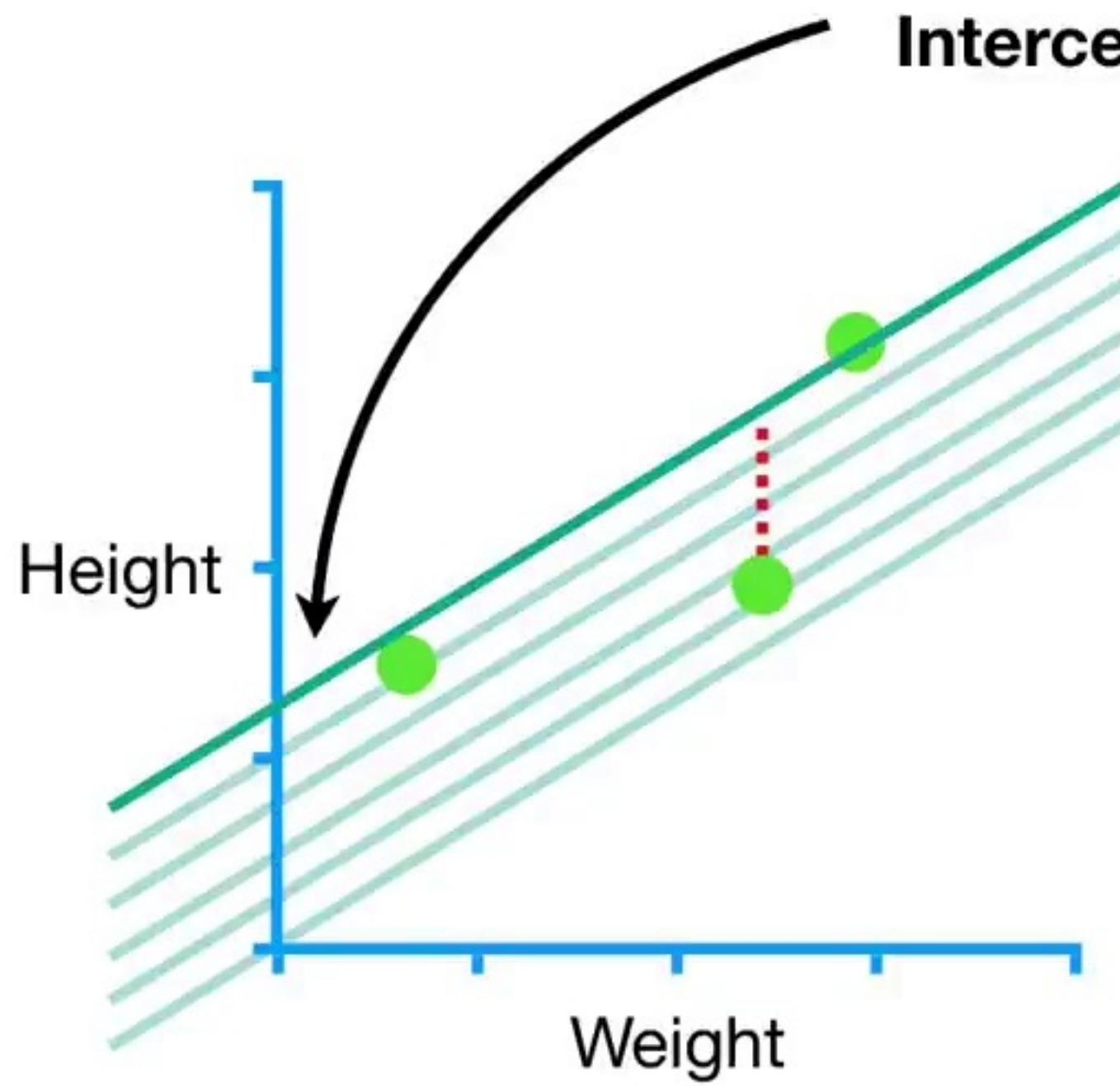
And for increasing values for the **Intercept**, we get these points.



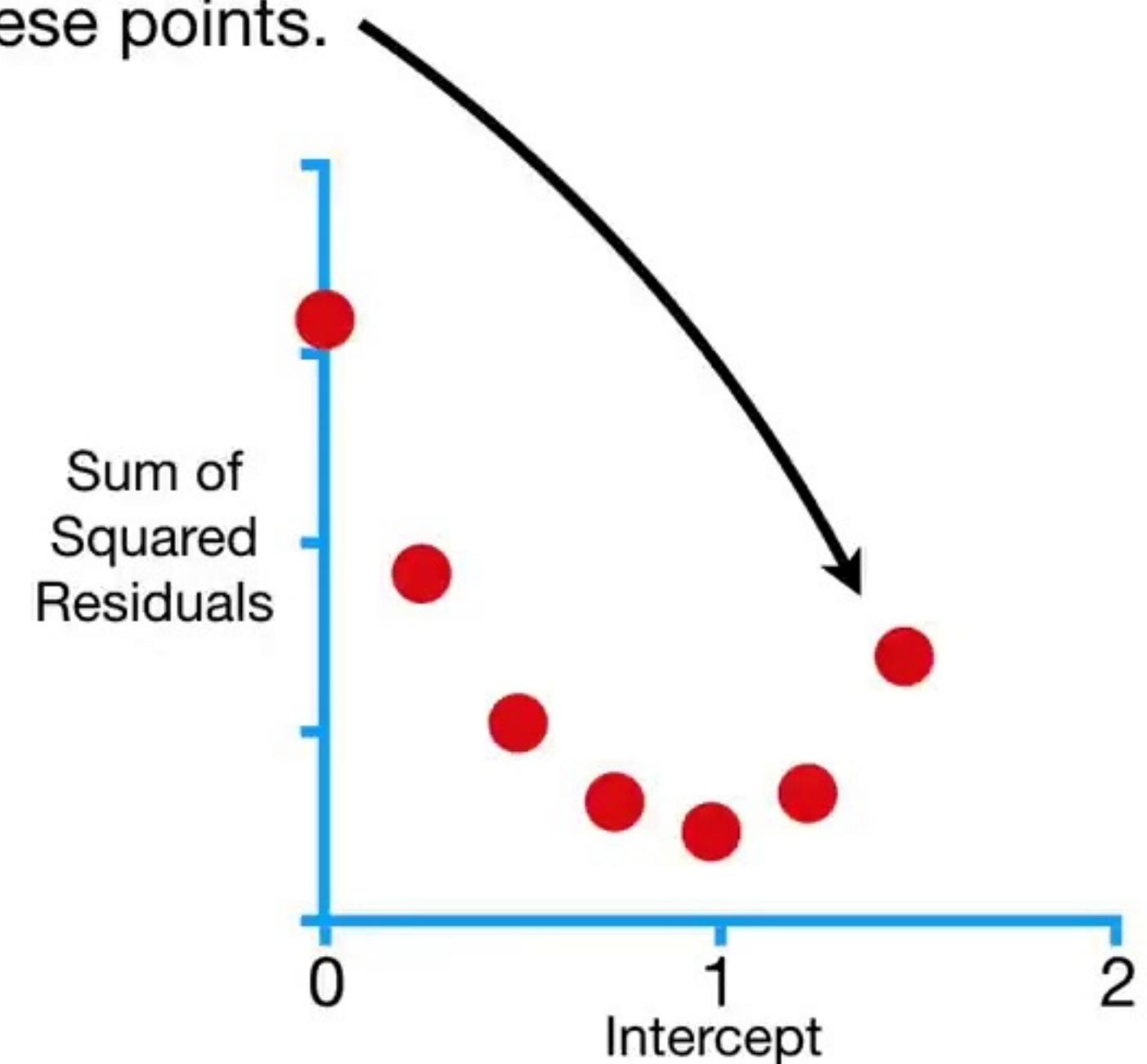
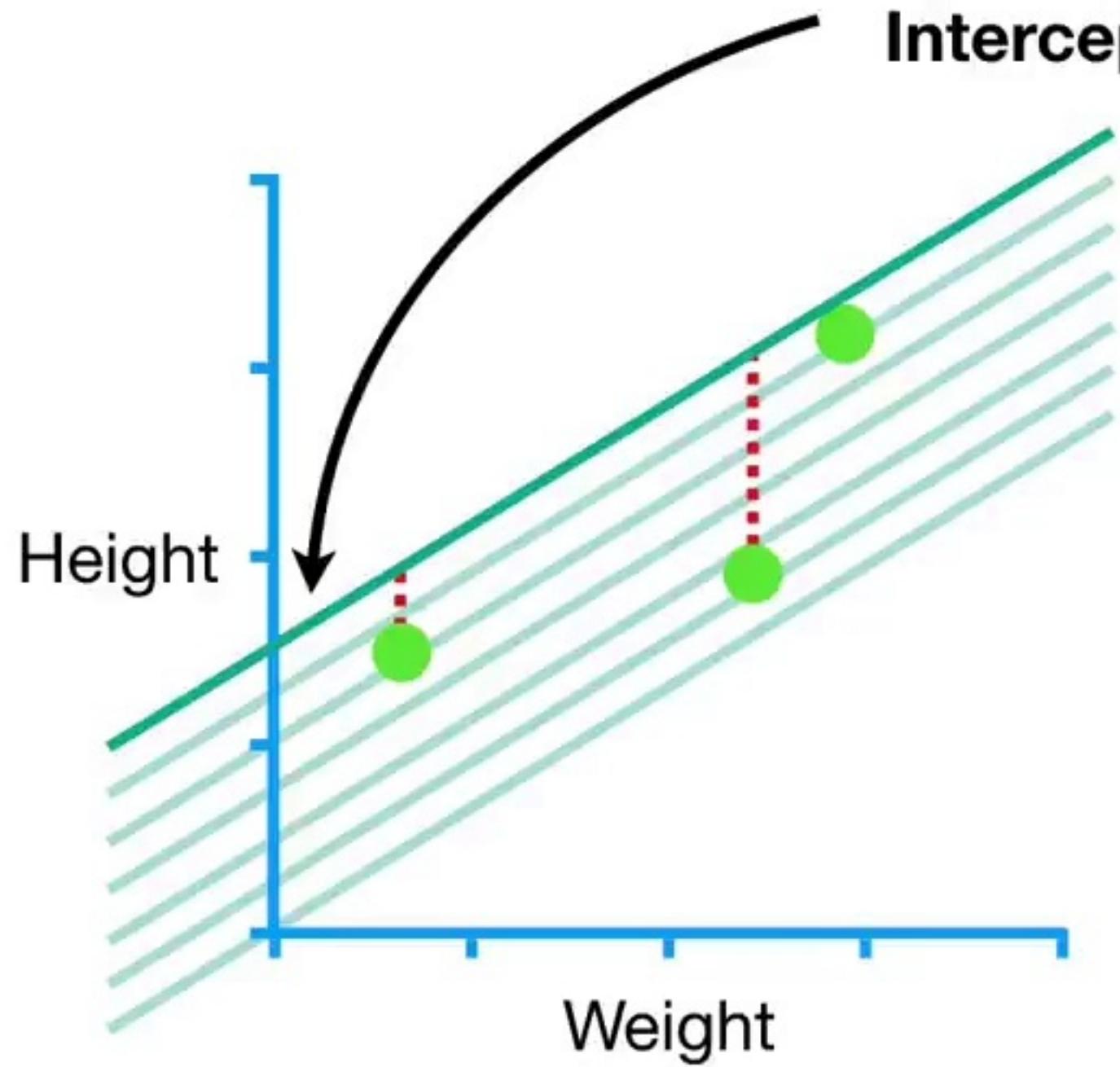
And for increasing values for the **Intercept**, we get these points.



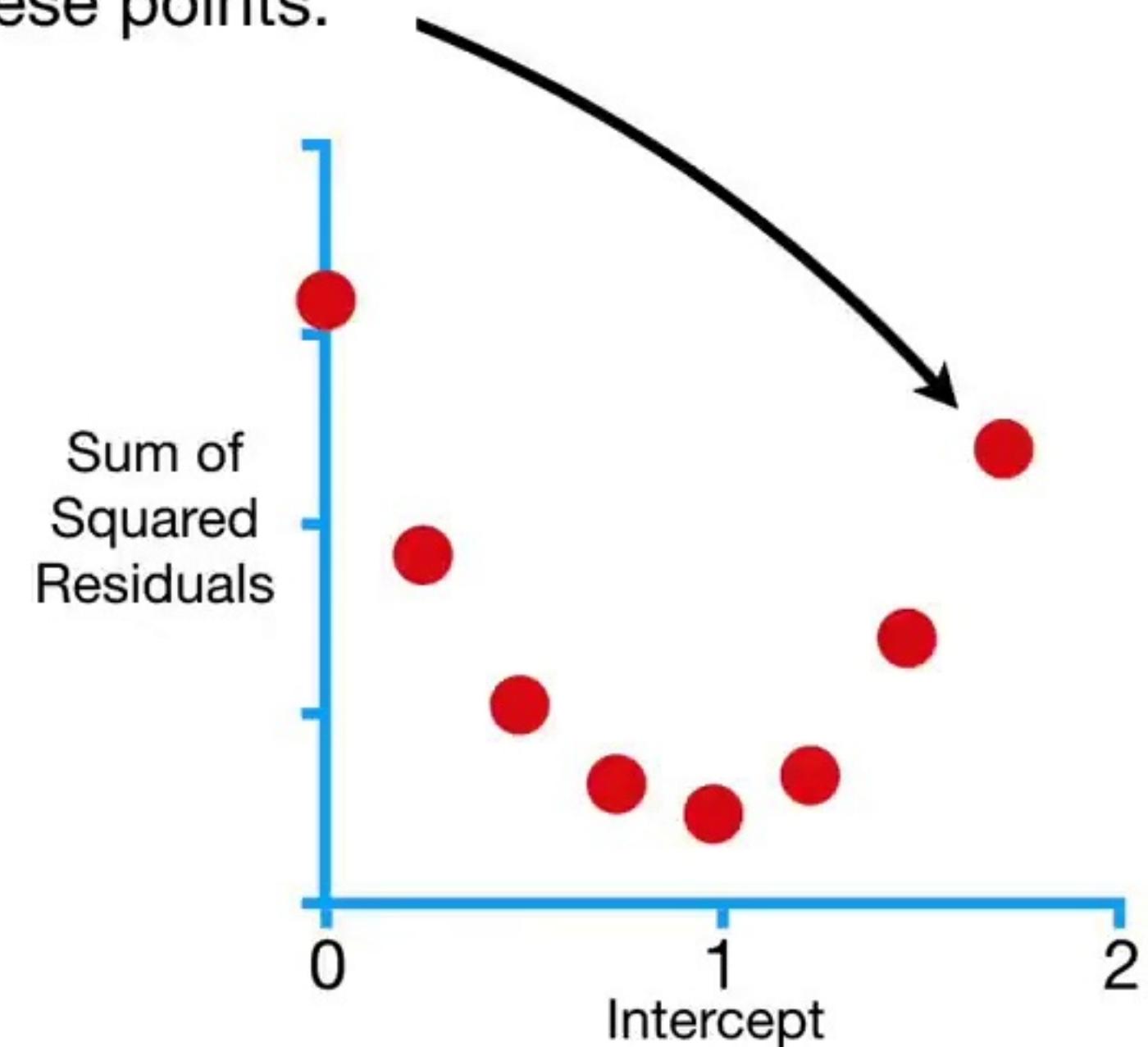
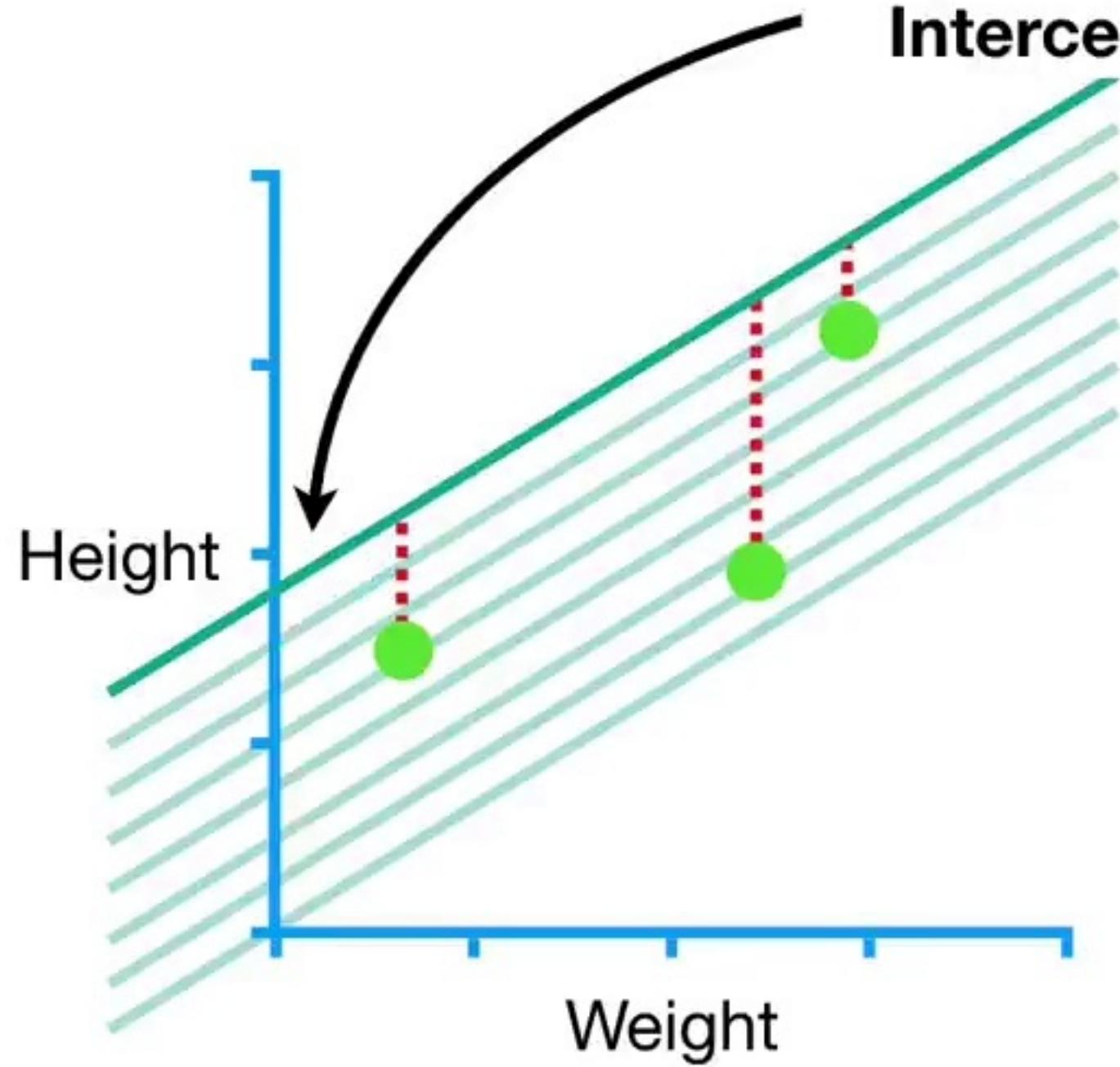
And for increasing values for the **Intercept**, we get these points.



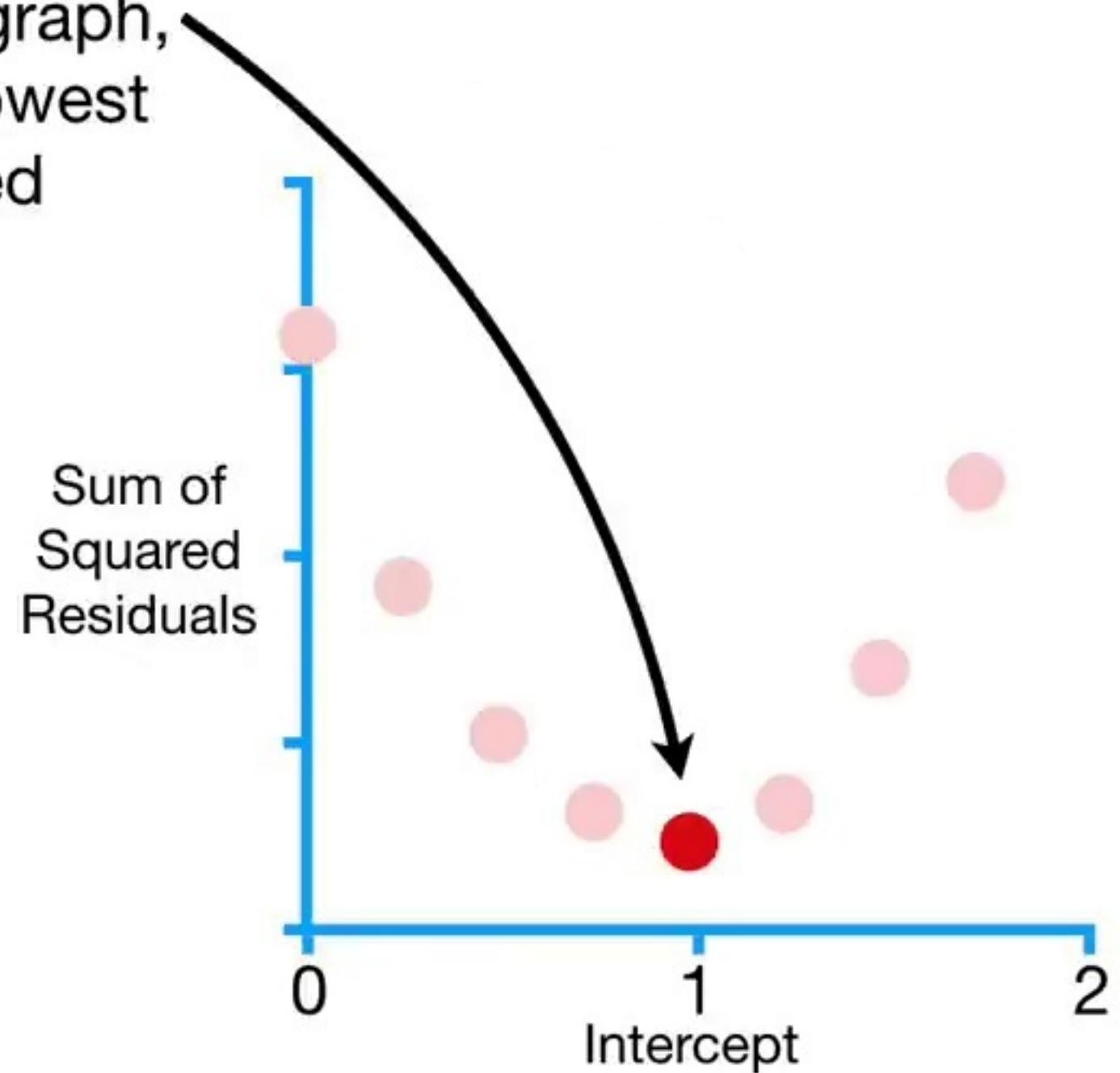
And for increasing values for the **Intercept**, we get these points.



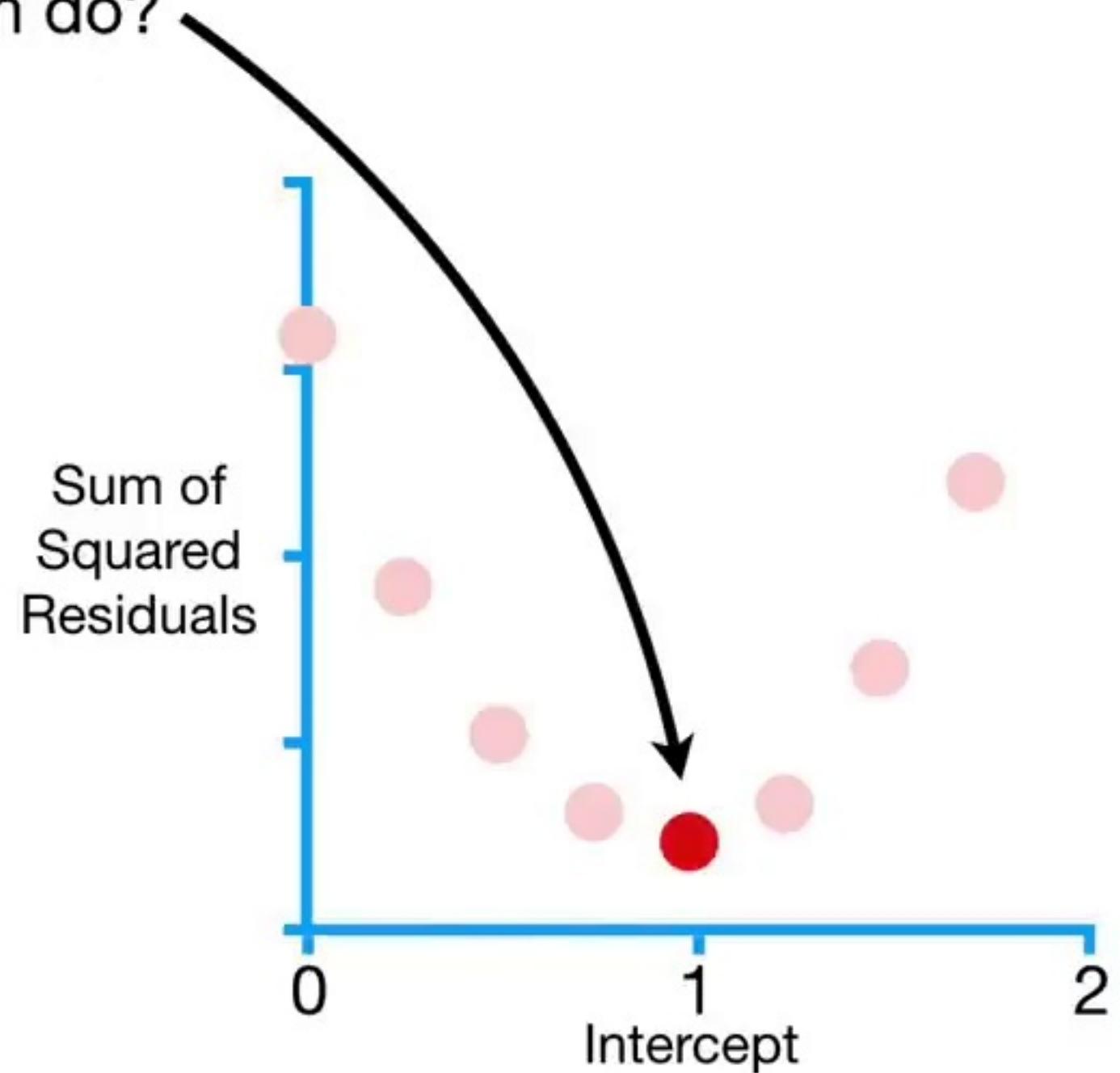
And for increasing values for the **Intercept**, we get these points.



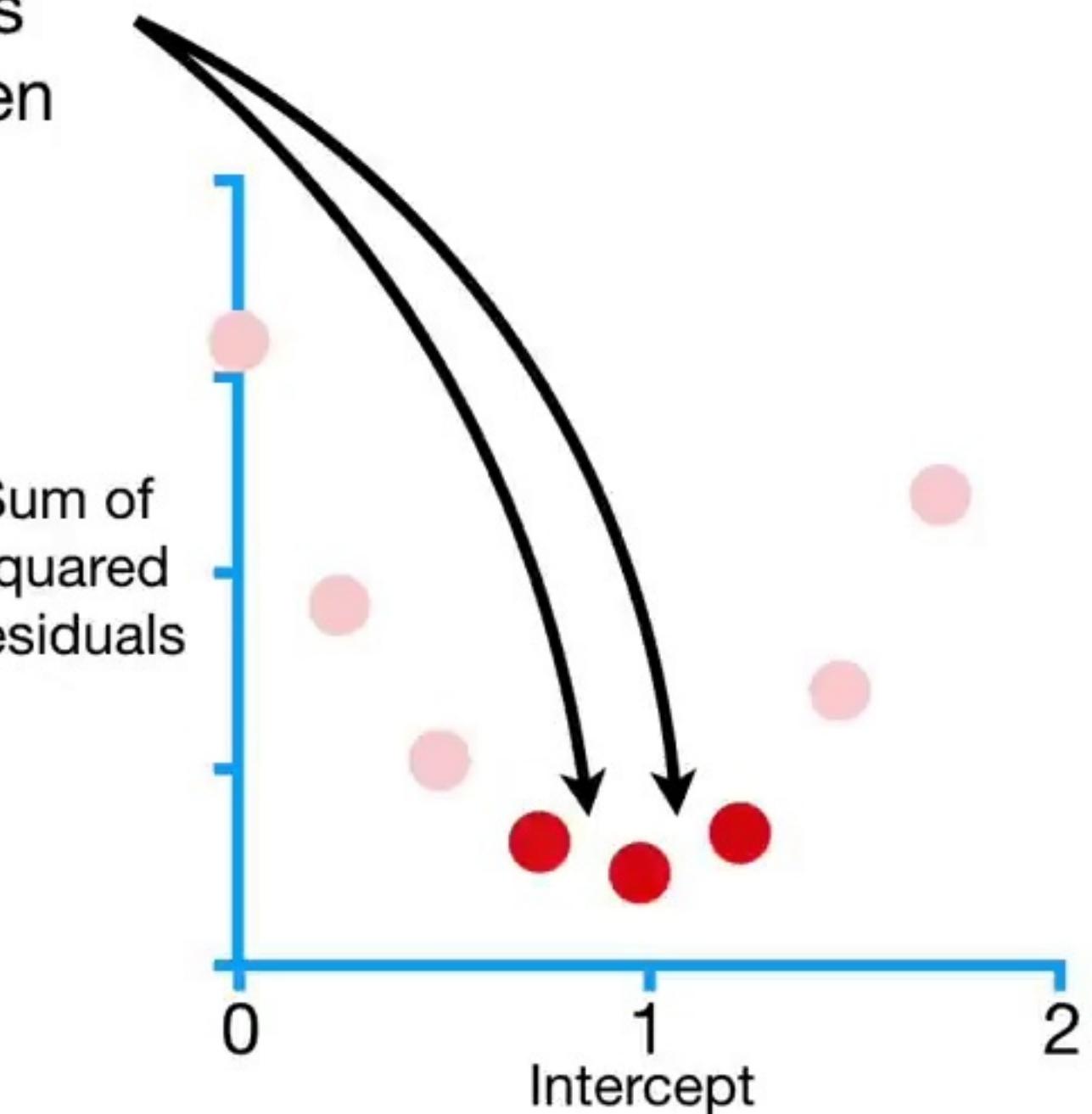
Of the points that we calculated for the graph, this one has the lowest Sum of Squared Residuals...



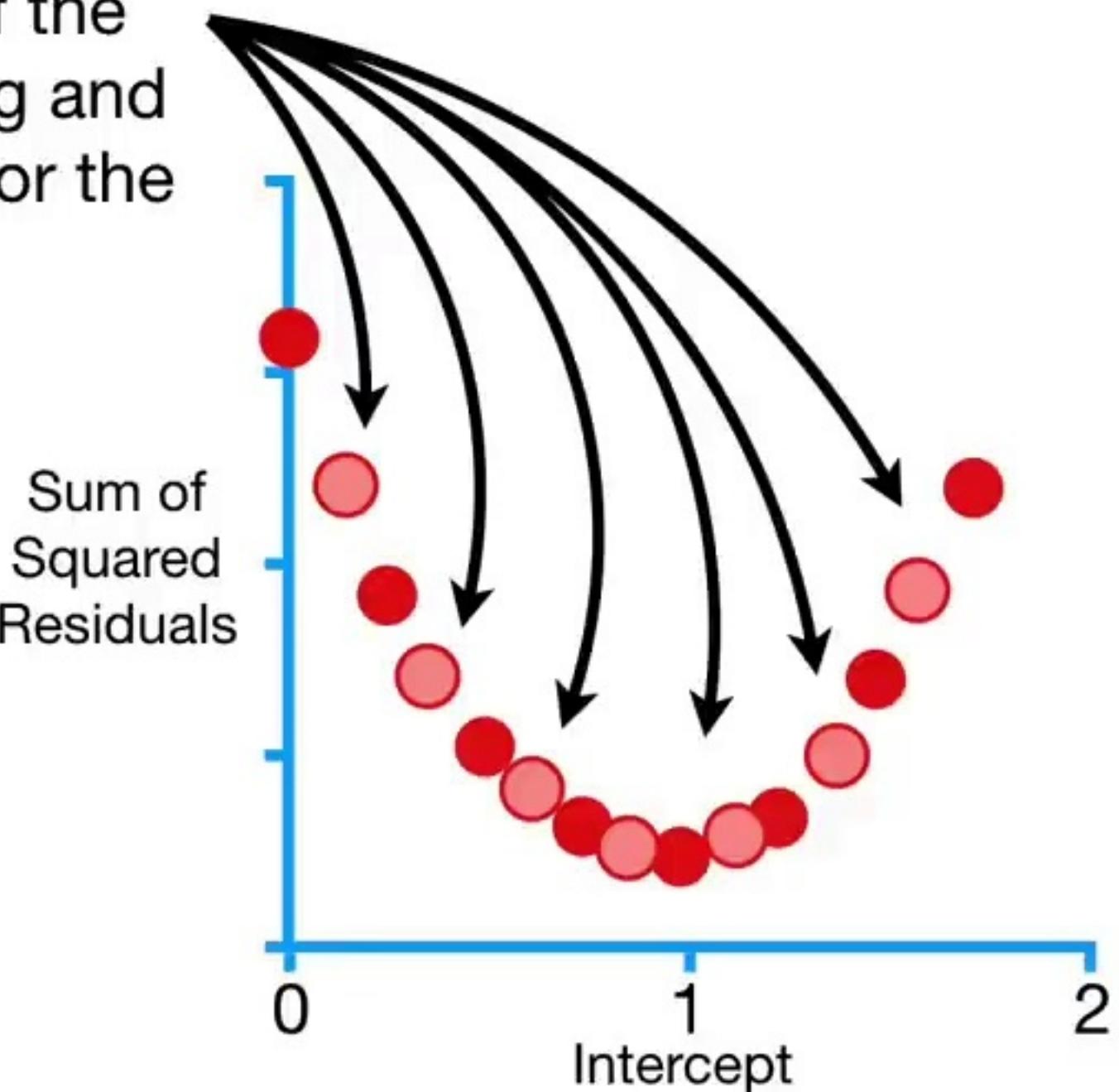
...but is it the best we can do?



What if the best value
for the **Intercept** is
somewhere between
these values?

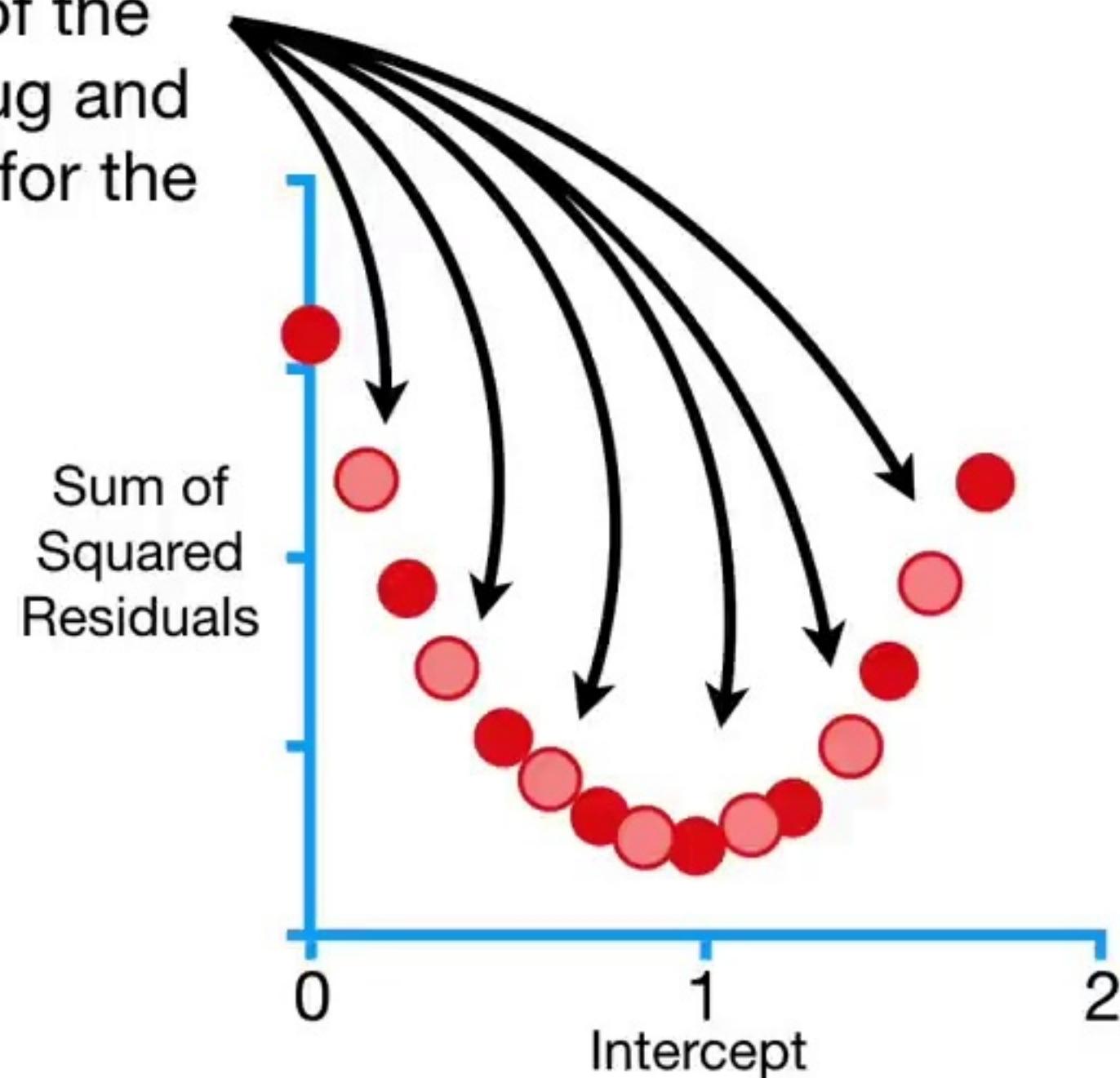


A slow and painful method for finding the minimal Sum of the Squared Residuals is to plug and chug a bunch more values for the **Intercept**.



A slow and painful method for finding the minimal Sum of the Squared Residuals is to plug and chug a bunch more values for the **Intercept**.

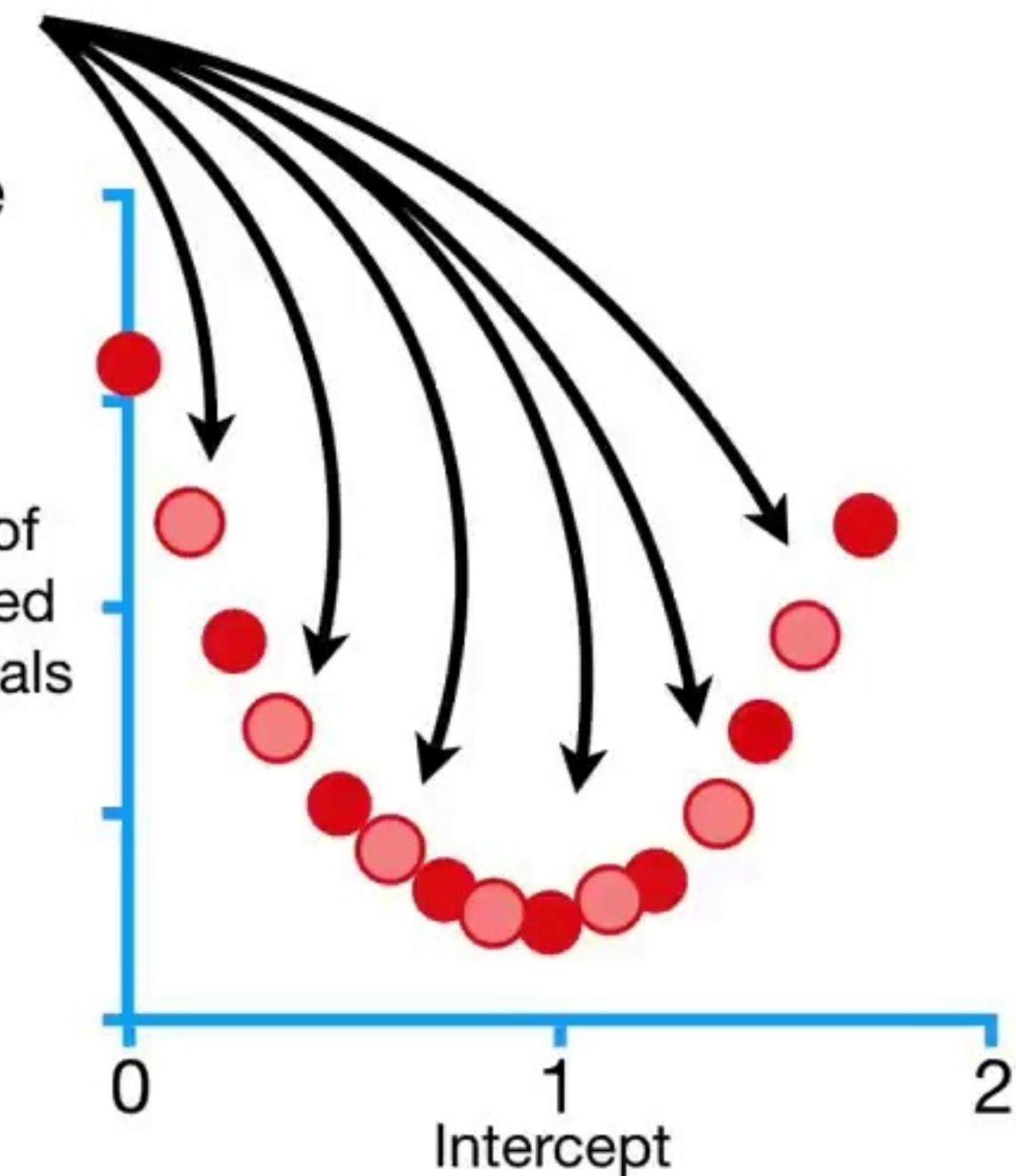
Ugh.



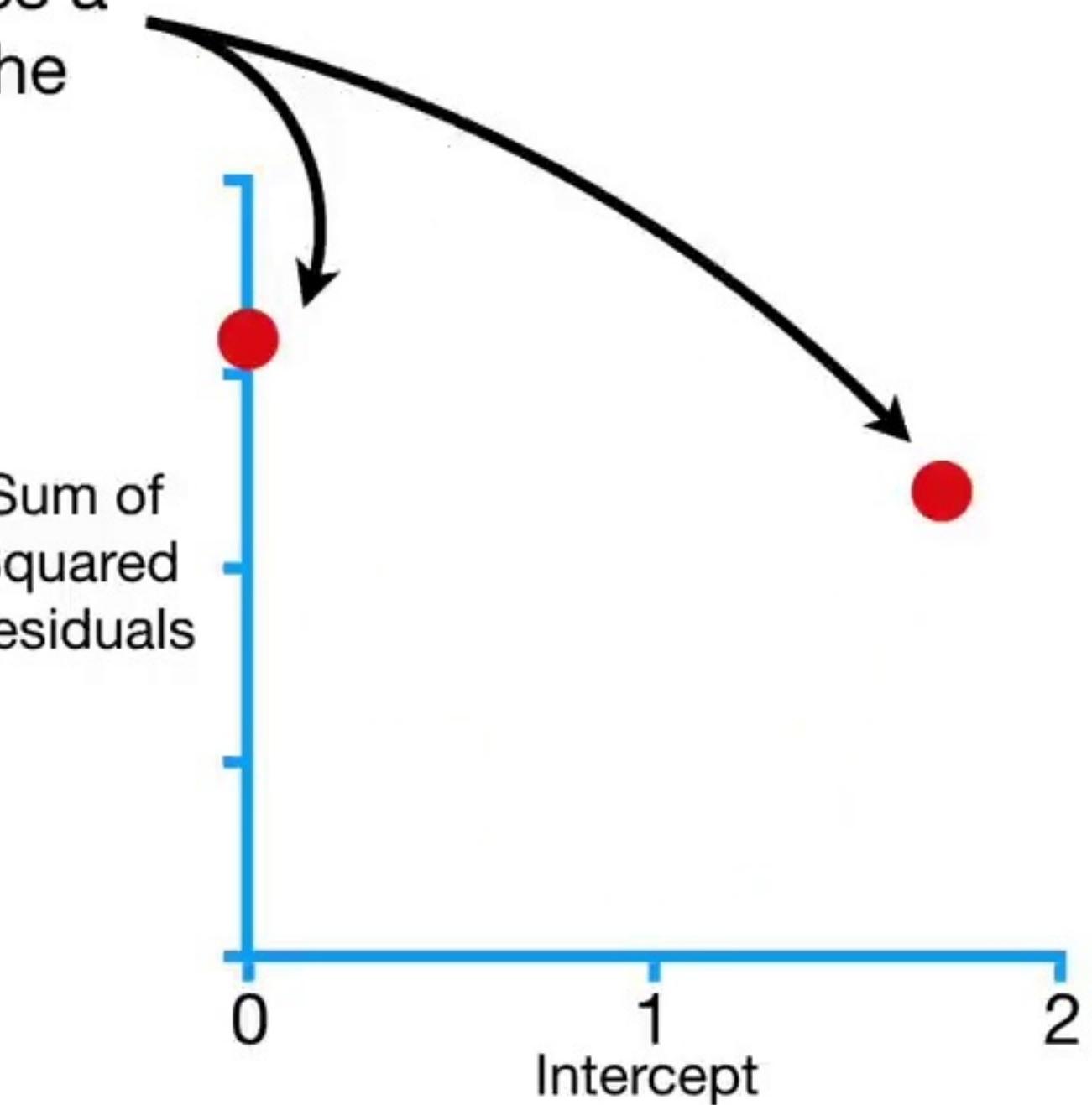
A slow and painful method for finding the minimal Sum of the Squared Residuals is to plug and chug a bunch more values for the **Intercept**.

Ugh.

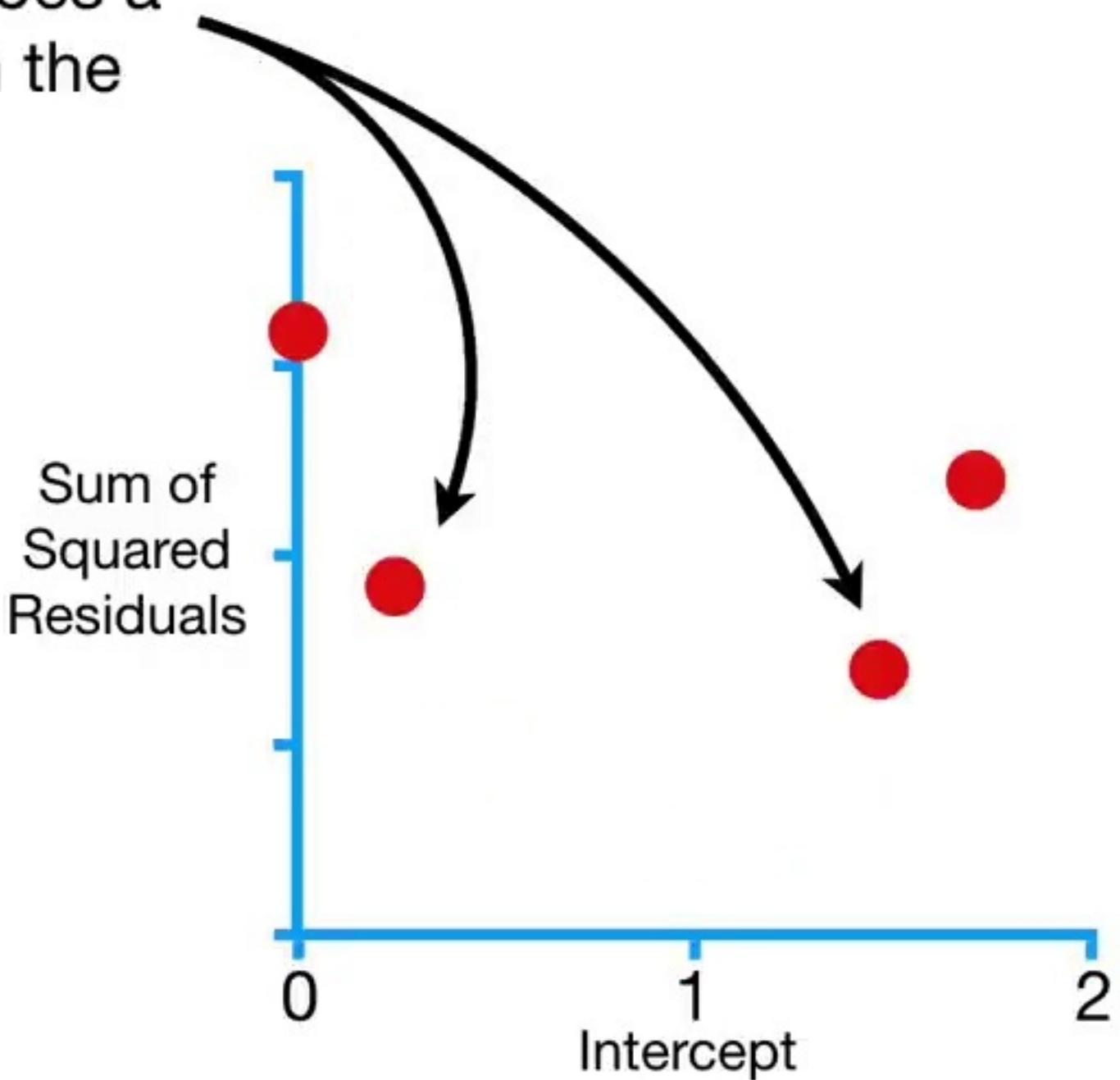
Don't despair!
Gradient Descent is
way more efficient!



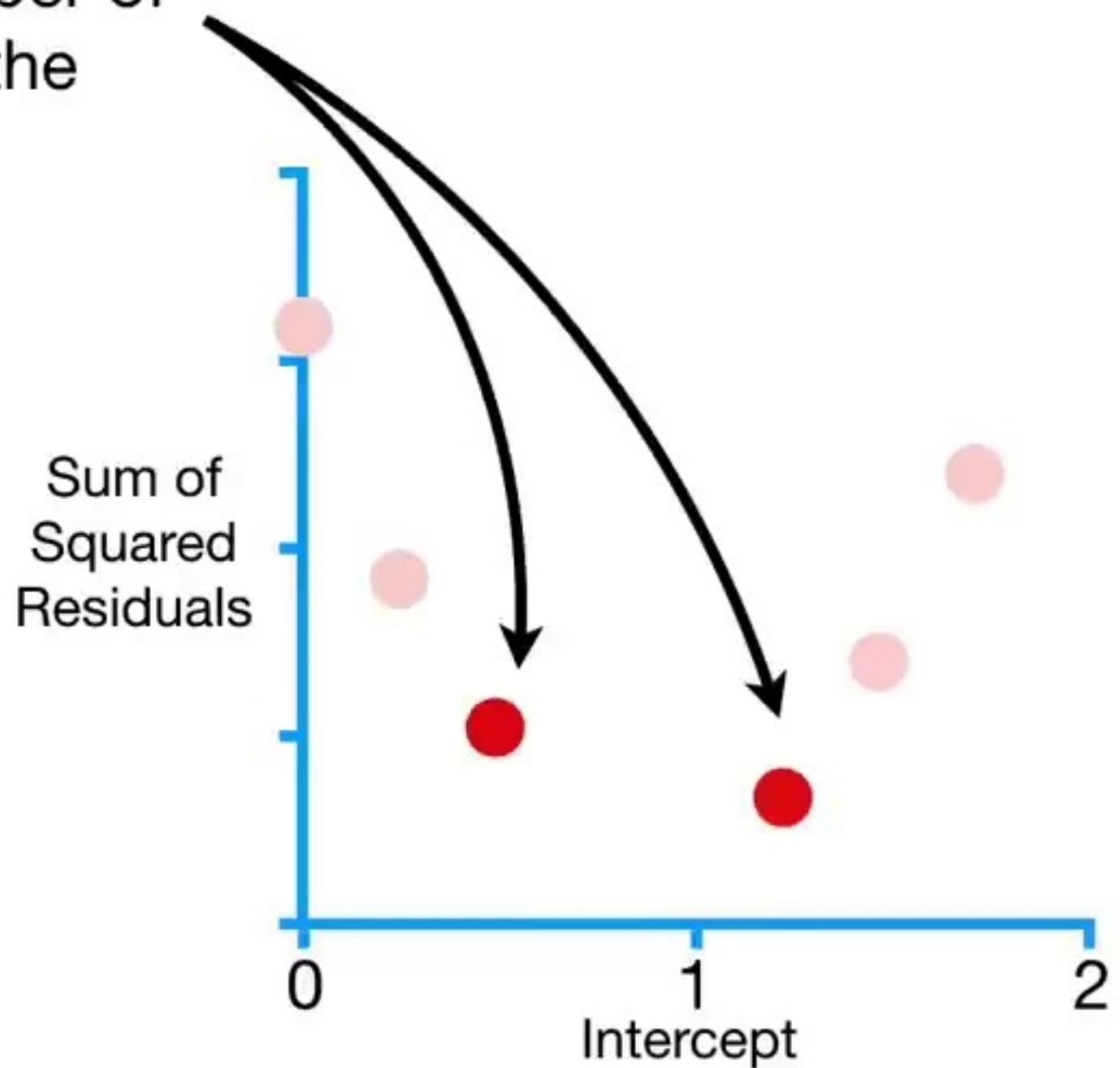
Gradient Descent only does a few calculations far from the optimal solution...



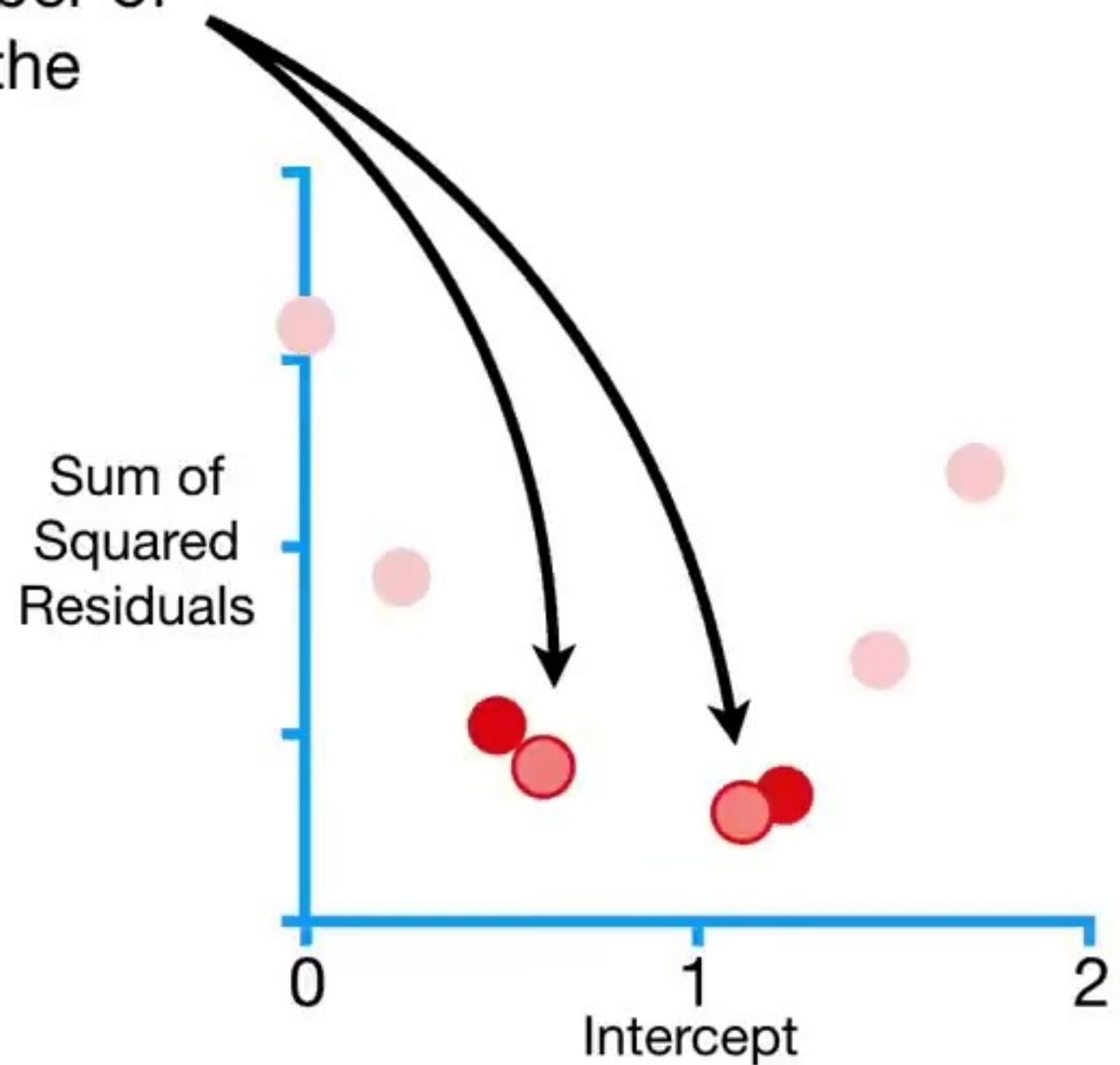
Gradient Descent only does a few calculations far from the optimal solution...



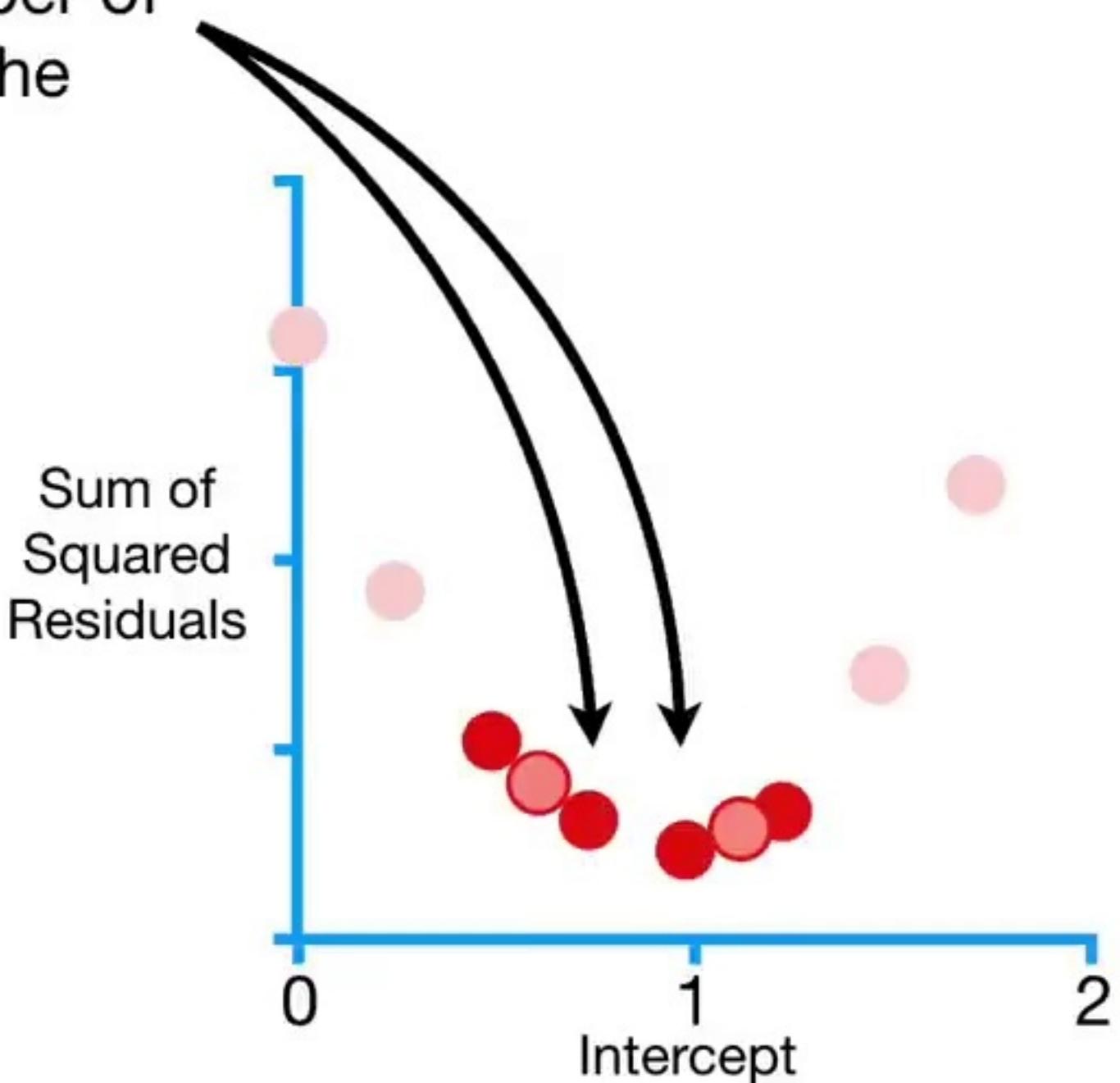
...and increases the number of calculations closer to the optimal value.



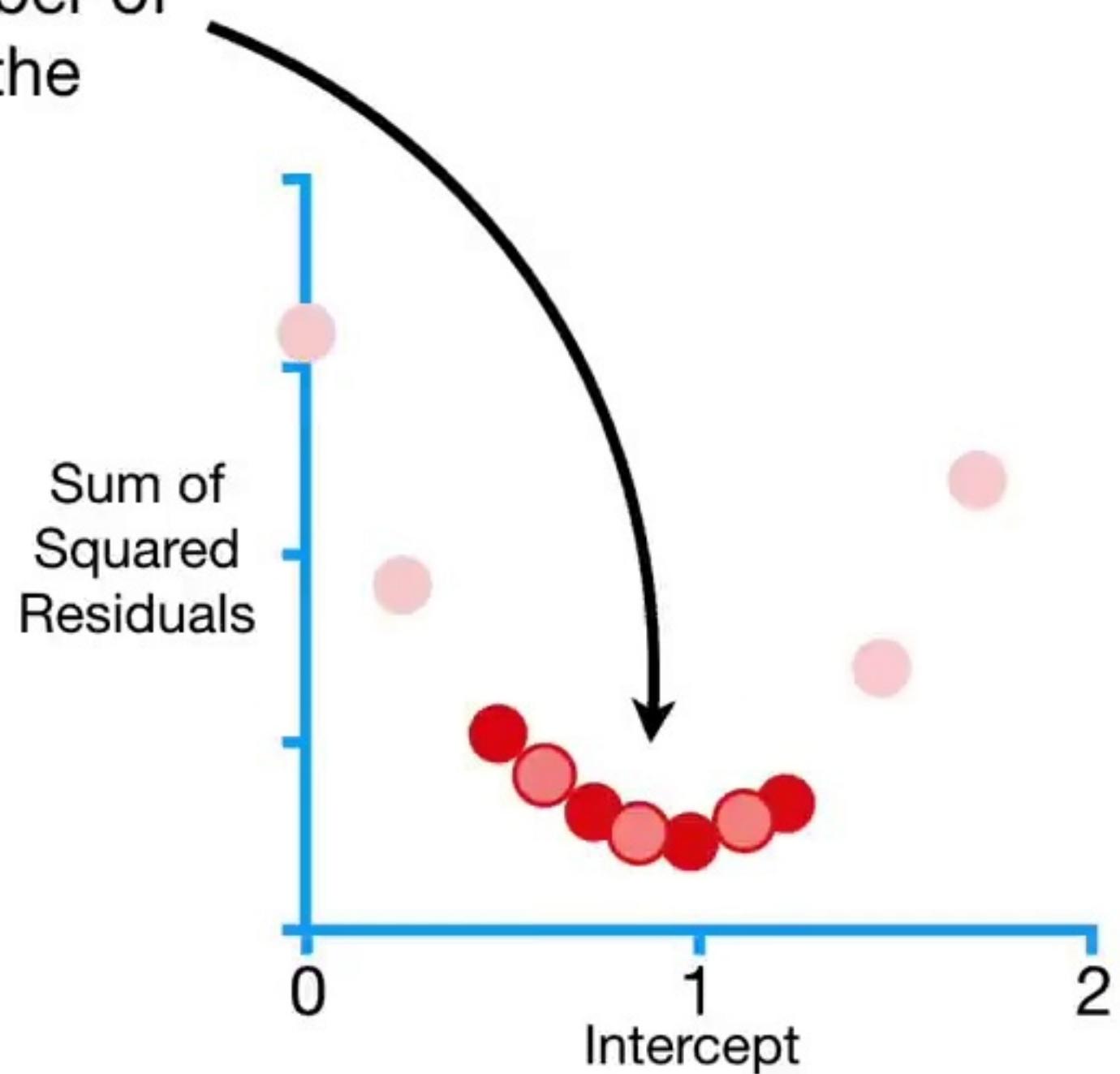
...and increases the number of calculations closer to the optimal value.



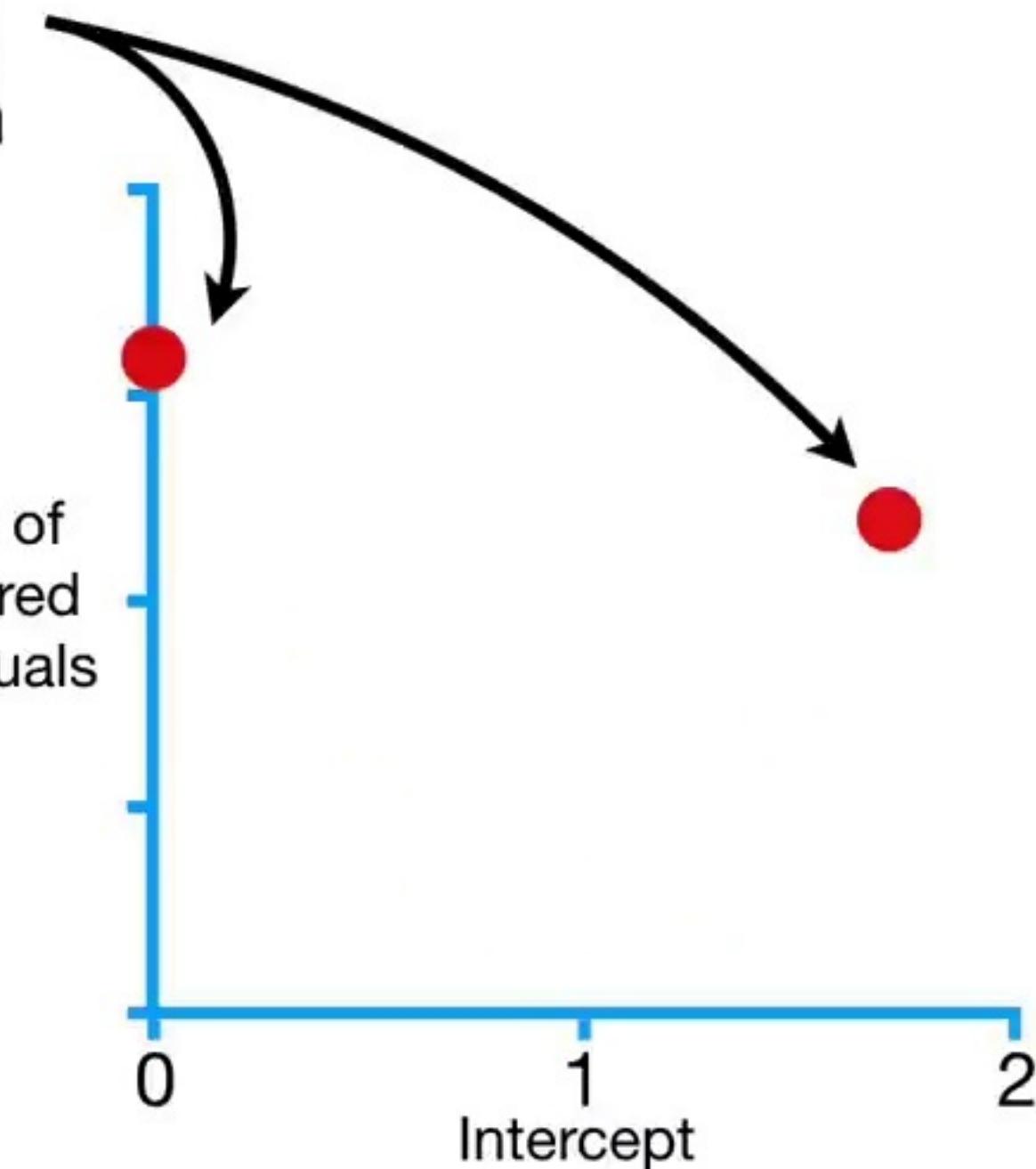
...and increases the number of calculations closer to the optimal value.



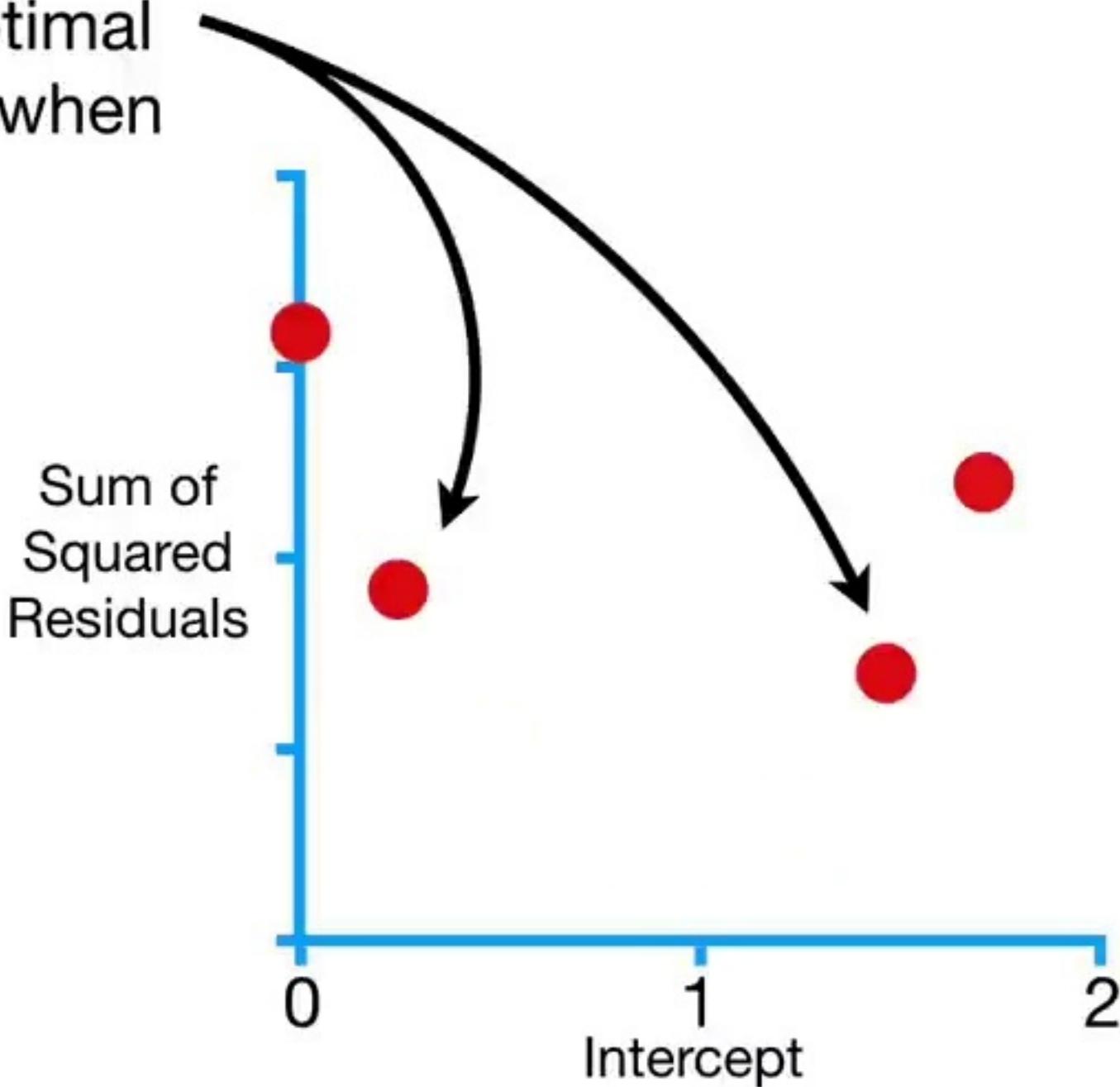
...and increases the number of calculations closer to the optimal value.



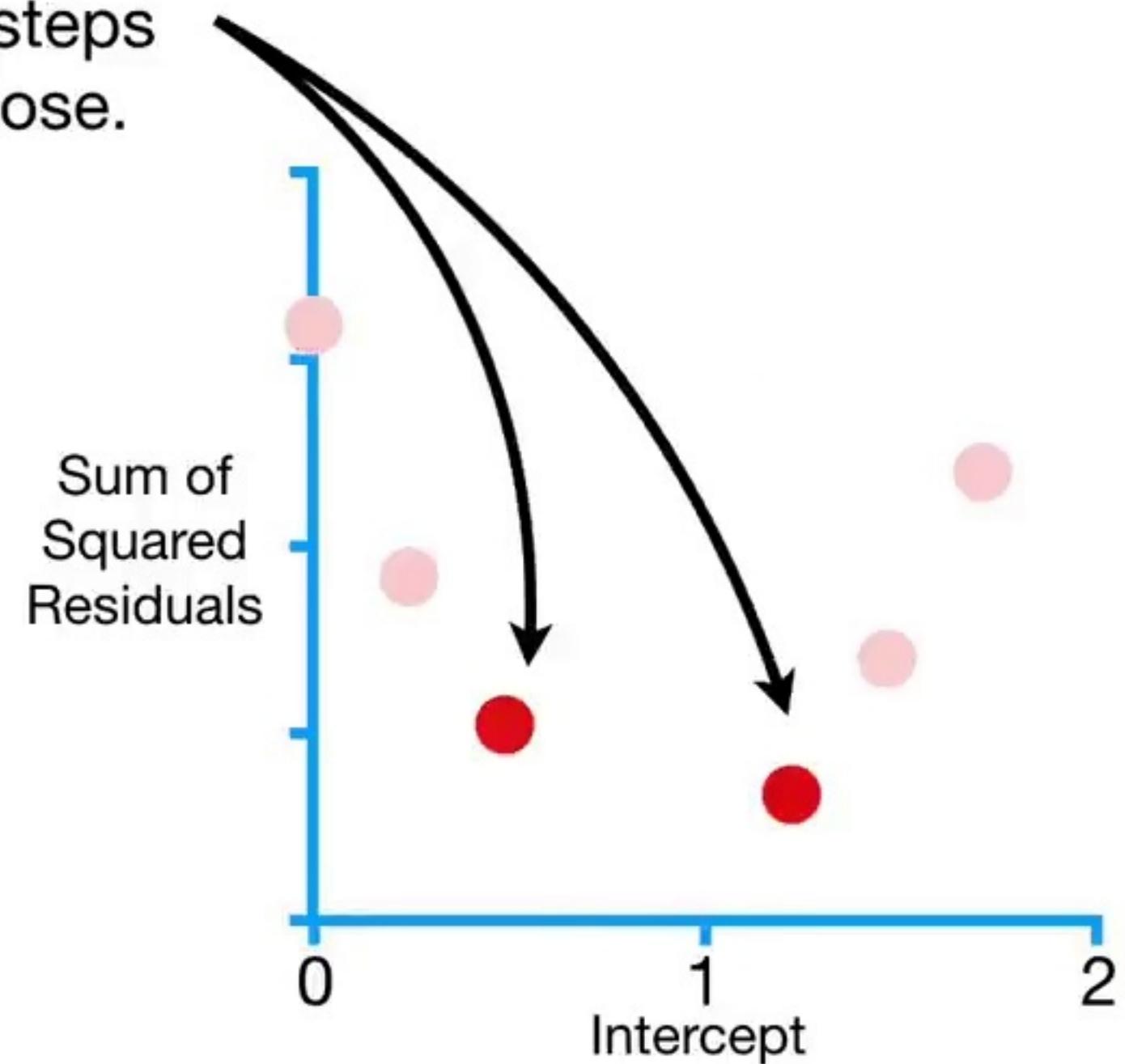
In other words, **Gradient Descent** identifies the optimal value by taking big steps when it is far away...



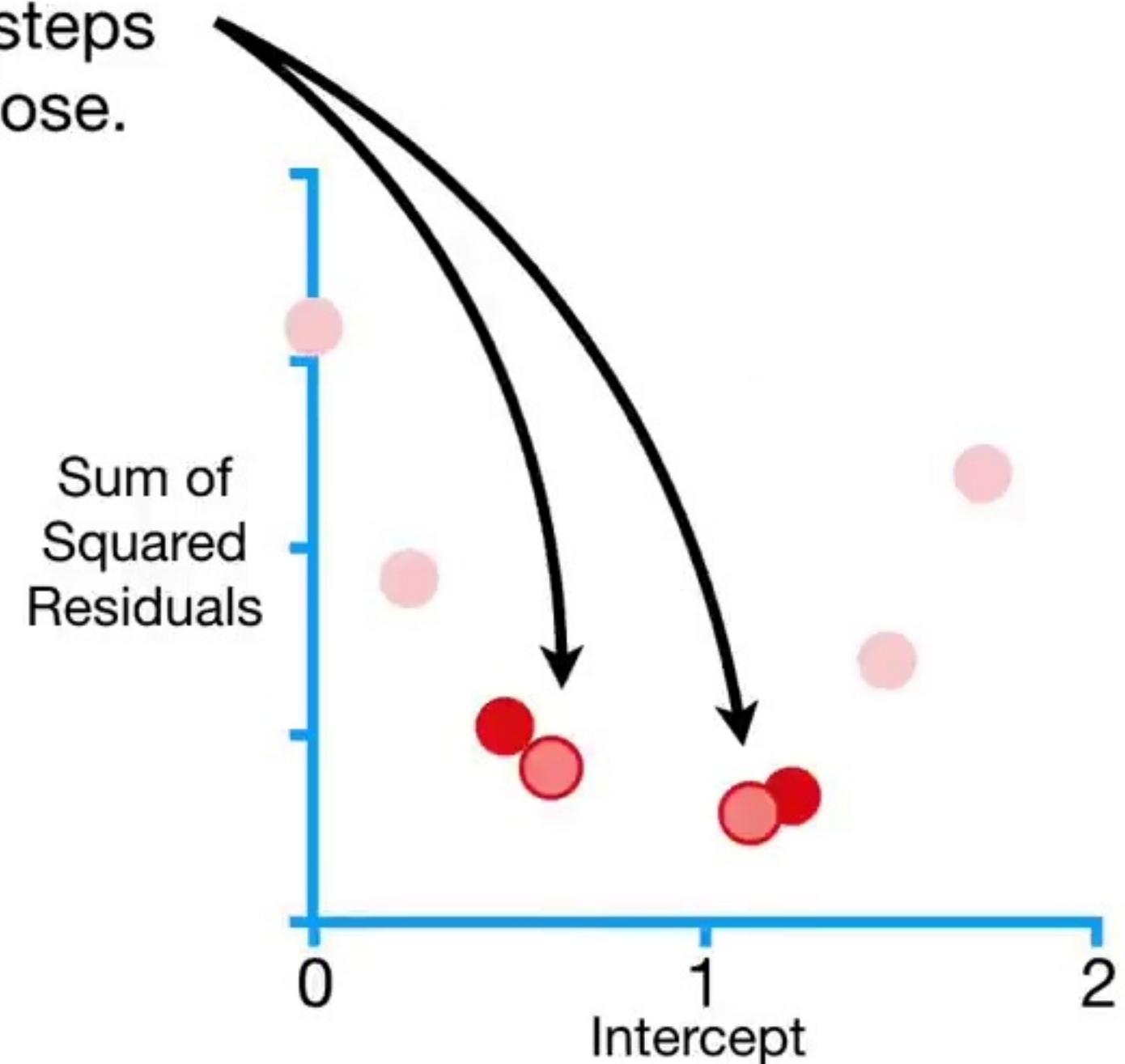
In other words, **Gradient Descent** identifies the optimal value by taking big steps when it is far away...



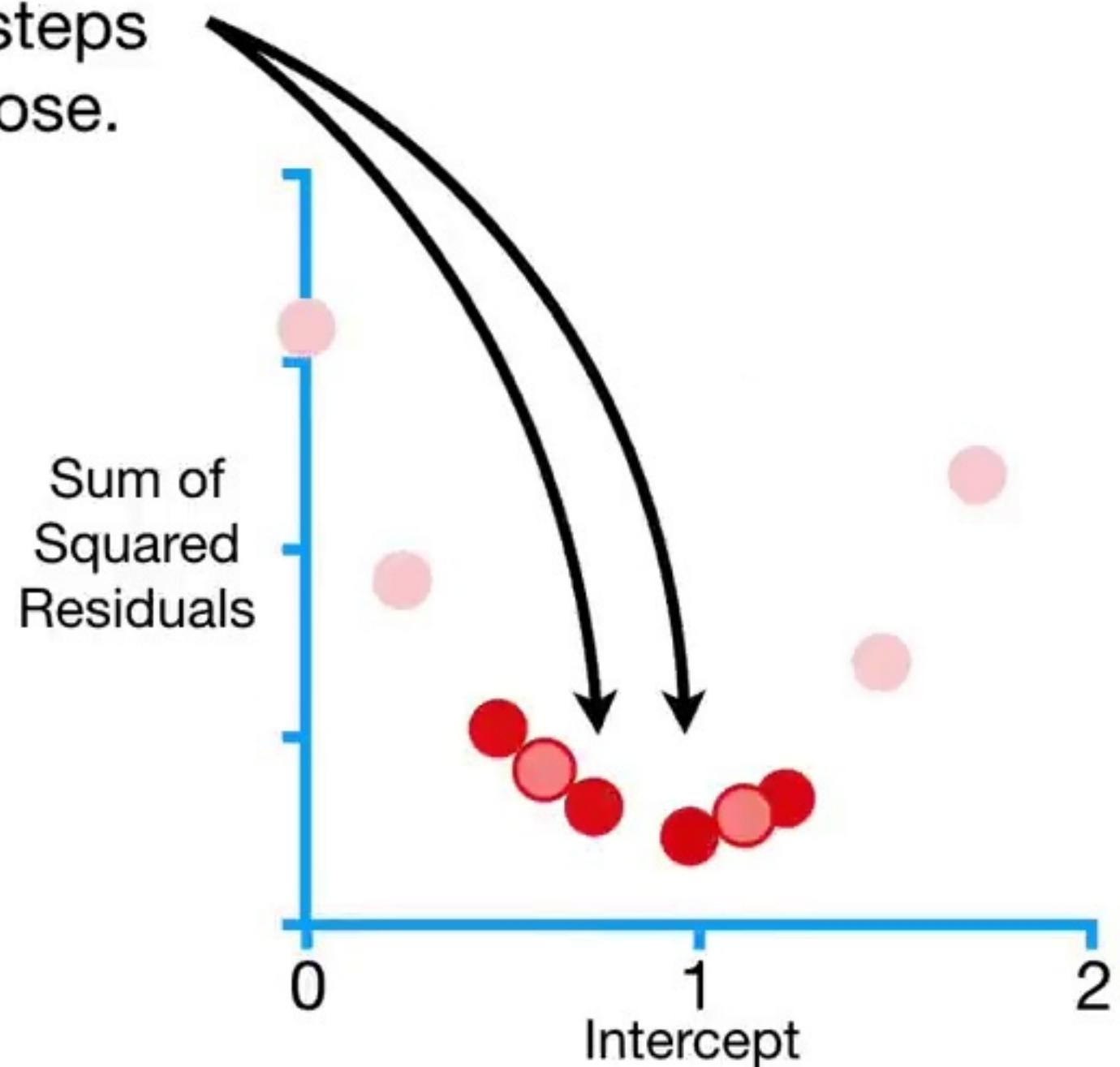
...and baby steps
when it is close.



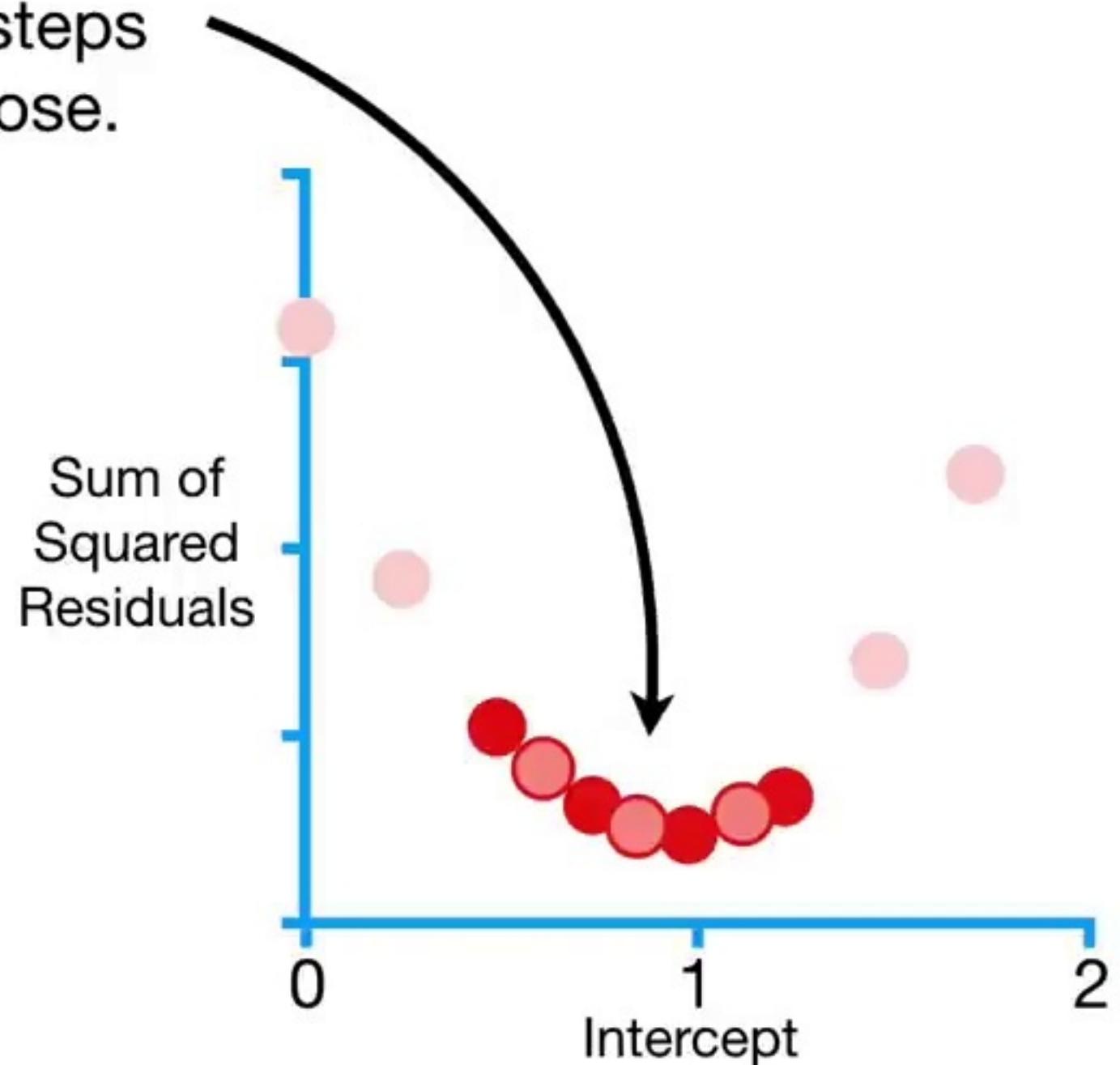
...and baby steps
when it is close.



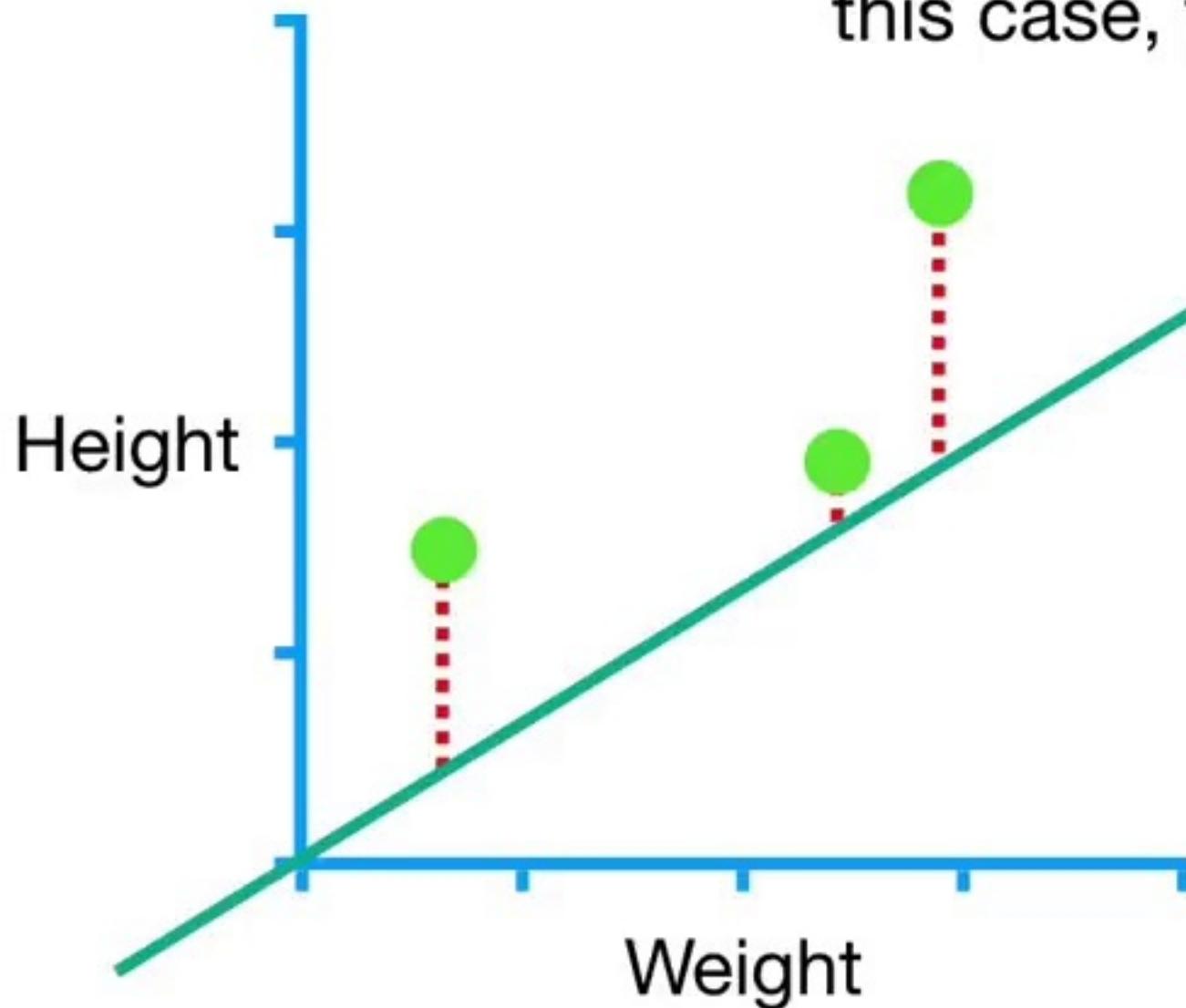
...and baby steps
when it is close.



...and baby steps
when it is close.

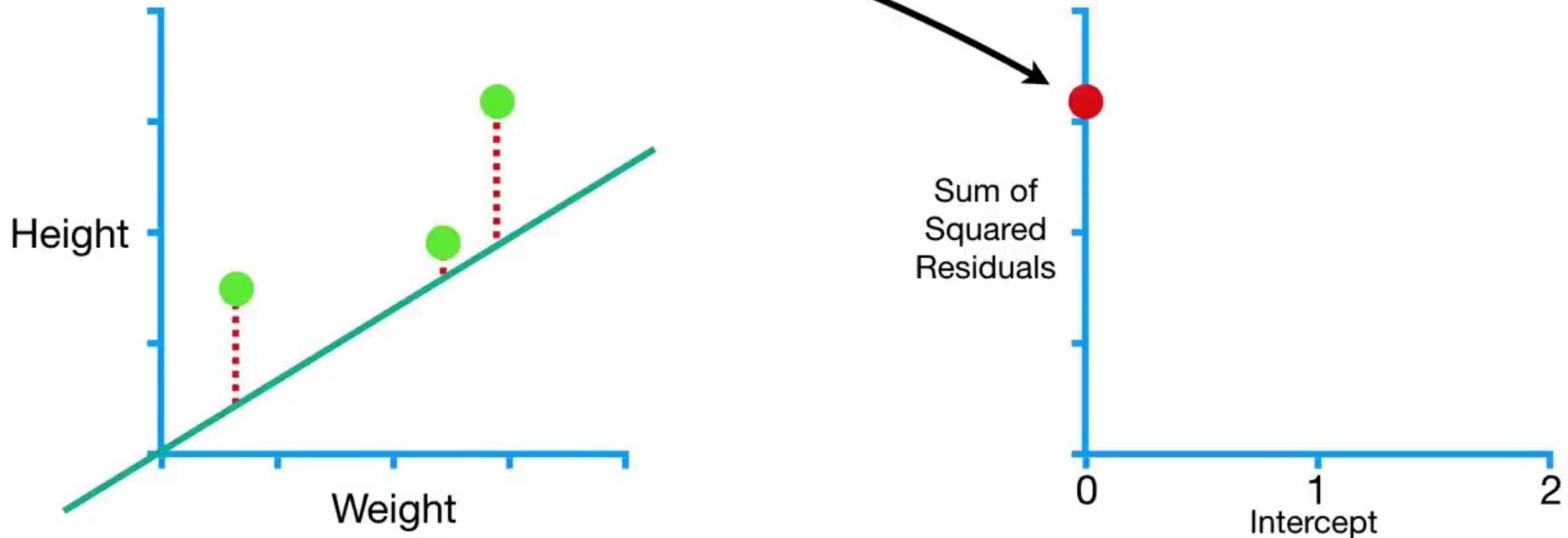


So let's get back to using **Gradient Descent** to find the optimal value for the **Intercept**, starting from a random value. In this case, the random value was **0**.

