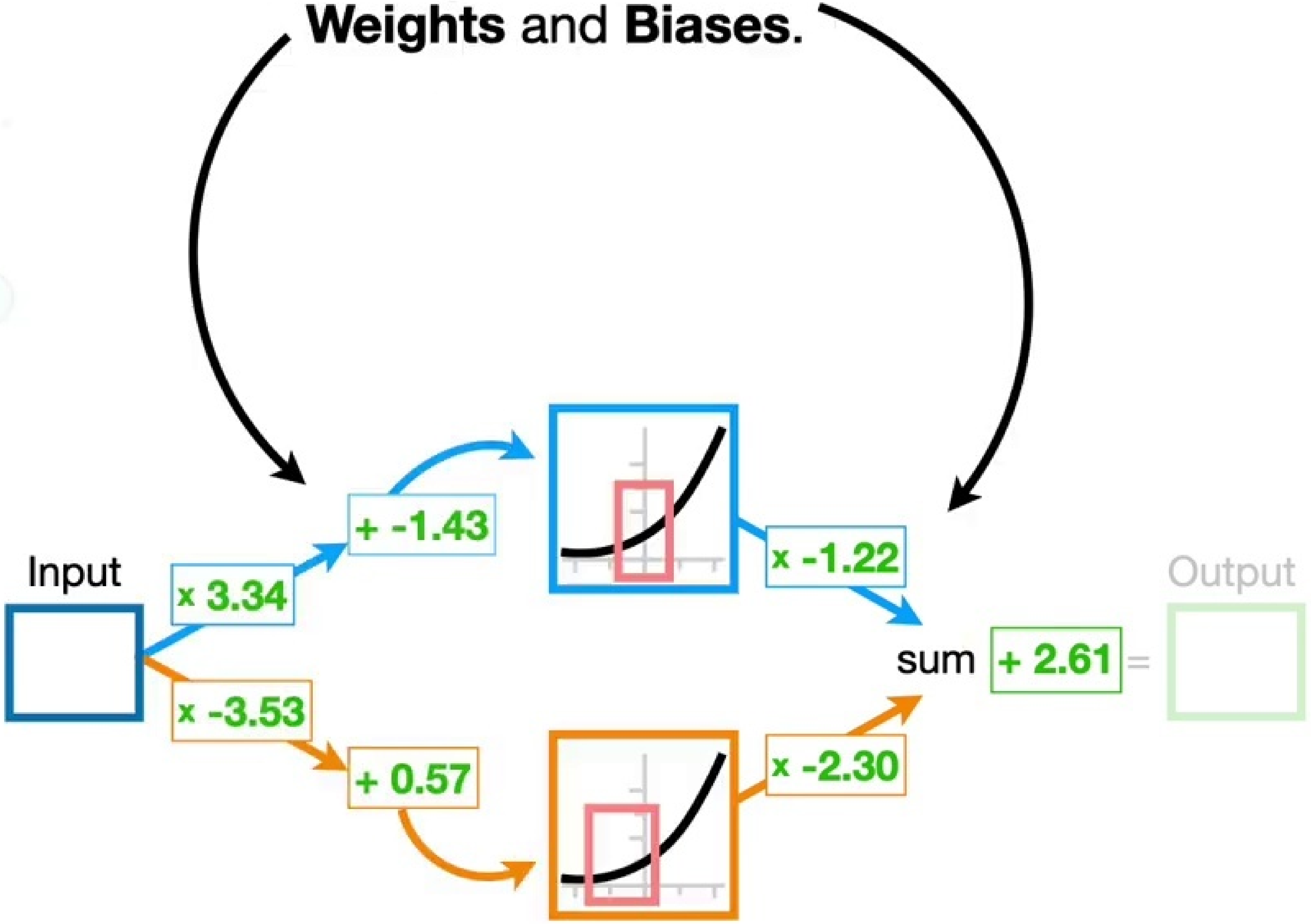
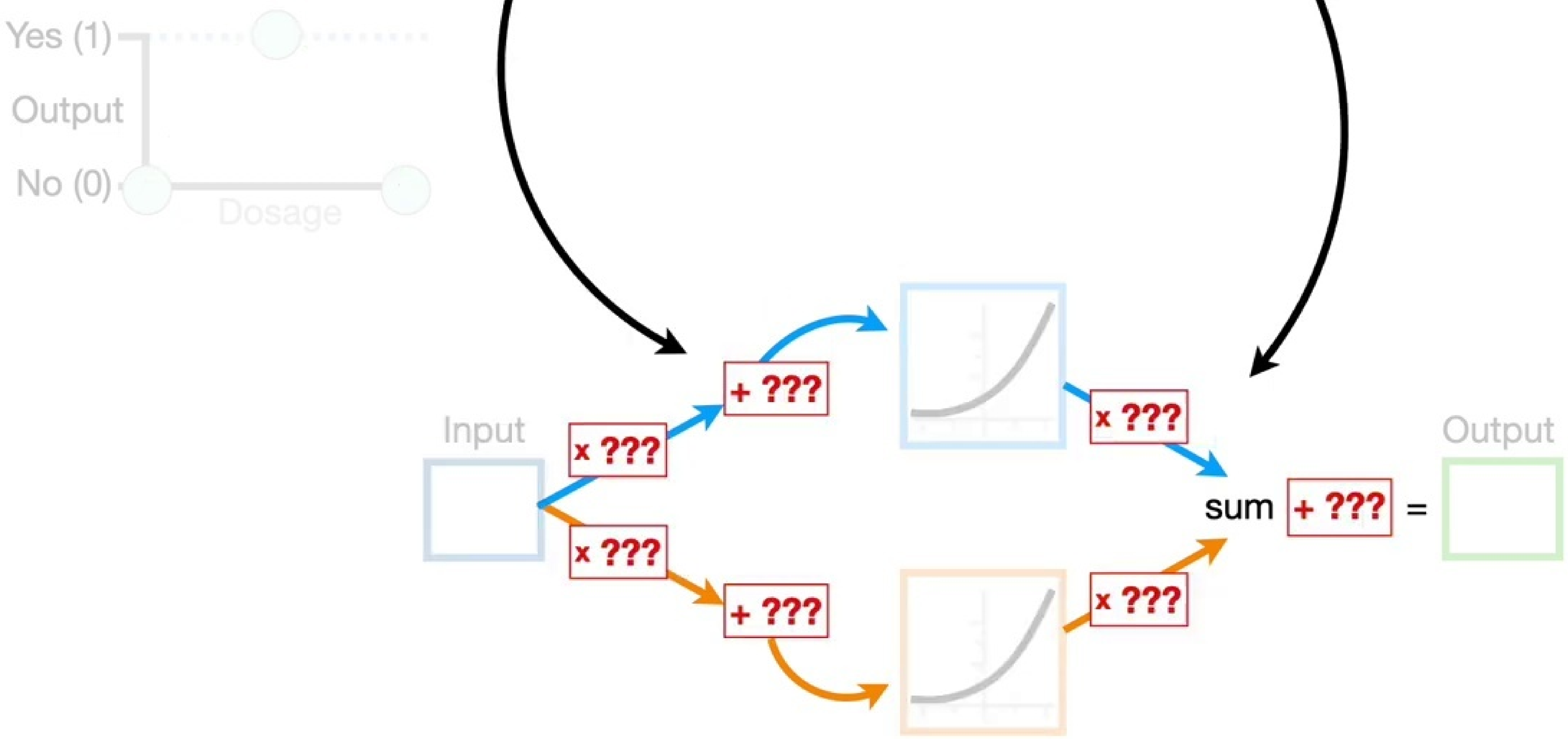
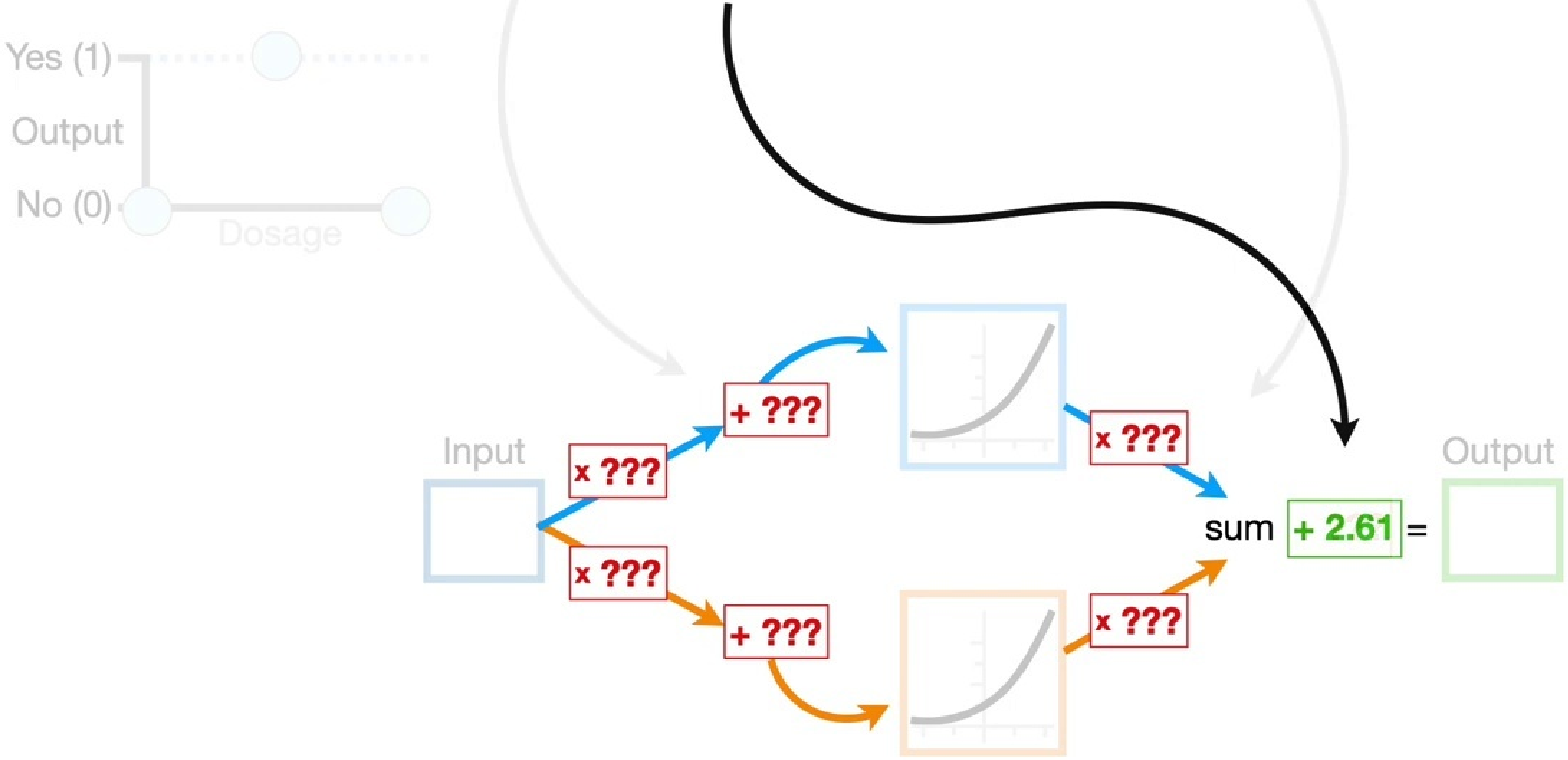
However, we did not talk about how to estimate the **Weights** and **Biases**.



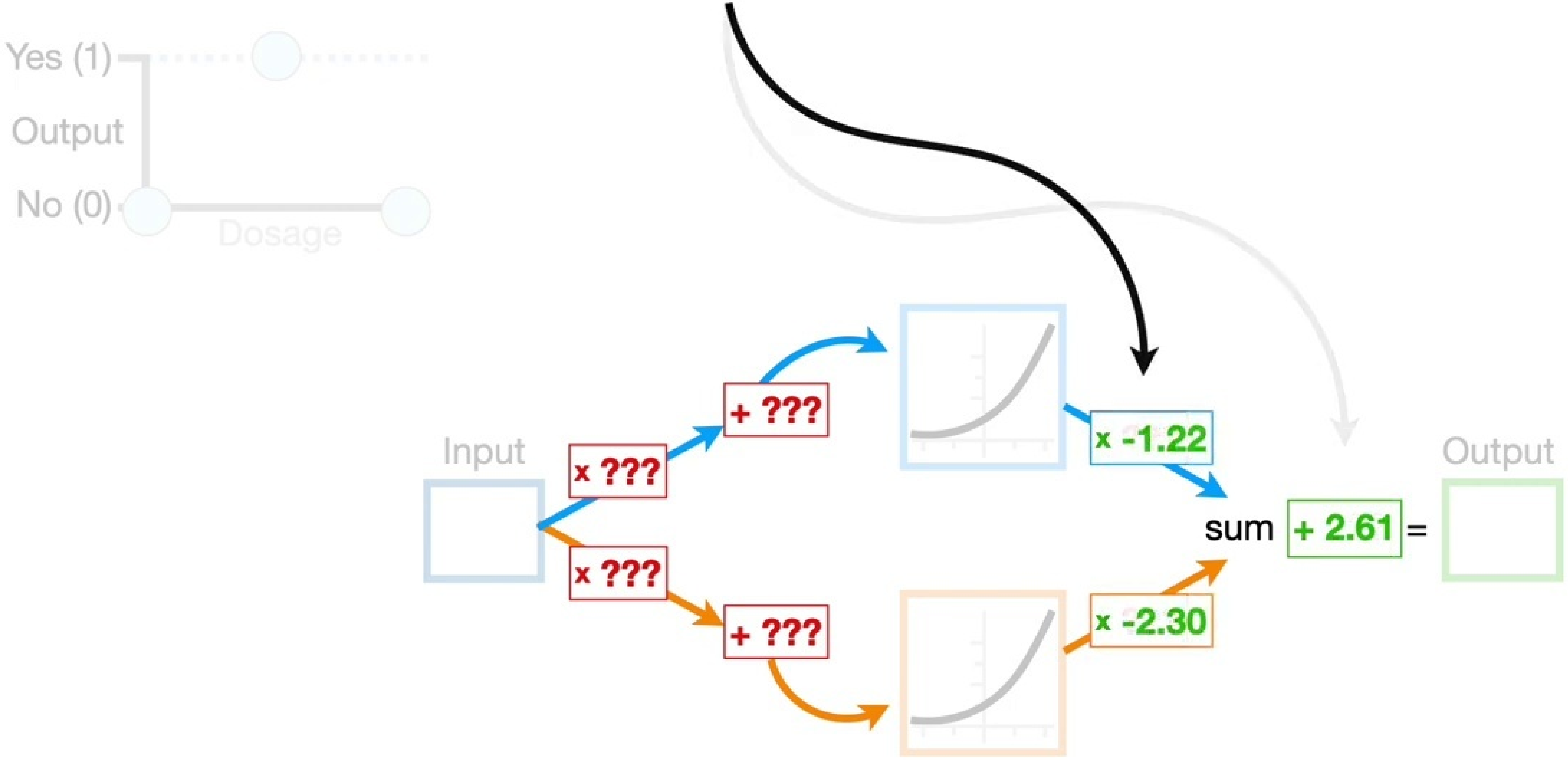
So let's talk about how **Backpropagation** optimizes the **Weights** and **Biases** in this and other **Neural Networks**.



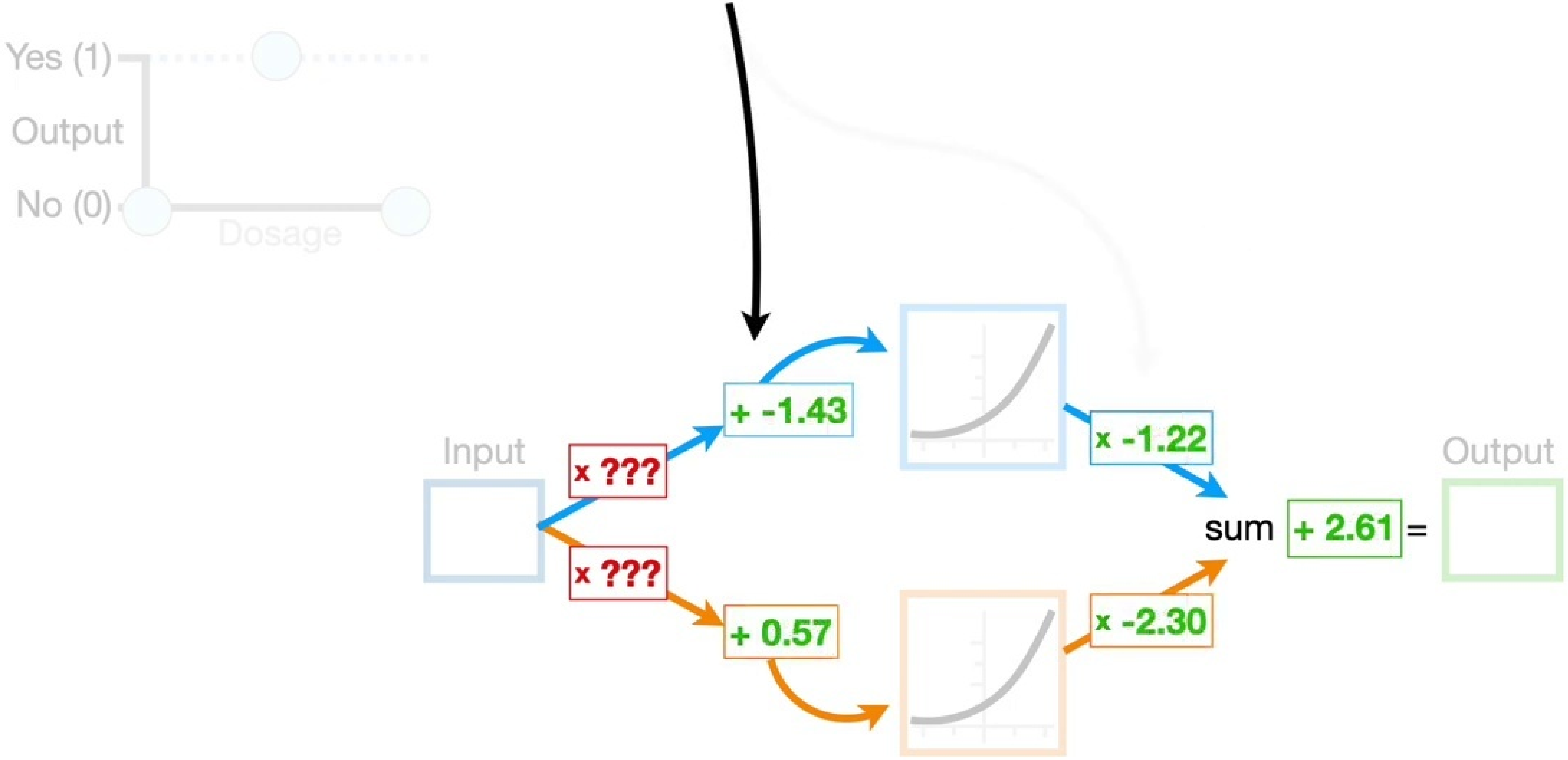
So let's talk about how **Backpropagation** optimizes the **Weights** and **Biases** in this and other **Neural Networks**.



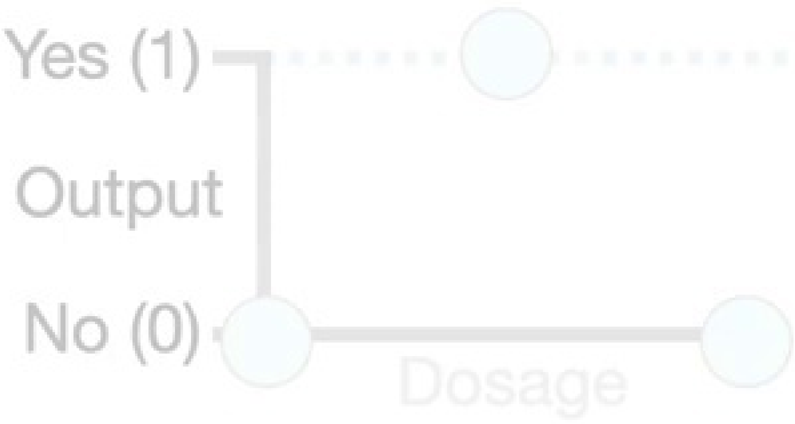
So let's talk about how **Backpropagation** optimizes the **Weights** and **Biases** in this and other **Neural Networks**.

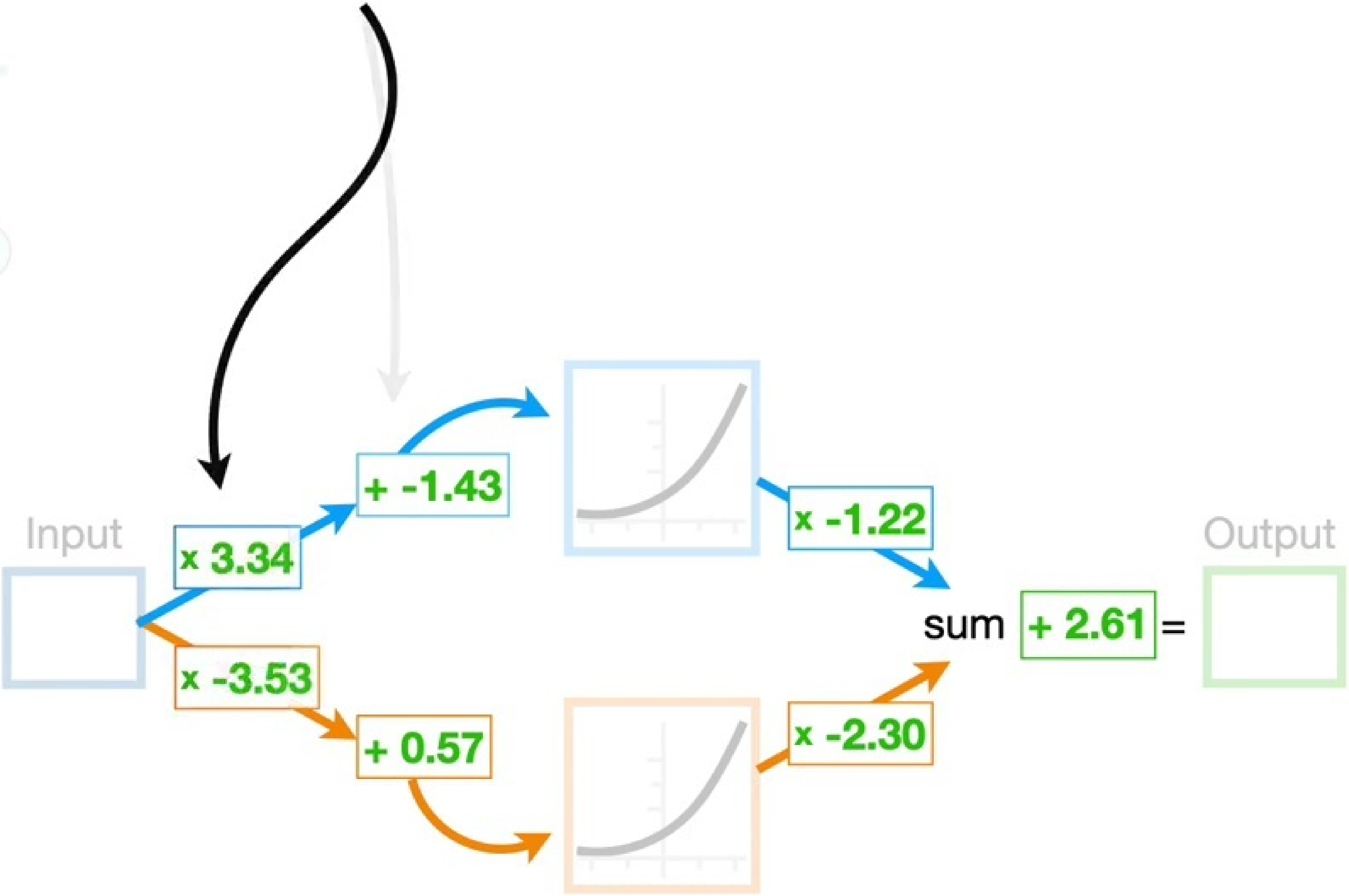


So let's talk about how **Backpropagation** optimizes the **Weights** and **Biases** in this and other **Neural Networks**.

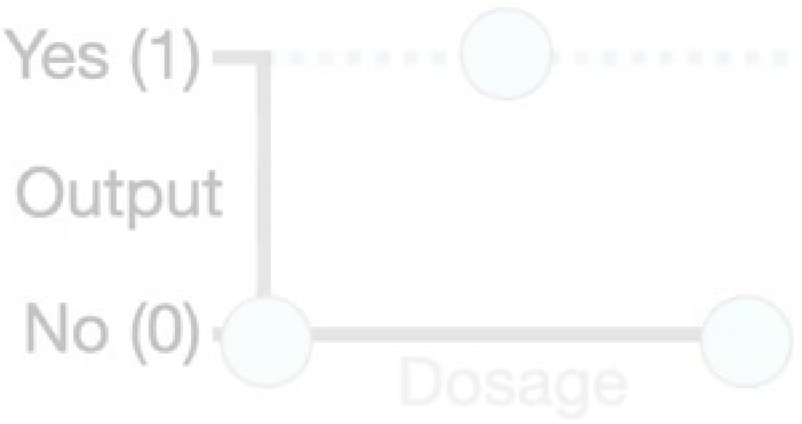


So let's talk about how **Backpropagation** optimizes the **Weights** and **Biases** in this and other **Neural Networks**.

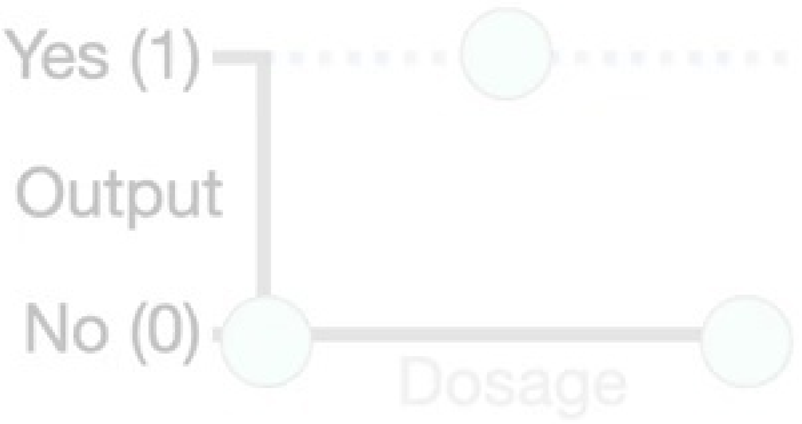




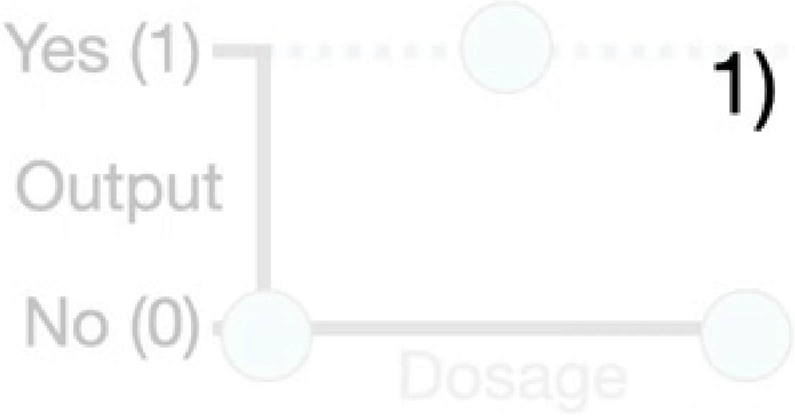
NOTE: Backpropagation is relatively simple, but there are a ton of details. So I've spit it up into bite sized pieces.



In this part, we talk about the **Main Ideas** of **Backpropagation:**



In this part, we talk about the **Main Ideas** of **Backpropagation:**

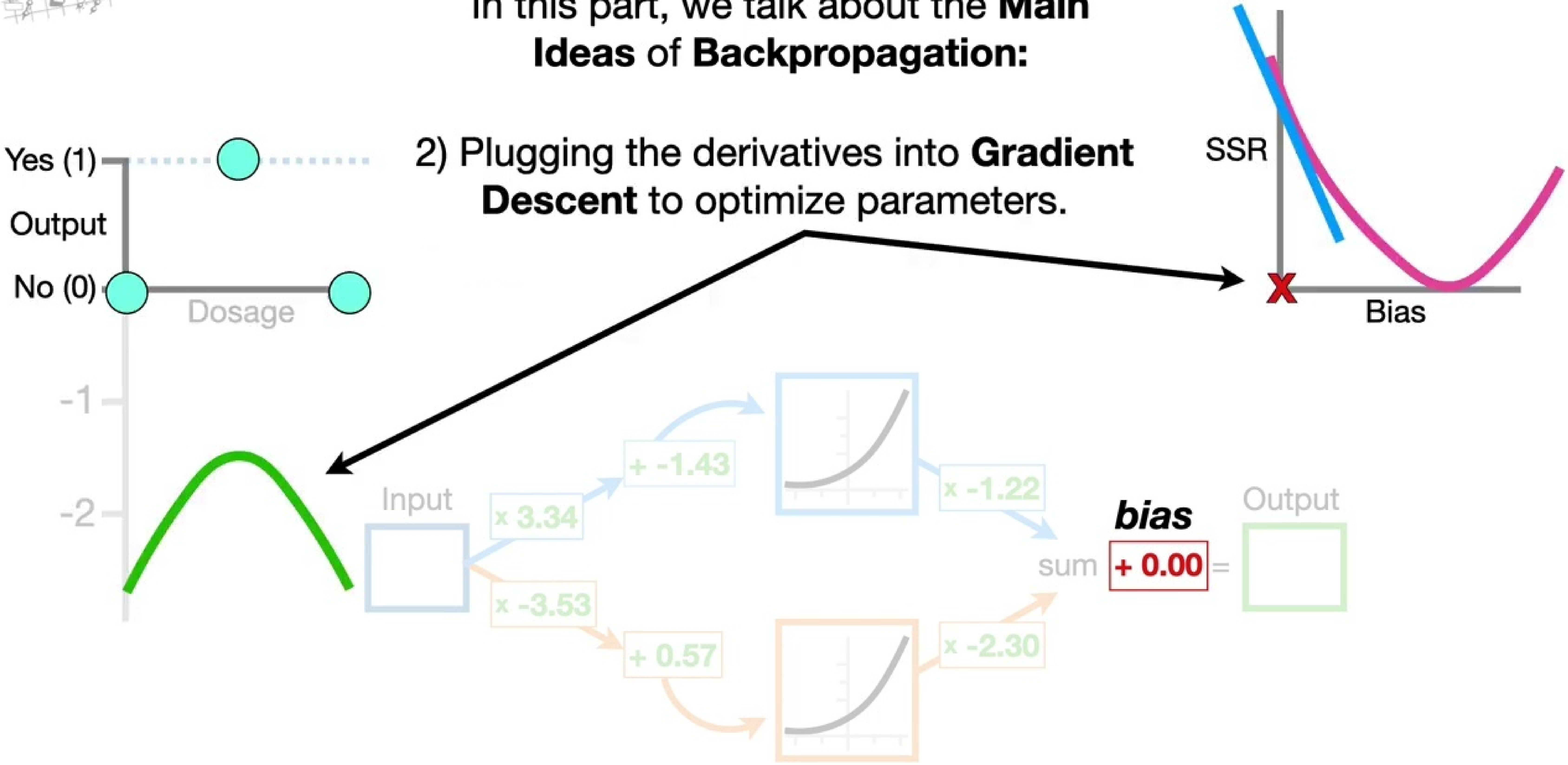


1) Using the **Chain Rule** to calculate derivatives…

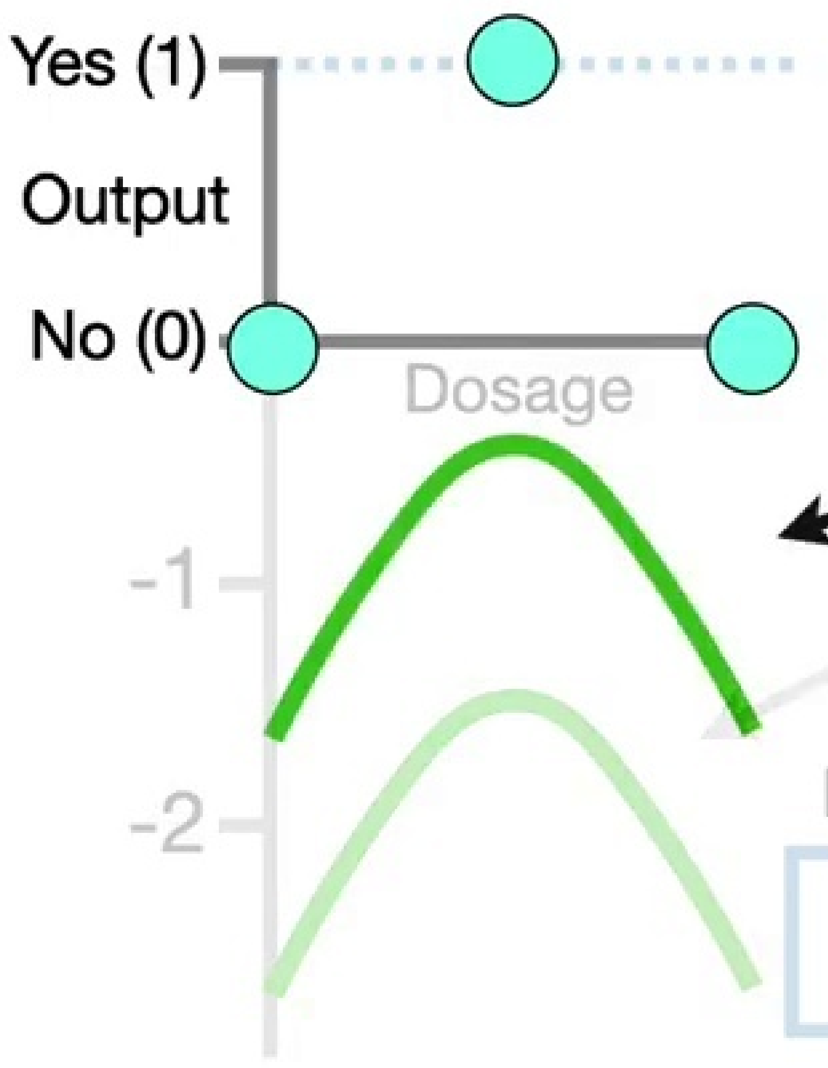
$$\frac{d\ SSR}{d\ bias} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ bias}$$

In this part, we talk about the **Main Ideas** of **Backpropagation:**

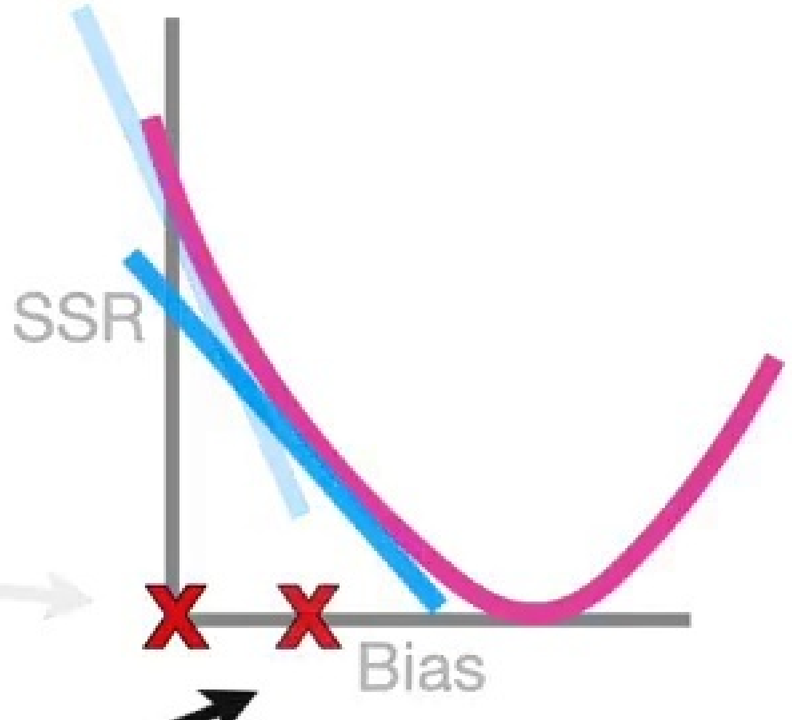
2) Plugging the derivatives into **Gradient Descent** to optimize parameters.



In this part, we talk about the **Main Ideas** of **Backpropagation:**

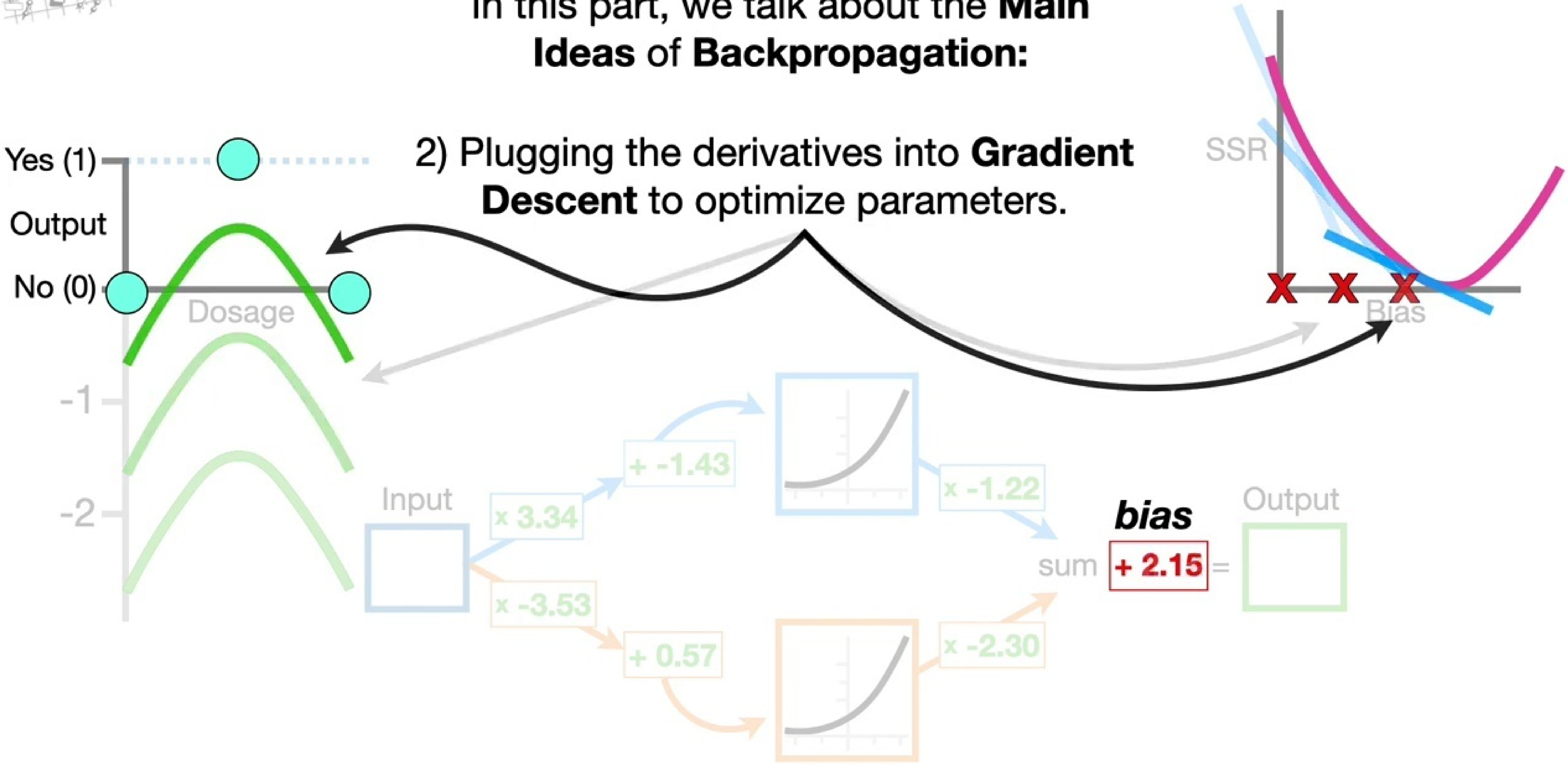


2) Plugging the derivatives into **Gradient Descent** to optimize parameters.



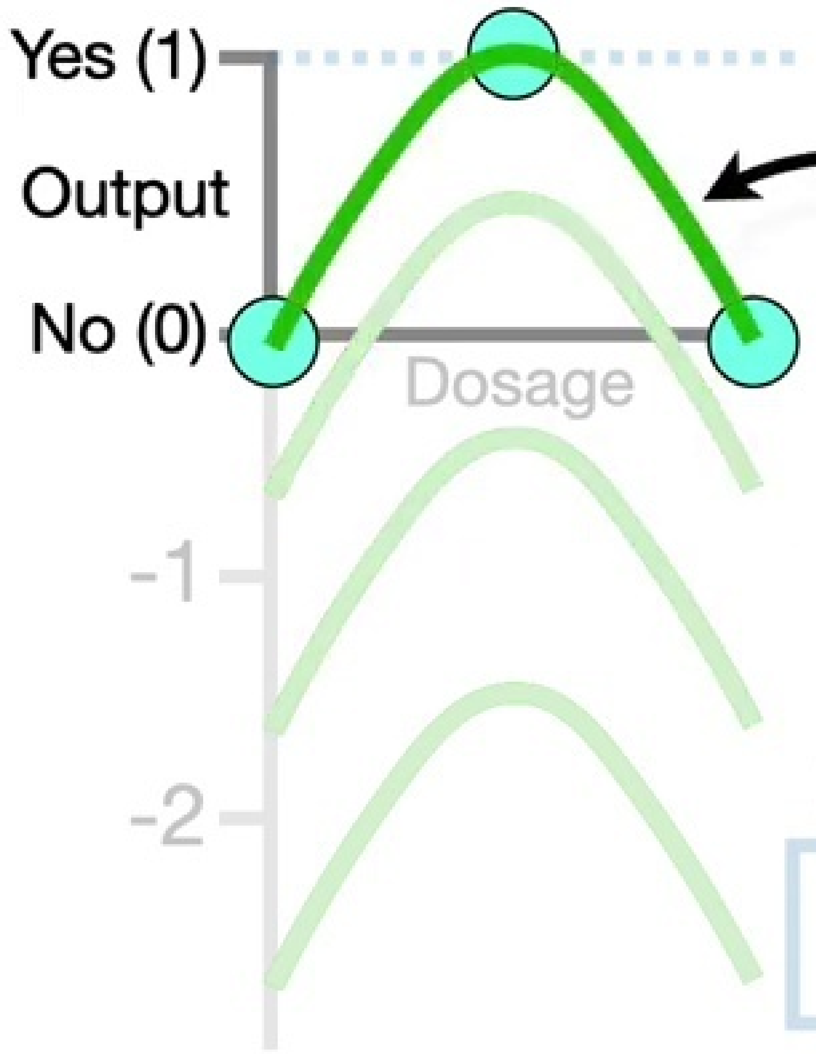
In this part, we talk about the **Main Ideas** of **Backpropagation:**

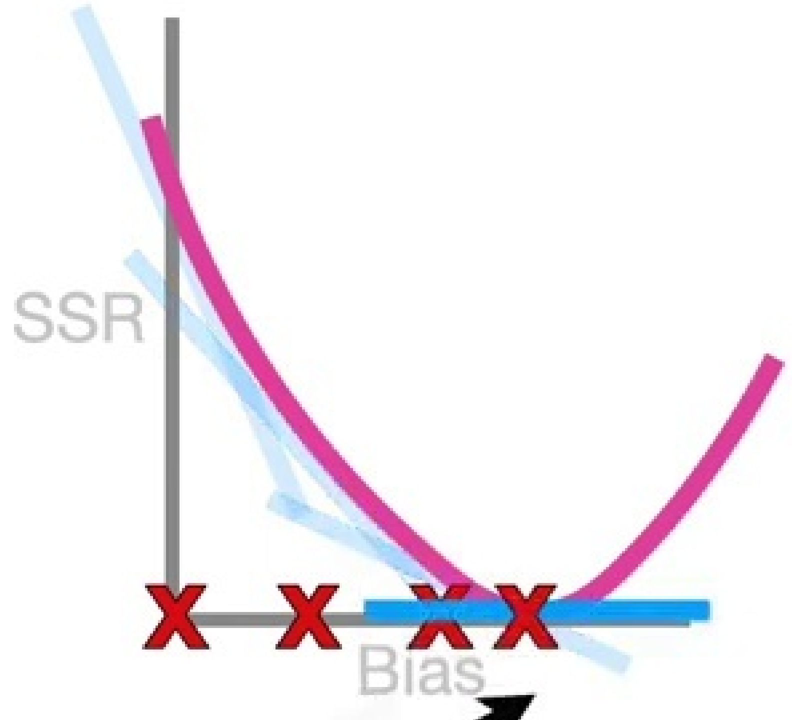
2) Plugging the derivatives into **Gradient Descent** to optimize parameters.



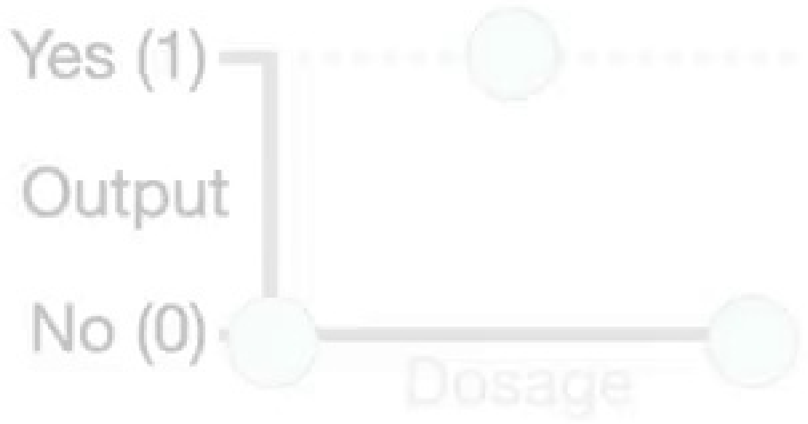
In this part, we talk about the **Main Ideas** of **Backpropagation:**

2) Plugging the derivatives into **Gradient Descent** to optimize parameters.

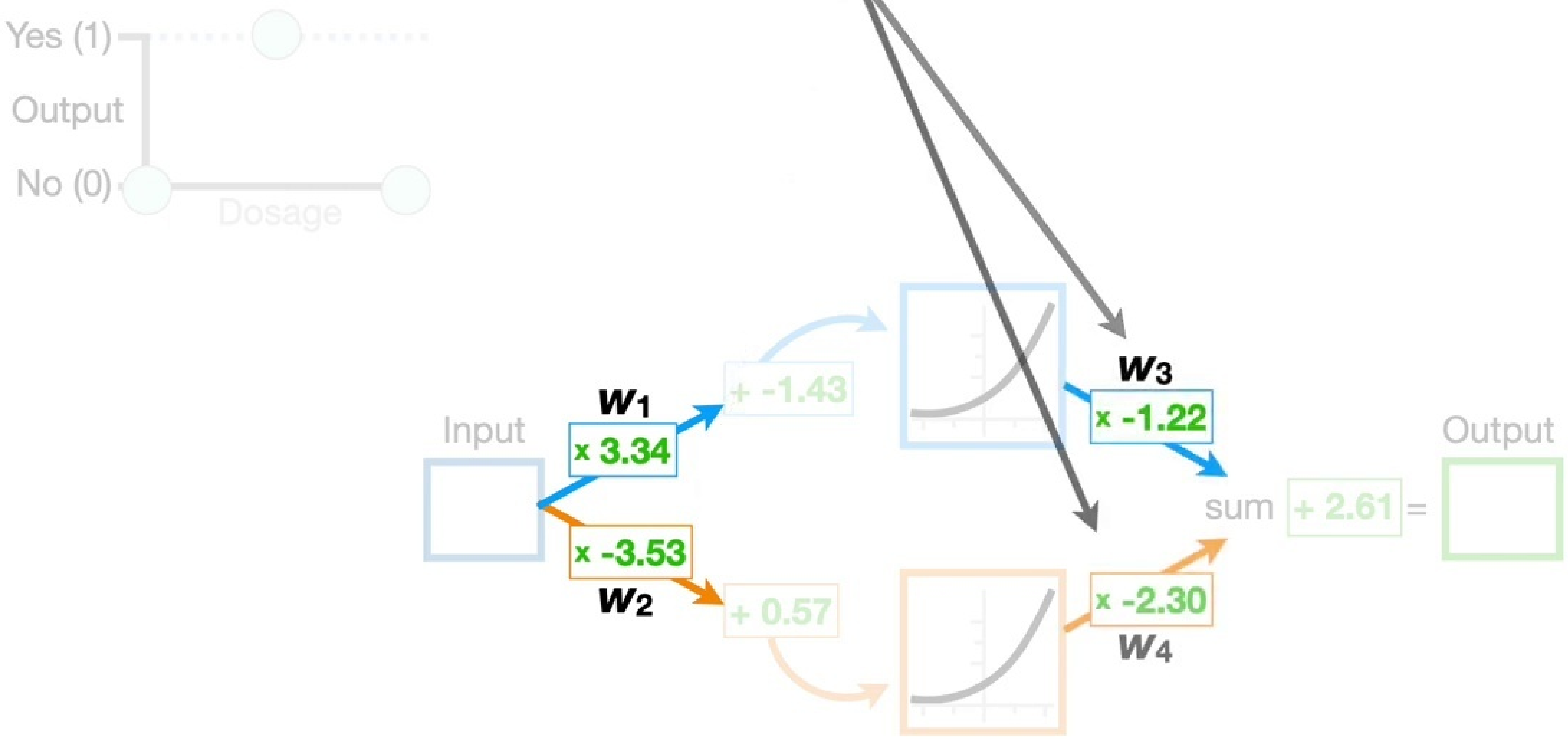




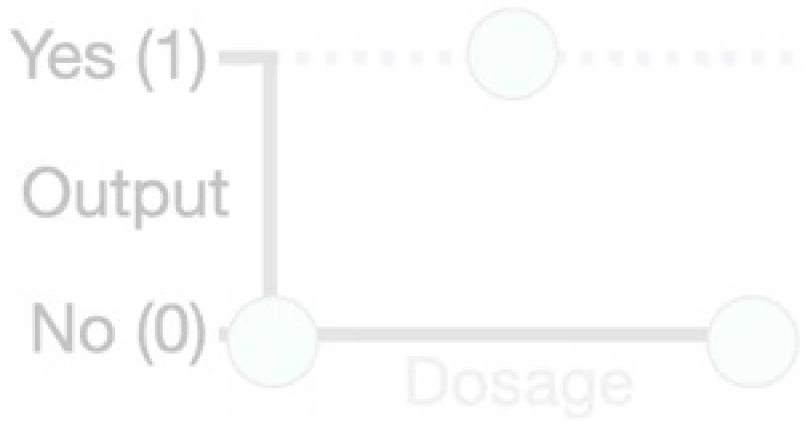
First, so we can be clear about which specific **Weights** we are talking about, let's give each one a name.



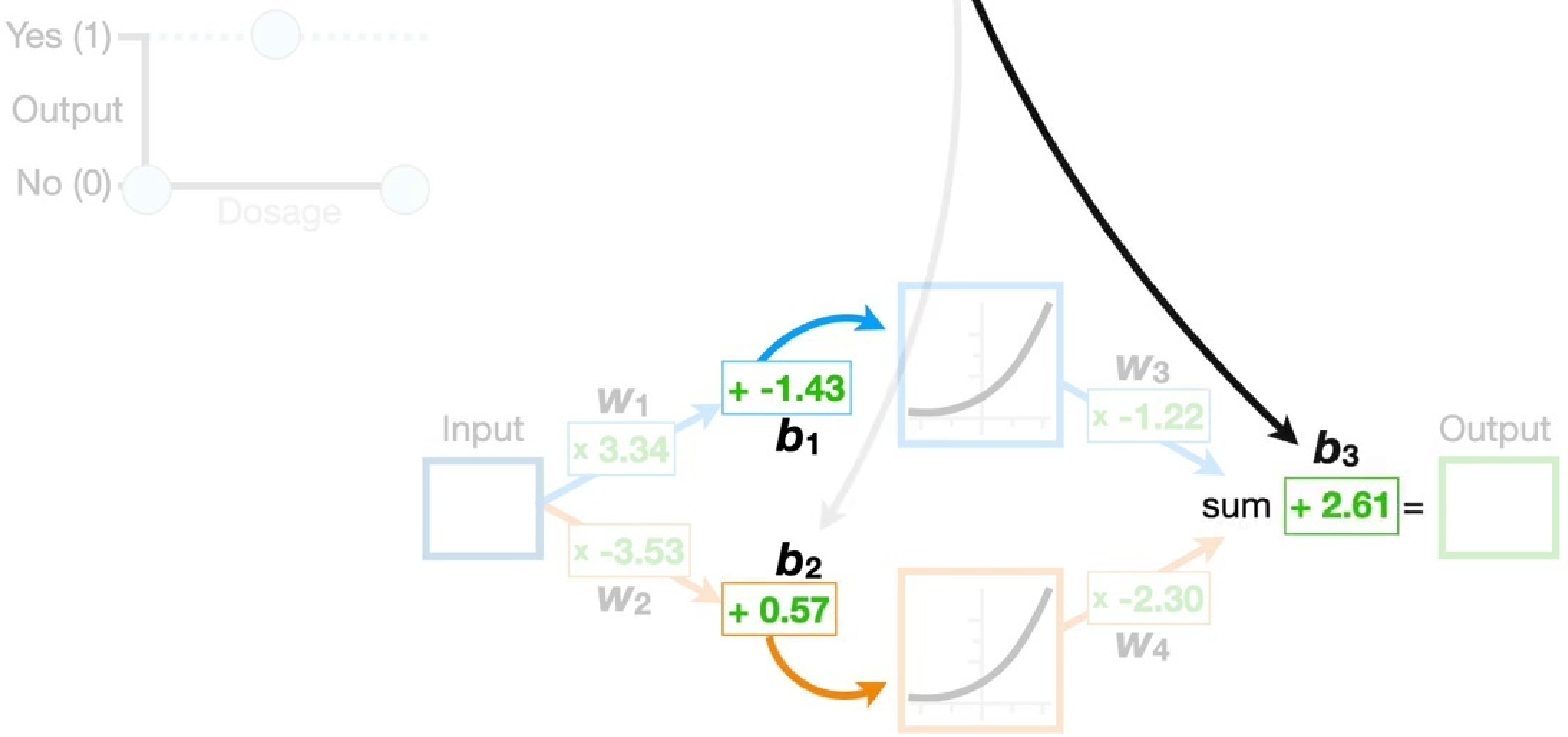
First, so we can be clear about which specific **Weights** we are talking about, let's give each one a name.



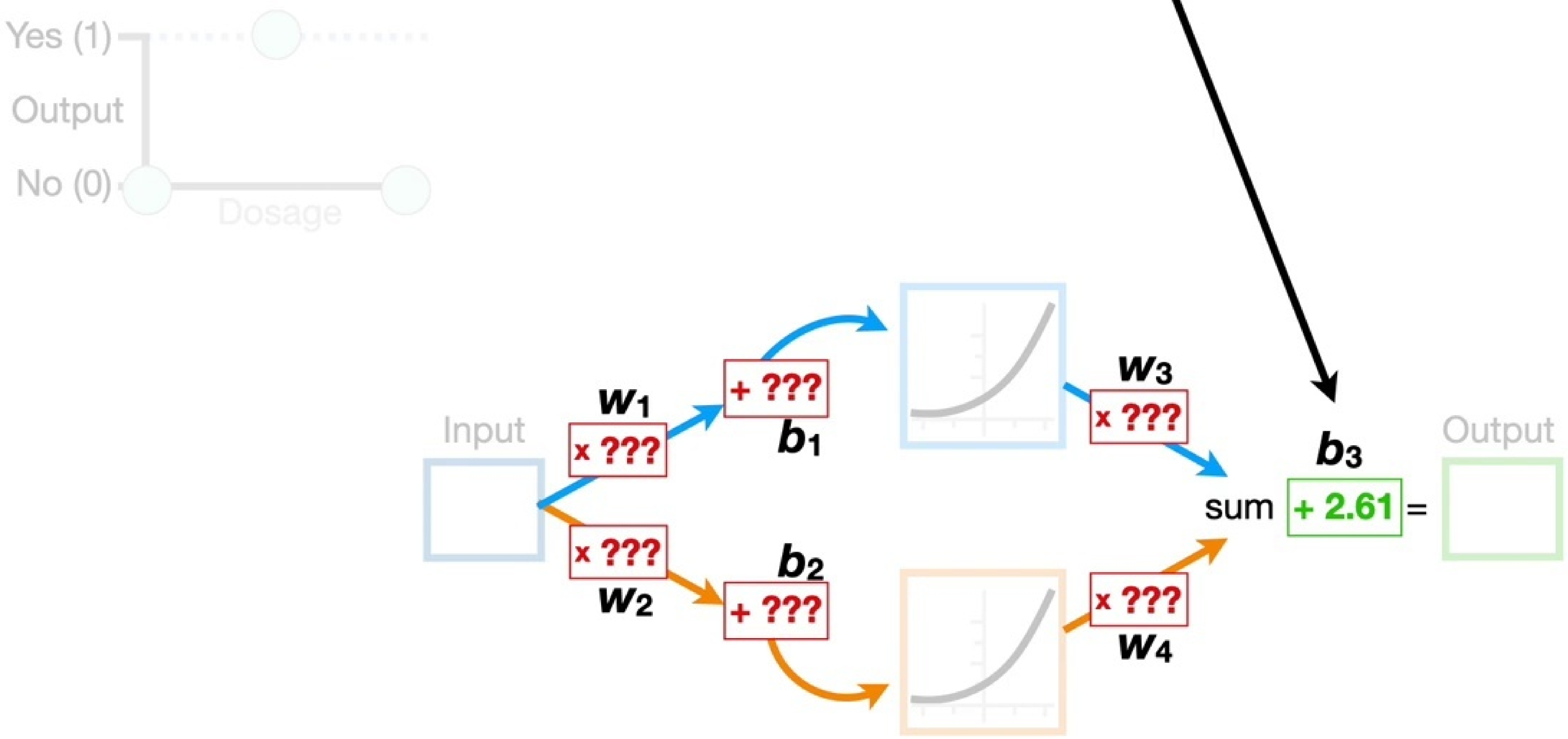
And let's name
each **Bias.**



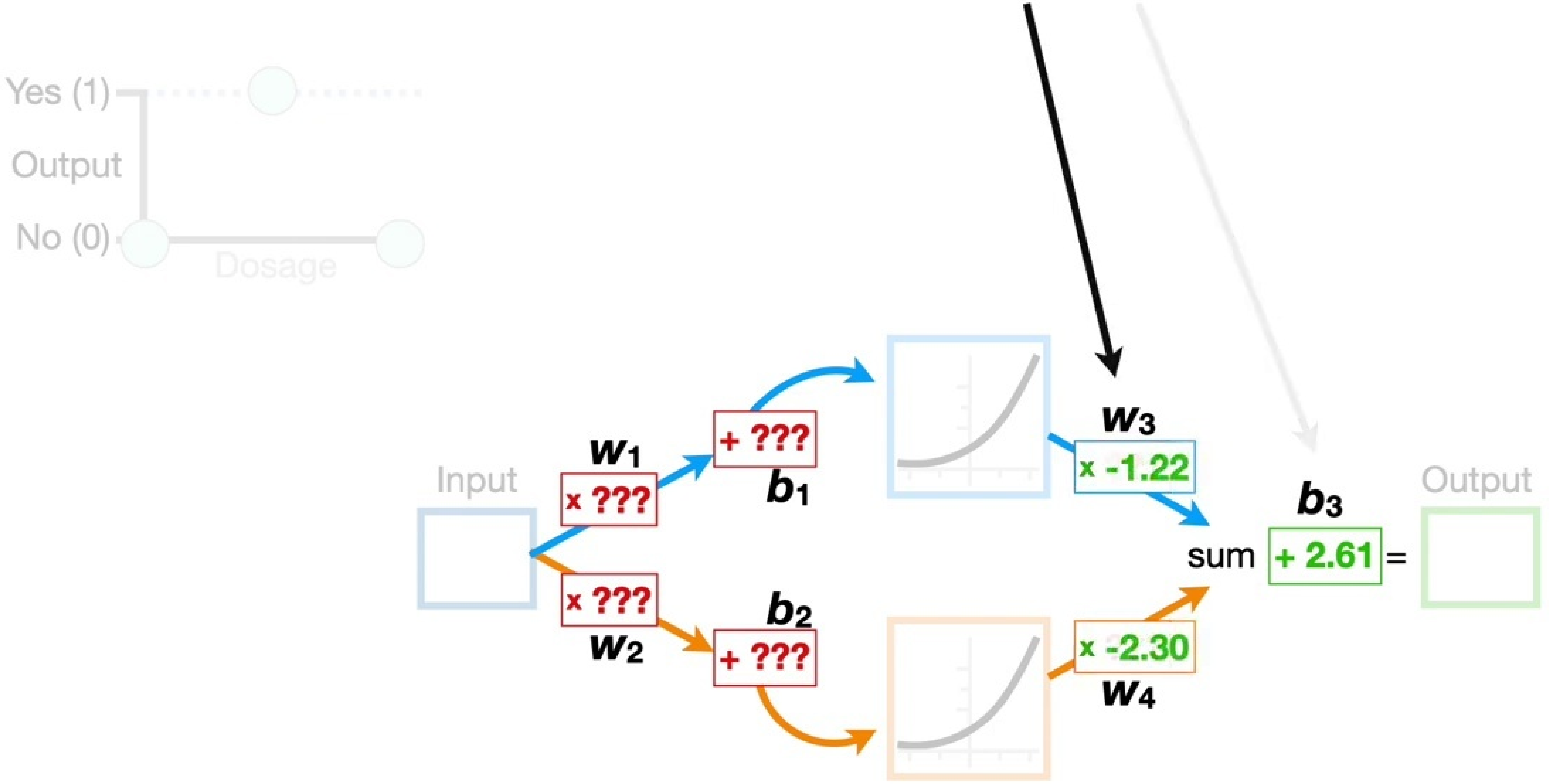
And let's name each **Bias.**



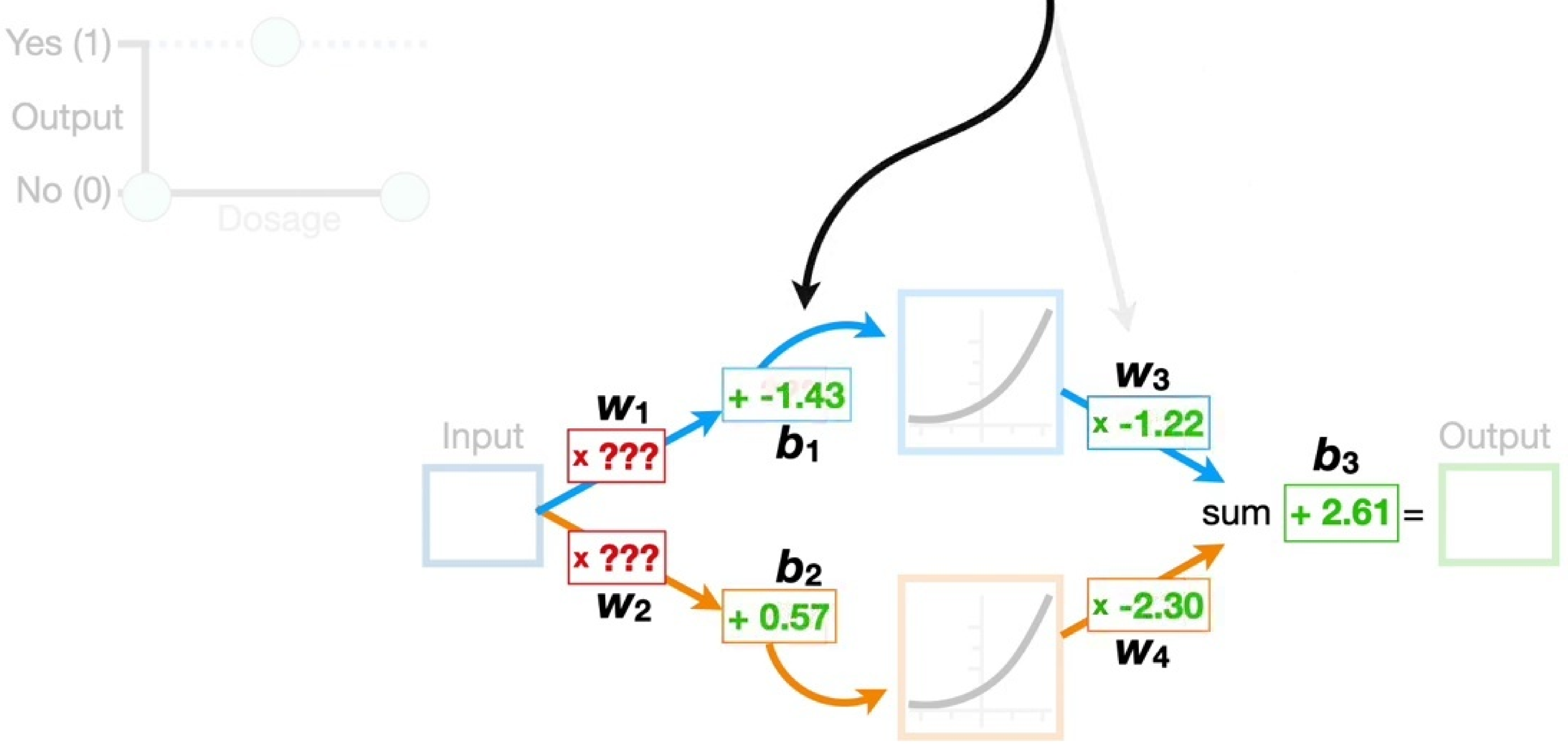
...and works its way backwards to estimate all of the other parameters.



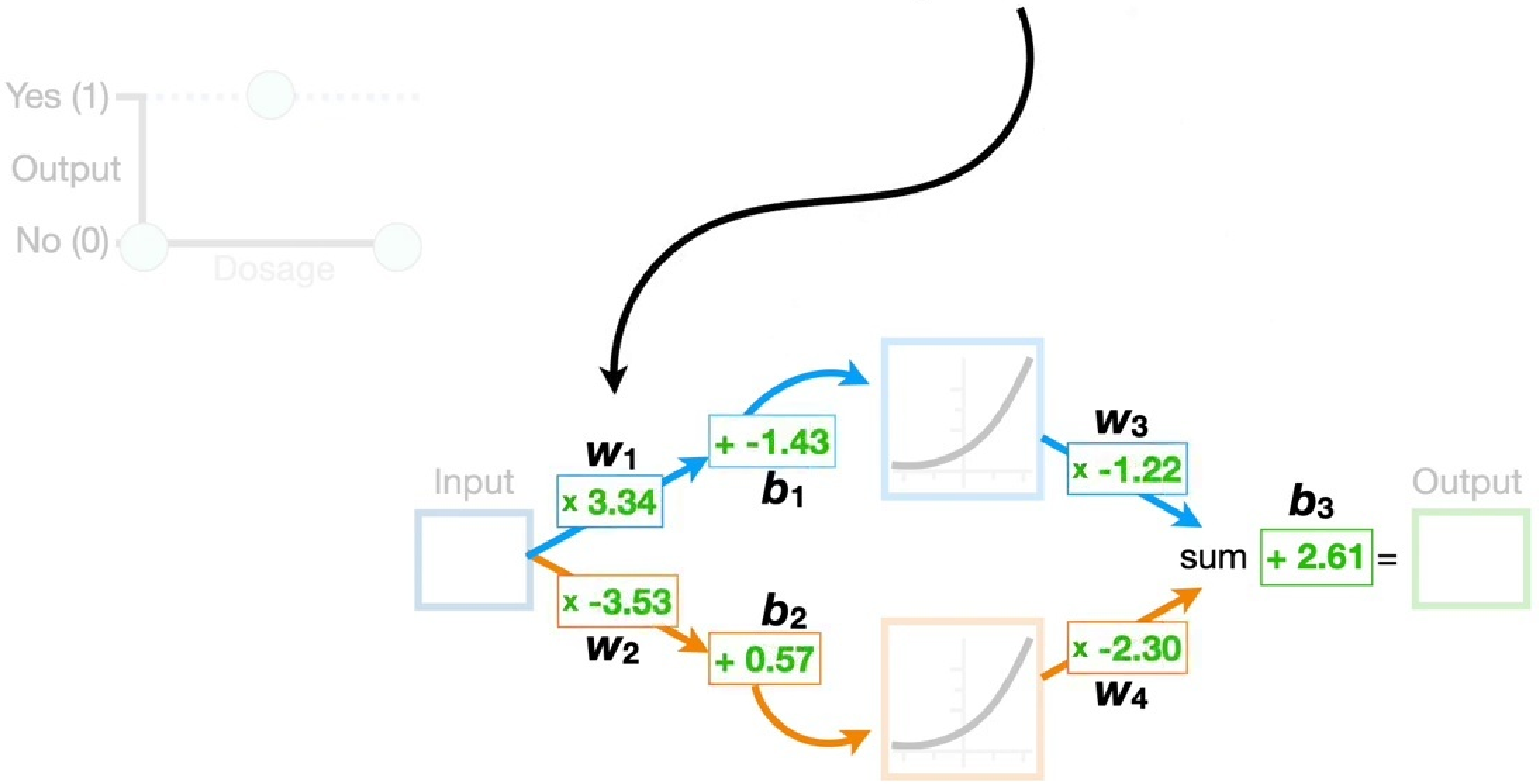
...and works its way backwards to estimate all of the other parameters.



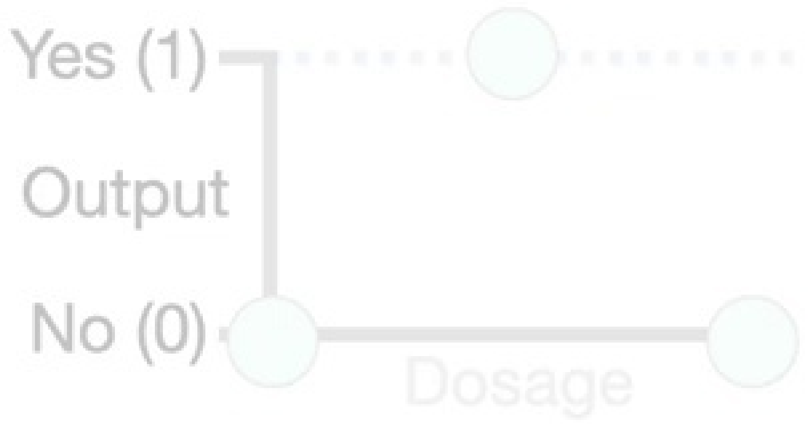
...and works its way backwards to estimate all of the other parameters.

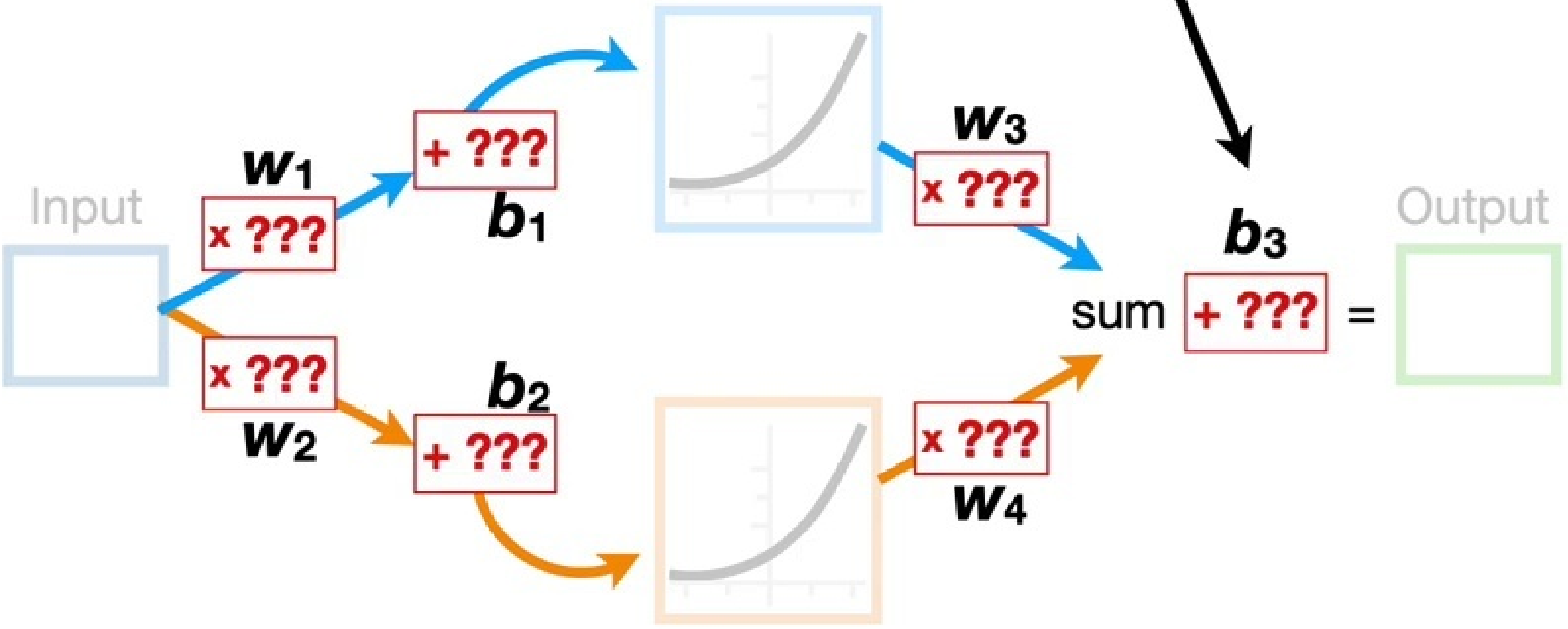


…and works its way backwards to estimate all of the other parameters.

