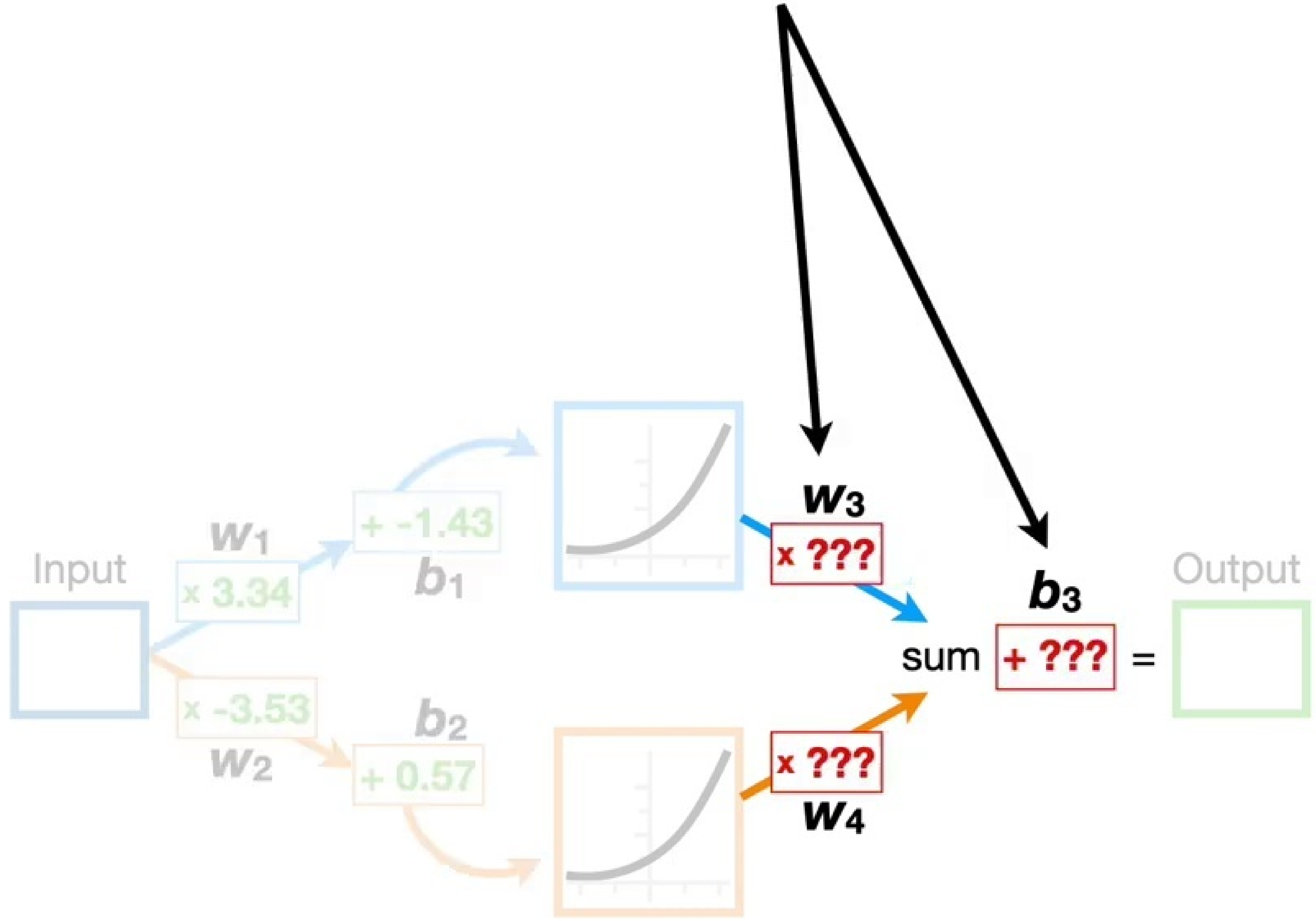
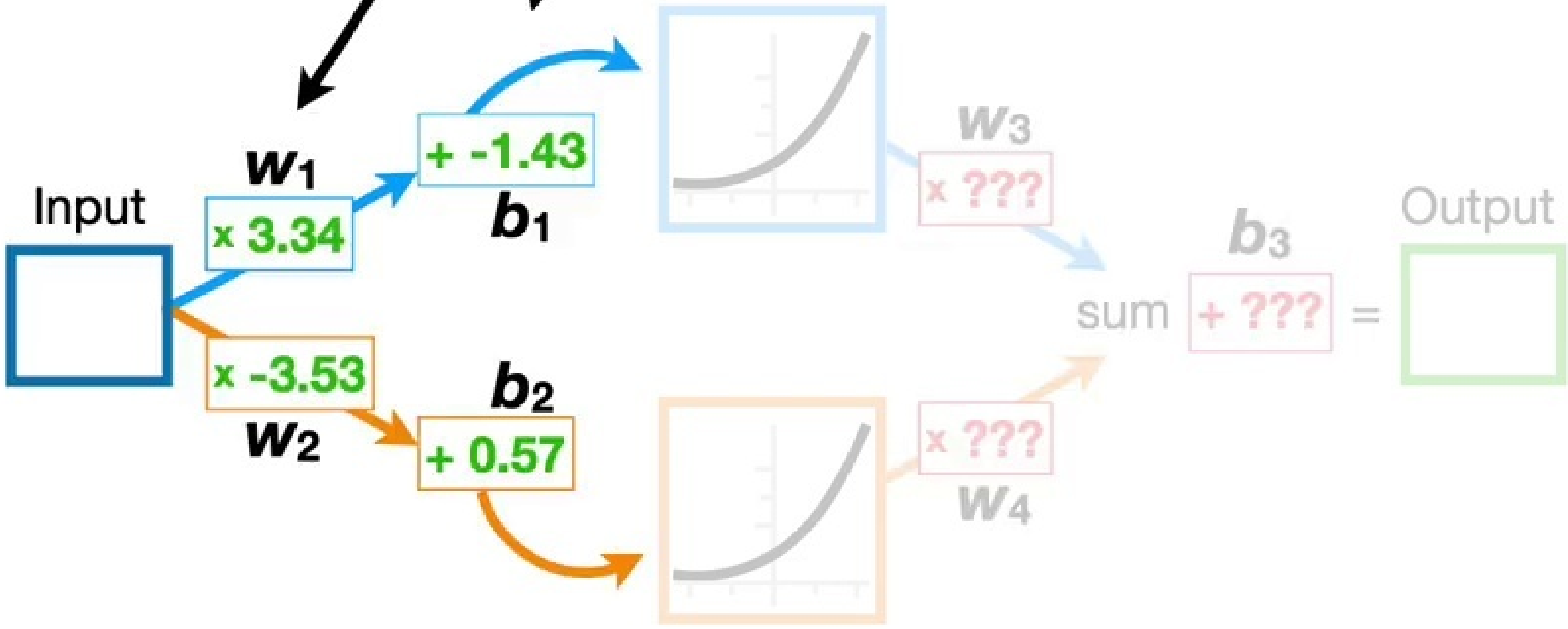
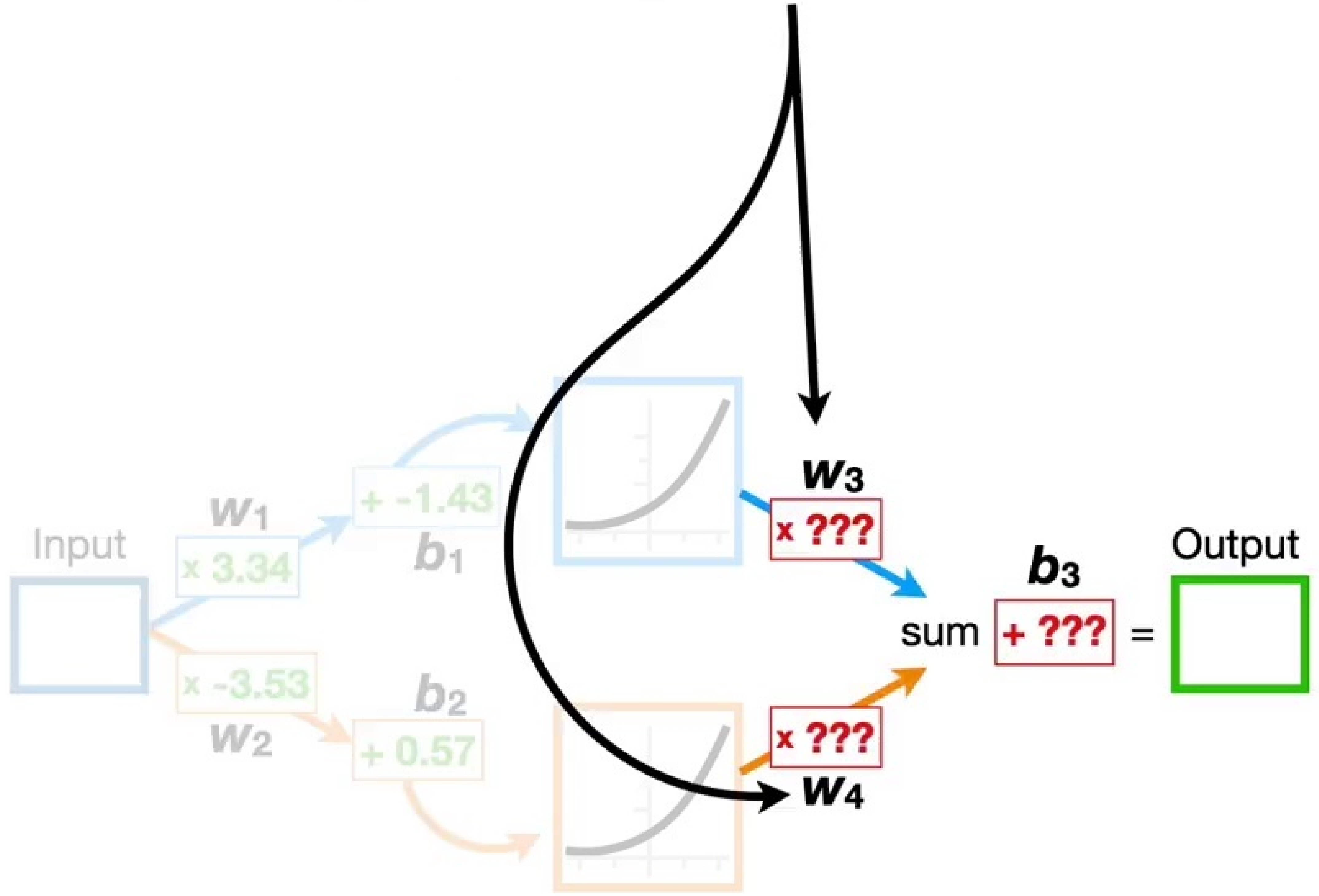
So let's go back to not knowing the optimal values for w_3, w_4 and b_3…



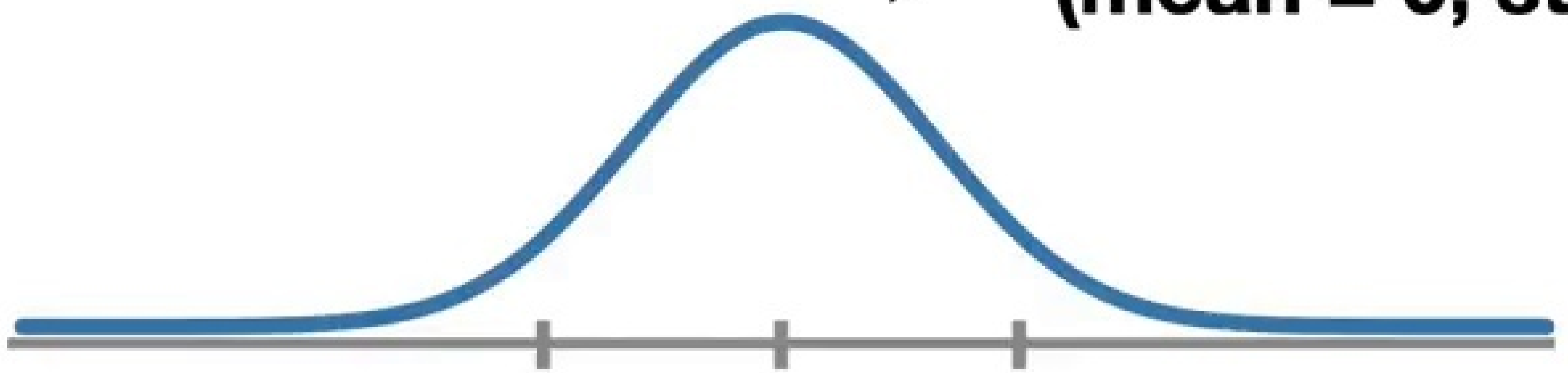
…and, just like before, we'll assume that the other **Weights** and **Biases** are already optimized.

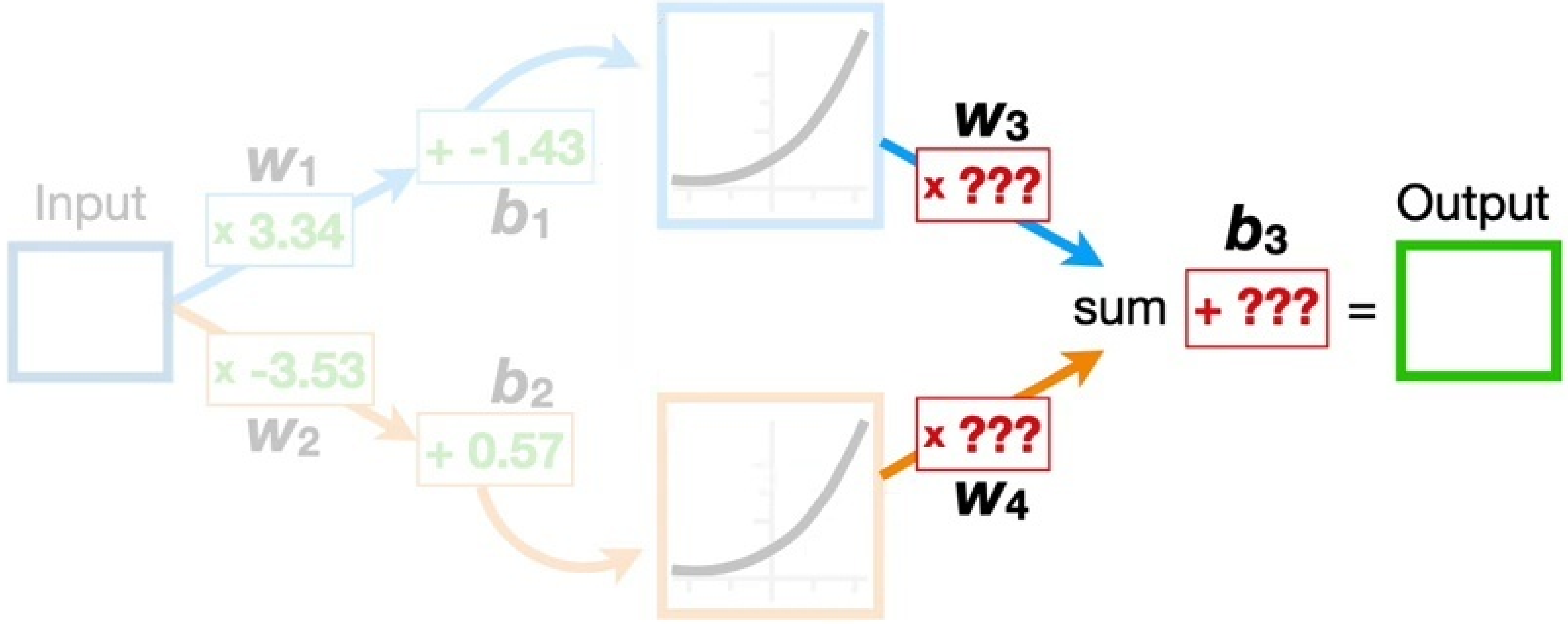


The first thing we do is initialize the **Weights**, ***w*₃** and ***w*₄**, with random starting values…

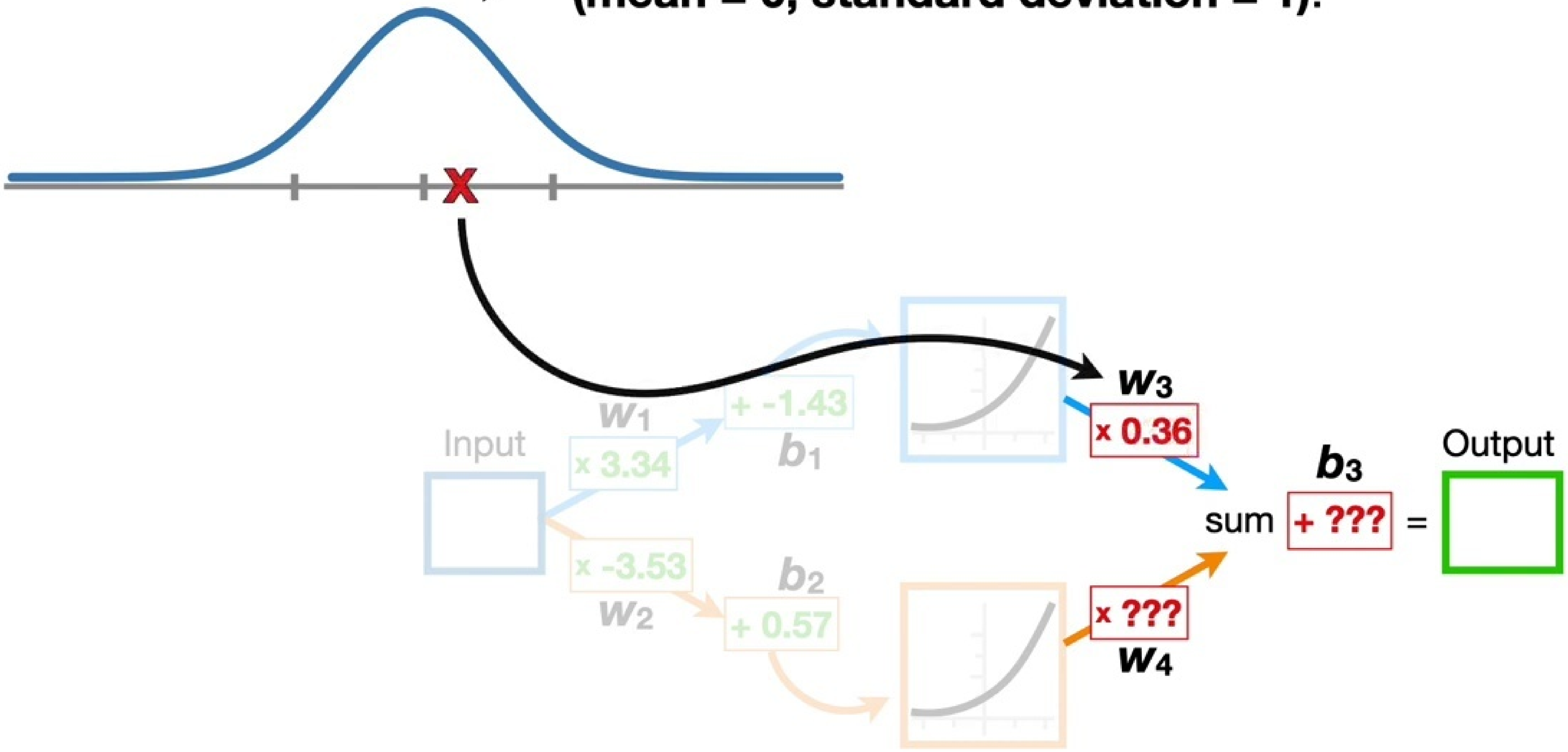


…and, in this example, that means we randomly select **2** values from a **Standard Normal Distribution (mean = 0, standard deviation = 1).**

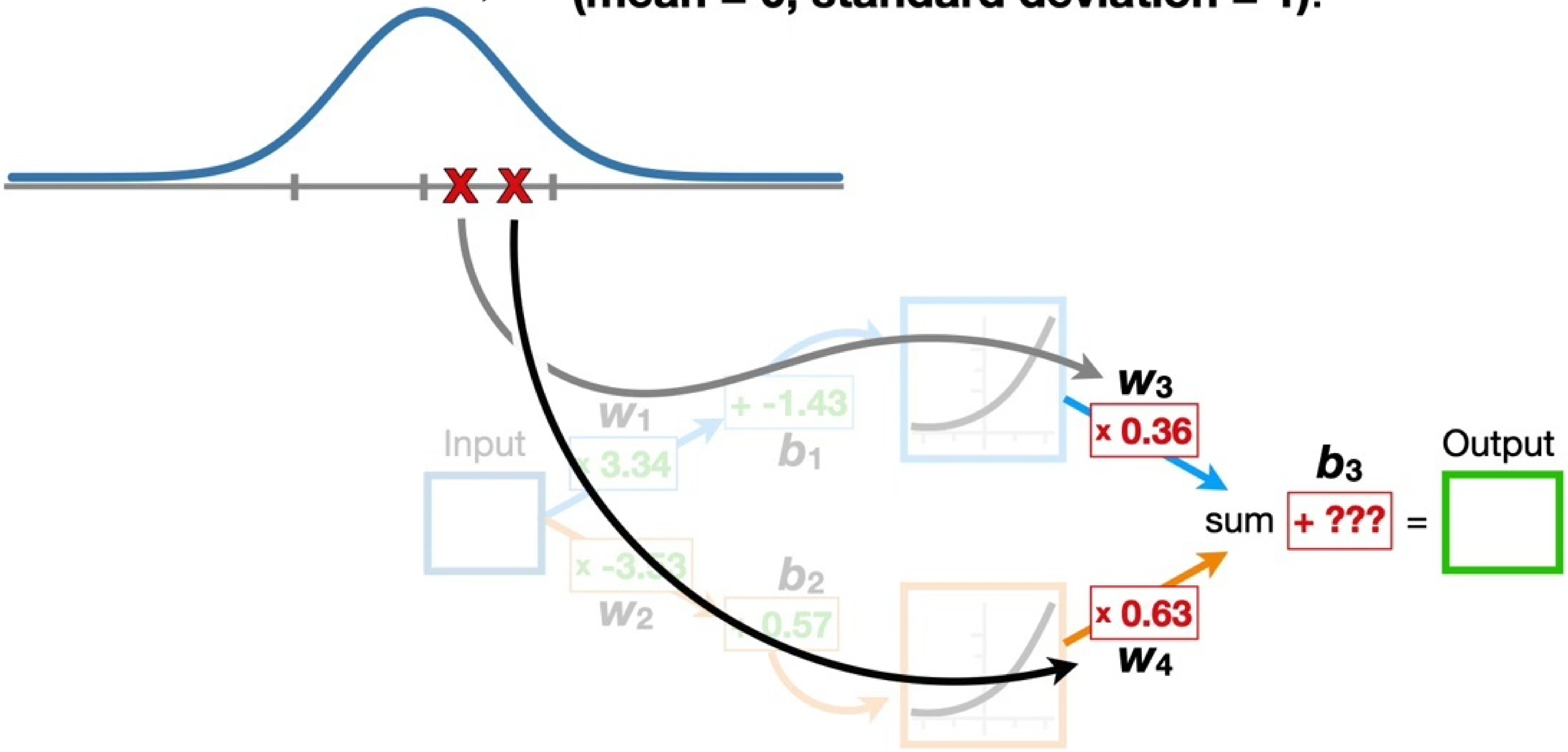




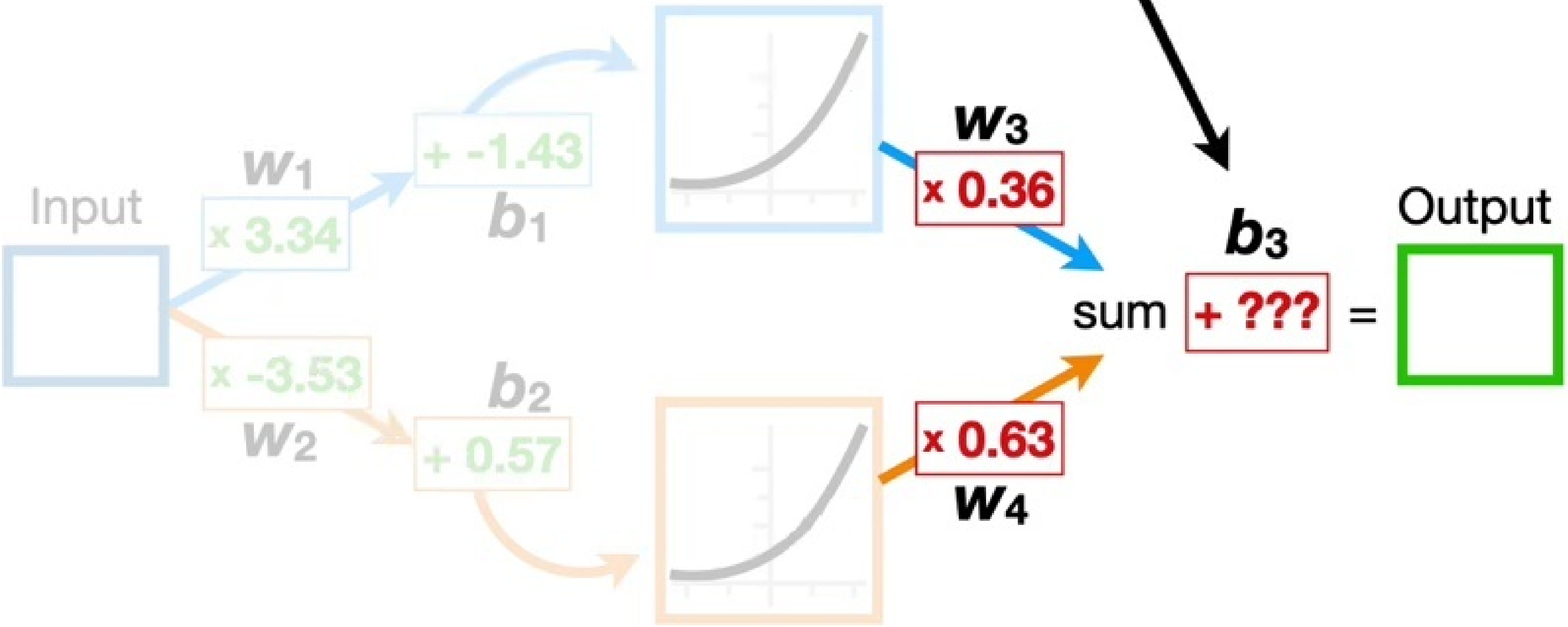
…and, in this example, that means we randomly select **2** values from a **Standard Normal Distribution (mean = 0, standard deviation = 1).**



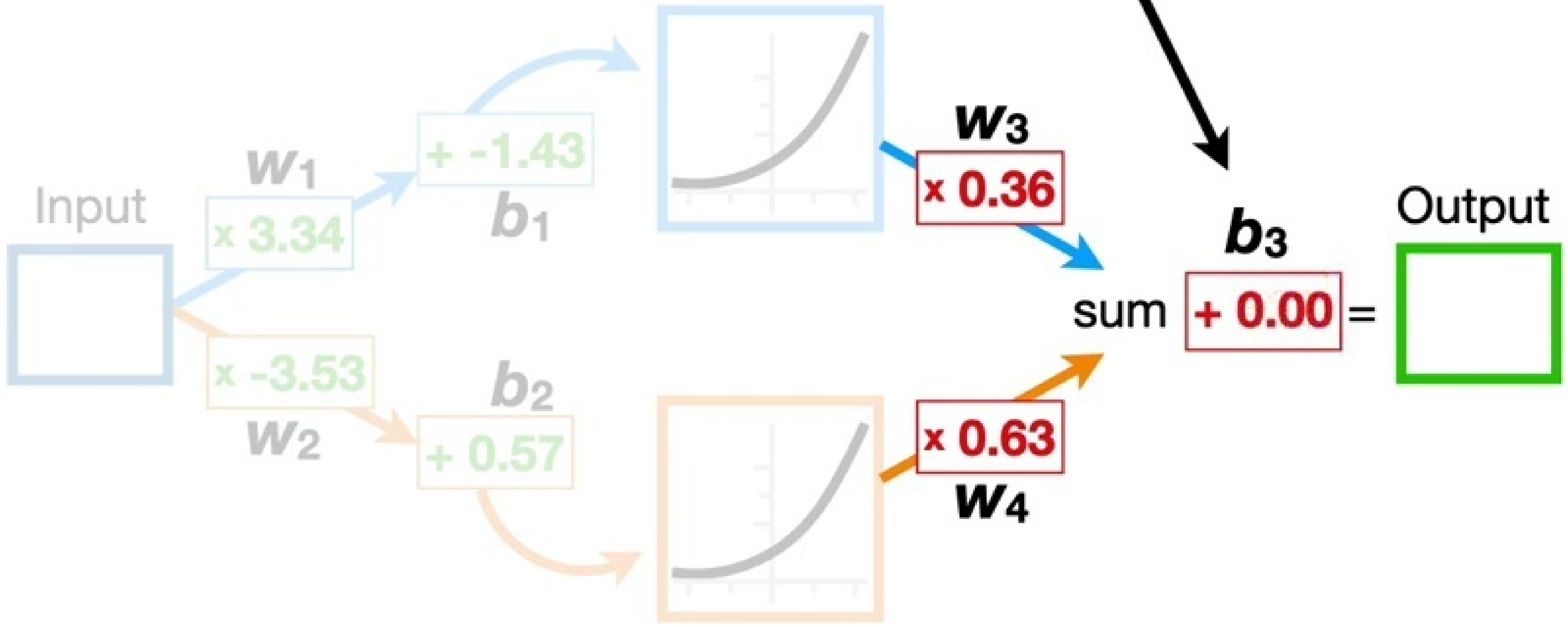
…and, in this example, that means we randomly select **2** values from a **Standard Normal Distribution (mean = 0, standard deviation = 1).**



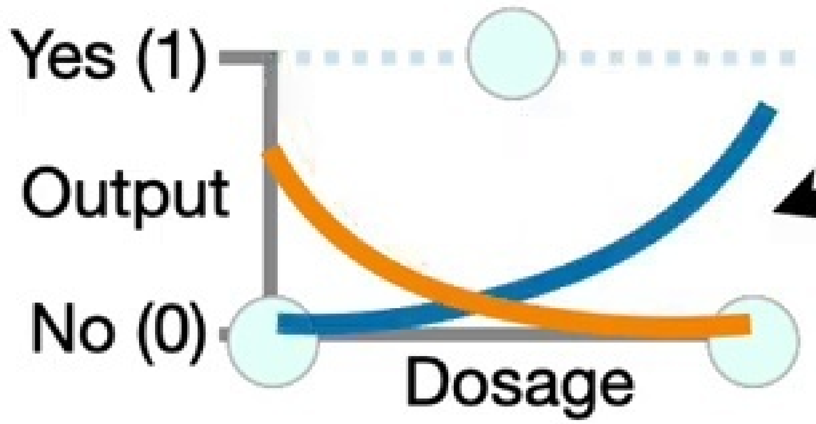
Then we initialize the last **Bias**, $\boldsymbol{b_3}$, to **0**, because **Bias** terms frequently start at **0**.

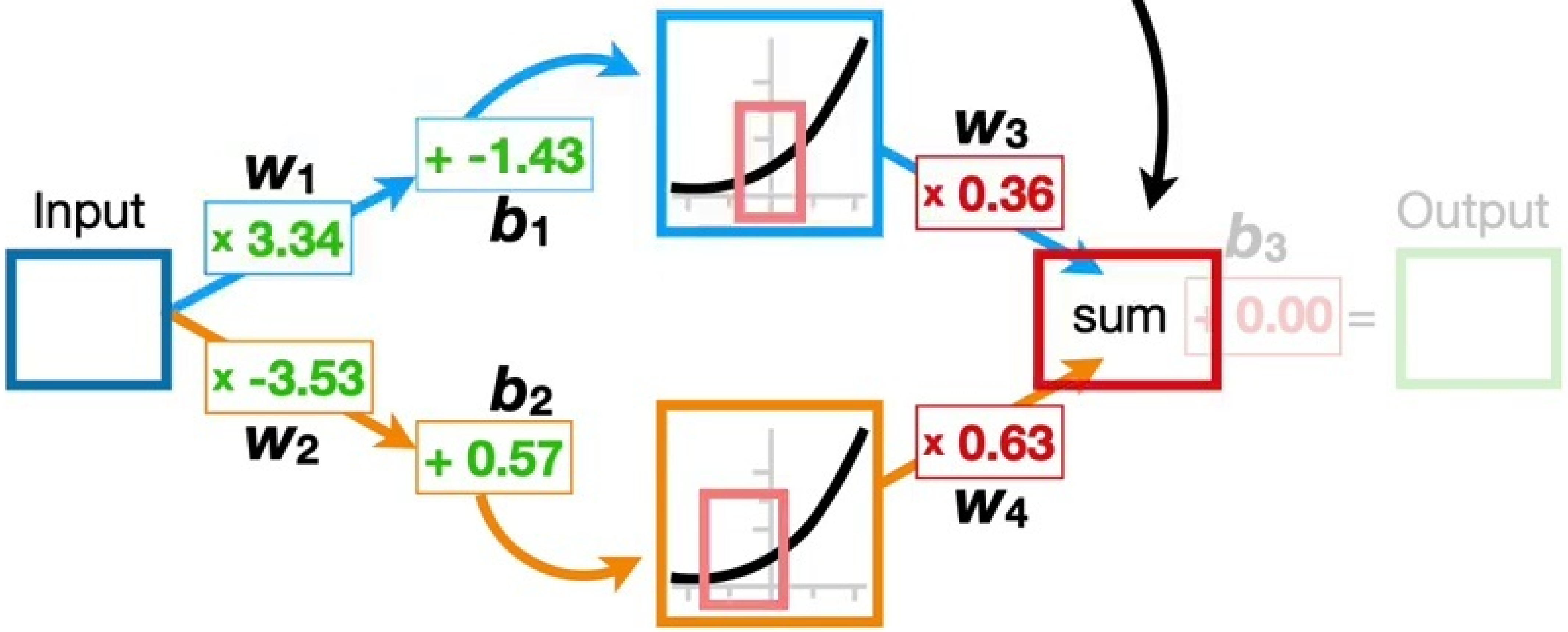


Then we initialize the last **Bias**, ***b*₃**, to **0**, because **Bias** terms frequently start at **0**.

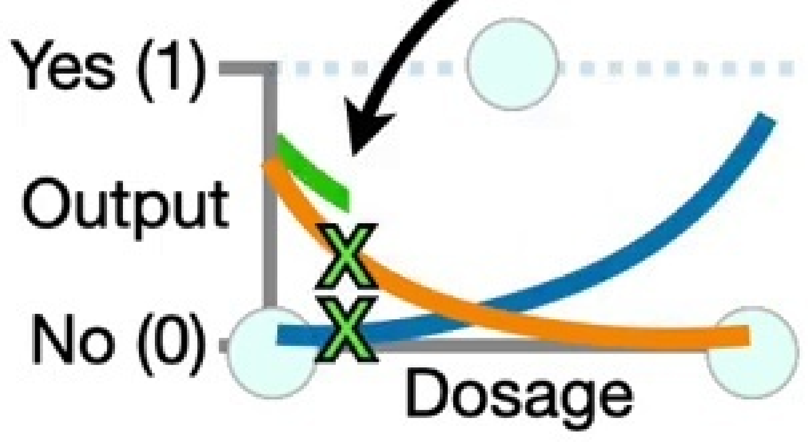


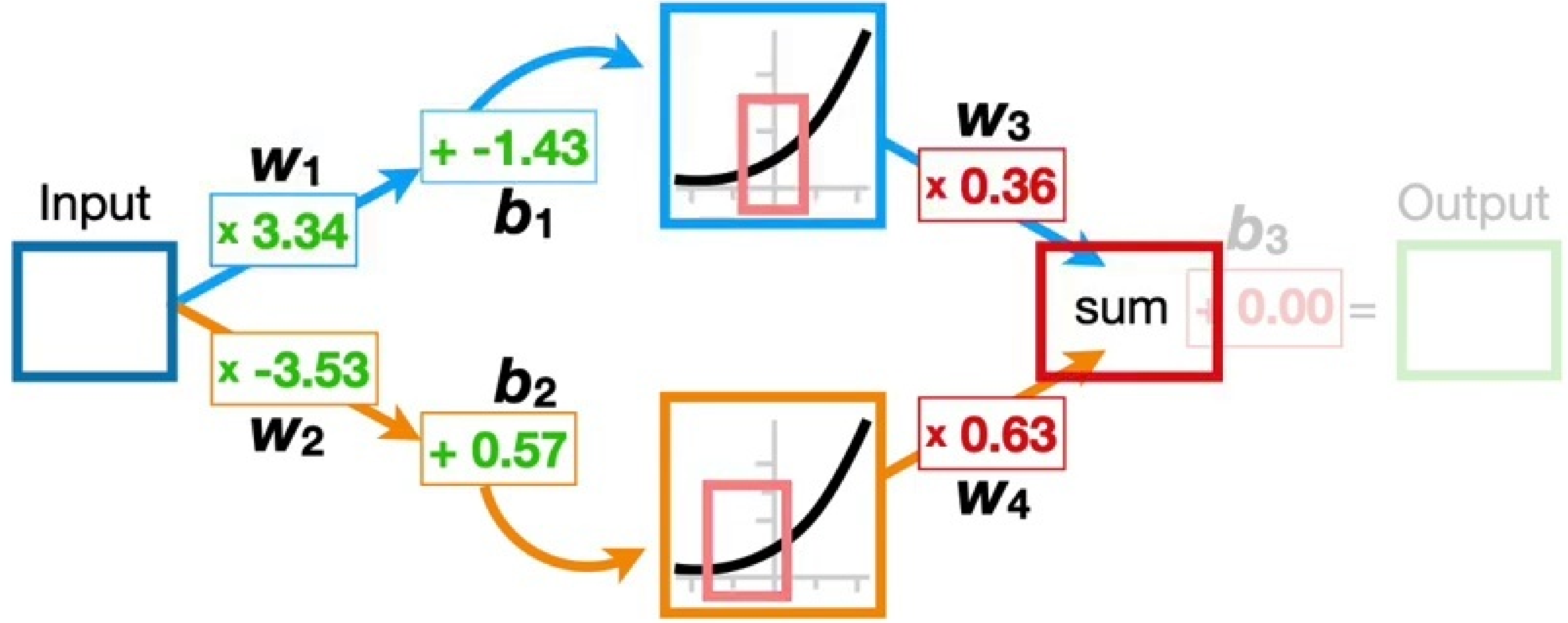
Now we add the **blue** and **orange** curves together…



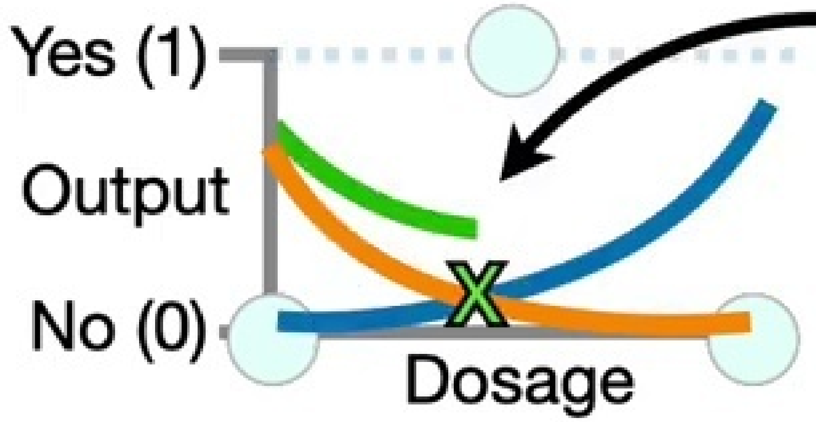


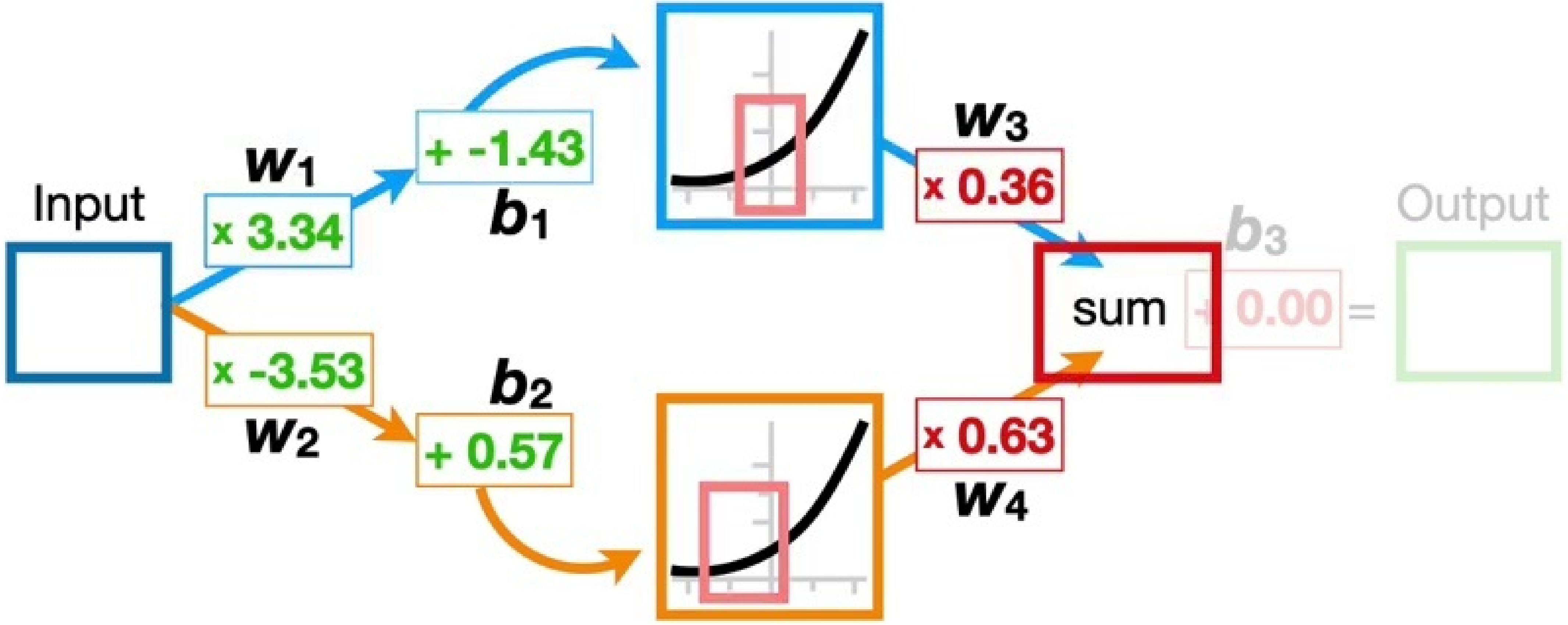
Now we add the **blue** and **orange** curves together…



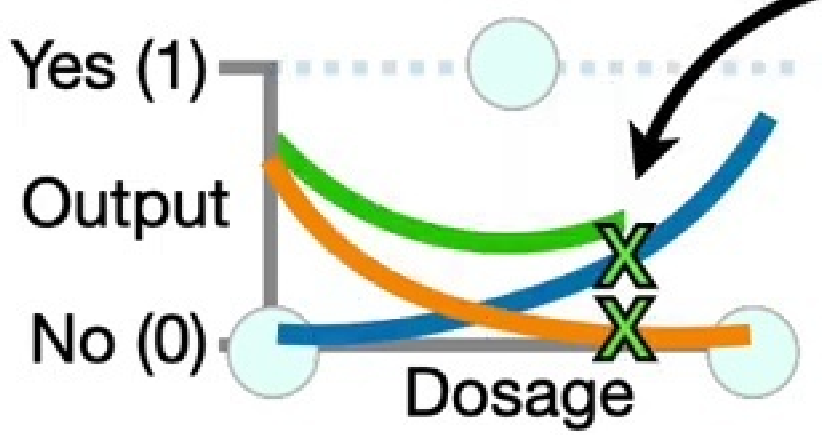


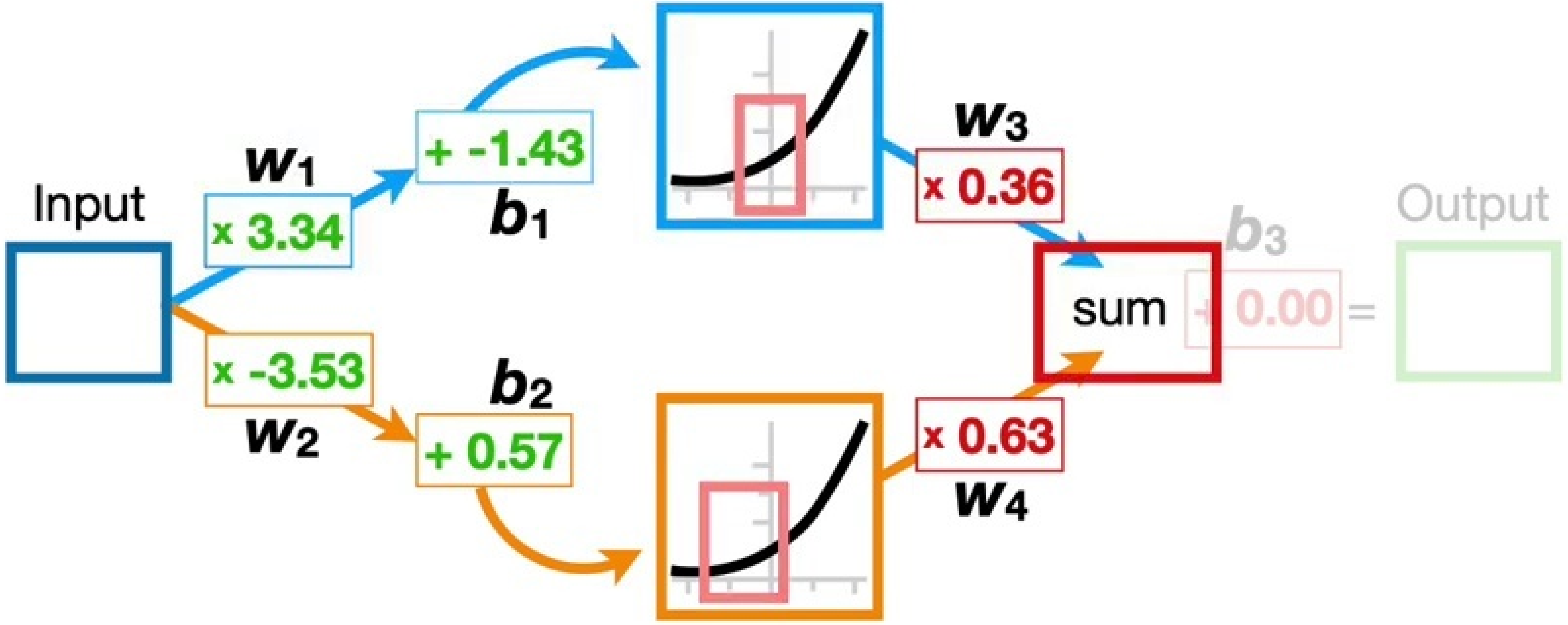
Now we add the **blue** and **orange** curves together…



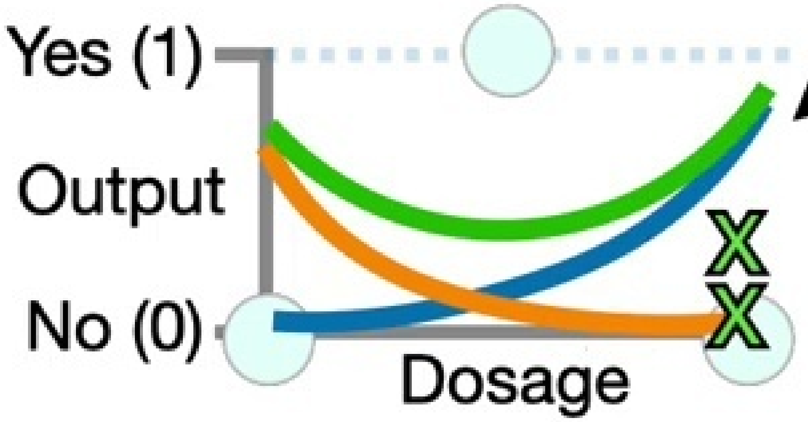


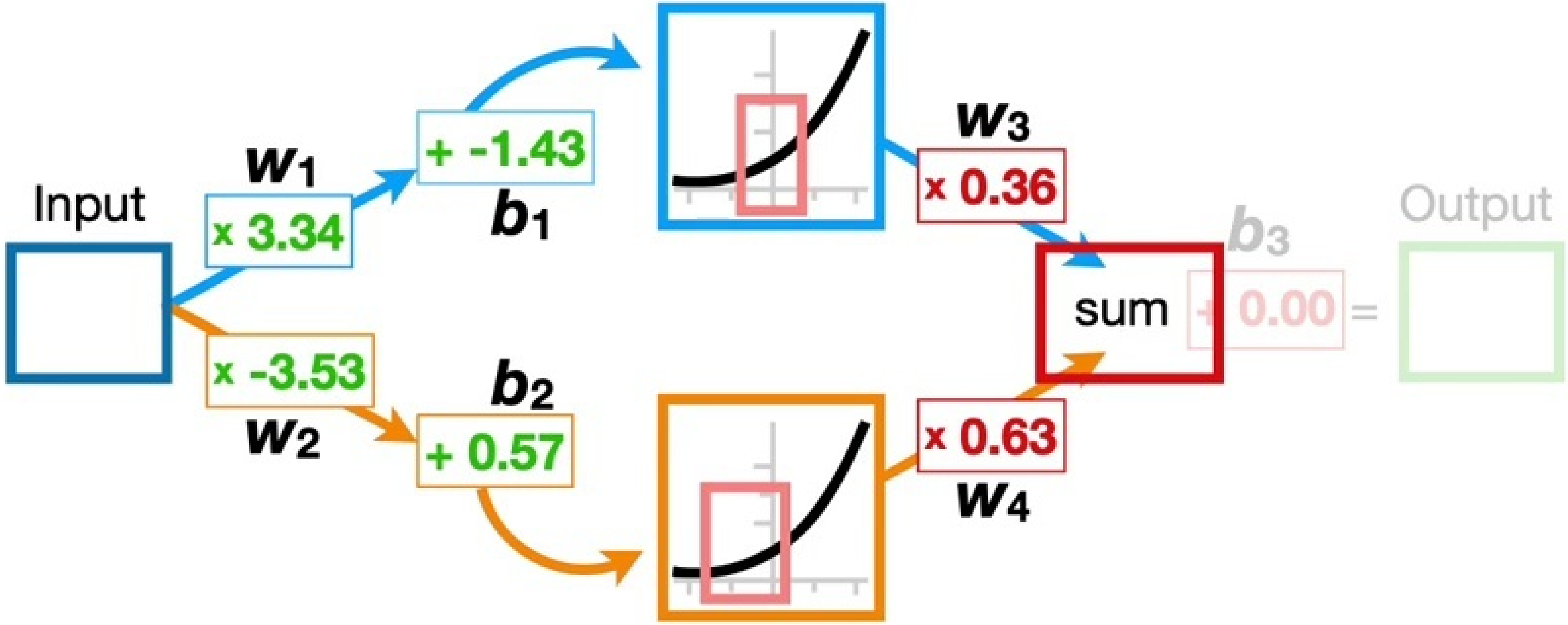
Now we add the **blue** and **orange** curves together…



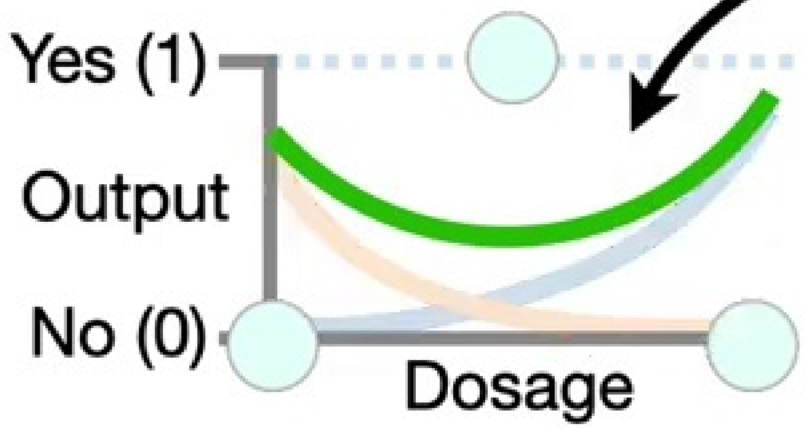


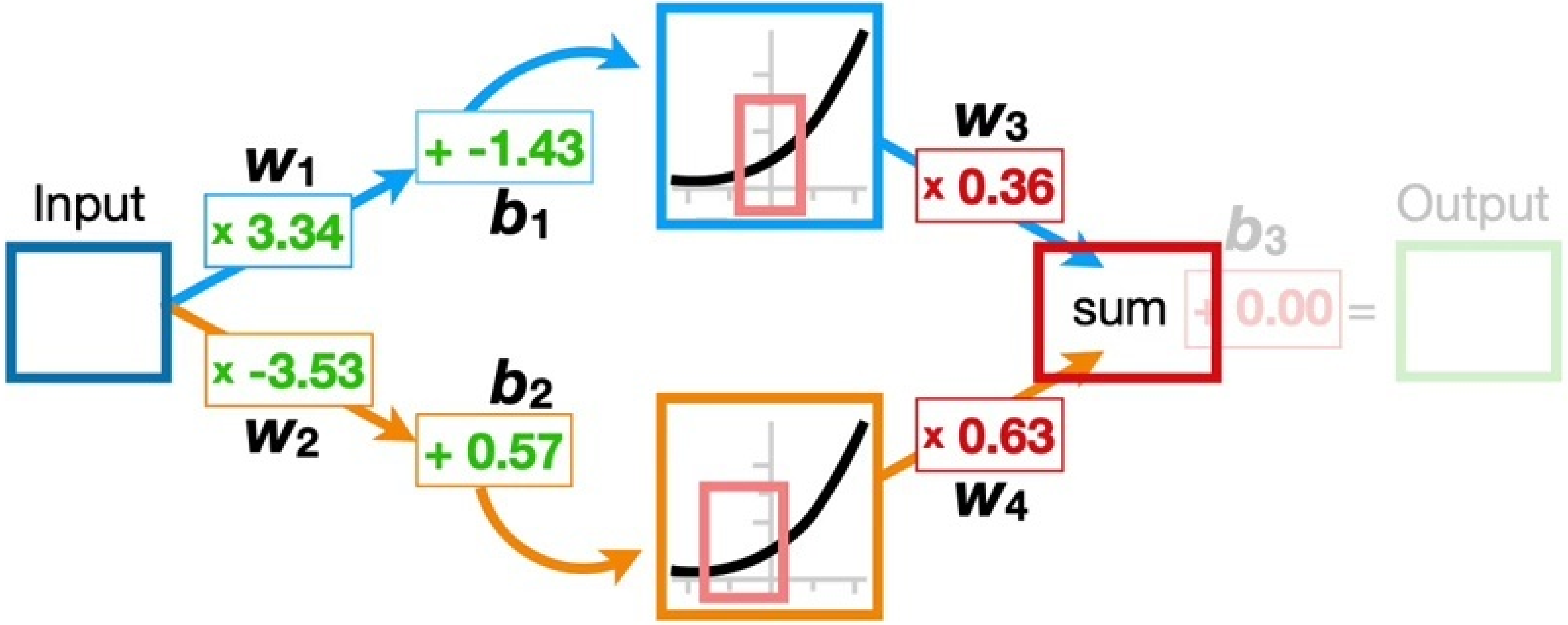
Now we add the **blue** and **orange** curves together…



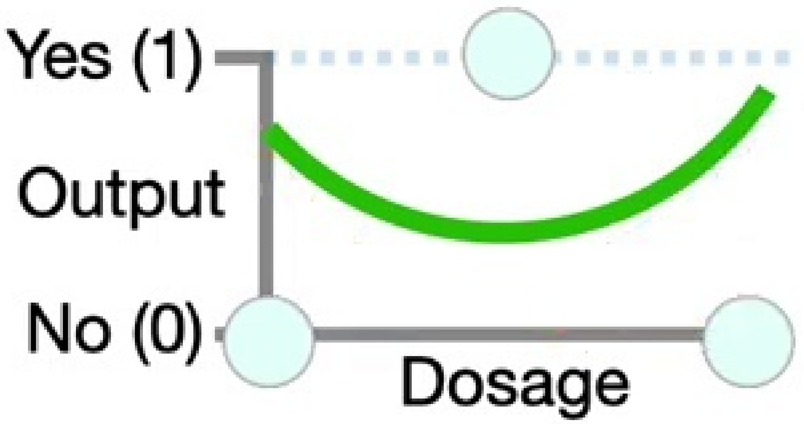


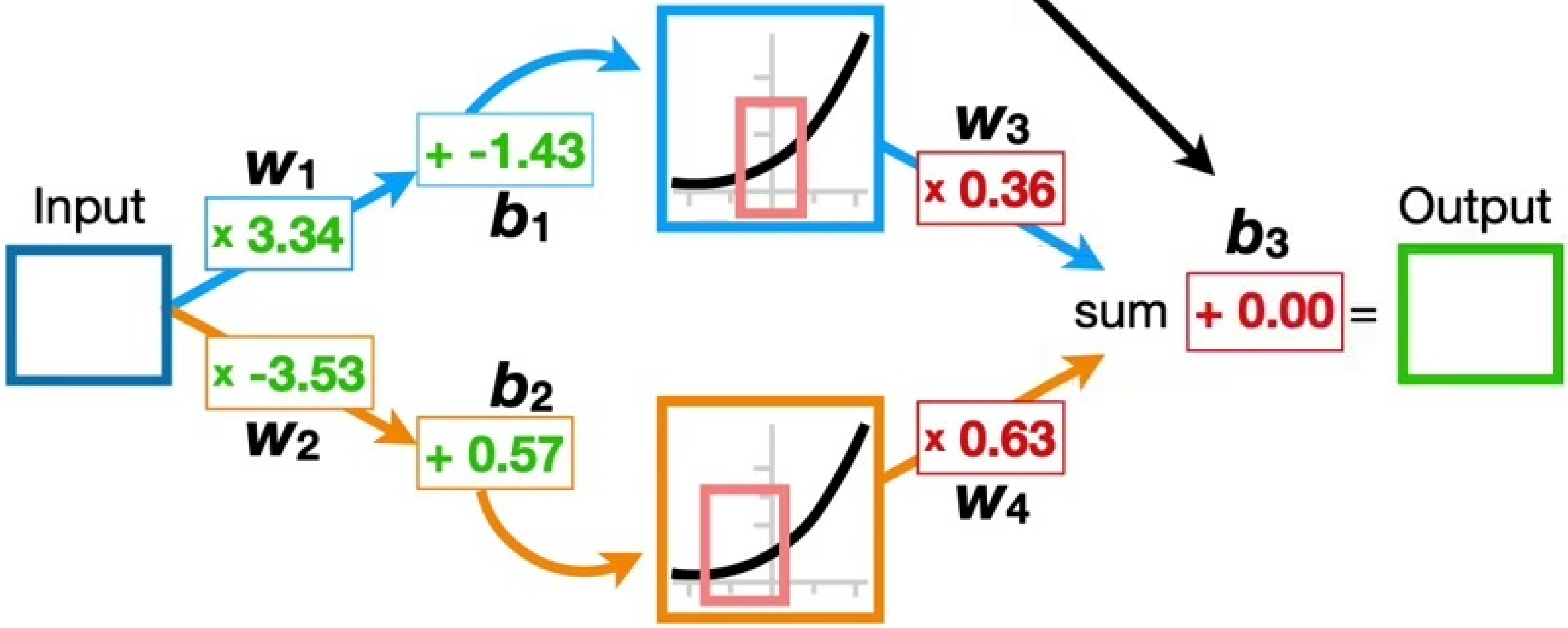
...and get this **green squiggle**.