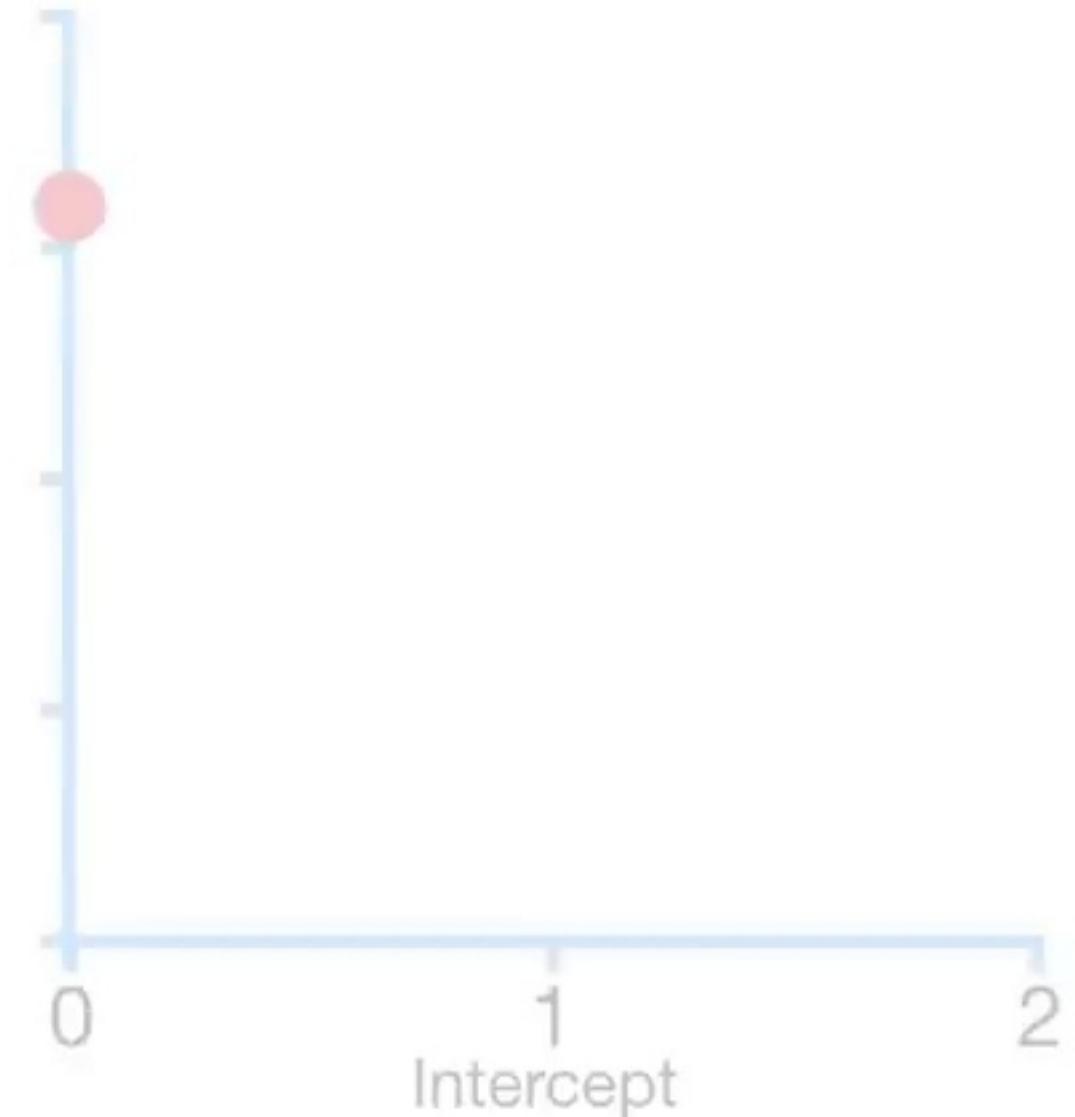
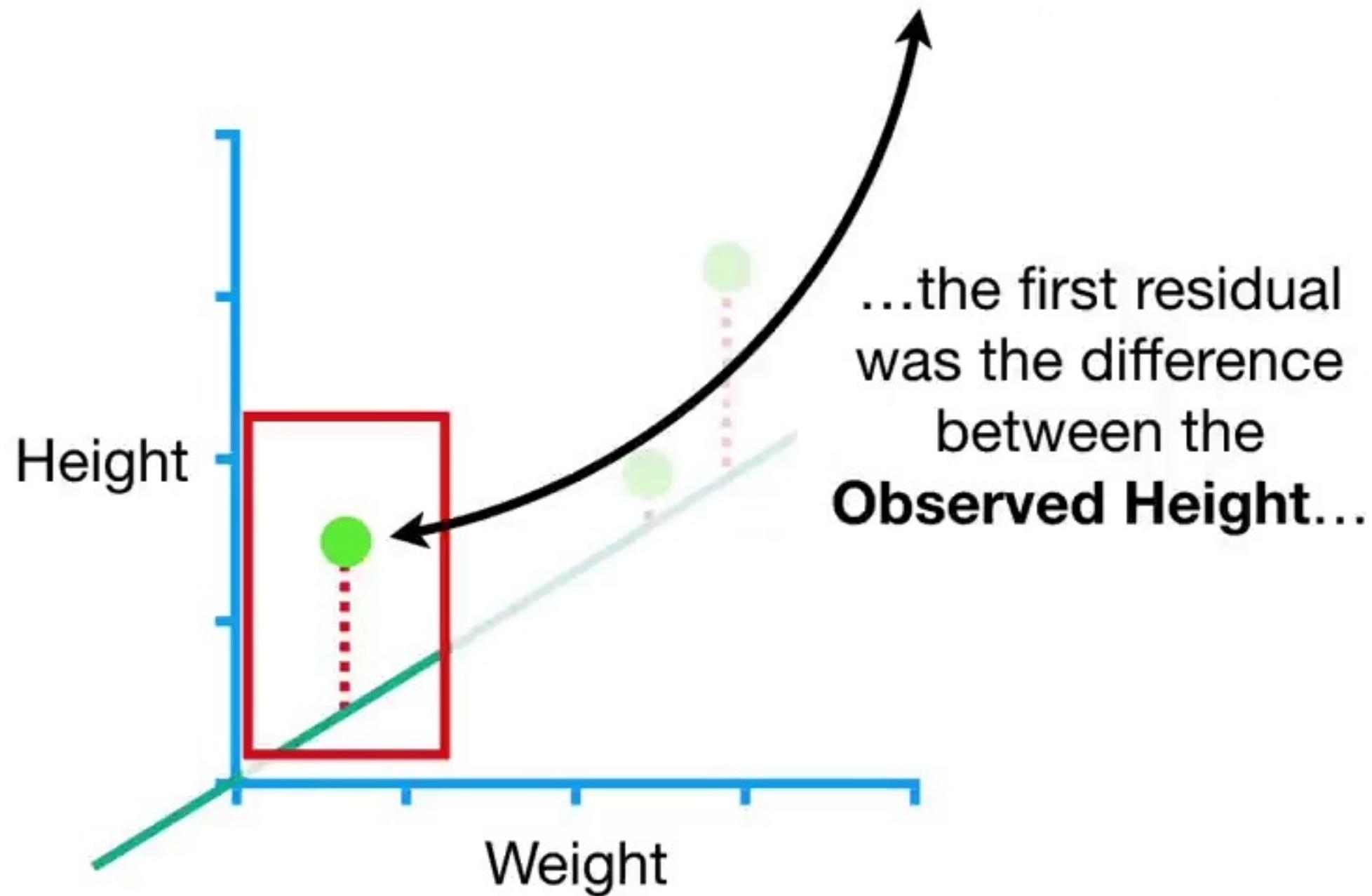


Sum of squared residuals

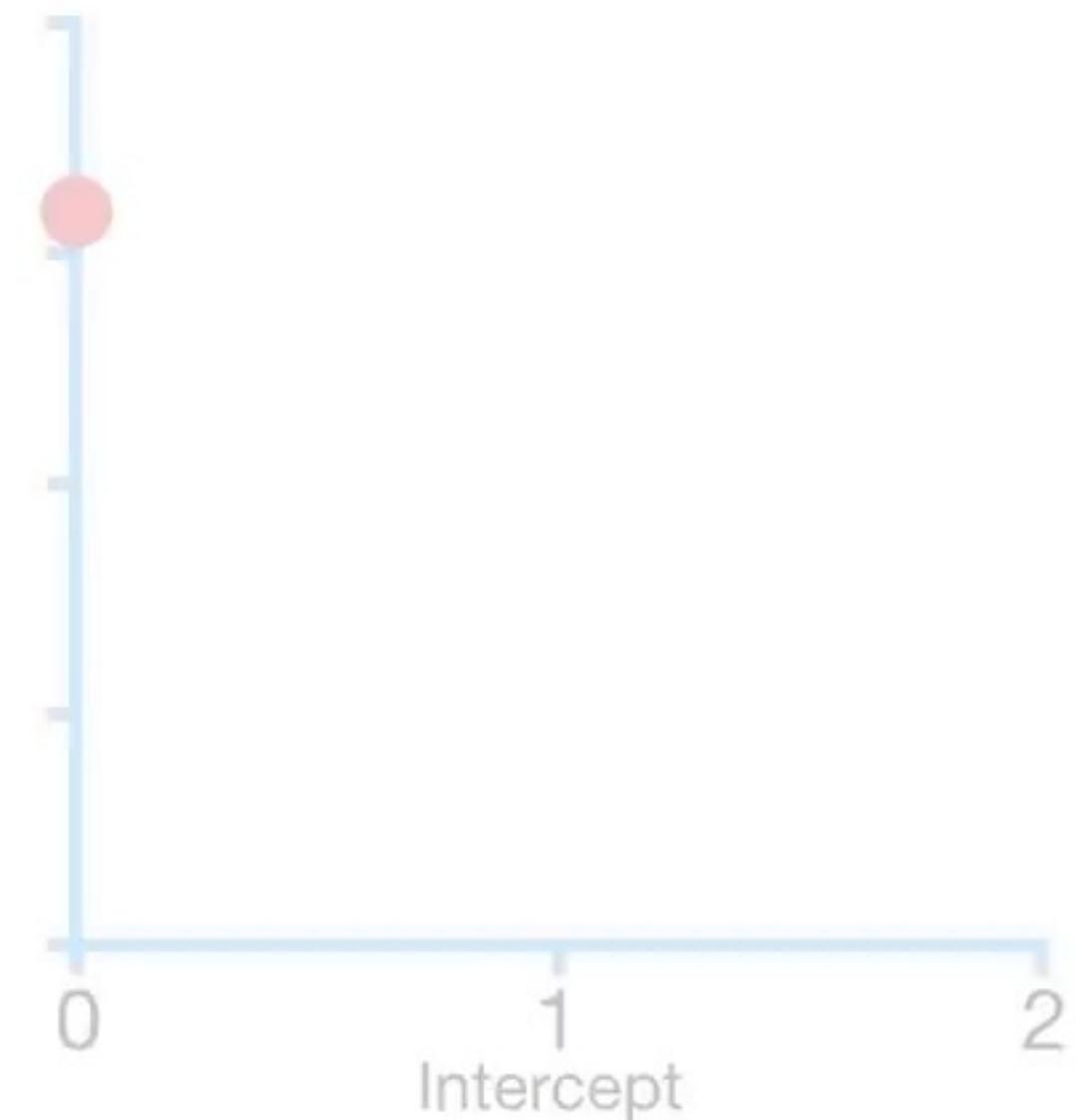
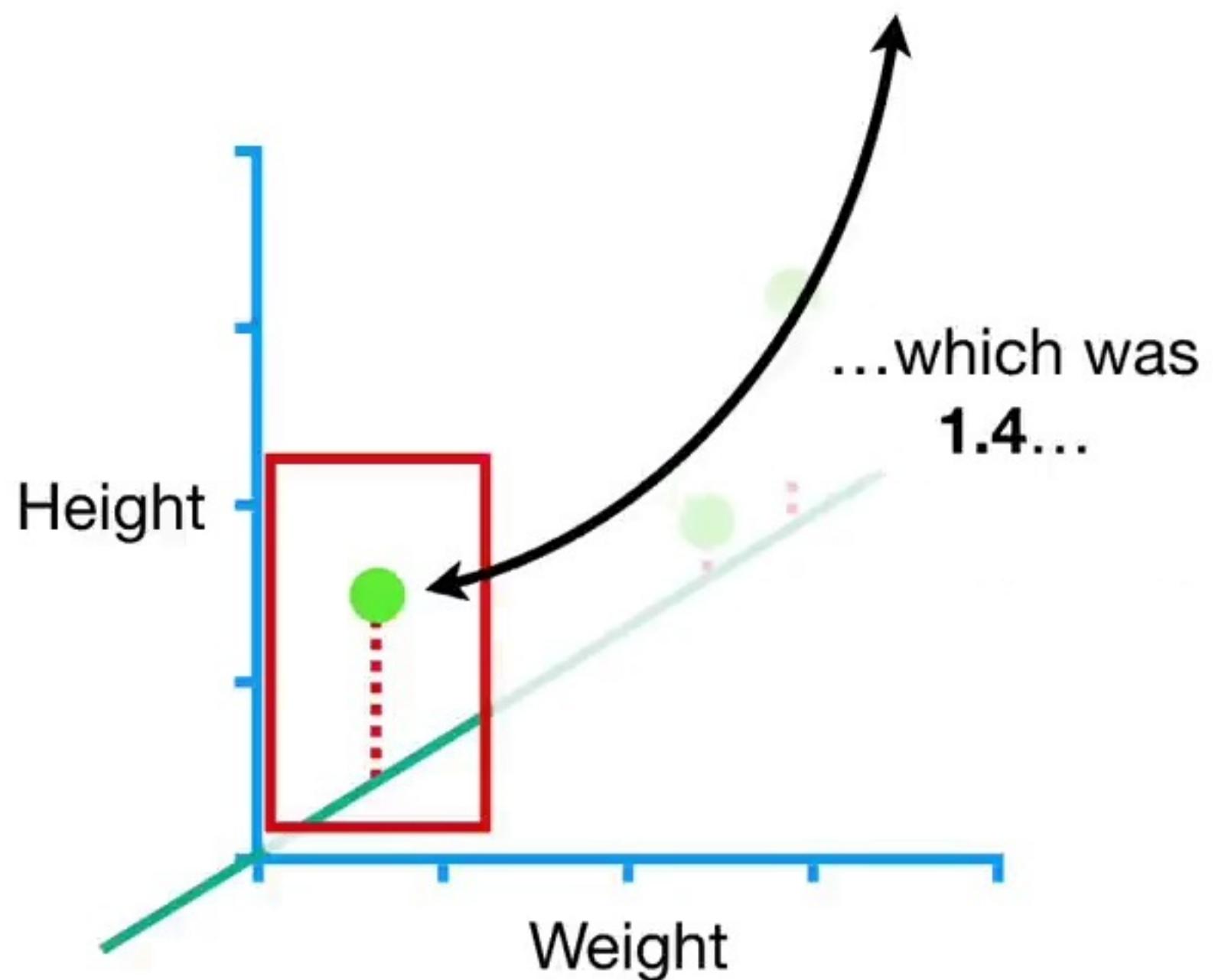
When we calculated the
Sum of the Squared
Residuals...



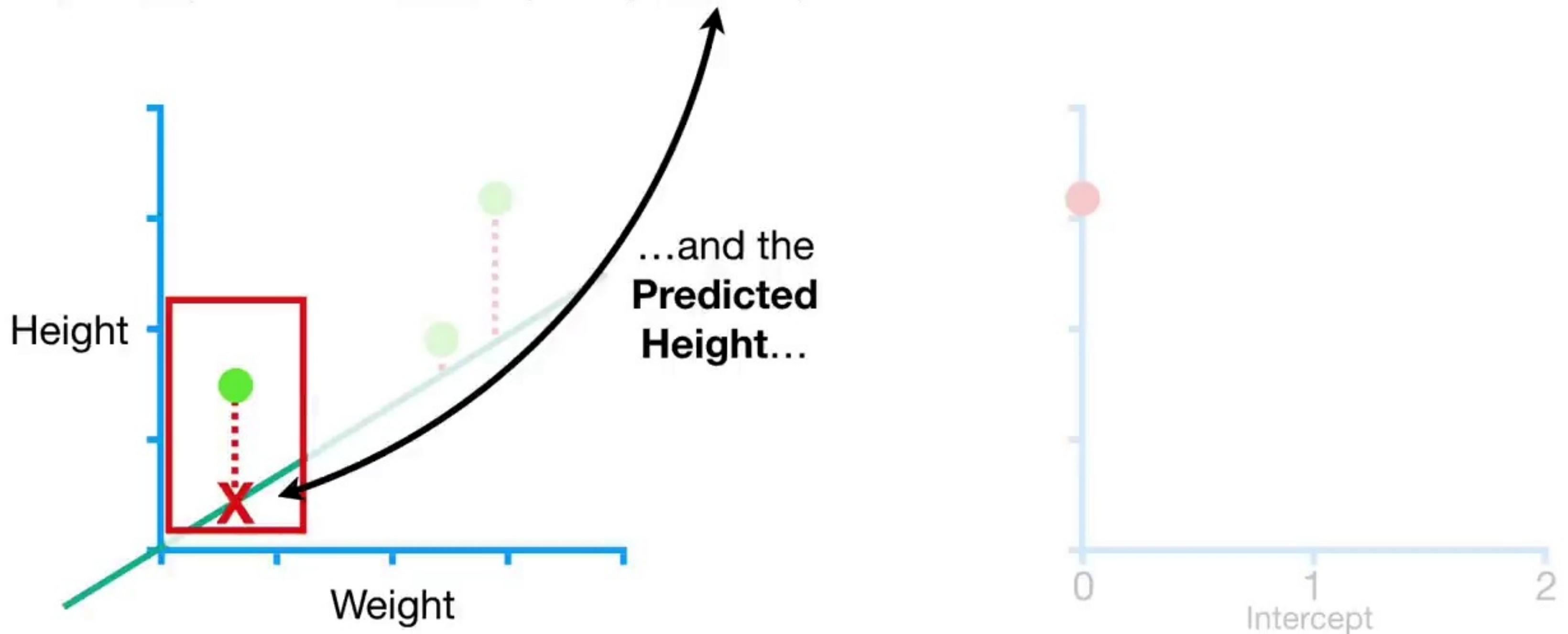
Sum of squared residuals = (observed - predicted)²



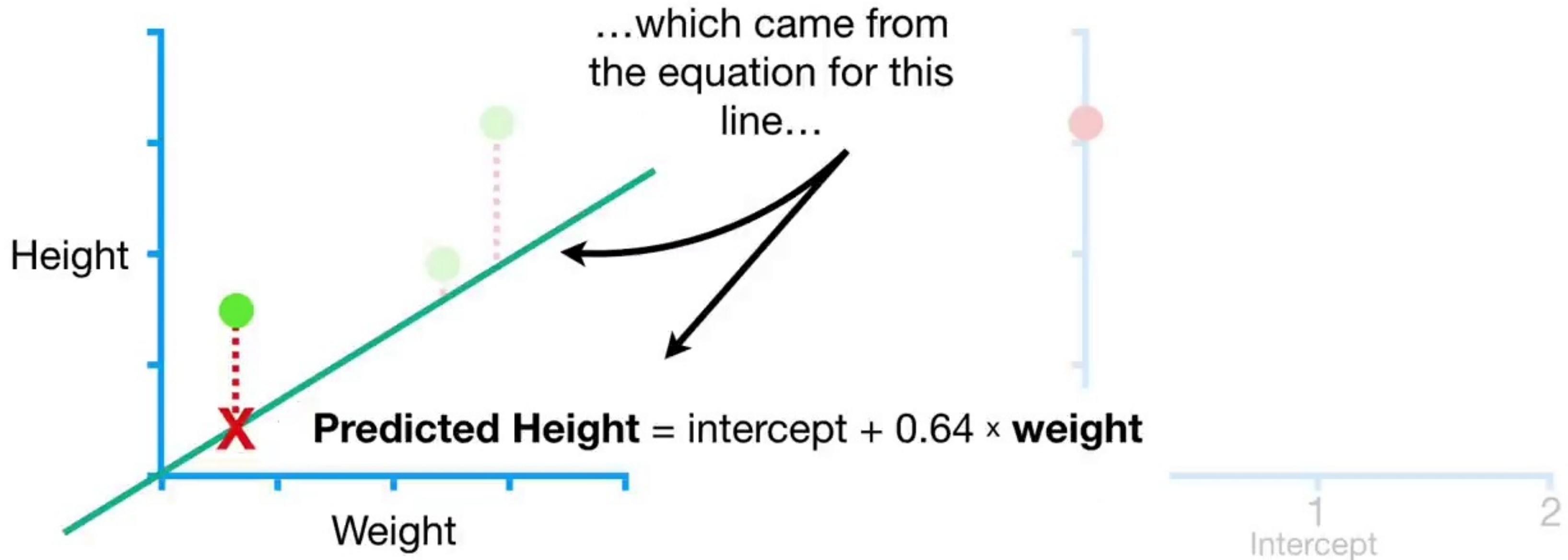
Sum of squared residuals = $(1.4 - \text{predicted})^2$



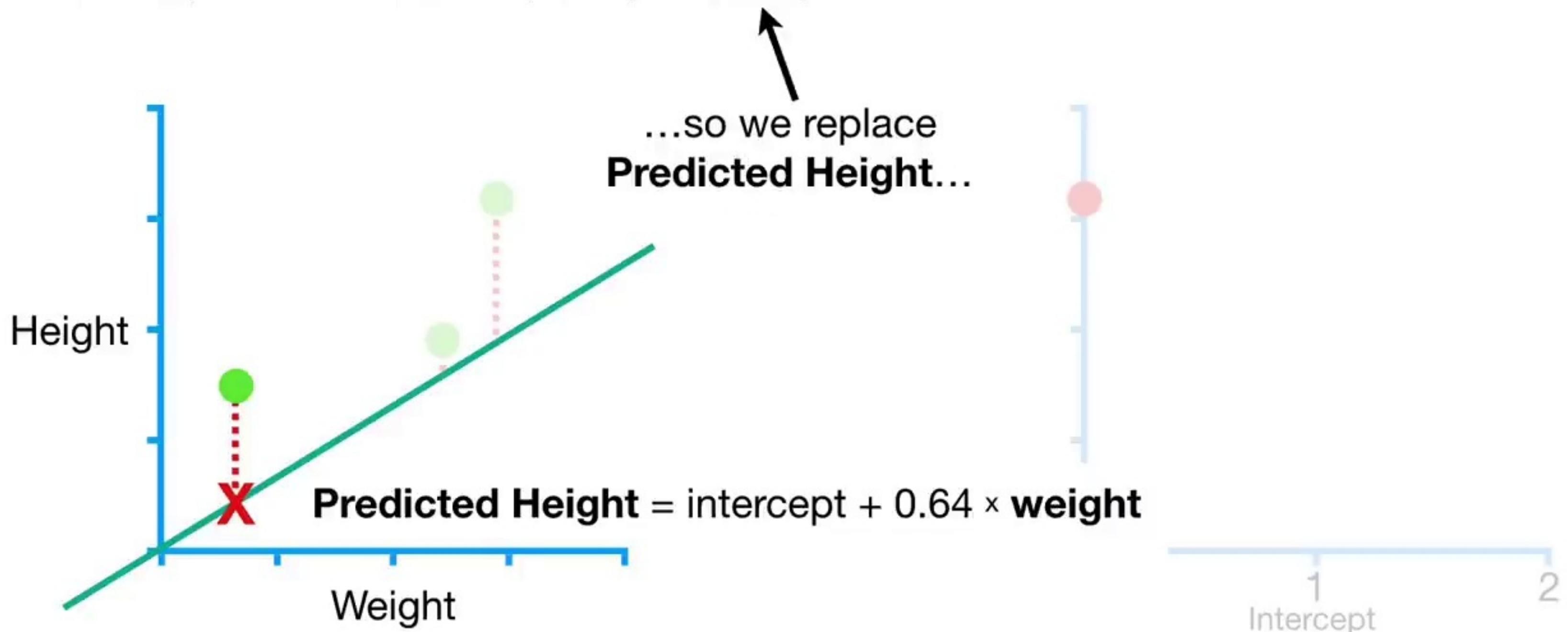
Sum of squared residuals = $(1.4 - \text{predicted})^2$



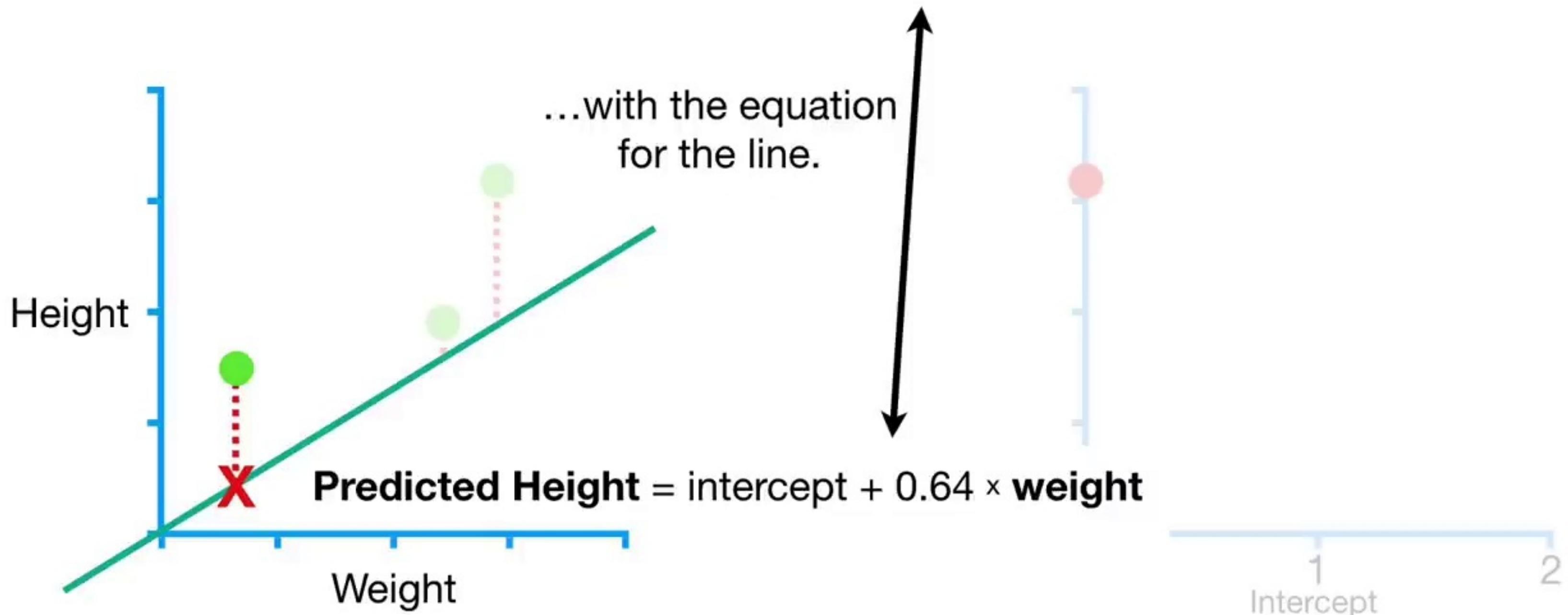
Sum of squared residuals = $(1.4 - \text{predicted})^2$



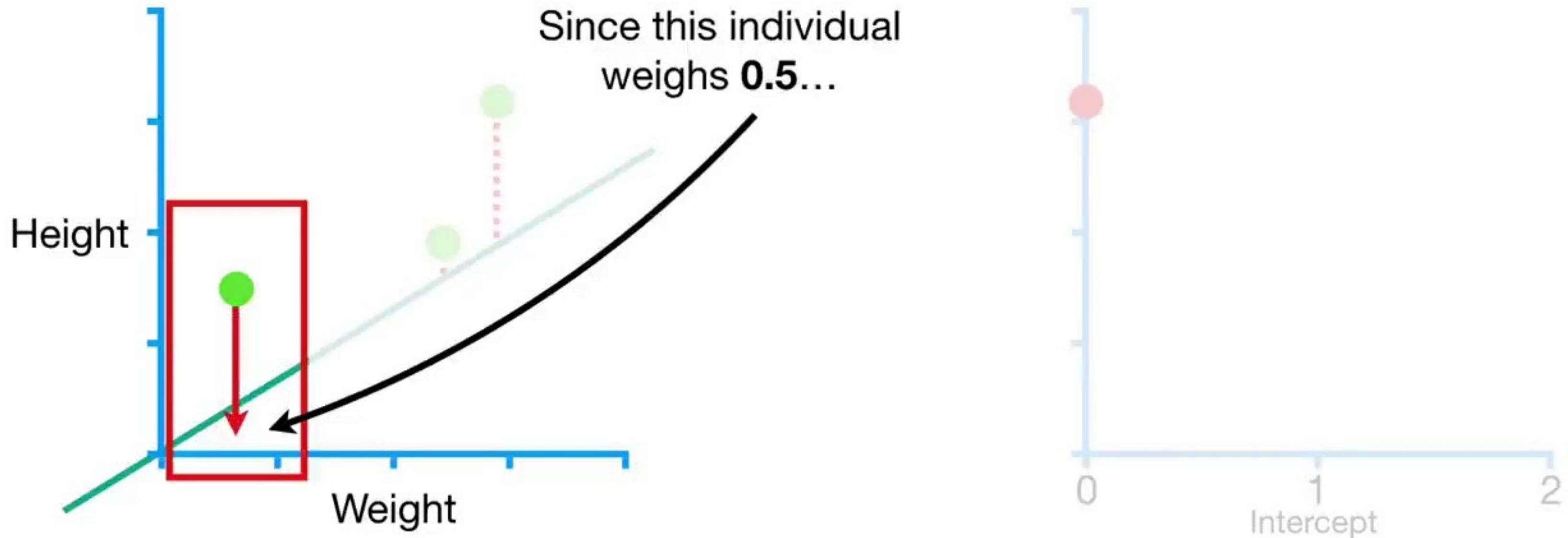
Sum of squared residuals = $(1.4 - \text{predicted})^2$



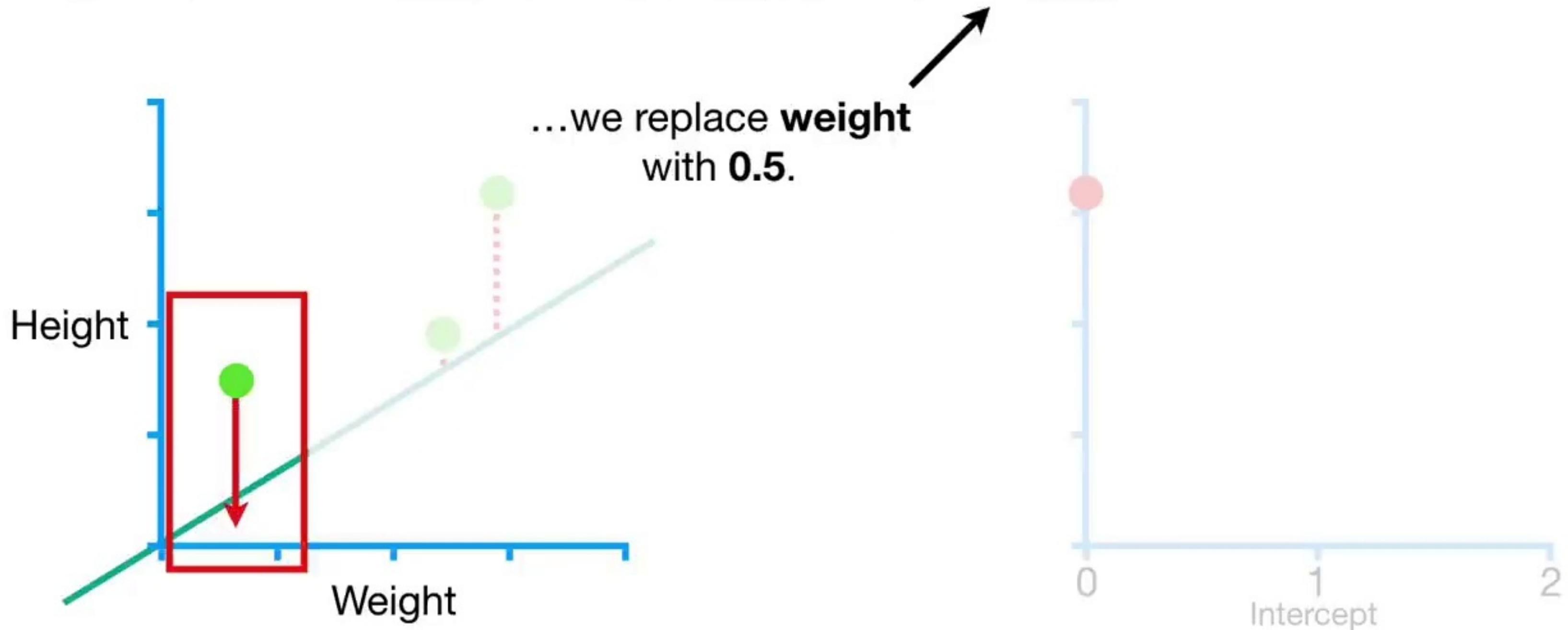
Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times \text{weight}))^2$



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times \text{weight}))^2$

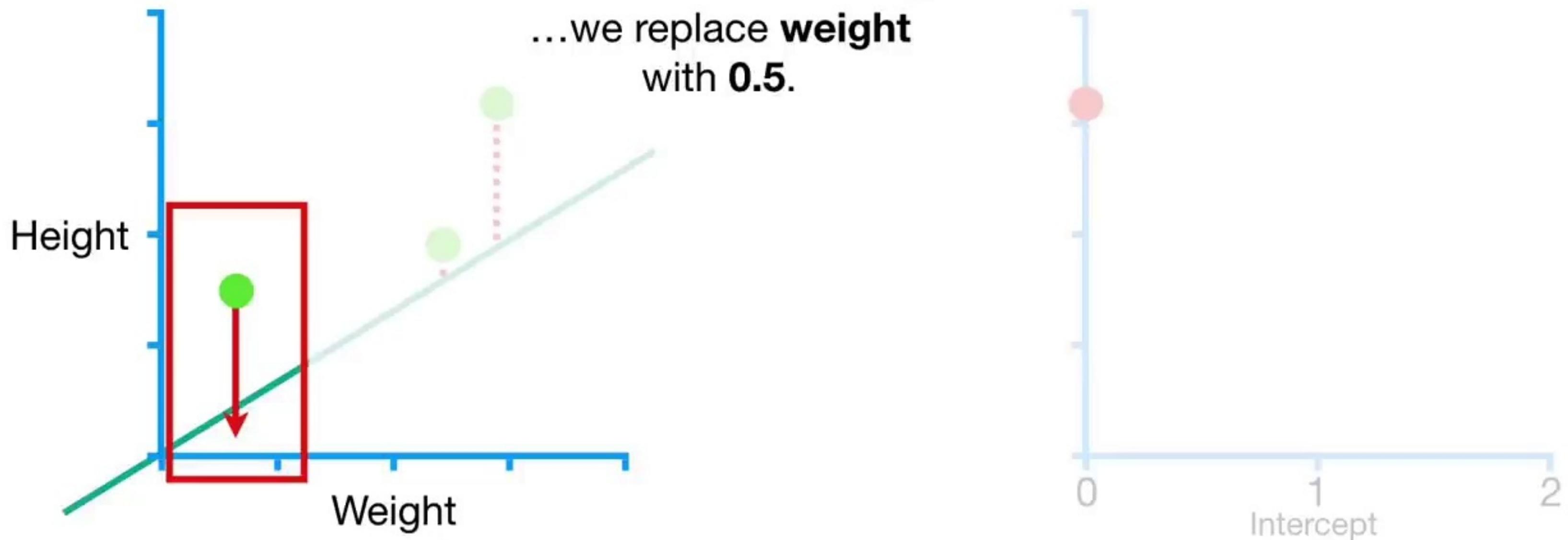


Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times \text{weight}))^2$

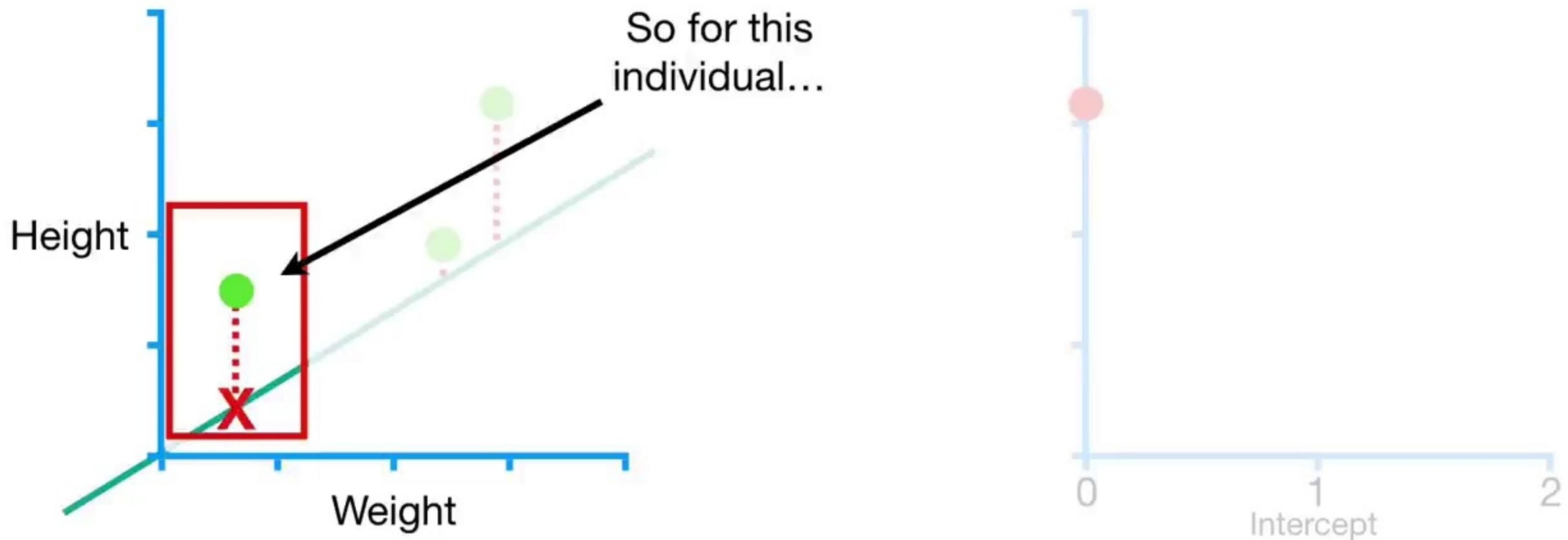


Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

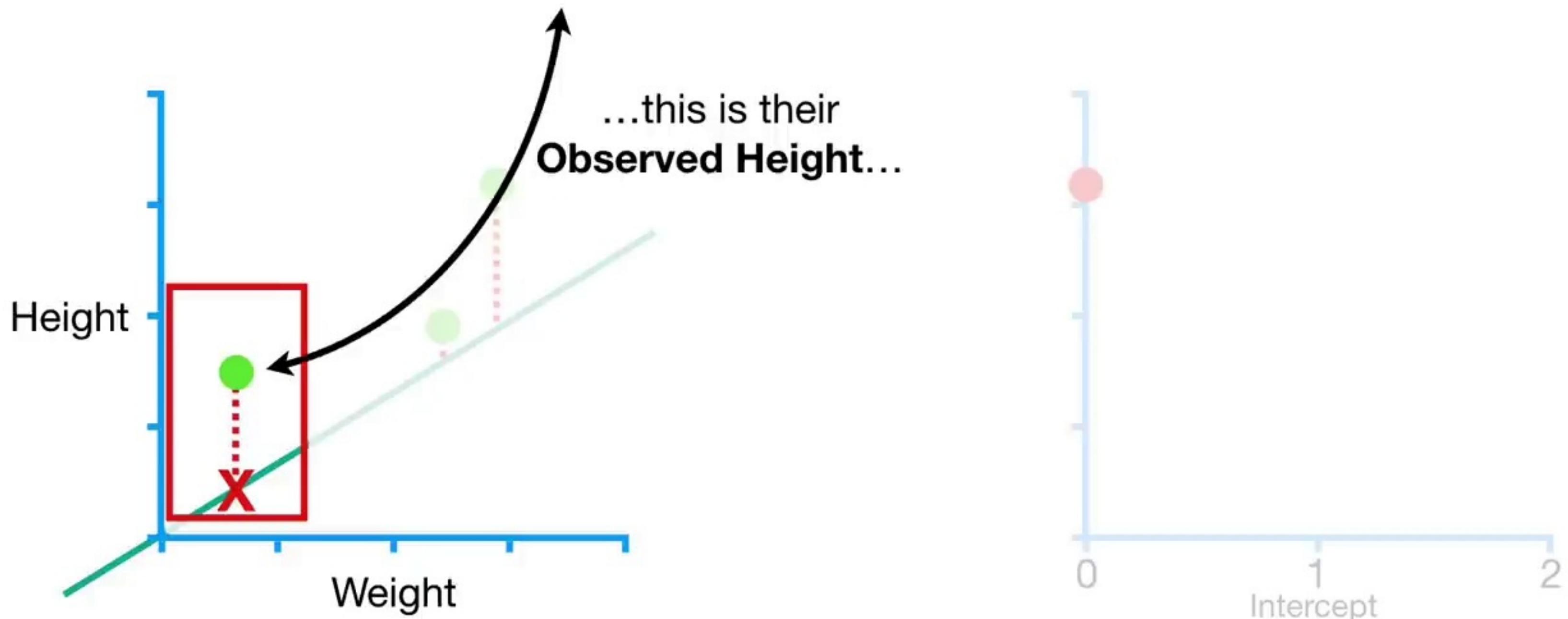
...we replace **weight**
with **0.5**.



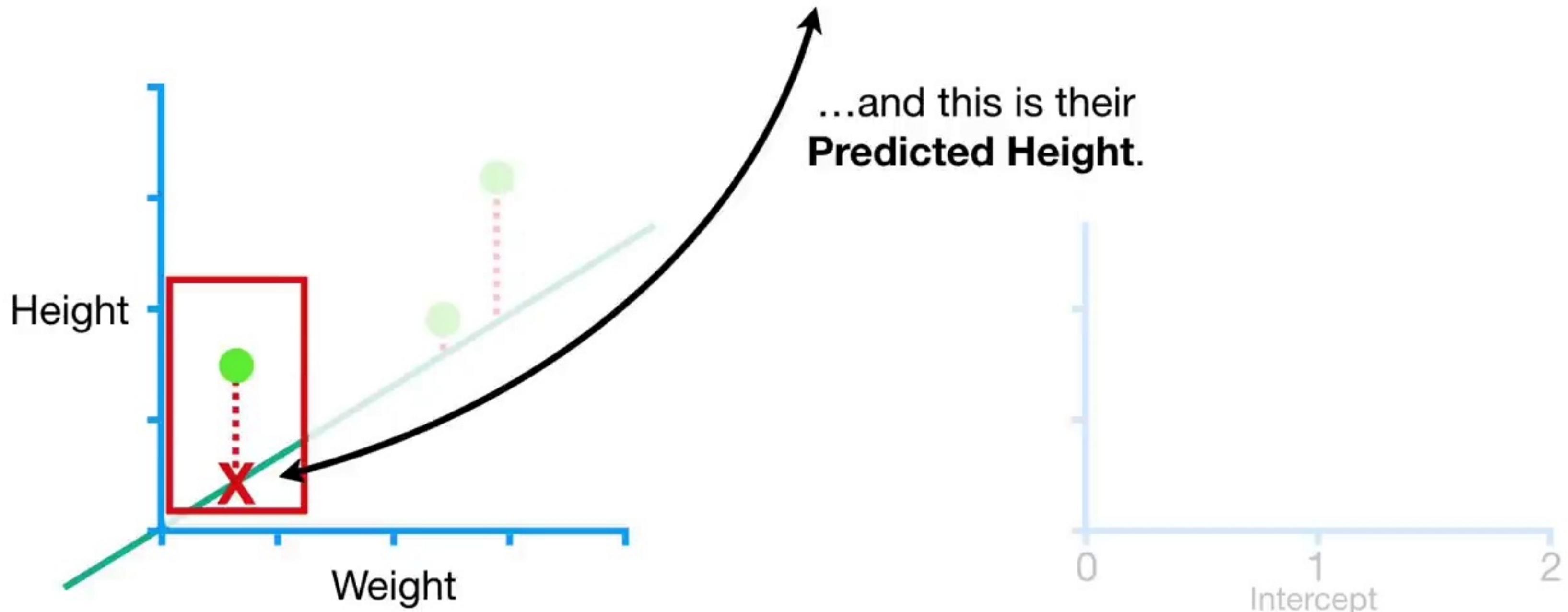
$$\text{Sum of squared residuals} = (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$



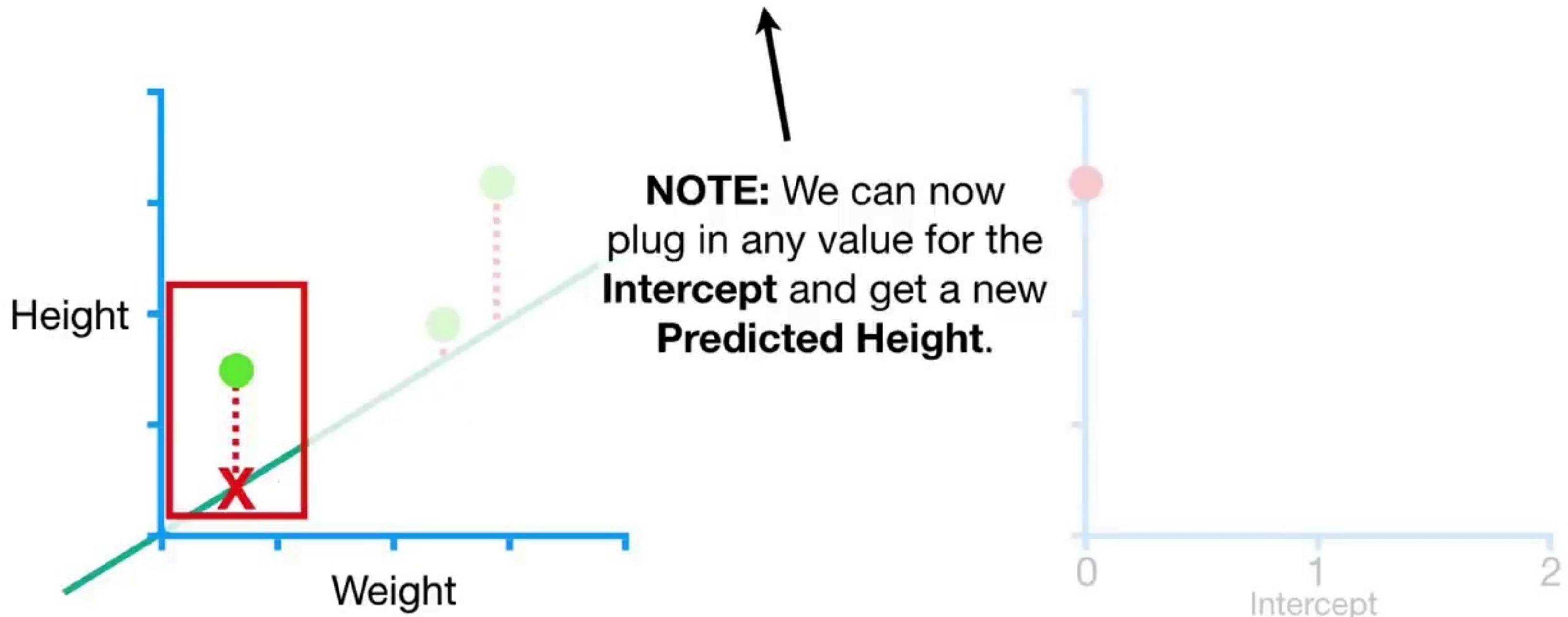
Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$



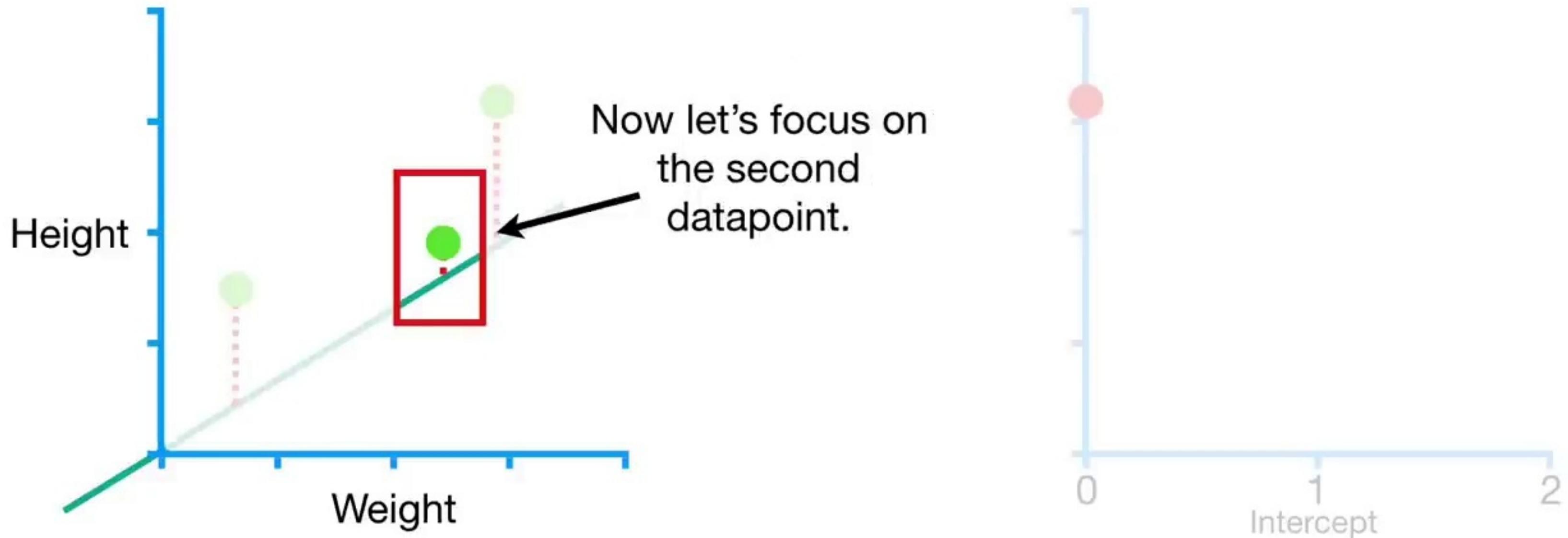
Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$



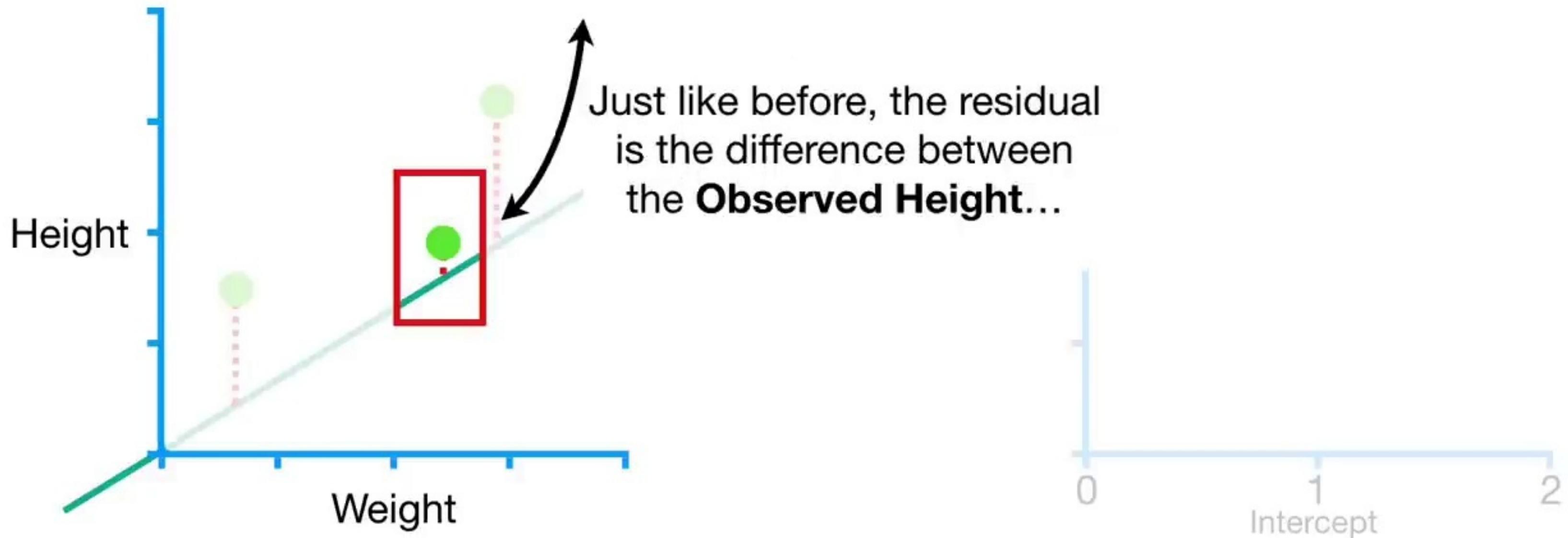
Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$



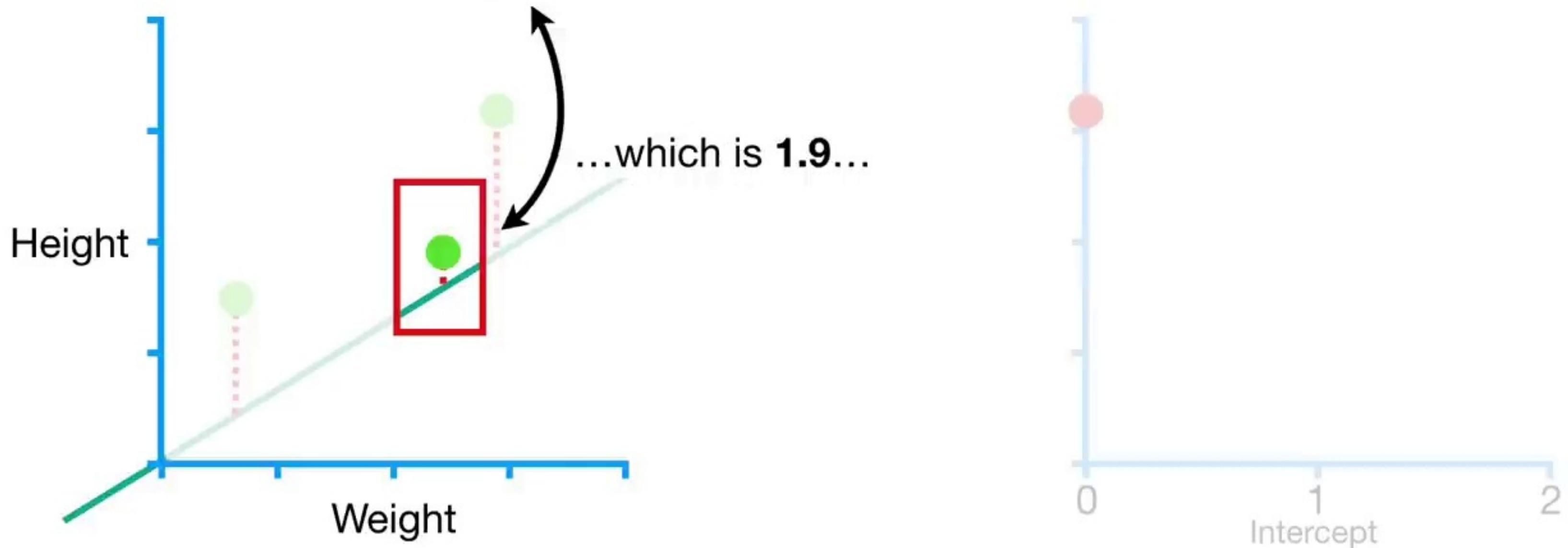
$$\text{Sum of squared residuals} = (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$



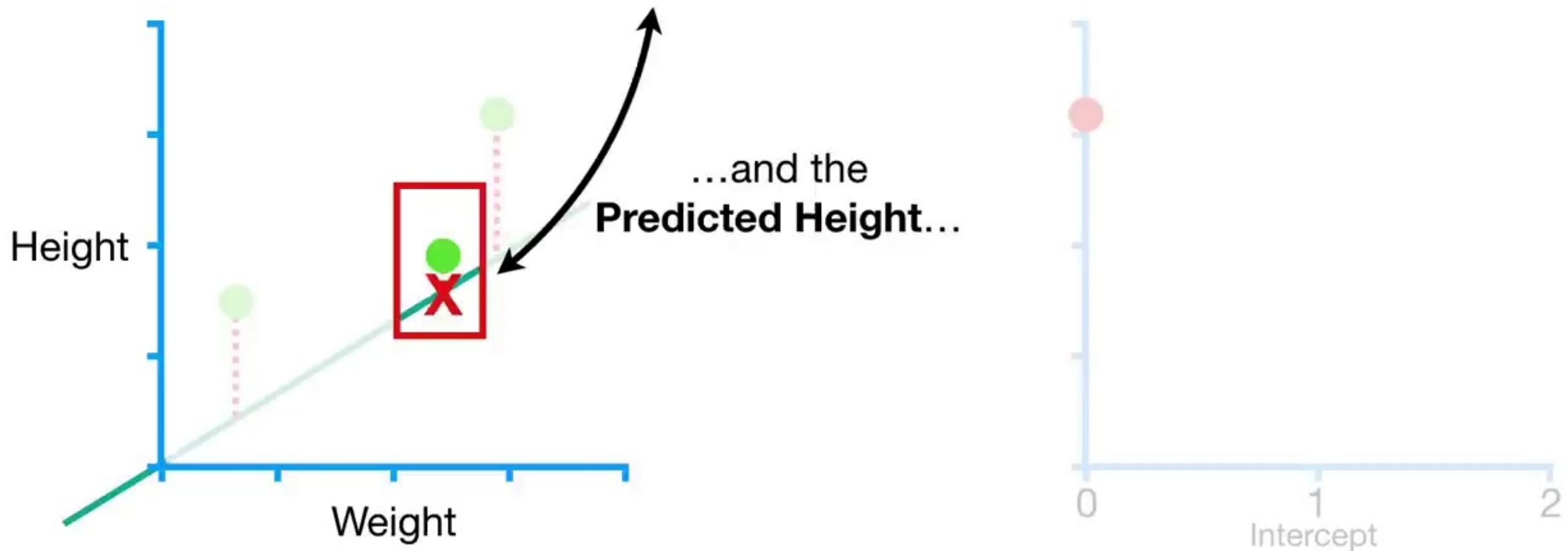
$$\begin{aligned}\text{Sum of squared residuals} = & (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ & + (\text{observed} - \text{predicted})^2\end{aligned}$$



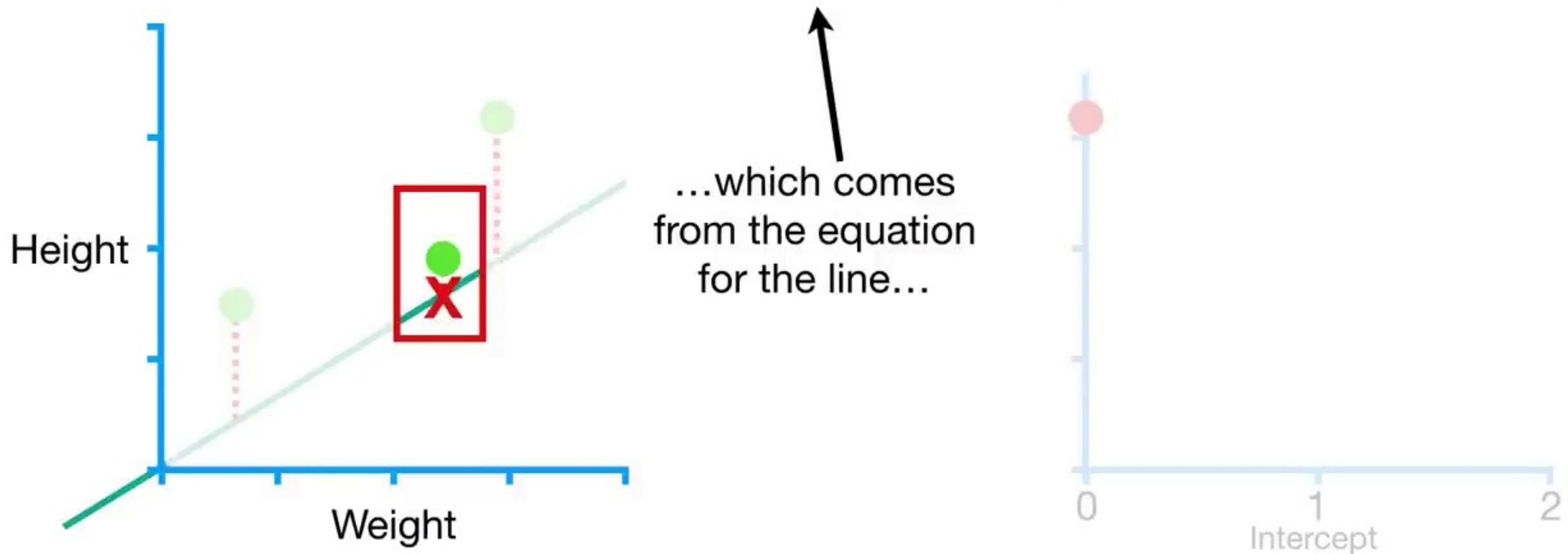
Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$
+ $(1.9 - \text{predicted})^2$



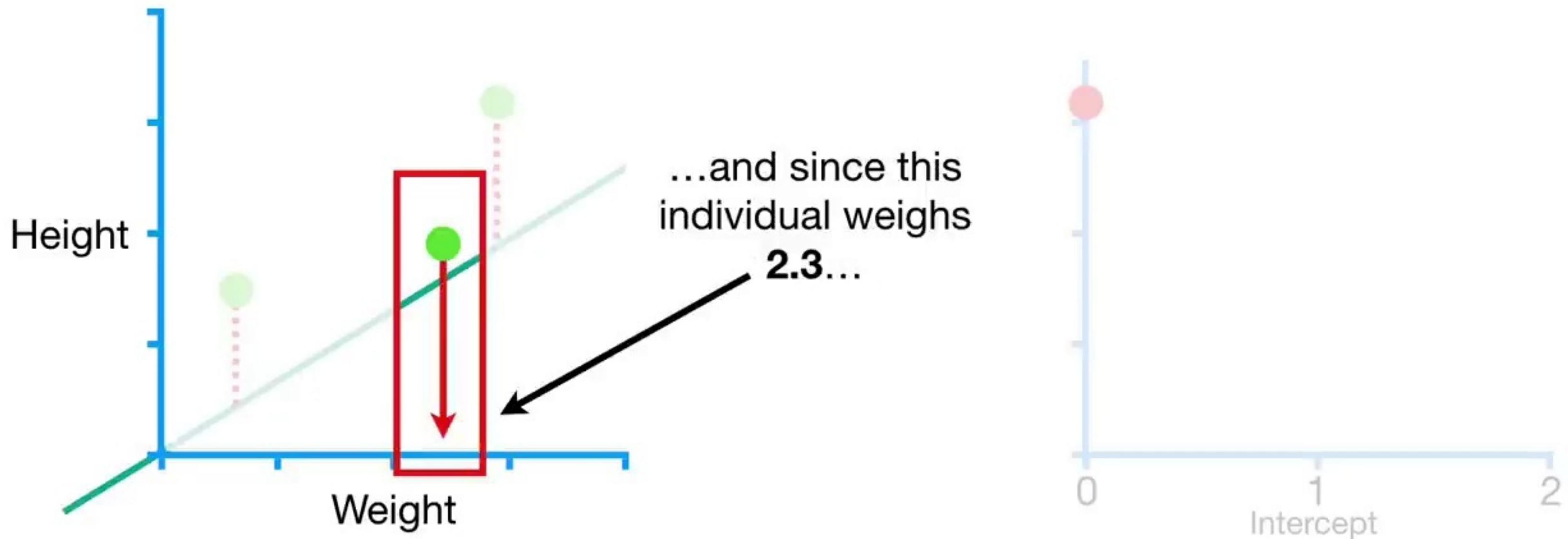
Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$
+ $(1.9 - \text{predicted})^2$



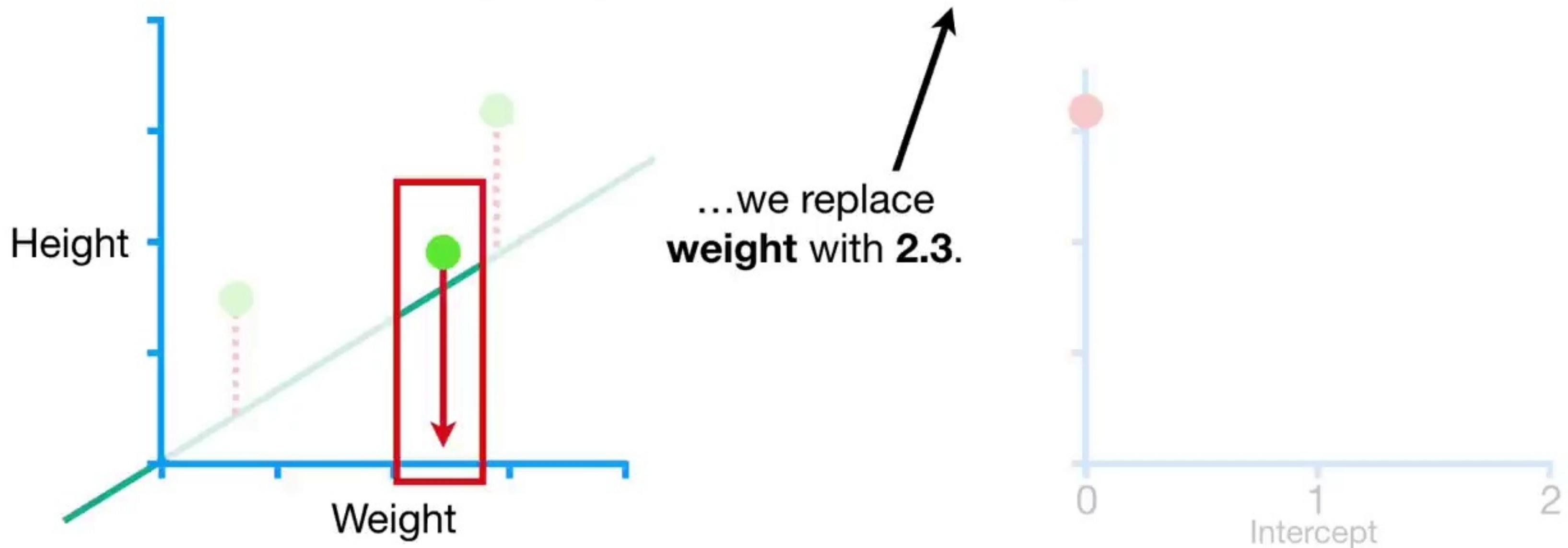
$$\begin{aligned}\text{Sum of squared residuals} = & (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ & + (1.9 - (\text{intercept} + 0.64 \times \text{weight}))^2\end{aligned}$$



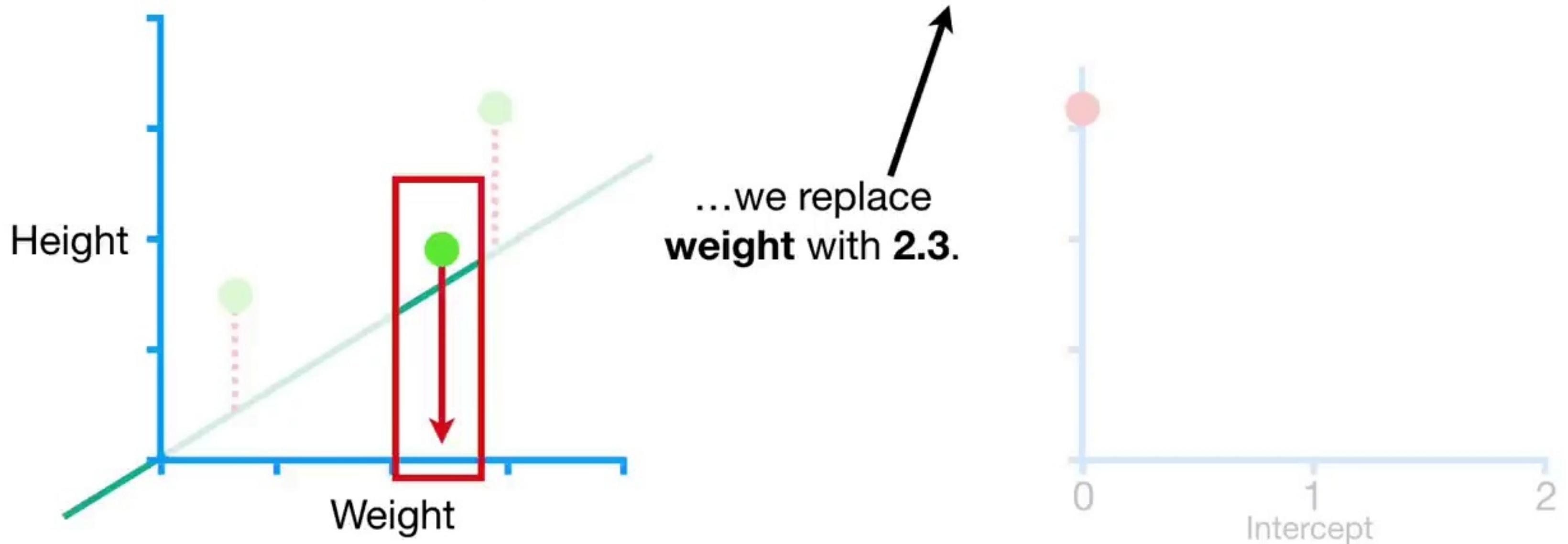
$$\begin{aligned}\text{Sum of squared residuals} = & (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ & + (1.9 - (\text{intercept} + 0.64 \times \text{weight}))^2\end{aligned}$$



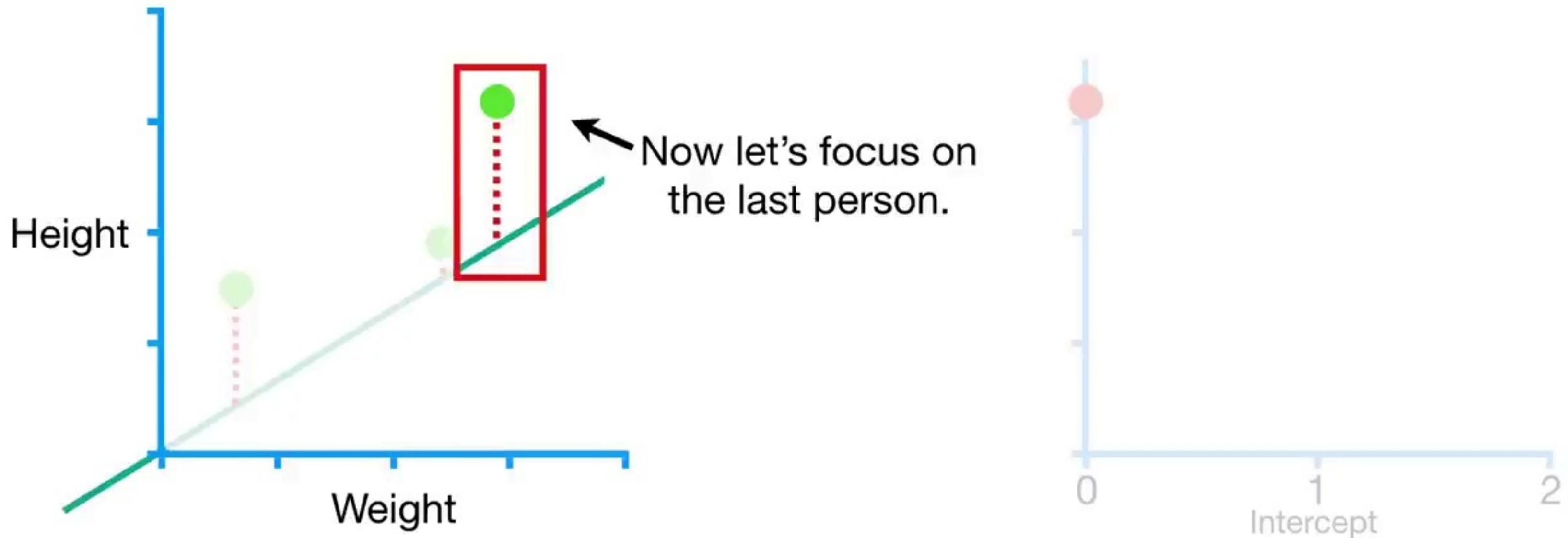
$$\begin{aligned}\text{Sum of squared residuals} = & (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ & + (1.9 - (\text{intercept} + 0.64 \times \text{weight}))^2\end{aligned}$$



$$\begin{aligned}\text{Sum of squared residuals} = & (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ & + (1.9 - (\text{intercept} + 0.64 \times 2.3))^2\end{aligned}$$



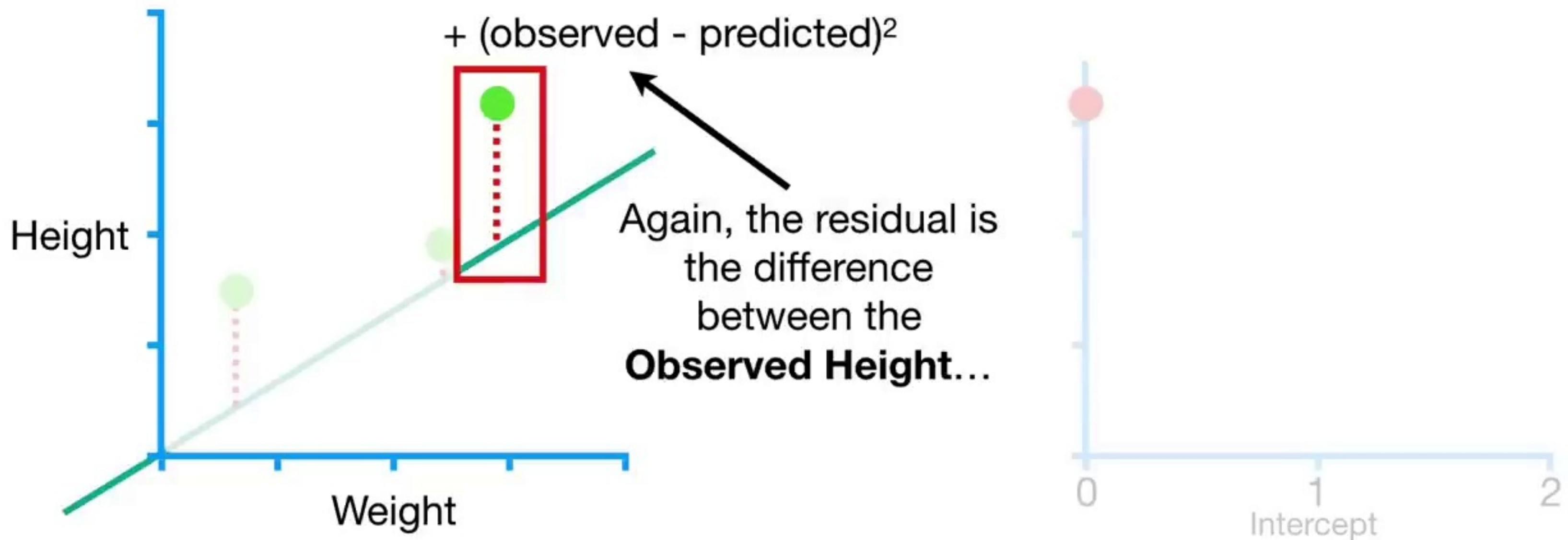
$$\begin{aligned}\text{Sum of squared residuals} = & (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ & + (1.9 - (\text{intercept} + 0.64 \times 2.3))^2\end{aligned}$$



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$

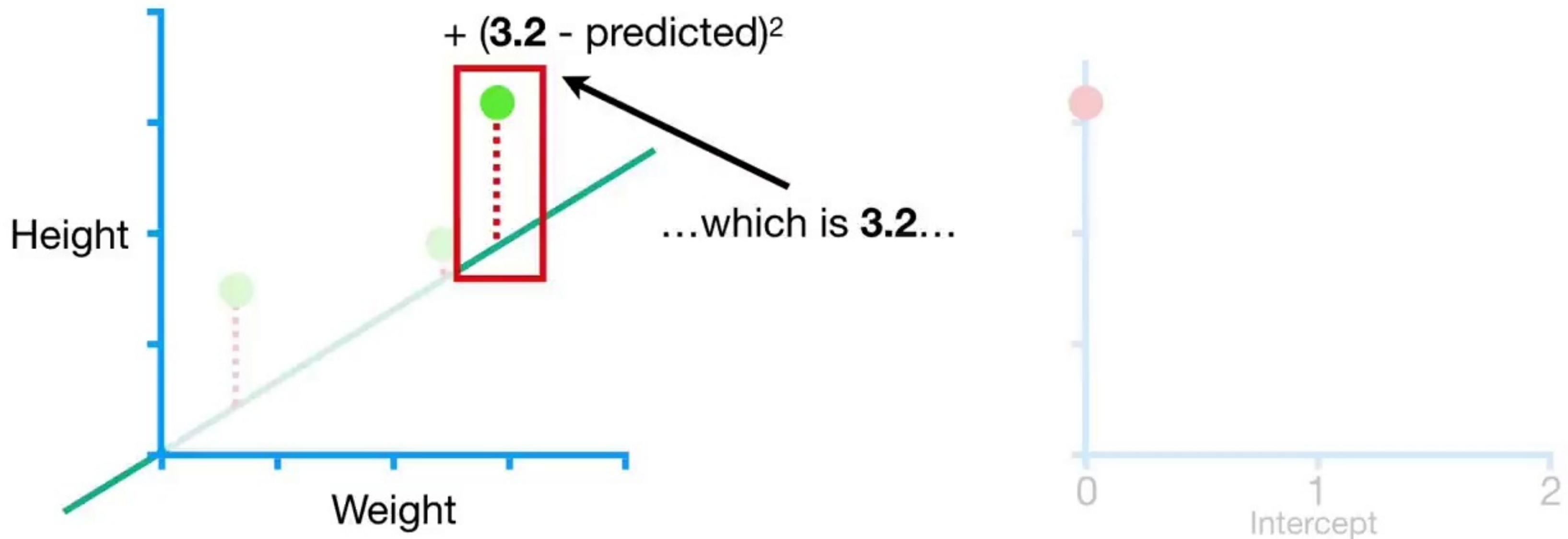
+ $(\text{observed} - \text{predicted})^2$



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$

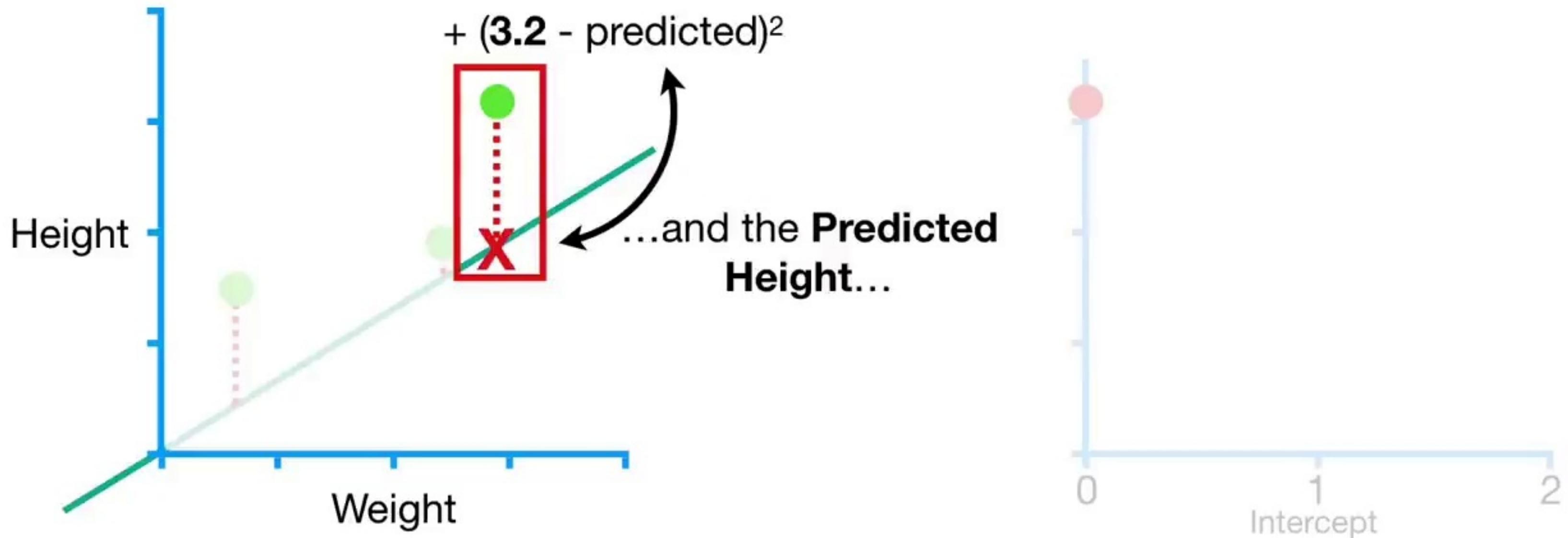
+ $(3.2 - \text{predicted})^2$



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$

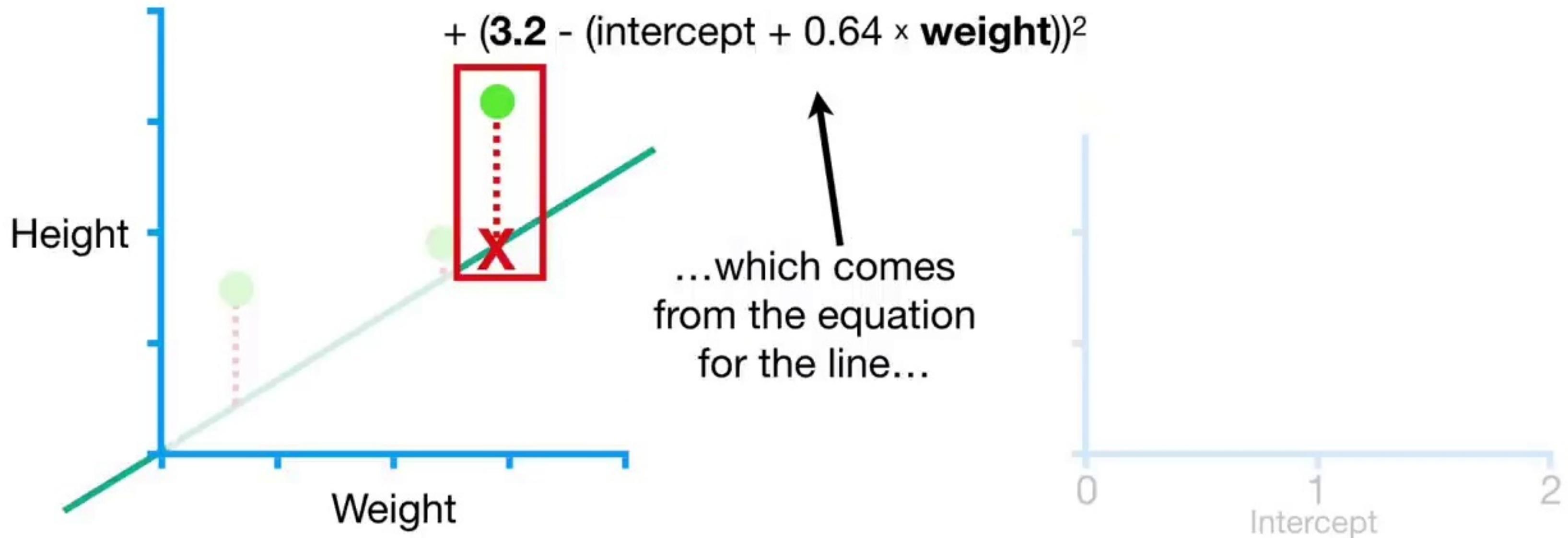
+ $(3.2 - \text{predicted})^2$



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$

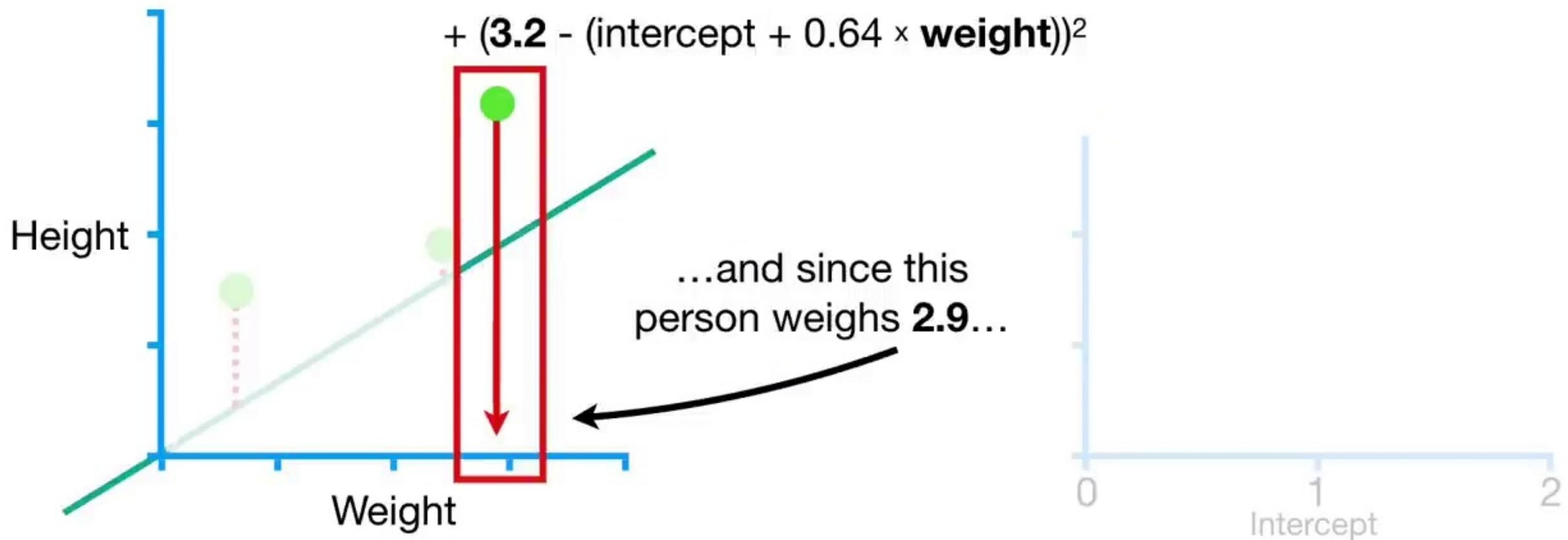
+ $(3.2 - (\text{intercept} + 0.64 \times \text{weight}))^2$



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$

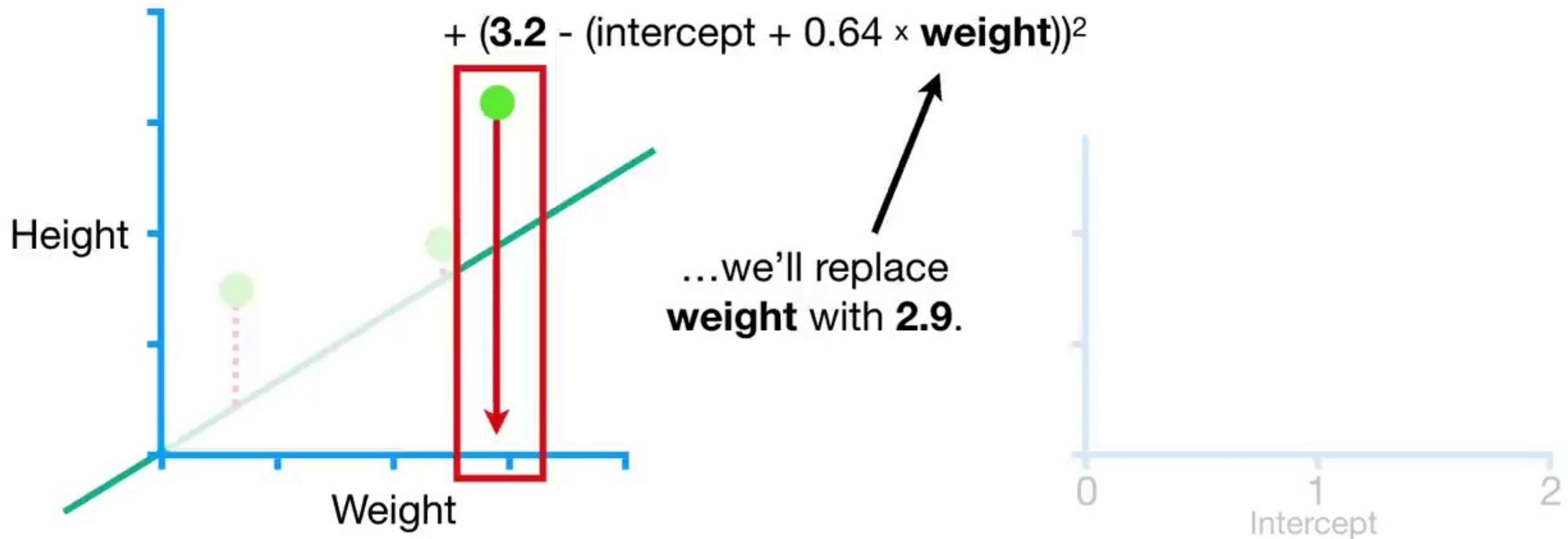
+ $(3.2 - (\text{intercept} + 0.64 \times \text{weight}))^2$



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$

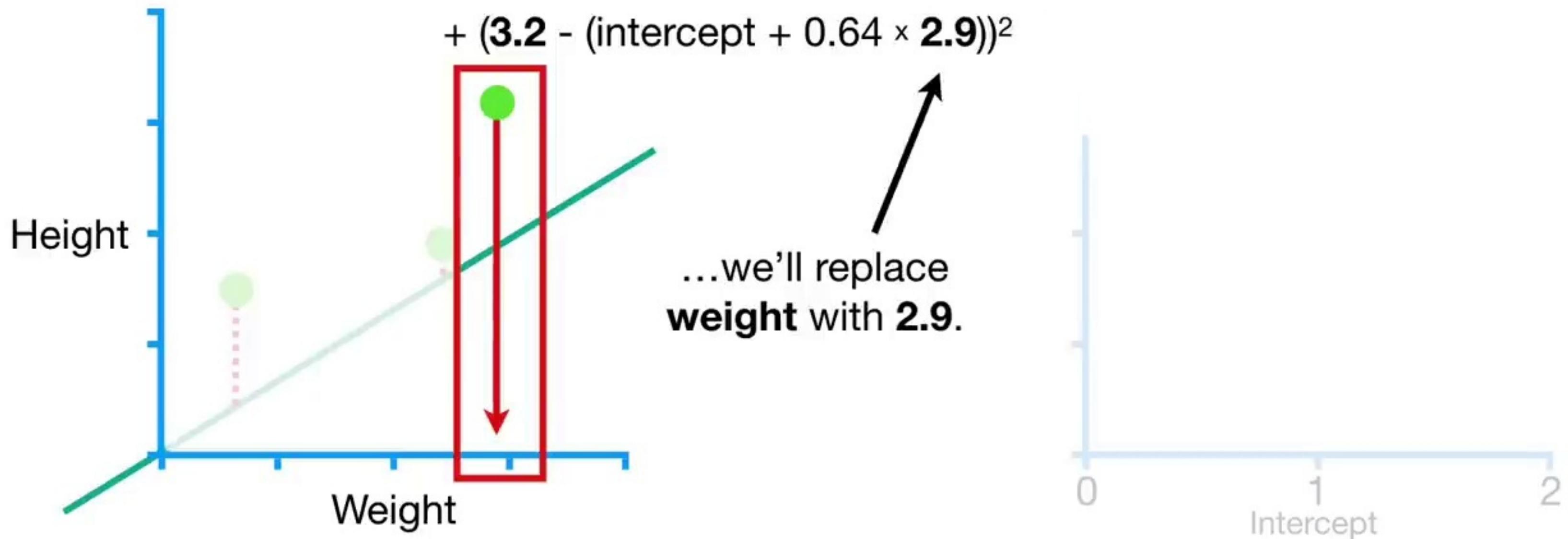
+ $(3.2 - (\text{intercept} + 0.64 \times \text{weight}))^2$



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$

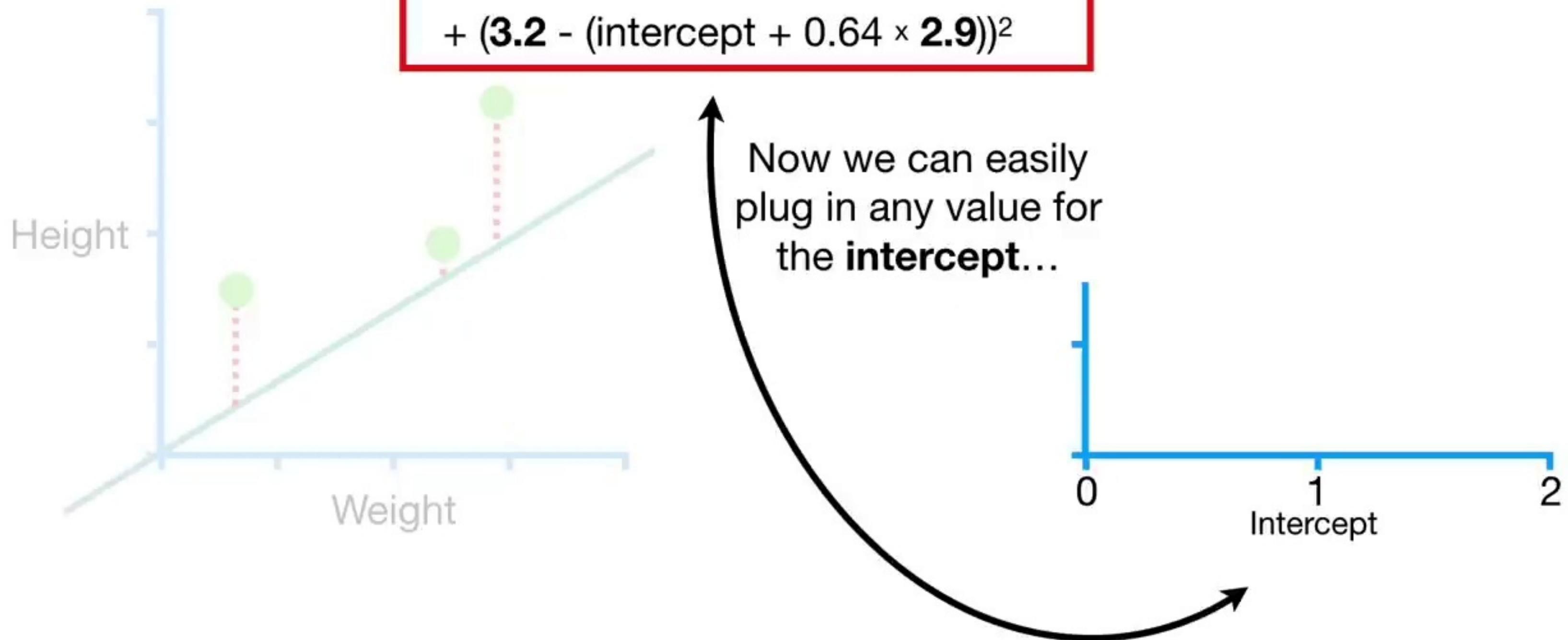
+ $(3.2 - (\text{intercept} + 0.64 \times 2.9))^2$



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$

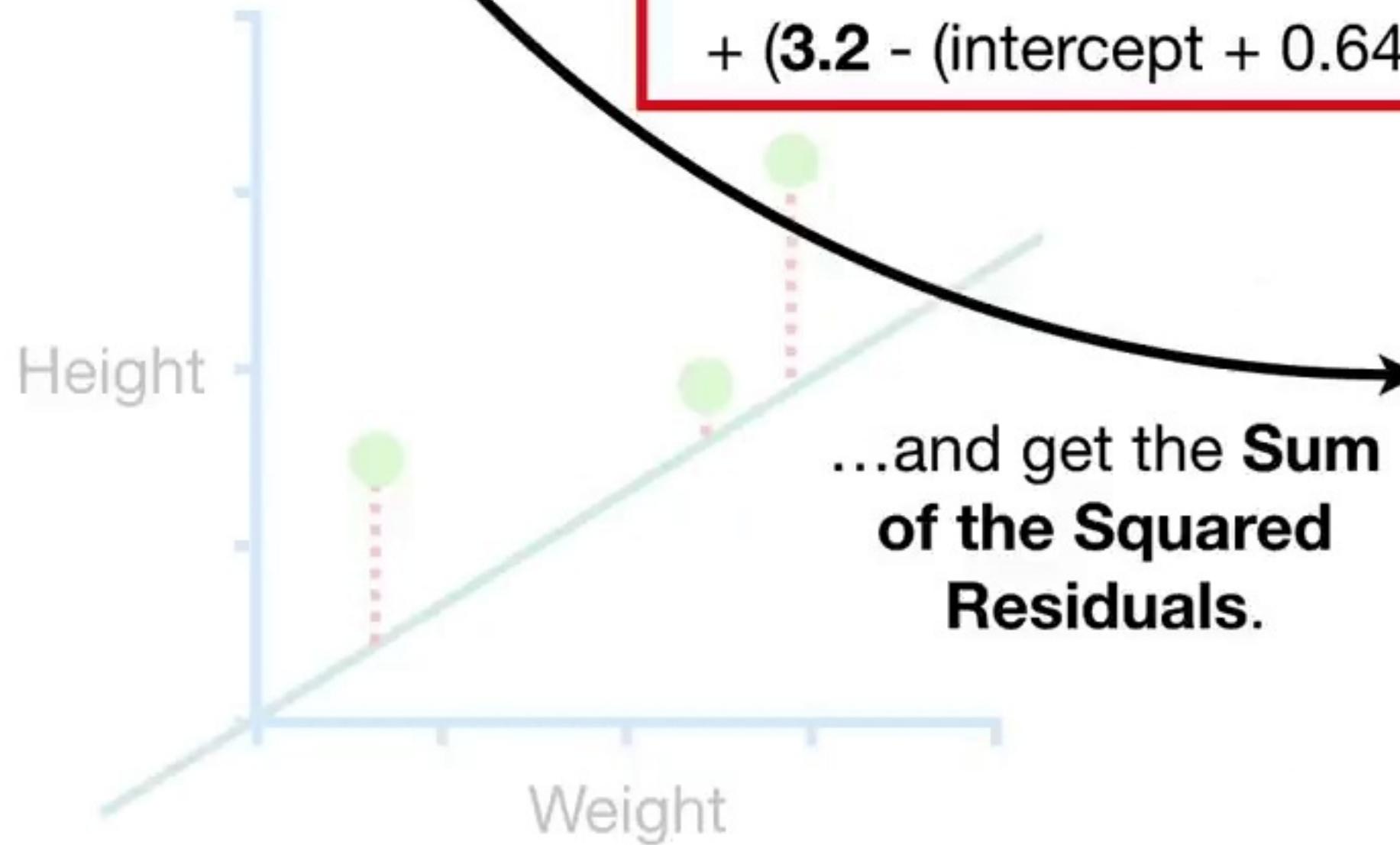
+ $(3.2 - (\text{intercept} + 0.64 \times 2.9))^2$



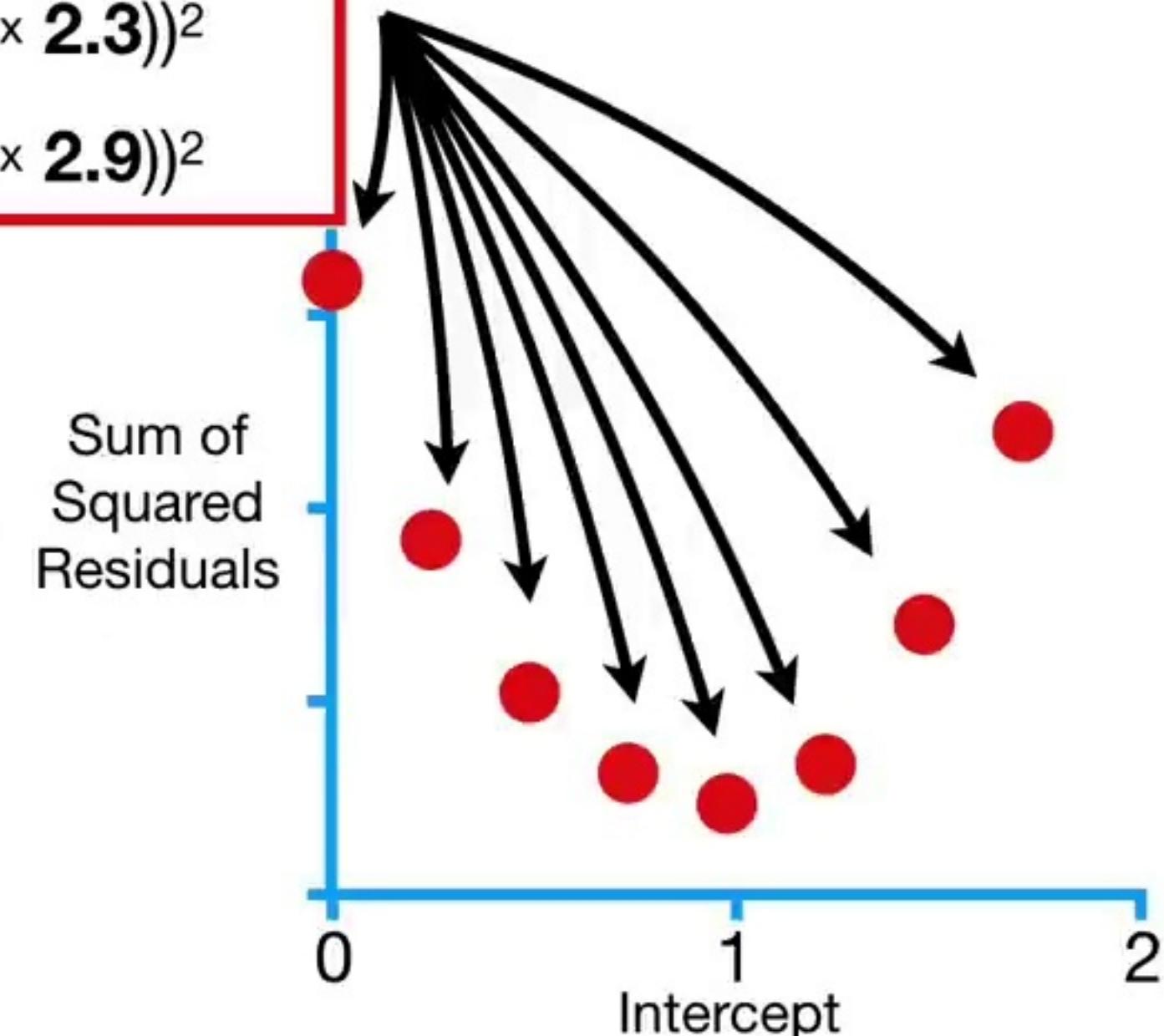
Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

$$+ (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$+ (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$



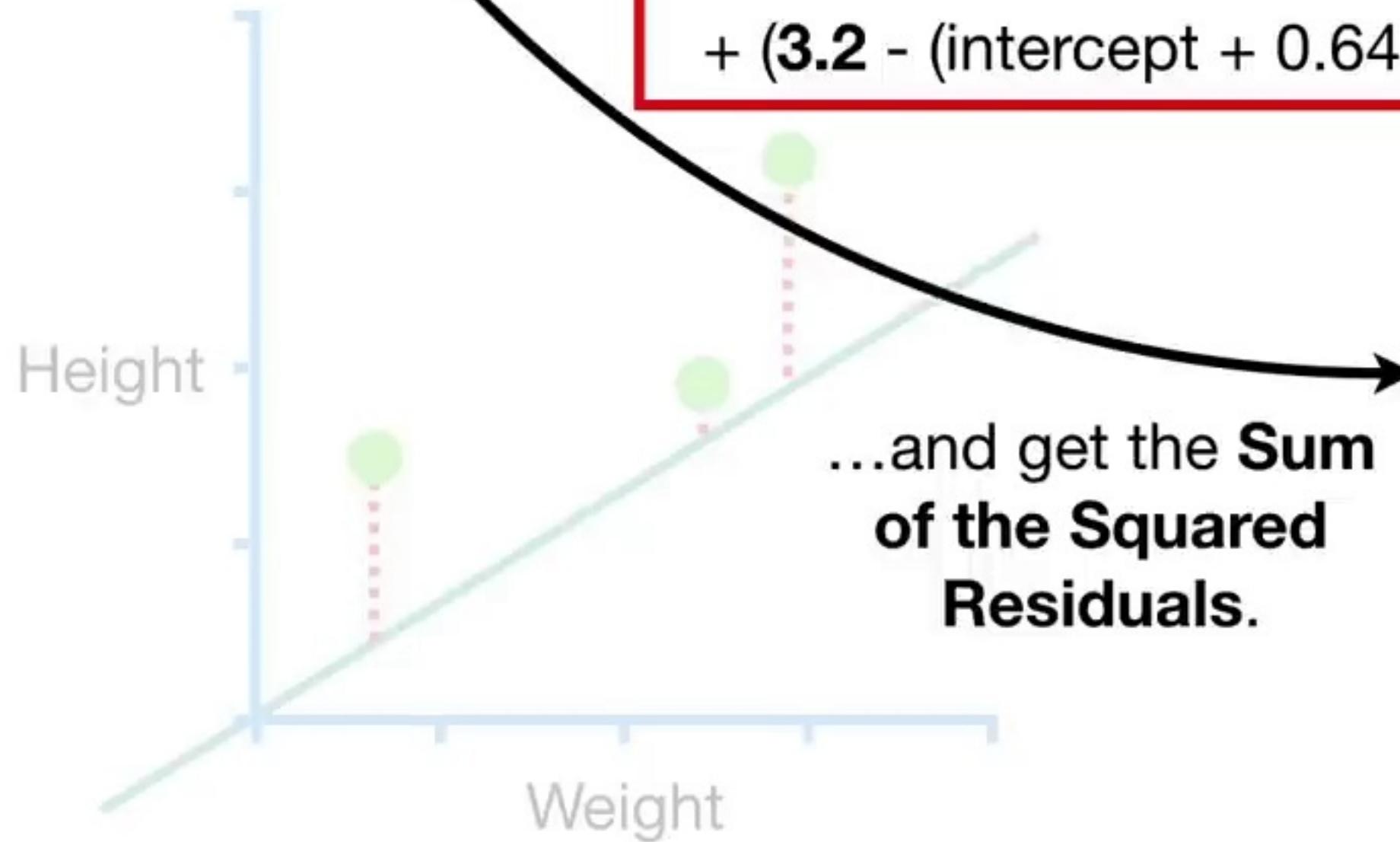
...and get the **Sum
of the Squared
Residuals.**



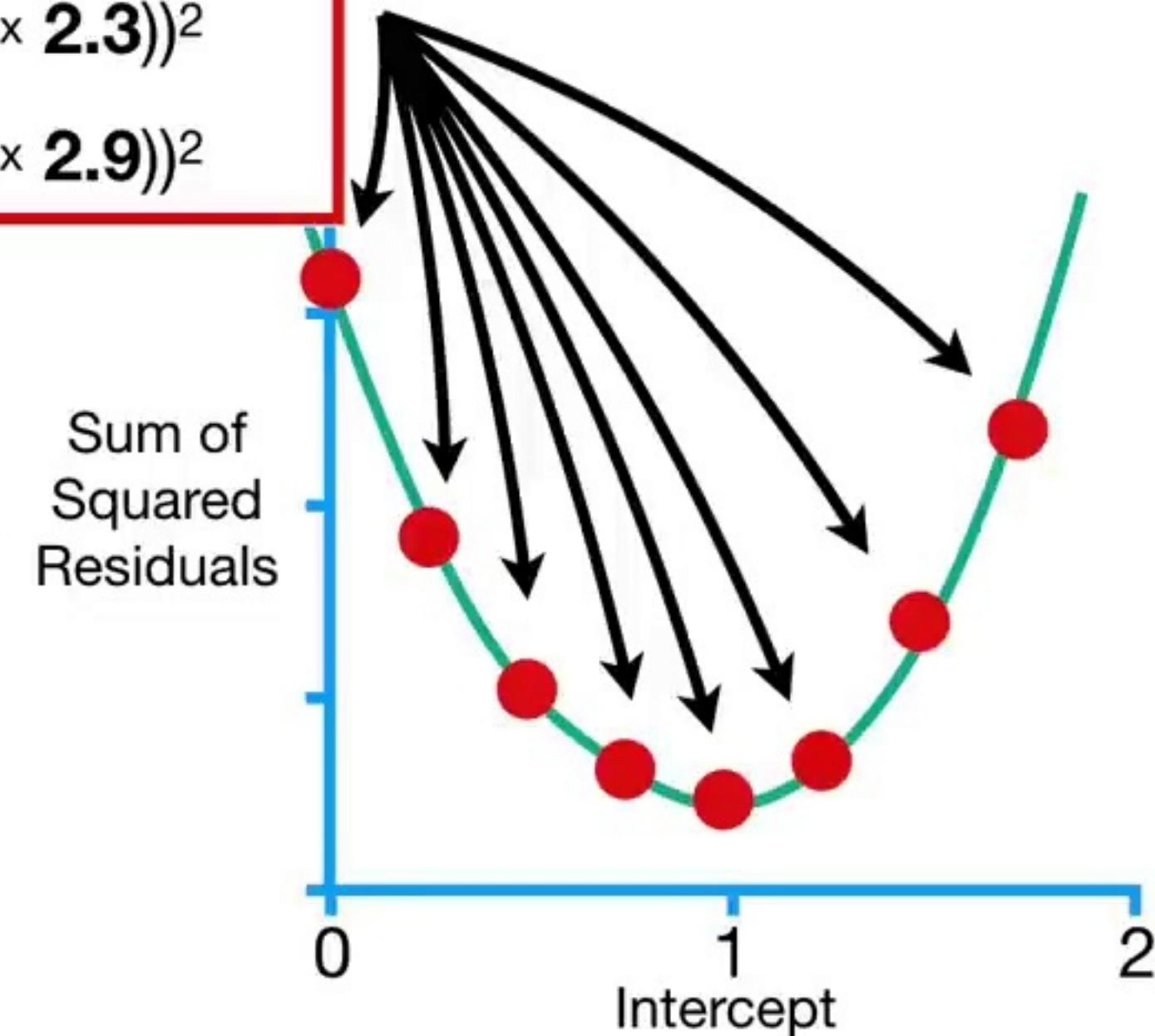
Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

$$+ (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$+ (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$



...and get the **Sum
of the Squared
Residuals.**

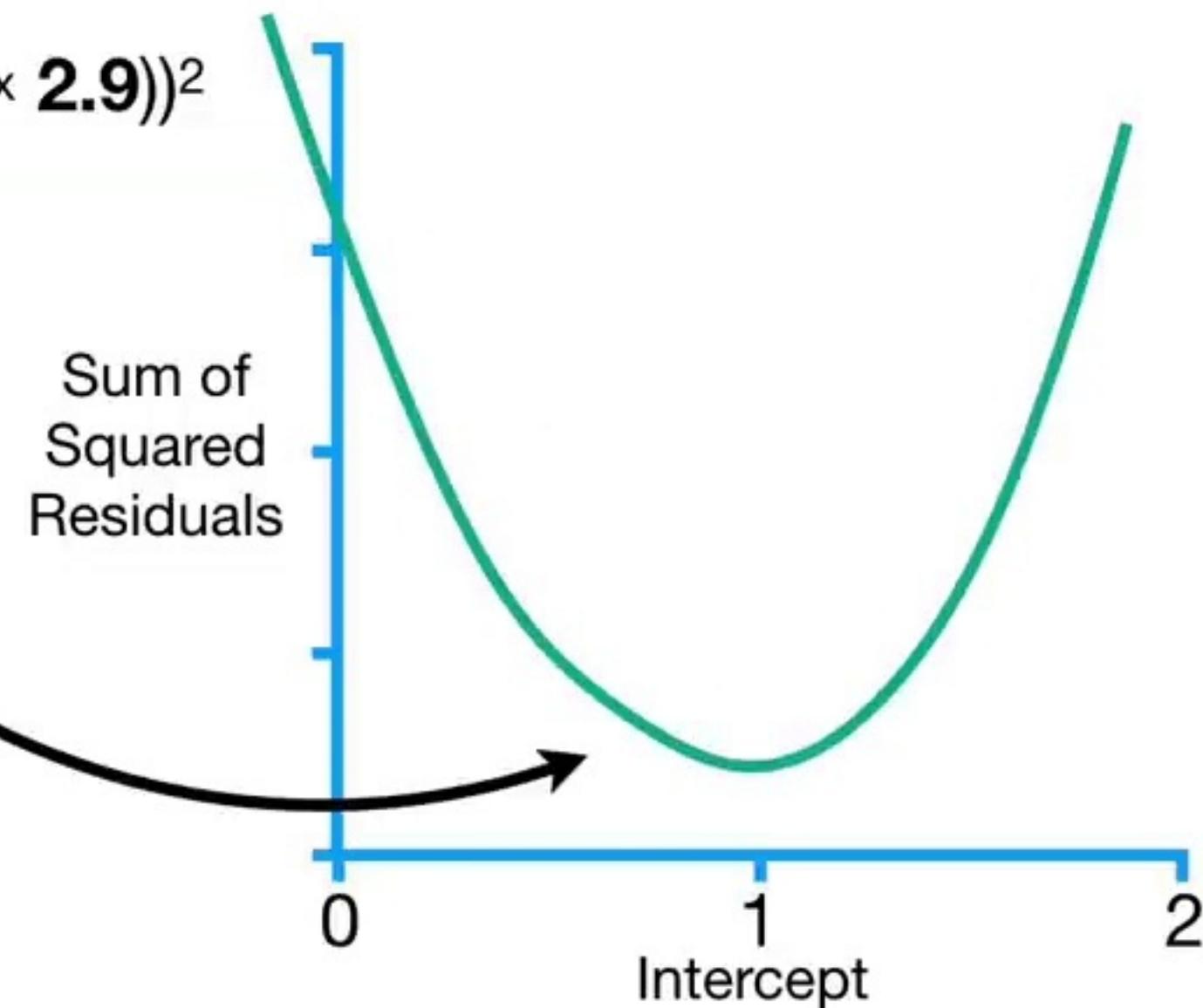


Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$

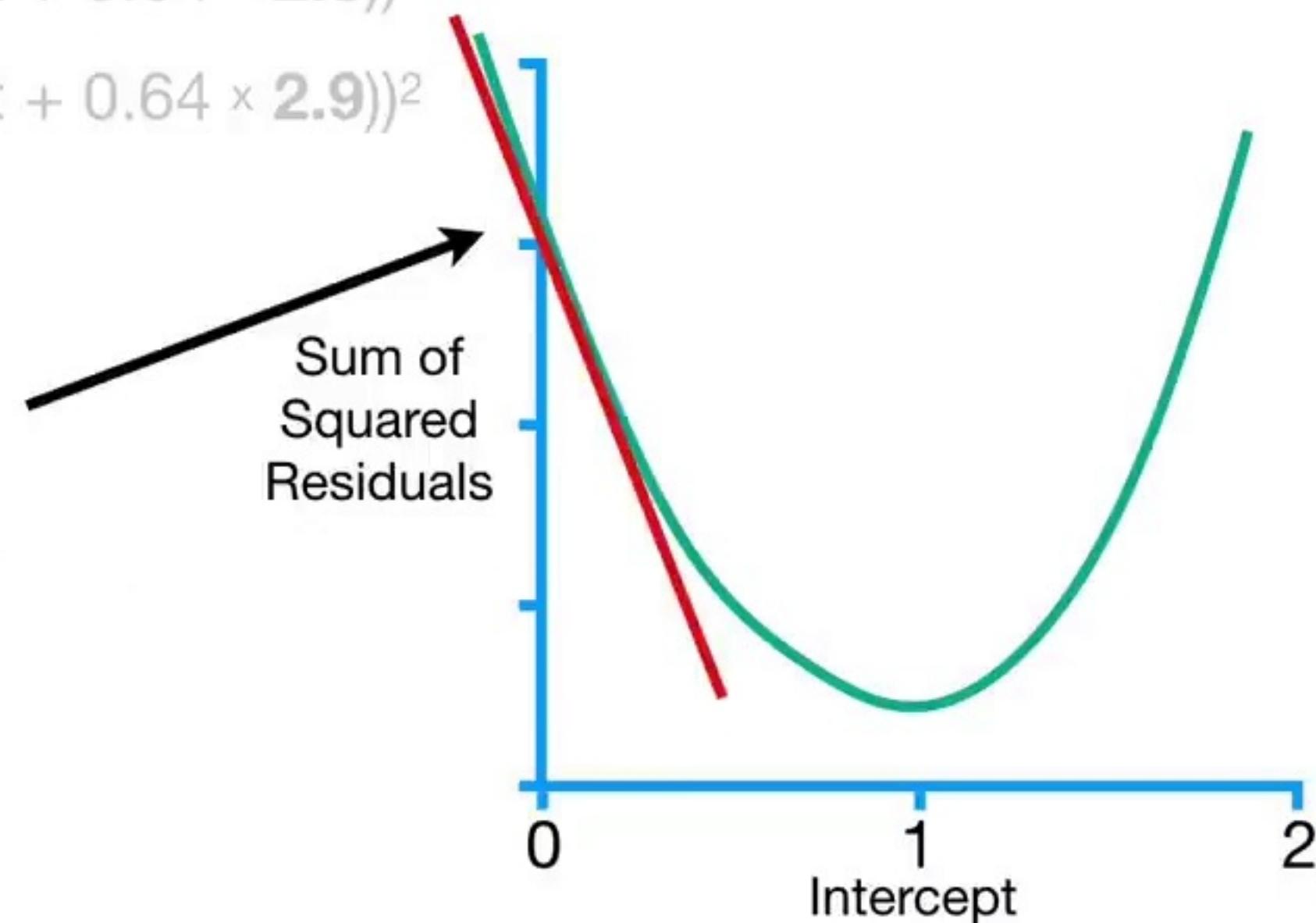
+ $(3.2 - (\text{intercept} + 0.64 \times 2.9))^2$

Thus, we now
have an equation
for this curve...