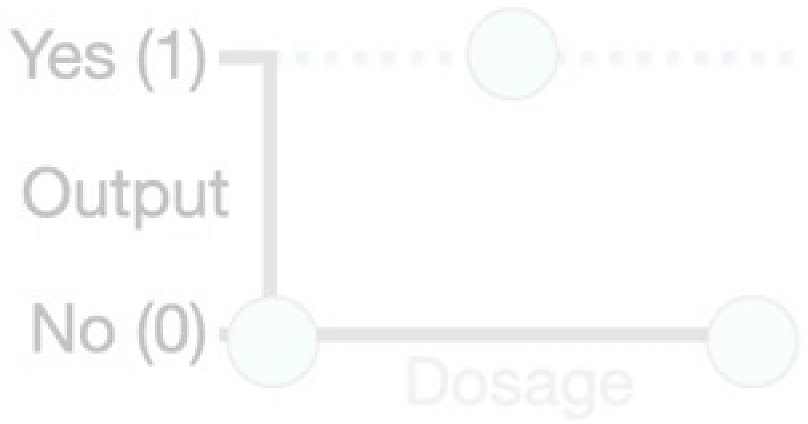


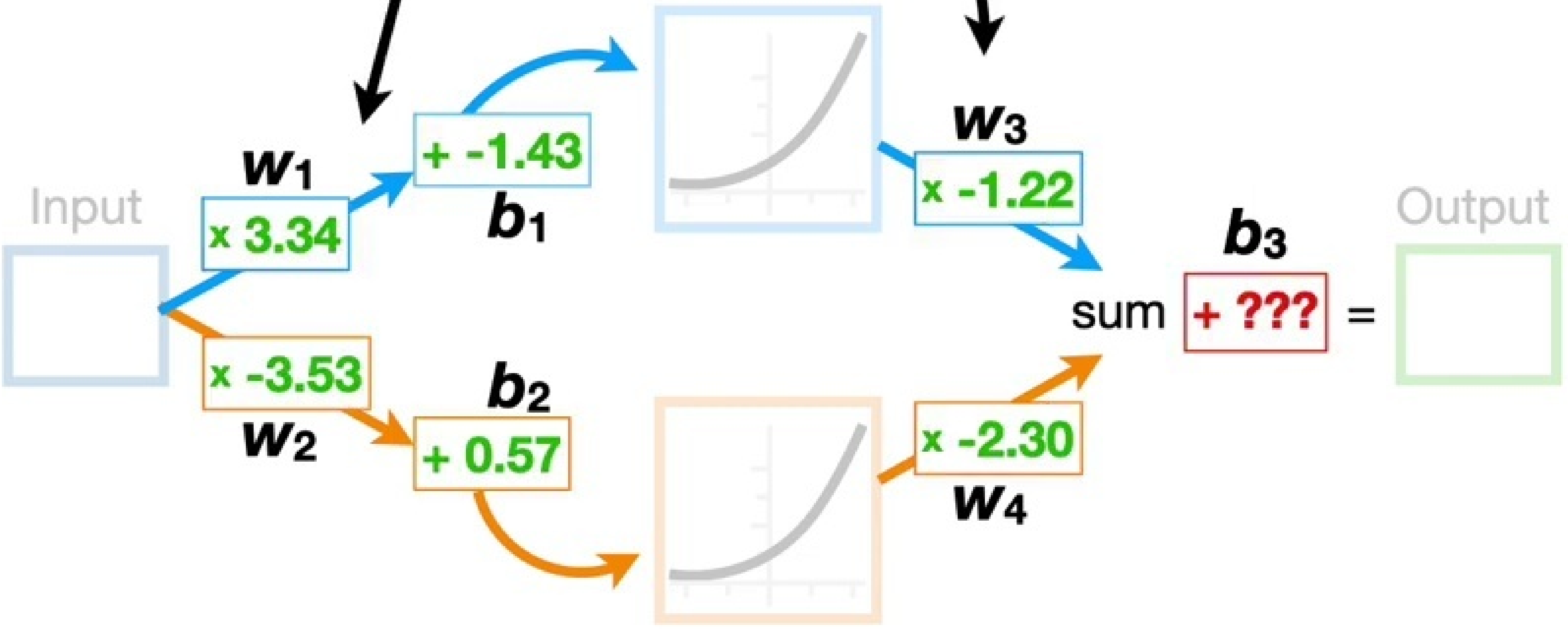
However, we can discuss all of the **Main Ideas** behind **Backpropagation** by just estimating the last **Bias**, ***b*₃**.





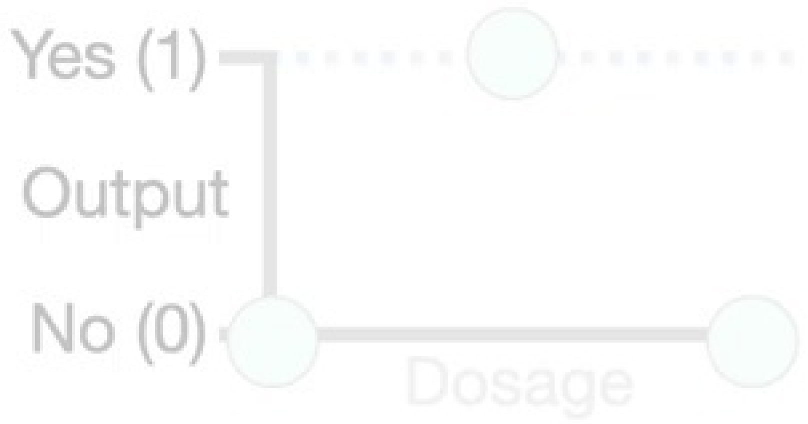
So, in order to start from the back, let's assume that we already have **optimal values** for all of the parameters…

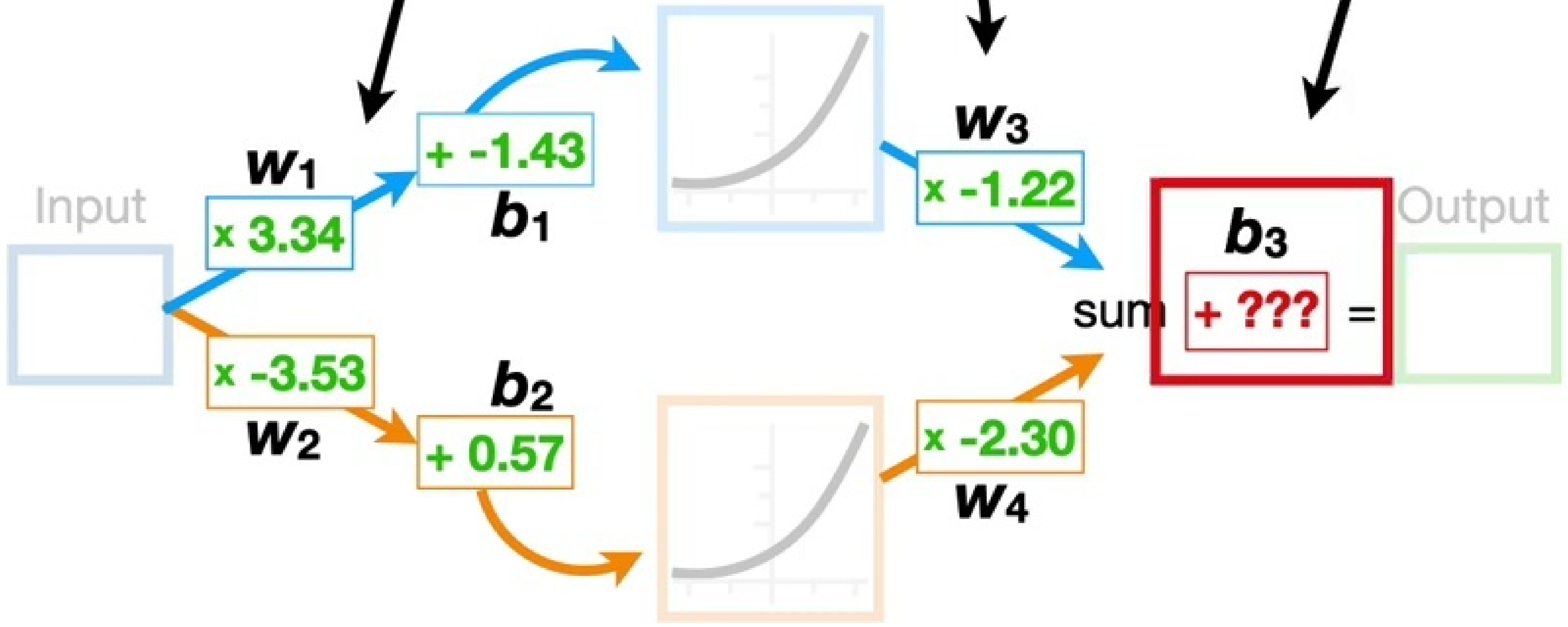




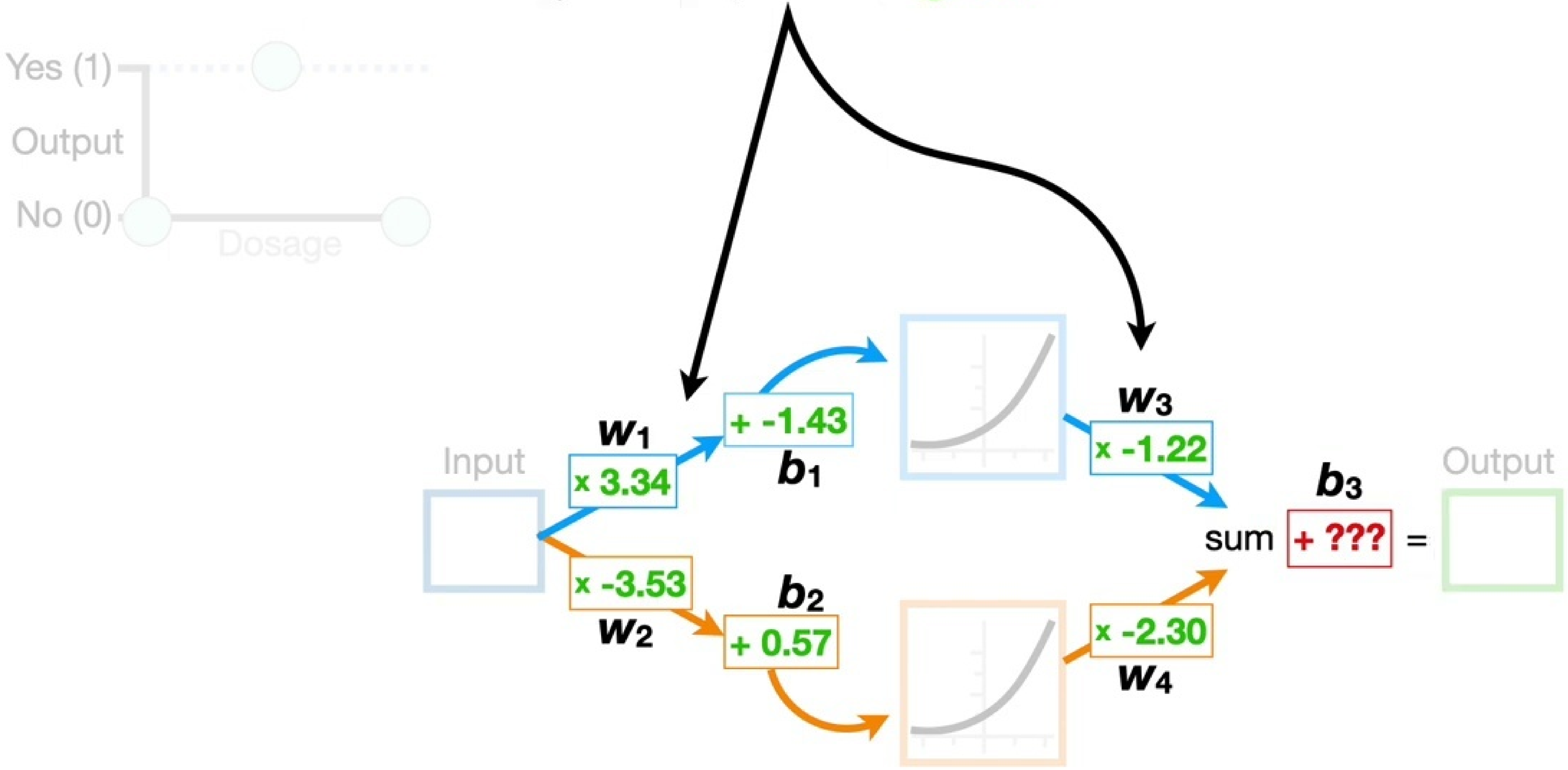
So, in order to start from the back, let's assume that we already have **optimal values** for all of the parameters…

…except for the last **Bias** term, ***b*₃**.



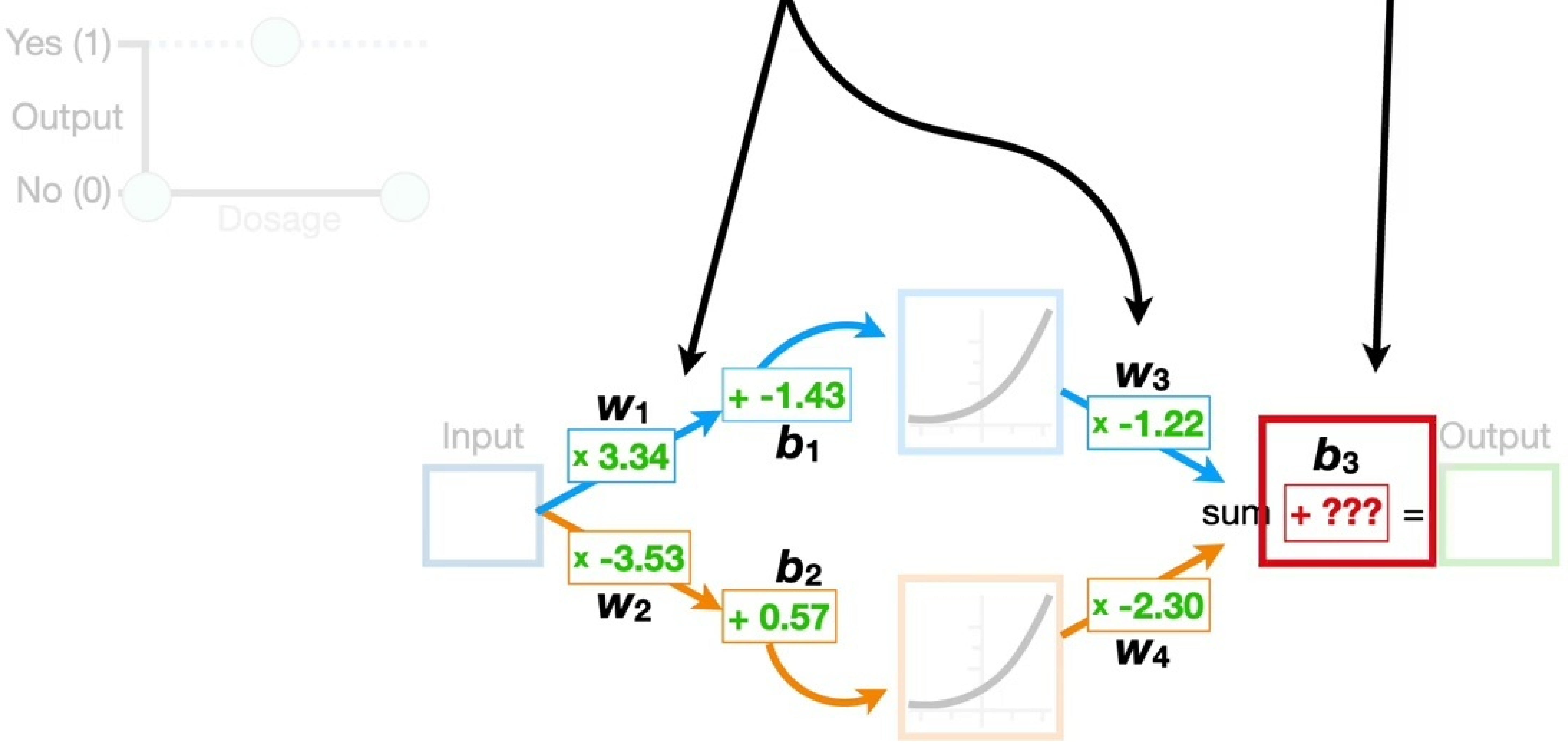


NOTE: Throughout this and the next **StatQuests**, I will make parameter values that have already been optimized **green**…

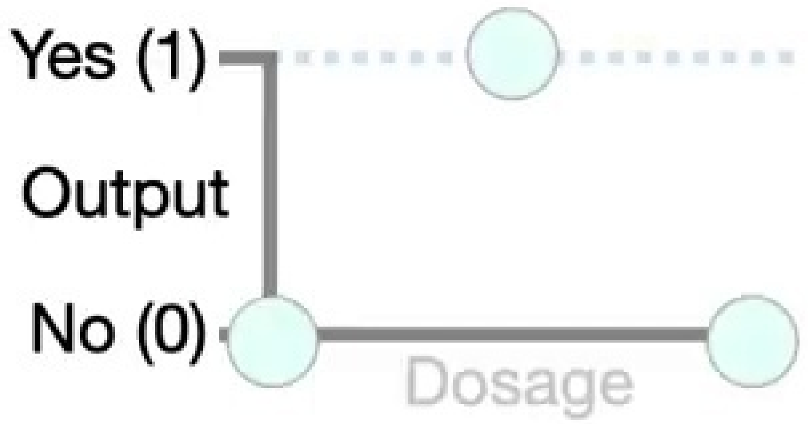


NOTE: Throughout this and the next **StatQuests**, I will make parameter values that have already been optimized green…

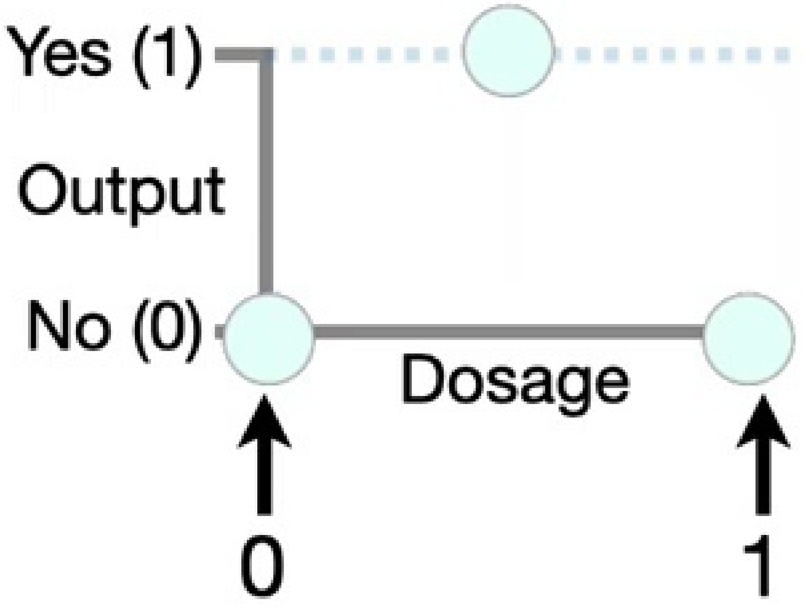
…and unoptimized parameters will be red.



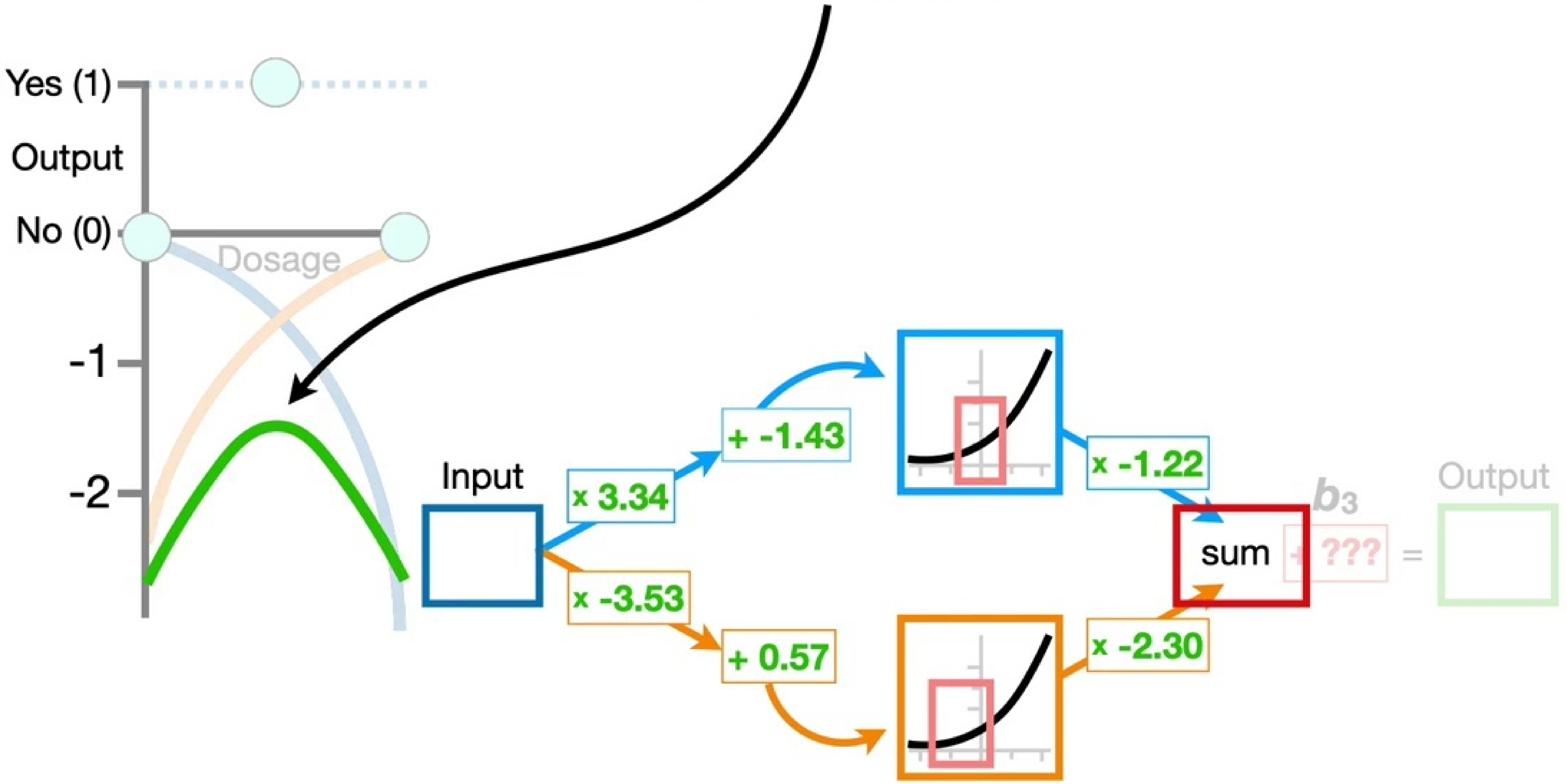
ALSO NOTE: To keep the math simple, let's assume **Dosages** go from **0** (low) to **1** (high).



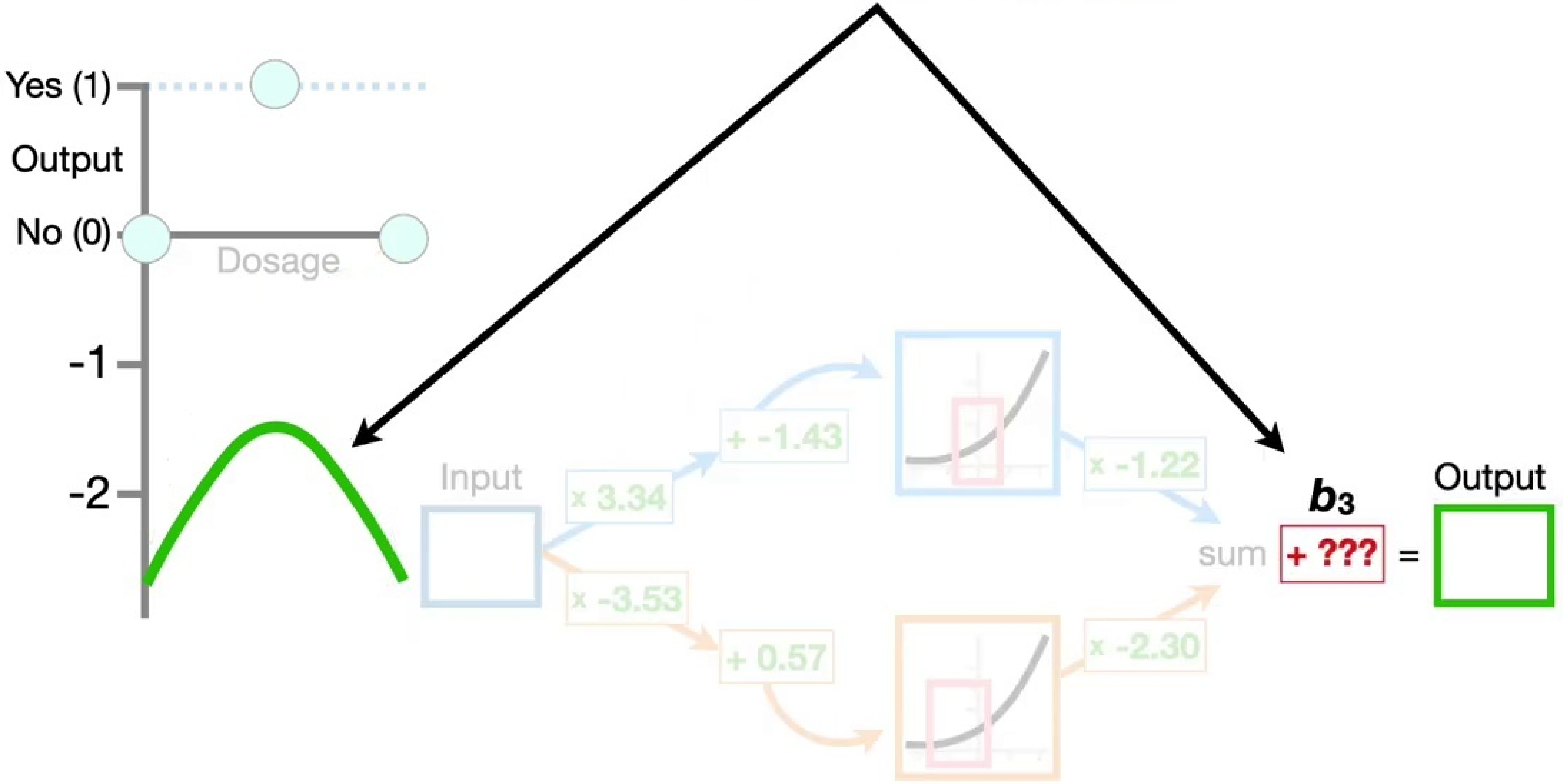
Now, if we run
Dosages from **0** to **1**…



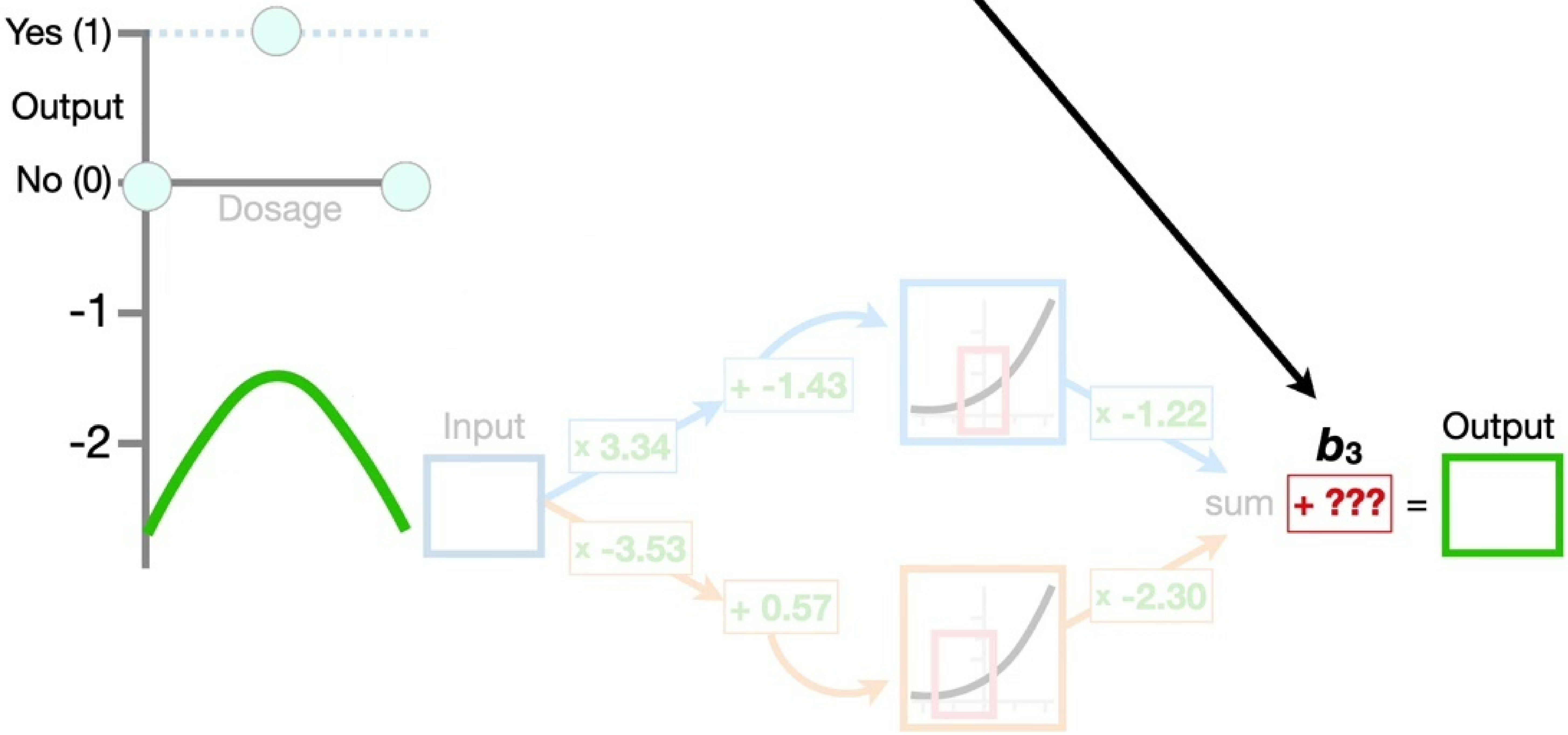
…to get this
green squiggle.



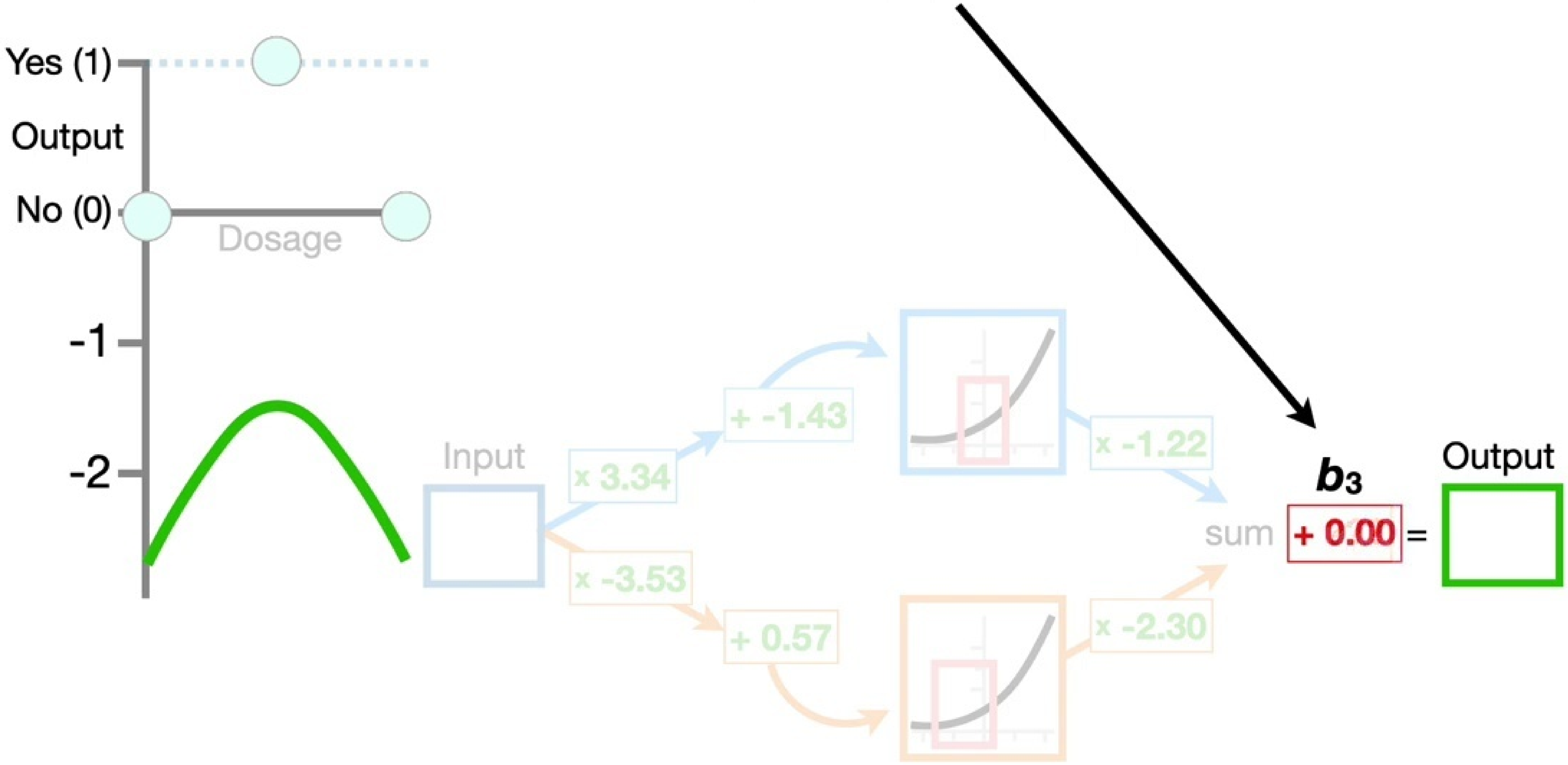
Now we are ready to add the final **Bias, b_3,** to the **green squiggle**.



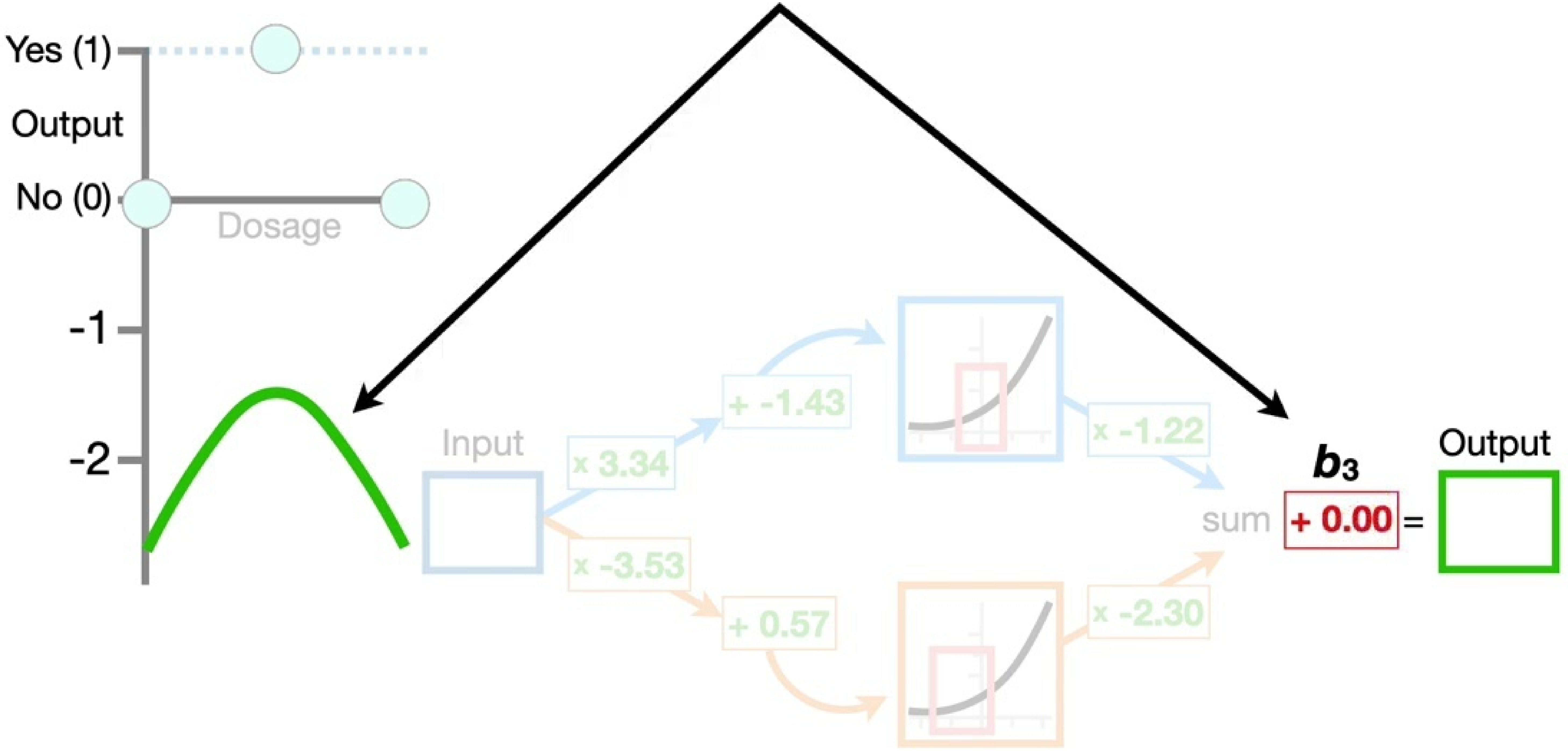
Because we don't yet know the optimal value for b_3, we have to give it an initial value…



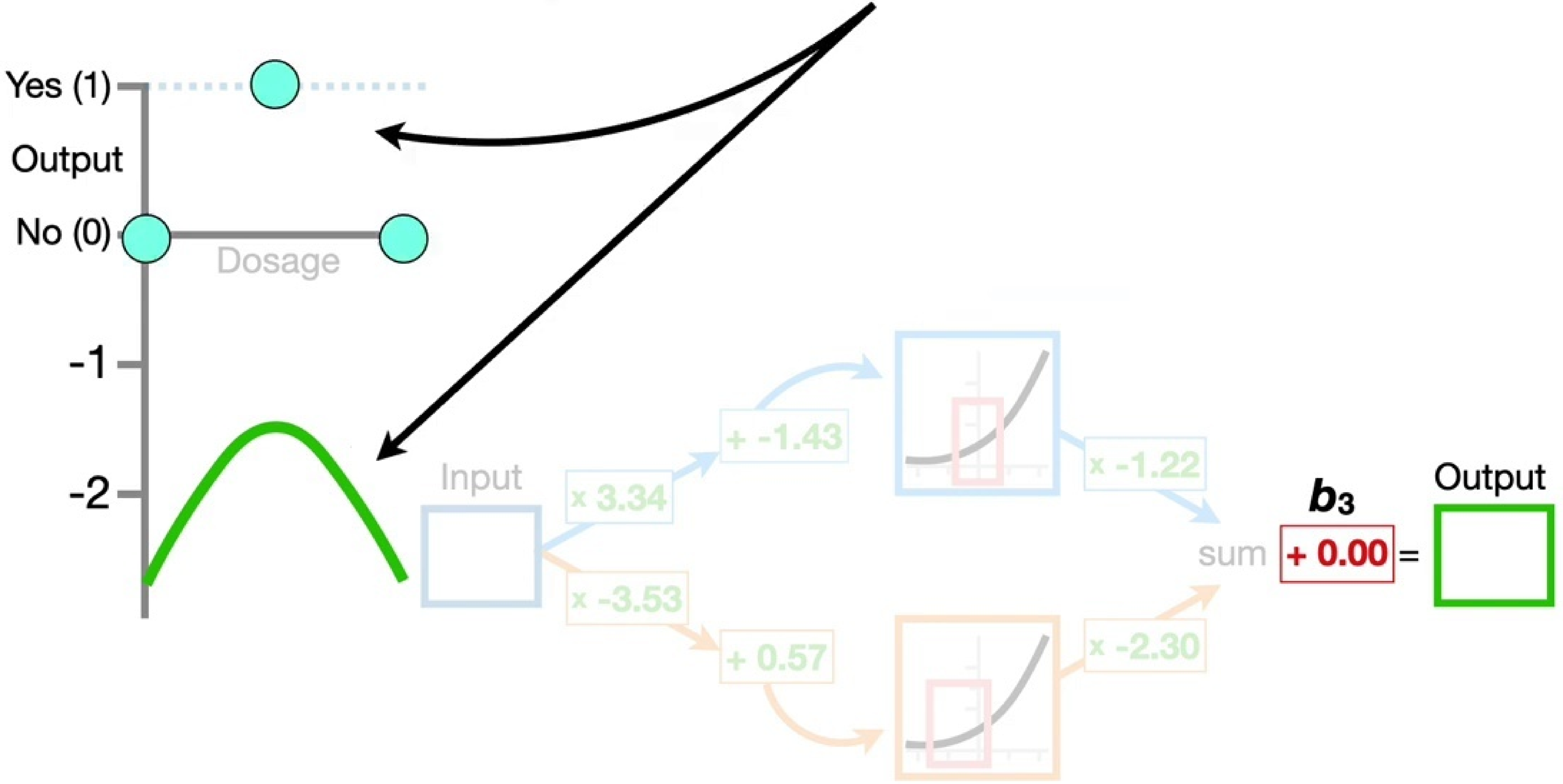
…and because **Bias** terms are frequently initialized to **0**, we will set $\boldsymbol{b_3} = \mathbf{0}$.



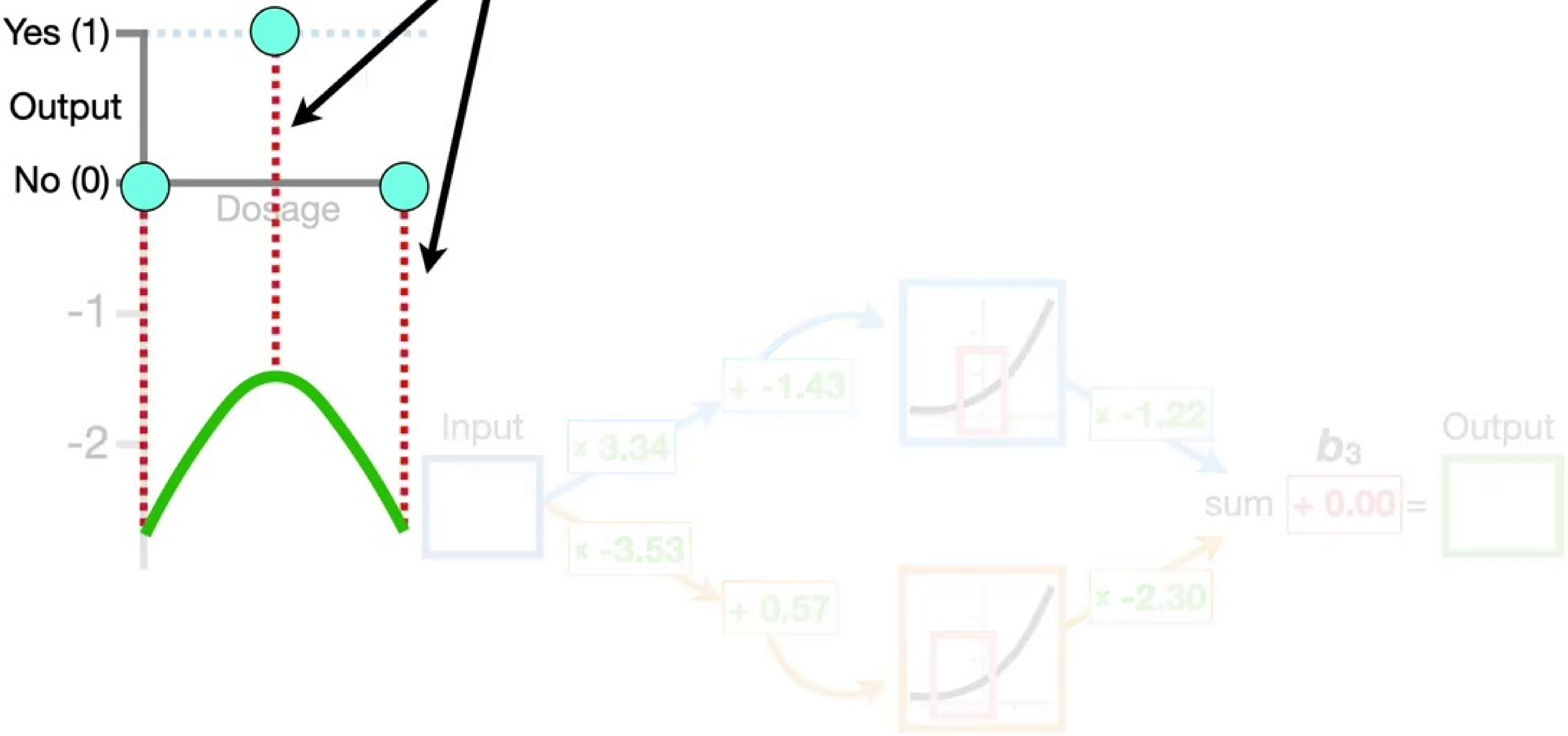
Now, adding **0** to all of the y-axis coordinates on the **green squiggle** leaves it right where it is.



However, that means the **green squiggle** is pretty far from the data that we observed.

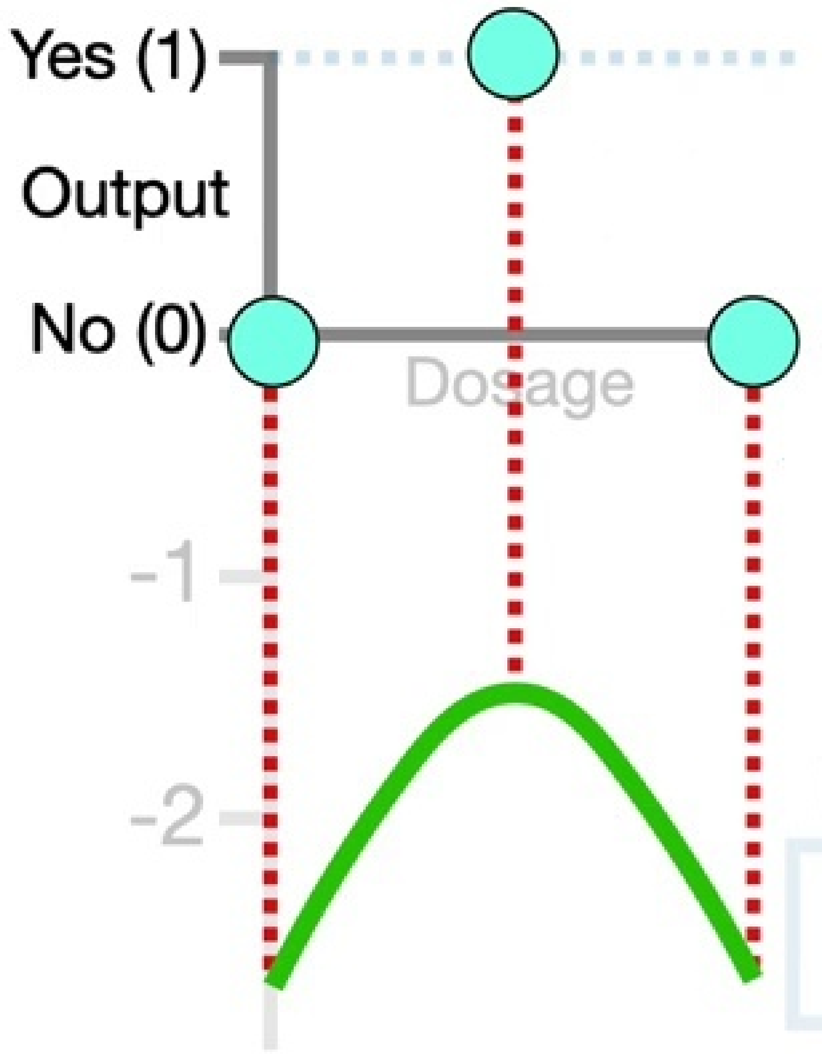


We can quantify how good the **green squiggle** fits the data by calculating the **Sum of the Squared Residuals.**

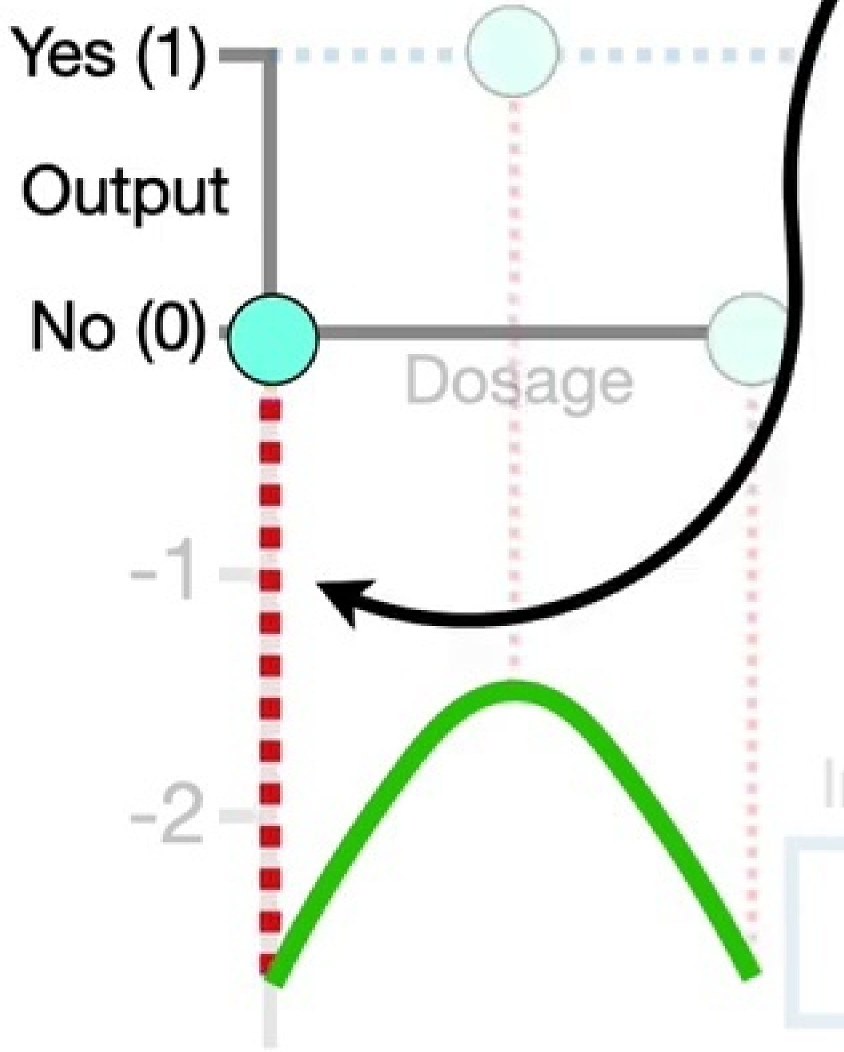


A **Residual** is the difference between the **Observed** and **Predicted** values.

Residual = (Observed - Predicted)



For example,
this **Residual**…

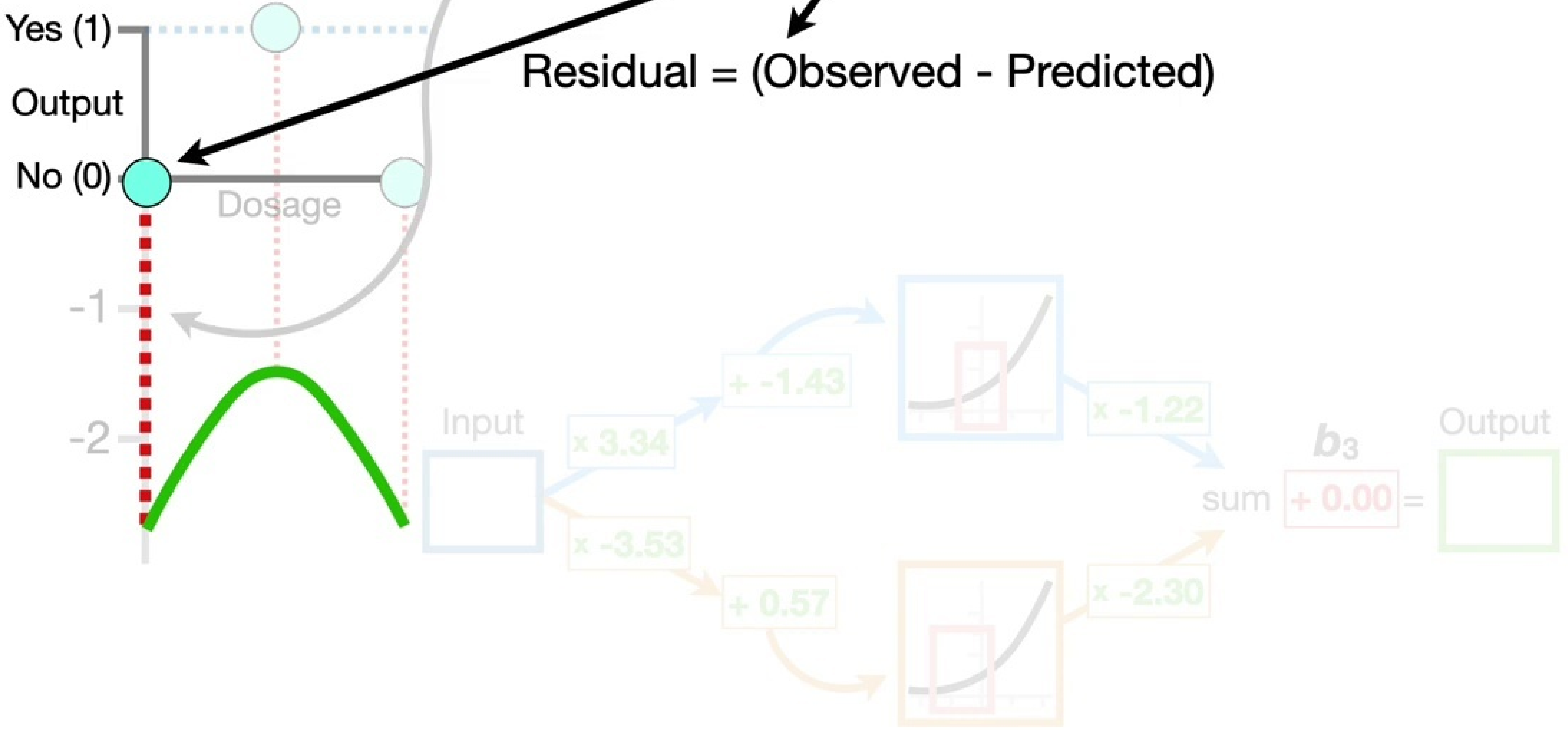


Residual = (Observed - Predicted)

For example, this **Residual**…

…is the **Observed** value, **0**…

Residual = (Observed - Predicted)

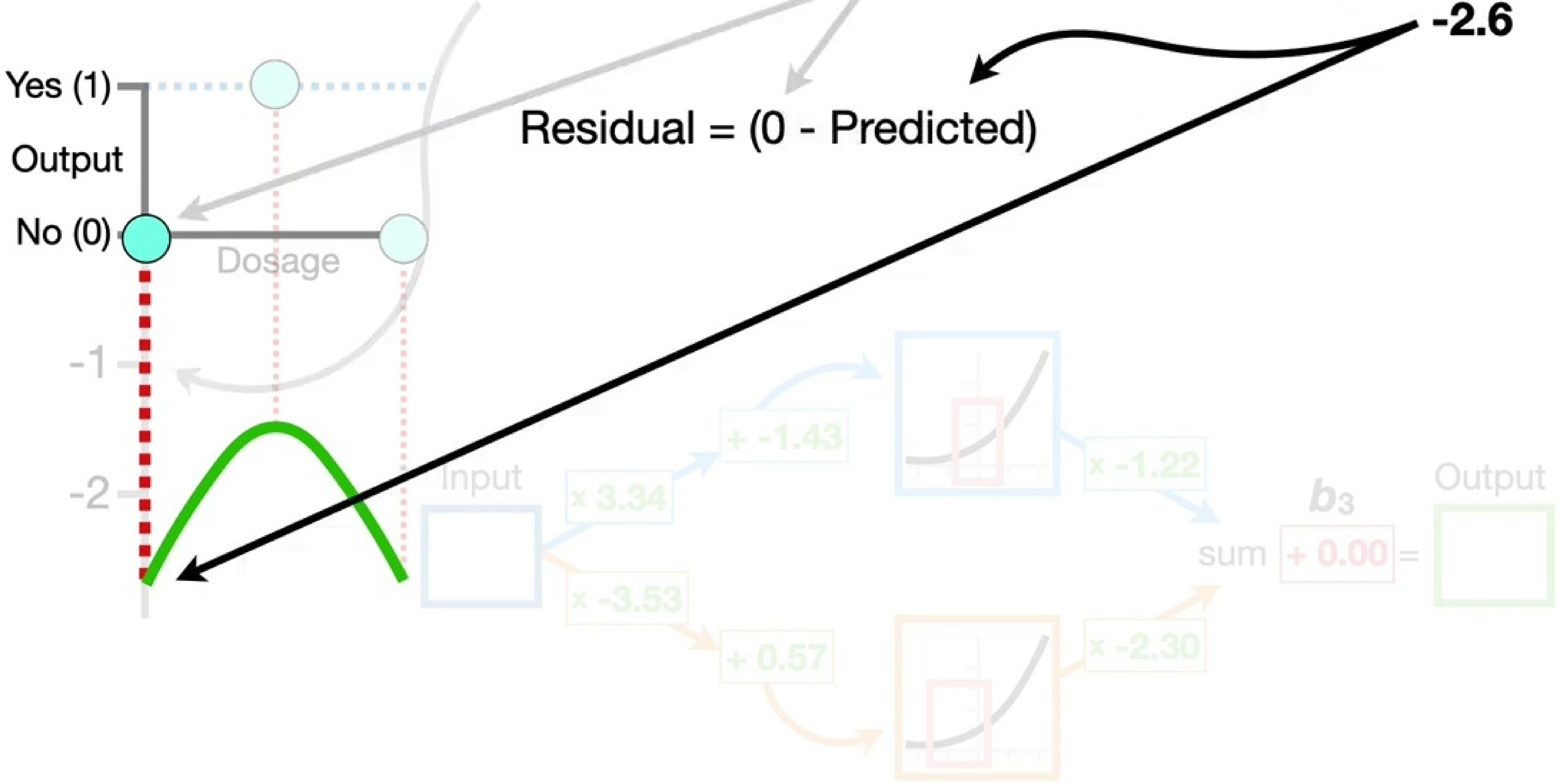


For example, this **Residual**…

…is the **Observed** value, **0**…

…minus the **Predicted** value from the **green squiggle**, **-2.6**

Residual = (0 - Predicted)

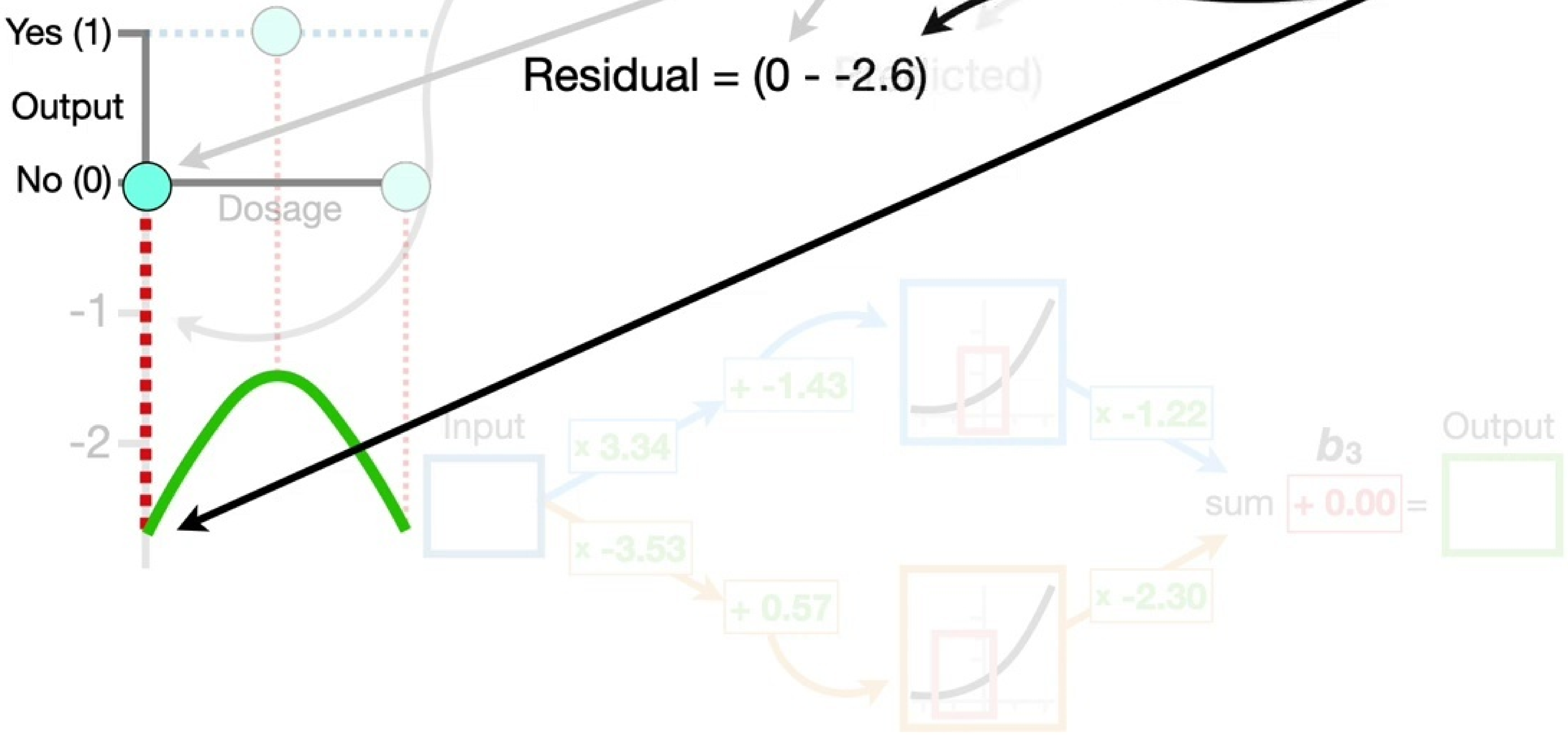


For example, this **Residual**…

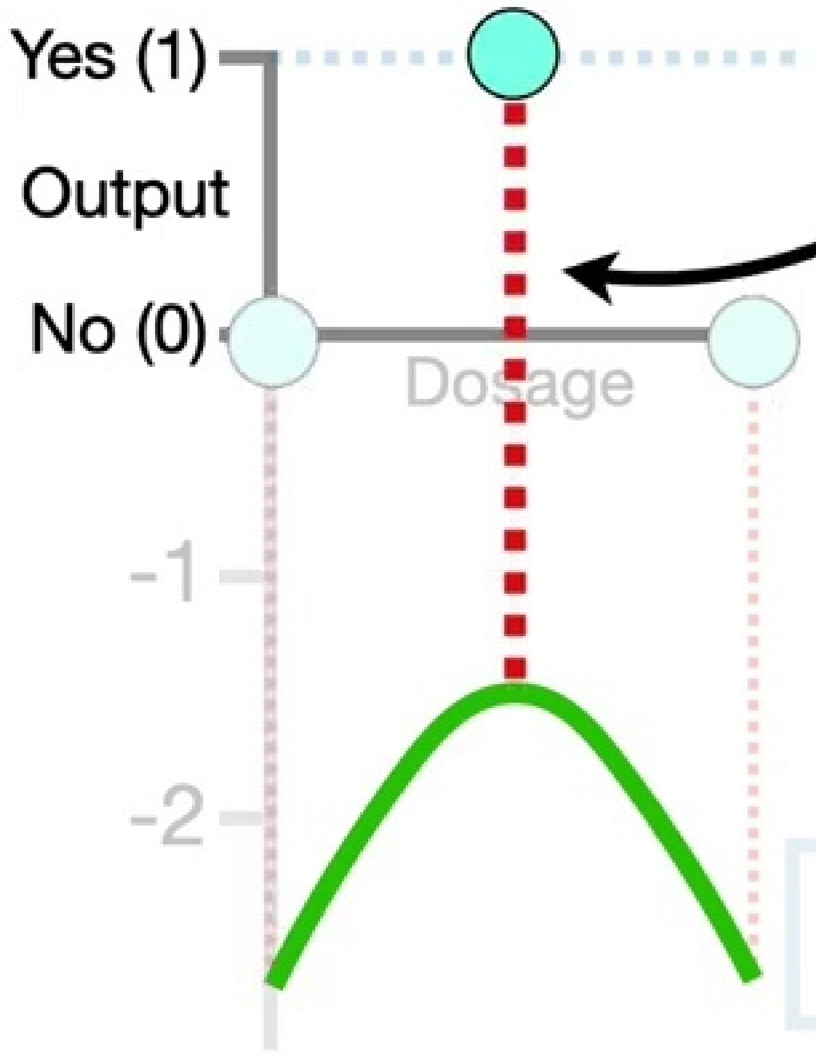
…is the **Observed** value, **0**…

…minus the **Predicted** value from the **green squiggle**, **-2.6**

Residual = (0 - -2.6)



This **Residual**…



(0 - -2.6)

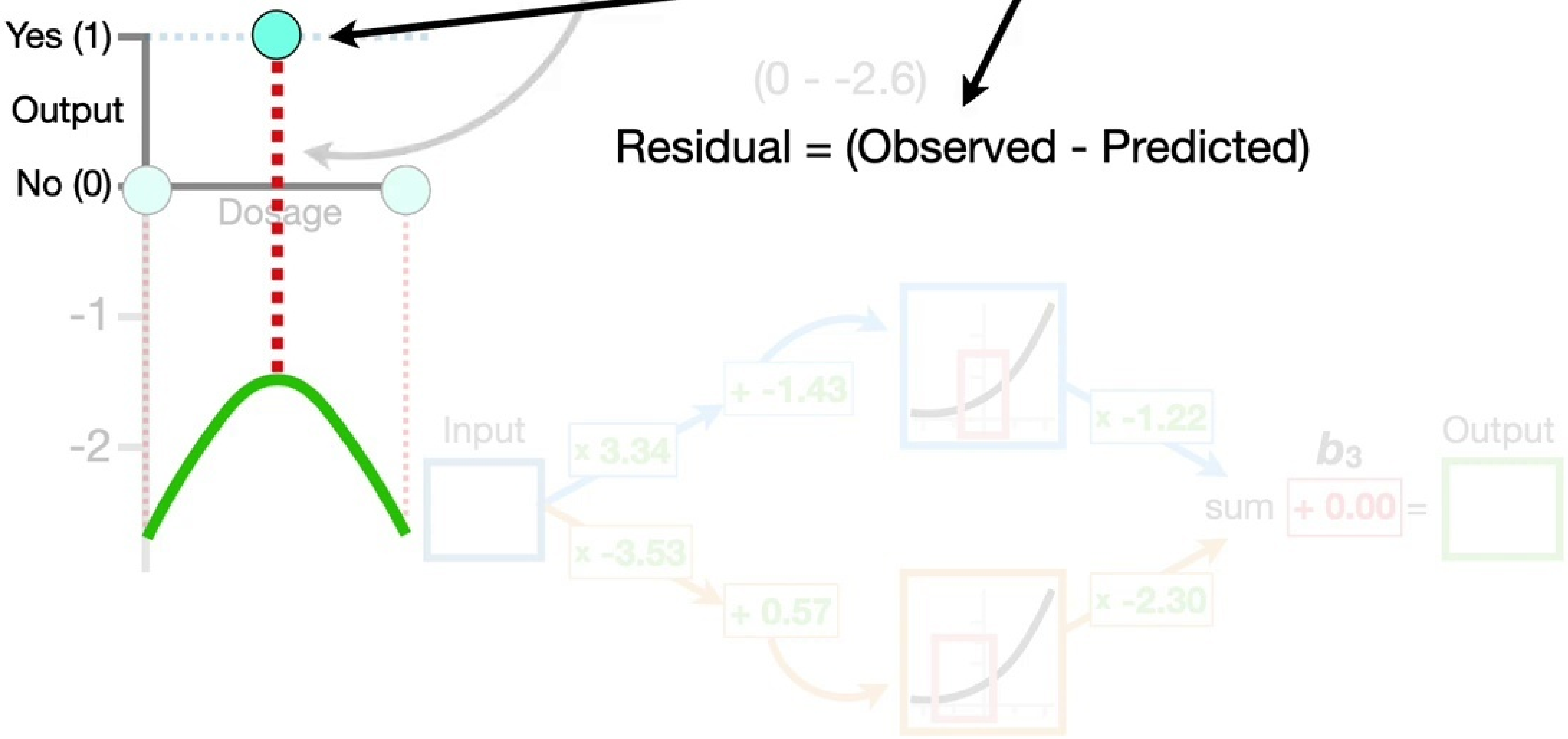
Residual = (Observed - Predicted)

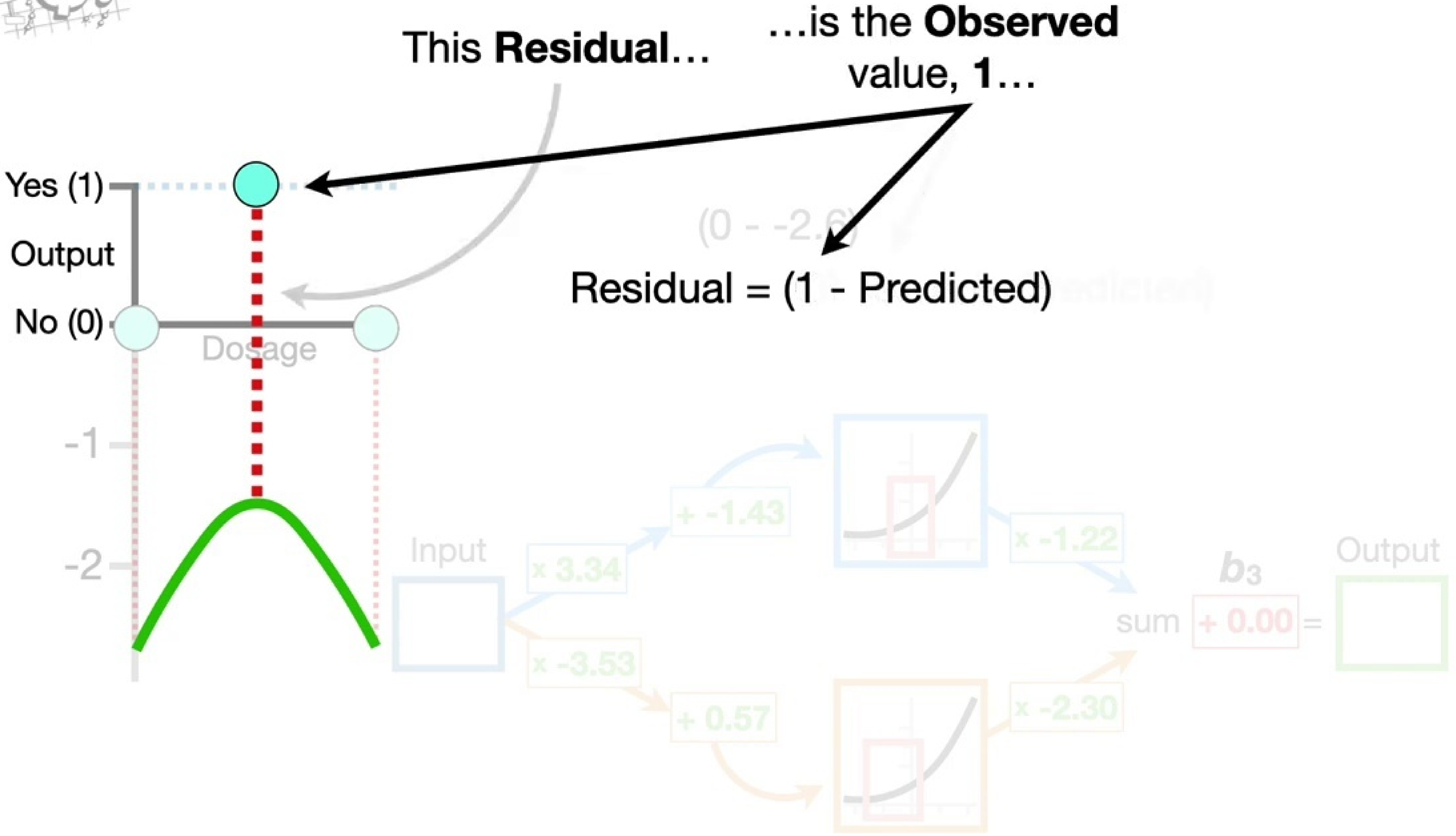
This **Residual**…

…is the **Observed** value, **1**…

(0 - -2.6)

Residual = (Observed - Predicted)



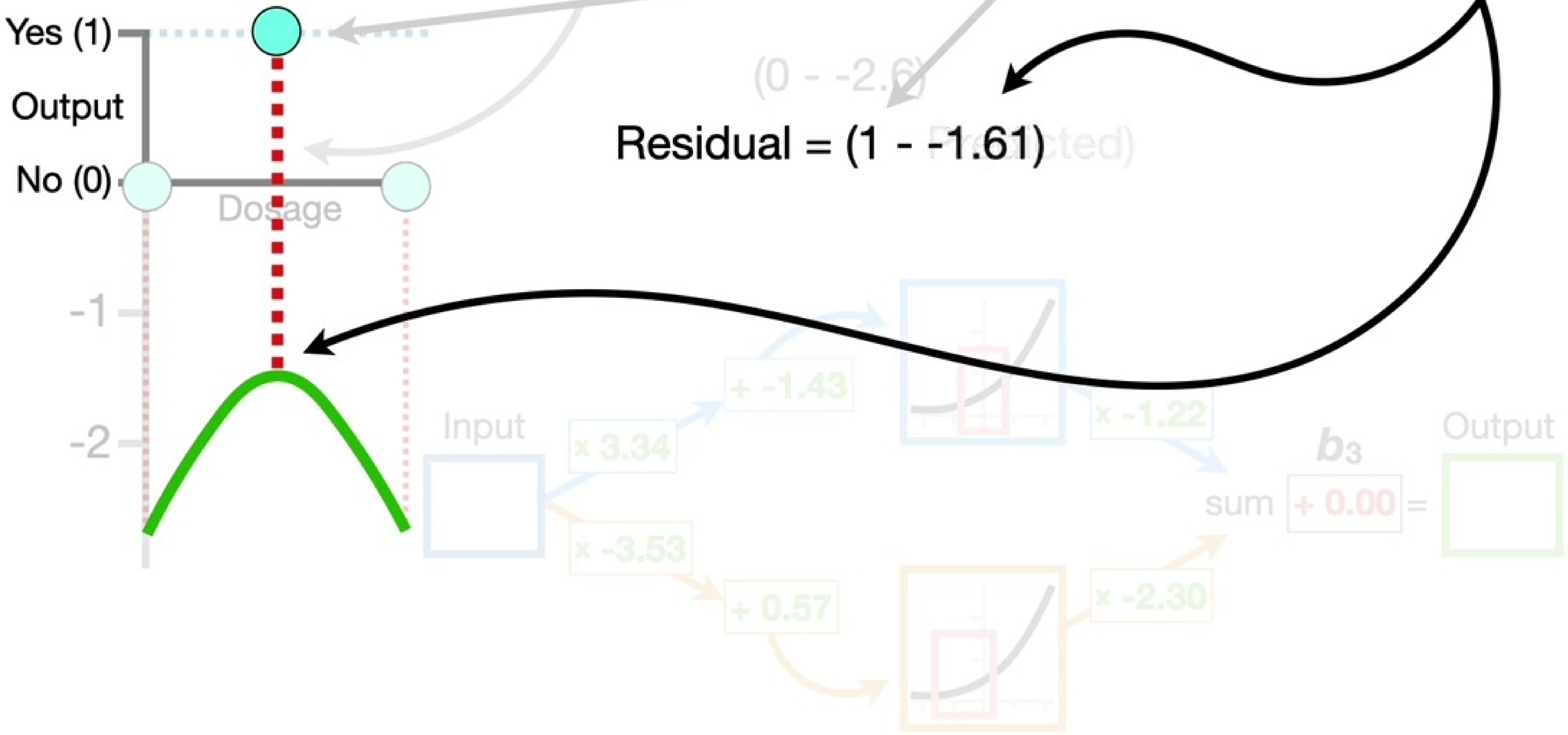

This **Residual**…
…is the **Observed** value, **1**…
Residual = (1 - Predicted)
Yes (1)
Output
No (0)
Dosage
-1
-2
Input
x 3.34
+ -1.43
x -1.22
x -3.53
+ 0.57
x -2.30
sum
b_3
+ 0.00 =
Output