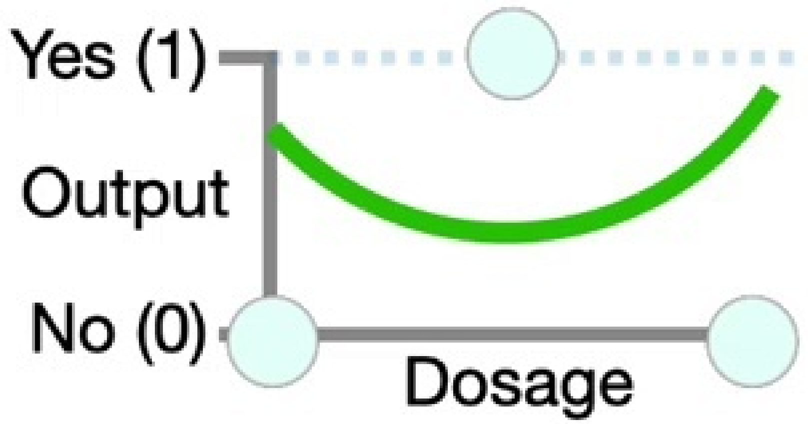


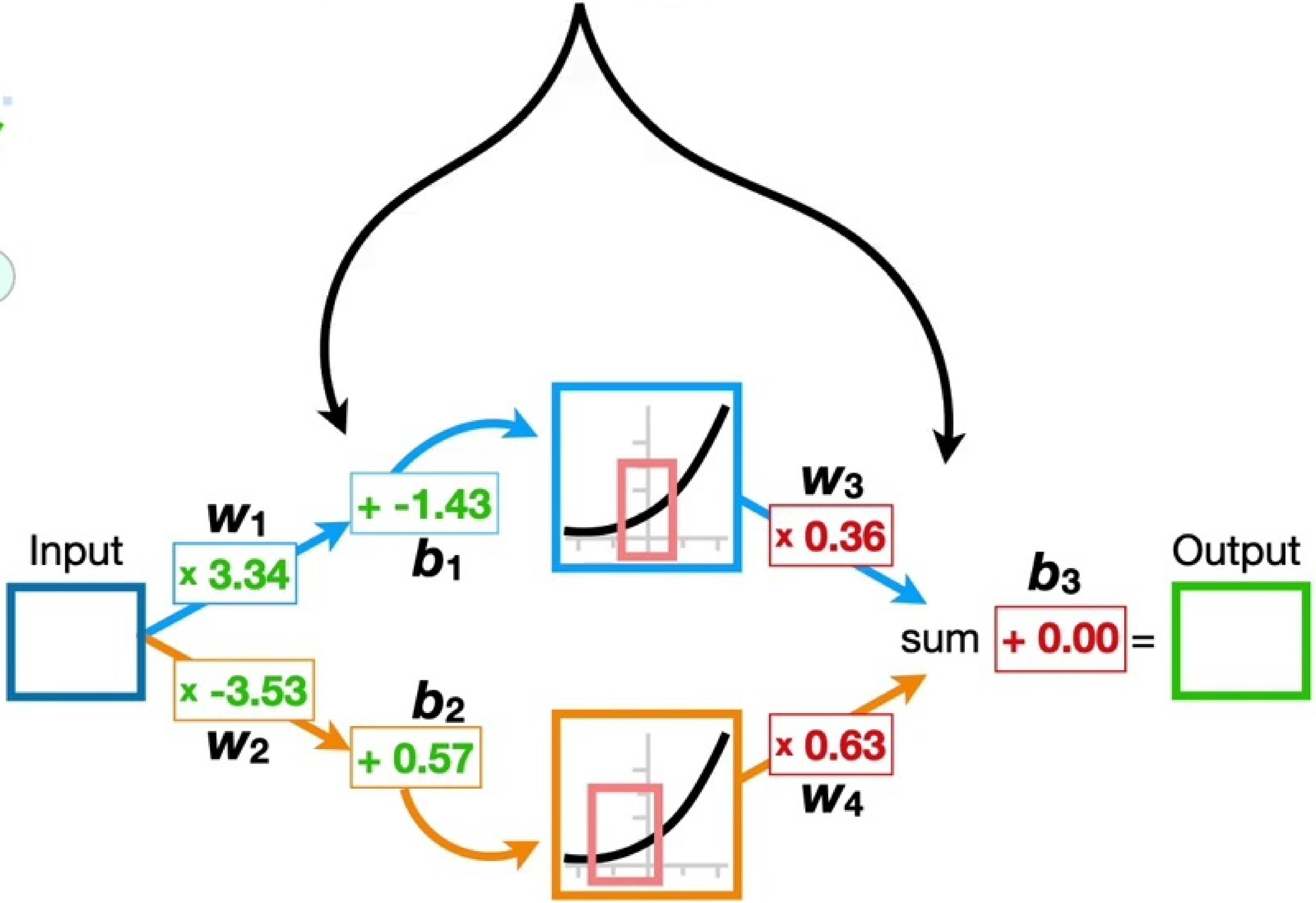
Lastly, since the initial value for b_3 is **0**, adding it to the y-axis values on the **green squiggle** does not change anything.



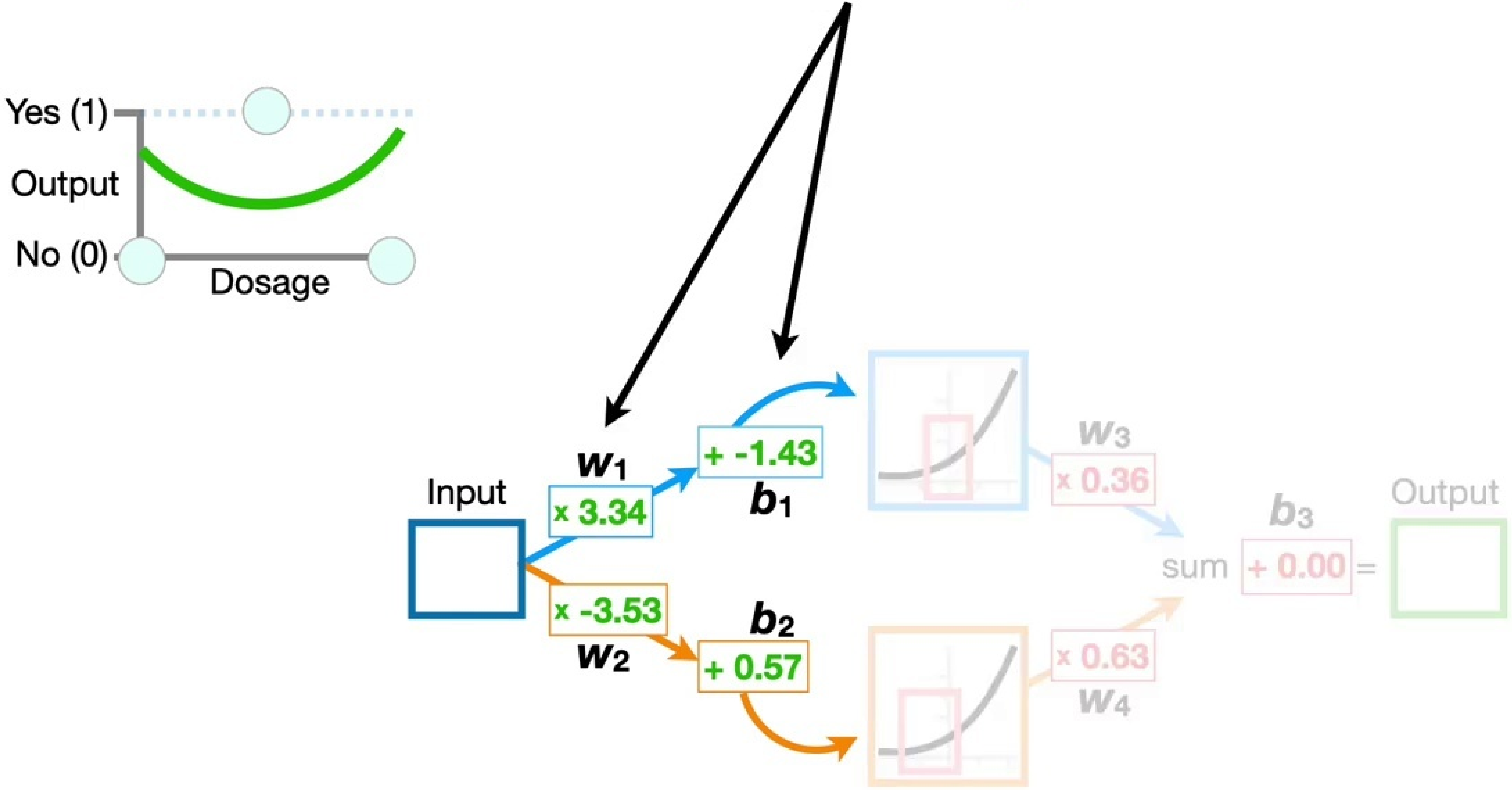


In other words, given the current parameters for this **Neural Network**…

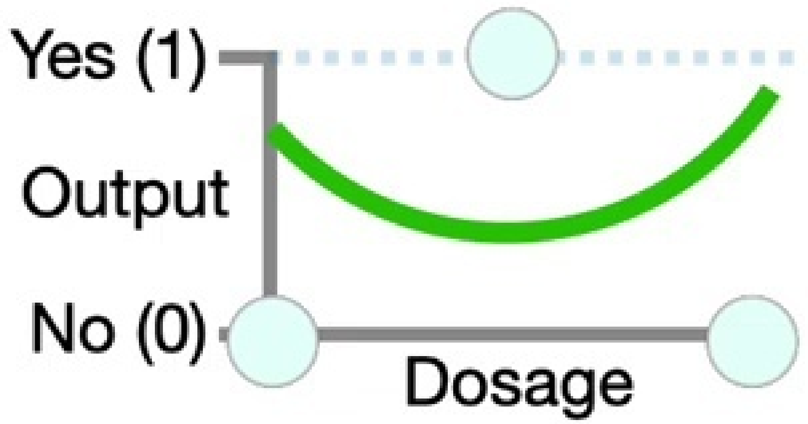


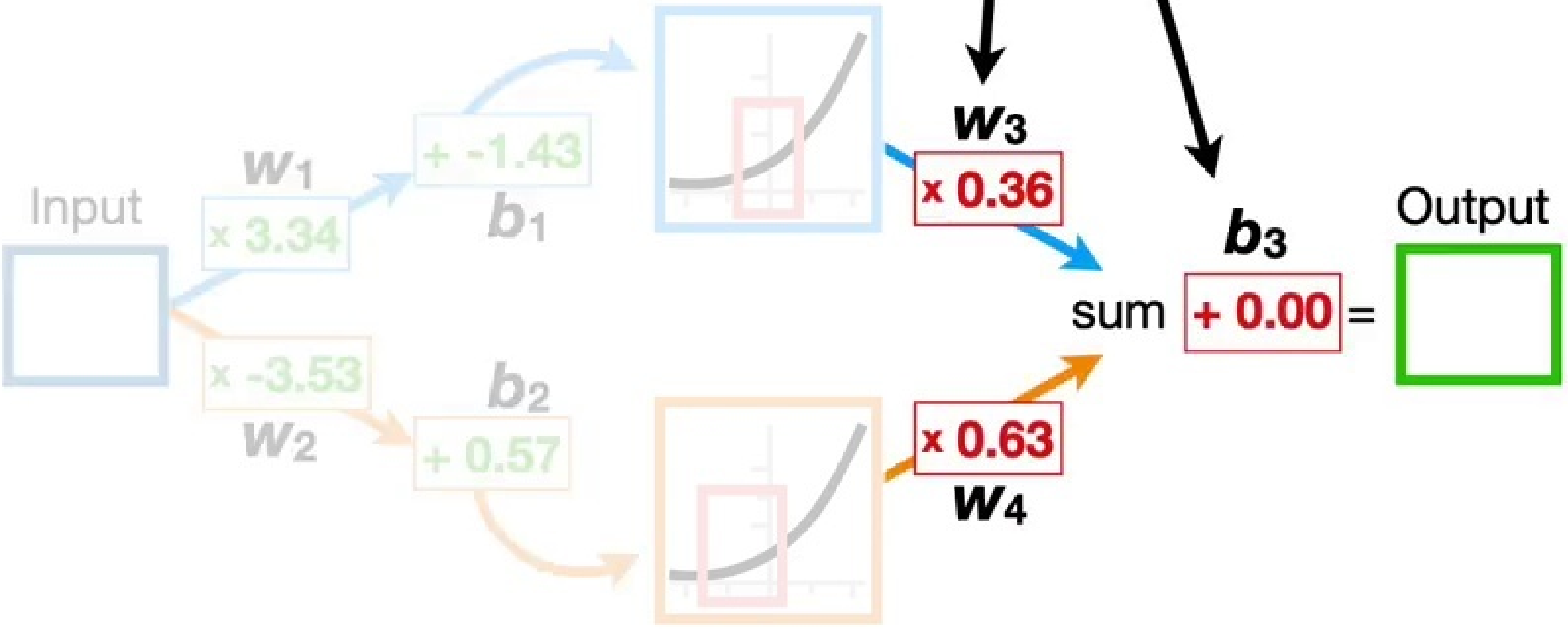


…some of which are **optimal**…

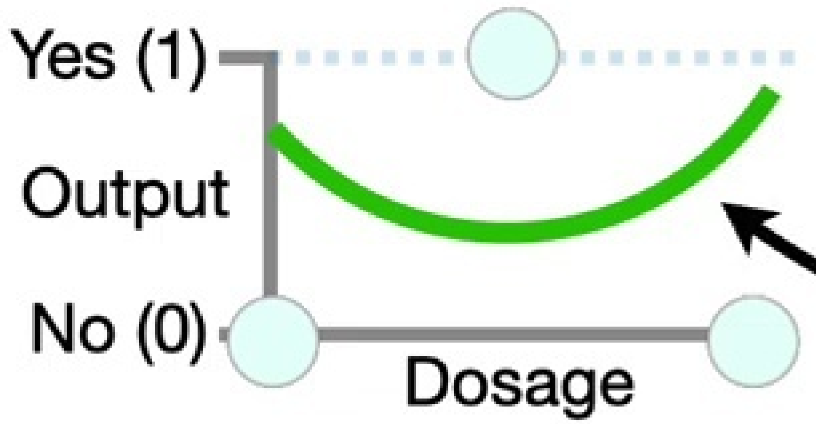


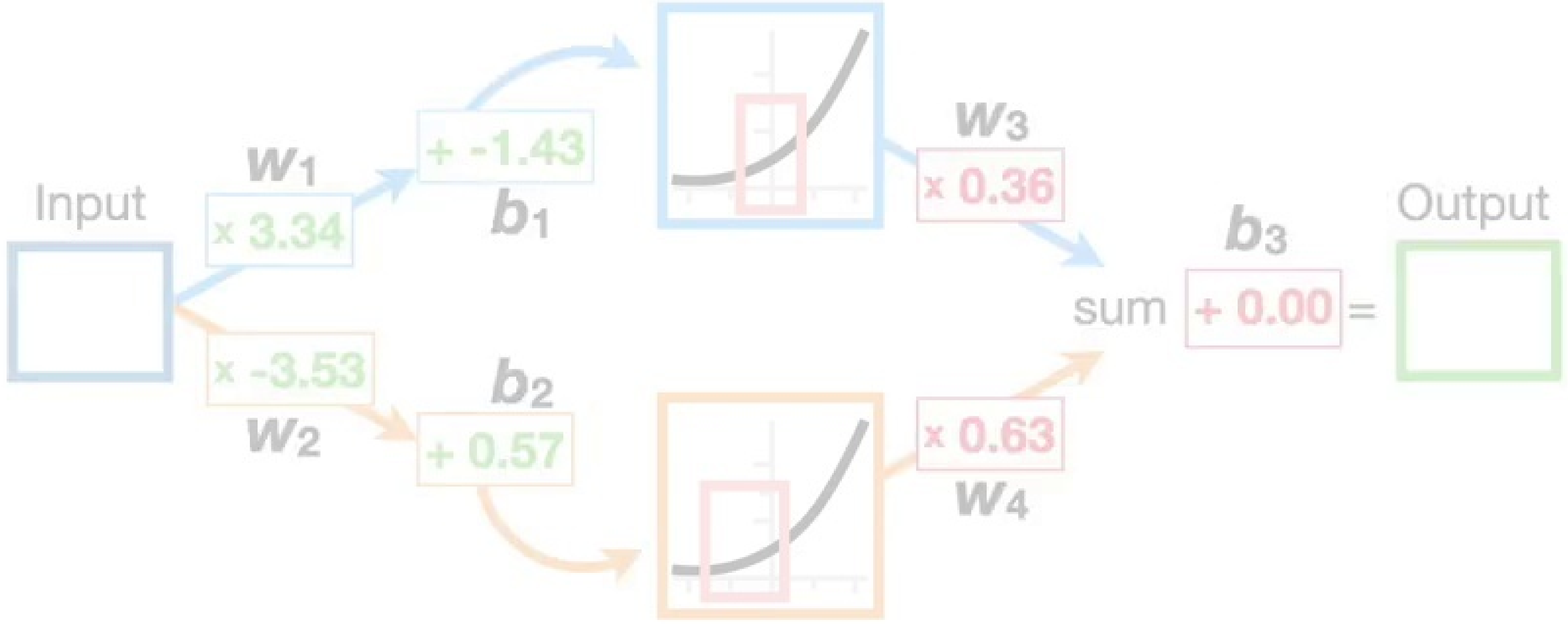
…and some of which are ***not* optimal**…



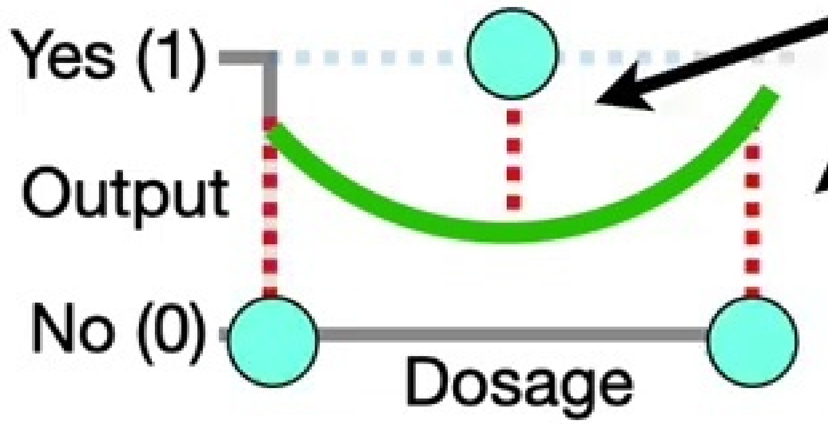


…we end up with this **green squiggle**.

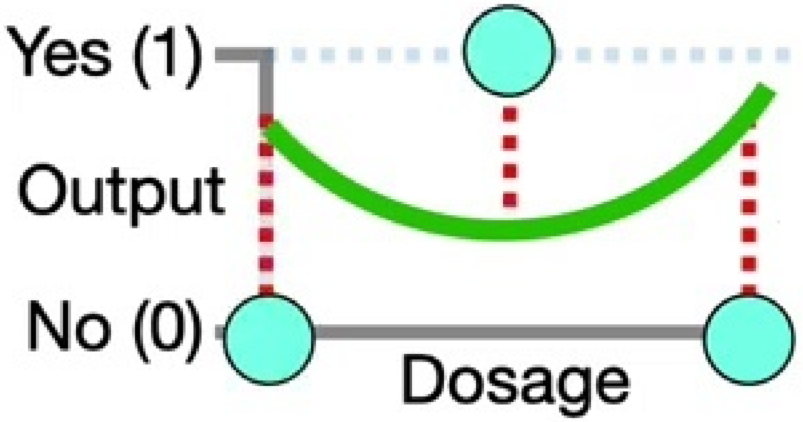




Now, just like before, we can quantify how well the **green squiggle** fits the data by calculating the **Sum of the Squared Residuals (SSR)**…

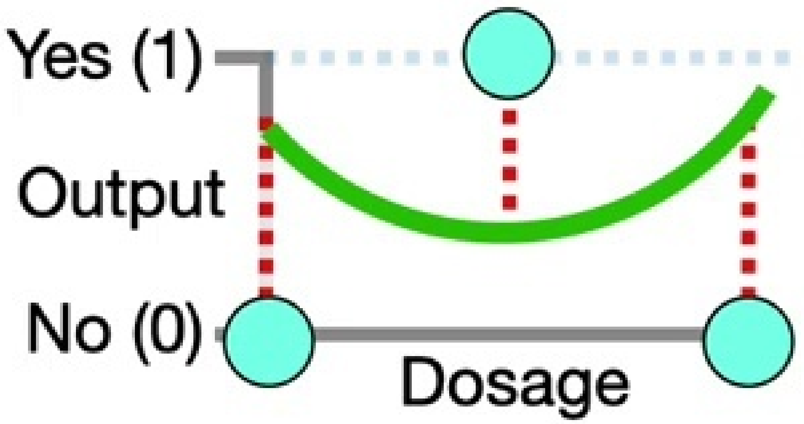


…and we get the
SSR = 1.4.

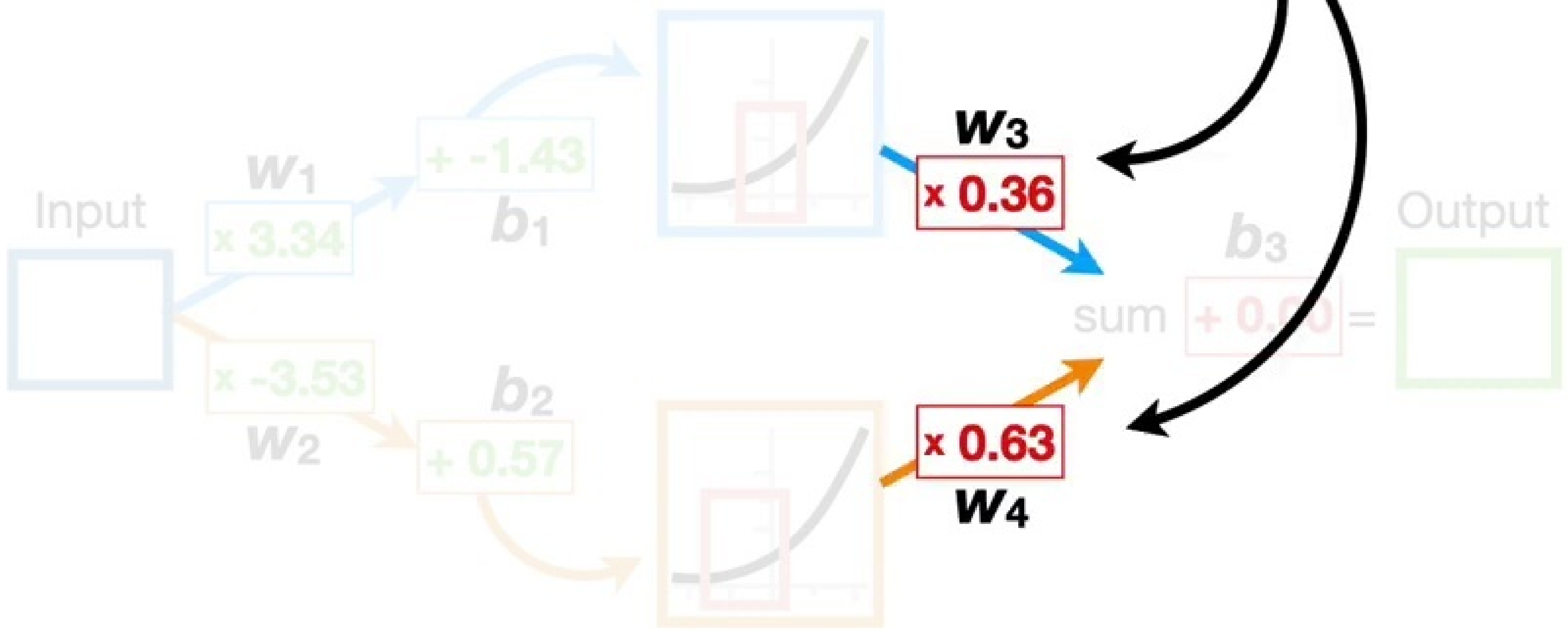


$$SSR = (0 - 0.72)^2 + (1 - 0.46)^2 + (0 - 0.77)^2 \boxed{= 1.4}$$

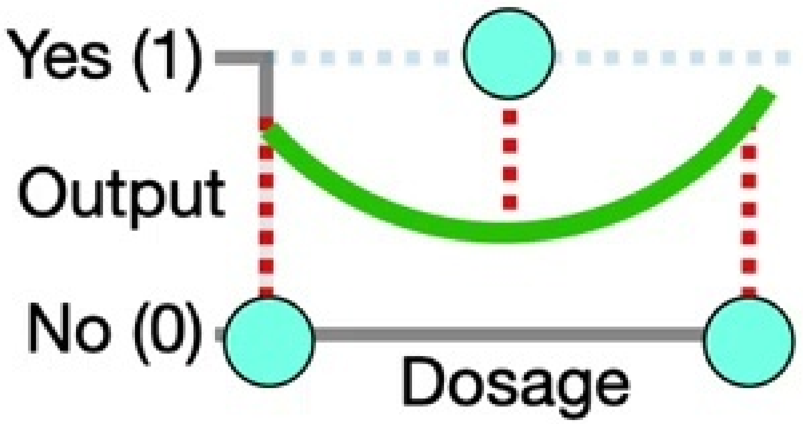
Now, even though we have not yet optimized w_3 and w_4…



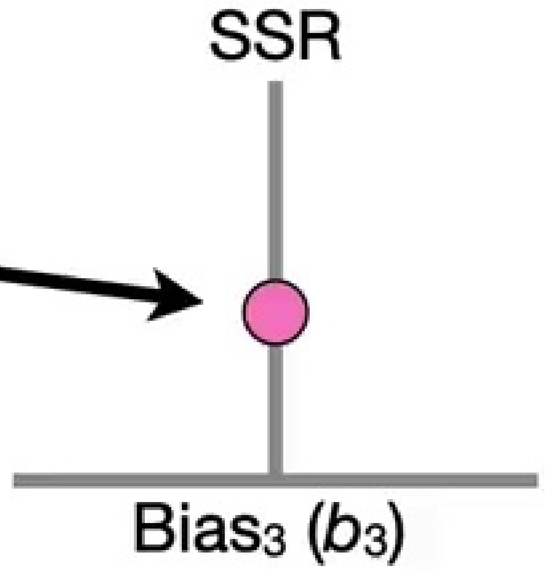
$$SSR = (0 - 0.72)^2 + (1 - 0.46)^2 + (0 - 0.77)^2 = 1.4$$



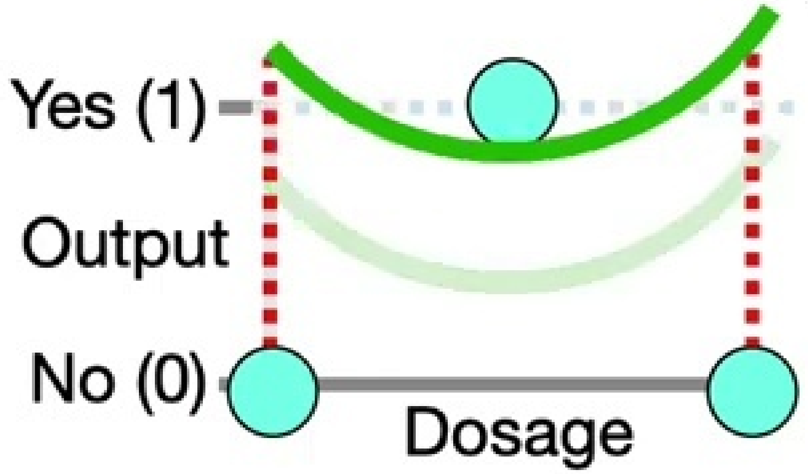
...we can still plot the **SSR** with respect to $\boldsymbol{b_3}$.



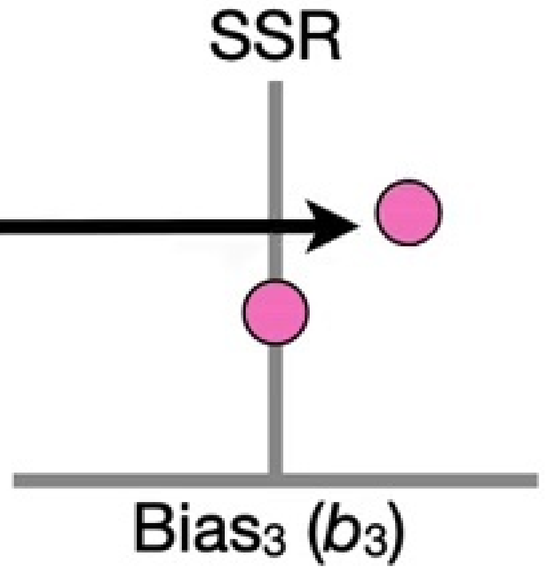
$$\text{SSR} = (0 - 0.72)^2 + (1 - 0.46)^2 + (0 - 0.77)^2 = 1.4$$

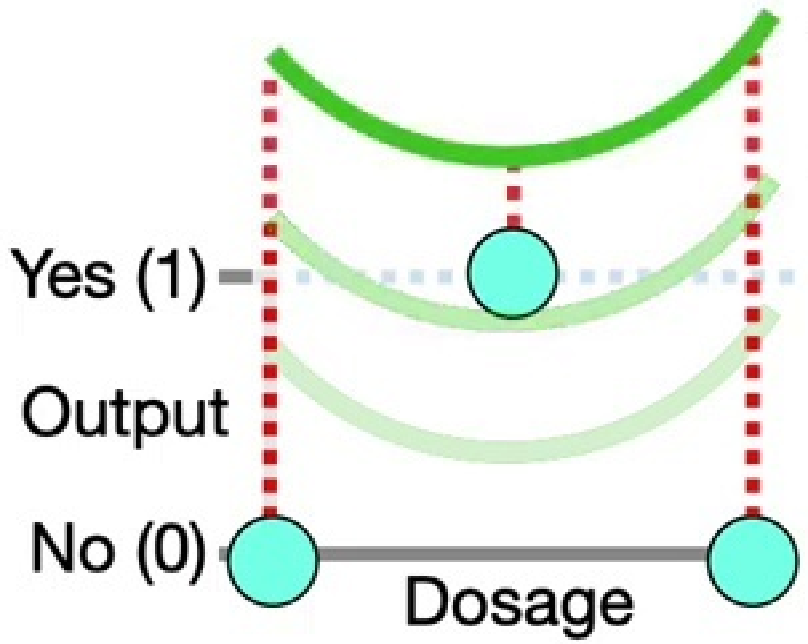


And just like before, if we change $\boldsymbol{b_3}$ then we will change the **SSR**…

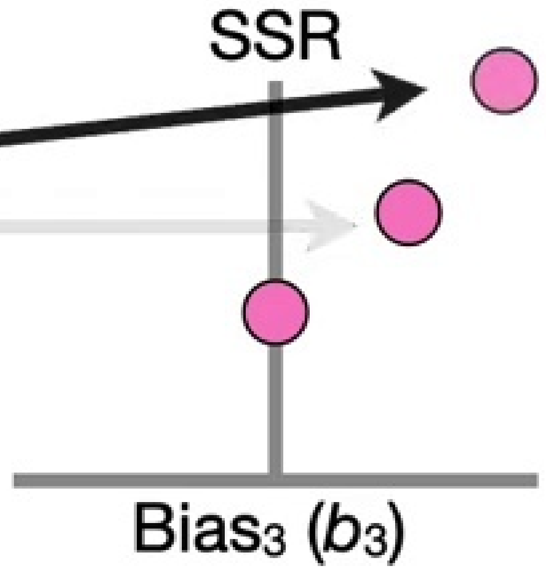


$$\text{SSR} = (0 - 1.22)^2 + (1 - 0.96)^2 + (0 - 1.27)^2 = 3.1$$

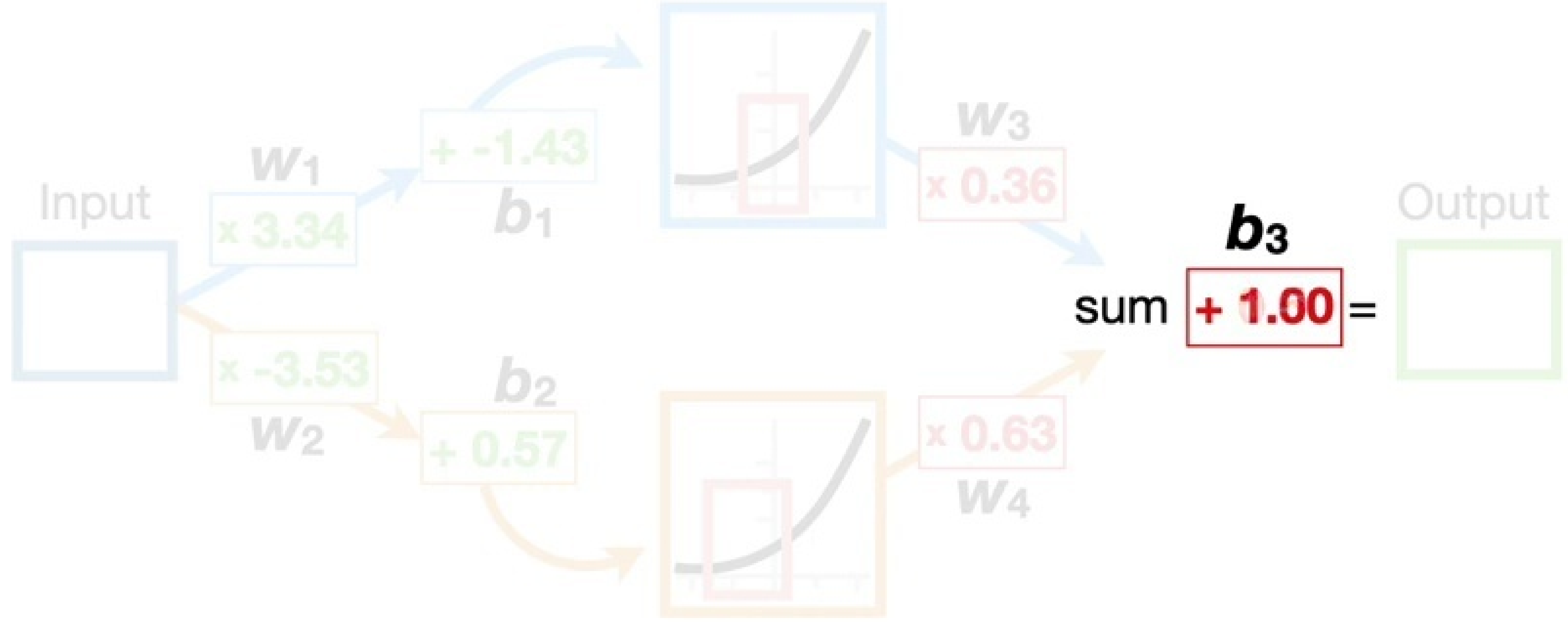




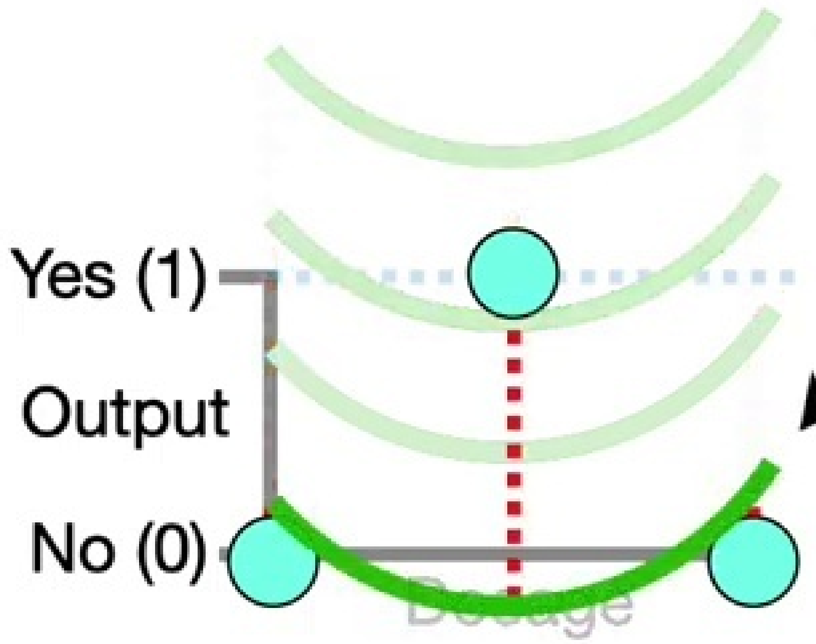
And just like before, if we change $\boldsymbol{b_3}$ then we will change the **SSR**…

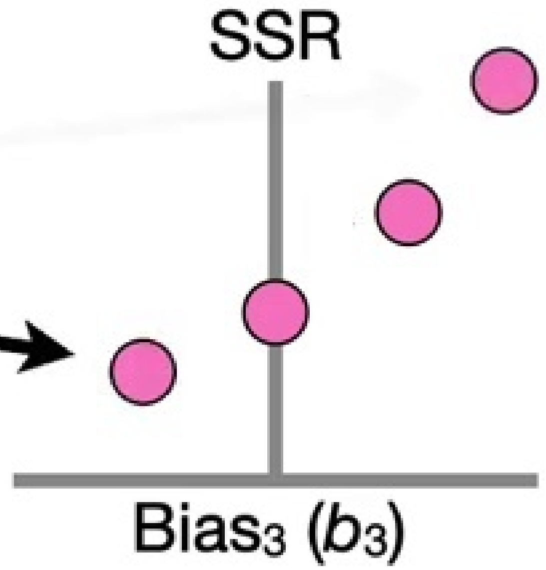


$$\text{SSR} = (0 - 1.72)^2 + (1 - 1.46)^2 + (0 - 1.97)^2 = 6.3$$

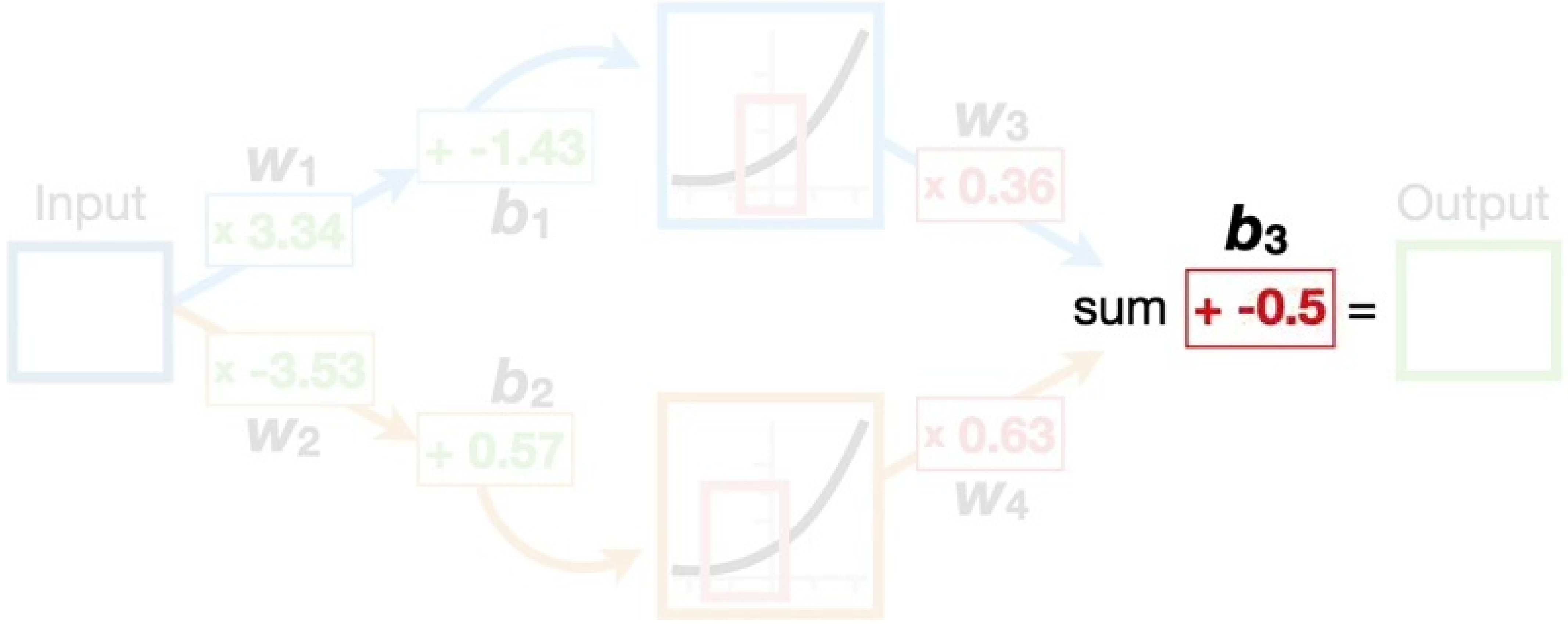


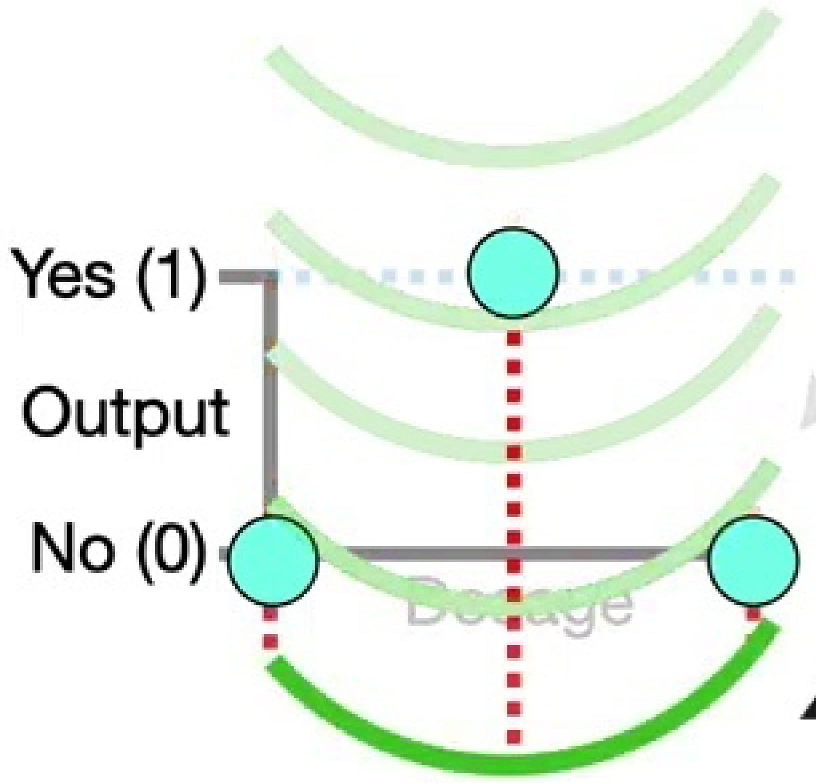
And just like before, if we change b_3 then we will change the **SSR**…



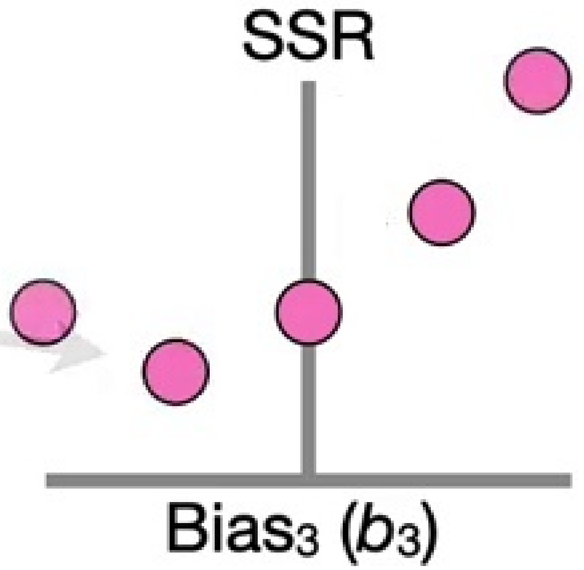


$$
\begin{aligned}
\text{SSR} = &(0 - 0.22)^2 \\
&+ (1 - -0.04)^2 \\
&+ (0 - 0.27)^2 = 1.2
\end{aligned}
$$

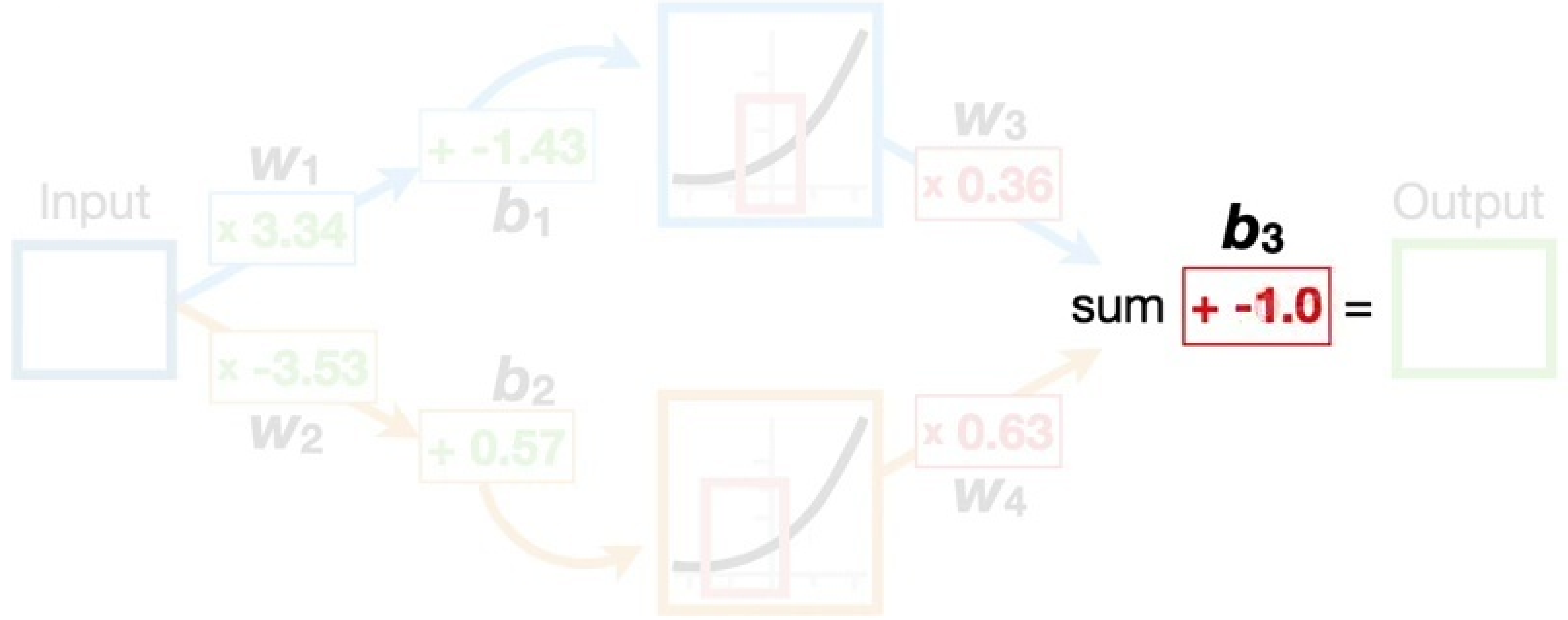




And just like before, if we change $\boldsymbol{b_3}$ then we will change the **SSR**…

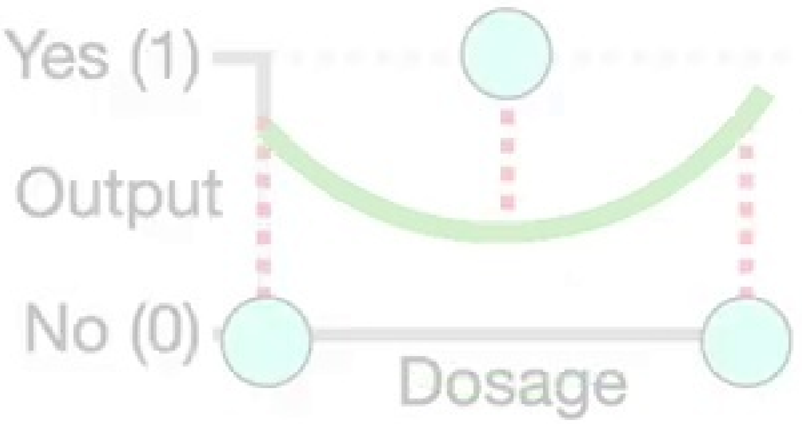


$$
\begin{aligned}
SSR = & (0 - -0.28)^2 \\
& + (1 - -0.54)^2 \\
& + (0 - -0.22)^2 = 2.5
\end{aligned}
$$



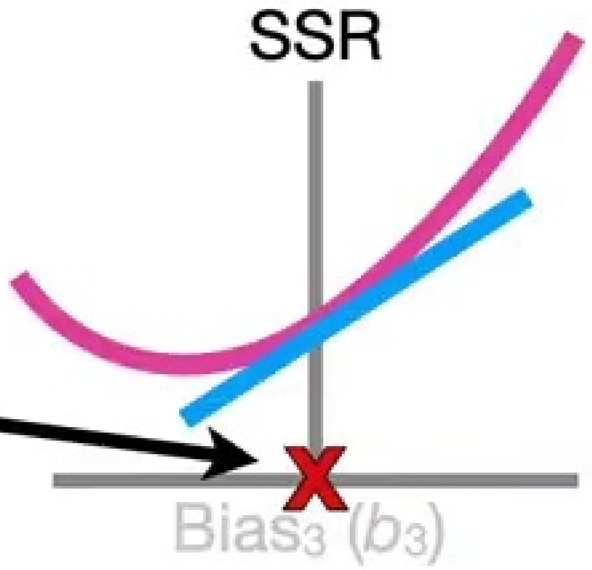
$$\frac{d\ SSR}{d\ b_3}$$

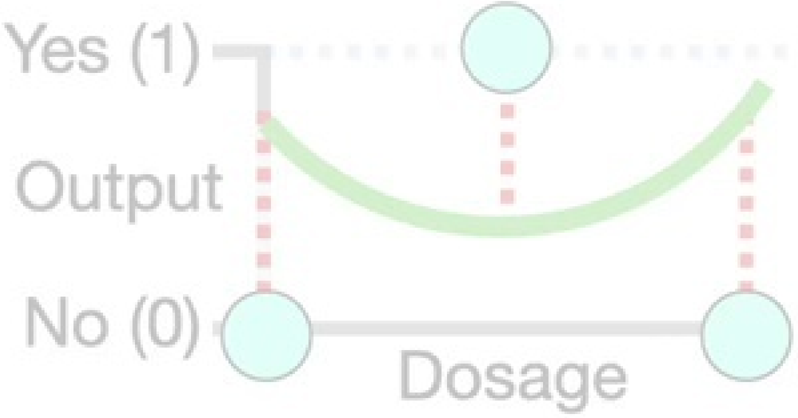
…and that means, just like before, we can optimize b_3 by finding the derivative of the **SSR** with respect to b_3…

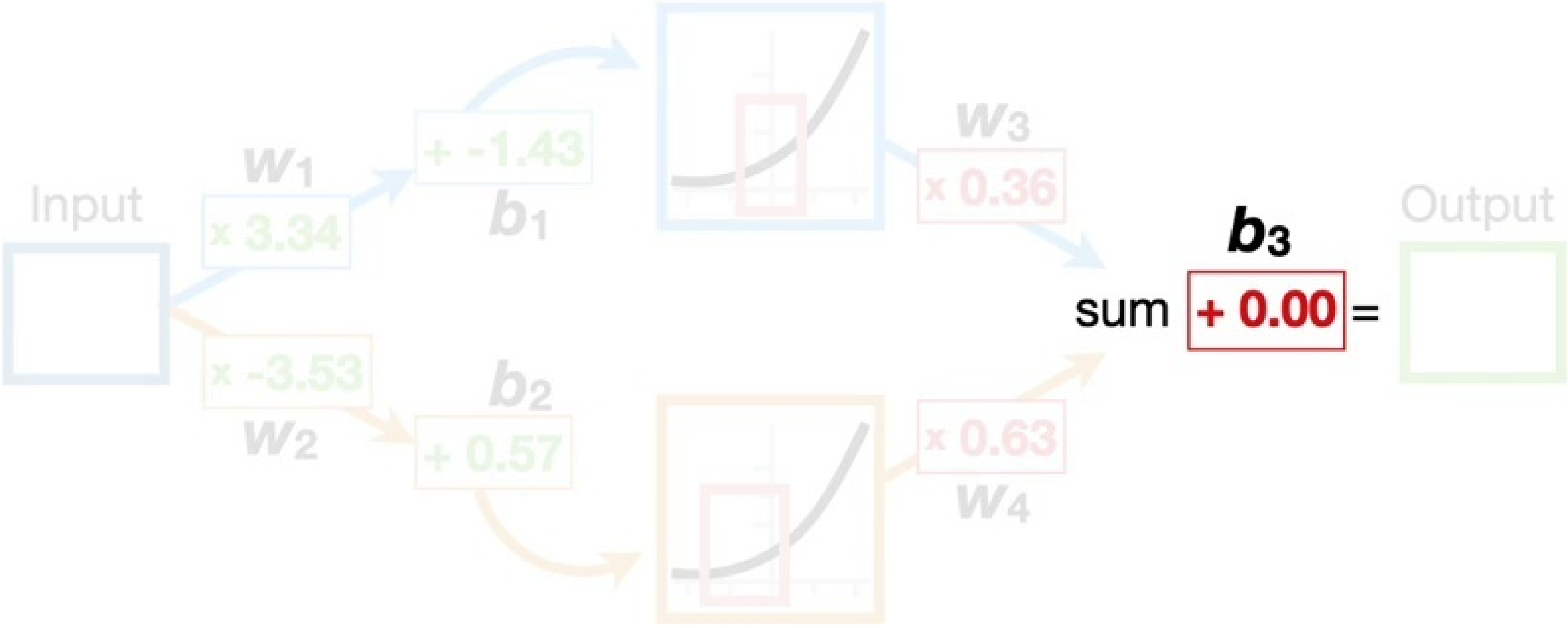


$$\frac{d\ SSR}{d\ b_3}$$

…and plugging the derivative into the **Gradient Descent** algorithm to find the optimal value for $\boldsymbol{b_3}$.