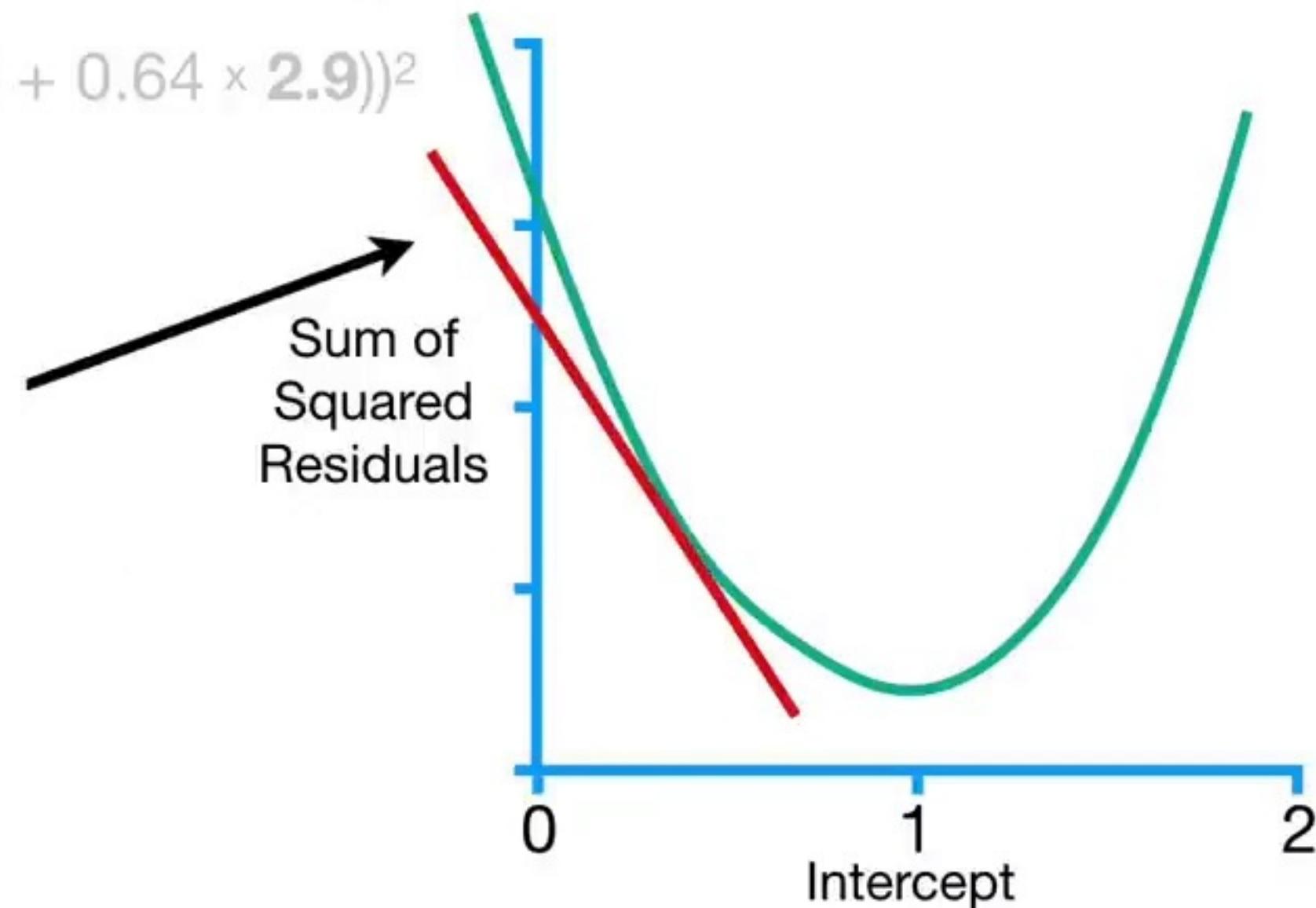
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ &\quad + (1.9 - (\text{intercept} + 0.64 \times 2.3))^2 \\ &\quad + (3.2 - (\text{intercept} + 0.64 \times 2.9))^2\end{aligned}$$

...and we can take the derivative of this function and determine the slope at any value for the **Intercept**.



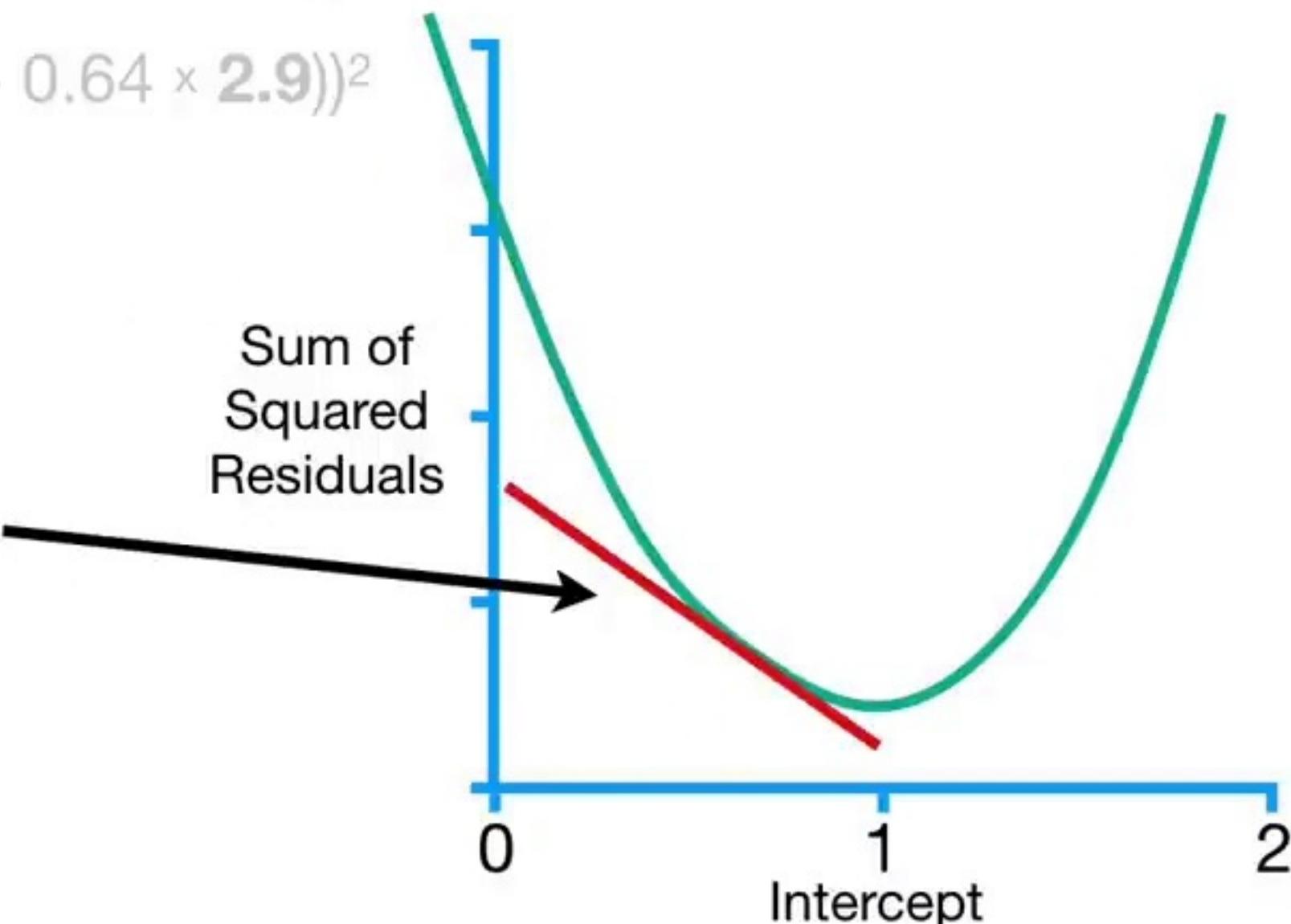
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ &\quad + (1.9 - (\text{intercept} + 0.64 \times 2.3))^2 \\ &\quad + (3.2 - (\text{intercept} + 0.64 \times 2.9))^2\end{aligned}$$

...and we can take the derivative of this function and determine the slope at any value for the **Intercept**.



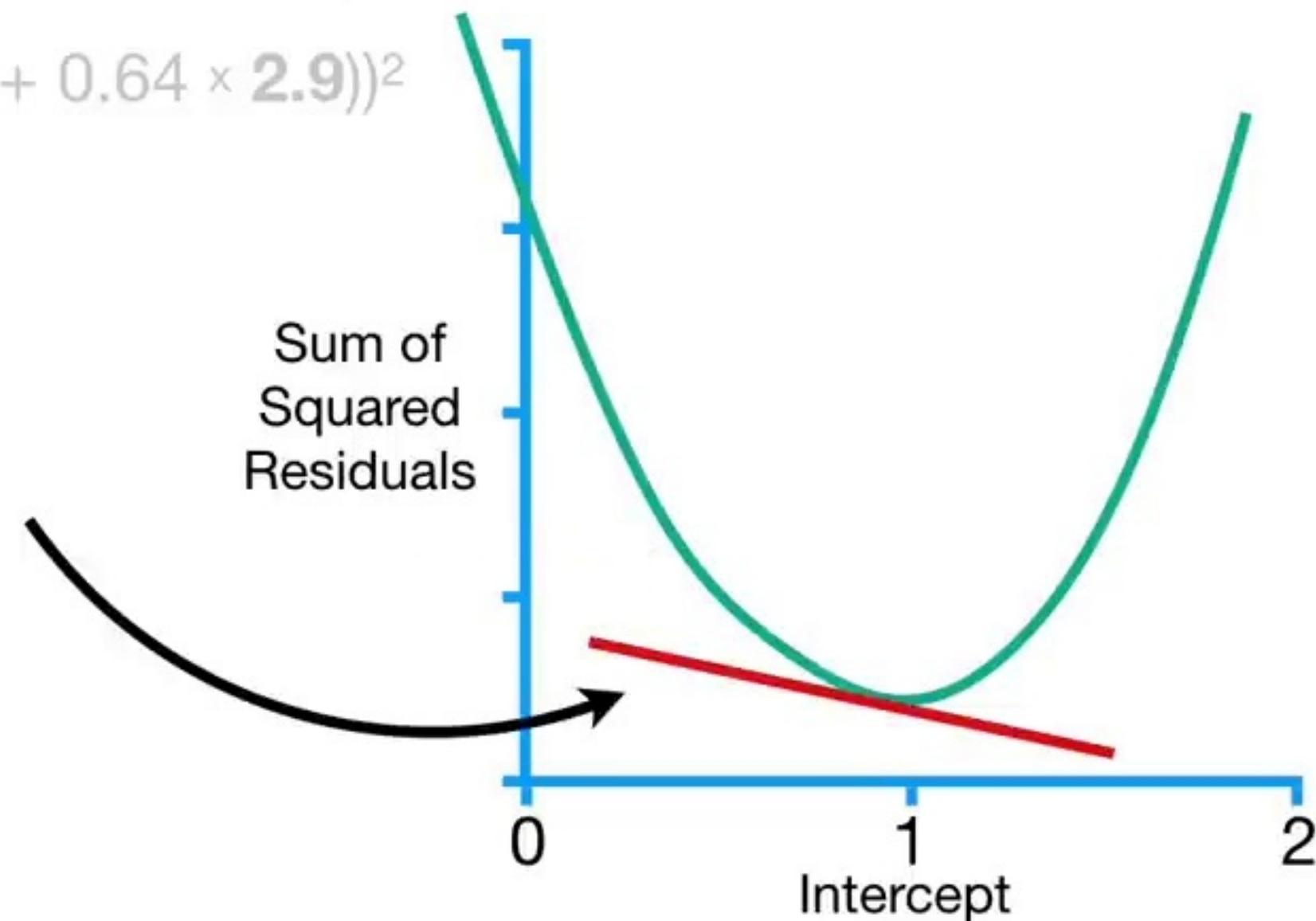
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + 0.64 \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + 0.64 \times 2.9))^2\end{aligned}$$

...and we can take the derivative of this function and determine the slope at any value for the **Intercept**.



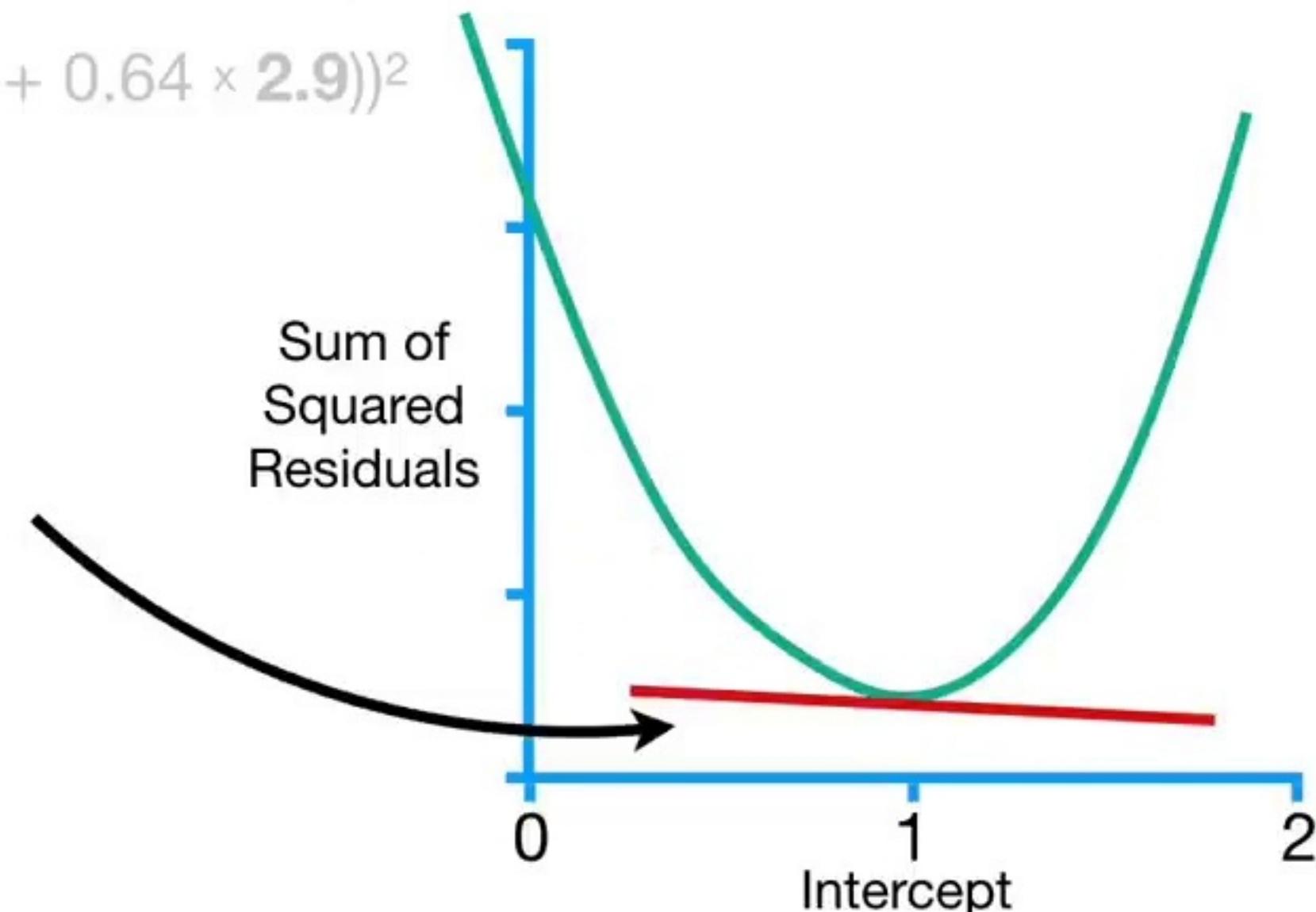
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + 0.64 \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + 0.64 \times 2.9))^2\end{aligned}$$

...and we can take the derivative of this function and determine the slope at any value for the **Intercept**.



$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + 0.64 \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + 0.64 \times 2.9))^2\end{aligned}$$

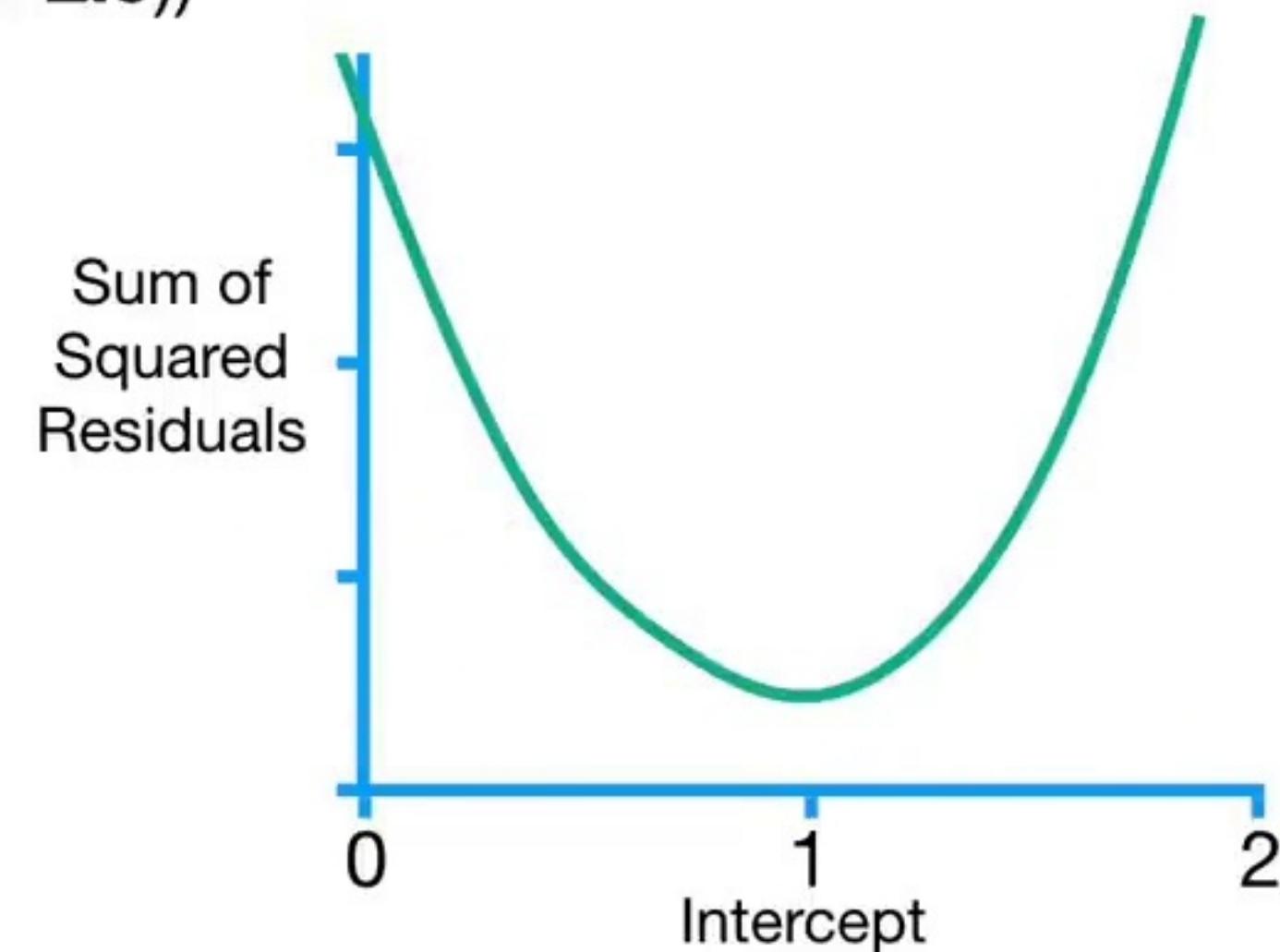
...and we can take the derivative of this function and determine the slope at any value for the **Intercept**.



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

$$+ (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$
$$+ (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

So let's take the derivative
of the Sum of the
Squared Residuals with
respect to the **Intercept**.



$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ &\quad + (1.9 - (\text{intercept} + 0.64 \times 2.3))^2 \\ &\quad + (3.2 - (\text{intercept} + 0.64 \times 2.9))^2\end{aligned}$$

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$

Sum of squared residuals = **(1.4 - (intercept + 0.64 × 0.5))²**

$$+ (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$+ (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

...the derivative of
the first part...

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$

$$\text{Sum of squared residuals} = (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$

$$+ (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$+ (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

...plus the derivative of the second part...

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$
$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$\text{Sum of squared residuals} = (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$

$$+ (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$+ (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

...plus the derivative
of the third part.

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$
$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$
$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ &\quad + (1.9 - (\text{intercept} + 0.64 \times 2.3))^2 \\ &\quad + (3.2 - (\text{intercept} + 0.64 \times 2.9))^2\end{aligned}$$

Let's start by taking the derivative
of the first part.



$$\begin{aligned}\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} &= \boxed{\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2} \\ &\quad + \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + 0.64 \times 2.3))^2 \\ &\quad + \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + 0.64 \times 2.9))^2\end{aligned}$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 = 2(1.4 - (\text{intercept} + 0.64 \times 0.5)) \times -1$$

...then we simplify by multiplying 2 by -1...

$$\begin{aligned}\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} &= \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ &\quad + \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + 0.64 \times 2.3))^2 \\ &\quad + \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + 0.64 \times 2.9))^2\end{aligned}$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 = 2(1.4 - (\text{intercept} + 0.64 \times 0.5)) \times -1$$

$$= -2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

...then we simplify by multiplying 2 by -1...

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 = 2(1.4 - (\text{intercept} + 0.64 \times 0.5)) \times -1$$

$$= -2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

...and this...

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 = 2(1.4 - (\text{intercept} + 0.64 \times 0.5)) \times -1$$

$$= -2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

...and this...

...is the derivative
of the first part...

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 = 2(1.4 - (\text{intercept} + 0.64 \times 0.5)) \times -1$$

$$= -2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

↓
...so we plug it in.

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = -2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

Now we need to take the derivative of the next two parts.

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} = -2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

$\frac{d}{d \text{ intercept}}$ Sum of squared residuals =

$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$

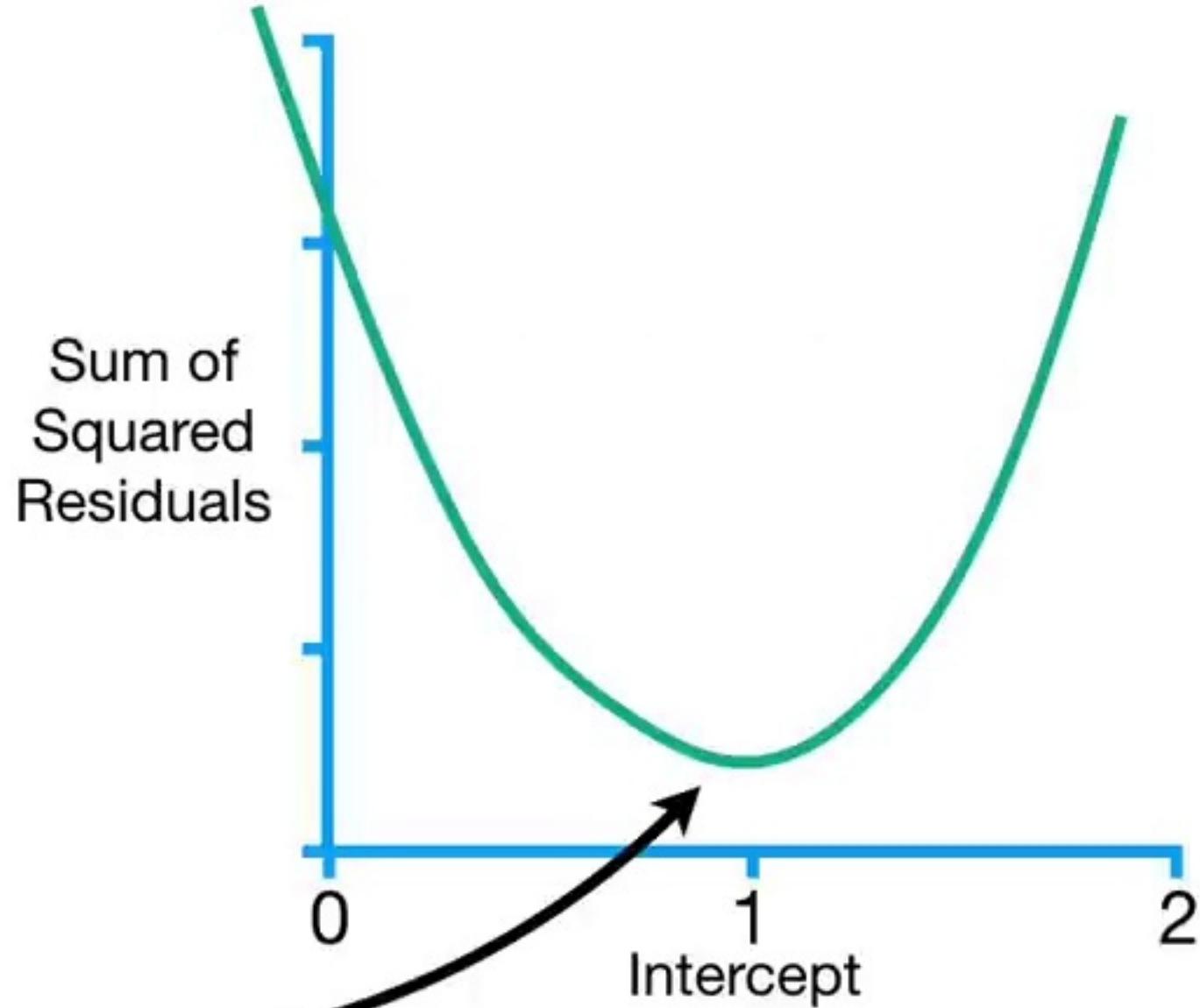
$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$



Let's move the derivative up here so that it's not taking up half of the screen.

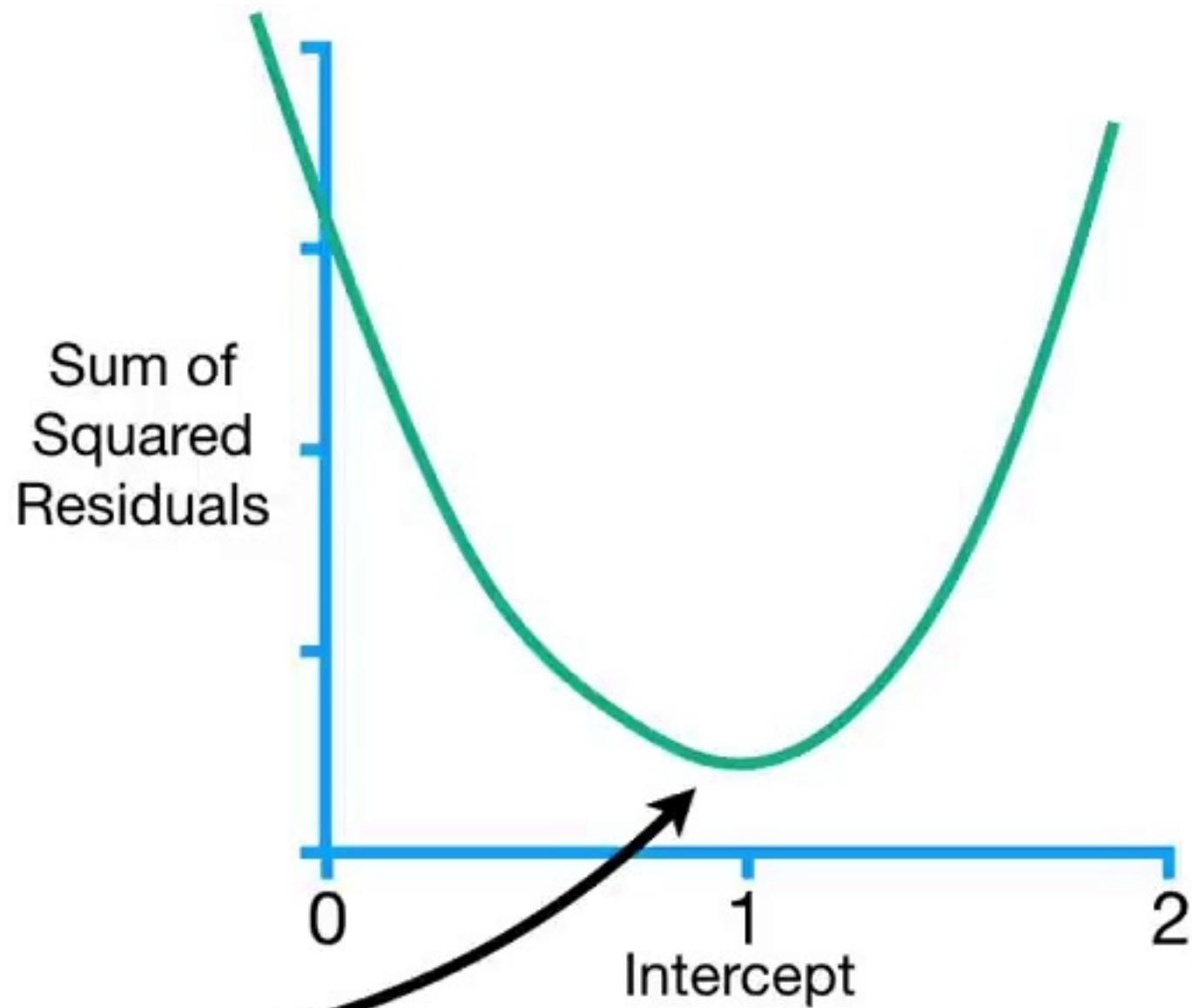
$$\frac{d}{d \text{intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

Now that we have the derivative,
Gradient Descent will use it to find
where the Sum of Squared
Residuals is lowest.



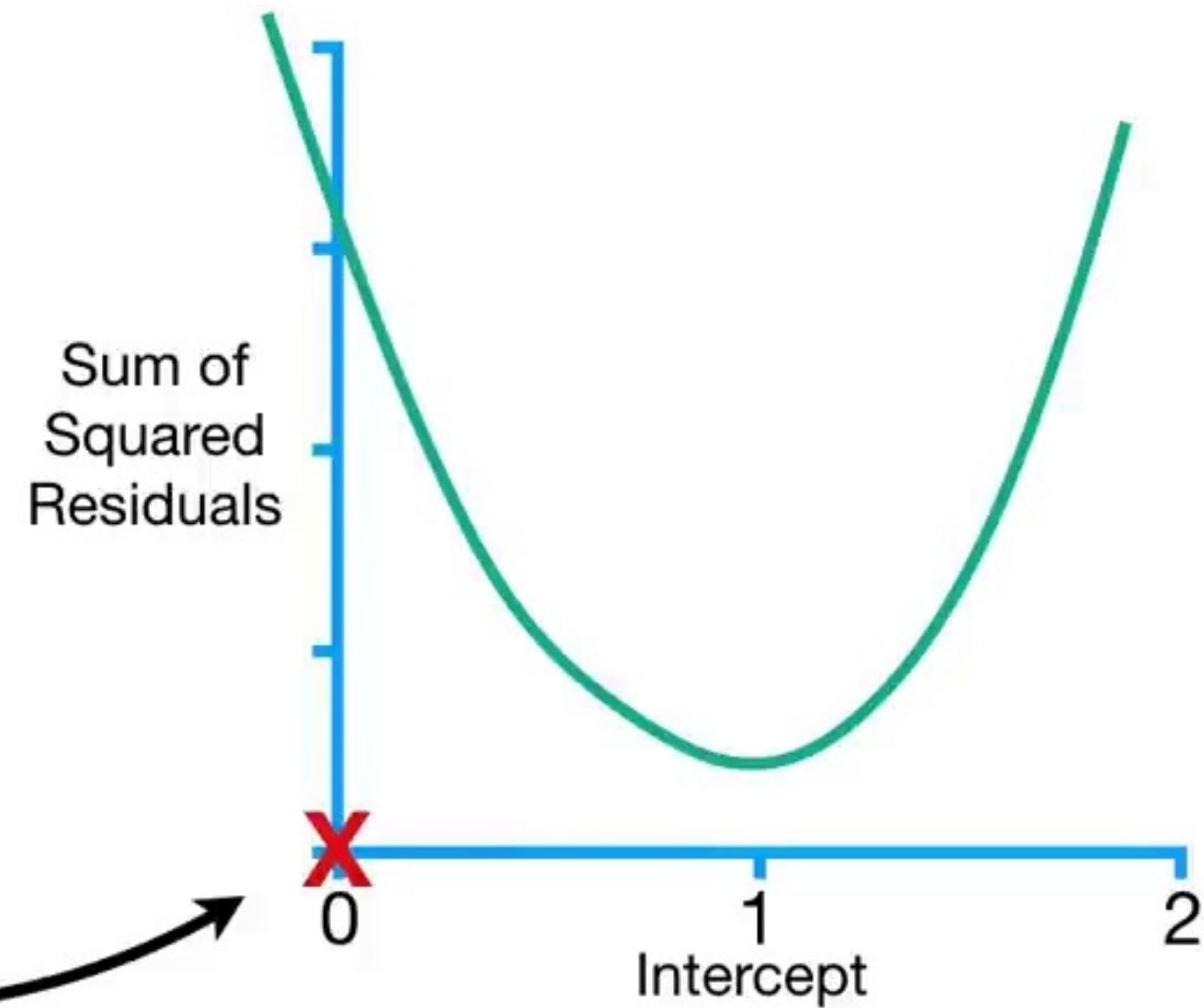
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

NOTE: If we were using **Least Squares** to solve for the optimal value for the **Intercept**, we would simply find where the slope of the curve = **0**.



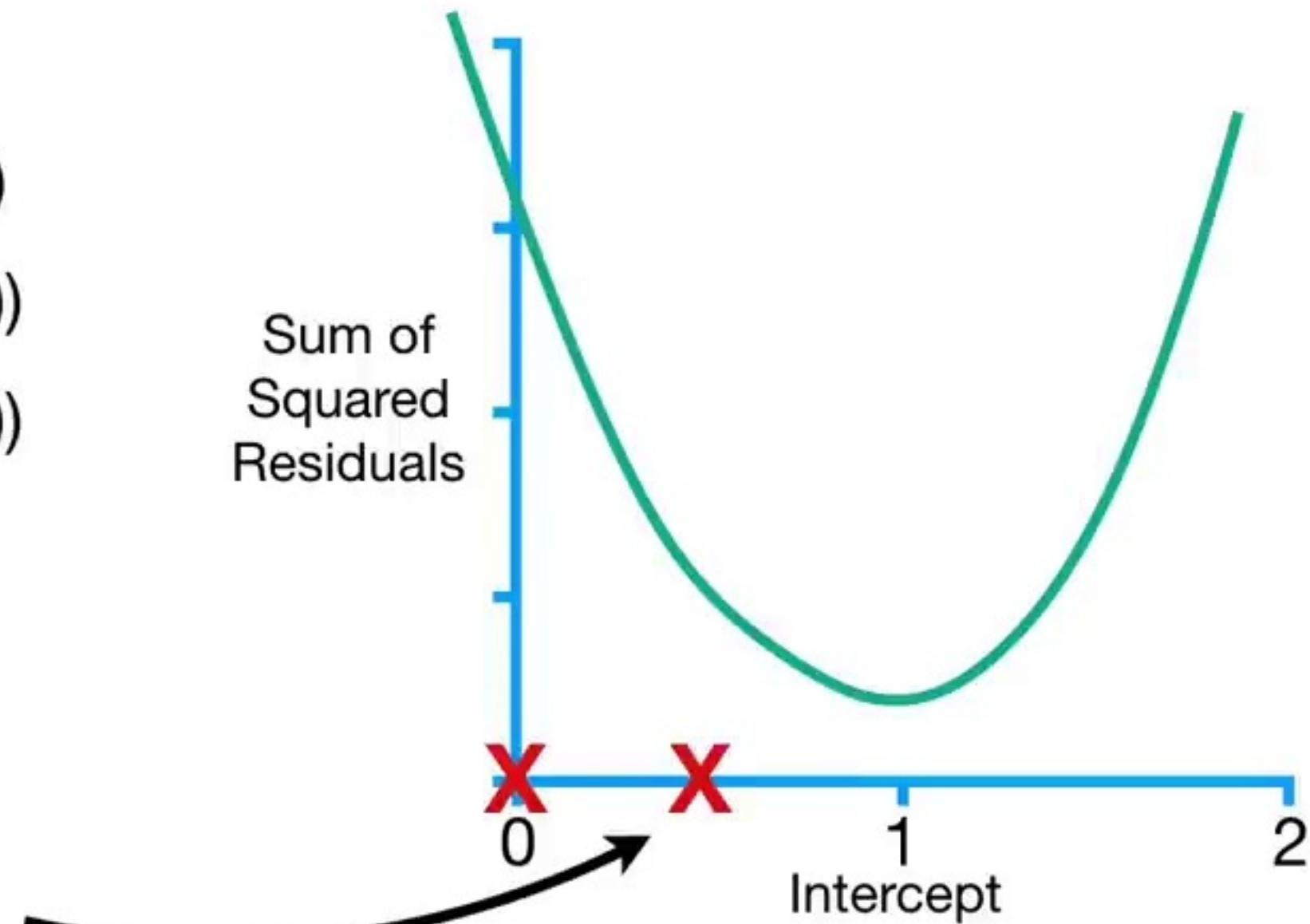
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

In contrast, **Gradient Descent** finds the minimum value by taking steps from an initial guess until it reaches the best value.



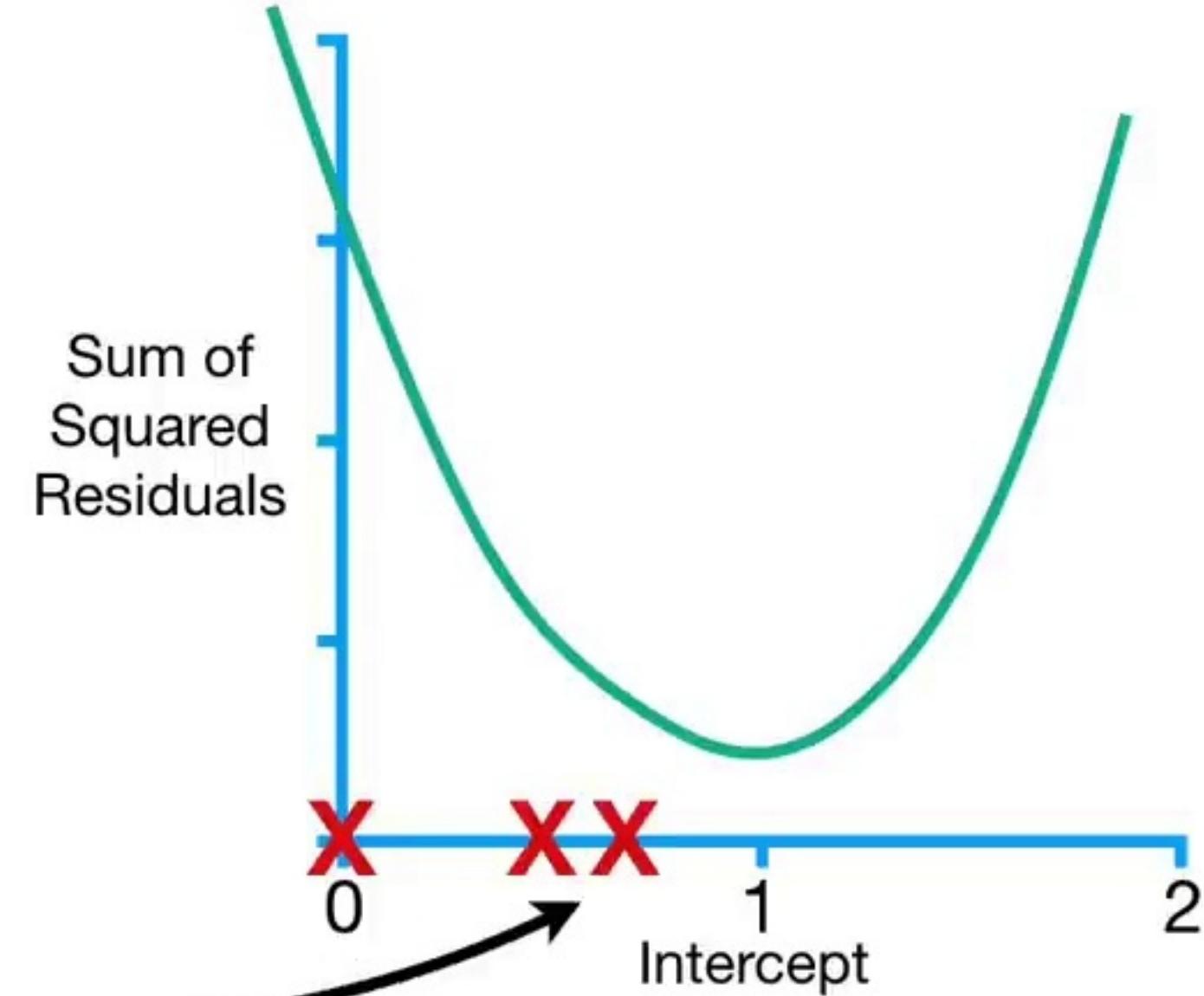
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

In contrast, **Gradient Descent** finds the minimum value by taking steps from an initial guess until it reaches the best value.



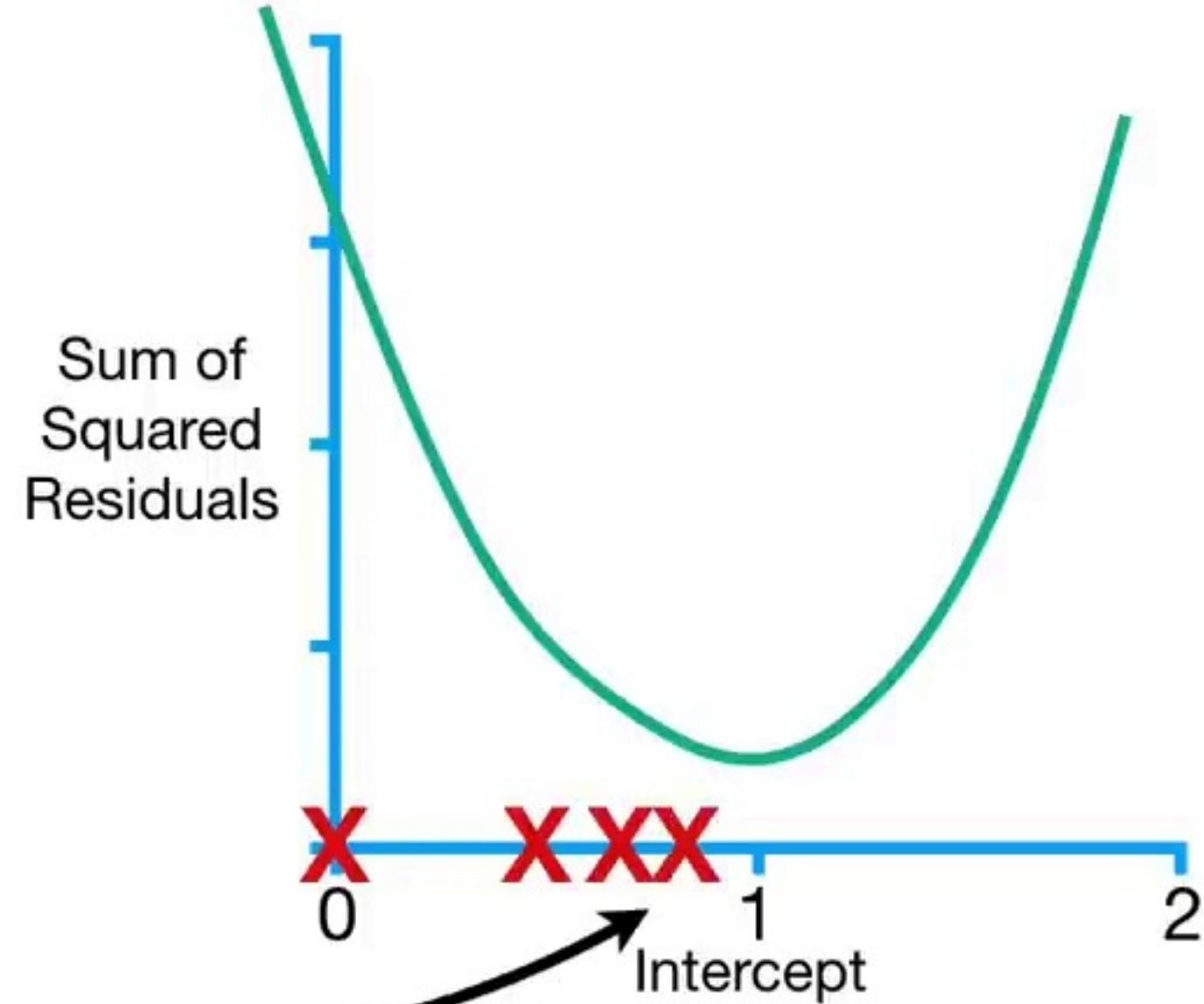
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

In contrast, **Gradient Descent** finds the minimum value by taking steps from an initial guess until it reaches the best value.



$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

In contrast, **Gradient Descent** finds the minimum value by taking steps from an initial guess until it reaches the best value.



$$\frac{d}{d \text{ intercept}}$$

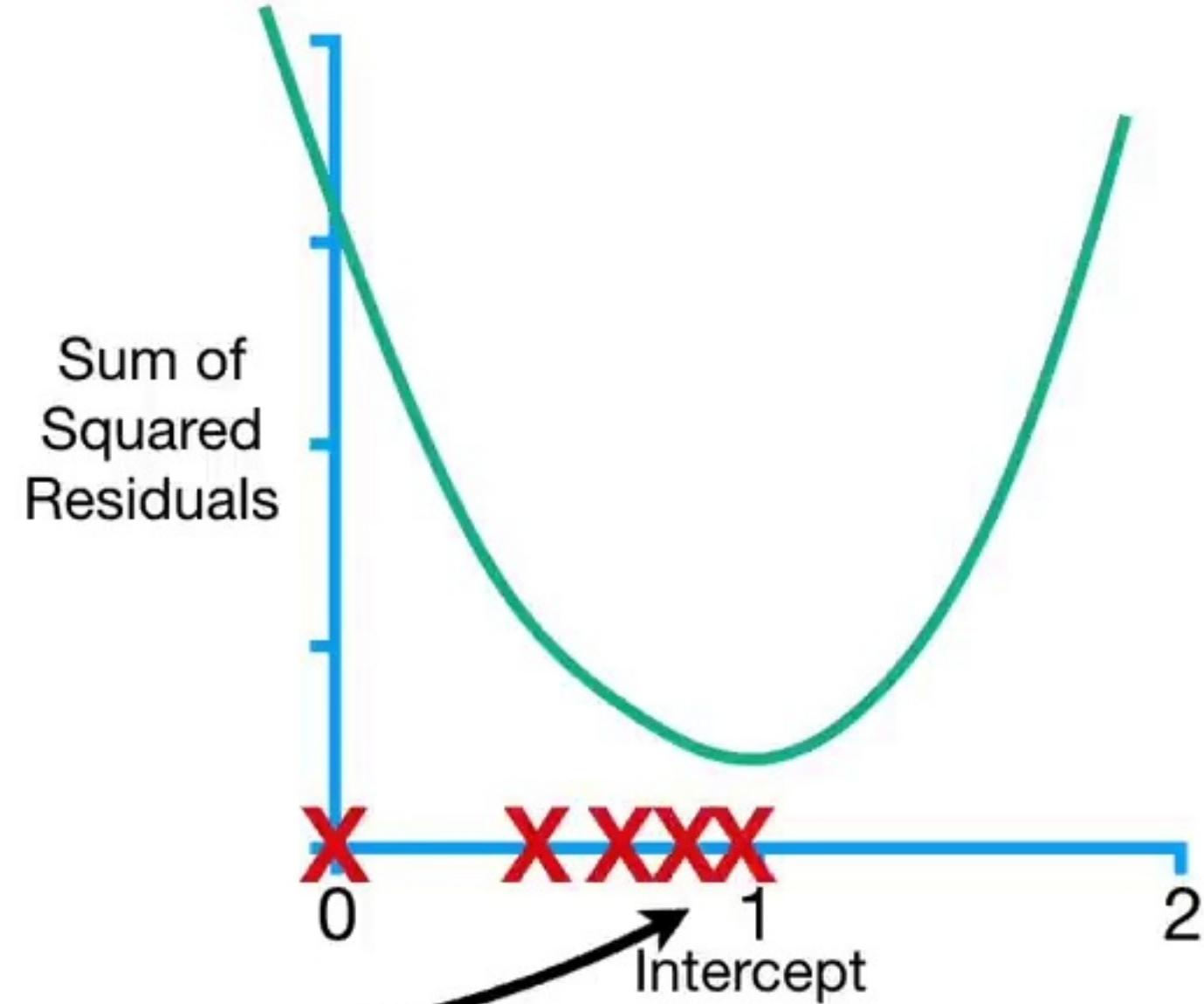
Sum of squared residuals =

$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$

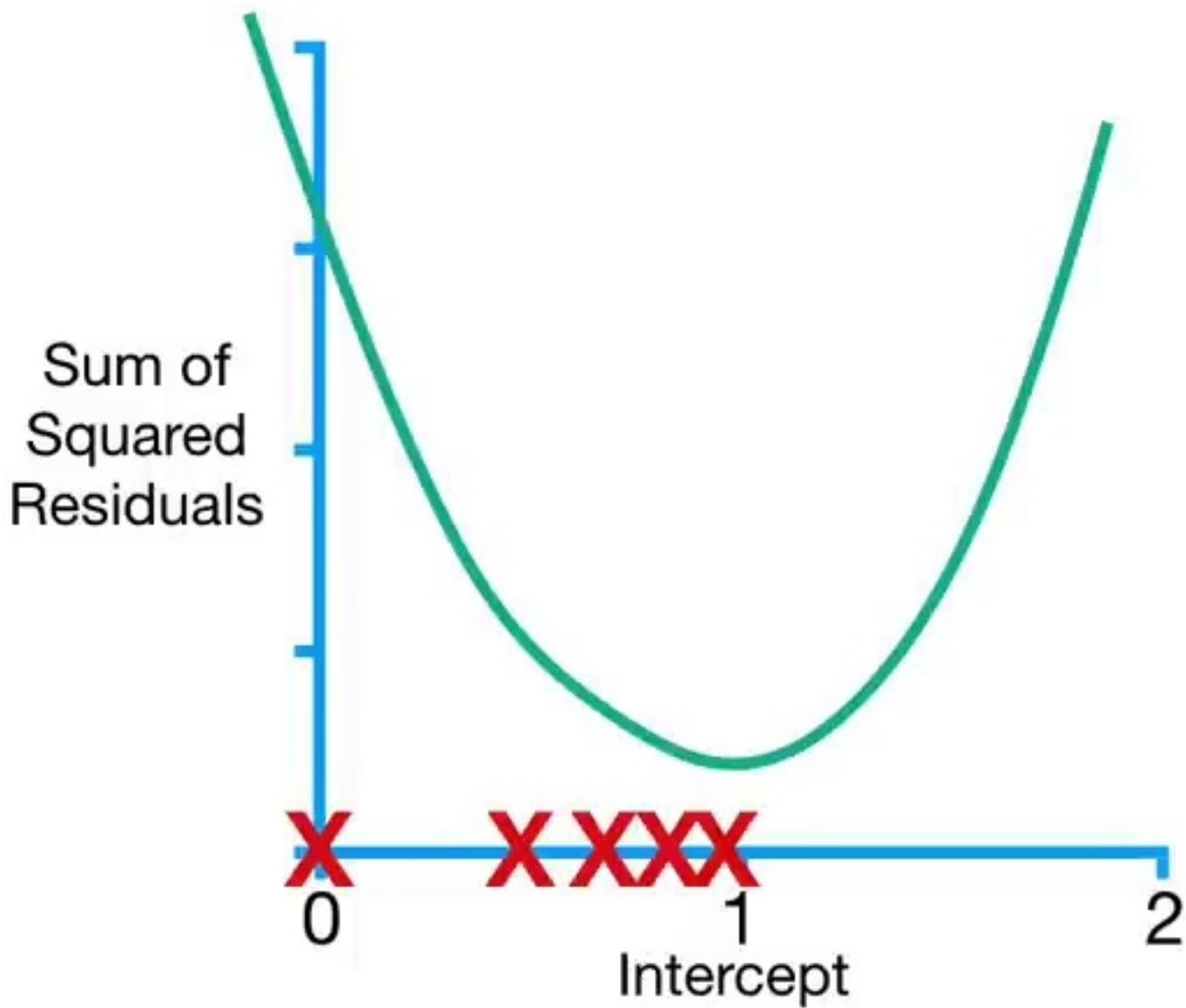
$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

In contrast, **Gradient Descent** finds the minimum value by taking steps from an initial guess until it reaches the best value.



$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

This makes **Gradient Descent** very useful when it is not possible to solve for where the derivative = **0**, and this is why **Gradient Descent** can be used in so many different situations.



$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =

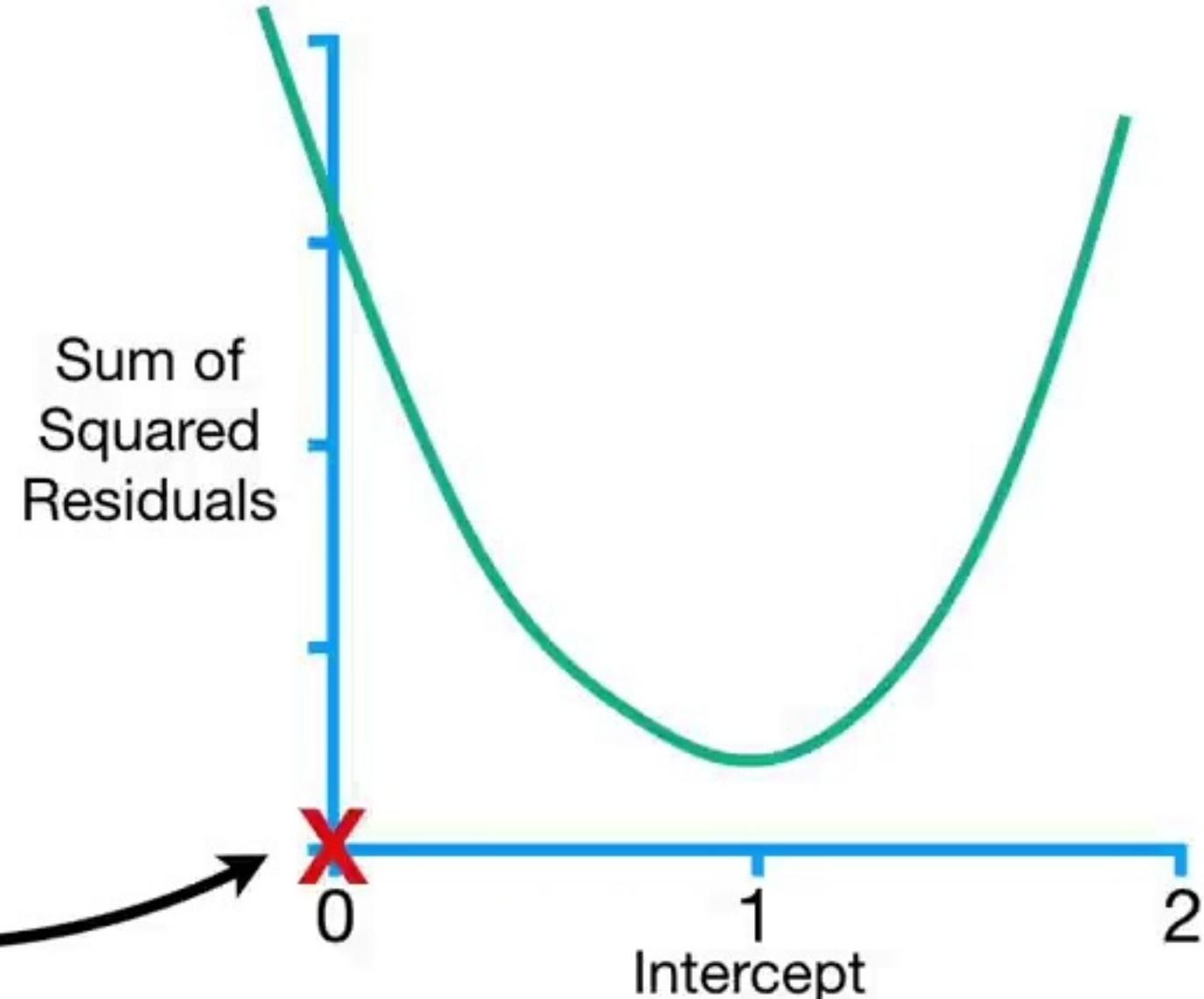
$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$

$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

Remember, we started by setting
the **Intercept** to a random number.

In this case, that was **0**.



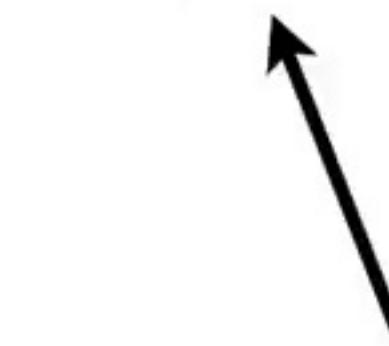
$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =

$$-2(1.4 - (0 + 0.64 \times 0.5))$$

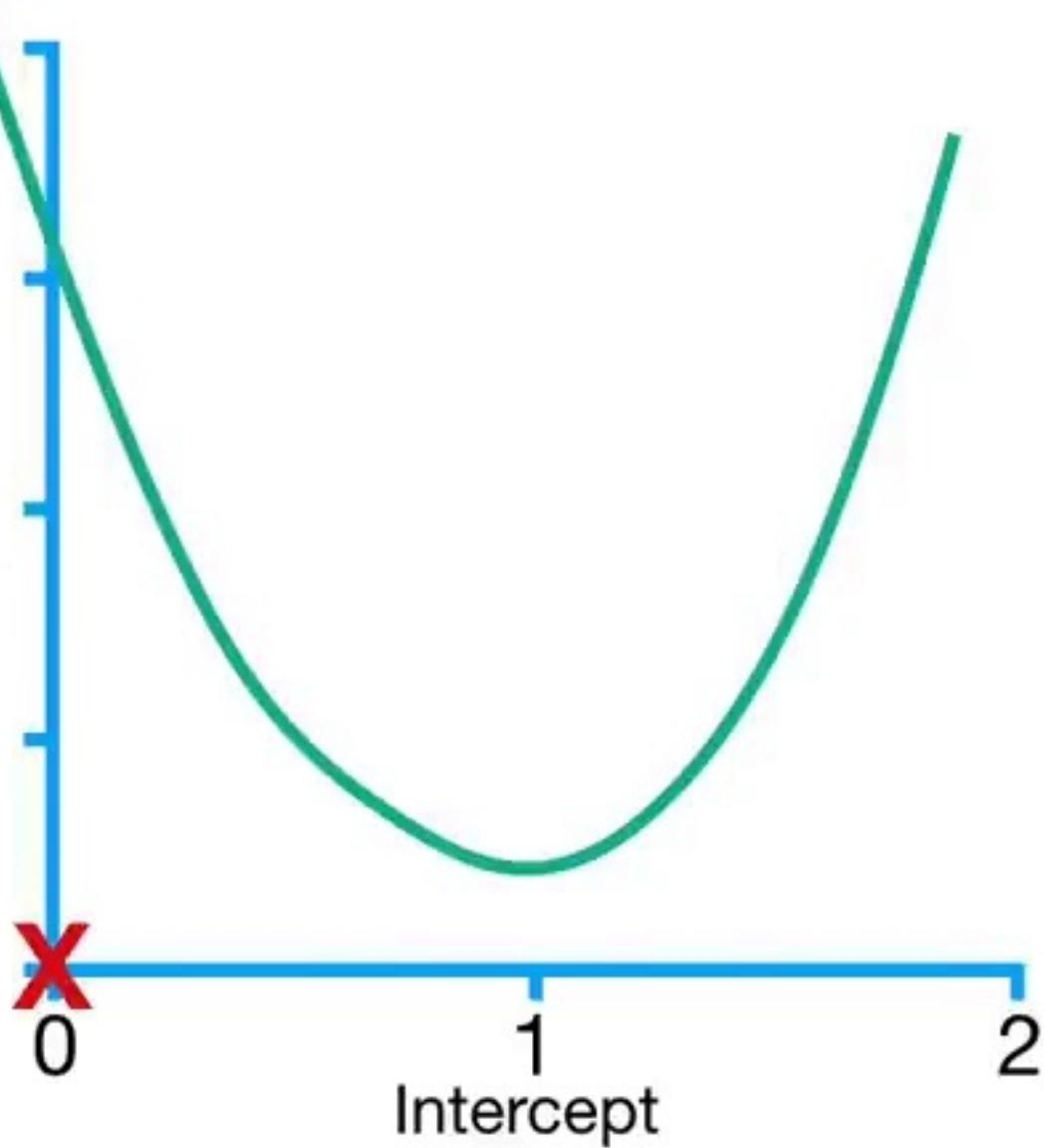
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$

$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$



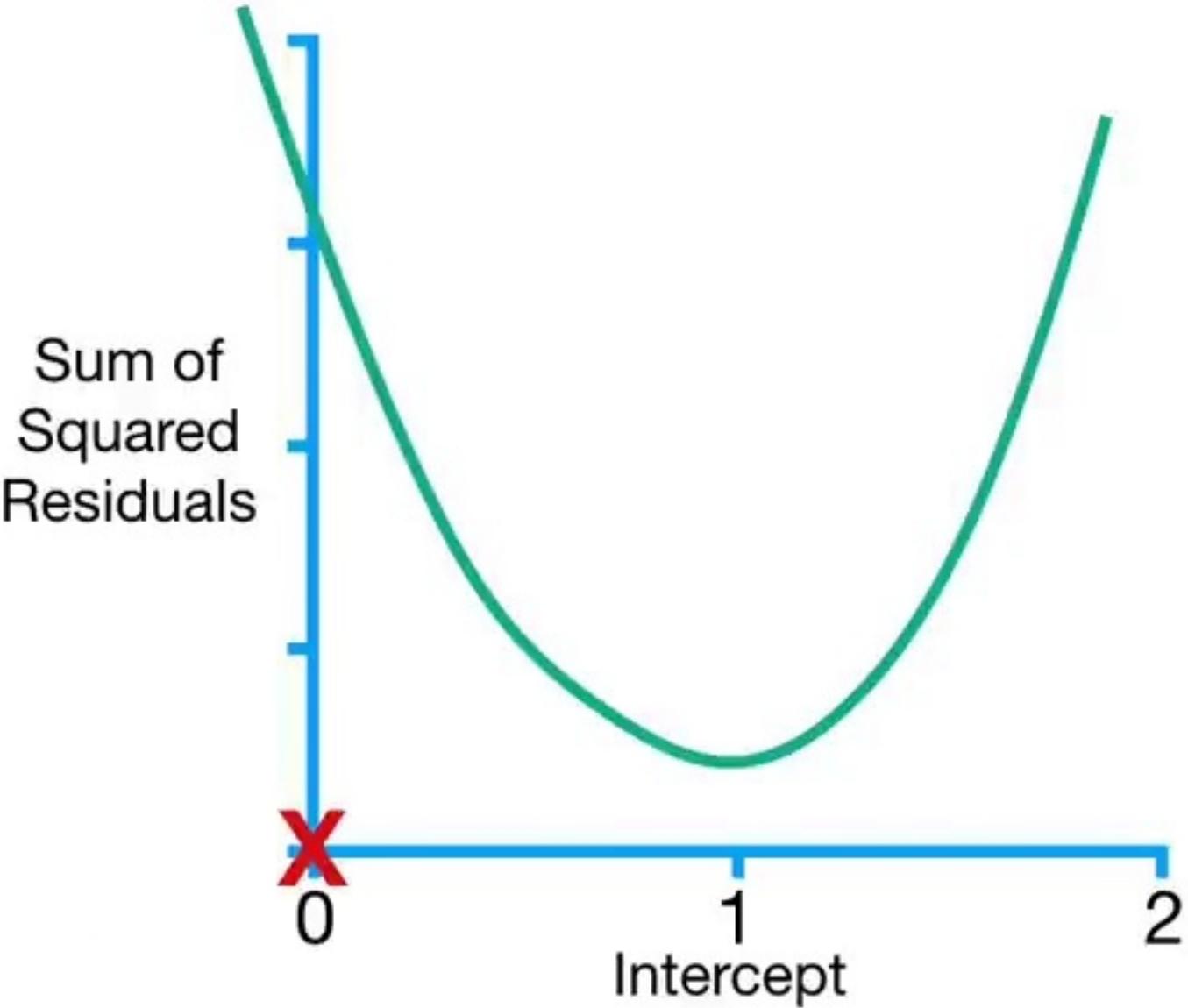
So we plug **0** into
the derivative...

Sum of
Squared
Residuals



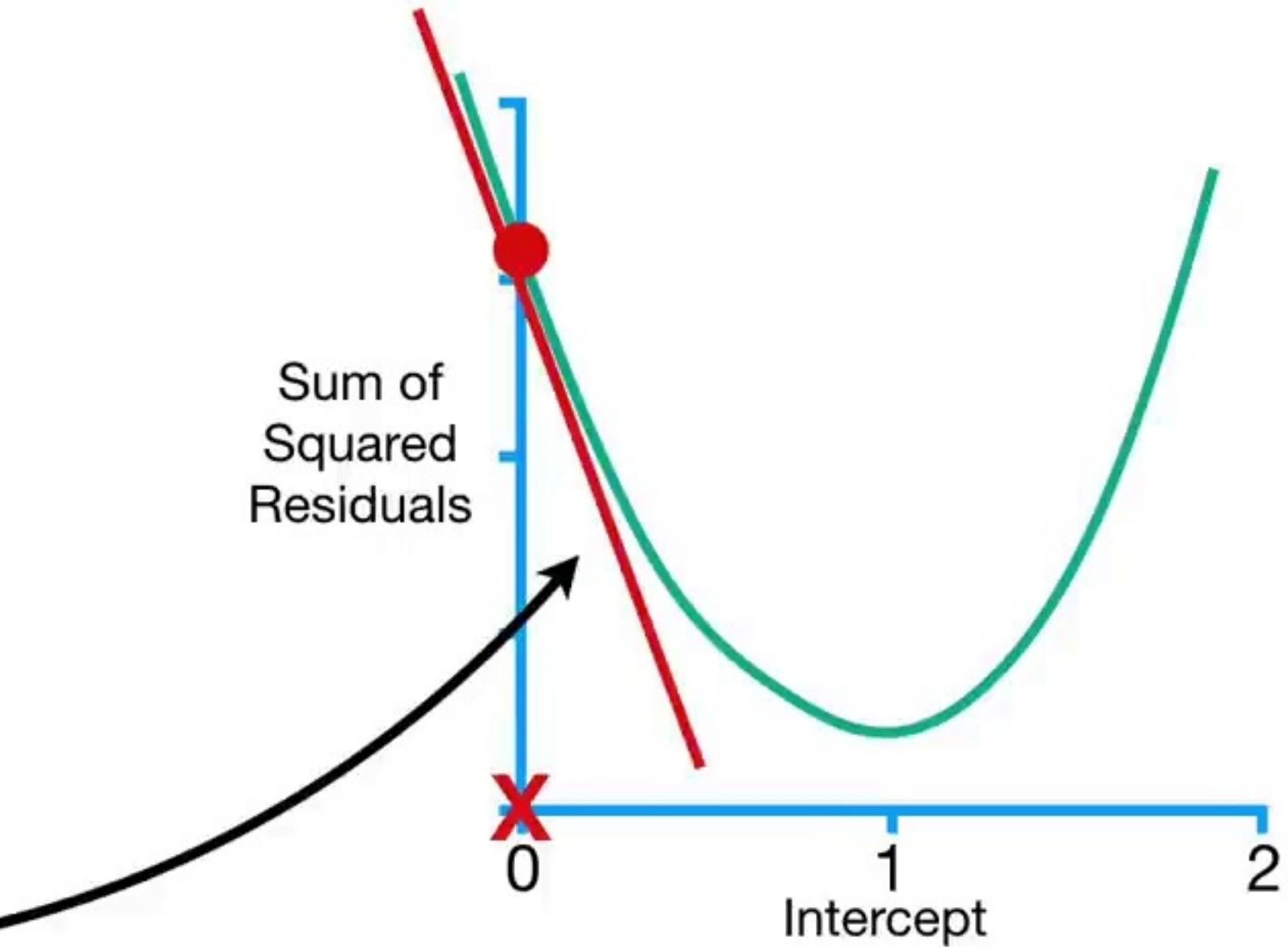
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

...and we get **-5.7**.



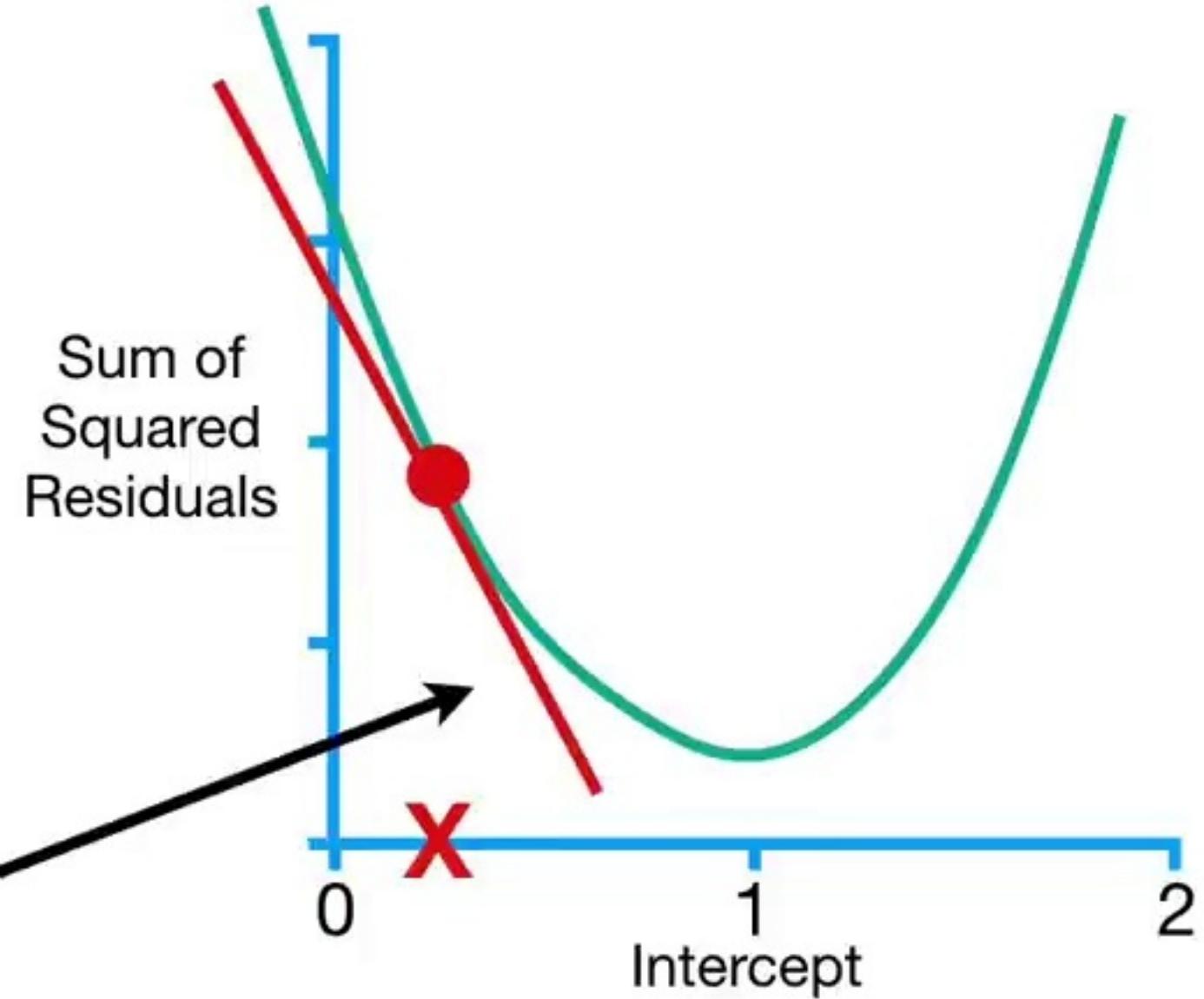
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

So when the **Intercept** = 0,
the slope of the curve = **-5.7**.



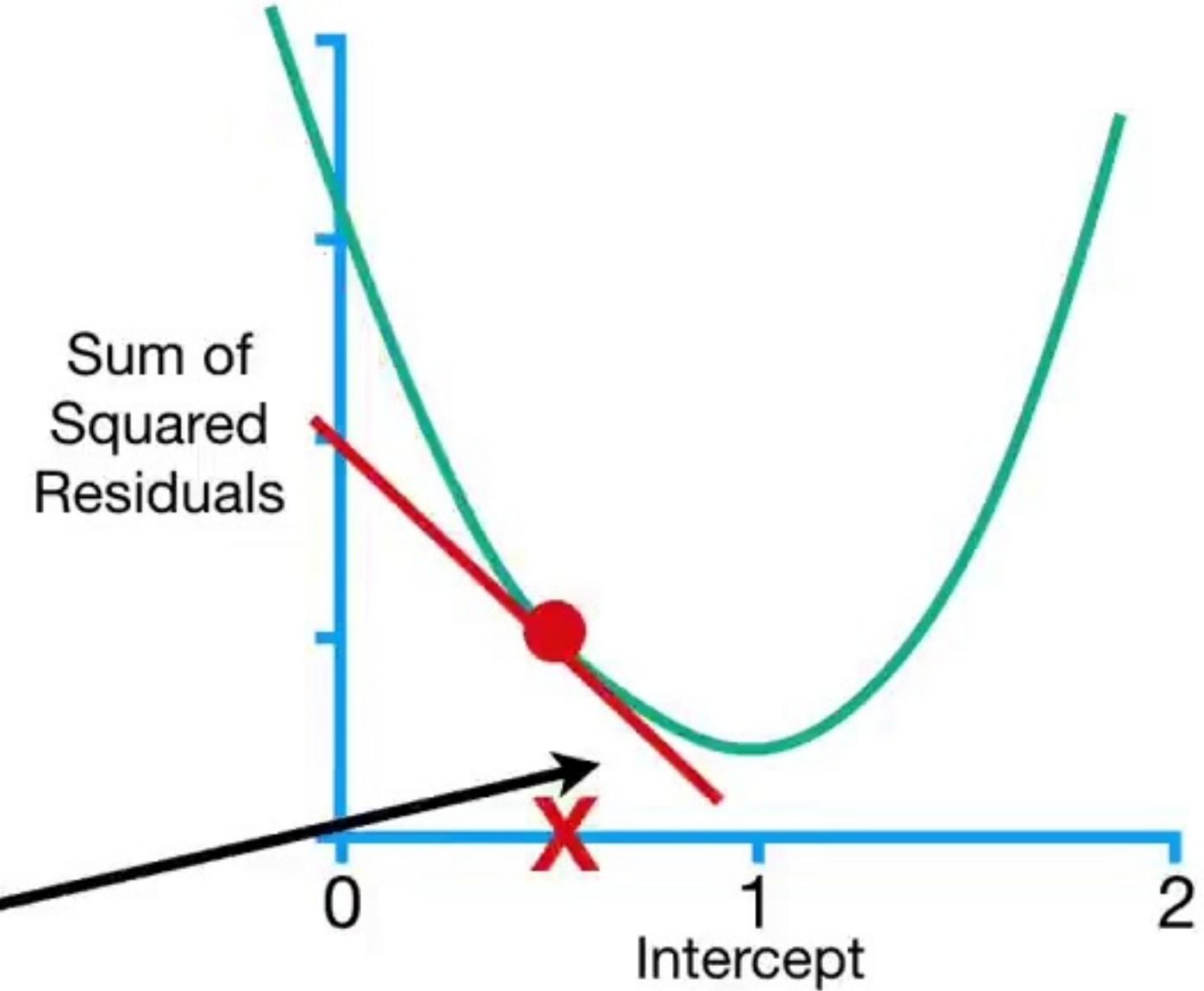
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

NOTE: The closer we get to the optimal value for the **Intercept**, the closer the slope of the curve gets to **0**.



$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

NOTE: The closer we get to the optimal value for the **Intercept**, the closer the slope of the curve gets to **0**.



$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =

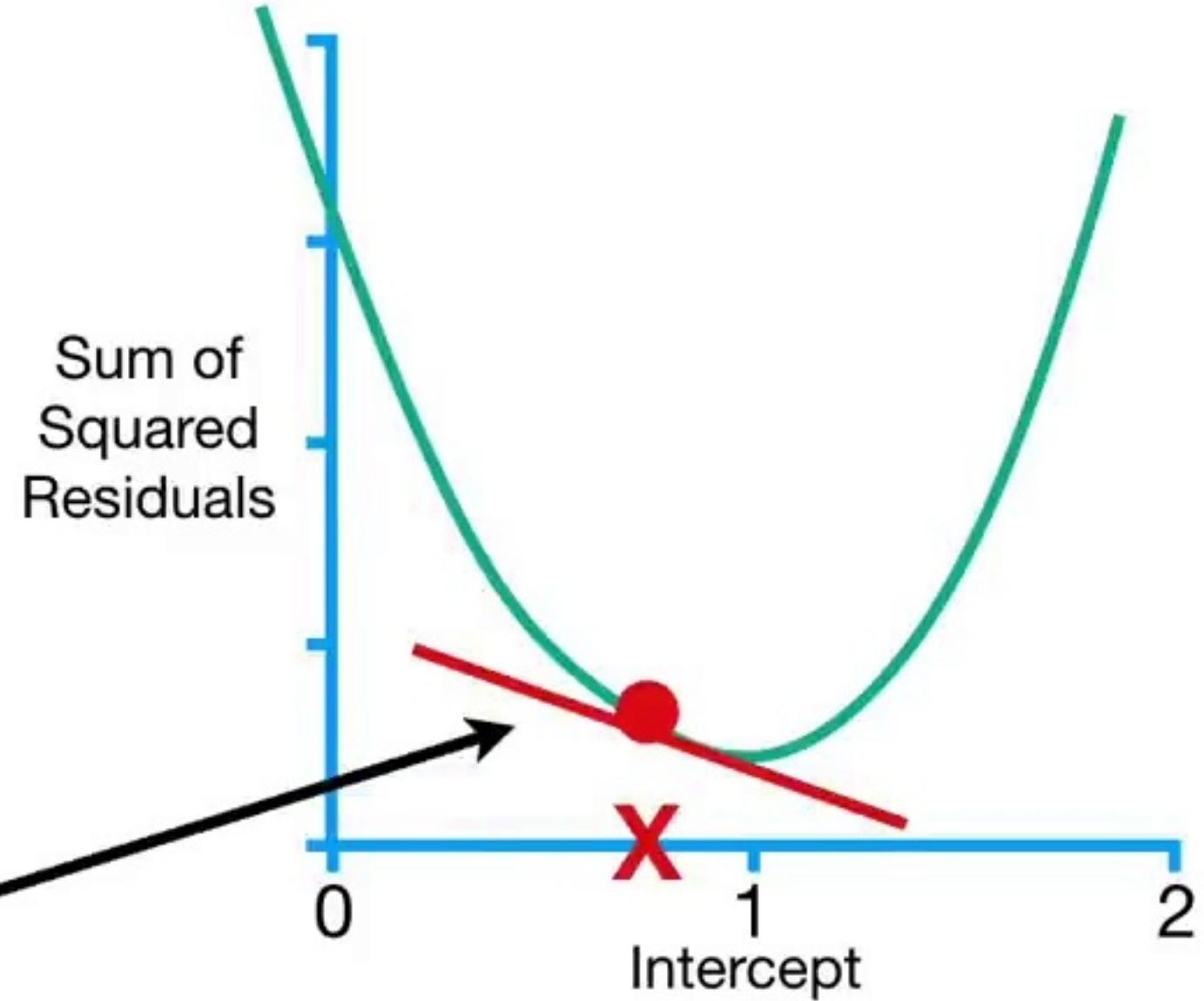
$$-2(1.4 - (0 + 0.64 \times 0.5))$$

$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$

$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$

$$= -5.7$$

NOTE: The closer we get to the optimal value for the **Intercept**, the closer the slope of the curve gets to **0**.



$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =

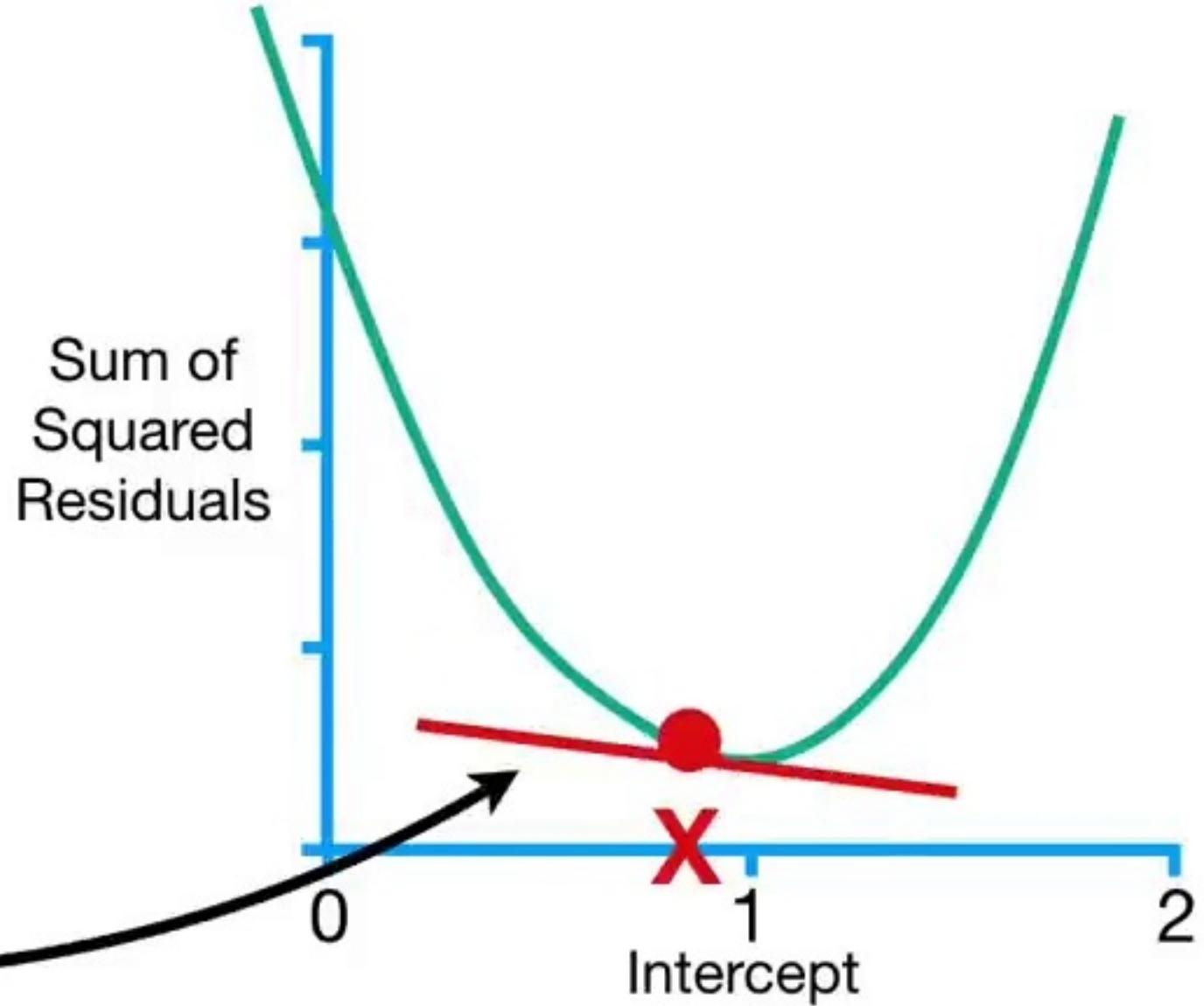
$$-2(1.4 - (0 + 0.64 \times 0.5))$$

$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$

$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$

$$= -5.7$$

NOTE: The closer we get to the optimal value for the **Intercept**, the closer the slope of the curve gets to **0**.



$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =

$$-2(1.4 - (0 + 0.64 \times 0.5))$$

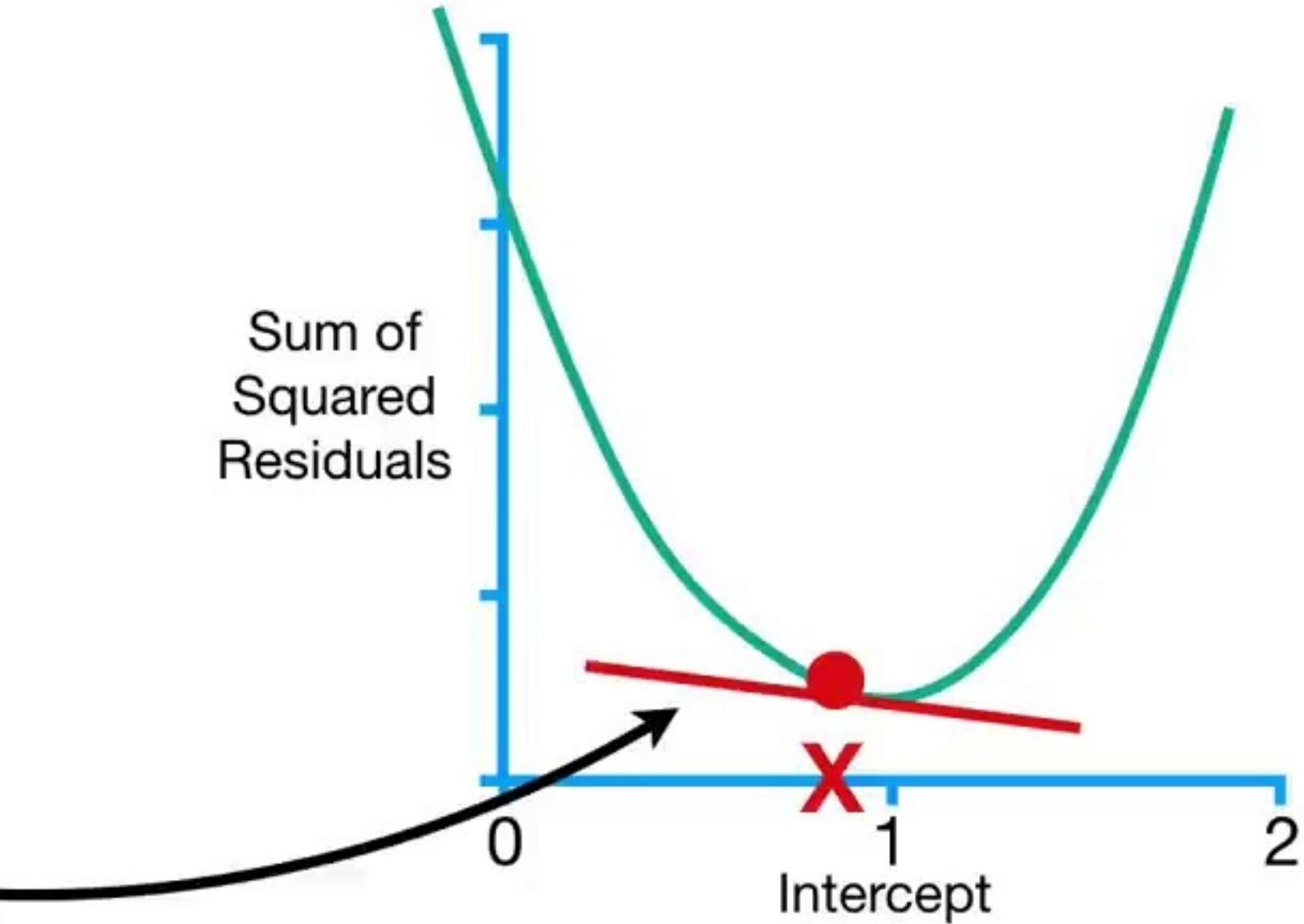
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$

$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$

$$= -5.7$$

This means that when
the slope of the curve is
close to **0**...

Sum of
Squared
Residuals

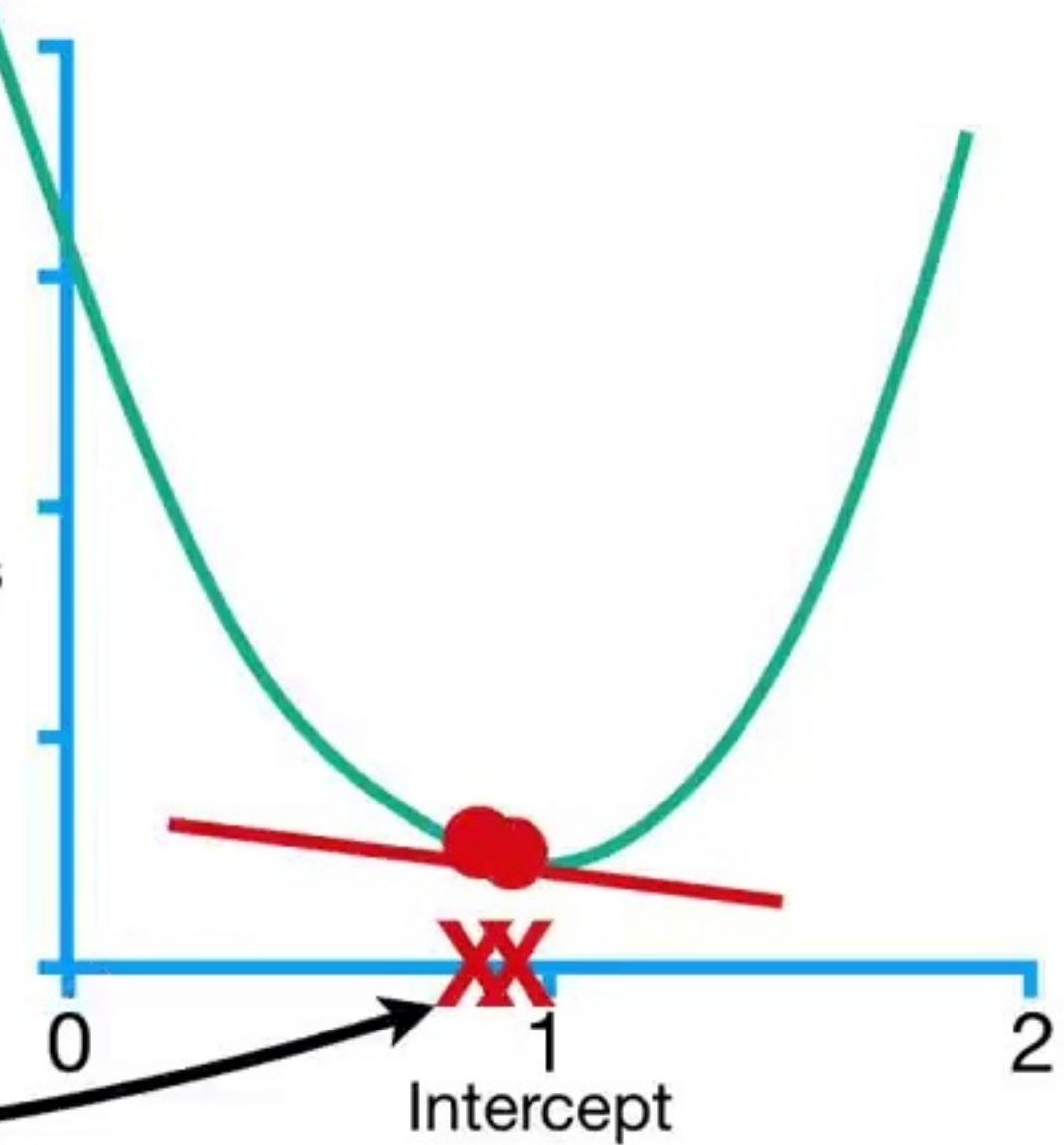


$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =
 $-2(1.4 - (0 + 0.64 \times 0.5))$
 $+ -2(1.9 - (0 + 0.64 \times 2.3))$
 $+ -2(3.2 - (0 + 0.64 \times 2.9))$
 $= -5.7$

...then we should take baby steps, because we are close to the optimal value...