This **Residual**…

…is the **Observed** value, **1**…

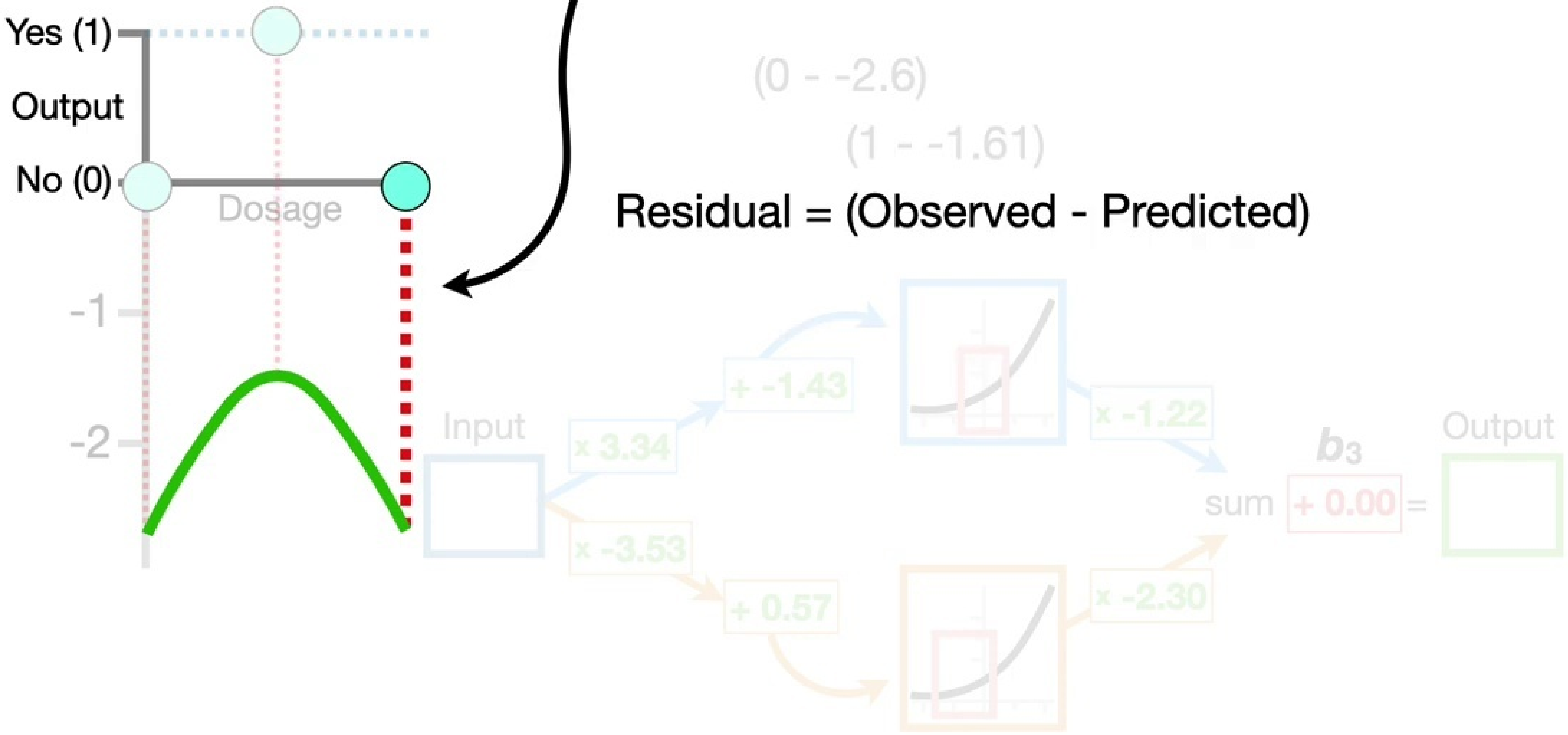
…minus the **Predicted** value from the green squiggle, **-1.61**

Residual = (1 - -1.61)



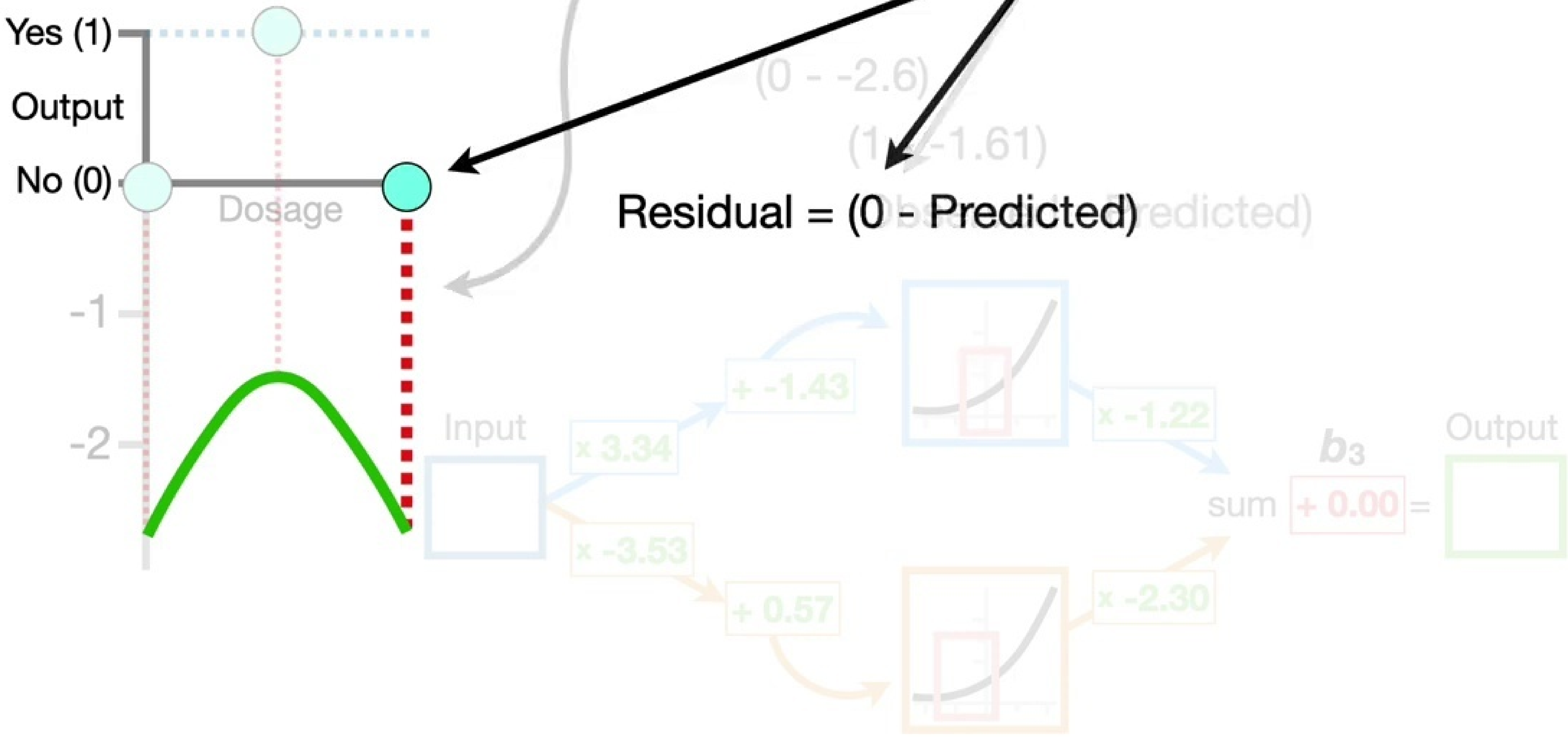
Lastly, this **Residual**…

Residual = (Observed - Predicted)

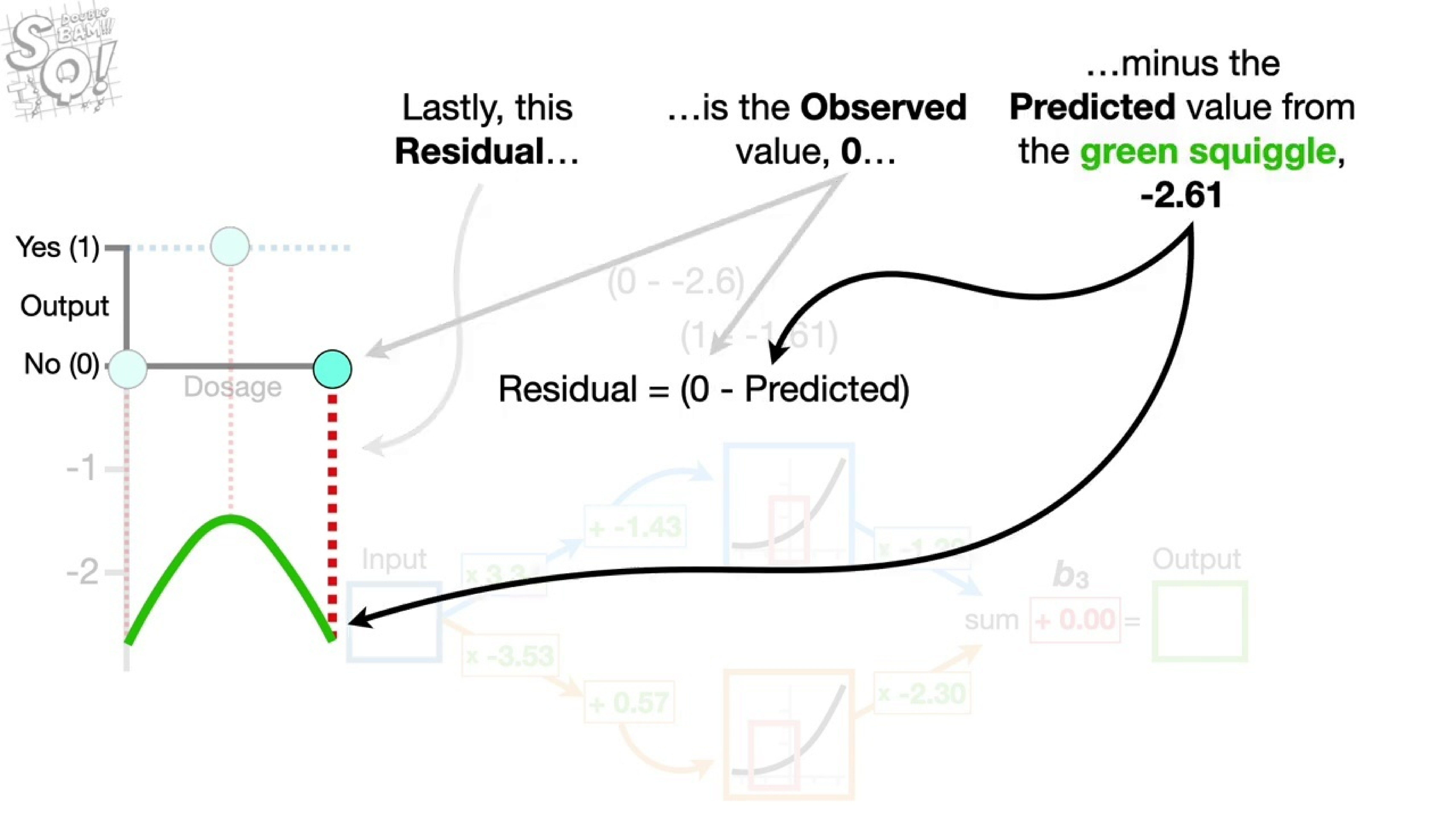


Lastly, this **Residual**…

…is the **Observed** value, **0**…



Residual = (0 - Predicted)

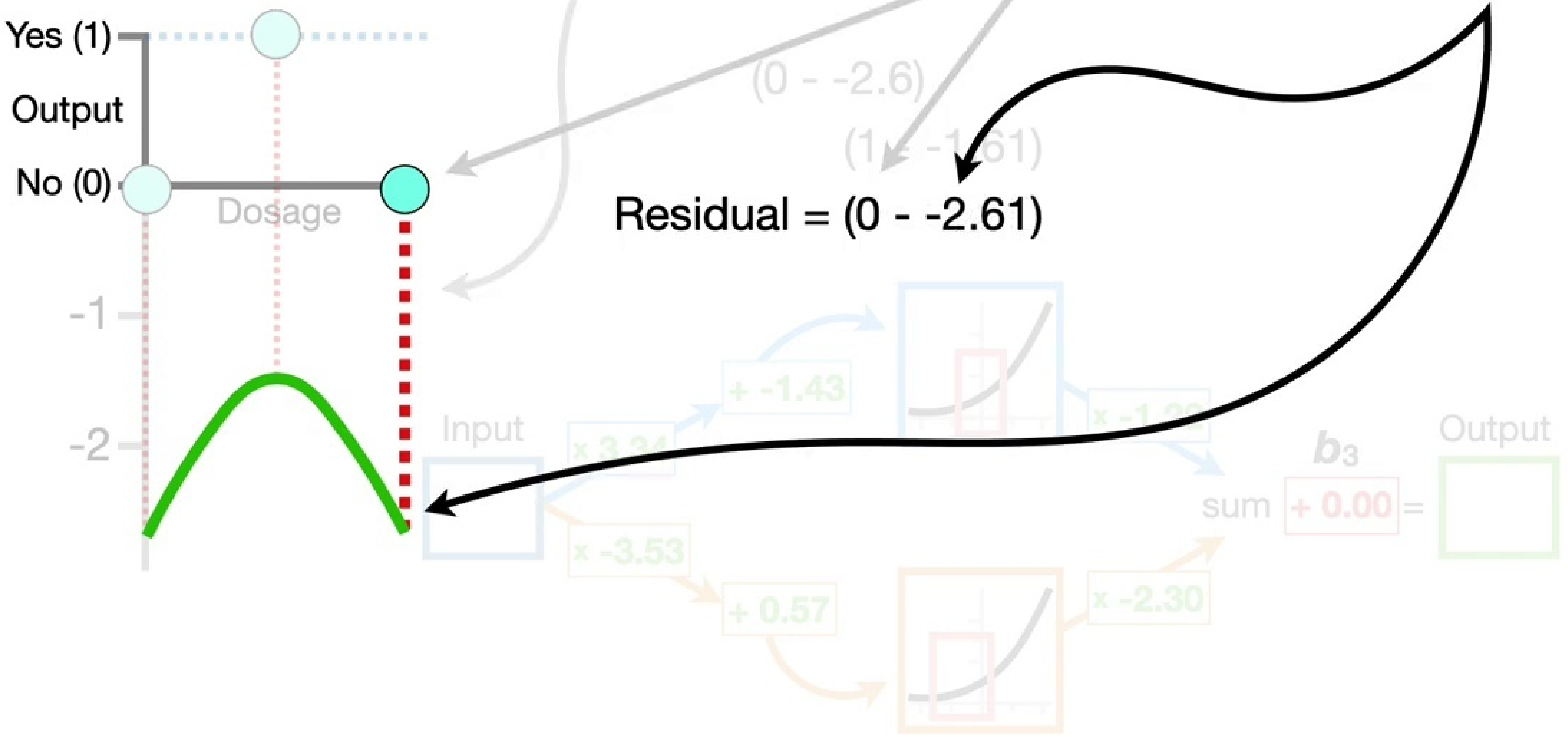

Lastly, this **Residual**…
…is the **Observed** value, **0**…
…minus the **Predicted** value from the **green squiggle**, **-2.61**
Yes (1)
Output
No (0)
Dosage
-1
-2
Residual = (0 - Predicted)
(0 - -2.6)
Input
+ -1.43
x -3.53
+ 0.57
x -2.30
sum
+ 0.00
=
Output

Lastly, this **Residual**…

…is the **Observed** value, **0**…

…minus the **Predicted** value from the **green squiggle**, **-2.61**

Residual = (0 - -2.61)

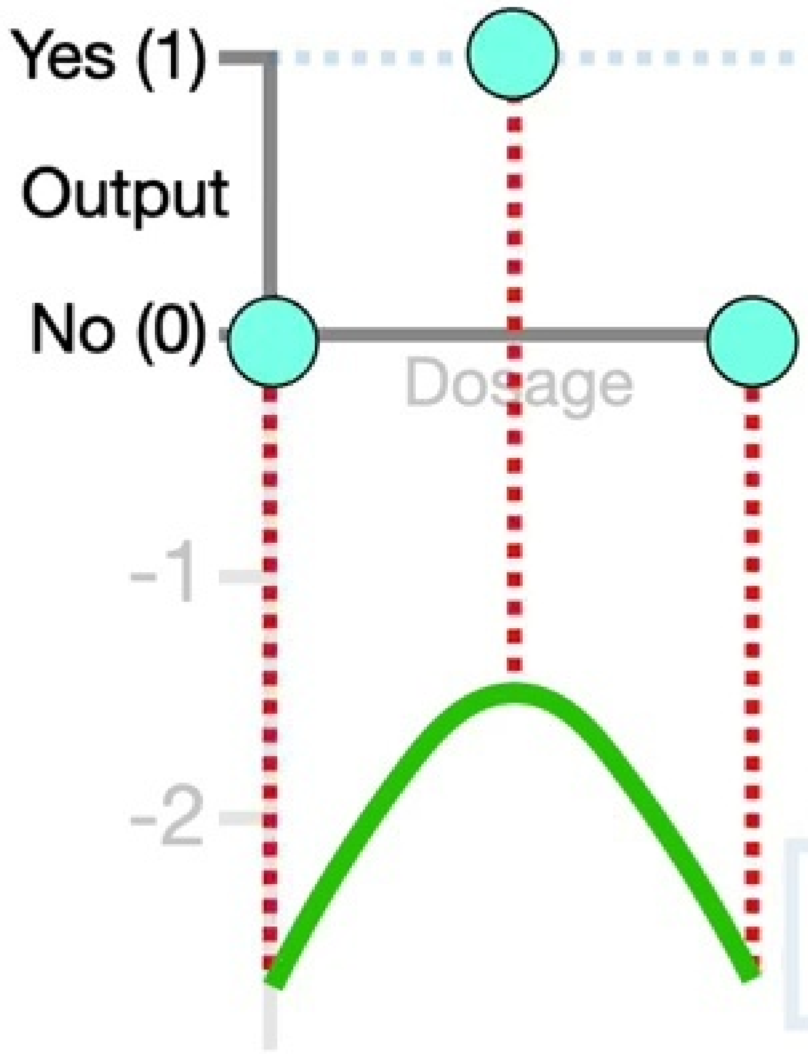


Now we square each **Residual**…

(0 - -2.6)

(1 - -1.61)

(0 - -2.61)

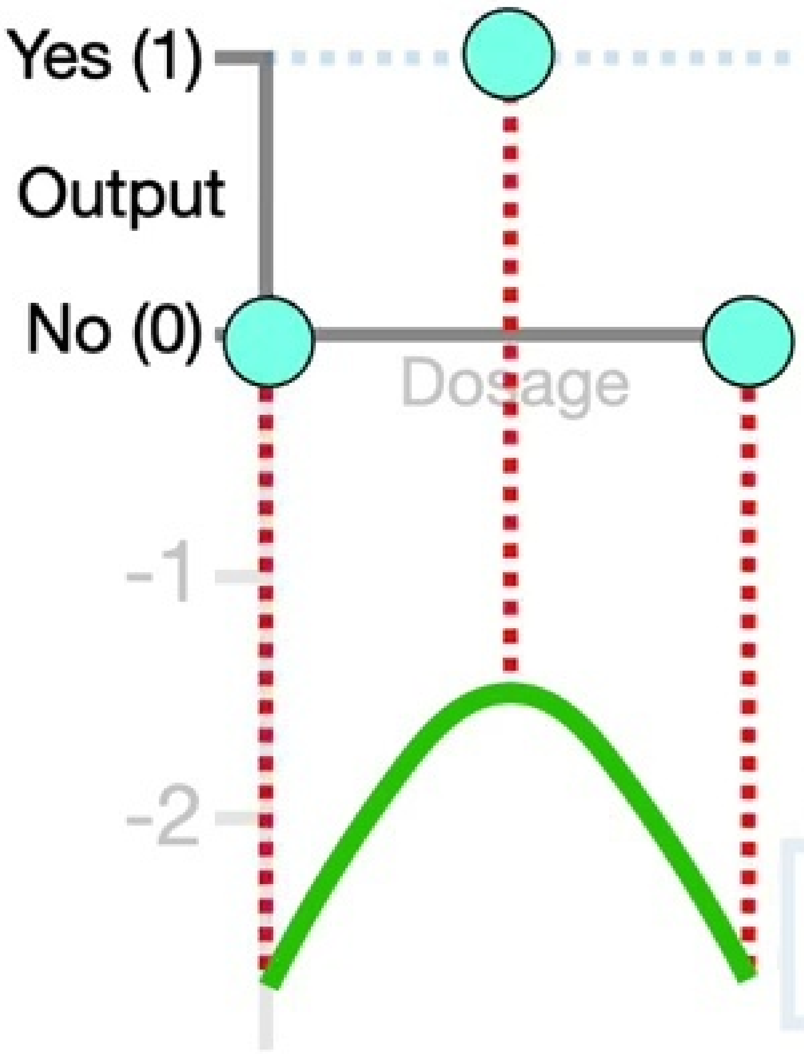


Now we square each **Residual**…

$(0 - -2.6)^2$

$(1 - -1.61)^2$

$(0 - -2.61)^2$

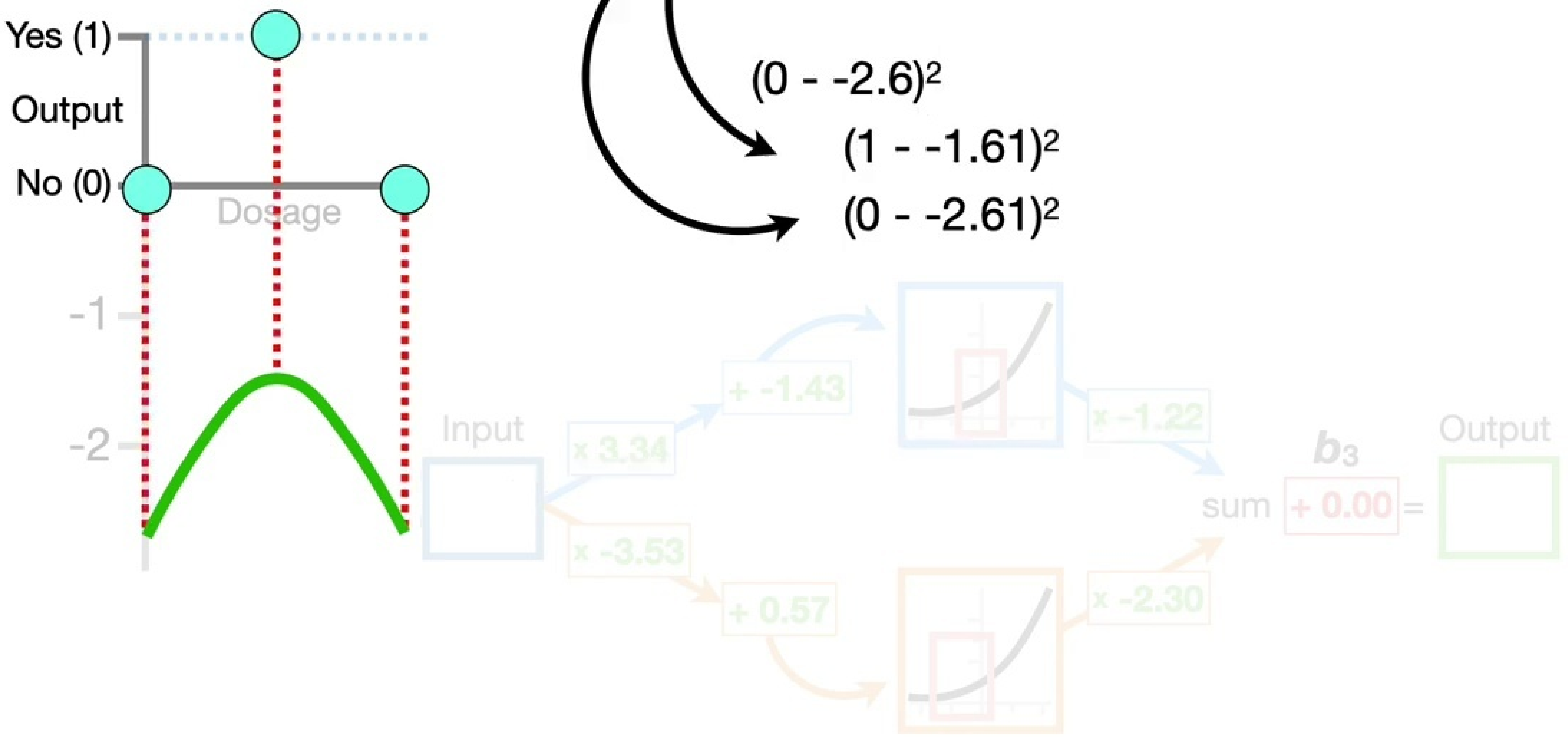


...and add them all together...

$(0 - -2.6)^2$

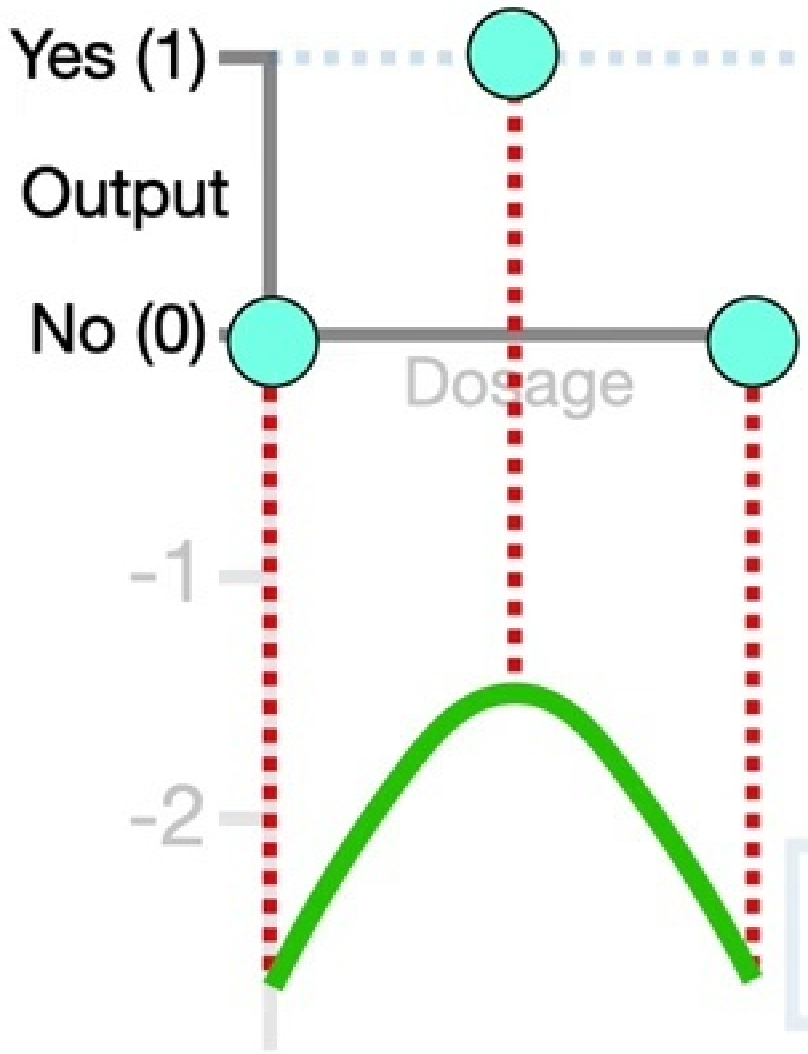
$(1 - -1.61)^2$

$(0 - -2.61)^2$



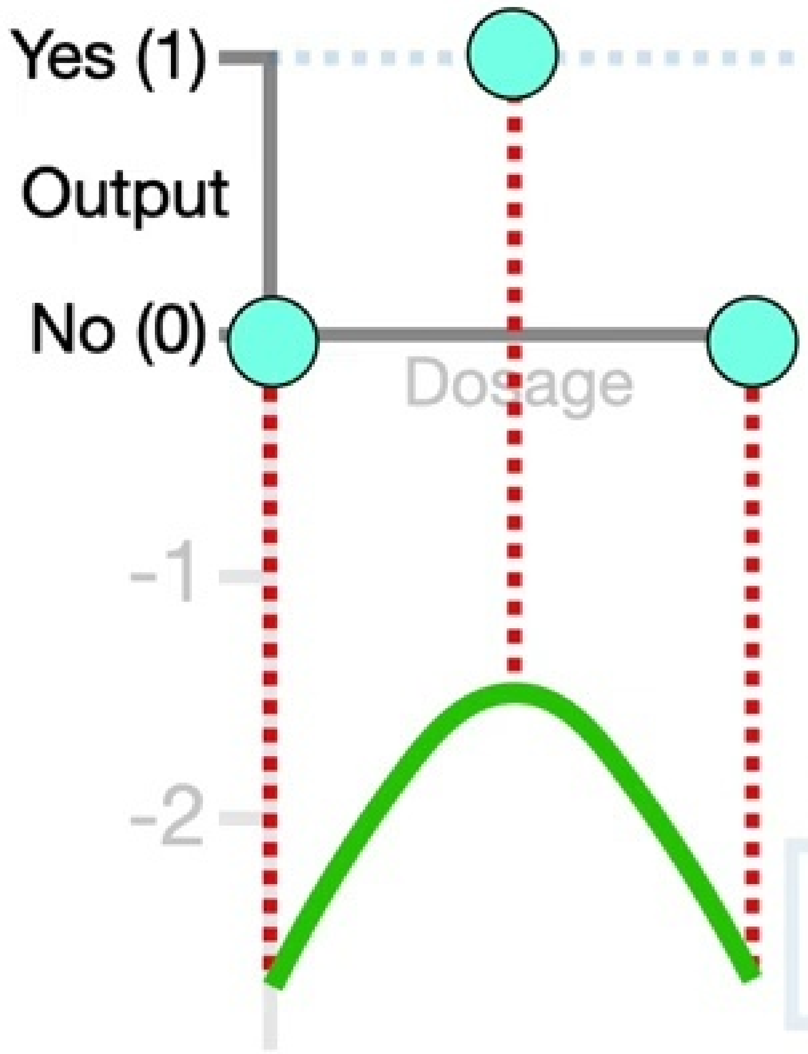
…and add them all together…

$$(0 - -2.6)^2$$
$$+ (1 - -1.61)^2$$
$$+ (0 - -2.61)^2$$

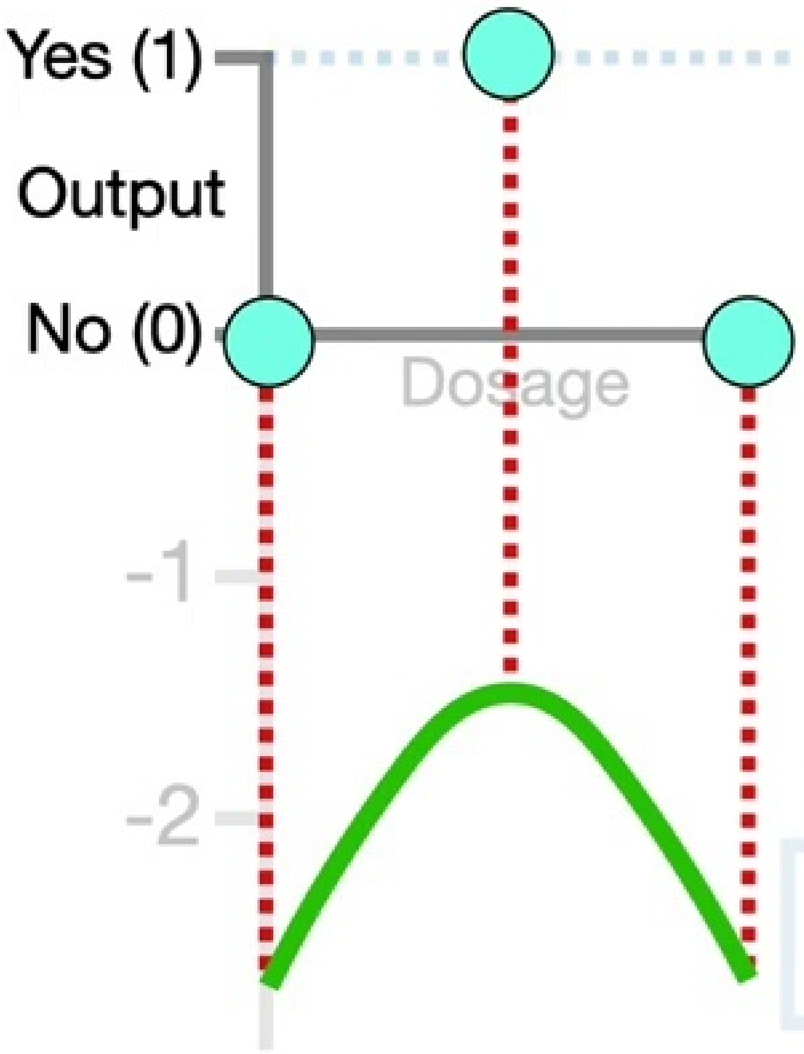


...to get **20.4** for the **Sum of the Squared Residuals (SSR)**.

$$SSR = (0 - -2.6)^2 + (1 - -1.61)^2 + (0 - -2.61)^2 \boxed{= 20.4}$$

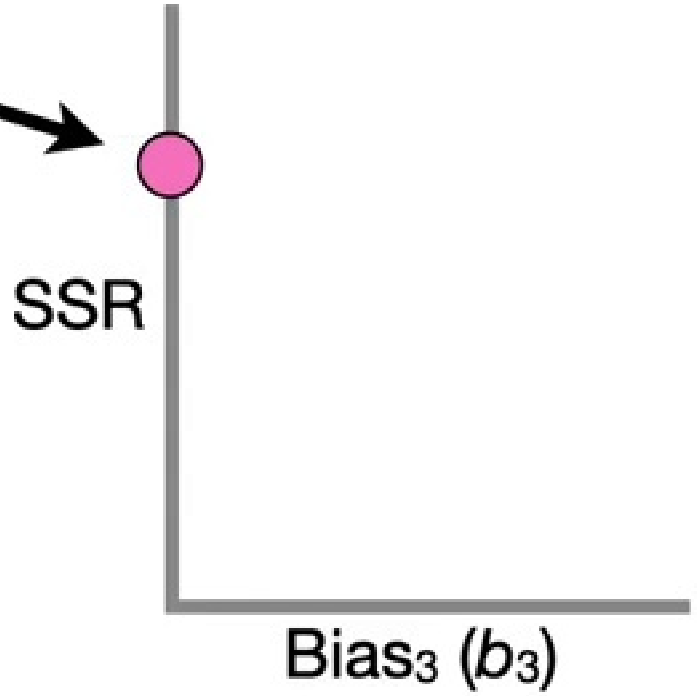


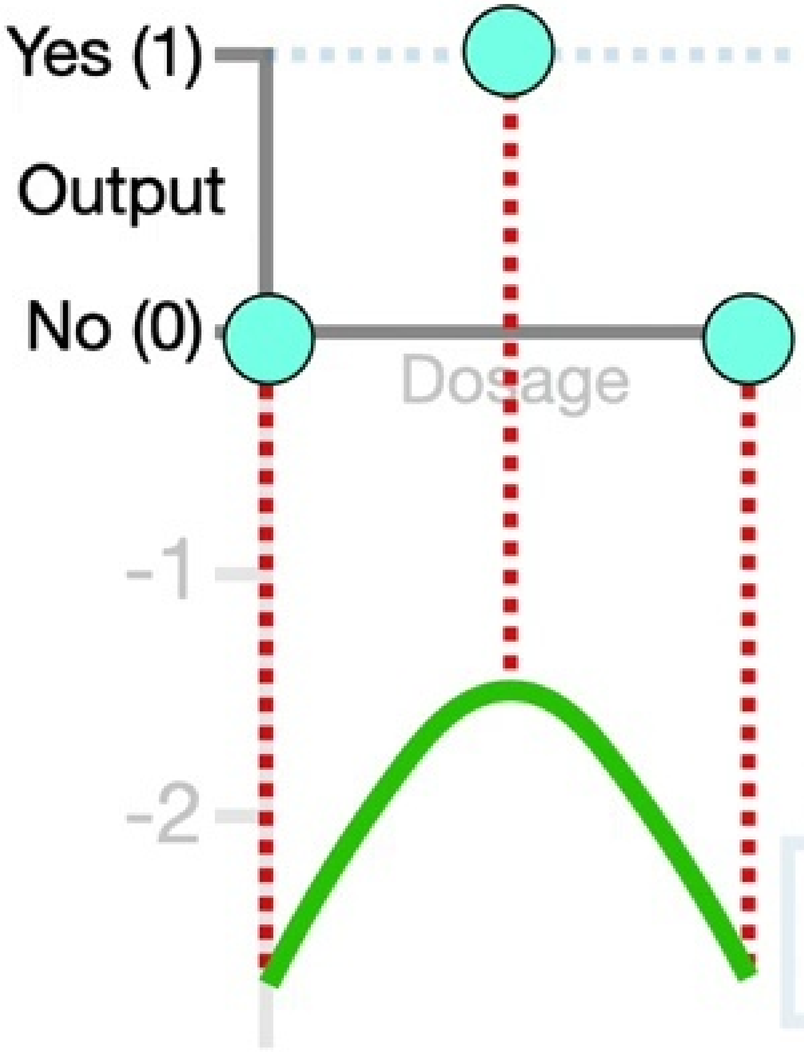
So, when $b_3 = 0$, the **SSR = 20.4**…



$$SSR = (0 - -2.6)^2 + (1 - -1.61)^2 + (0 - -2.61)^2 \boxed{= 20.4}$$

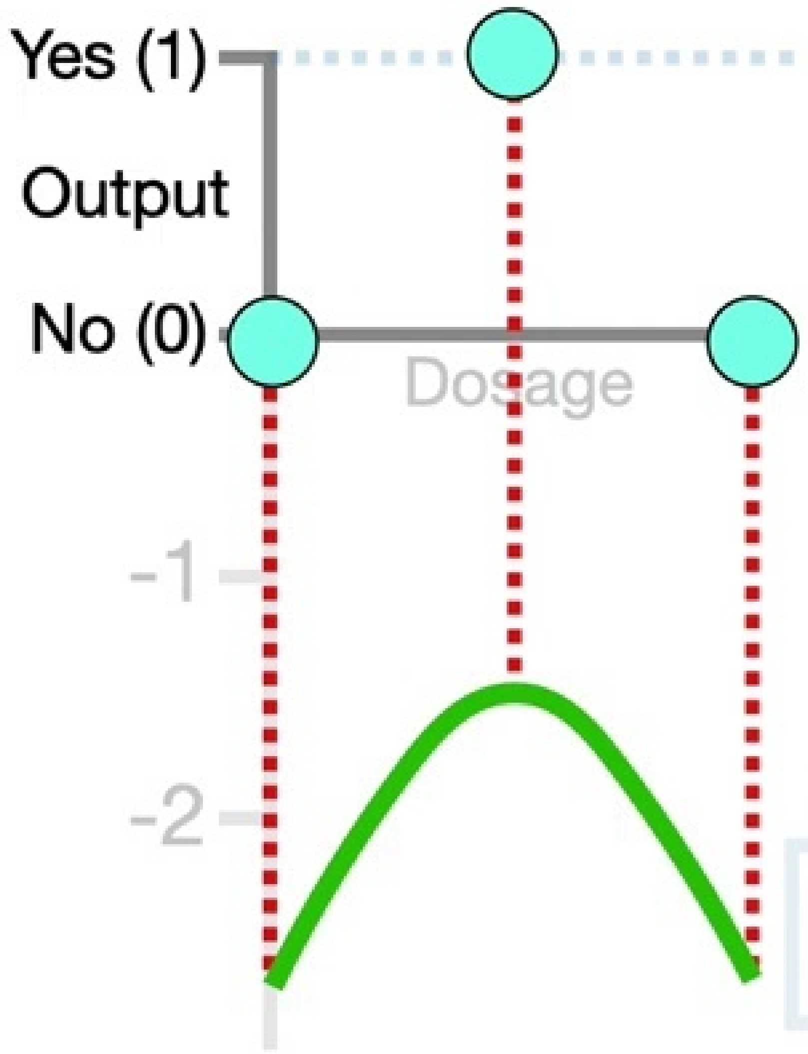
…and that corresponds to this location on this graph…



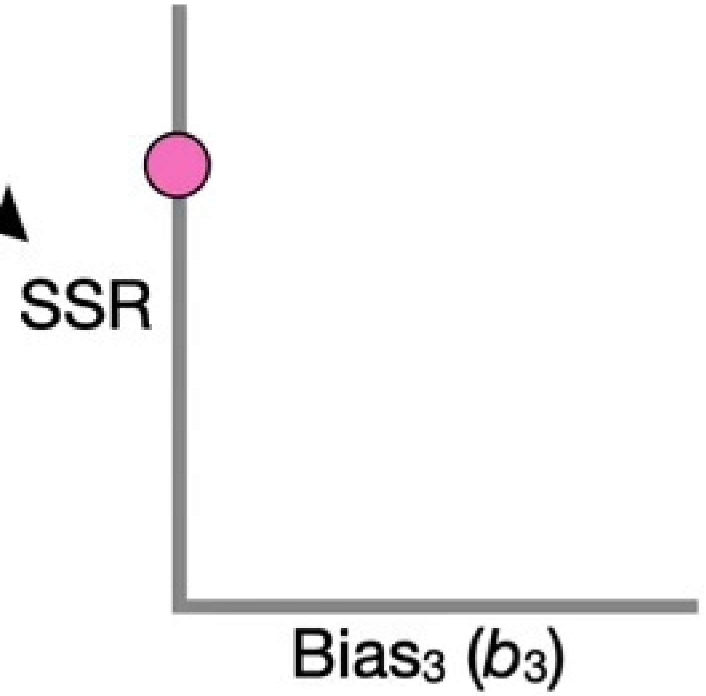


$$SSR = (0 - -2.6)^2 + (1 - -1.61)^2 + (0 - -2.61)^2 = 20.4$$

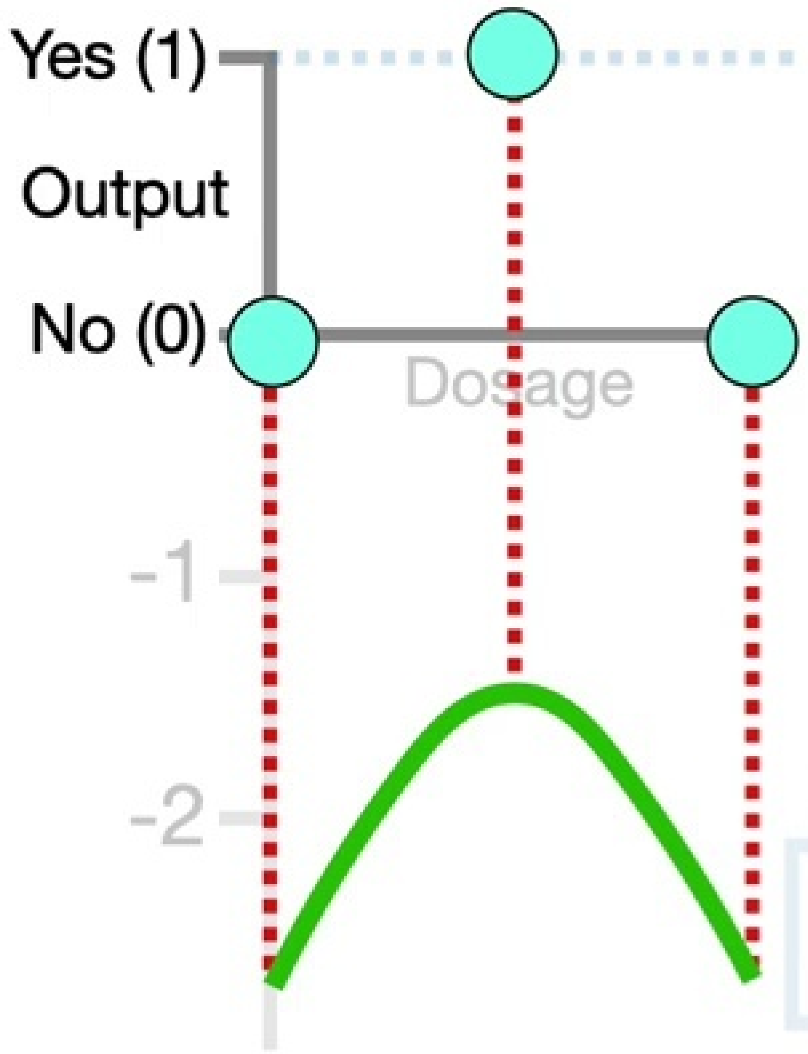
…that has the **SSR** on the y-axis…



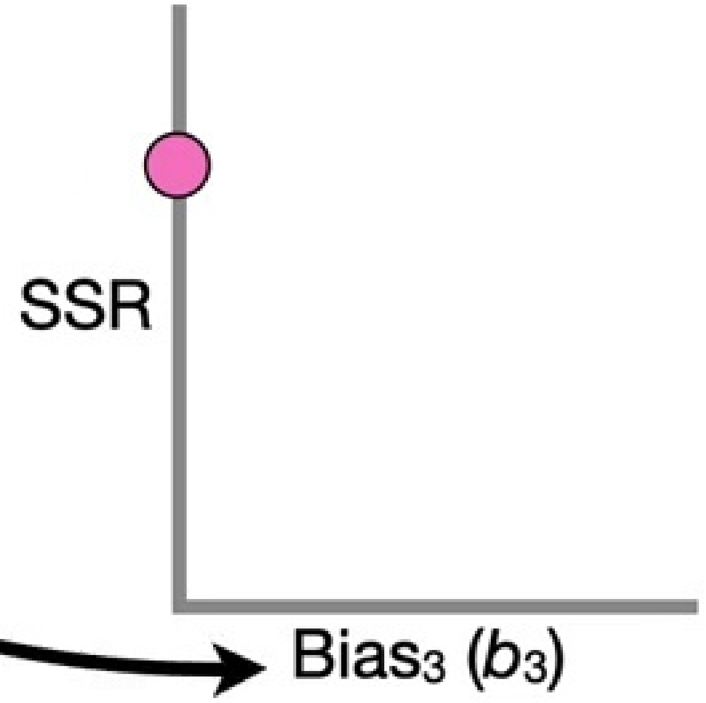
$$\text{SSR} = (0 - -2.6)^2 + (1 - -1.61)^2 + (0 - -2.61)^2 = 20.4$$



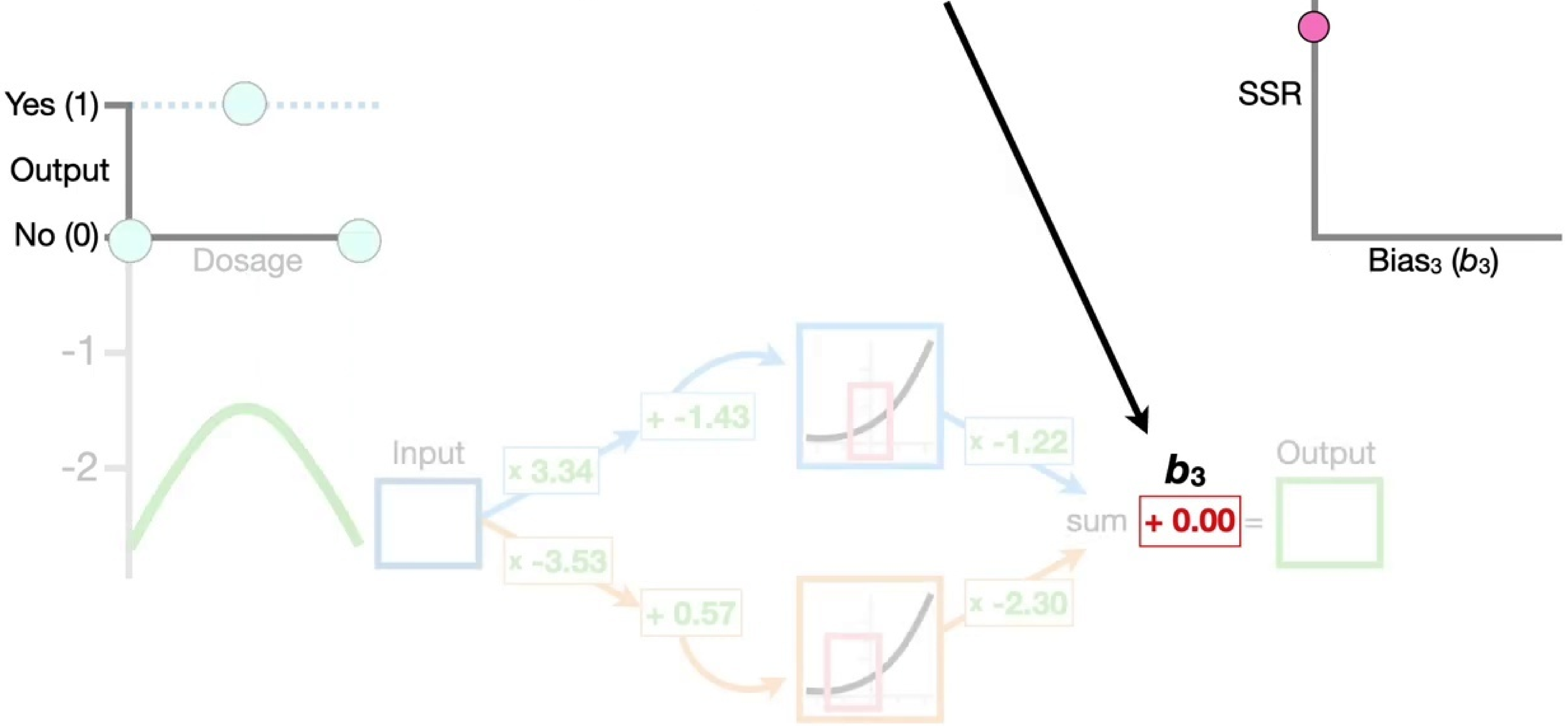
...and the **Bias, *b_3*,** on the x-axis.



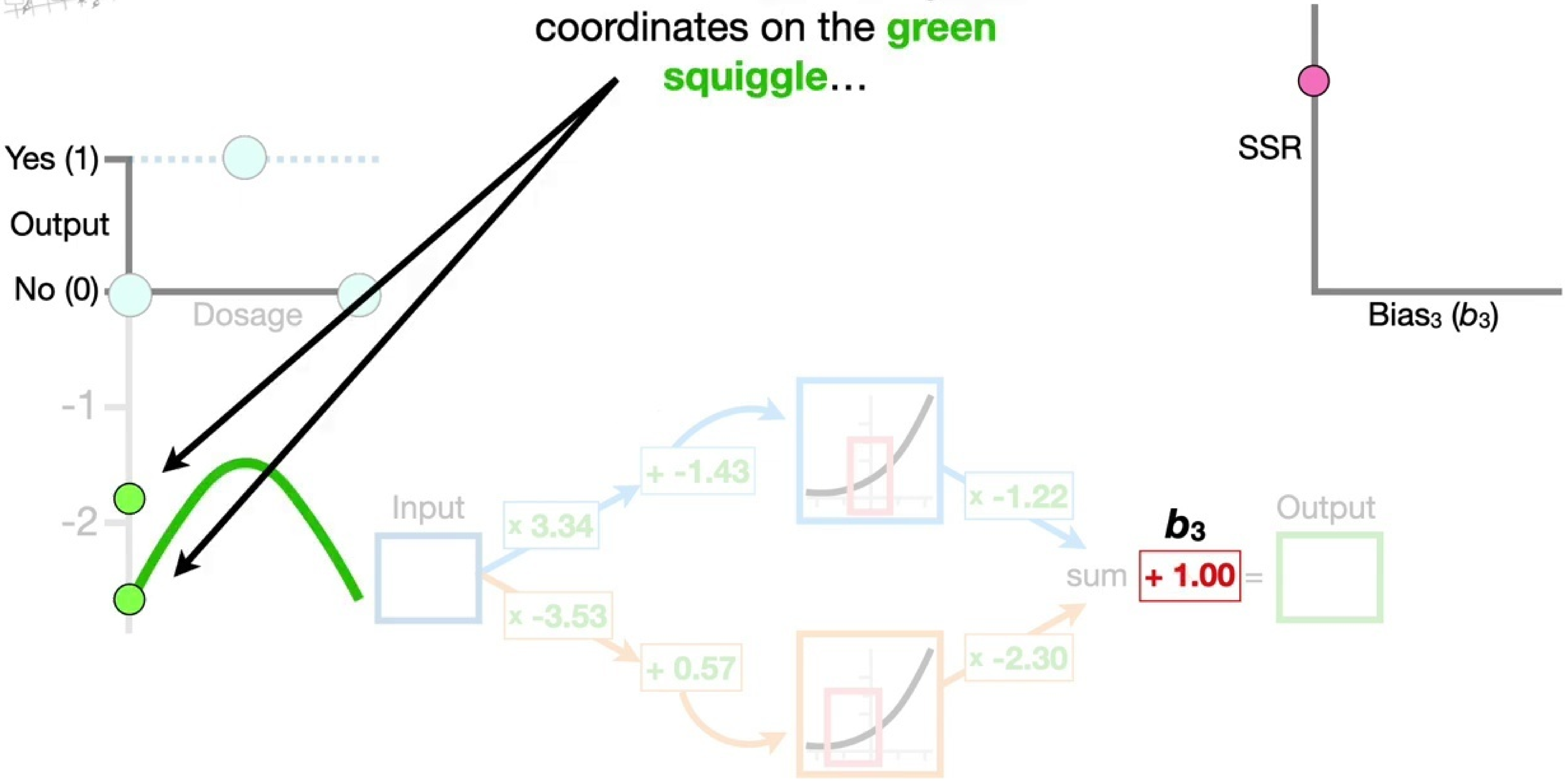
$$SSR = (0 - -2.6)^2 + (1 - -1.61)^2 + (0 - -2.61)^2 = 20.4$$



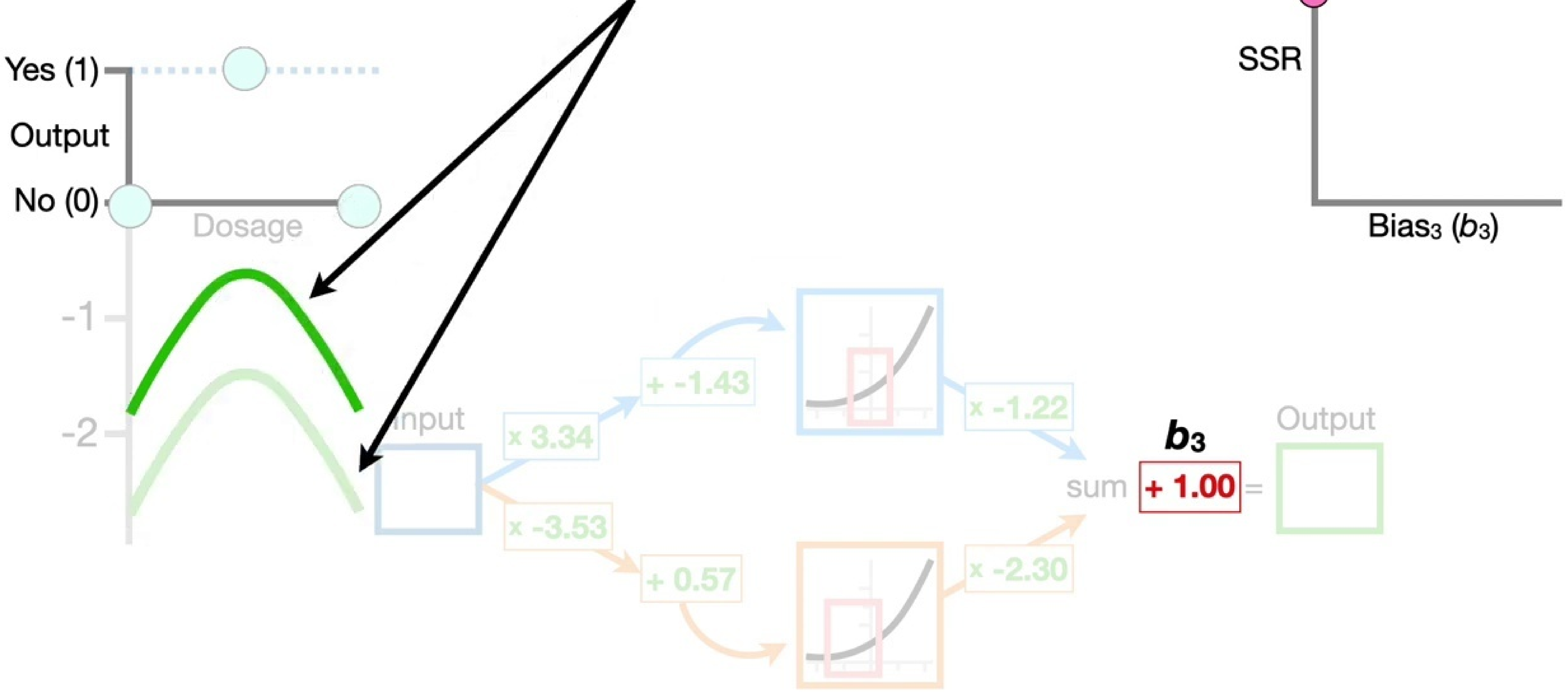
Now, if we increase b_3 to 1…



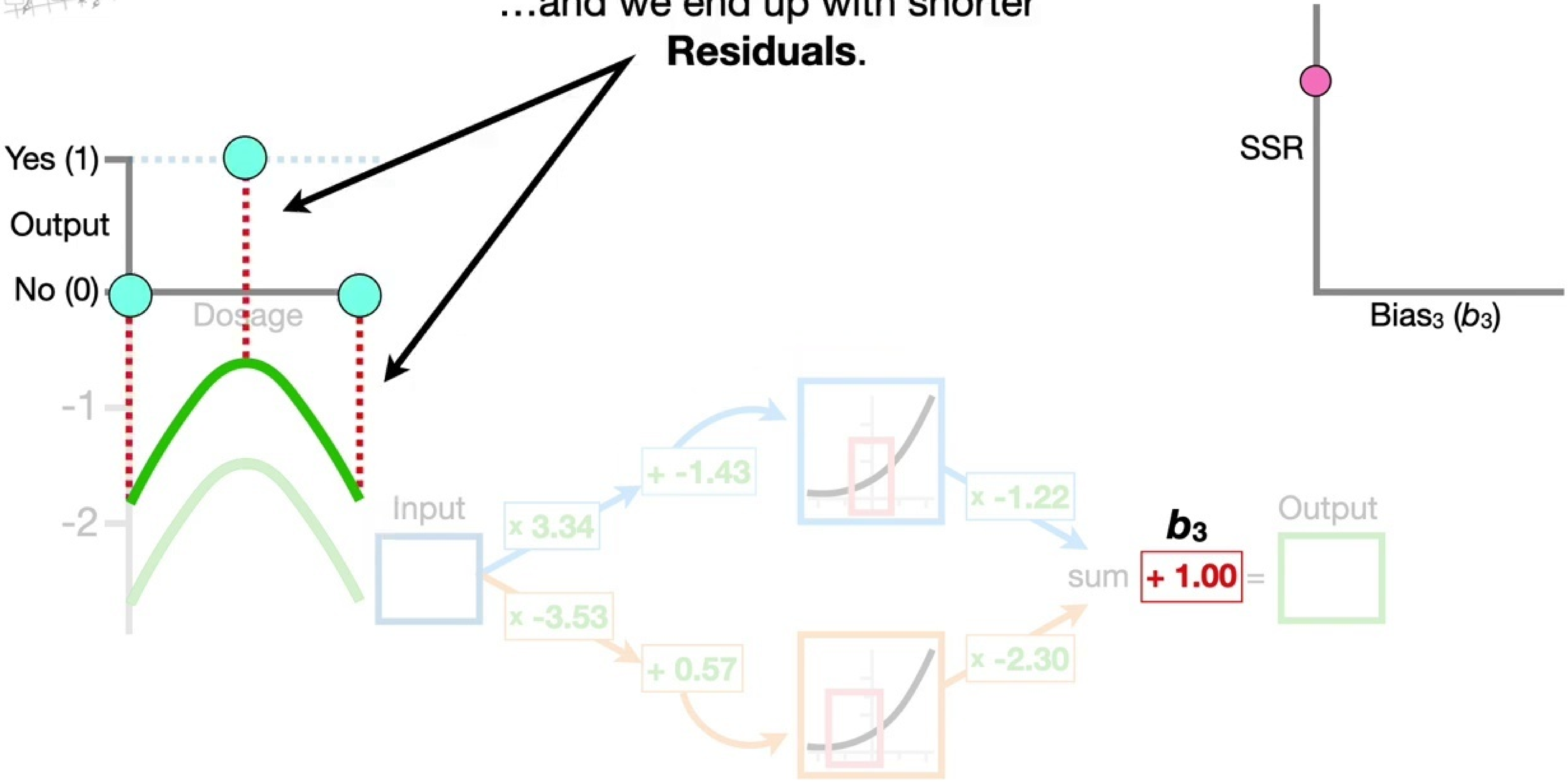
…then we add **1** to the y-axis coordinates on the **green squiggle**…



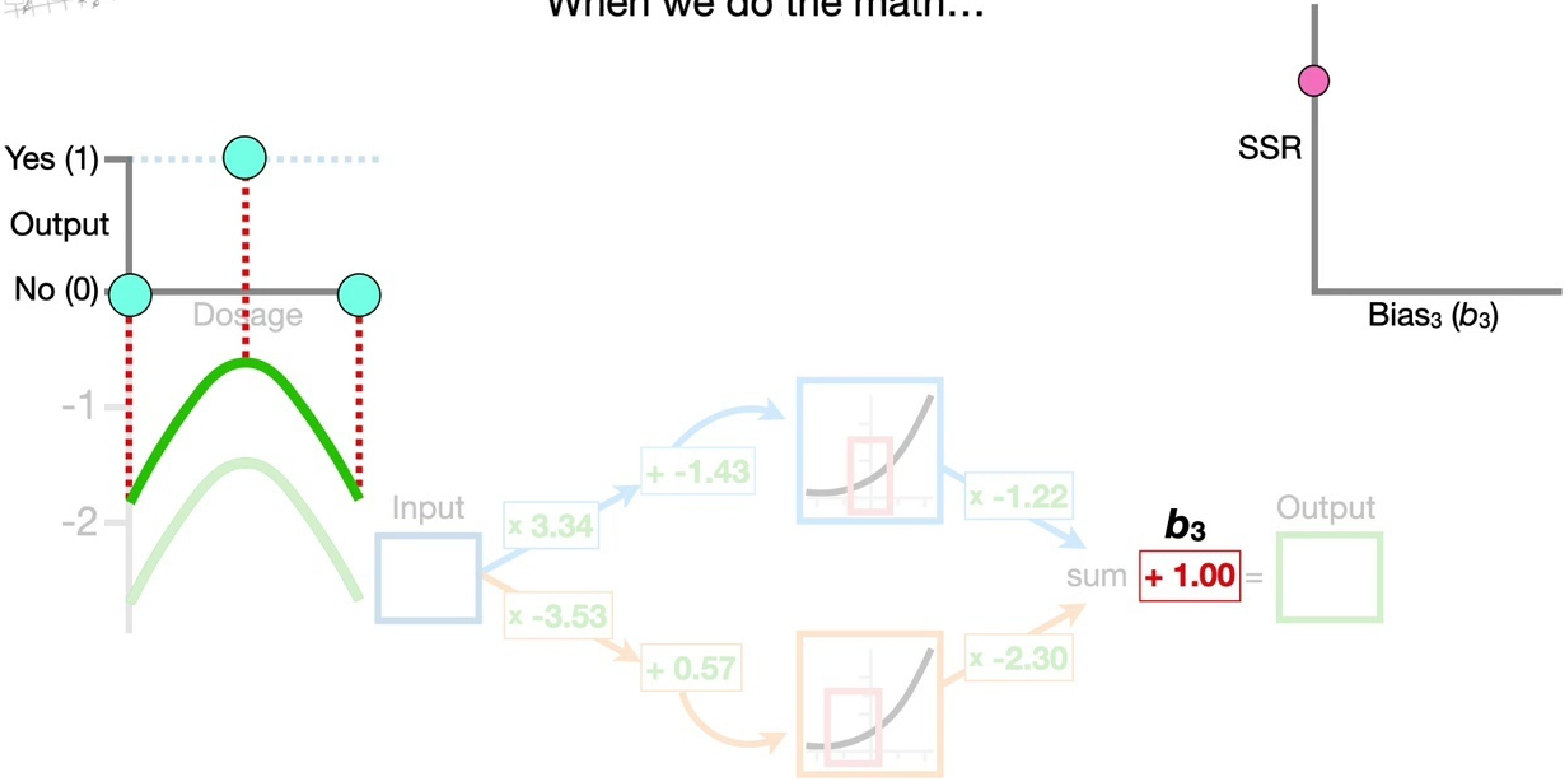
...and shift the **green squiggle** up **1**...



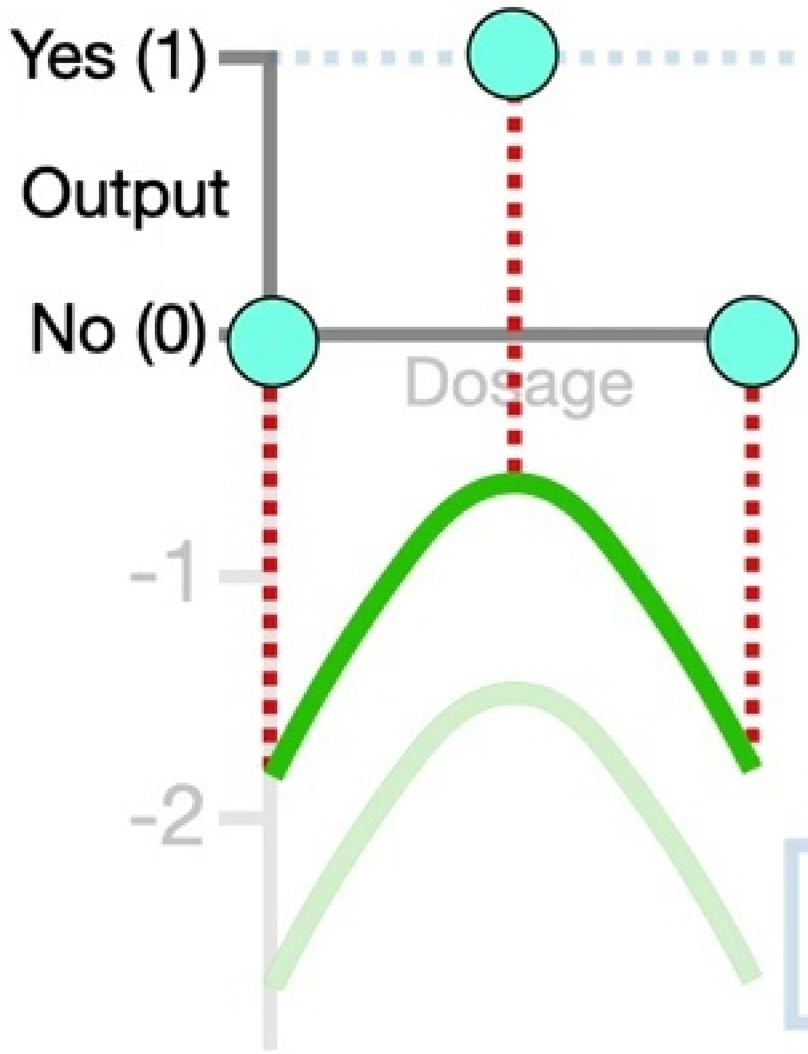
...and we end up with shorter **Residuals.**



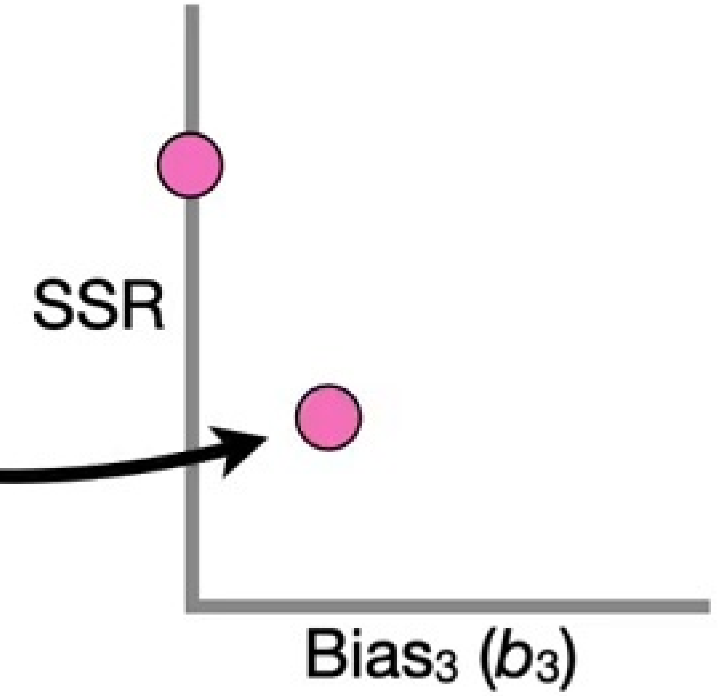
When we do the math…



...and that corresponds to this point on our graph.