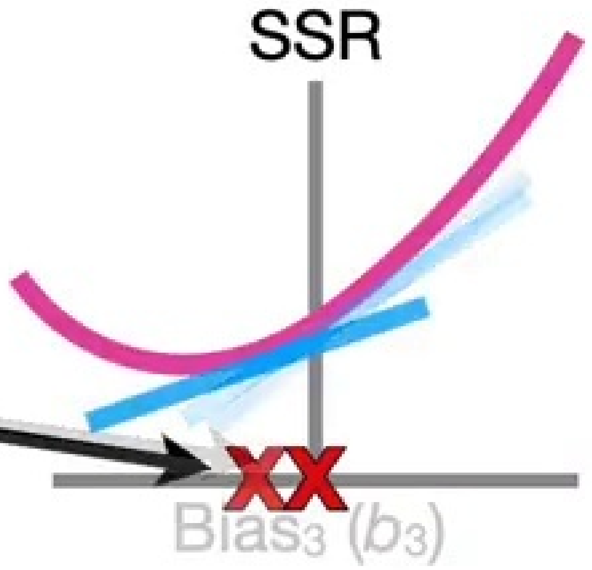


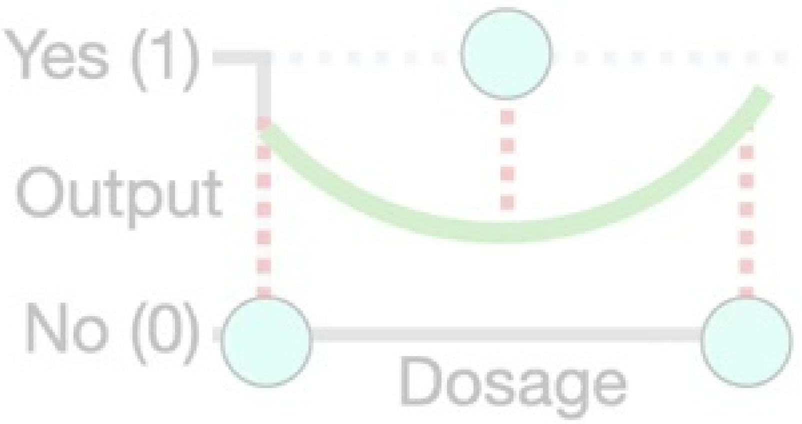


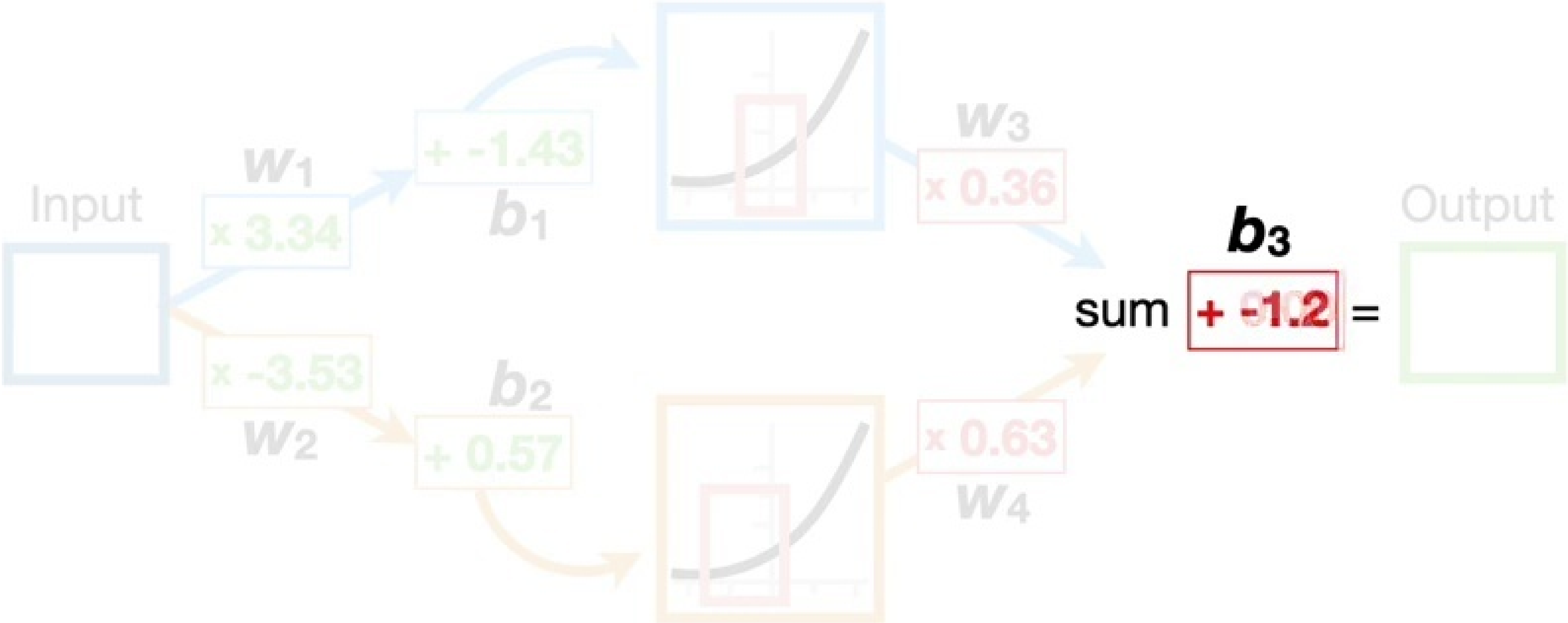


$$\frac{d\ SSR}{d\ b_3}$$

…and plugging the derivative into the **Gradient Descent** algorithm to find the optimal value for b_3.

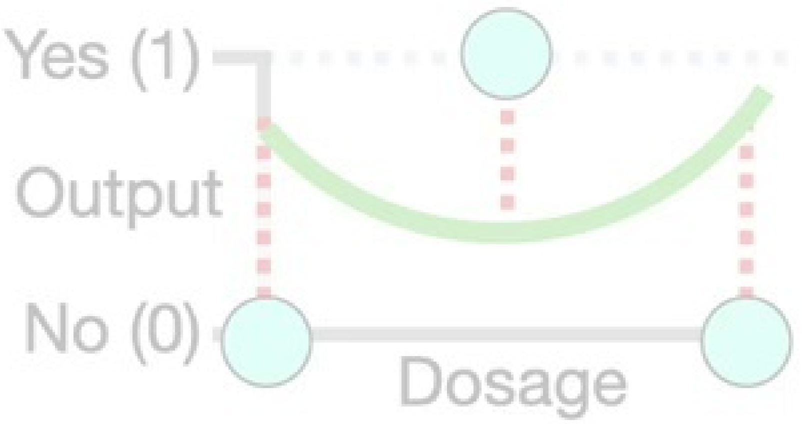


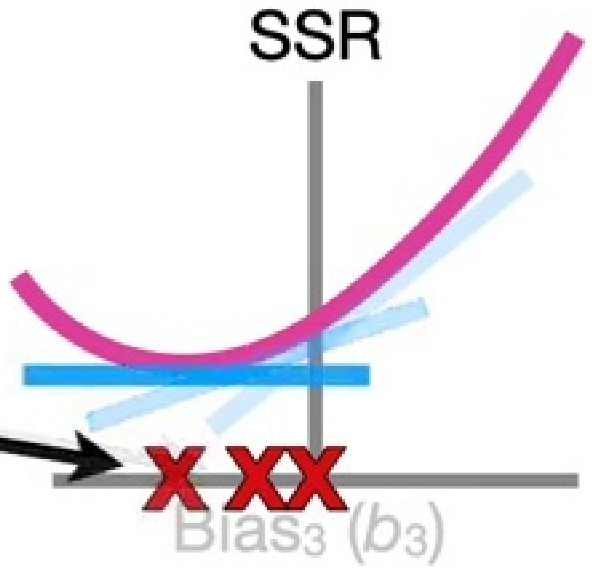


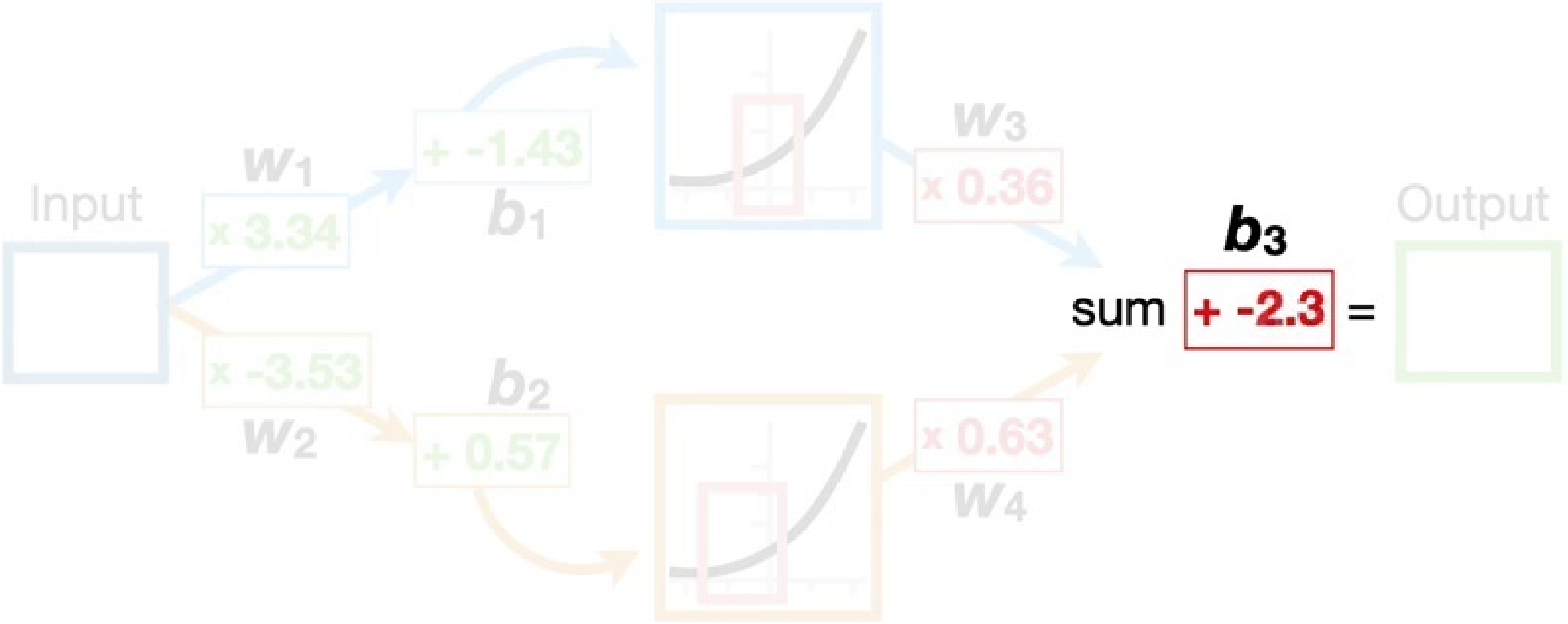


$$\frac{d\ SSR}{d\ b_3}$$

...and plugging the derivative into the **Gradient Descent** algorithm to find the optimal value for b_3.



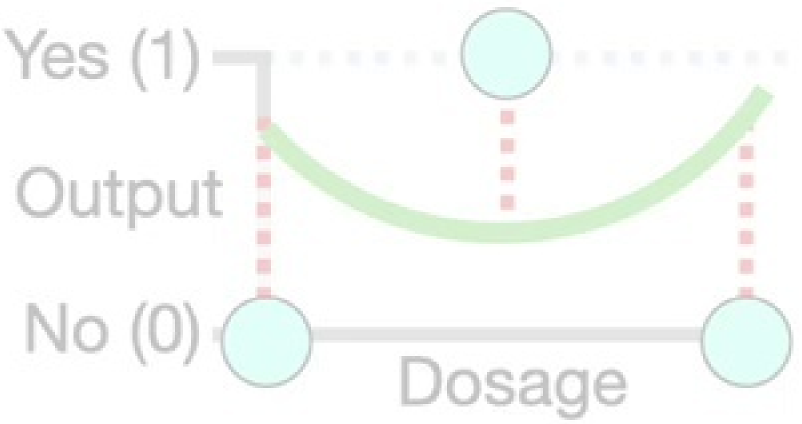


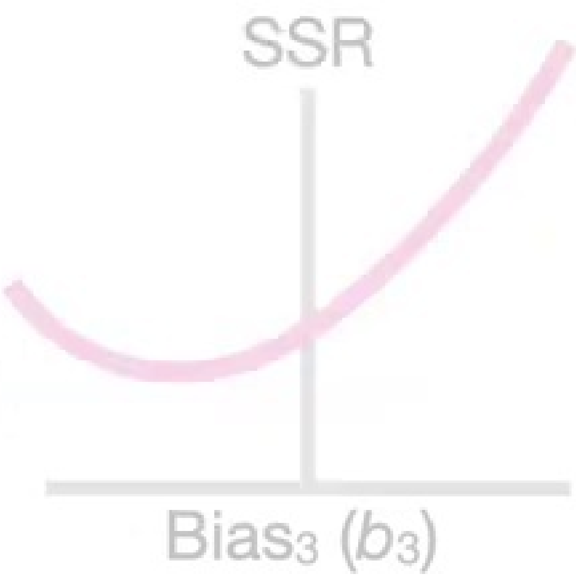


$$\frac{d\ SSR}{d\ b_3}$$

And, just like before, because the **Predicted** values in the **SSR**…

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

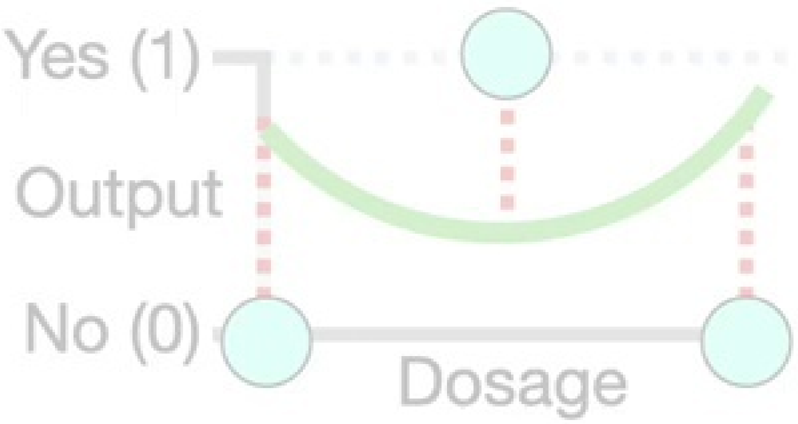


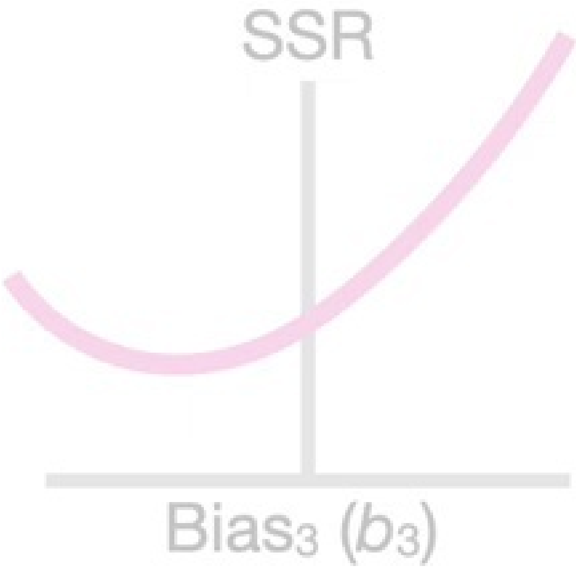


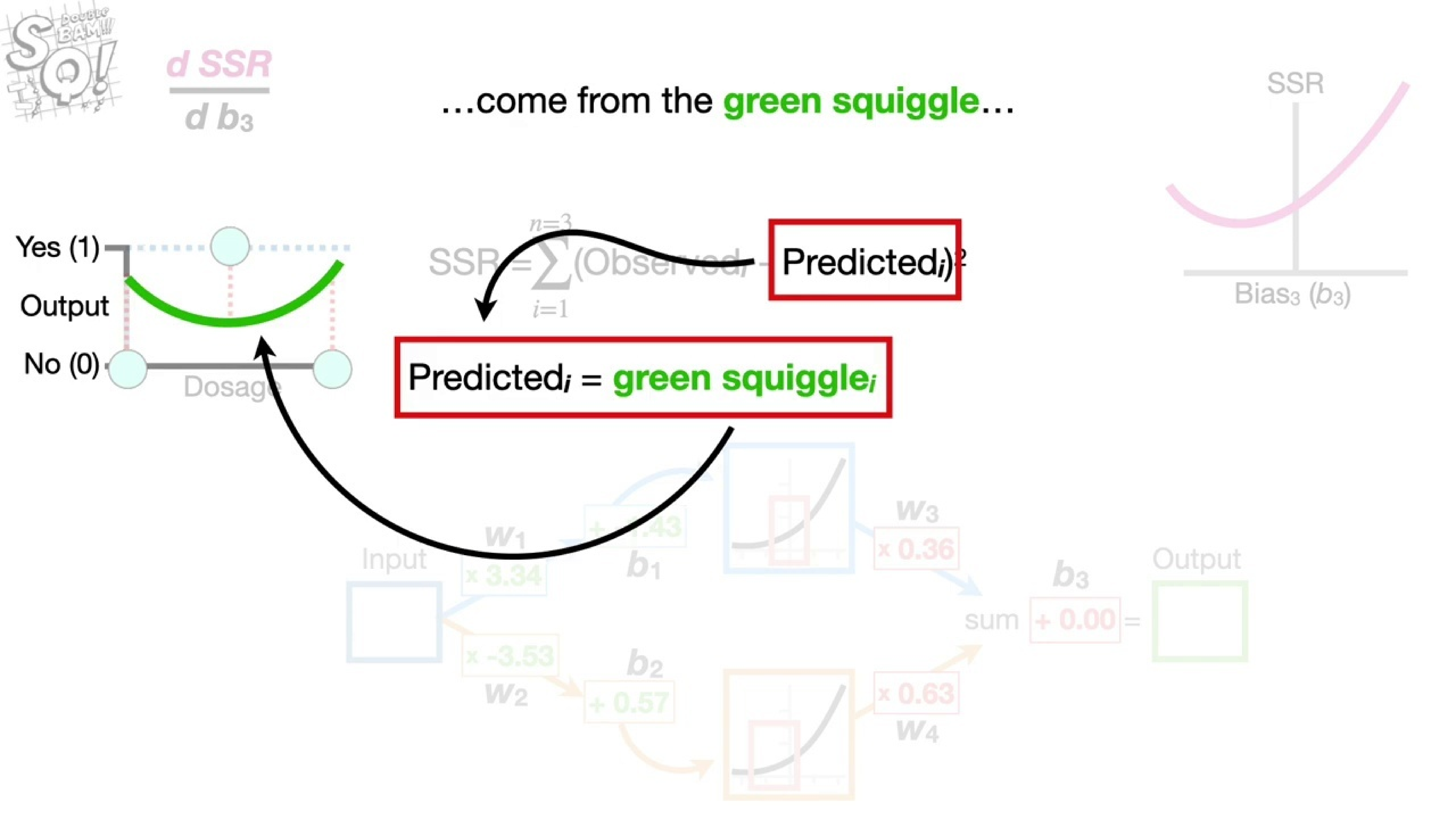
$$\frac{d\ \text{SSR}}{d\ b_3}$$

And, just like before, because the **Predicted** values in the **SSR**…

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



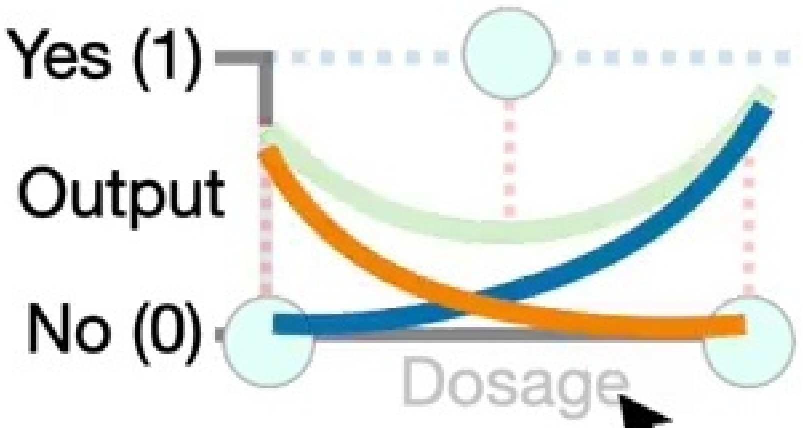




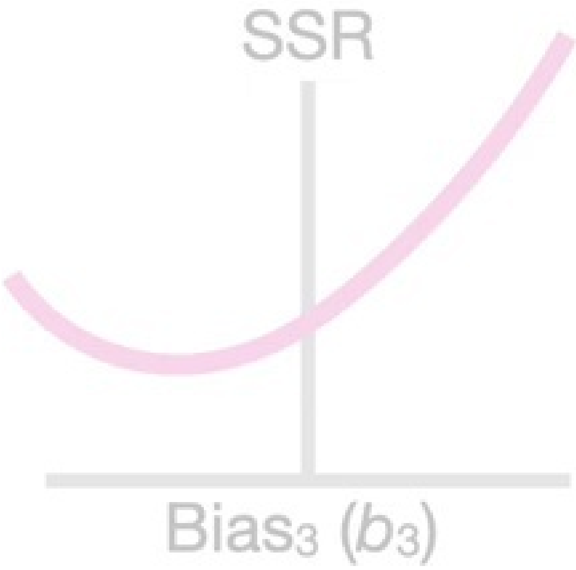
d SSR / d b3
…come from the green squiggle…
SSR
Bias3 (b3)
Yes (1)
Output
No (0)
Dosage
SSR = ∑ (Observedi – Predictedi)², i=1 to n=3
Predictedi = green squiggle_i
Input
w1
x 3.34
+ -1.43
b1
w3
x 0.36
x -3.53
w2
b2
+ 0.57
x 0.63
w4
b3
sum + 0.00 =
Output

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

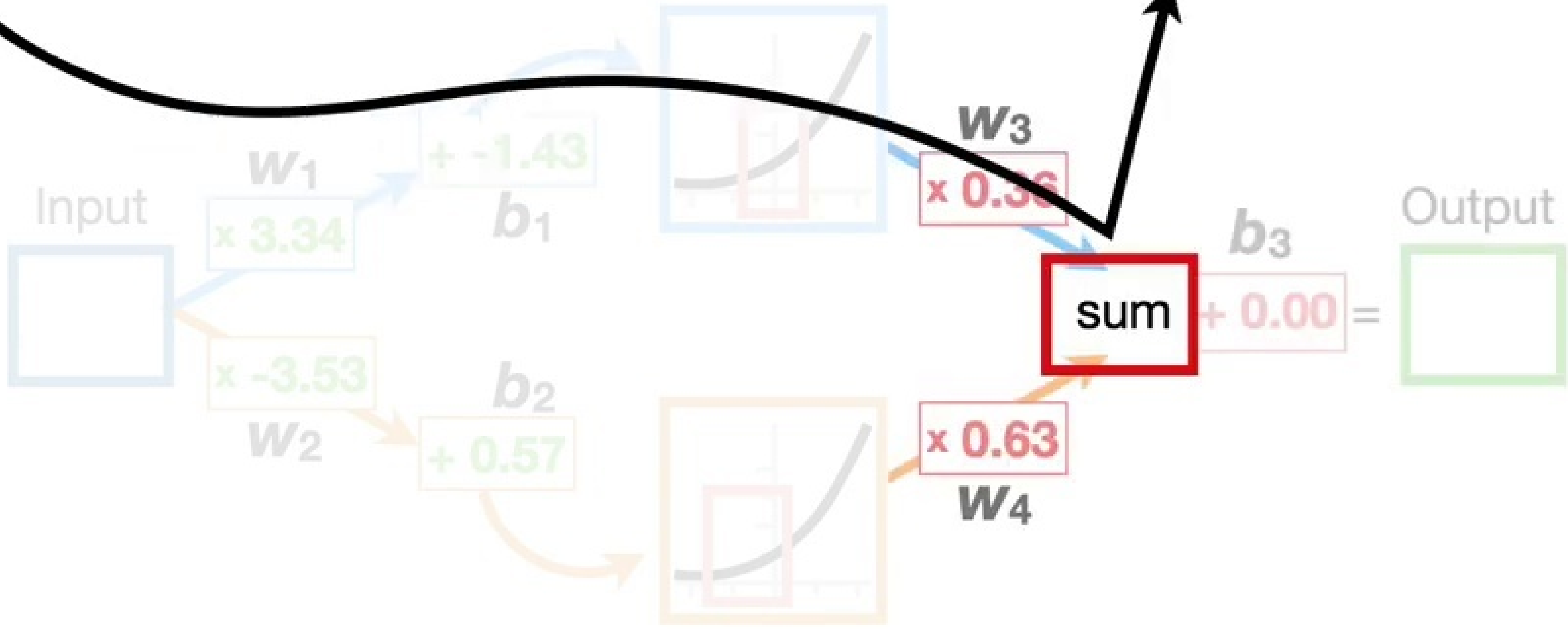
...and the **green squiggle** is the **sum** of the **blue** and **orange** curves...



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

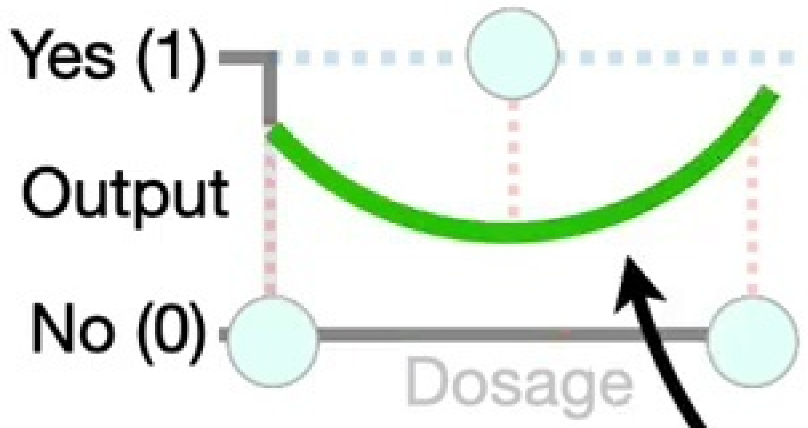


$$\text{Predicted}_i = \boxed{\text{green squiggle}_i = \text{blue} + \text{orange}} + b_3$$

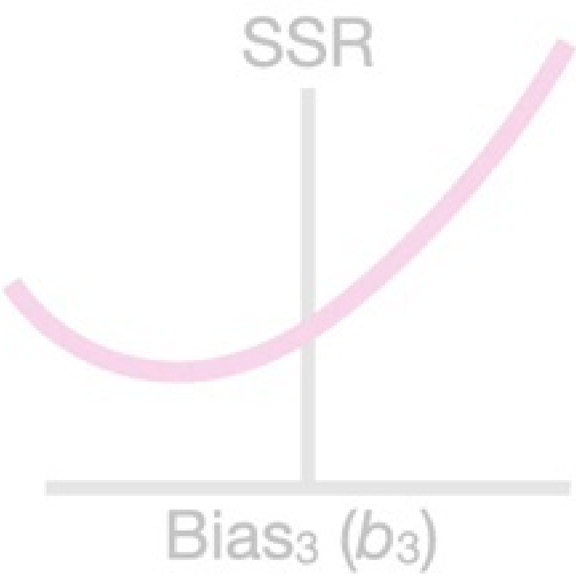


$$\frac{d\ SSR}{d\ b_3}$$

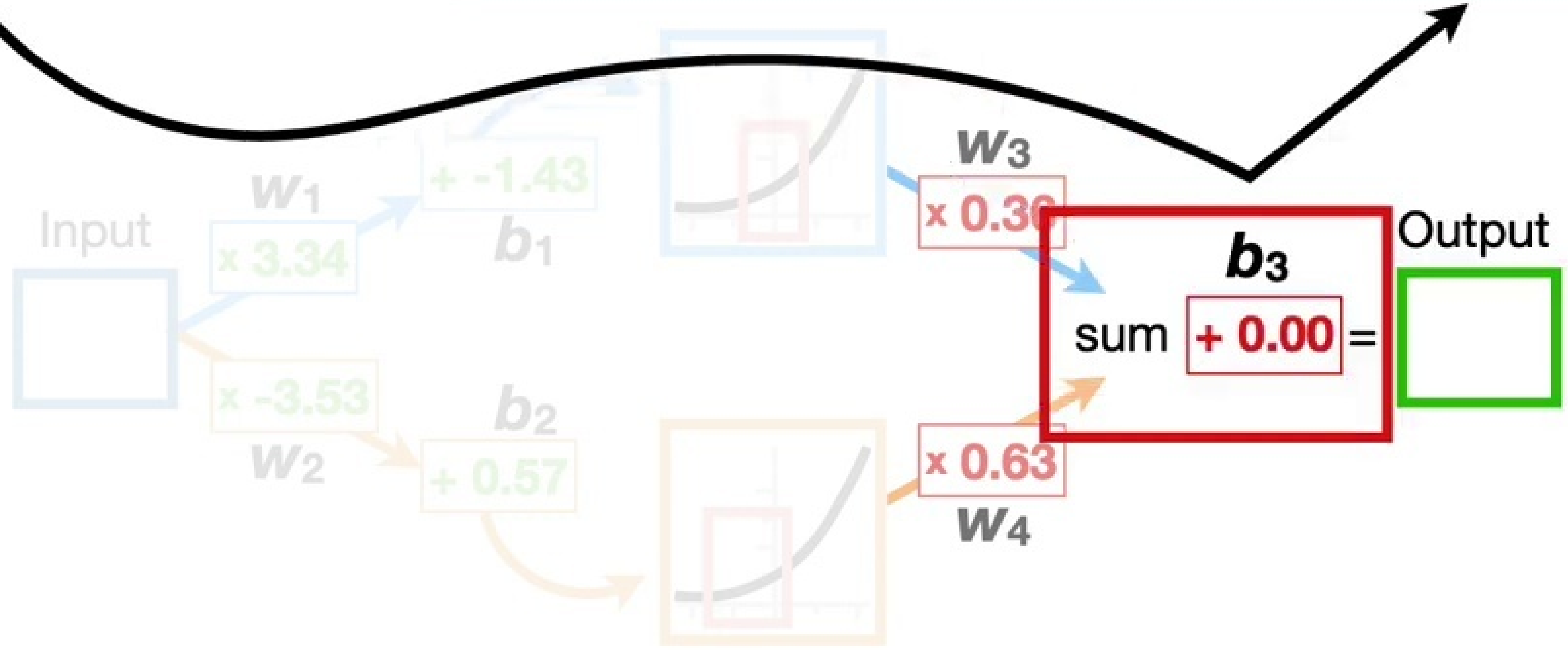
…plus b_3…



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



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

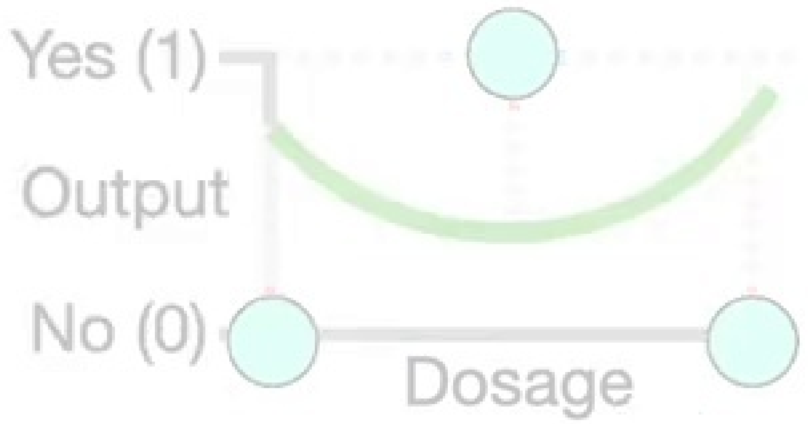


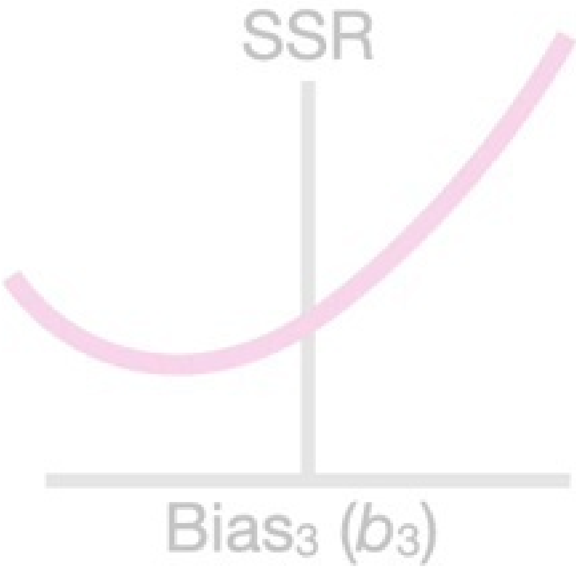
$$\frac{d\ SSR}{d\ b_3}$$

...then the **SSR** are linked to b_3...

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

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



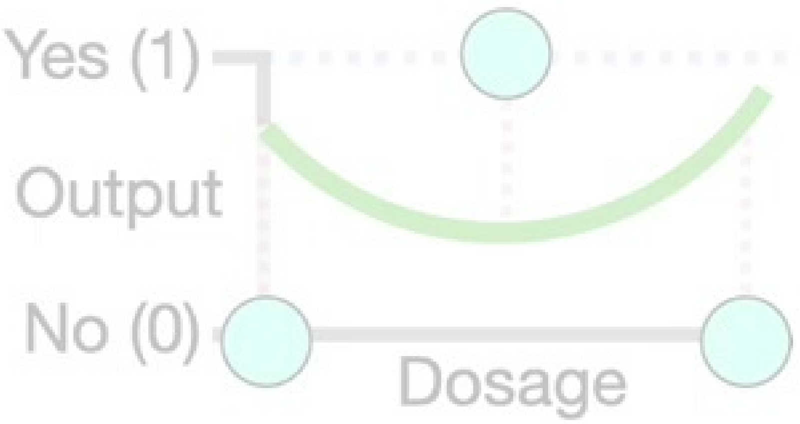


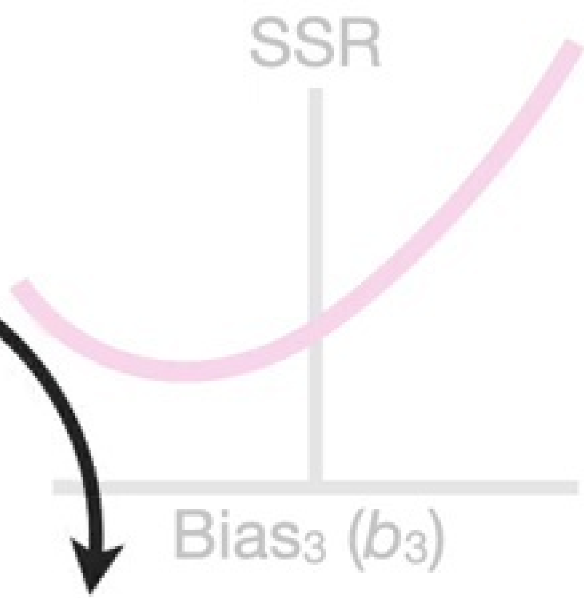
$$\frac{d\ SSR}{d\ b_3}$$

…then the **SSR** are linked to $\boldsymbol{b_3}$…

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

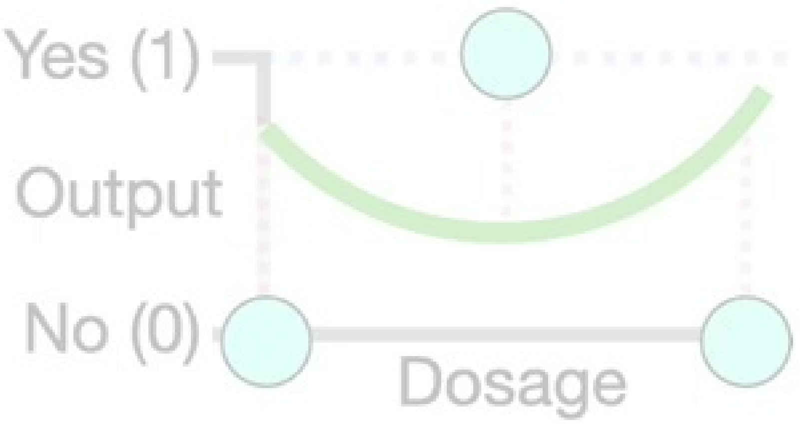
$$\text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} \boxed{+ b_3}$$





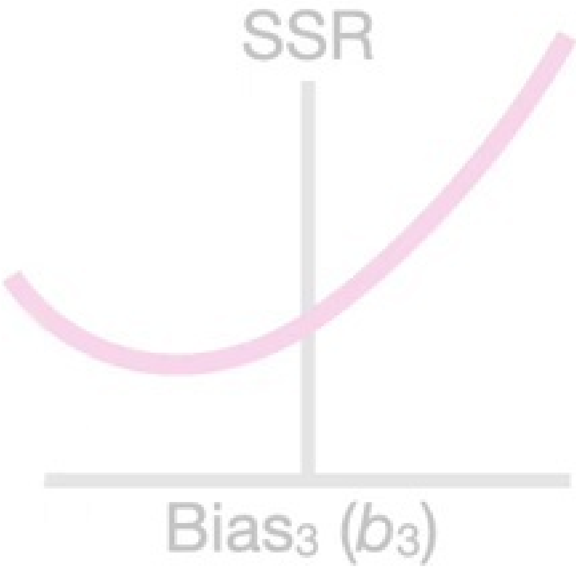
$$\frac{d\ SSR}{d\ b_3}$$

…by the **Predicted** values.



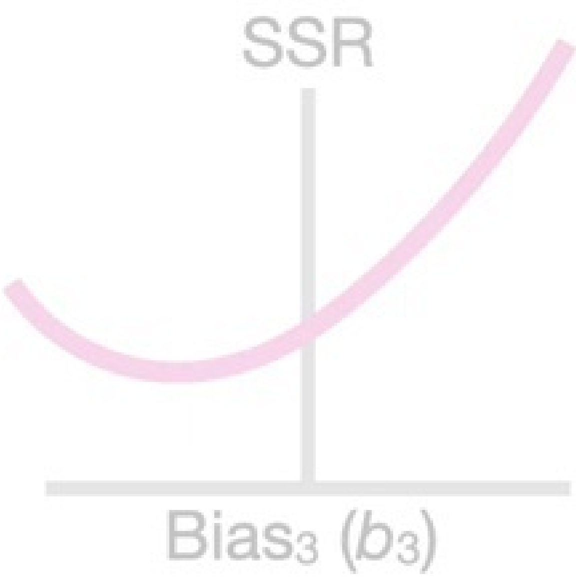
$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \boxed{\text{Predicted}_i})^2$$

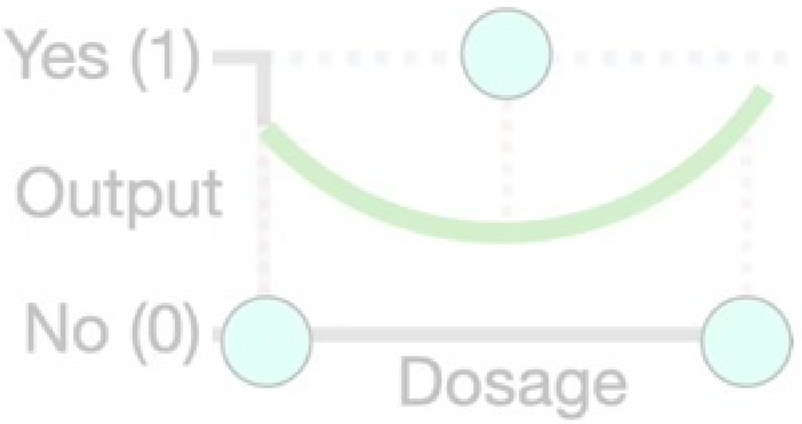
$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$



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

So, by **The Chain Rule**, the derivative of the **SSR** with respect to b_3…





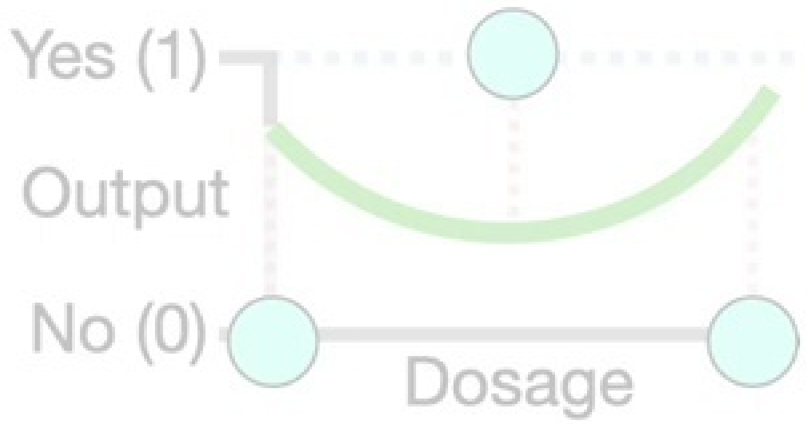
$$\mathrm{SSR} = \sum_{i=1}^{n=3} (\mathrm{Observed}_i - \boxed{\mathrm{Predicted}_i})^2$$

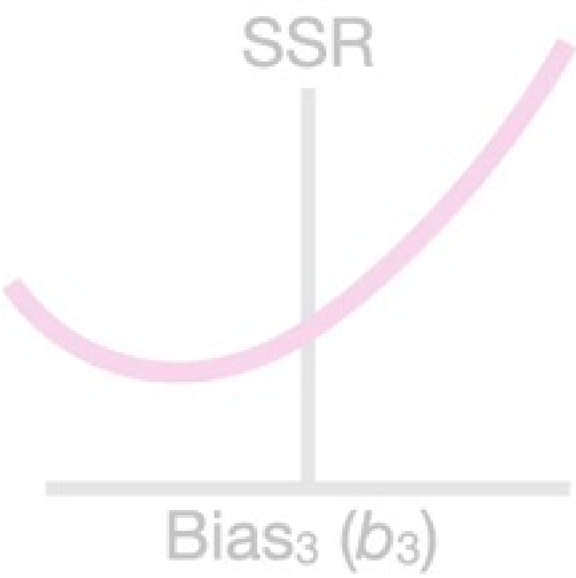
$\boxed{\mathrm{Predicted}_i}$ = **green squiggle$_i$** = **blue** + **orange** + b_3

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

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

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



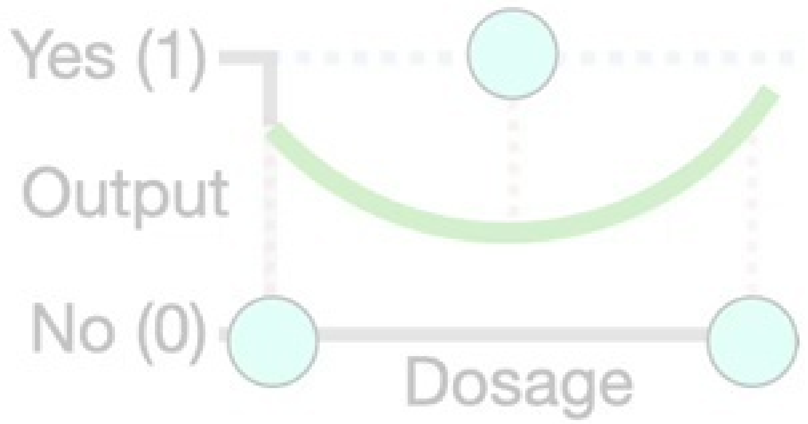


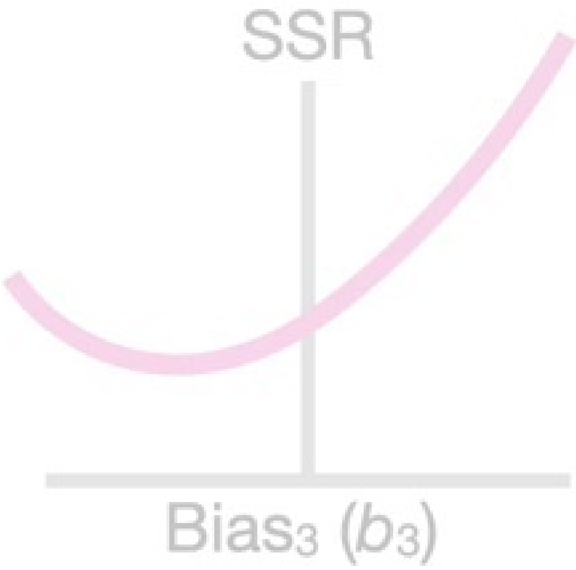
$\text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$

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

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

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

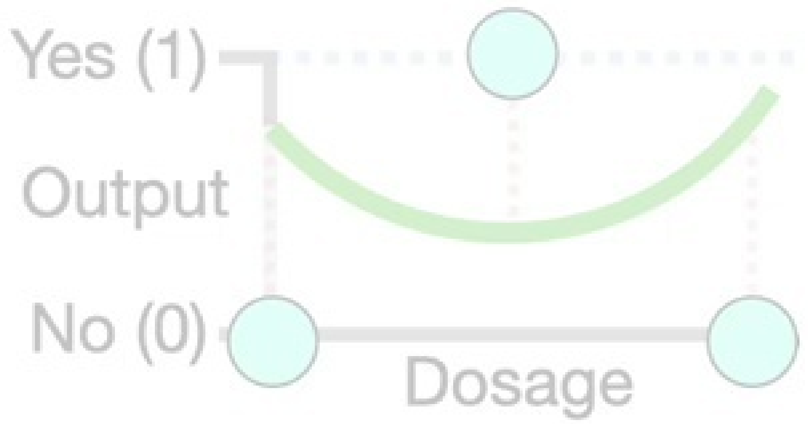


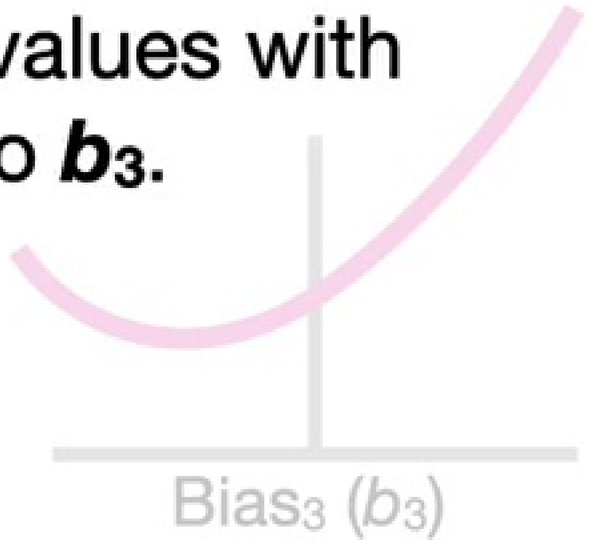


Predicted_i = green squiggle_i = blue + orange + b_3

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

…times the derivative of the **Predicted** values with respect to b_3.