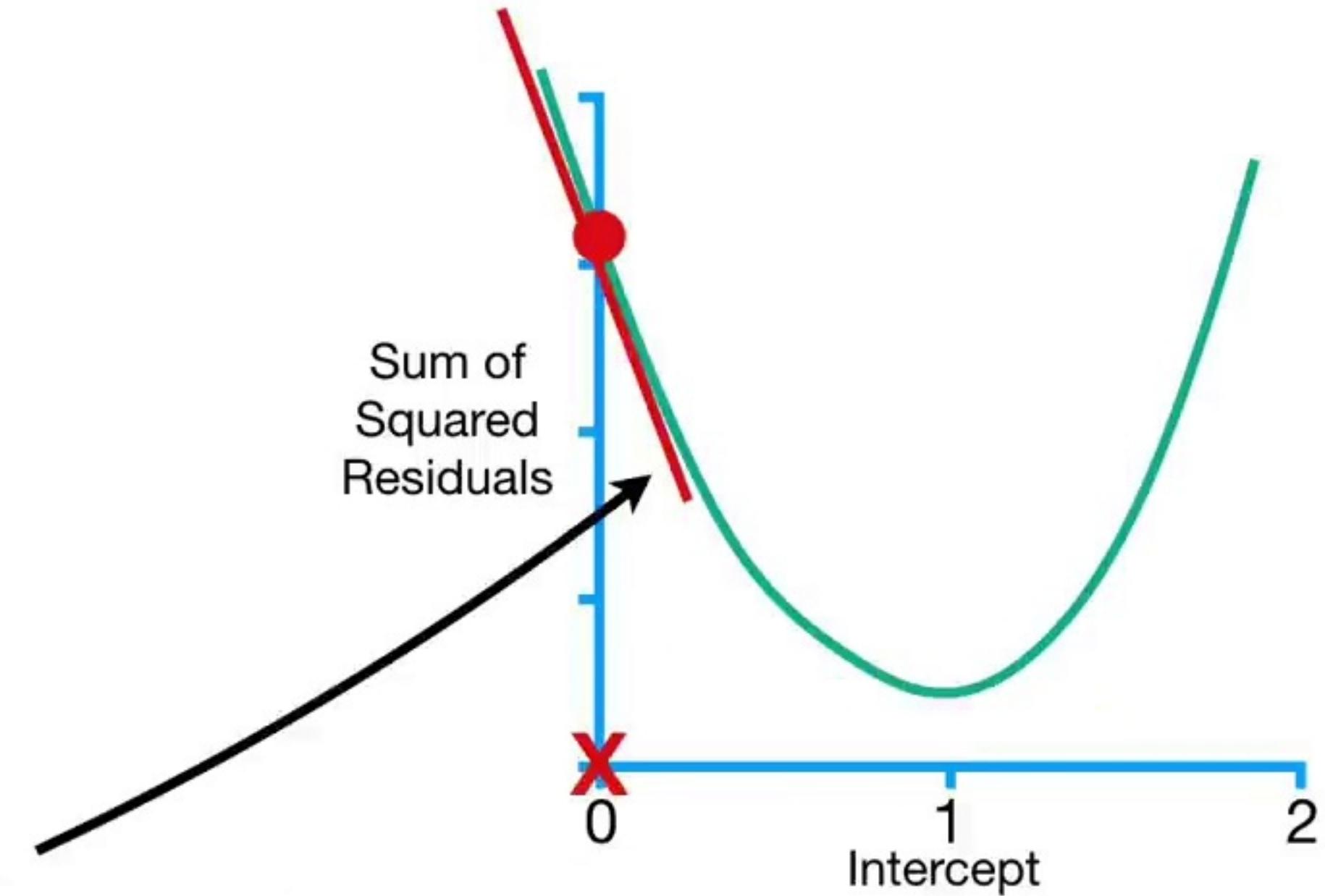
Sum of Squared Residuals



$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =
 $-2(1.4 - (0 + 0.64 \times 0.5))$
 $+ -2(1.9 - (0 + 0.64 \times 2.3))$
 $+ -2(3.2 - (0 + 0.64 \times 2.9))$
 $= -5.7$

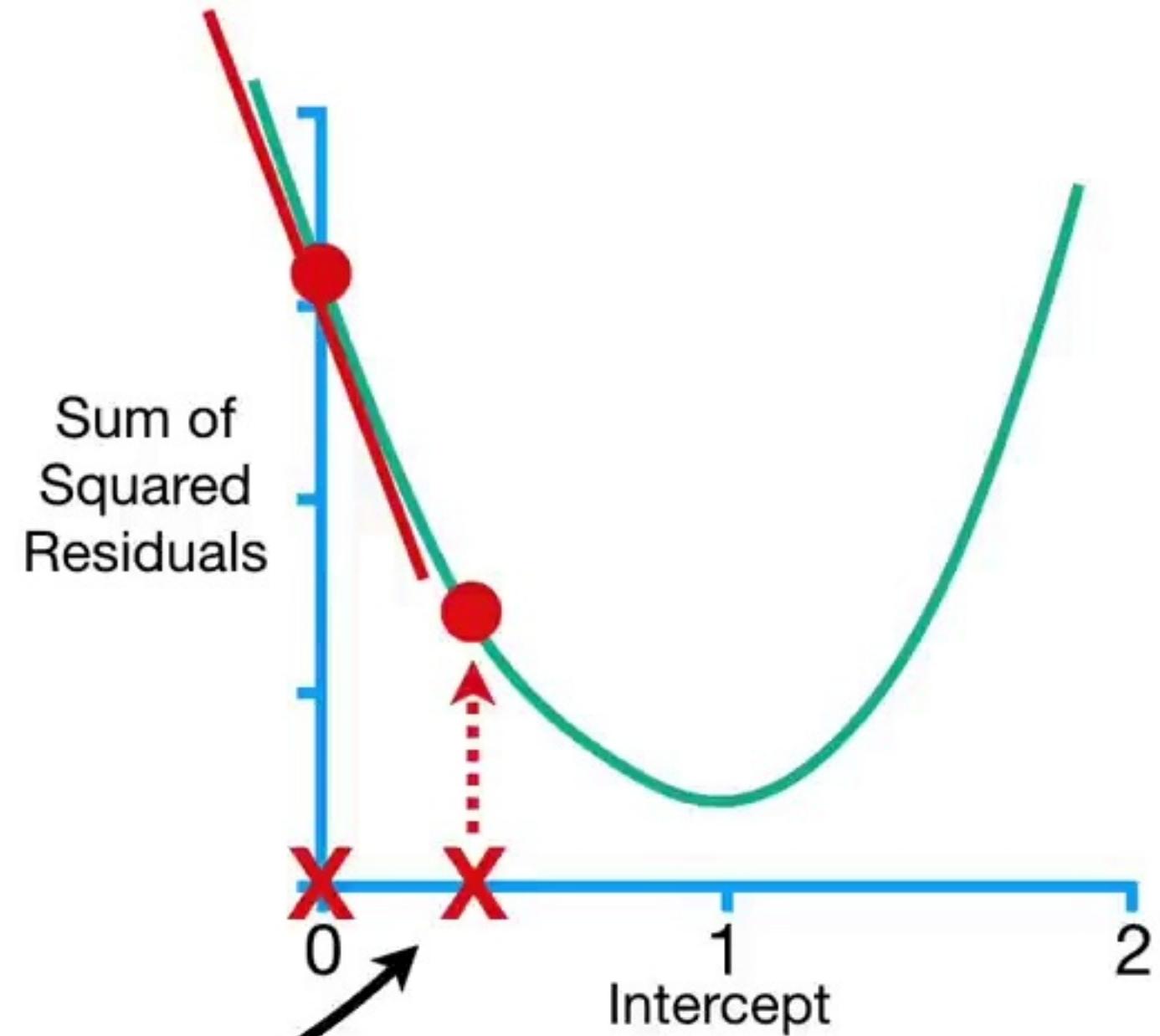
...and when the slope is
far from 0...



$$\frac{d}{d \text{ intercept}}$$

$$\begin{aligned}\text{Sum of squared residuals} &= \\ &-2(1.4 - (0 + 0.64 \times 0.5)) \\ &+ -2(1.9 - (0 + 0.64 \times 2.3)) \\ &+ -2(3.2 - (0 + 0.64 \times 2.9)) \\ &= -5.7\end{aligned}$$

...then we should take big steps,
because we are far from the
optimal value.



$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =

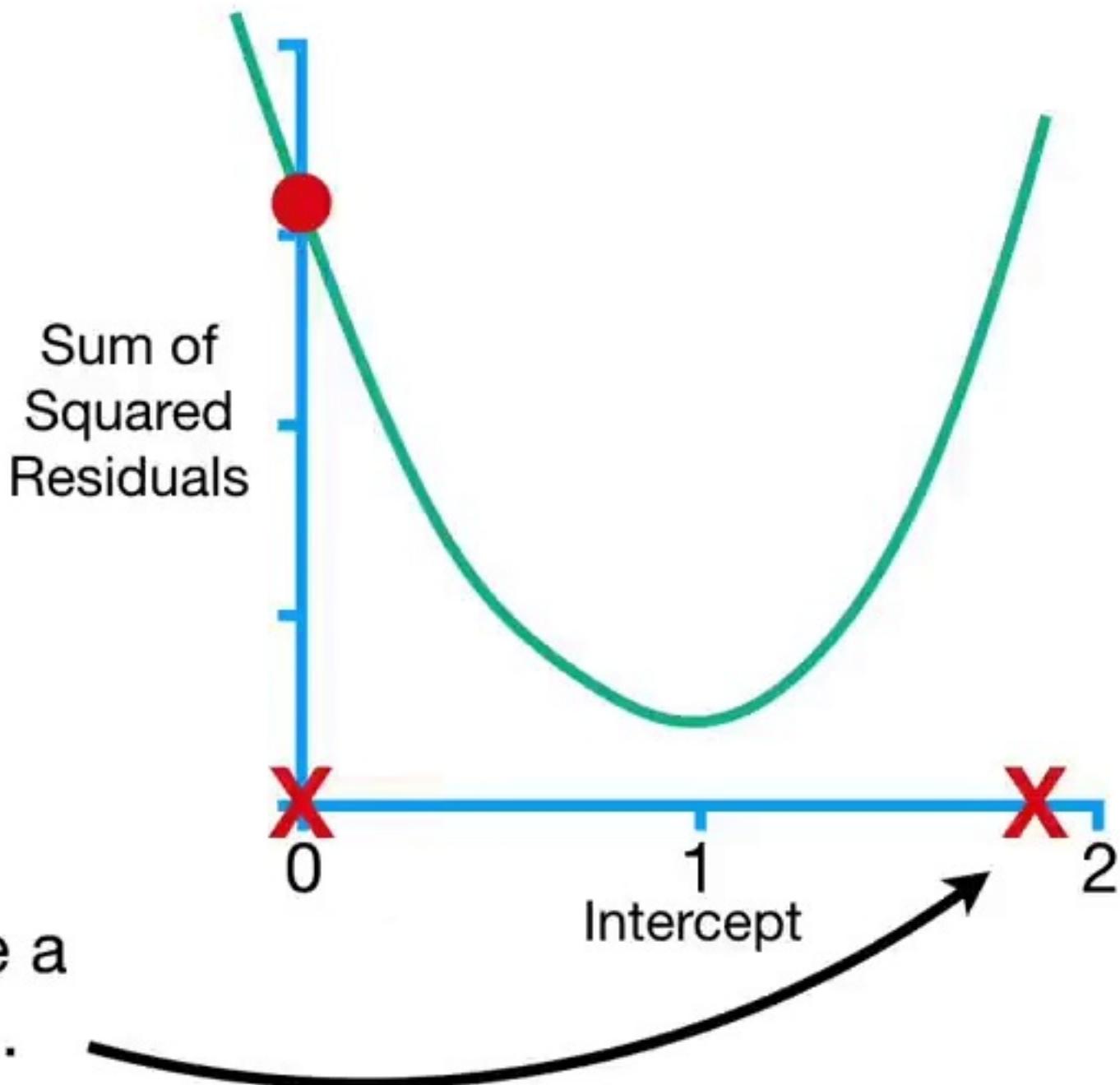
$$-2(1.4 - (0 + 0.64 \times 0.5))$$

$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$

$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$

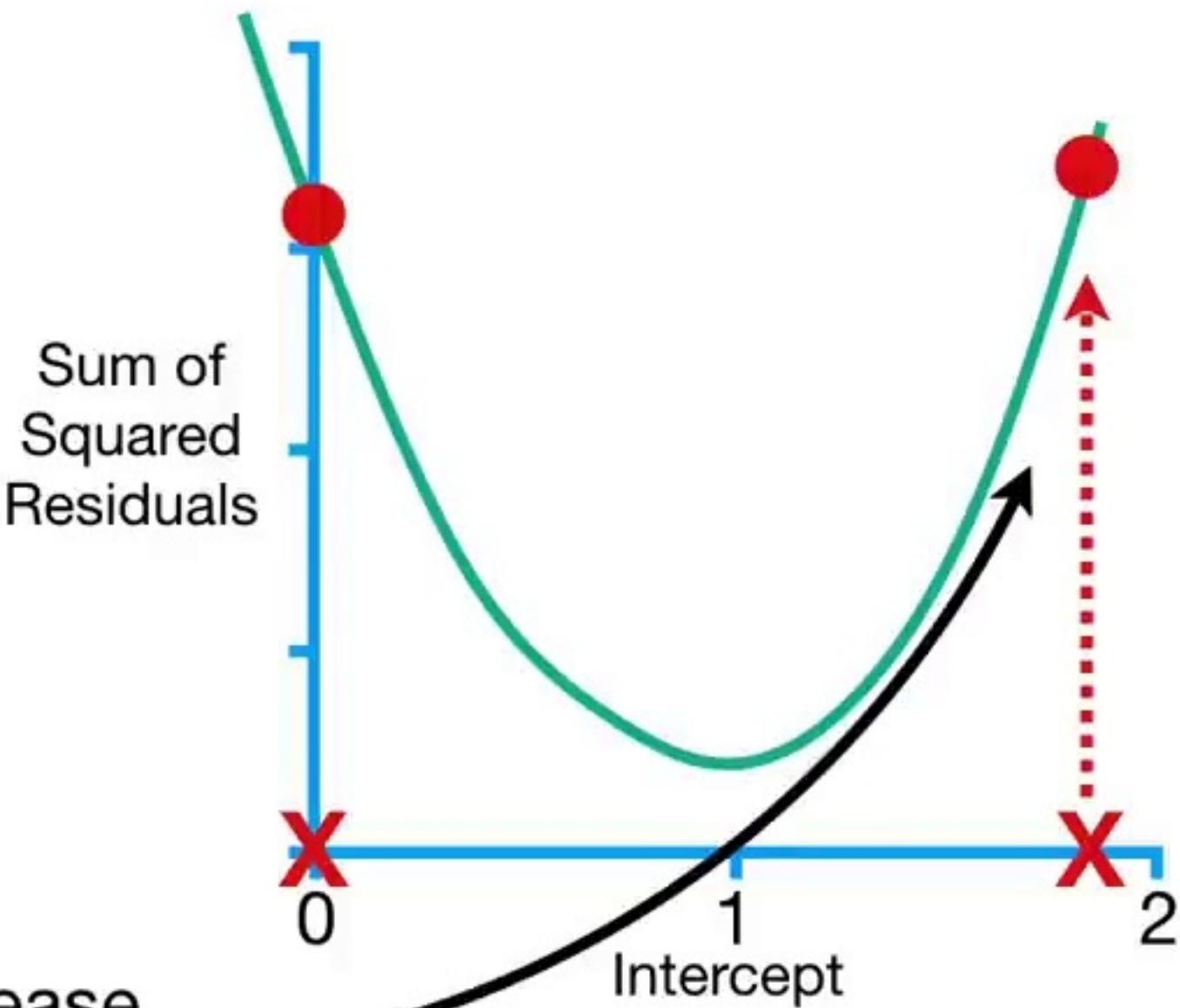
$$= -5.7$$

However, if we take a
super huge step...



$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

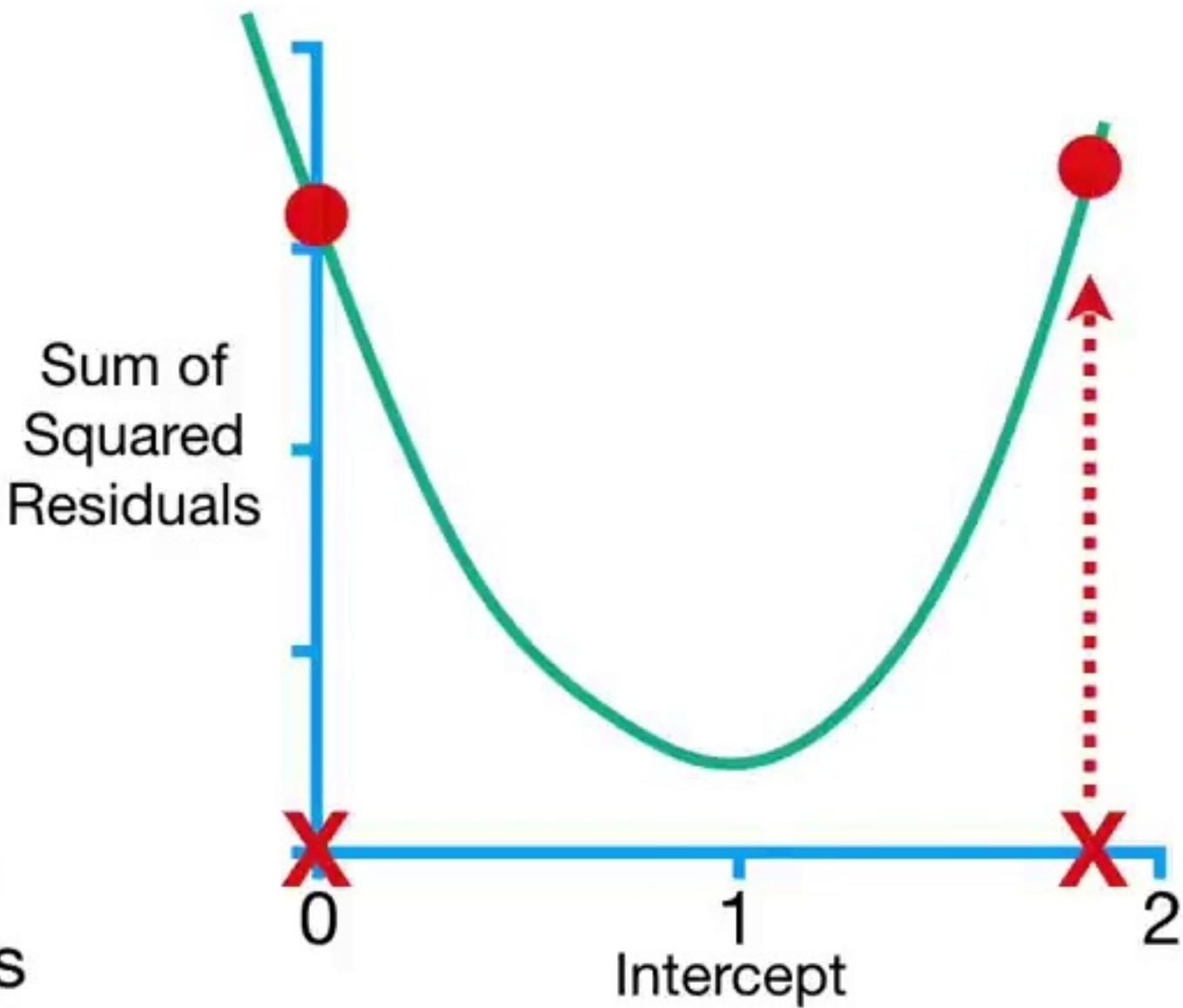
...then we would increase
the Sum of the Squared
Residuals!



$$\frac{d}{d \text{ intercept}}$$

$$\begin{aligned}\text{Sum of squared residuals} &= \\ &-2(1.4 - (0 + 0.64 \times 0.5)) \\ &+ -2(1.9 - (0 + 0.64 \times 2.3)) \\ &+ -2(3.2 - (0 + 0.64 \times 2.9)) \\ &= -5.7\end{aligned}$$

So the size of the step should be related to the slope, since it tells us if we should take a baby step or a big step, but we need to make sure the big step is not too big.



$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =

$$-2(1.4 - (0 + 0.64 \times 0.5))$$

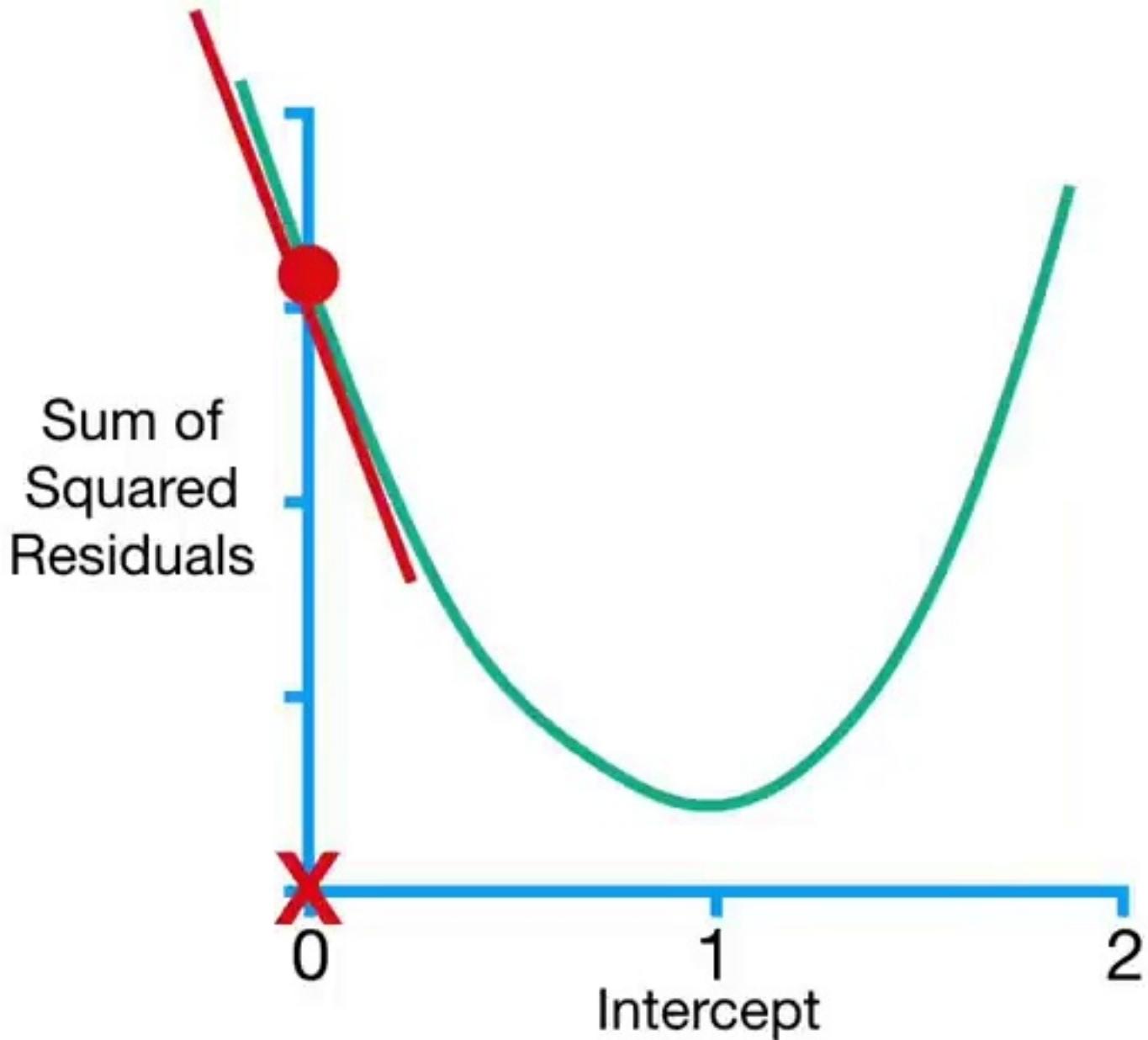
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$

$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$

$$= -5.7$$

Step Size = -5.7

Gradient Descent determines the
Step Size by multiplying the **slope**...



$$\frac{d}{d \text{ intercept}}$$

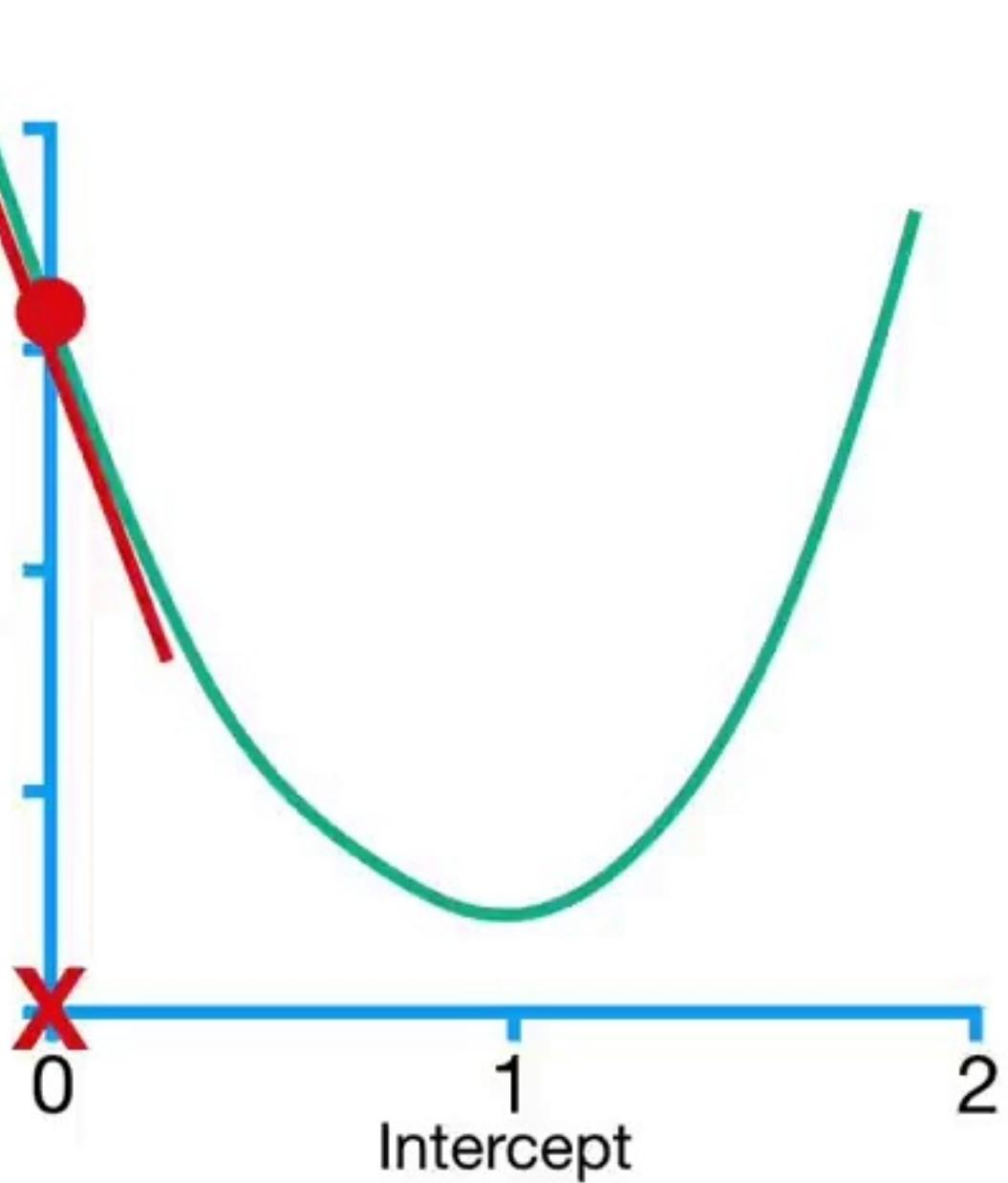
Sum of squared residuals =
 $-2(1.4 - (0 + 0.64 \times 0.5))$
 $+ -2(1.9 - (0 + 0.64 \times 2.3))$
 $+ -2(3.2 - (0 + 0.64 \times 2.9))$
 $= -5.7$

Step Size = -5.7×0.1



...by a small number called
The Learning Rate.

Sum of
Squared
Residuals



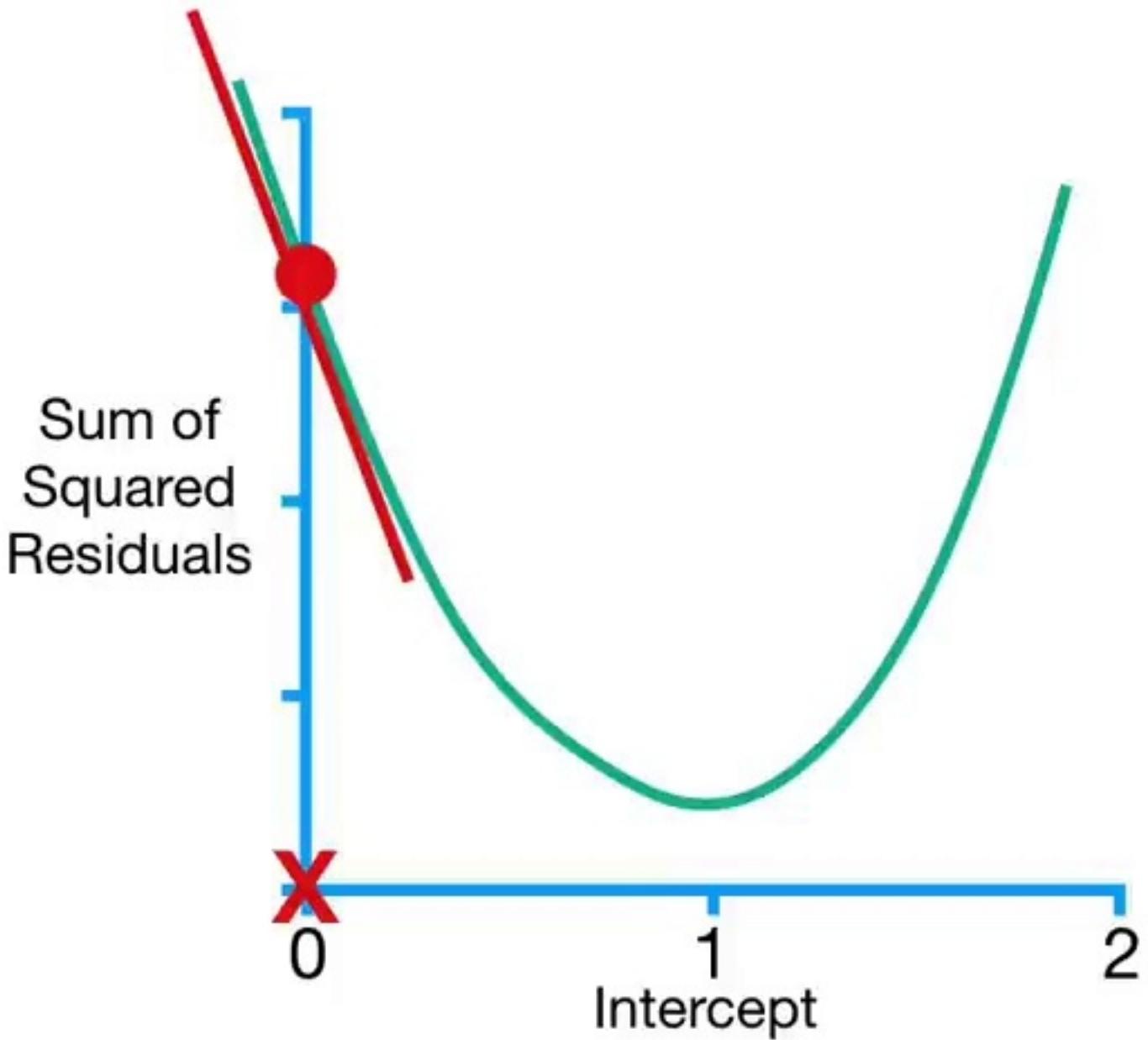
$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =
 $-2(1.4 - (0 + 0.64 \times 0.5))$
 $+ -2(1.9 - (0 + 0.64 \times 2.3))$
 $+ -2(3.2 - (0 + 0.64 \times 2.9))$
 $= -5.7$

Step Size = $-5.7 \times 0.1 = -0.57$



When the **Intercept** = 0, the
Step Size = **-0.57**.

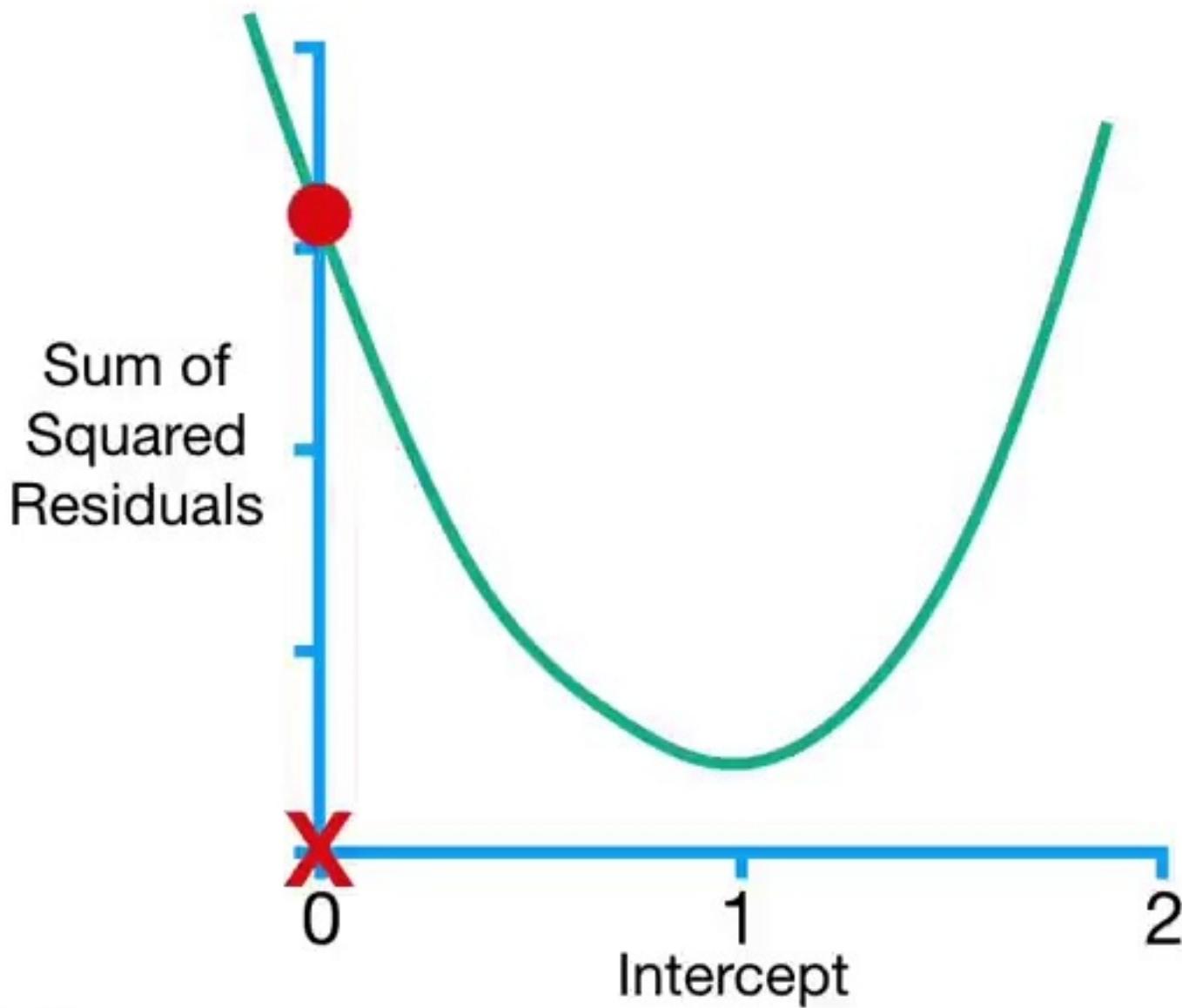


$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

$$\text{Step Size} = -5.7 \times 0.1 = -0.57$$

New Intercept = 

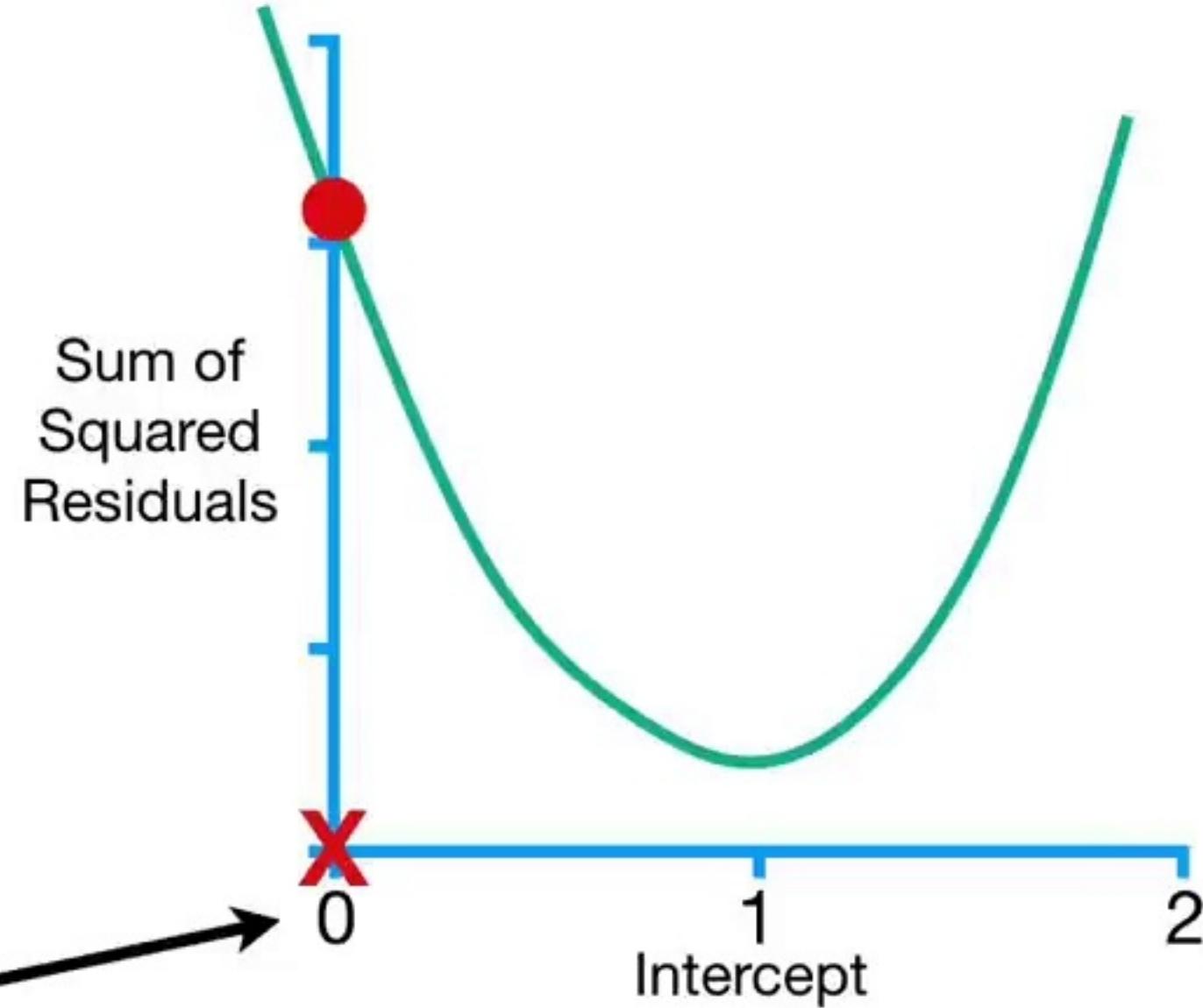
With the **Step Size**,
we can calculate a
New Intercept.



$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

$$\text{Step Size} = -5.7 \times 0.1 = -0.57$$

$$\text{New Intercept} = \text{Old Intercept} - \text{Step Size}$$

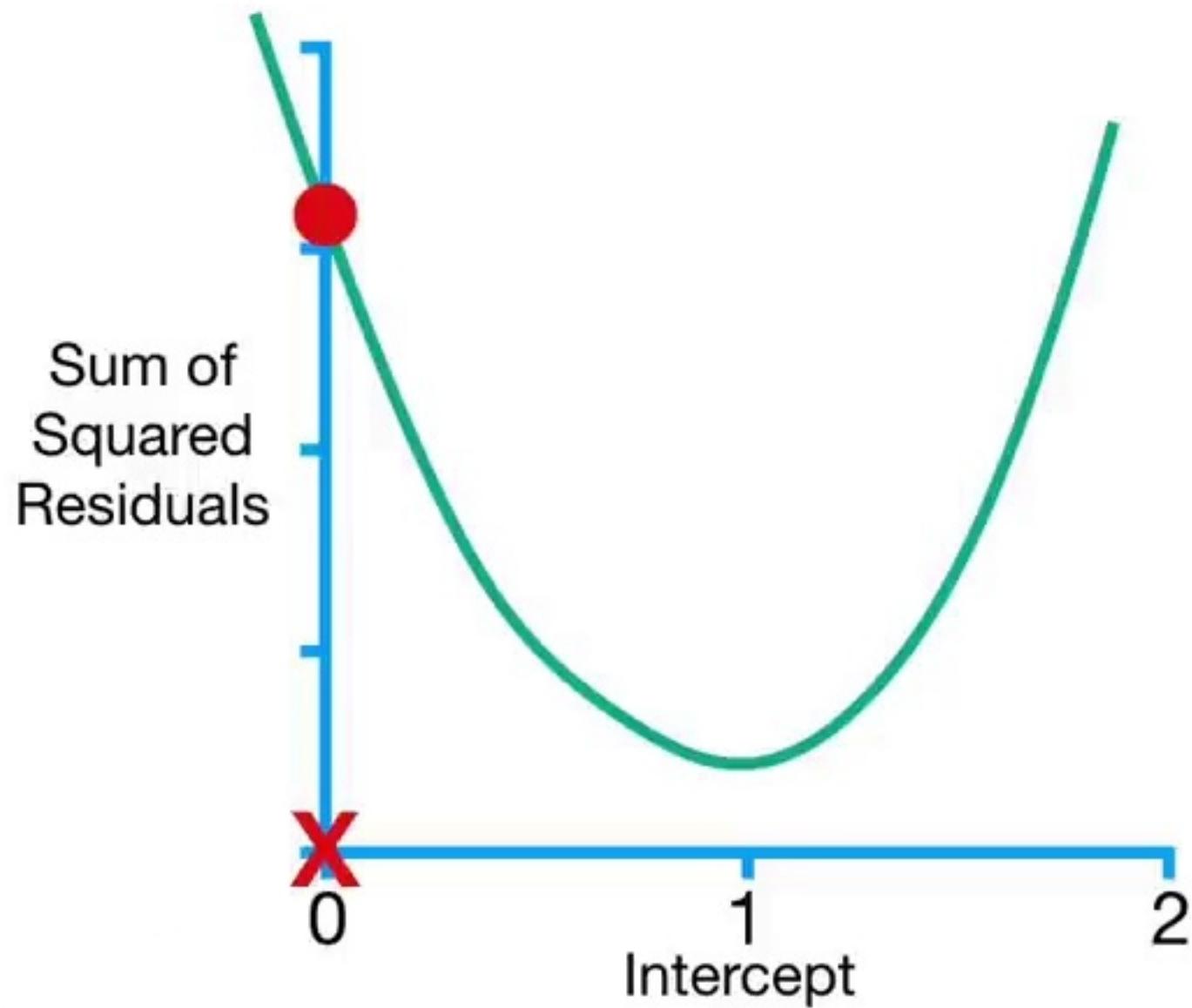


The **New Intercept** is
the **Old Intercept**...

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

$$\text{Step Size} = -5.7 \times 0.1 = \boxed{-0.57}$$

...minus the **Step Size**.

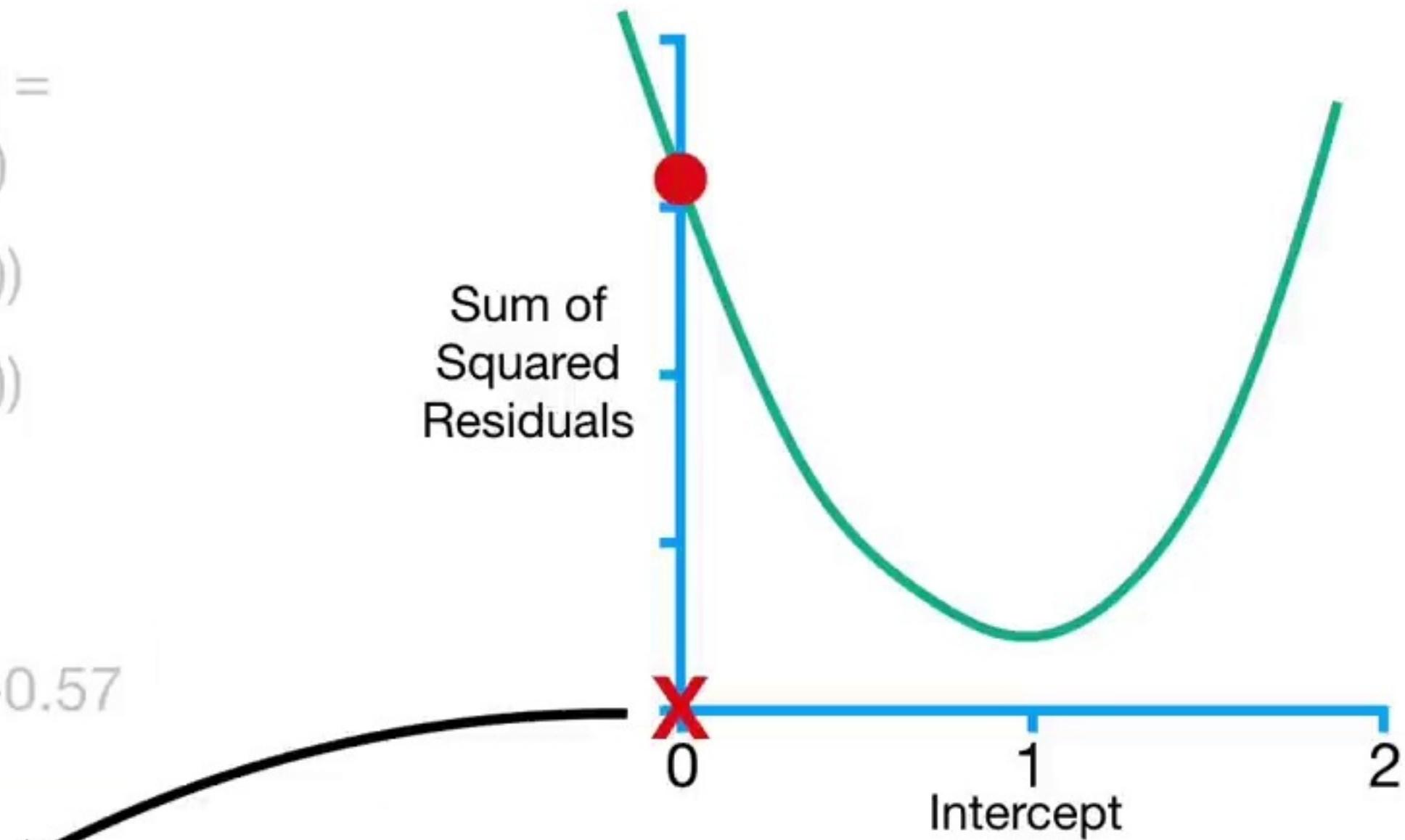


$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

$$\text{Step Size} = -5.7 \times 0.1 = -0.57$$

$$\text{New Intercept} = \text{Old Intercept} - \text{Step Size}$$

So we plug in the numbers...



$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =

$$-2(1.4 - (0 + 0.64 \times 0.5))$$

$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$

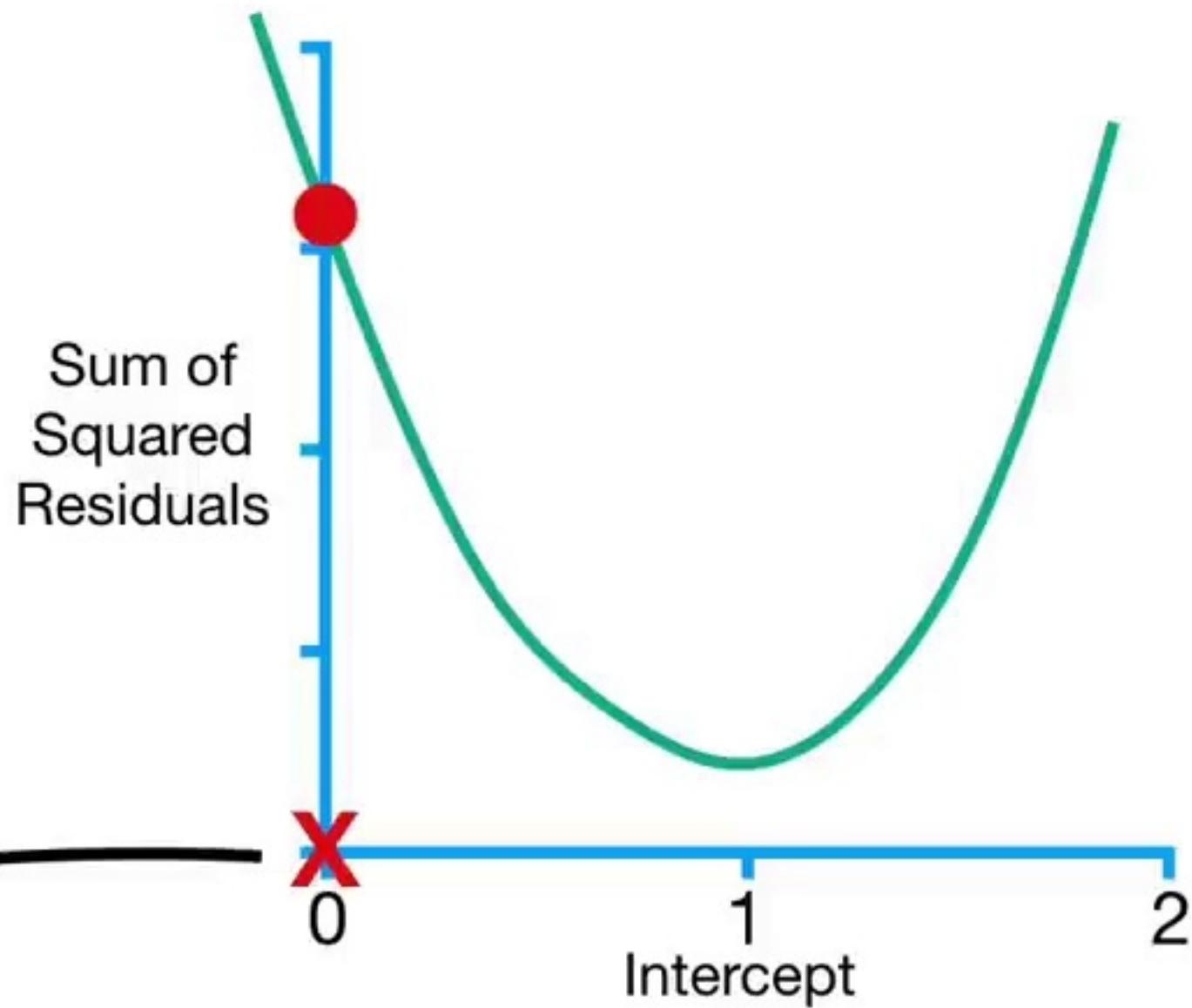
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$

$$= -5.7$$

$$\text{Step Size} = -5.7 \times 0.1 = -0.57$$

New Intercept = 0 - Step Size

So we plug in the numbers...



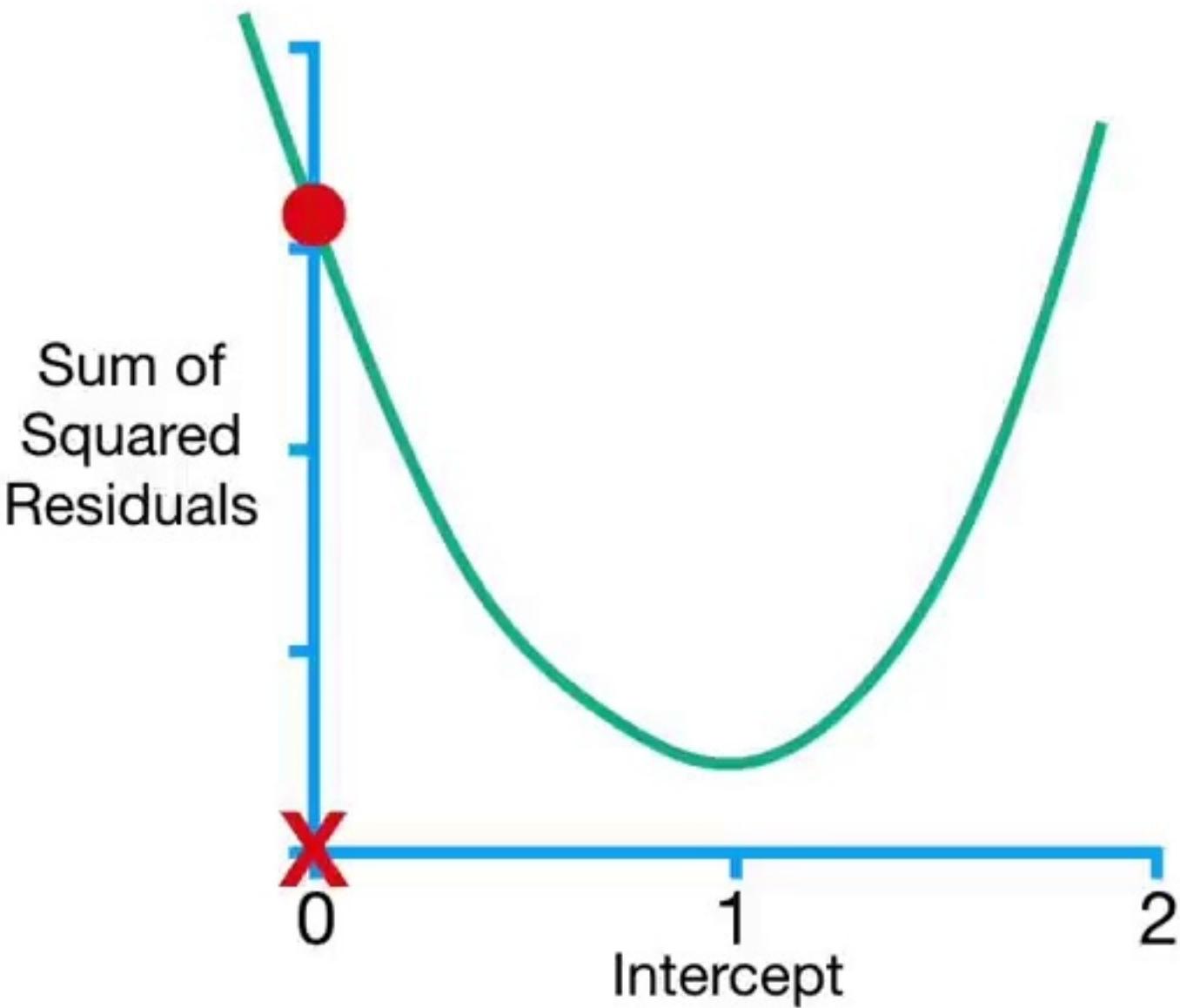
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

$$\text{Step Size} = -5.7 \times 0.1 = \boxed{-0.57}$$



$$\text{New Intercept} = 0 - \text{Step Size}$$

So we plug in the numbers...

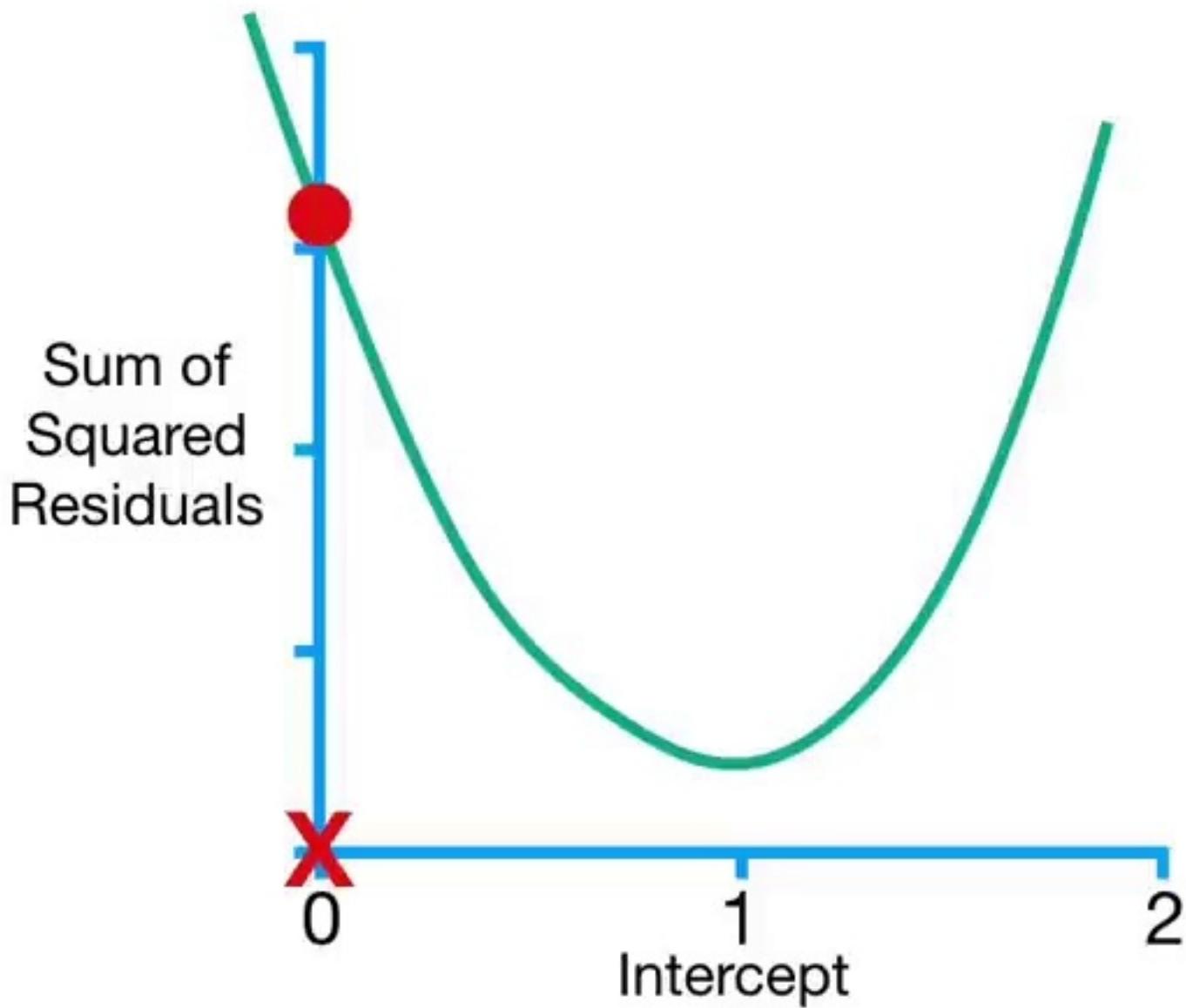


$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

$$\text{Step Size} = -5.7 \times 0.1 = \boxed{-0.57}$$

$$\text{New Intercept} = 0 - (-0.57)$$

So we plug in the numbers...

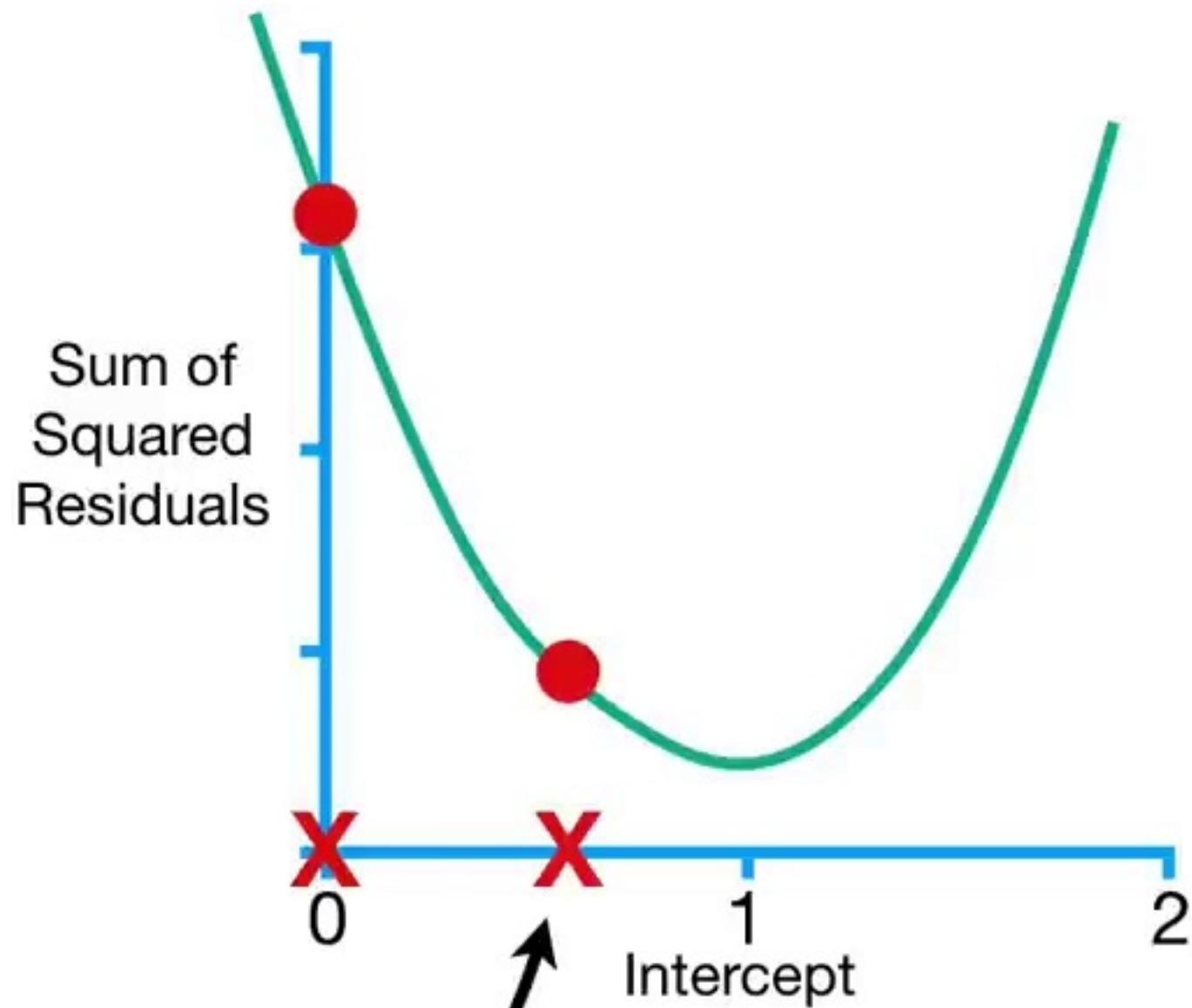


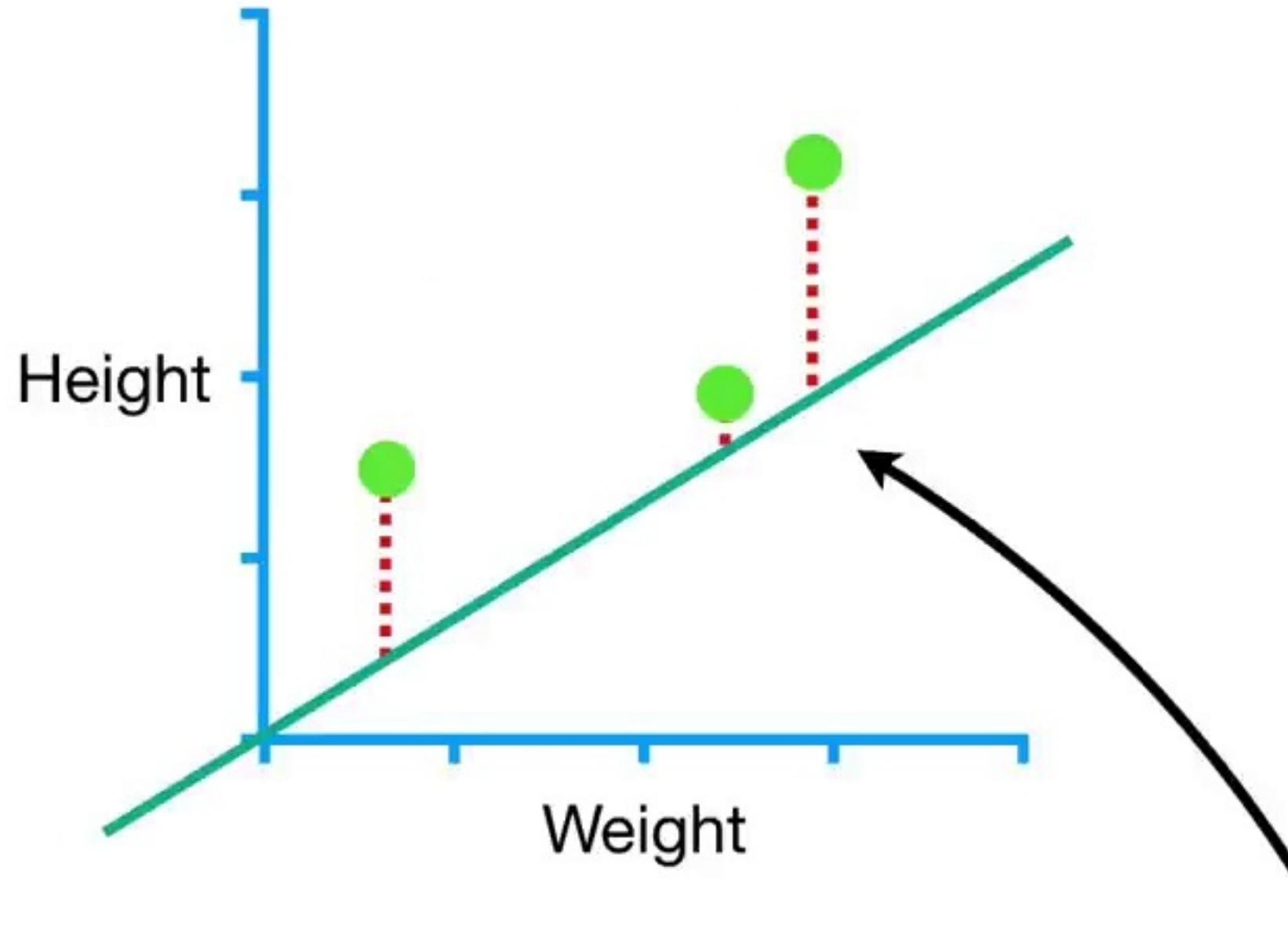
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0 + 0.64 \times 2.9))$$
$$= -5.7$$

$$\text{Step Size} = -5.7 \times 0.1 = -0.57$$

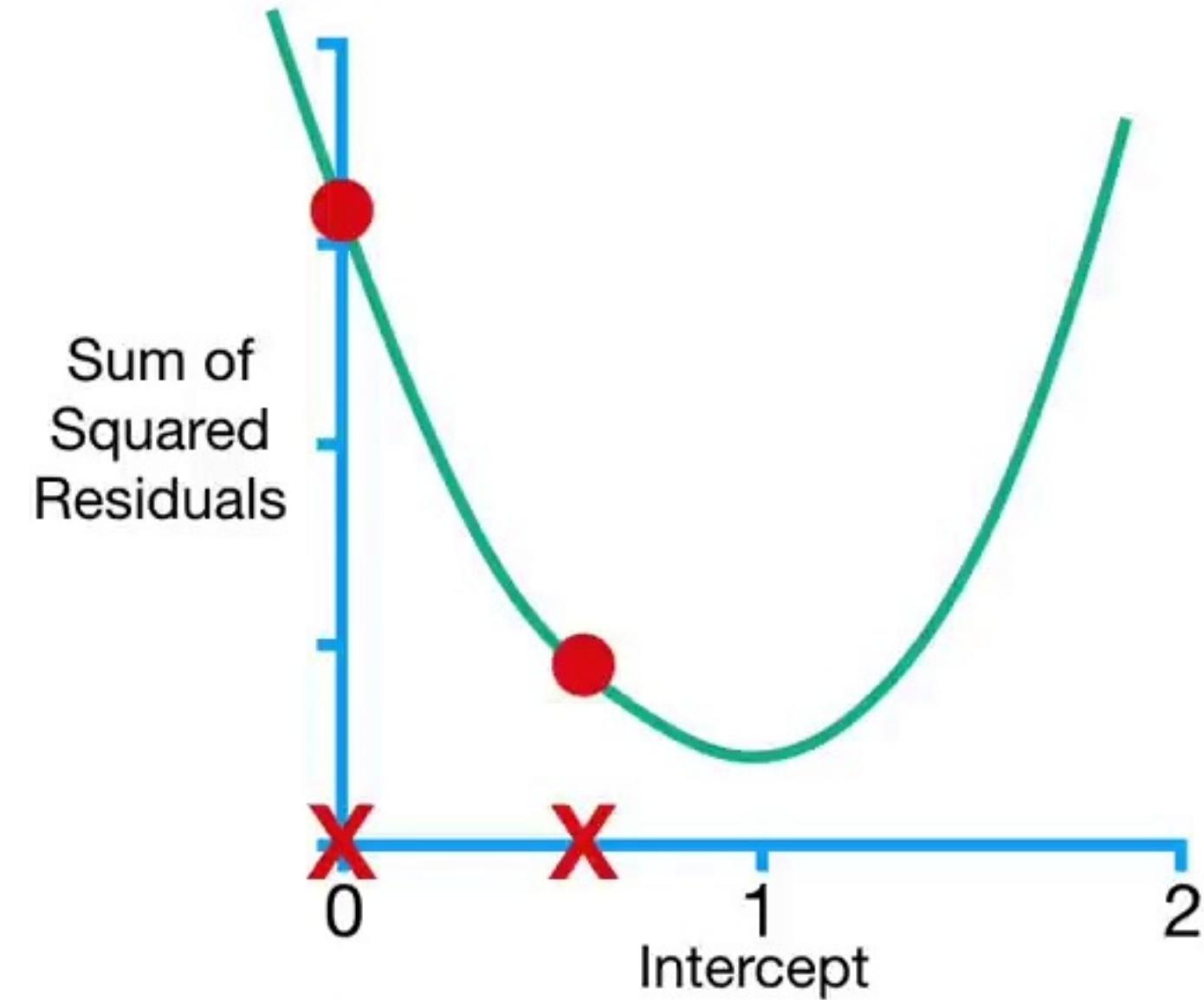
$$\text{New Intercept} = 0 - (-0.57) = \boxed{0.57}$$

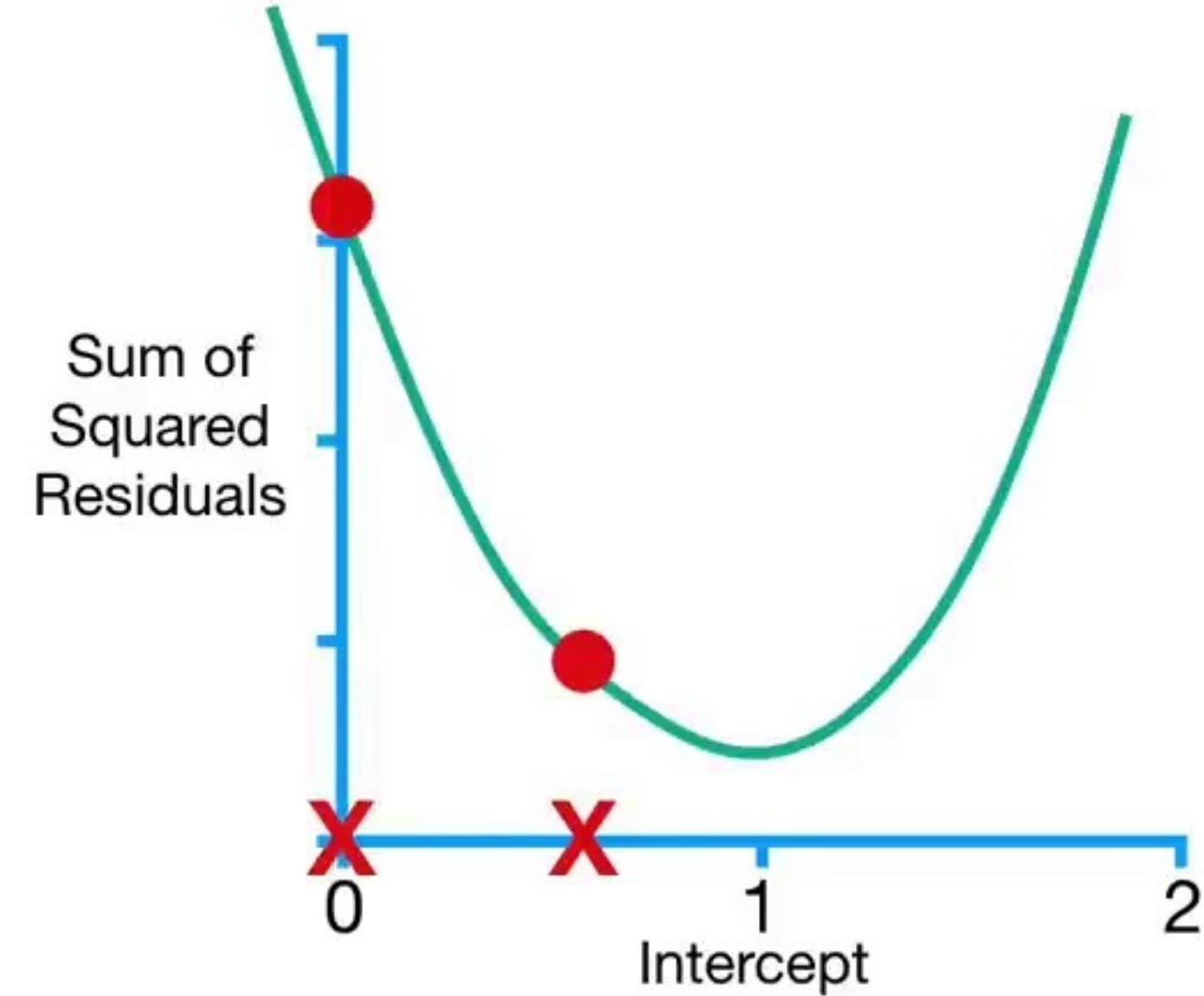
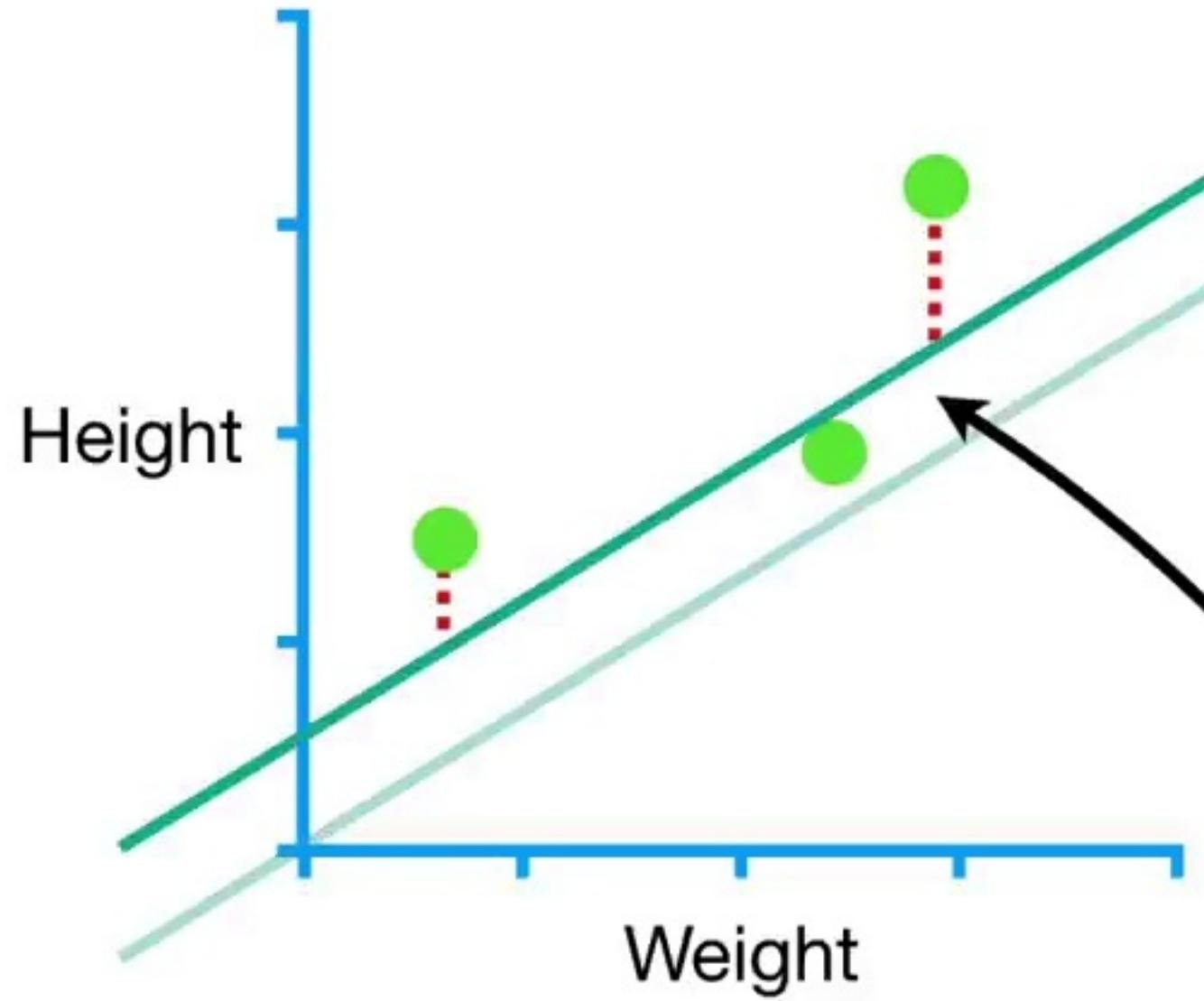
...and the New Intercept = 0.57.



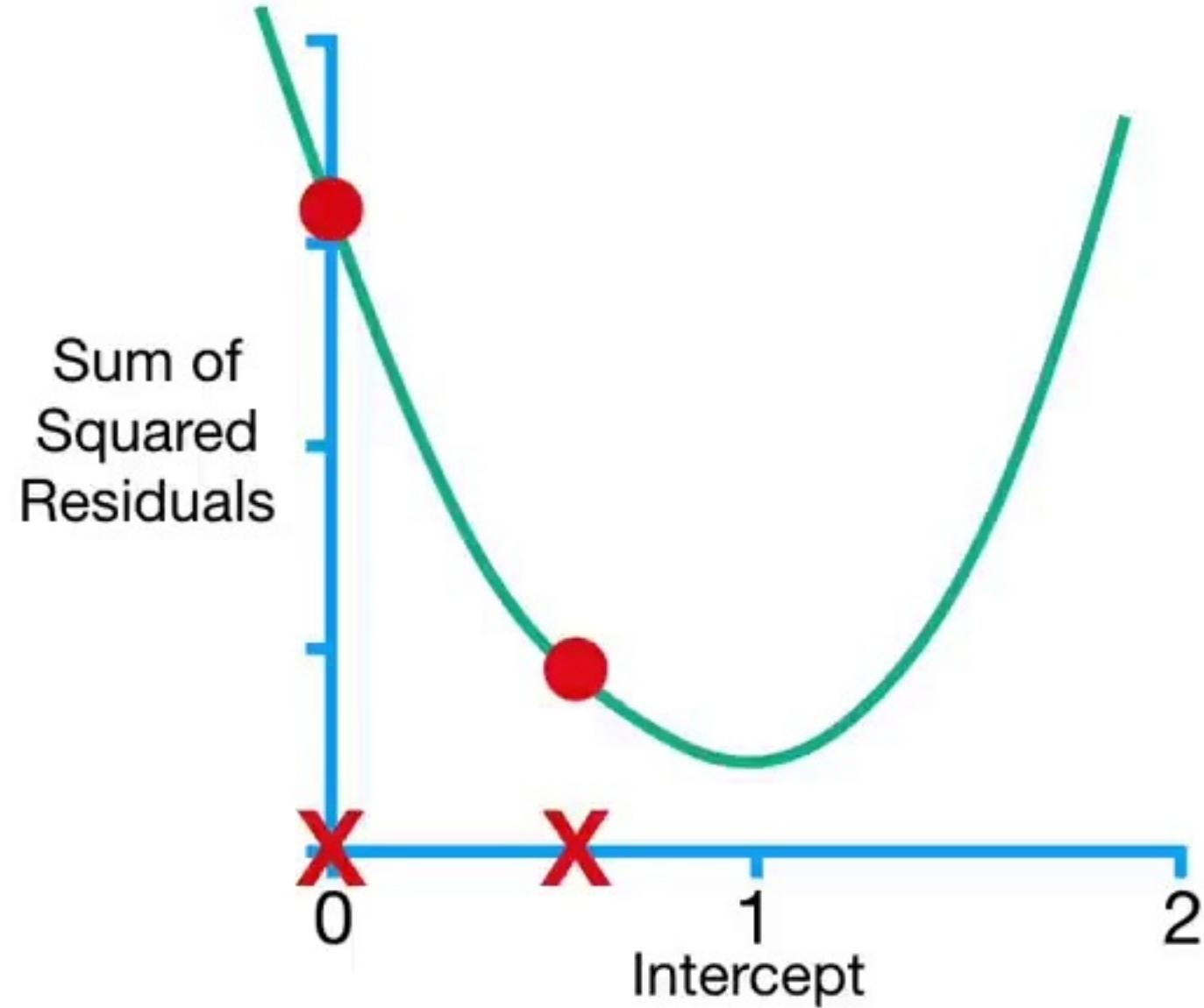
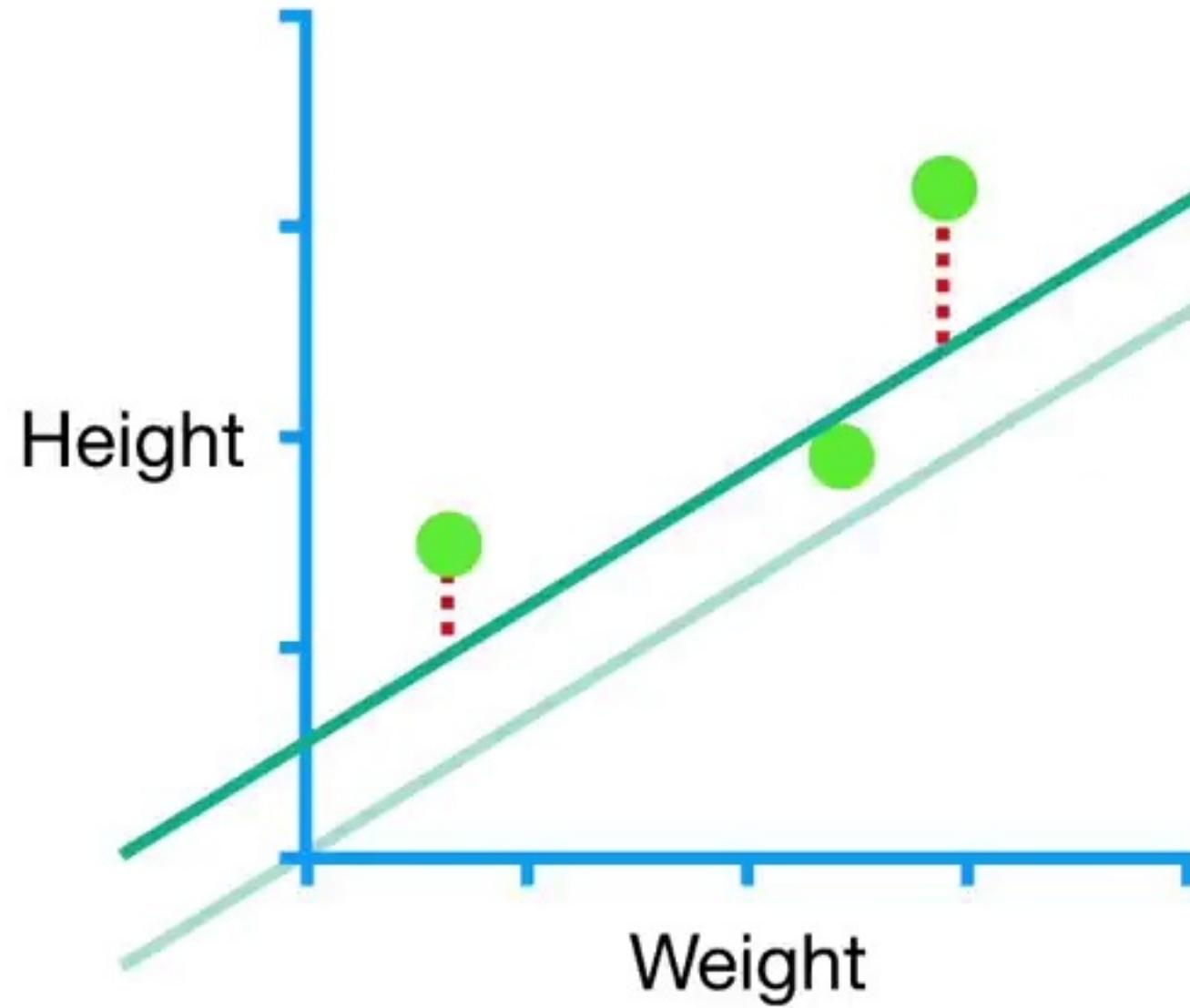


Going back to the original data and the original line, with the **Intercept = 0**...





...we can see how much the residuals shrink when the
Intercept = 0.57.



Now let's take another step
closer to the optimal value
for the **Intercept**.

$$\frac{d}{d \text{ intercept}}$$

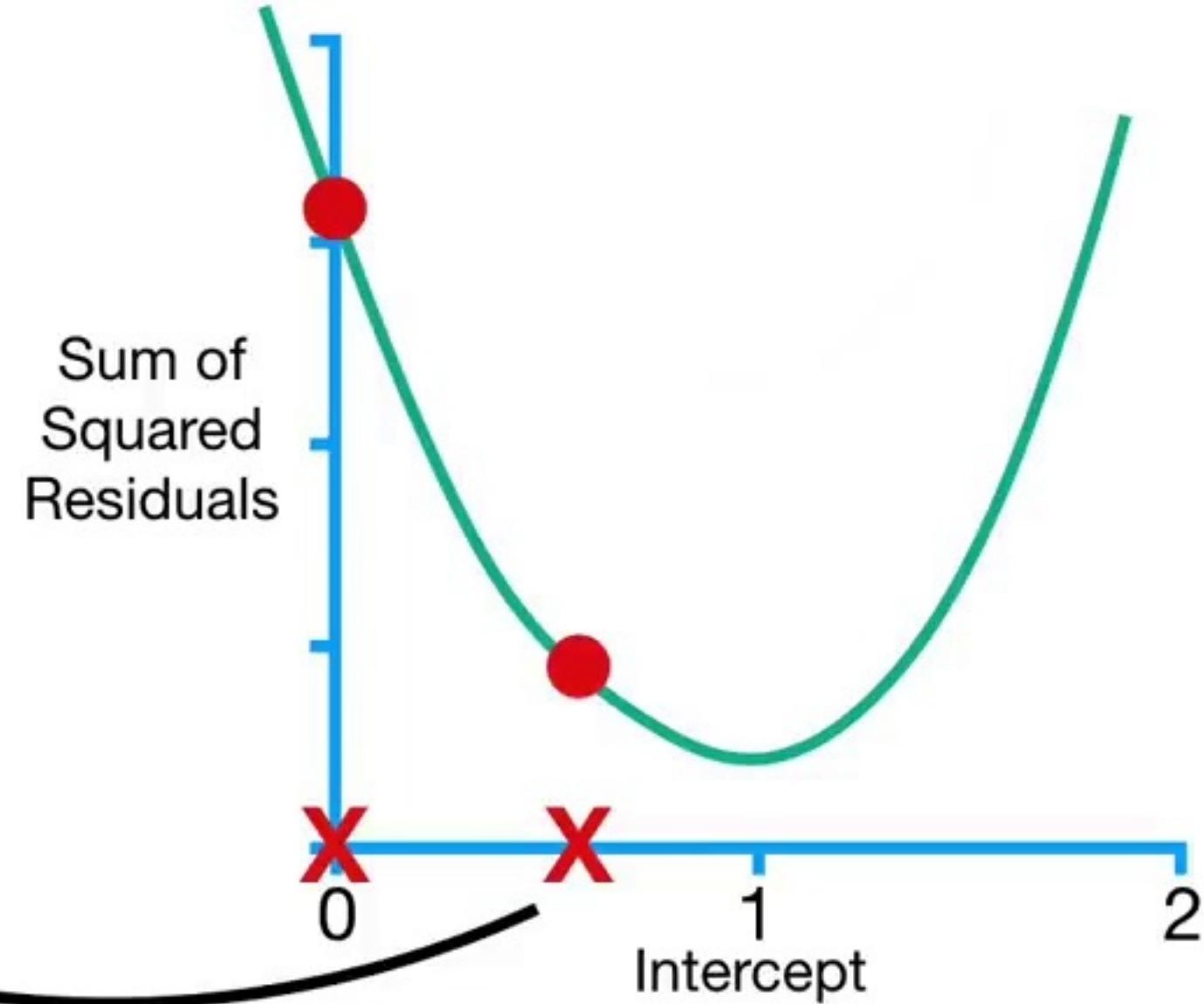
Sum of squared residuals =

$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$

$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

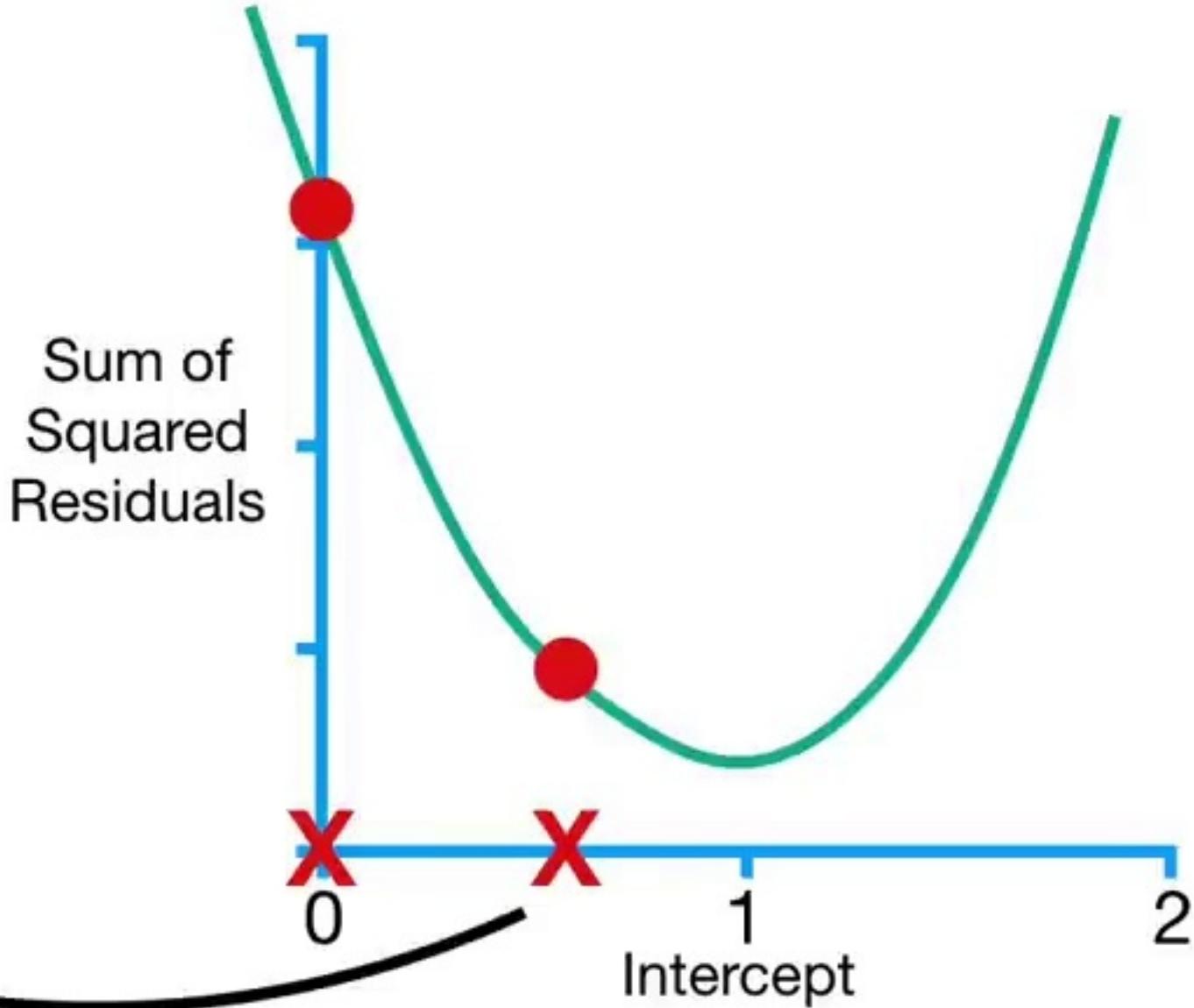
To take another step, we go back to the derivative and plug in the **New Intercept (0.57)**...



$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =
 $-2(1.4 - (0.57 + 0.64 \times 0.5))$
 $+ -2(1.9 - (0.57 + 0.64 \times 2.3))$
 $+ -2(3.2 - (0.57 + 0.64 \times 2.9))$

To take another step, we go back to the derivative and plug in the **New Intercept (0.57)**...



$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =

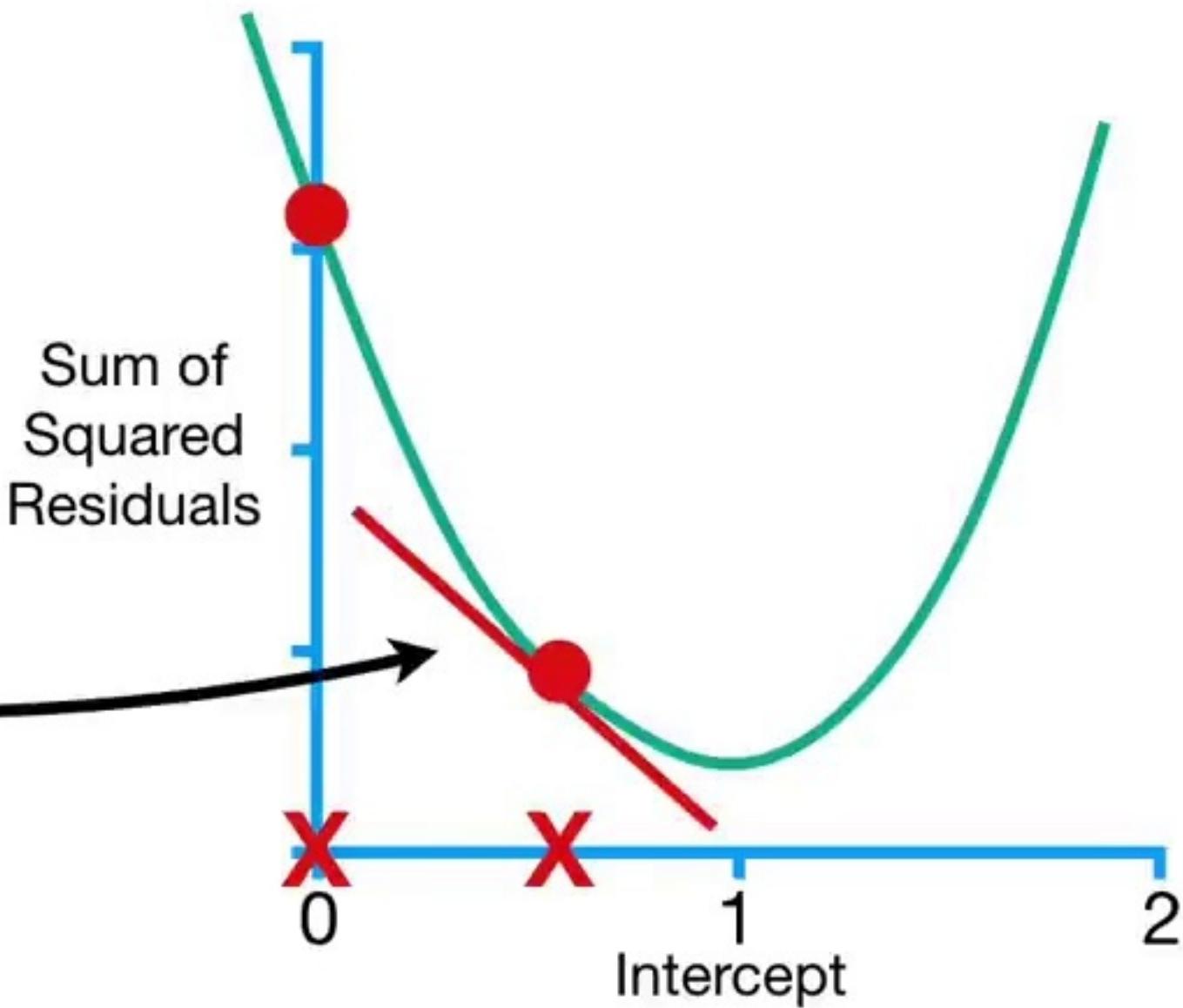
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$

$$+ -2(1.9 - (0.57 + 0.64 \times 2.3))$$

$$+ -2(3.2 - (0.57 + 0.64 \times 2.9))$$

$$= -2.3$$

...and that tells us the slope of the curve = **-2.3**.