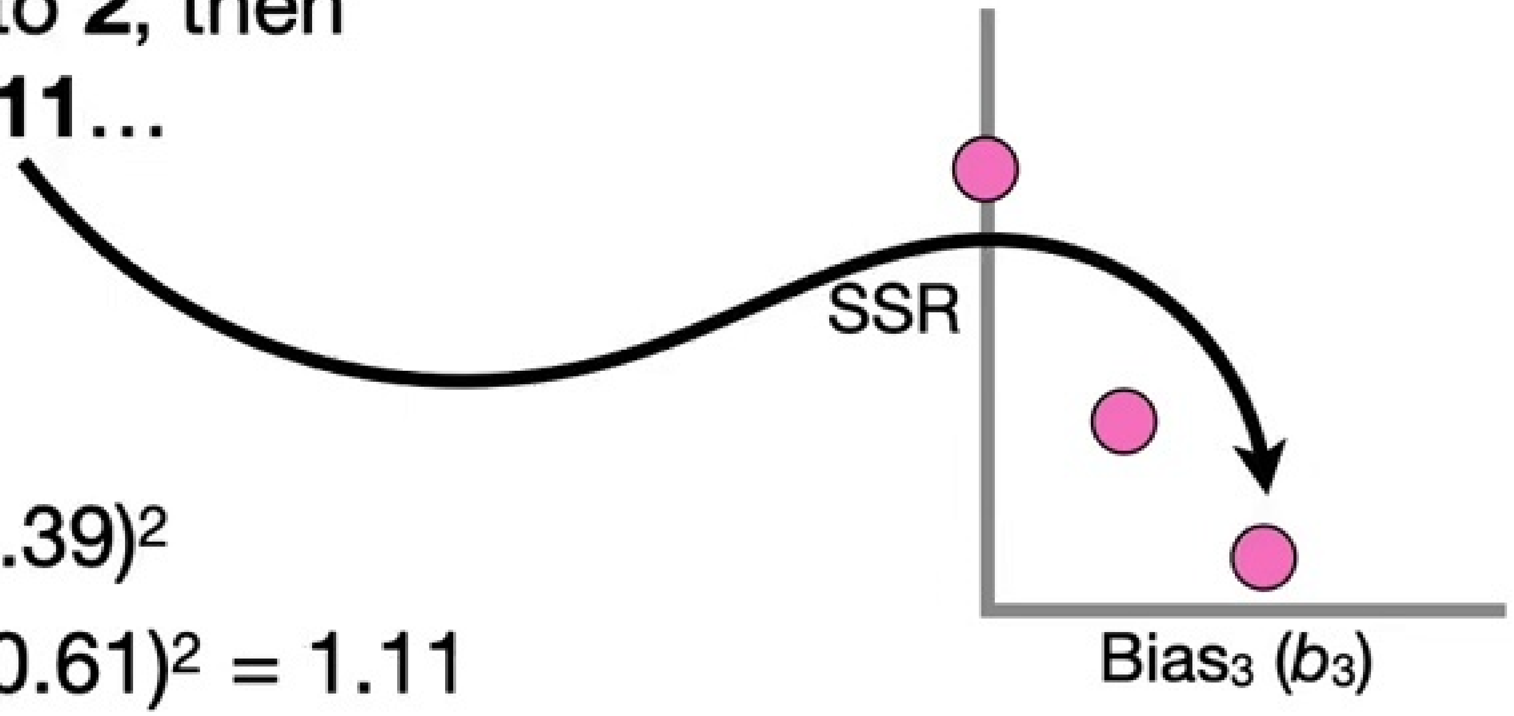


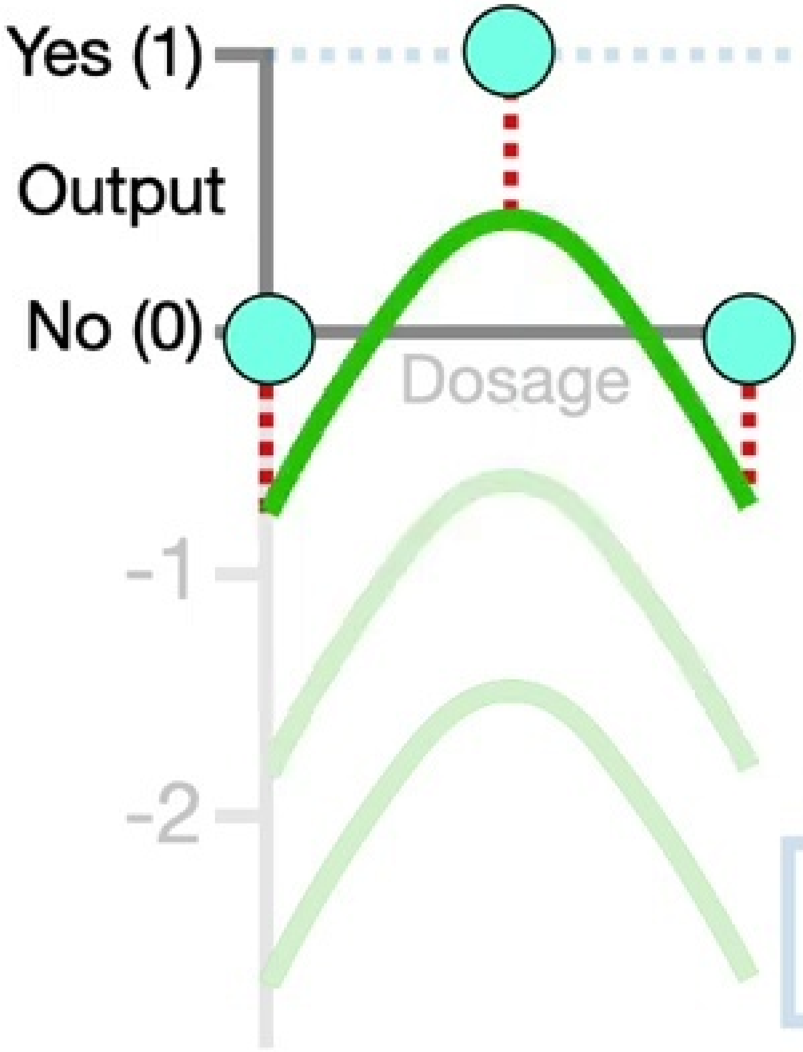
$$SSR = (0 - -1.6)^2 + (1 - -0.61)^2 + (0 - -1.61)^2 = 7.8$$



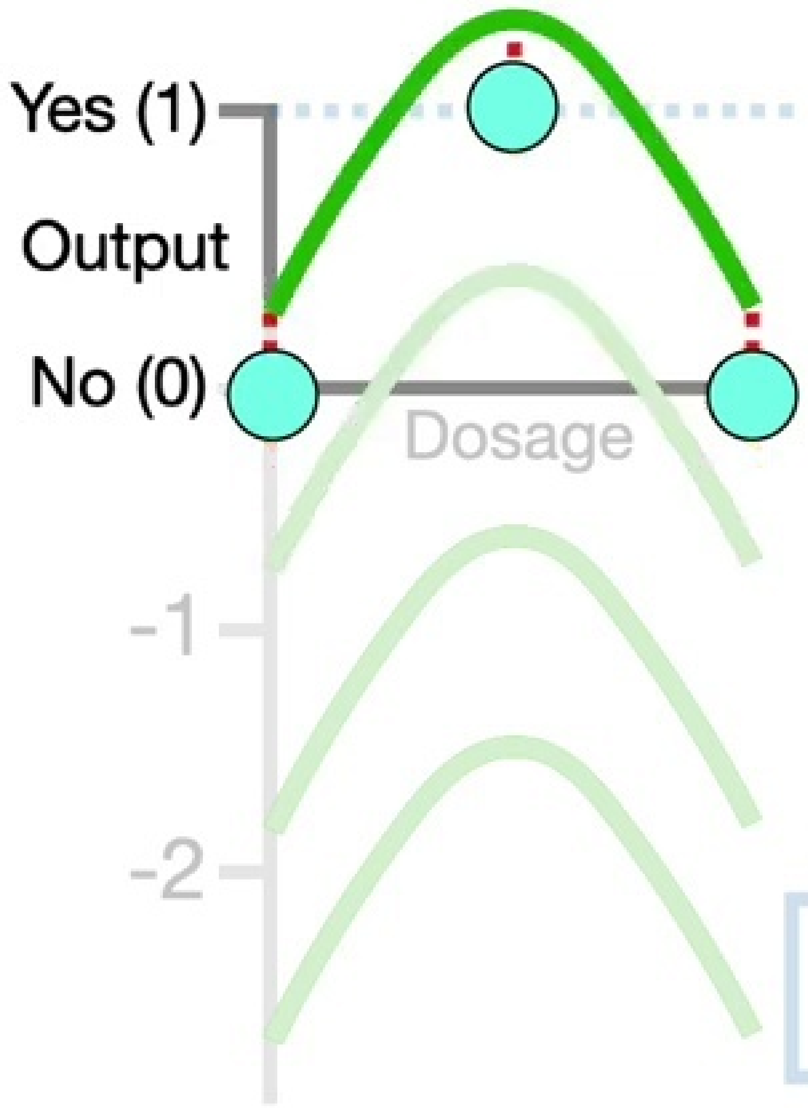
If we increase b_3 to 2, then the SSR = 1.11…

$$SSR = (0 - -0.6)^2 + (1 - 0.39)^2 + (0 - -0.61)^2 = 1.11$$

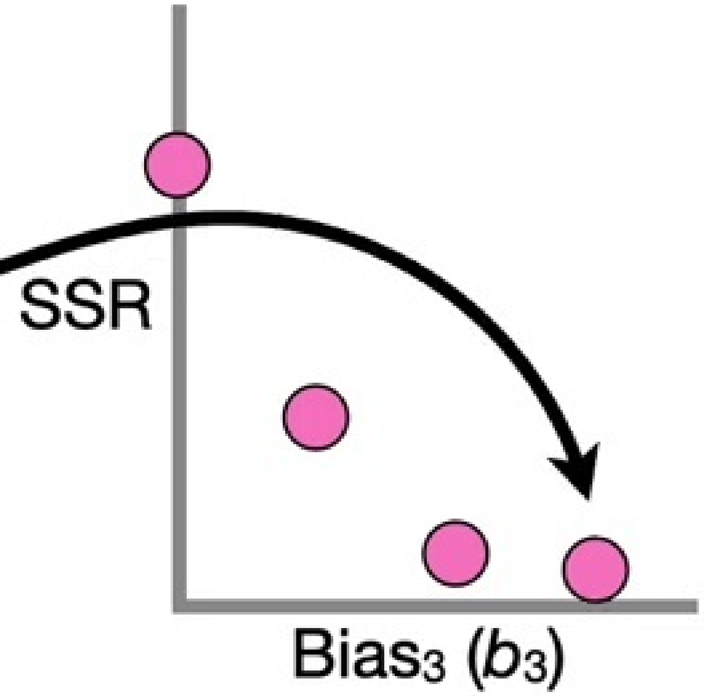




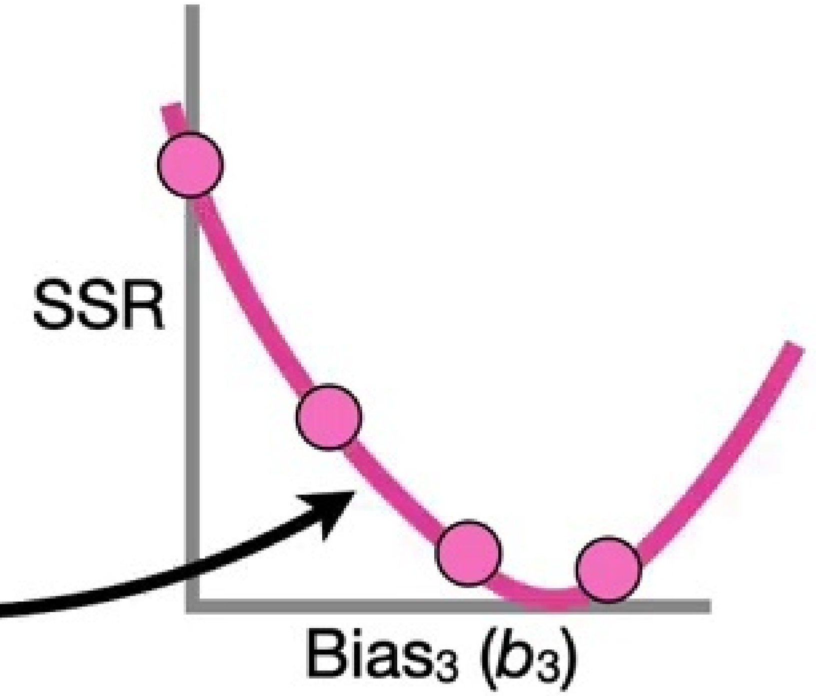
…and if we increase $\boldsymbol{b_3}$ to **3**, then the **SSR = 0.46.**

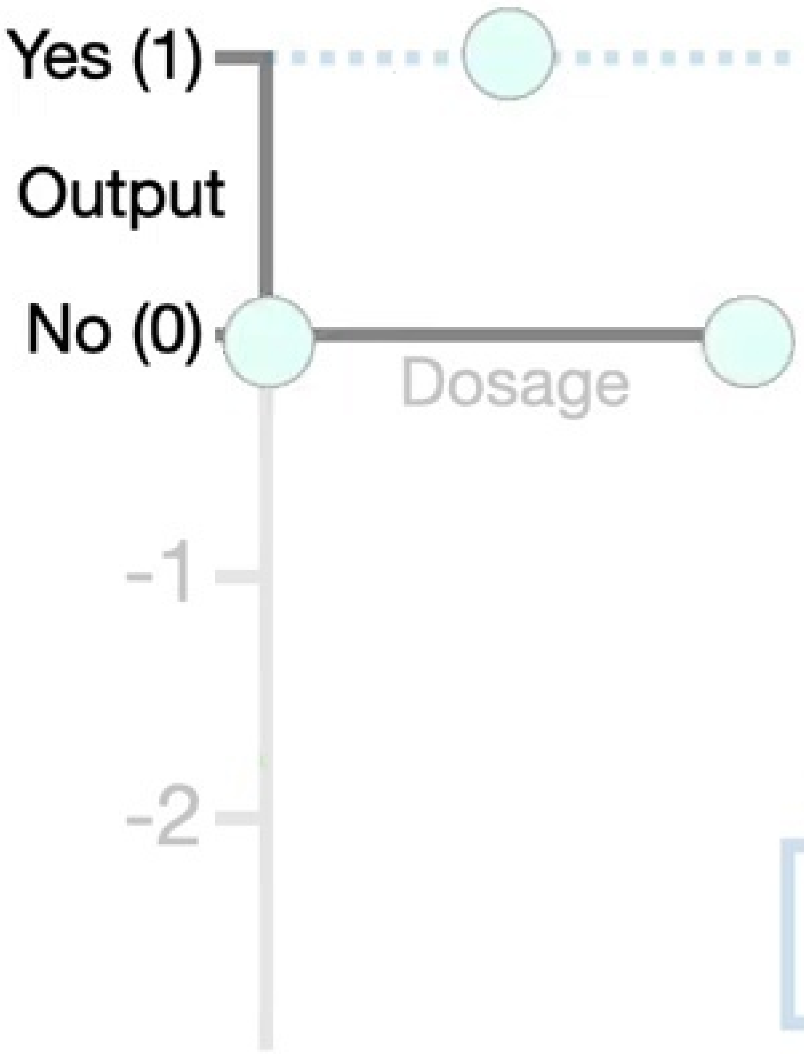


$$\mathrm{SSR} = (0 - 0.4)^2 + (1 - 1.39)^2 + (0 - 0.39)^2 = 0.46$$

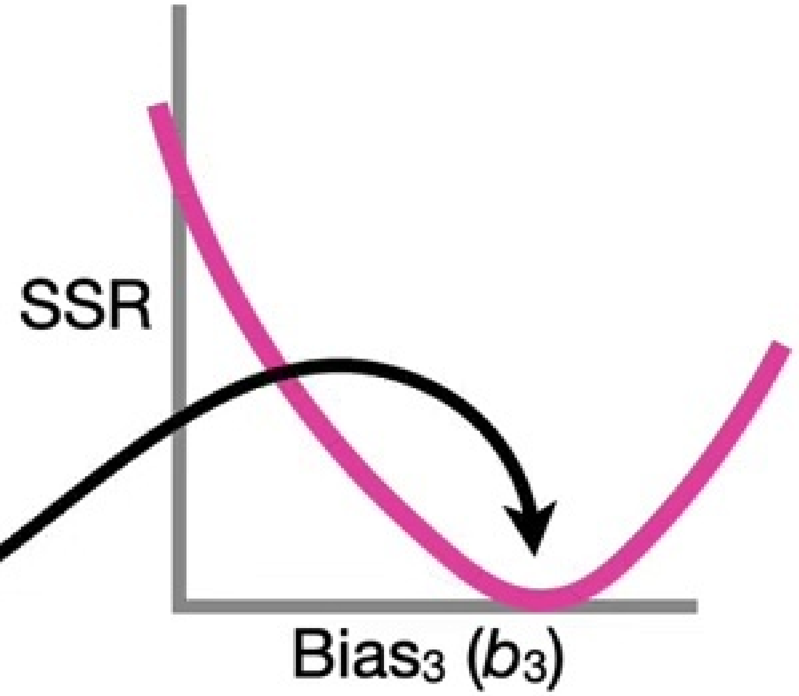


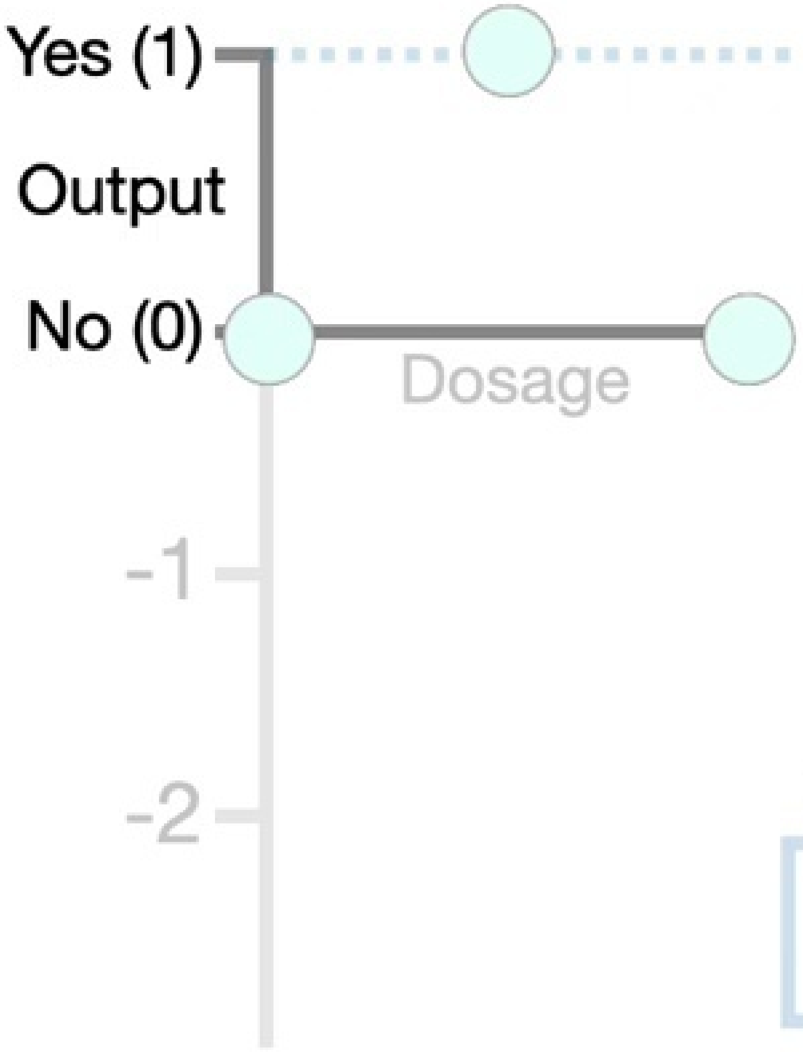
And if we had time to plug in tons of values for $\boldsymbol{b_3}$, we would get this **pink curve**…



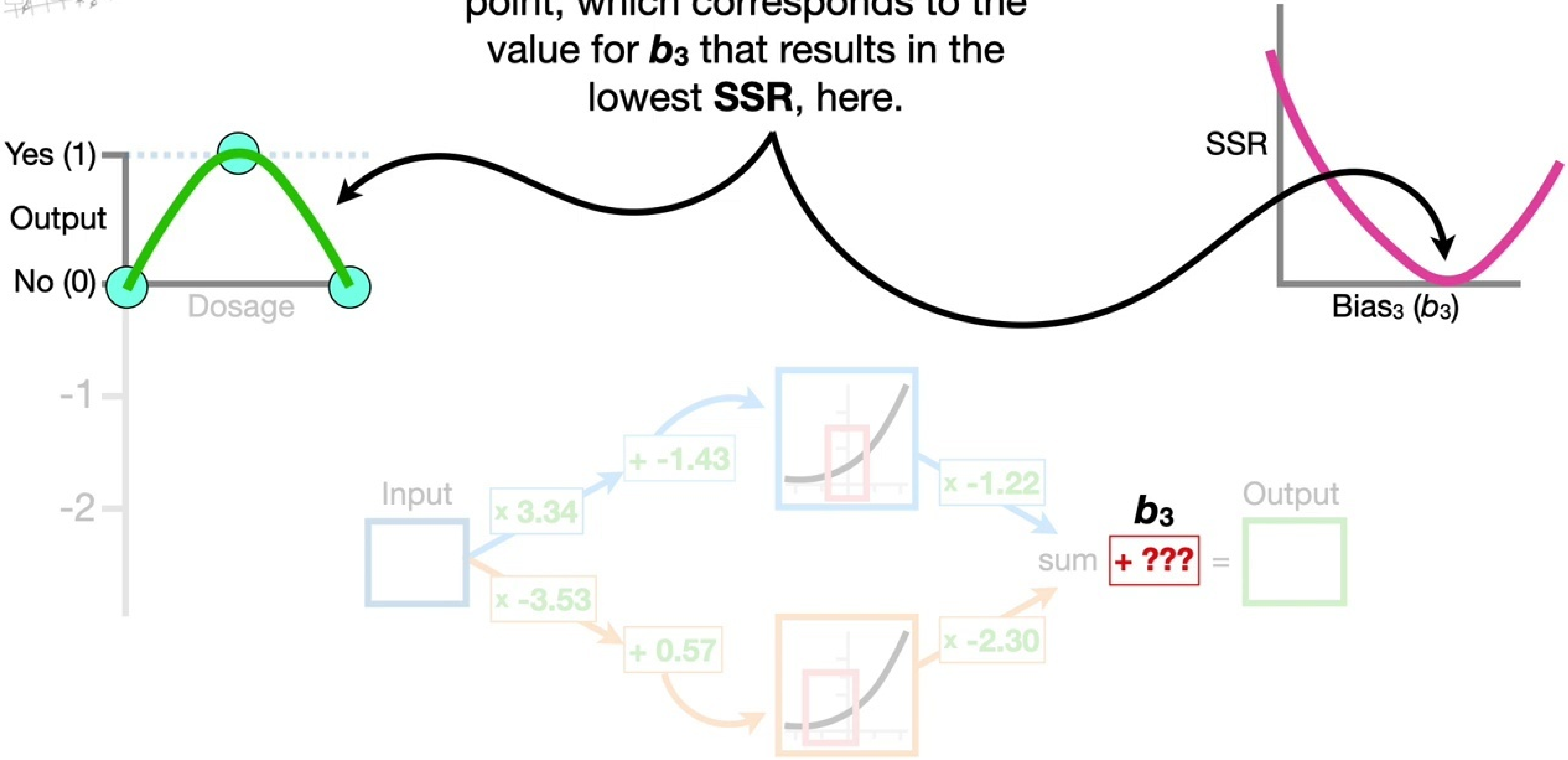


...and we could find the lowest point, which corresponds to the value for b_3 that results in the lowest **SSR**, here.

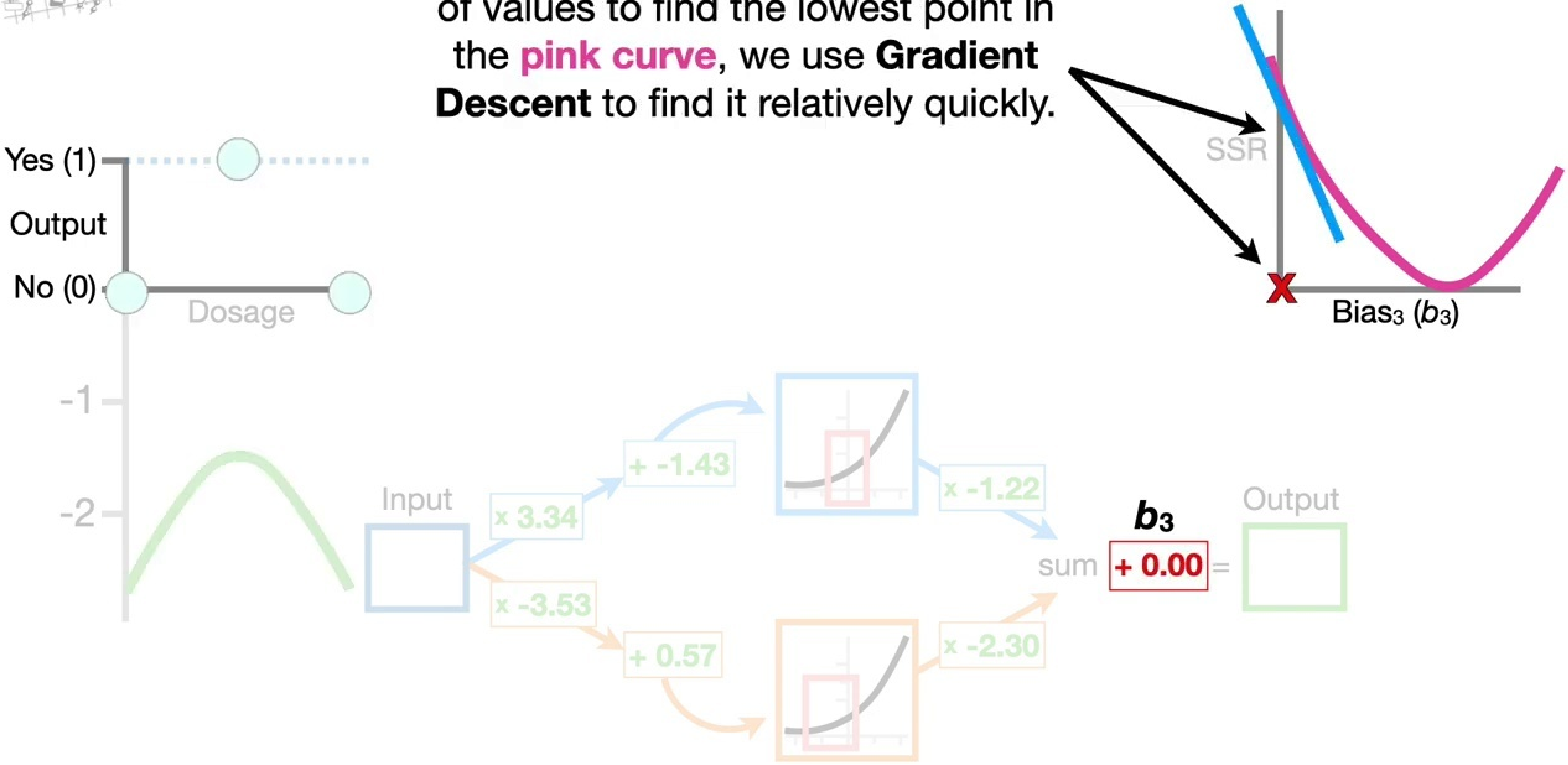




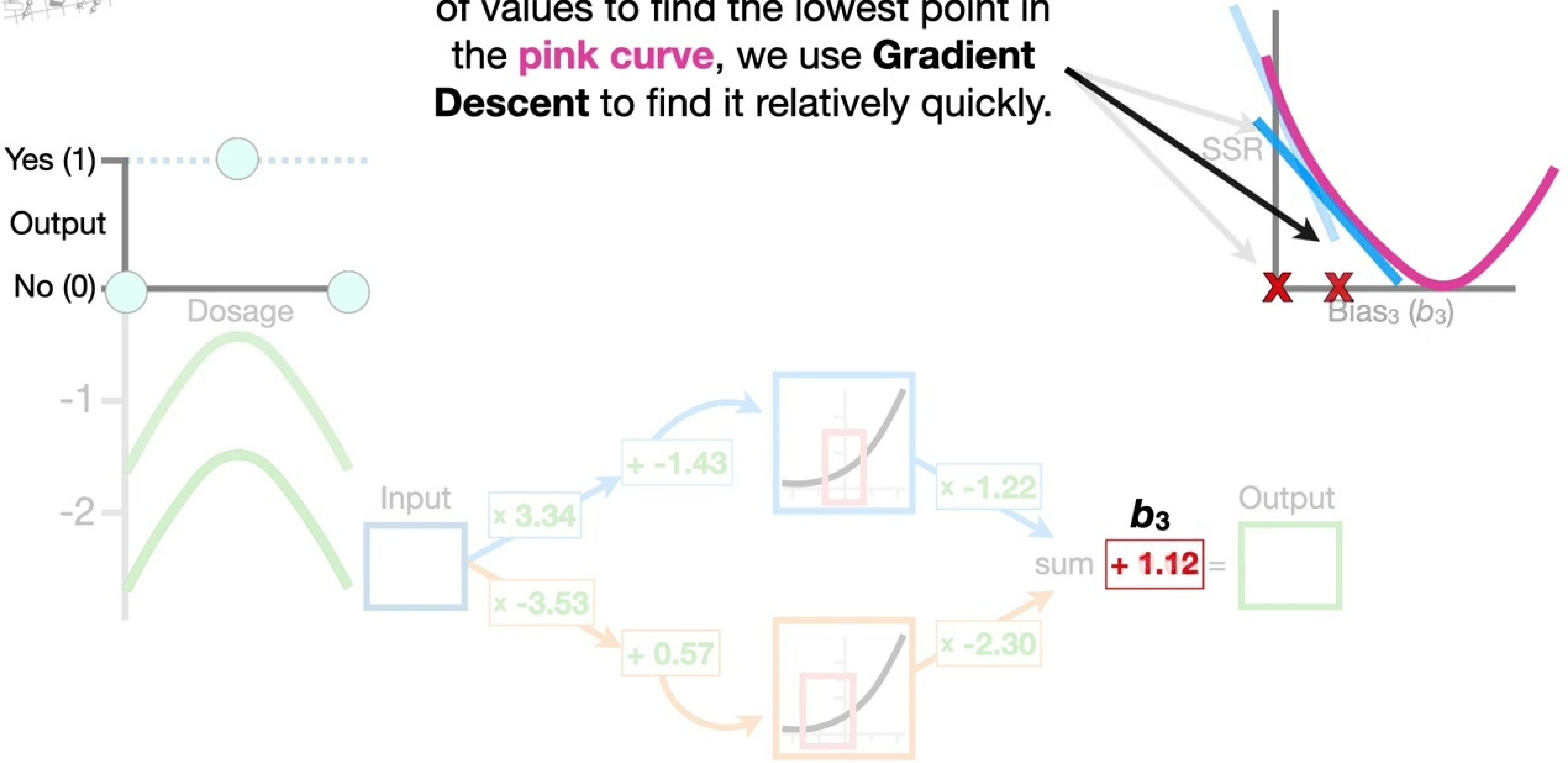
…and we could find the lowest point, which corresponds to the value for $\boldsymbol{b_3}$ that results in the lowest **SSR**, here.



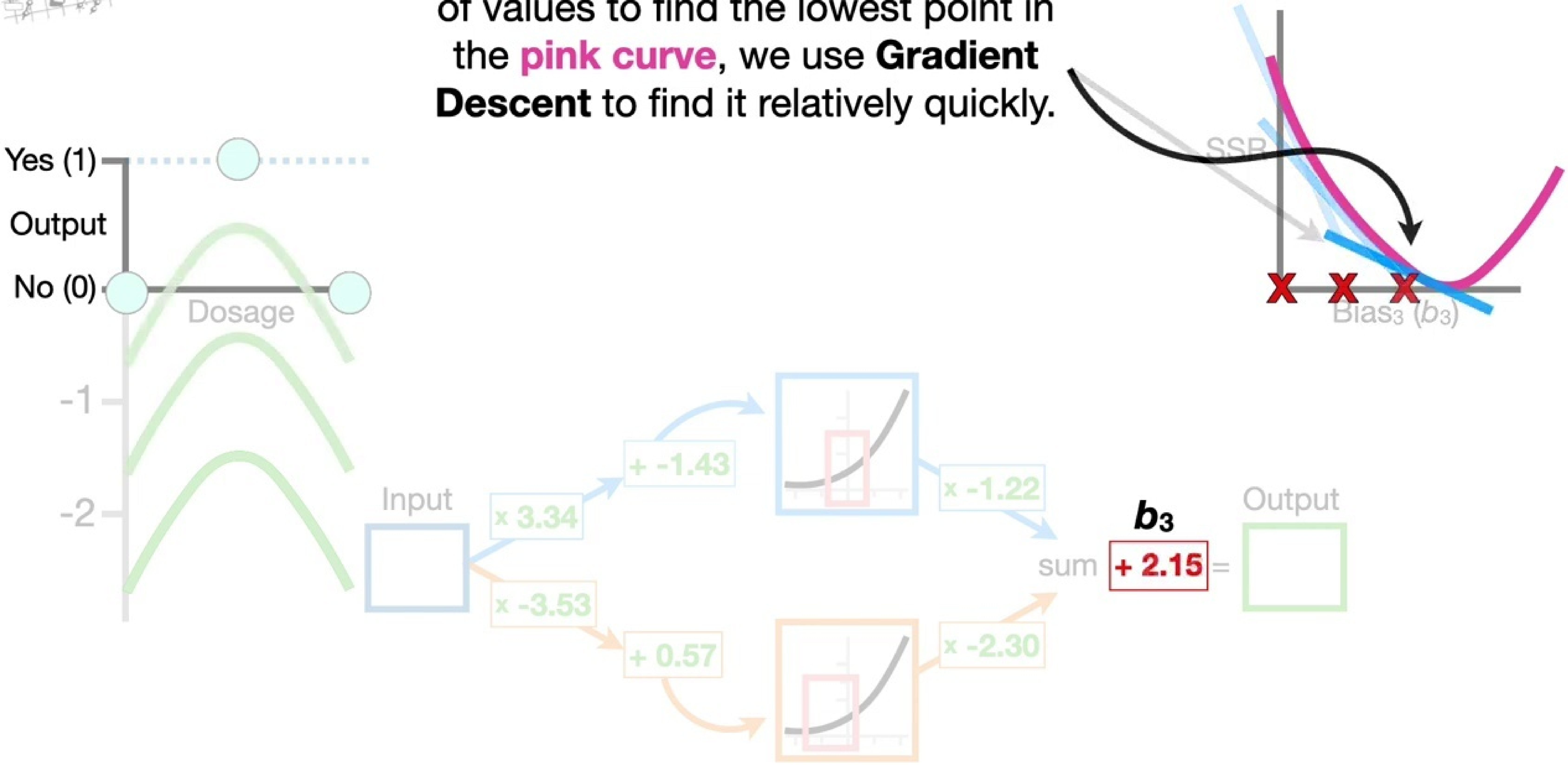
However, instead of plugging in tons of values to find the lowest point in the **pink curve**, we use **Gradient Descent** to find it relatively quickly.



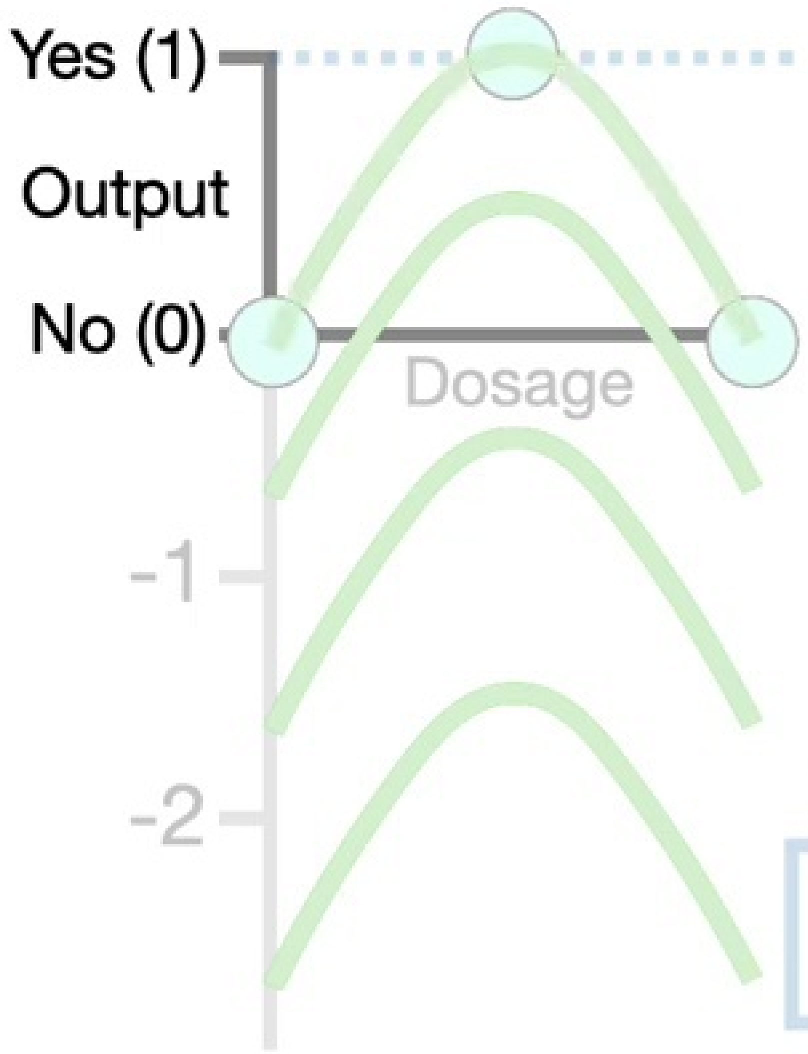
However, instead of plugging in tons of values to find the lowest point in the **pink curve**, we use **Gradient Descent** to find it relatively quickly.

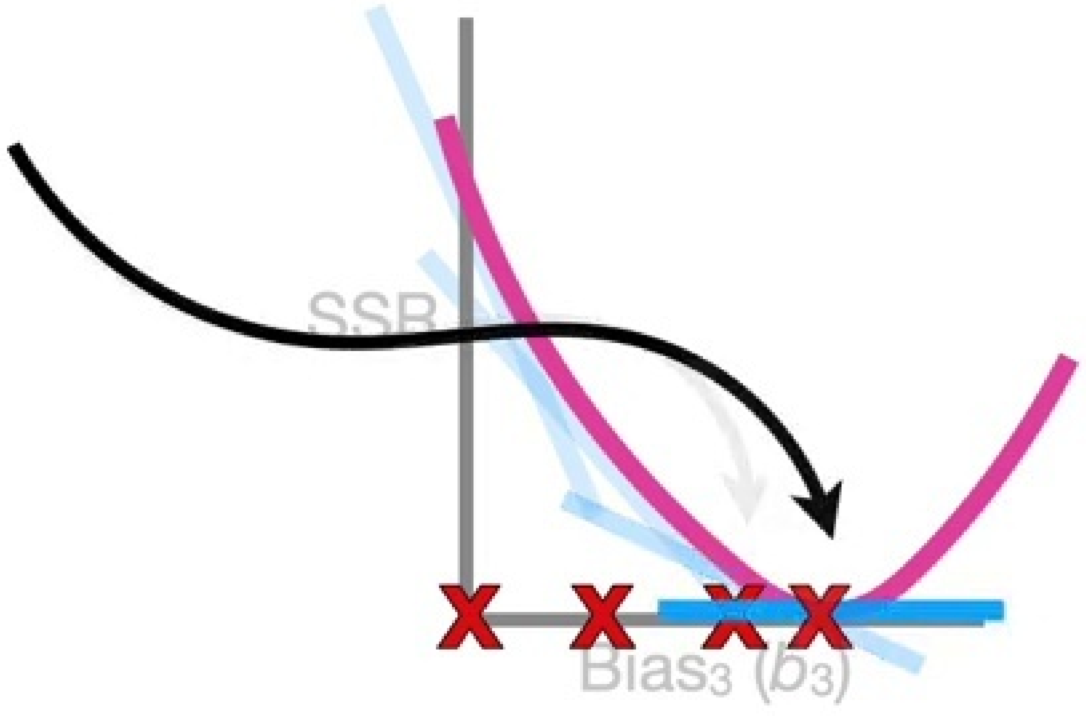


However, instead of plugging in tons of values to find the lowest point in the **pink curve**, we use **Gradient Descent** to find it relatively quickly.

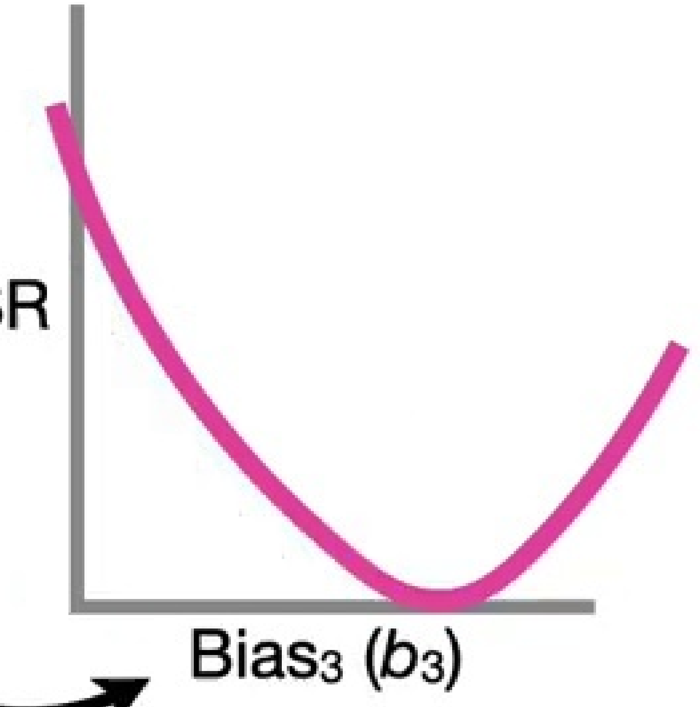


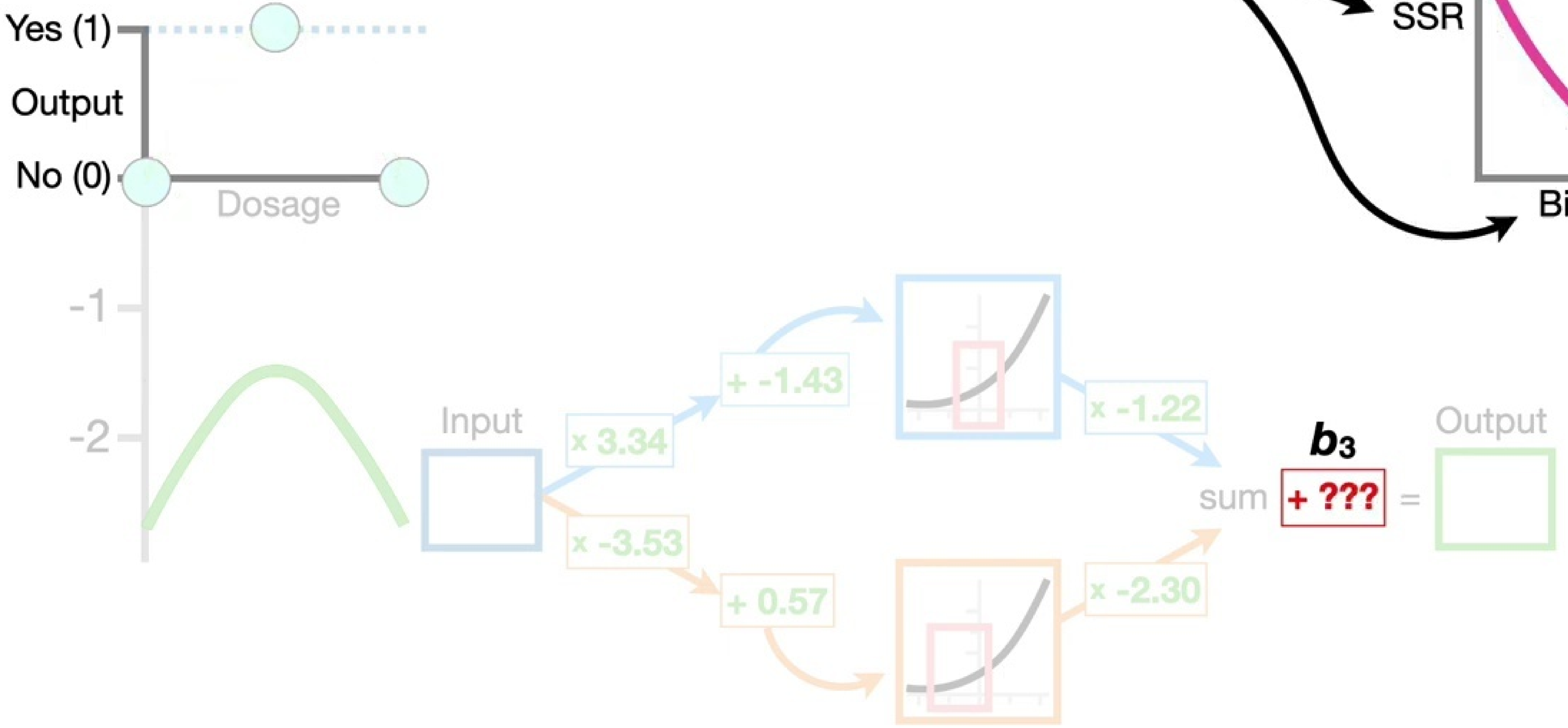
However, instead of plugging in tons of values to find the lowest point in the **pink curve**, we use **Gradient Descent** to find it relatively quickly.





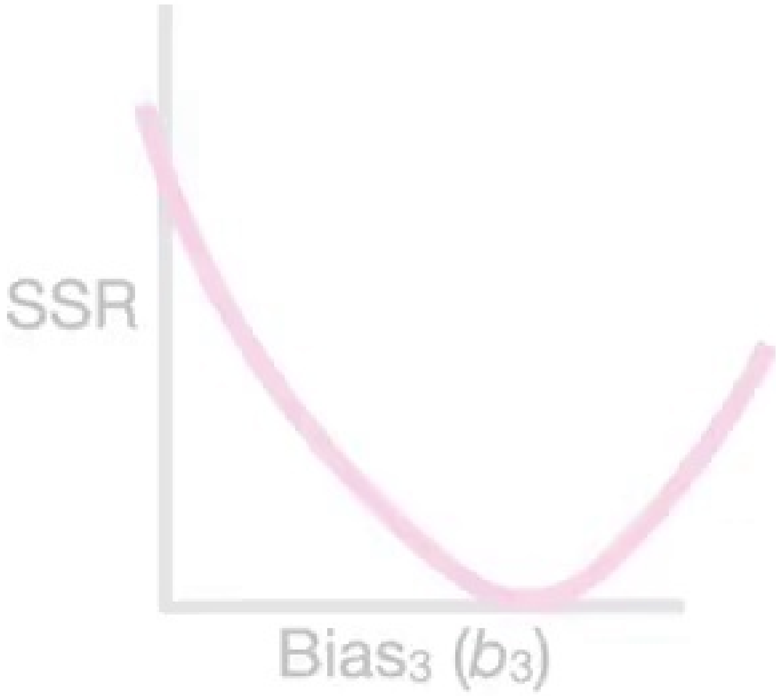
And that means we need to find the derivative of the **Sum of the Squared Residuals** with respect to ***b*₃**.





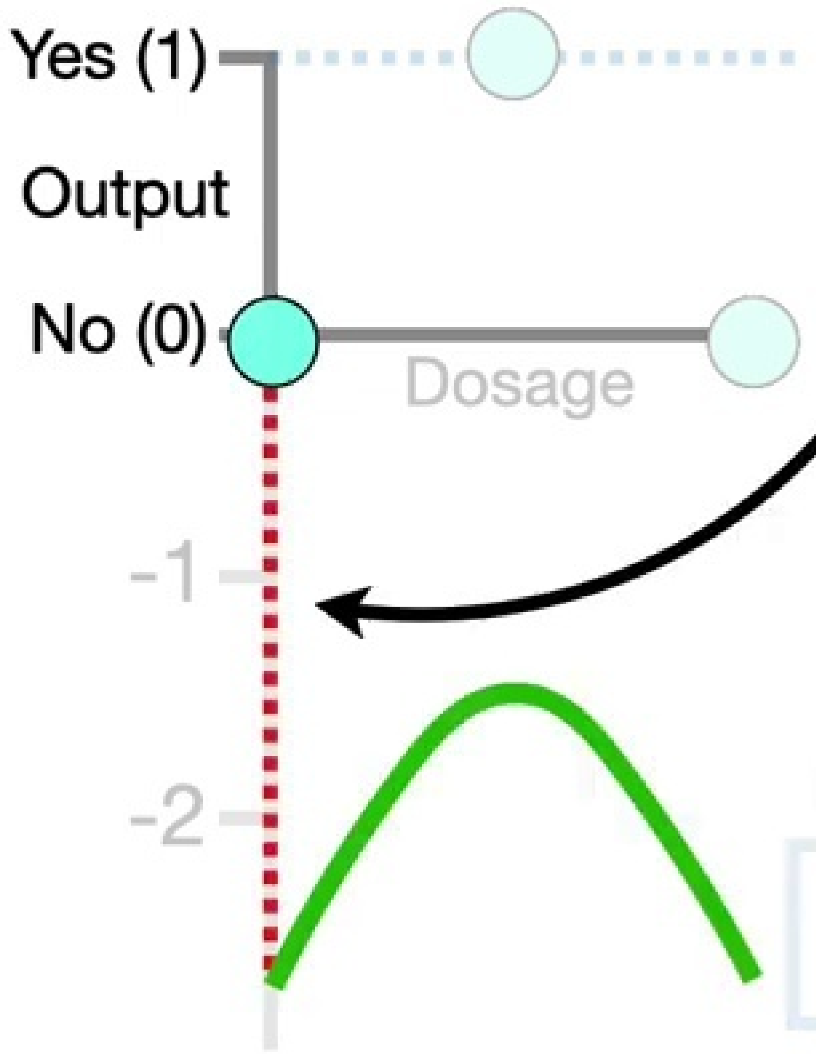
Now, remember the **Sum of the Squared Residuals** equals…

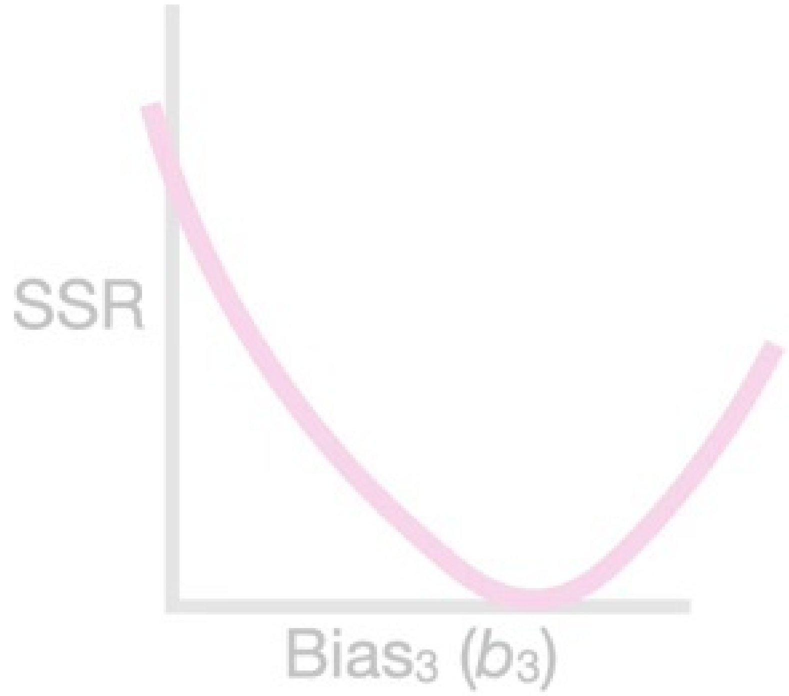
SSR =



…the first **Residual** squared…

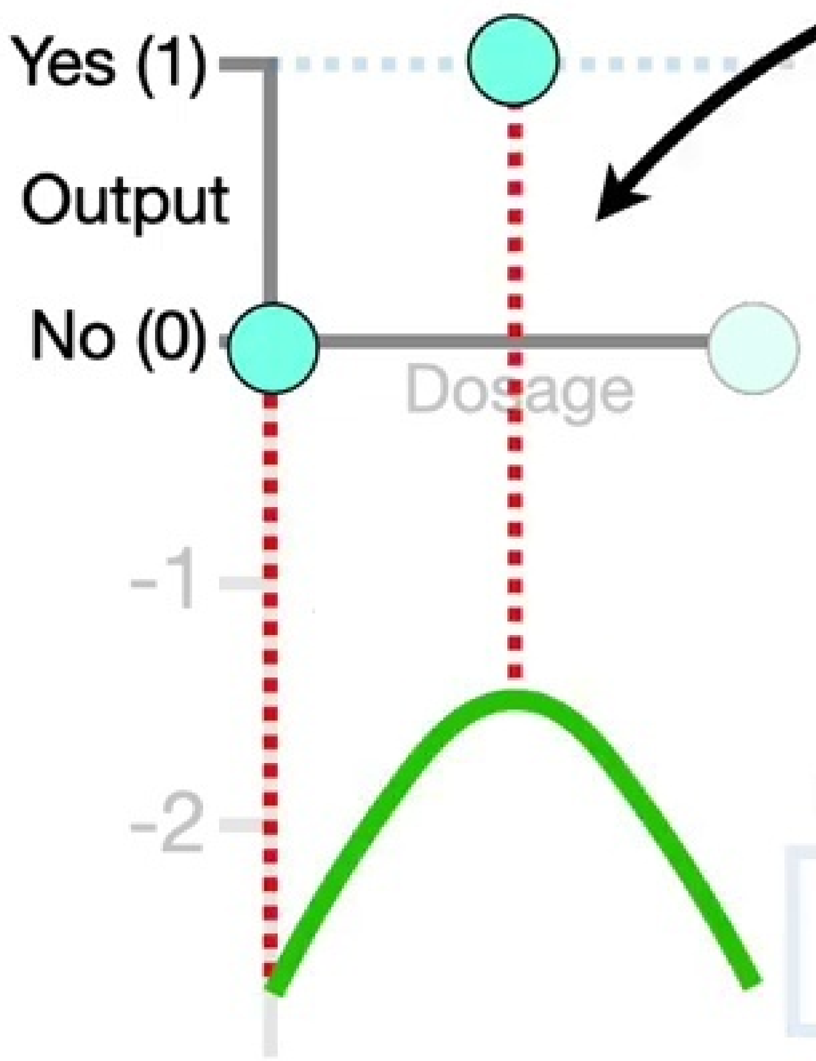
$$SSR = (Observed_1 - Predicted_1)^2$$

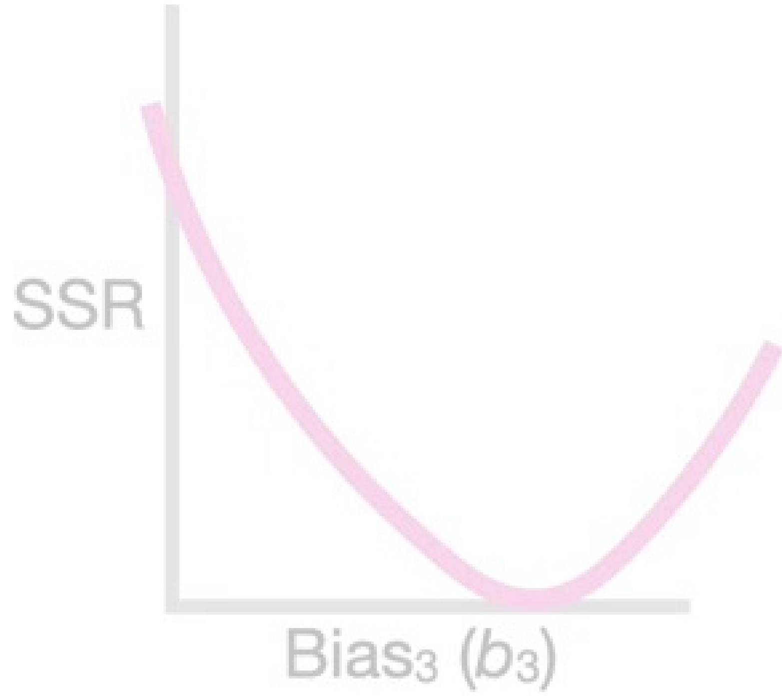




...plus all of the other squared **Residuals.**

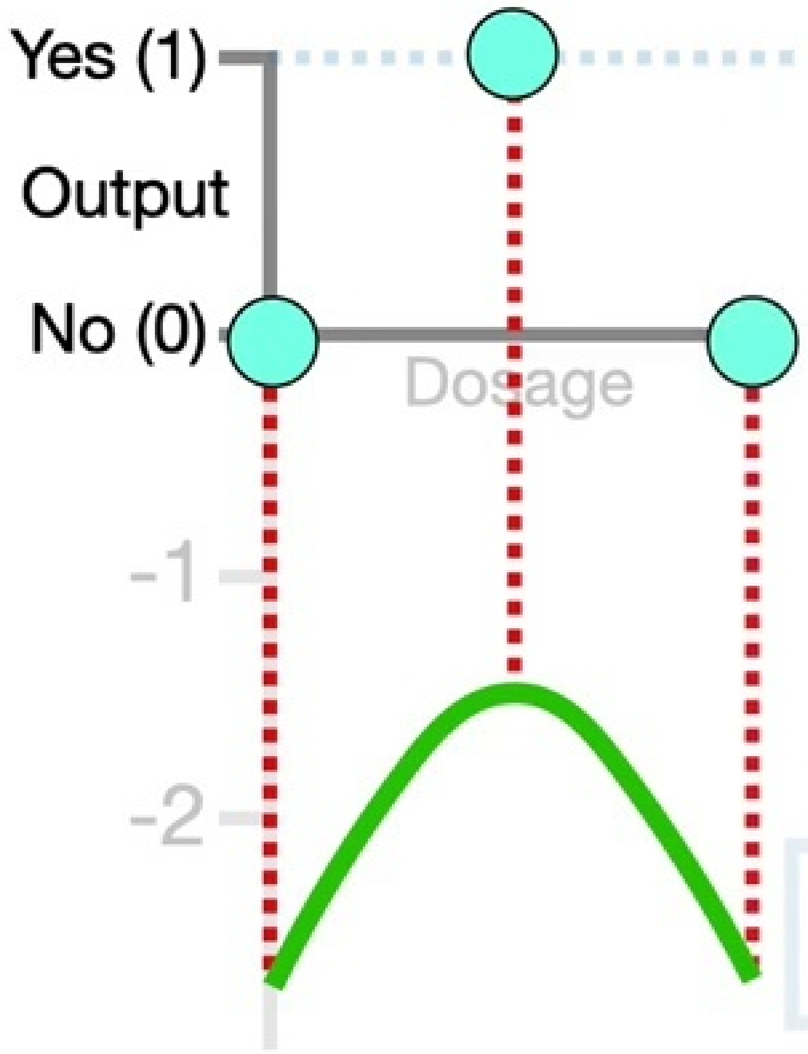
$$SSR = (Observed_1 - Predicted_1)^2$$
$$+ (Observed_2 - Predicted_2)^2$$

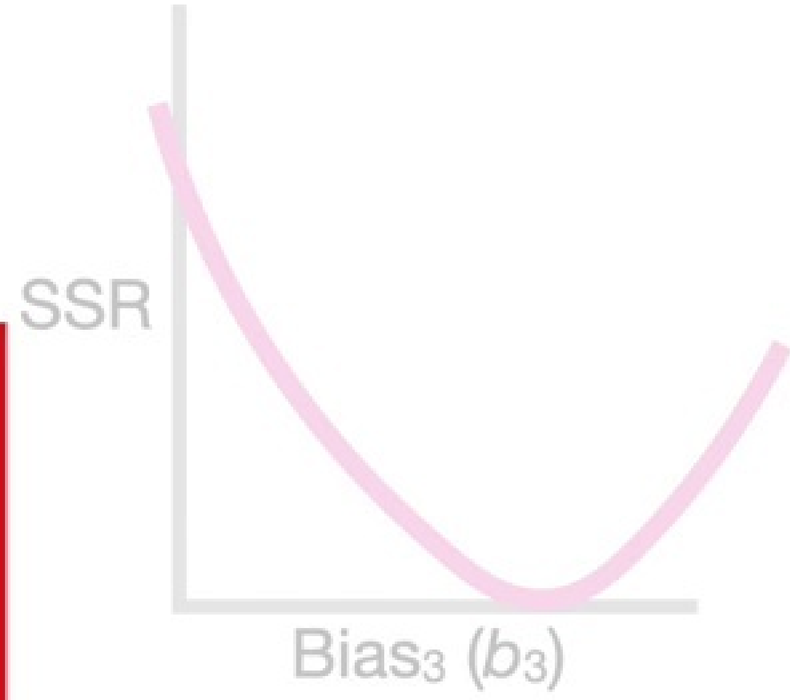




Now, because this equation takes up a lot of space…

$$
\begin{aligned}
\text{SSR} = &(\text{Observed}_1 - \text{Predicted}_1)^2 \\
&+ (\text{Observed}_2 - \text{Predicted}_2)^2 \\
&+ (\text{Observed}_3 - \text{Predicted}_3)^2
\end{aligned}
$$

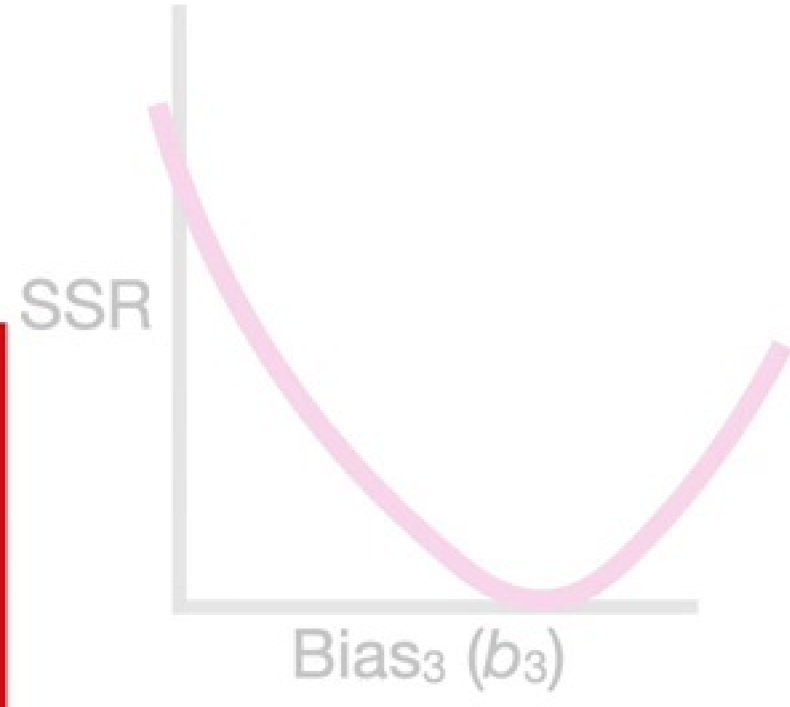


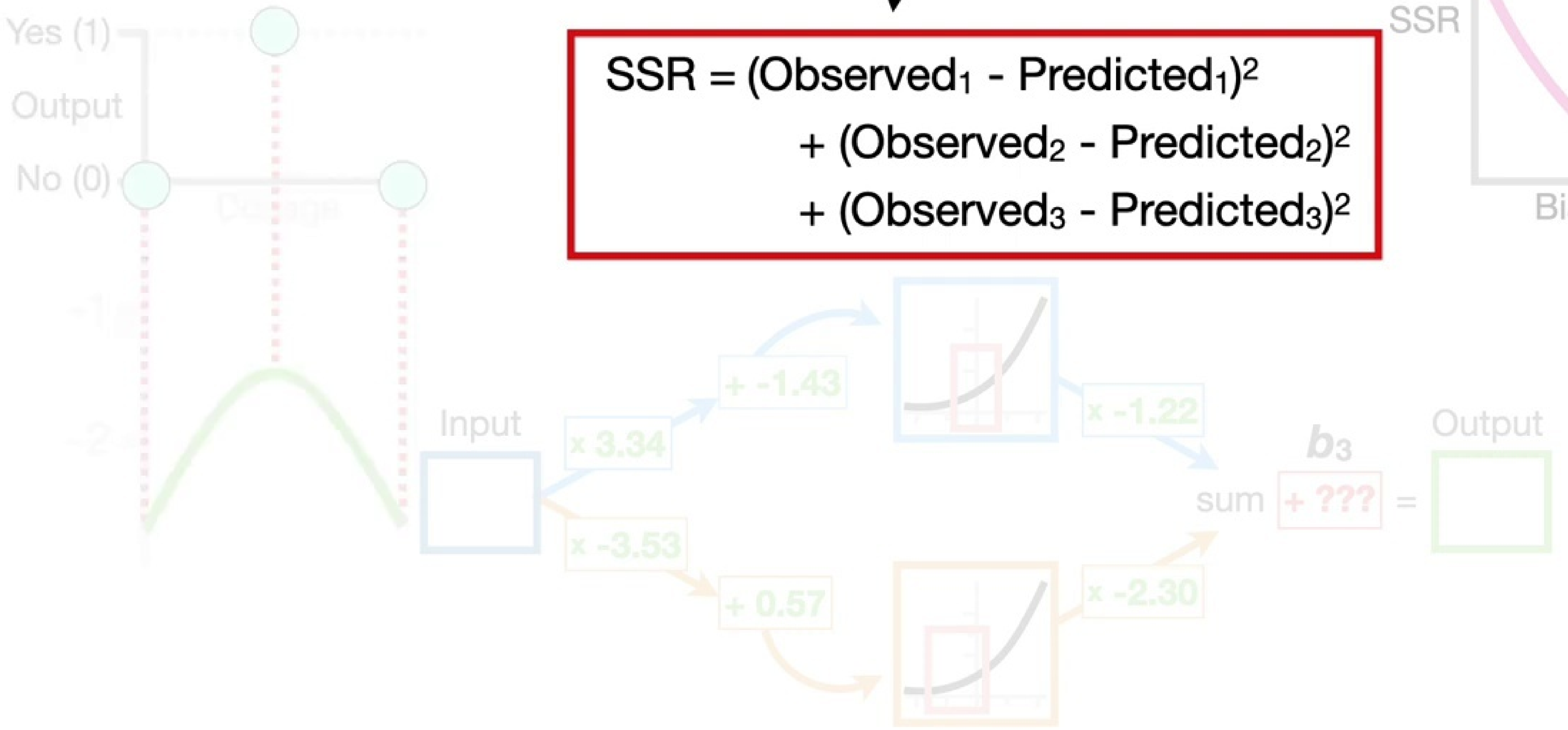


…we can make it smaller by using **summation notation.**

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

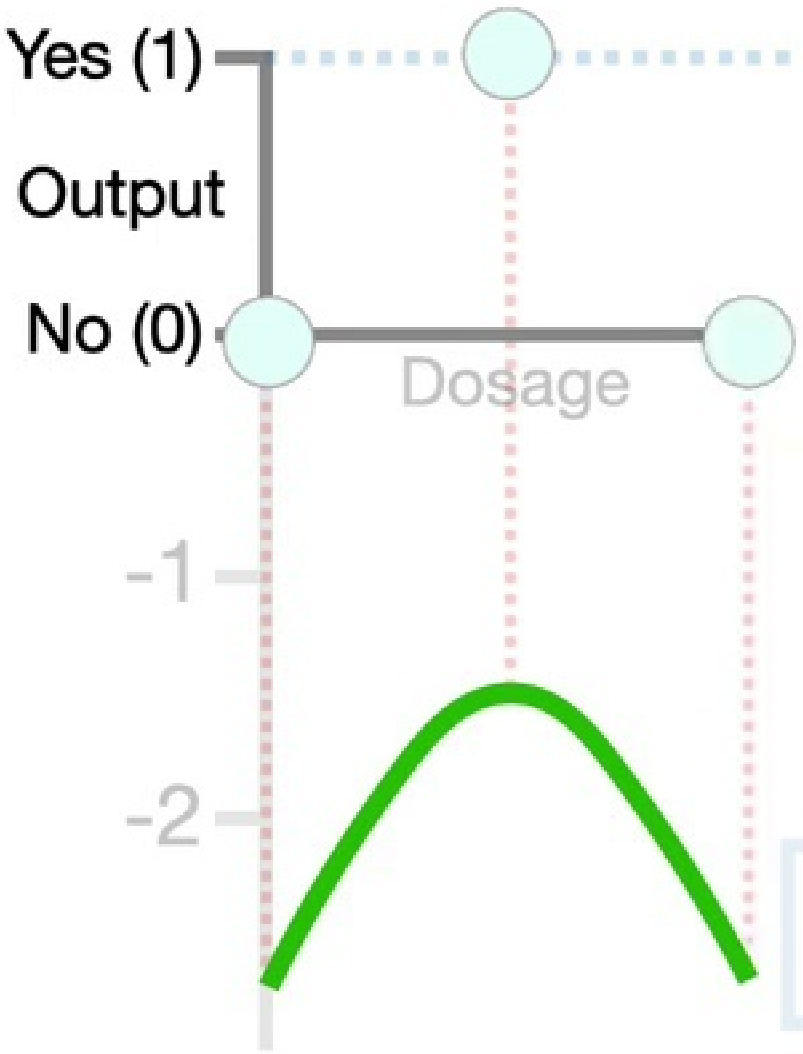
$$\boxed{\begin{aligned}\text{SSR} = &(\text{Observed}_1 - \text{Predicted}_1)^2 \\ &+ (\text{Observed}_2 - \text{Predicted}_2)^2 \\ &+ (\text{Observed}_3 - \text{Predicted}_3)^2\end{aligned}}$$

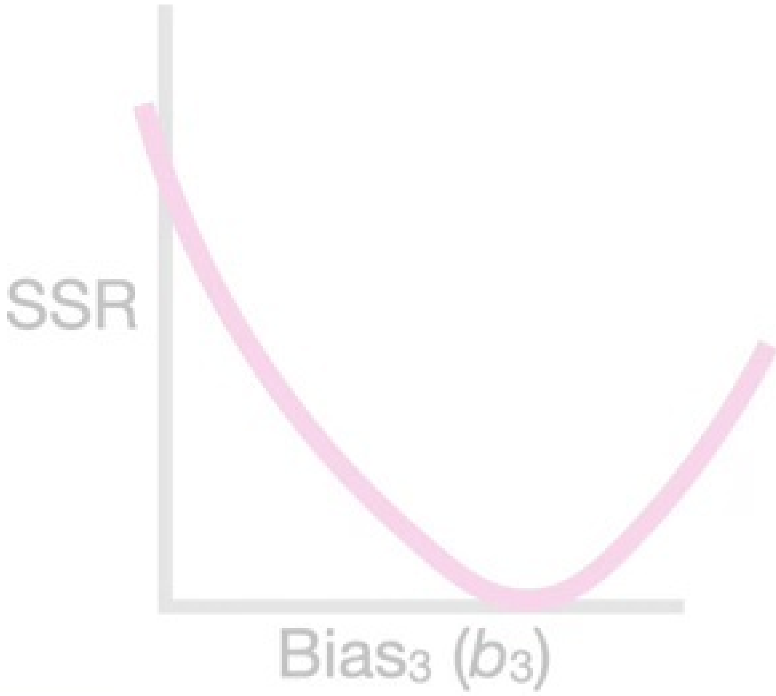




Now let's talk a little bit more about the **Predicted** values.