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

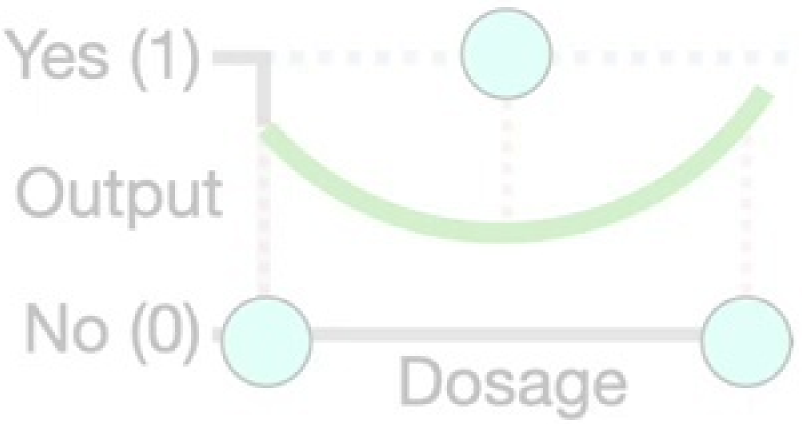
$\boxed{Predicted_i}$ = green squiggle$_i$ = blue + orange + b_3

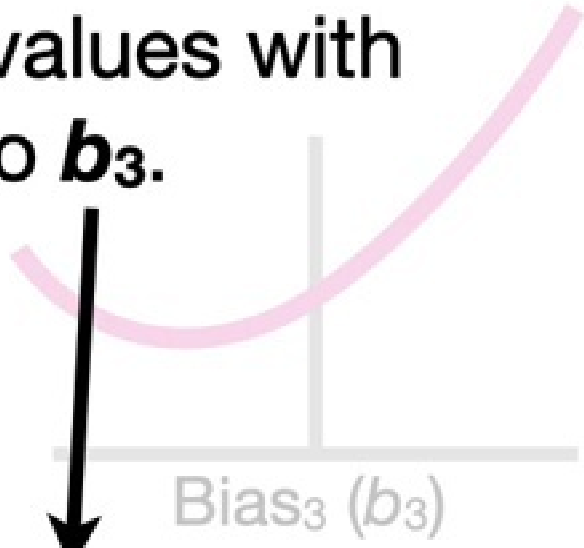
$$\frac{d\,SSR}{d\,b_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,b_3}$$

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

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

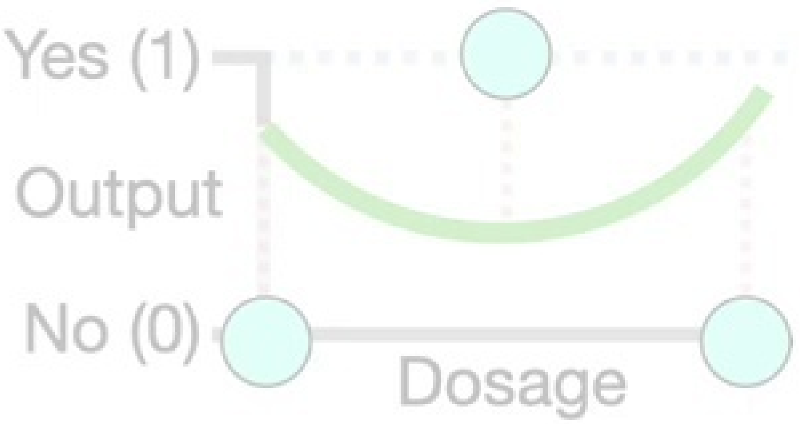
$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = \text{blue} + \text{orange} \boxed{+\, b_3}$$





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

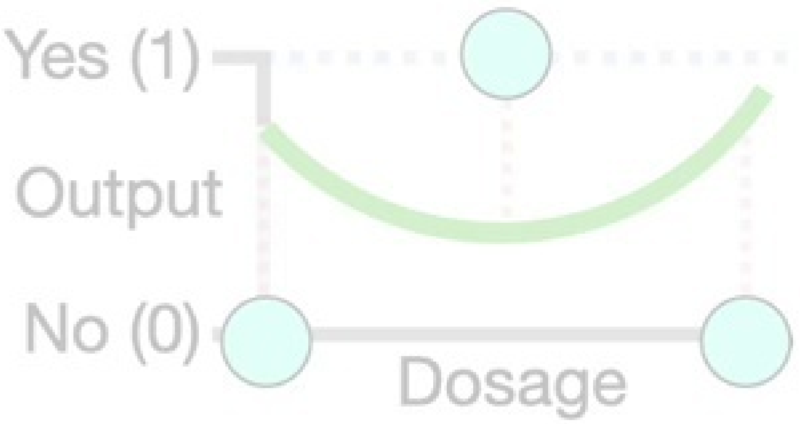
NOTE: This is the exact same derivative that we calculated in **Backpropagation Main ideas.**

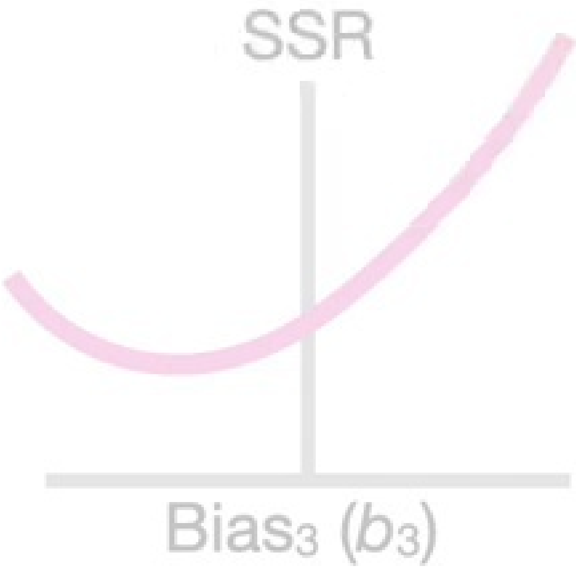


$$SSR = \sum_{i=1}^{n=3} (Observed_i - Predicted_i)^2$$

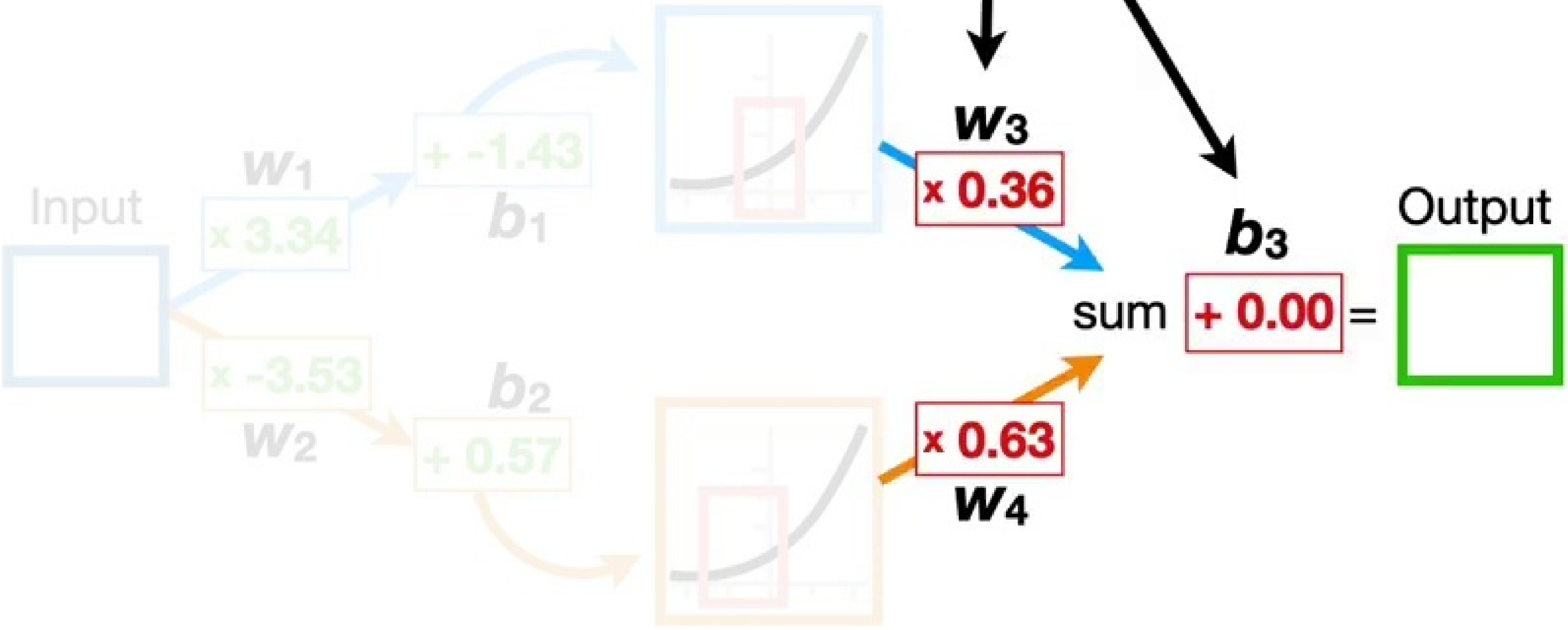
$$Predicted_i = \text{green squiggle}_i = \text{blue} + \text{orange} + b_3$$

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

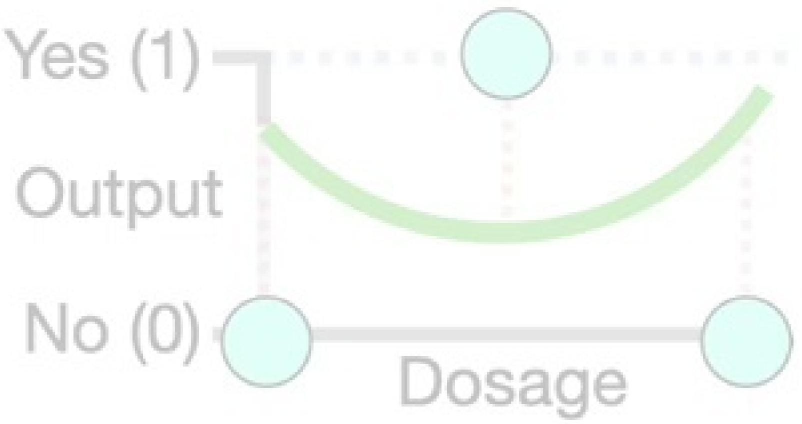


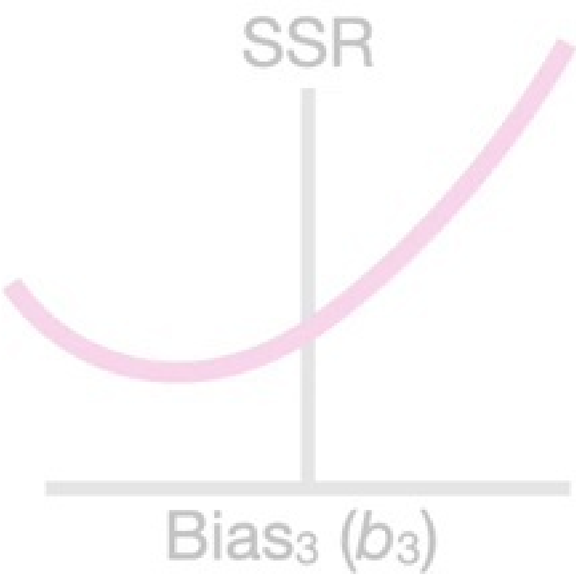


The point of this, is that even though we are now optimizing more than one parameter…

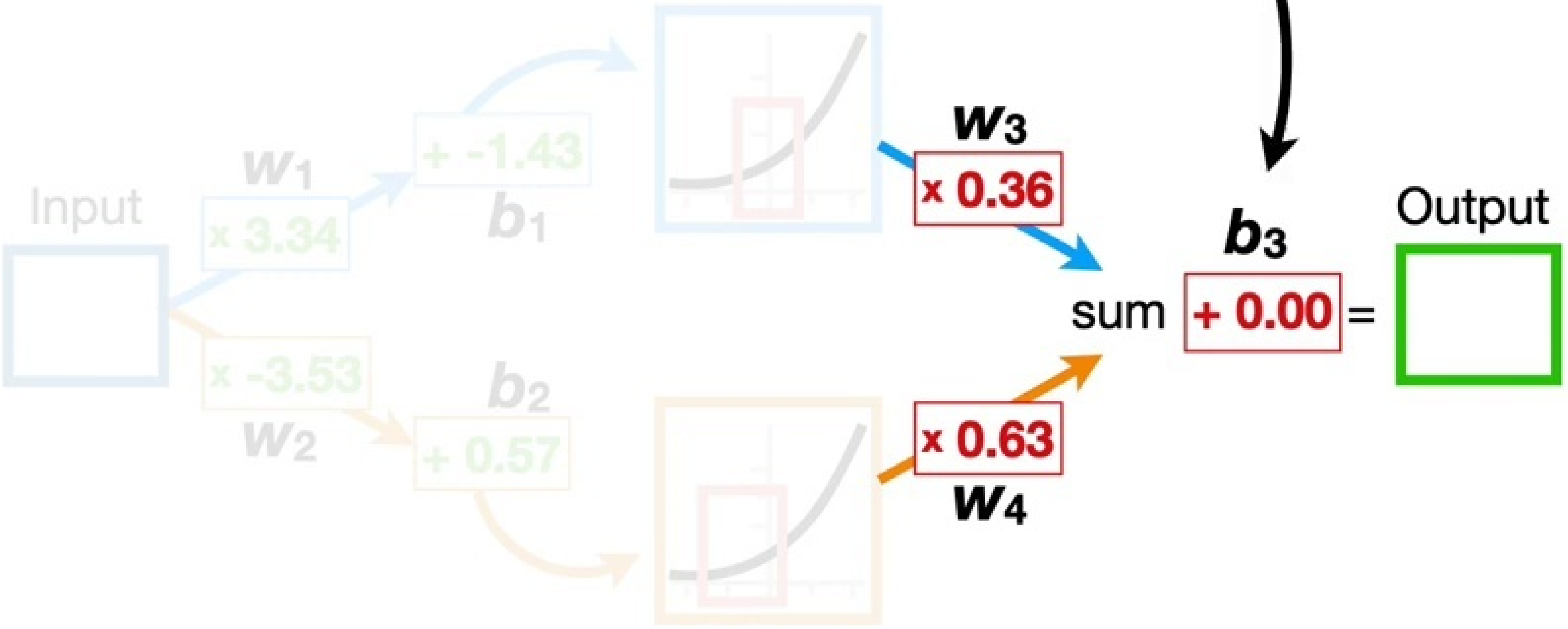


$$\frac{d\ SSR}{d\ b_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ b_3}$$

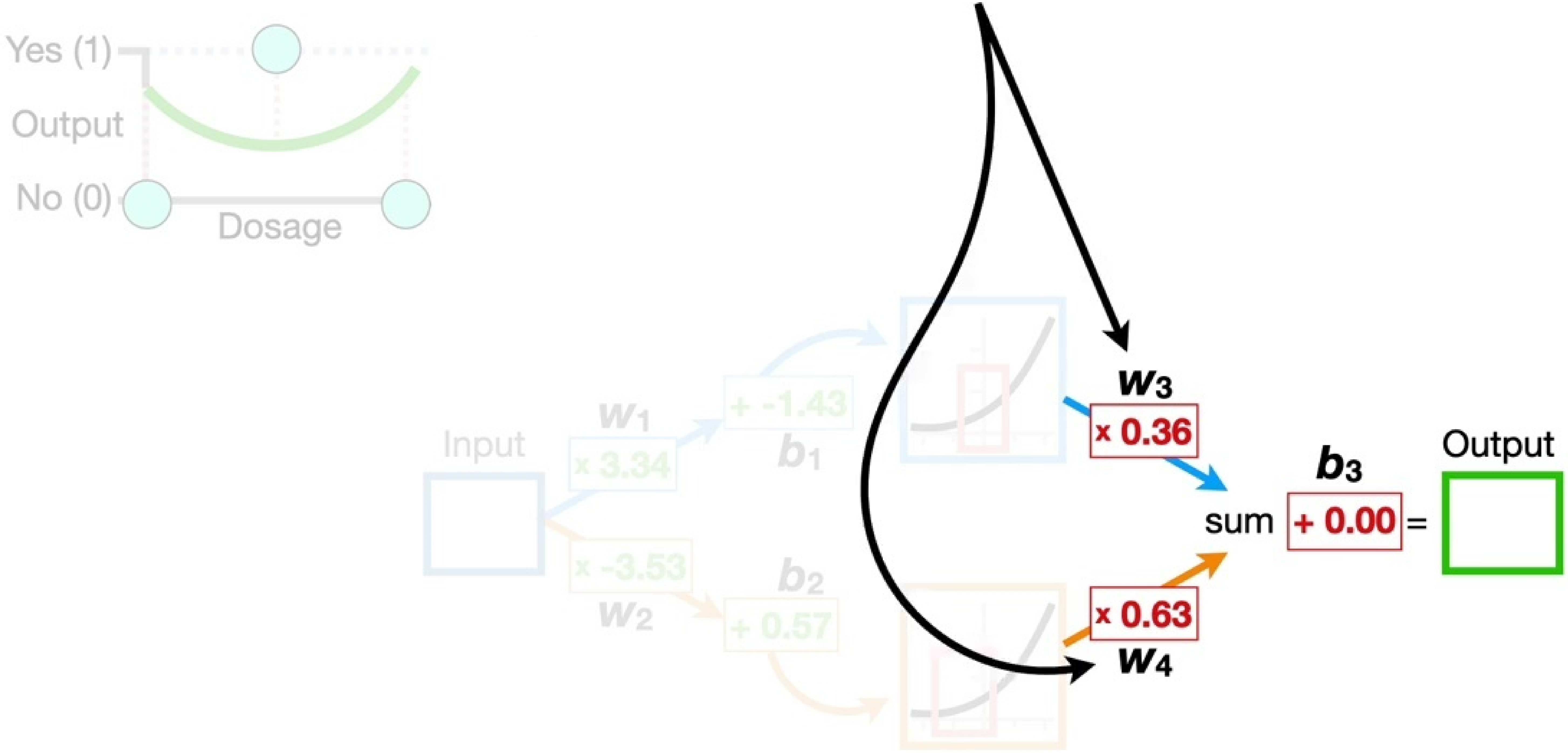




…the derivatives that we have already calculated with respect to the **SSR** do not change.

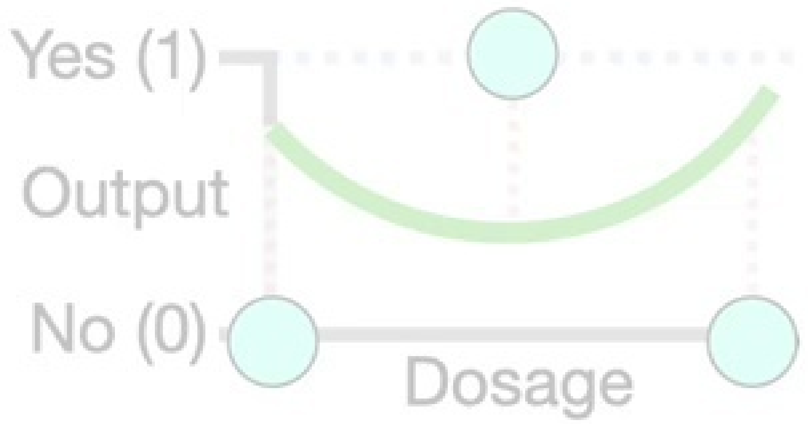


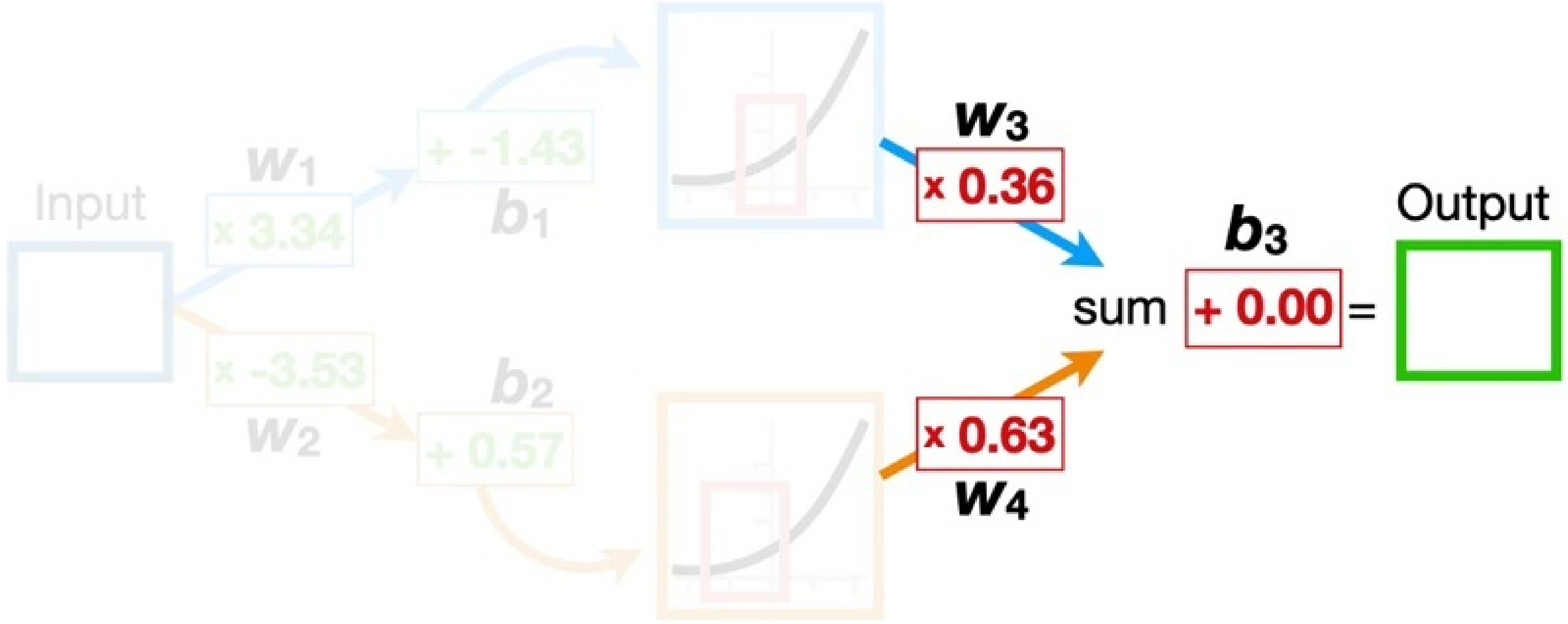
Now let's talk about how to calculate the derivatives of the **SSR** with respect to the **Weights** w_3 and w_4.



Unfortunately, before we can do that we have to introduce some…

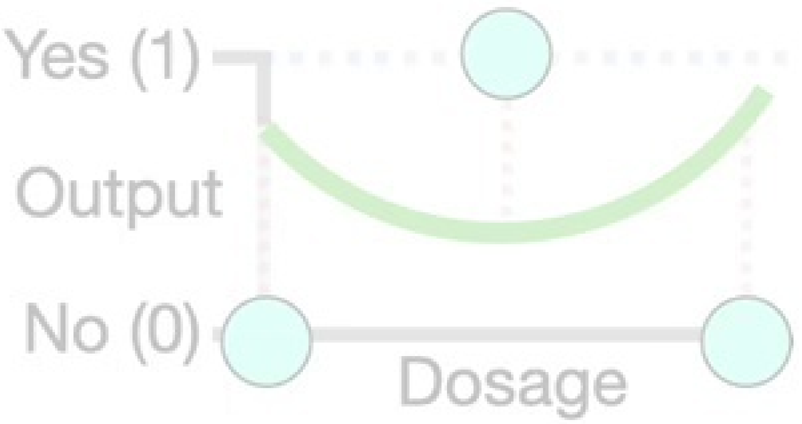
Fancy Notation!!!



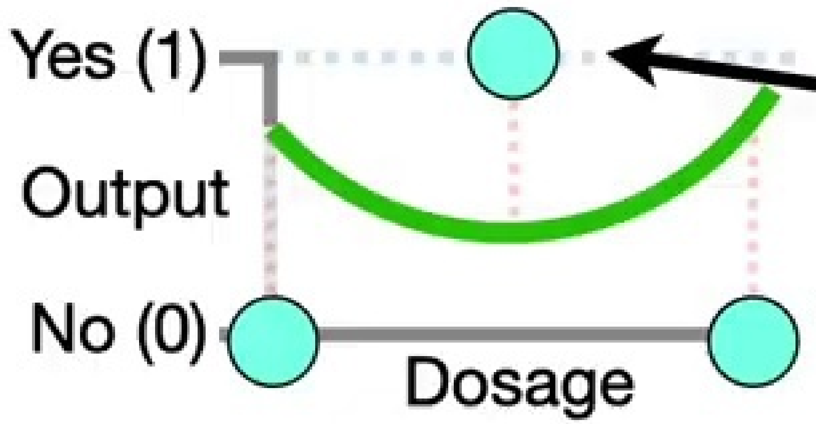


$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

First, let's remember that the ***i*** in this summation notation…



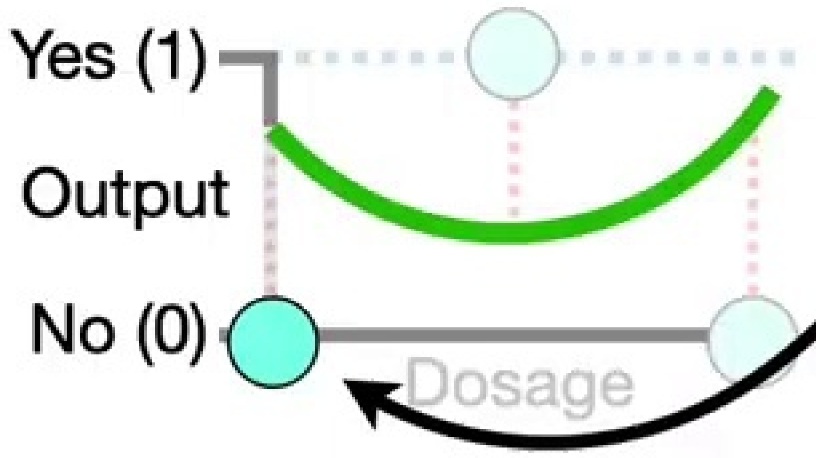
$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$



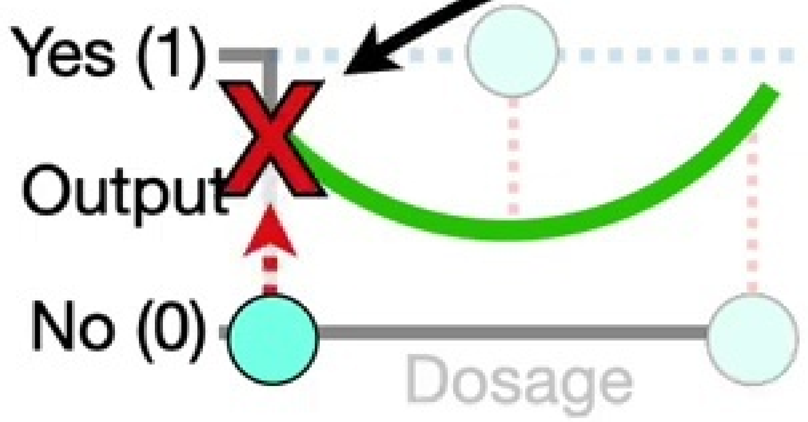
...is an **index** for the data in the dataset.

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

For example, when $\boldsymbol{i = 1}$, we are talking about **Observed$_1$**, which is **0**…

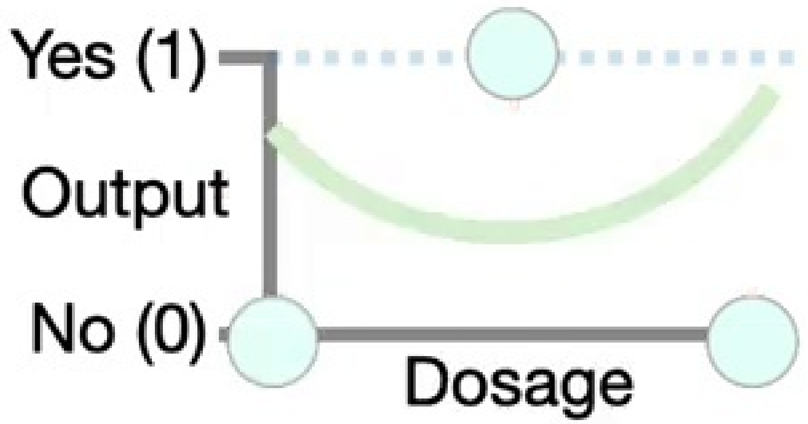


$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$



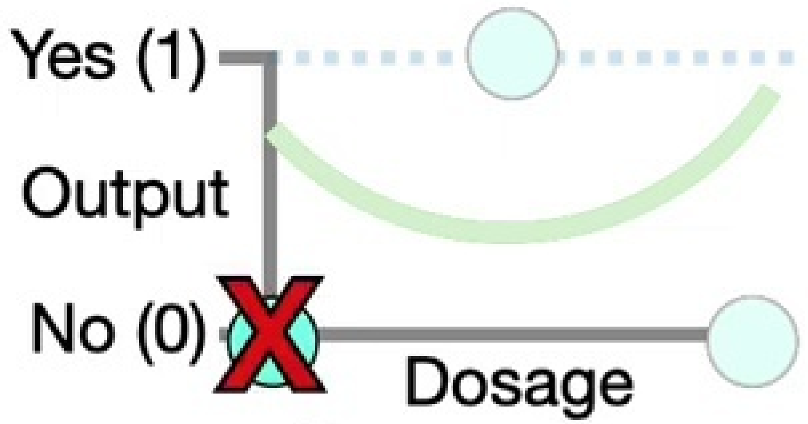
...and we are talking about **Predicted$_1$**, which is **0.72.**

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



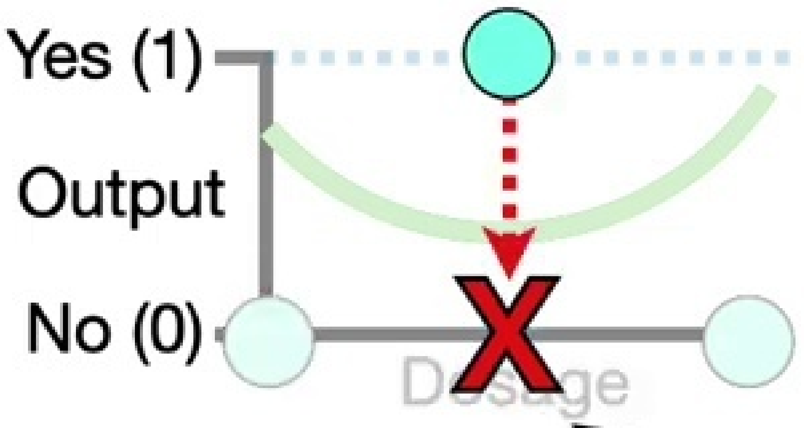
However, we can also talk about **Dosage$_i$**…

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



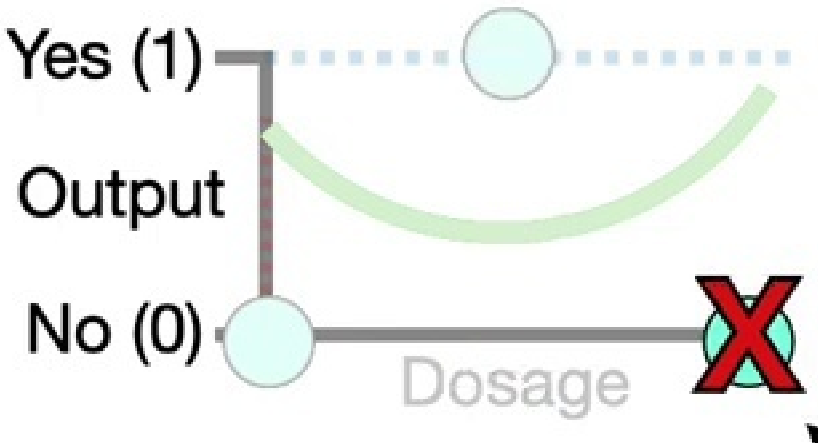
…and when ***i* = 1**, we are talking about **Dosage$_1$**, which is **0**.

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

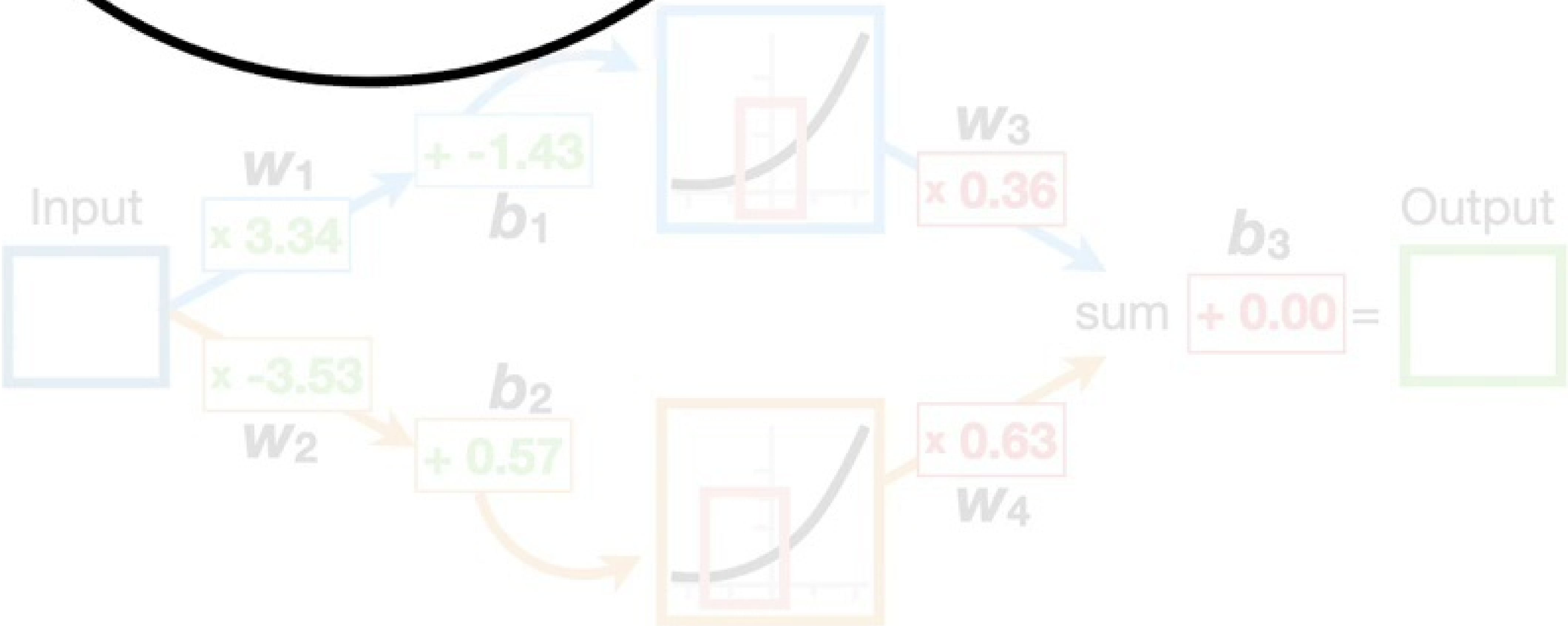


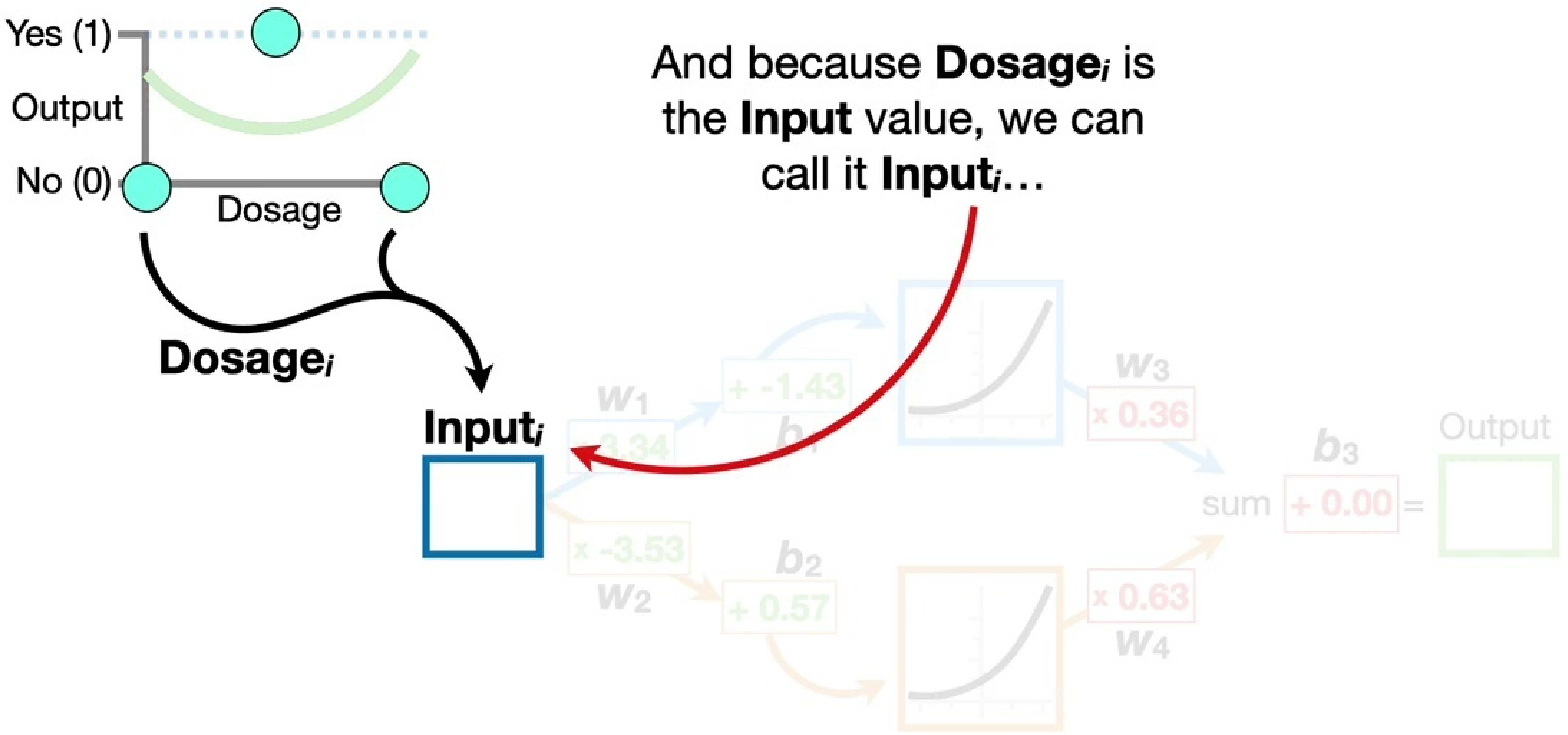
When ***i* = 2**, we are talking about **Dosage$_2$**, which is **0.5**…

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

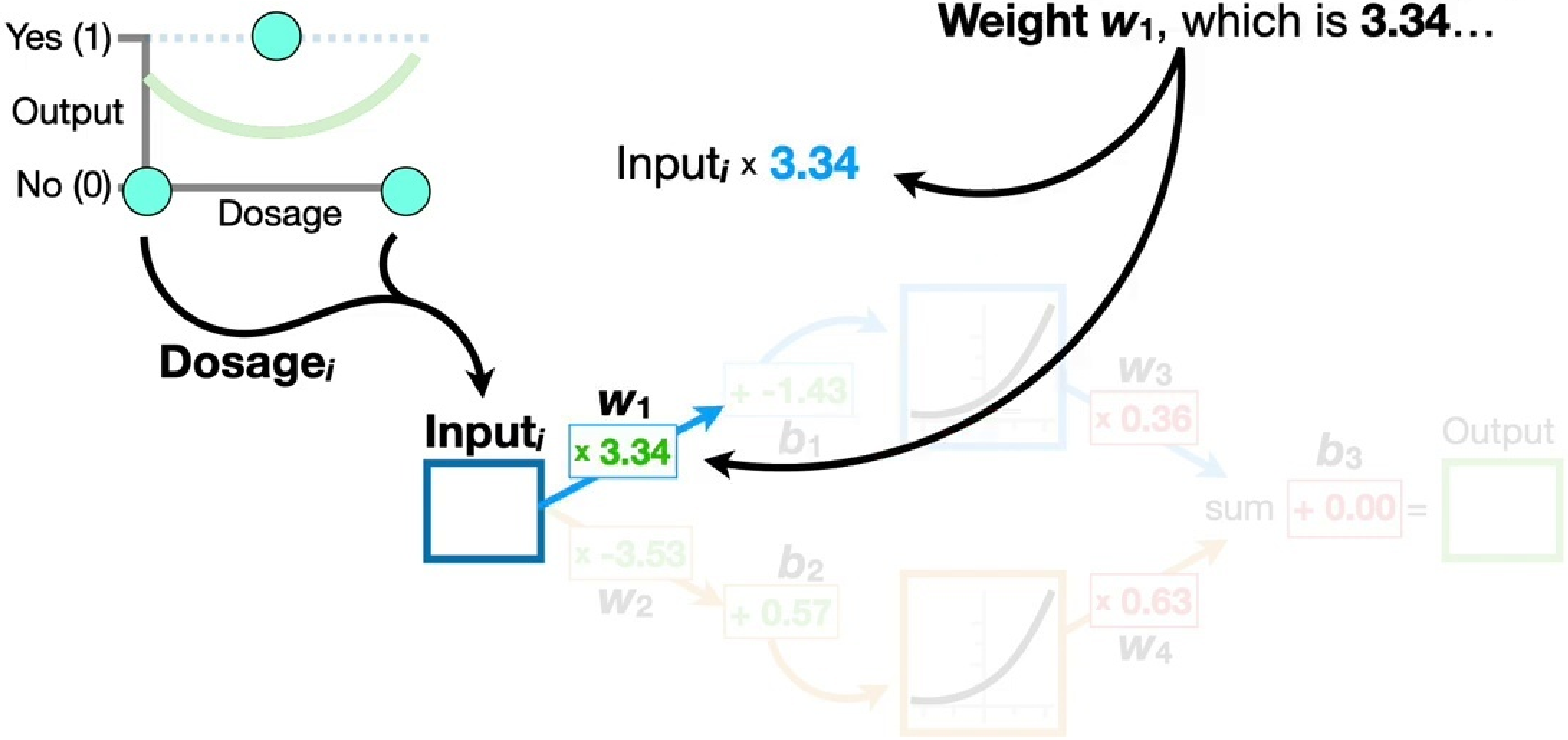


…and when ***i* = 3**, we are talking about **Dosage$_3$**, which is **1**.

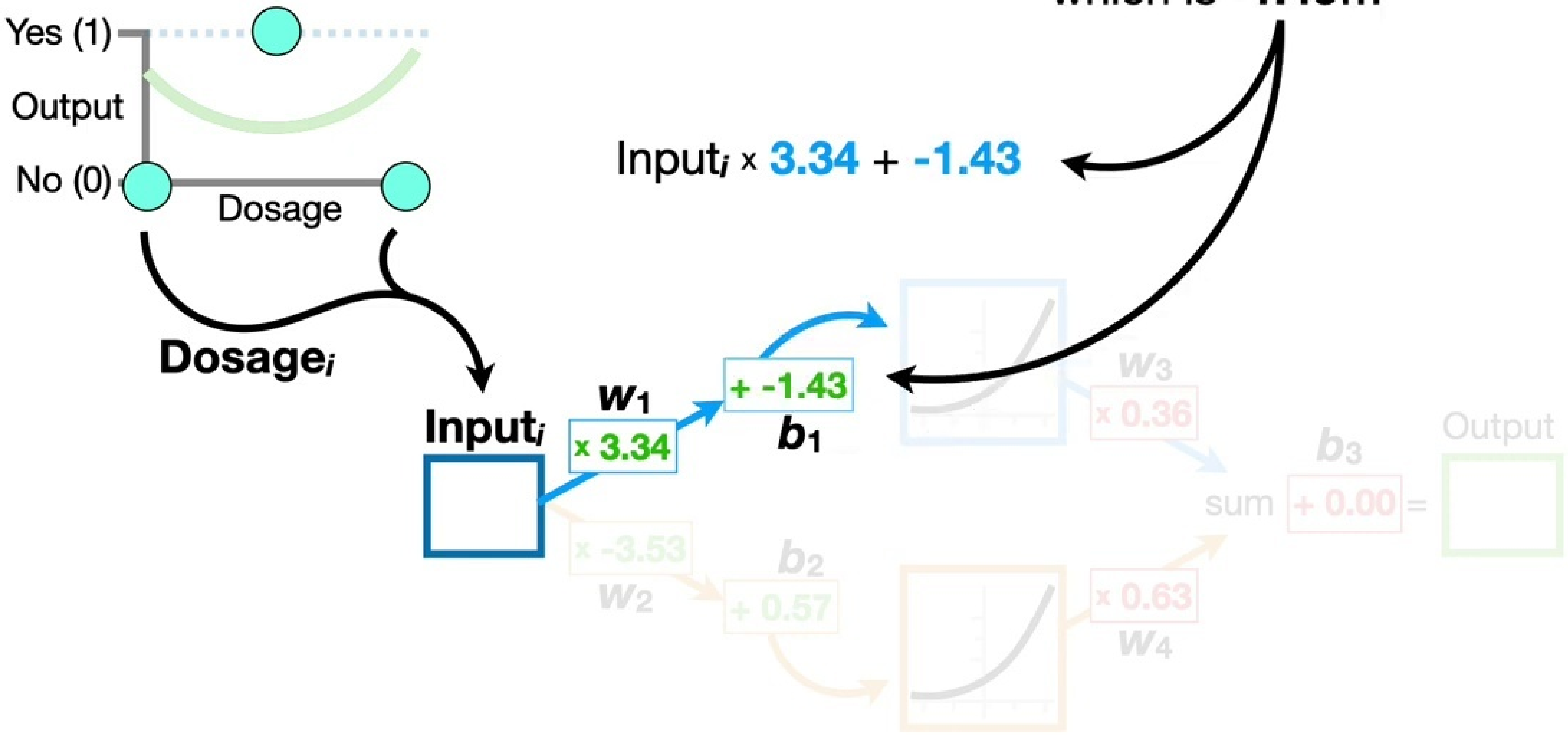



And because **Dosage*i*** is the **Input** value, we can call it **Input*i***…
Yes (1)
Output
No (0)
Dosage
Dosage*i*
Input*i*
+ -1.43
x 0.36
x -3.53
+ 0.57
x 0.63
sum
+ 0.00 =
Output

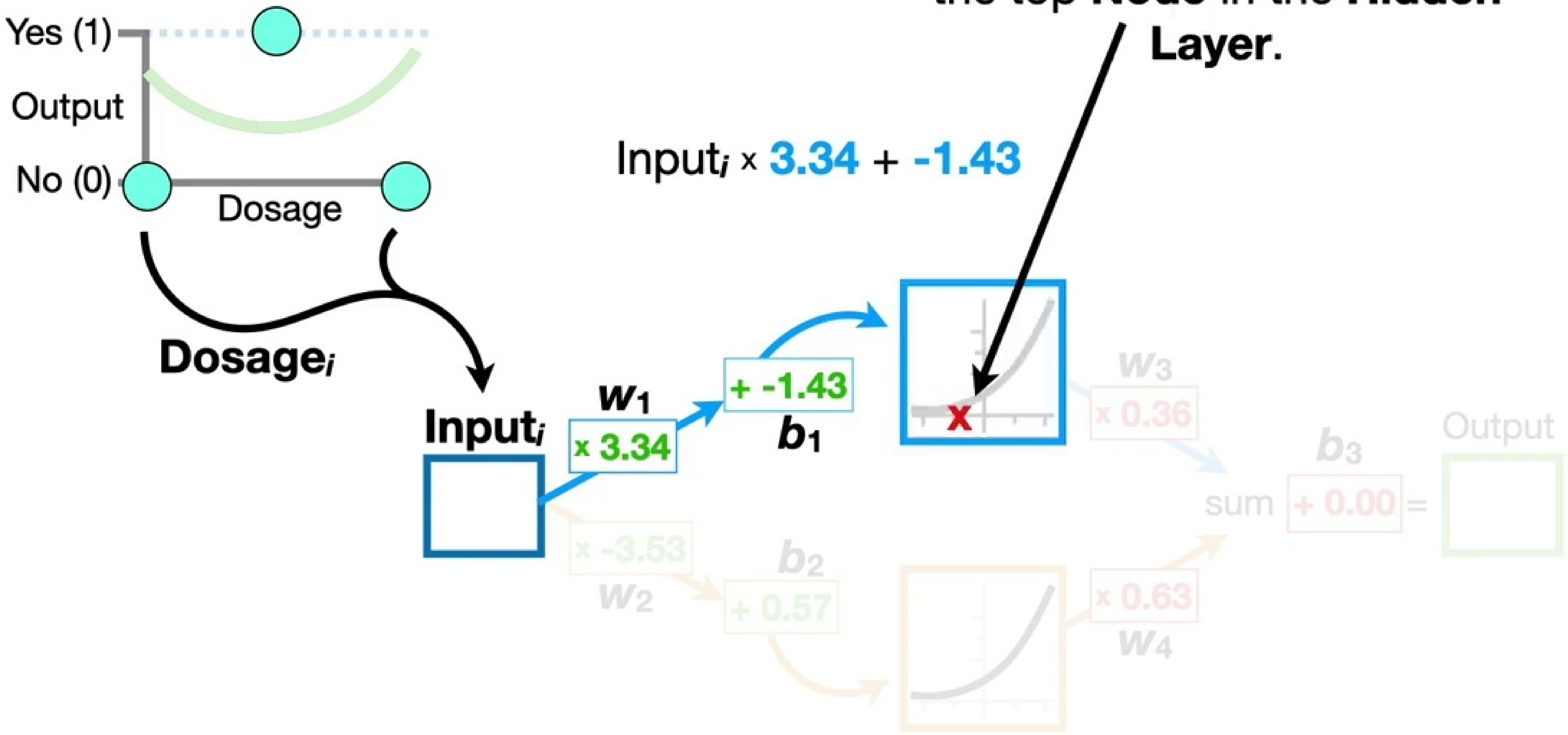
…and that means this connection multiplies **Input$_i$** by **Weight w_1**, which is **3.34**…



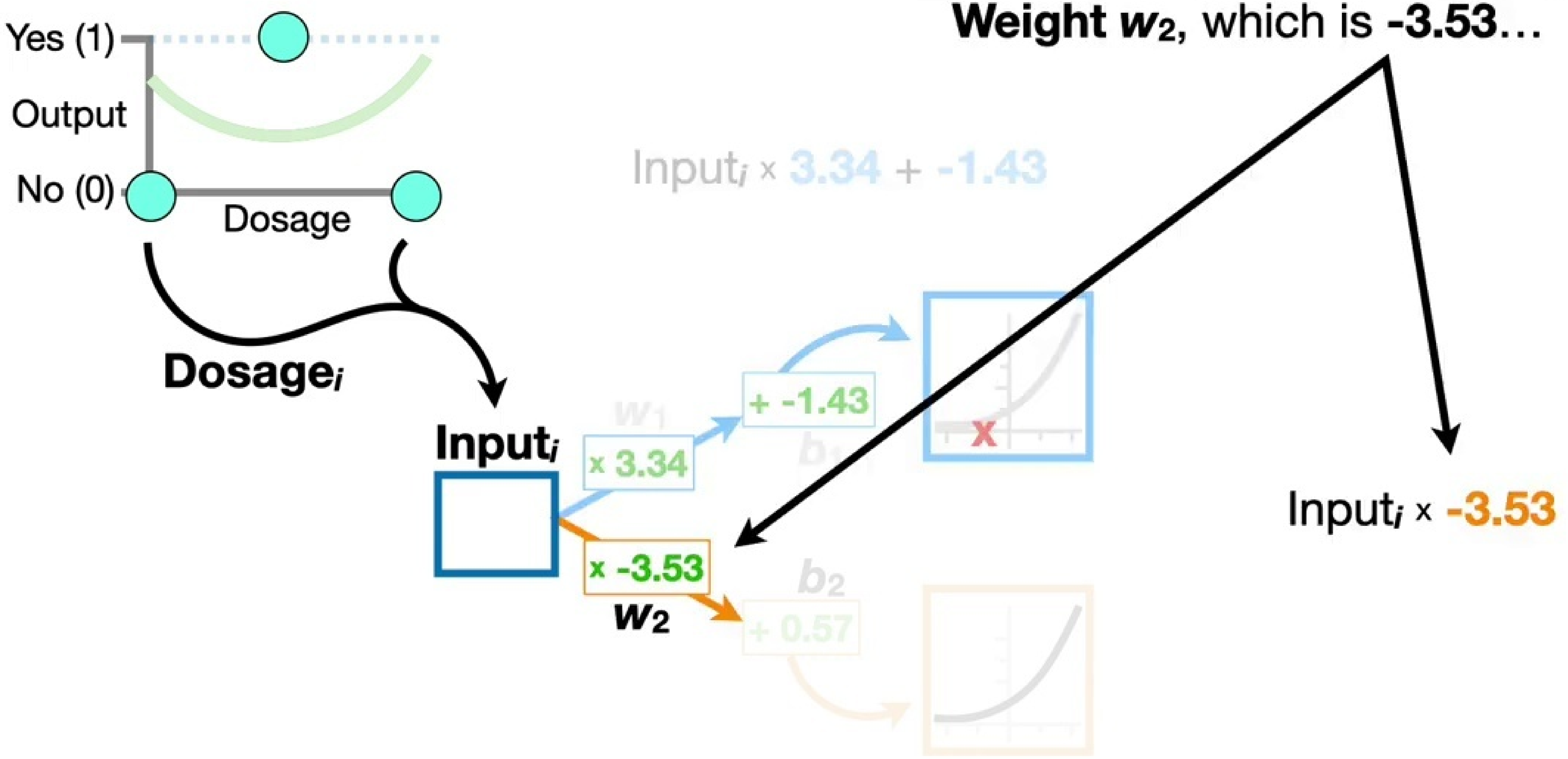
…and adds **Bias** b_1,
which is **-1.43**…



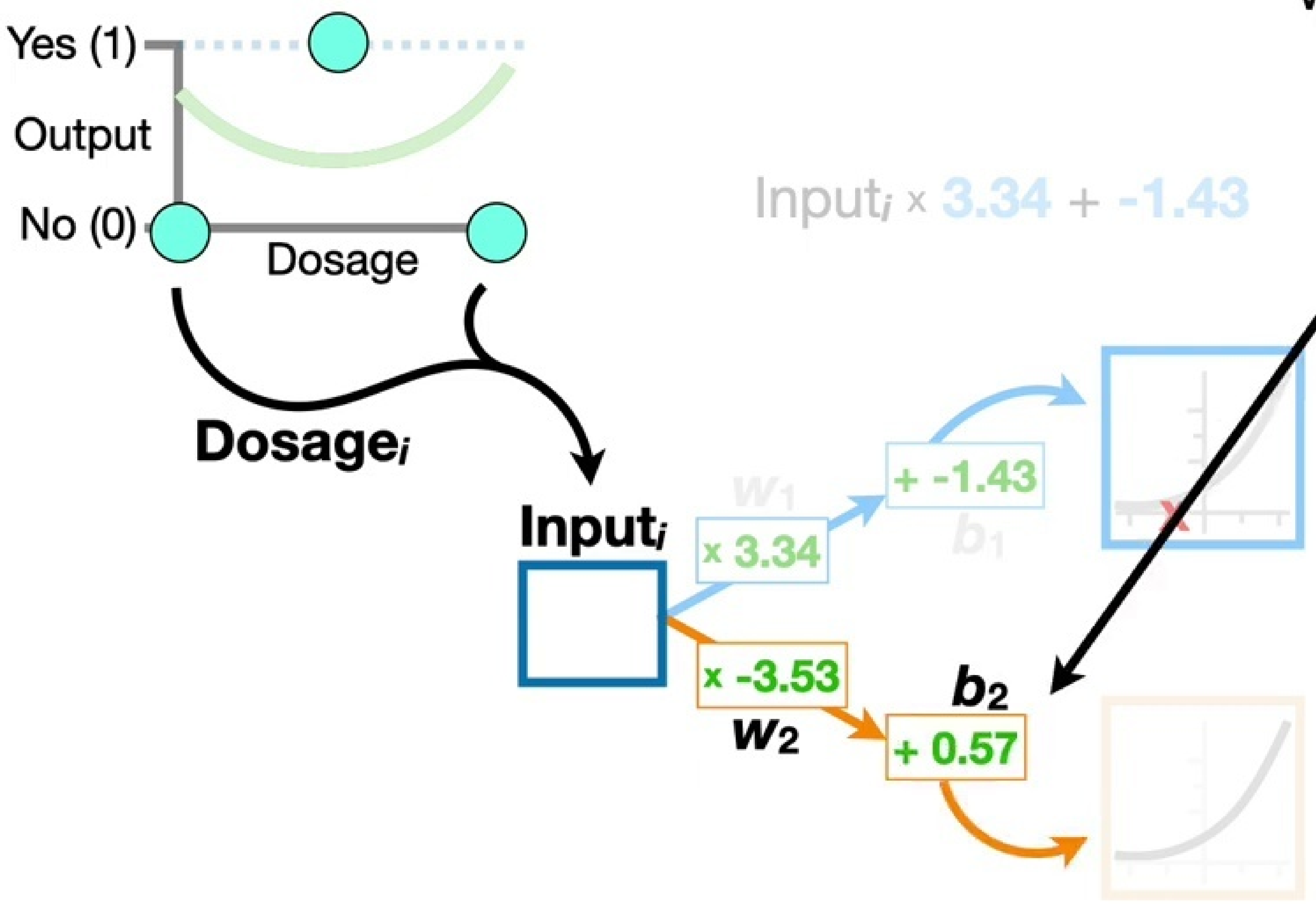
…to get an x-axis coordinate for the **Activation Function** in the top **Node** in the **Hidden Layer**.