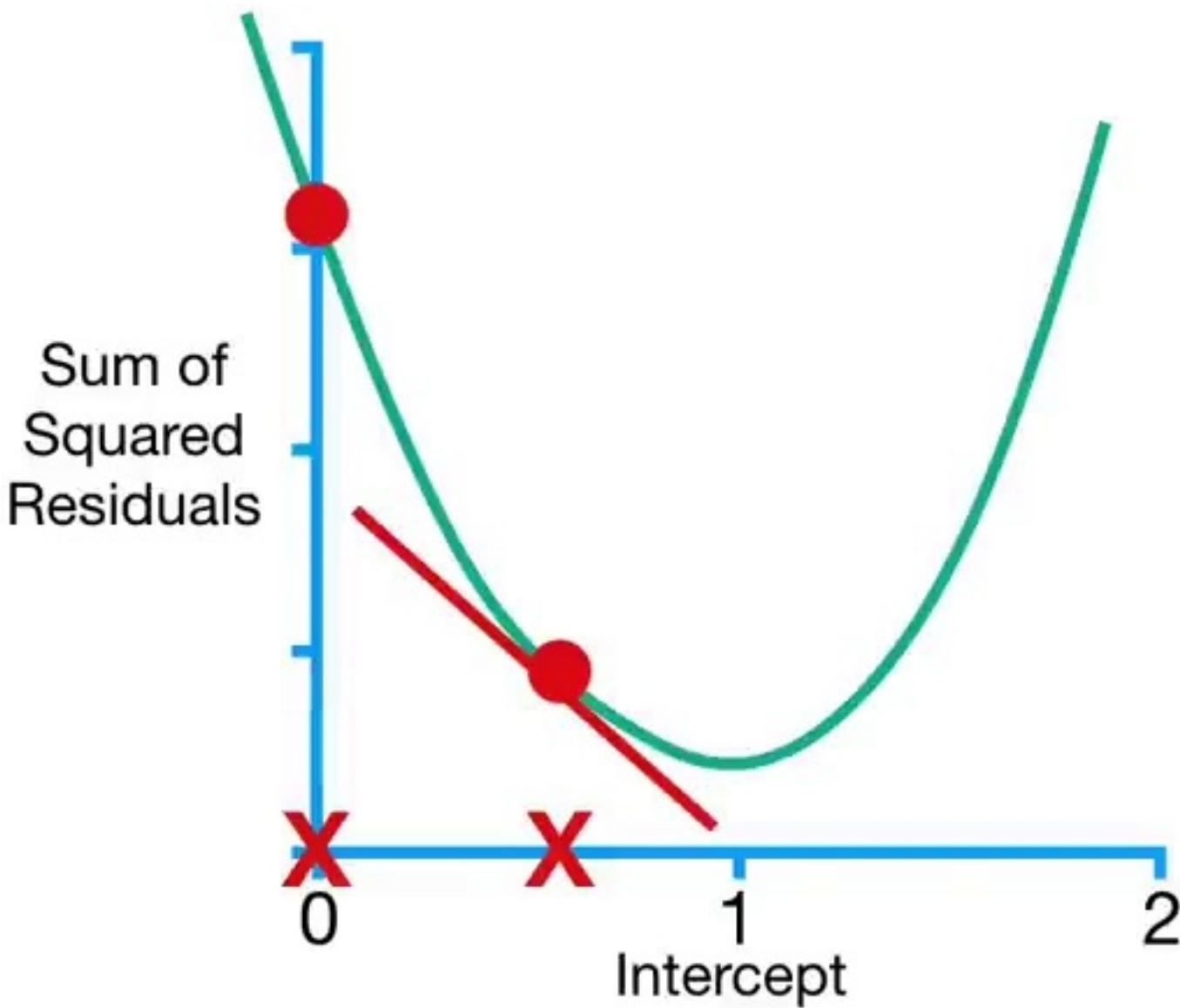


$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.57 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.57 + 0.64 \times 2.9))$$
$$= -2.3$$

Step Size = Slope × Learning Rate



Now let's calculate the
Step Size...

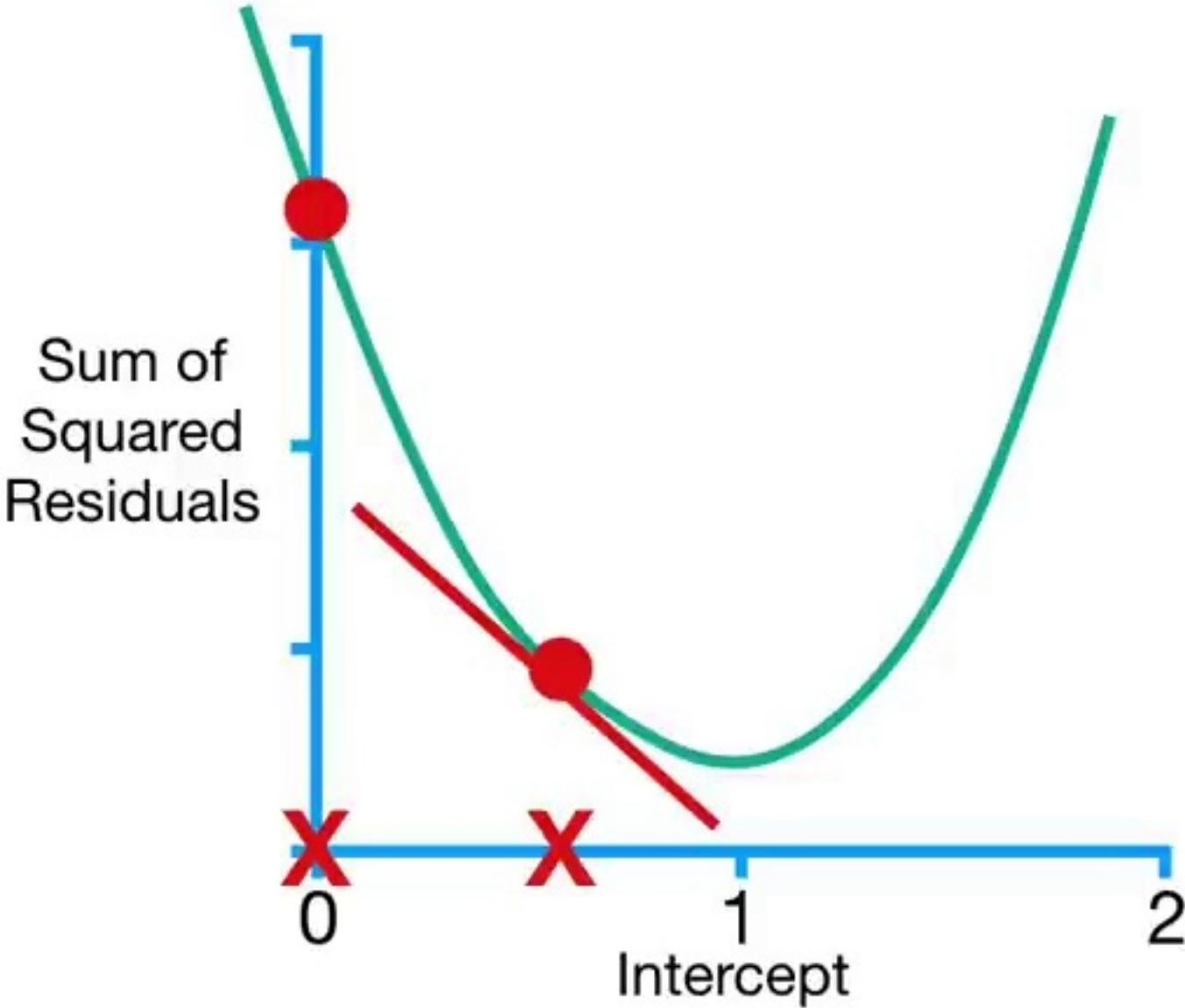


$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =
 $-2(1.4 - (0.57 + 0.64 \times 0.5))$
 $+ -2(1.9 - (0.57 + 0.64 \times 2.3))$
 $+ -2(3.2 - (0.57 + 0.64 \times 2.9))$
 $= -2.3$

Step Size = $-2.3 \times \text{Learning Rate}$

...by plugging in **-2.3** for
the **Slope**...

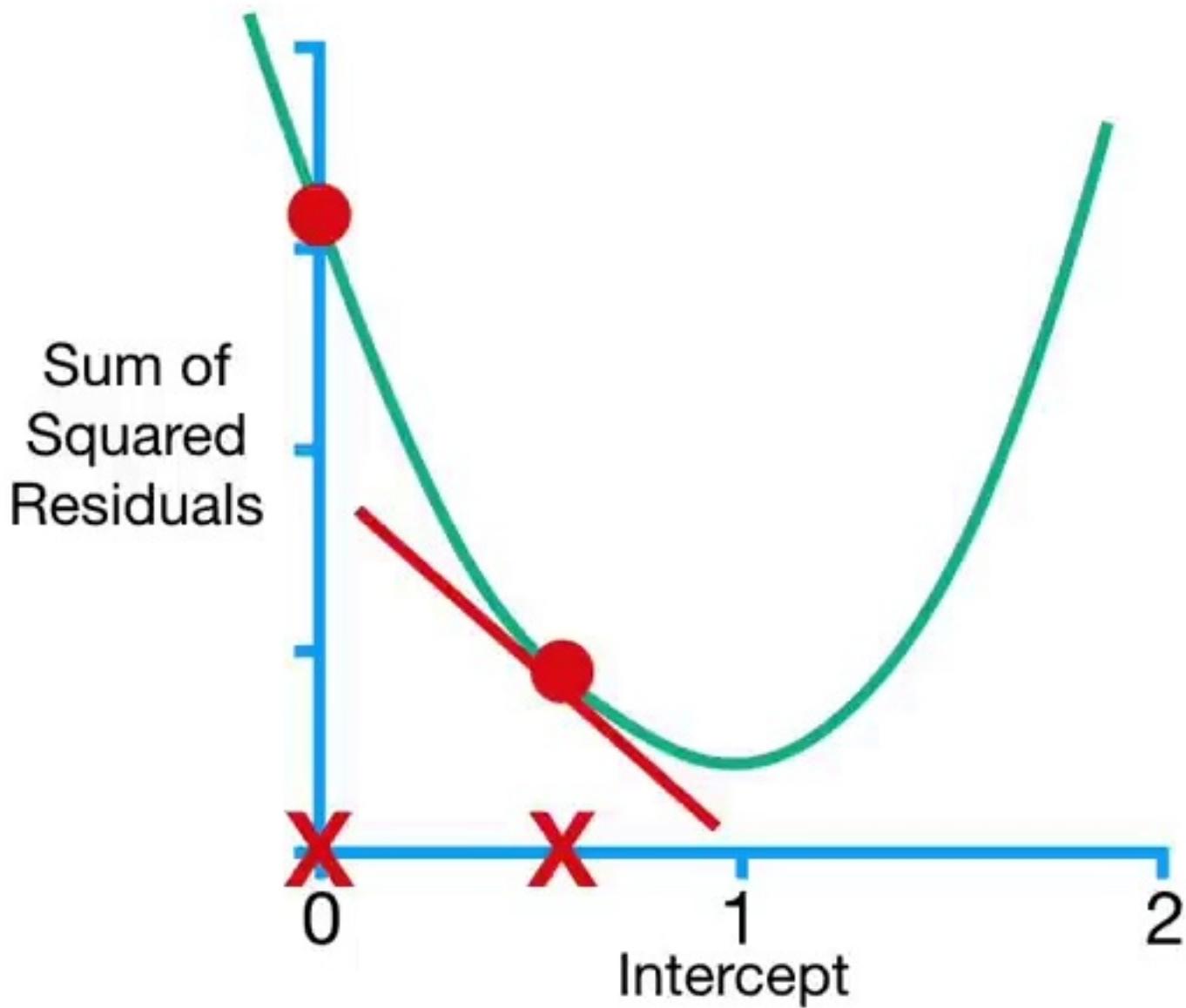


$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.57 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.57 + 0.64 \times 2.9))$$
$$= -2.3$$

Step Size = -2.3×0.1



...and **0.1** for the
Learning Rate.

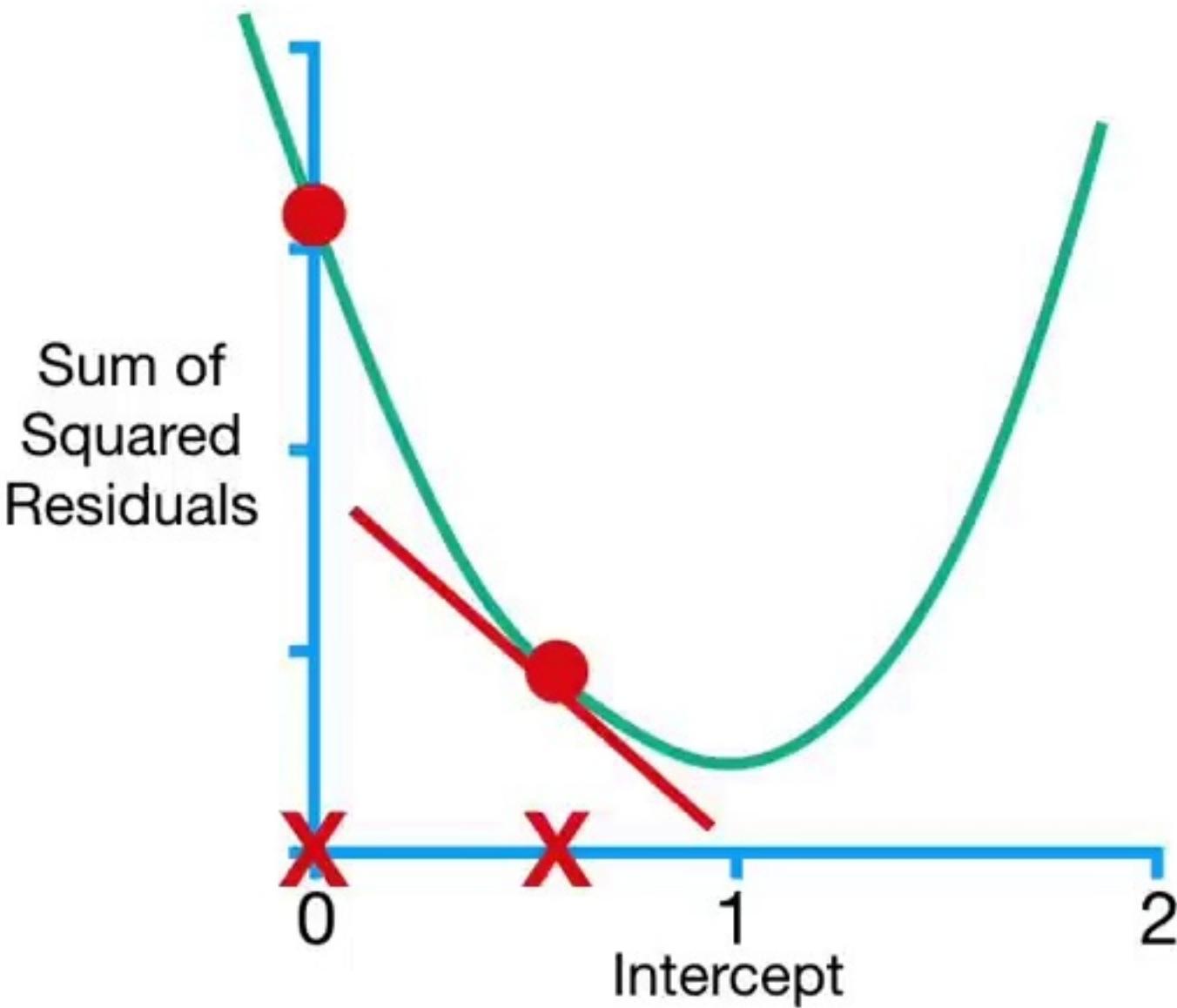


$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.57 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.57 + 0.64 \times 2.9))$$
$$= -2.3$$

$$\text{Step Size} = -2.3 \times 0.1 = -0.23$$



Ultimately, the **Step Size** is **-0.23**...

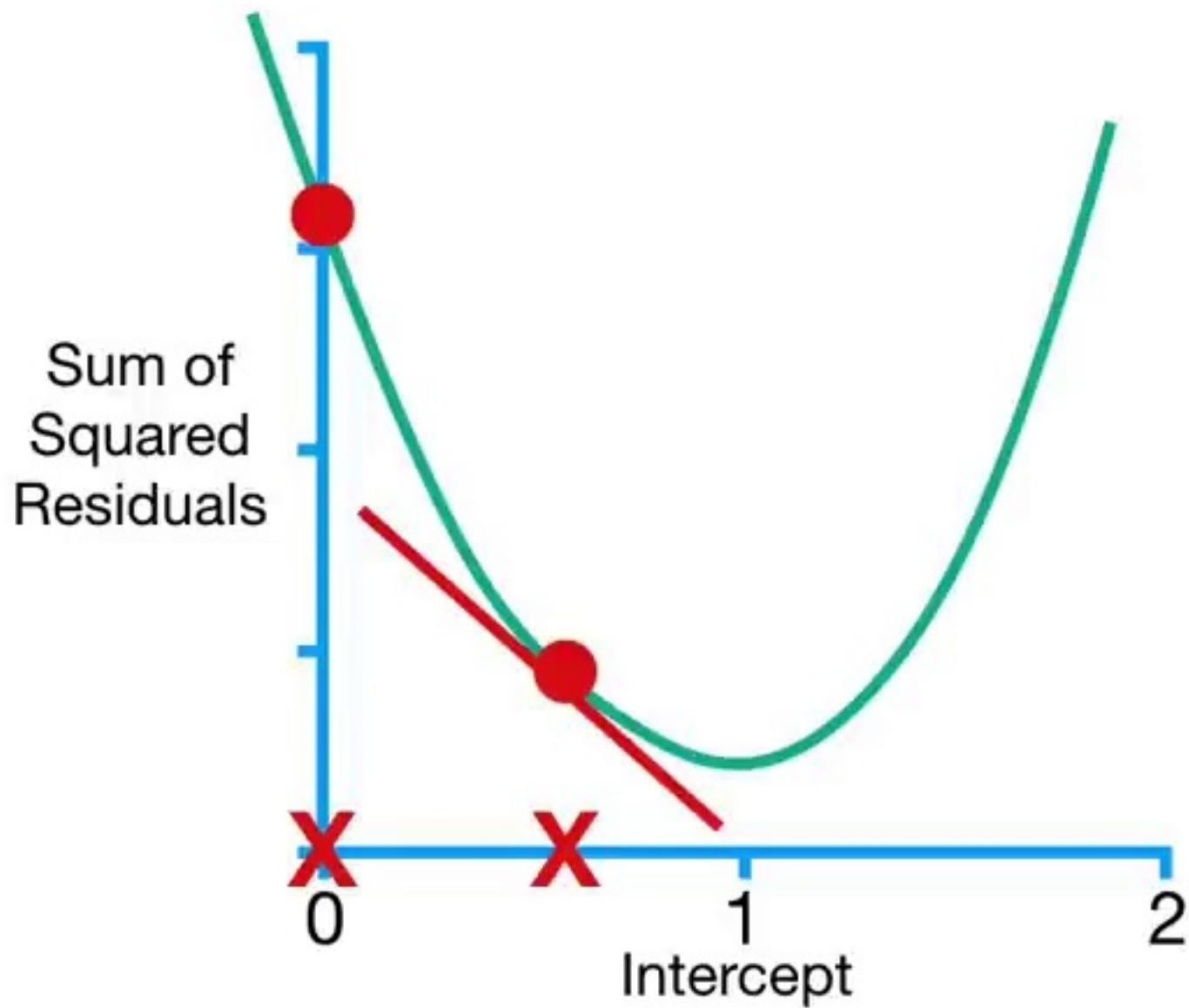


$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.57 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.57 + 0.64 \times 2.9))$$
$$= -2.3$$

$$\text{Step Size} = -2.3 \times 0.1 = -0.23$$

$$\text{New Intercept} = \text{Old Intercept} - \text{Step Size}$$

...and the **New Intercept...**

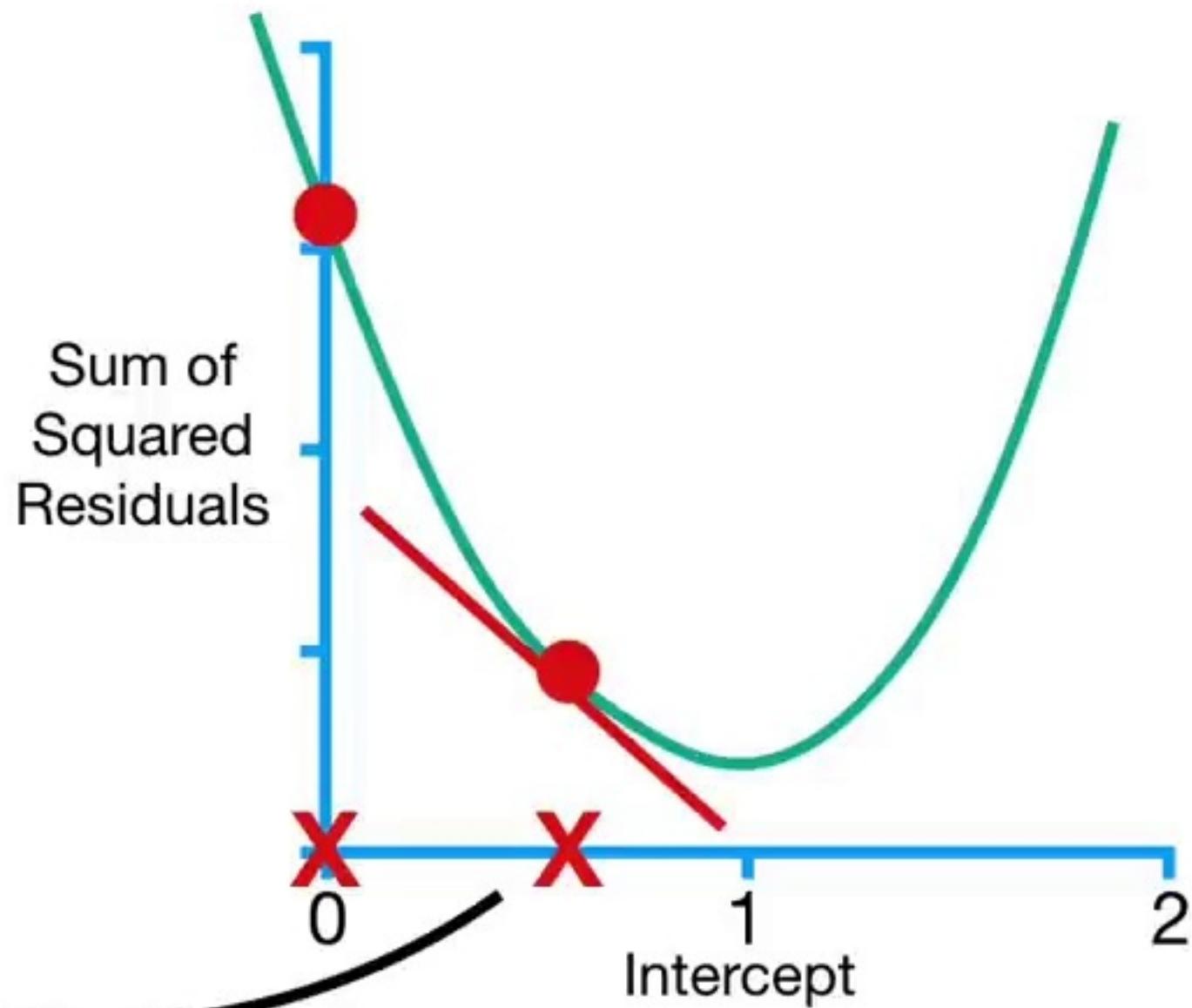


$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.57 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.57 + 0.64 \times 2.9))$$
$$= -2.3$$

$$\text{Step Size} = -2.3 \times 0.1 = -0.23$$

$$\text{New Intercept} = \text{Old Intercept} - \text{Step Size}$$

...and the **New Intercept**...

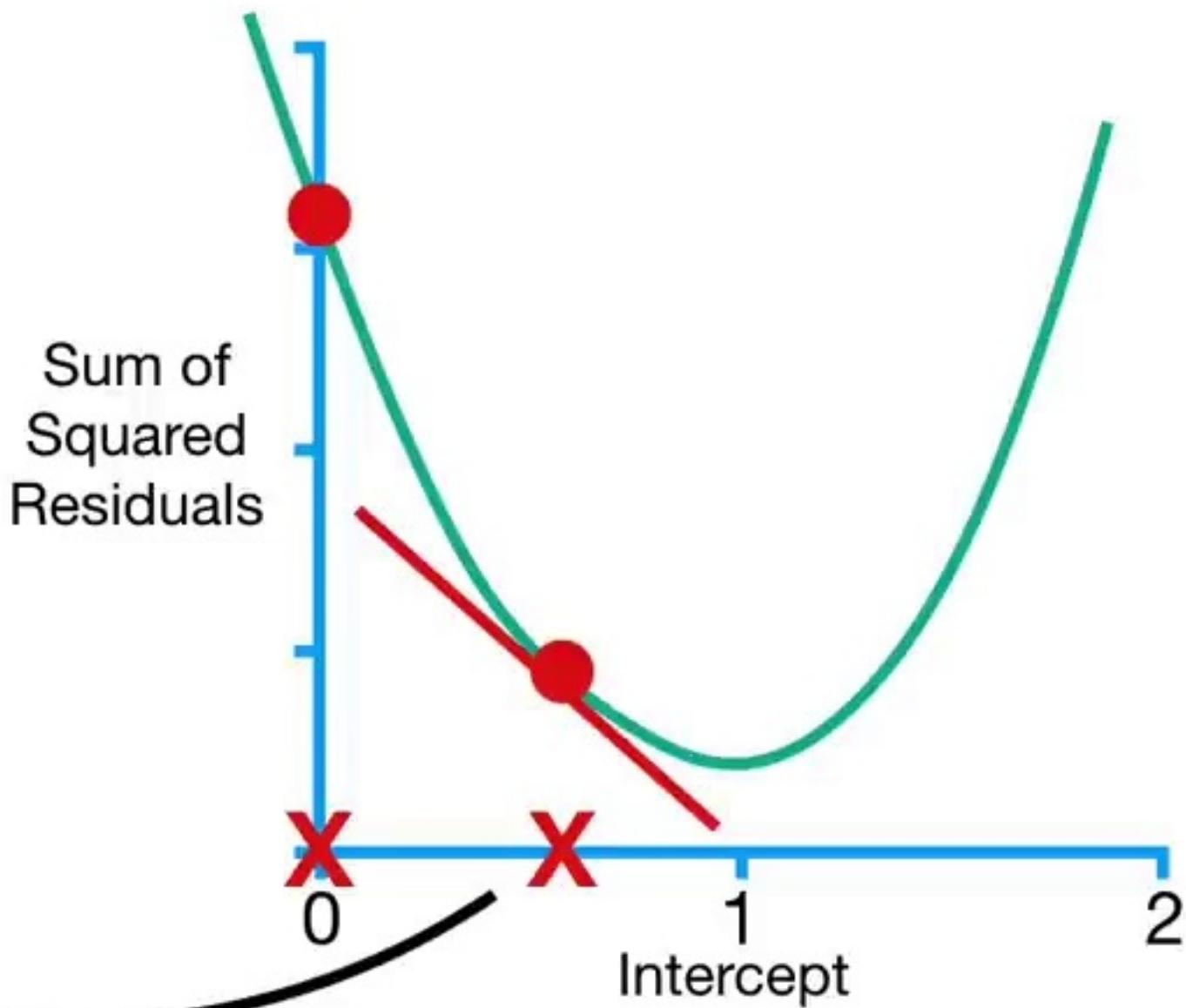


$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.57 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.57 + 0.64 \times 2.9))$$
$$= -2.3$$

$$\text{Step Size} = -2.3 \times 0.1 = -0.23$$

$$\text{New Intercept} = 0.57 - \text{Step Size}$$

...and the **New Intercept**...



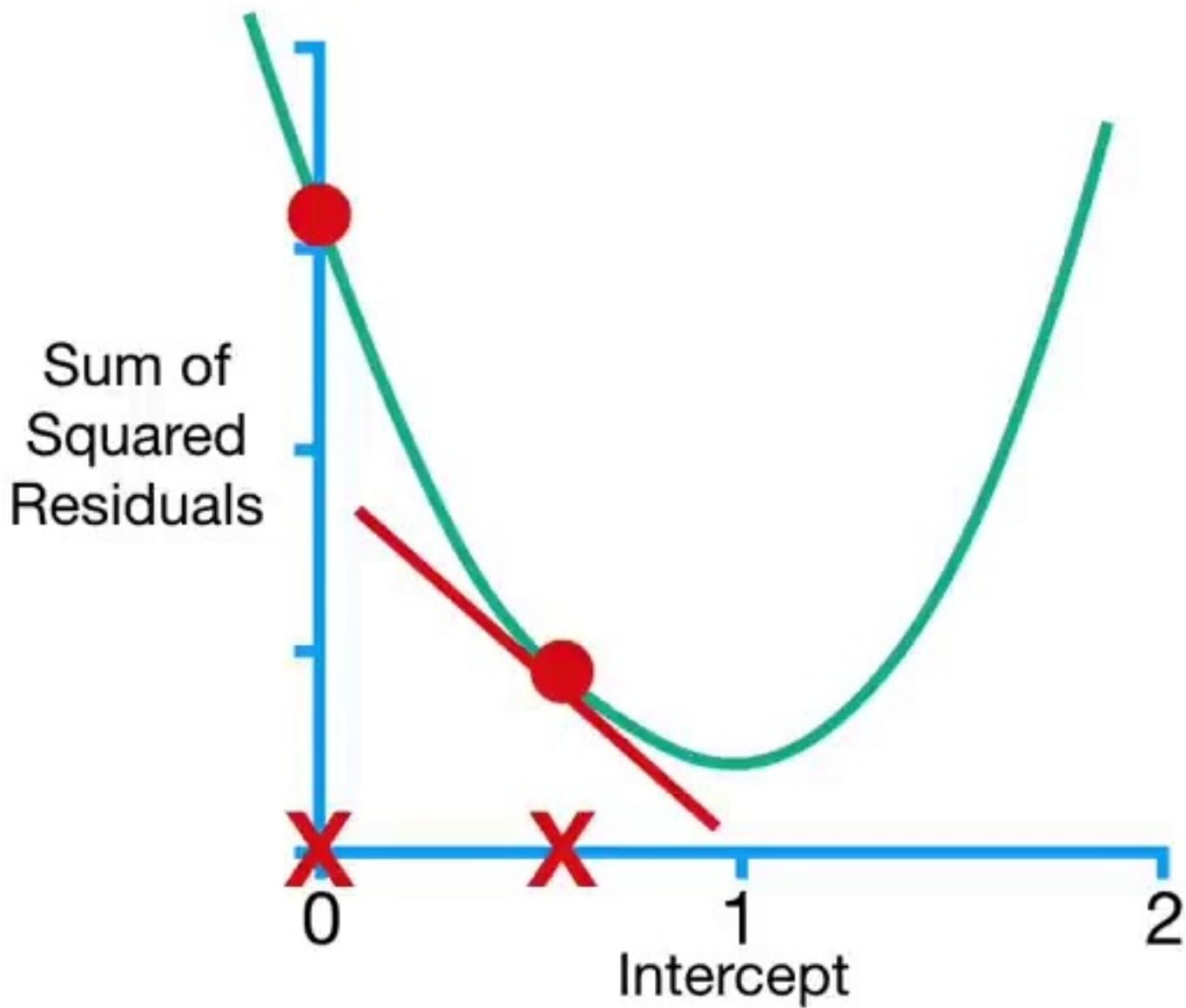
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.57 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.57 + 0.64 \times 2.9))$$
$$= -2.3$$

$$\text{Step Size} = -2.3 \times 0.1 = \boxed{-0.23}$$



$$\text{New Intercept} = 0.57 - \text{Step Size}$$

...and the **New Intercept**...



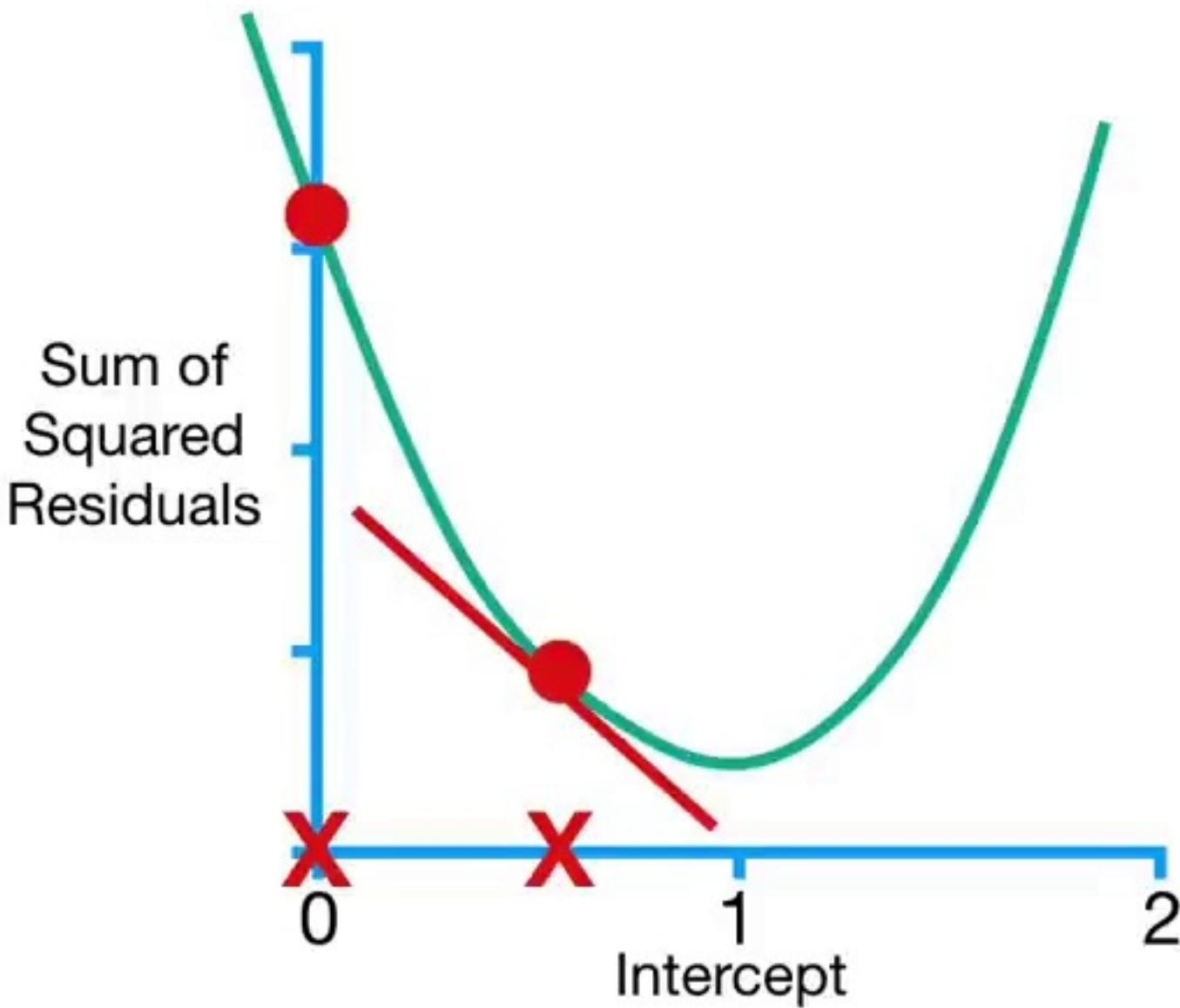
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.57 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.57 + 0.64 \times 2.9))$$
$$= -2.3$$

Step Size = $-2.3 \times 0.1 = \boxed{-0.23}$



New Intercept = $0.57 - (-0.23)$

...and the **New Intercept**...

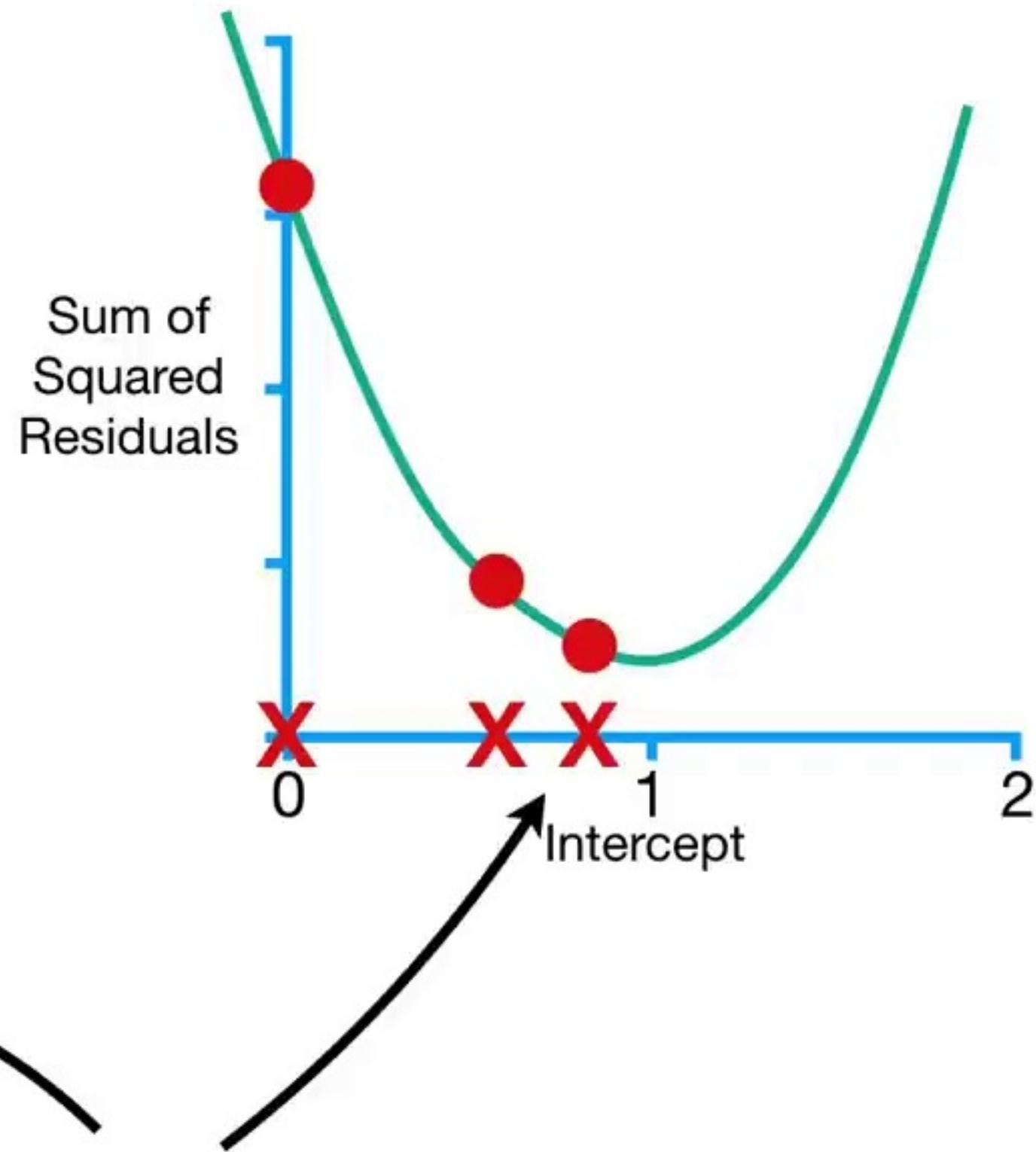


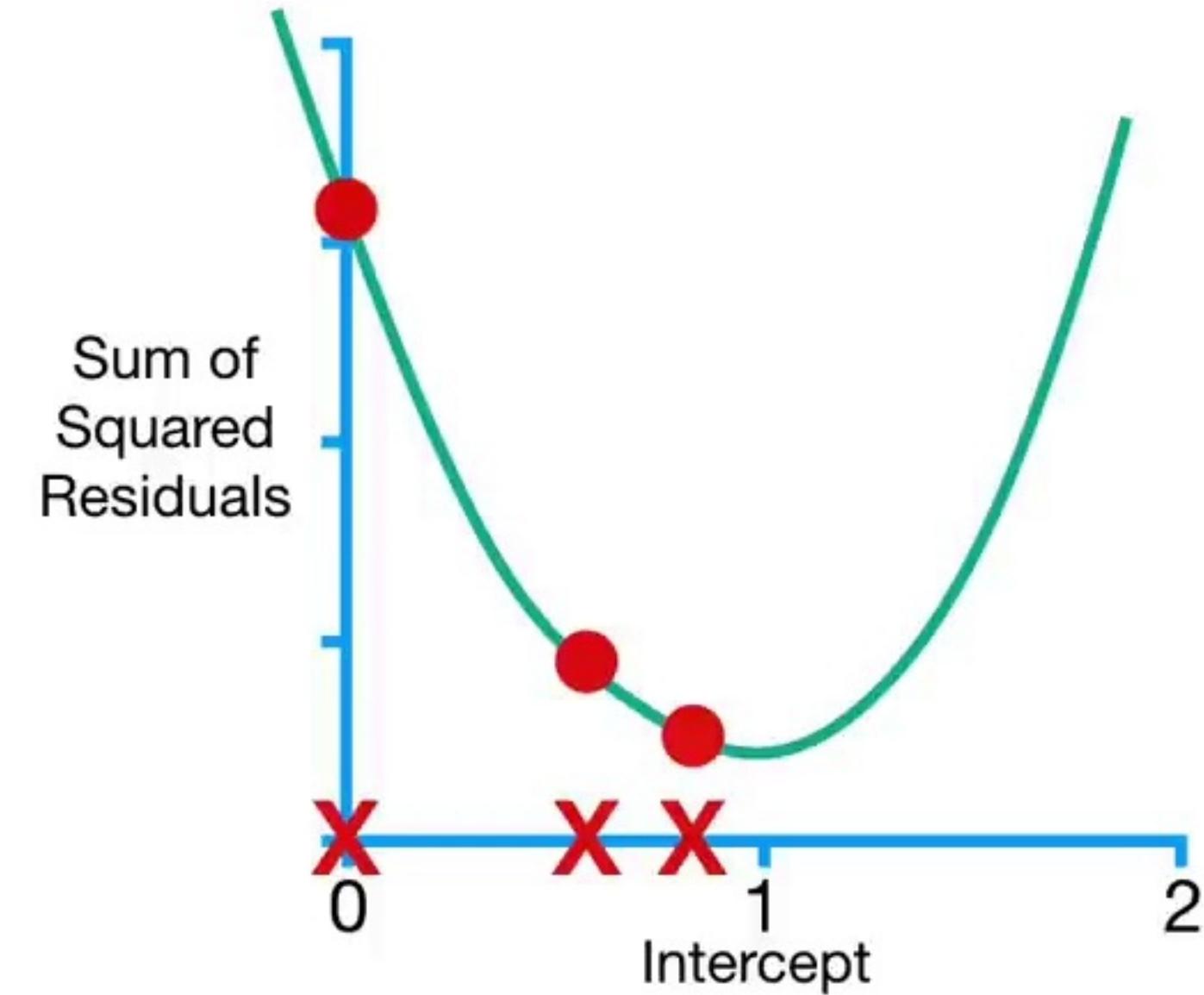
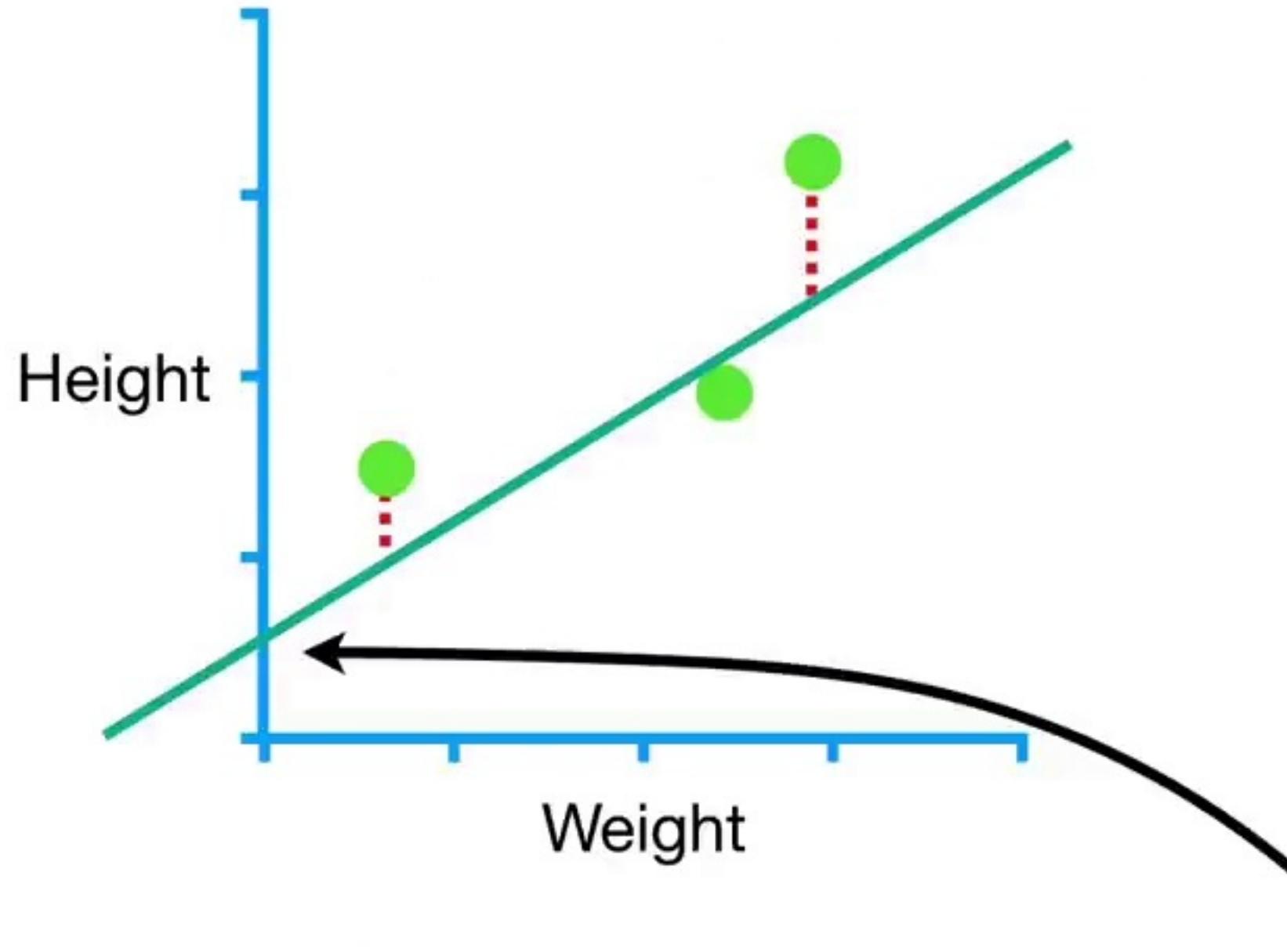
$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.57 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.57 + 0.64 \times 2.9))$$
$$= -2.3$$

$$\text{Step Size} = -2.3 \times 0.1 = -0.23$$

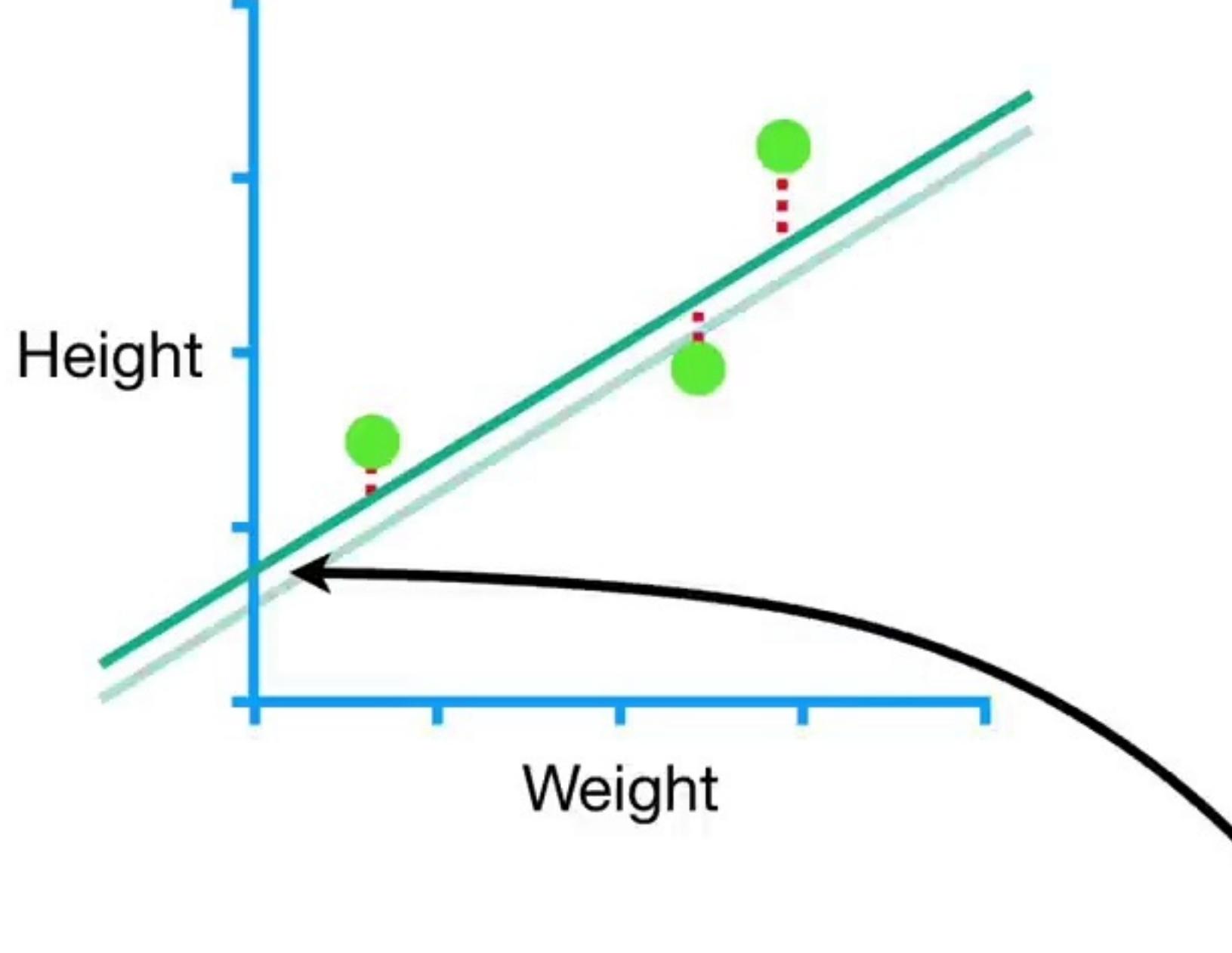
$$\text{New Intercept} = 0.57 - (-0.23) = \boxed{0.8}$$

...and the **New Intercept = 0.8**

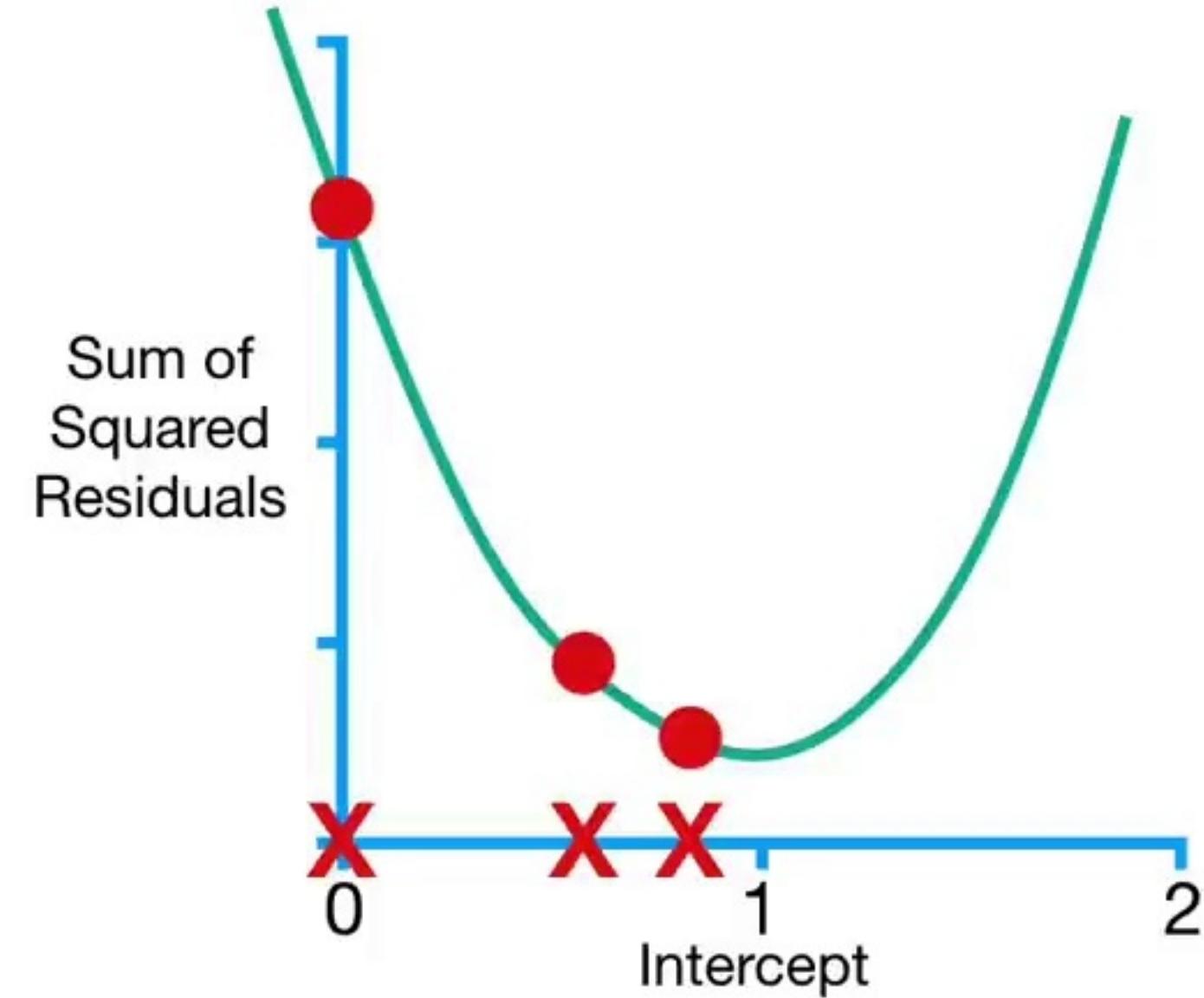


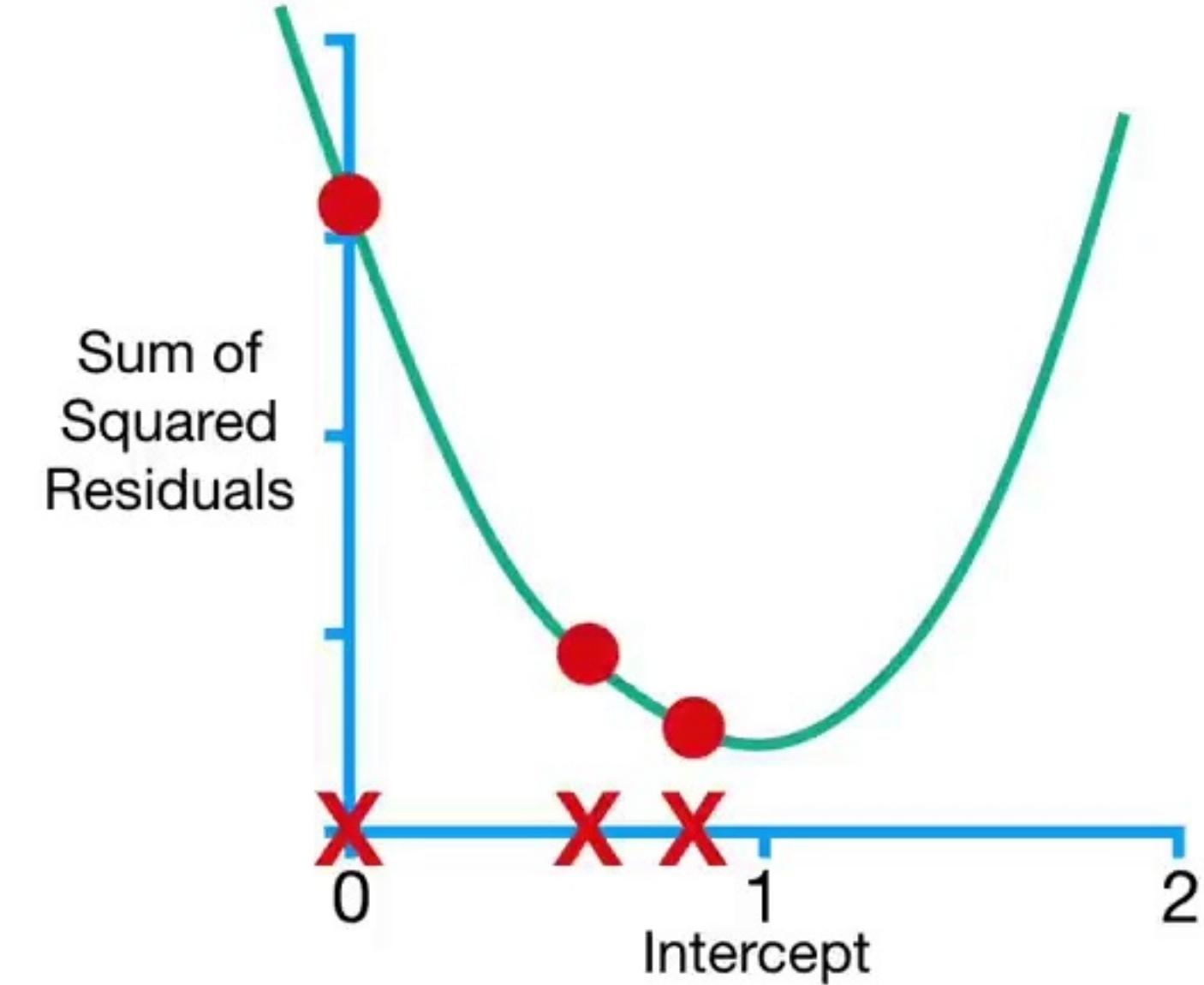
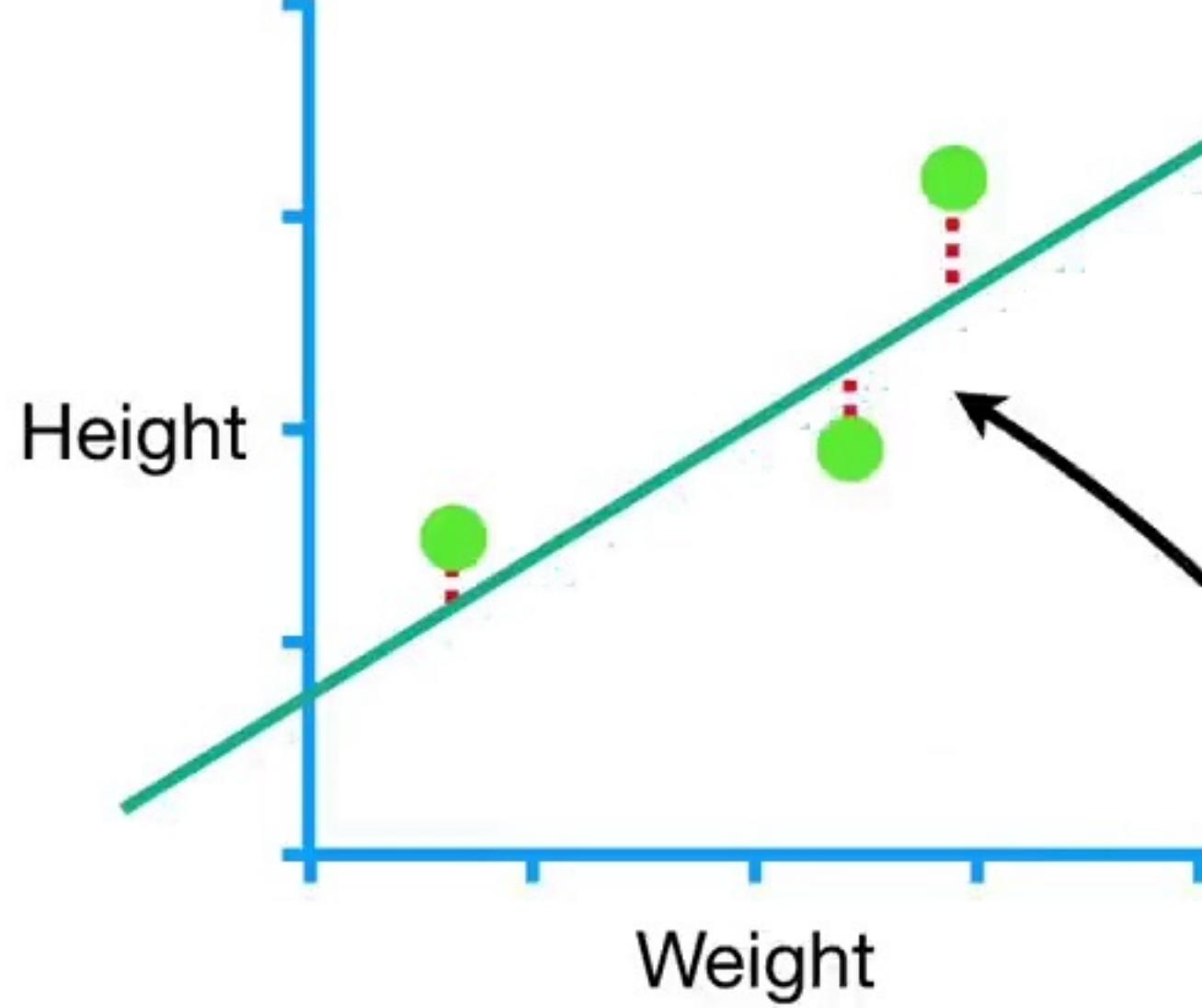


Now we can compare the
residuals when the
Intercept = 0.57...

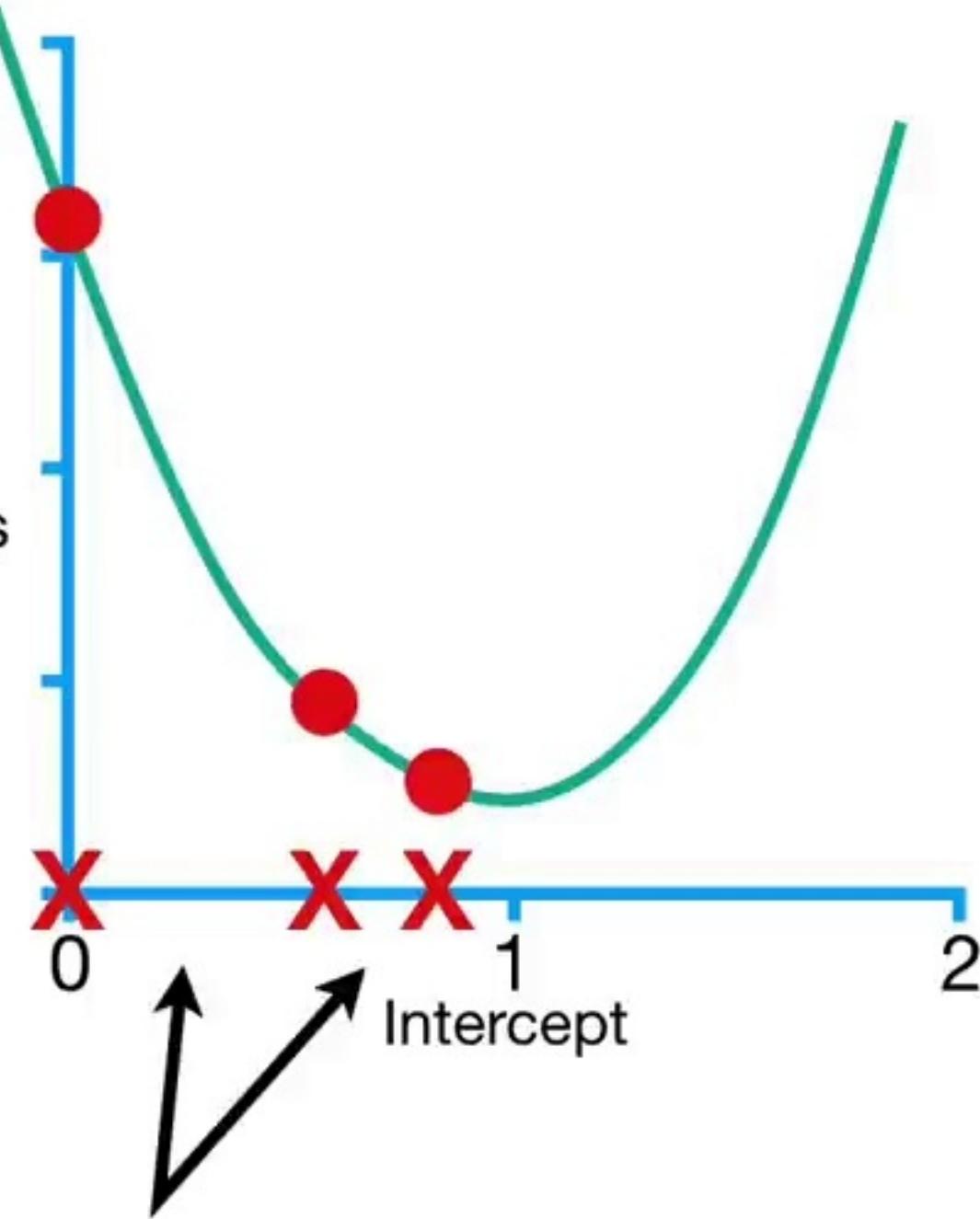


...to when the
Intercept = 0.8





Overall, the Sum of the
Squared Residuals is getting
smaller.



Notice that the first step was relatively large compared to the second step.

$$\frac{d}{d \text{ intercept}}$$

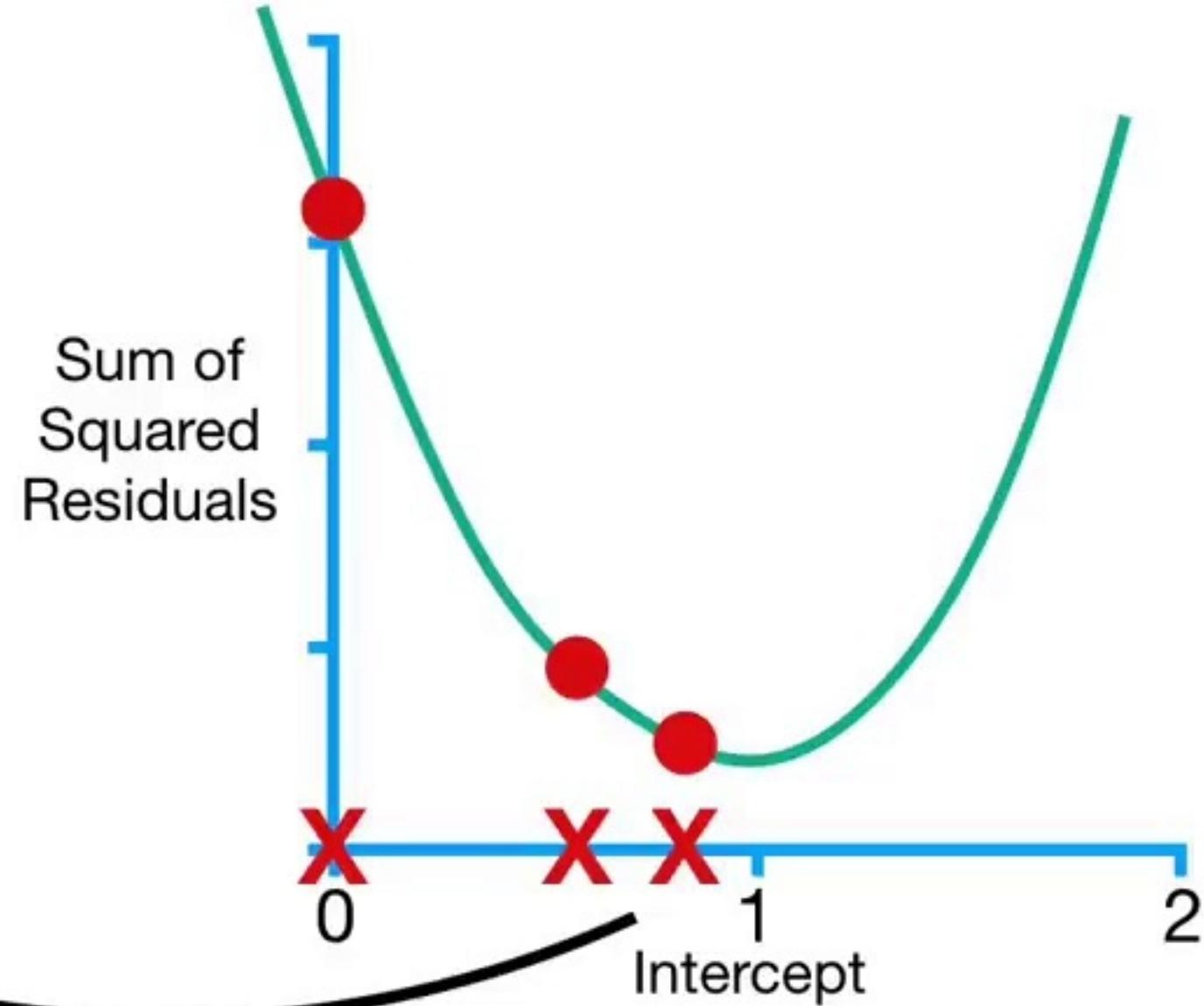
Sum of squared residuals =

$$-2(1.4 - (\text{intercept} + 0.64 \times 0.5))$$

$$+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3))$$

$$+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9))$$

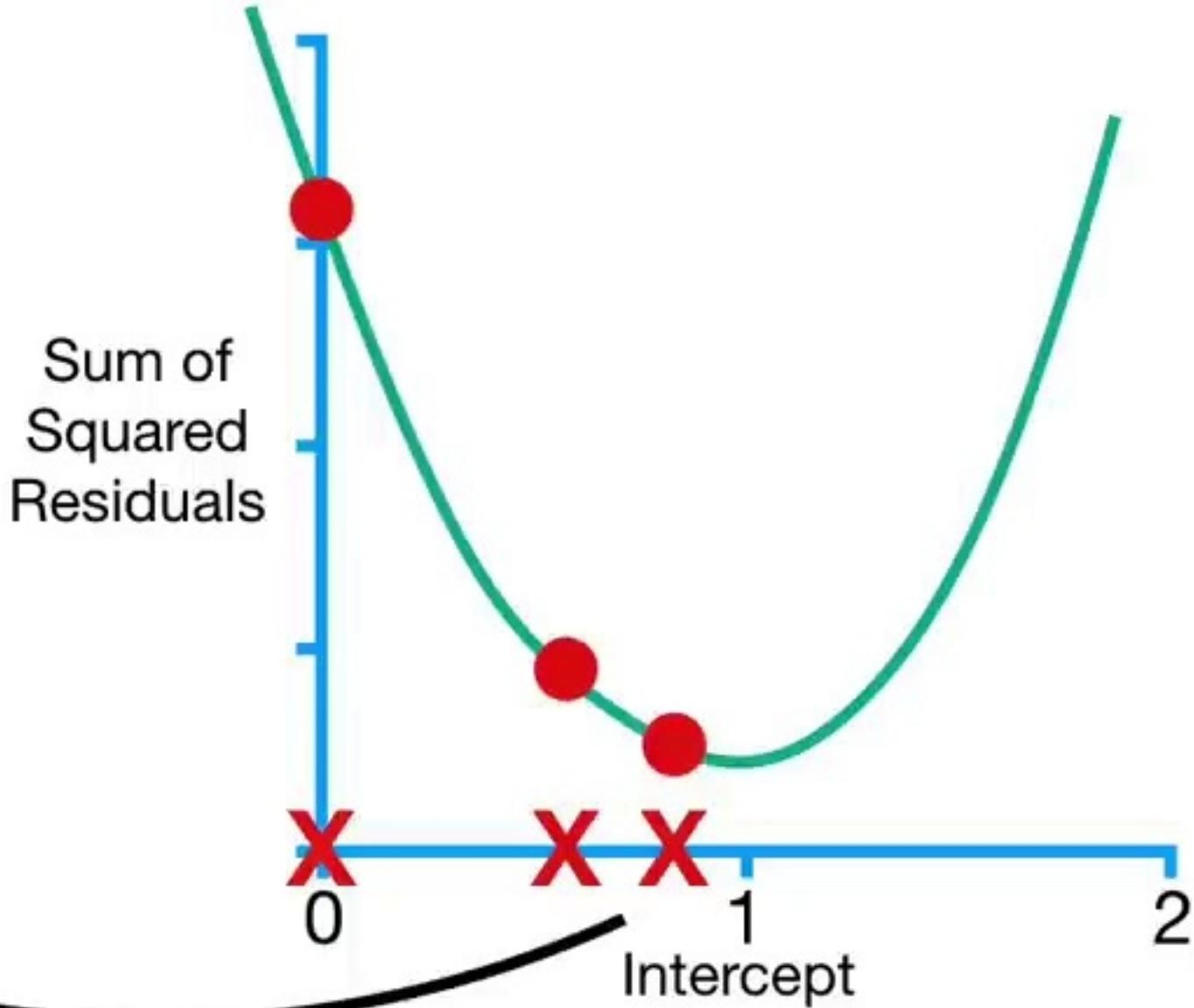
Now let's calculate the derivative at the
New Intercept (0.8)...



$$\frac{d}{d \text{ intercept}}$$

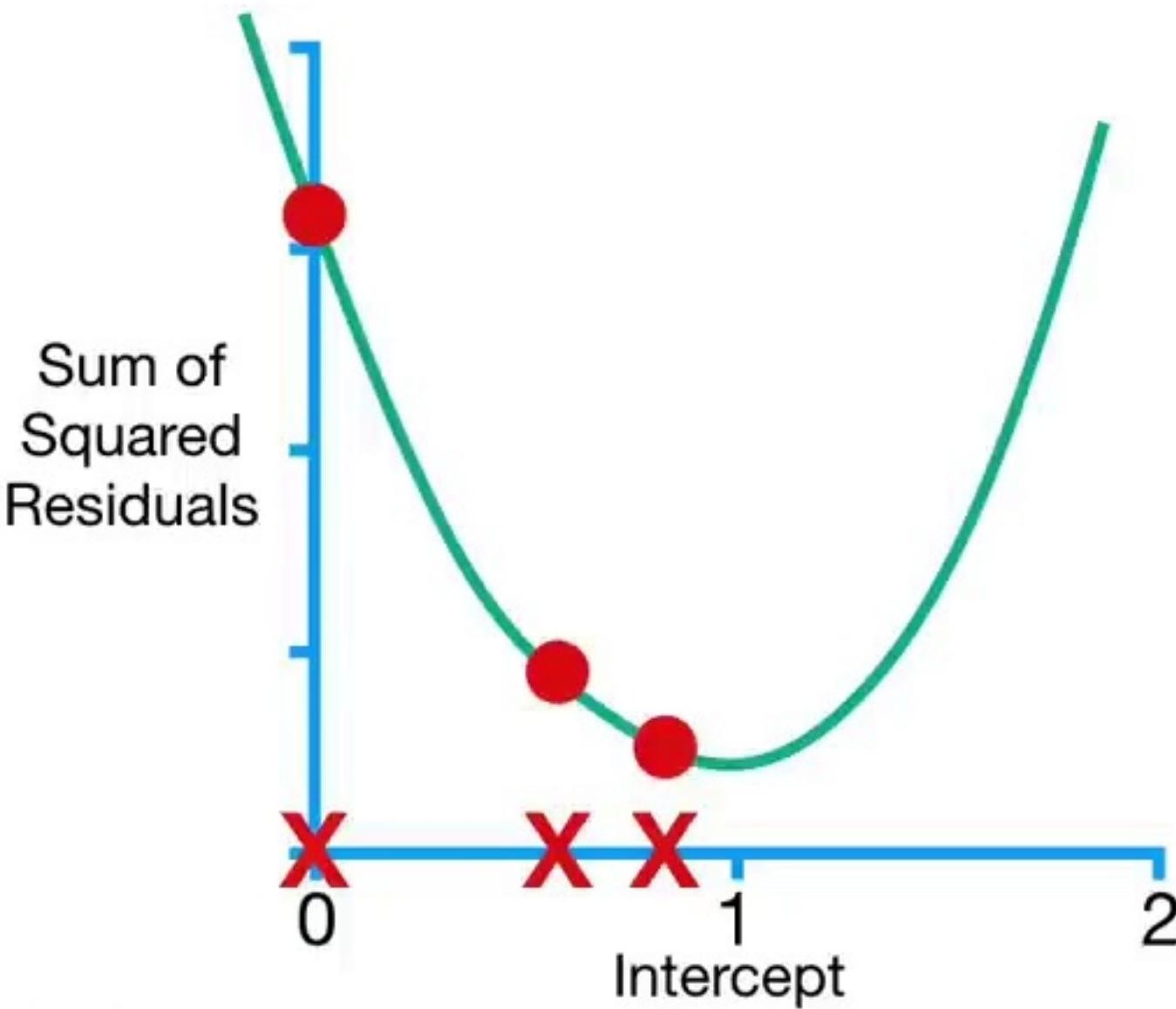
Sum of squared residuals =
 $-2(1.4 - (0.8 + 0.64 \times 0.5))$
 $+ -2(1.9 - (0.8 + 0.64 \times 2.3))$
 $+ -2(3.2 - (0.8 + 0.64 \times 2.9))$

Now let's calculate the derivative at the
New Intercept (0.8)...



$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.8 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.8 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.8 + 0.64 \times 2.9))$$
$$= \boxed{-0.9}$$

...and we get **-0.9**.

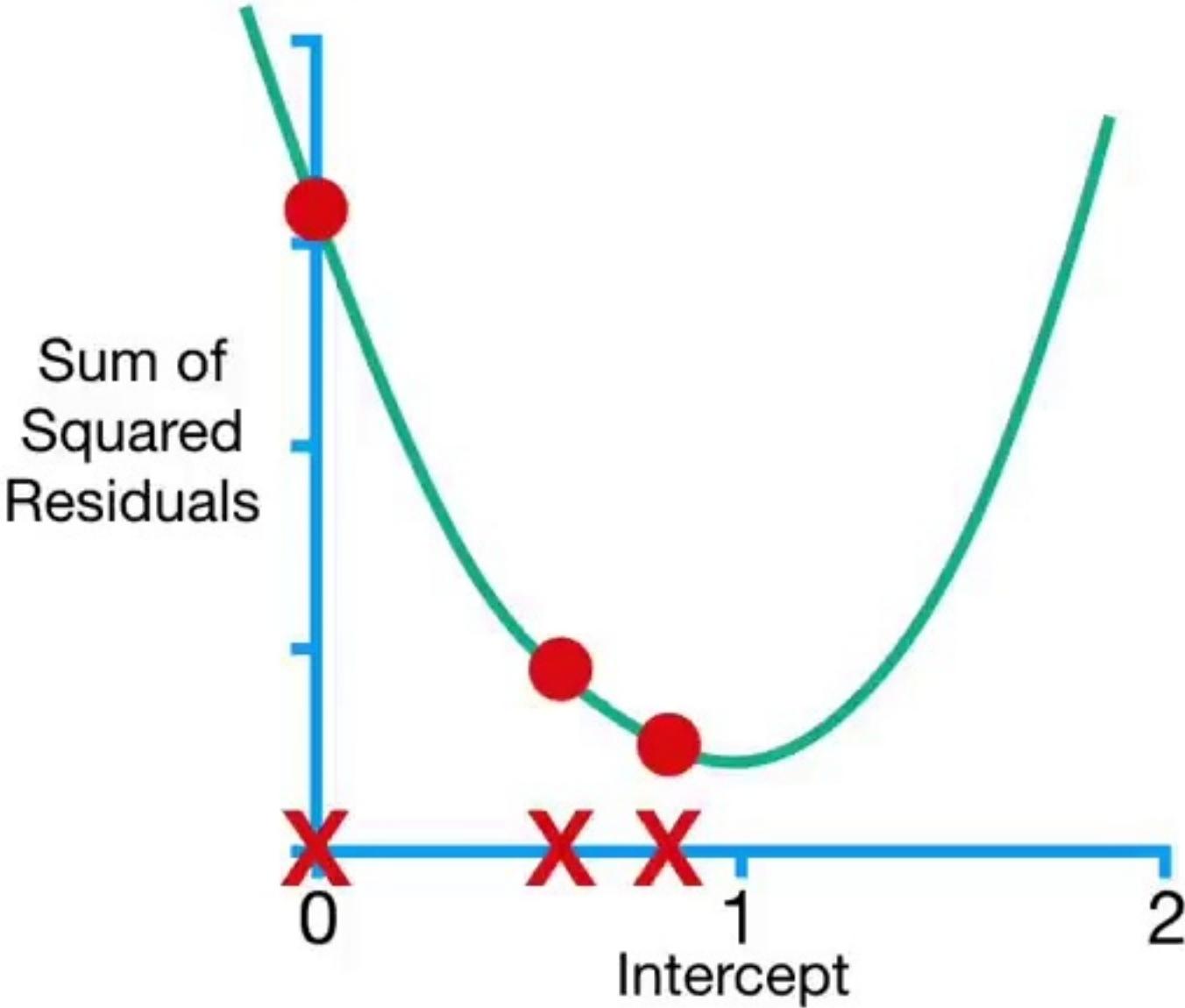


$$\frac{d}{d \text{ intercept}}$$

Sum of squared residuals =
-2(1.4 - (0.8 + 0.64 × 0.5))
+ -2(1.9 - (0.8 + 0.64 × 2.3))
+ -2(3.2 - (0.8 + 0.64 × 2.9))
= -0.9

Step Size = Slope × Learning Rate

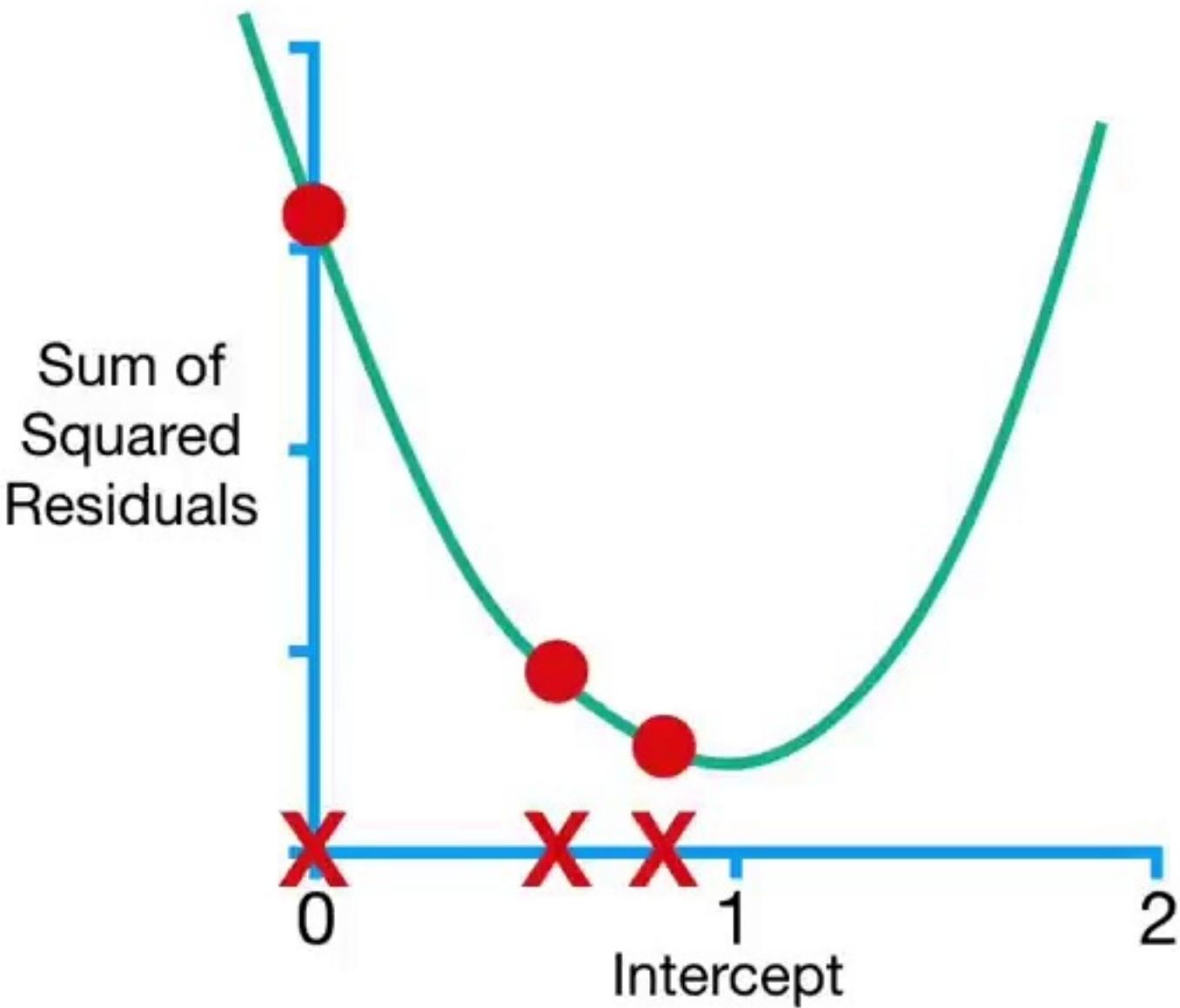
The **Step Size**...



$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.8 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.8 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.8 + 0.64 \times 2.9))$$
$$= -0.9$$

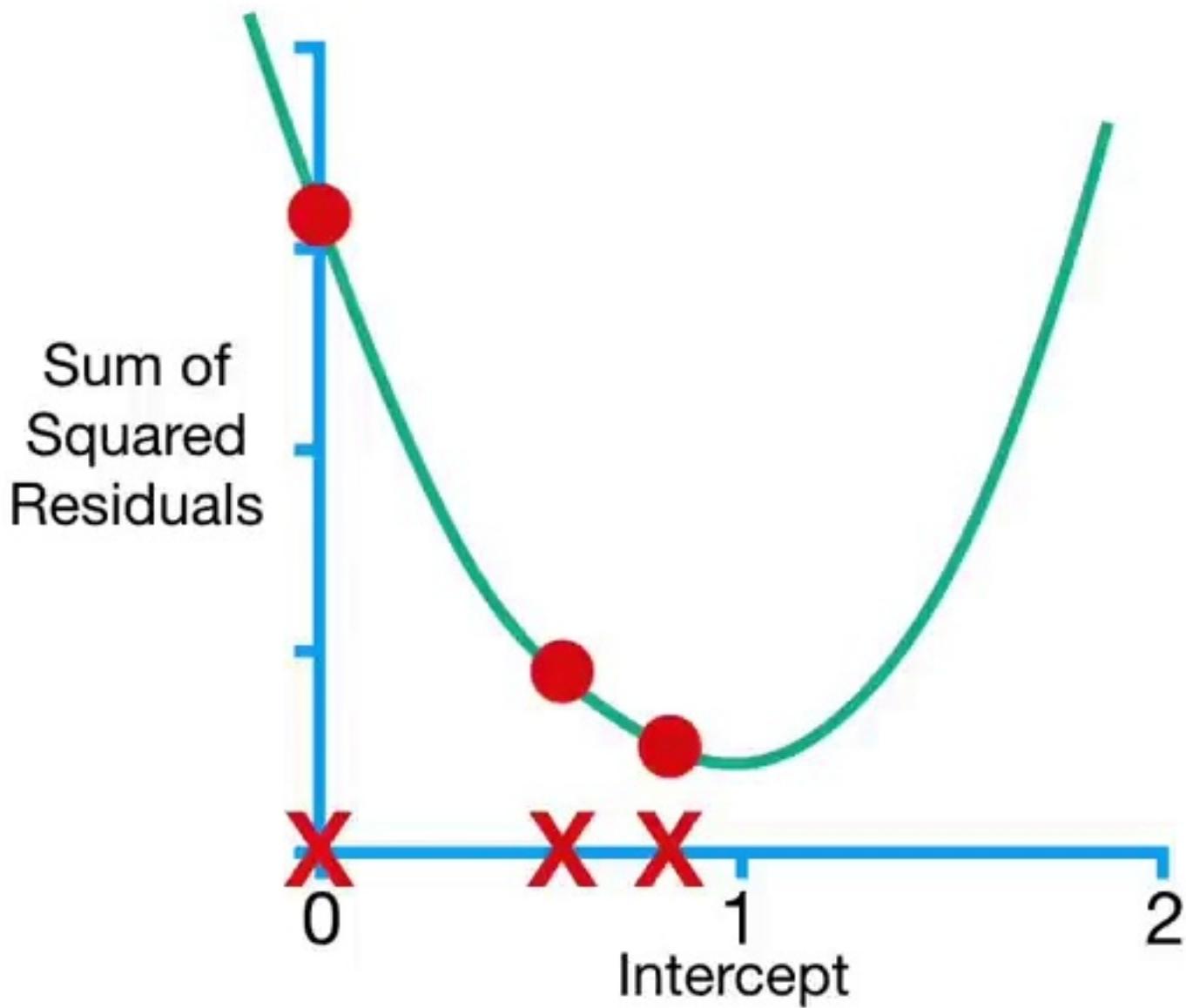
Step Size = $-0.9 \times 0.1 = -0.09$

The **Step Size** = **-0.09...**



$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} =$$
$$-2(1.4 - (0.8 + 0.64 \times 0.5))$$
$$+ -2(1.9 - (0.8 + 0.64 \times 2.3))$$
$$+ -2(3.2 - (0.8 + 0.64 \times 2.9))$$
$$= -0.9$$

$$\text{Step Size} = -0.9 \times 0.1 = -0.09$$



New Intercept = Old Intercept - Step Size

...and the **New Intercept**...