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \boxed{\text{Predicted}_i})^2$$

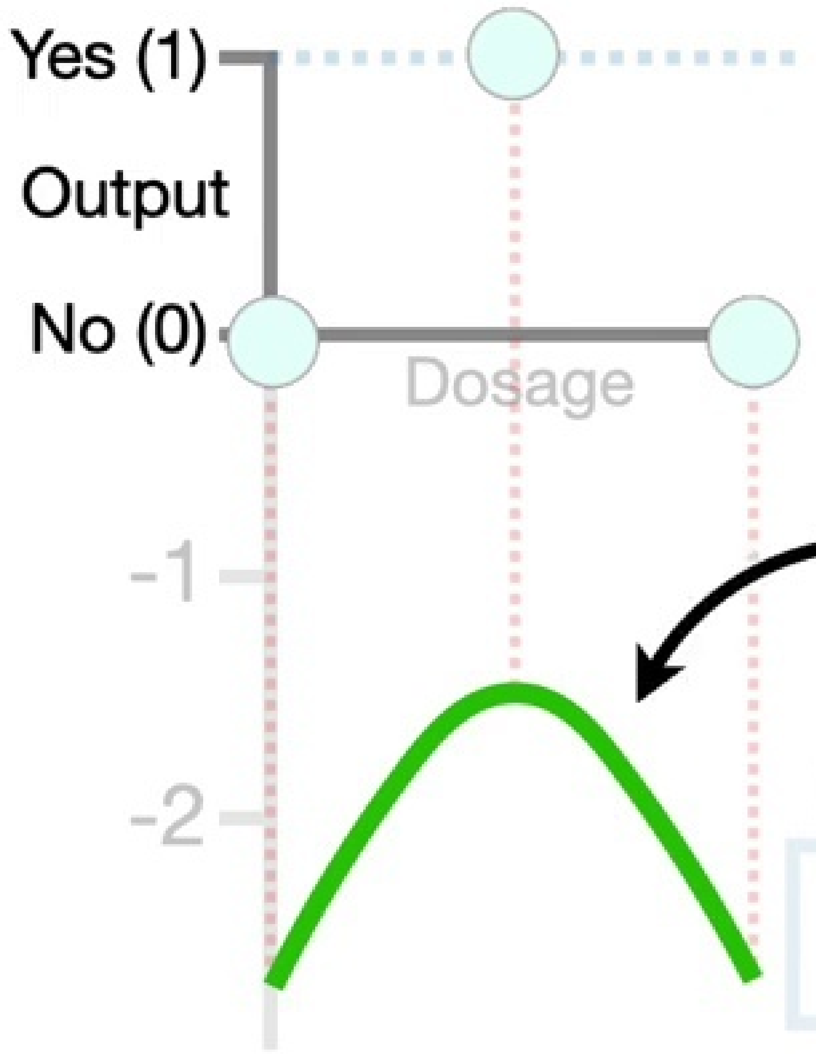


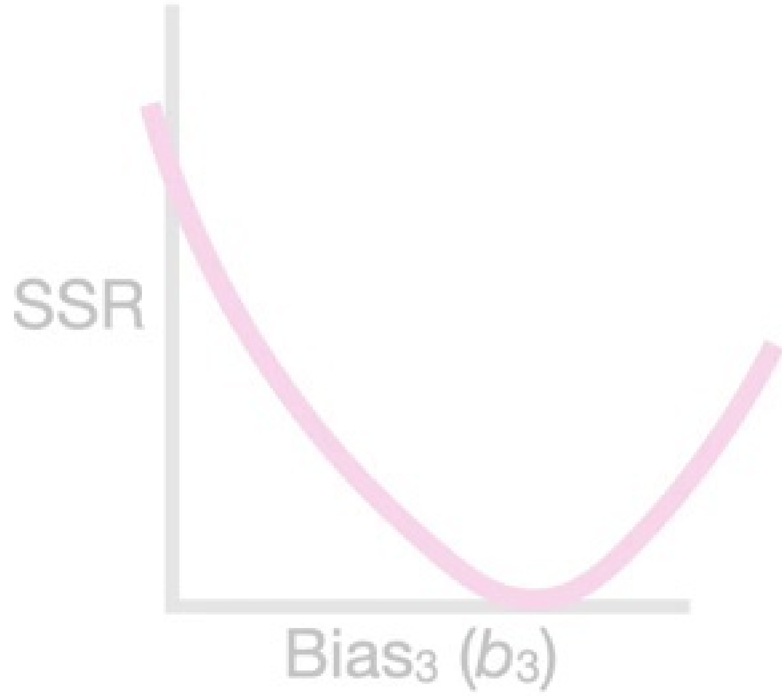


Each **Predicted** value comes from the **green squiggle**…

$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \boxed{\text{Predicted}_i})^2$$

$$\boxed{\text{Predicted}_i = \text{green squiggle}_i}$$

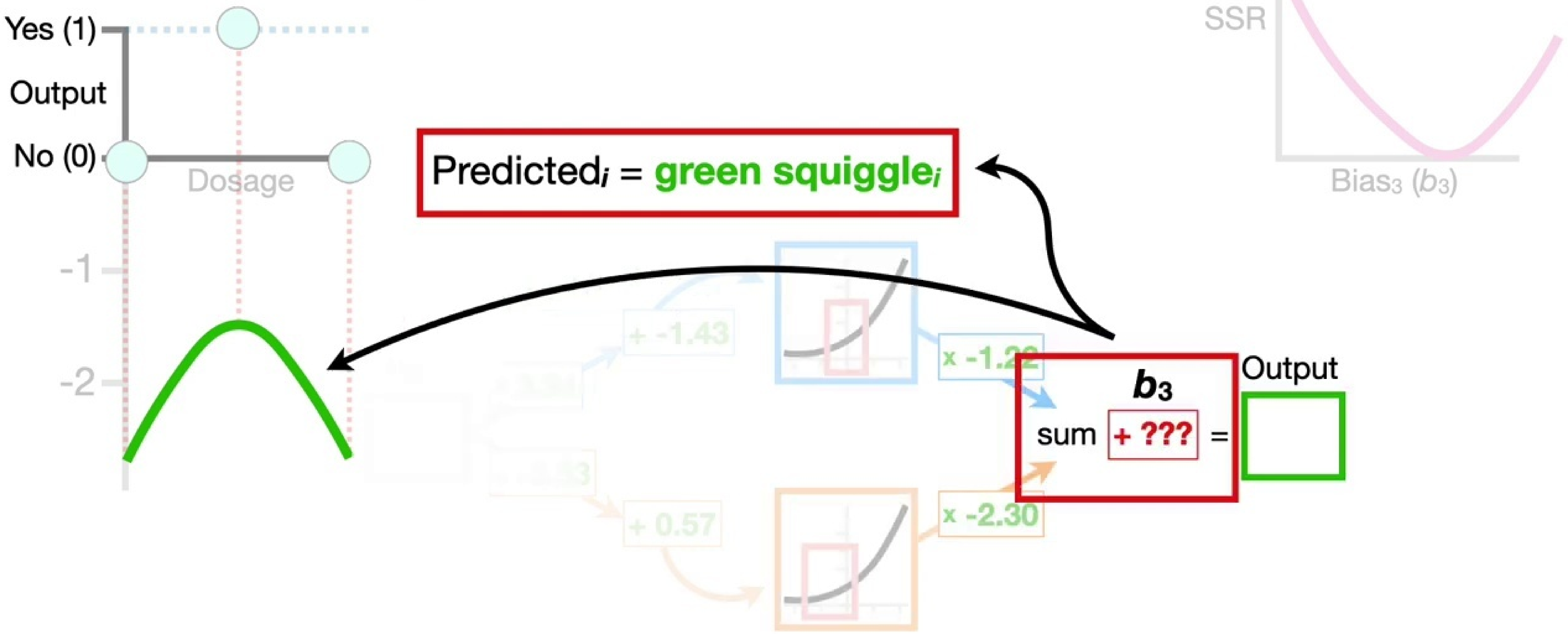




...and the **green squiggle** comes from the last part of the **Neural Network.**

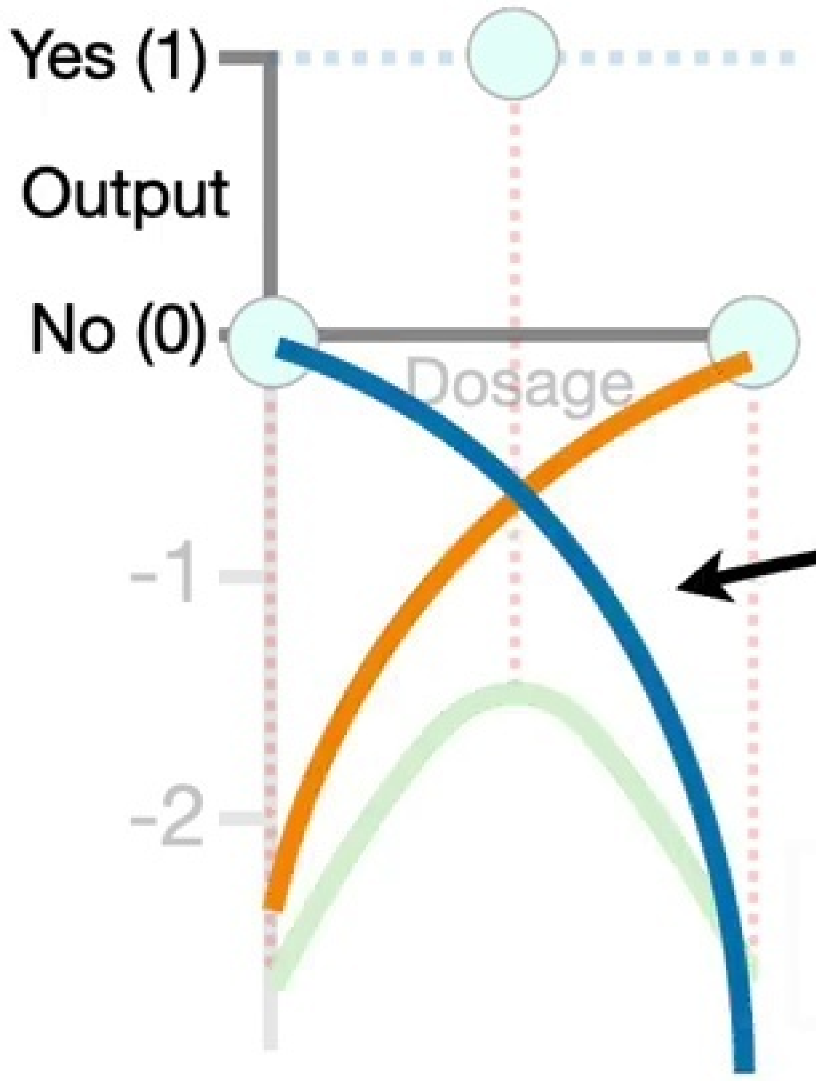
$$SSR = \sum_{i=1}^{n=3}(\text{Observed}_i - \text{Predicted}_i)^2$$

$$\text{Predicted}_i = \text{green squiggle}_i$$

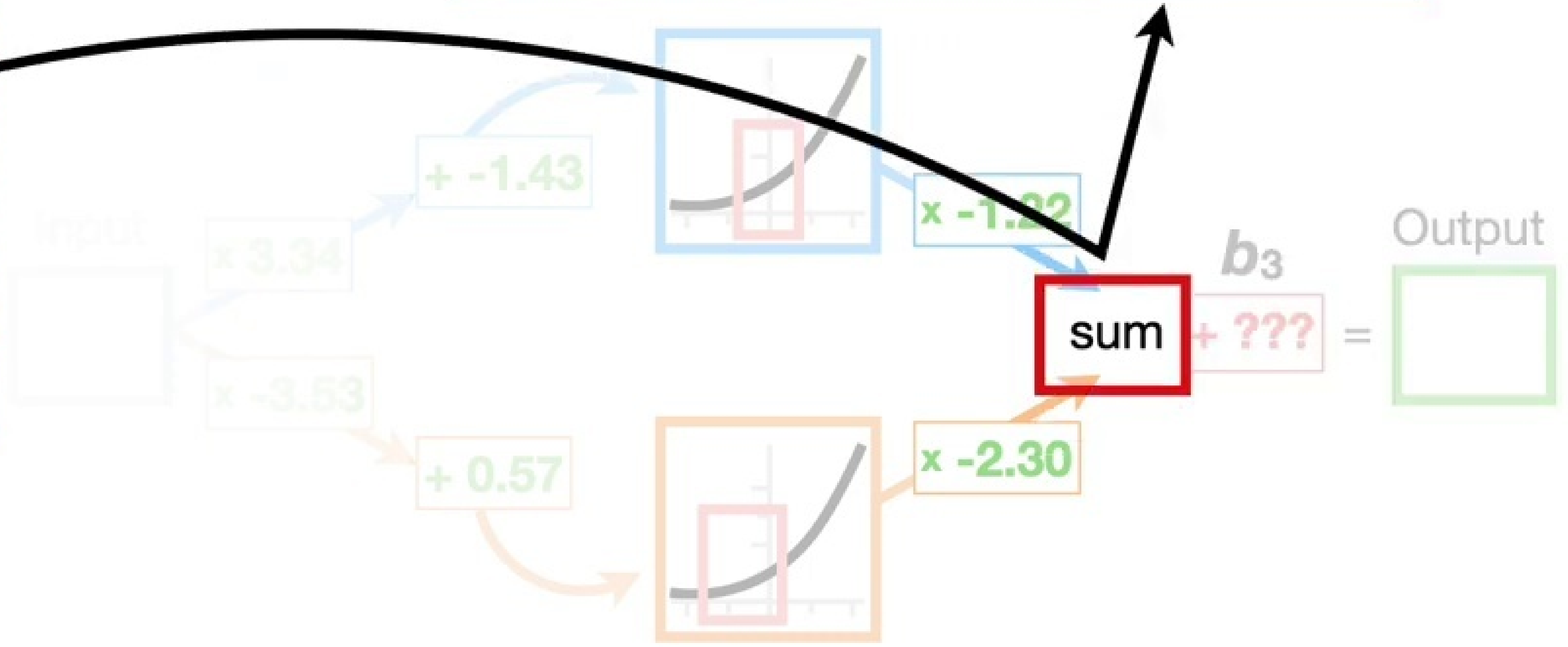


In other words, the **green squiggle** is the **sum** of the **blue** and **orange** curves…

$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

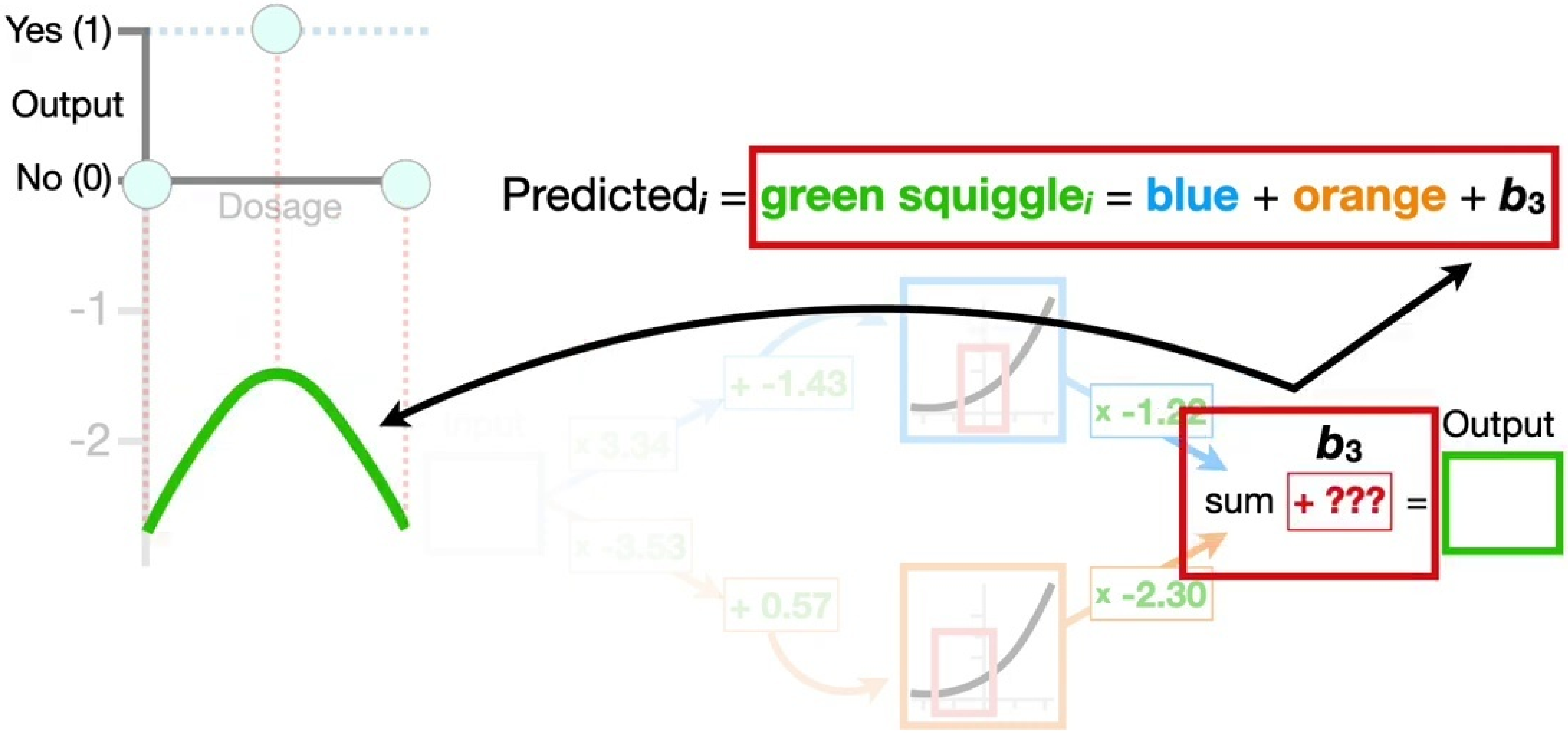


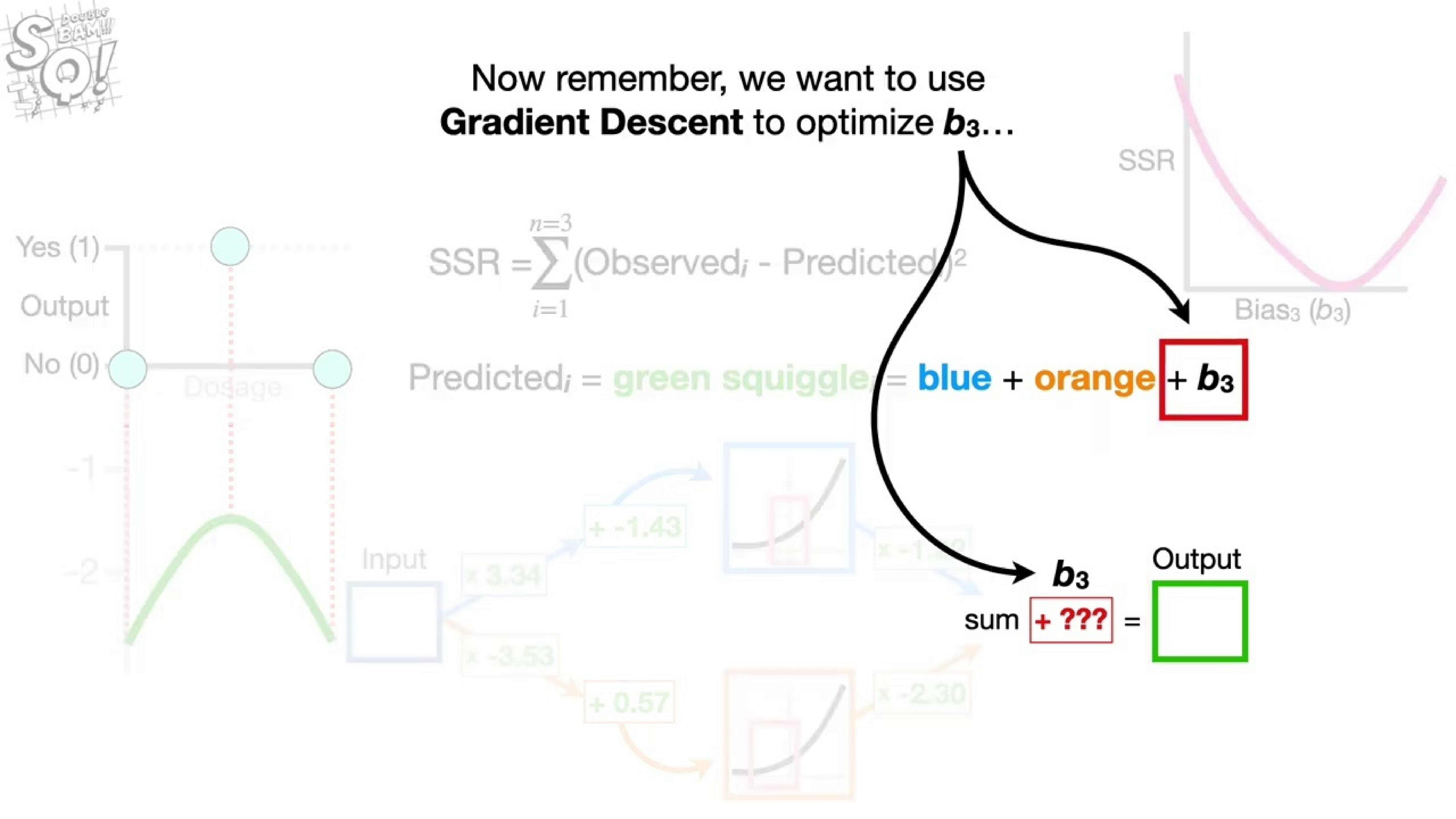
$$\text{Predicted}_i = \boxed{\text{green squiggle}_i = \text{blue} + \text{orange}} + b_3$$



…plus b_3.

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

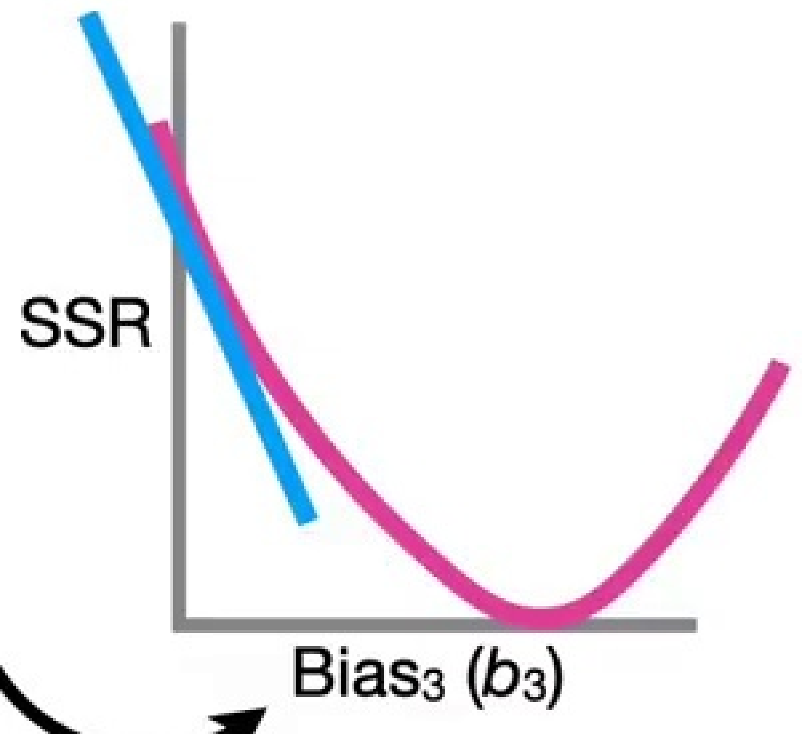


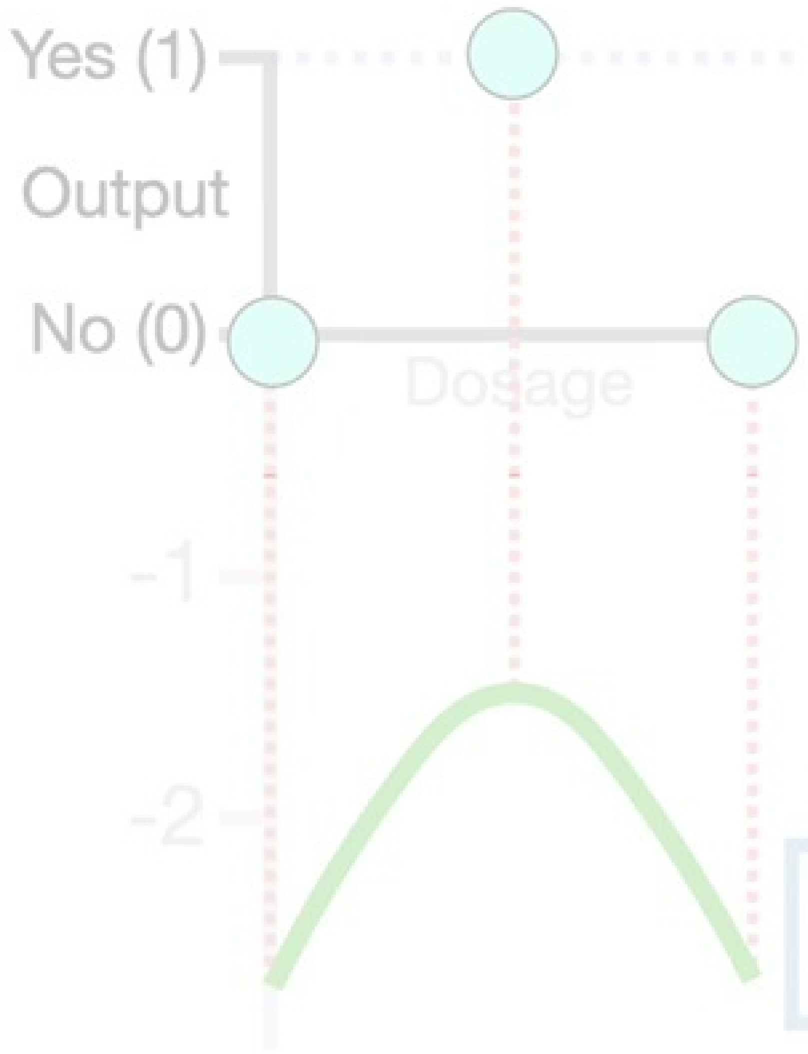


Now remember, we want to use
Gradient Descent to optimize b_3…
SSR
Bias₃ (b_3)
Yes (1)
Output
No (0)
Dosage
$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$
Predicted$_i$ = green squiggle$_i$ = blue + orange $\boxed{+ b_3}$
-1
-2
Input
x 3.34
+ -1.43
x -3.53
+ 0.57
x -2.30
b_3
Output
sum $\boxed{+ ???}$ =

$$\frac{d\ SSR}{d\ b_3}$$

...and that means we need to take the derivative of the **SSR** with respect to $\boldsymbol{b_3}$.



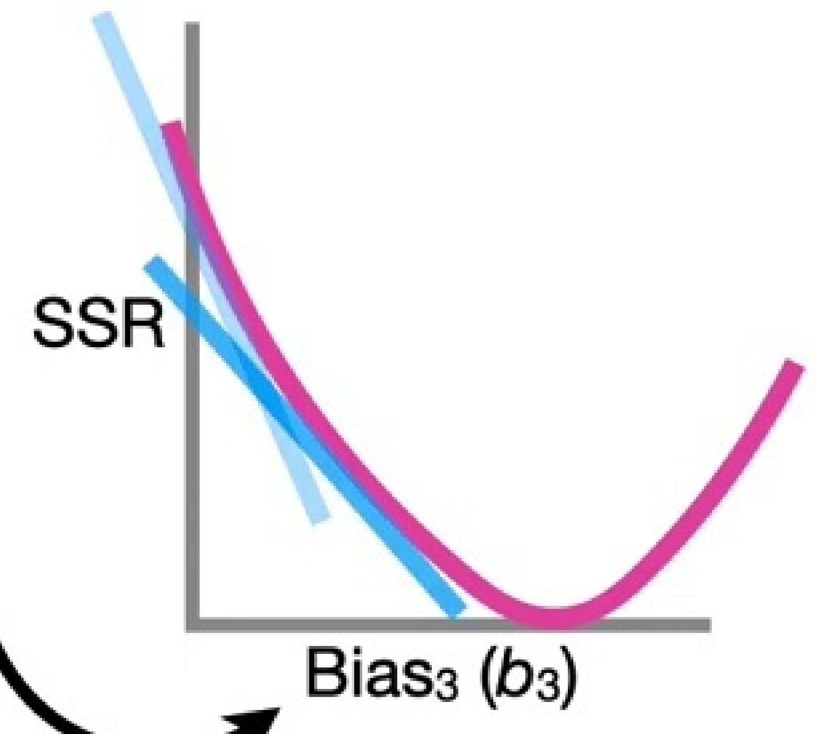


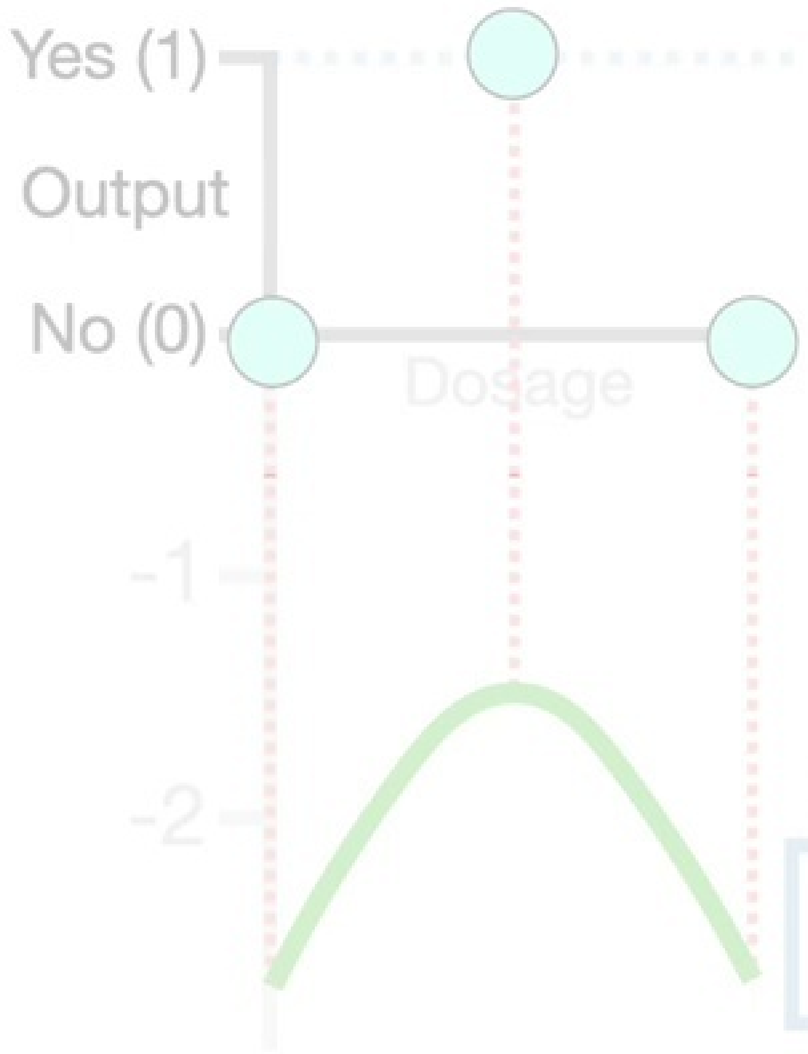
$$\text{SSR} = \sum_{i=1}^{n=3}(\text{Observed}_i - \text{Predicted}_i)^2$$

$$\text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3}$$

…and that means we need to take the derivative of the **SSR** with respect to b_3.



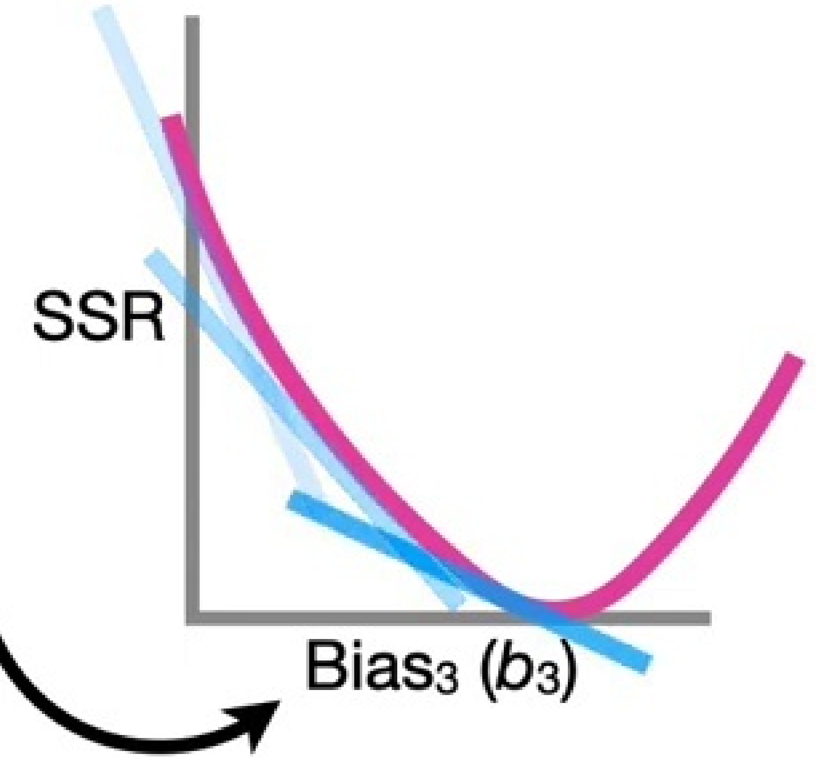


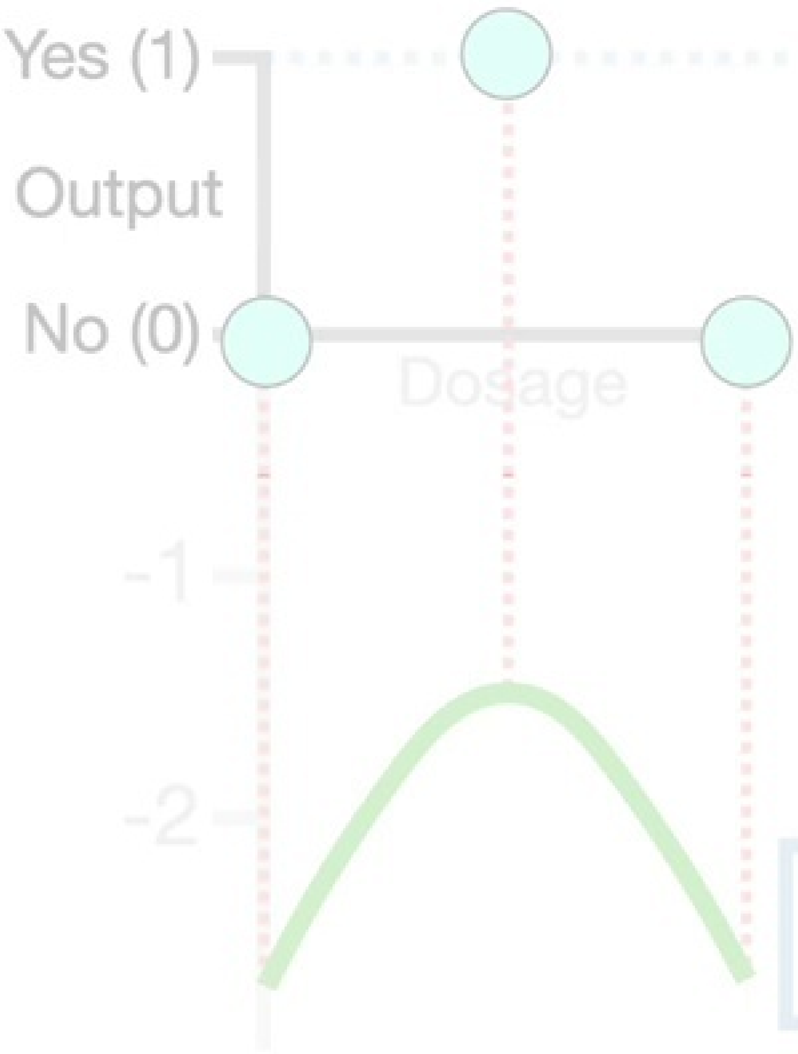
$$SSR = \sum_{i=1}^{n=3}(Observed_i - Predicted_i)^2$$

$$Predicted_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3}$$

…and that means we need to take the derivative of the **SSR** with respect to b_3.



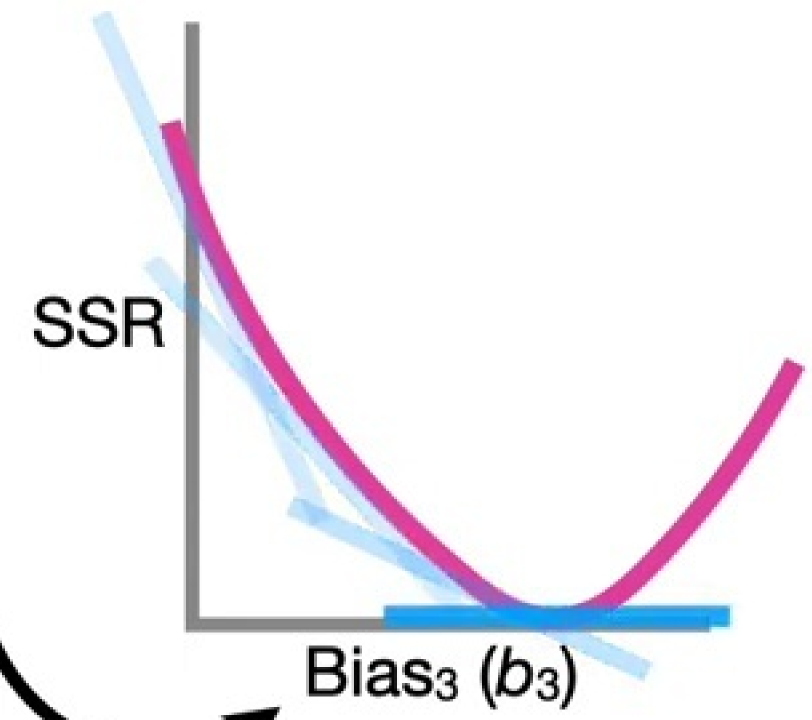


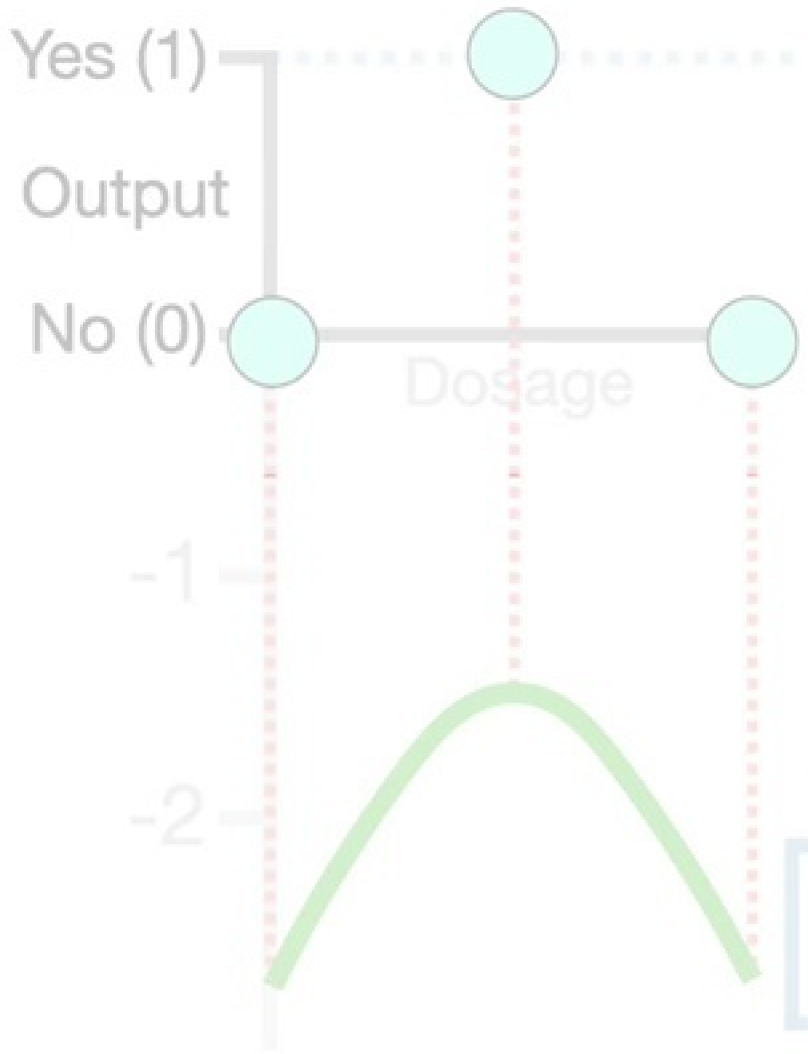
$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

$$\text{Predicted}_i = \textbf{green squiggle}_i = \textbf{blue} + \textbf{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3}$$

…and that means we need to take the derivative of the **SSR** with respect to $\boldsymbol{b_3}$.





$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

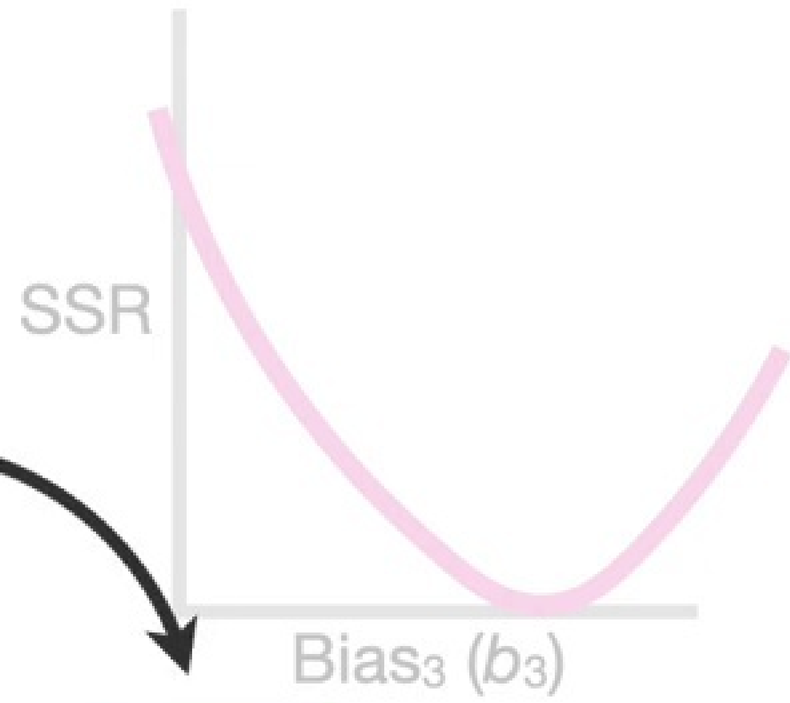
$$\text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

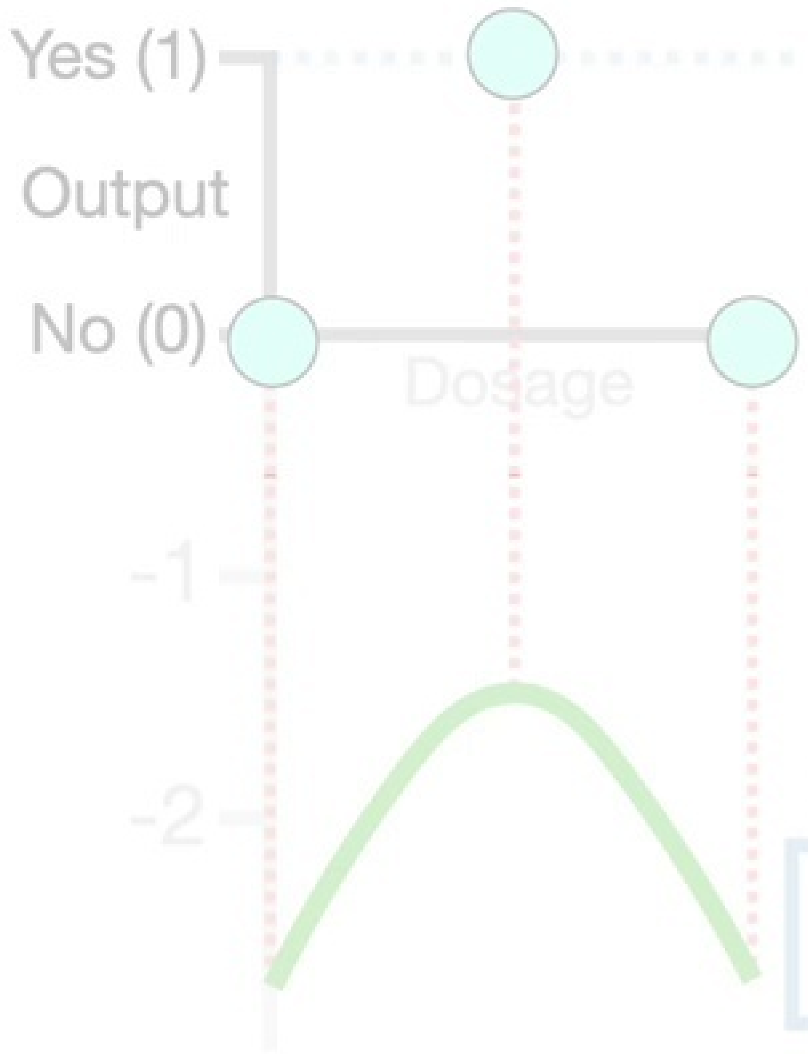
$$\frac{d\ SSR}{d\ b_3}$$

And because the **SSR** are linked to ***b*₃**…

$$\boxed{\text{SSR}} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

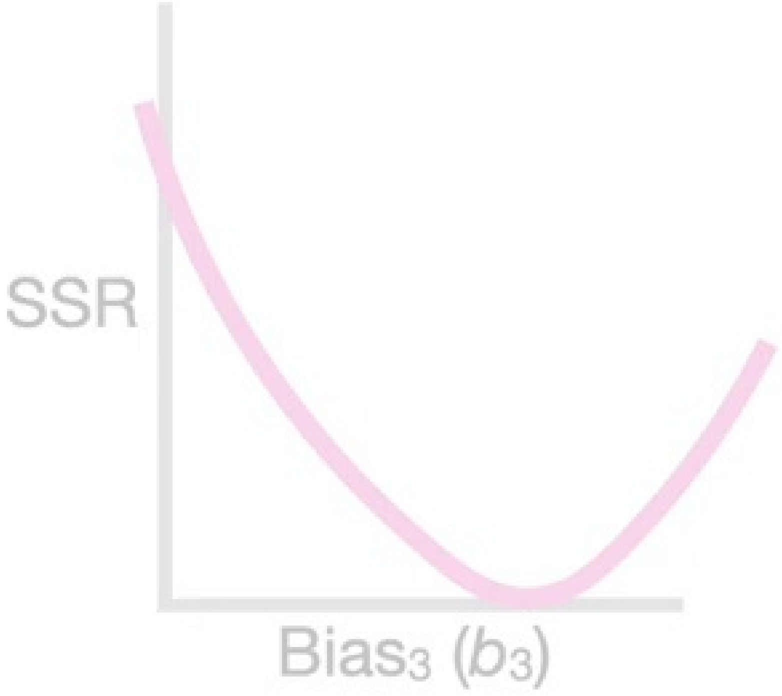
$$\text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} \boxed{+\ b_3}$$





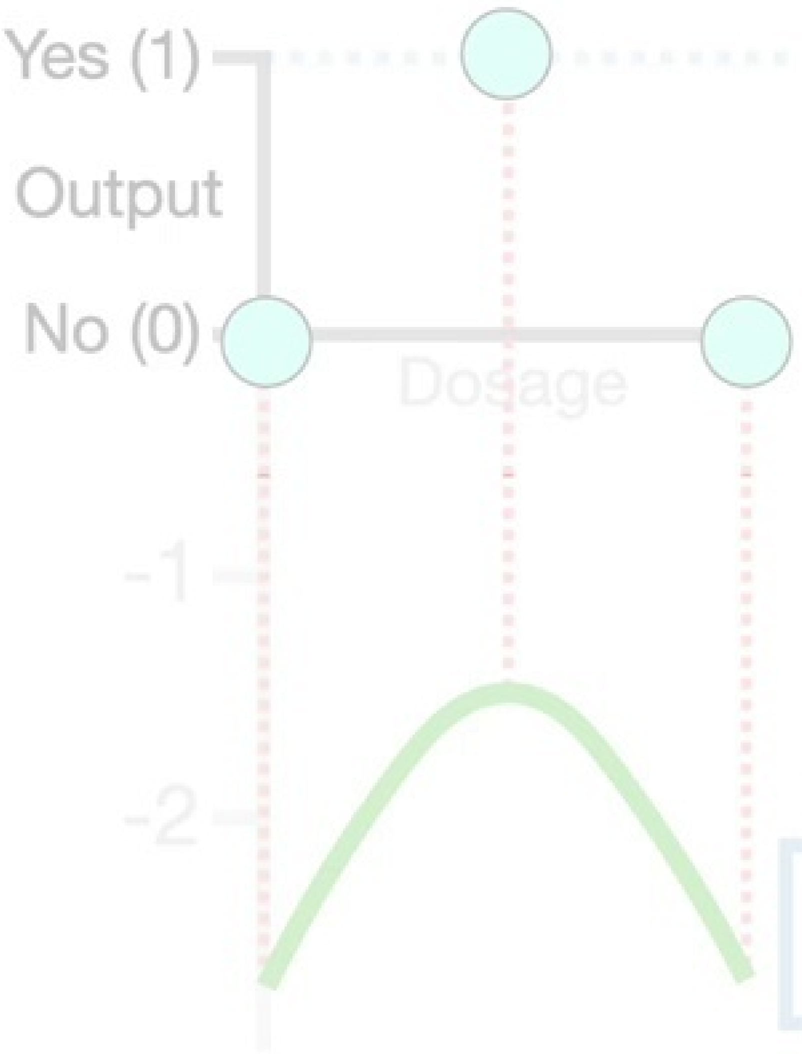
$$\frac{d\ SSR}{d\ b_3}$$

...by the **Predicted** values...



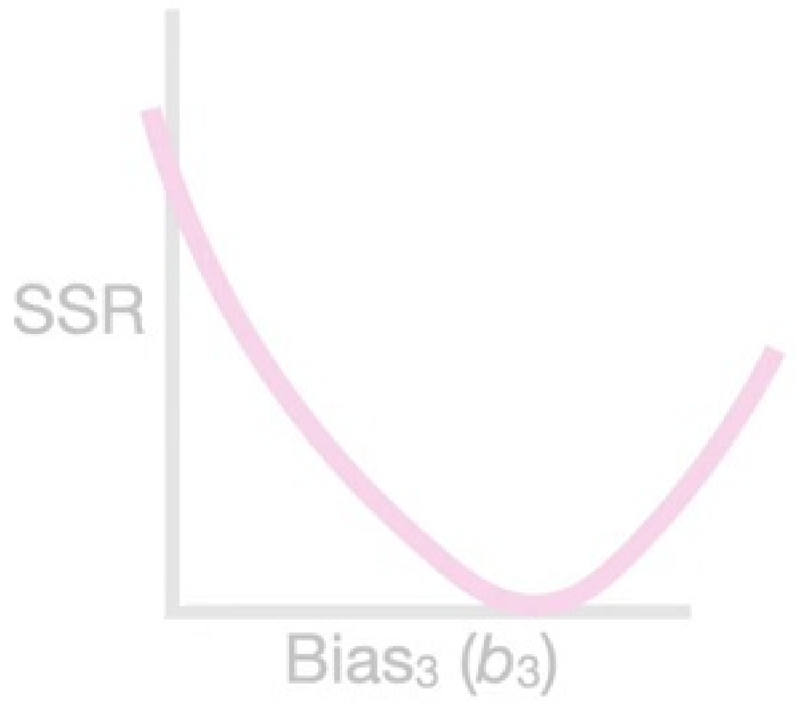
$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \boxed{\text{Predicted}_i})^2$$

$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$



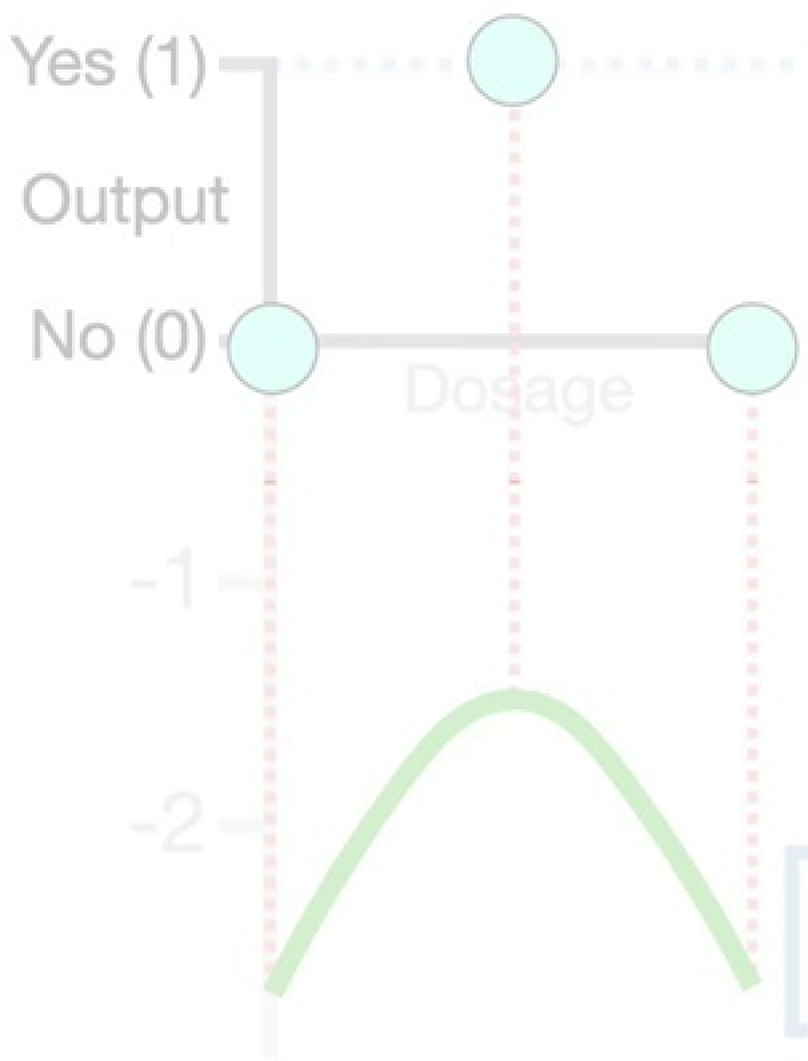
$$\frac{d\ SSR}{d\ b_3}$$

...we can use
The Chain Rule...



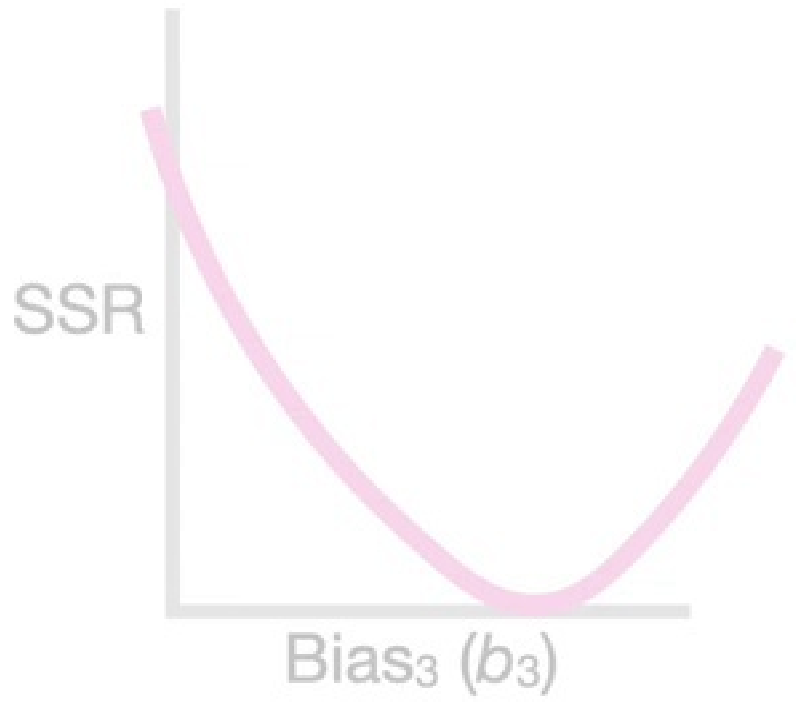
$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \boxed{\text{Predicted}_i})^2$$

$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$



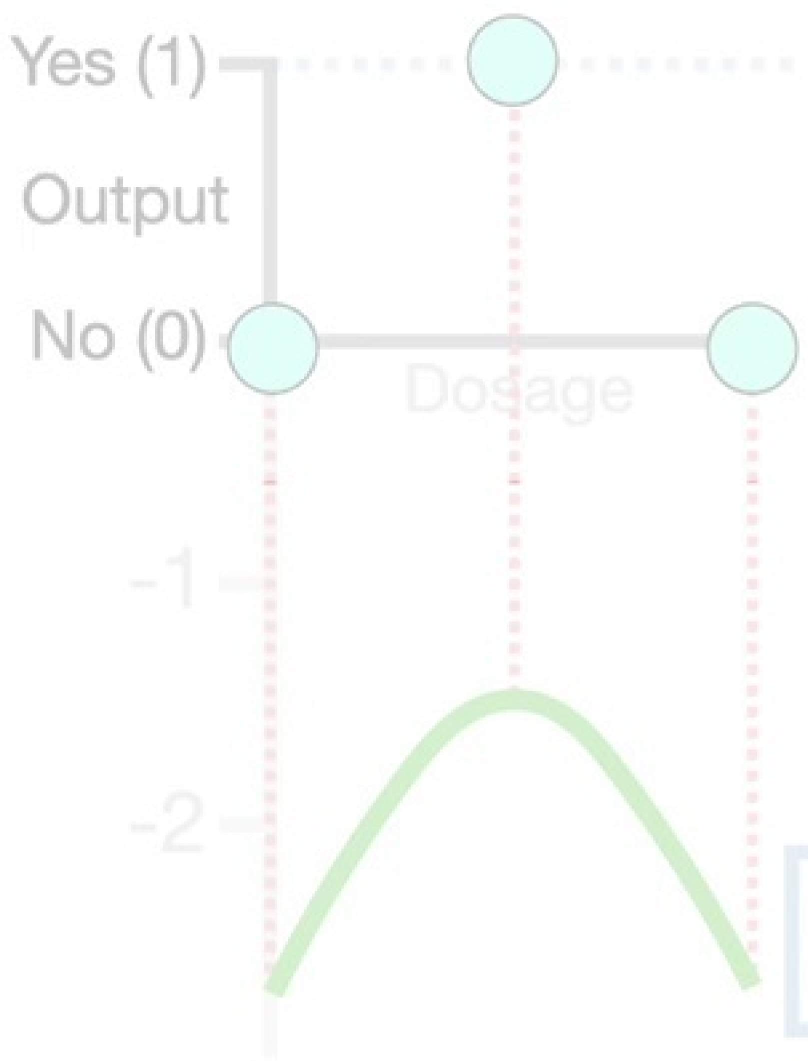
$$\frac{d\ SSR}{d\ b_3} =$$

...to solve for the derivative of the **SSR** with respect to b_3.



$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \boxed{\text{Predicted}_i})^2$$

$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

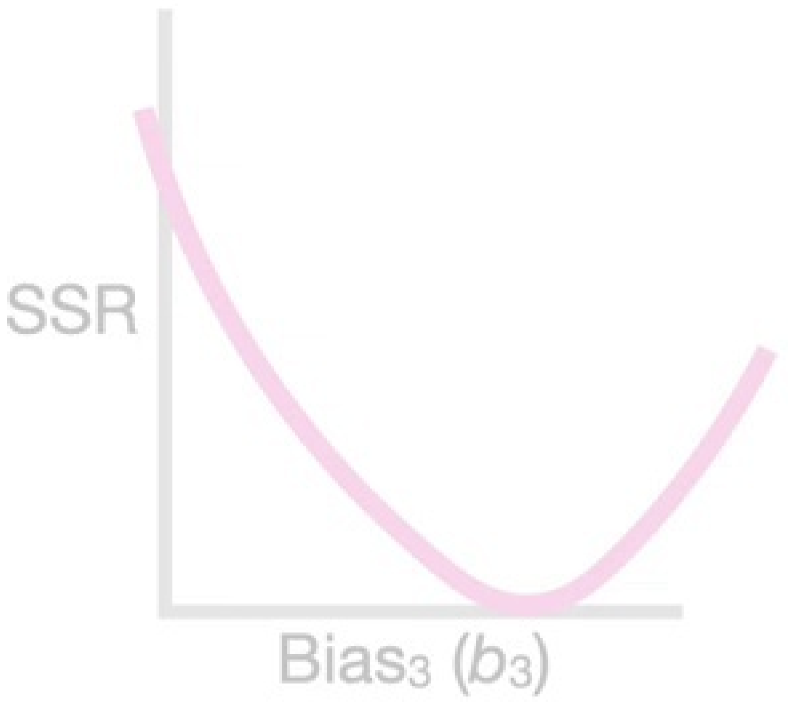


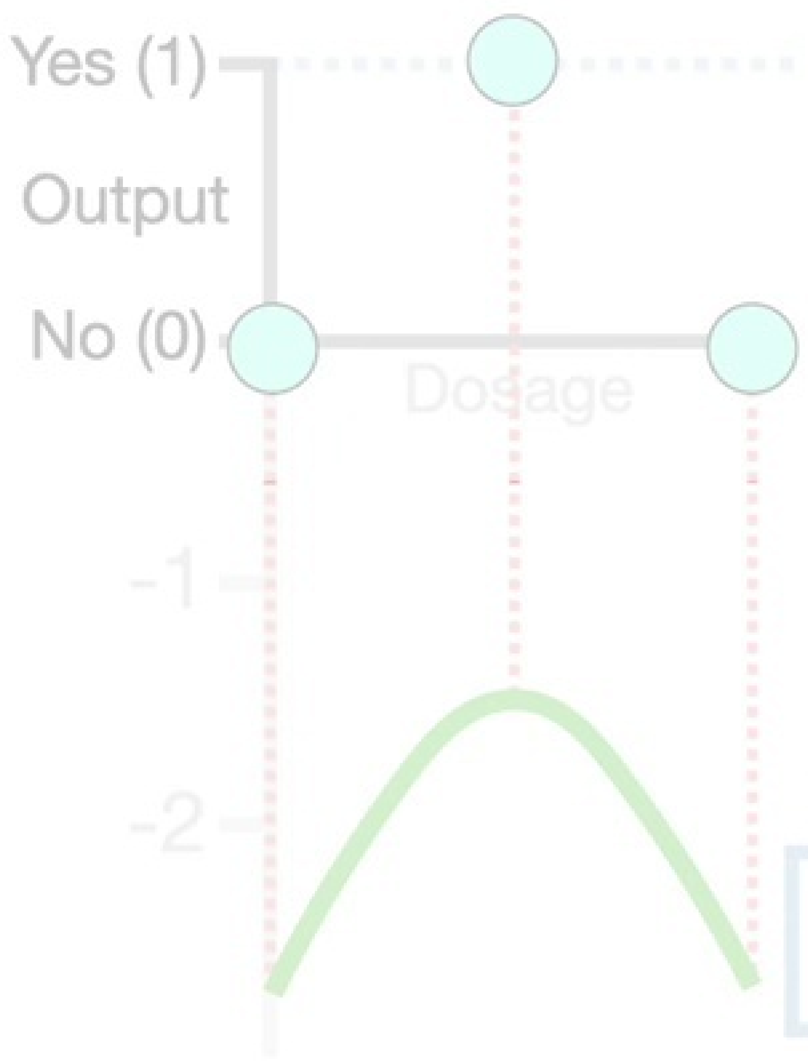
$$\frac{d\,SSR}{d\,b_3} =$$

The Chain Rule says that the derivative of the **SSR** with respect to ***b*₃**...

$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \boxed{\text{Predicted}_i})^2$$

$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

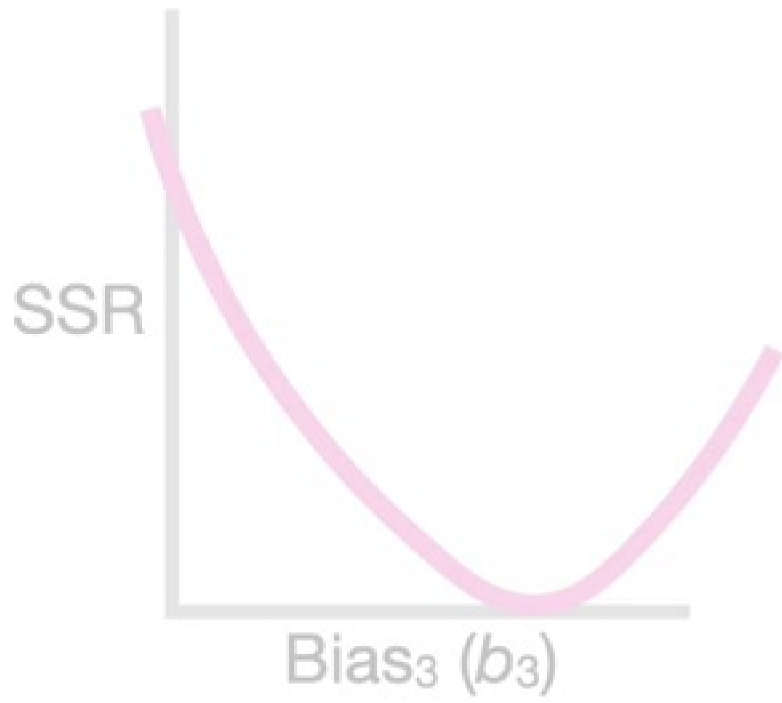




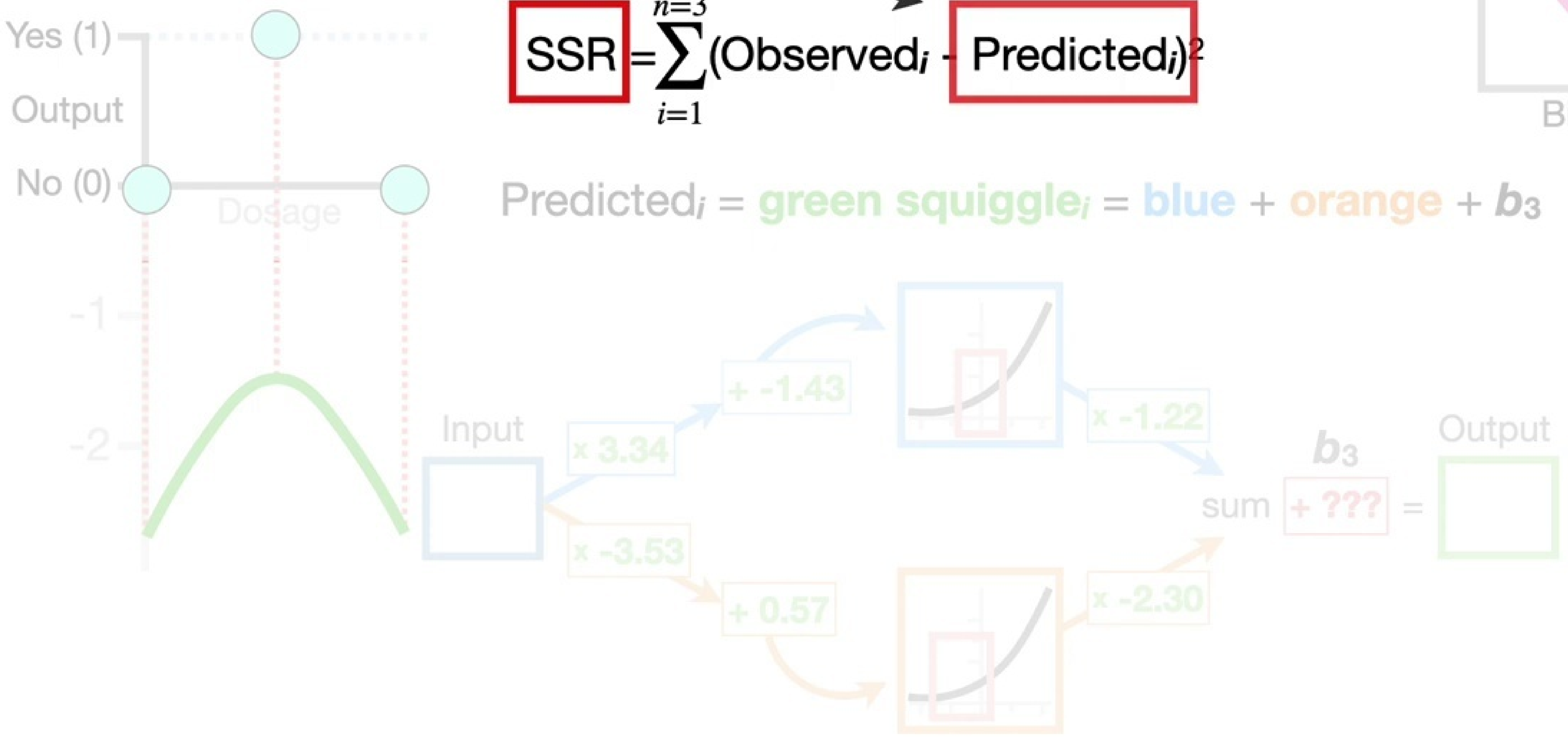
$$\frac{d\,SSR}{d\,b_3} = \frac{d\,SSR}{d\,Predicted}$$

...is the derivative of the **SSR** with respect to the **Predicted** values...

$$\boxed{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \boxed{\text{Predicted}_i})^2$$

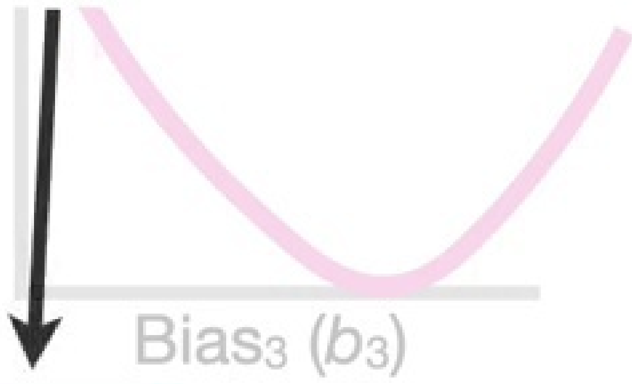


Predicted$_i$ = green squiggle$_i$ = blue + orange + b_3



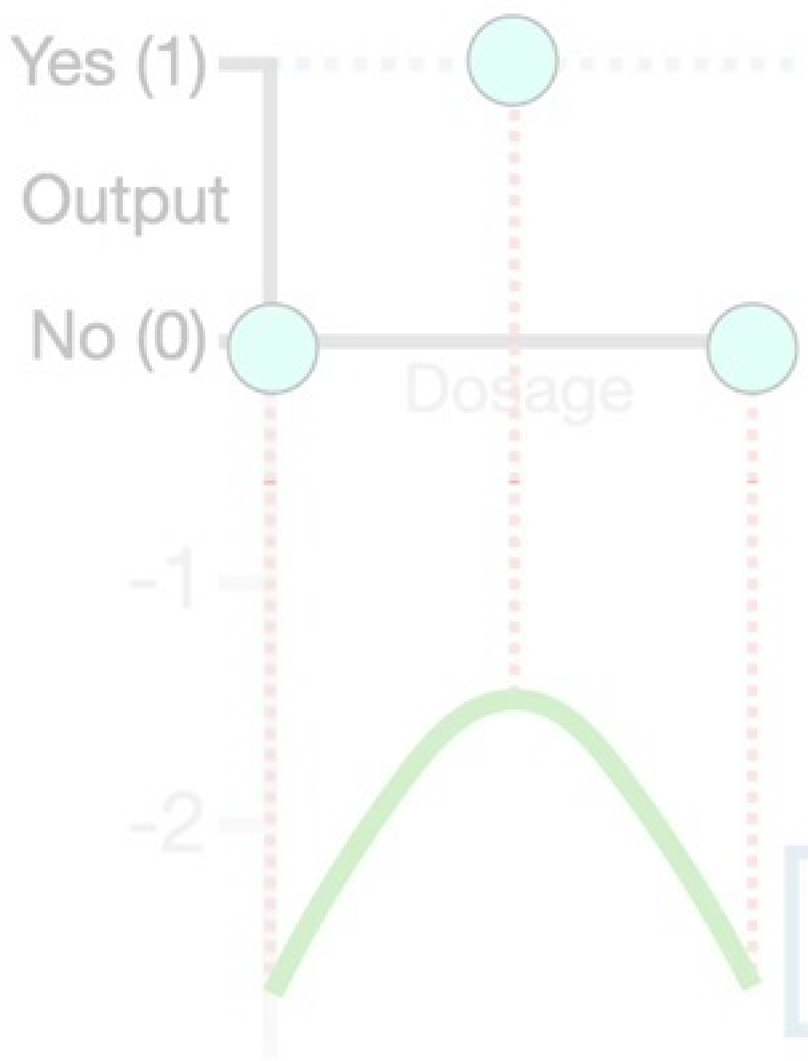
$$\frac{d\,SSR}{d\,b_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,b_3}$$

…times the derivative of the **Predicted** values with respect to $\boldsymbol{b_3}$.



$$SSR = \sum_{i=1}^{n=3}(Observed_i - \boxed{Predicted_i})^2$$

$$\boxed{Predicted_i} = \text{green squiggle}_i = \text{blue} + \text{orange} \boxed{+\, b_3}$$



$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

Now, we can solve for the derivative of the **SSR** with respect to the **Predicted** values by first substituting in the equation…

$$\frac{d\ SSR}{d\ Predicted} =$$

$$\text{SSR} = \sum_{i=1}^{n=3}(\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

Now, we can solve for the derivative of the **SSR** with respect to the **Predicted** values by first substituting in the equation…

$$\frac{d\ SSR}{d\ Predicted} = \frac{d}{d\ Predicted} \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

$$\boxed{\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2}$$

$\text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

...and then use **The Chain Rule**...

$$\frac{d\ SSR}{d\ Predicted} = \frac{d}{d\ Predicted} \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

...and we have the derivative of the **SSR** with respect to the **Predicted** values.

$$\frac{d\ SSR}{d\ Predicted} = \frac{d}{d\ Predicted} \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 = \boxed{\sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)}$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \textbf{green squiggle}_i = \textbf{blue} + \textbf{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

So let's move that up here…

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)$$

$$\frac{d\ SSR}{d\ Predicted} = \frac{d}{d\ Predicted} \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \boxed{= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)}$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \boxed{\frac{d\ SSR}{d\ Predicted}} \times \frac{d\ Predicted}{d\ b_3}$$

...and now we are done with the first part.

$$\boxed{= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)}$$

$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

Now let's solve for the second part, the derivative of the **Predicted** values with respect to b_3.

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

We start by plugging in the equation for the **Predicted** values.

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)$$

$$\frac{d\ Predicted}{d\ b_3}$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

Predicted_i = **green squiggle$_i$** = **blue** + **orange** + b_3

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

We start by plugging in the equation for the **Predicted** values.

$$= \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\ Predicted}{d\ b_3} = \frac{d}{d\ b_3}\ \text{green squiggle}$$

$$SSR = \sum_{i=1}^{n=3} (Observed_i - Predicted_i)^2$$

$$\boxed{Predicted_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3}$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

We start by plugging in the equation for the **Predicted** values.

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)$$

$$\frac{d\ Predicted}{d\ b_3} = \frac{d}{d\ b_3}\ \text{green squiggle} = \frac{d}{d\ b_3}(\text{blue} + \text{orange} + b_3)$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

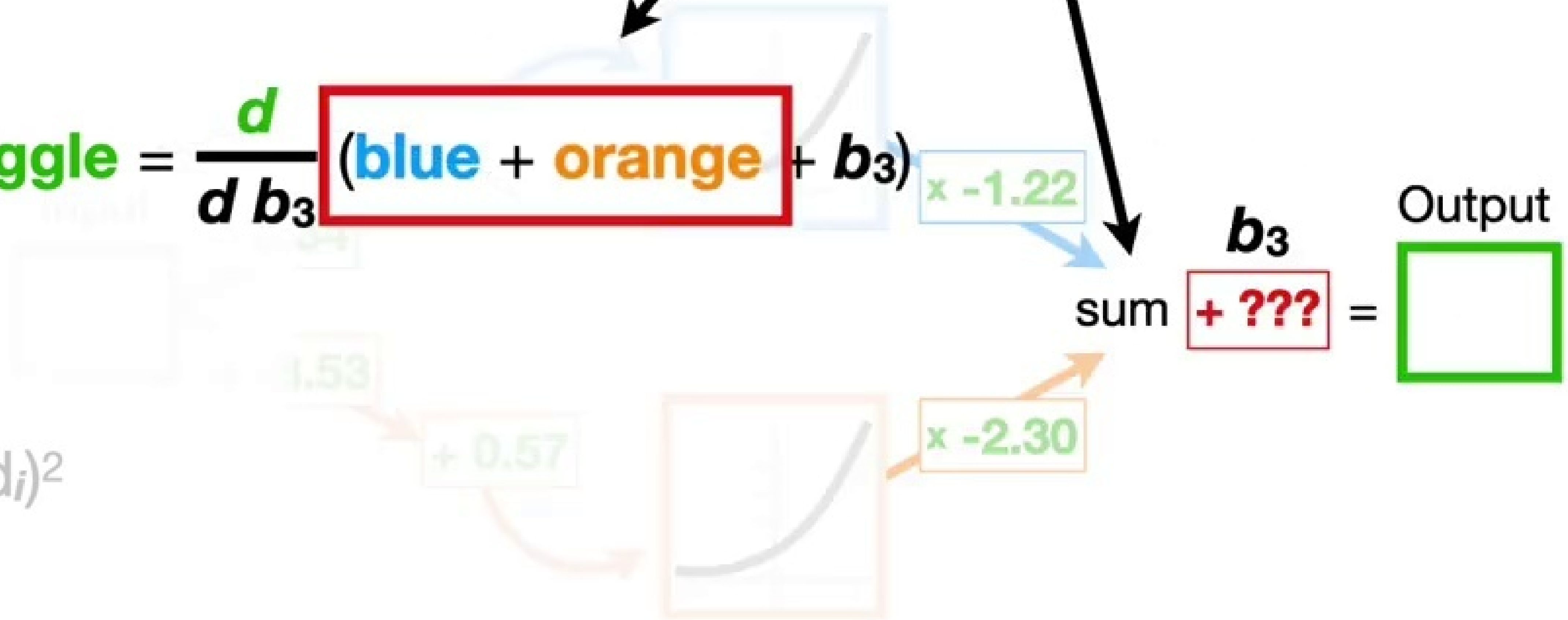
$$\boxed{\text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3}$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

Remember, the **blue** and **orange** curves were created before we got to b_3…

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)$$

$$\frac{d\ Predicted}{d\ b_3} = \frac{d}{d\ b_3}\ \text{green squiggle} = \frac{d}{d\ b_3}(\boxed{\text{blue} + \text{orange}} + b_3)$$



$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

$$\frac{d\,SSR}{d\,b_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,b_3}$$

…so the derivative of the **blue curve** with respect to ***b*₃** is **0**, because the **blue curve** is independent of ***b*₃**…

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)$$

$$\frac{d\,Predicted}{d\,b_3} = \frac{d}{d\,b_3}\,\text{green squiggle} = \frac{d}{d\,b_3}(\boxed{\text{blue}} + \text{orange} + b_3) \boxed{= 0}$$

$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

...and the derivative of the **orange curve** with respect to b_3 is also **0**.

$$= \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\ Predicted}{d\ b_3} = \frac{d}{d\ b_3}\ \text{green squiggle} = \frac{d}{d\ b_3}(\text{blue} \boxed{+ \text{orange}} + b_3) = 0 \boxed{+ 0}$$

$$SSR = \sum_{i=1}^{n=3}(Observed_i - Predicted_i)^2 \quad Predicted_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

Lastly, the derivative of b_3 with respect to b_3 is **1**.

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)$$

$$\frac{d\ Predicted}{d\ b_3} = \frac{d}{d\ b_3}\ \text{green squiggle} = \frac{d}{d\ b_3}(\text{blue} + \text{orange}\ \boxed{+\ b_3}) = 0 + 0\ \boxed{+\ 1}$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

Now we just add everything up…

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)$$

$$\frac{d\ Predicted}{d\ b_3} = \frac{d}{d\ b_3}\ \text{green squiggle} = \frac{d}{d\ b_3}(\text{blue} + \text{orange} + b_3) \boxed{= 0 + 0 + 1}$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

...and the derivative of the **Predicted** values with respect to ***b*₃** is **1**.

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)$$

$$\frac{d\ Predicted}{d\ b_3} = \frac{d}{d\ b_3}\ \text{green squiggle} = \frac{d}{d\ b_3}(\text{blue} + \text{orange} + b_3) \boxed{= 1}$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

So we multiply the derivative of the **SSR** with respect to the **Predicted** values by **1**.

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \boxed{\times\ 1}$$

$$\frac{d\ Predicted}{d\ b_3} = \frac{d}{d\ b_3} \text{ green squiggle} = \frac{d}{d\ b_3} (\text{blue} + \text{orange} + b_3) = 1$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

NOTE: This "**times 1**" part in the equation doesn't do anything…

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \boxed{\times 1}$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times \mathbf{1}$$

…but I'm leaving it in to remind us that the derivative of the **SSR** with respect to b_3 consists of two parts.

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times 1$$

The derivative of the **SSR** with respect to the **Predicted** values…

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \boxed{\frac{d\ Predicted}{d\ b_3}}$$

...and the derivative of the **Predicted** values with respect to $\boldsymbol{b_3}$.

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \boxed{\times\ 1}$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

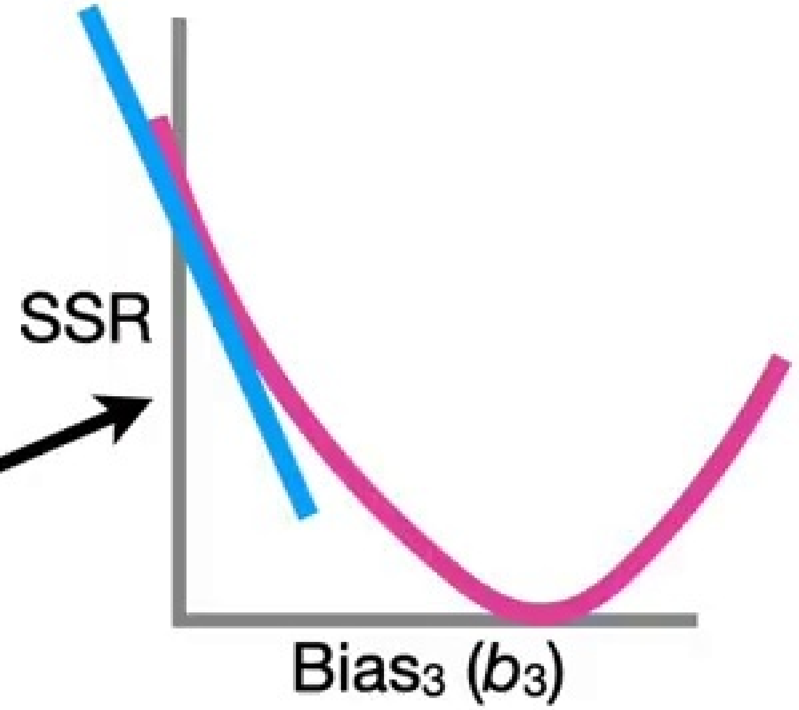
$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times \mathbf{1}$$

And at long last, we have the derivative of the **SSR** with respect to b_3.