Meanwhile, the other connection multiplies **Input$_i$** by **Weight w_2**, which is **-3.53**…

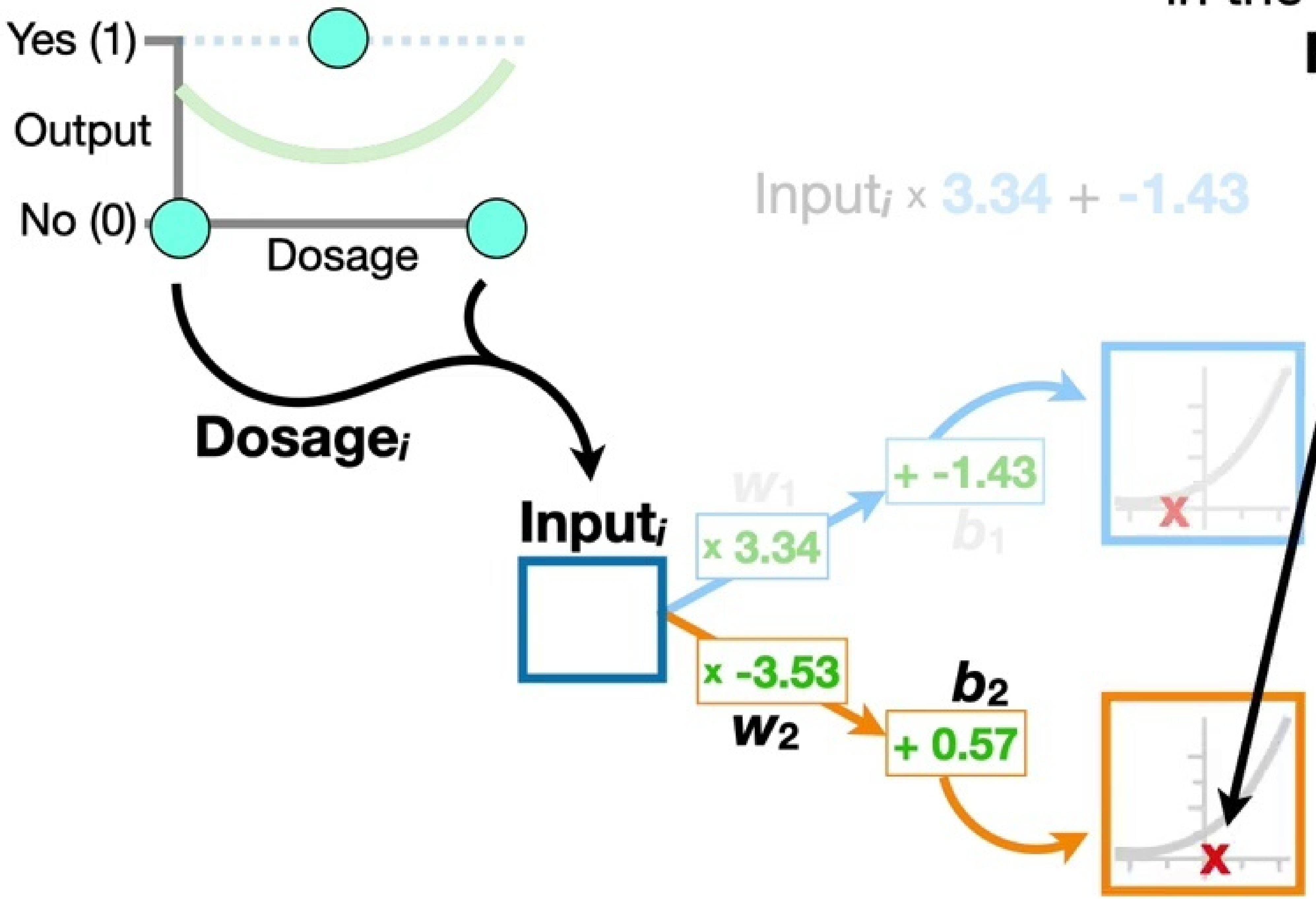


…and adds **Bias b_2**,
which is **0.57**…



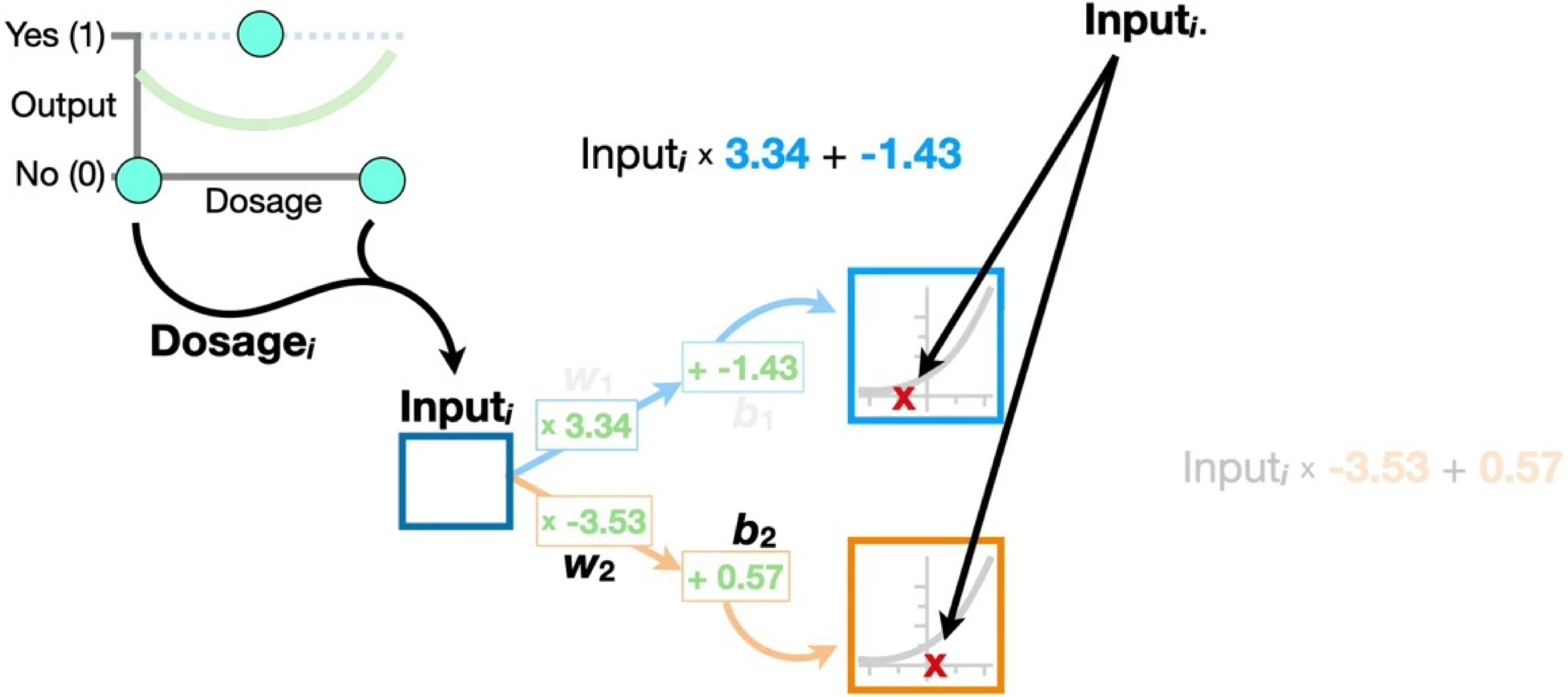
Input$_i$ × -3.53 + 0.57

…to get an x-axis coordinate for the **Activation Function** in the bottom **Node** in the **Hidden Layer**.



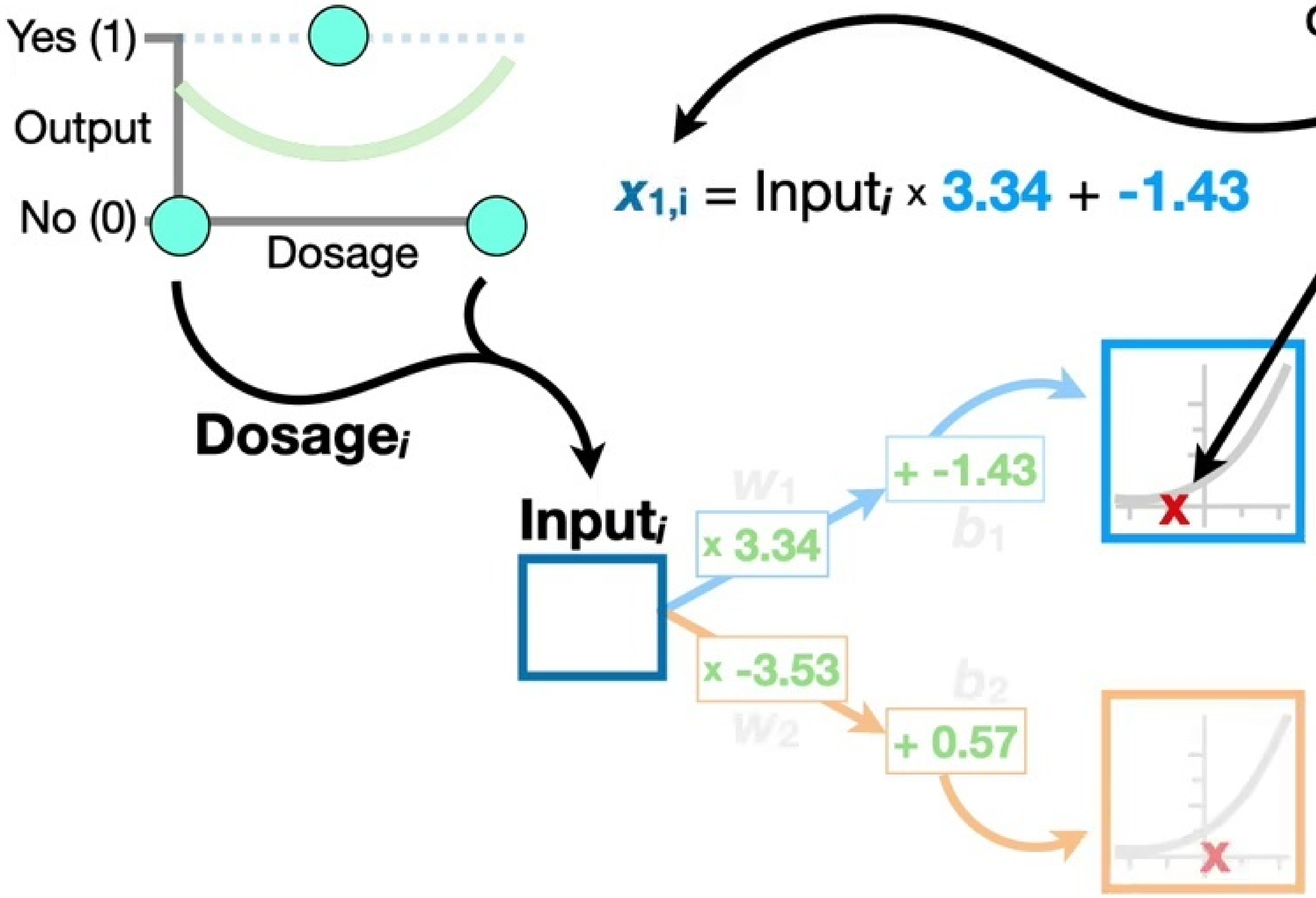
Input$_i$ × -3.53 + 0.57

So we have **2** different x-axis coordinates for **Input$_i$**.



In order to keep track of things, let's call this x-axis coordinate $x_{1,i}$.

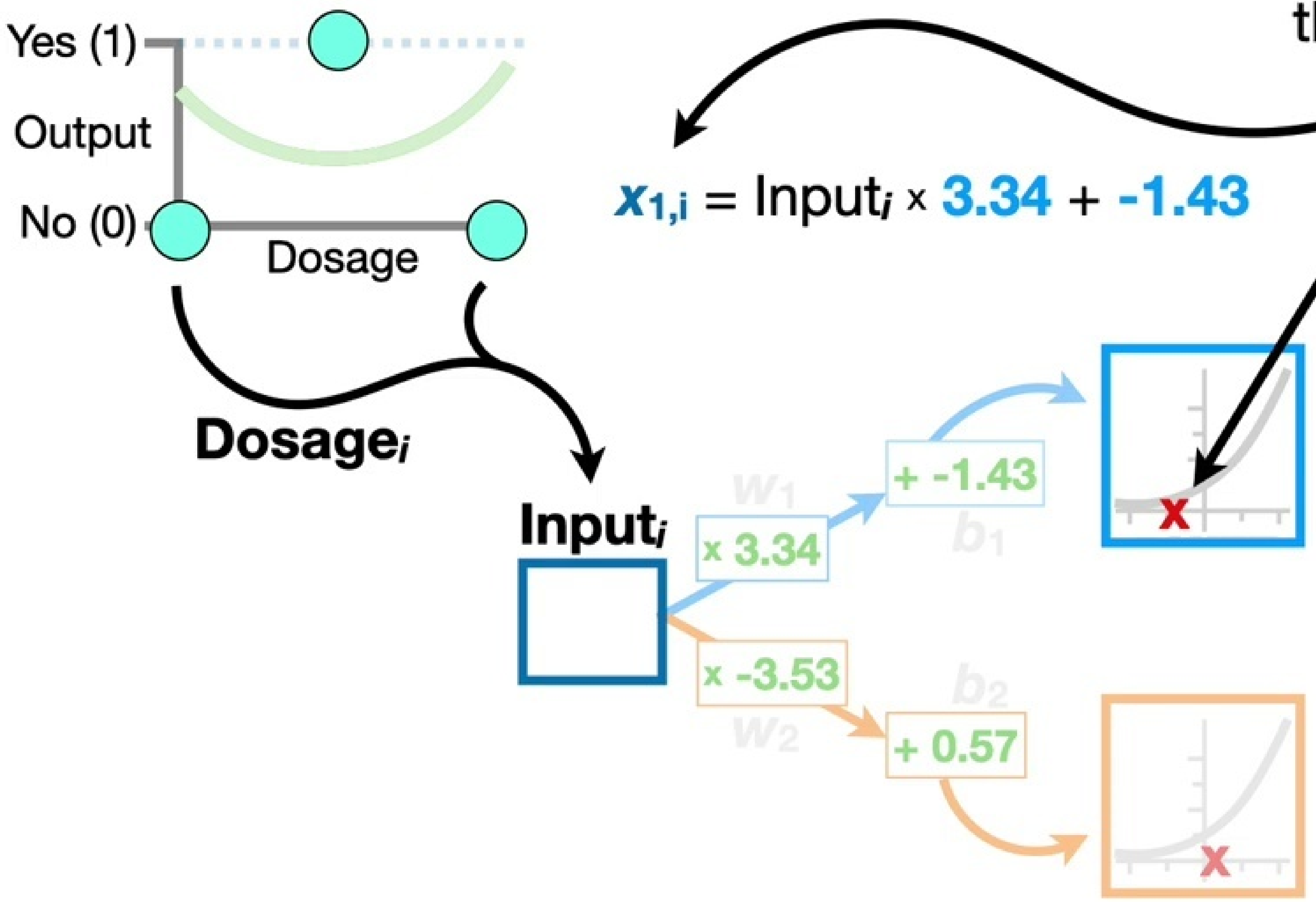
$$x_{1,i} = \text{Input}_i \times 3.34 + -1.43$$



$\text{Input}_i \times -3.53 + 0.57$

Where the **1,** in **1,i** refers to the **Activation Function** in the top **Node**…

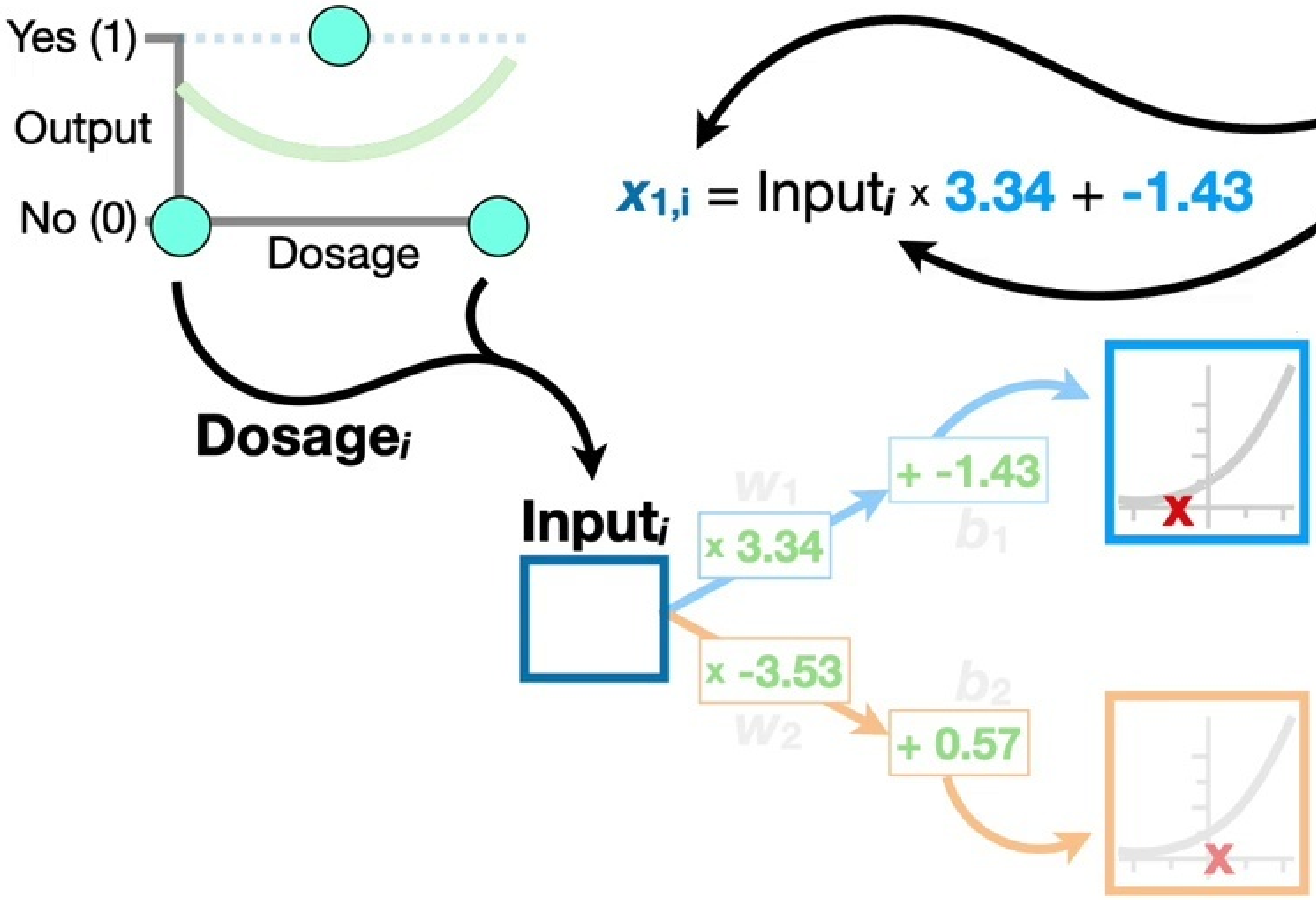
$x_{1,i}$ = Input$_i$ × 3.34 + -1.43



Input$_i$ × -3.53 + 0.57

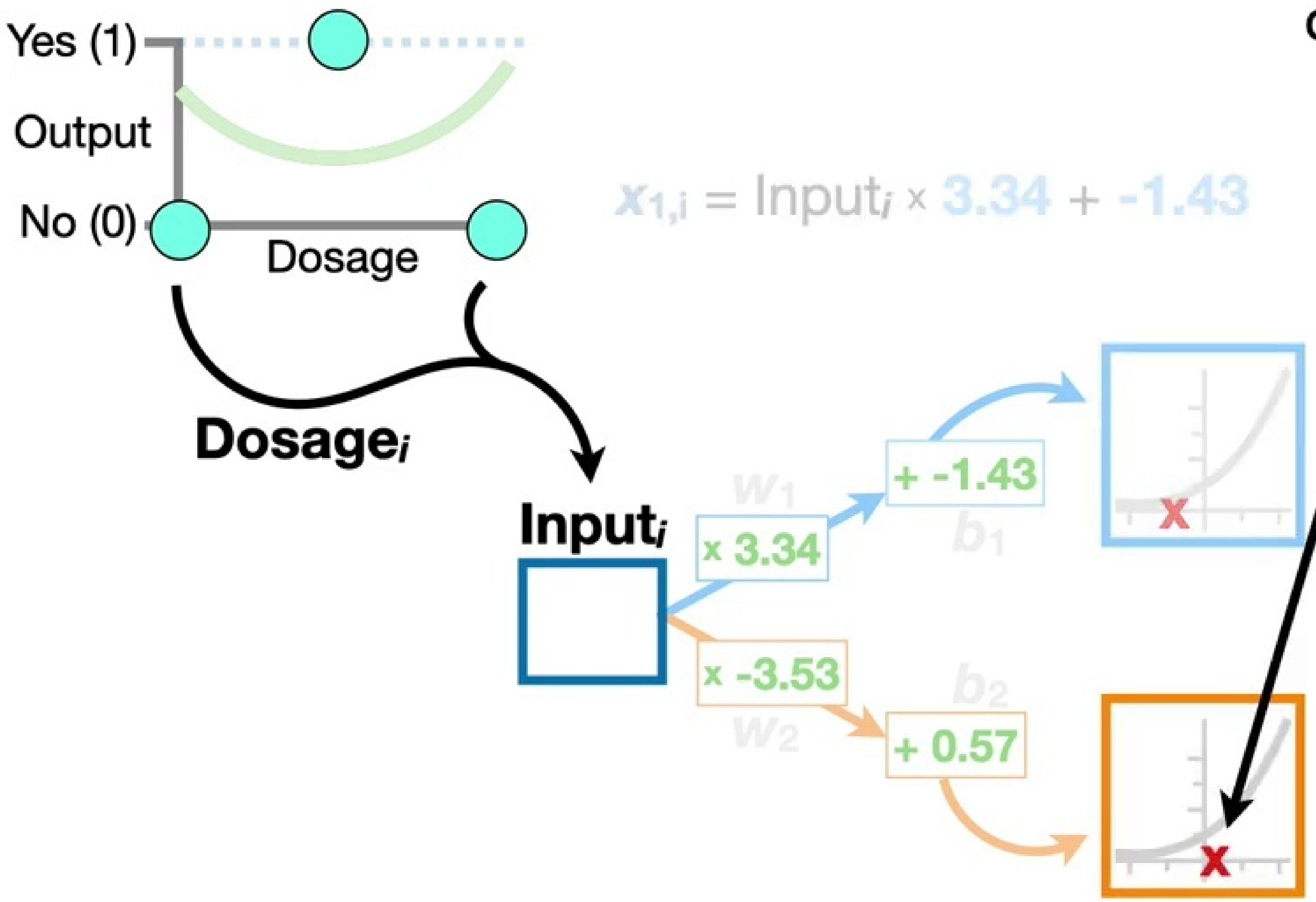
…and the ***i***, in **1,i** tells us that it corresponds to **Input*i*.**

$$x_{1,i} = \text{Input}_i \times 3.34 + -1.43$$



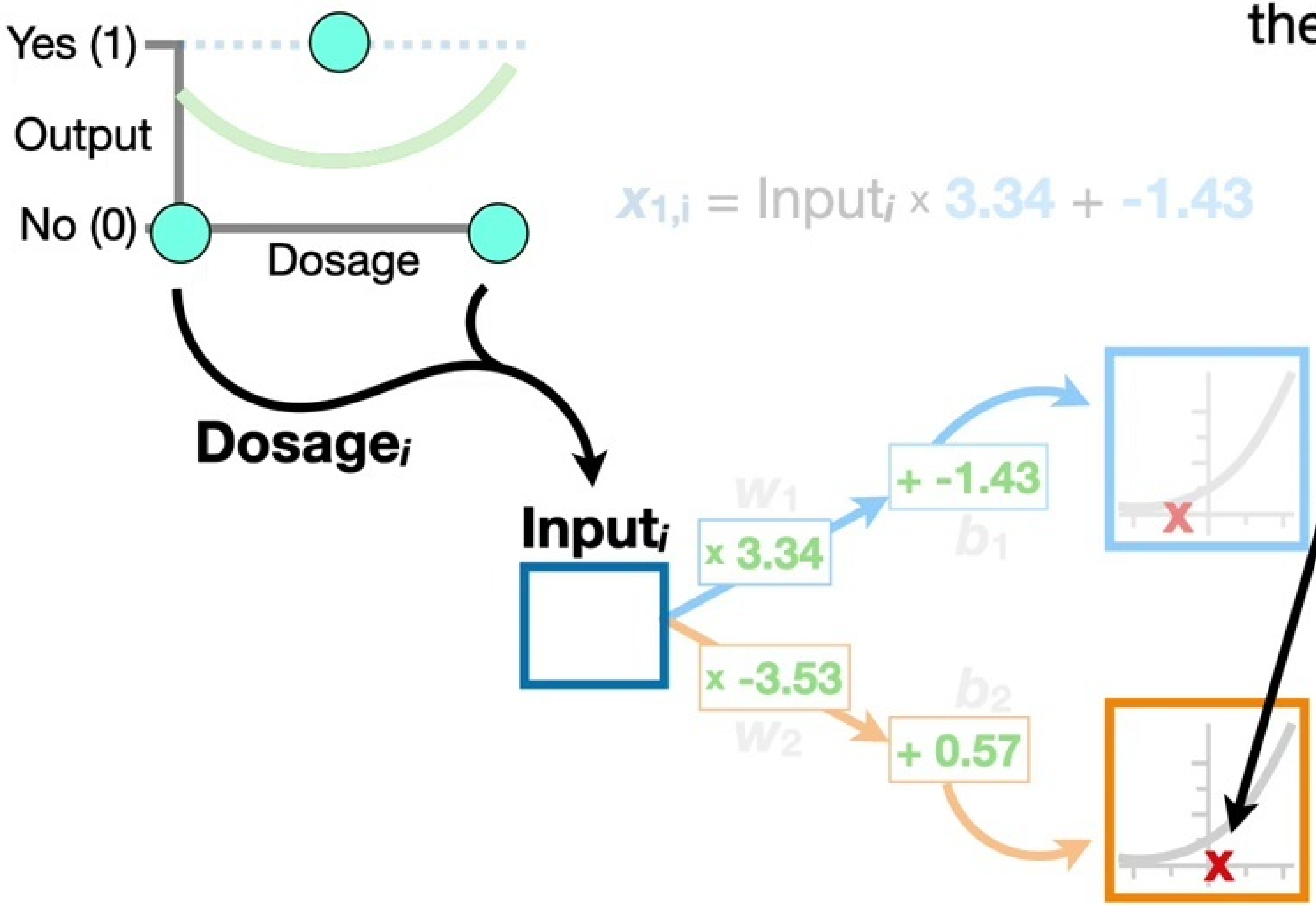
$\text{Input}_i \times -3.53 + 0.57$

Likewise, let's call this x-axis coordinate $x_{2,i}$.



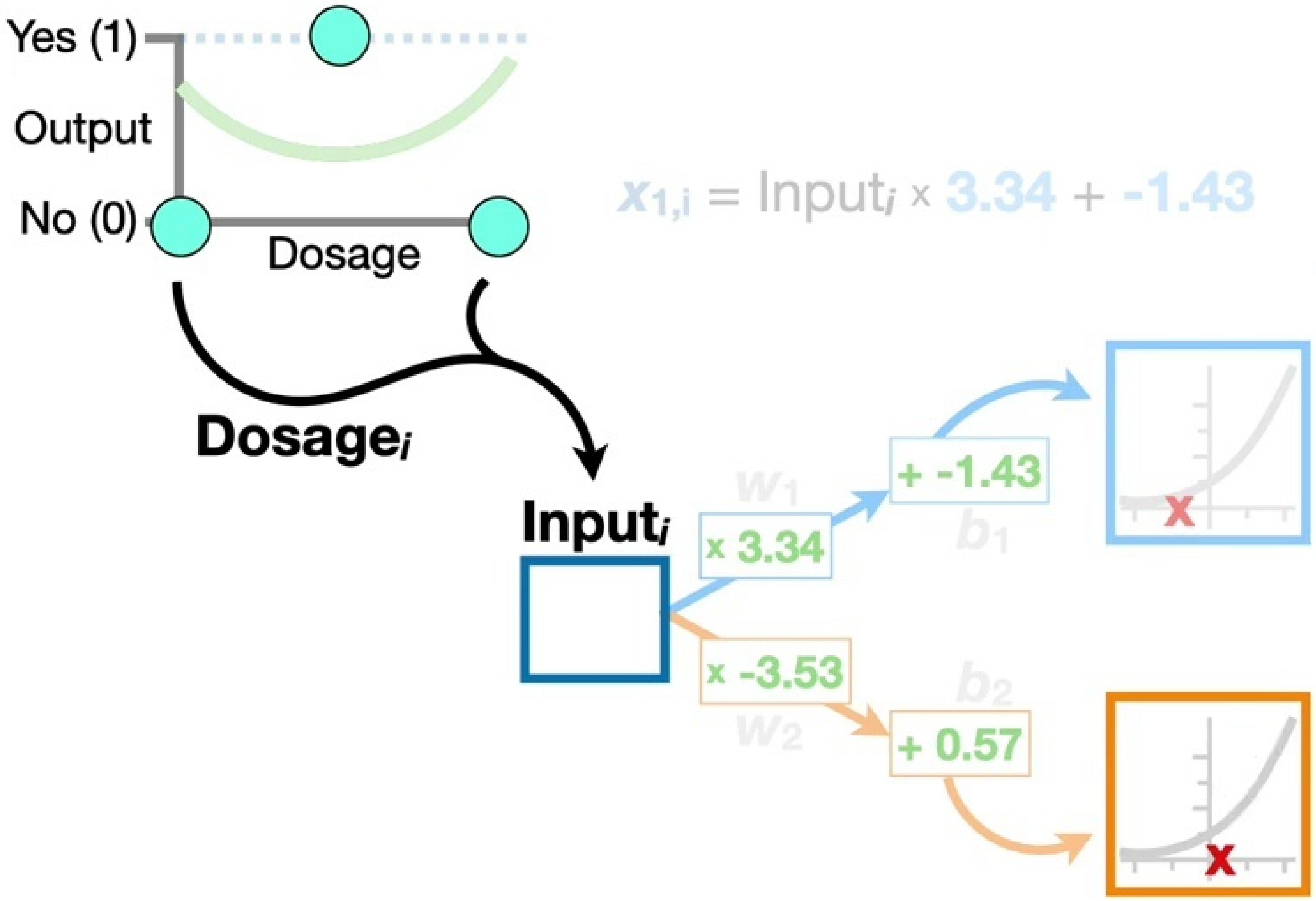
$x_{2,i}$ = Input$_i$ × -3.53 + 0.57

Where the **2** in **2,i** refers to the **Activation Function** in the bottom **Node**…



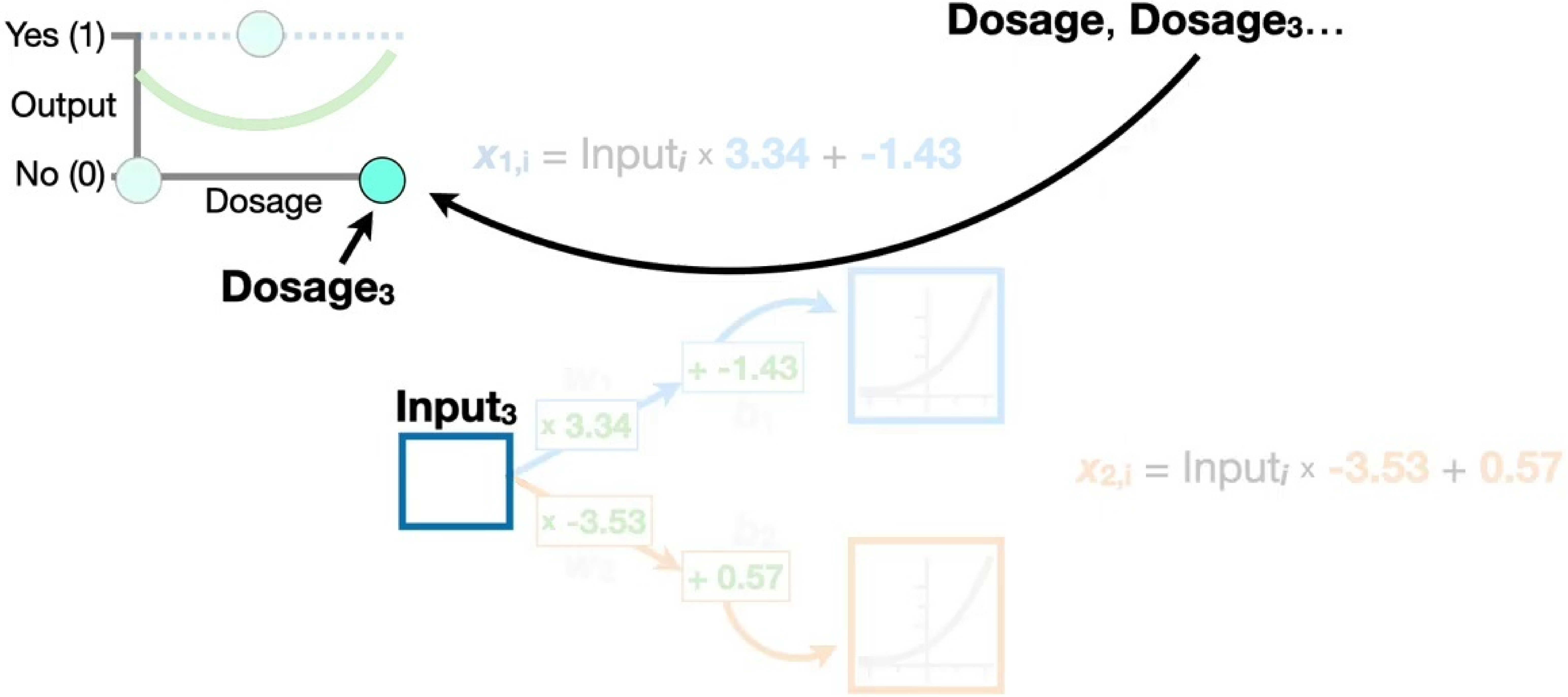
$x_{2,i}$ = Input$_i$ × -3.53 + 0.57

…and the ***i*** in **2,i** tells us that it corresponds to **Input*ᵢ*.**

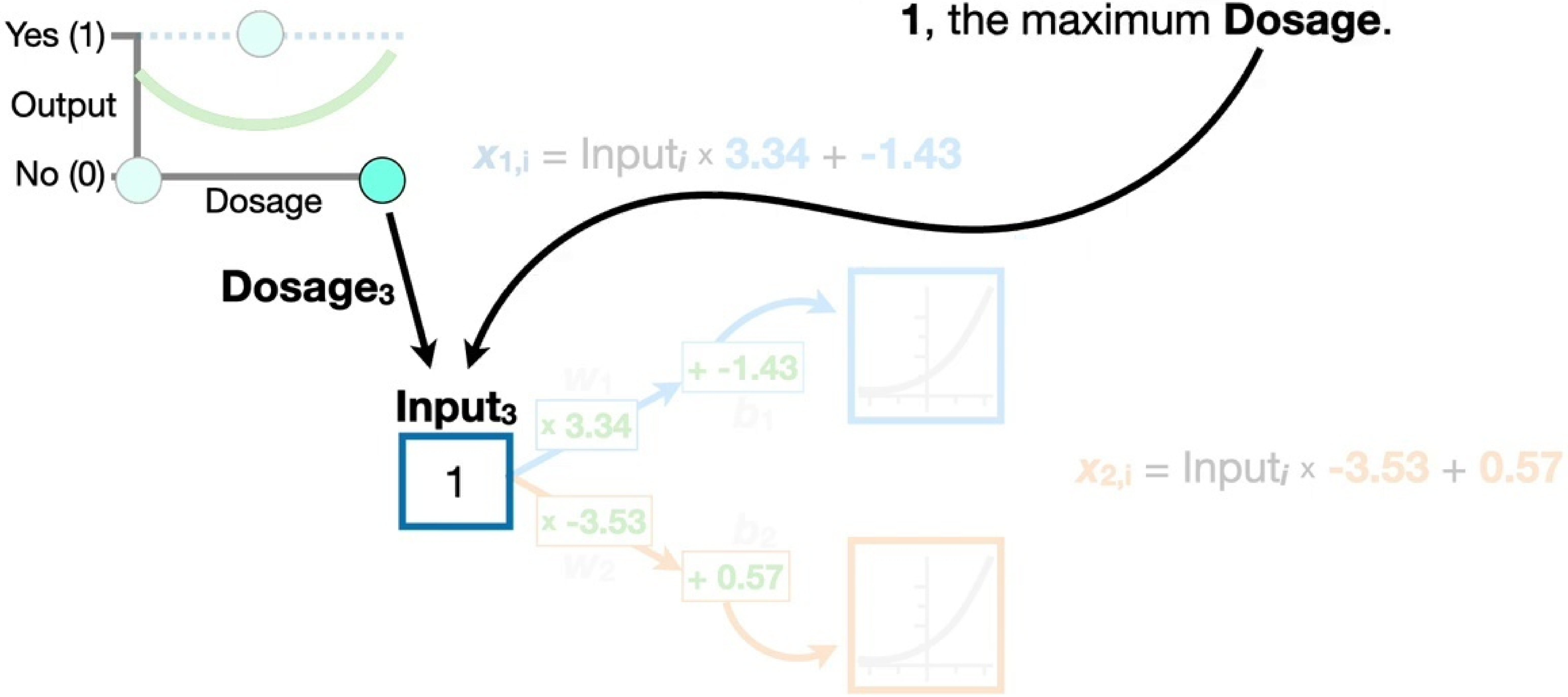


$x_{2,i}$ = Input$_i$ × -3.53 + 0.57

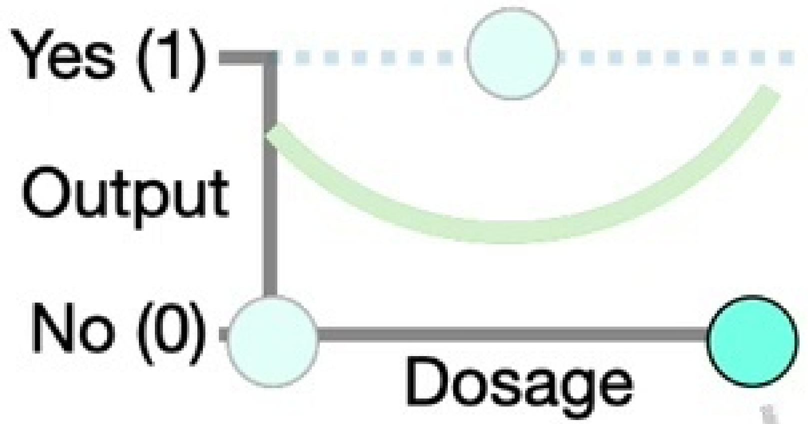
For example if $i = 3$, then we are talking about the third **Dosage, Dosage₃**…



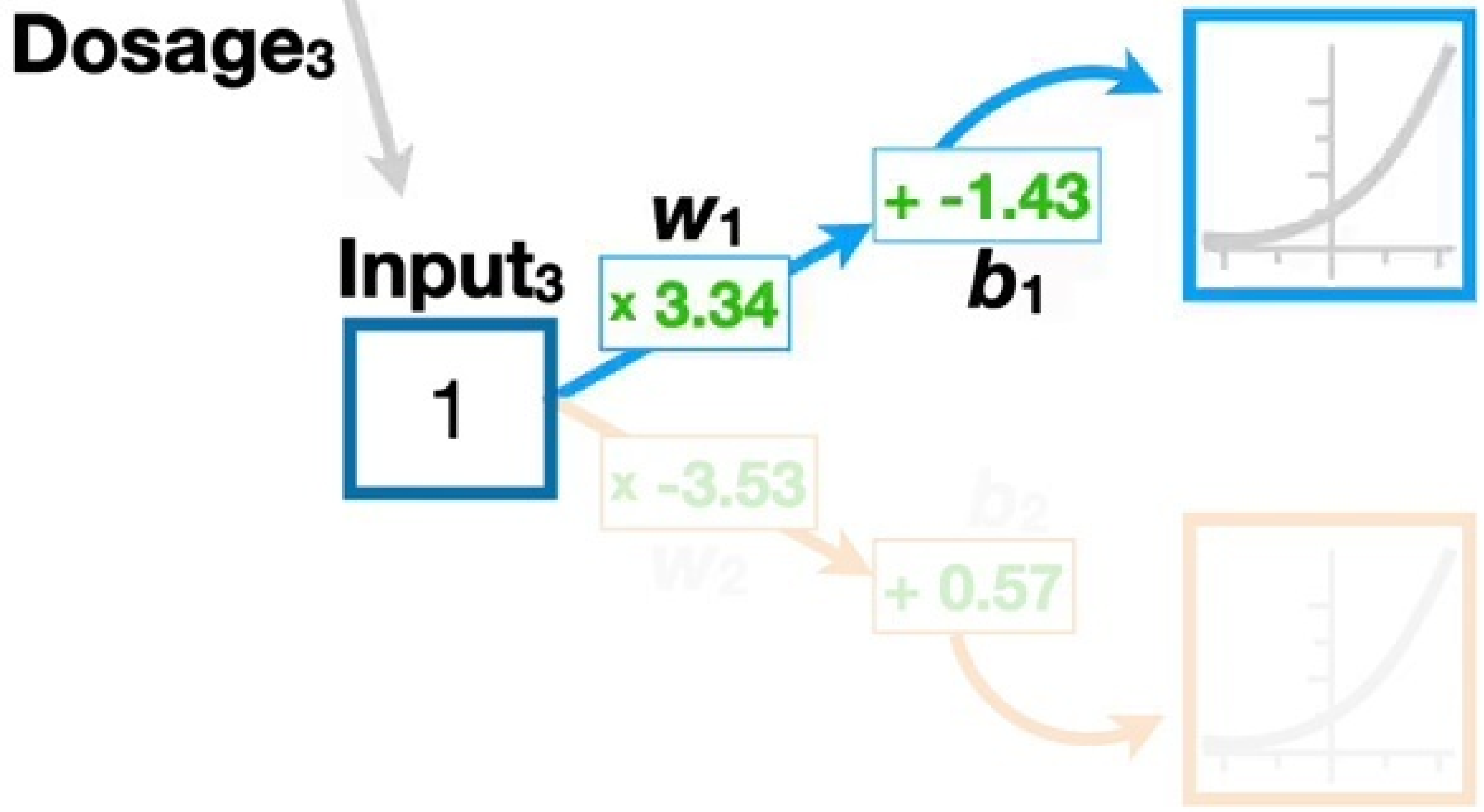
...and that means we're talking about **Input₃**, which is **1**, the maximum **Dosage**.