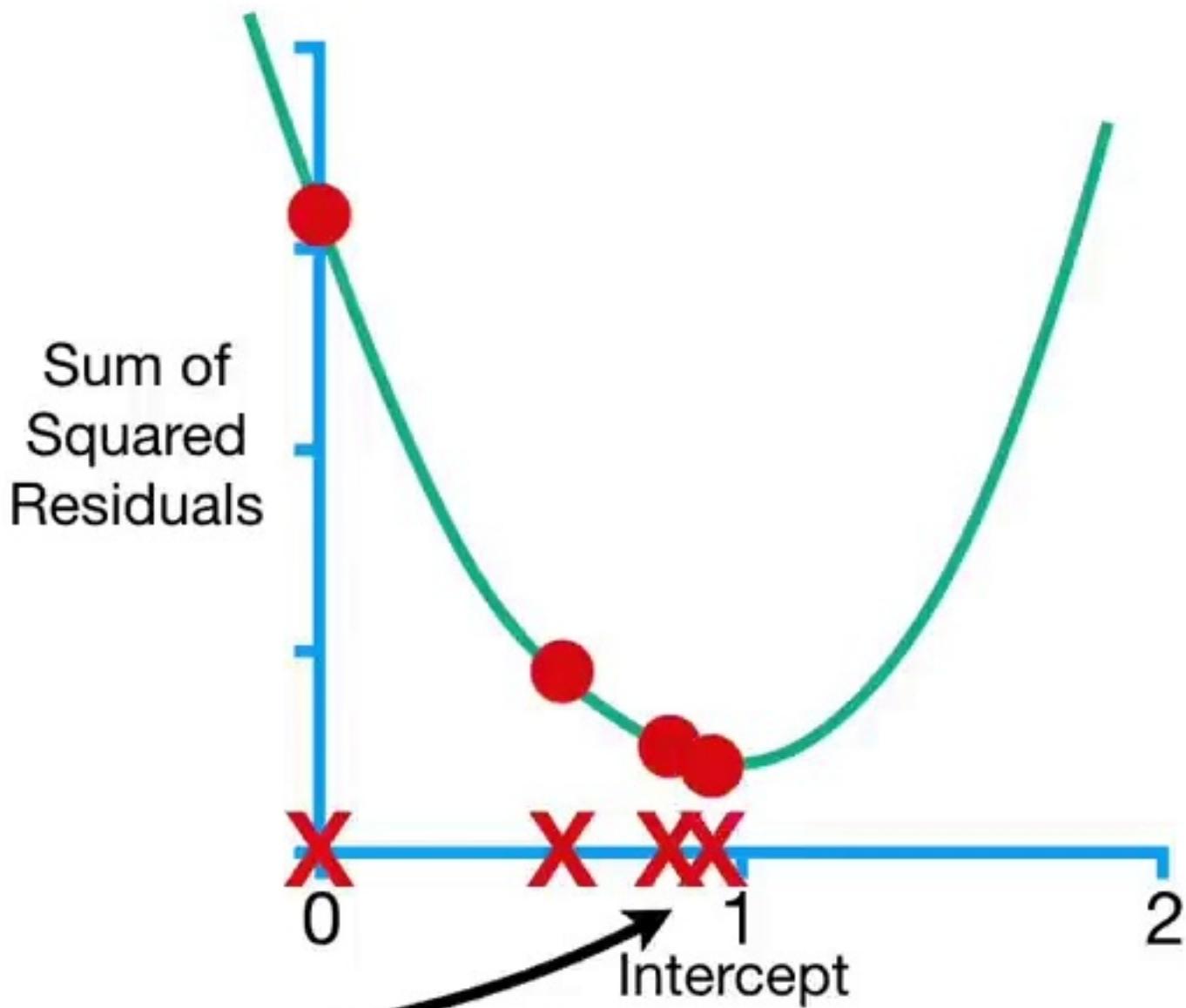
$$\frac{d}{d \text{ intercept}}$$

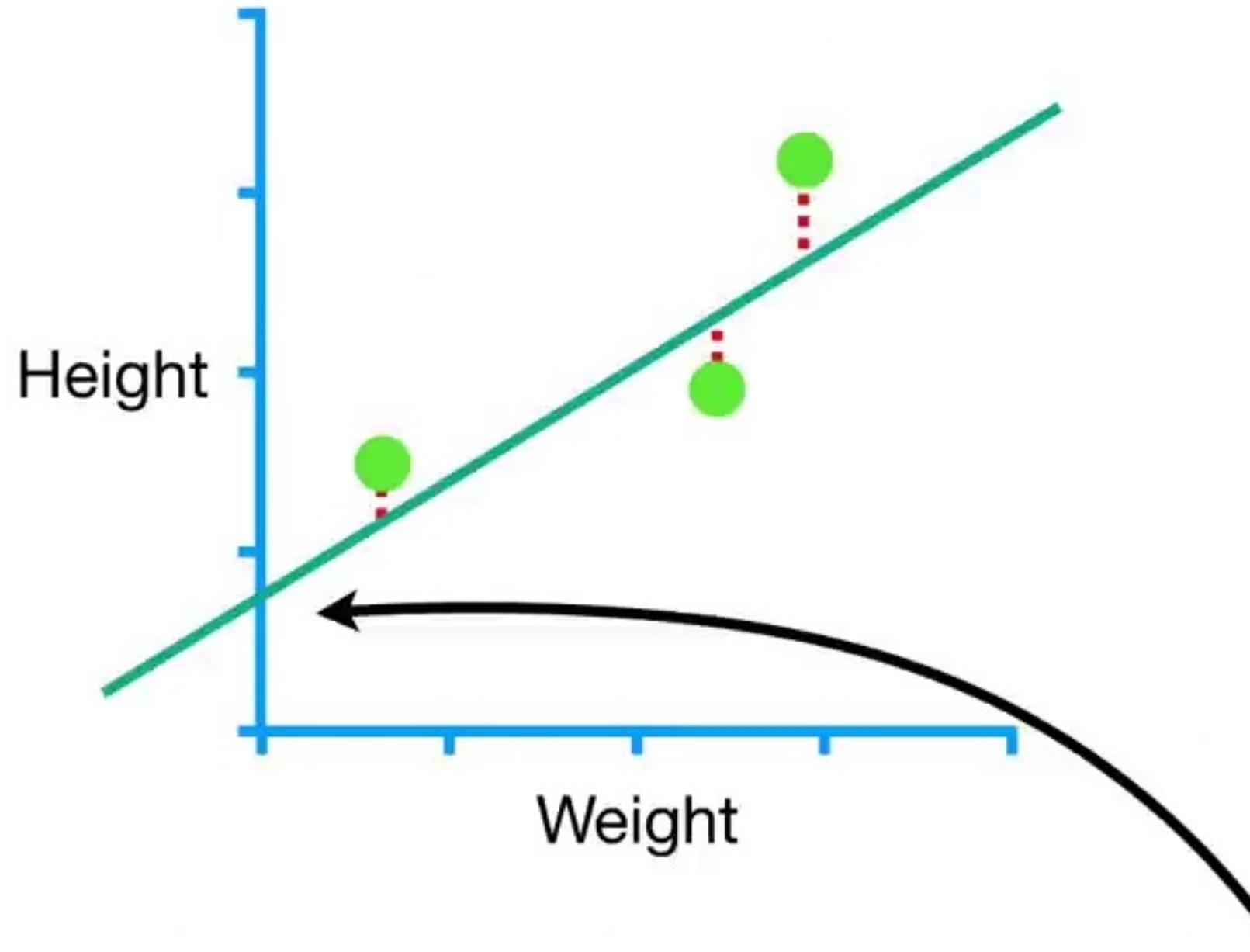
Sum of squared residuals =
 $-2(1.4 - (0.8 + 0.64 \times 0.5))$
 $+ -2(1.9 - (0.8 + 0.64 \times 2.3))$
 $+ -2(3.2 - (0.8 + 0.64 \times 2.9))$
 $= -0.9$

Step Size = $-0.9 \times 0.1 = -0.09$

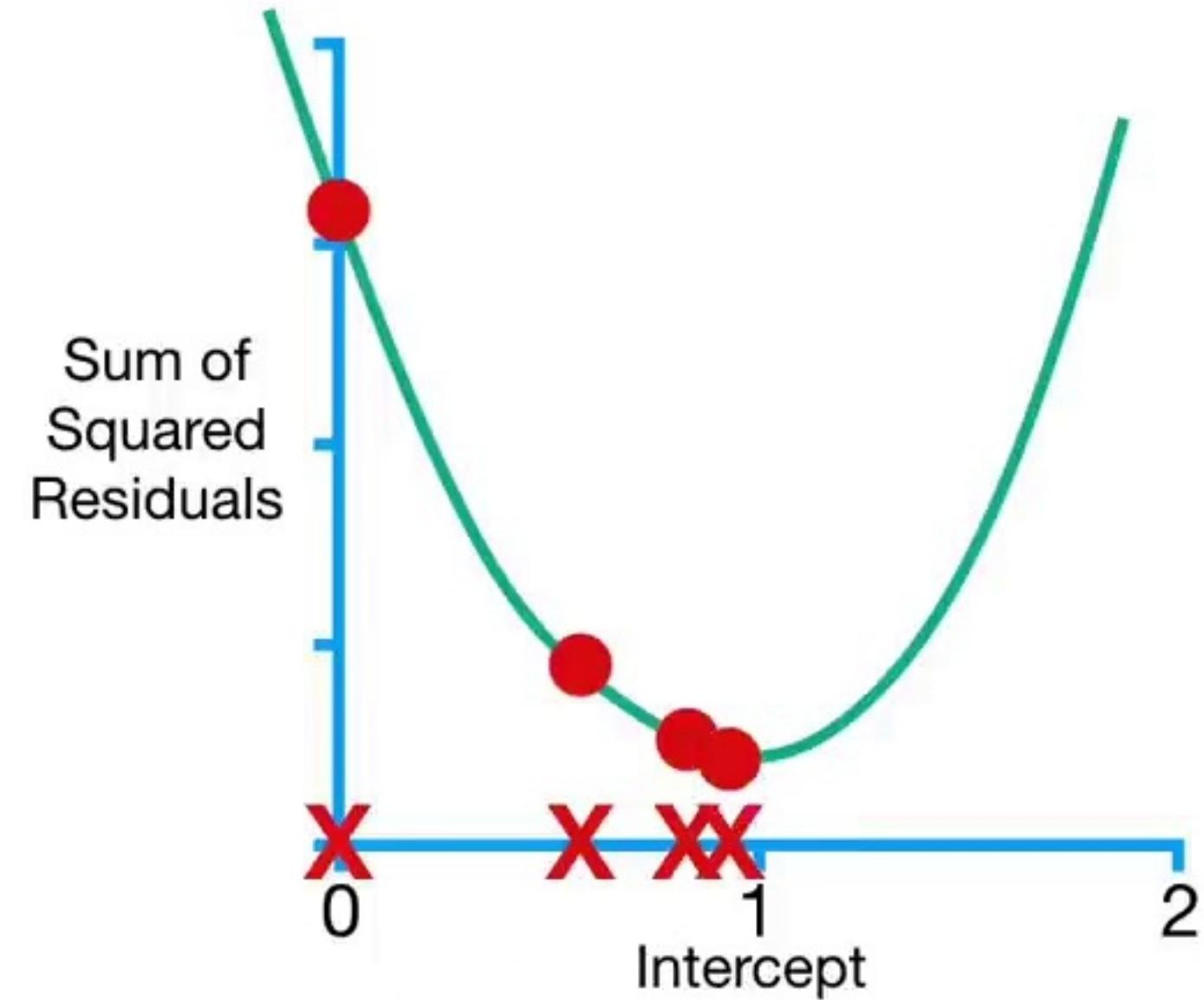
New Intercept = $0.8 - (-0.09) = 0.89$

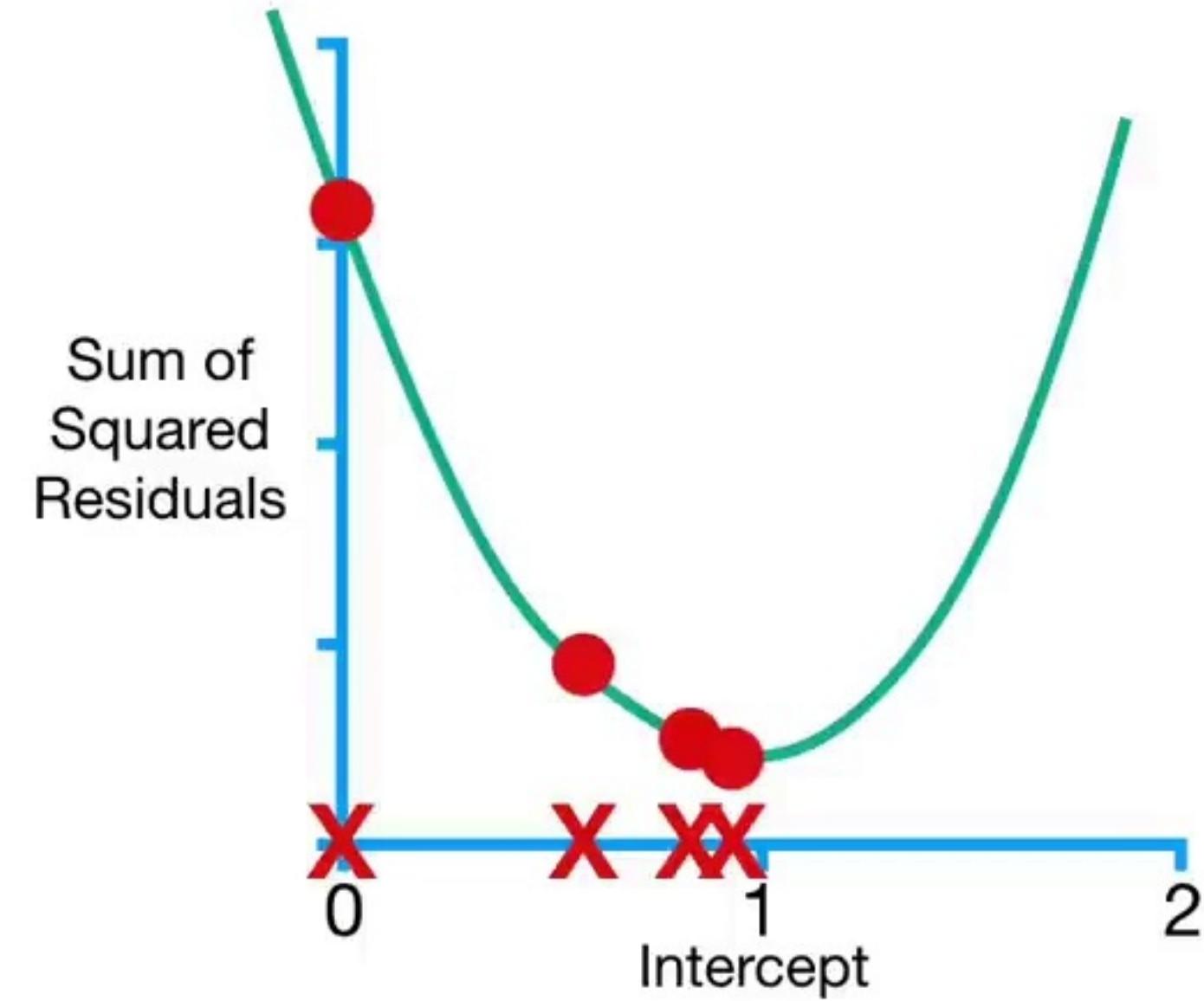
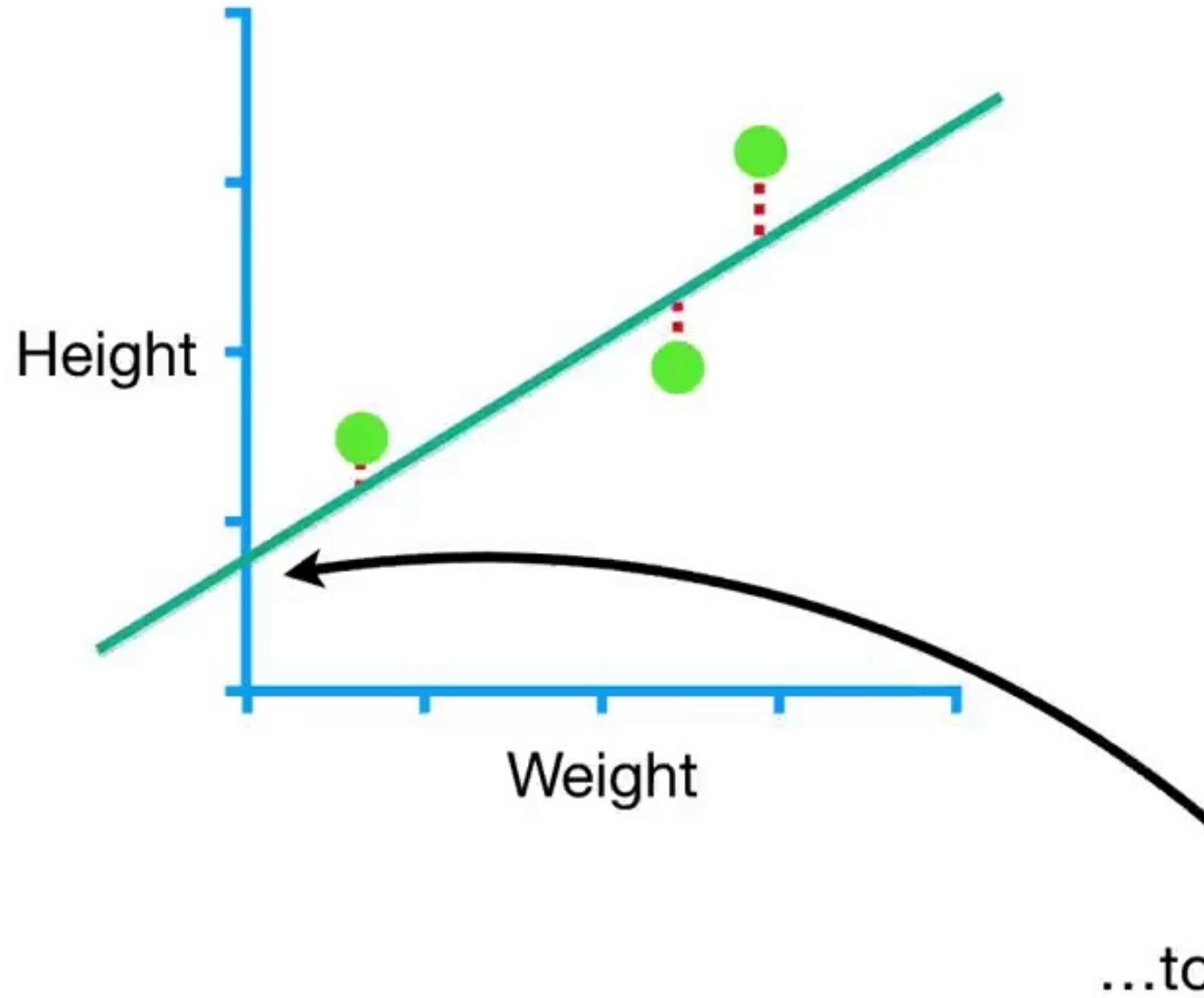
...and the New Intercept = 0.89

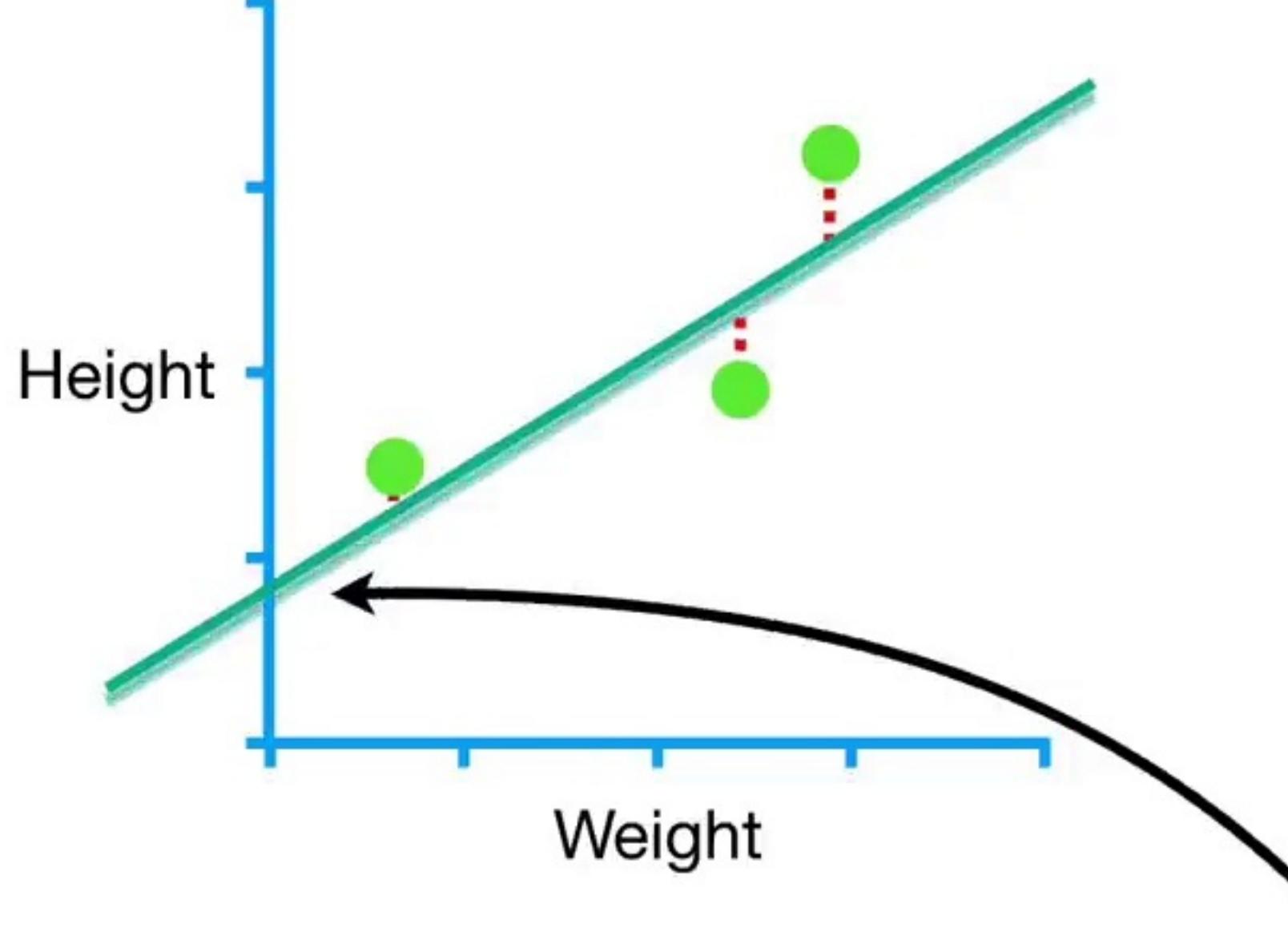




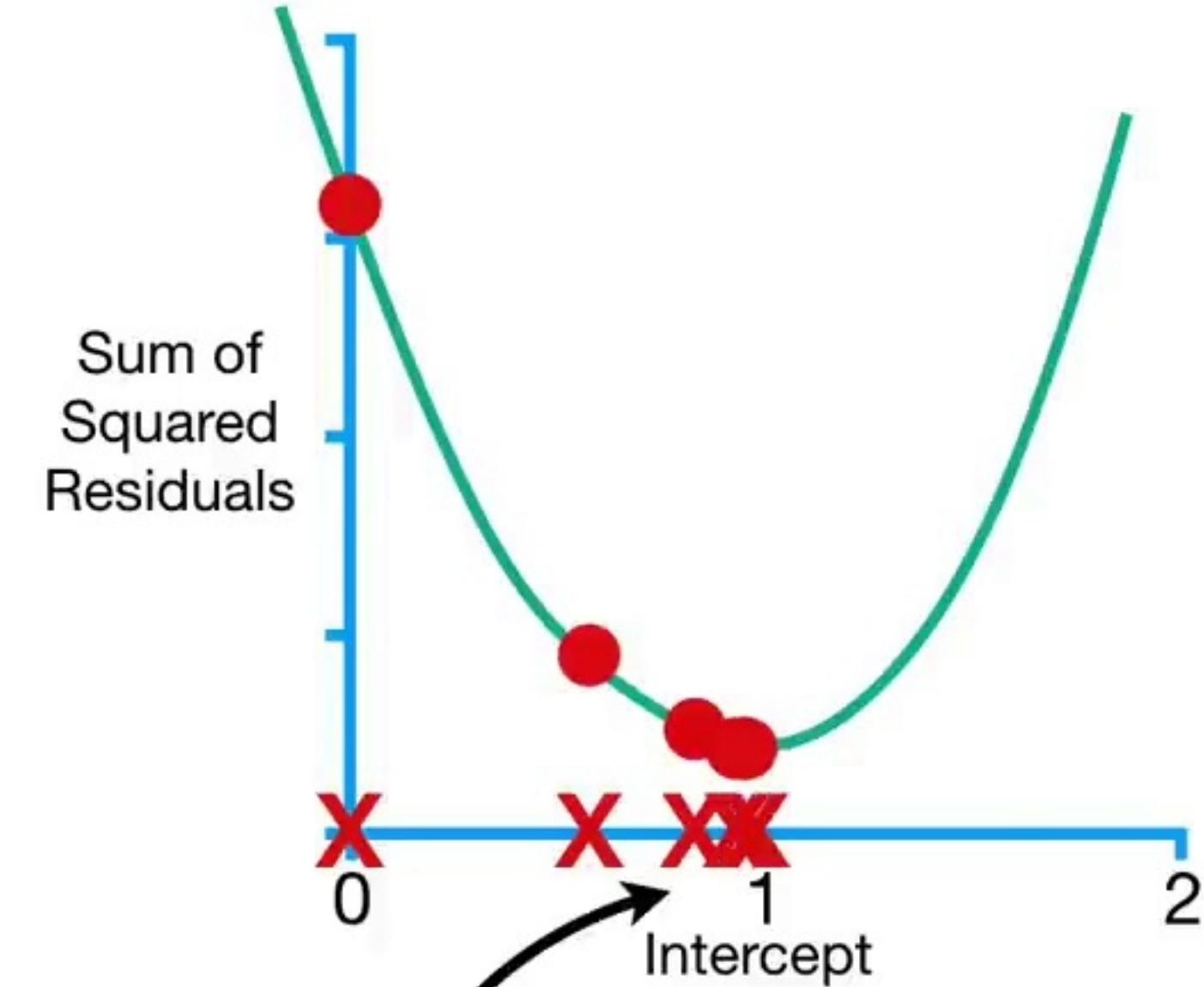
Now we increase the
Intercept from 0.8...

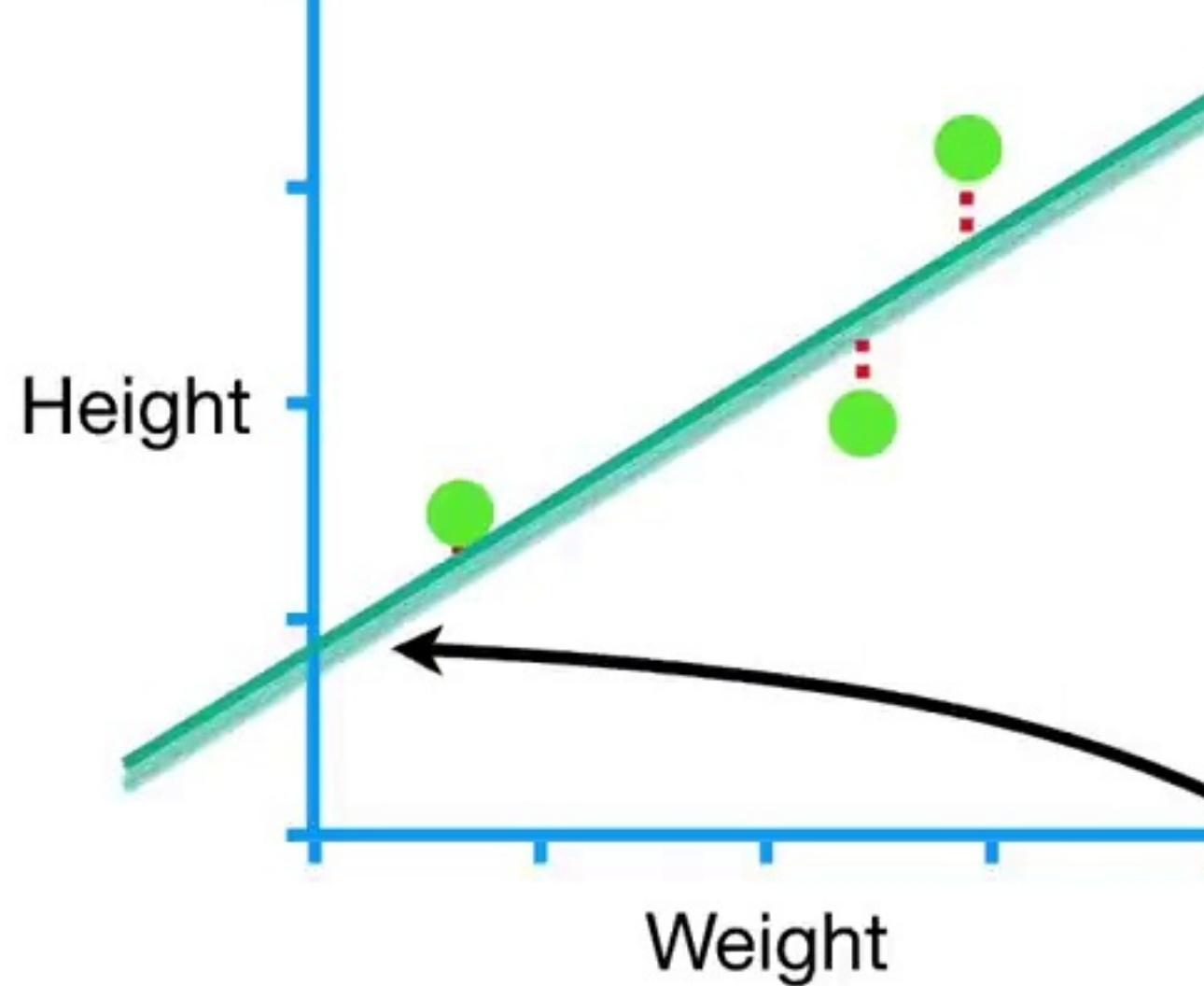




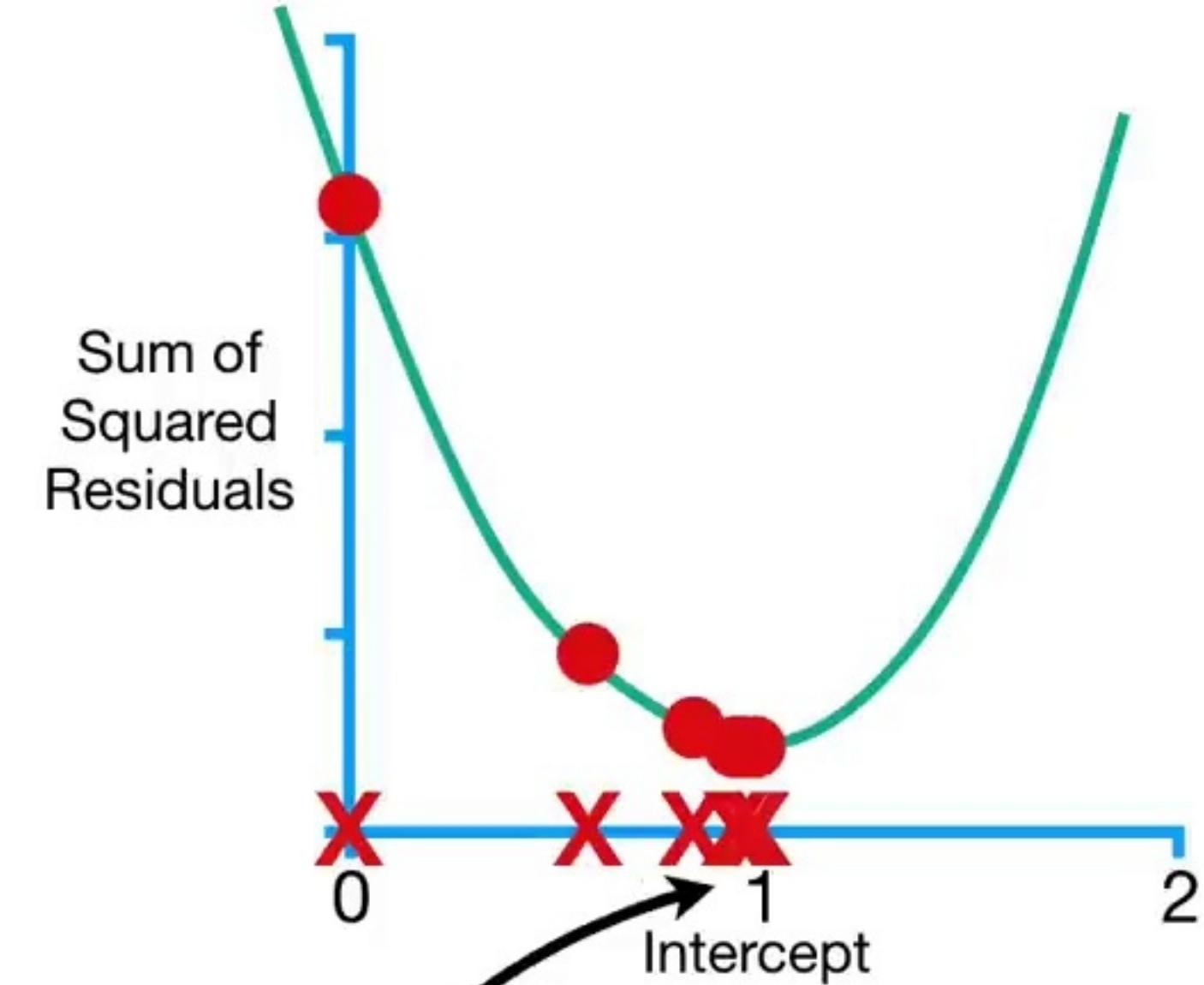


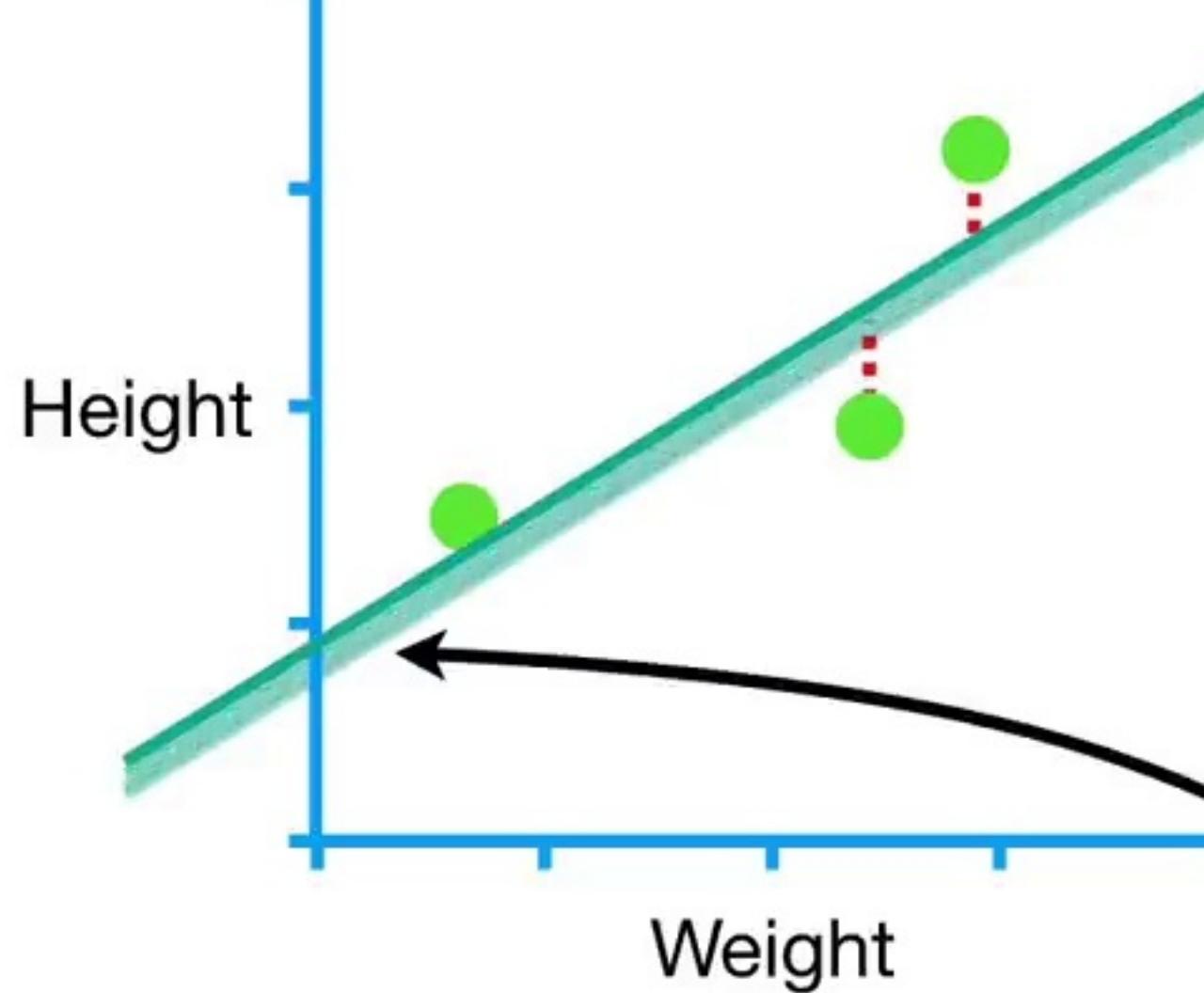
Then we take another step and
the **New Intercept** = 0.92...



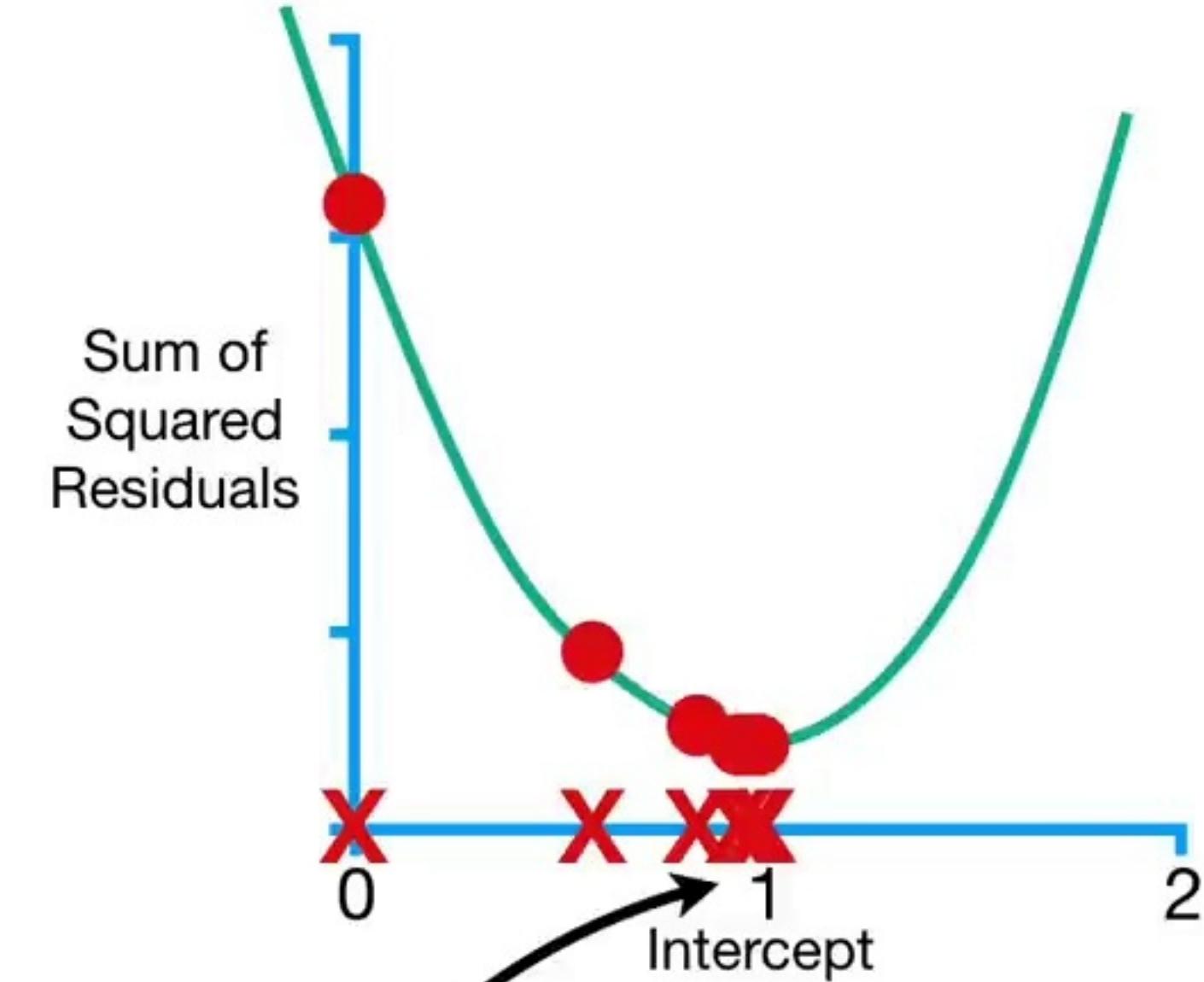


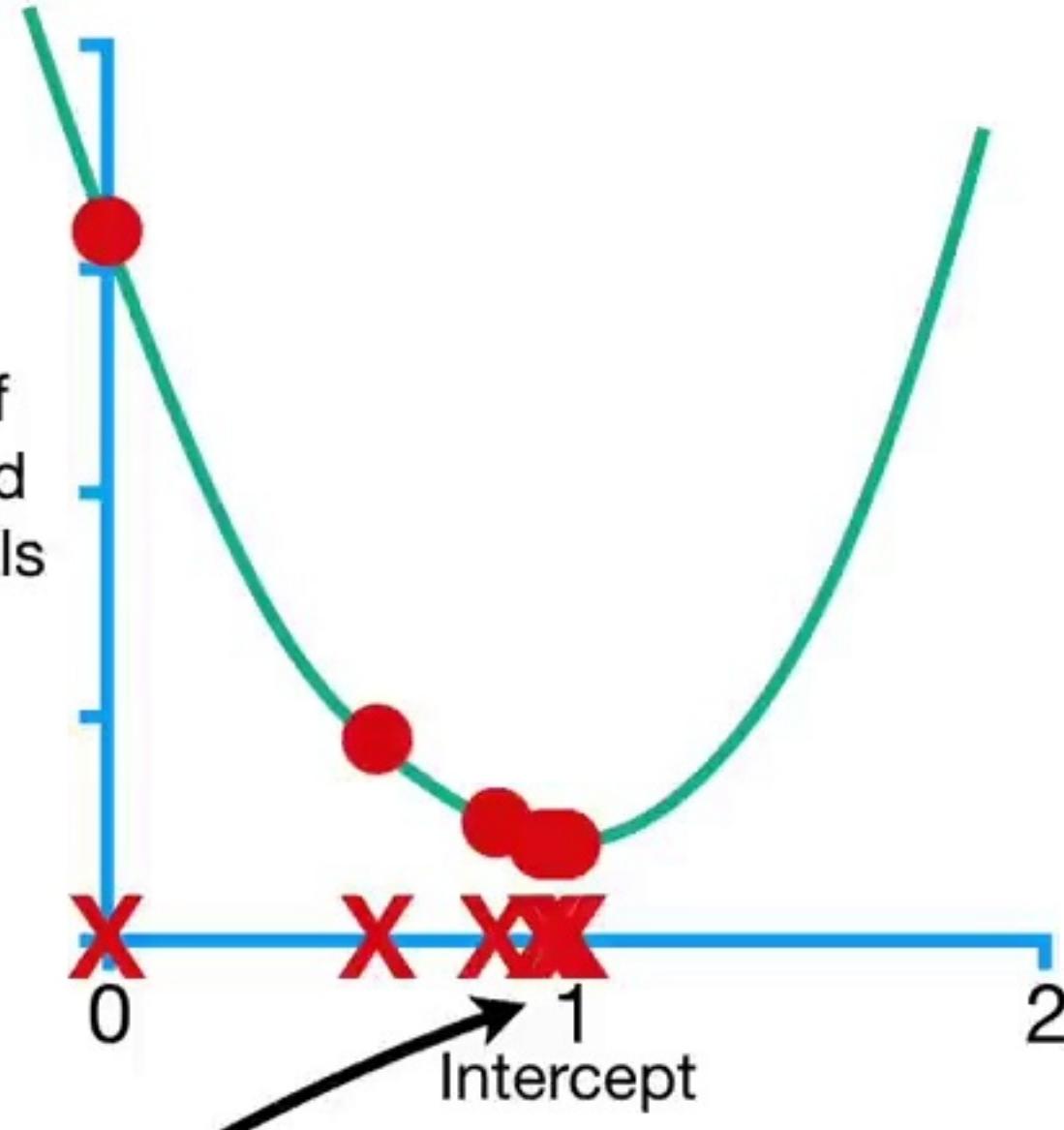
...and then we take another
step and the
New Intercept = 0.94...





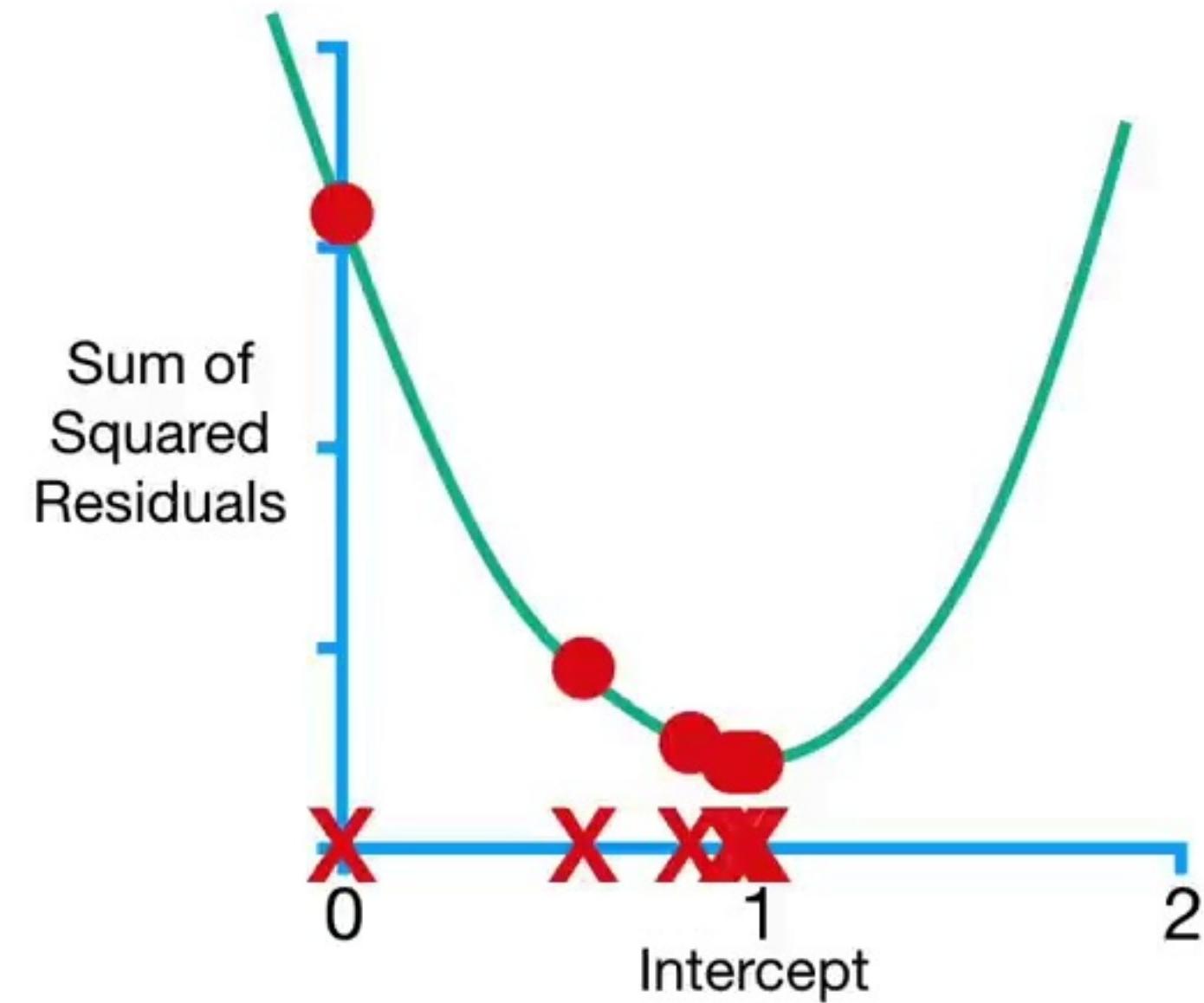
...and then we take another
step and the
New Intercept = 0.95.





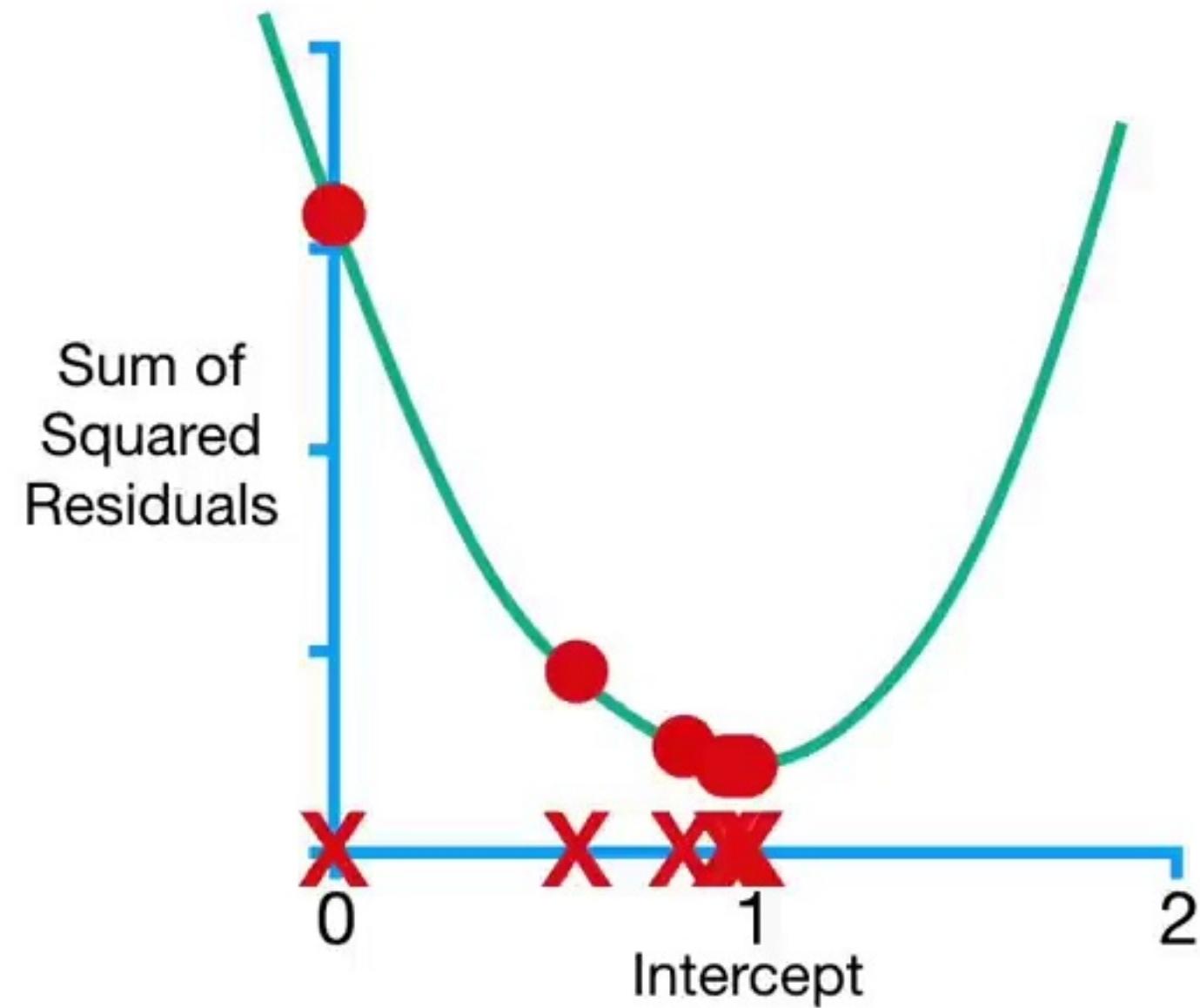
Notice how each step gets smaller and smaller the closer we get to the bottom of the curve.

After 6 steps, the **Gradient Descent** estimate for the **Intercept** is **0.95**.



After 6 steps, the **Gradient Descent** estimate for the **Intercept** is **0.95**.

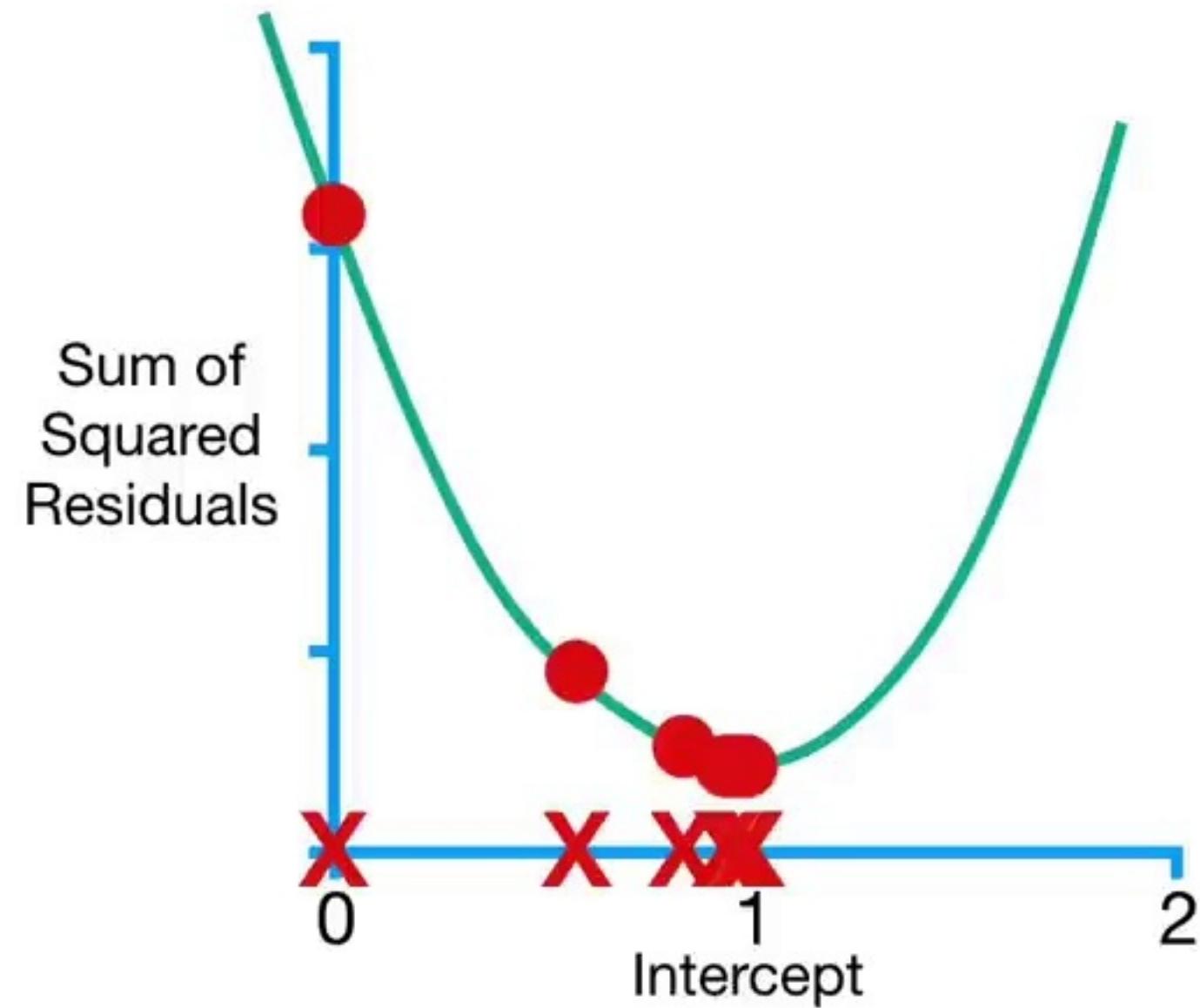
NOTE: The **Least Squares** estimate for the intercept is also **0.95**.



After 6 steps, the **Gradient Descent** estimate for the **Intercept** is **0.95**.

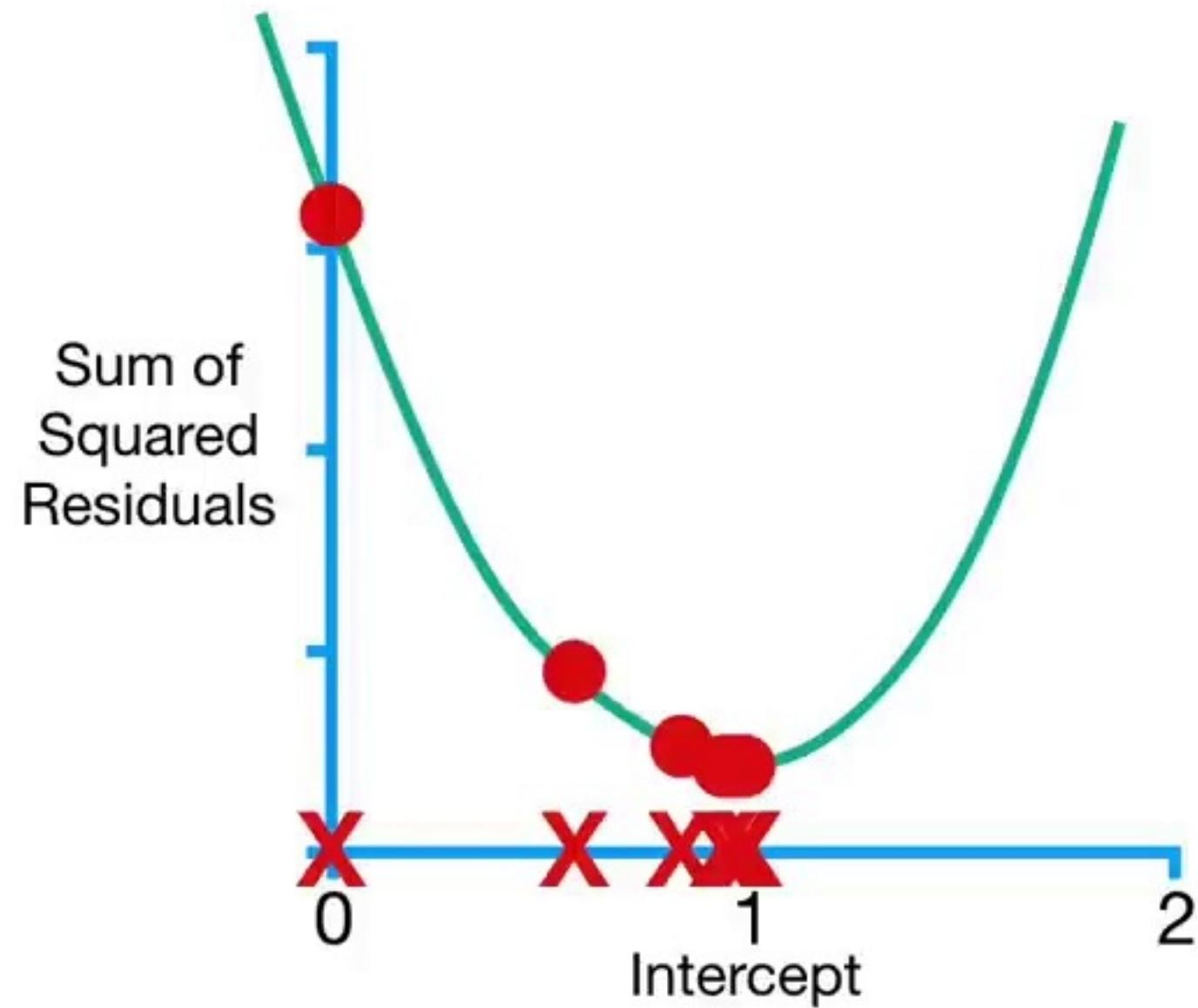
NOTE: The **Least Squares** estimate for the intercept is also **0.95**.

So we know that **Gradient Descent** has done its job, but without comparing its solution to a gold standard, how does **Gradient Descent** know to stop taking steps?



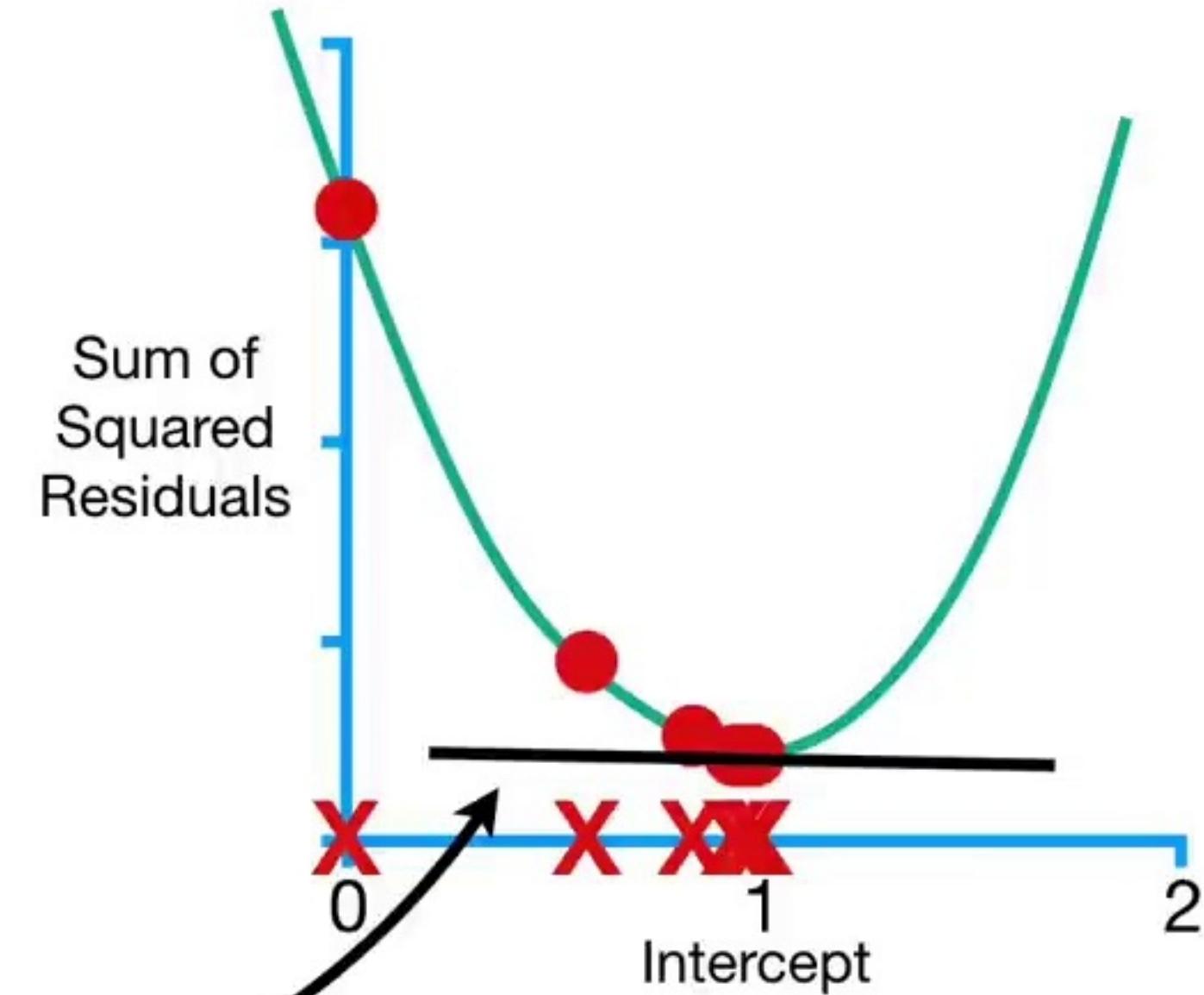
Gradient Descent stops
when the **Step Size** is **Very**
Close To 0.

Step Size = Slope × Learning Rate



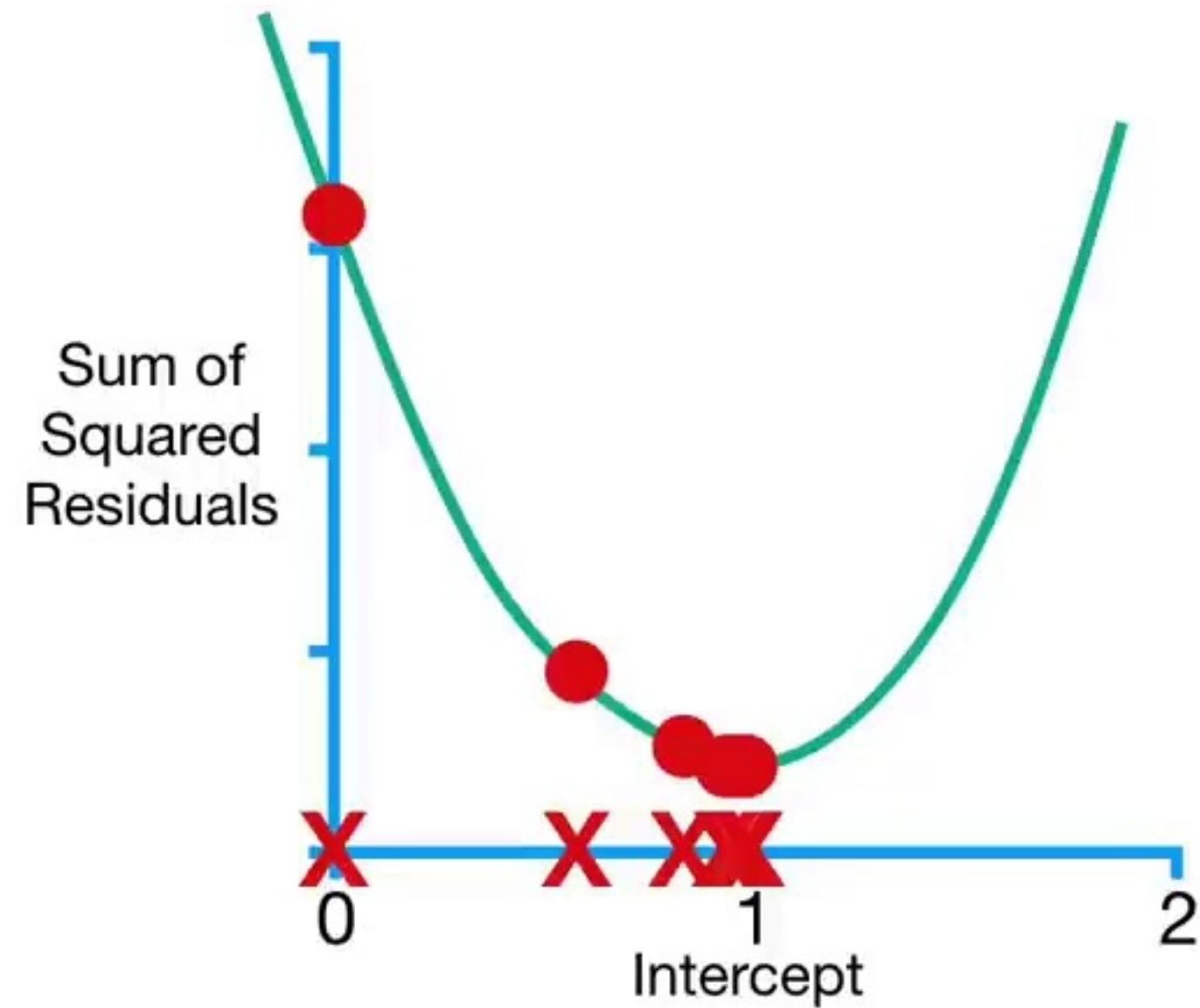
The **Step Size** will be **Very Close to 0** when the **Slope** is very close to **0**.

$$\text{Step Size} = \boxed{\text{Slope}} \times \text{Learning Rate}$$



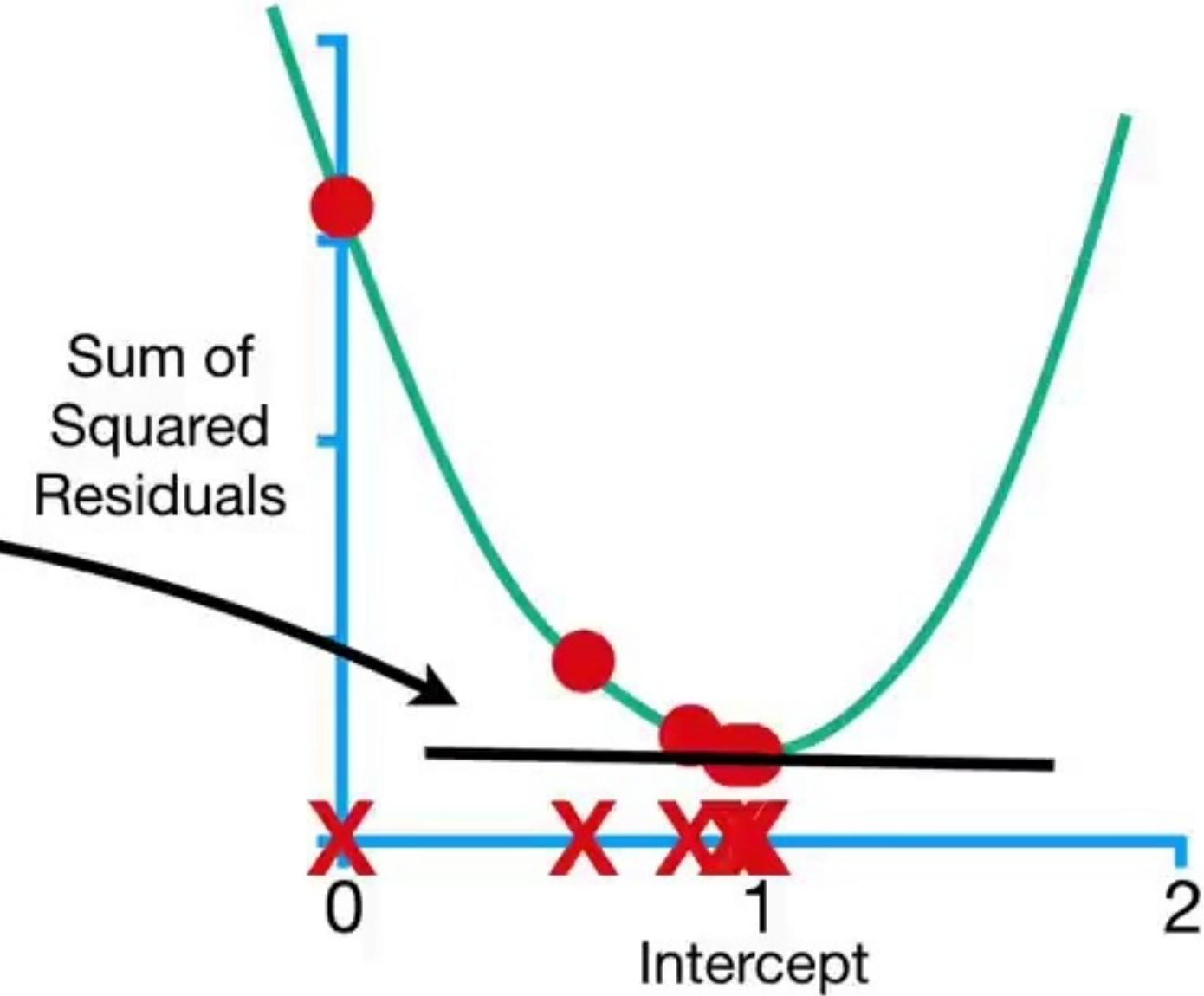
In practice, the
Minimum Step Size = 0.001
or smaller.

Step Size = Slope × Learning Rate



So if this **slope** = 0.009...

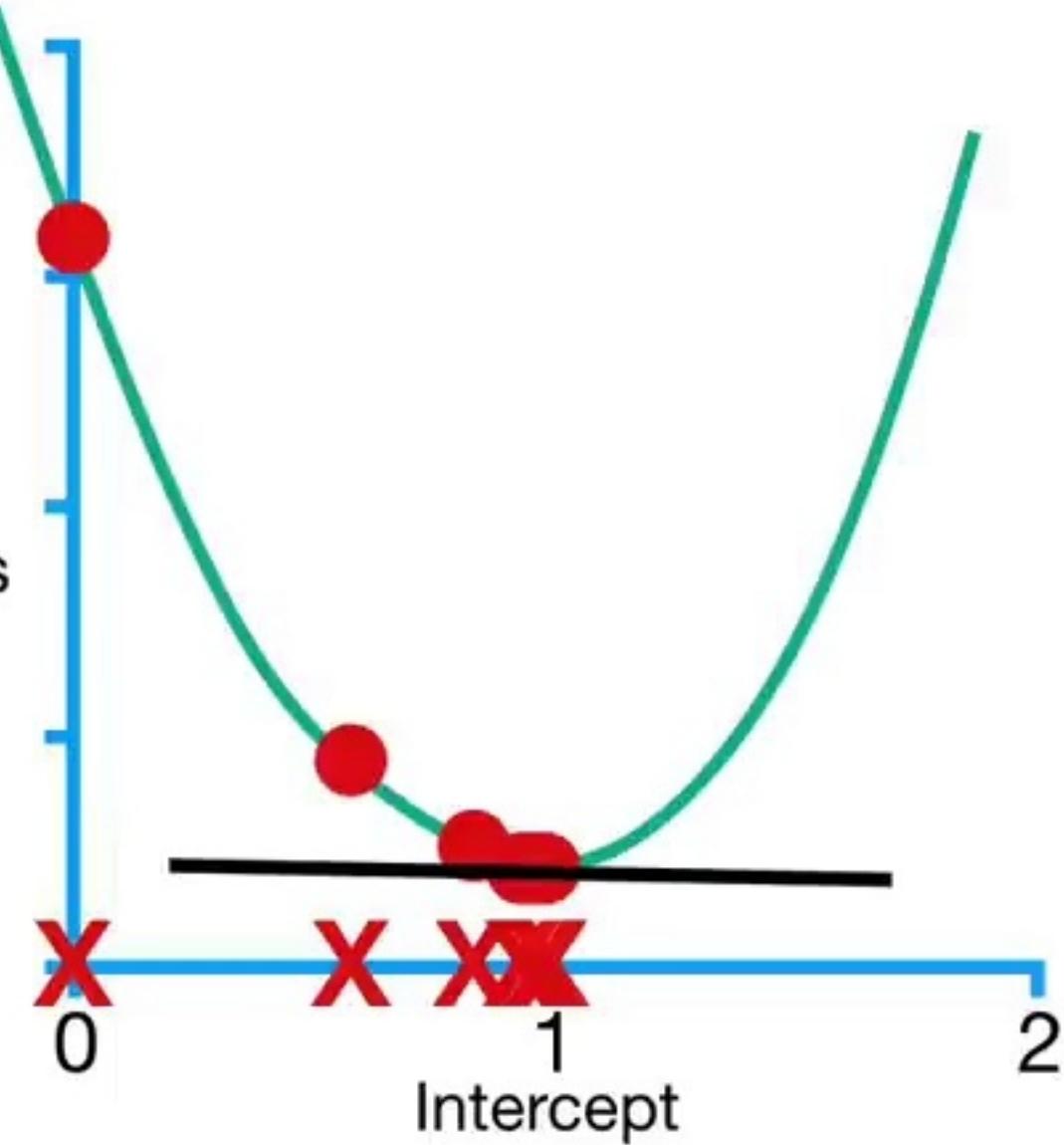
Step Size = Slope × Learning Rate



Then we would plug in
0.009 for the **Slope** and **0.1**
for the **Learning Rate..**

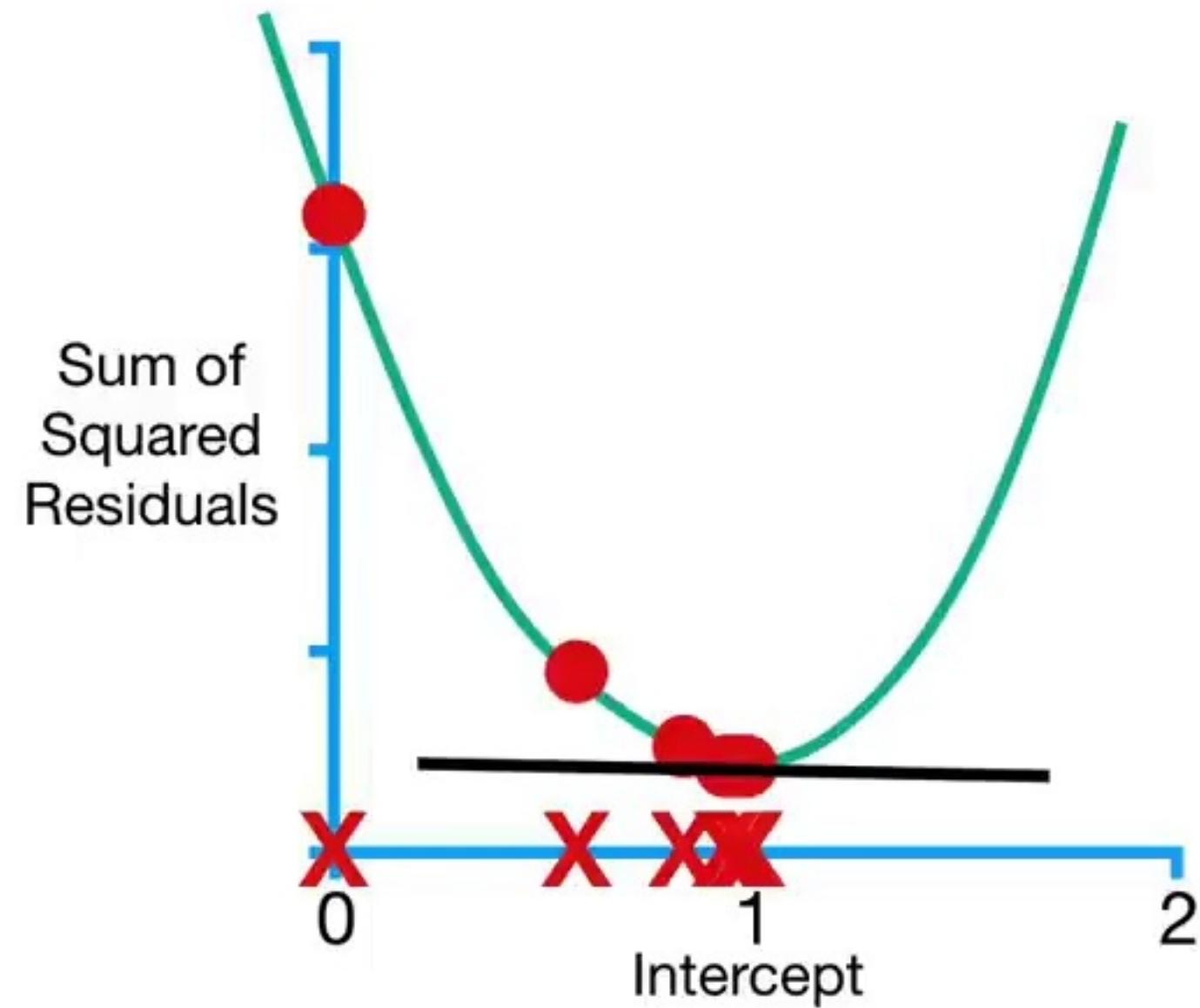
$$\text{Step Size} = \text{Slope} \times \text{Learning Rate}$$


Sum of
Squared
Residuals



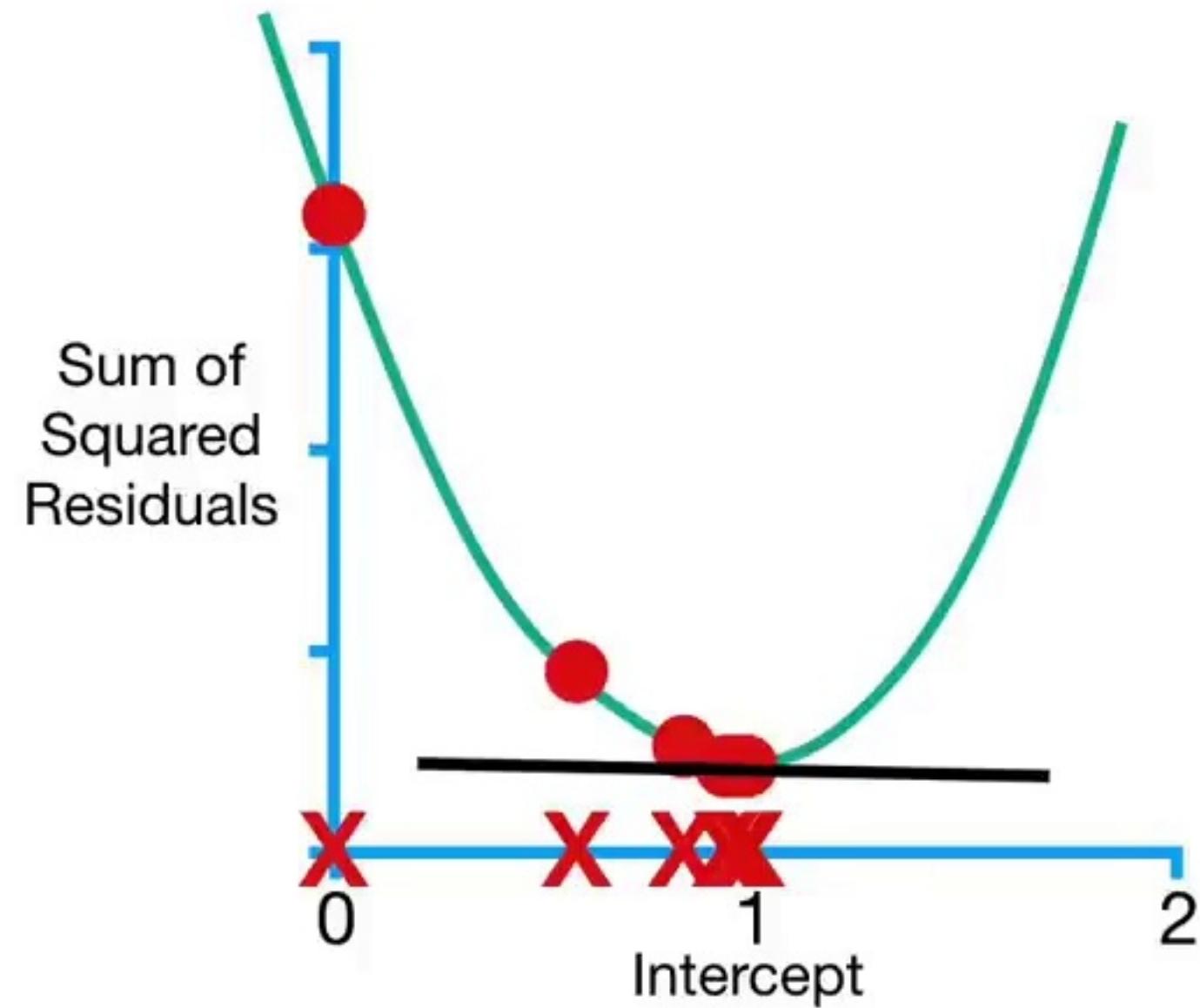
Then we would plug in
0.009 for the **Slope** and **0.1**
for the **Learning Rate..**

$$\text{Step Size} = 0.009 \times 0.1$$

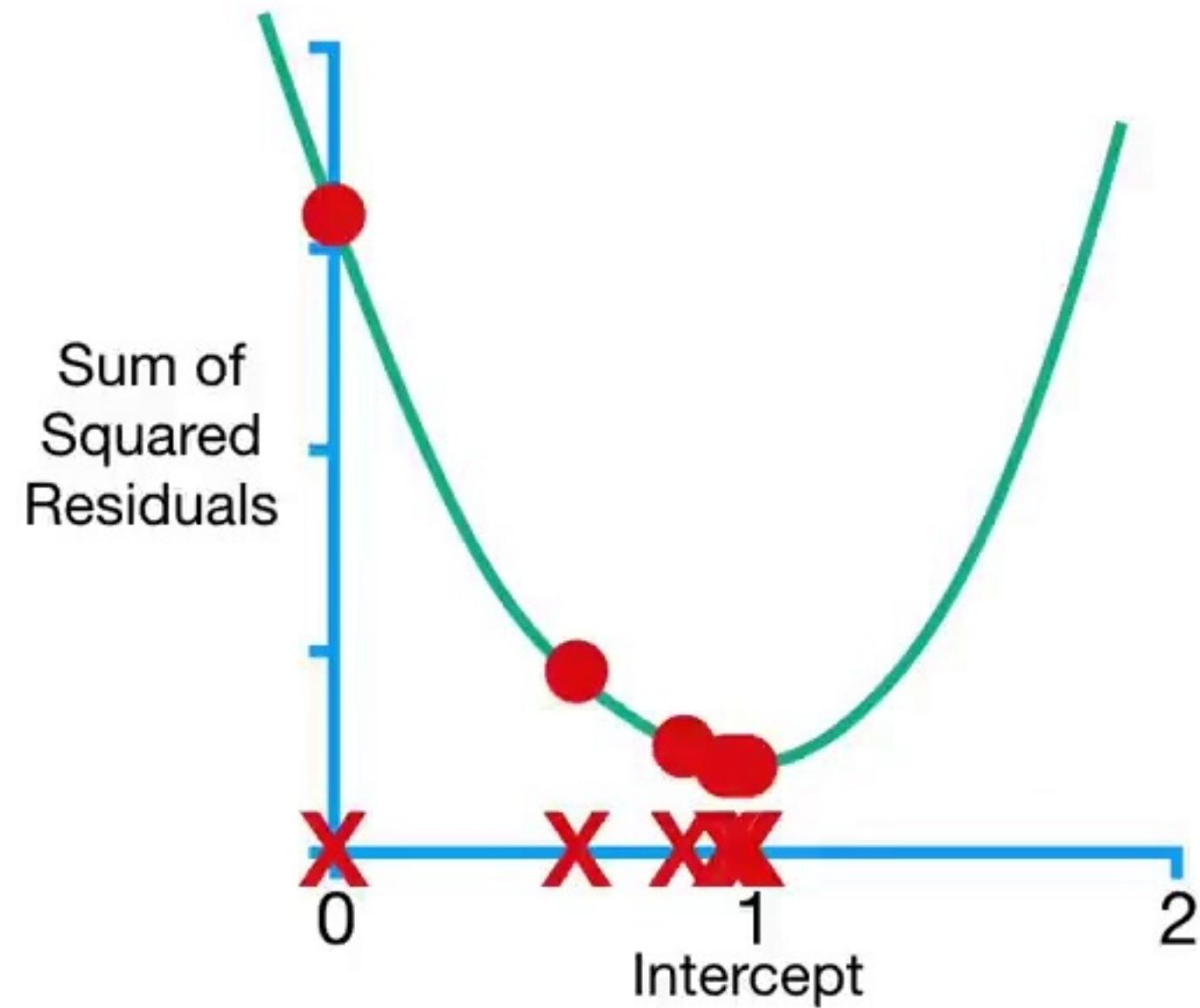


...and get **0.0009**, which is smaller than **0.001**, so **Gradient Descent** would stop.

$$\text{Step Size} = 0.009 \times 0.1 = 0.0009$$

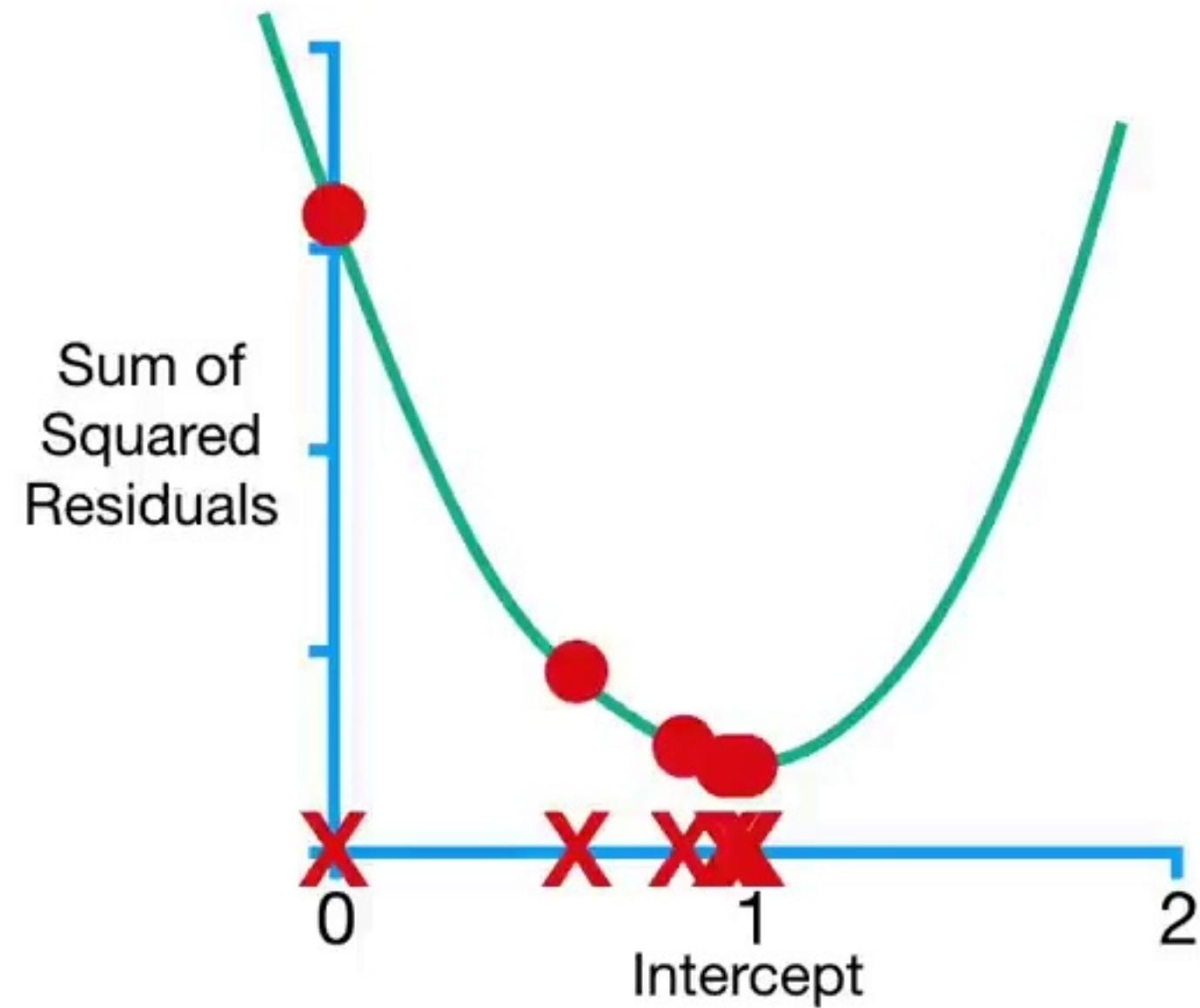


That said, **Gradient Descent** also includes a limit on the number of steps it will take before giving up.

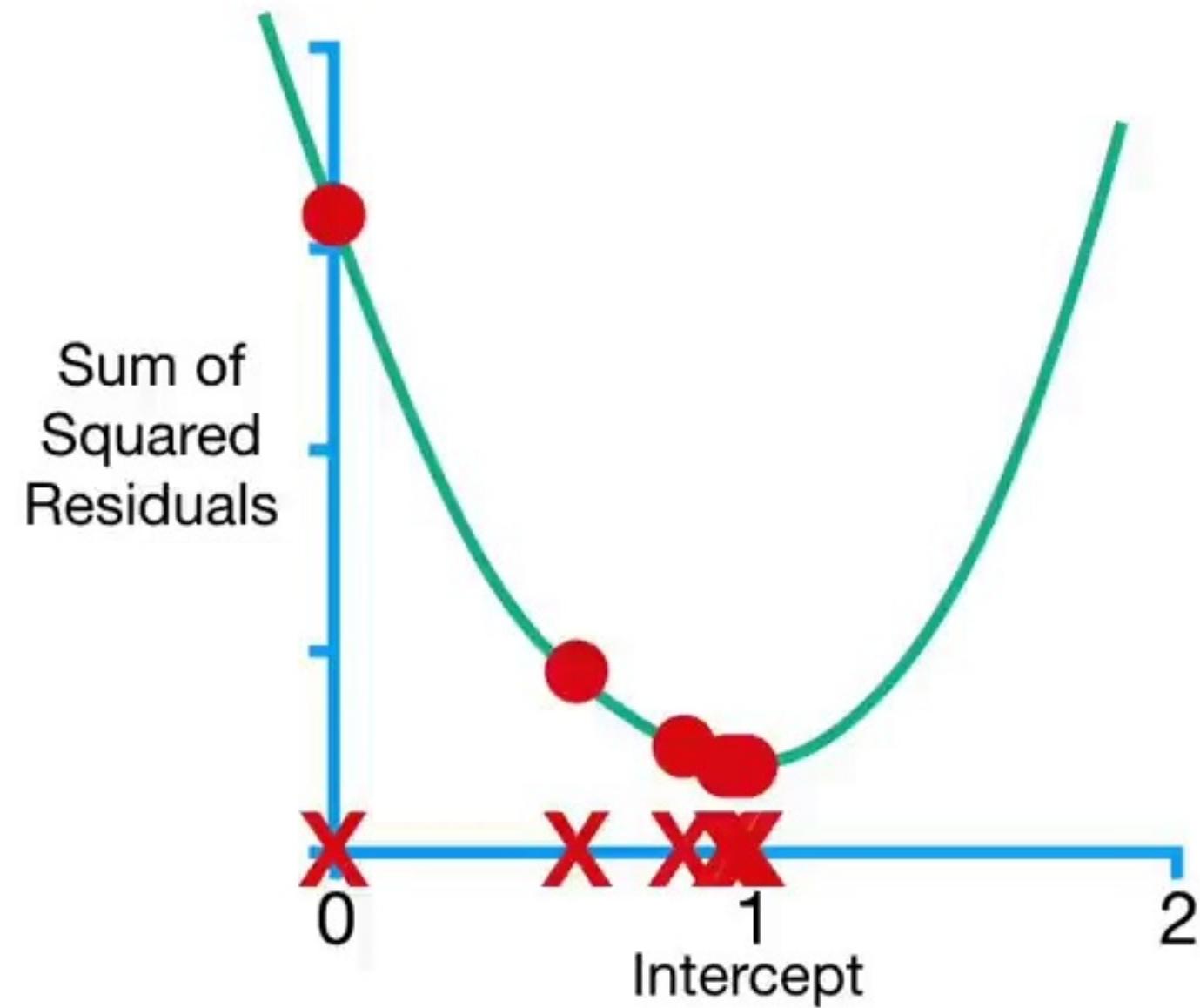


That said, **Gradient Descent** also includes a limit on the number of steps it will take before giving up.

In practice, the **Maximum Number of Steps = 1,000** or greater.

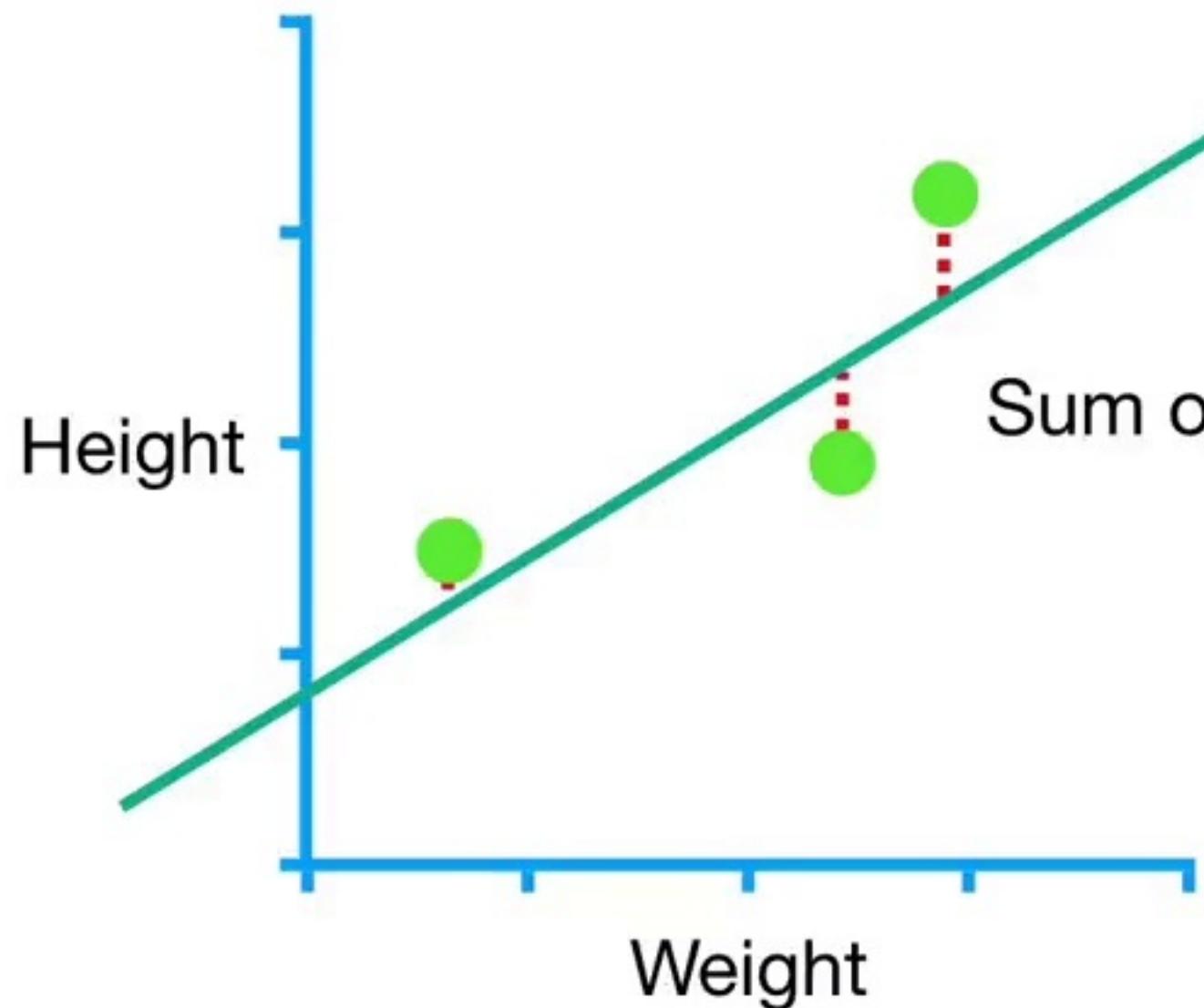


So, even if the **Step Size** is large, if there have been more than the **Maximum Number of Steps, Gradient Descent** will stop.



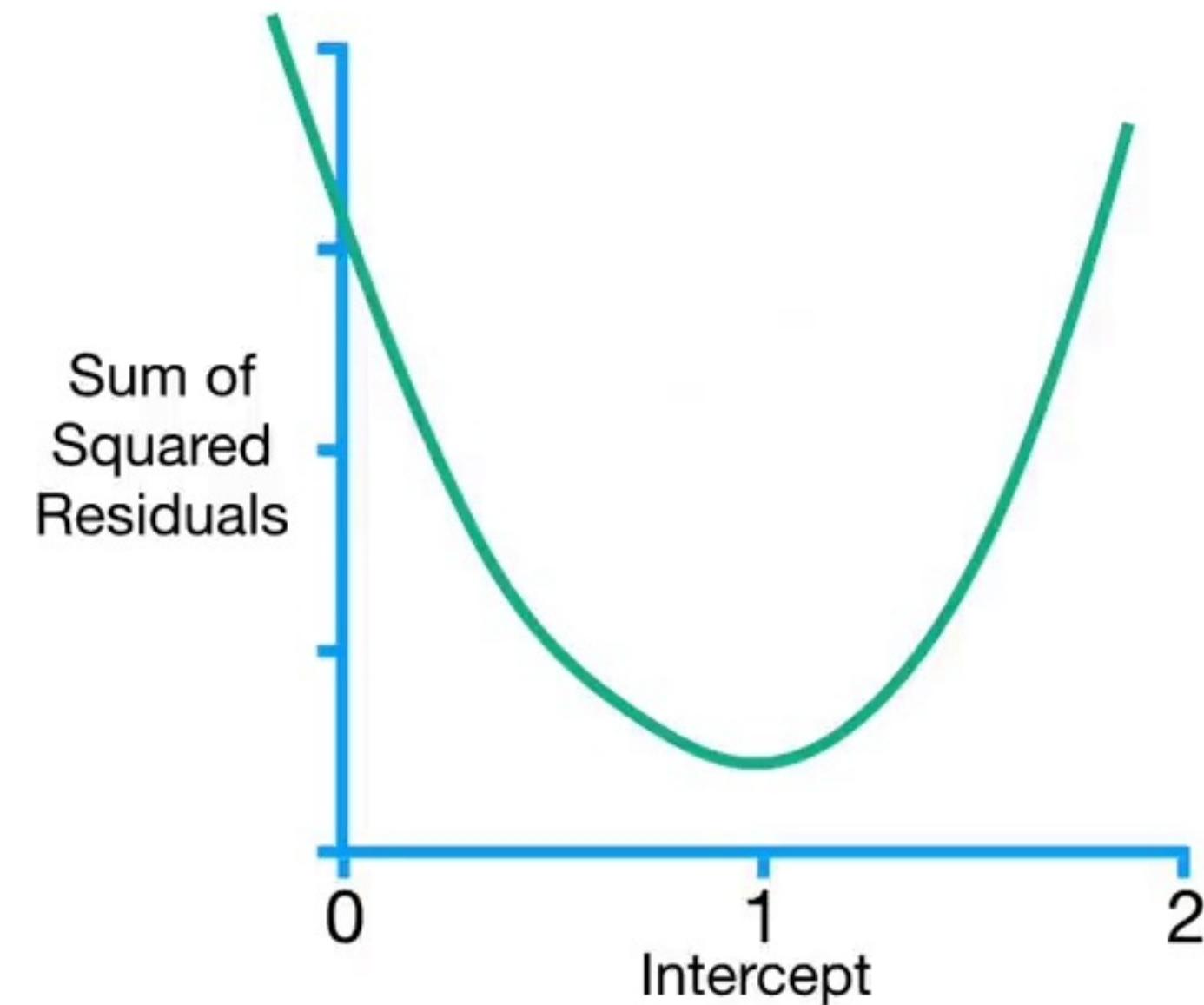
OK, let's review what we've learned so far...

The first thing we did is decide to use the Sum of the Squared Residuals as the **Loss Function** to evaluate how well a line fits the data...



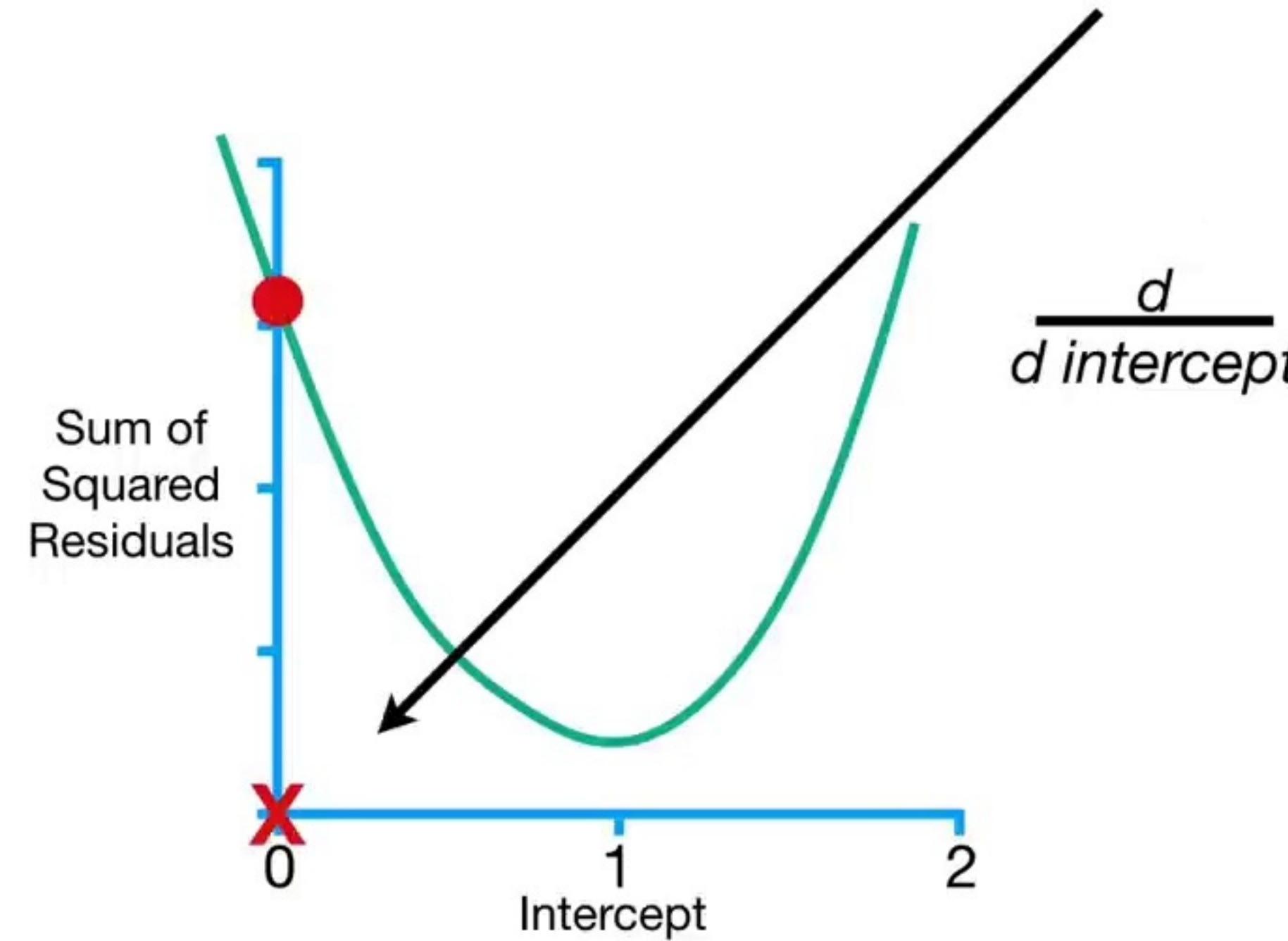
Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$
+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$
+ $(3.2 - (\text{intercept} + 0.64 \times 2.9))^2$

...then we took the derivative of the Sum of the Squared Residuals. In other words, we took the derivative of the **Loss Function**...



$\frac{d}{d \text{ intercept}}$ Sum of squared residuals =
-2(1.4 - (intercept + 0.64 × 0.5))
+ -2(1.9 - (intercept + 0.64 × 2.3))
+ -2(3.2 - (intercept + 0.64 × 2.9))

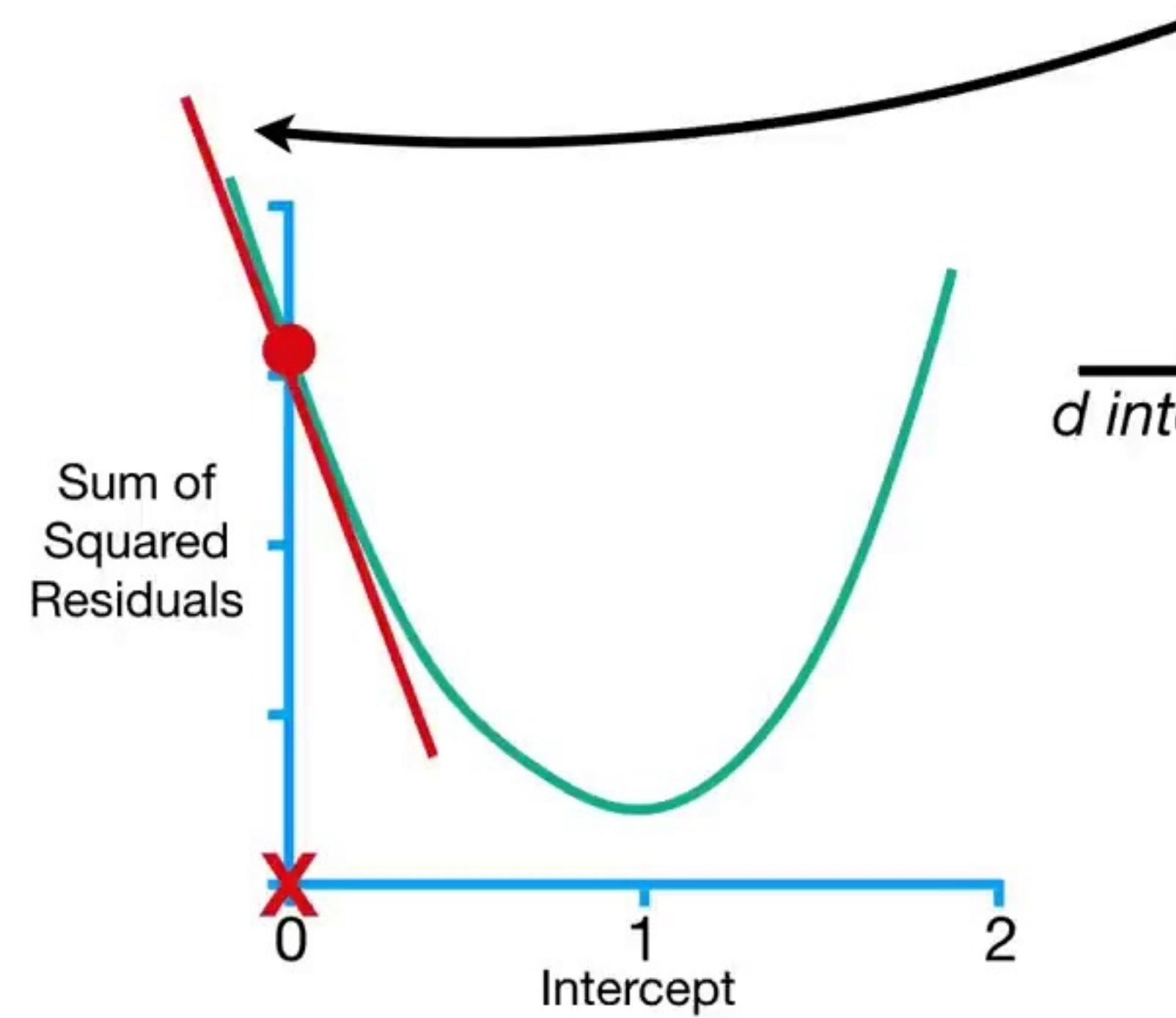
...then we picked a random value for the **Intercept**, in this case we set the **Intercept = 0...**



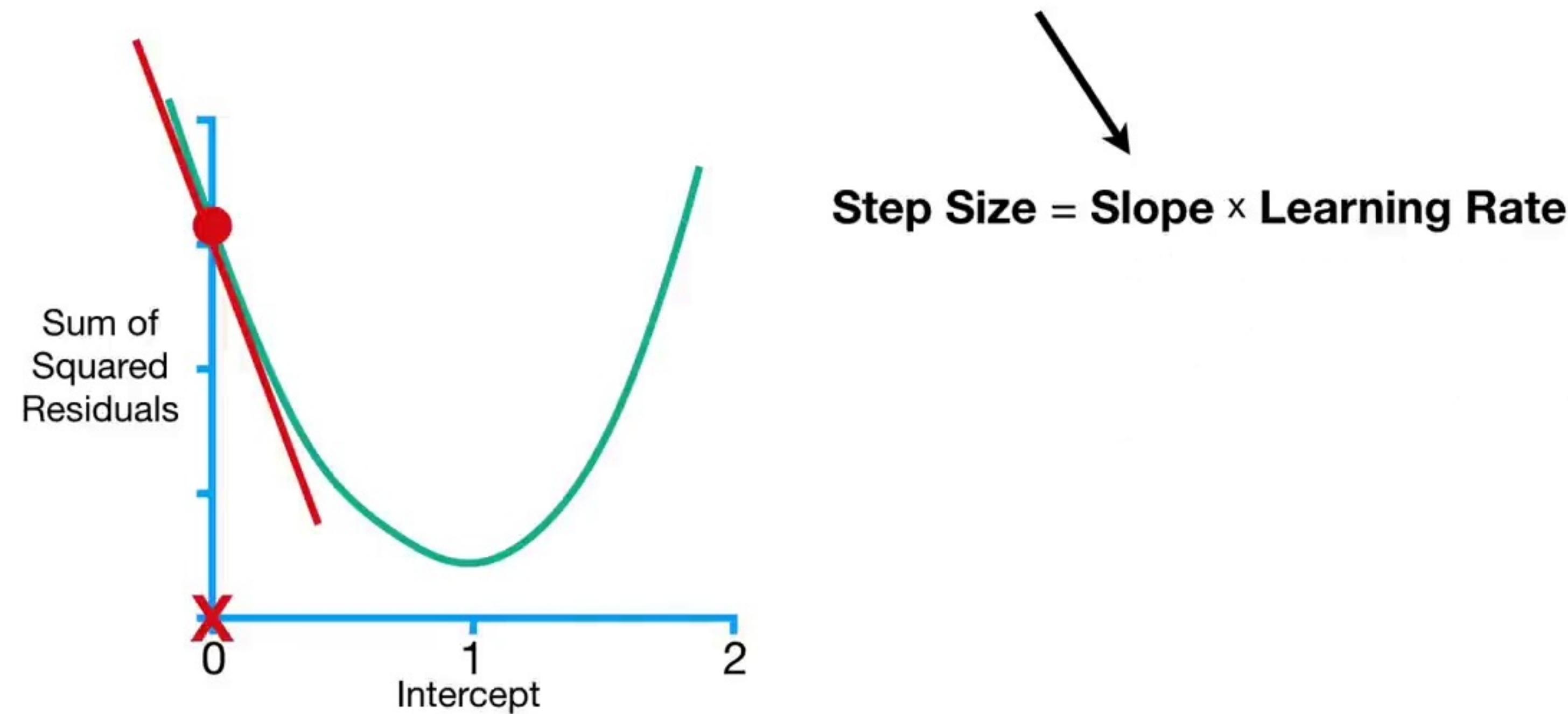
$$\frac{d}{d \text{ intercept}}$$

$$\begin{aligned} \text{Sum of squared residuals} &= \\ &-2(1.4 - (\text{intercept} + 0.64 \times 0.5)) \\ &+ -2(1.9 - (\text{intercept} + 0.64 \times 2.3)) \\ &+ -2(3.2 - (\text{intercept} + 0.64 \times 2.9)) \end{aligned}$$

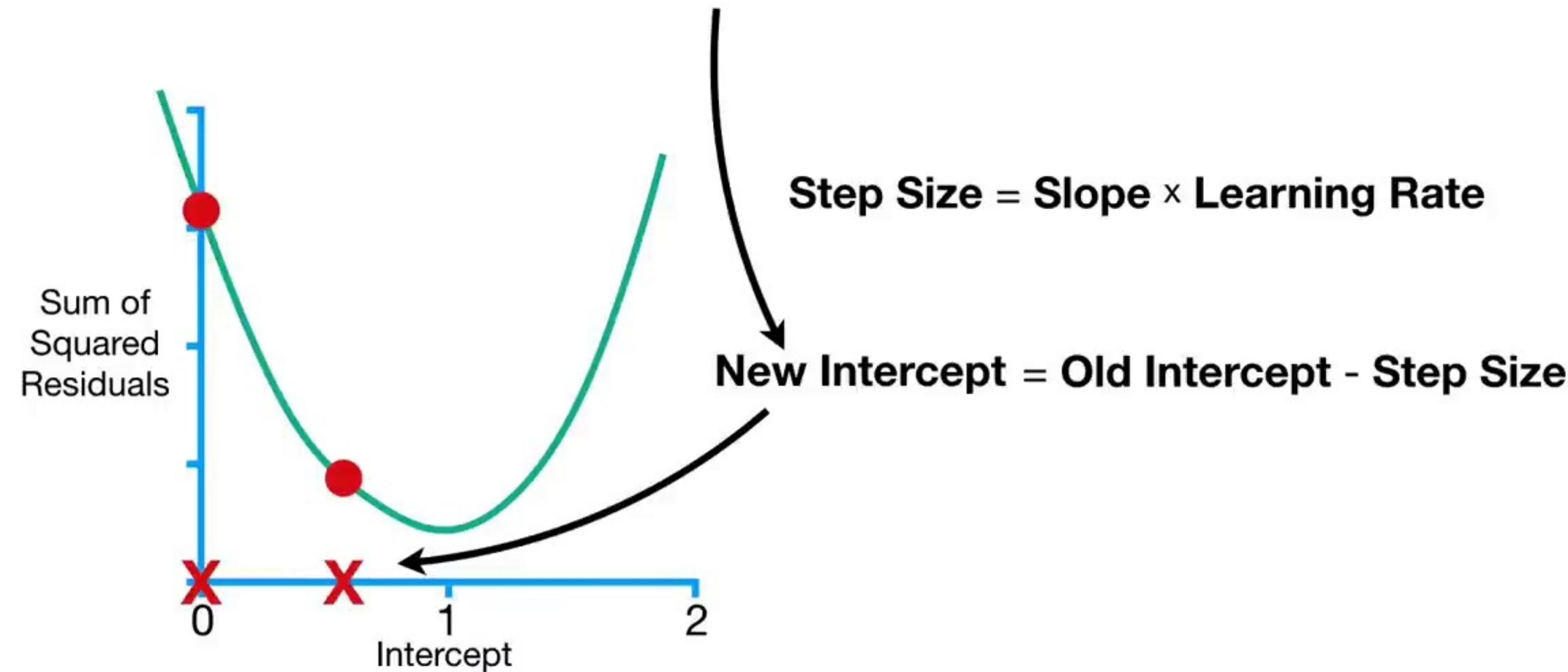
...then we calculated the derivative
when the **Intercept** = 0...