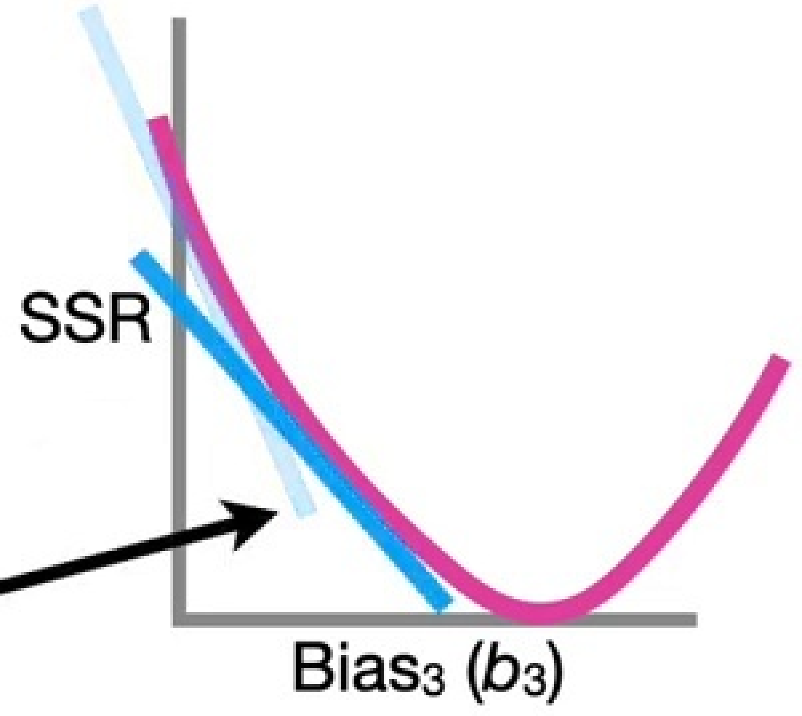
$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times \mathbf{1}$$



And that means we can plug this derivative into **Gradient Descent** to find the optimal value for b_3.

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

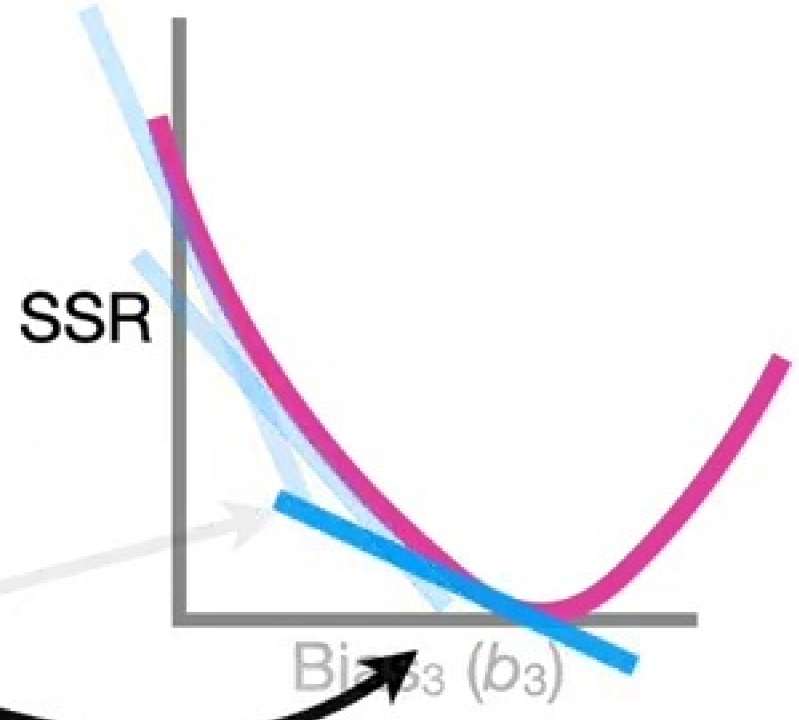


$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times \mathbf{1}$$

And that means we can plug this derivative into **Gradient Descent** to find the optimal value for b_3.

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

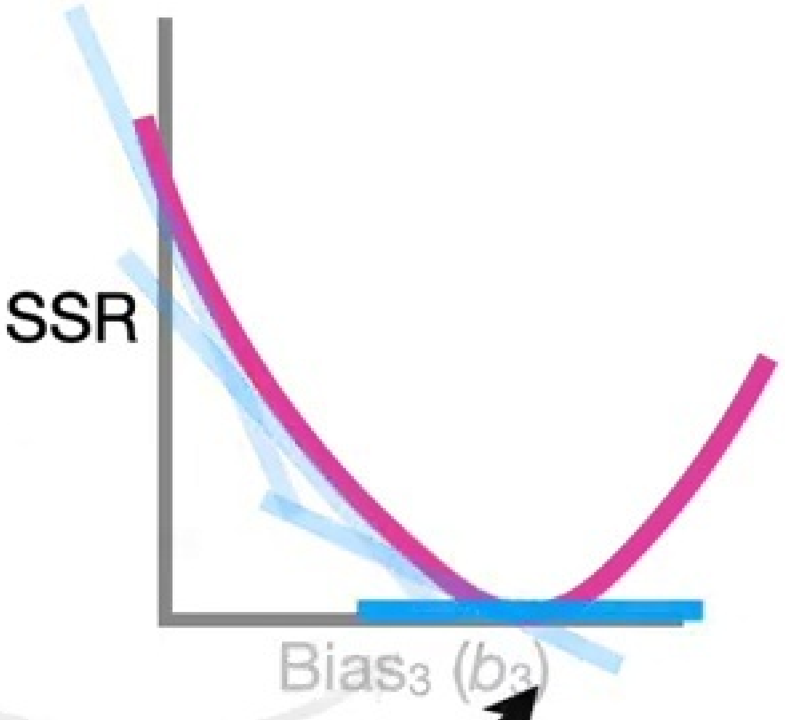
$$\boxed{= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times \mathbf{1}}$$



And that means we can plug this derivative into **Gradient Descent** to find the optimal value for $\boldsymbol{b_3}$.

$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

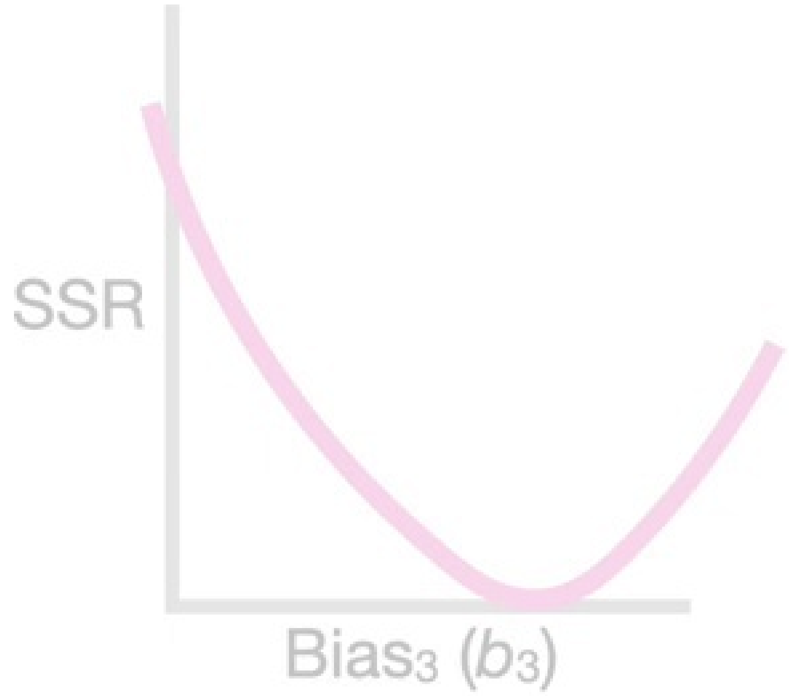
$$= \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times \mathbf{1}$$



And that means we can plug this derivative into **Gradient Descent** to find the optimal value for b_3.

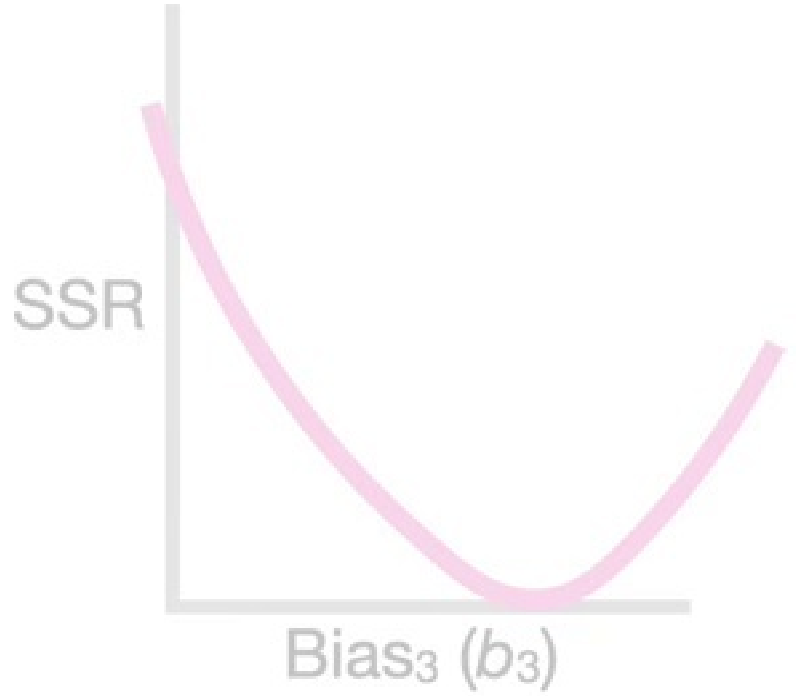
$$\frac{d\,SSR}{d\,b_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times \mathbf{1}$$

Anyway, first, we expand the summation.

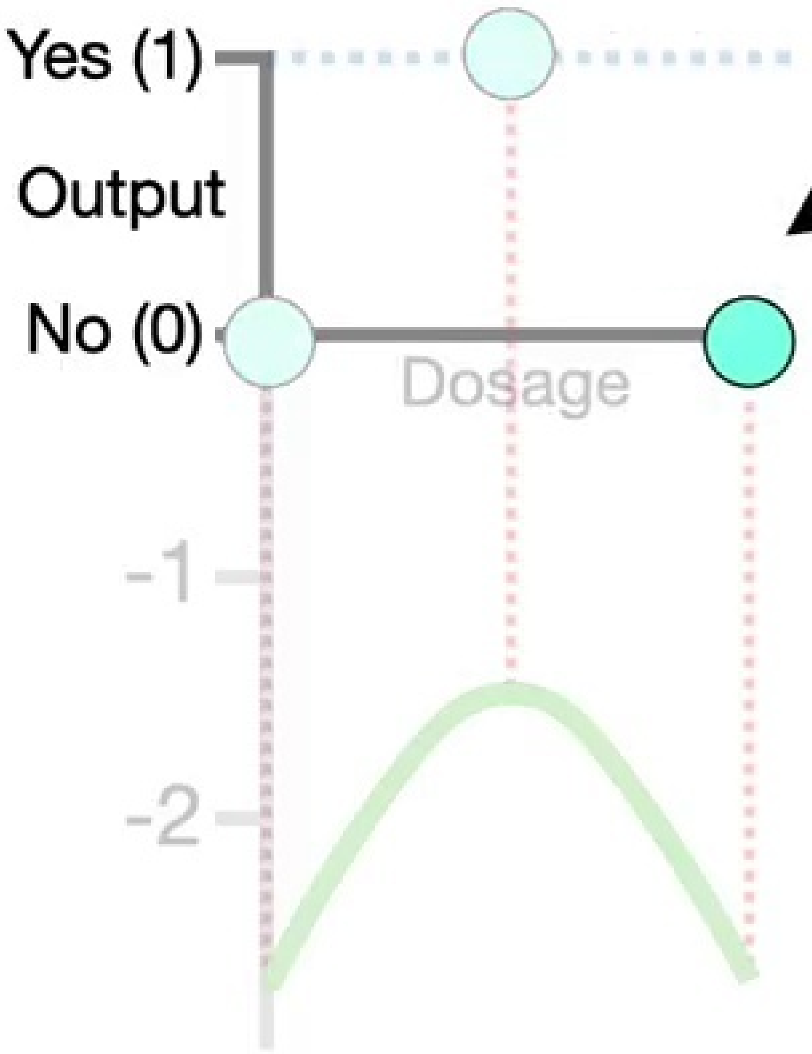


$$\frac{d\ SSR}{d\ b_3} = -2 \times (\text{Observed}_1 - \text{Predicted}_1) \times 1 + -2 \times (\text{Observed}_2 - \text{Predicted}_2) \times 1 + -2 \times (\text{Observed}_3 - \text{Predicted}_3) \times 1$$

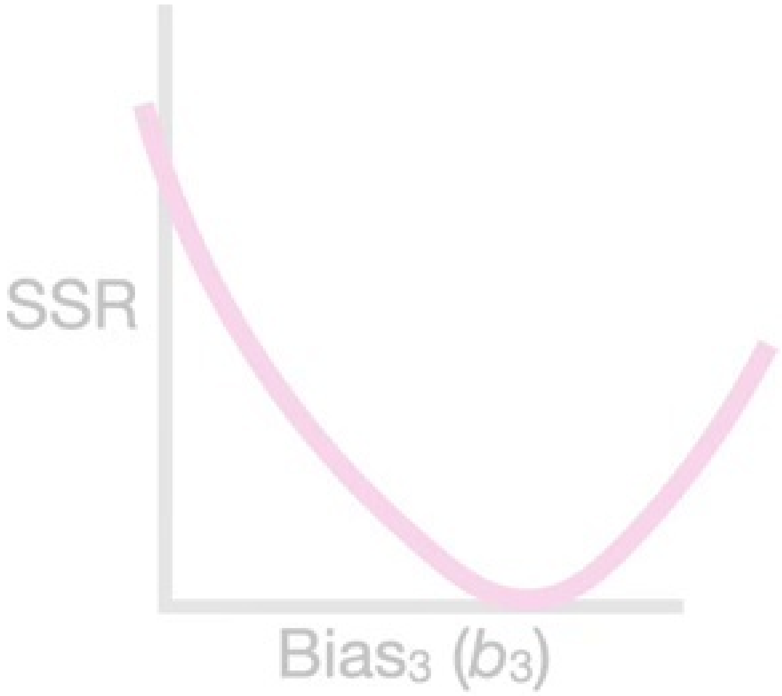
Anyway, first, we expand the summation.



$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - \text{Predicted}_1) \times 1$$
$$+ -2 \times (1 - \text{Predicted}_2) \times 1$$
$$+ -2 \times (0 - \text{Predicted}_3) \times 1$$

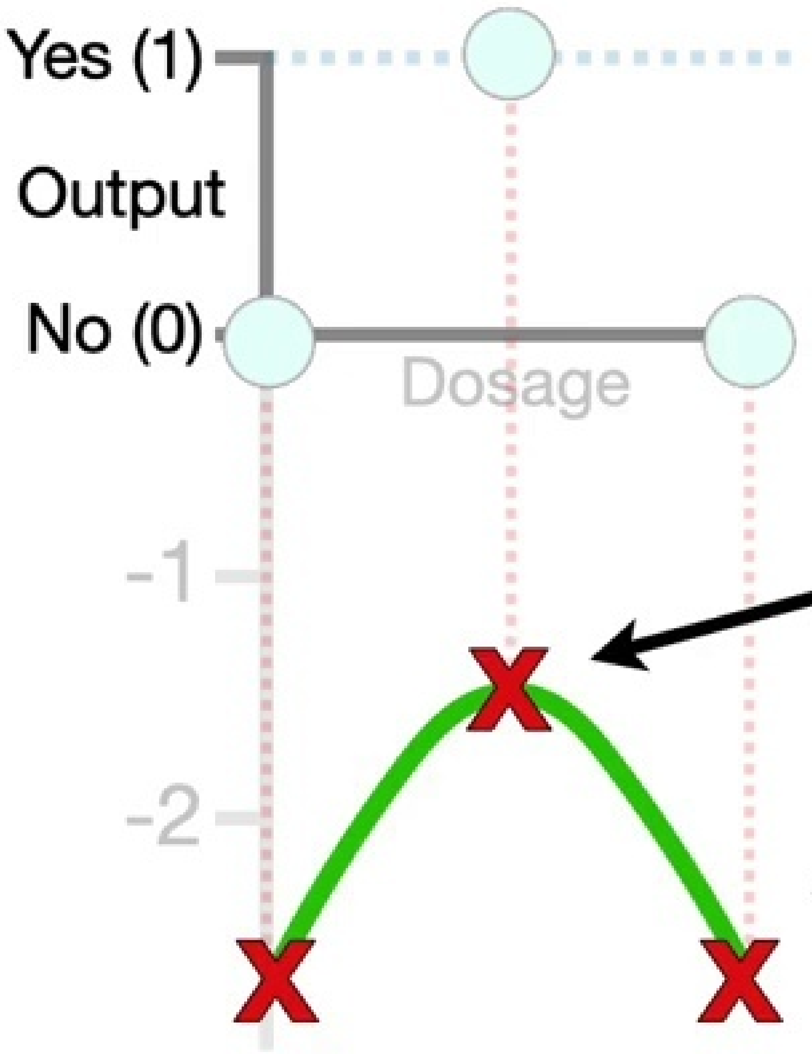


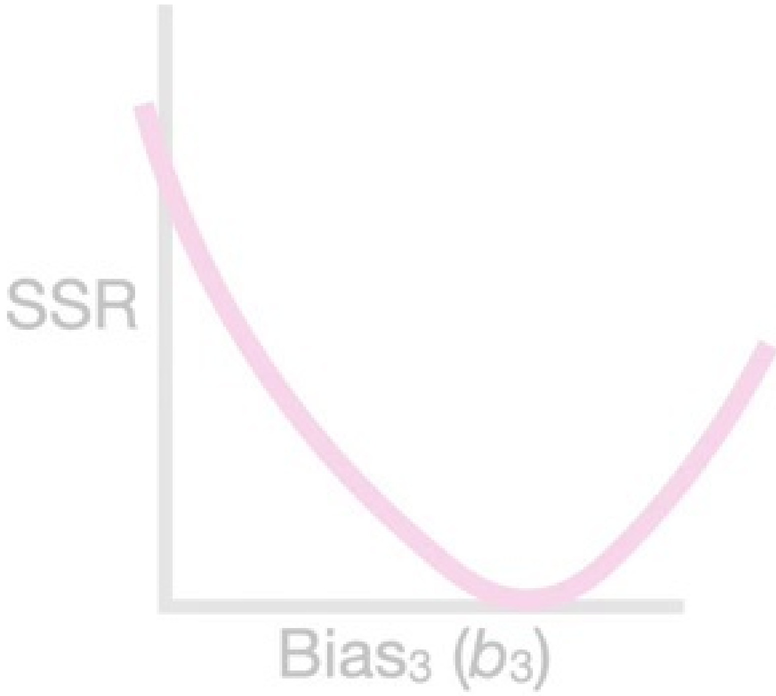
Then we plug in the **Observed** values….



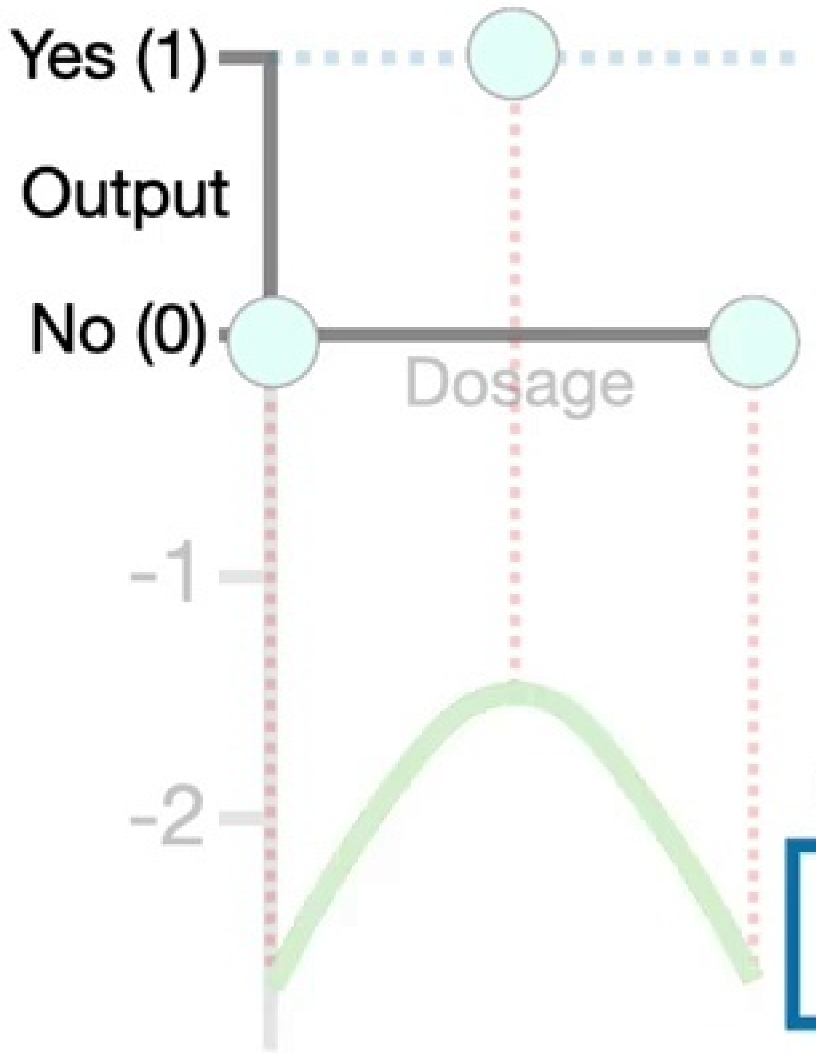
$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - \text{Predicted}_1) \times 1 + -2 \times (1 - \text{Predicted}_2) \times 1 + -2 \times (0 - \text{Predicted}_3) \times 1$$

...and the values **Predicted** by the **green squiggle**.

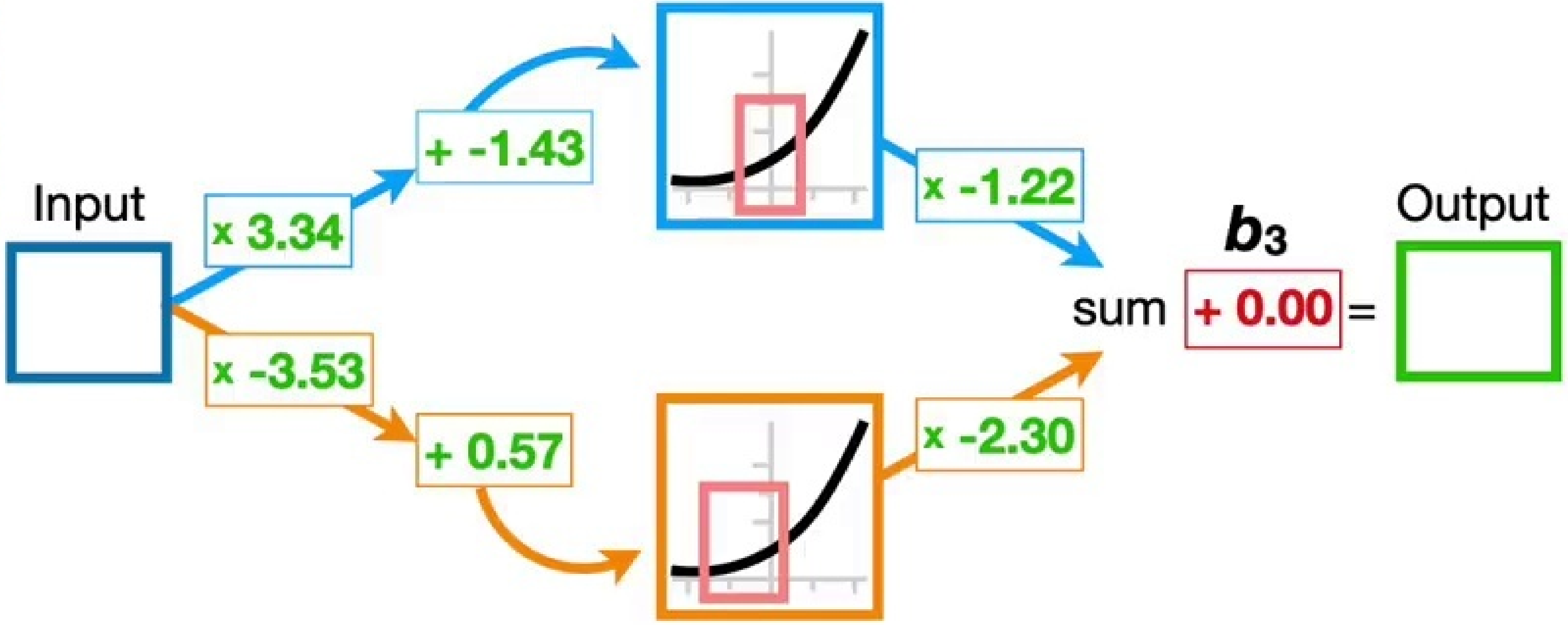




$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - \text{Predicted}_1) \times 1 + -2 \times (1 - \text{Predicted}_2) \times 1 + -2 \times (0 - \text{Predicted}_3) \times 1$$

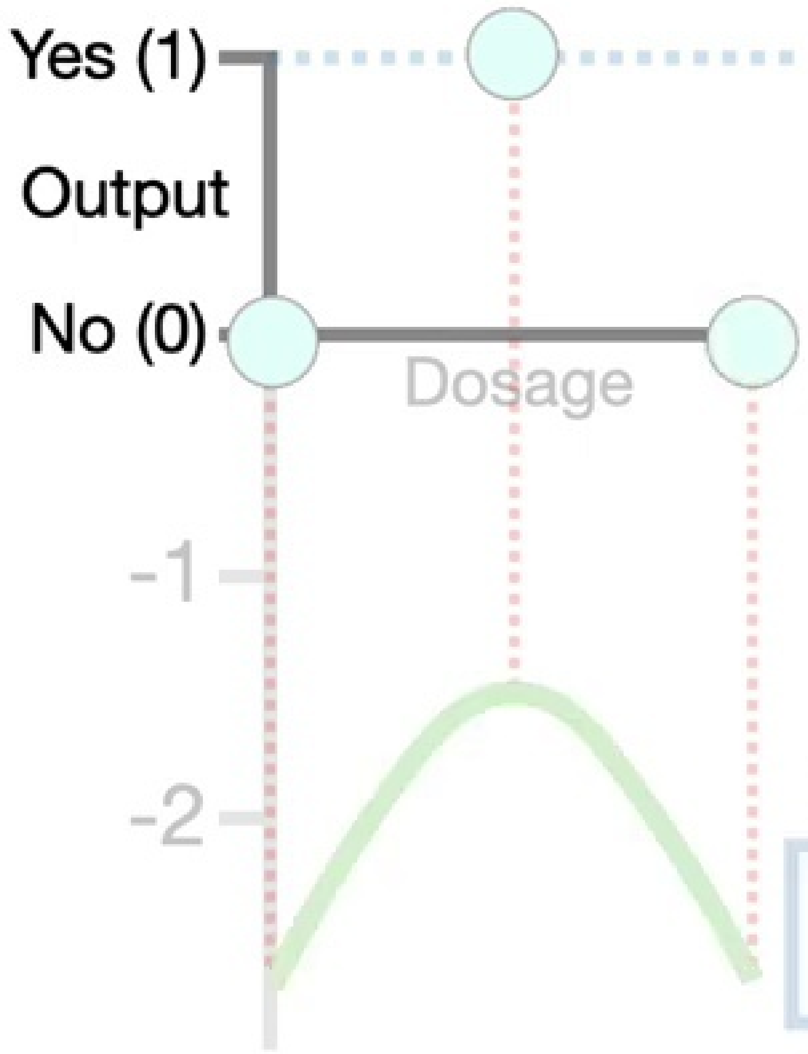


Remember, we get the **Predicted** values on the **green squiggle** by running the **Dosages** through the **Neural Network**.



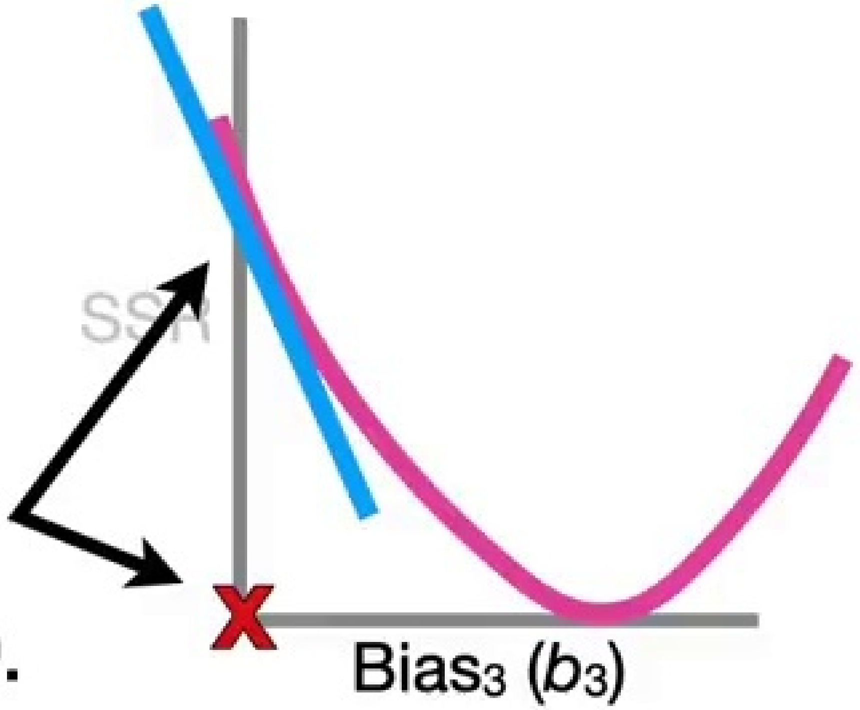
$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -2.6) \times \mathbf{1} + -2 \times (1 - -1.6) \times \mathbf{1} + -2 \times (0 - -2.61) \times \mathbf{1} \boxed{= -15.7}$$

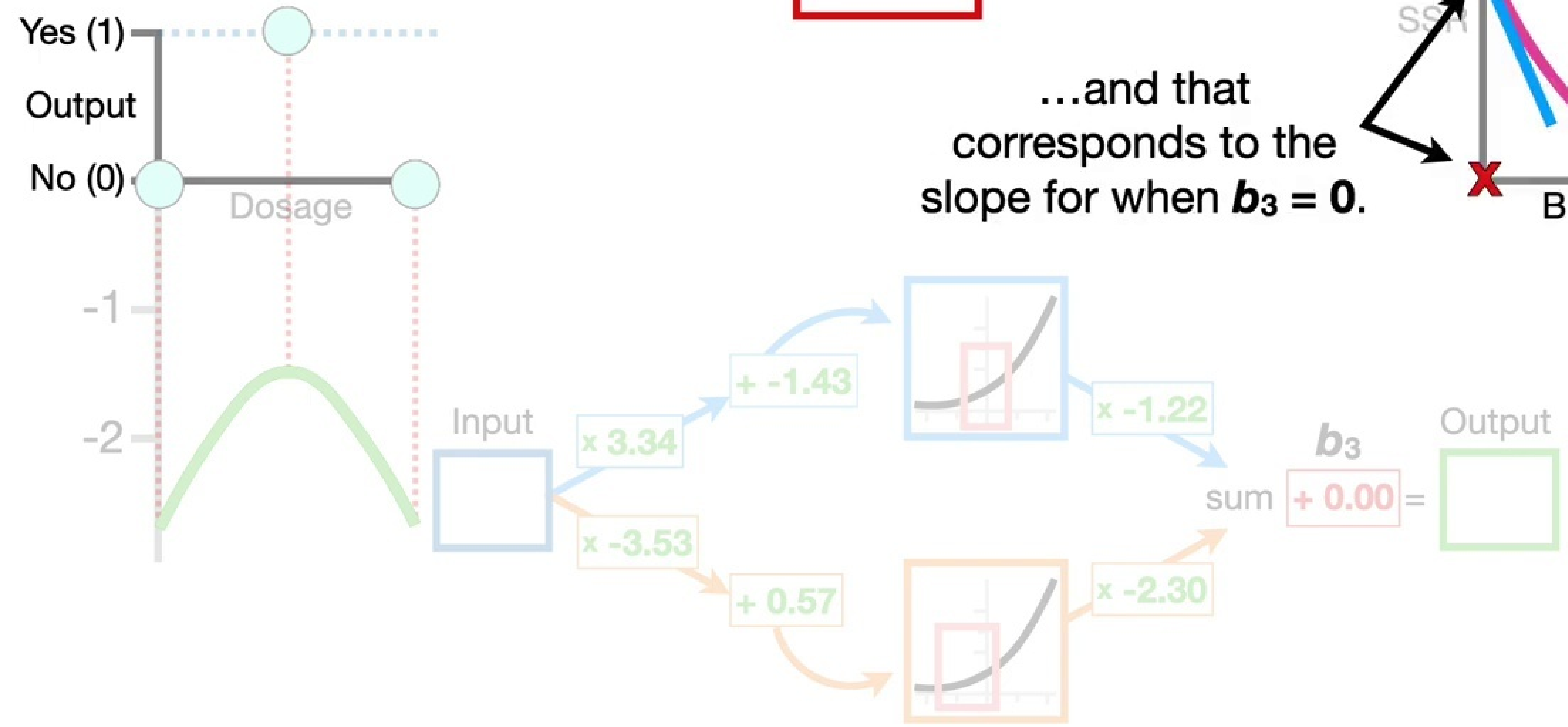
Now we just do the math and get **-15.7**…



$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -2.6) \times \mathbf{1} + -2 \times (1 - -1.6) \times \mathbf{1} + -2 \times (0 - -2.61) \times \mathbf{1} \boxed{= -15.7}$$

...and that corresponds to the slope for when $\boldsymbol{b_3} = \mathbf{0}$.

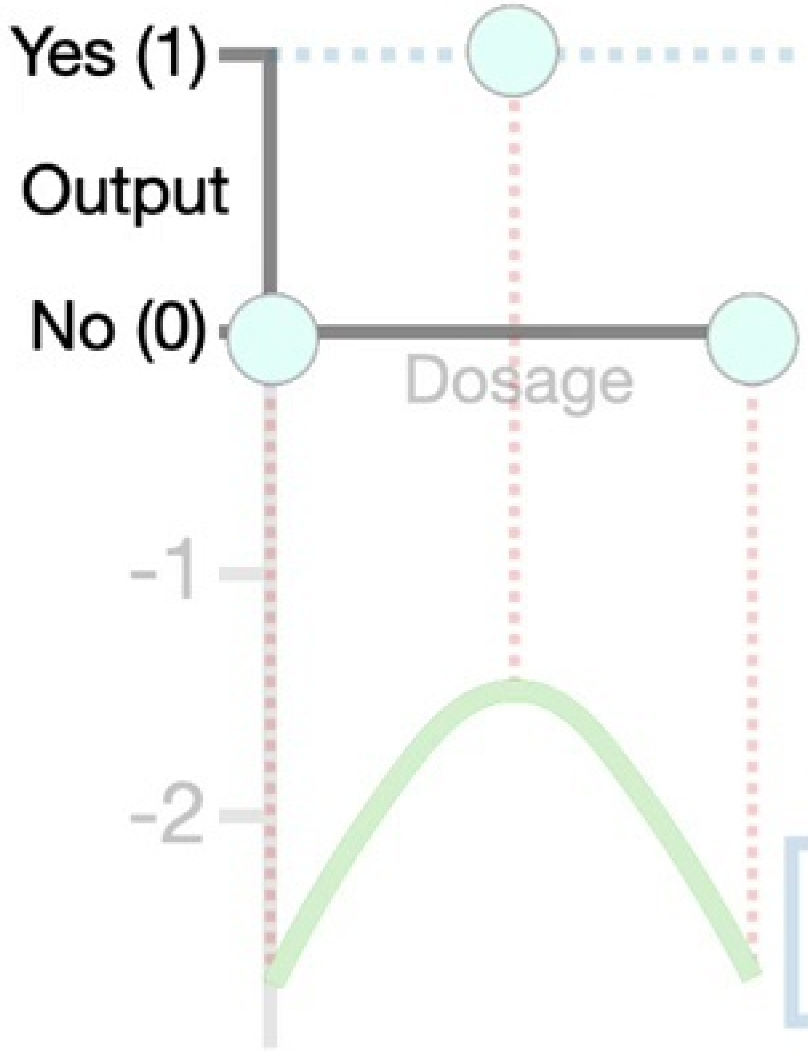


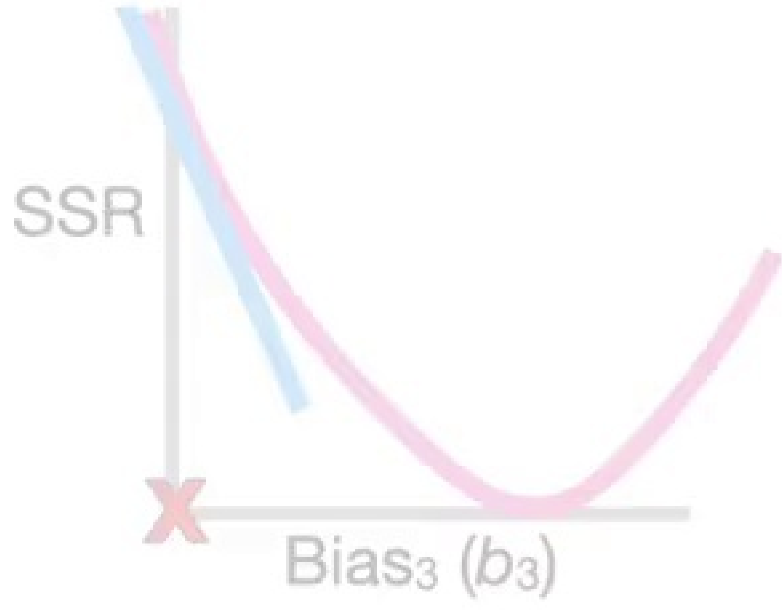


Now we plug the slope into the **Gradient Descent** equation for **Step Size**…

$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -2.6) \times 1 + -2 \times (1 - -1.6) \times 1 + -2 \times (0 - -2.61) \times 1 = \boxed{-15.7}$$

Step Size = $\boxed{\textbf{Slope}}$ × Learning Rate

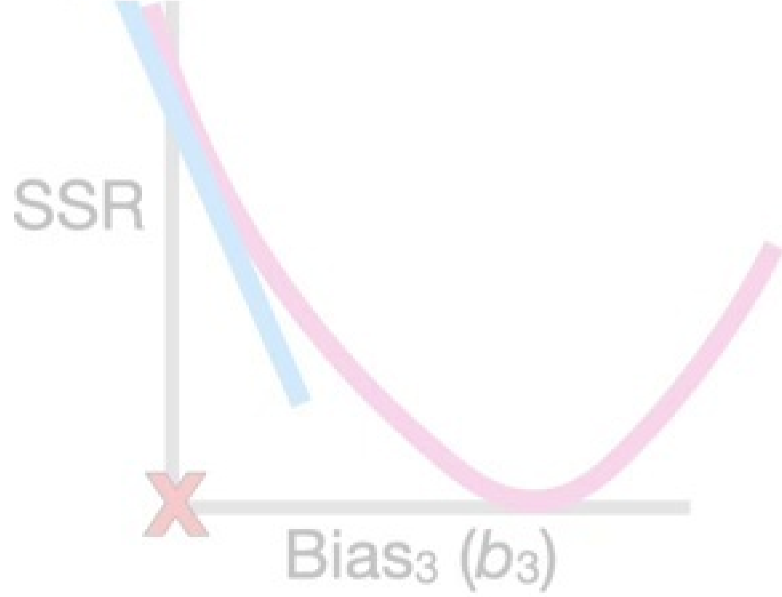


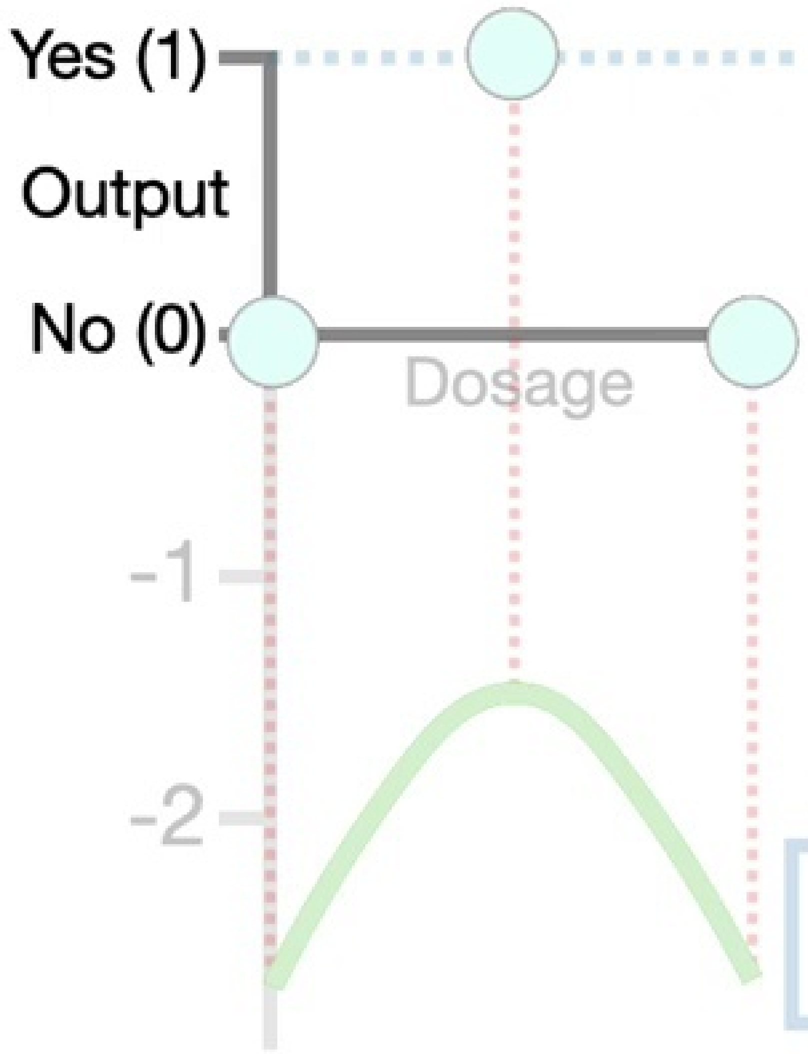


Now we plug the slope into the **Gradient Descent** equation for **Step Size**…

$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -2.6) \times 1 + -2 \times (1 - -1.6) \times 1 + -2 \times (0 - -2.61) \times 1 = \boxed{-15.7}$$

$$\textbf{Step Size} = \boxed{-15.7} \times \textbf{Learning Rate}$$

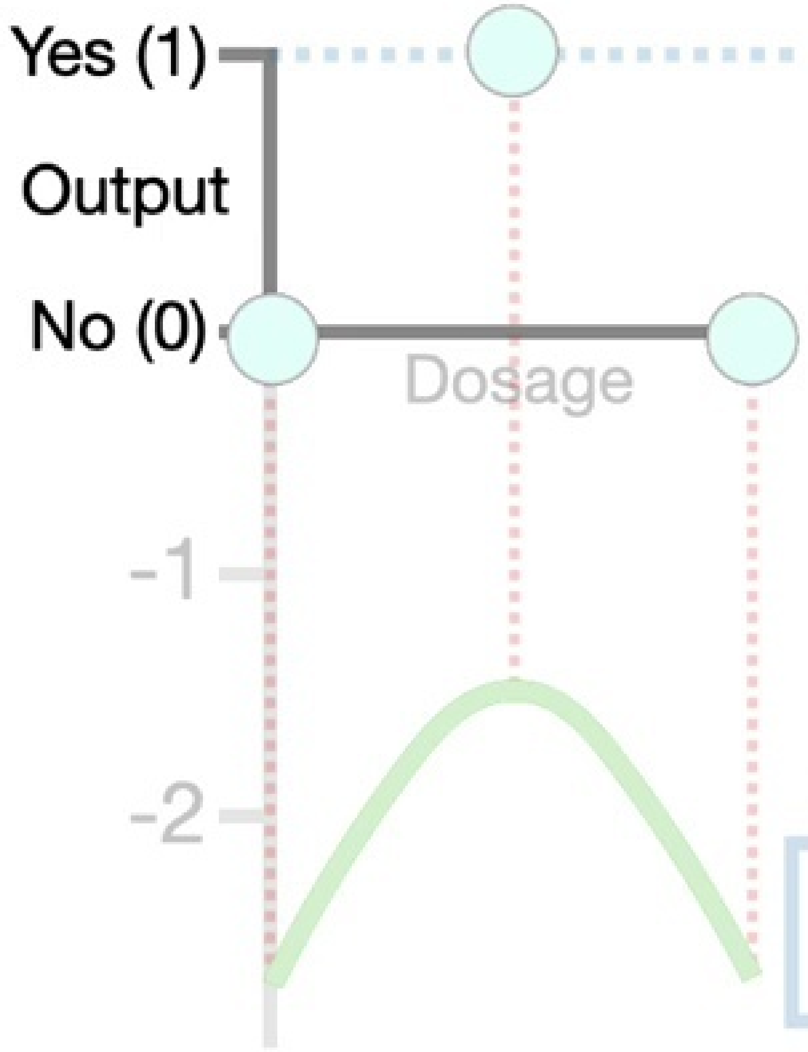


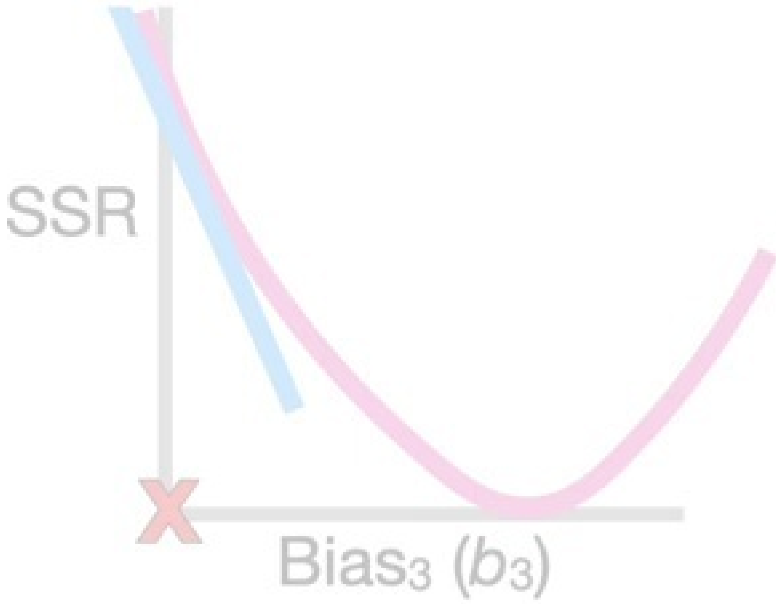


$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -2.6) \times 1 + -2 \times (1 - -1.6) \times 1 + -2 \times (0 - -2.61) \times 1 = -15.7$$

…and in this example, we'll set the **Learning Rate** to **0.1**…

Step Size = -15.7 $\boxed{\times\ 0.1}$

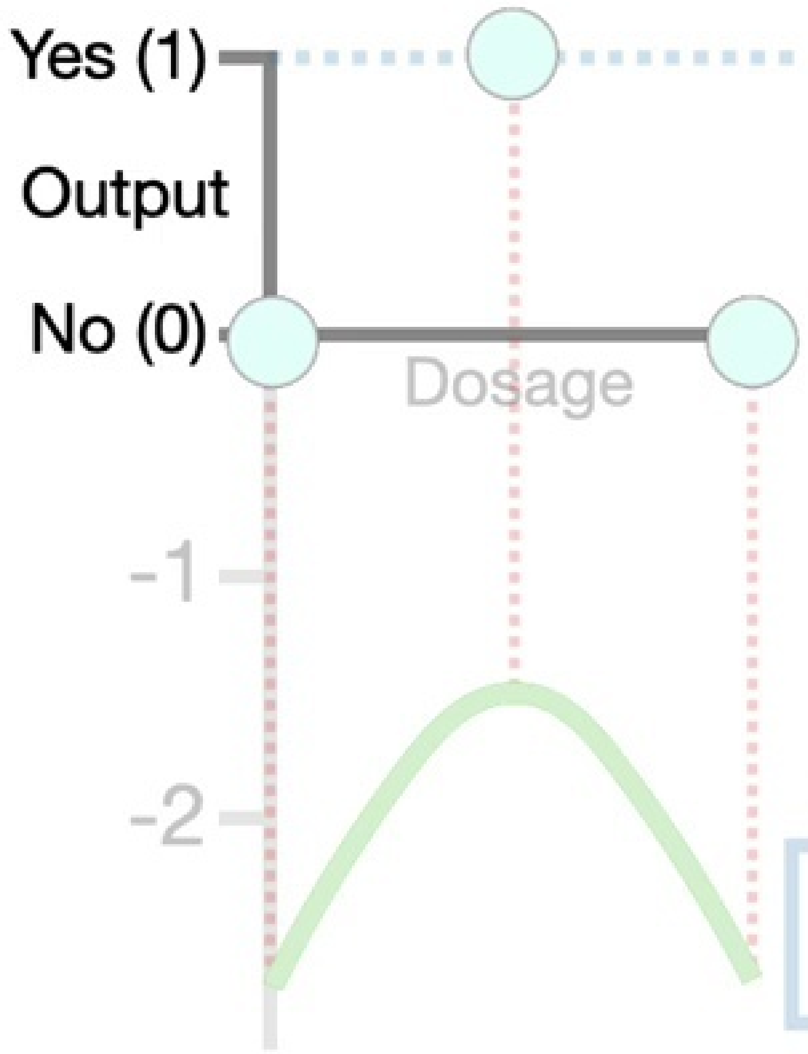


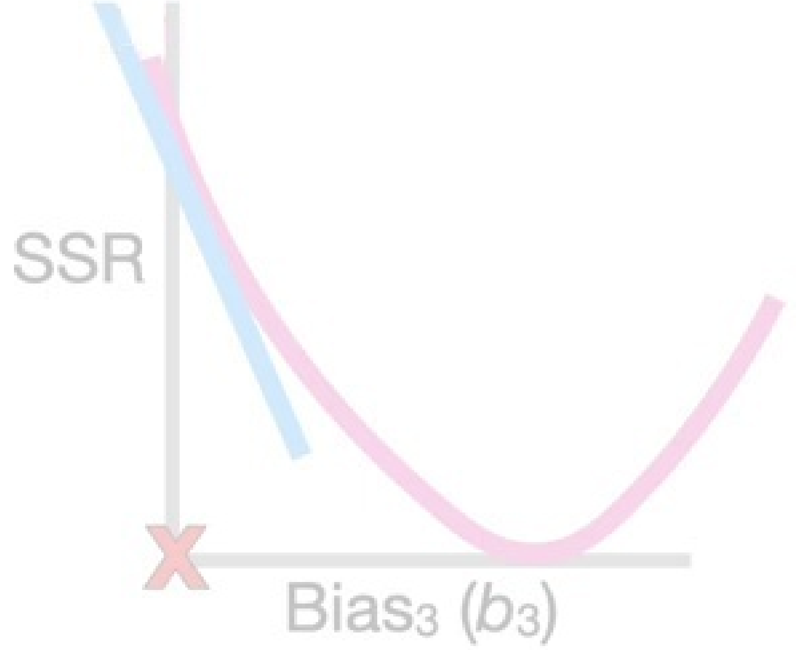


$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -2.6) \times 1 + -2 \times (1 - -1.6) \times 1 + -2 \times (0 - -2.61) \times 1 = -15.7$$

…and that means the **Step Size** is **-1.57**.

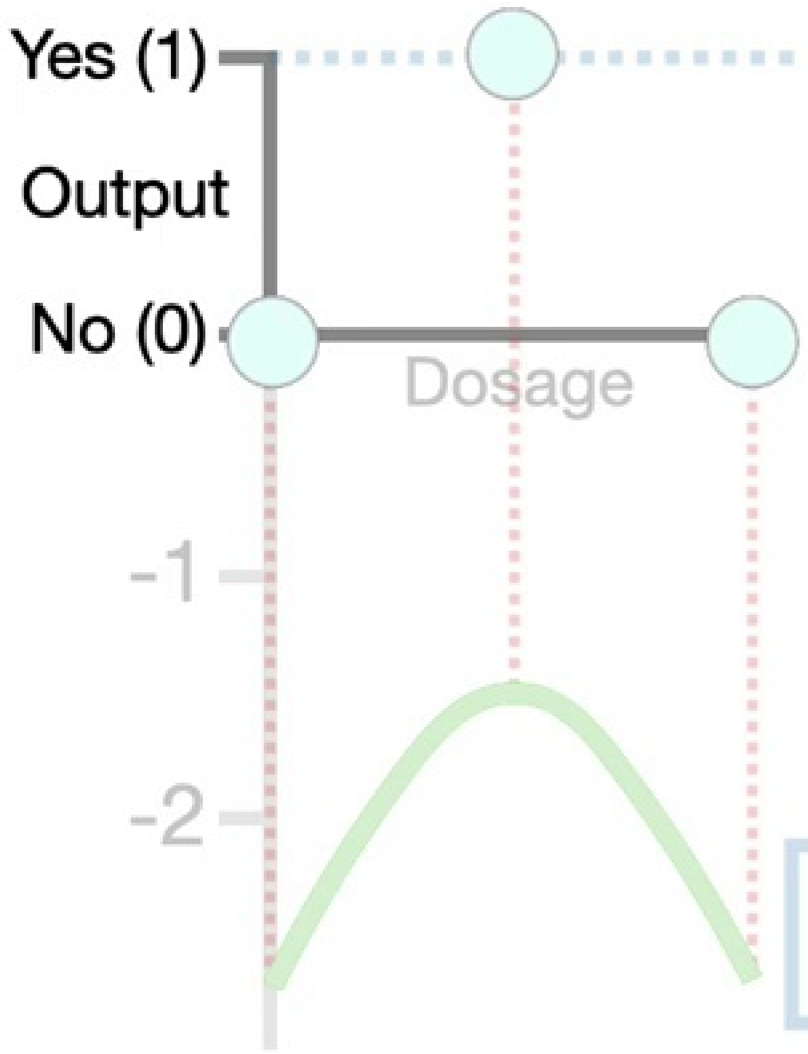
Step Size = $-15.7 \times 0.1 \boxed{= -1.57}$





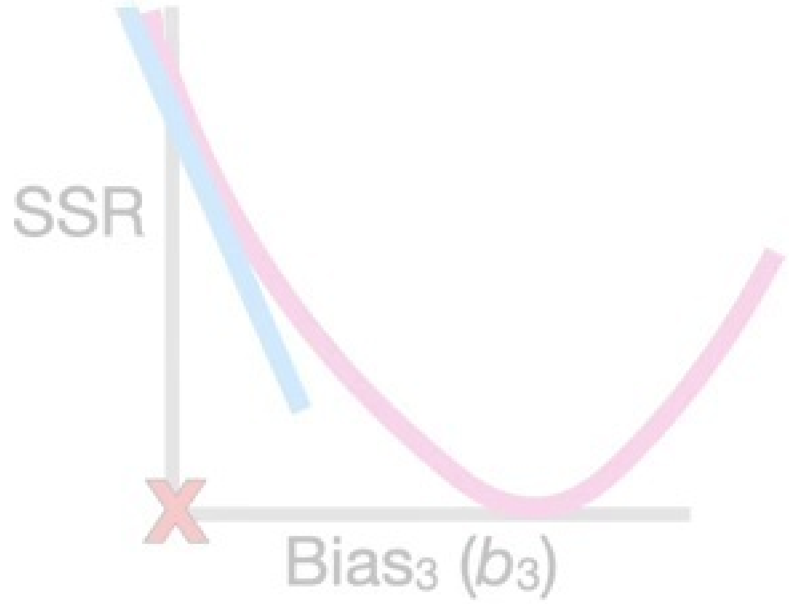
$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -2.6) \times 1 + -2 \times (1 - -1.6) \times 1 + -2 \times (0 - -2.61) \times 1 = -15.7$$

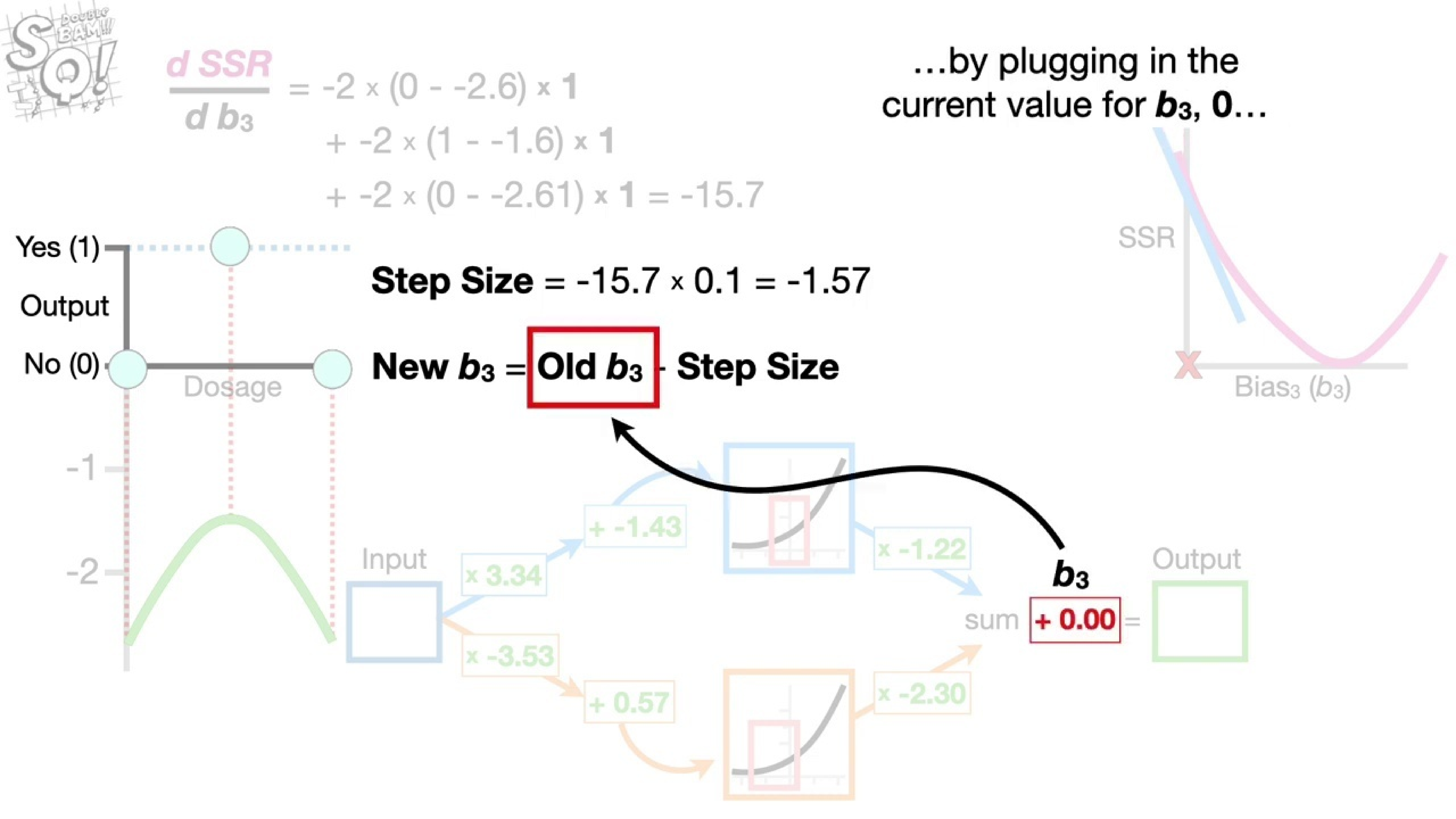
Now we use the **Step Size** to calculate the new value for b_3…



Step Size = $-15.7 \times 0.1 = -1.57$

New b_3 = Old b_3 - Step Size

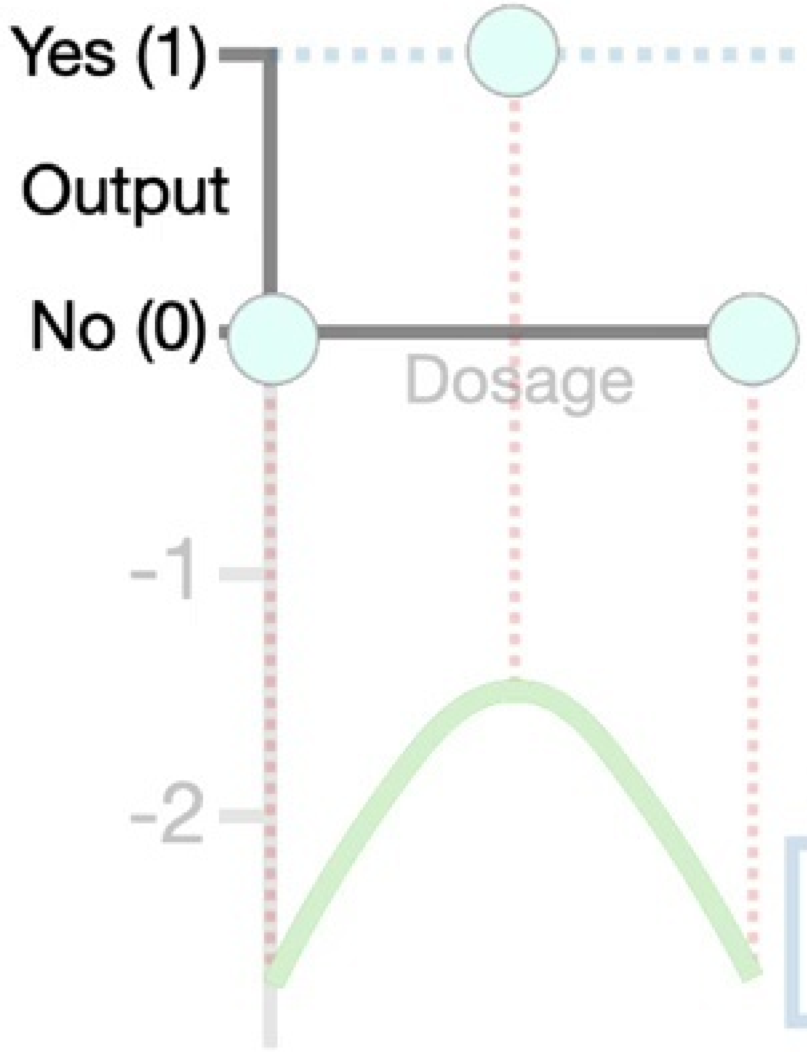




DOUBLE BAM!!!
$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -2.6) \times 1$
$+ -2 \times (1 - -1.6) \times 1$
$+ -2 \times (0 - -2.61) \times 1 = -15.7$
…by plugging in the current value for b_3, 0…
Step Size = -15.7 x 0.1 = -1.57
New b_3 = Old b_3 - Step Size
Yes (1)
Output
No (0)
Dosage
-1
-2
SSR
Bias$_3$ (b_3)
Input
x 3.34
+ -1.43
x -1.22
x -3.53
+ 0.57
x -2.30
b_3
sum
+ 0.00
=
Output

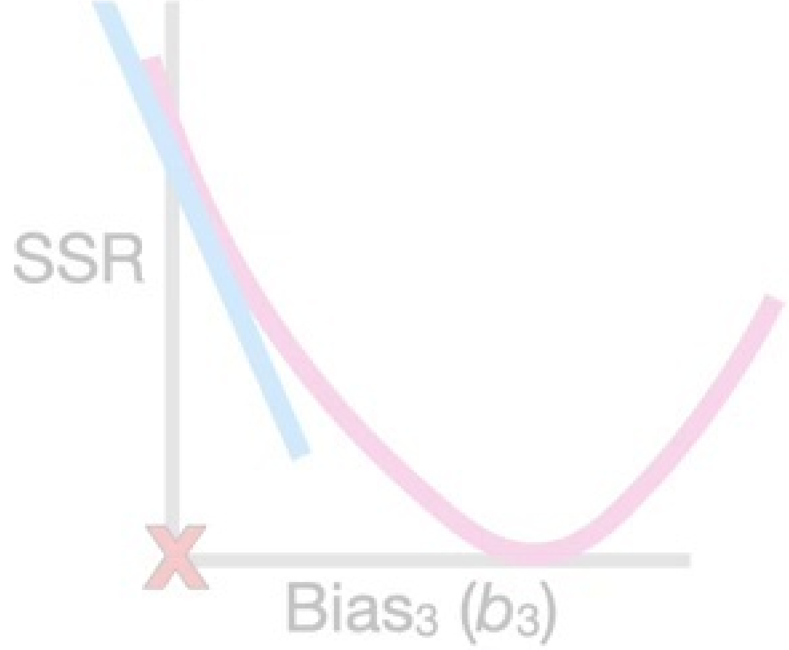
$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -2.6) \times 1 + -2 \times (1 - -1.6) \times 1 + -2 \times (0 - -2.61) \times 1 = -15.7$$

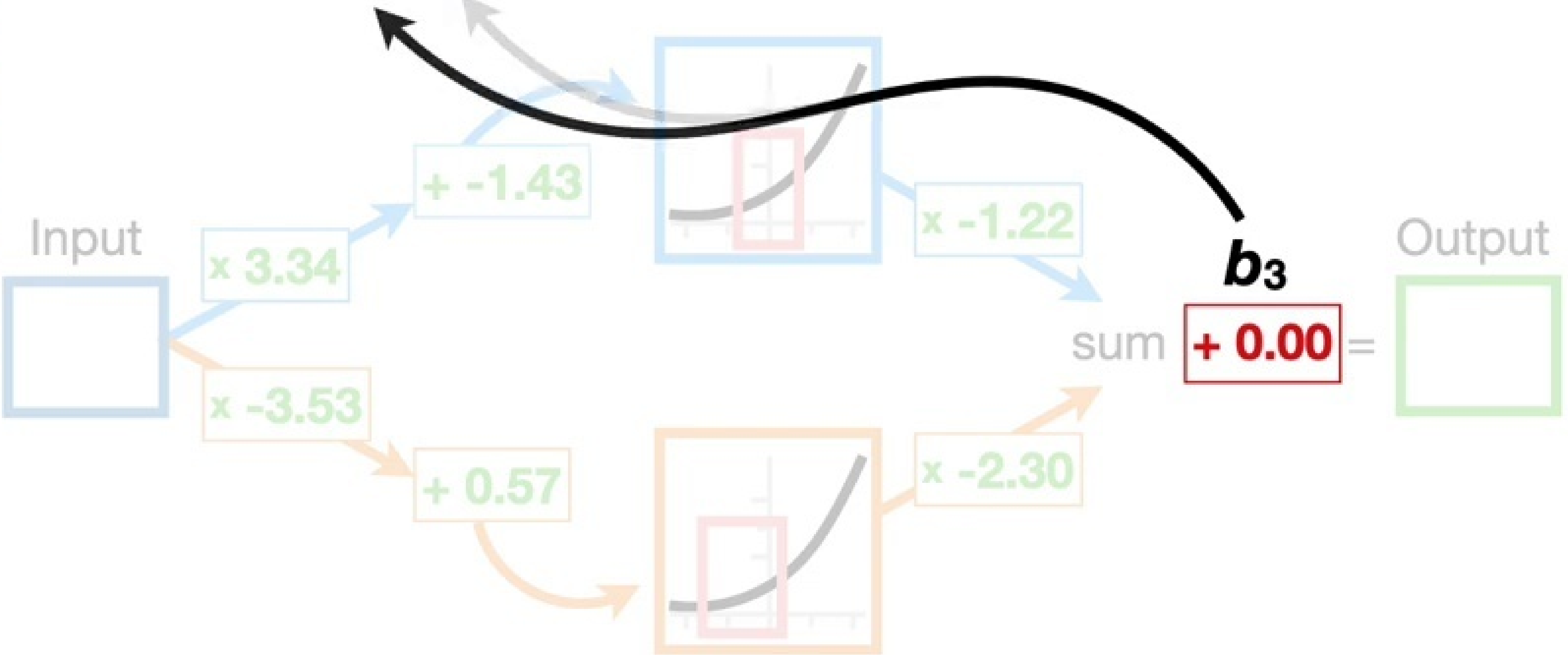
…by plugging in the current value for b_3, **0**…



Step Size = -15.7 × 0.1 = -1.57

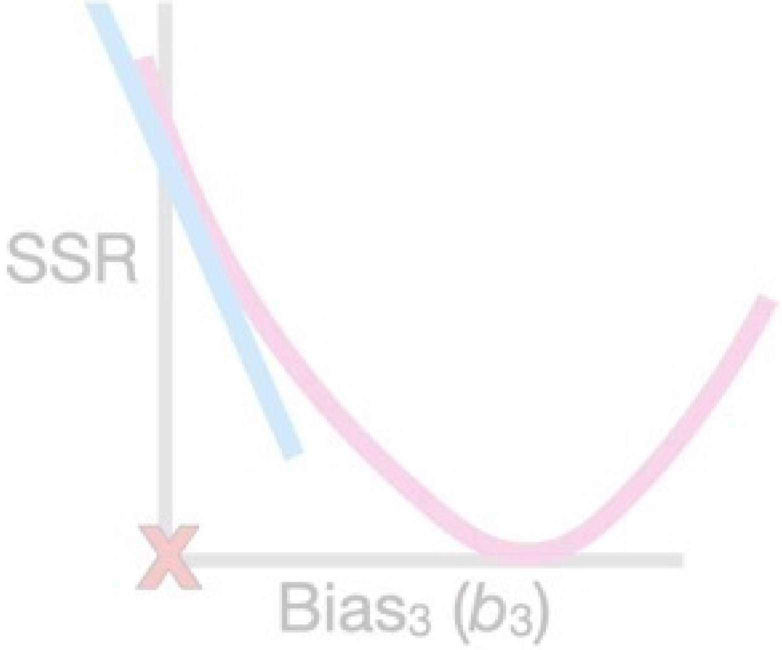
New b_3 = $\boxed{0}$ - **Step Size**

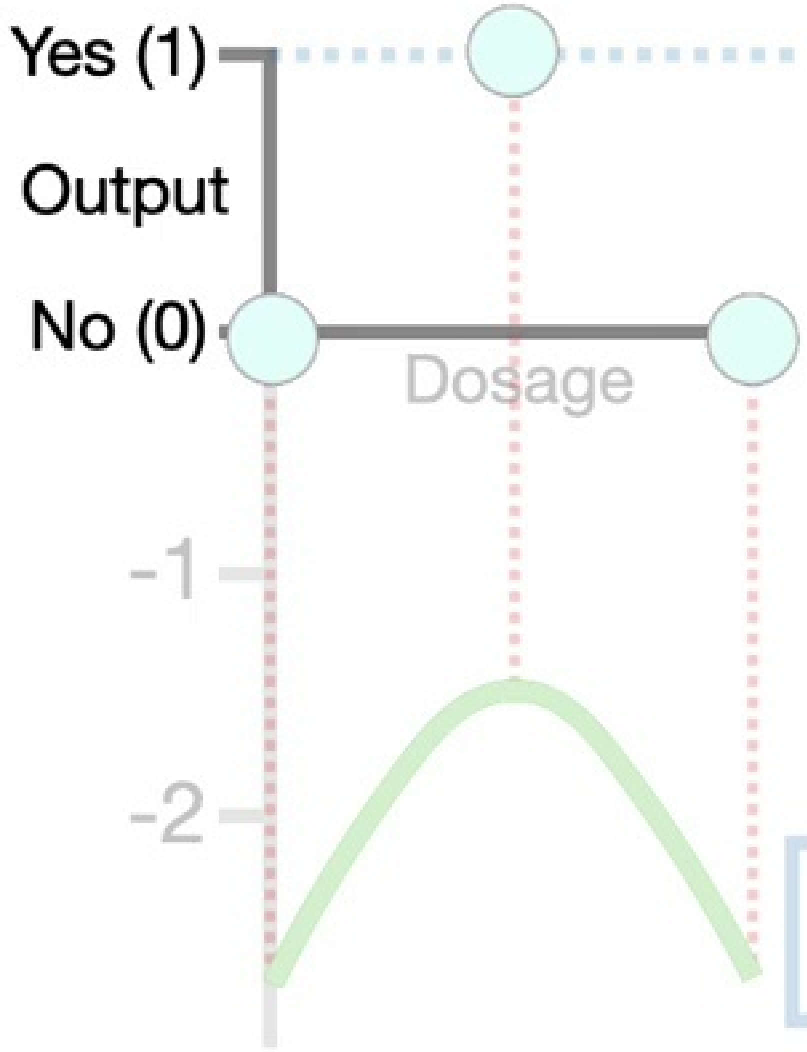




$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -2.6) \times 1 + -2 \times (1 - -1.6) \times 1 + -2 \times (0 - -2.61) \times 1 = -15.7$$

…and the **Step Size**, **-1.57**.