And that means the x-axis coordinate for the **Activation Function** in the top **Node, $x_{1,3}$,** is equal to…



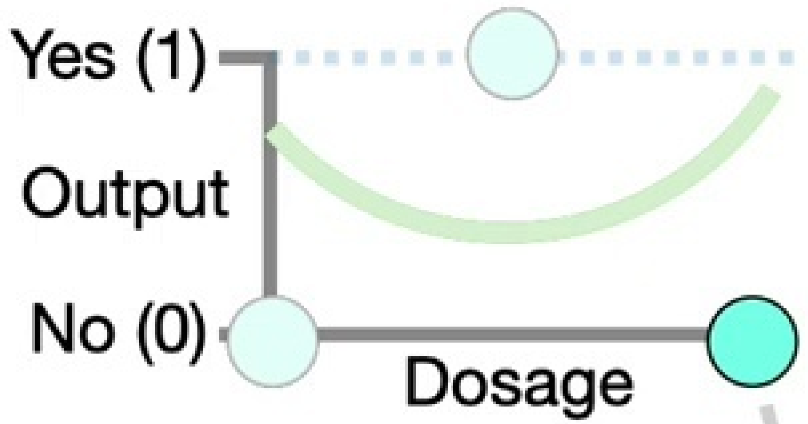
$x_{1,3}$ = Input$_3$ × 3.34 + -1.43

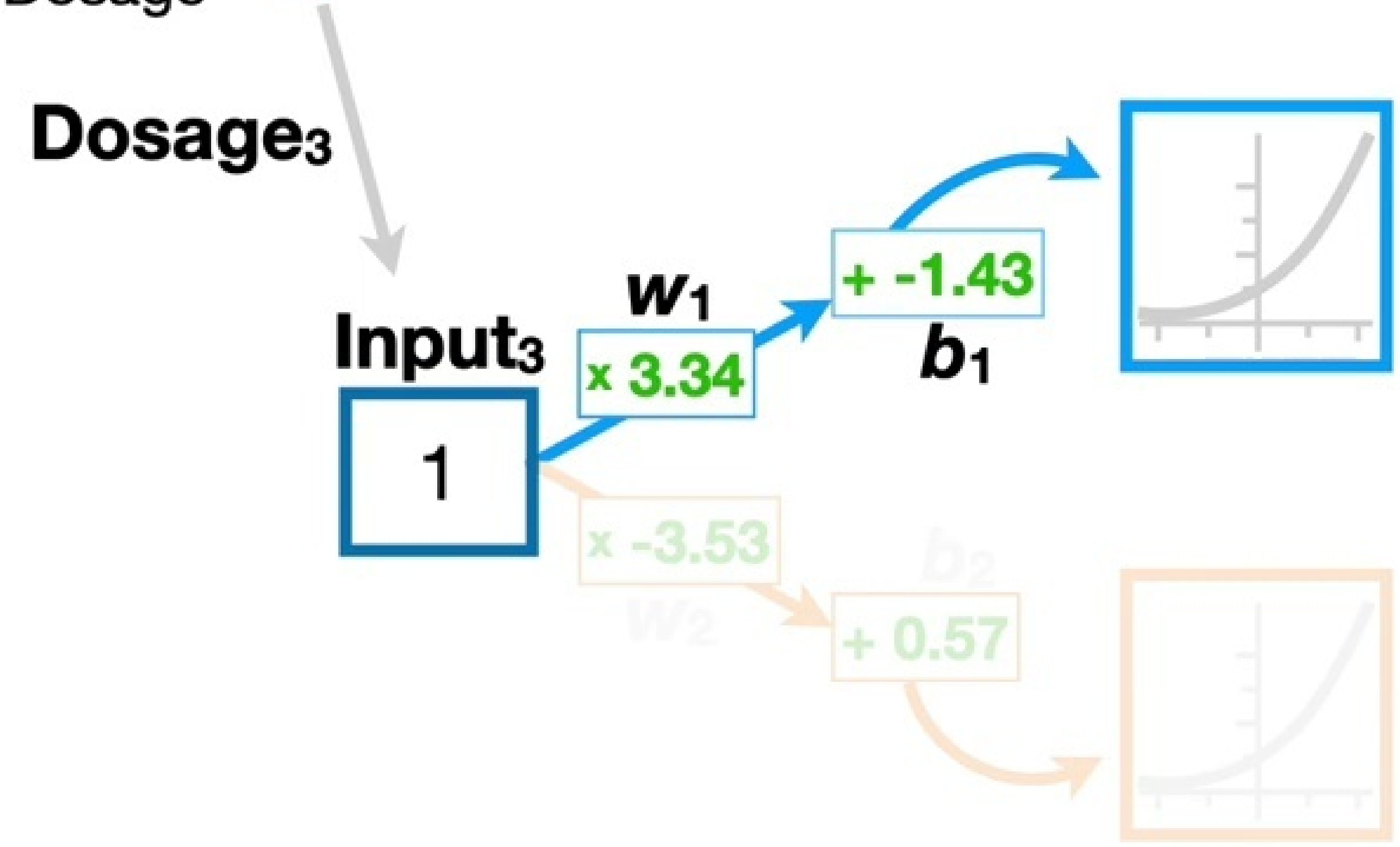


$x_{2,i}$ = Input$_i$ × -3.53 + 0.57

And that means the x-axis coordinate for the **Activation Function** in the top **Node, $x_{1,3}$,** is equal to…

$$x_{1,3} = \text{Input}_3 \times 3.34 + -1.43$$

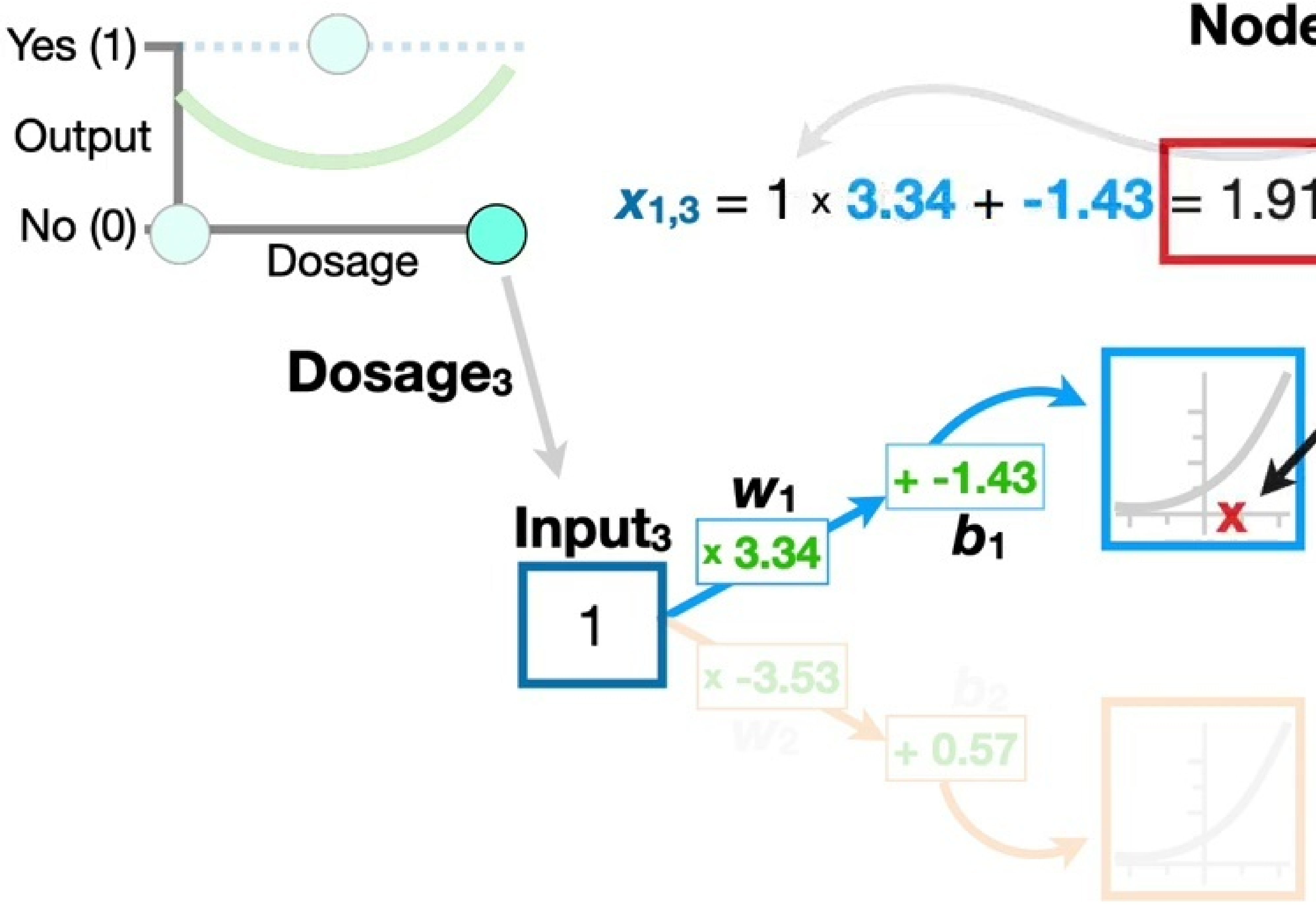




$$x_{2,i} = \text{Input}_i \times -3.53 + 0.57$$

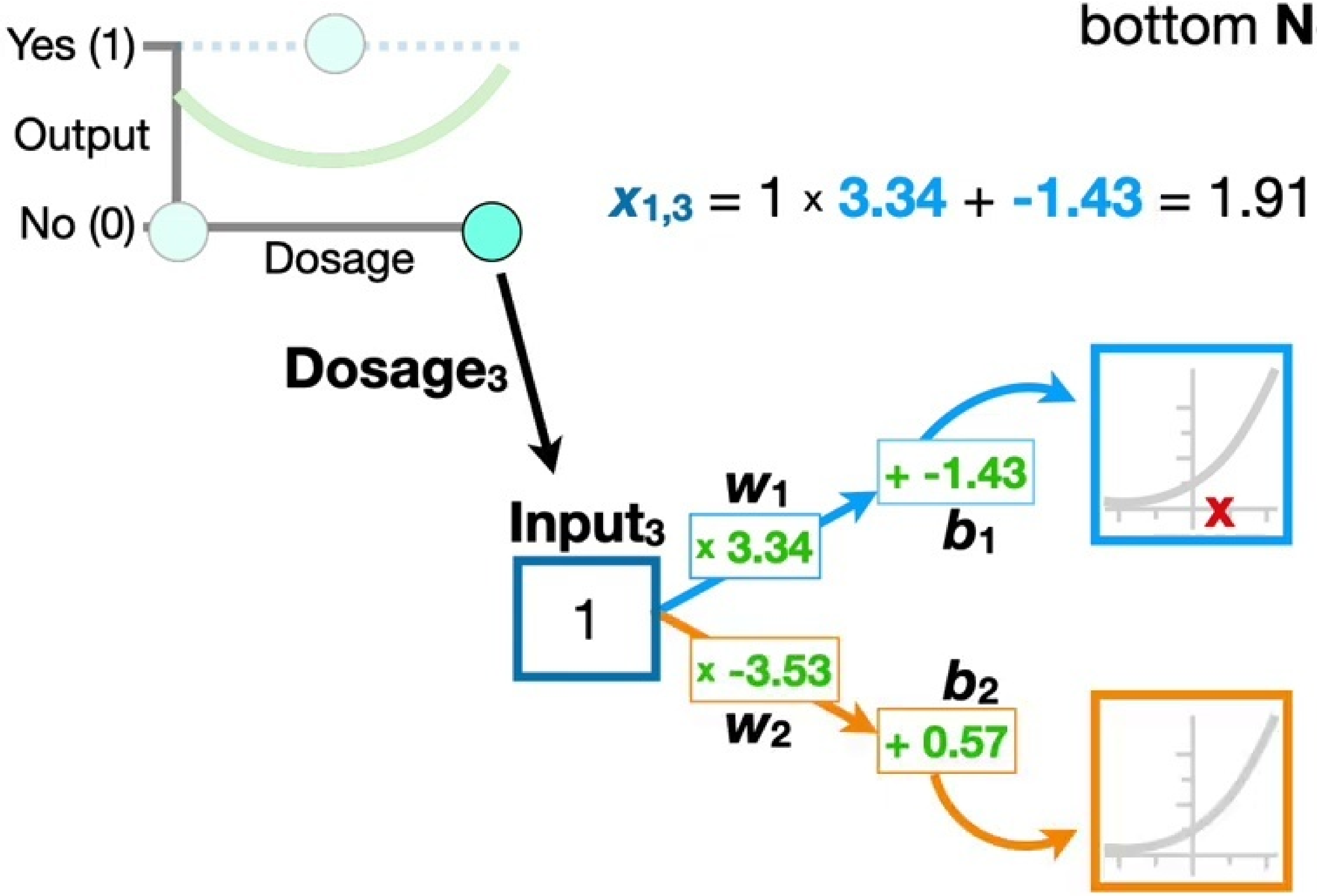
And that means the x-axis coordinate for the **Activation Function** in the top **Node, $x_{1,3}$,** is equal to…

$$x_{1,3} = 1 \times 3.34 + -1.43 \boxed{= 1.91}$$



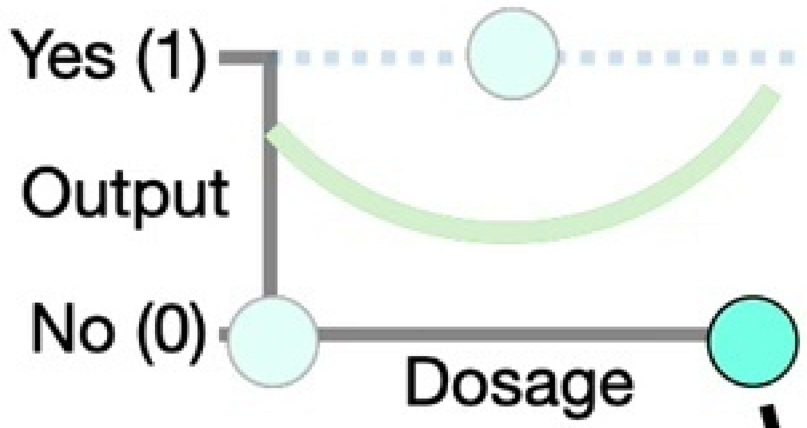
$x_{2,i}$ = Input$_i$ × -3.53 + 0.57

And the x-axis coordinate for the **Activation Function** in the bottom **Node**, $x_{2,3}$, is equal to…

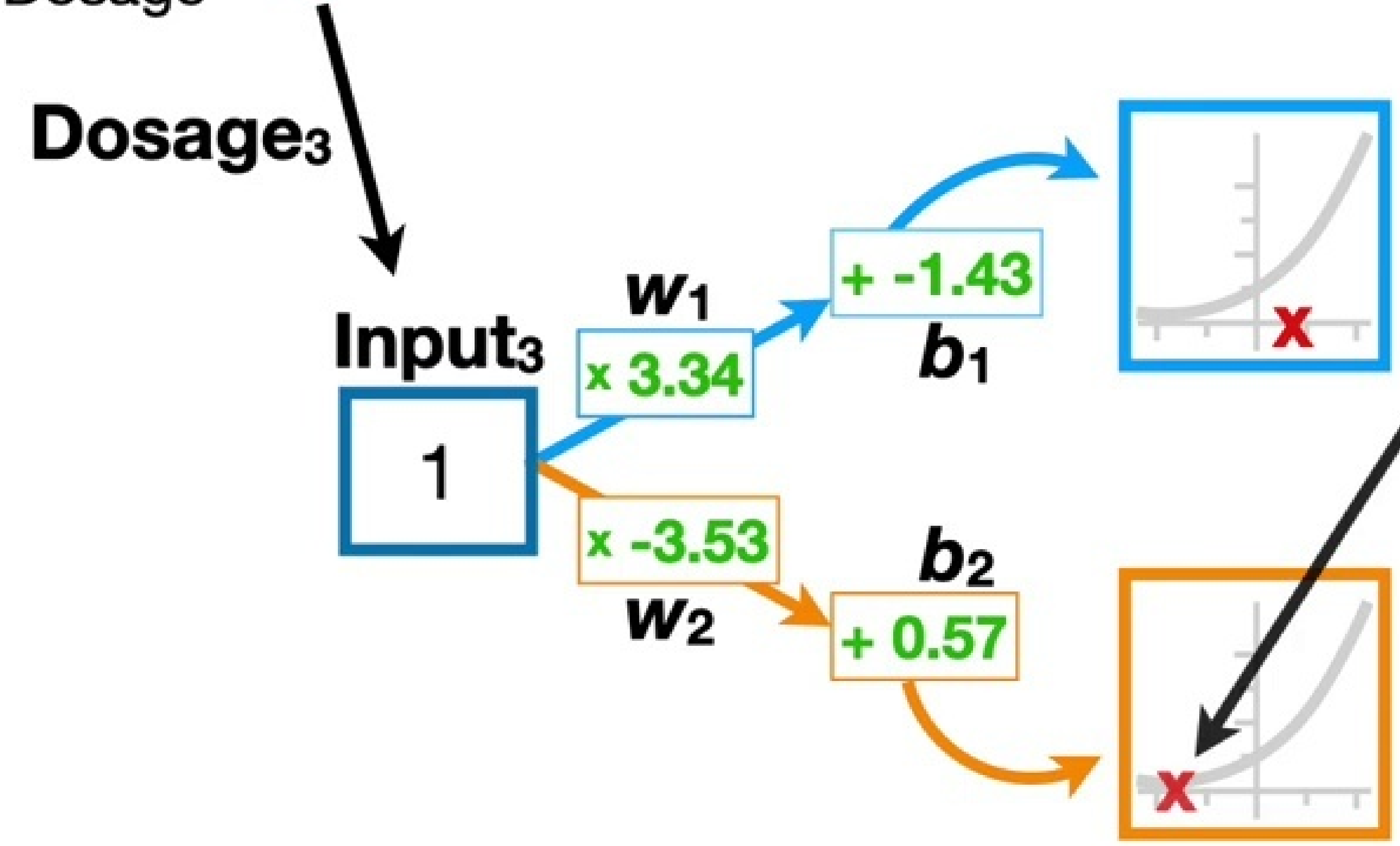


$x_{2,3} = \text{Input}_i \times -3.53 + 0.57$

And the x-axis coordinate for the **Activation Function** in the bottom **Node**, $x_{2,3}$, is equal to…



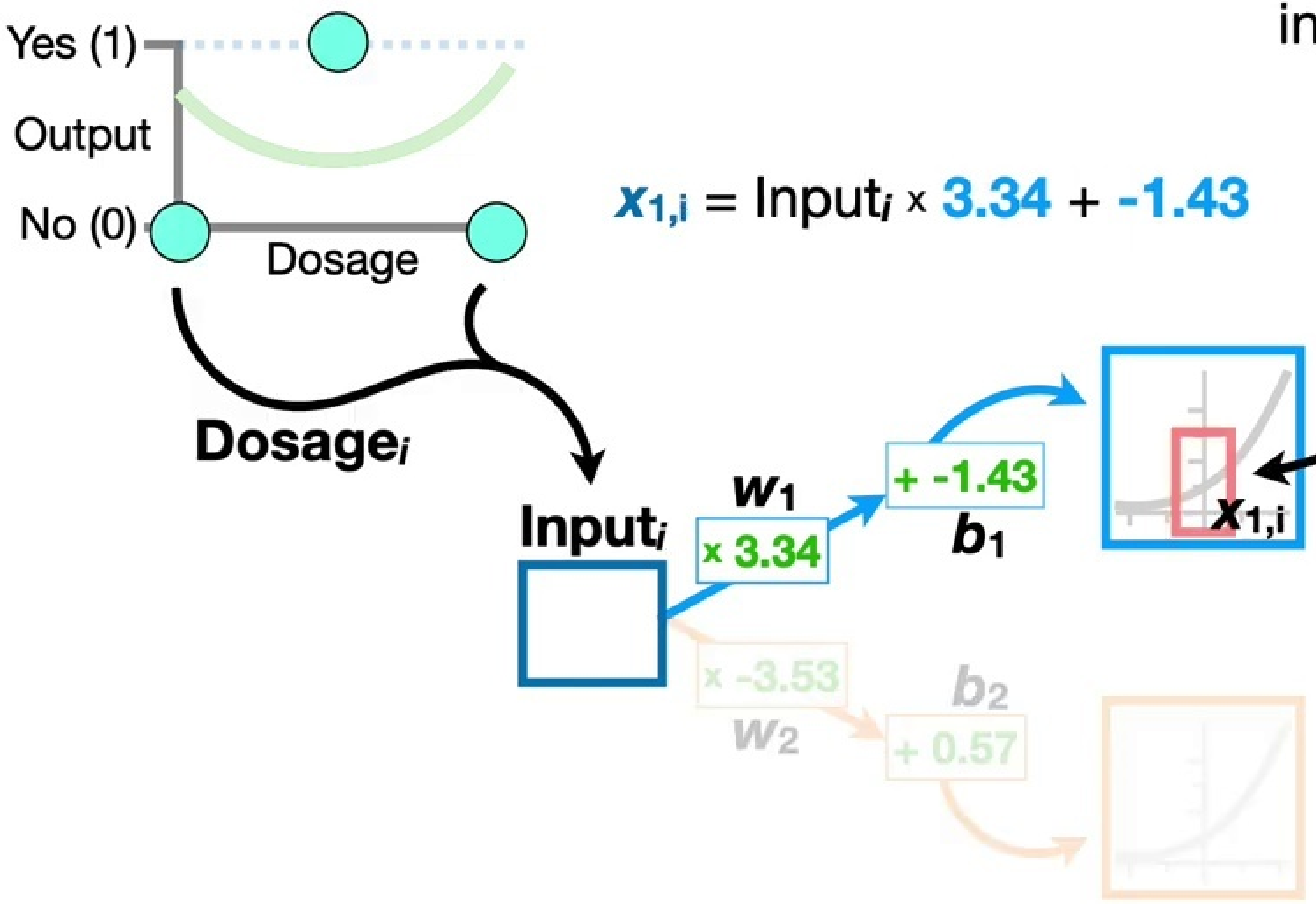
$$x_{1,3} = 1 \times 3.34 + -1.43 = 1.91$$



$$x_{2,3} = 1 \times -3.53 + 0.57 \boxed{= -2.96}$$

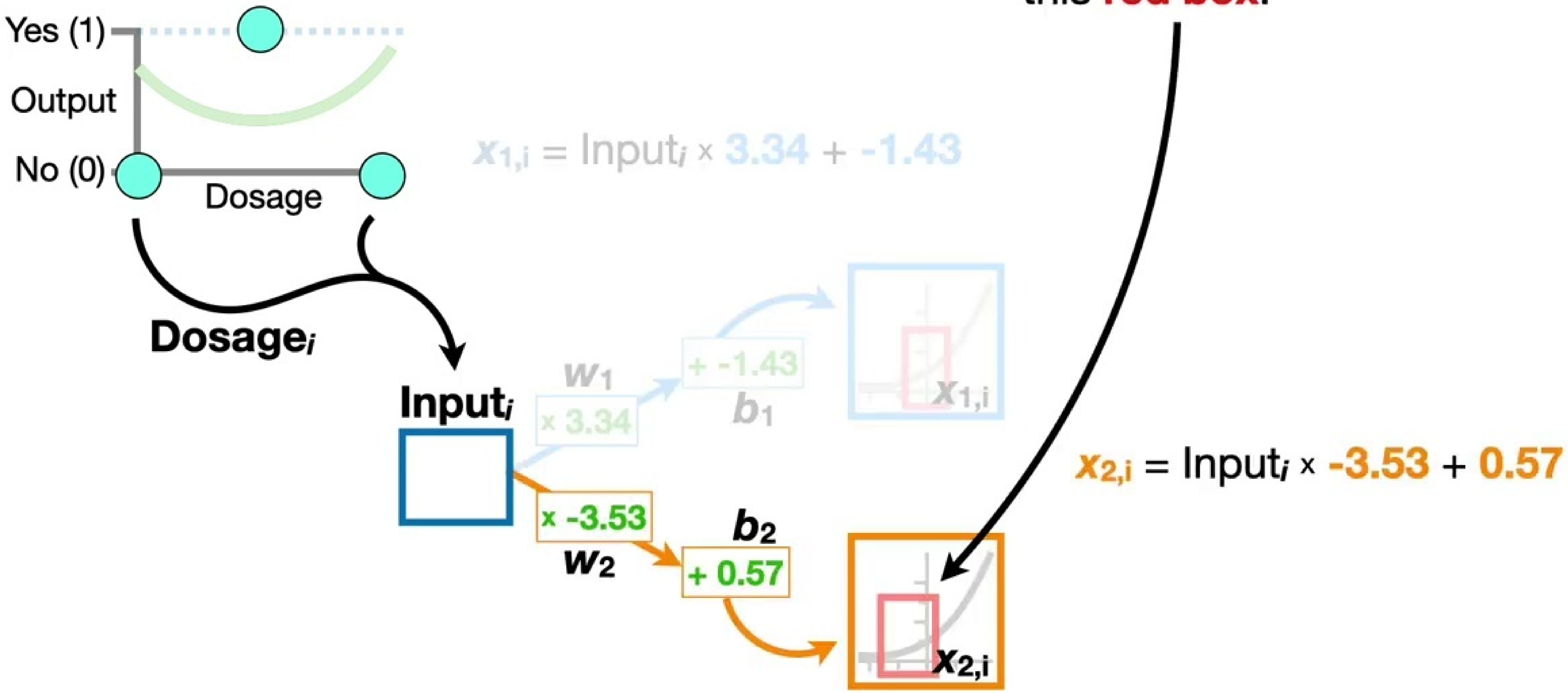
If we plugged in all values for ***i*** into **Dosage*ᵢ***, we get $\mathbf{x_{1,i}}$ values in this **red box**…

$$x_{1,i} = \text{Input}_i \times 3.34 + -1.43$$

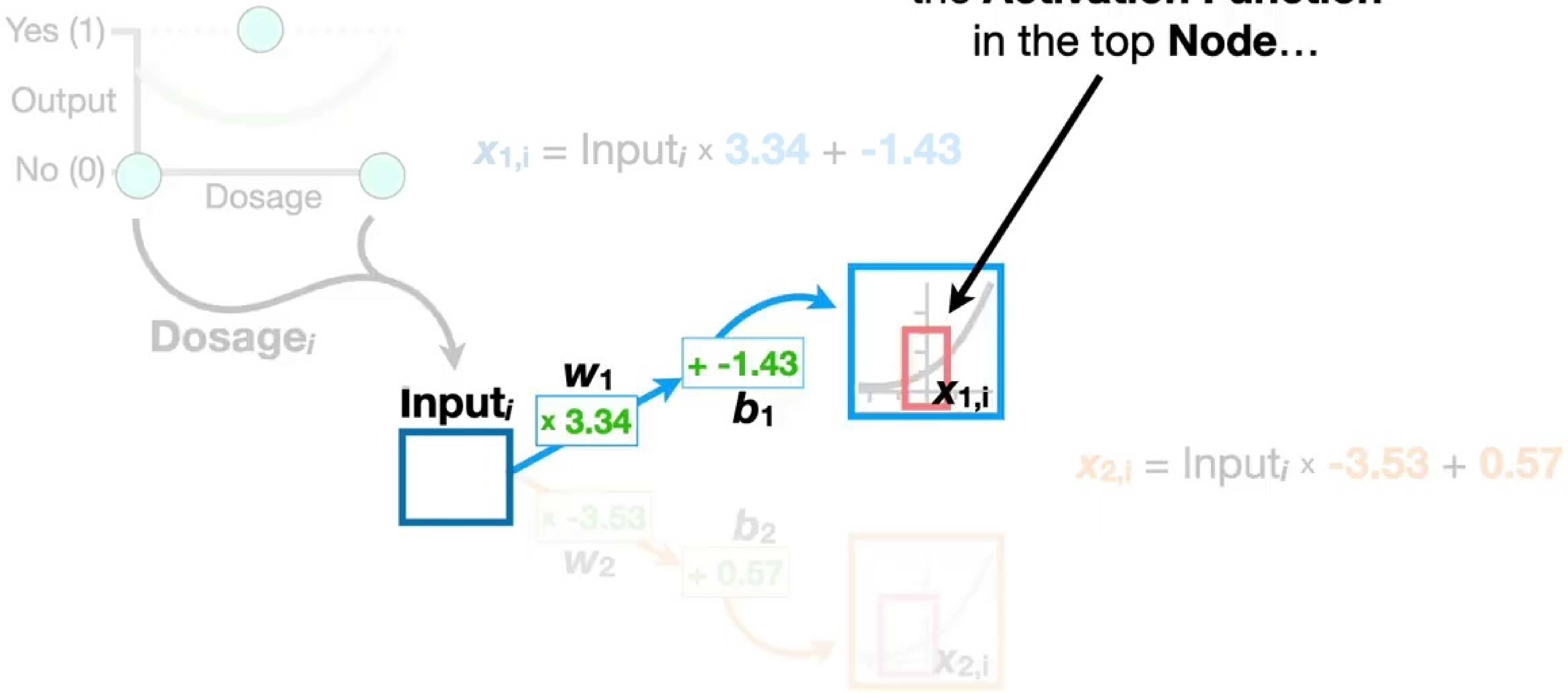


$$x_{2,i} = \text{Input}_i \times -3.53 + 0.57$$

…and $\mathbf{x_{2,i}}$ values in this **red box**.



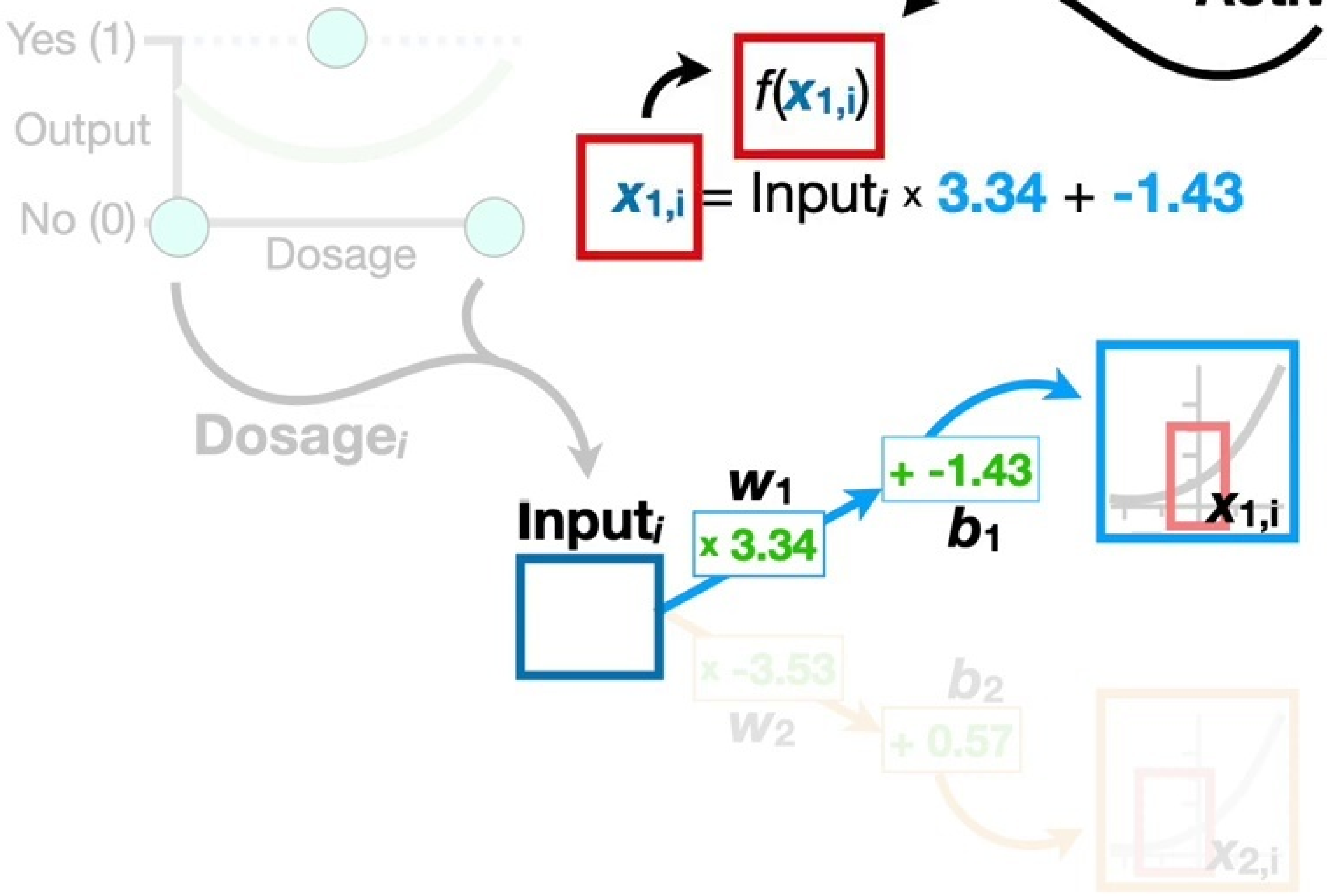
Now, in order to get the y-axis coordinates for the **Activation Function** in the top **Node**…



…we plug $x_{1,i}$ into the **Activation Function**…

$f(x_{1,i})$

$x_{1,i}$ = Input$_i$ × 3.34 + -1.43

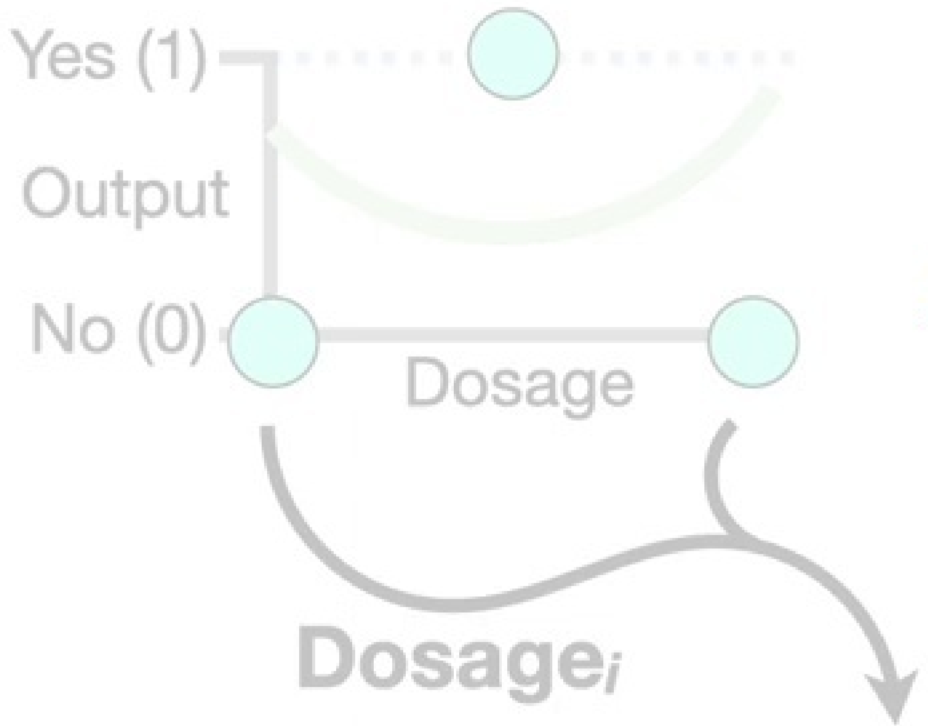


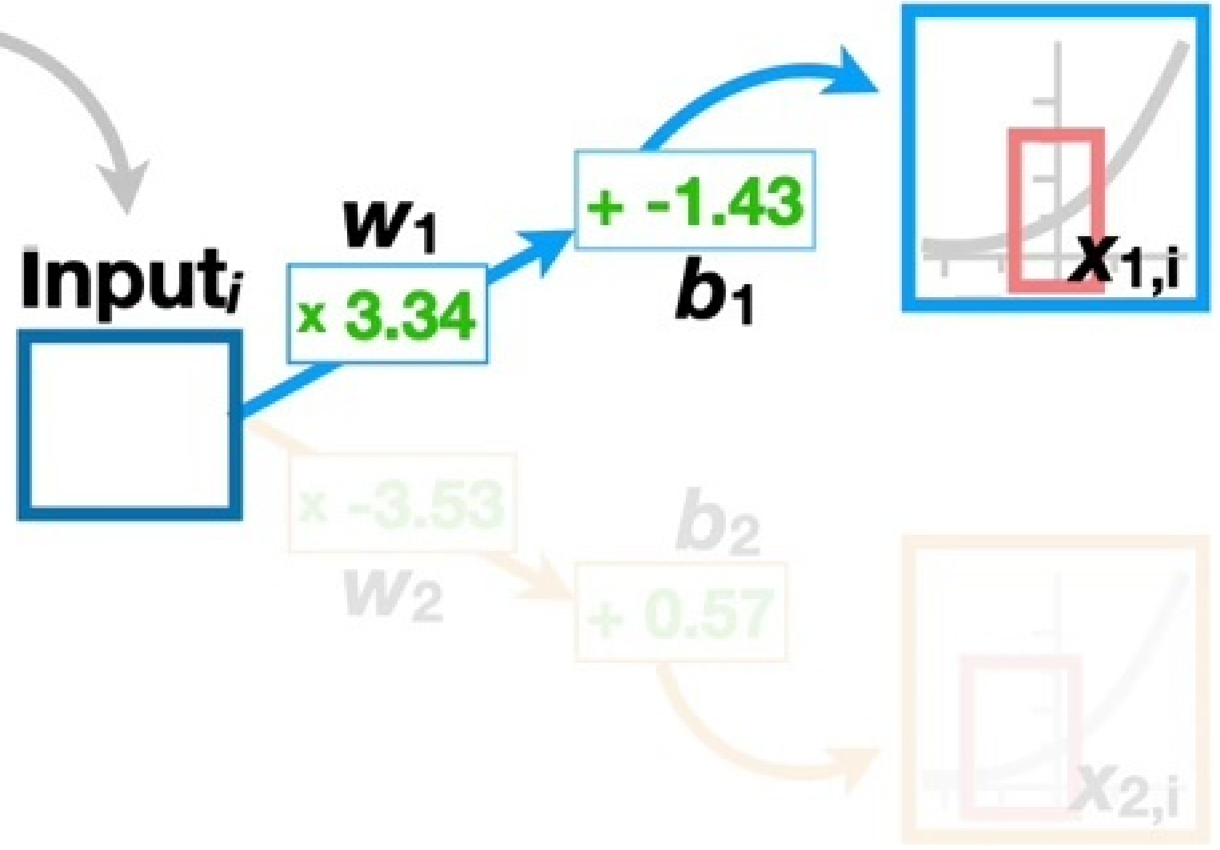
$x_{2,i}$ = Input$_i$ × -3.53 + 0.57

…which, in this example, is the **softplus** function…

$$f(x_{1,i}) = \log(1 + e^{x})$$

$$x_{1,i} = \text{Input}_i \times 3.34 + -1.43$$



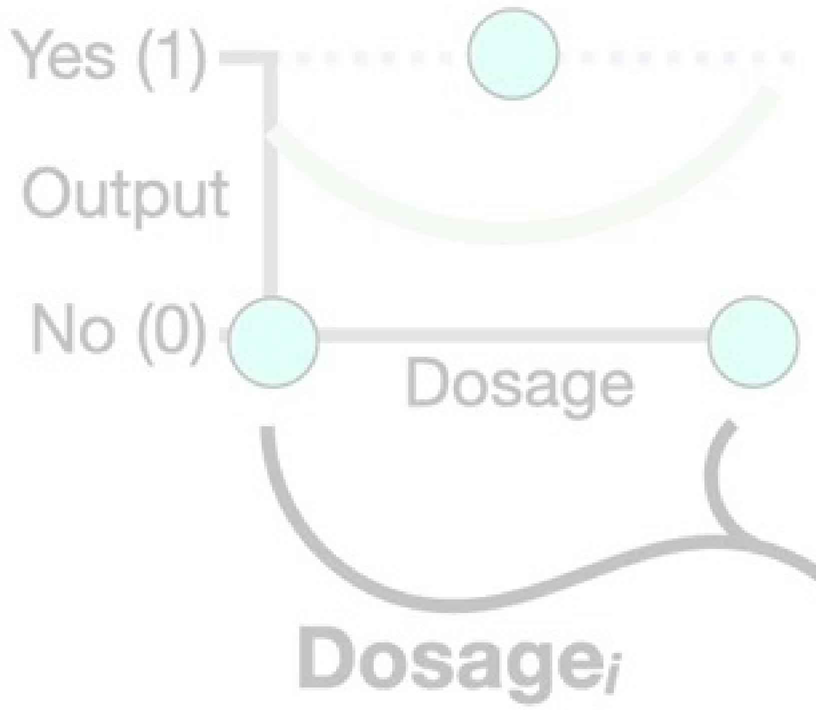


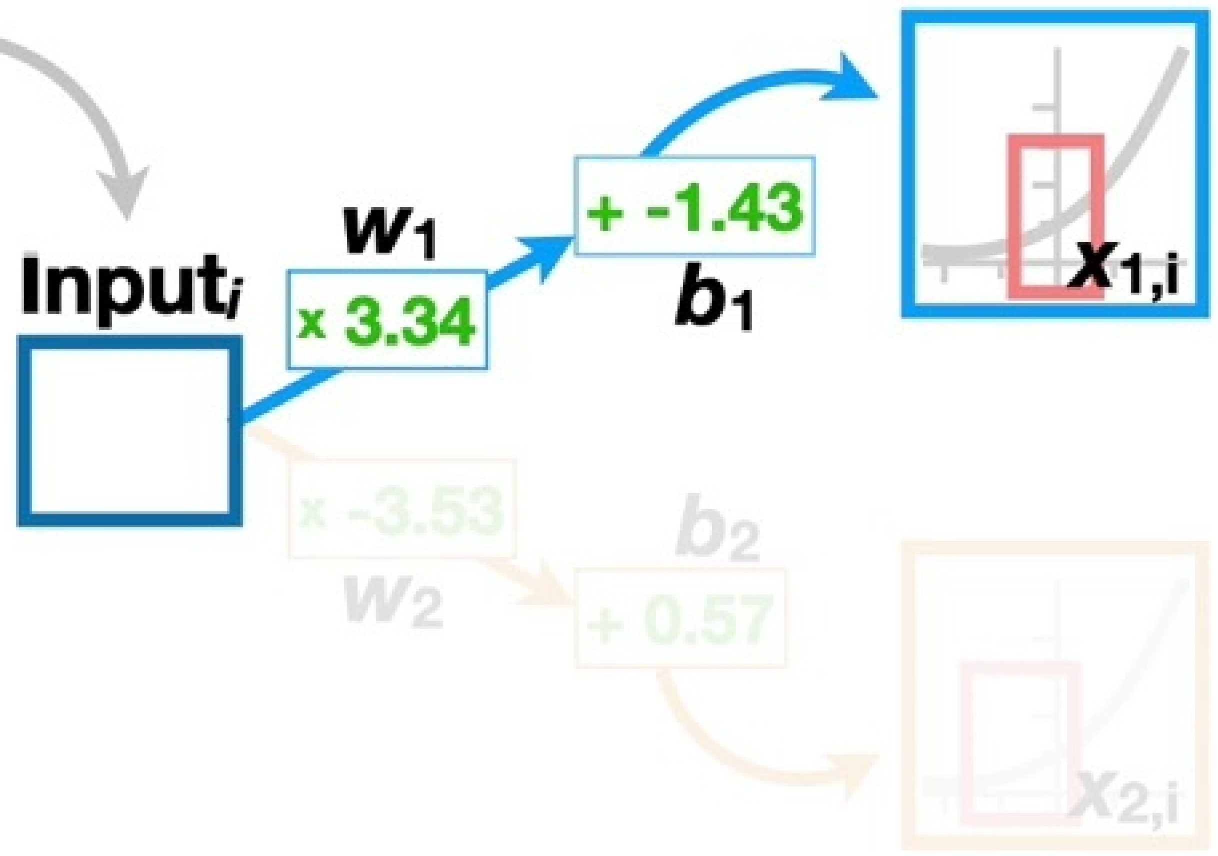
$$x_{2,i} = \text{Input}_i \times -3.53 + 0.57$$

…and that gives us $\boldsymbol{y_{1,i}}$.

$$y_{1,i} = f(x_{1,i}) = \log(1 + e^x)$$

$$x_{1,i} = \text{Input}_i \times 3.34 + -1.43$$

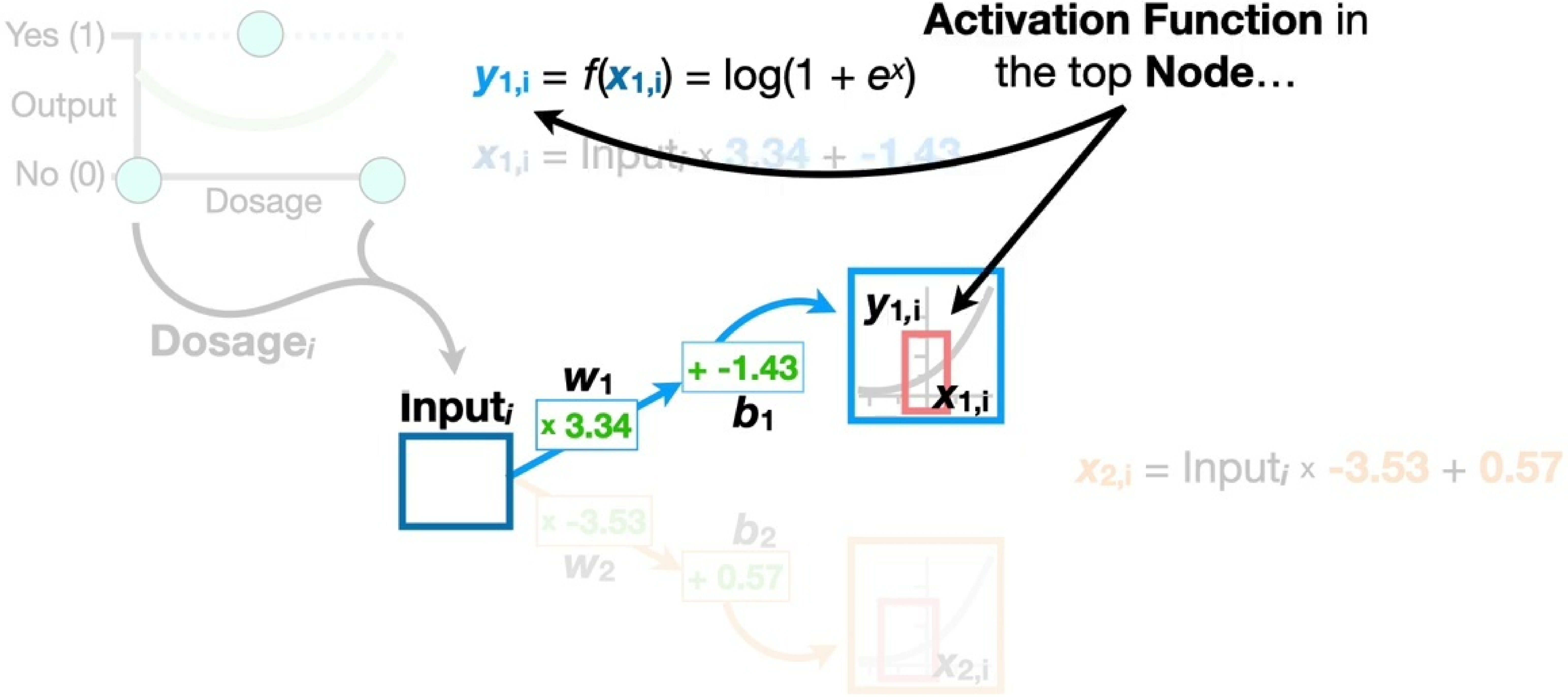




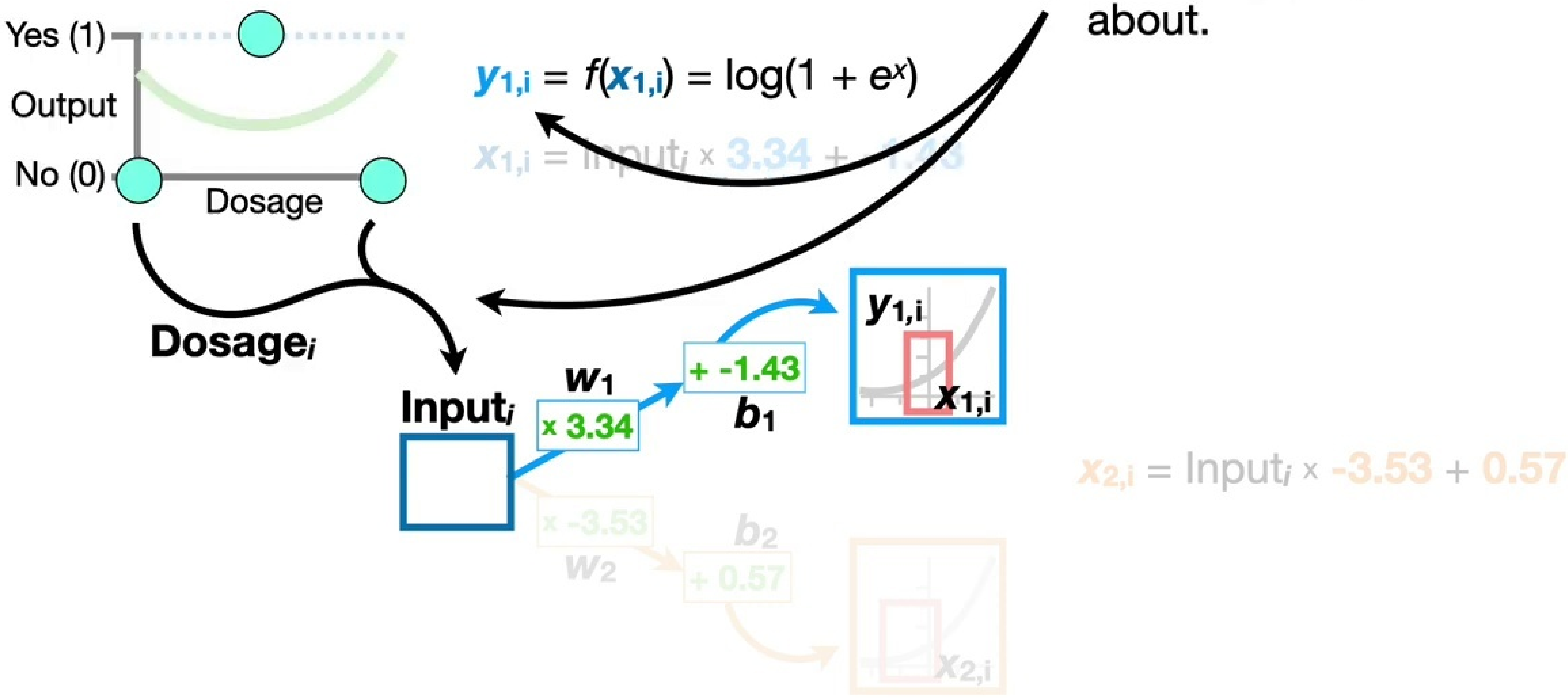
$$x_{2,i} = \text{Input}_i \times -3.53 + 0.57$$

Just like before, the **1** in **1,i** tells us that we are talking about the **Activation Function** in the top **Node**…

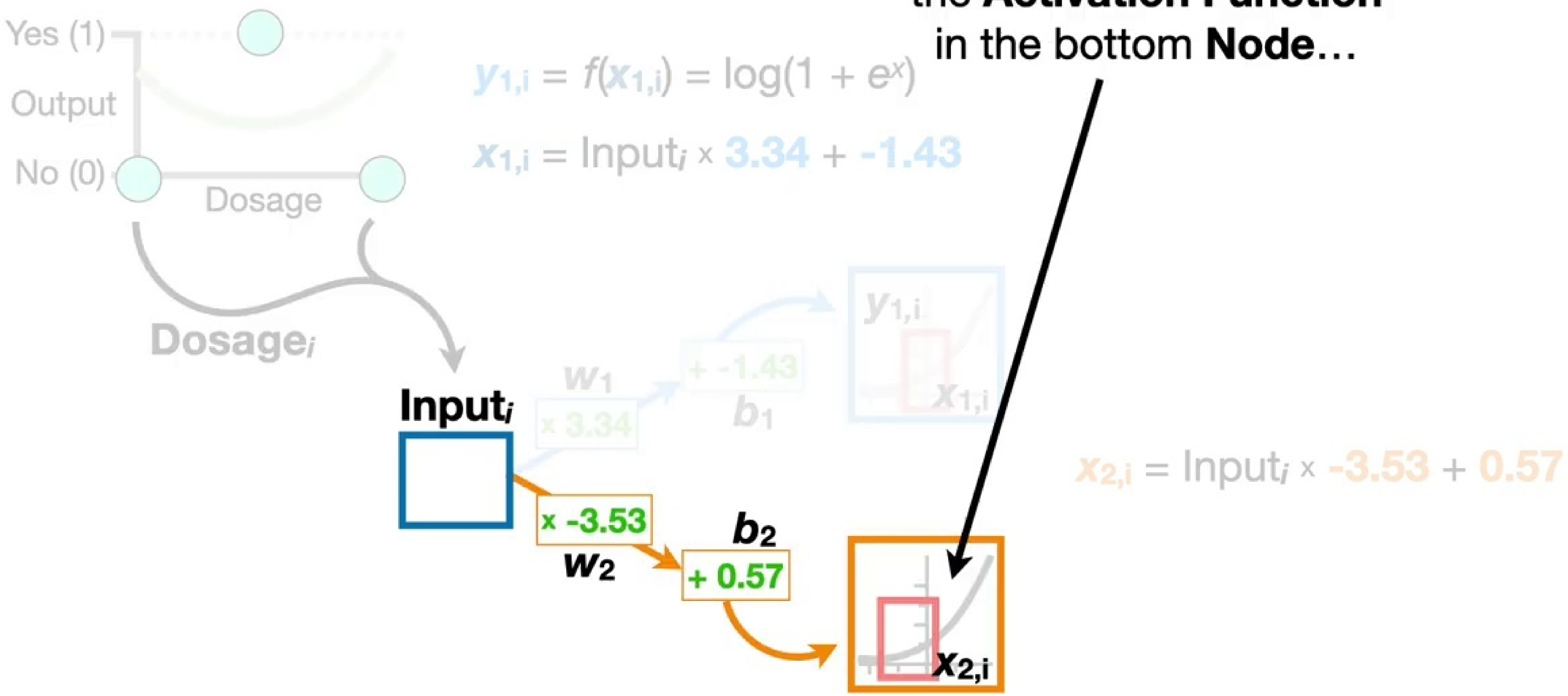
$$y_{1,i} = f(x_{1,i}) = \log(1 + e^x)$$



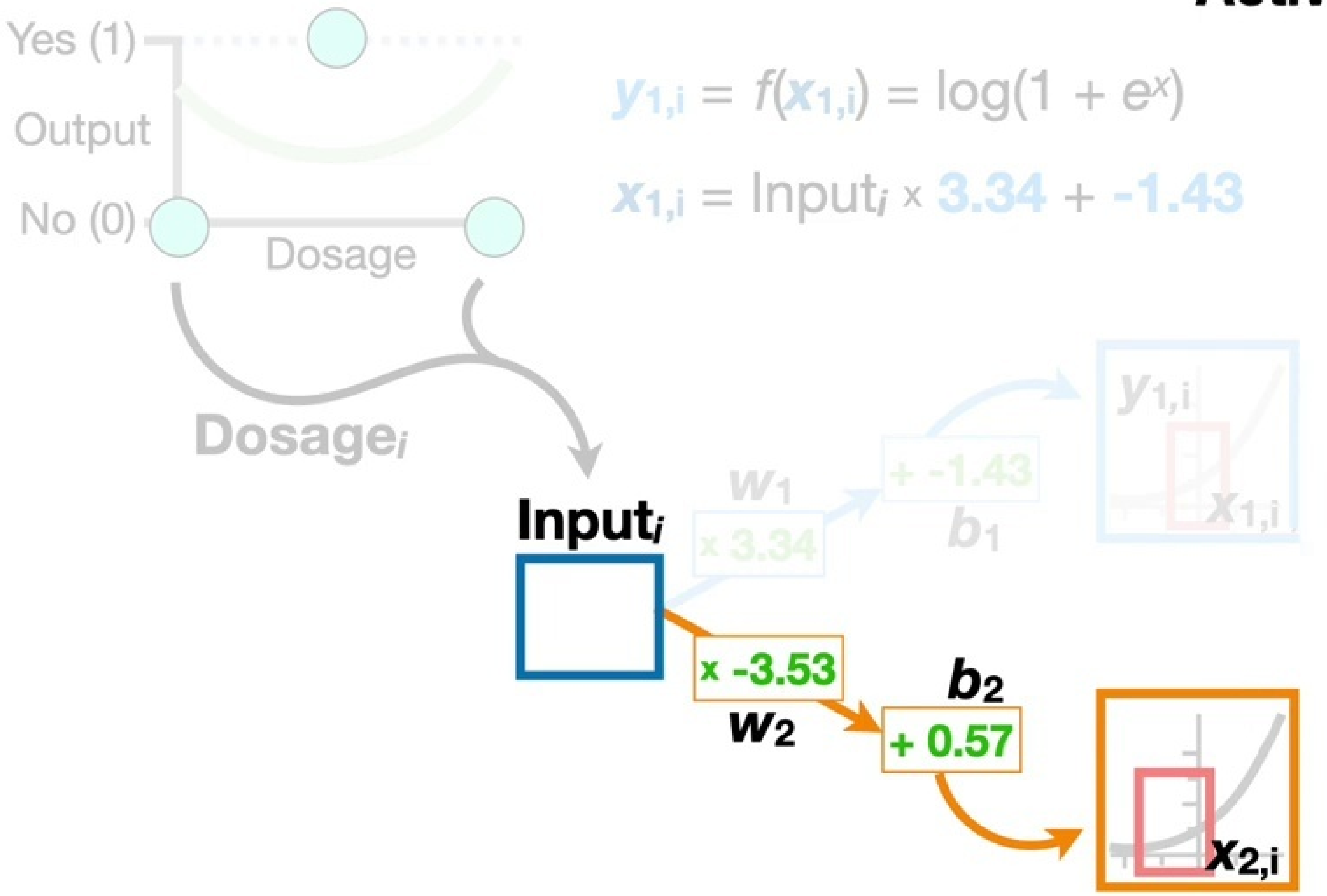
…and the ***i*** tells us which **Dosage** we are talking about.



Likewise, in order to get the y-axis coordinates for the **Activation Function** in the bottom **Node**…



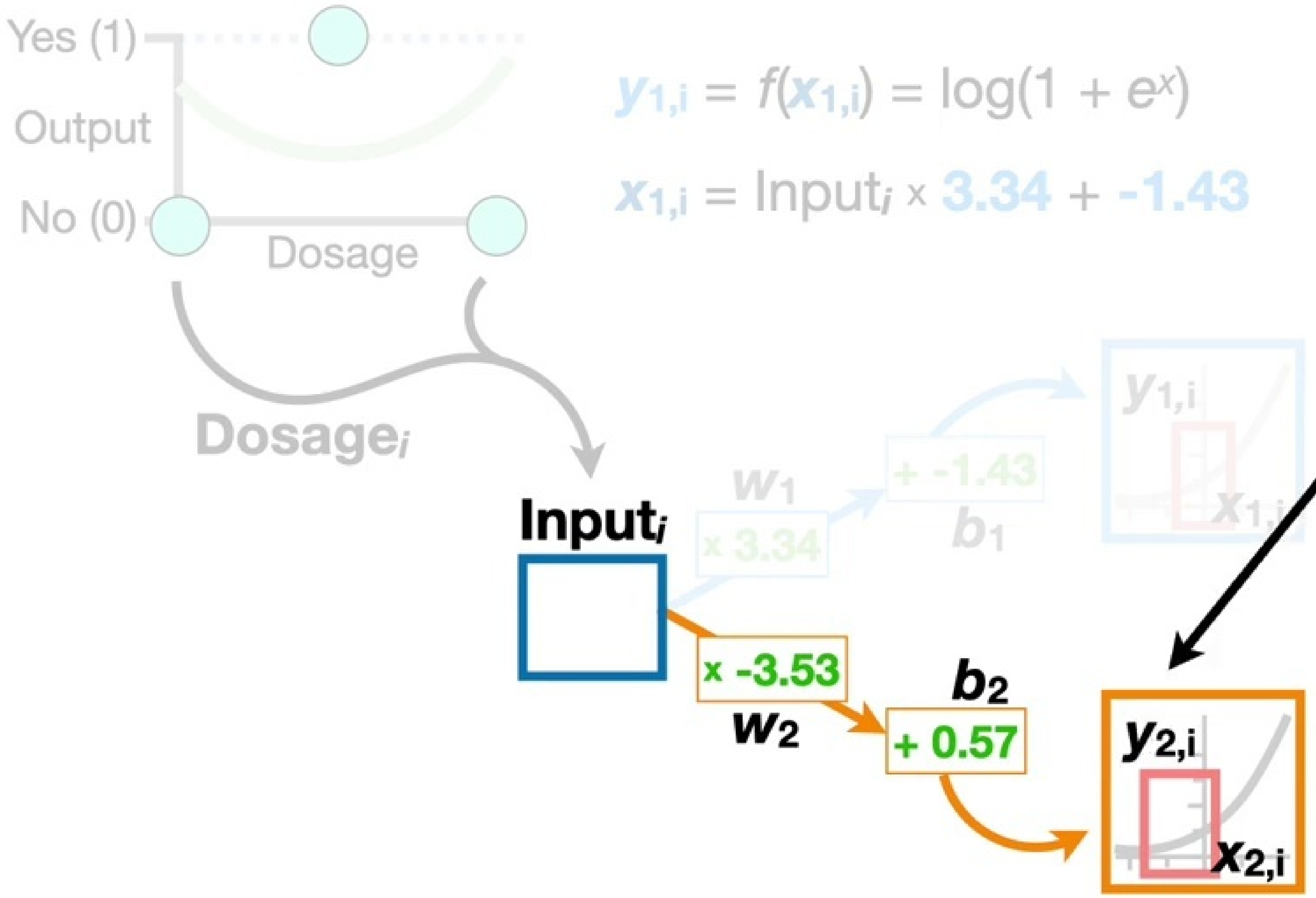
…we plug $\boldsymbol{x_{2,i}}$ into the **Activation Function…**



$$f(x_{2,i}) = \log(1 + e^x)$$

$$x_{2,i} = \text{Input}_i \times -3.53 + 0.57$$

…and that gives us $y_{2,i}$.