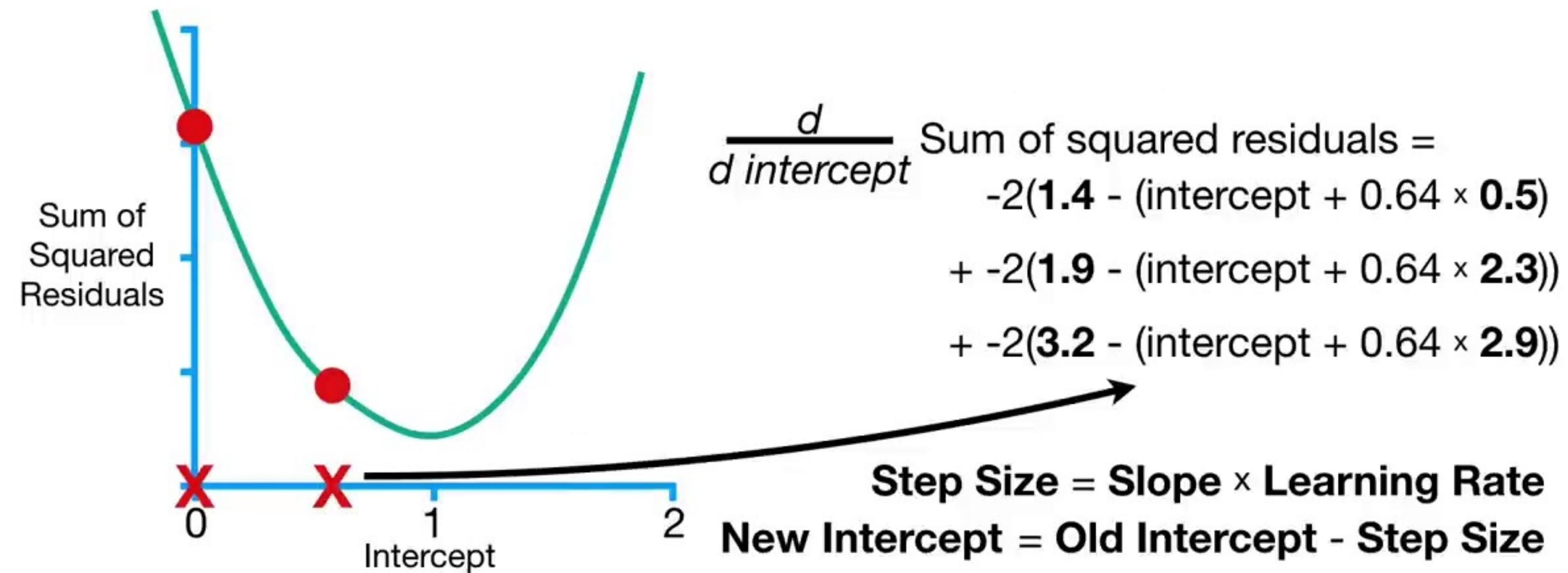
...plugged that slope into the **Step Size** calculation...



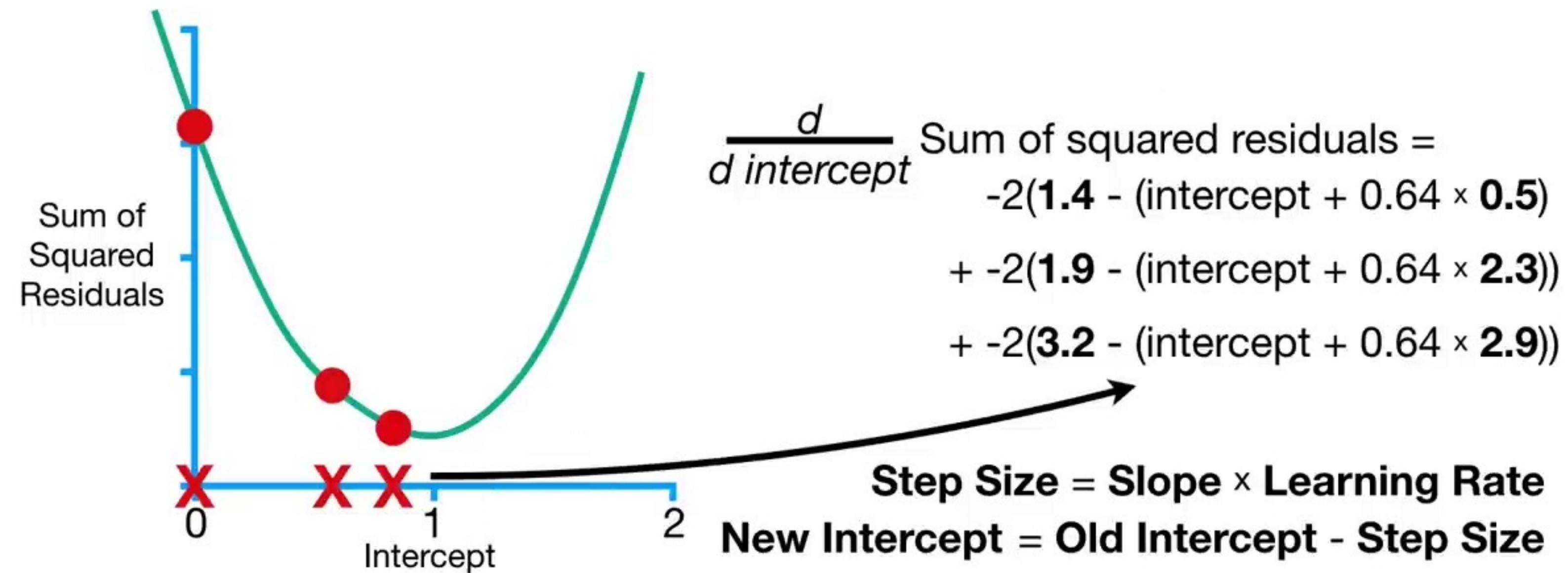
...then calculated the **New Intercept**,
the difference between the **Old
Intercept** and the **Step Size**.



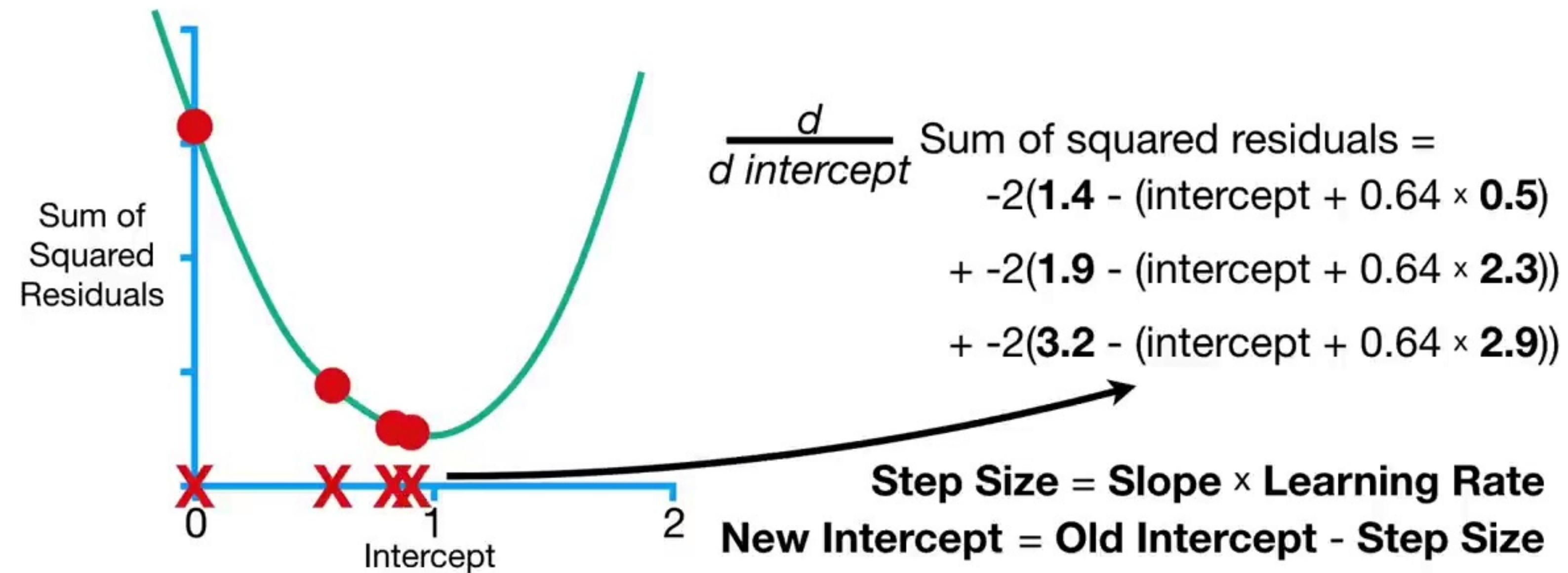
Lastly, we plugged the **New Intercept** into the derivative and repeated everything until **Step Size** was close to 0.



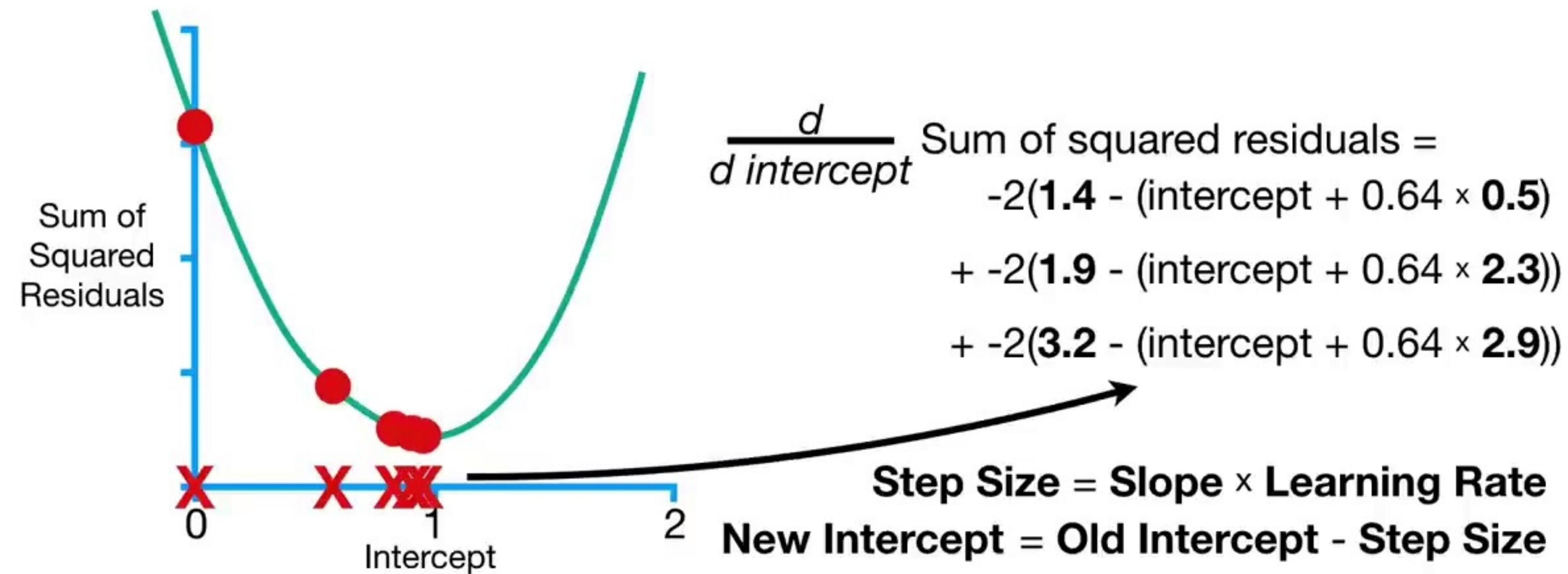
Lastly, we plugged the **New Intercept** into the derivative and repeated everything until **Step Size** was close to 0.



Lastly, we plugged the **New Intercept** into the derivative and repeated everything until **Step Size** was close to 0.



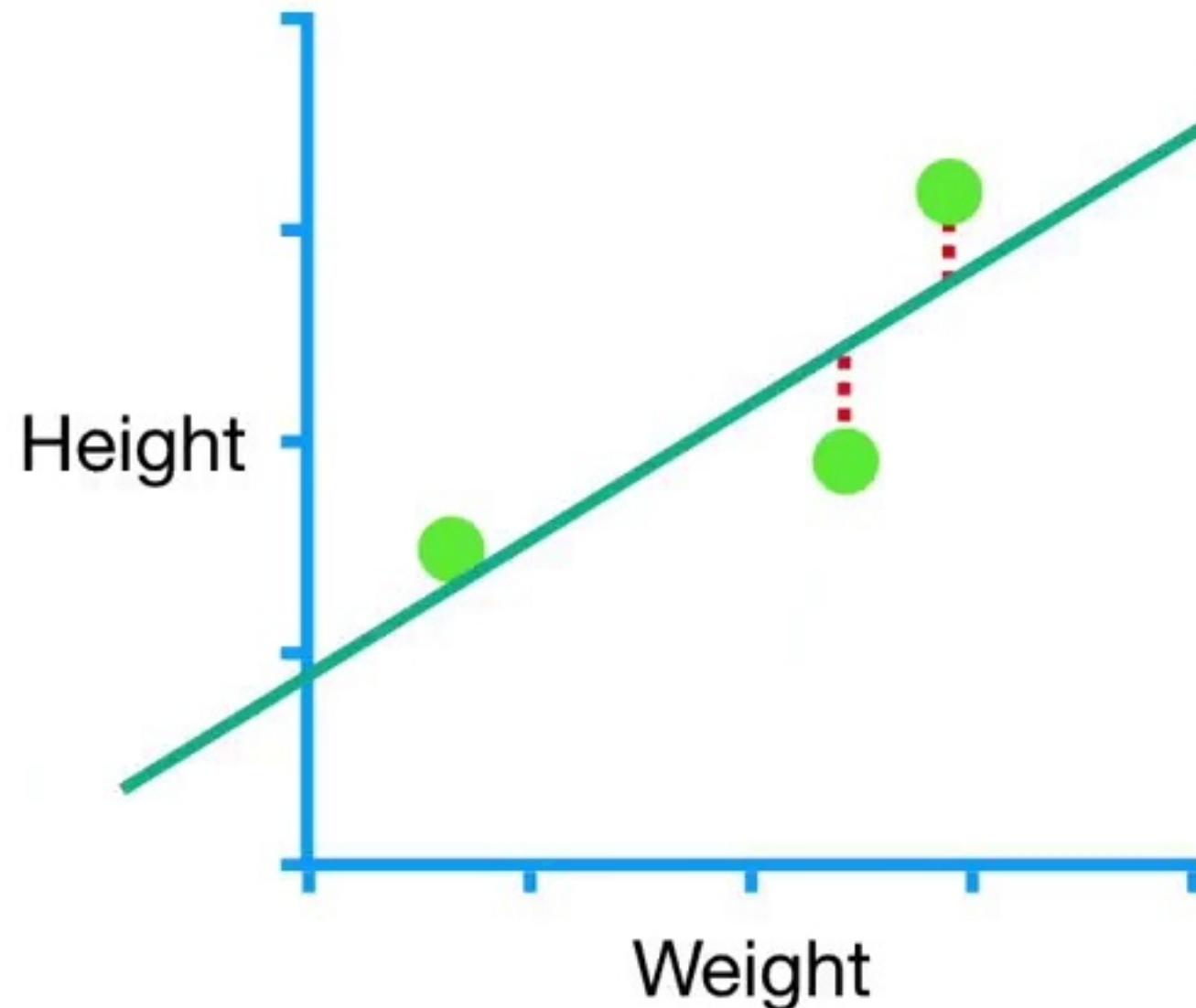
Lastly, we plugged the **New Intercept** into the derivative and repeated everything until **Step Size** was close to 0.



Predicted Height = intercept + $0.64 \times \text{Weight}$



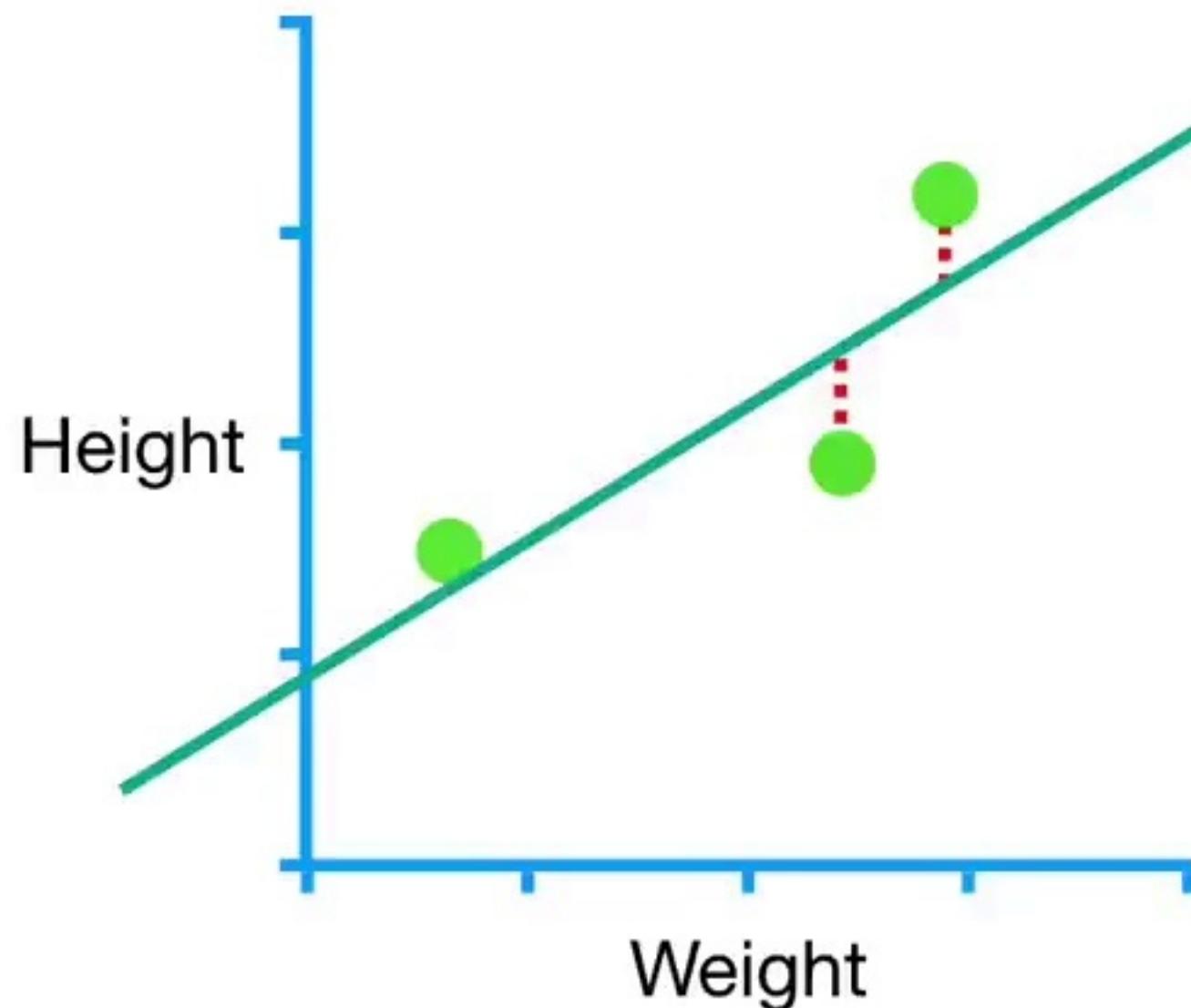
Now that we understand
how **Gradient Descent** can
estimate the **Intercept**...



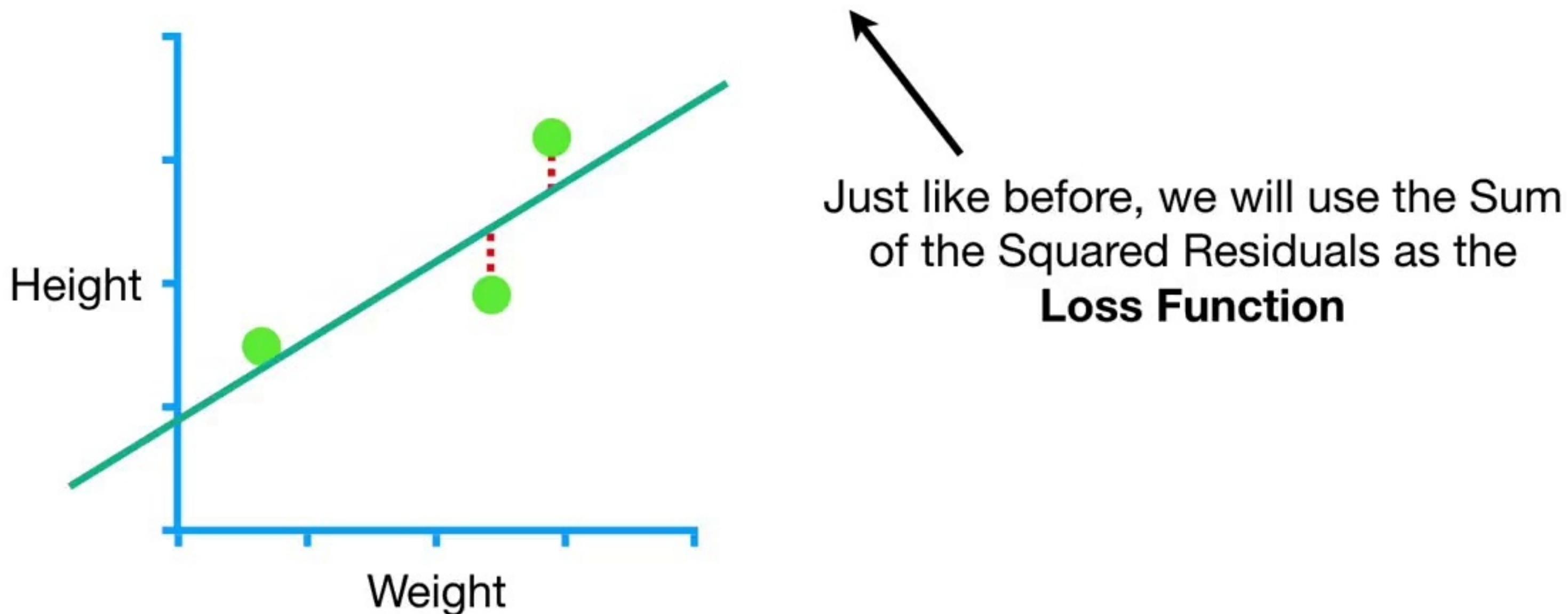
Predicted Height = intercept + slope × **Weight**



...let's talk about how to
estimate the **Intercept** and
the **Slope**.

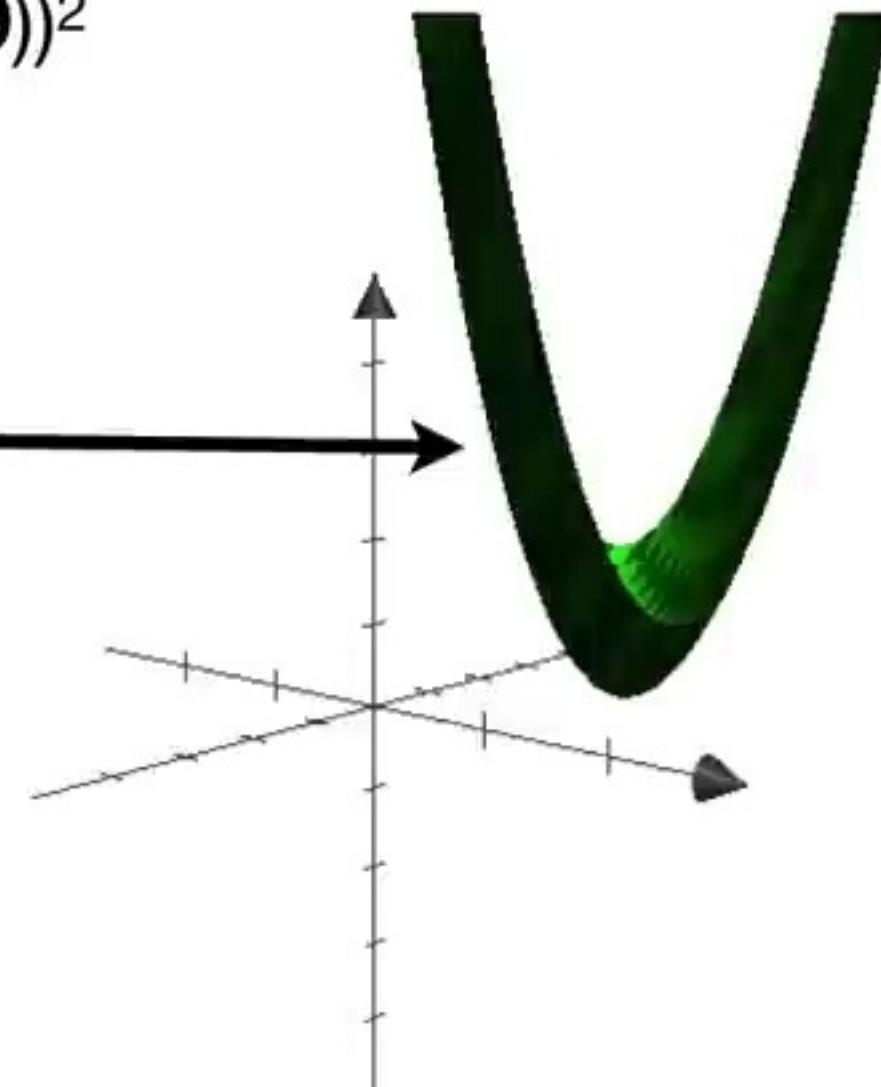


Sum of squared residuals = $(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$
+ $(1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$
+ $(3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$



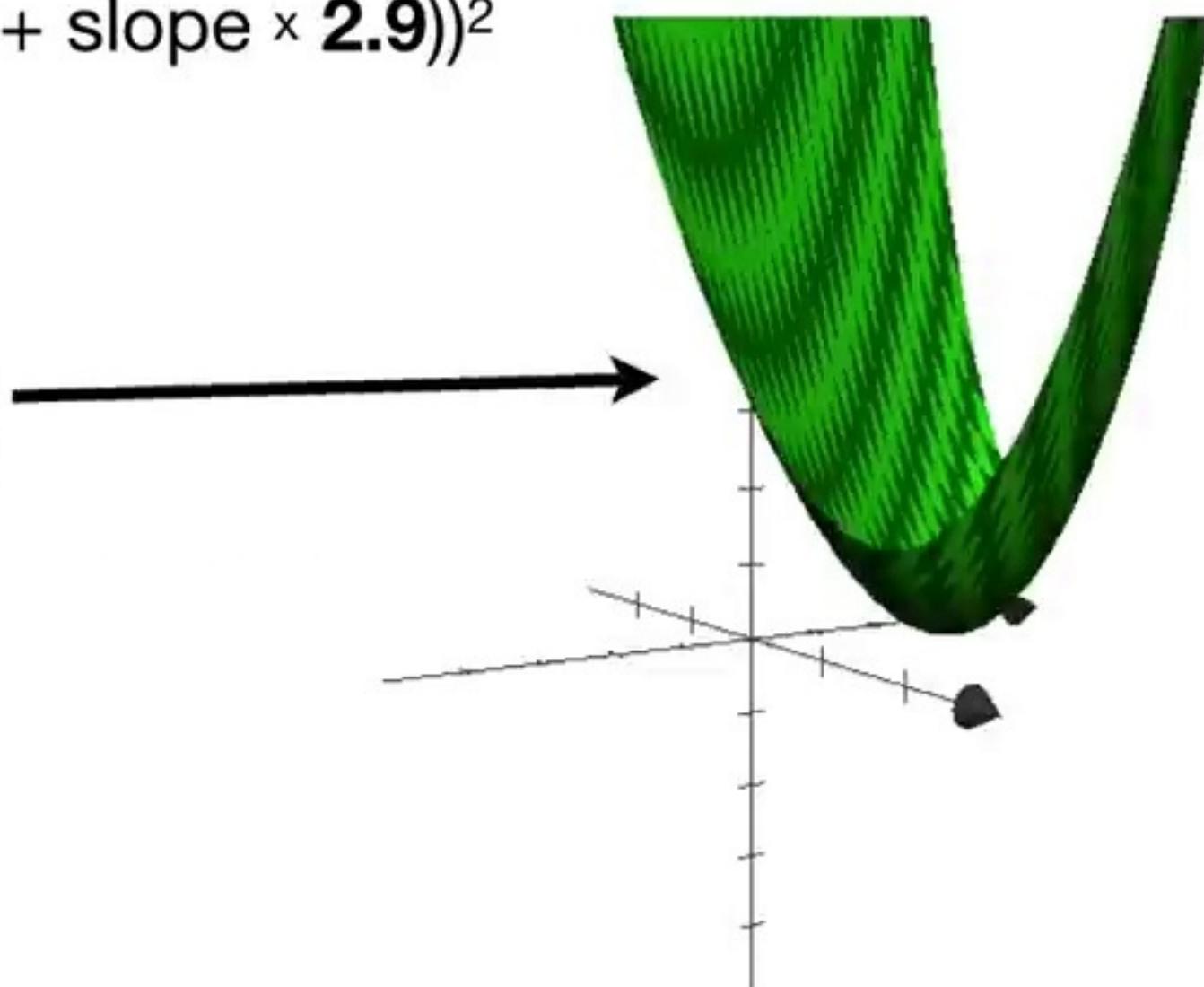
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



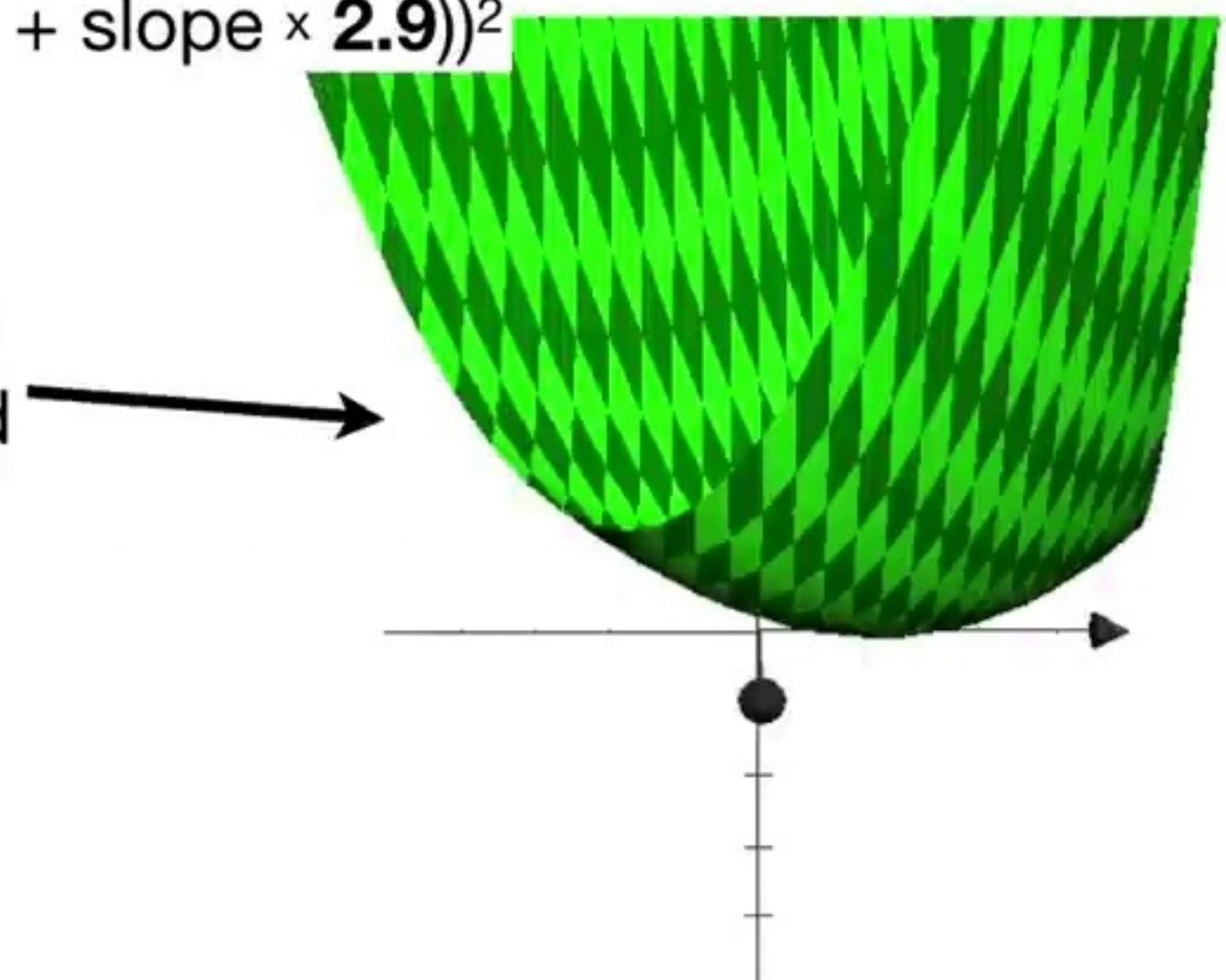
$$\begin{aligned}\text{Sum of squared residuals} = & (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ & + (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ & + (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



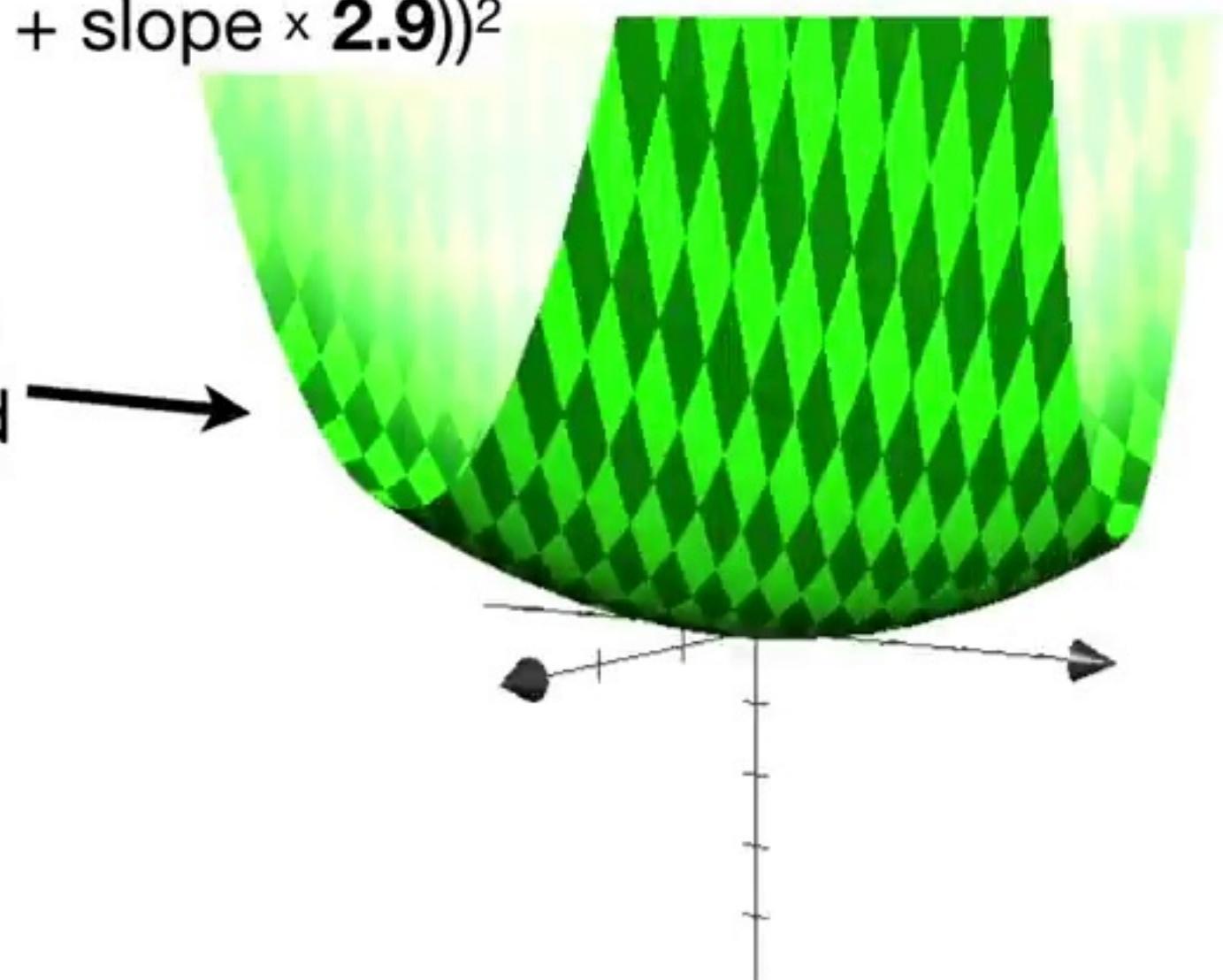
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



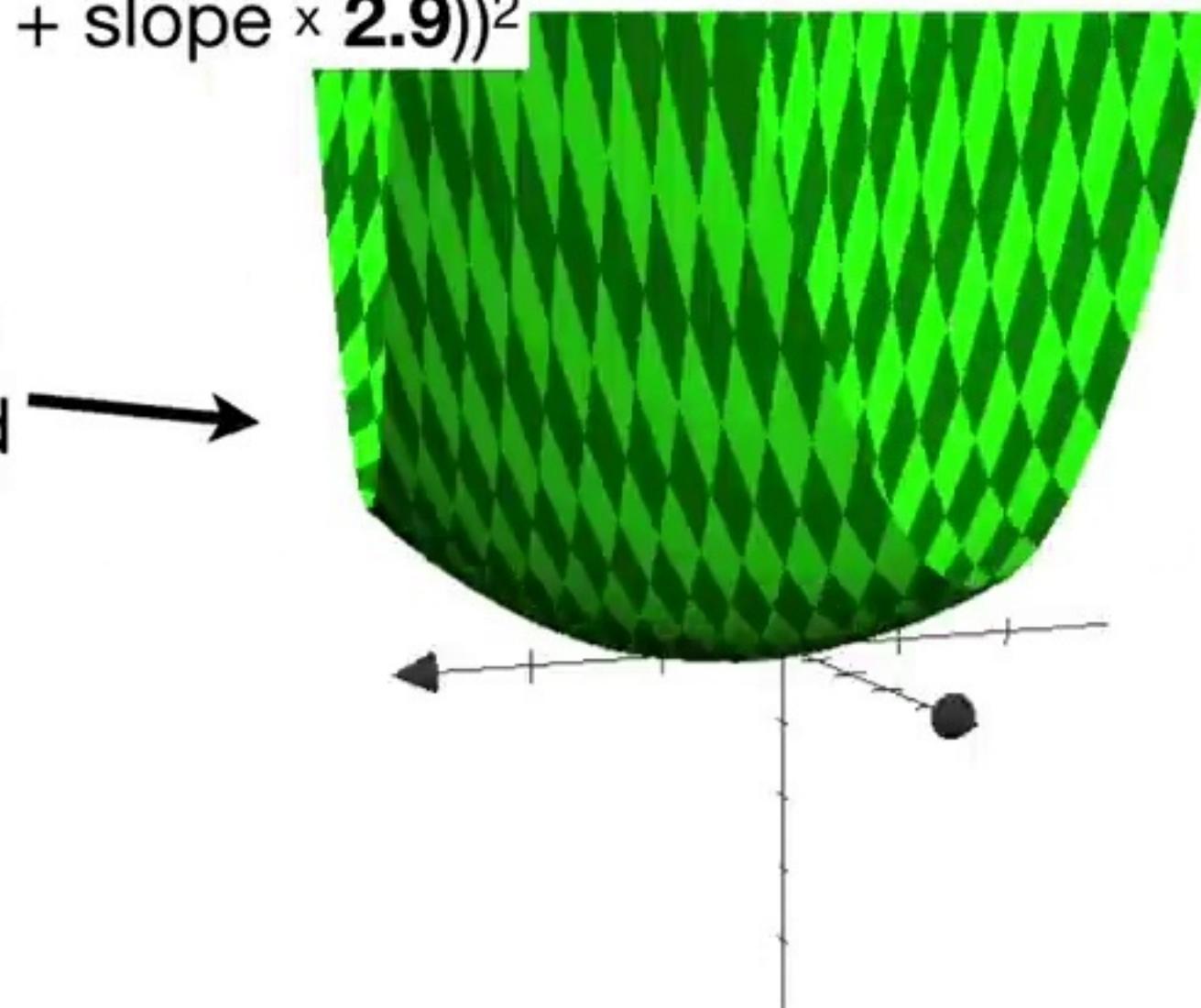
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



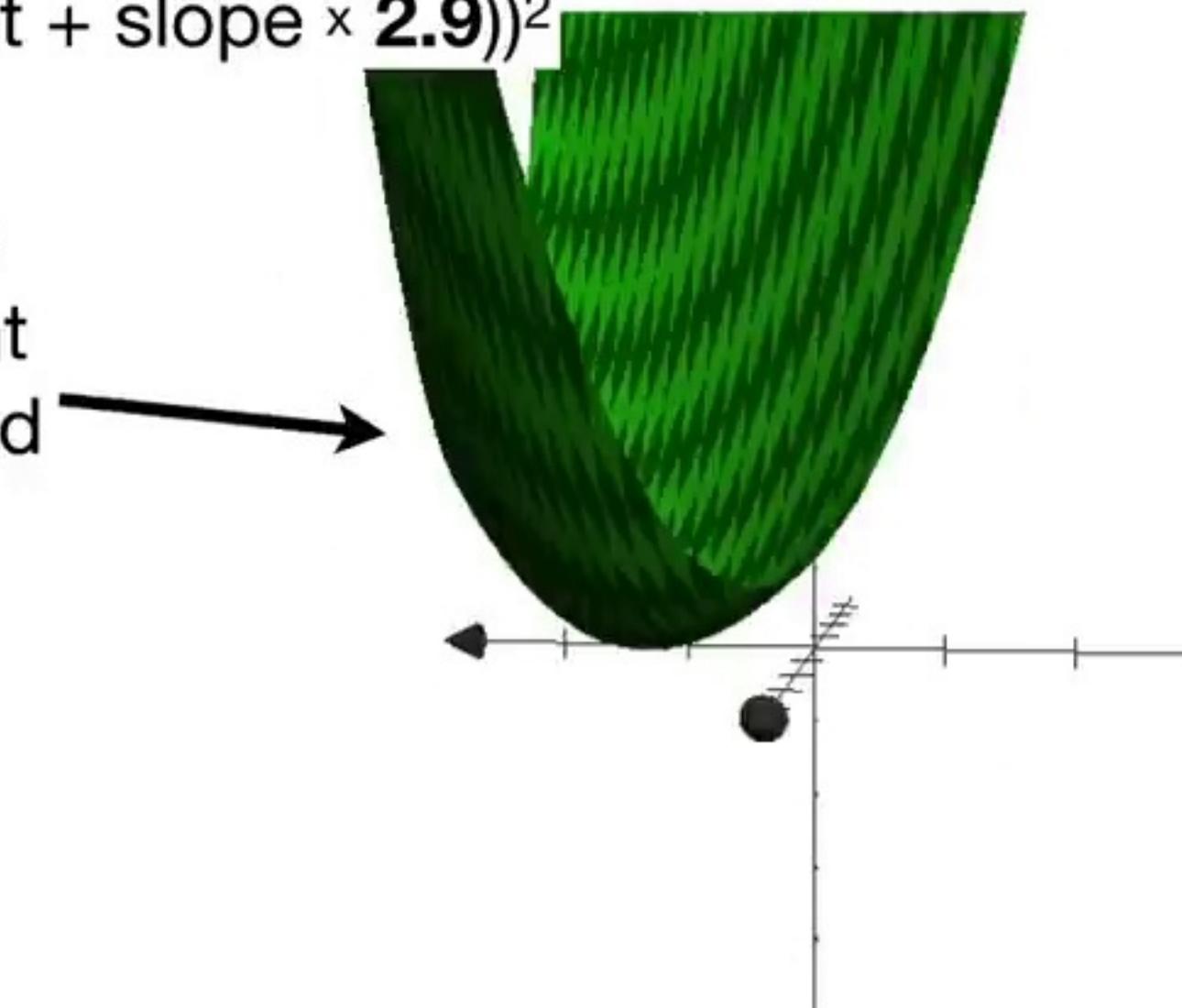
$$\begin{aligned}\text{Sum of squared residuals} = & (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ & + (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ & + (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



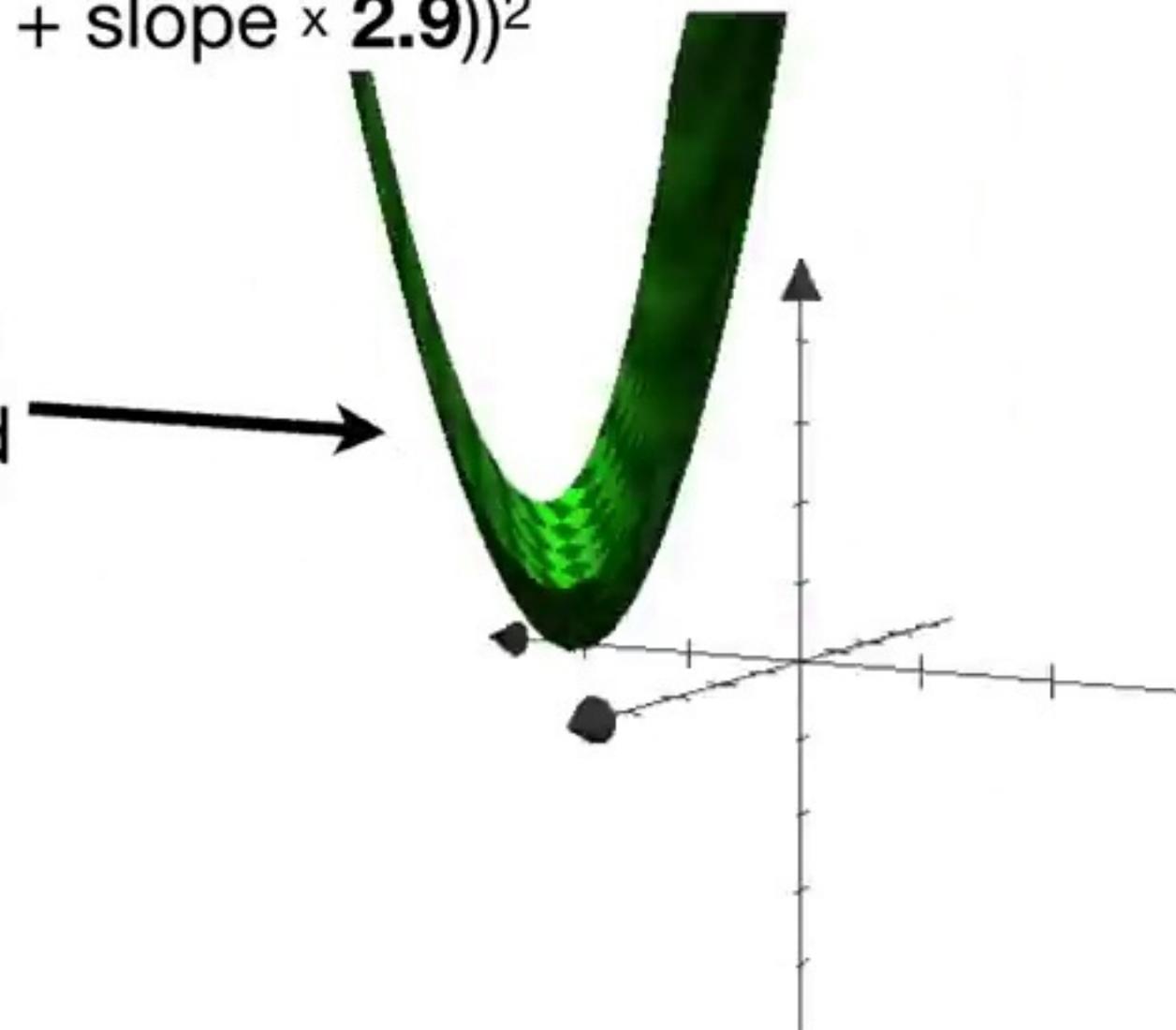
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



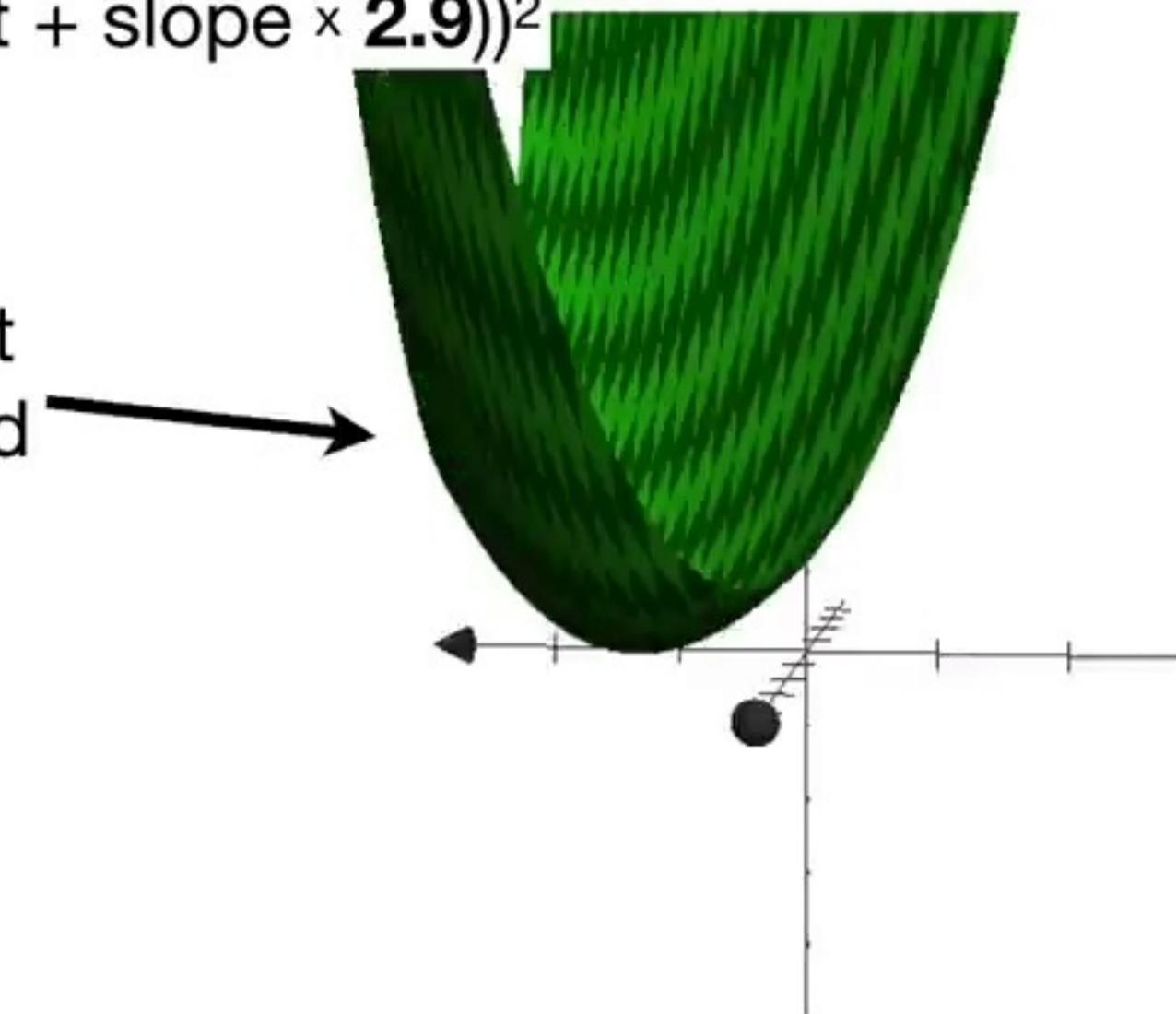
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



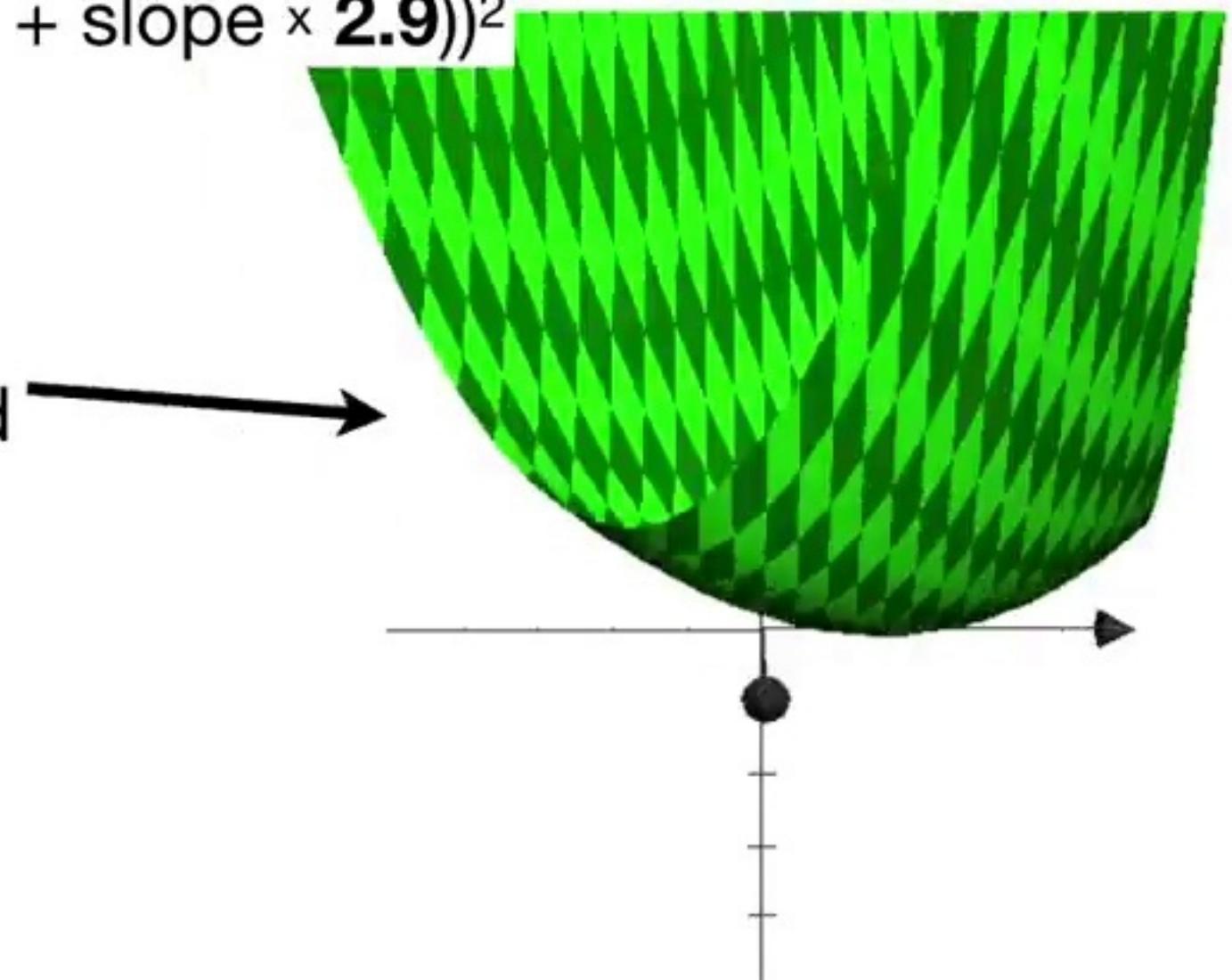
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



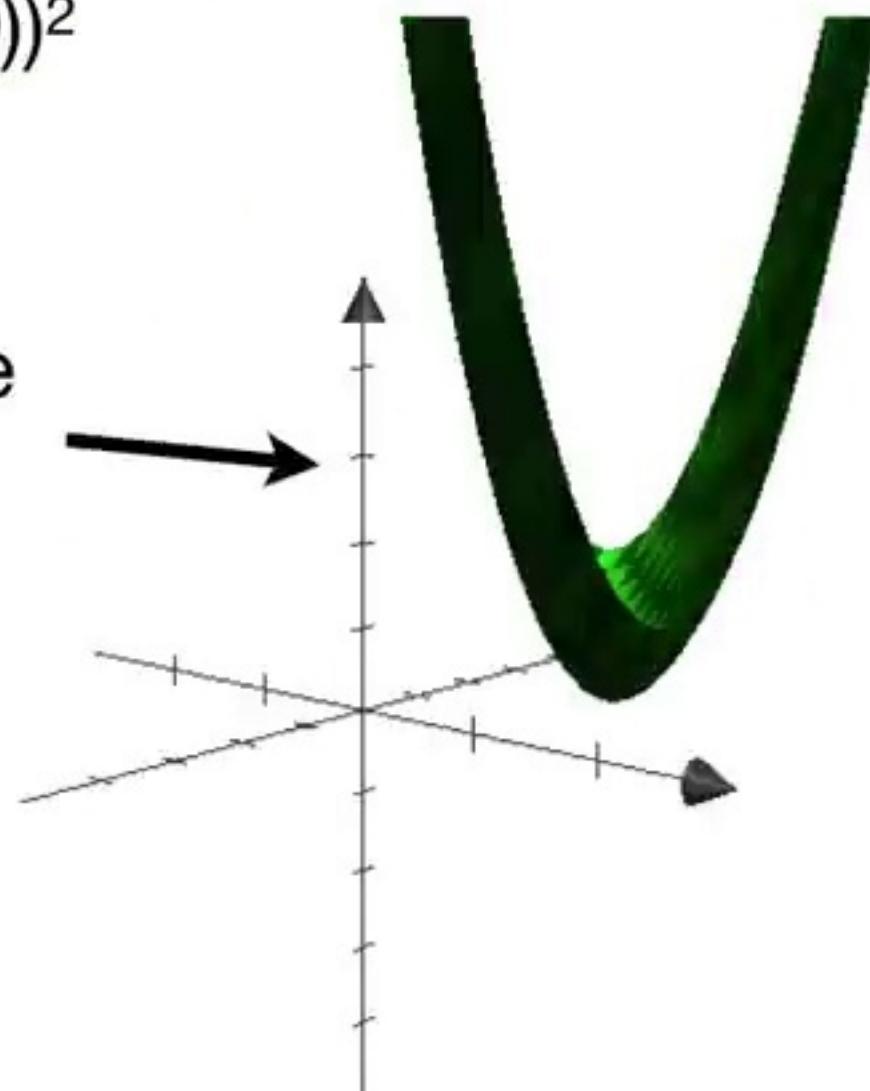
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



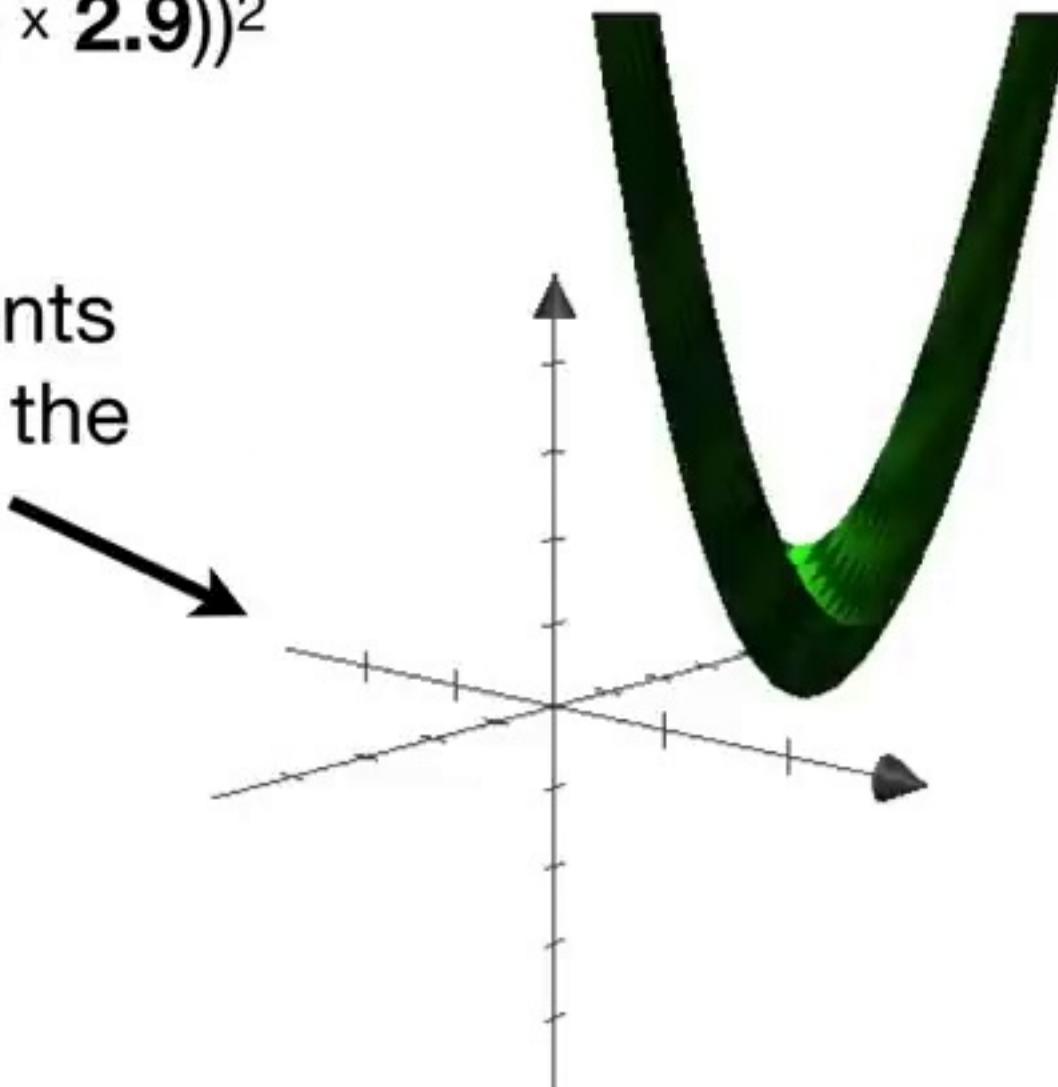
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

This axis is the Sum of the
Squared Residuals...



$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

...this axis represents
different values for the
Slope...

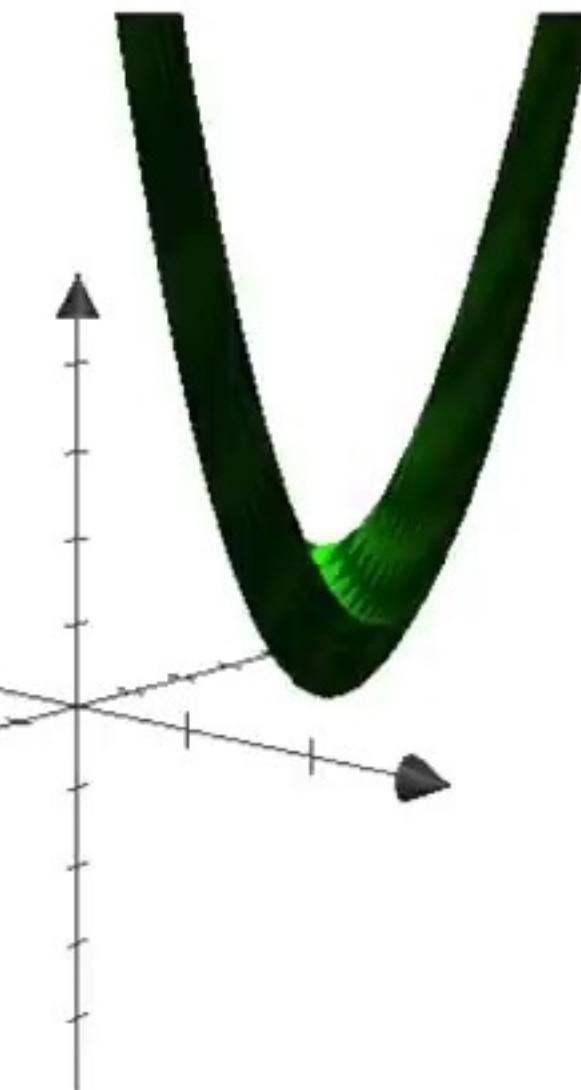


Sum of squared residuals = $(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$

+ $(1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$

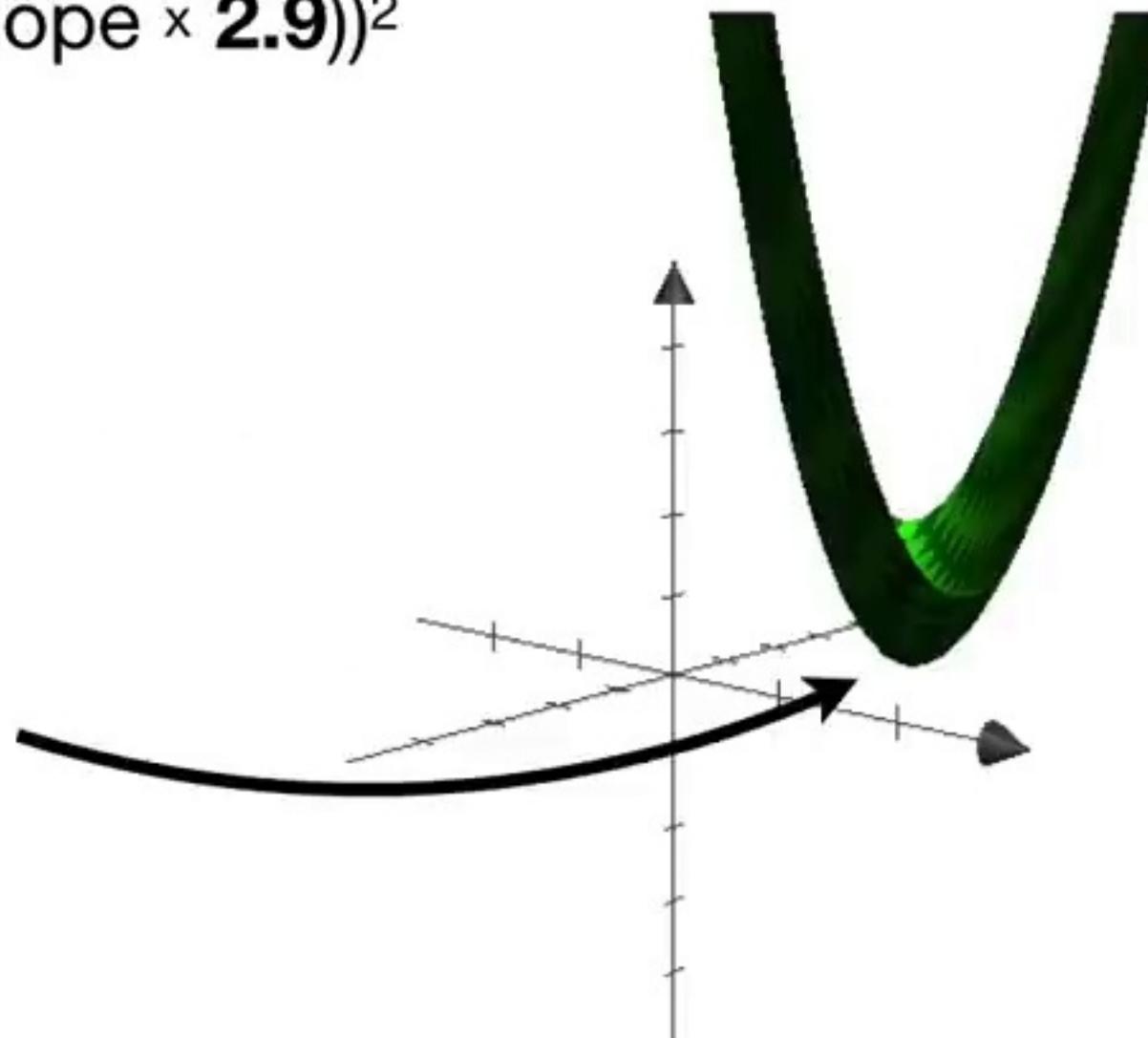
+ $(3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$

...and this axis represents
different values for the
Intercept.



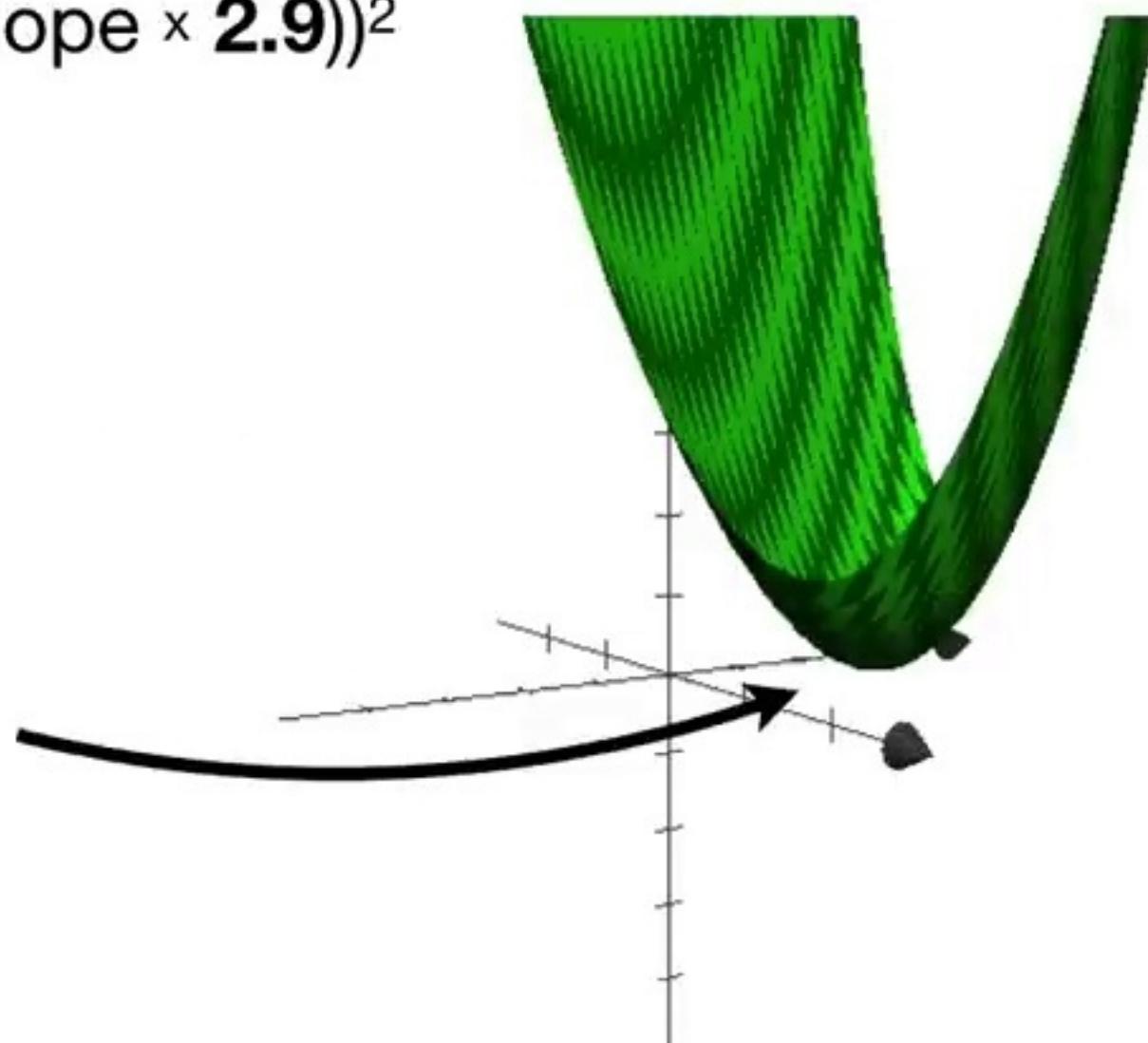
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

We want to find the values for the **Intercept** and **Slope** that give us the minimum Sum of the Squared Residuals.



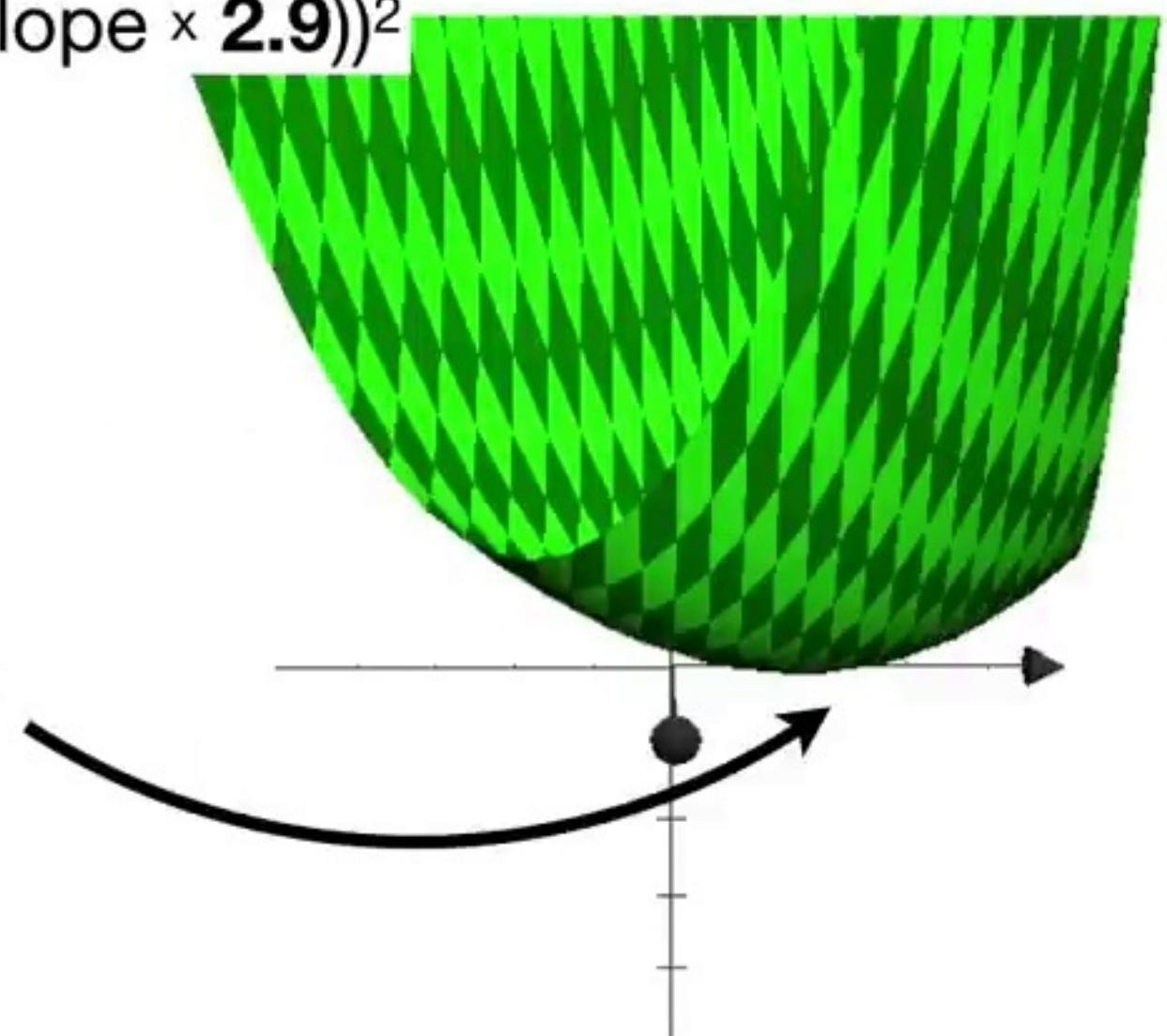
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

We want to find the values for the **Intercept** and **Slope** that give us the minimum Sum of the Squared Residuals.



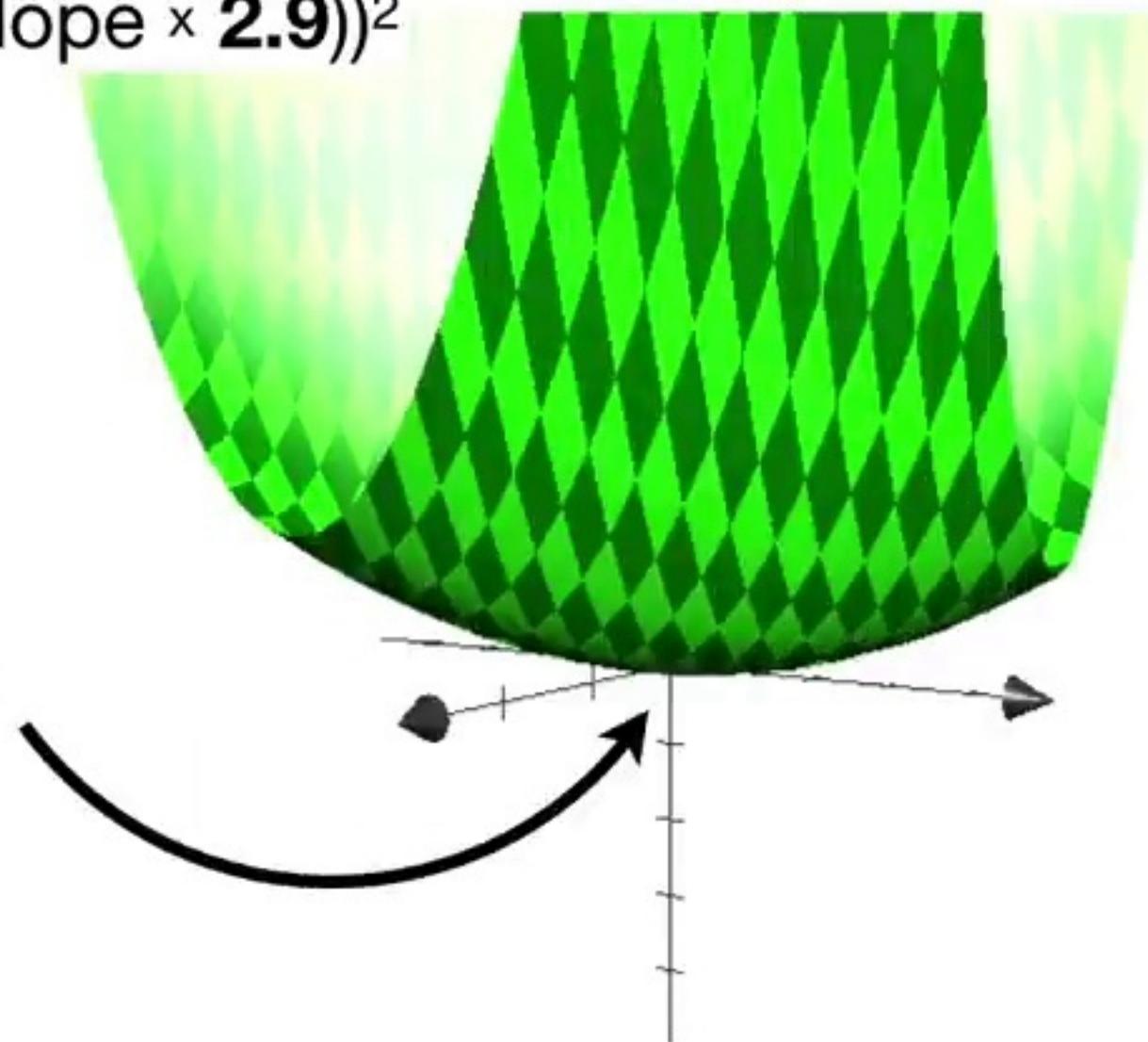
Sum of squared residuals = $(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$
+ $(1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$
+ $(3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$

We want to find the values
for the **Intercept** and **Slope**
that give us the minimum
Sum of the Squared
Residuals.



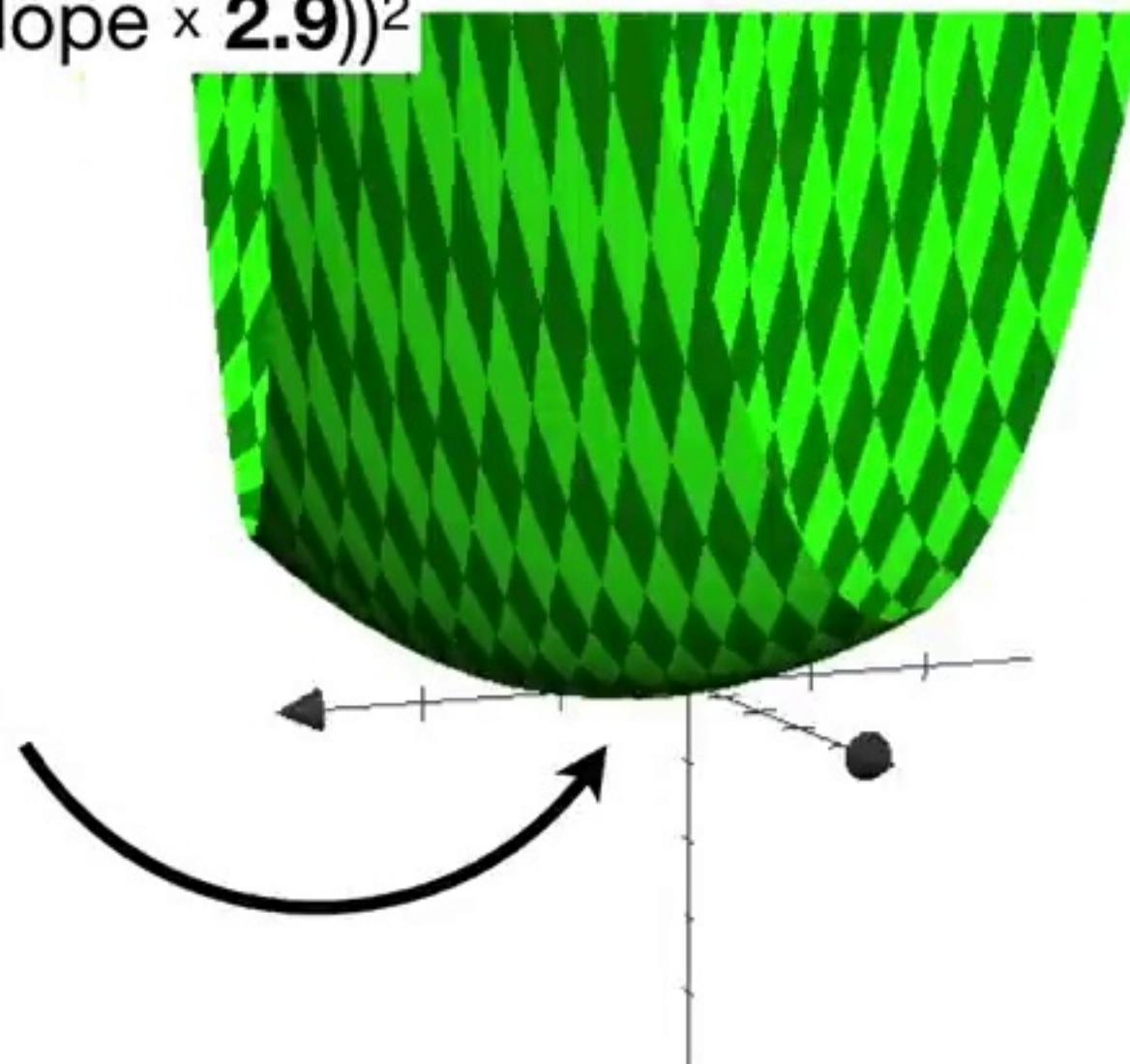
$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

We want to find the values for the **Intercept** and **Slope** that give us the minimum Sum of the Squared Residuals.



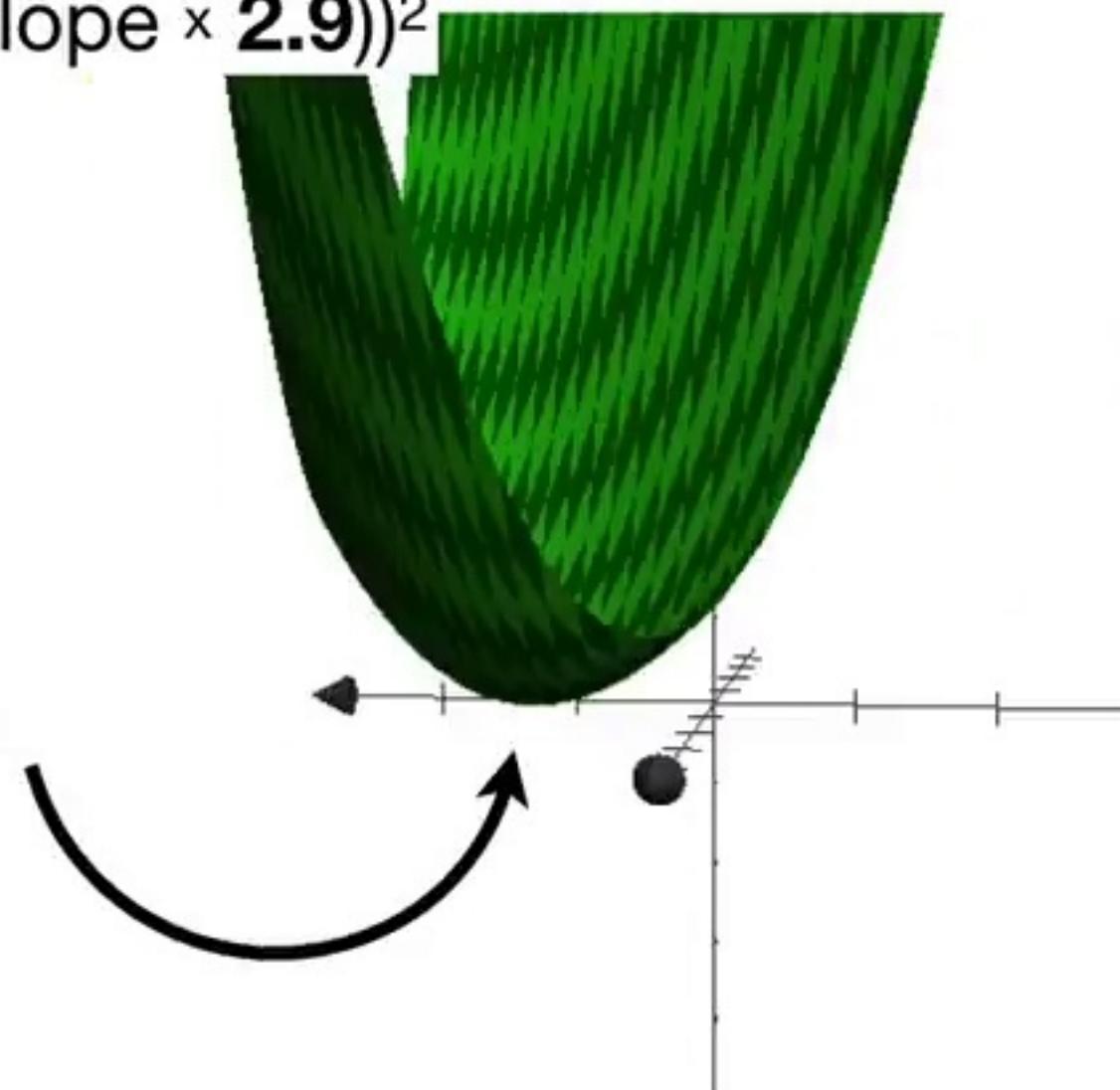
Sum of squared residuals = $(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$
+ $(1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$
+ $(3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$

We want to find the values
for the **Intercept** and **Slope**
that give us the minimum
Sum of the Squared
Residuals.

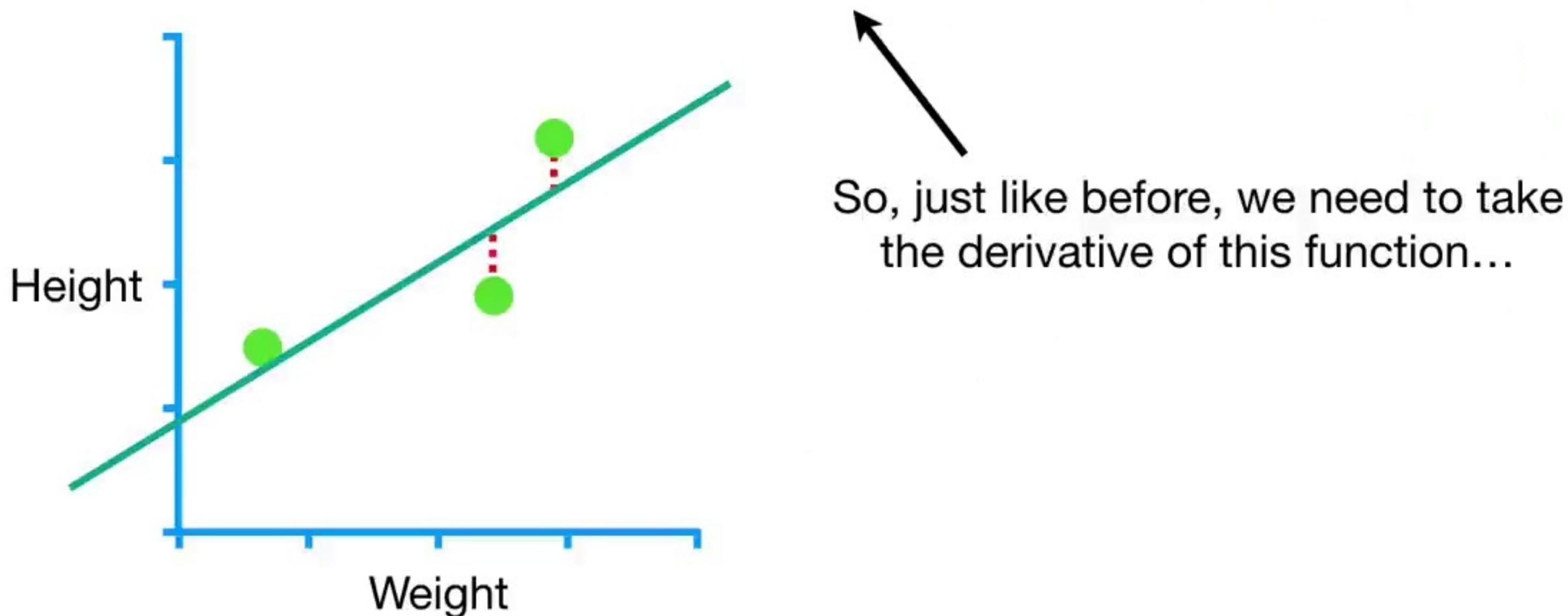


$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

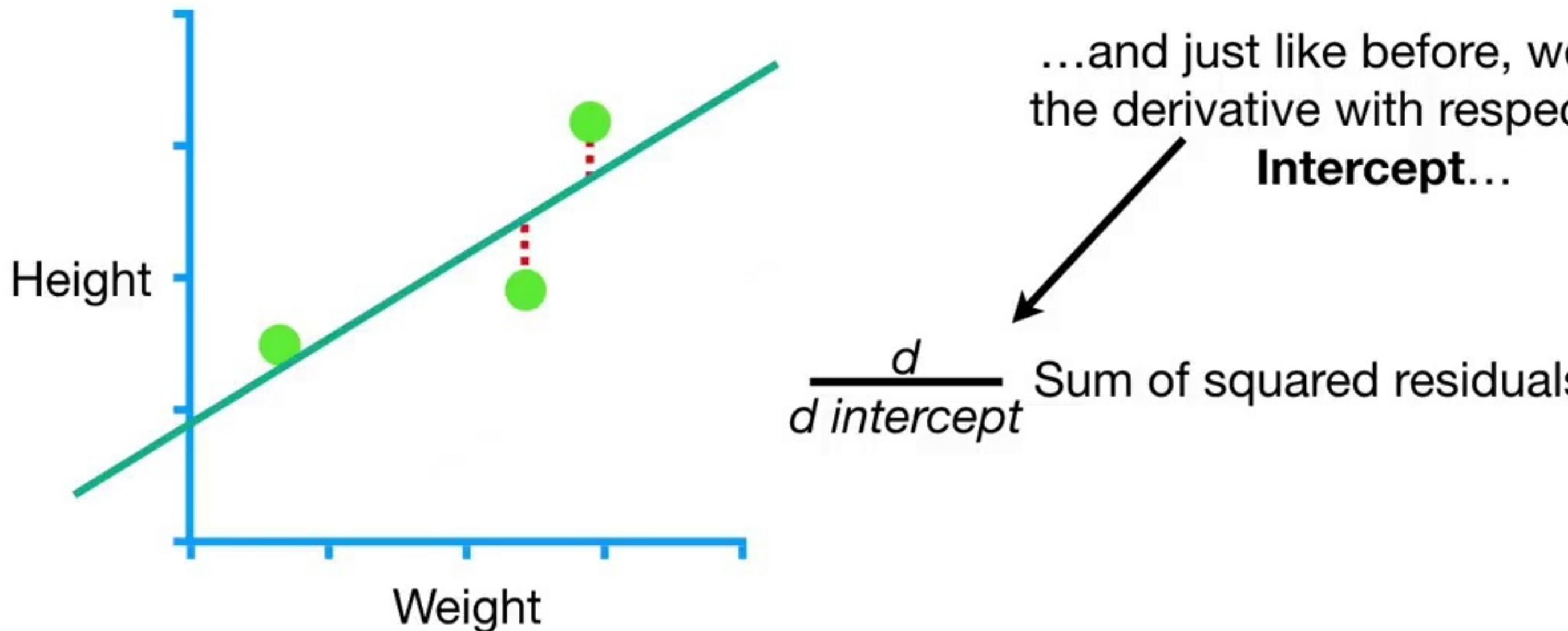
We want to find the values for the **Intercept** and **Slope** that give us the minimum Sum of the Squared Residuals.