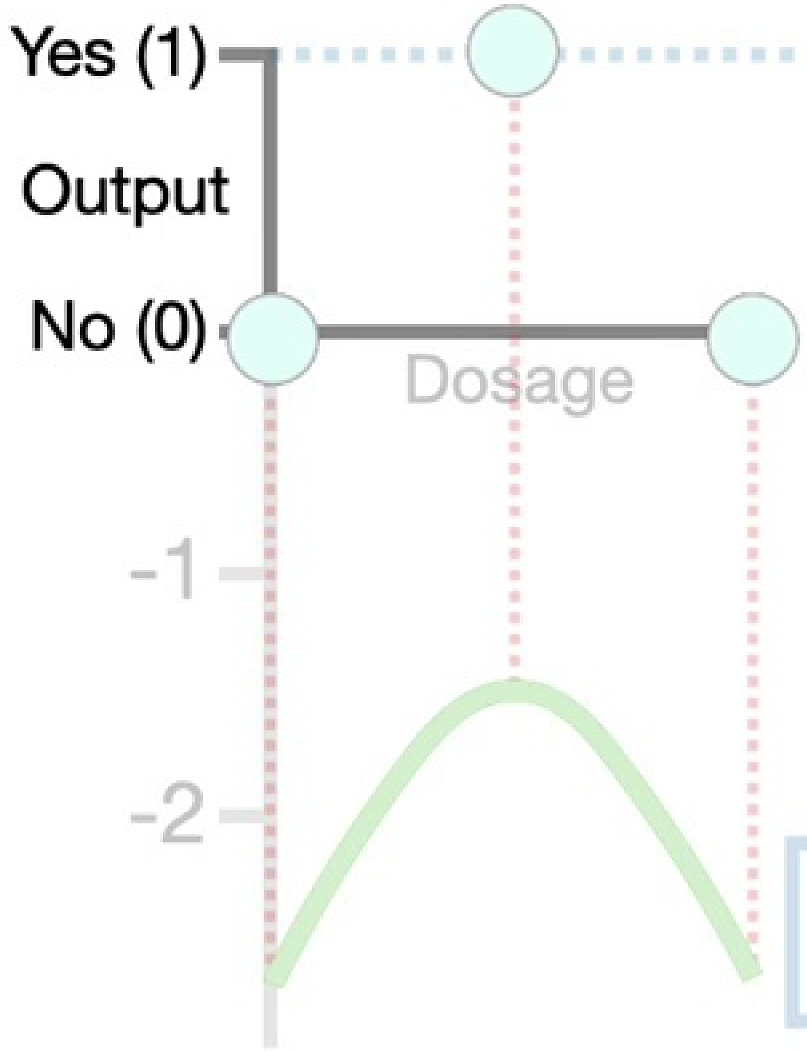


Step Size = $-15.7 \times 0.1 \boxed{= -1.57}$

New b_3 = **0** - $\boxed{\textbf{Step Size}}$

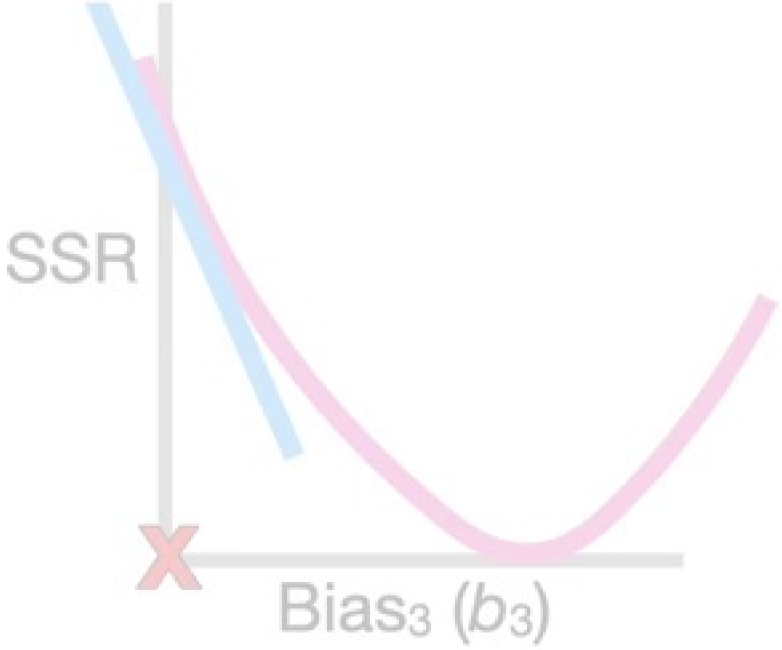
$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -2.6) \times 1 + -2 \times (1 - -1.6) \times 1 + -2 \times (0 - -2.61) \times 1 = -15.7$$

...and the **Step Size**, **-1.57**.



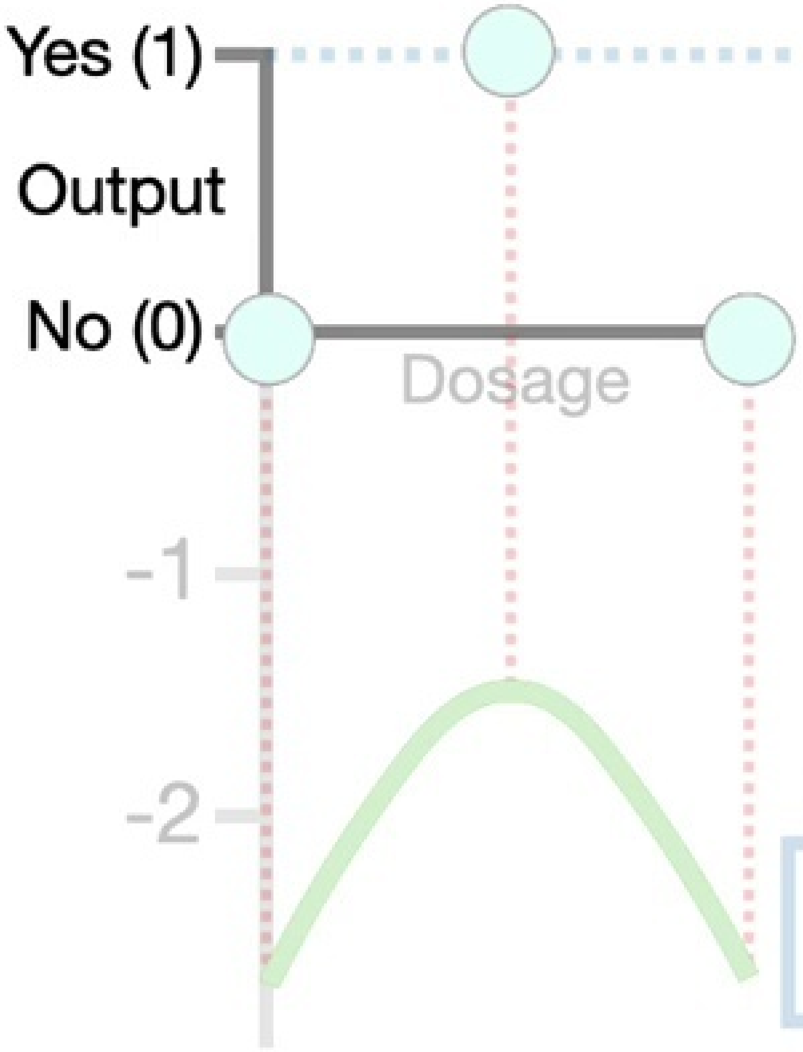
Step Size = $-15.7 \times 0.1 \boxed{= -1.57}$

New $\boldsymbol{b_3}$ = $\color{red}{\mathbf{0}}$ - $\boxed{(-1.57)}$



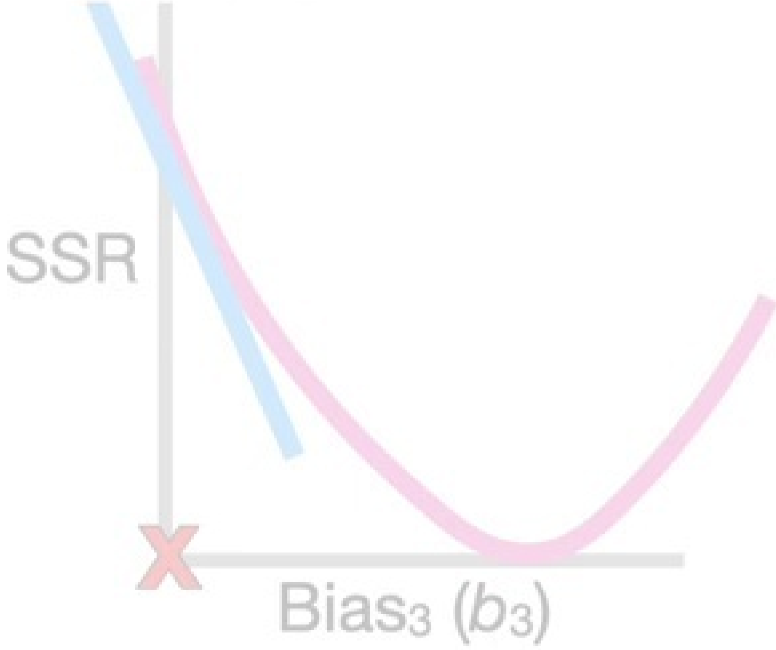
$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -2.6) \times \mathbf{1} + -2 \times (1 - -1.6) \times \mathbf{1} + -2 \times (0 - -2.61) \times \mathbf{1} = -15.7$$

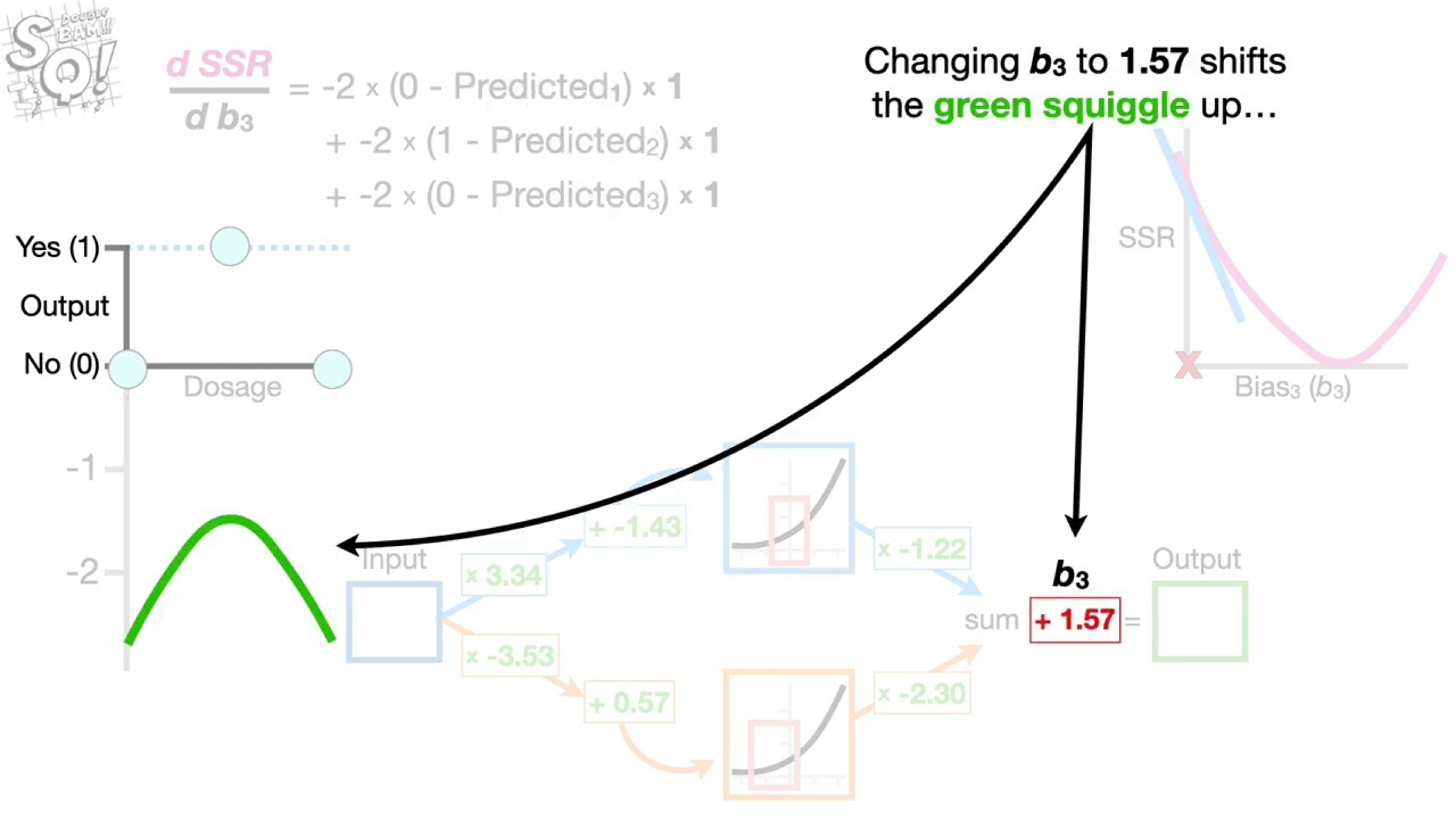
Step Size = -15.7 × 0.1 = -1.57



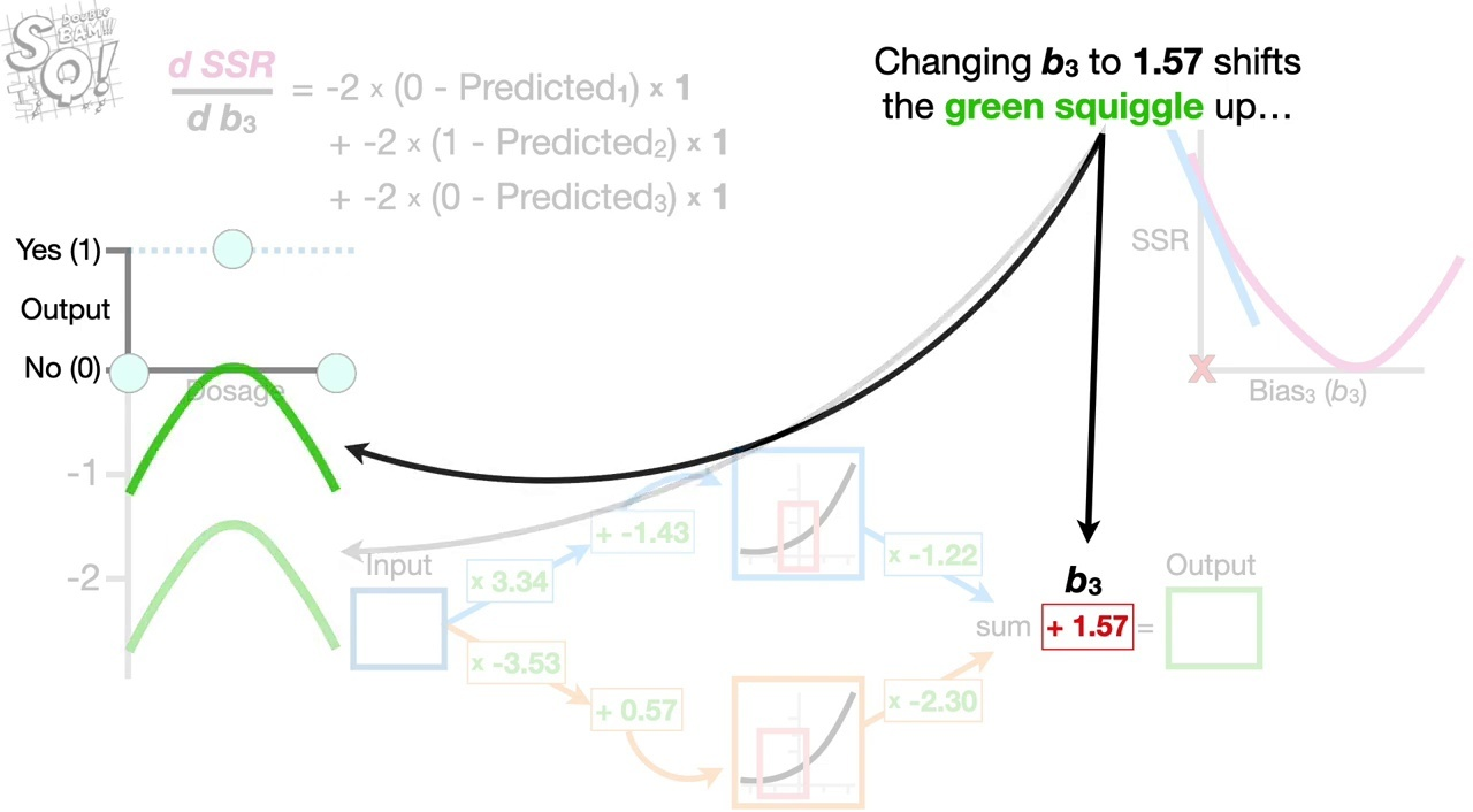
New b_3 = **0** - (-1.57) $\boxed{= 1.57}$

And the new value for b_3 is **1.57**.

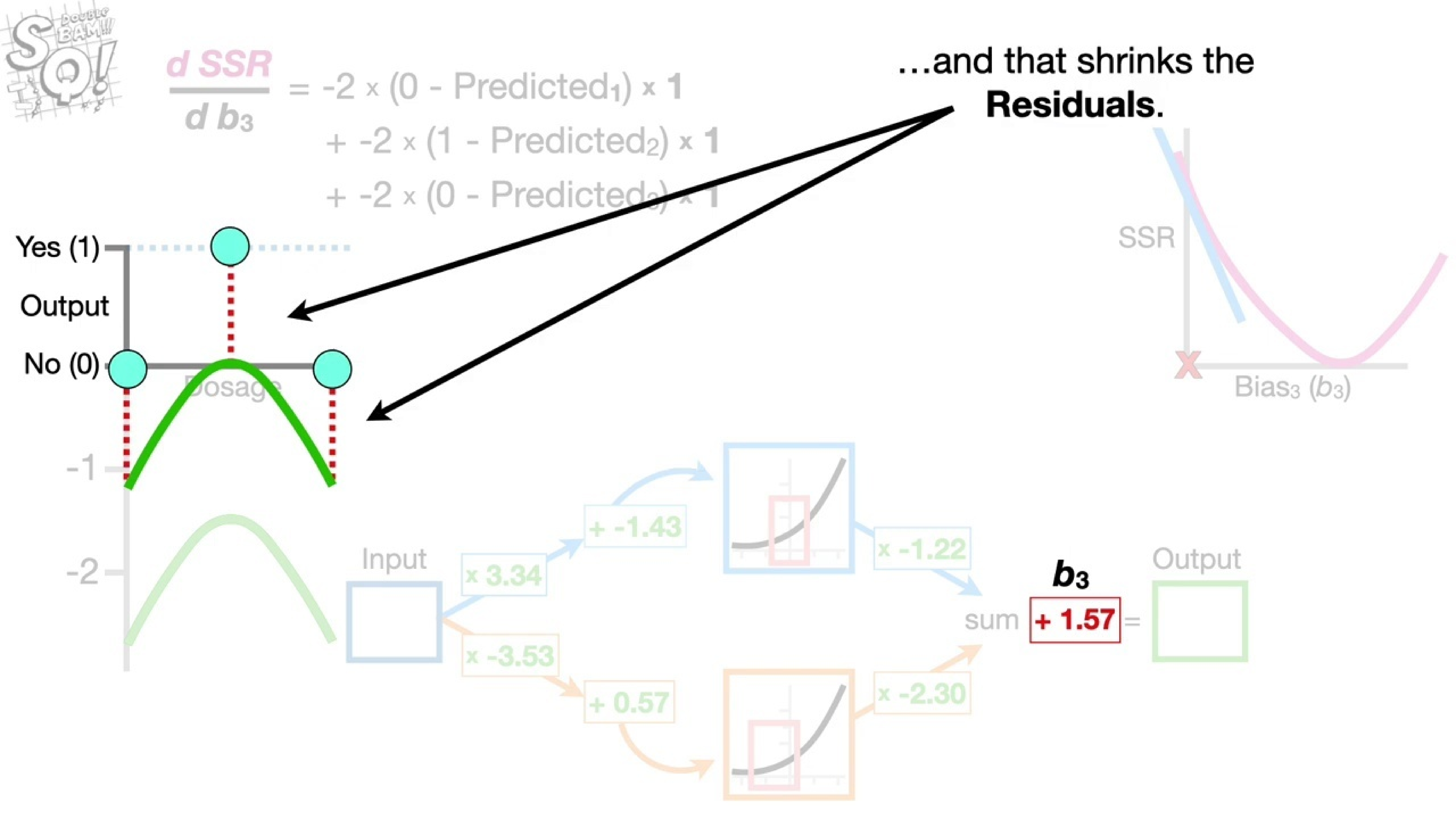




Changing b_3 to 1.57 shifts the green squiggle up…
$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - \text{Predicted}_1) \times 1 + -2 \times (1 - \text{Predicted}_2) \times 1 + -2 \times (0 - \text{Predicted}_3) \times 1$
Yes (1)
Output
No (0)
Dosage
-1
-2
Input
x 3.34
+ -1.43
x -1.22
x -3.53
+ 0.57
x -2.30
b_3
sum
+ 1.57
=
Output
SSR
Bias$_3$ (b_3)



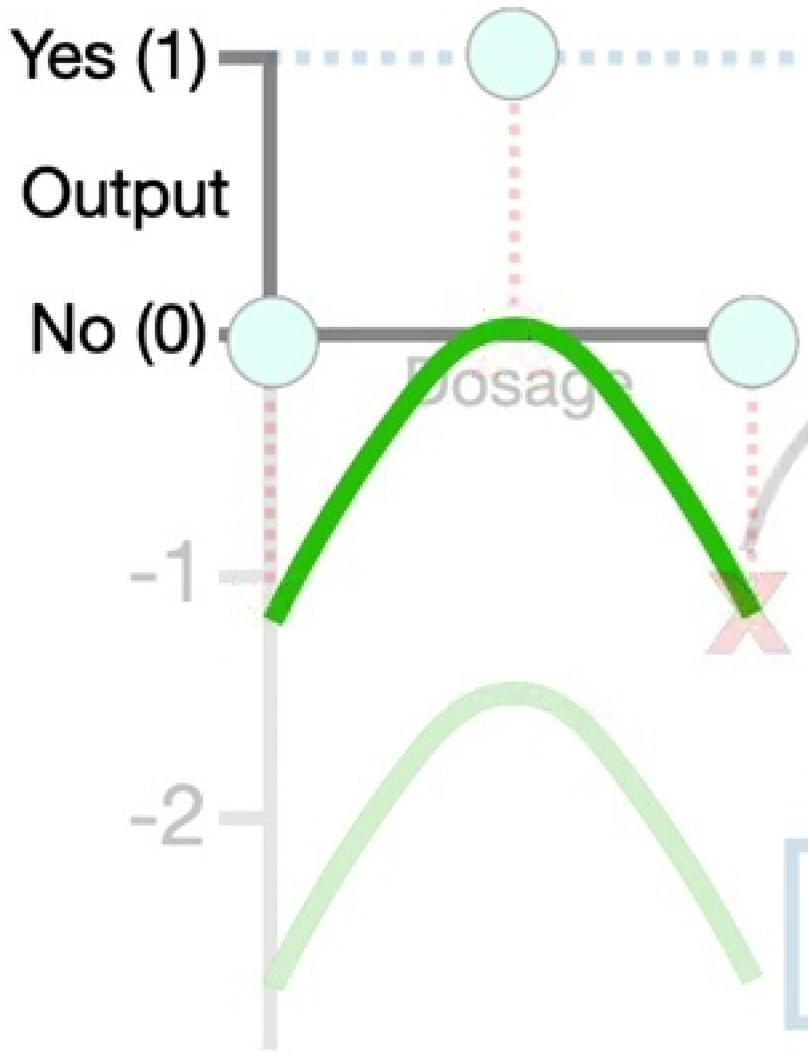
$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - \text{Predicted}_1) \times 1 + -2 \times (1 - \text{Predicted}_2) \times 1 + -2 \times (0 - \text{Predicted}_3) \times 1$$
Changing b_3 to 1.57 shifts the green squiggle up…
Yes (1)
Output
No (0)
Dosage
-1
-2
SSR
Bias$_3$ (b_3)
Input
x 3.34
+ -1.43
x -1.22
x -3.53
+ 0.57
x -2.30
b_3
sum + 1.57 =
Output

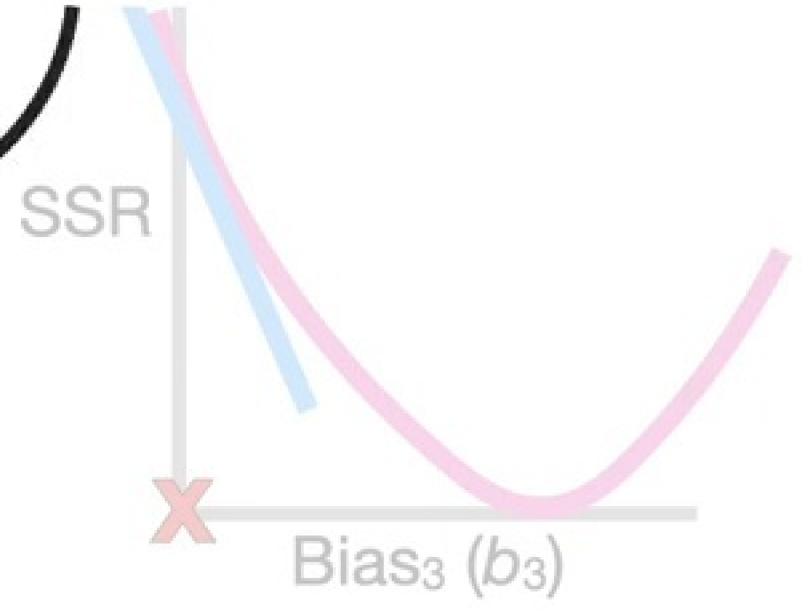


d SSR / d b3 = -2 x (0 - Predicted1) x 1
+ -2 x (1 - Predicted2) x 1
+ -2 x (0 - Predicted3) x 1
...and that shrinks the
Residuals.
Yes (1)
Output
No (0)
Dosage
-1
-2
SSR
Bias3 (b3)
Input
x 3.34
+ -1.43
x -1.22
x -3.53
+ 0.57
x -2.30
b3
sum
+ 1.57
=
Output

Now, plugging in the new **Predicted** values and doing the math gives us **-6.26**…

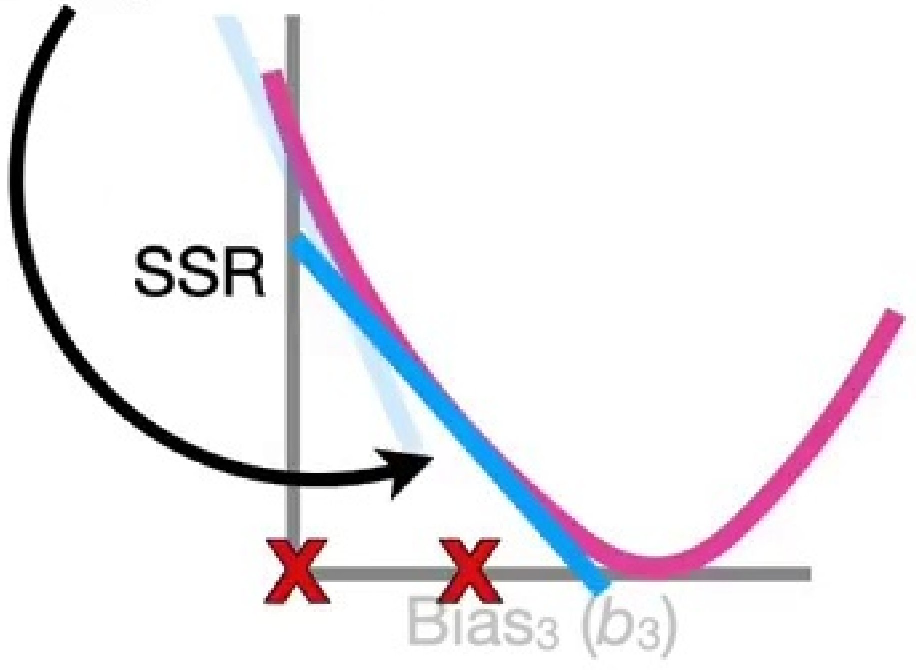
$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 \ \boxed{= -6.26}$$

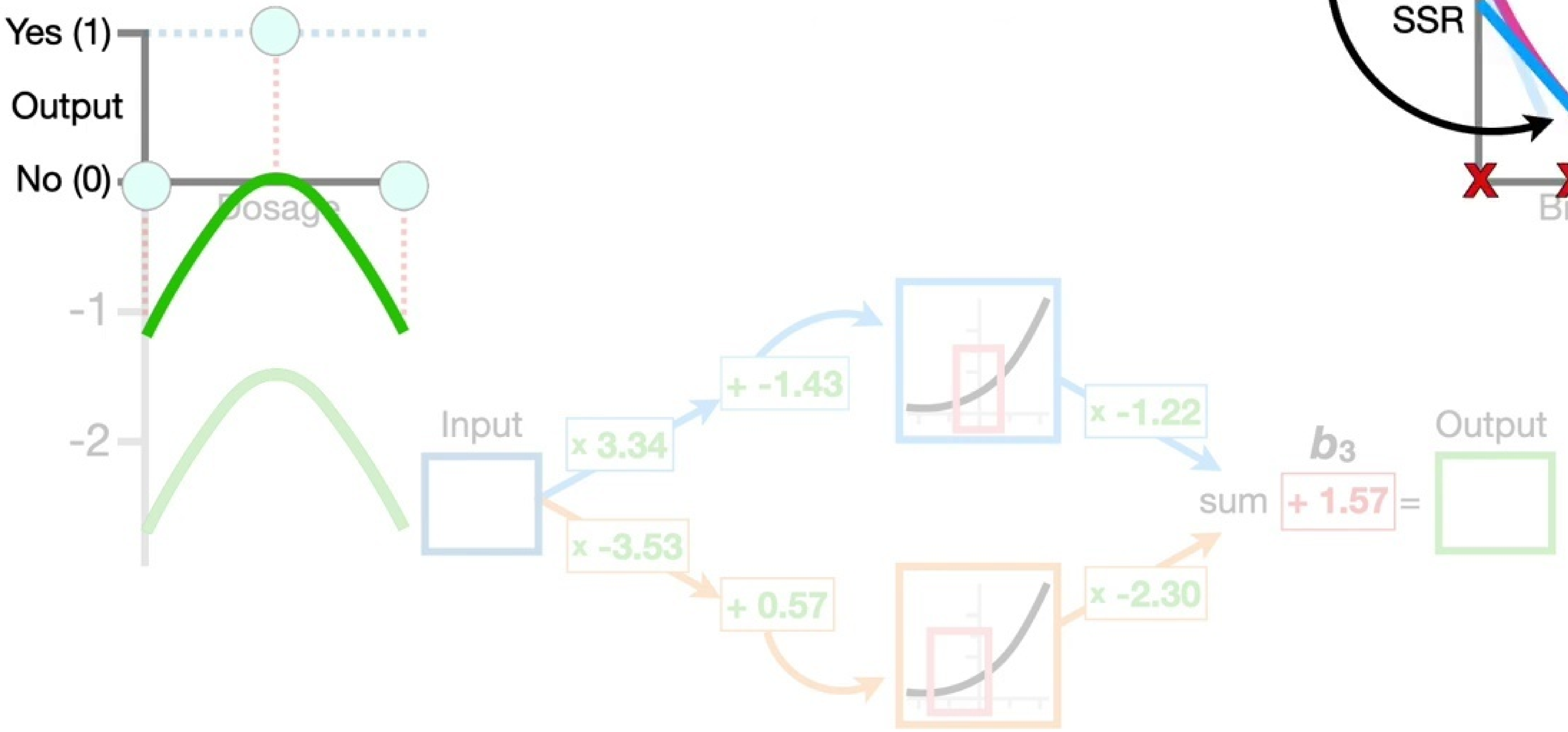




$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -1.03) \times \mathbf{1} + -2 \times (1 - -0.03) \times \mathbf{1} + -2 \times (0 - -1.04) \times \mathbf{1} = -6.26$$

…which corresponds to the slope when $\boldsymbol{b_3} = \mathbf{1.57}$.

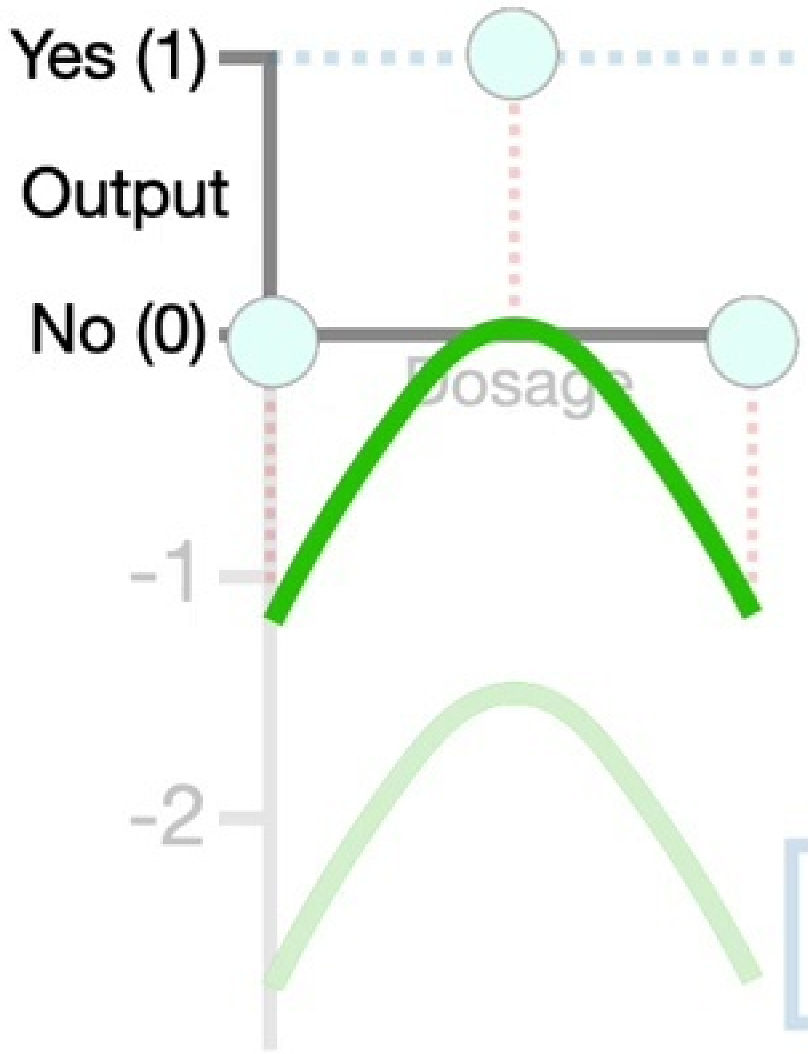


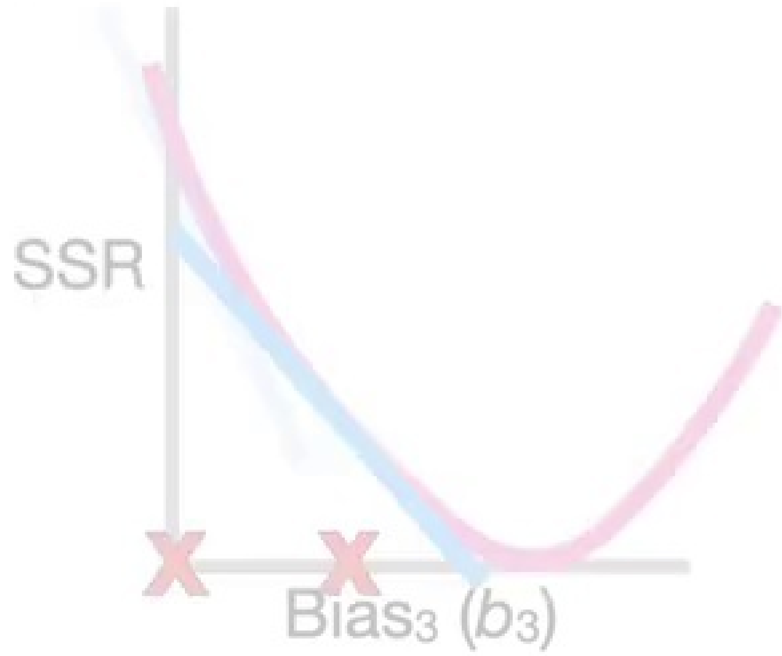


$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = -6.26$$

Then we calculate the **Step Size**…

Step Size = Slope × Learning Rate

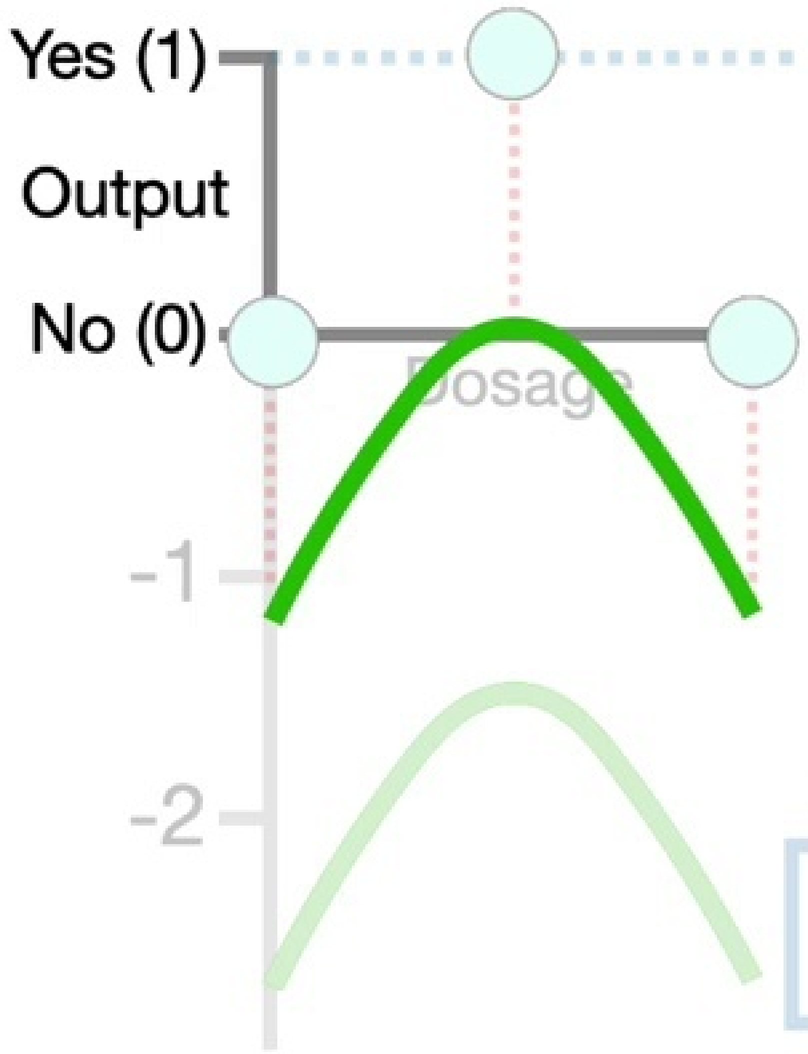


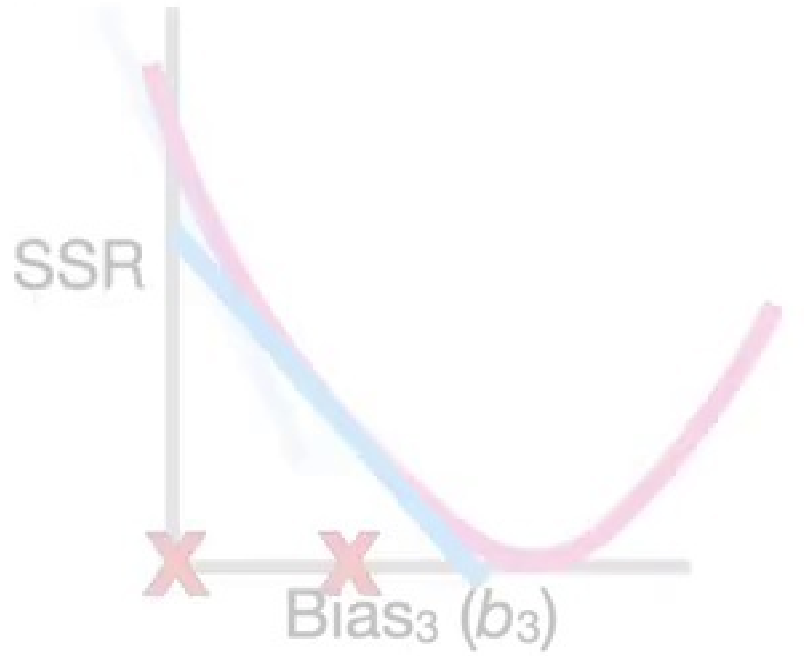


Then we calculate the **Step Size**…

$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = \boxed{-6.26}$$

Step Size = $\boxed{\textbf{Slope}}$ × **Learning Rate**

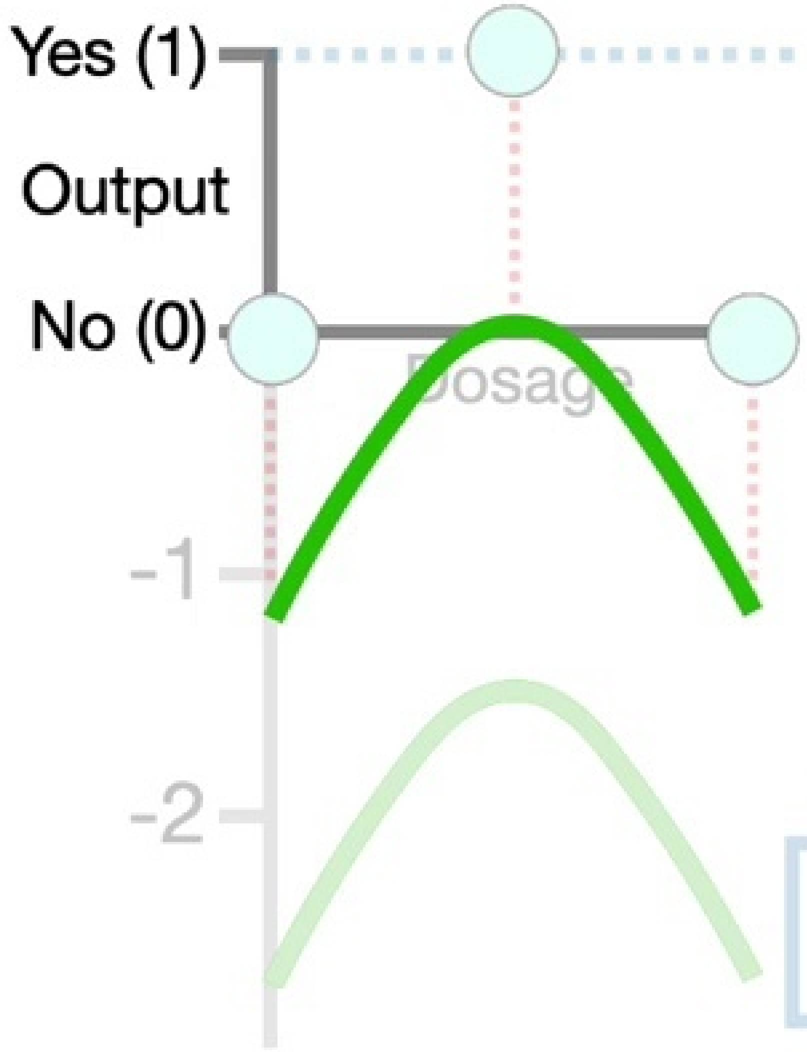


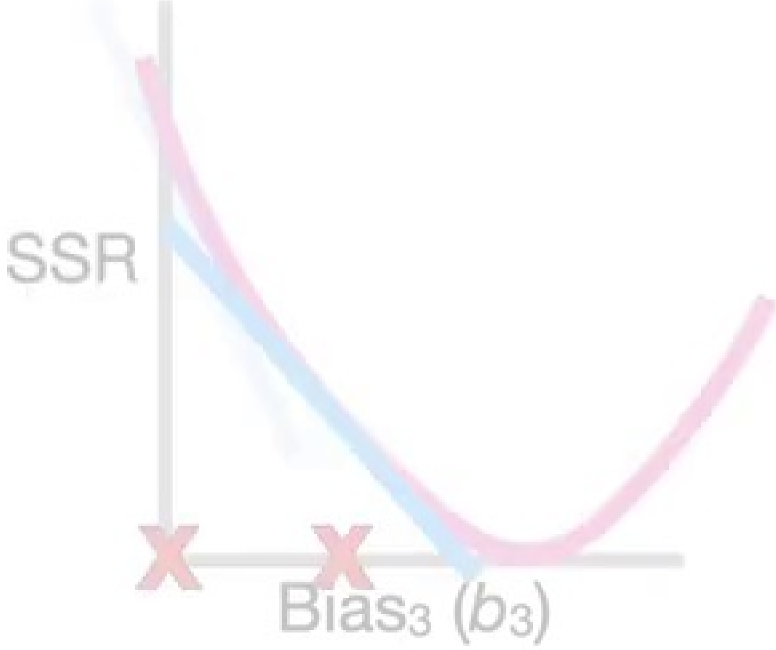


$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = \boxed{-6.26}$$

Then we calculate the **Step Size**…

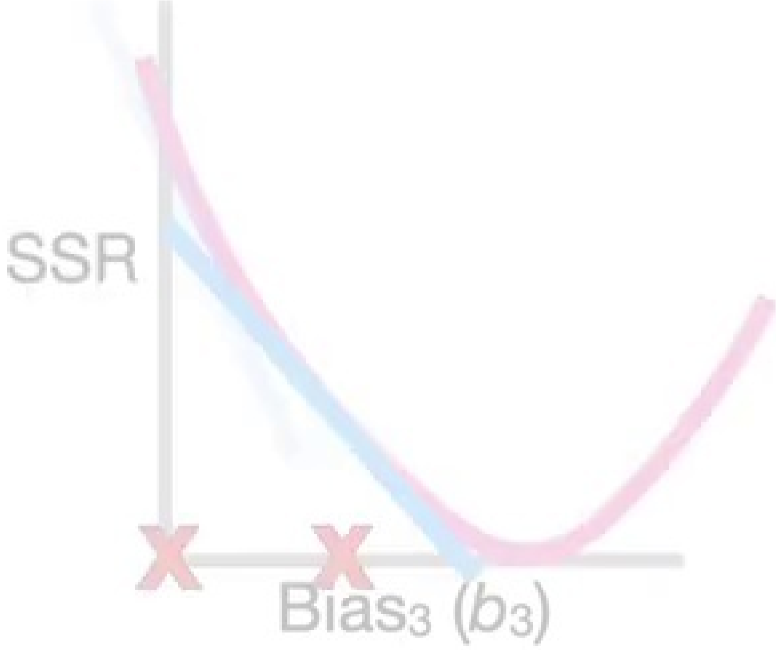
Step Size = $\boxed{-6.26}$ × **Learning Rate**



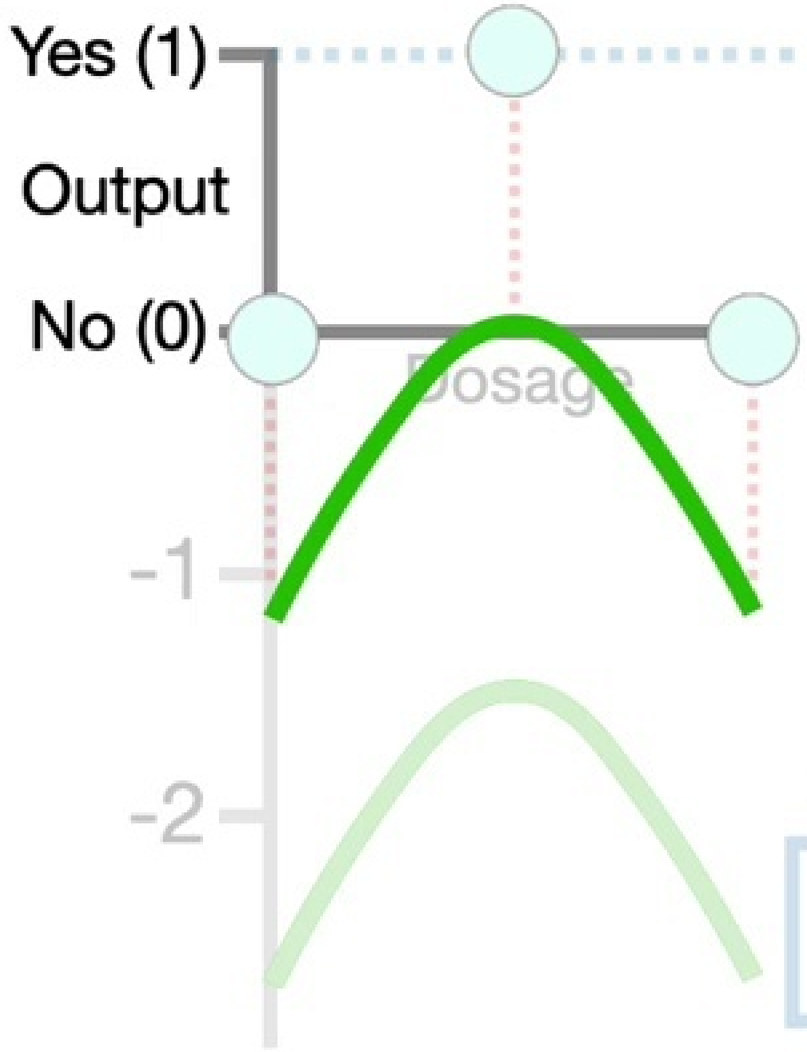


$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = -6.26$$

Then we calculate the **Step Size**…

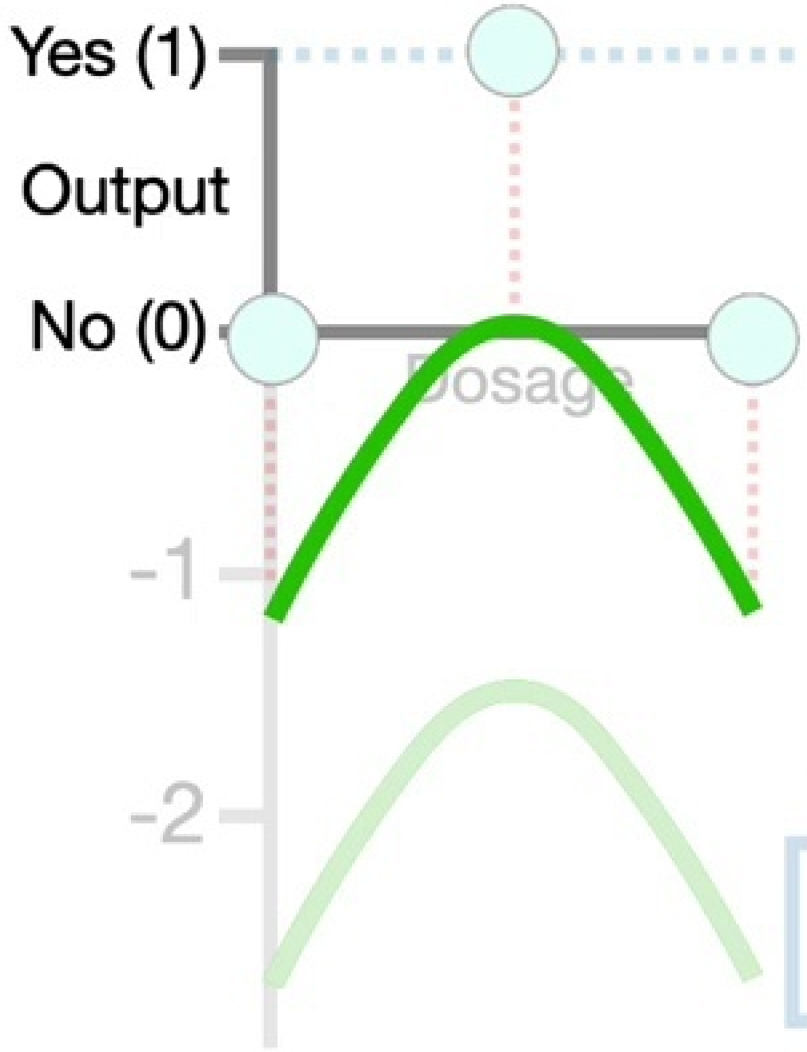


Step Size = -6.26×0.1

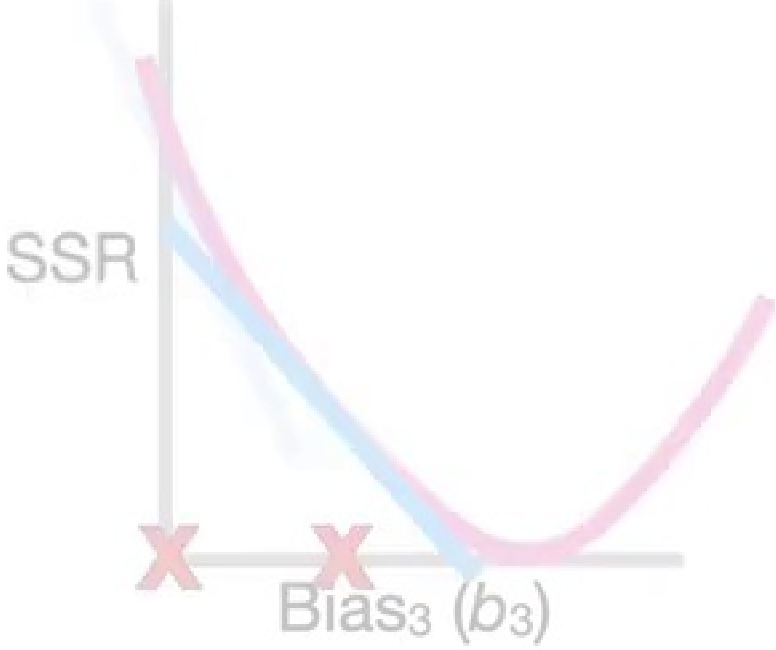


$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = -6.26$$

Then we calculate the **Step Size**…

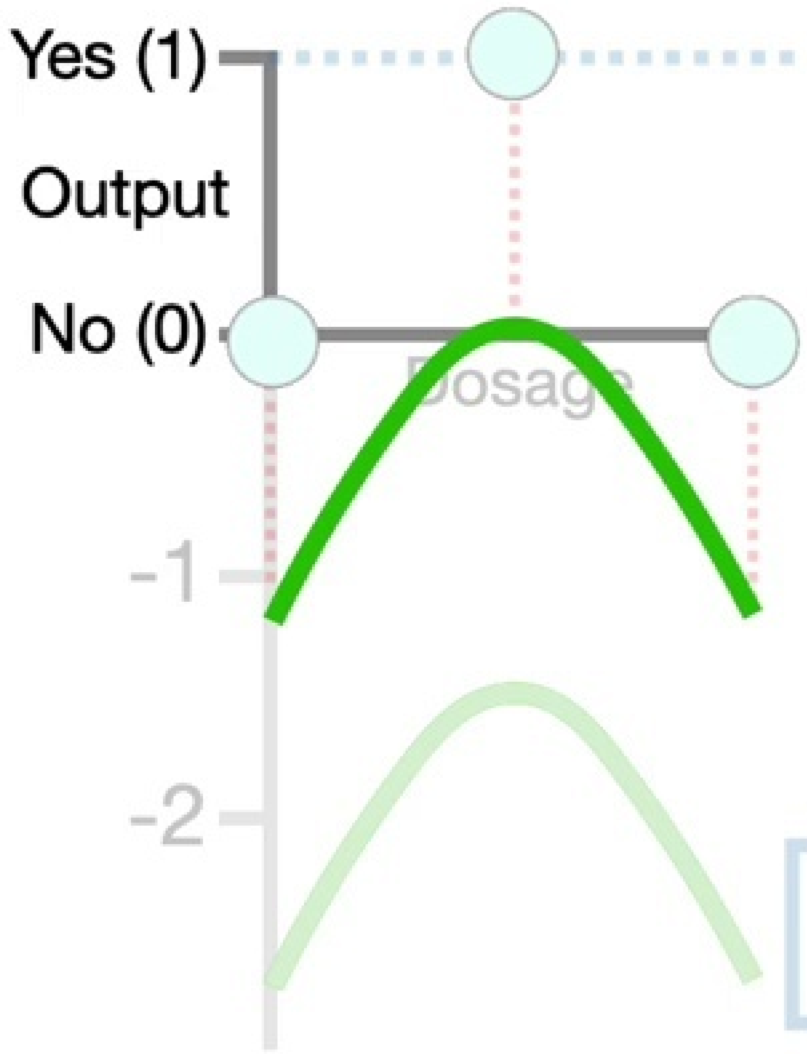


Step Size = $-6.26 \times 0.1 = -0.626$



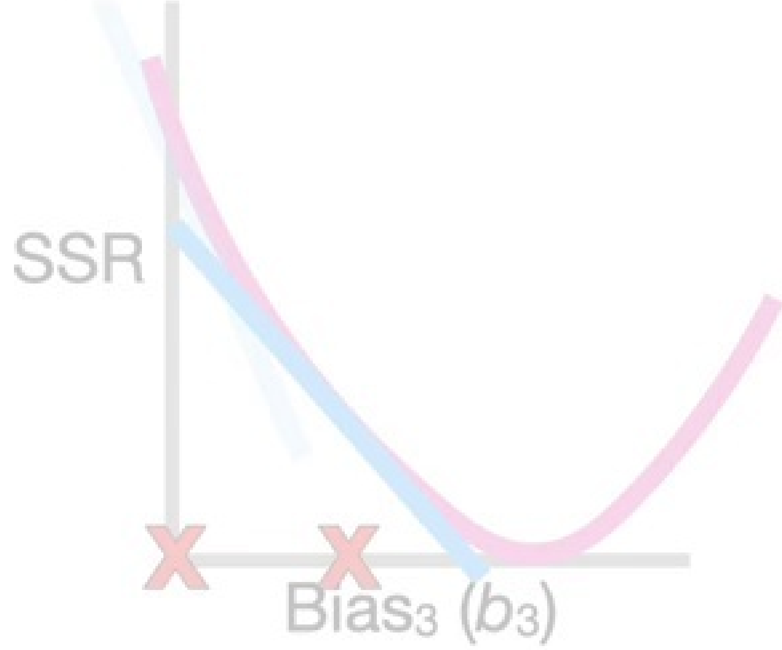
$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = -6.26$$

...and the new value for $\boldsymbol{b_3}$...



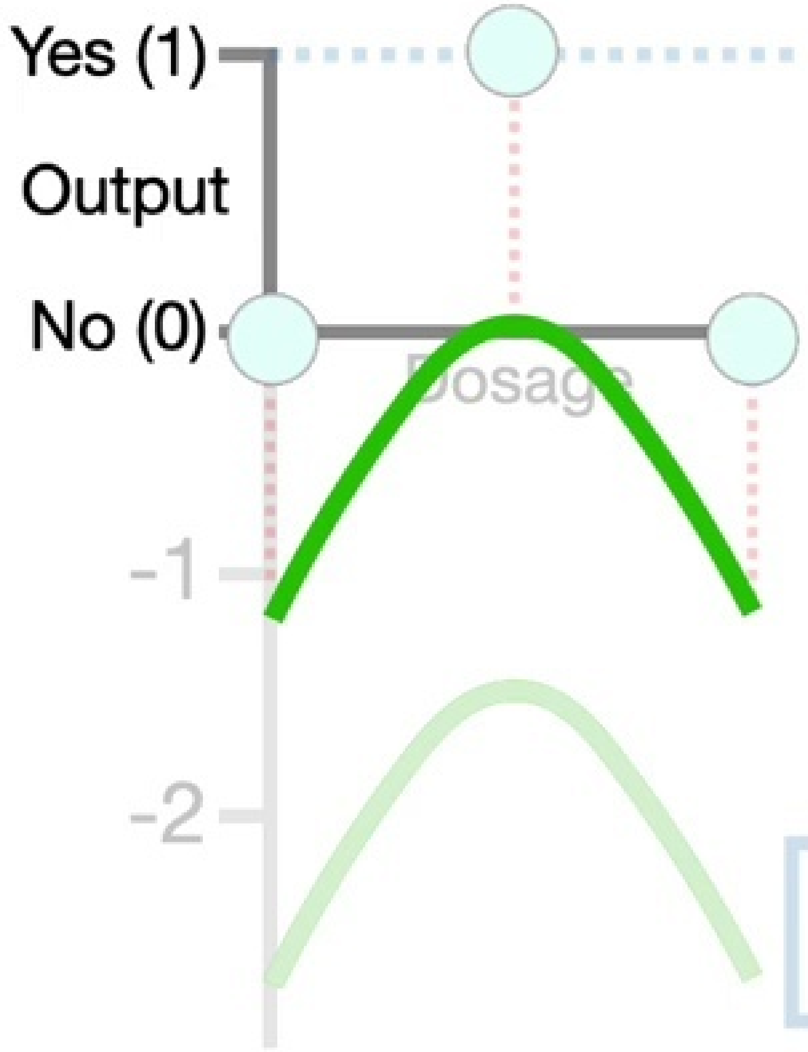
Step Size = $-6.26 \times 0.1 = -0.626$

New b_3 = Old b_3 - Step Size



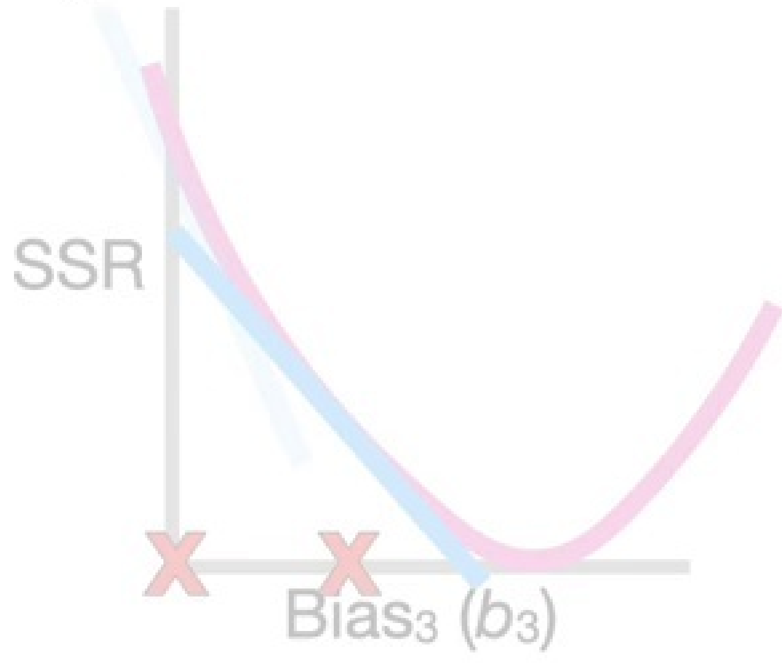
$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = -6.26$$

…and the new value for b_3…



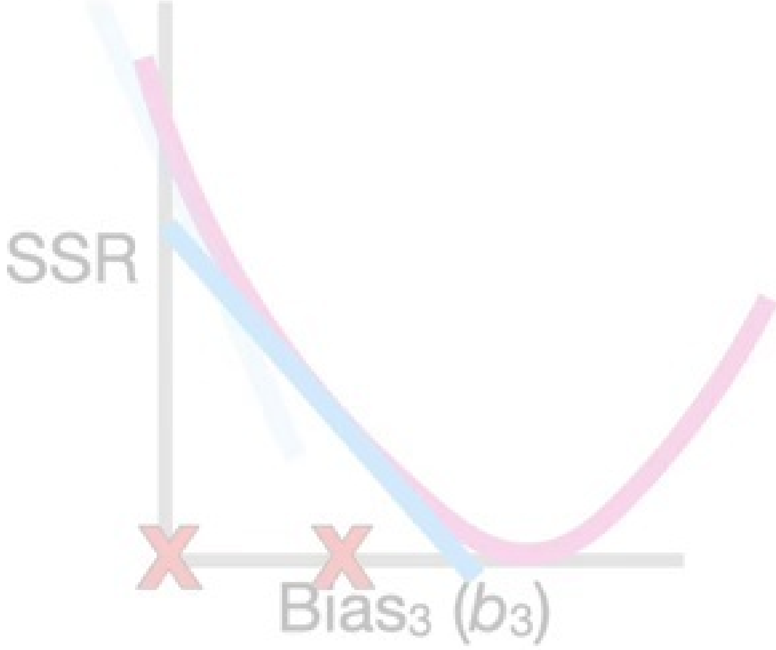
Step Size = -6.26 × 0.1 = -0.626

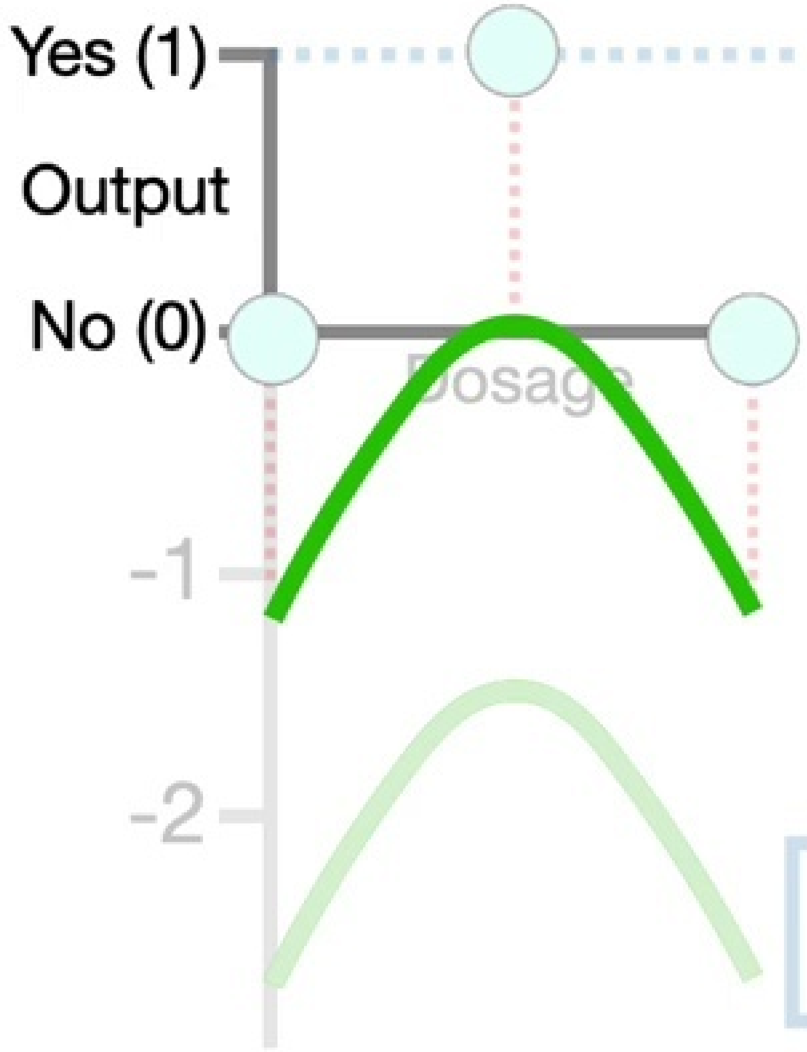
$$\textbf{New } b_3 = \boxed{\textbf{Old } b_3} - \textbf{Step Size}$$



$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = -6.26$$

…and the new value for b_3…



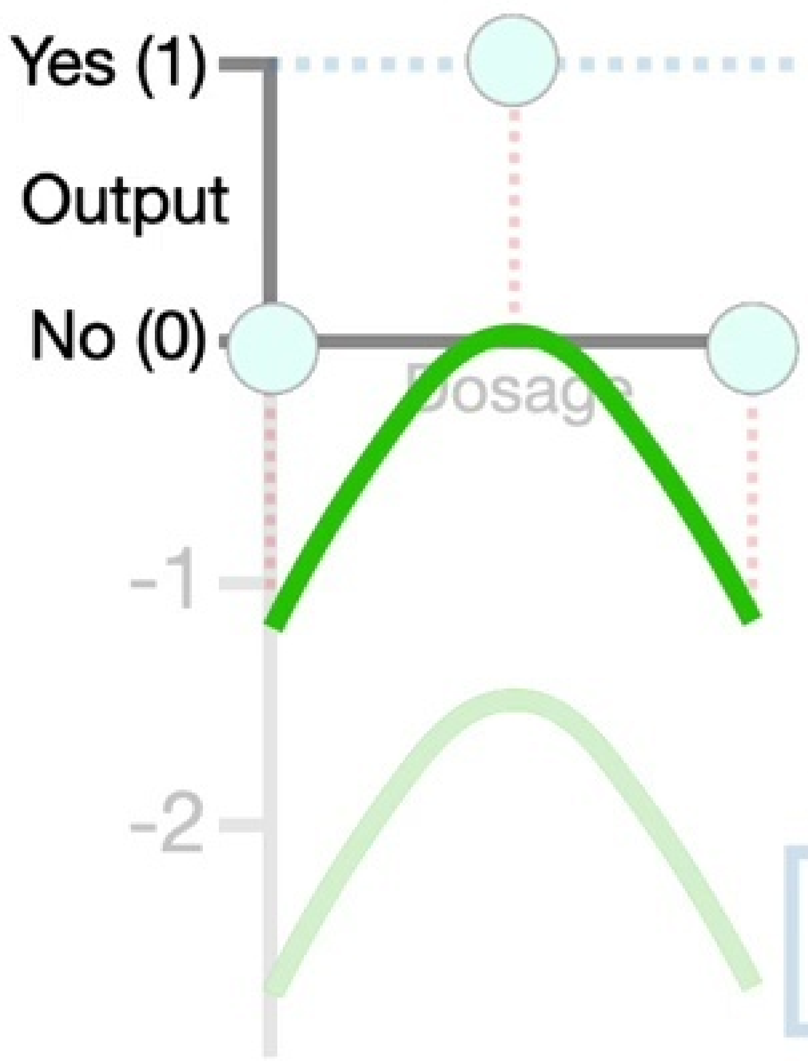


Step Size = -6.26 × 0.1 = -0.626

New b_3 = $\boxed{1.57}$ - Step Size

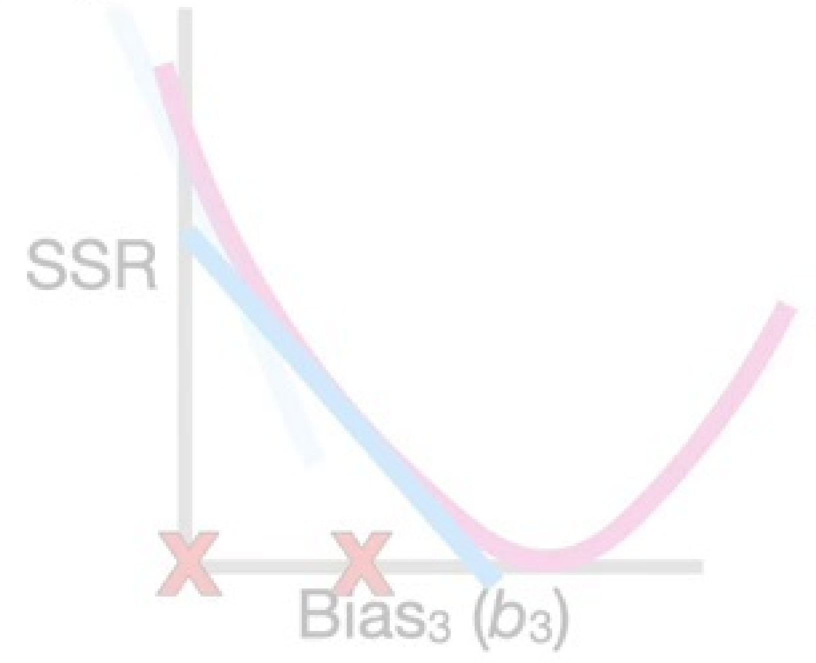
$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = -6.26$$

…and the new value for $\boldsymbol{b_3}$…



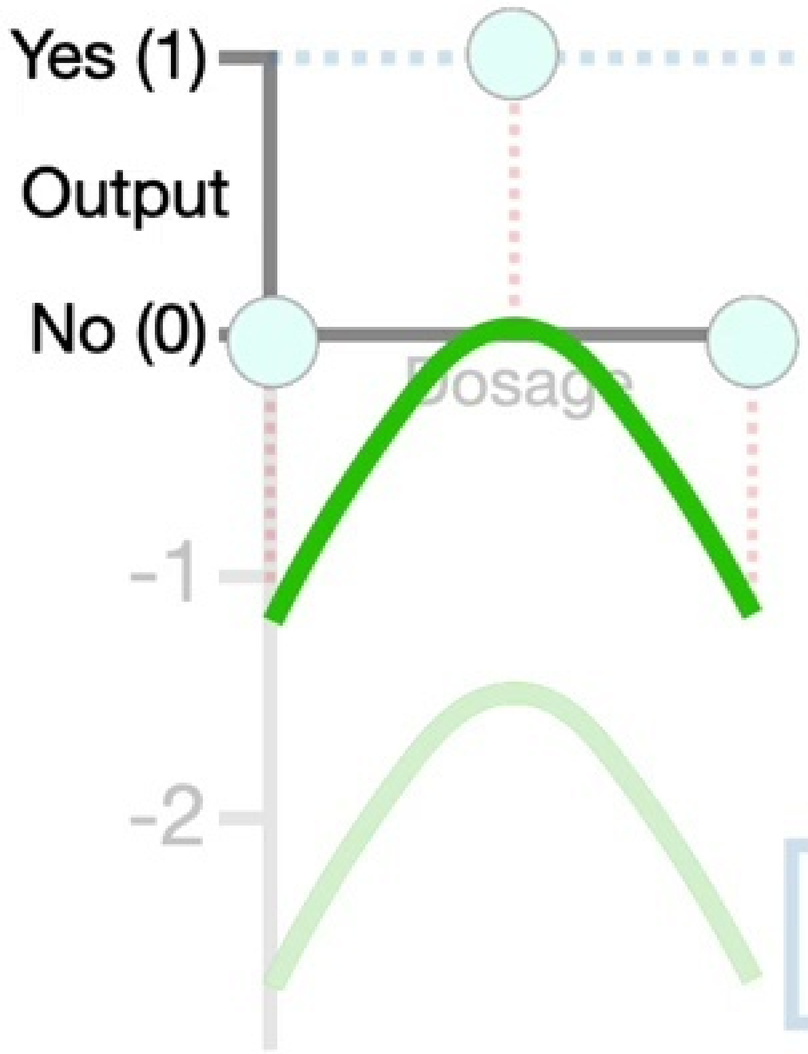
Step Size = $-6.26 \times 0.1 \boxed{= -0.626}$

New b_3 = 1.57 $\boxed{\text{- Step Size}}$



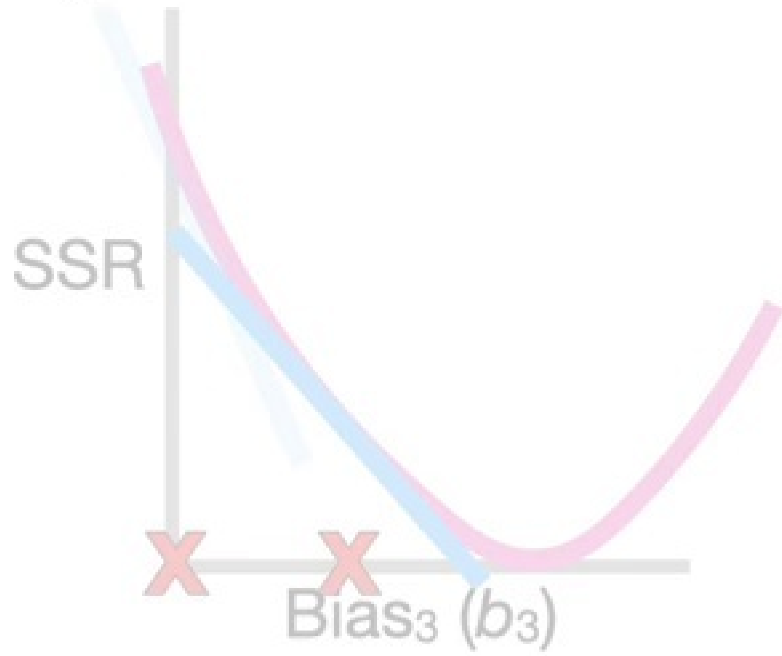
$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = -6.26$$

…and the new value for $\boldsymbol{b_3}$…



Step Size = -6.26 × 0.1 $\boxed{= -0.626}$

New $\boldsymbol{b_3}$ **= 1.57** $\boxed{- (-0.626)}$

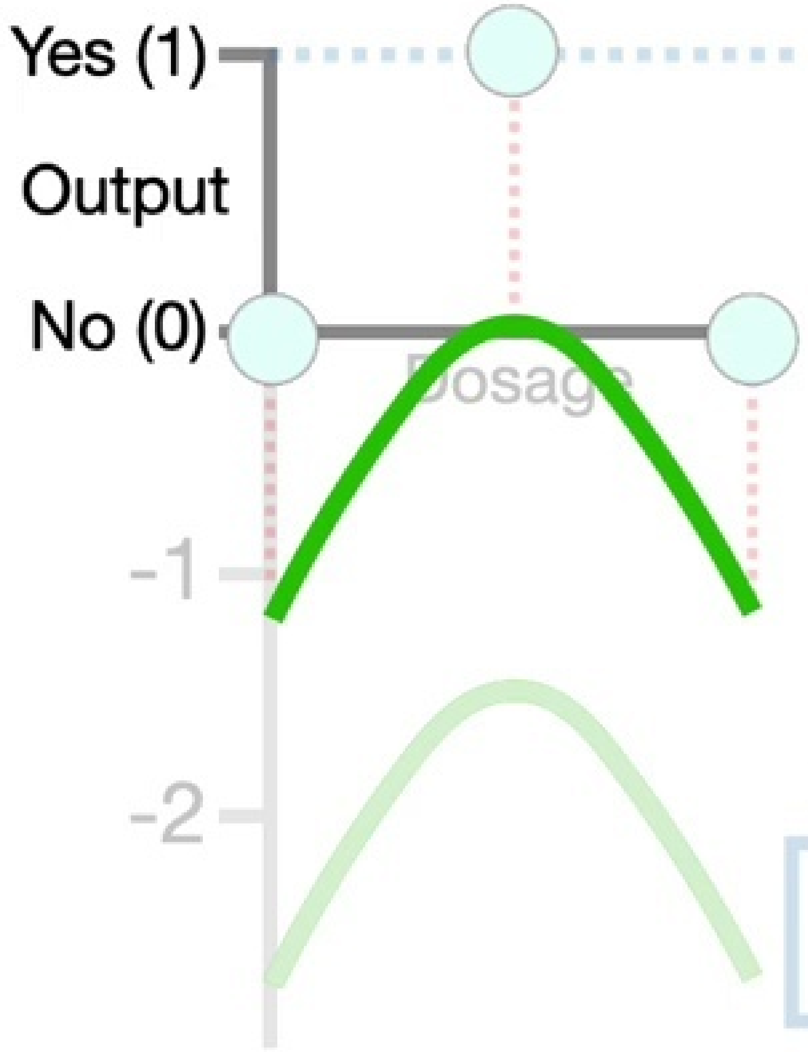


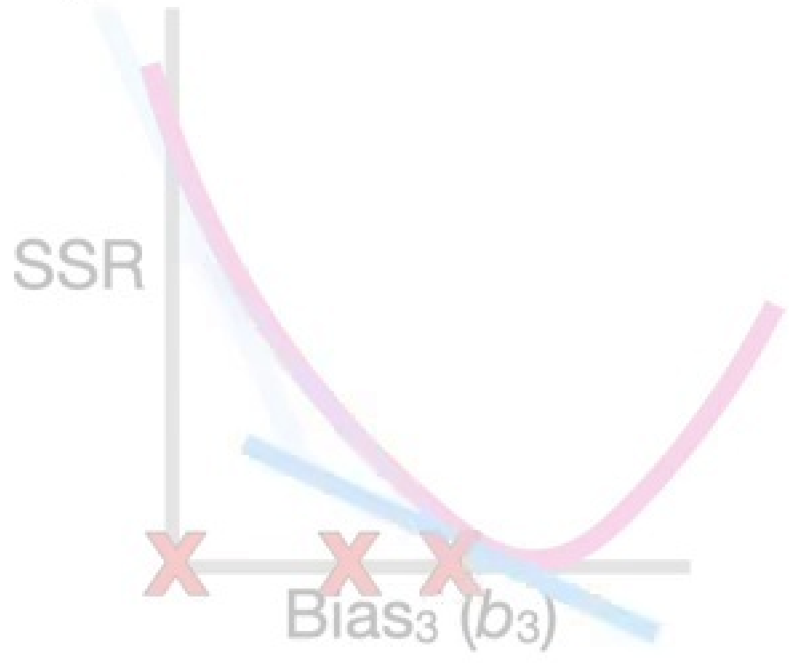
...and the new value for $\boldsymbol{b_3}$...