$y_{1,i} = f(x_{1,i}) = \log(1 + e^x)$

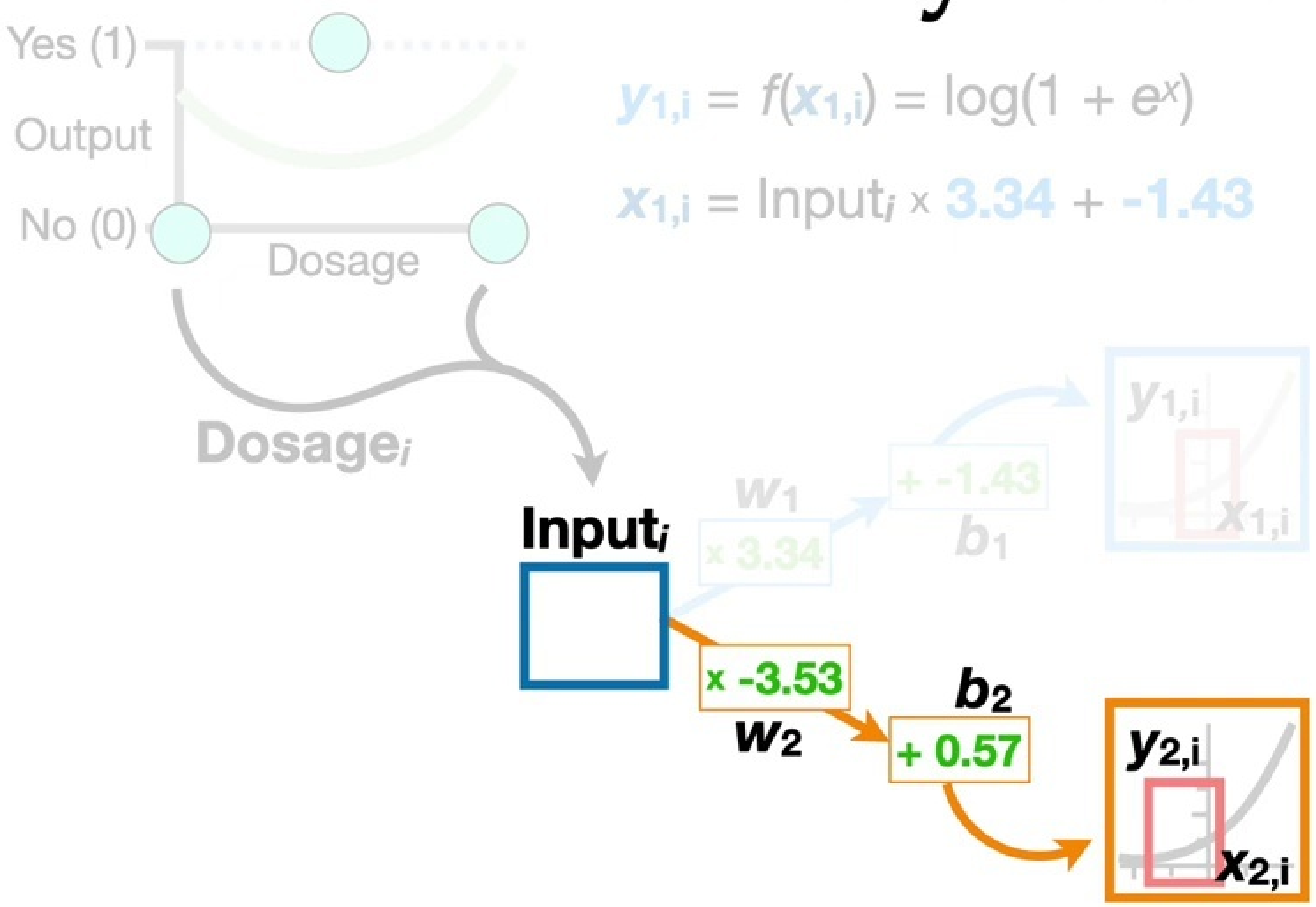
$x_{1,i} = \text{Input}_i \times 3.34 + -1.43$

$y_{2,i} = f(x_{2,i}) = \log(1 + e^x)$

$x_{2,i} = \text{Input}_i \times -3.53 + 0.57$

Now that we understand the

Fancy Notation…



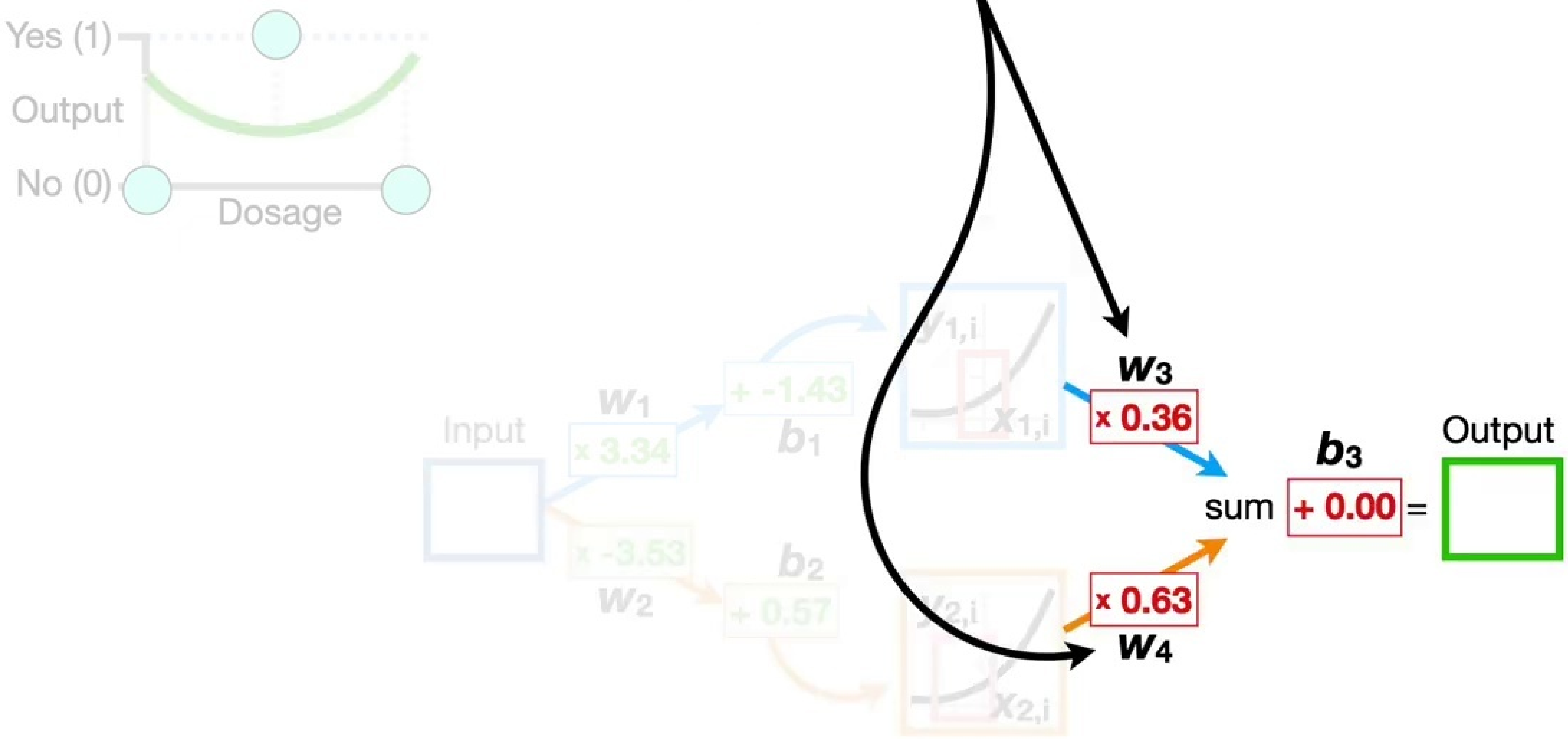
$y_{1,i} = f(x_{1,i}) = \log(1 + e^x)$

$x_{1,i} = \text{Input}_i \times 3.34 + -1.43$

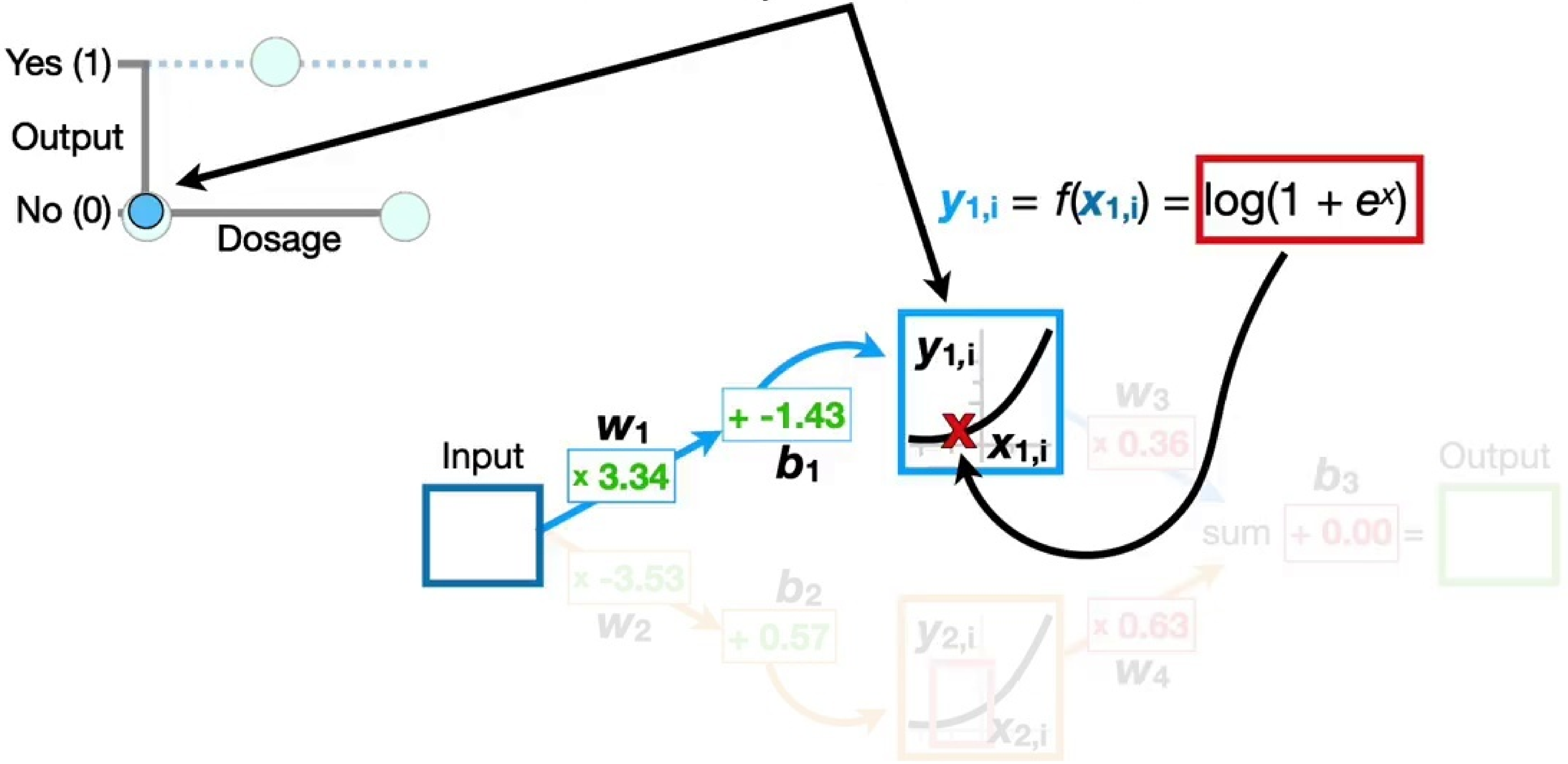
$y_{2,i} = f(x_{2,i}) = \log(1 + e^x)$

$x_{2,i} = \text{Input}_i \times -3.53 + 0.57$

…we can talk about how to calculate the derivatives of the **SSR** with respect to the **Weights** $\boldsymbol{w_3}$ and $\boldsymbol{w_4}$.

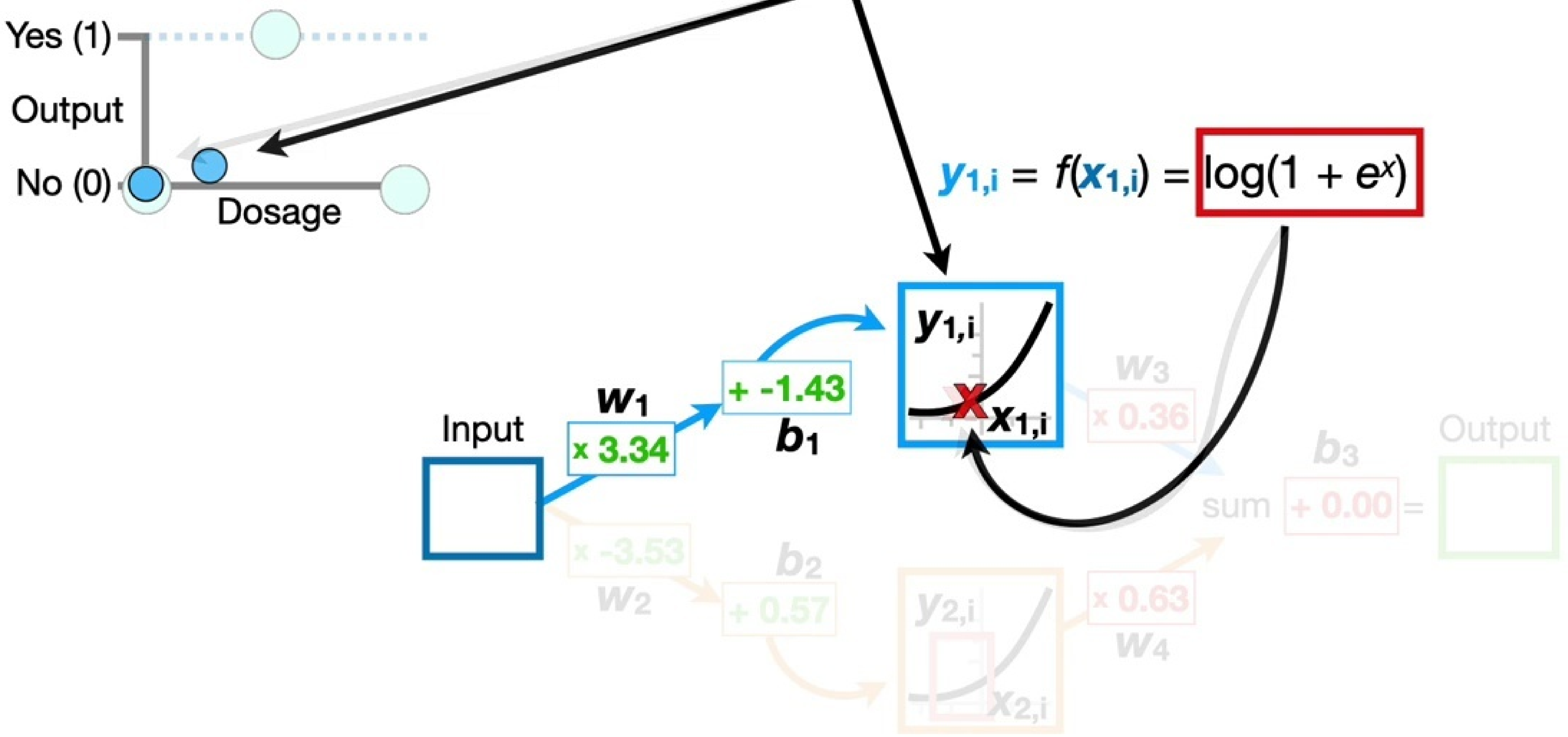


First, remember that $\boldsymbol{y}_{1,i}$ represents the y-axis coordinates for the top **Activation Function**…

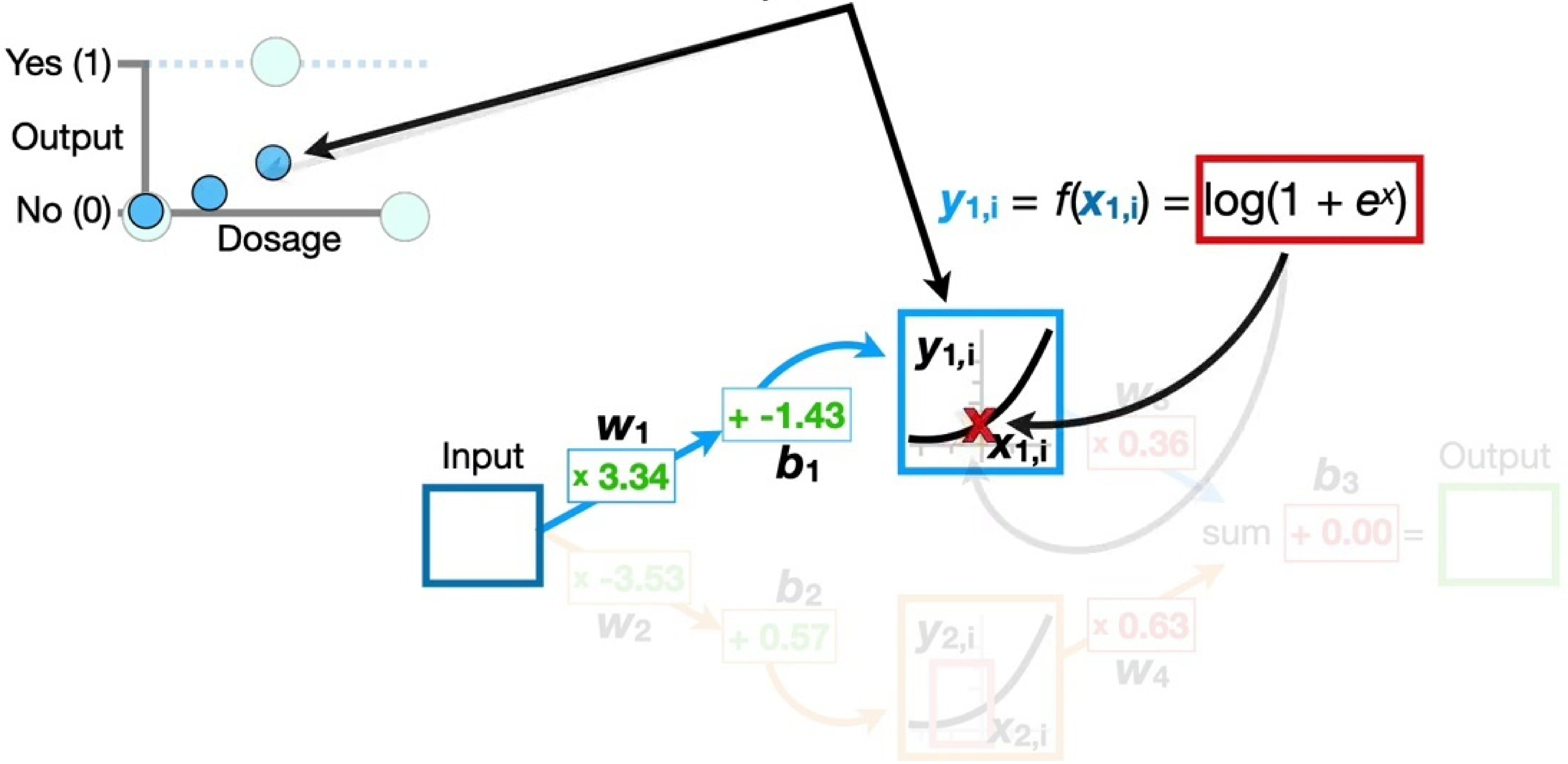


First, remember that $y_{1,i}$ represents the y-axis coordinates for the top **Activation Function**…

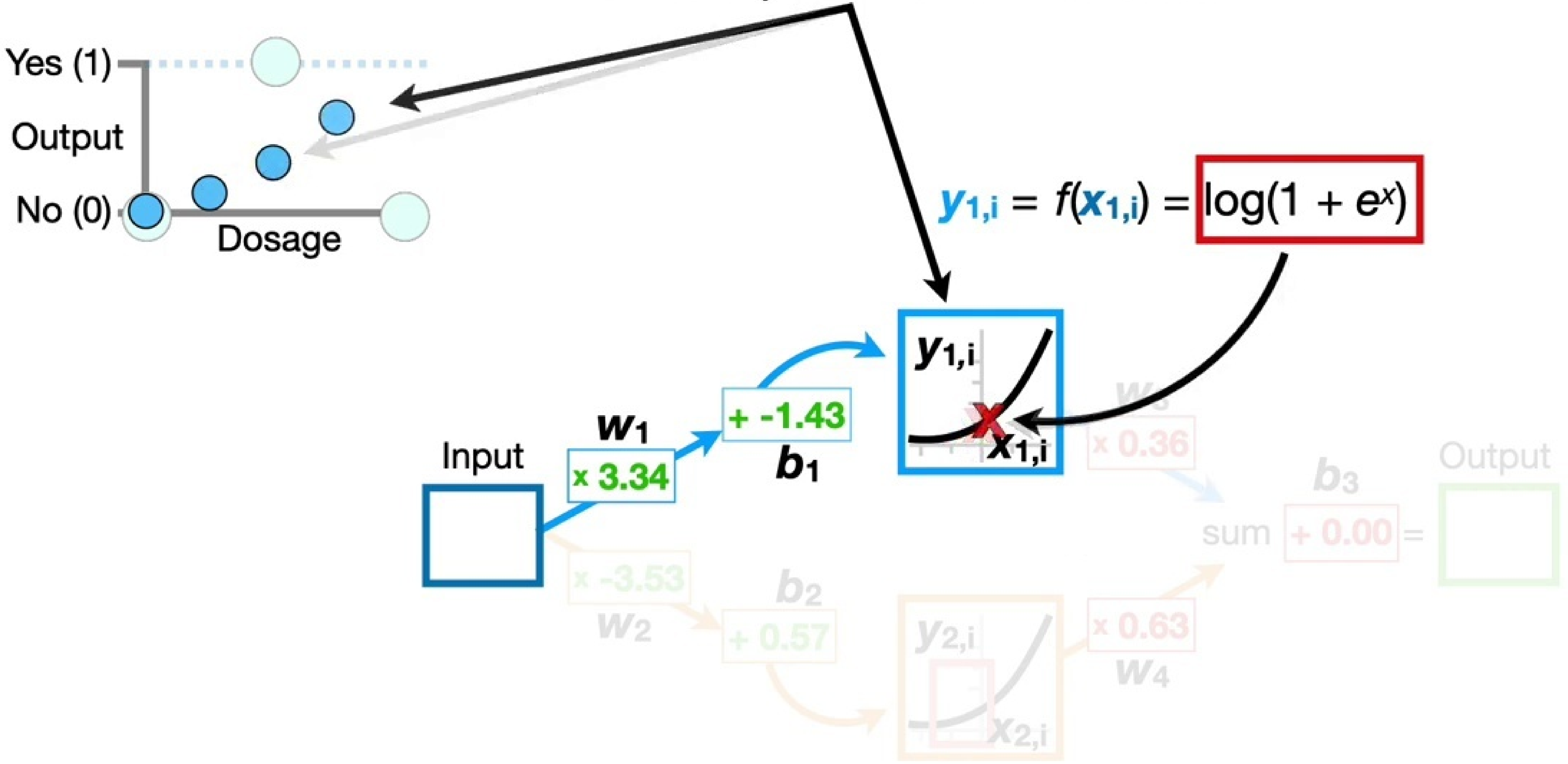
$$y_{1,i} = f(x_{1,i}) = \boxed{\log(1 + e^x)}$$



First, remember that $\boldsymbol{y_{1,i}}$ represents the y-axis coordinates for the top **Activation Function**…

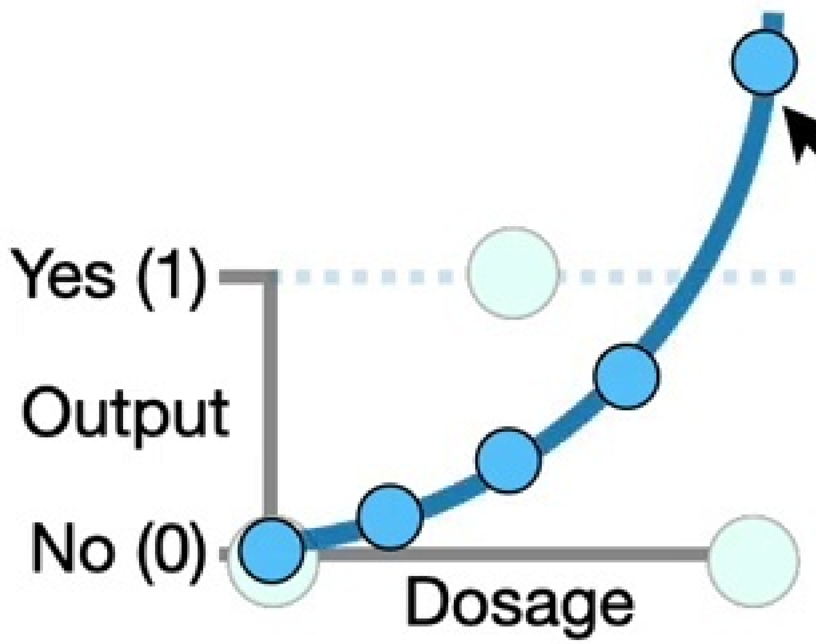


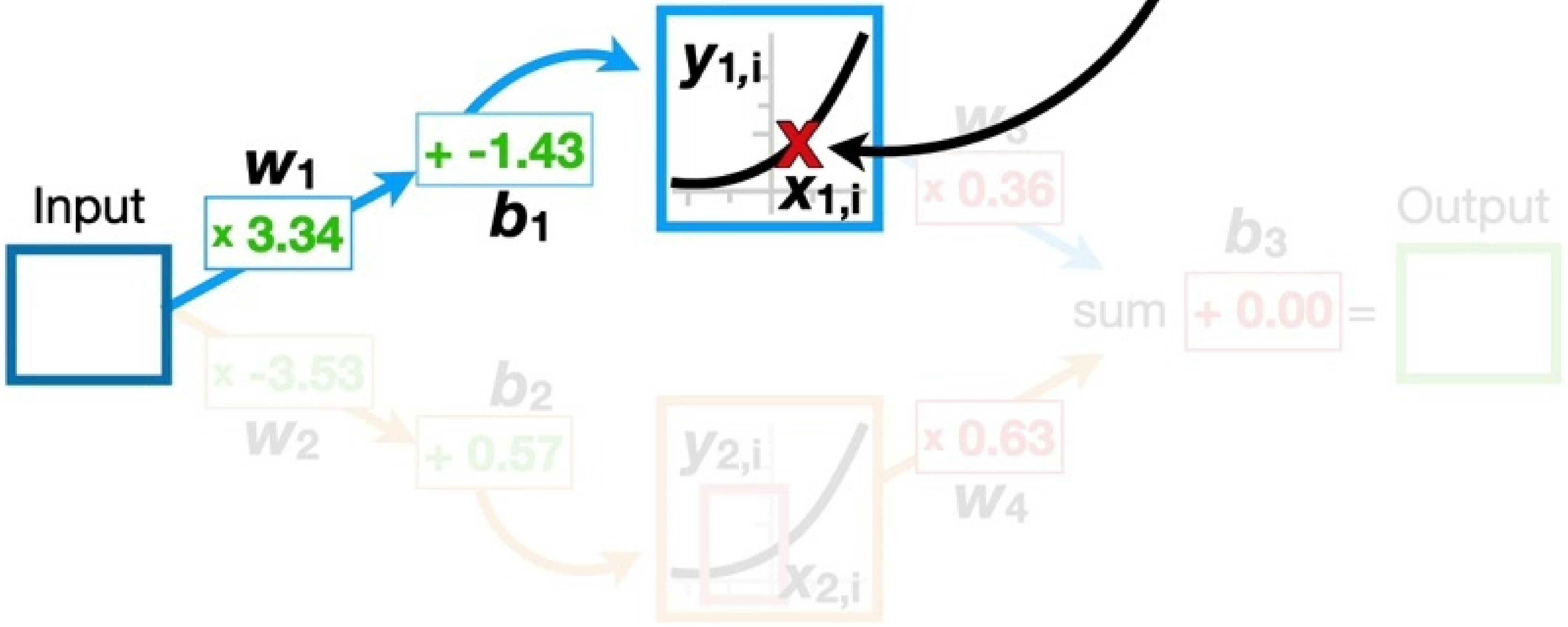
First, remember that $\boldsymbol{y_{1,i}}$ represents the y-axis coordinates for the top **Activation Function**…



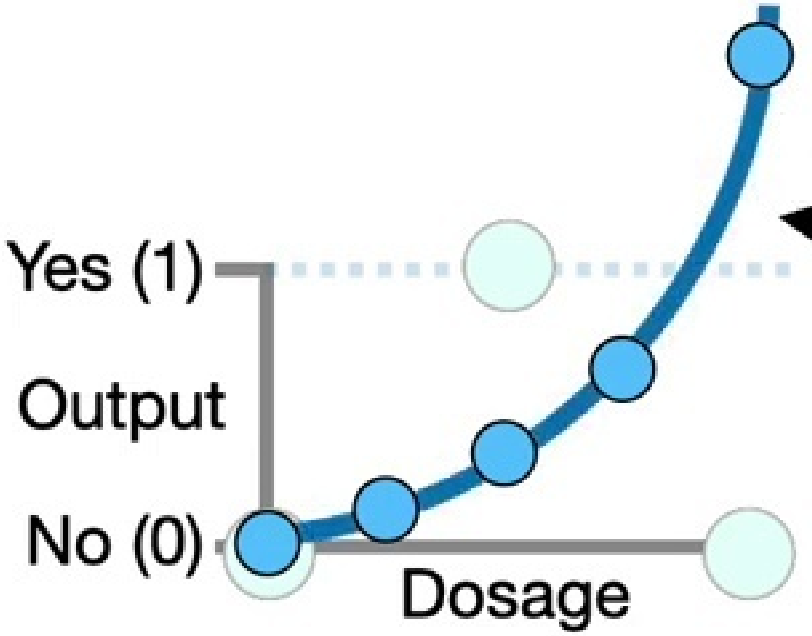
First, remember that $y_{1,i}$ represents the y-axis coordinates for the top **Activation Function**…

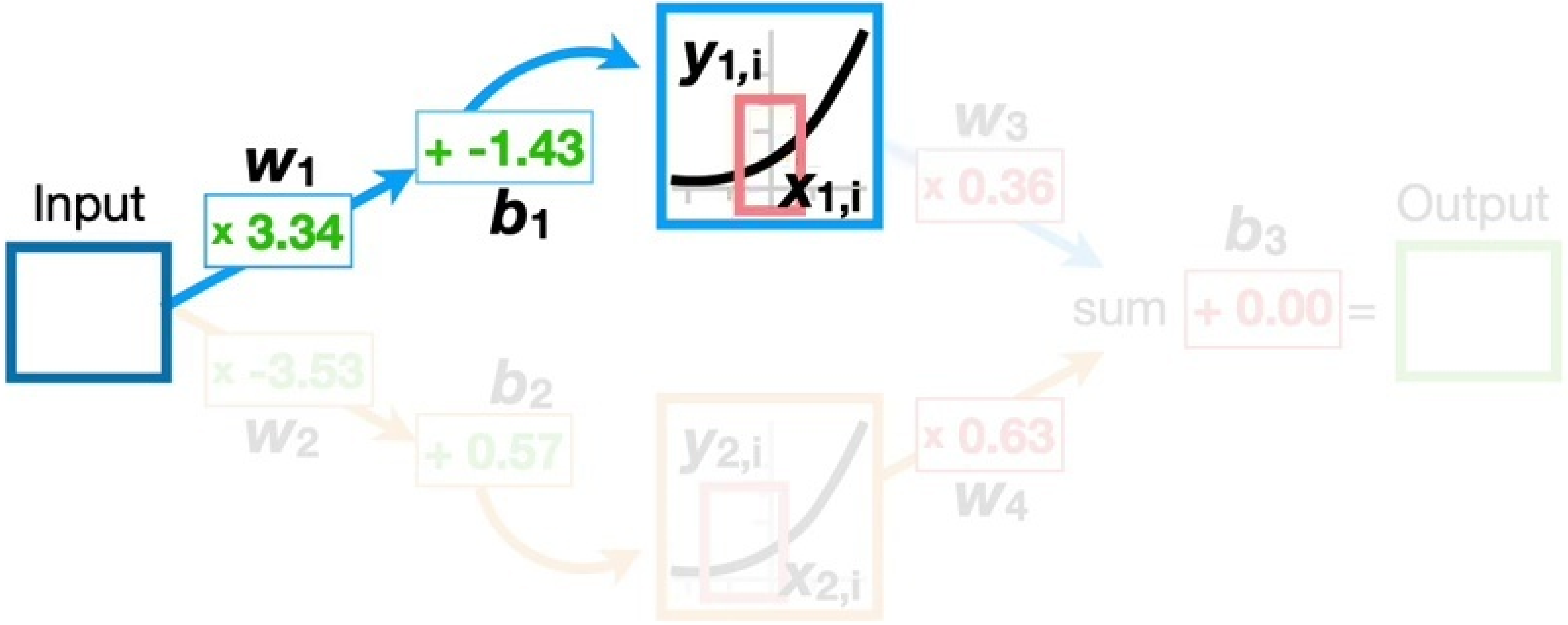
$$y_{1,i} = f(x_{1,i}) = \boxed{\log(1 + e^x)}$$



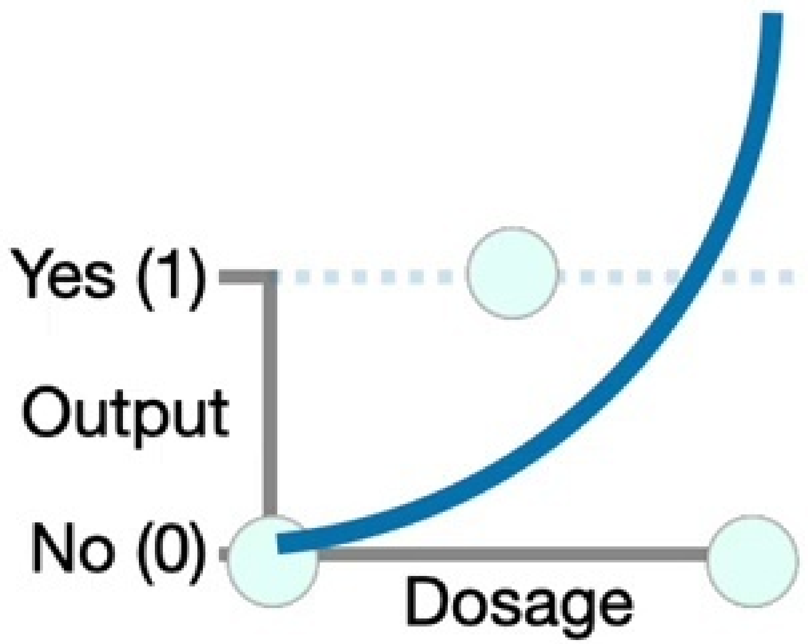


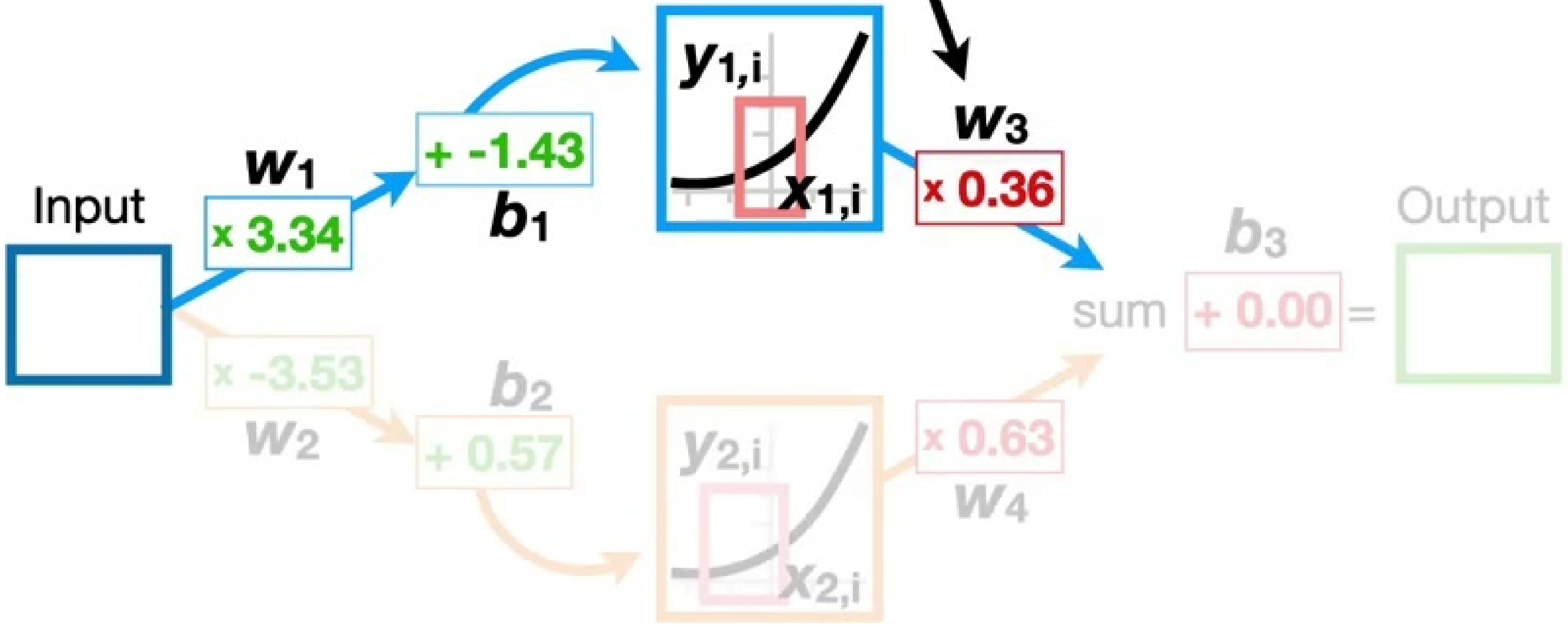
...and they form this initial **blue curve**.

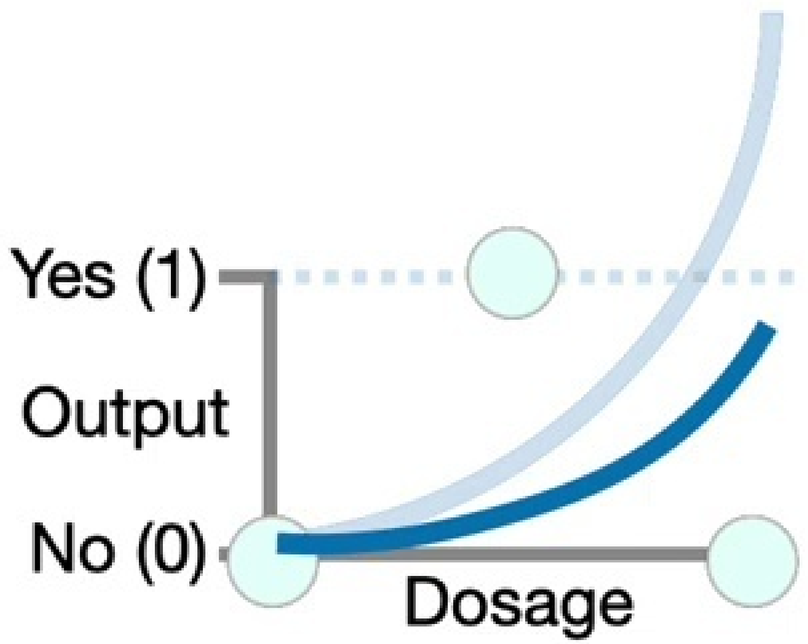




However, we get the final **blue curve** by multiplying the y-axis coordinates, $y_{1,i}$, by w_3.

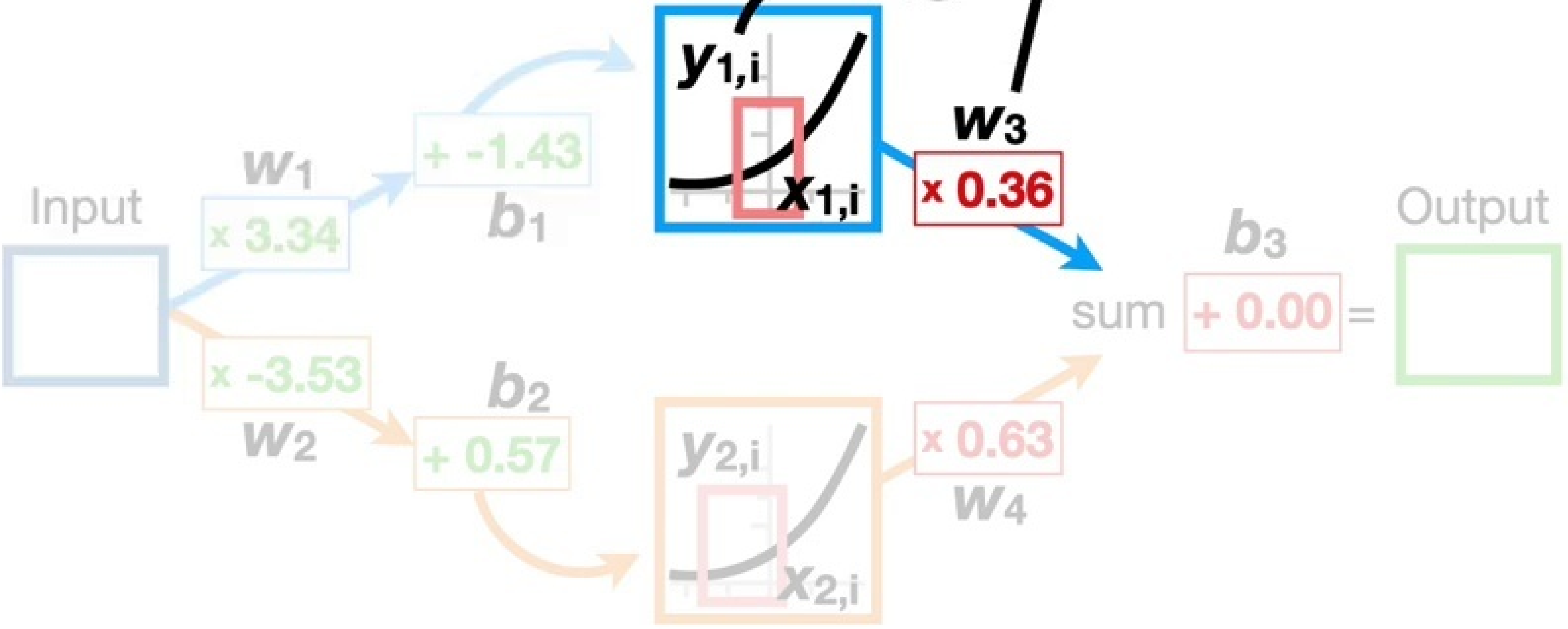




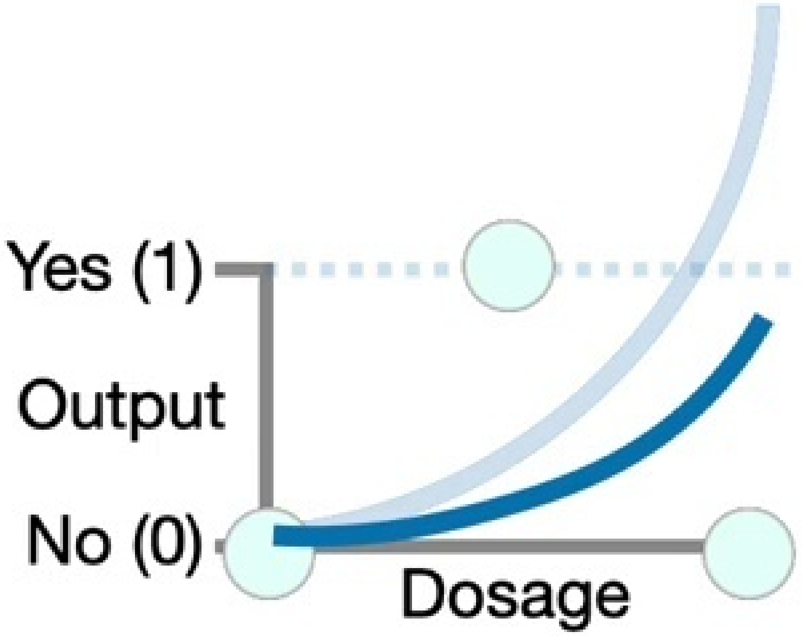


And that means we can plug $\boldsymbol{y_{1,i}}$ times $\boldsymbol{w_3}$ into the equation for the **Predicted** values.

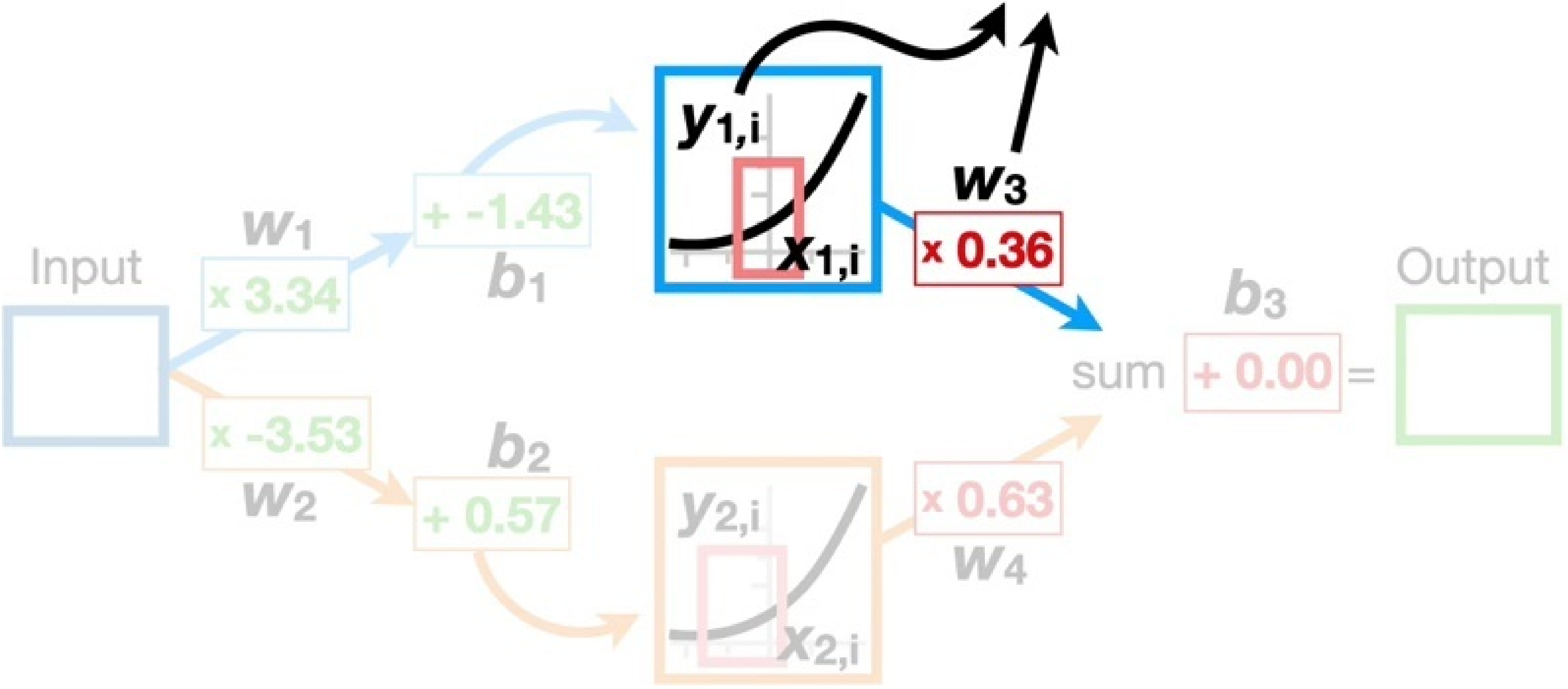
$$\text{Predicted}_i = \text{green squiggle}_i = \text{blue} + \text{orange} + \boldsymbol{b_3}$$



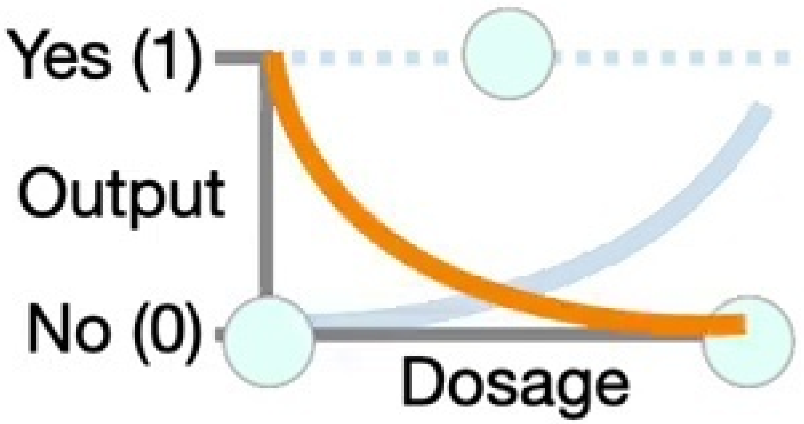
And that means we can plug **$y_{1,i}$** times **w_3** into the equation for the **Predicted** values.