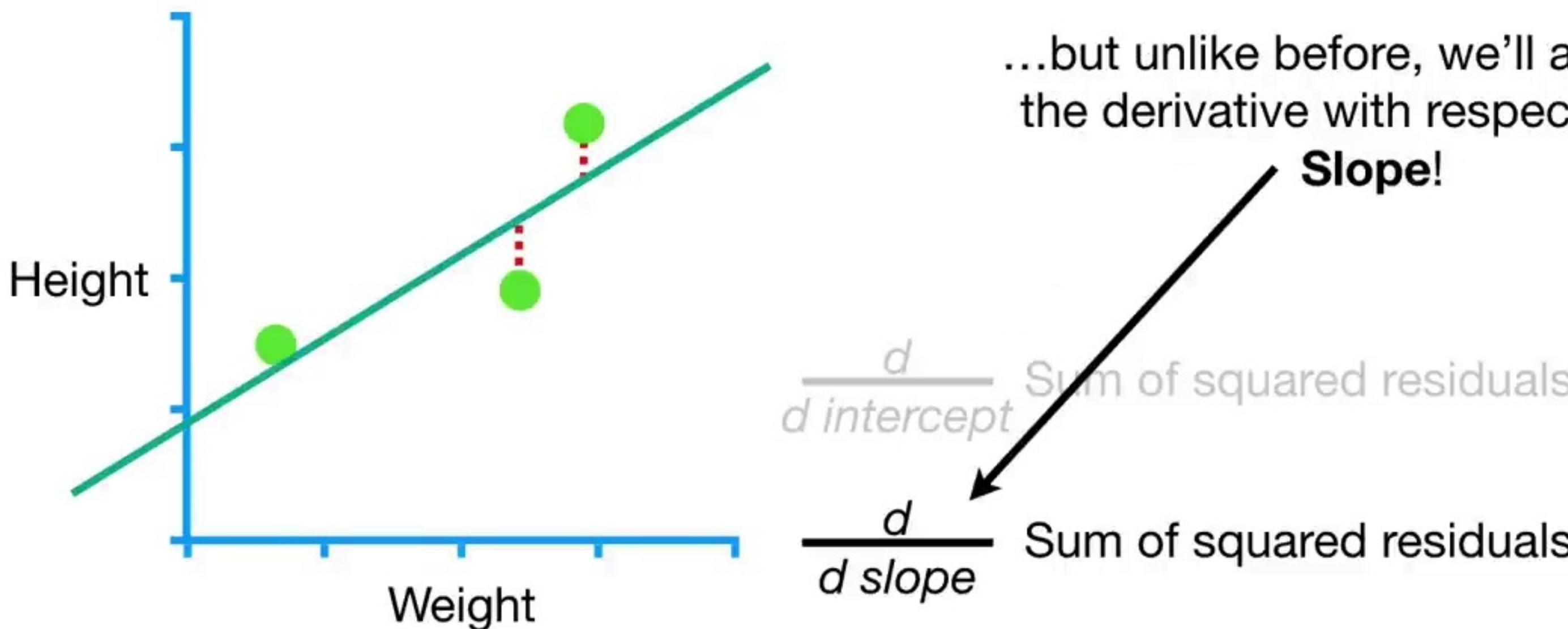
Sum of squared residuals = $(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$
+ $(1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$
+ $(3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$



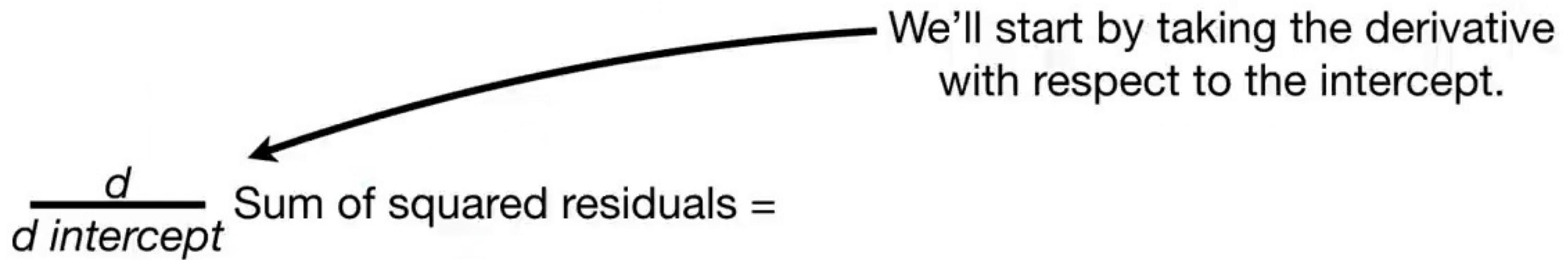
Sum of squared residuals = $(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$
+ $(1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$
+ $(3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$



$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$



$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$



Sum of squared residuals = $(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$

$$+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

Just like before, we take the derivative of each part...

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

Sum of squared residuals = $(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$

$$+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

Just like before, we take the derivative of each part...

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &\quad + (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &\quad + \boxed{(3.2 - (\text{intercept} + \text{slope} \times 2.9))^2}\end{aligned}$$

Just like before, we take the derivative of each part...

$$\begin{aligned}\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} &= \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &\quad + \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &\quad + \boxed{\frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2}\end{aligned}$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

...and just like before,
we'll use...

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = \boxed{\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2}$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$



...and move the square to
the front...

$$\begin{aligned} \frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} &= \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &\quad + \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &\quad + \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2 \end{aligned}$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1$$



...and multiply that by the derivative of the stuff inside the parentheses.

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1$$



$$\frac{d}{d \text{ intercept}} 1.4 - (\text{intercept} + \text{slope} \times 0.5)$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1$$



$$\frac{d}{d \text{ intercept}} 1.4 - (\text{intercept} + \text{slope} \times 0.5)$$



$$\frac{d}{d \text{ intercept}} 1.4 + (-1)\text{intercept} - \text{slope} \times 0.5$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1$$

$$\frac{d}{d \text{ intercept}} 1.4 - (\text{intercept} + \text{slope} \times 0.5)$$

$$\frac{d}{d \text{ intercept}} 1.4 + (-1)\text{intercept} - \boxed{\text{slope} \times 0.5}$$

Since we are taking the derivative with respect to the **Intercept**, we treat the **Slope** like a constant, and the derivative of a constant is **0**.

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1$$

$$\frac{d}{d \text{ intercept}} 1.4 - (\text{intercept} + \text{slope} \times 0.5)$$

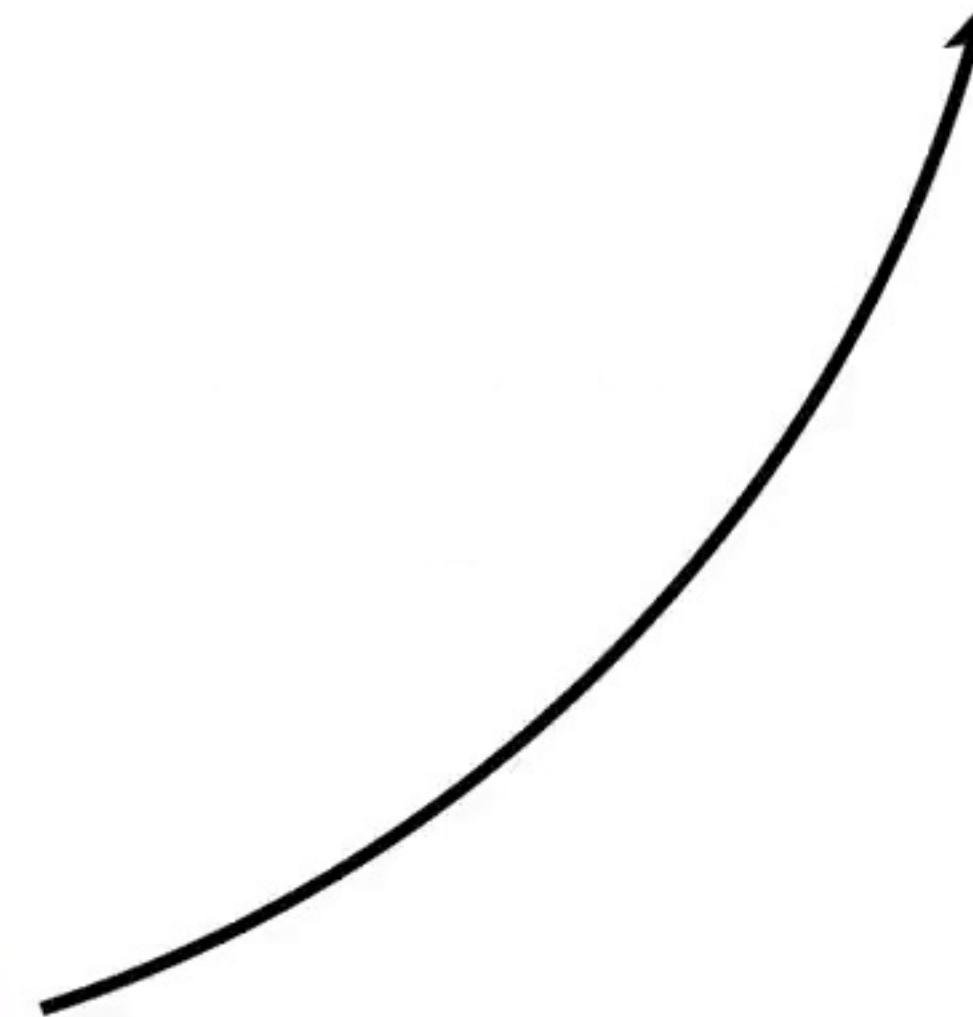
$$\frac{d}{d \text{ intercept}} 1.4 + (-1)\text{intercept} - \text{slope} \times 0.5 = -1$$

So we end up with -1 ,
just like before.

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1$$

$$\frac{d}{d \text{ intercept}} 1.4 - (\text{intercept} + \text{slope} \times 0.5)$$

$$\frac{d}{d \text{ intercept}} 1.4 + (-1)\text{intercept} - \text{slope} \times 0.5 = -1$$



$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = \boxed{2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1}$$

...then we simplify by multiplying 2 by -1...

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1$$

$$= -2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

...then we simplify by multiplying 2 by -1...

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1$$

$$= -2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

...and this...

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1$$

$$= -2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

...and this...

...is the derivative
of the first part...

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = \frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

$$\frac{d}{d \text{ intercept}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -1$$

$$= -2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$



...so we plug it in.

$$\frac{d}{d \text{ intercept}} \text{Sum of squared residuals} = -2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

Likewise, we replace these terms with their derivatives...

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} = -2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$



$$+ \frac{d}{d \text{ intercept}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ intercept}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

Likewise, we replace these terms with their derivatives...

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} = -2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

...and this whole thing is the derivative of
the Sum of the Squared Residuals with
respect to the **Intercept**.



$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} = -2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

Now let's take the derivative of the Sum of the Squared Residuals with respect to the **Slope**.

$$\frac{d}{d \text{ slope}} \text{Sum of squared residuals} =$$


Sum of squared residuals = $(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$

$$+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

Just like before, we take the derivative of each part...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} = \frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

Sum of squared residuals = $(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$

$$+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

Just like before, we take the derivative of each part...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} = \frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

$$+ \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$\text{Sum of squared residuals} = (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

$$+ (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$\boxed{+ (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2}$$

Just like before, we take the derivative of each part...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} = \frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

$$+ \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$\boxed{+ \frac{d}{d \text{ slope}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2}$$

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

...and just like
before, we'll use...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} = \boxed{\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2}$$

$$+ \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ slope}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$



...to move the square to
the front...

$$\begin{aligned}\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} &= \frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &\quad + \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &\quad + \frac{d}{d \text{ slope}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

$$\frac{d}{d \text{ slope}}$$

$$(1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$



...and multiply that by the derivative of the stuff inside the parentheses.

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$



$$\frac{d}{d \text{ slope}} 1.4 - (\text{intercept} + \text{slope} \times 0.5)$$

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$



$$\frac{d}{d \text{ slope}} 1.4 - (\text{intercept} + \text{slope} \times 0.5)$$



$$\frac{d}{d \text{ slope}} 1.4 + (-1)\text{intercept} - \text{slope} \times 0.5$$

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$

$$\frac{d}{d \text{ slope}} 1.4 - (\text{intercept} + \text{slope} \times 0.5)$$

$$\frac{d}{d \text{ slope}} 1.4 + (-1)\text{intercept} - \text{slope} \times 0.5$$

Since we are taking the derivative with respect to the **Slope**, we treat the **Intercept** like a constant, and the derivative of a constant is 0.

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$

$$\frac{d}{d \text{ slope}} 1.4 - (\text{intercept} + \text{slope} \times 0.5)$$

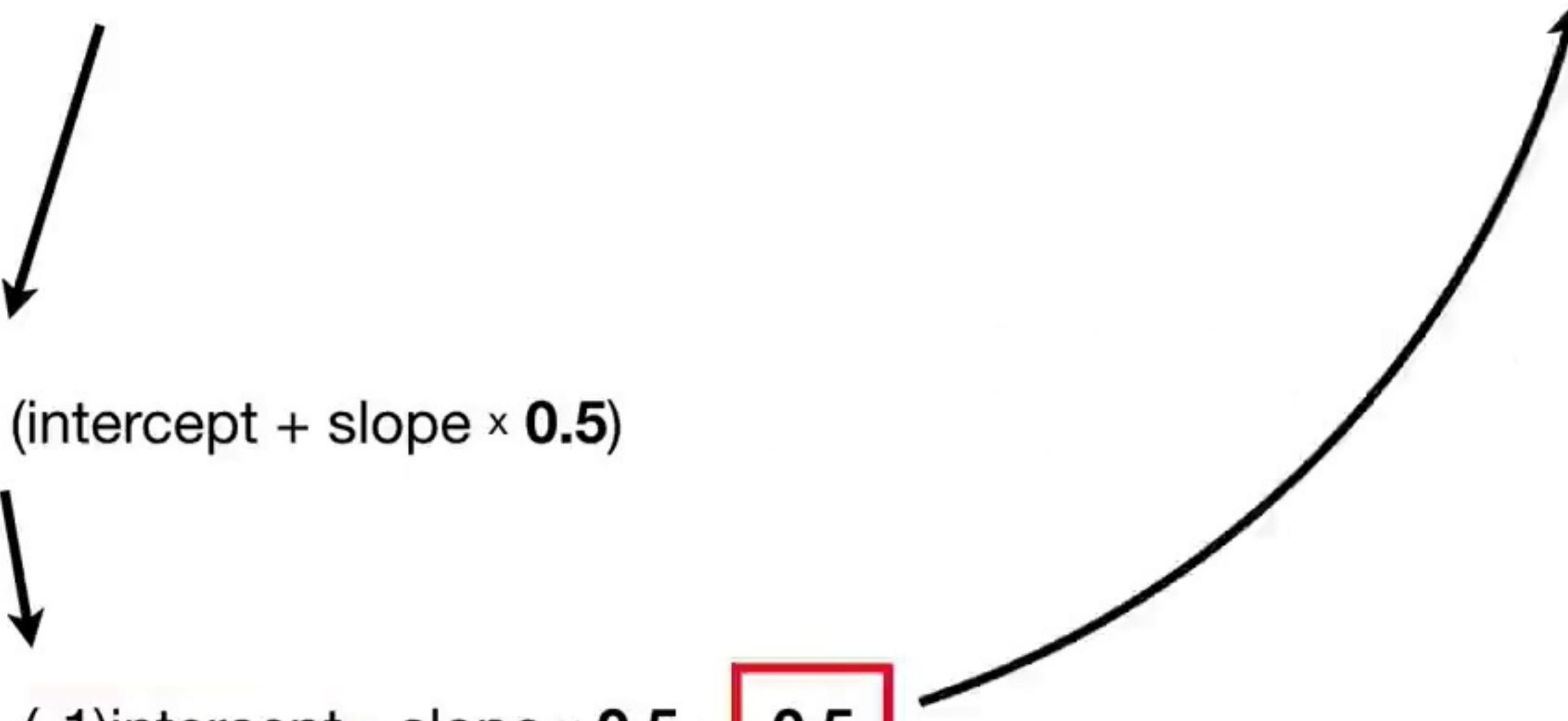
$$\frac{d}{d \text{ slope}} 1.4 + (-1)\text{intercept} - \text{slope} \times 0.5 = -0.5$$

So we end up with **-0.5**.

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$

$$\frac{d}{d \text{ slope}} 1.4 - (\text{intercept} + \text{slope} \times 0.5)$$

$$\frac{d}{d \text{ slope}} 1.4 + (-1)\text{intercept} - \text{slope} \times 0.5 = -0.5$$



$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$

...then we simplify by moving the -0.5 to the front

$$\begin{aligned}\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} &= \frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &\quad + \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &\quad + \frac{d}{d \text{ slope}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$

$$= -2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

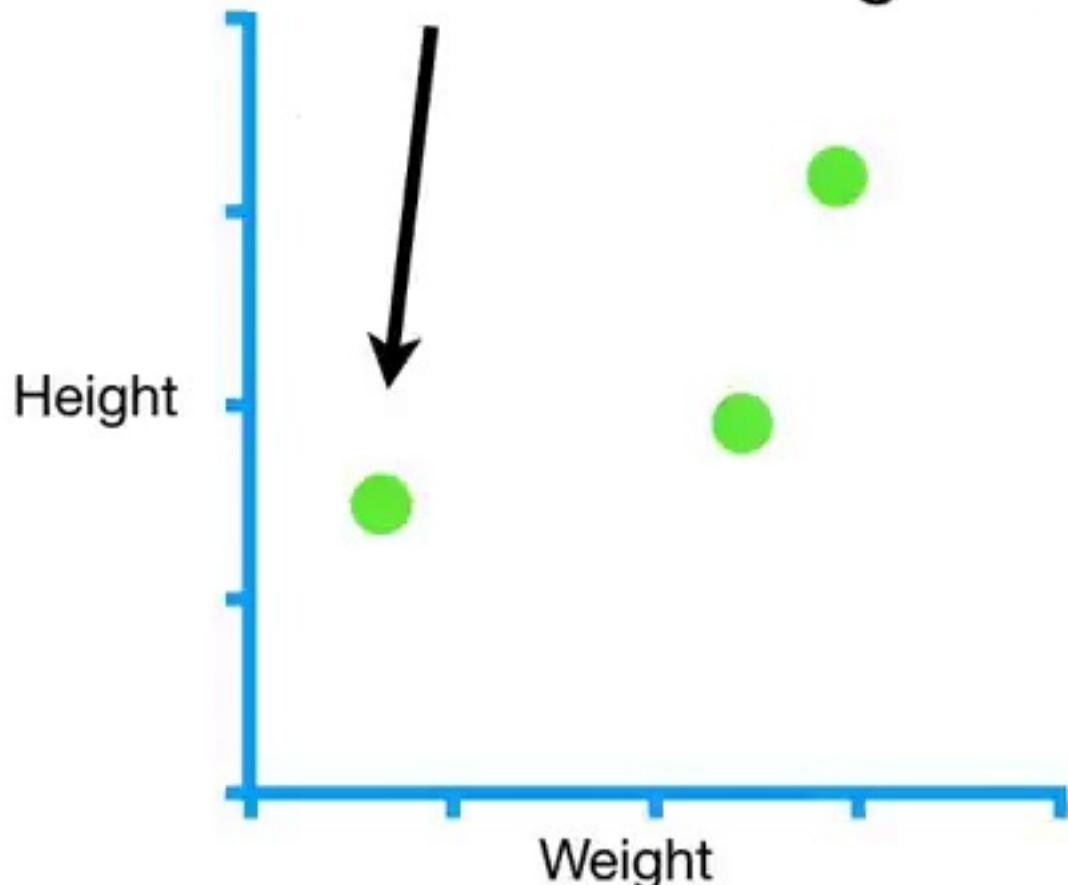
...then we simplify by moving the -0.5 to the front

$$\begin{aligned}\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} &= \frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &\quad + \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &\quad + \frac{d}{d \text{ slope}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$

$$= -2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

NOTE: I left the **0.5** in bold instead of multiplying it by 2 to remind us that **0.5** is the weight for the first sample.



$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$

$$= -2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

...and this...

$$\begin{aligned}\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} &= \frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 \\ &\quad + \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2 \\ &\quad + \frac{d}{d \text{ slope}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2\end{aligned}$$

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$

$$= -2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

...and this...

...is the derivative
of the first part...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} = \frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

$$+ \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ slope}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$

$$= -2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$



...so we plug it in.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} = \frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2$$

$$+ \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ slope}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

$$\frac{d}{d \text{ slope}} (1.4 - (\text{intercept} + \text{slope} \times 0.5))^2 = 2(1.4 - (\text{intercept} + \text{slope} \times 0.5)) \times -0.5$$

$$= -2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$



...so we plug it in.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} = -2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

$$+ \frac{d}{d \text{ slope}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

Likewise, we replace these terms with their derivatives.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} = -2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ \frac{d}{d \text{ slope}} (1.9 - (\text{intercept} + \text{slope} \times 2.3))^2$$

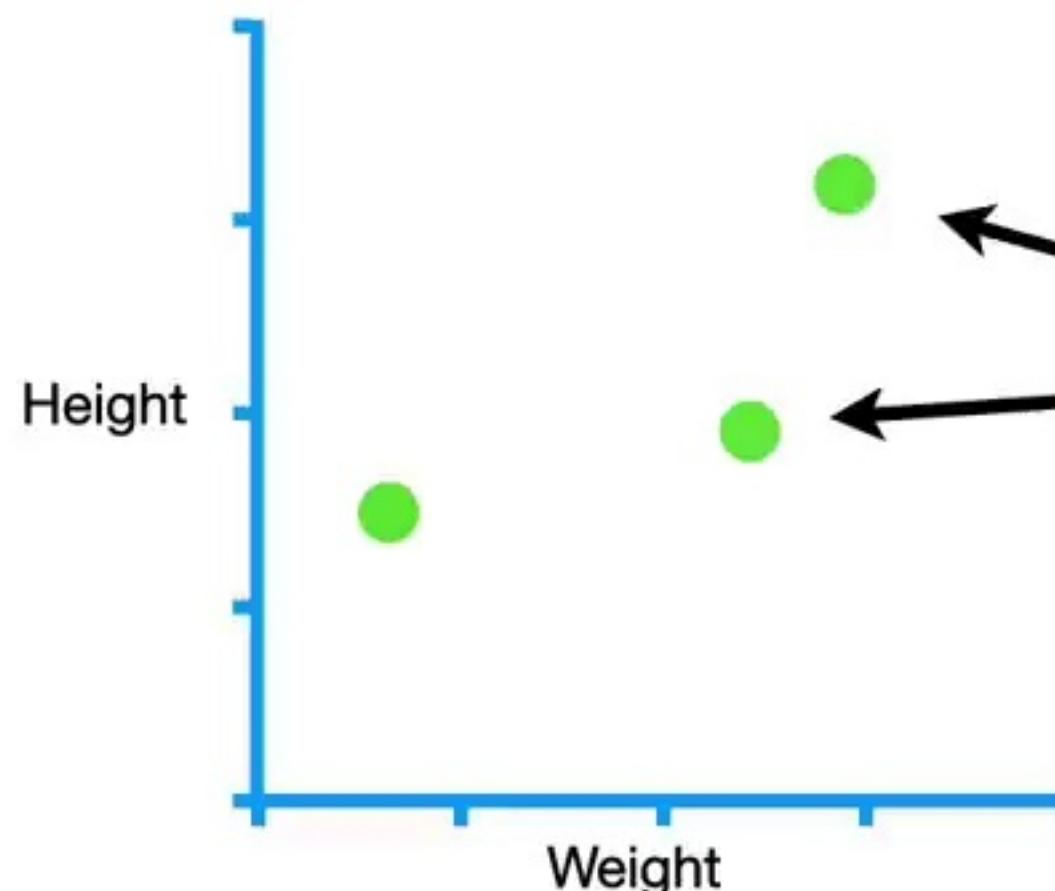
$$+ \frac{d}{d \text{ slope}} (3.2 - (\text{intercept} + \text{slope} \times 2.9))^2$$

Likewise, we replace these terms with their derivatives.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} = -2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ -2 \times 2.3(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$+ -2 \times 2.9(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$



Again, **2.3** and **2.9** are in bold to remind us that they are the weights of the second and third samples.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} = -2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2 \times 2.3(2.3 - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$
$$+ -2 \times 2.9(3.2 - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$



Here's the derivative of the
Sum of the Squared
Residuals with respect to
the **Intercept**...

$$\frac{d}{d \text{ intercept}}$$
 Sum of squared residuals =

$$-2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

Here's the derivative of the
Sum of the Squared
Residuals with respect to
the Intercept...



...and here's the derivative
with respect to the Slope.


$$\frac{d}{d \text{ slope}}$$
 Sum of squared residuals =

$$-2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ -2 \times 2.9(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

$$+ -2 \times 2.3(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

NOTE: When you have two or more derivatives of the same function, they are called a **Gradient**.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

We will use this **Gradient** to descend to lowest point in the **Loss Function**, which, in this case, is the Sum of the Squared Residuals...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

We will use this **Gradient** to descend to lowest point in the **Loss Function**, which, in this case, is the Sum of the Squared Residuals...

...thus, this is why this algorithm is called **Gradient Descent!**

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

Just like before, we will start by picking a random number for the **Intercept**. In this case we'll set the **Intercept = 0...**

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$\frac{d}{d \text{ intercept}}$ Sum of squared residuals =

$$-2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

Just like before, we will start by picking a random number for the **Intercept**. In this case we'll set the **Intercept = 0...**

...and we'll pick a random number for the **Slope**. In this case we'll set the **Slope = 1.**

$\frac{d}{d \text{ slope}}$ Sum of squared residuals =

$$-2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

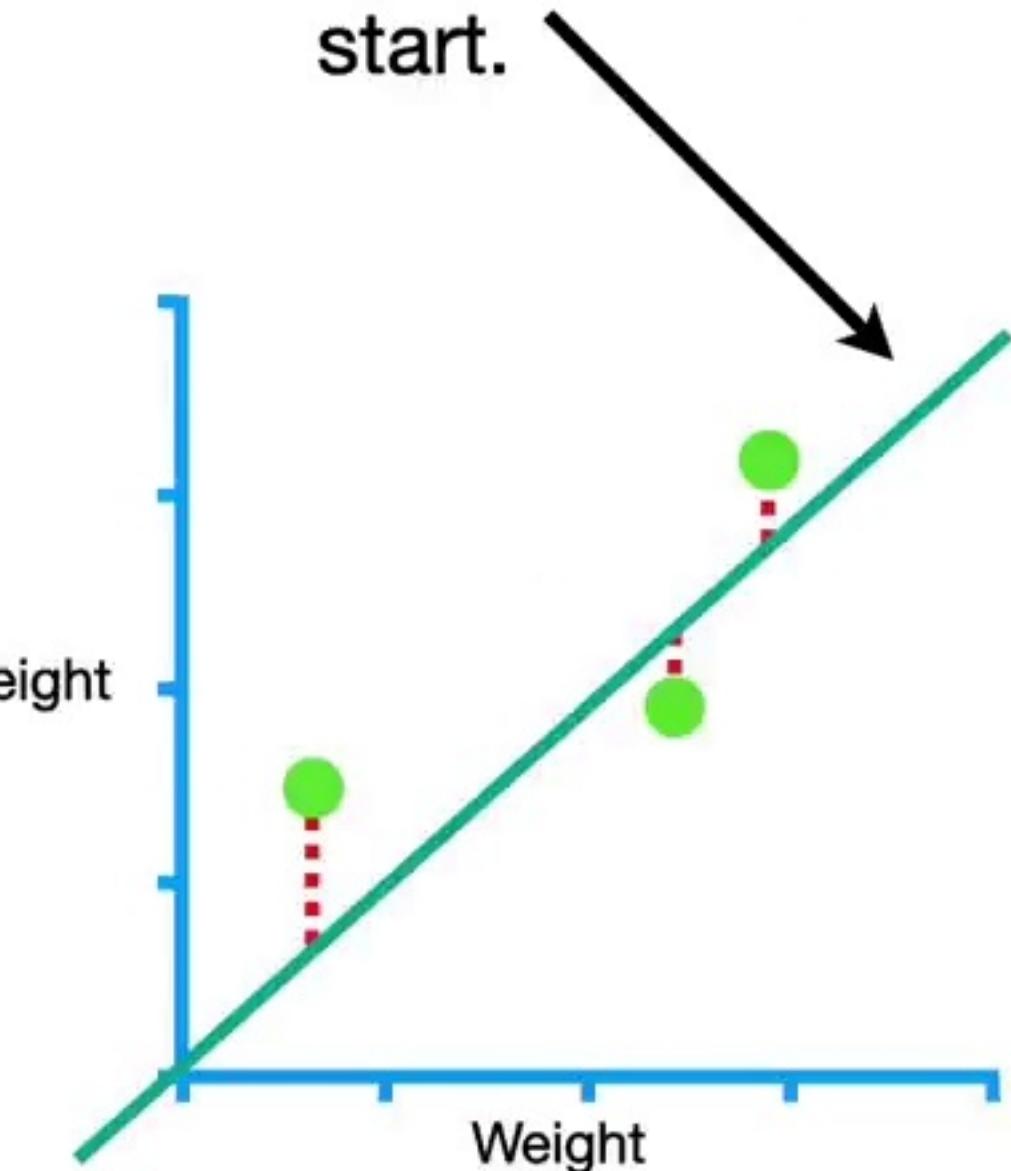
$$+ -2 \times 2.9(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

$$+ -2 \times 2.3(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$
$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

Thus, this line, with **Intercept = 0** and **Slope = 1**, is where we will start.

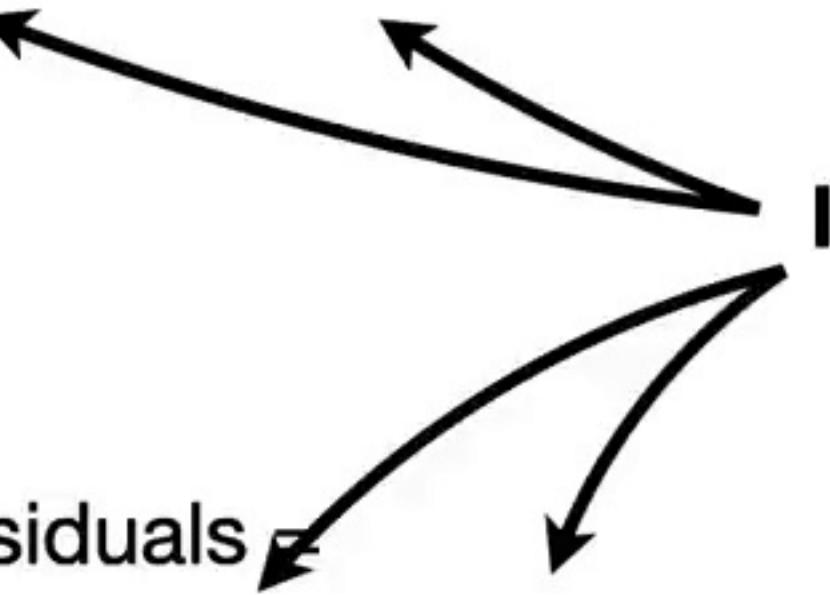


$\frac{d}{d \text{ intercept}}$ Sum of squared residuals =

$$-2(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ -2(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$+ -2(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$



Now let's plug in **0** for the
Intercept and **1** for the **Slope**...

$\frac{d}{d \text{ slope}}$ Sum of squared residuals =

$$-2 \times 0.5(1.4 - (\text{intercept} + \text{slope} \times 0.5))$$

$$+ -2 \times 2.9(3.2 - (\text{intercept} + \text{slope} \times 2.9))$$

$$+ -2 \times 2.3(1.9 - (\text{intercept} + \text{slope} \times 2.3))$$

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (0 + 1 \times 0.5))$$
$$+ -2(1.9 - (0 + 1 \times 2.3))$$
$$+ -2(3.2 - (0 + 1 \times 2.9))$$

Now let's plug in **0** for the
Intercept and **1** for the **Slope**...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3))$$

$\frac{d}{d \text{ intercept}}$ Sum of squared residuals =

$$-2(1.4 - (0 + 1 \times 0.5))$$

$$+ -2(1.9 - (0 + 1 \times 2.3))$$

$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

...and that gives us
two **Slopes**...

$\frac{d}{d \text{ slope}}$ Sum of squared residuals =

$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$

$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$

$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

$\frac{d}{d \text{ intercept}}$ Sum of squared residuals =

$$-2(1.4 - (0 + 1 \times 0.5))$$

$$+ -2(1.9 - (0 + 1 \times 2.3))$$

$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

Step Size_{Intercept} = Slope × Learning Rate



...now we plug the
Slopes into the **Step
Size** formulas...

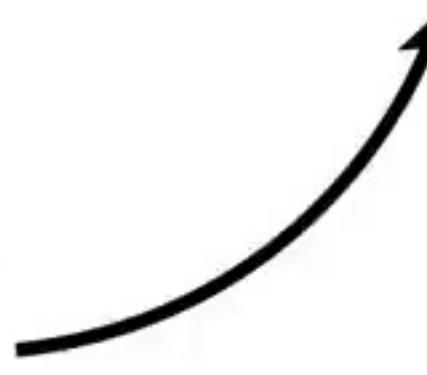
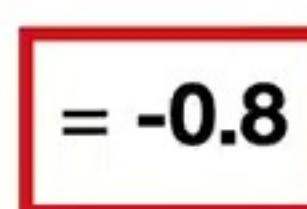
$\frac{d}{d \text{ slope}}$ Sum of squared residuals =

$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$

$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$

$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

Step Size_{Slope} = Slope × Learning Rate



$\frac{d}{d \text{ intercept}}$ Sum of squared residuals =

$$-2(1.4 - (0 + 1 \times 0.5))$$

$$+ -2(1.9 - (0 + 1 \times 2.3))$$

$$+ -2(3.2 - (0 + 1 \times 2.9))$$

Step Size_{Intercept} = $-1.6 \times \text{Learning Rate}$

$$= -1.6$$



...now we plug the
Slopes into the **Step
Size** formulas...

$\frac{d}{d \text{ slope}}$ Sum of squared residuals =

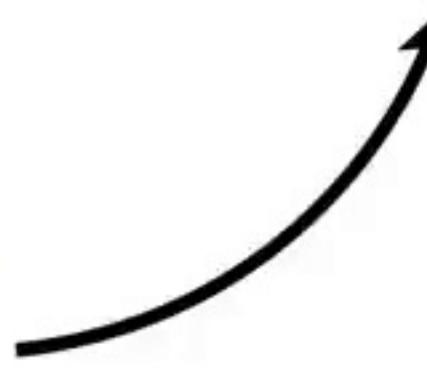
$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$

$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$

$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3))$$

Step Size_{Slope} = $-0.8 \times \text{Learning Rate}$

$$= -0.8$$



$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (0 + 1 \times 0.5))$$
$$+ -2(1.9 - (0 + 1 \times 2.3))$$
$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

Step Size_{Intercept} = -1.6 × Learning Rate



...and multiply by the
Learning Rate, which
this time we set to **0.01**...



$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

Step Size_{Slope} = -0.8 × Learning Rate

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (0 + 1 \times 0.5))$$
$$+ -2(1.9 - (0 + 1 \times 2.3))$$
$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

Step Size_{Intercept} = -1.6×0.01

NOTE: The larger **Learning Rate** that we used in the first example doesn't work this time. Even after a bunch of steps, **Gradient Descent** doesn't arrive at the correct answer.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

Step Size_{Slope} = -0.8×0.01

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (0 + 1 \times 0.5))$$
$$+ -2(1.9 - (0 + 1 \times 2.3))$$
$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

Step Size_{Intercept} = -1.6×0.01

This means that **Gradient Descent** can be very sensitive to the **Learning Rate**.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

Step Size_{Slope} = -0.8×0.01

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (0 + 1 \times 0.5))$$
$$+ -2(1.9 - (0 + 1 \times 2.3))$$
$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

Step Size_{Intercept} = -1.6×0.01

The good news is that in practice, a reasonable **Learning Rate** can be determined automatically by starting large and getting smaller with each step.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

Step Size_{Slope} = -0.8×0.01

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (0 + 1 \times 0.5))$$
$$+ -2(1.9 - (0 + 1 \times 2.3))$$
$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

Step Size_{Intercept} = -1.6×0.01

So, in theory, you shouldn't have to worry too much about the **Learning Rate**.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

Step Size_{Slope} = -0.8×0.01

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (0 + 1 \times 0.5))$$
$$+ -2(1.9 - (0 + 1 \times 2.3))$$
$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

Step Size_{Intercept} = $-1.6 \times 0.01 = -0.016$

Anyway, we do the math
and get two **Step Sizes**.

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

Step Size_{Slope} = $-0.8 \times 0.01 = -0.008$

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$

$$-2(1.4 - (0 + 1 \times 0.5))$$

$$+ -2(1.9 - (0 + 1 \times 2.3))$$

$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

$$\text{Step Size}_{\text{Intercept}} = -1.6 \times 0.01 = -0.016$$

New Intercept = Old Intercept - Step Size

Now we calculate the
New Intercept and **New Slope** by plugging in the
Old Intercept and the
Old Slope...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$

$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$

$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$

$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

$$\text{Step Size}_{\text{Slope}} = -0.8 \times 0.01 = -0.008$$

New Slope = Old Slope - Step Size

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$

$$-2(1.4 - (0 + 1 \times 0.5))$$

$$+ -2(1.9 - (0 + 1 \times 2.3))$$

$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

$$\text{Step Size}_{\text{Intercept}} = -1.6 \times 0.01 = -0.016$$

New Intercept = 0 - Step Size

Now we calculate the
New Intercept and **New Slope** by plugging in the
Old Intercept and the
Old Slope...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$

$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$

$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$

$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

$$\text{Step Size}_{\text{Slope}} = -0.8 \times 0.01 = -0.008$$

New Slope = 1 - Step Size

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (0 + 1 \times 0.5))$$
$$+ -2(1.9 - (0 + 1 \times 2.3))$$
$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

Step Size_{Intercept} = $-1.6 \times 0.01 = -0.016$

New Intercept = $0 - \text{Step Size}$ ←

...and the
Step Sizes...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

Step Size_{Slope} = $-0.8 \times 0.01 = -0.008$

New Slope = $1 - \text{Step Size}$ ←

$$\frac{d}{d \text{ intercept}} \text{ Sum of squared residuals} =$$
$$-2(1.4 - (0 + 1 \times 0.5))$$
$$+ -2(1.9 - (0 + 1 \times 2.3))$$
$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

Step Size_{Intercept} = $-1.6 \times 0.01 = -0.016$

New Intercept = $0 - (-0.016)$ ←

...and the
Step Sizes...

$$\frac{d}{d \text{ slope}} \text{ Sum of squared residuals} =$$
$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$
$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$
$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

Step Size_{Slope} = $-0.8 \times 0.01 = -0.008$

New Slope = $1 - (-0.008)$ ←

$\frac{d}{d \text{ intercept}}$ Sum of squared residuals =

$$-2(1.4 - (0 + 1 \times 0.5))$$

$$+ -2(1.9 - (0 + 1 \times 2.3))$$

$$+ -2(3.2 - (0 + 1 \times 2.9)) = -1.6$$

Step Size_{Intercept} = $-1.6 \times 0.01 = -0.016$

New Intercept = $0 - (-0.016) = 0.016$



...and we end up
with a **New Intercept**
and a **New Slope**.

$\frac{d}{d \text{ slope}}$ Sum of squared residuals =

$$-2 \times 0.5(1.4 - (0 + 1 \times 0.5))$$

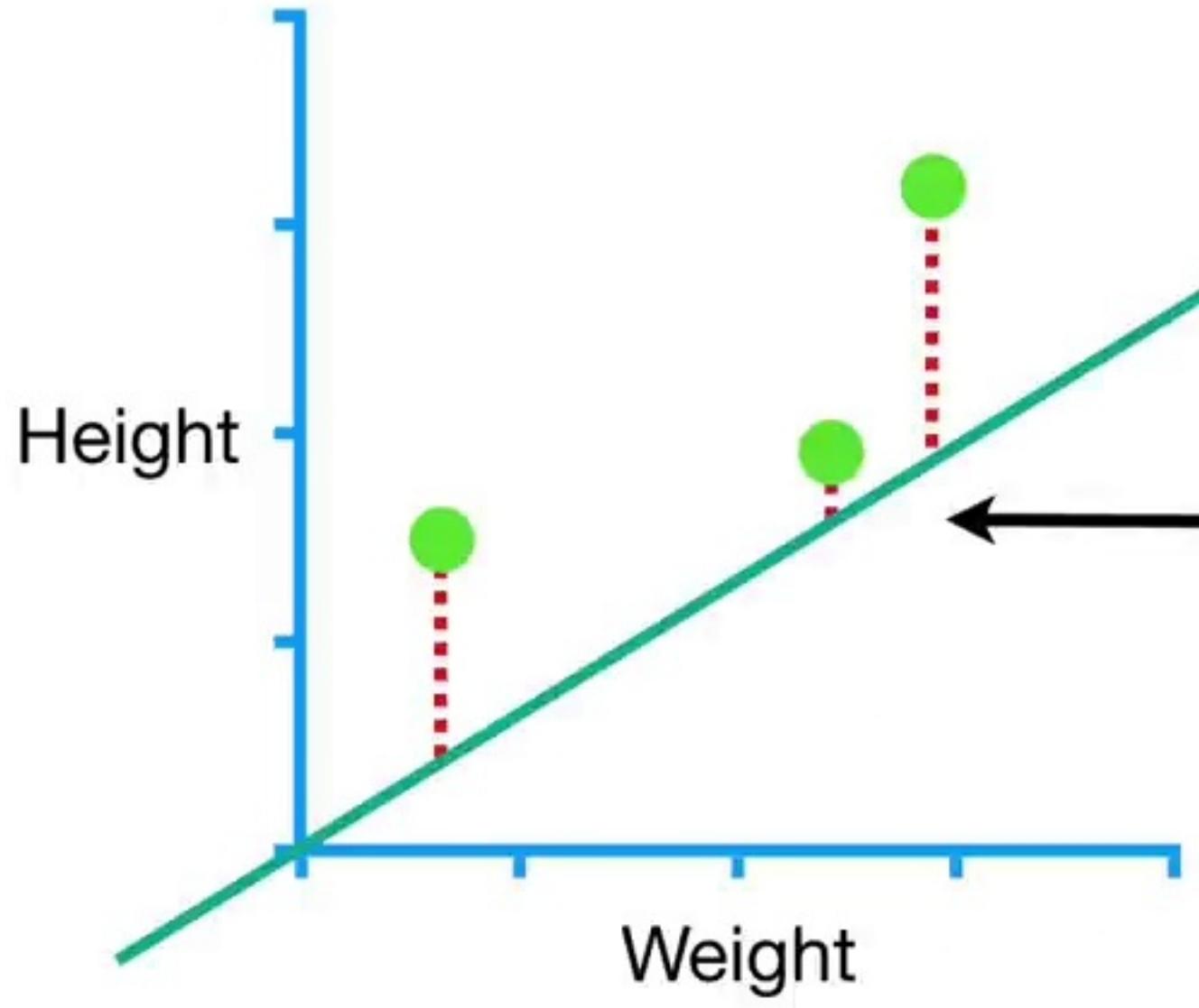
$$+ -2 \times 2.9(3.2 - (0 + 1 \times 2.9))$$

$$+ -2 \times 2.3(1.9 - (0 + 1 \times 2.3)) = -0.8$$

Step Size_{Slope} = $-0.8 \times 0.01 = -0.008$

New Slope = $1 - (-0.008) = 1.008$

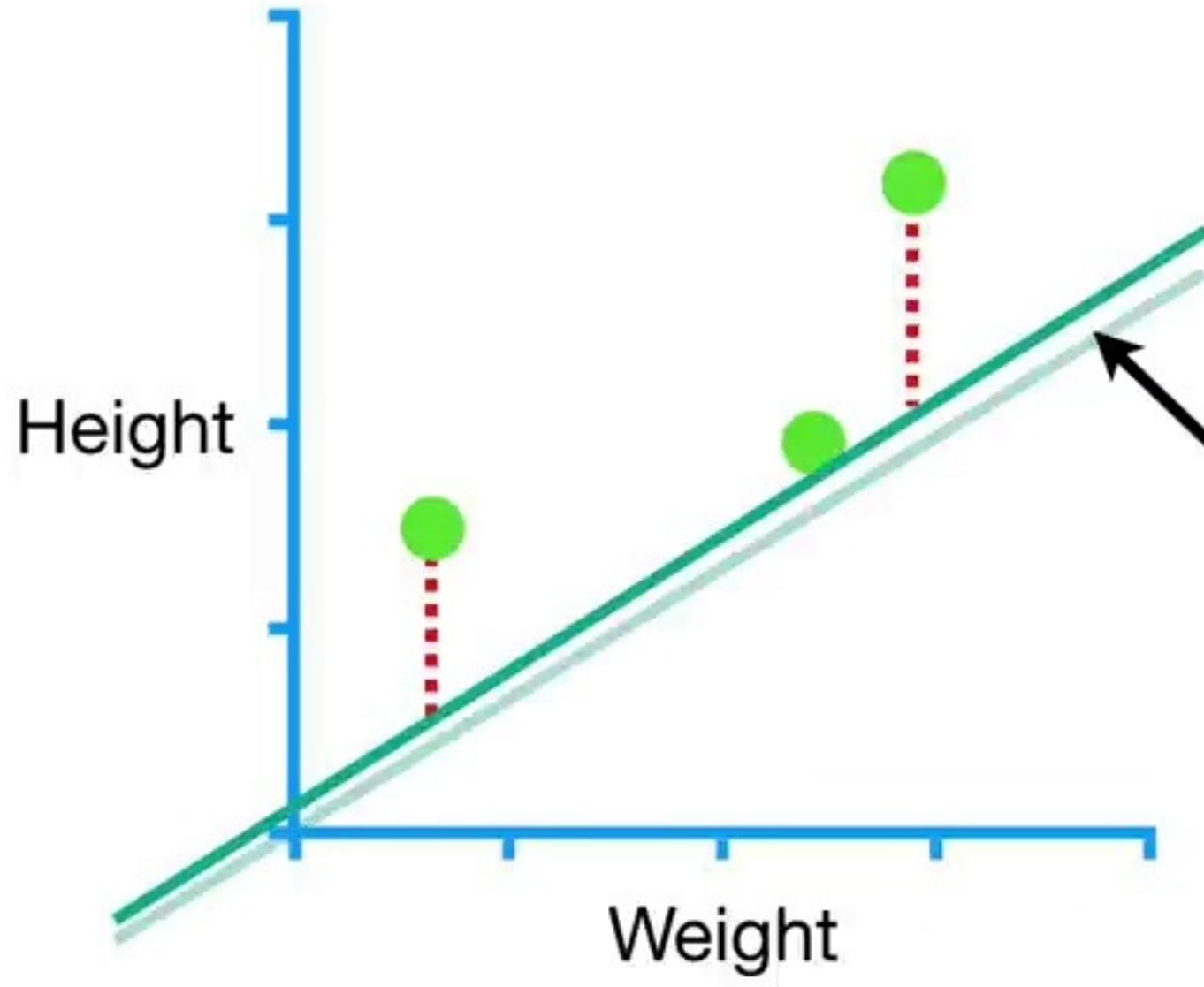




$$\text{New Intercept} = 0 - (-0.016) = 0.016$$

This is the line we
started with...
**(Slope = 1 and
Intercept = 0)**

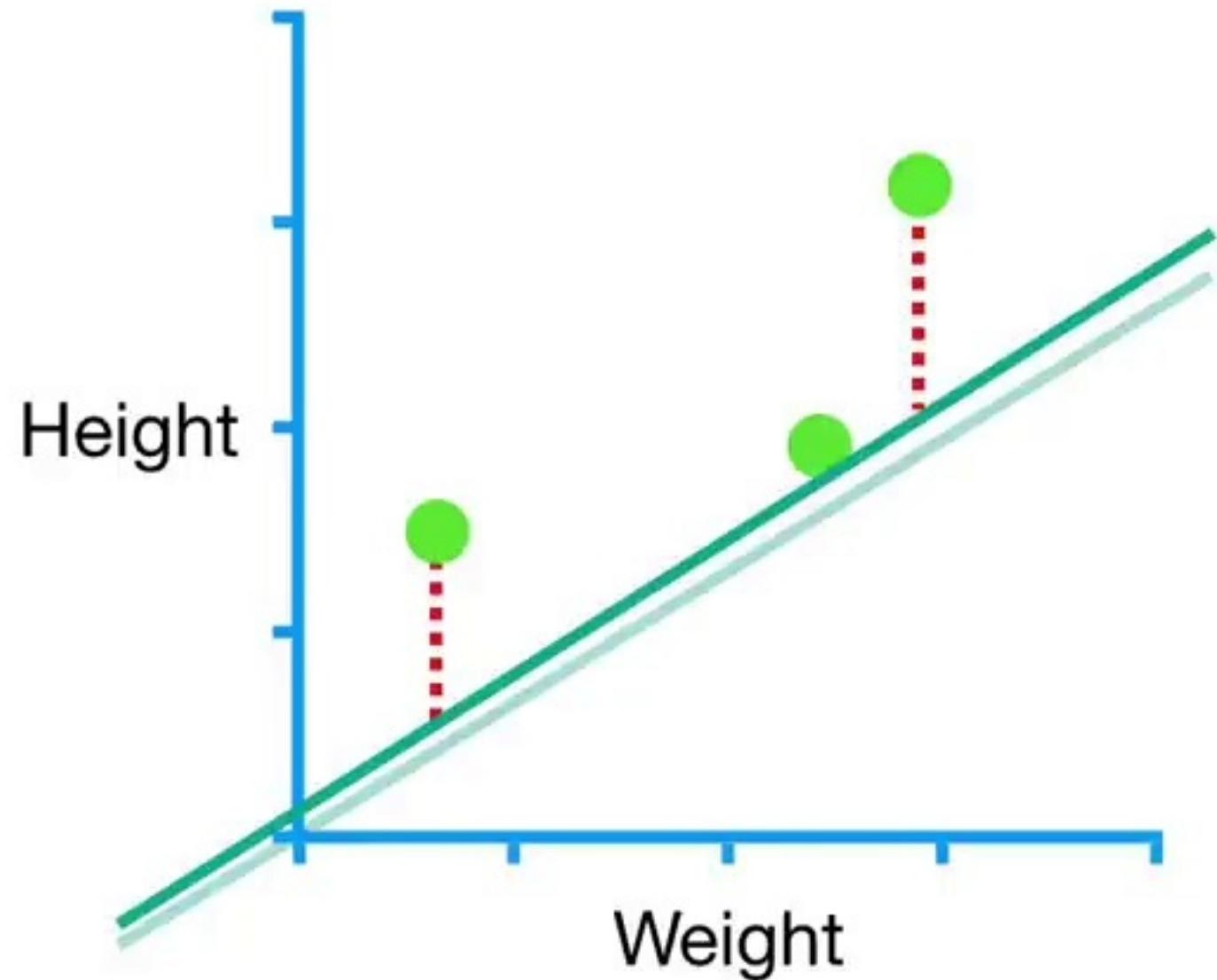
$$\text{New Slope} = 1 - (-0.008) = 1.008$$



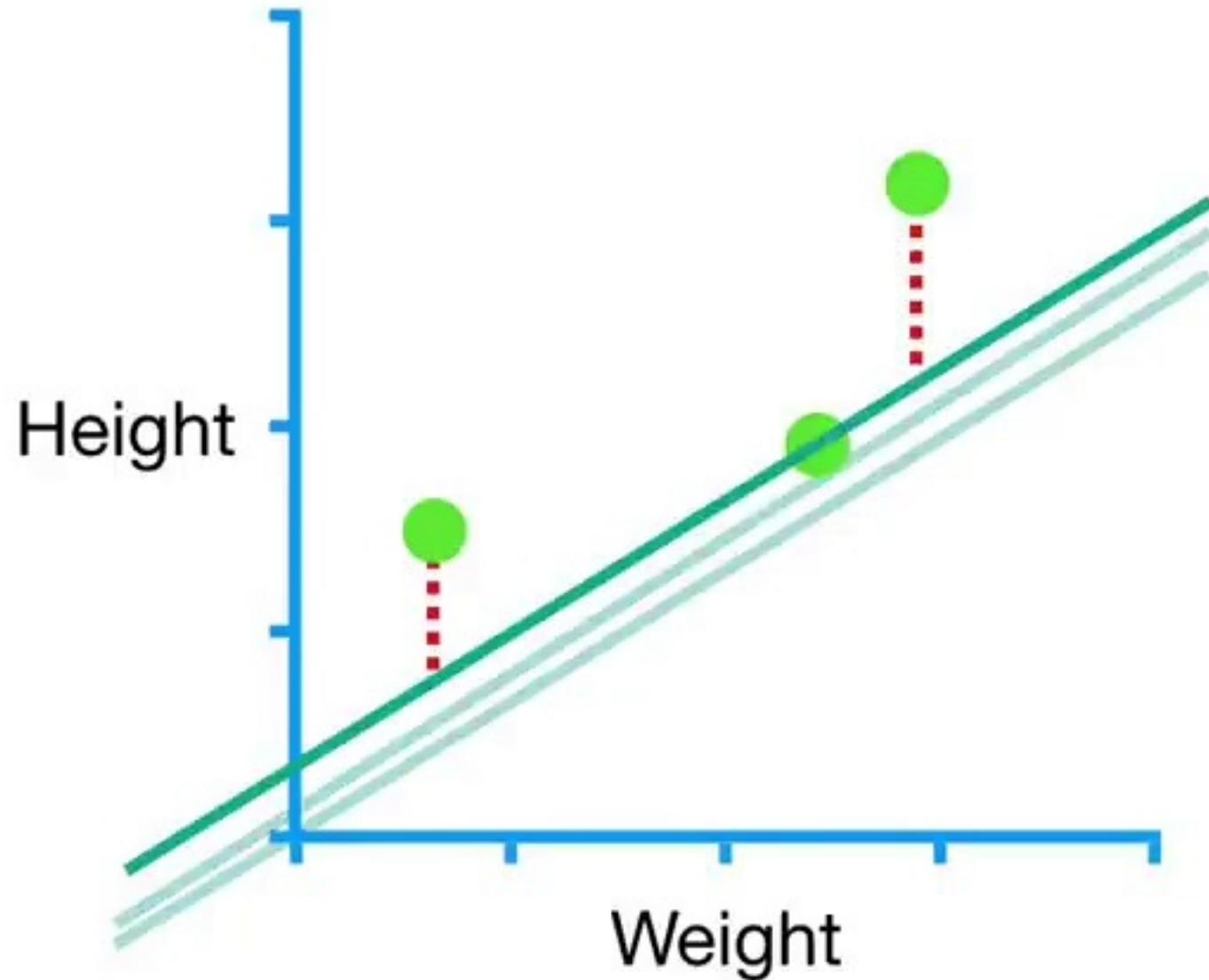
$$\text{New Intercept} = 0 - (-0.016) = 0.016$$

...and this is the new line
(with **Slope = 1.008** and
Intercept = 0.016) after
the first step.

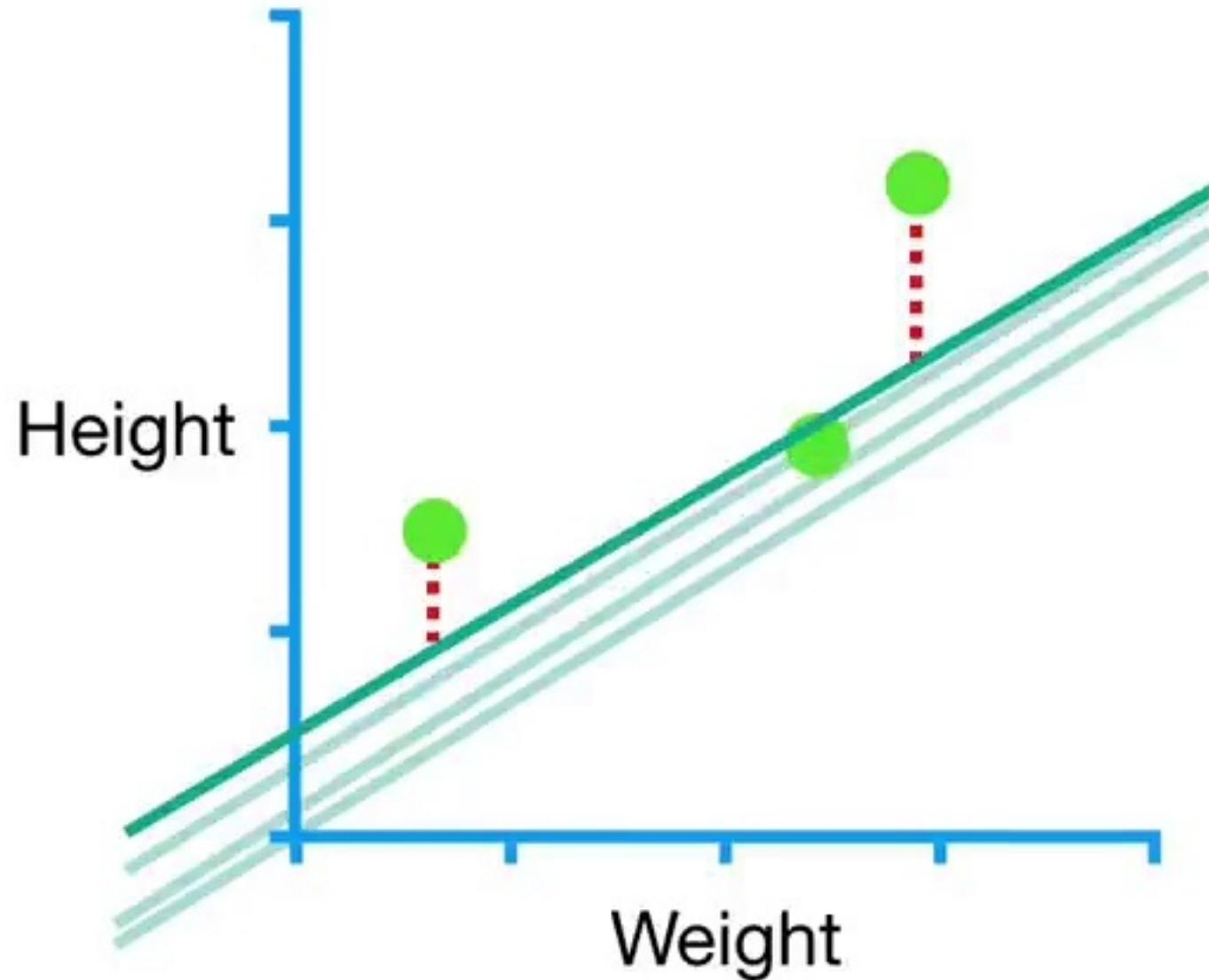
$$\text{New Slope} = 1 - (-0.008) = 1.008$$



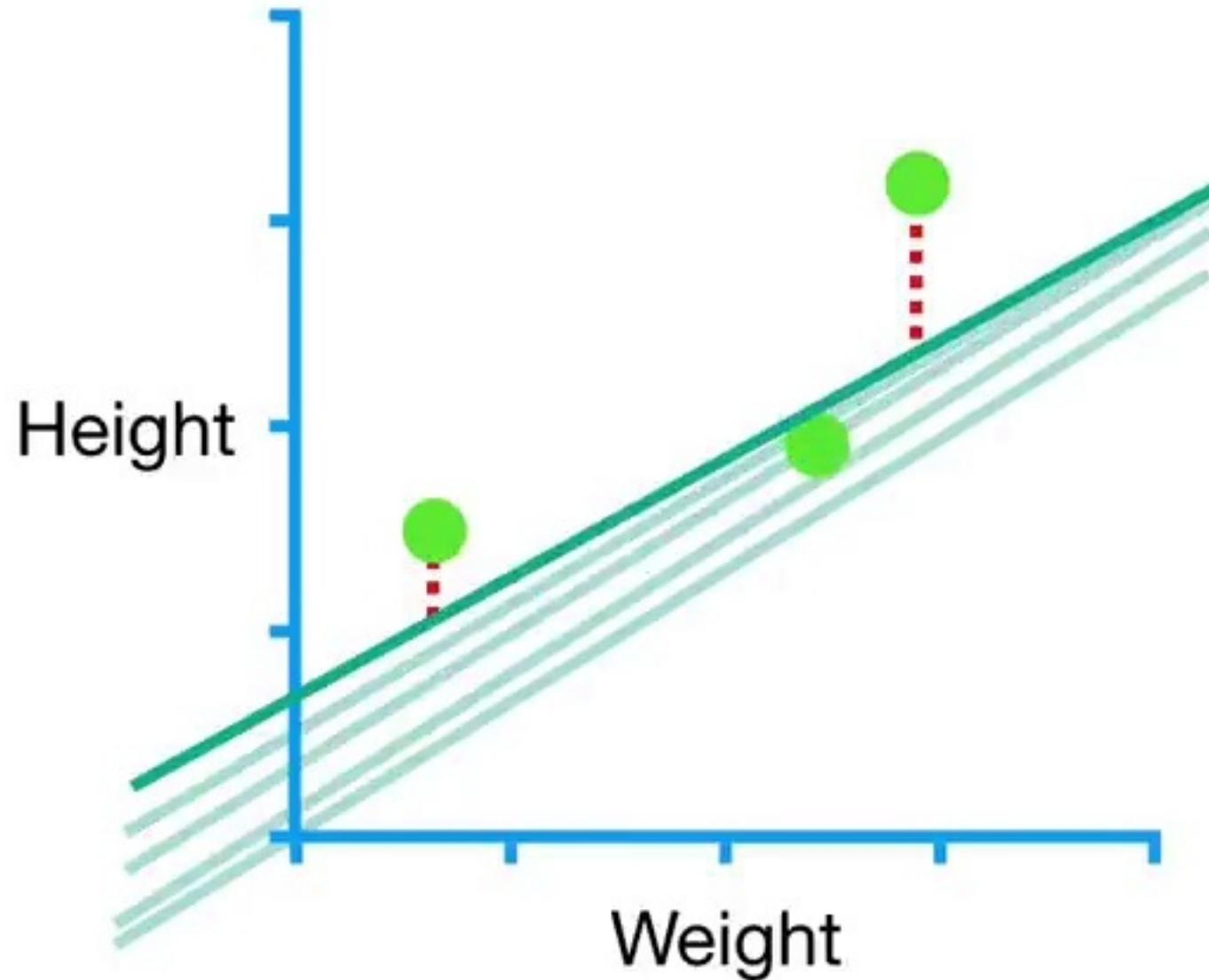
Now we just repeat what we did until all of the **Steps Sizes** are very small or we reach the **Maximum Number of Steps**.



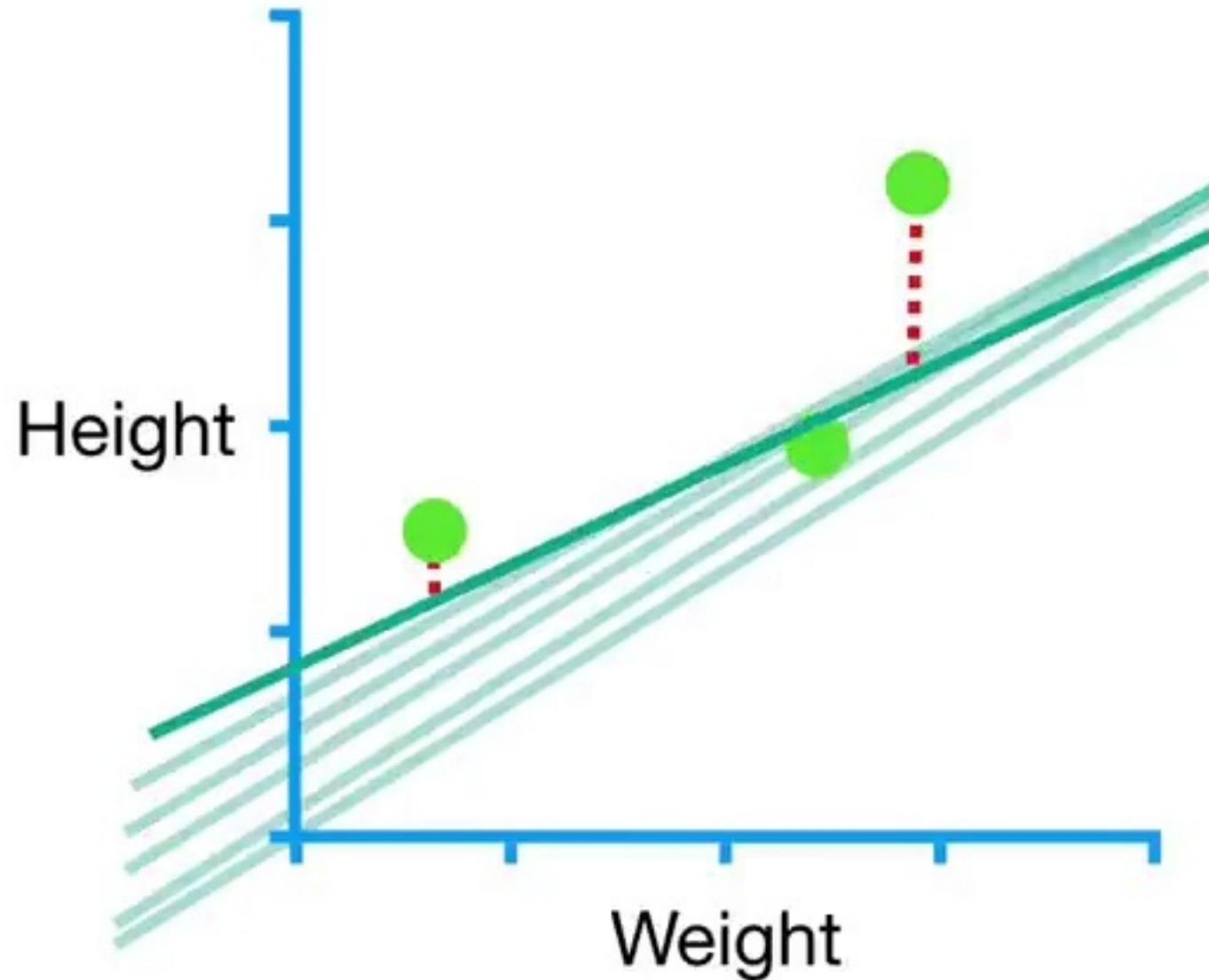
Now we just repeat what we did until all of the **Steps Sizes** are very small or we reach the **Maximum Number of Steps**.



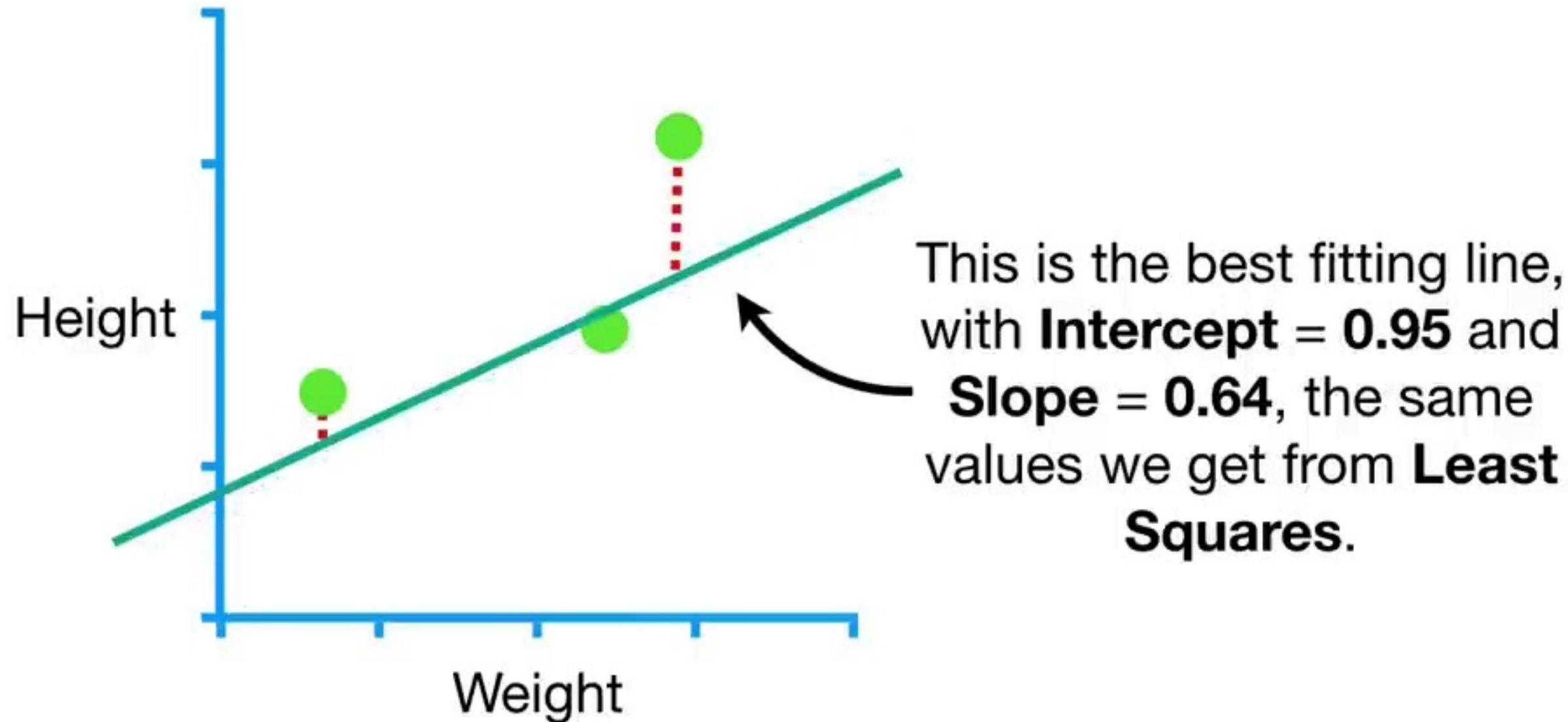
Now we just repeat what we did until all of the **Steps Sizes** are very small or we reach the **Maximum Number of Steps**.

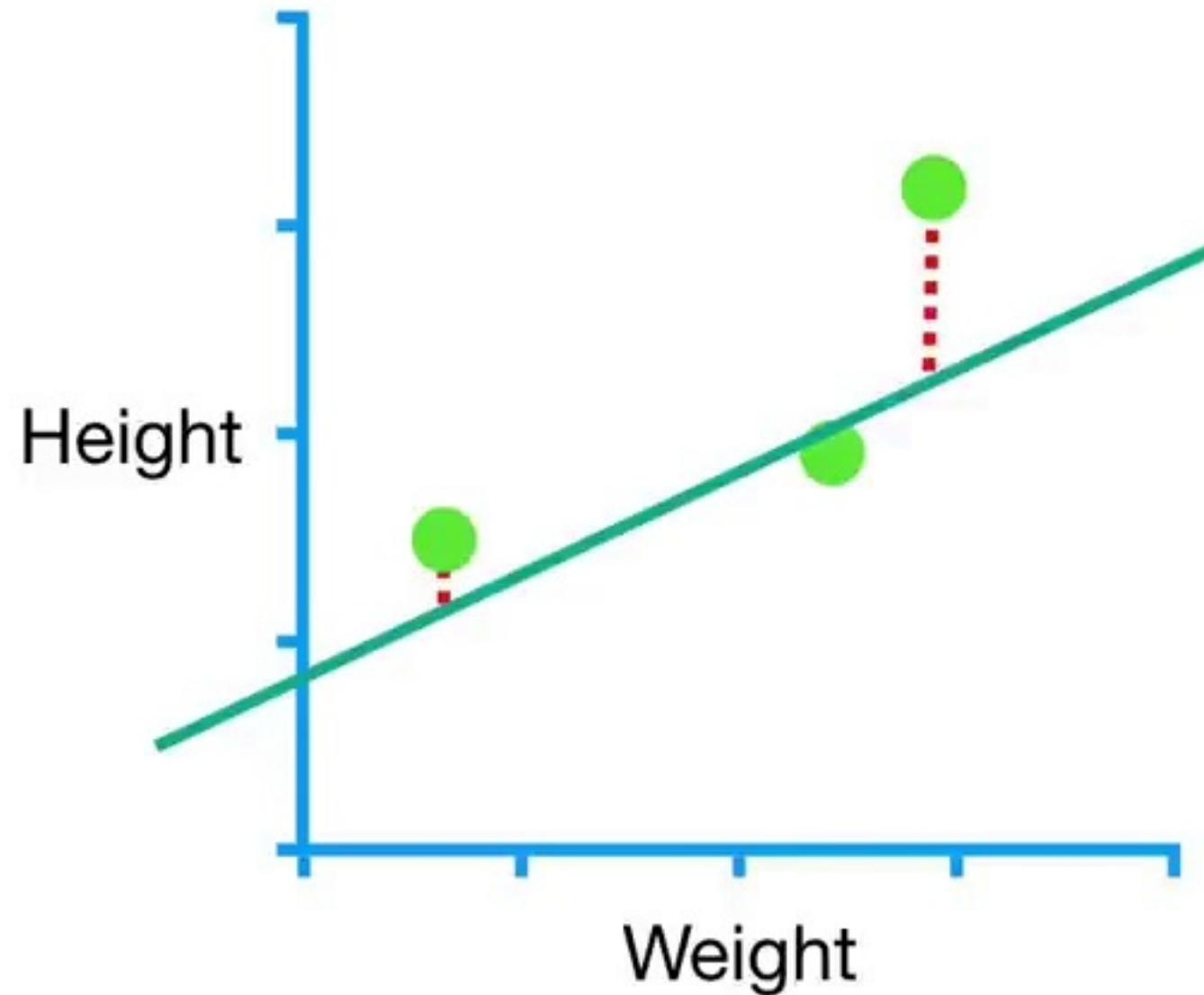


Now we just repeat what we did until all of the **Steps Sizes** are very small or we reach the **Maximum Number of Steps**.

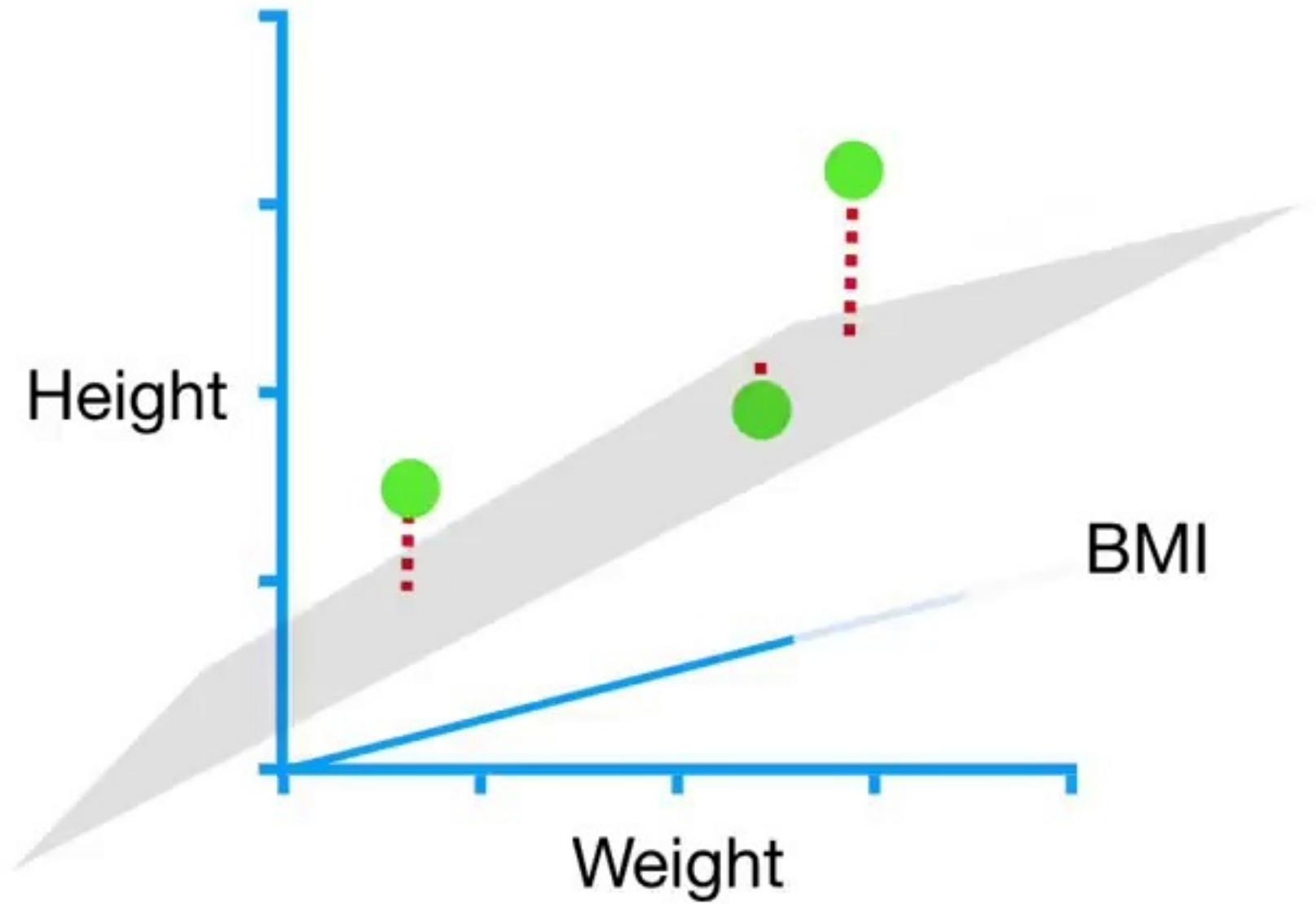


Now we just repeat what we did until all of the **Steps Sizes** are very small or we reach the **Maximum Number of Steps**.





We now know how **Gradient Descent** optimizes two parameters, the **Slope** and **Intercept**.



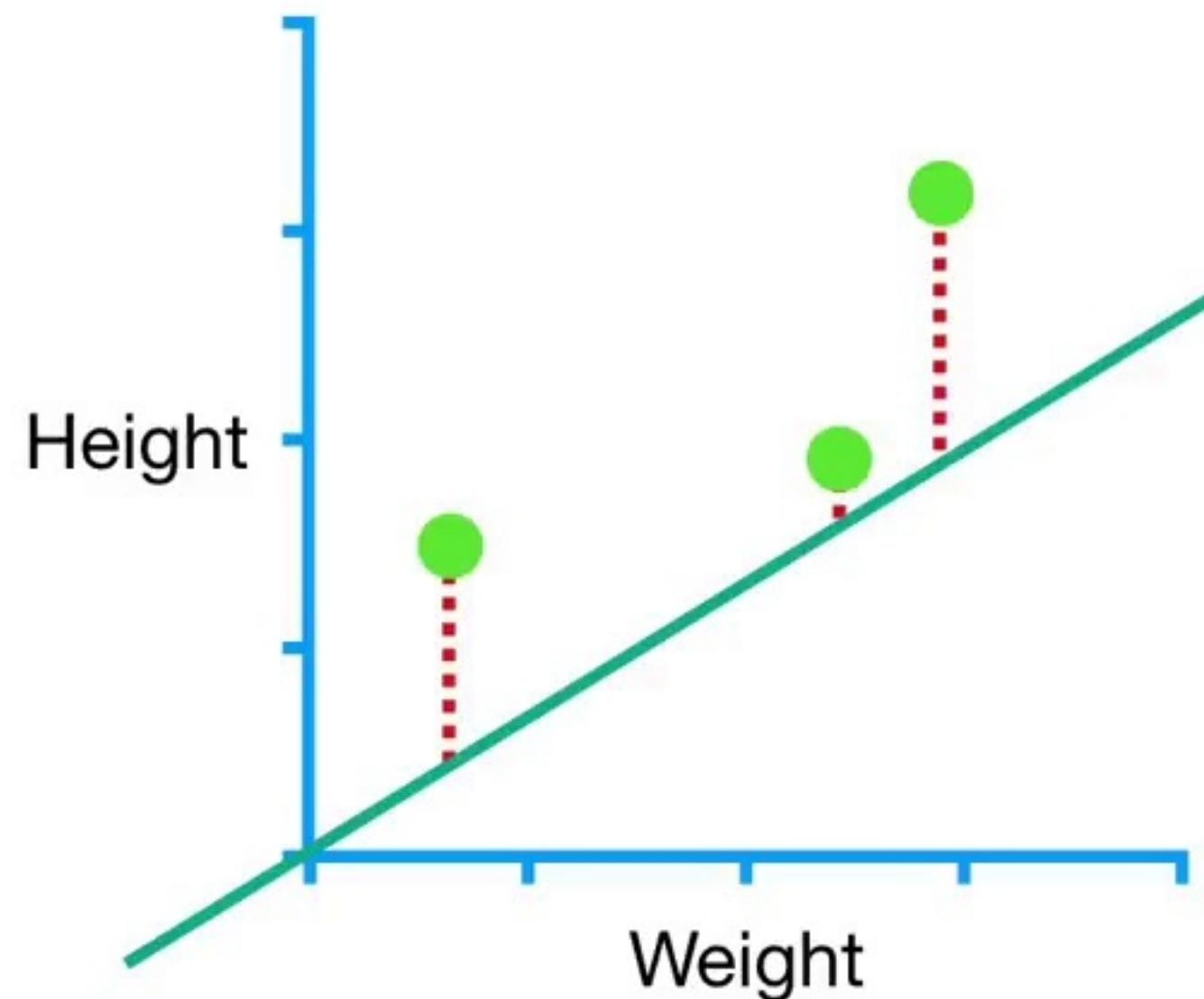
If we had more parameters,
then we'd just take more
derivatives and everything else
stays the same.

Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$

$$+ (1.9 - (\text{intercept} + 0.64 \times 2.3))^2$$

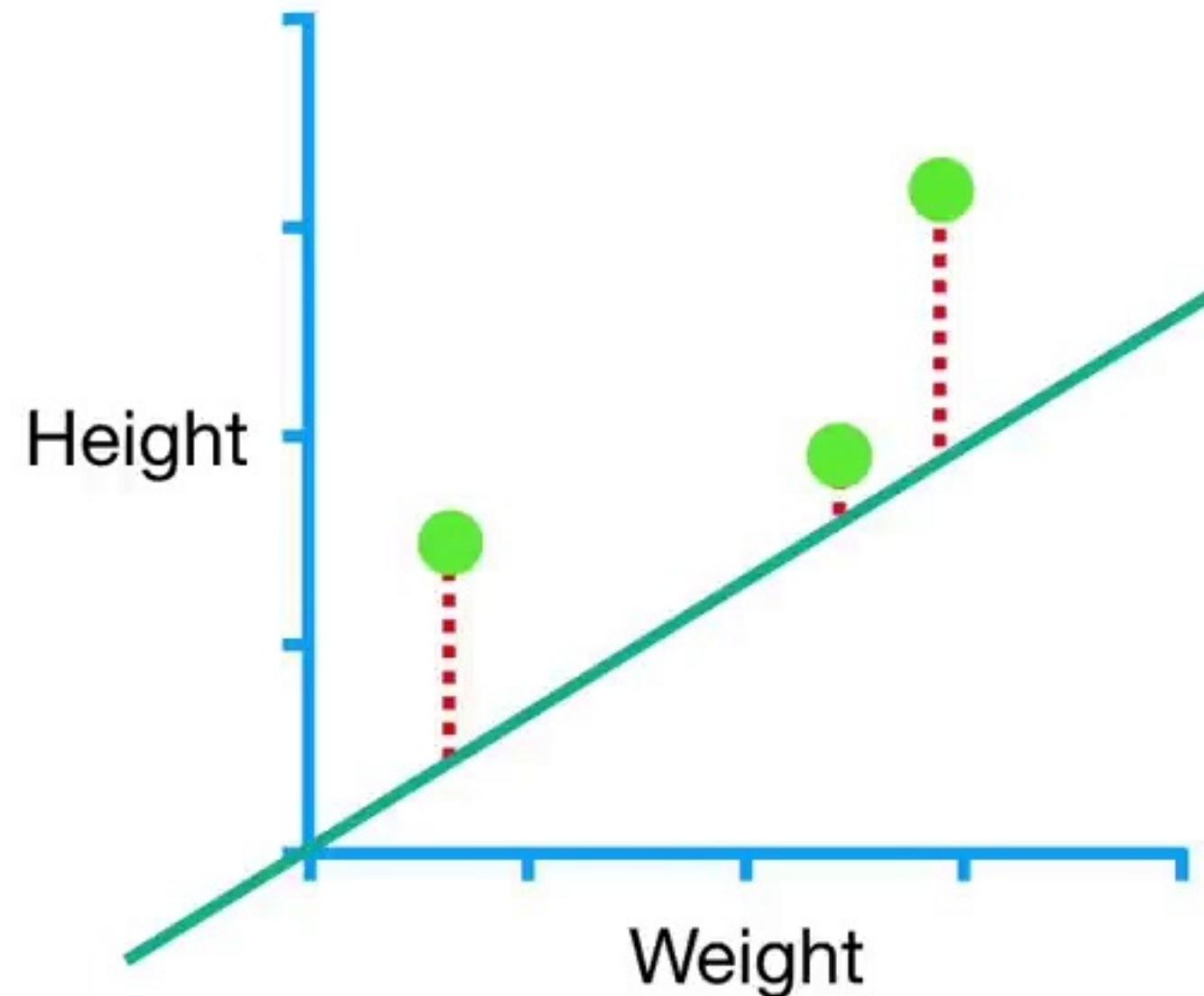
$$+ (3.2 - (\text{intercept} + 0.64 \times 2.9))^2$$

NOTE: The Sum of the Squared Residuals is just one type of **Loss Function**.



$$\begin{aligned}\text{Sum of squared residuals} &= (1.4 - (\text{intercept} + 0.64 \times 0.5))^2 \\ &\quad + (1.9 - (\text{intercept} + 0.64 \times 2.3))^2 \\ &\quad + (3.2 - (\text{intercept} + 0.64 \times 2.9))^2\end{aligned}$$

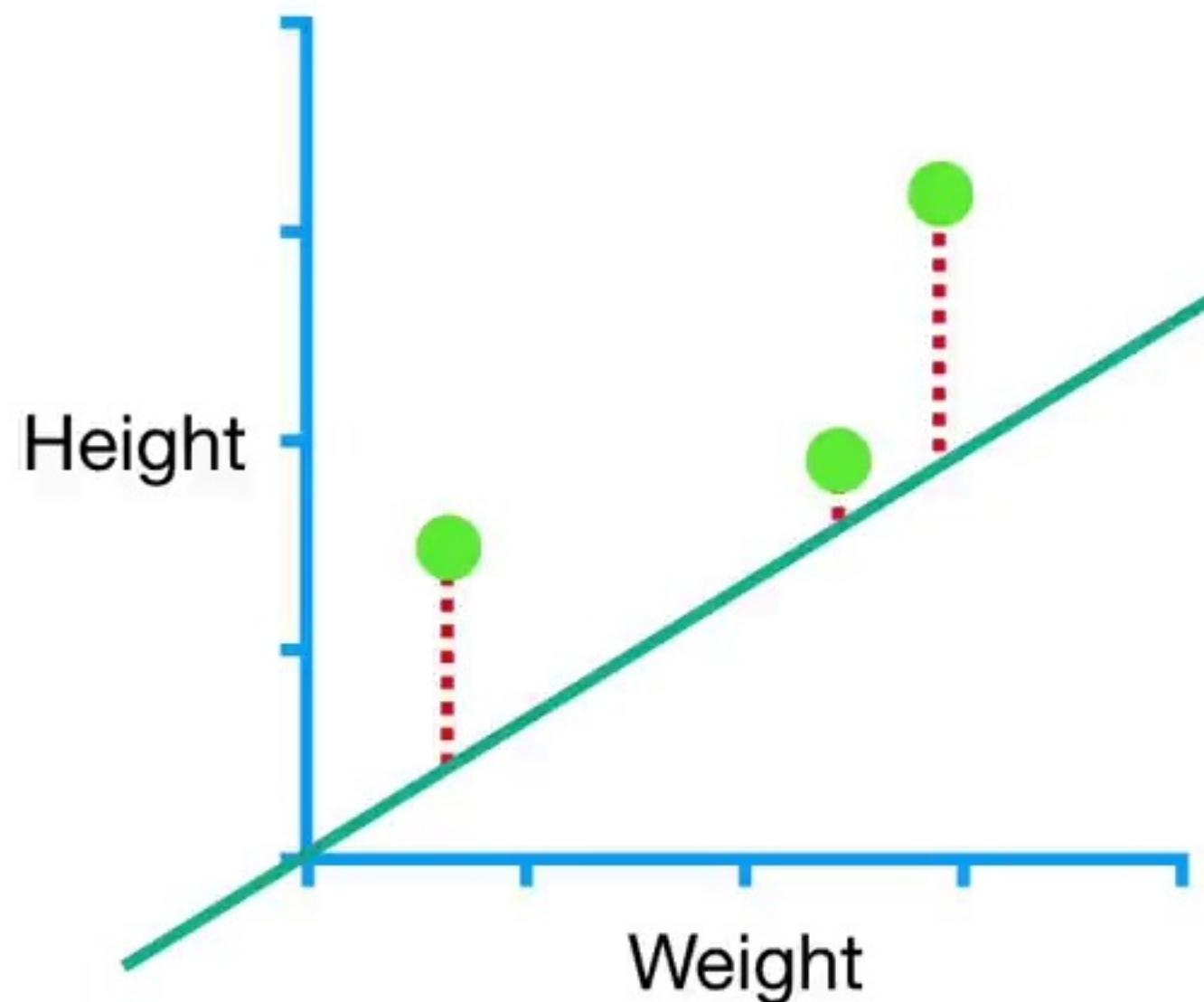
However, there are tons of other
Loss Functions that work with
other types of data.



Sum of squared residuals = $(1.4 - (\text{intercept} + 0.64 \times 0.5))^2$
+ $(1.9 - (\text{intercept} + 0.64 \times 2.3))^2$
+ $(3.2 - (\text{intercept} + 0.64 \times 2.9))^2$

However, there are tons of other
Loss Functions that work with
other types of data.

Regardless of which **Loss Function** you use, **Gradient Descent** works the same way.



Step 1: Take the derivative of the **Loss Function** for each parameter in it.

Step 1: Take the derivative of the **Loss Function** for each parameter in it.
In fancy Machine Learning Lingo, take the **Gradient** of the **Loss Function**.

Step 1: Take the derivative of the **Loss Function** for each parameter in it.
In fancy Machine Learning Lingo, take the **Gradient** of the **Loss Function**.

Step 2: Pick random values for the parameters.

Step 1: Take the derivative of the **Loss Function** for each parameter in it.
In fancy Machine Learning Lingo, take the **Gradient** of the **Loss Function**.

Step 2: Pick random values for the parameters.

Step 3: Plug the parameter values into the derivatives (ahem, the **Gradient**).

Step 1: Take the derivative of the **Loss Function** for each parameter in it.
In fancy Machine Learning Lingo, take the **Gradient** of the **Loss Function**.

Step 2: Pick random values for the parameters.

Step 3: Plug the parameter values into the derivatives (ahem, the **Gradient**).

Step 4: Calculate the Step Sizes: **Step Size = Slope × Learning Rate**

Step 1: Take the derivative of the **Loss Function** for each parameter in it.
In fancy Machine Learning Lingo, take the **Gradient** of the **Loss Function**.

Step 2: Pick random values for the parameters.

Step 3: Plug the parameter values into the derivatives (ahem, the **Gradient**).

Step 4: Calculate the Step Sizes: **Step Size = Slope × Learning Rate**

Step 5: Calculate the New Parameters:

New Parameter = Old Parameter - Step Size

Now go back to **Step 3** and repeat until
Step Size is very small, or you reach
the **Maximum Number of Steps**.

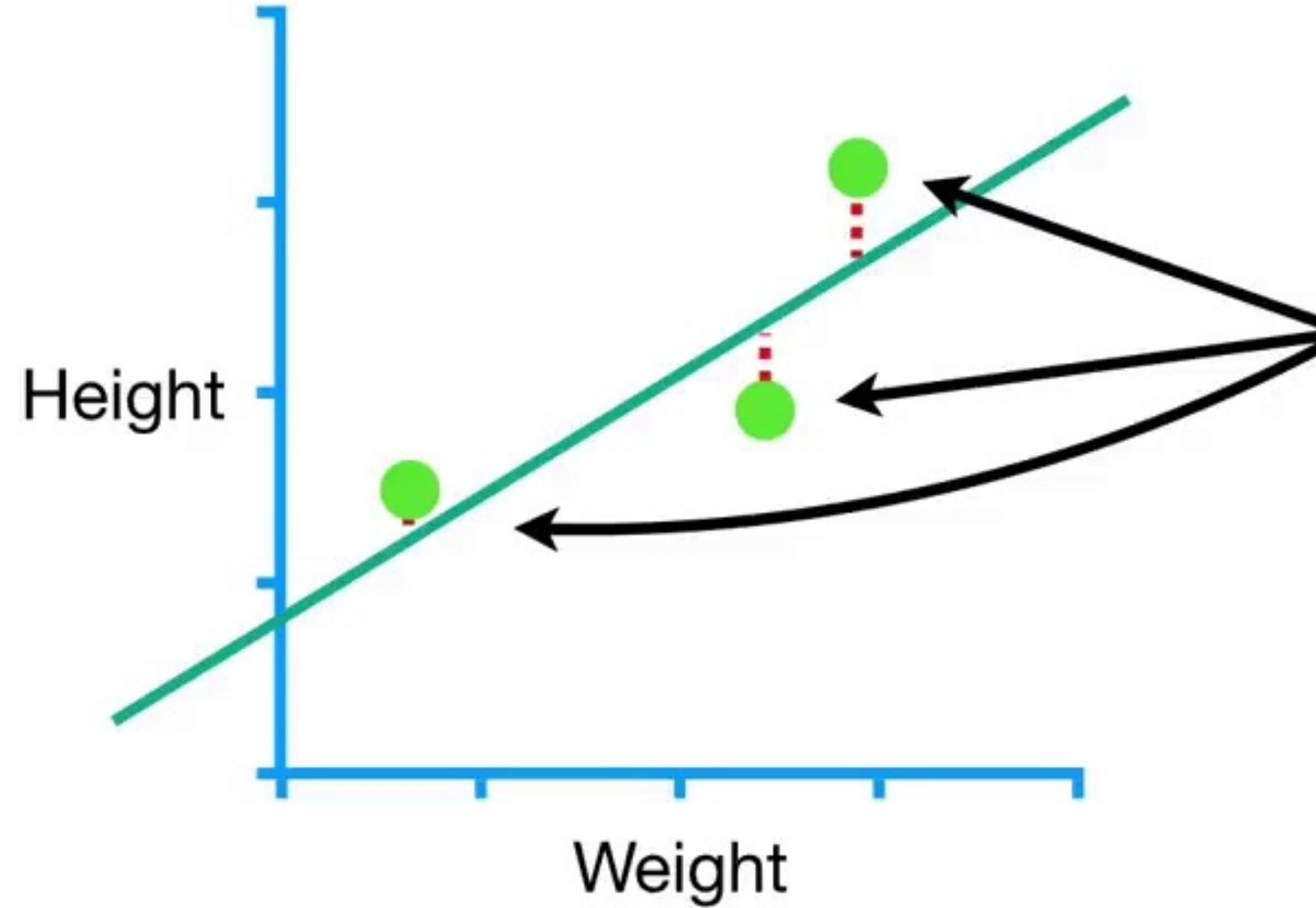
Step 3: Plug the parameter values into the derivatives (ahem, the **Gradient**).

Step 4: Calculate the Step Sizes: **Step Size = Slope × Learning Rate**

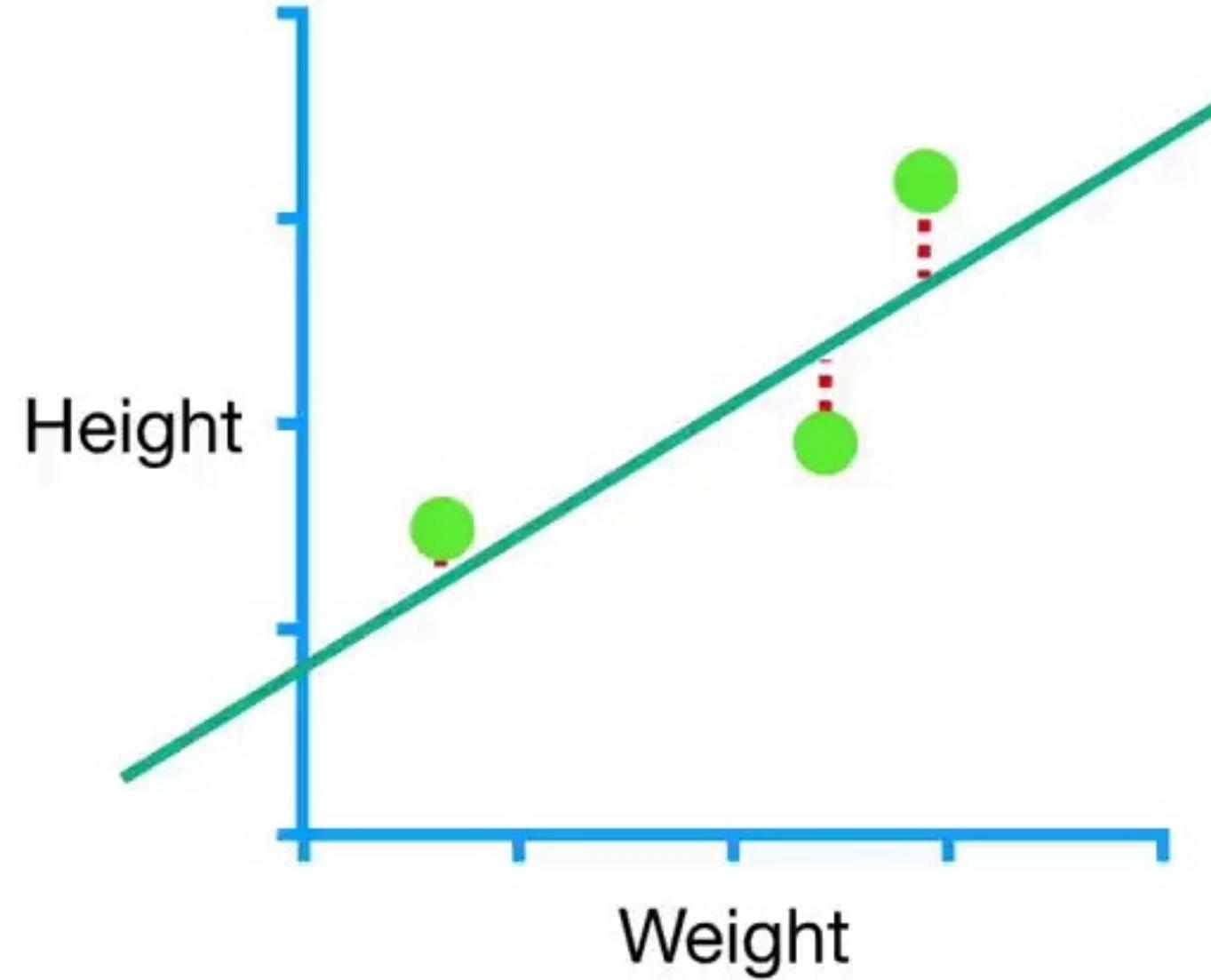
Step 5: Calculate the New Parameters:

New Parameter = Old Parameter - Step Size

One last thing before we're done...

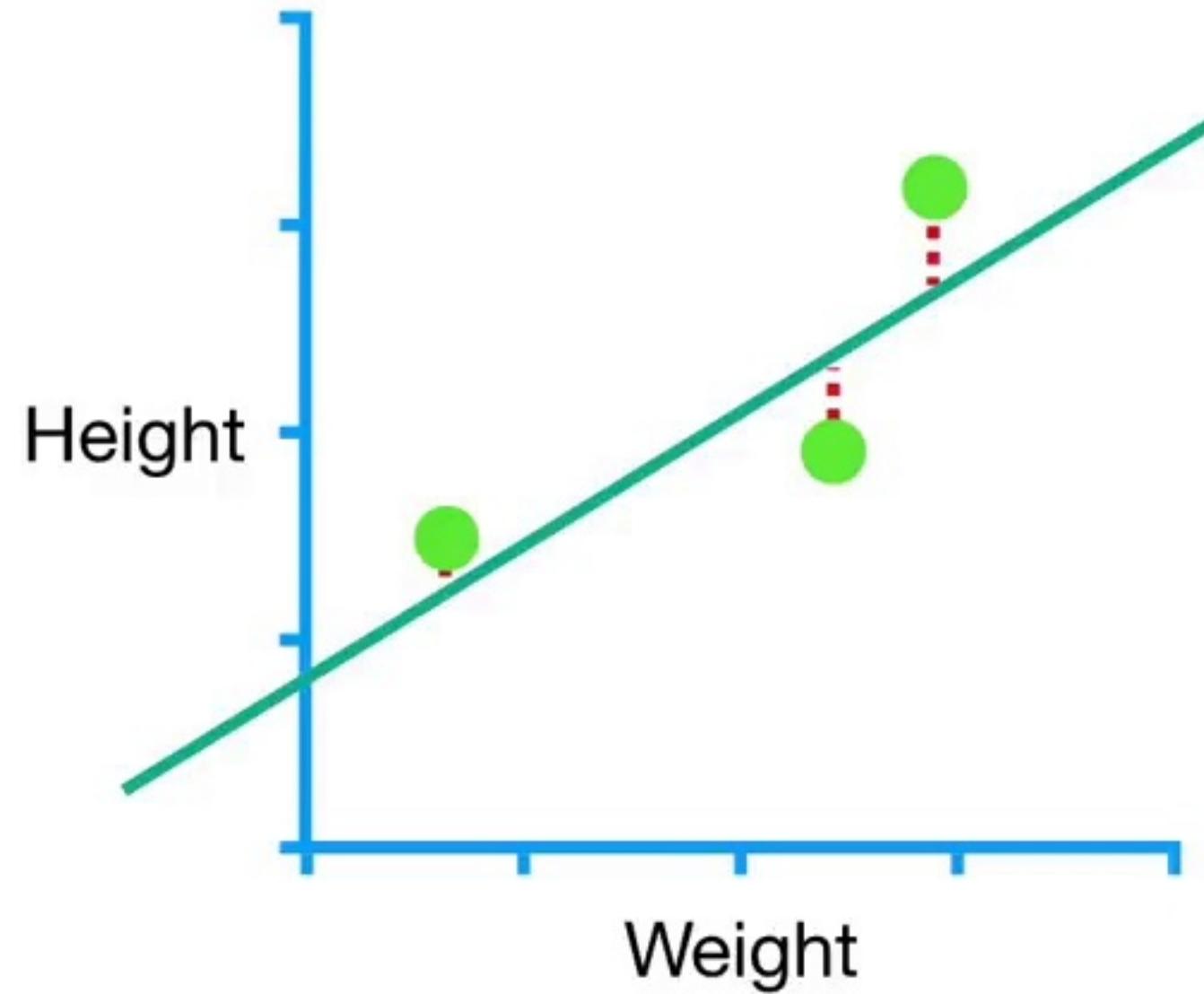


In our example, we only had three data points, so the math didn't take very long...



...but when you have millions of
data points, it can take a long
time.

So there is a thing called
Stochastic Gradient Descent
that uses a randomly selected
subset of the data at every step
rather than the full dataset.



So there is a thing called
Stochastic Gradient Descent
that uses a randomly selected
subset of the data at every step
rather than the full dataset.

This reduces the time spent
calculating the derivatives of the
Loss Function.