$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = -6.26$$

Step Size = -6.26 × 0.1 = -0.626

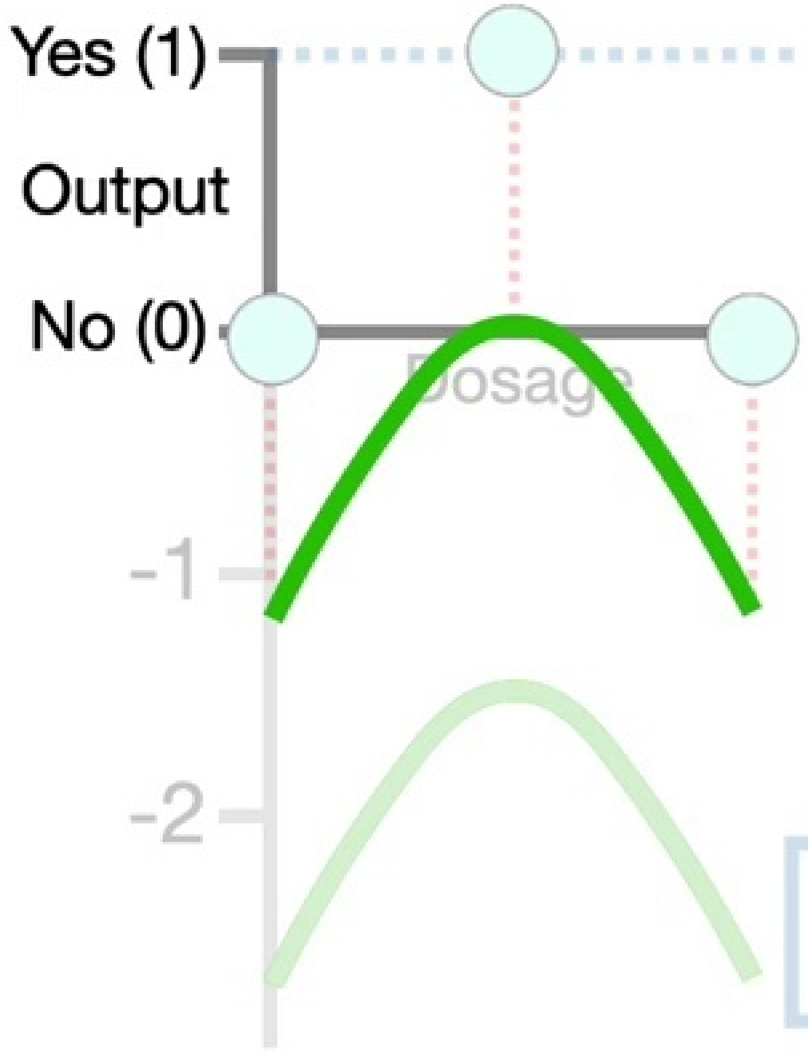
New b_3 = 1.57 - (-0.626) = 2.19

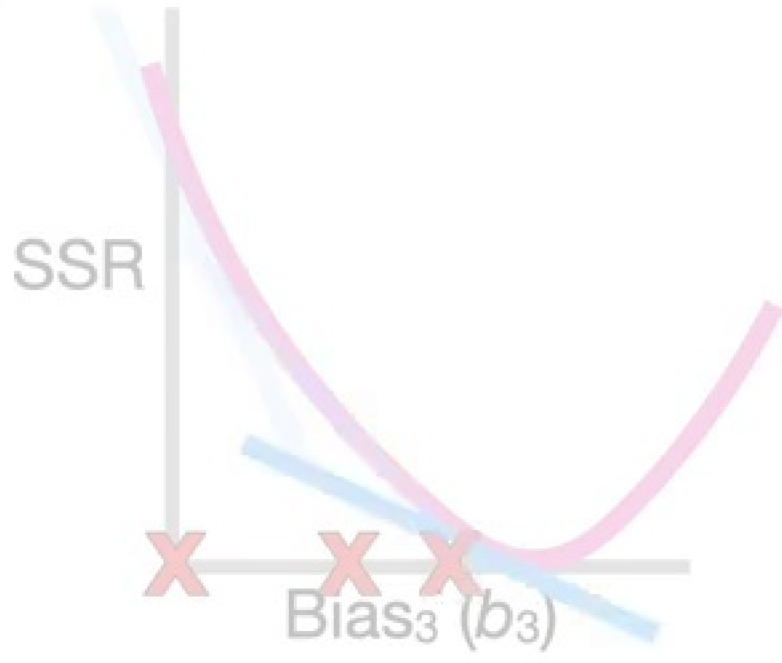




$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - -1.03) \times 1 + -2 \times (1 - -0.03) \times 1 + -2 \times (0 - -1.04) \times 1 = -6.26$$

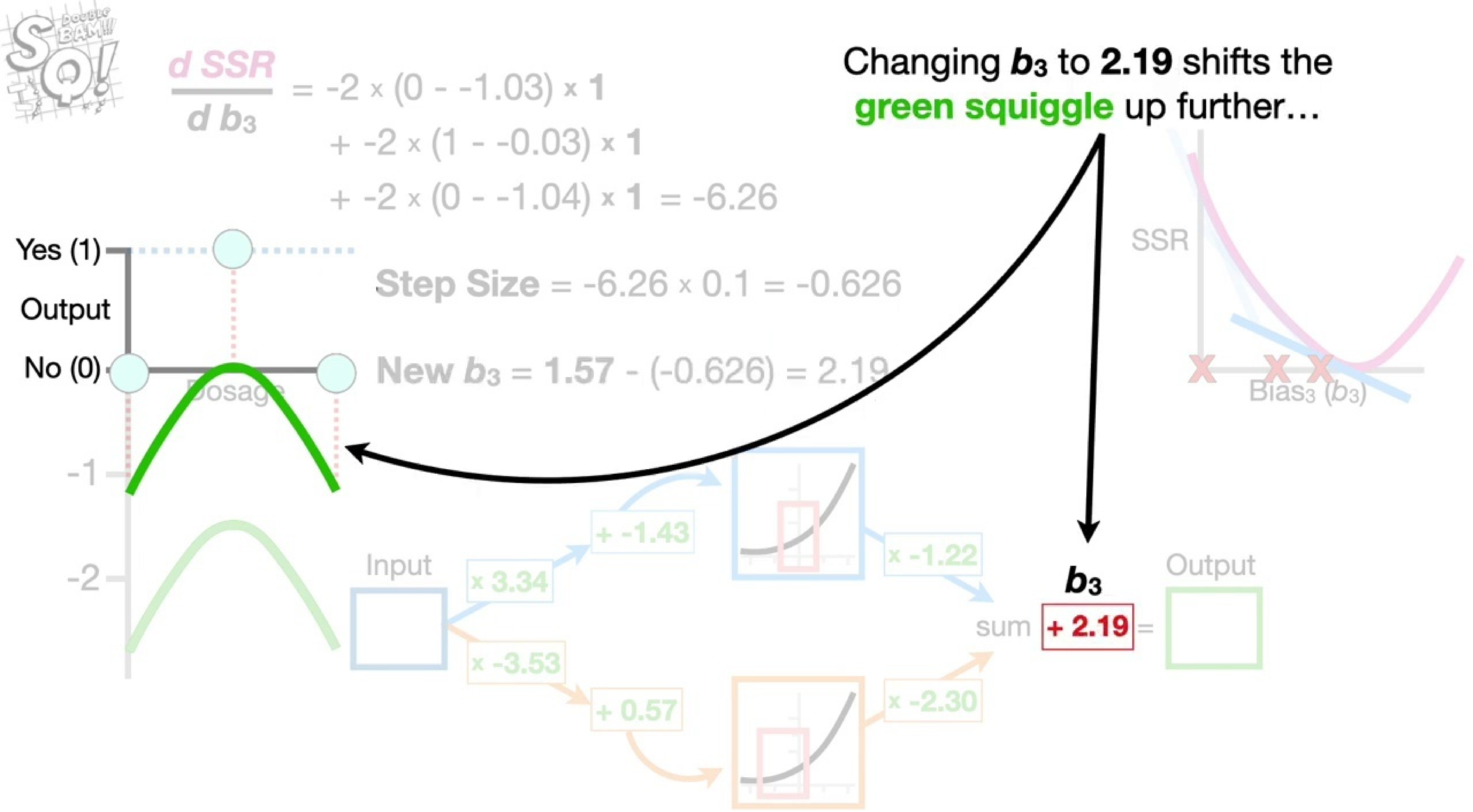
…which is **2.19.**

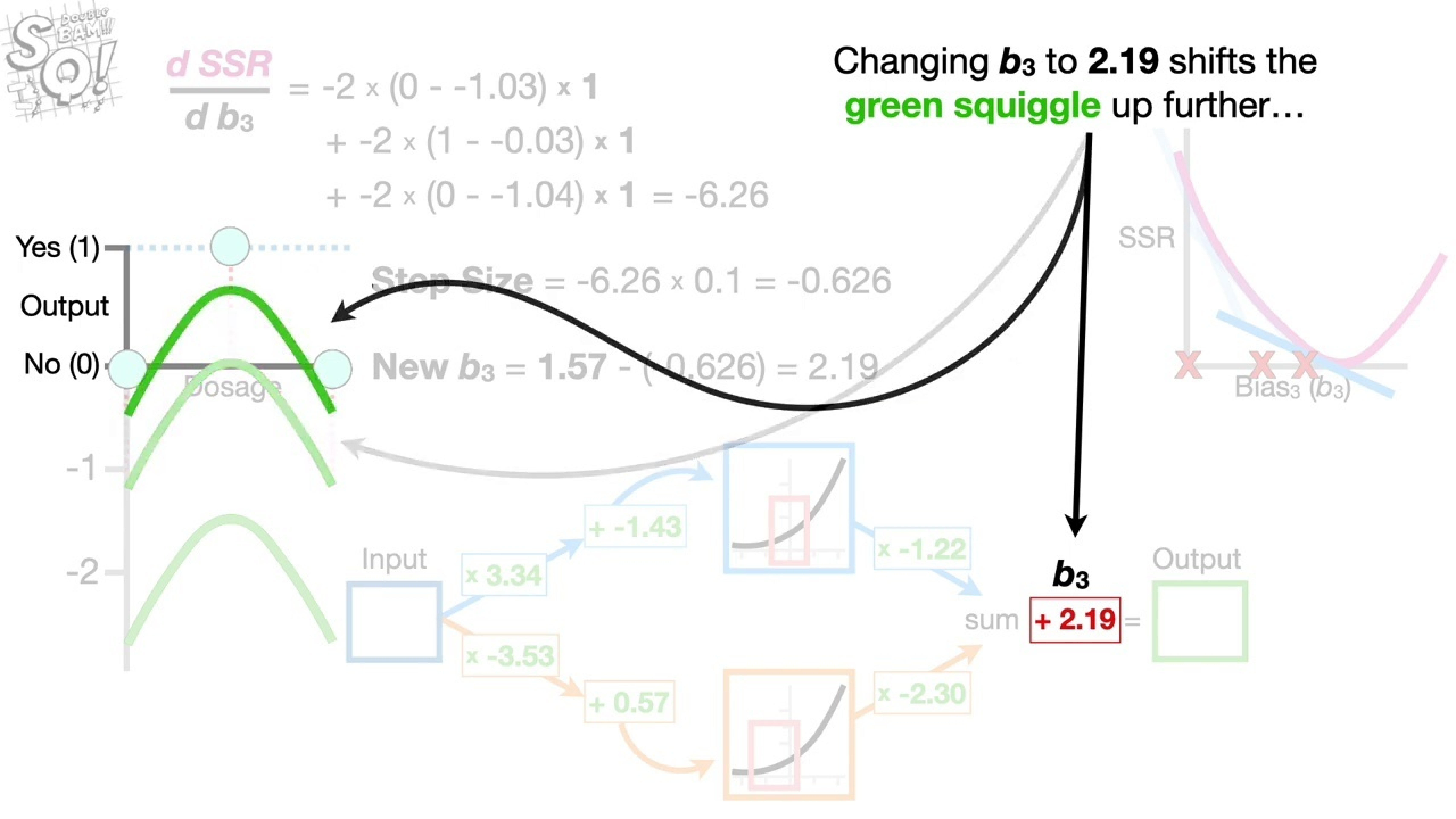


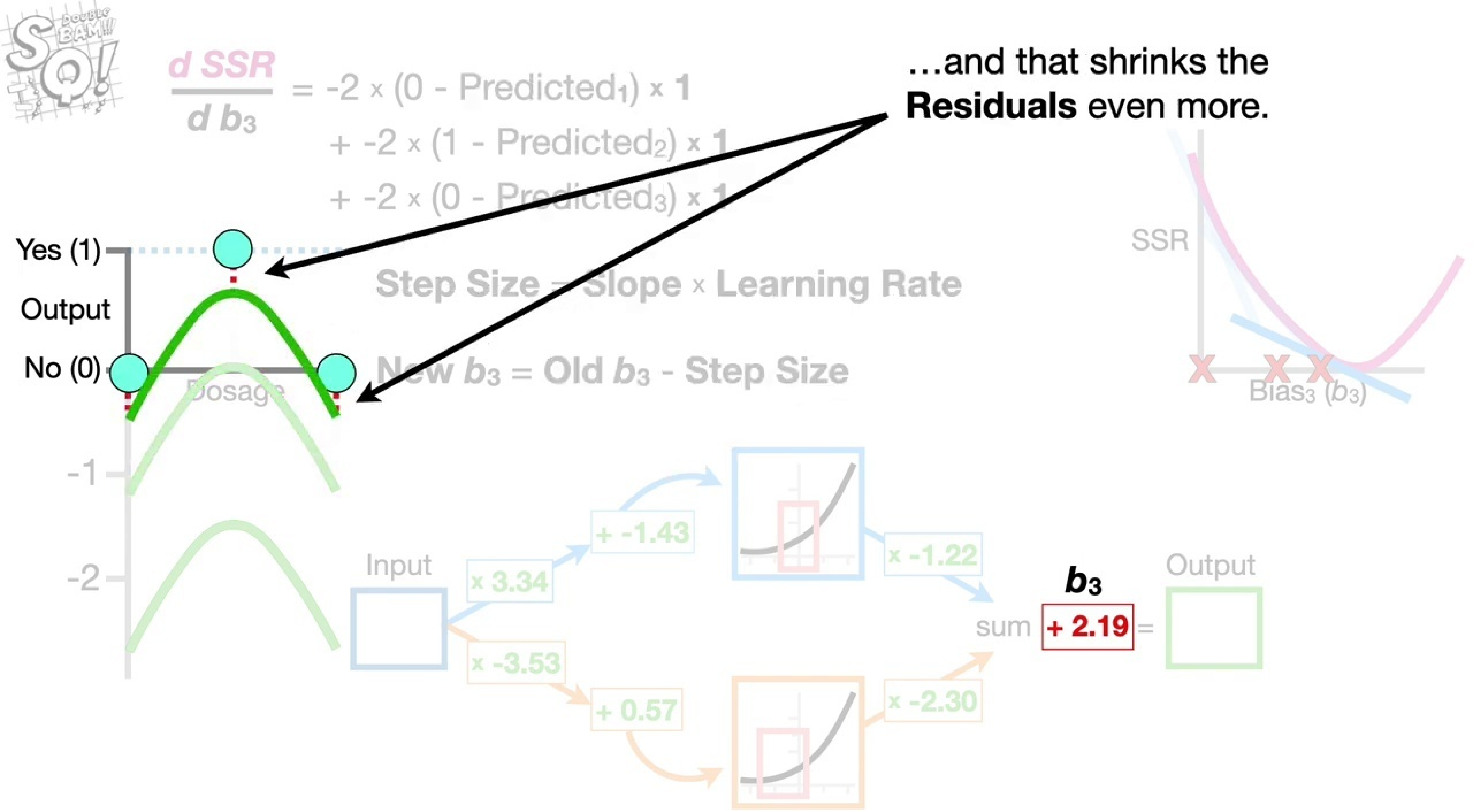


Step Size = -6.26 × 0.1 = -0.626

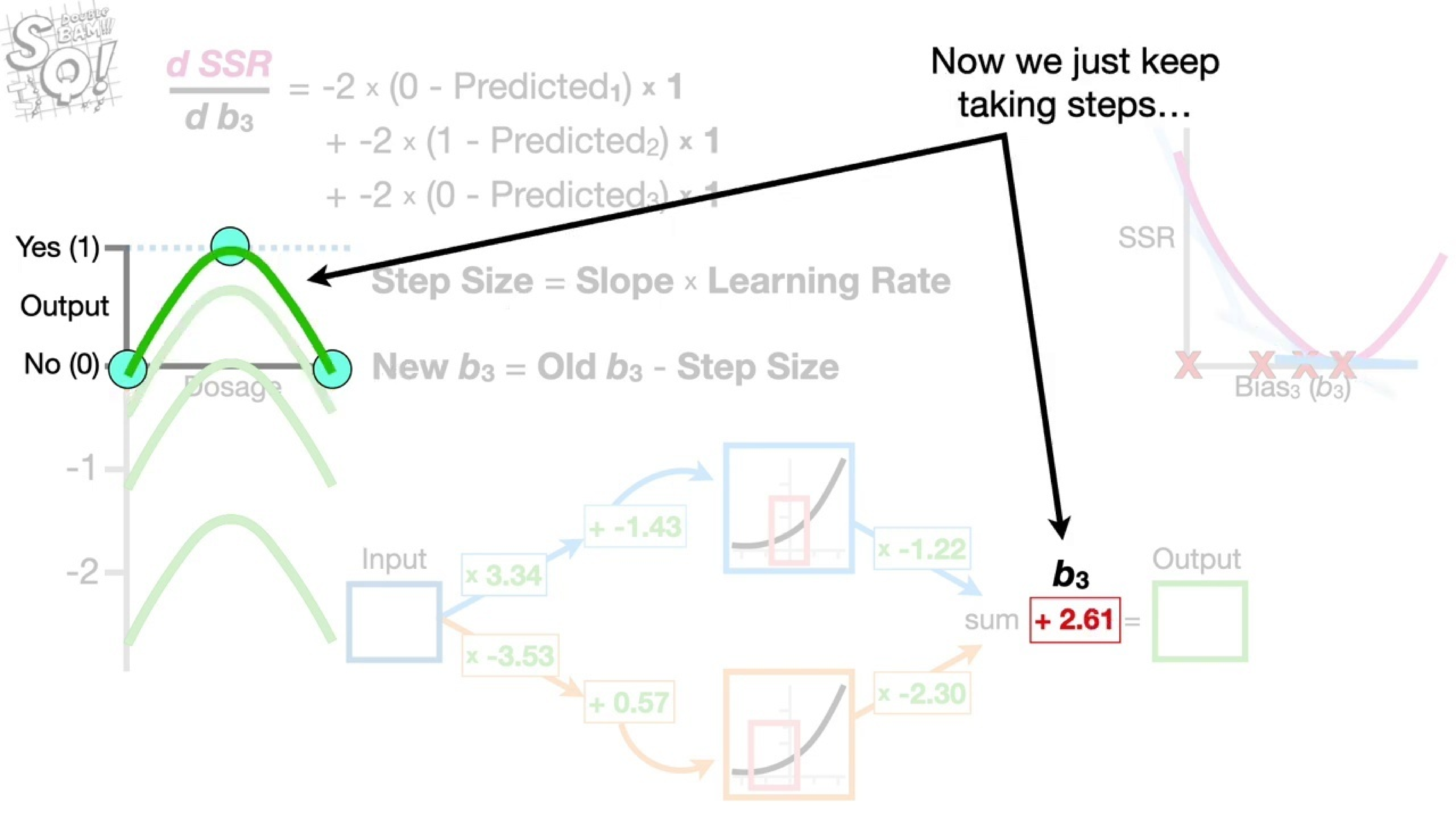
New b_3 = 1.57 - (-0.626) $\boxed{= 2.19}$


Changing b_3 to 2.19 shifts the green squiggle up further…
$\frac{d\ SSR}{d\ b_3}$ = -2 x (0 - -1.03) x 1
+ -2 x (1 - -0.03) x 1
+ -2 x (0 - -1.04) x 1 = -6.26
Step Size = -6.26 x 0.1 = -0.626
New b_3 = 1.57 - (-0.626) = 2.19
Yes (1)
Output
No (0)
Dosage
-1
-2
SSR
Bias3 (b_3)
Input
x 3.34
+ -1.43
x -1.22
x -3.53
+ 0.57
x -2.30
b_3
sum + 2.19 =
Output


Changing b_3 to 2.19 shifts the green squiggle up further…
$\frac{d\ SSR}{d\ b_3}$ = -2 x (0 - -1.03) x 1
+ -2 x (1 - -0.03) x 1
+ -2 x (0 - -1.04) x 1 = -6.26
Step Size = -6.26 x 0.1 = -0.626
New b_3 = 1.57 - (-0.626) = 2.19
Yes (1)
Output
No (0)
Dosage
-1
-2
SSR
Bias3 (b_3)
Input
x 3.34
+ -1.43
x -1.22
x -3.53
+ 0.57
x -2.30
b_3
sum + 2.19 =
Output



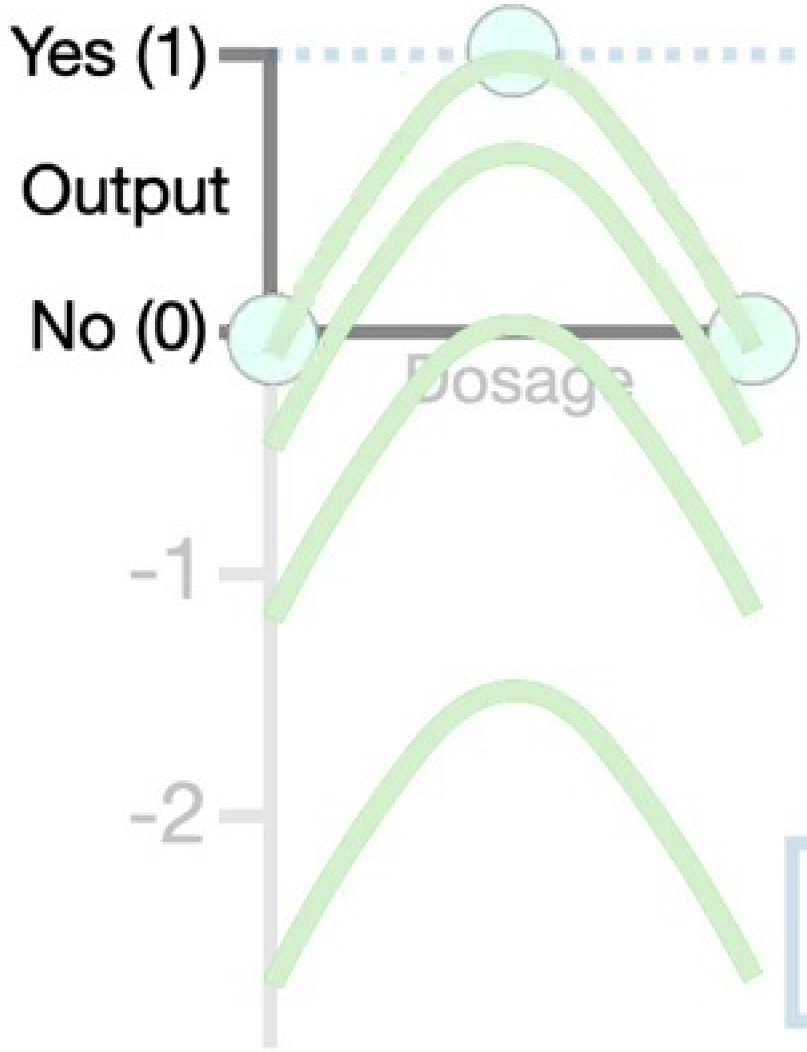
...and that shrinks the **Residuals** even more.
$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - \text{Predicted}_1) \times 1$
$+ -2 \times (1 - \text{Predicted}_2) \times 1$
$+ -2 \times (0 - \text{Predicted}_3) \times 1$
Step Size = Slope × Learning Rate
New b_3 = Old b_3 - Step Size
Yes (1)
Output
No (0)
Dosage
-1
-2
SSR
Bias$_3$ (b_3)
Input
x 3.34
+ -1.43
x -1.22
x -3.53
+ 0.57
x -2.30
b_3
sum
+ 2.19
=
Output



Now we just keep taking steps…
$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - \text{Predicted}_1) \times 1 + -2 \times (1 - \text{Predicted}_2) \times 1 + -2 \times (0 - \text{Predicted}_3) \times 1$
Step Size = Slope x Learning Rate
New b_3 = Old b_3 - Step Size
Yes (1)
Output
No (0)
Dosage
-1
-2
SSR
Bias$_3$ (b_3)
Input
x 3.34
+ -1.43
x -1.22
x -3.53
+ 0.57
x -2.30
b_3
sum
+ 2.61 =
Output

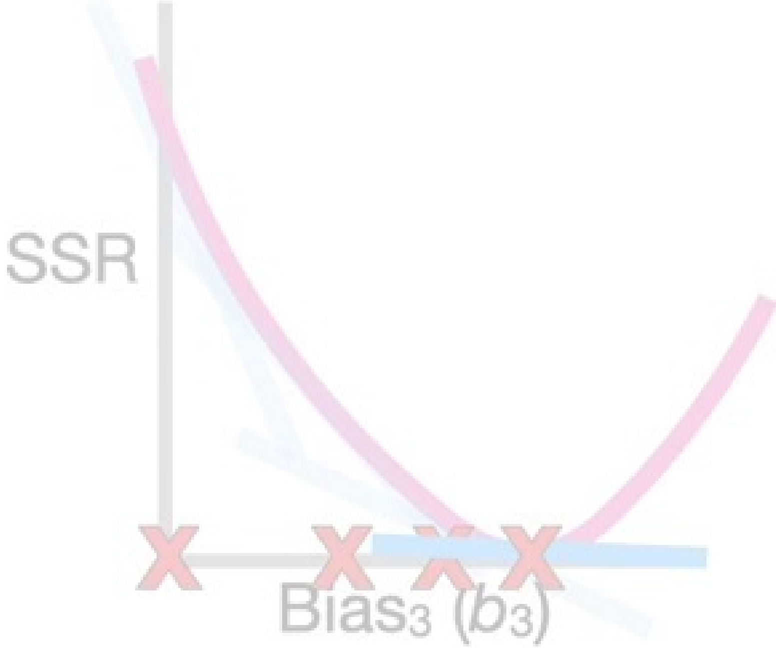
$$\frac{d\,SSR}{d\,b_3} = -2 \times (0 - 0.01) + -2 \times (1 - 1.01) + -2 \times (0 - 0) = \text{close to } 0$$

…until the **Step Size** is close to **0**.



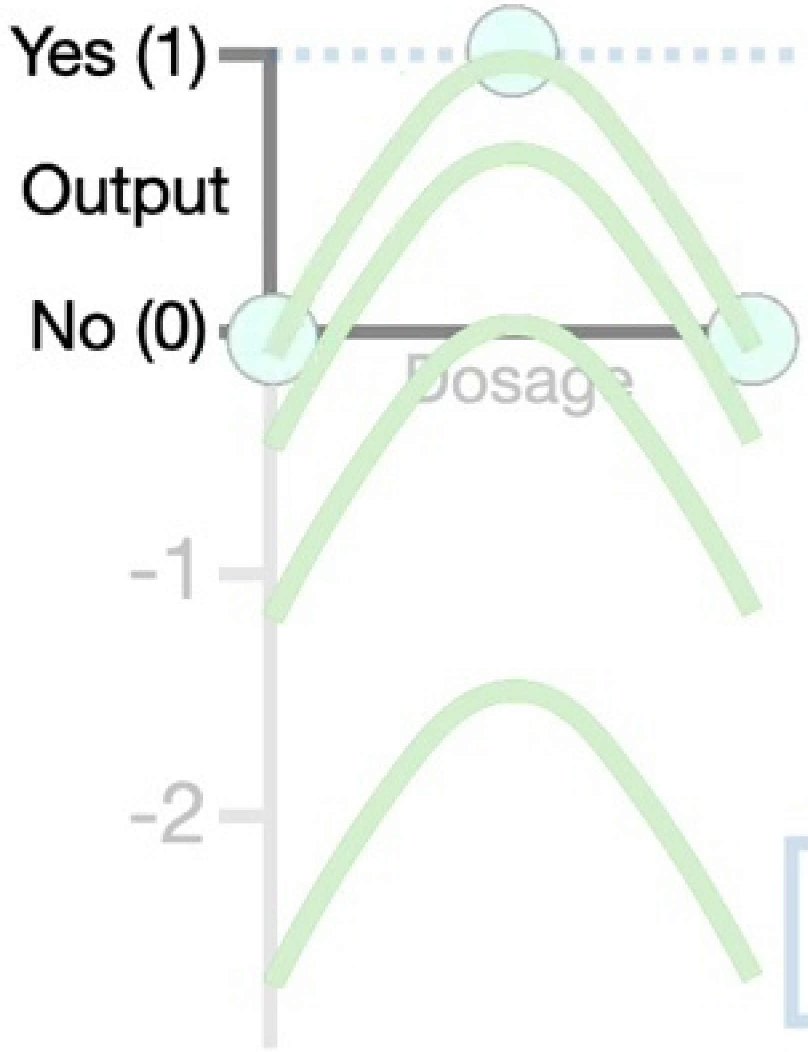
Step Size = close to 0 × 0.1 = close to 0

New b_3 = Old b_3 - Step Size



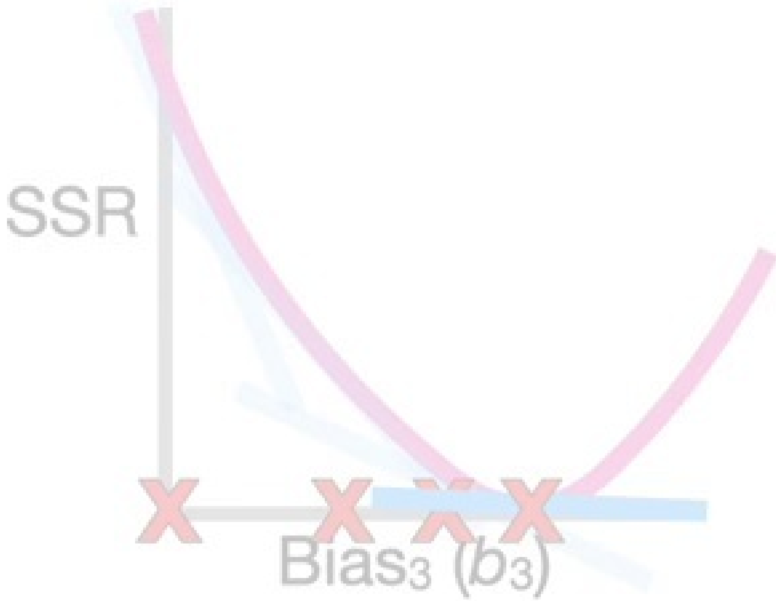
$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - 0.01) + -2 \times (1 - 1.01) + -2 \times (0 - 0) = \text{close to } 0$$

And because the **Step Size** is close to **0** when $\boldsymbol{b_3}$ **= 2.61**…



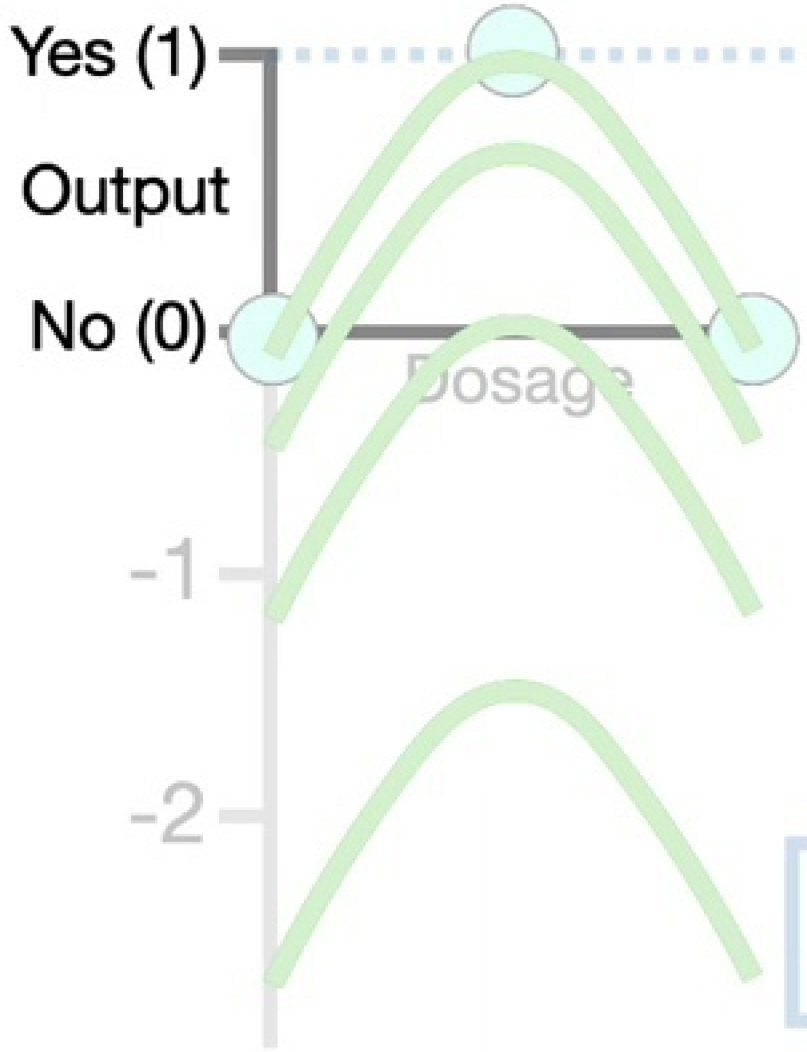
Step Size = close to 0 × 0.1 $\boxed{= \text{close to } 0}$

New $\boldsymbol{b_3}$ **= Old** $\boldsymbol{b_3}$ **- Step Size**



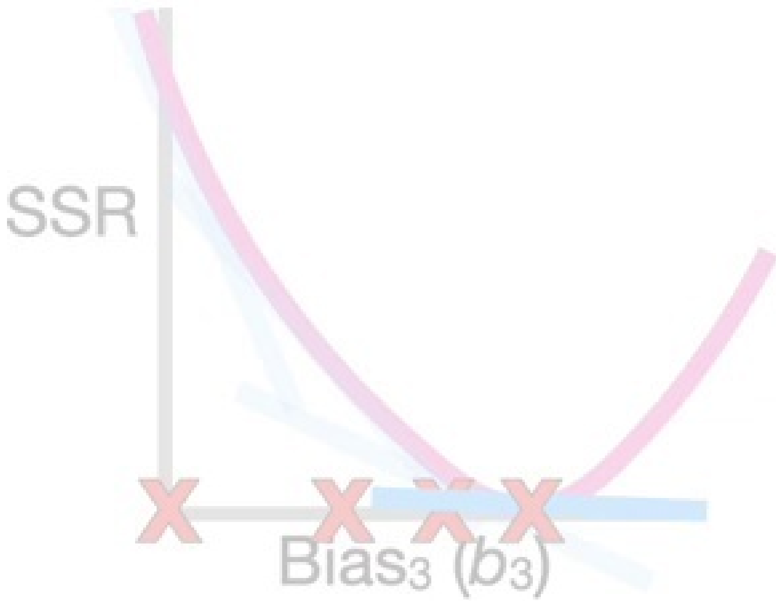
$$\frac{d\ SSR}{d\ b_3} = -2 \times (0 - 0.01) + -2 \times (1 - 1.01) + -2 \times (0 - 0) = \text{close to } 0$$

…we decide that **2.61** is the optimal value for ***b*₃**.

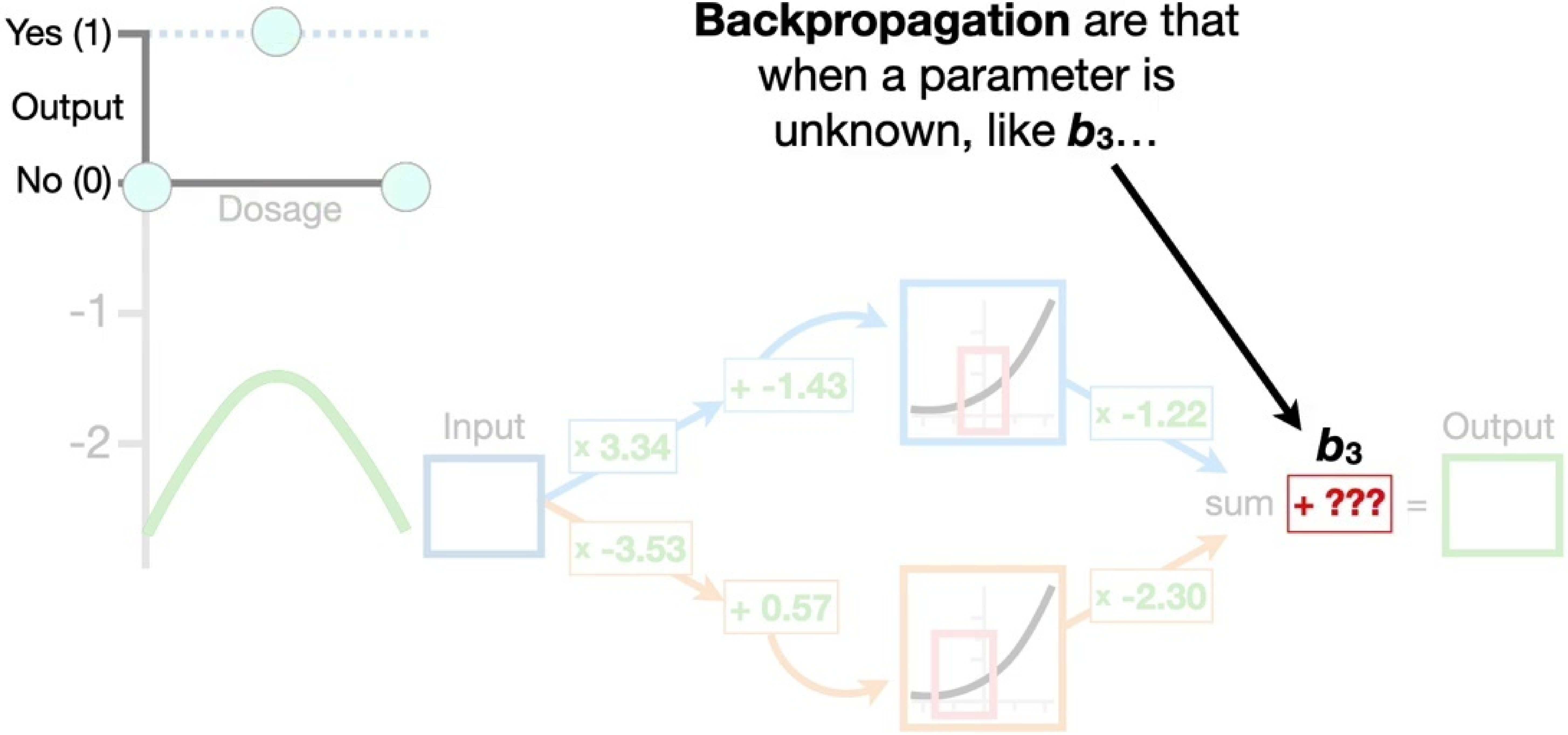


Step Size = close to 0 × 0.1 $\boxed{= \text{close to } 0}$

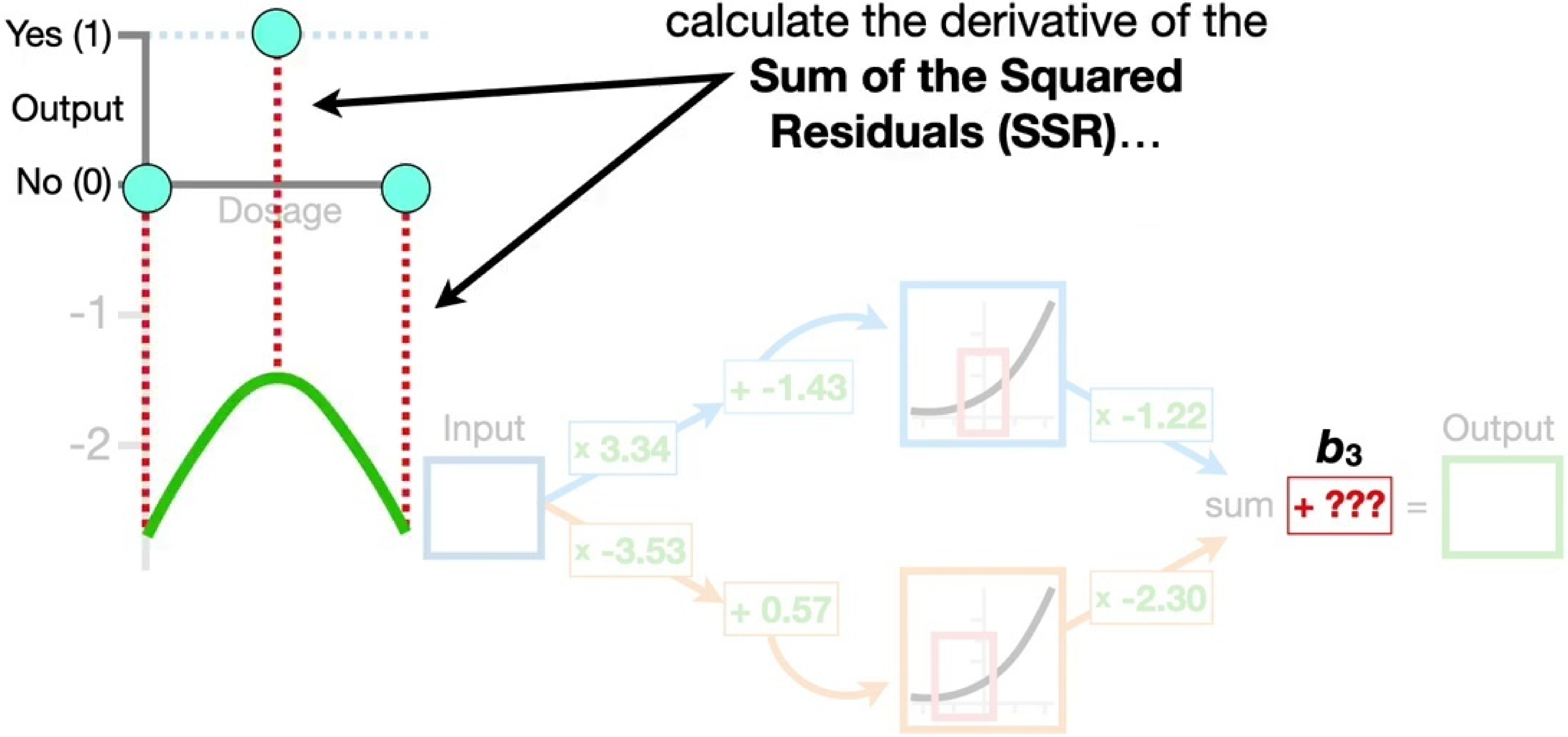
New b_3 = Old b_3 - Step Size



So the **Main Ideas** for **Backpropagation** are that when a parameter is unknown, like b_3…

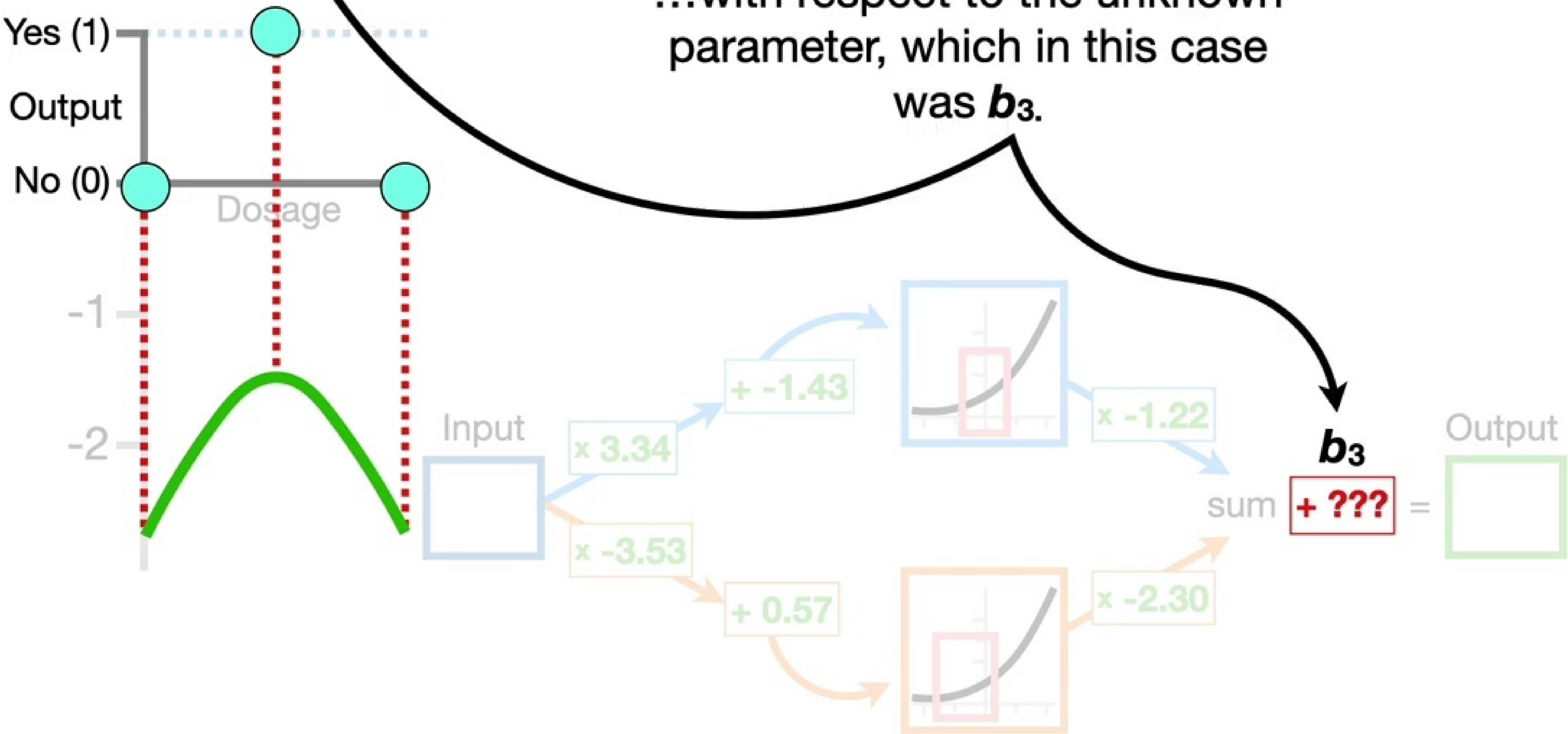


…we use **The Chain Rule** to calculate the derivative of the **Sum of the Squared Residuals (SSR)…**



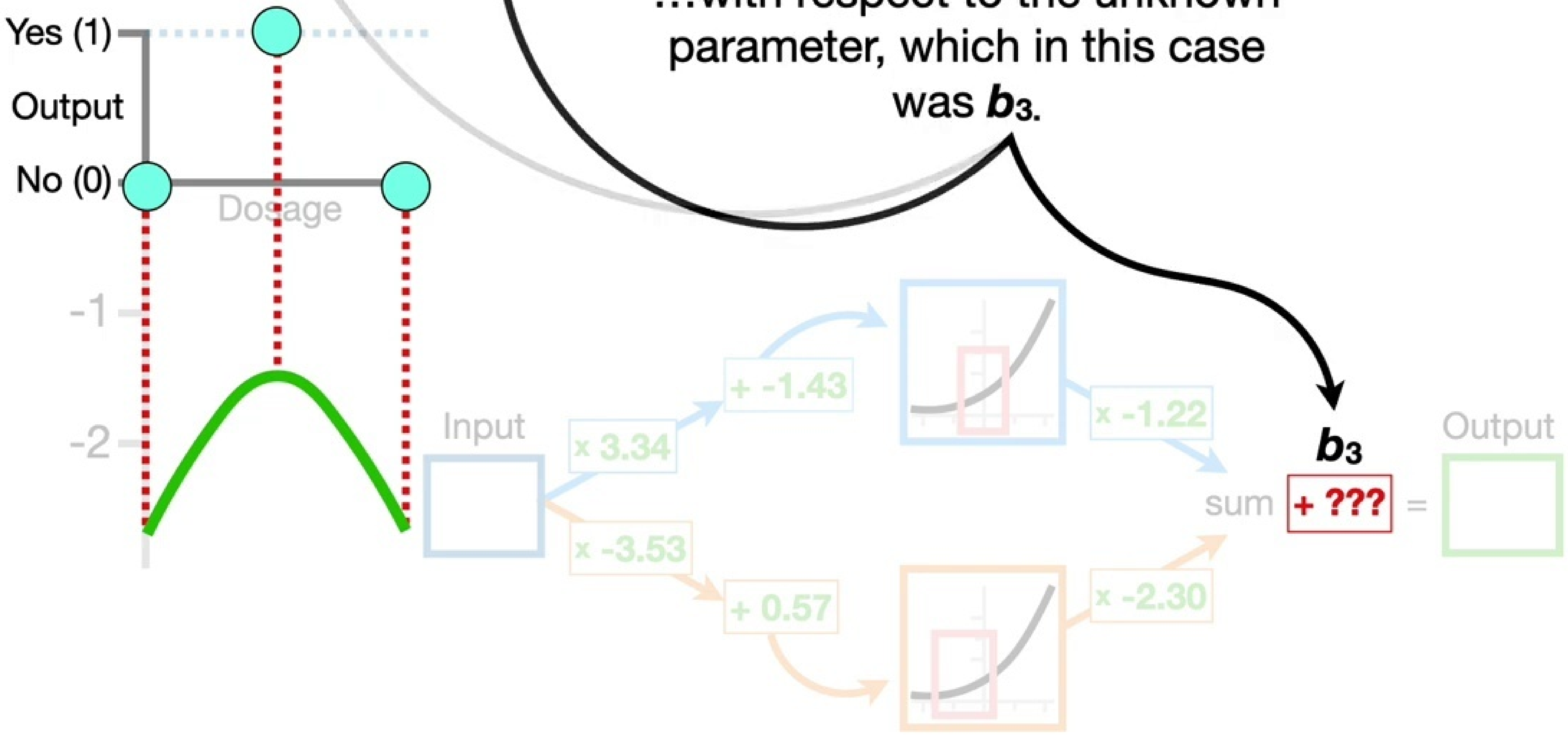
$$\frac{d\ SSR}{d\ b_3}$$

...with respect to the unknown parameter, which in this case was b_3.



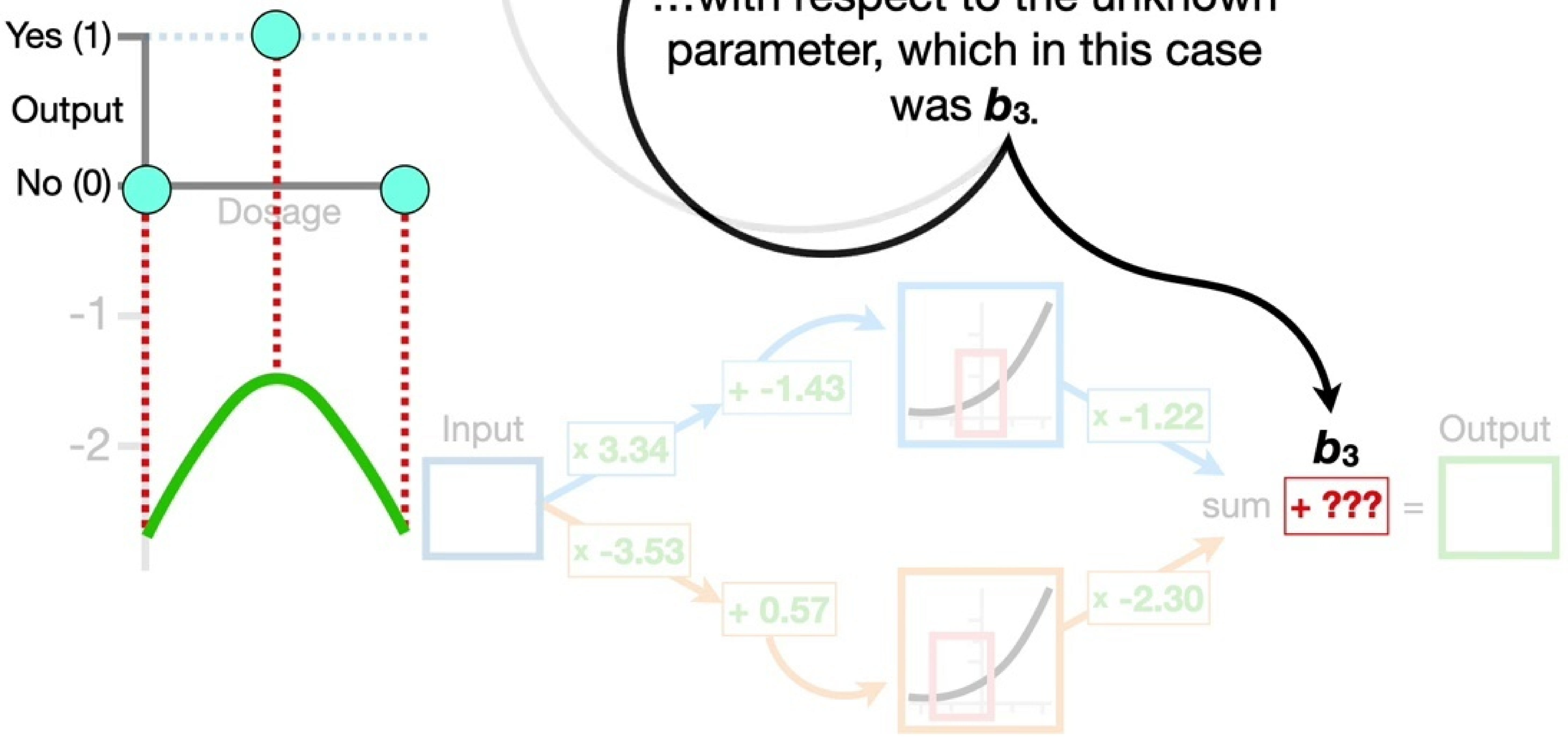
$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted}$$

…with respect to the unknown parameter, which in this case was b_3.



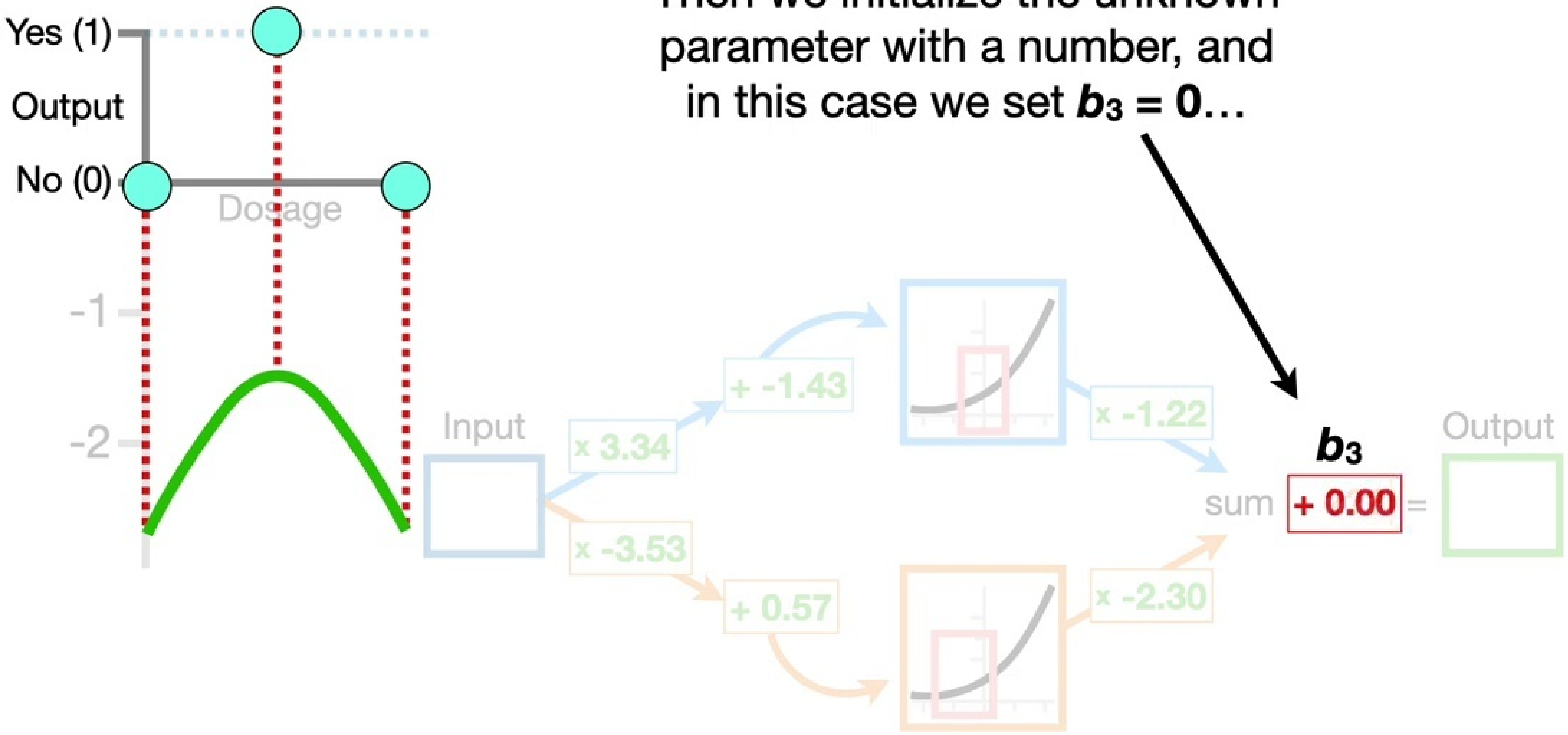
$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

…with respect to the unknown parameter, which in this case was b_3.



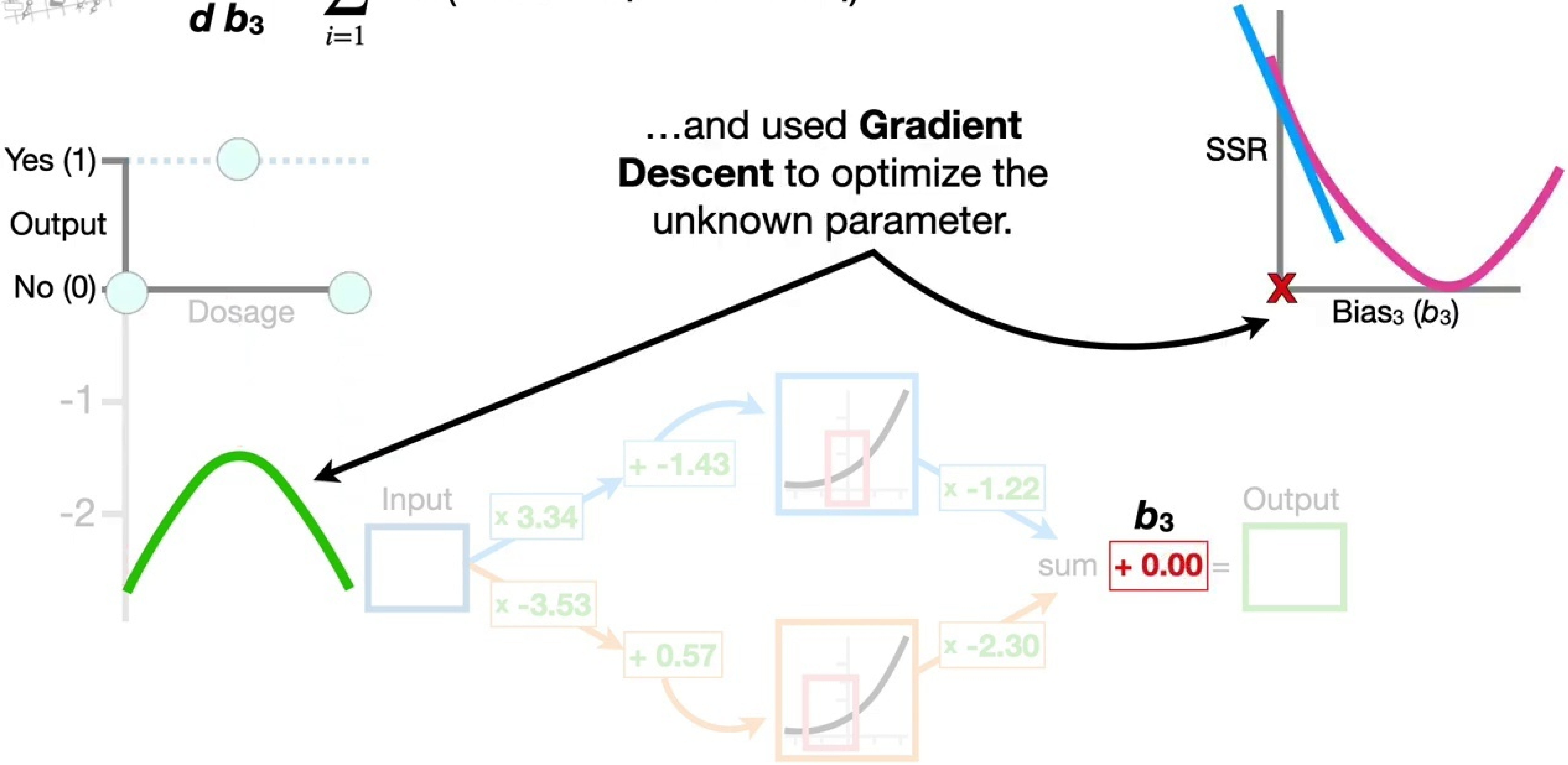
$$\frac{d\ SSR}{d\ b_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times 1$$

Then we initialize the unknown parameter with a number, and in this case we set $b_3 = 0$…



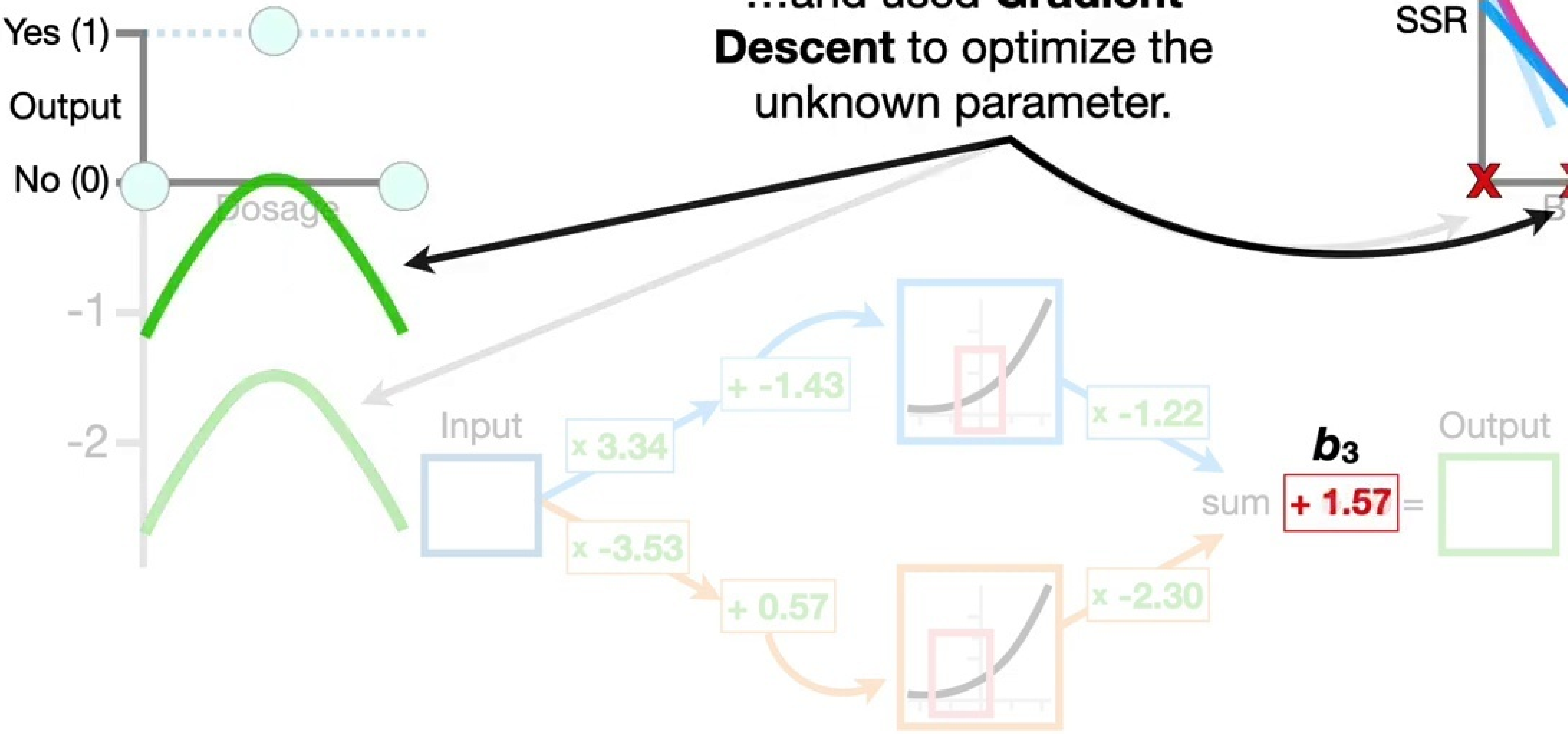
$$\frac{d\ SSR}{d\ b_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times \mathbf{1}$$

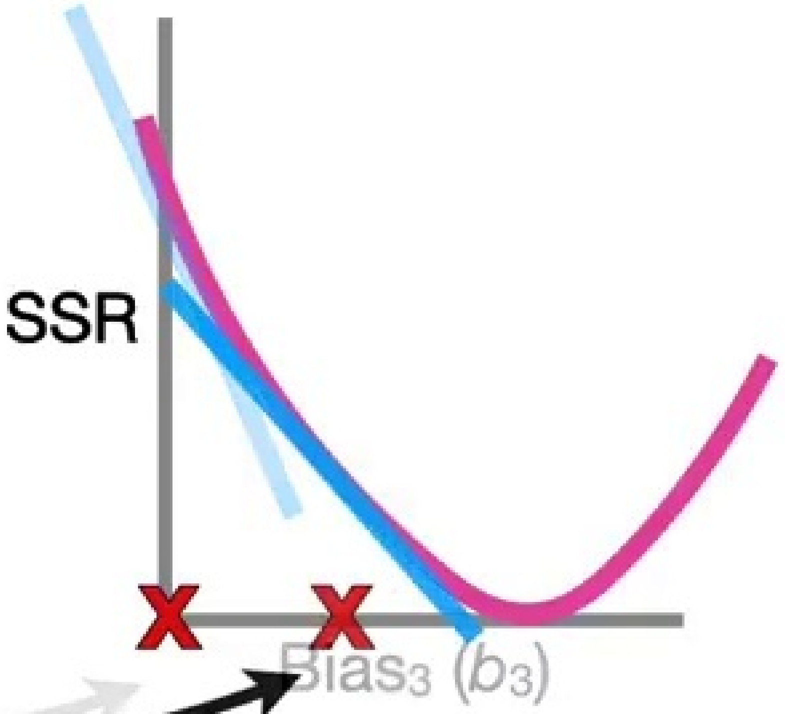
...and used **Gradient Descent** to optimize the unknown parameter.



$$\frac{d\ SSR}{d\ b_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times \mathbf{1}$$

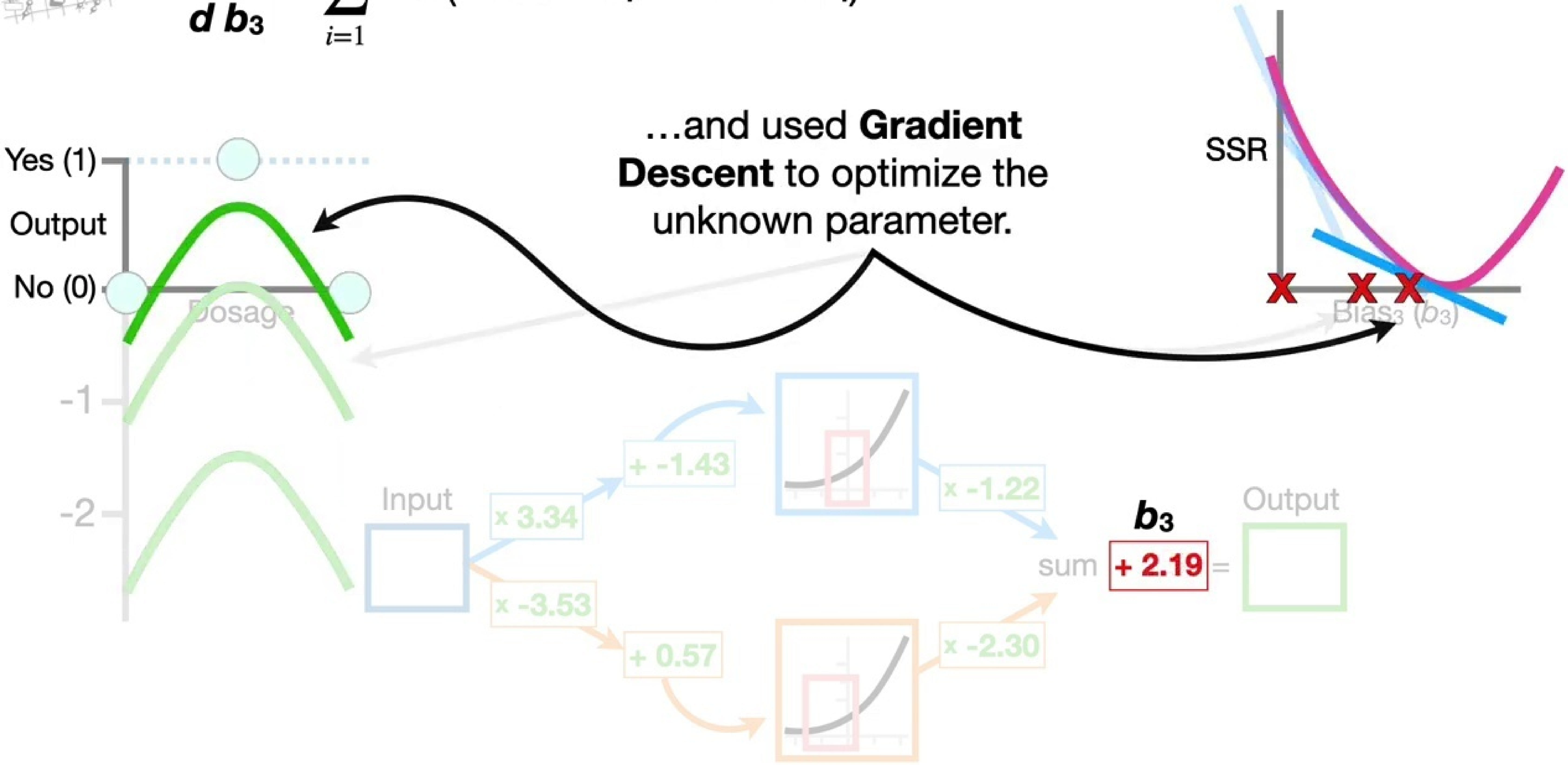
...and used **Gradient Descent** to optimize the unknown parameter.





$$\frac{d\ SSR}{d\ b_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times 1$$

…and used **Gradient Descent** to optimize the unknown parameter.



$$\frac{d\ SSR}{d\ b_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times \mathbf{1}$$

...and used **Gradient Descent** to optimize the unknown parameter.