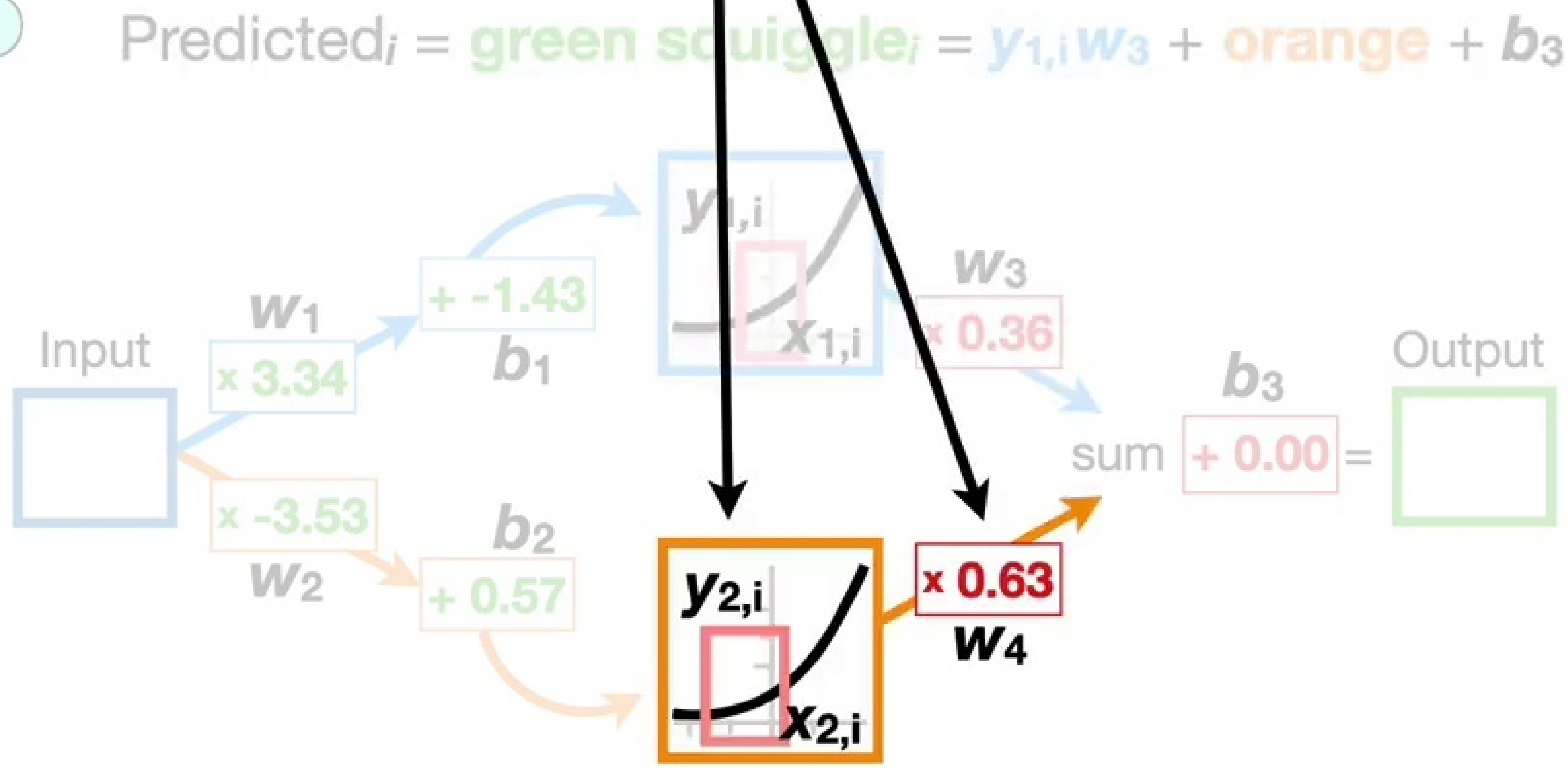


$$\text{Predicted}_i = \text{green squiggle}_i = y_{1,i}\, w_3 + \text{orange} + b_3$$

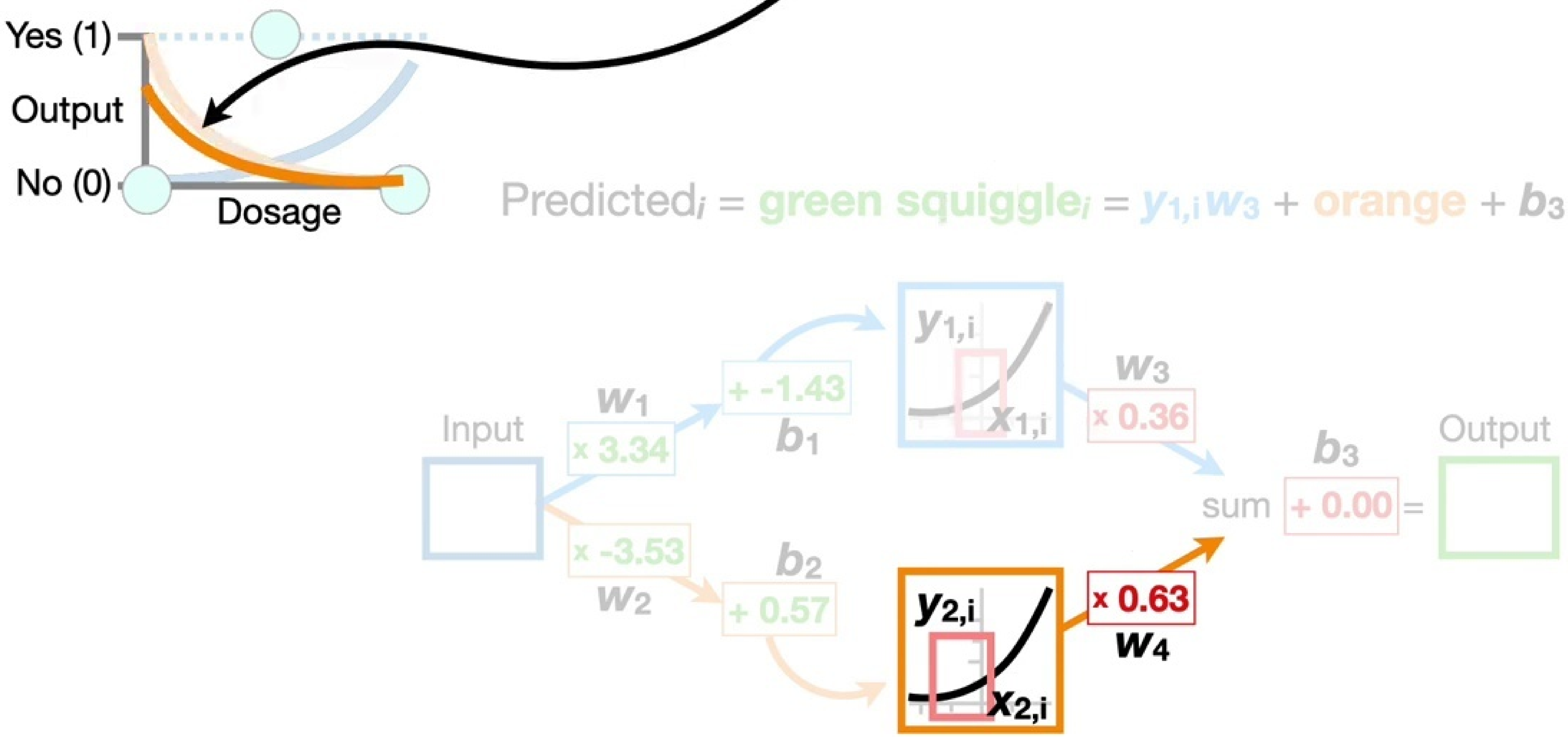


Likewise $\boldsymbol{w_4}$ multiplies the y-axis coordinates, $\boldsymbol{y_{2,i}}$, from the bottom **Activation Function**…

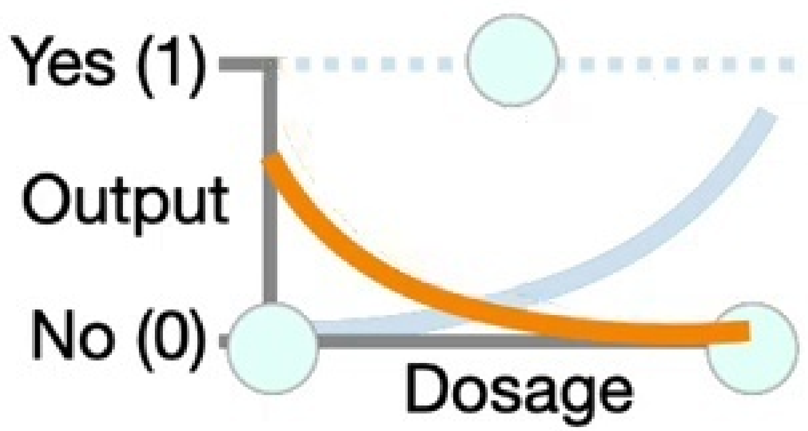




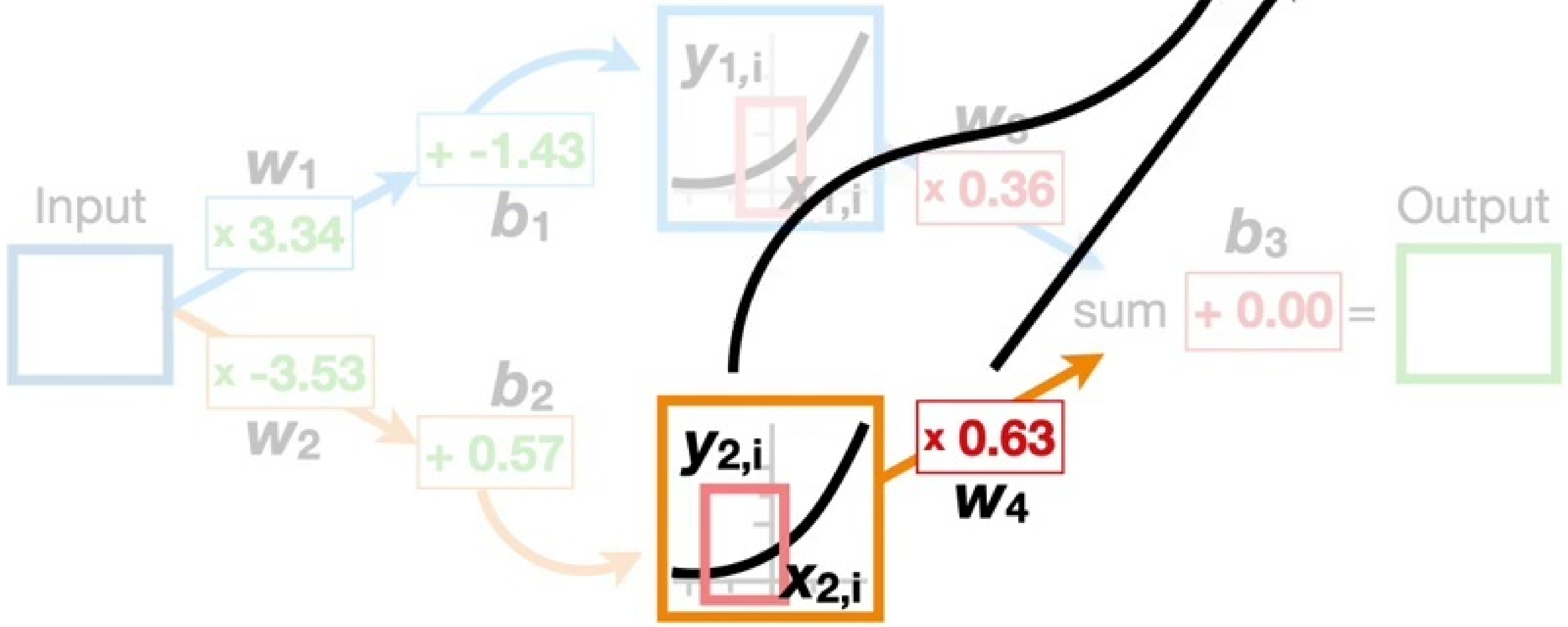
…to create the final **orange curve**.



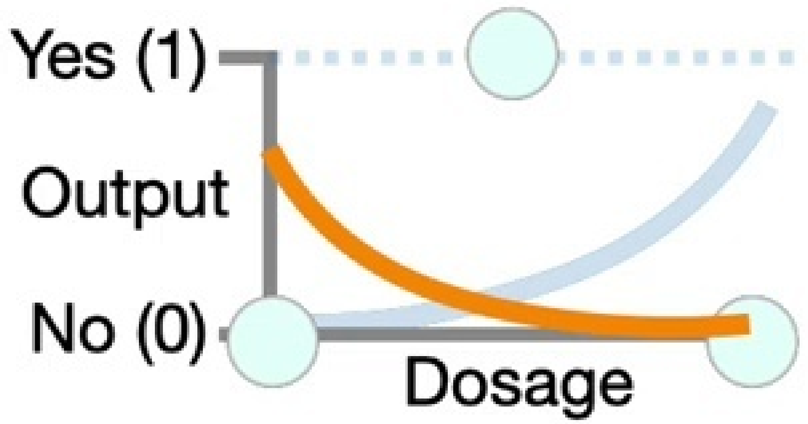
And that means we can plug $y_{2,i}$ times w_4 into the equation for the **Predicted** values.



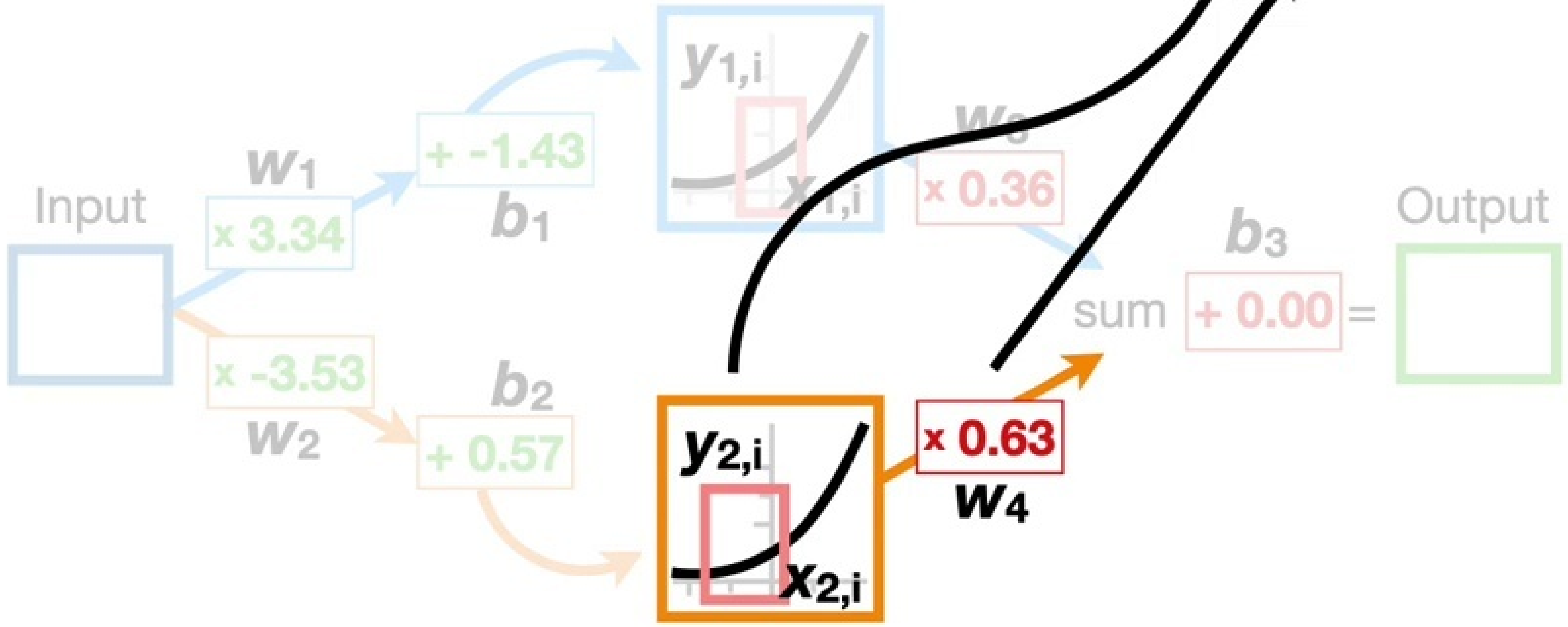
$$\text{Predicted}_i = \text{green squiggle}_i = y_{1,i} w_3 + \text{orange} + b_3$$



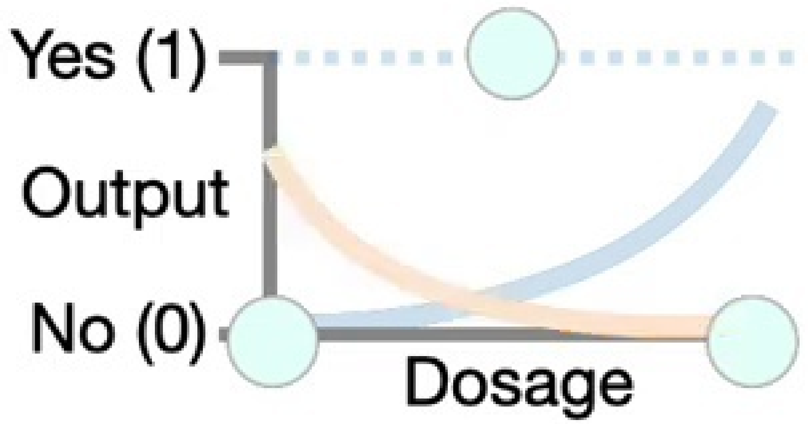
And that means we can plug $y_{2,i}$ times w_4 into the equation for the **Predicted** values.



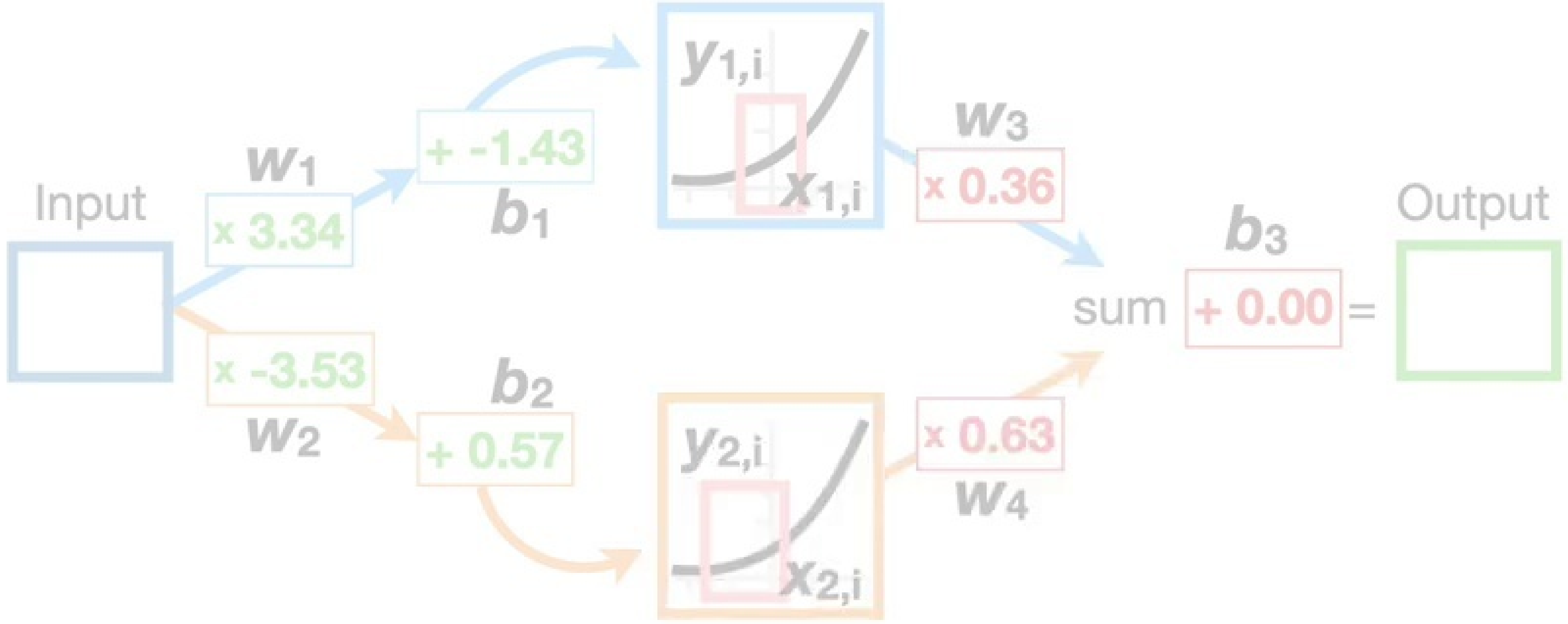
$$\text{Predicted}_i = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$



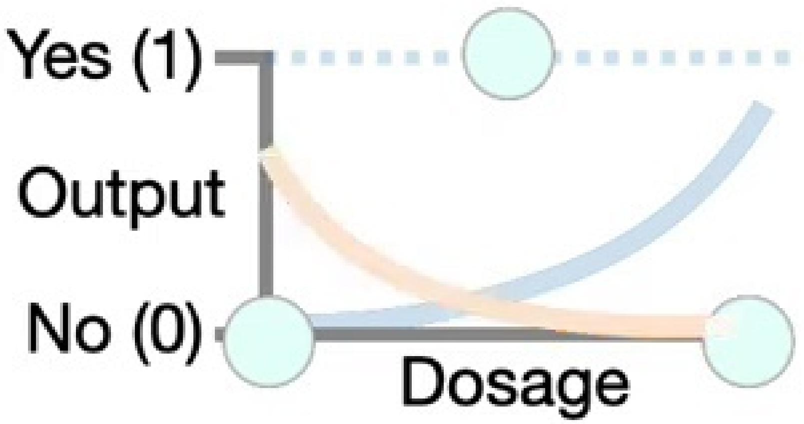
Now, since this sum…



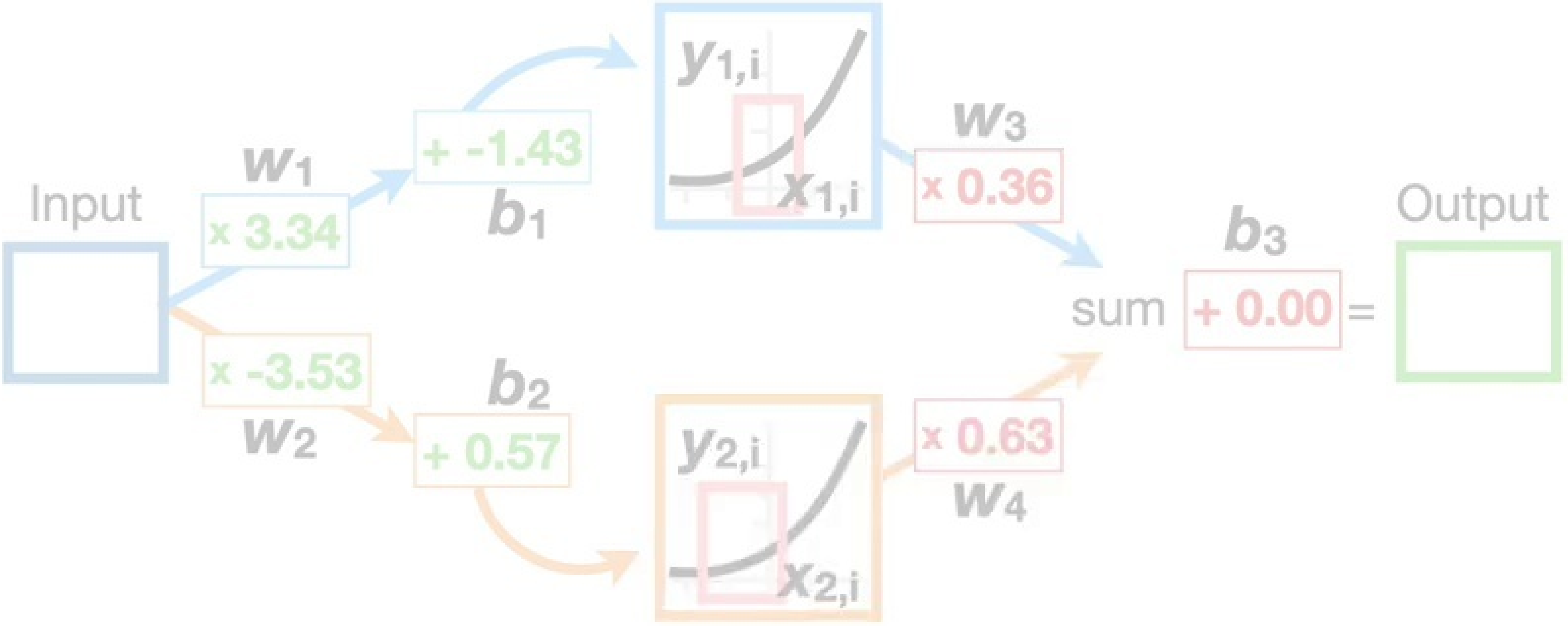
$$\text{Predicted}_i = \textbf{green squiggle}_i = \boxed{y_{1,i}w_3 + y_{2,i}w_4 + b_3}$$



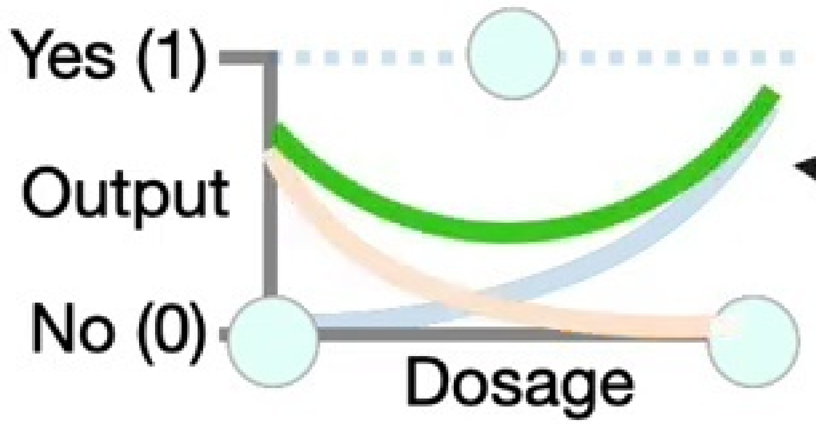
…creates the **green squiggle**…



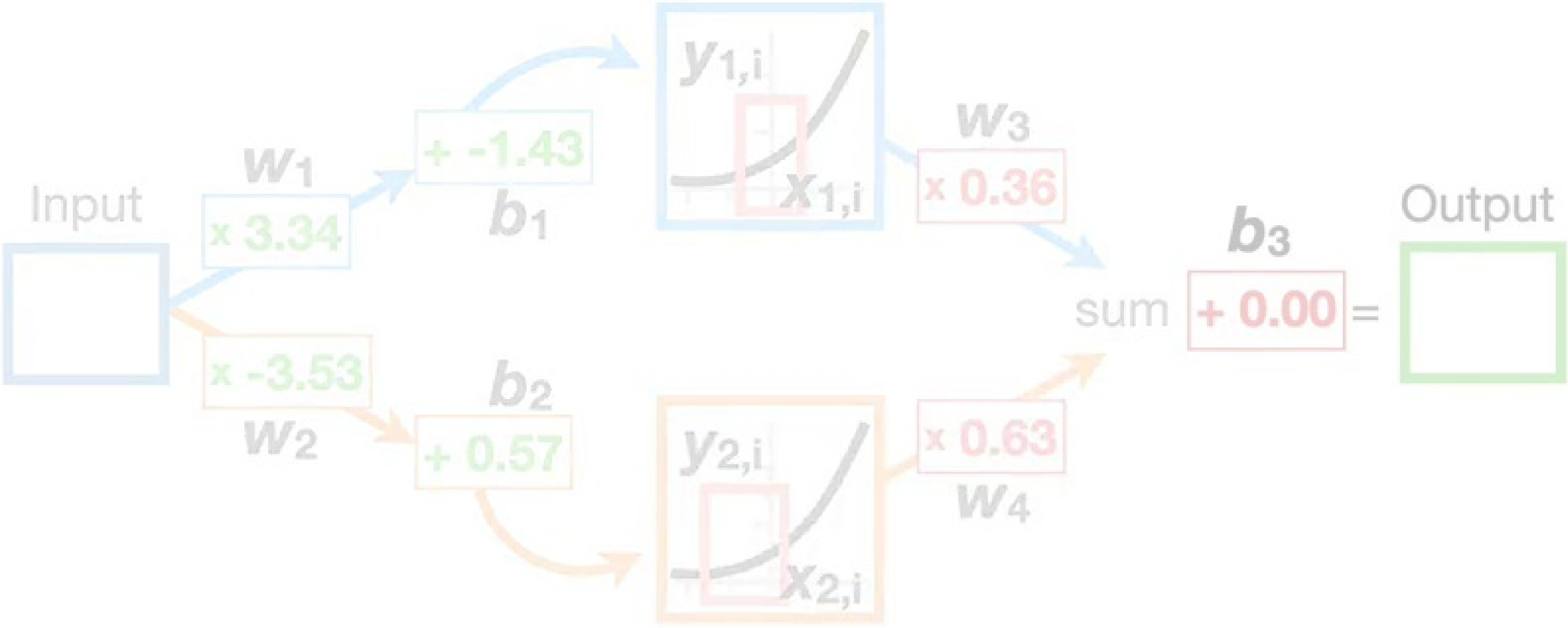
$$\text{Predicted}_i = \boxed{\text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3}$$



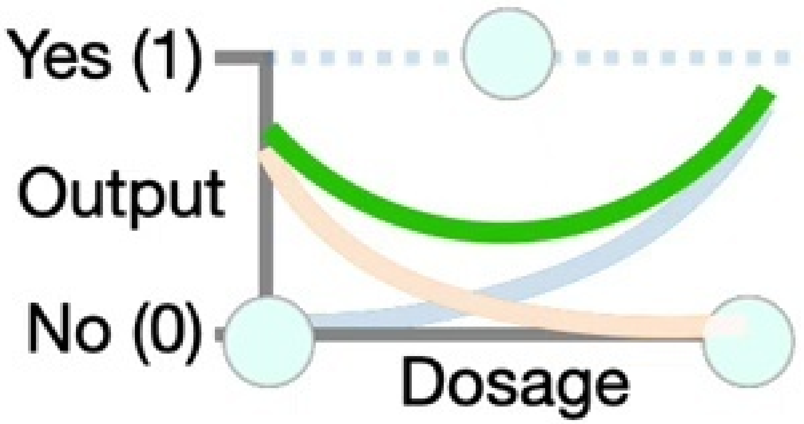
…creates the **green squiggle**…



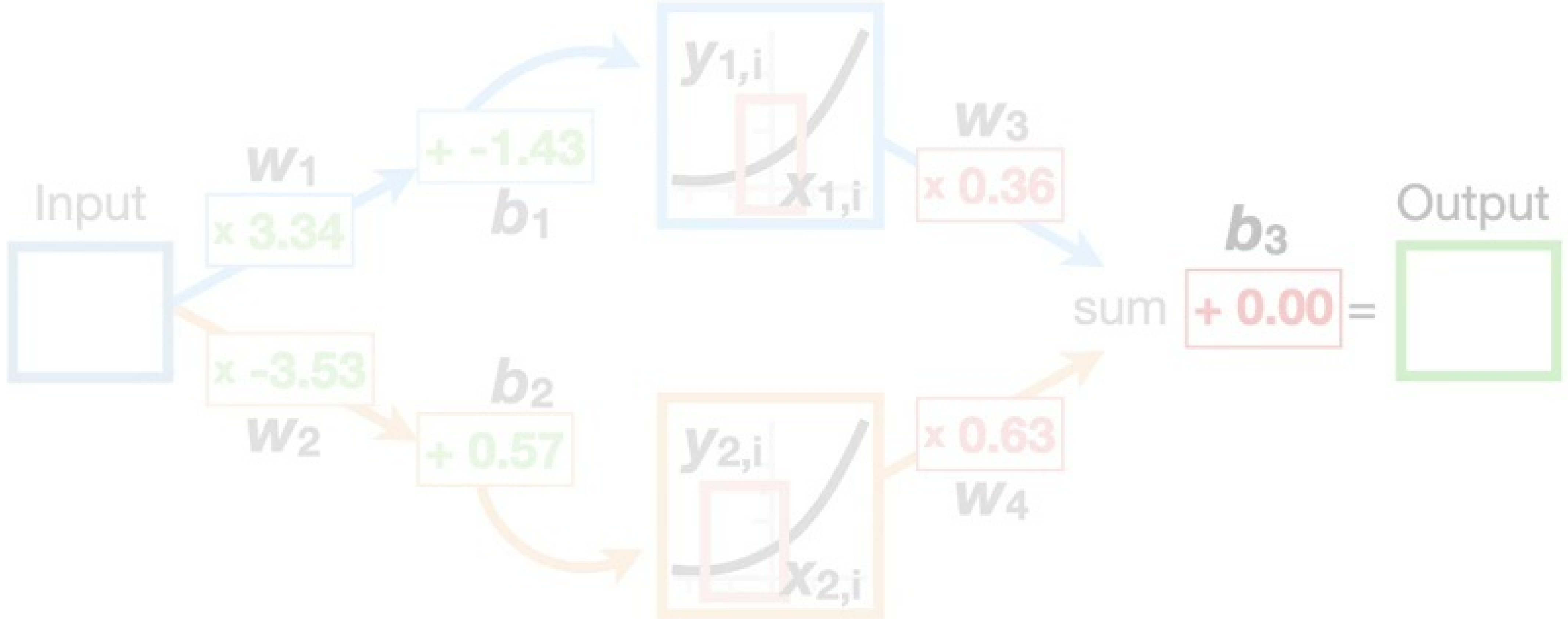
$$\text{Predicted}_i = \boxed{\text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3}$$



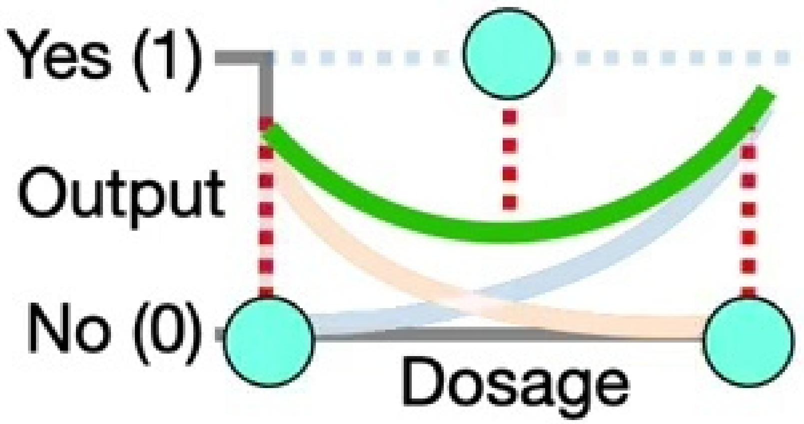
…and the **green squiggle** gives us **Predictions**…



$$\boxed{\text{Predicted}_i = \text{green squiggle}_i} = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

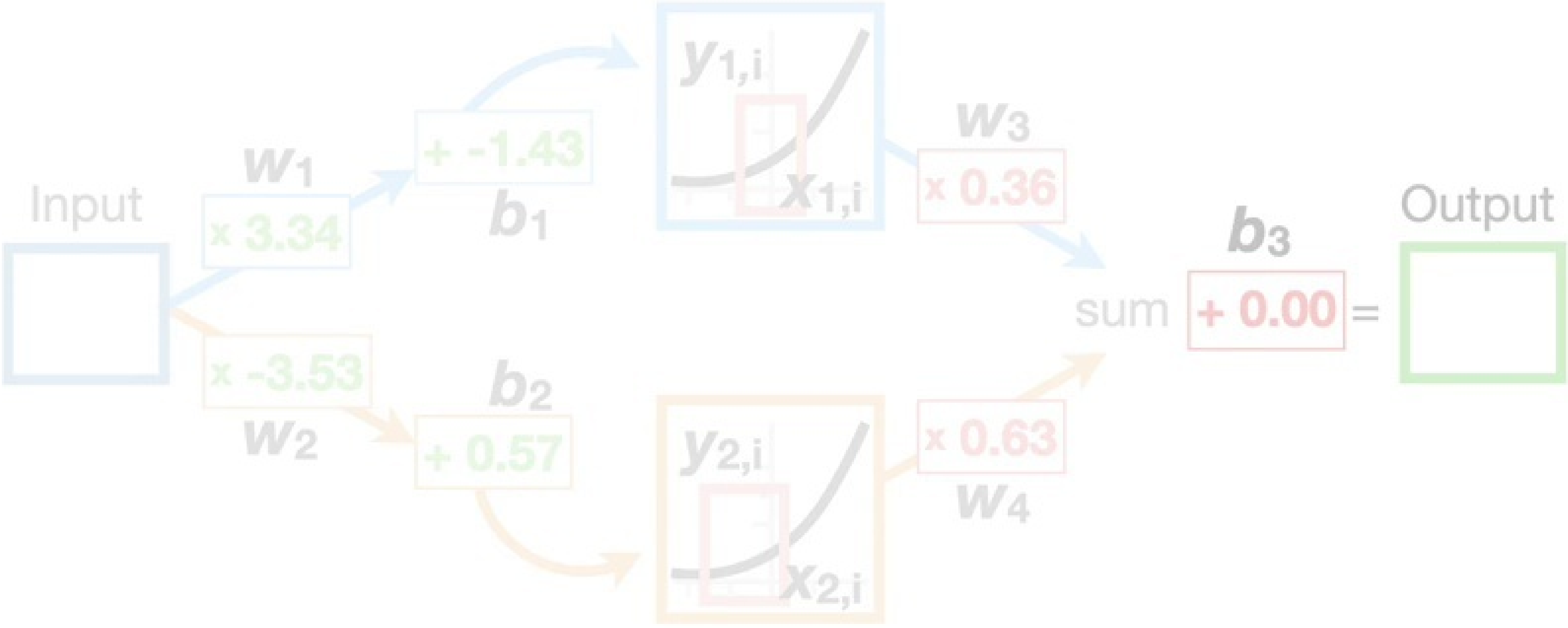


…that we evaluate with the **SSR**…

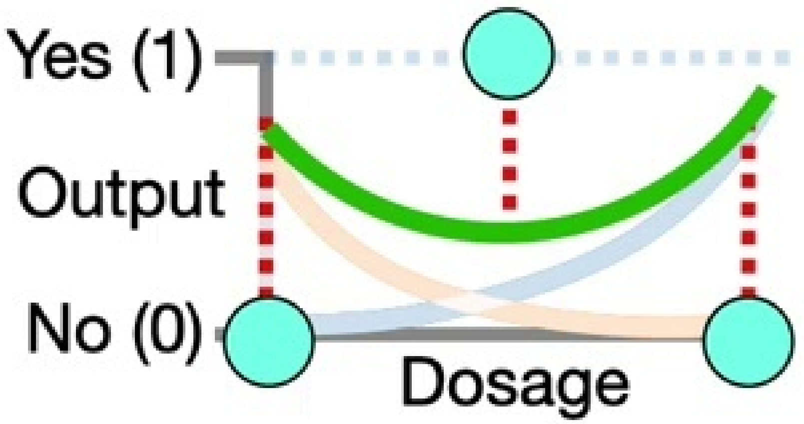


$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

$$\text{Predicted}_i = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

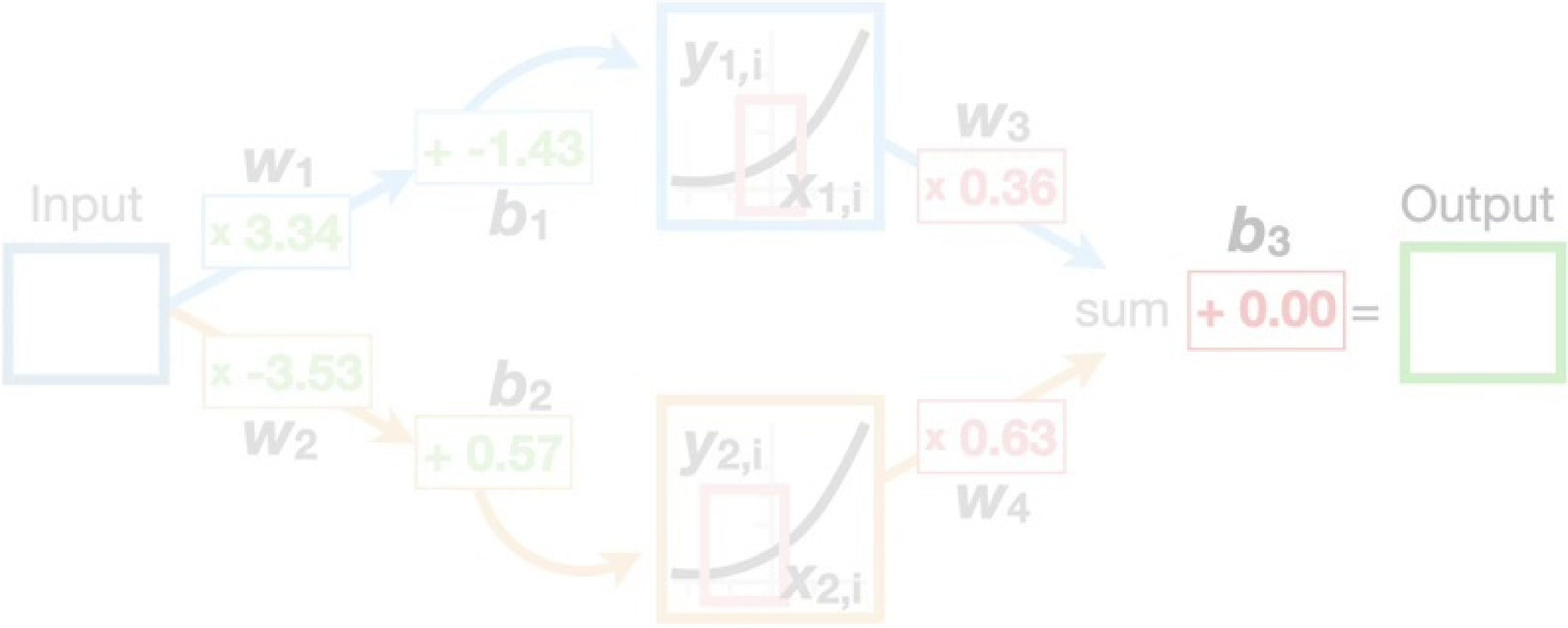


…then the **SSR** are linked to w_3 and w_4…

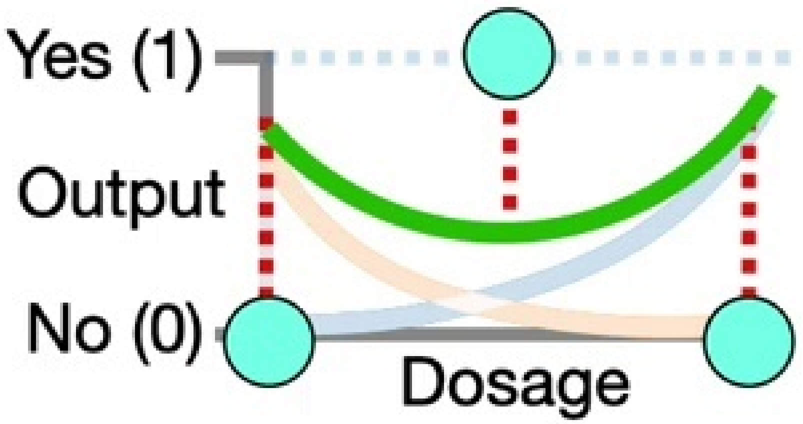


$$\boxed{\text{SSR}} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2$$

$$\text{Predicted}_i = \text{green squiggle}_i = y_{1,i} \boxed{w_3} + y_{2,i} \boxed{w_4} + b_3$$

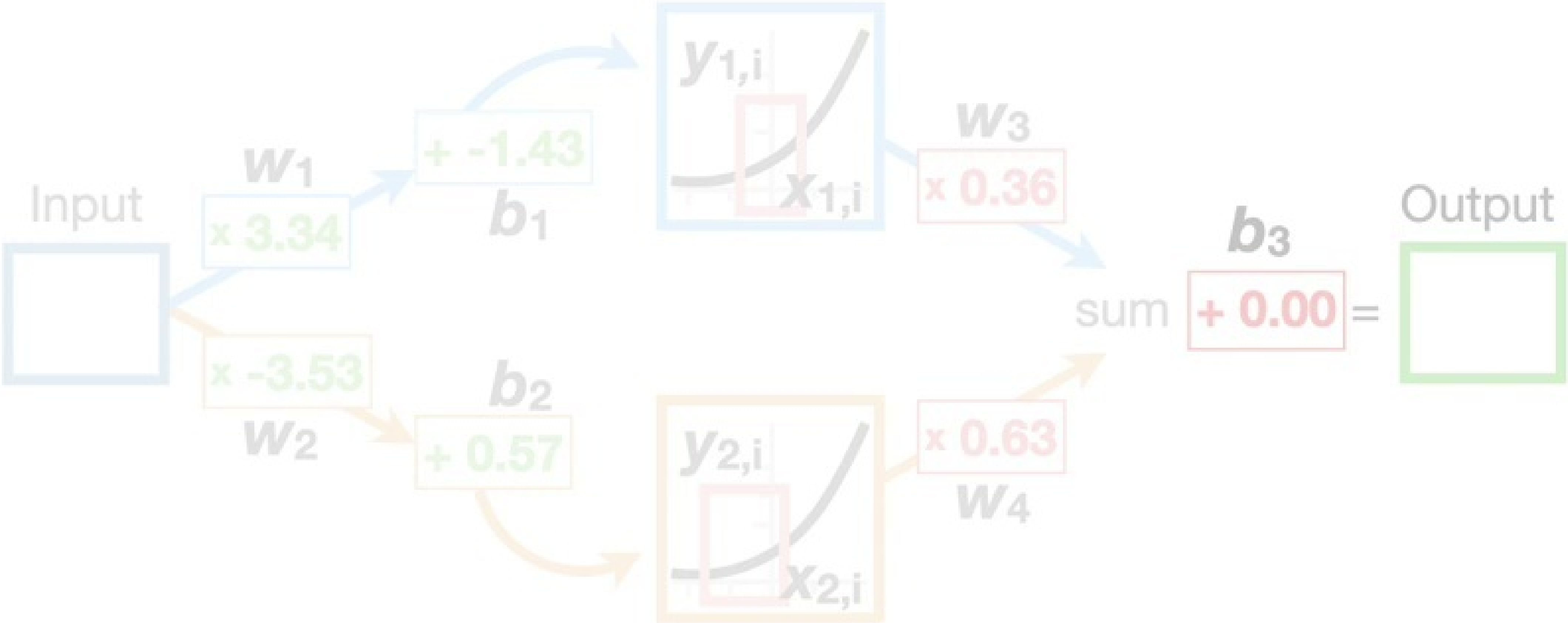


…by the **Predicted** values.

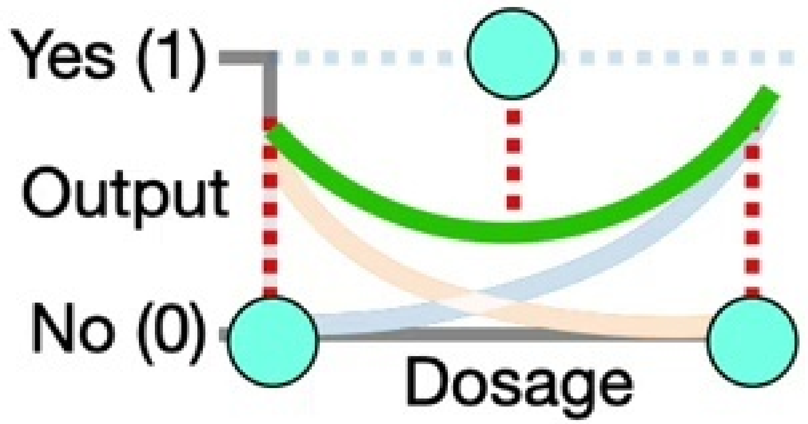


$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \boxed{\text{Predicted}_i})^2$$

$$\boxed{\text{Predicted}_i} = \textbf{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

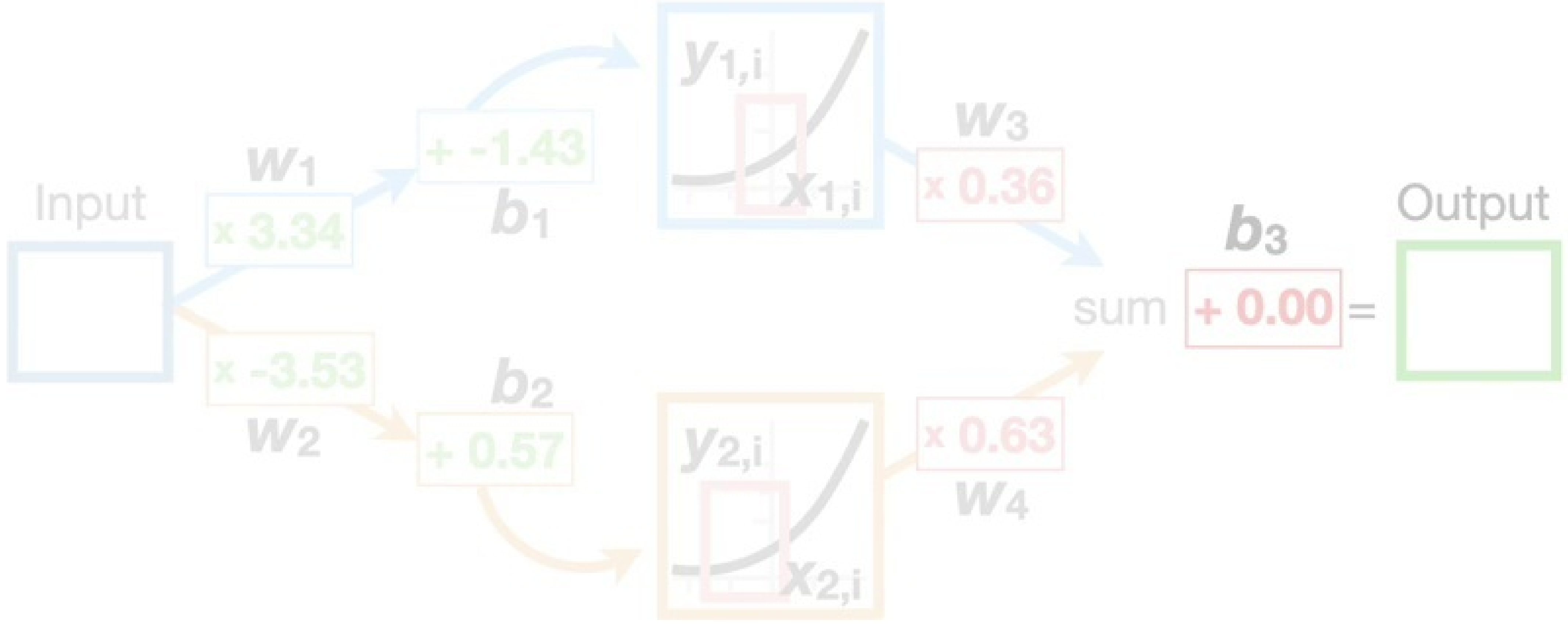


That means we can use
The Chain Rule…



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

$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

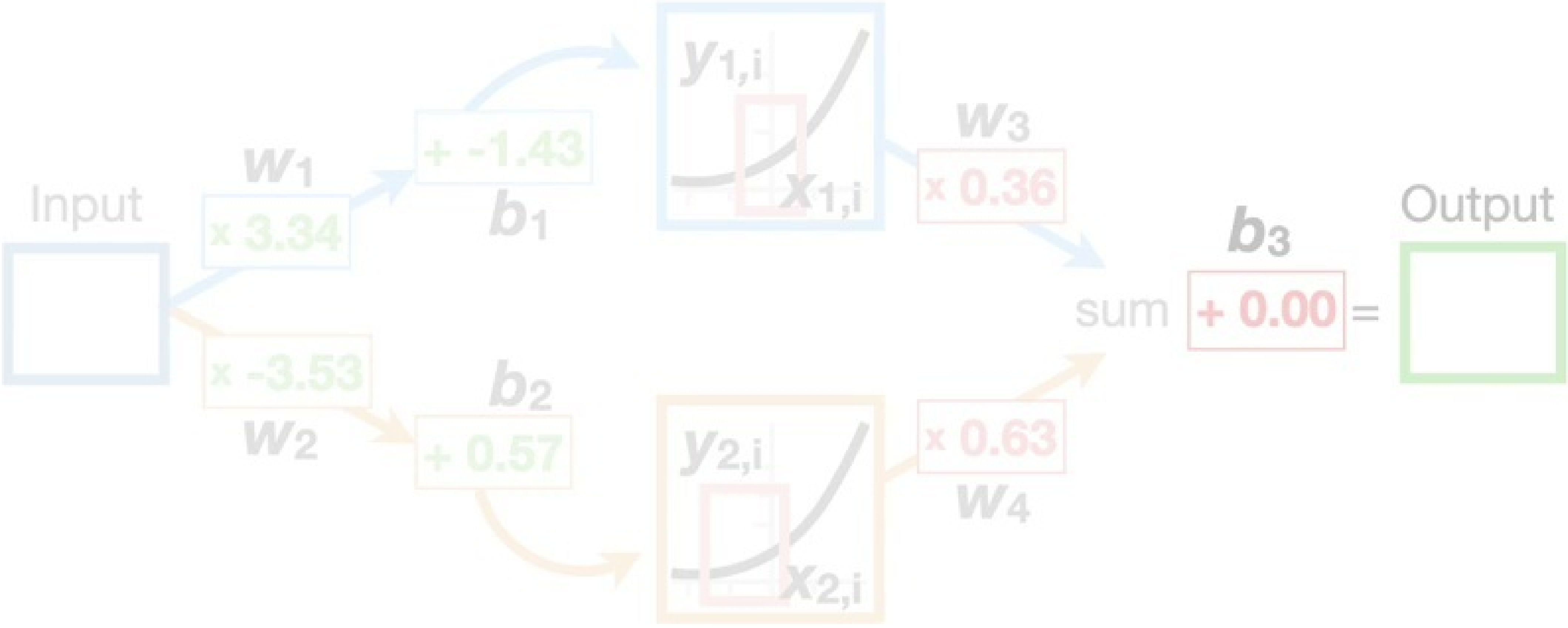


$$\frac{d\ SSR}{d\ w_3}$$

...to determine the derivative of the **SSR** with respect to ***w*₃**...

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

$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$



$$\frac{d\,SSR}{d\,w_3}$$

...and with respect to w_4.

$$\frac{d\,SSR}{d\,w_4}$$

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

$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

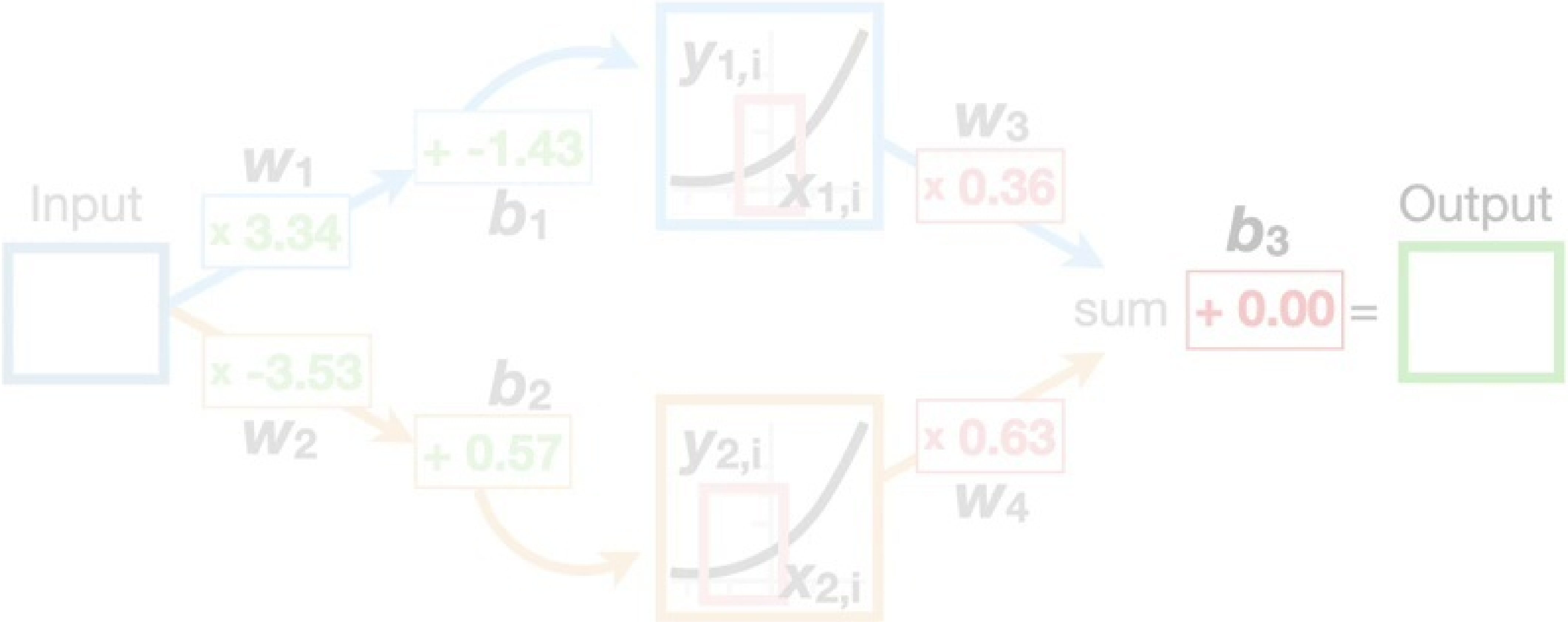
$$\frac{d\,SSR}{d\,w_3} =$$

The Chain Rule says that the derivative of the **SSR** with respect to $\boldsymbol{w_3}$…

$$\frac{d\,SSR}{d\,w_4}$$

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

$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$



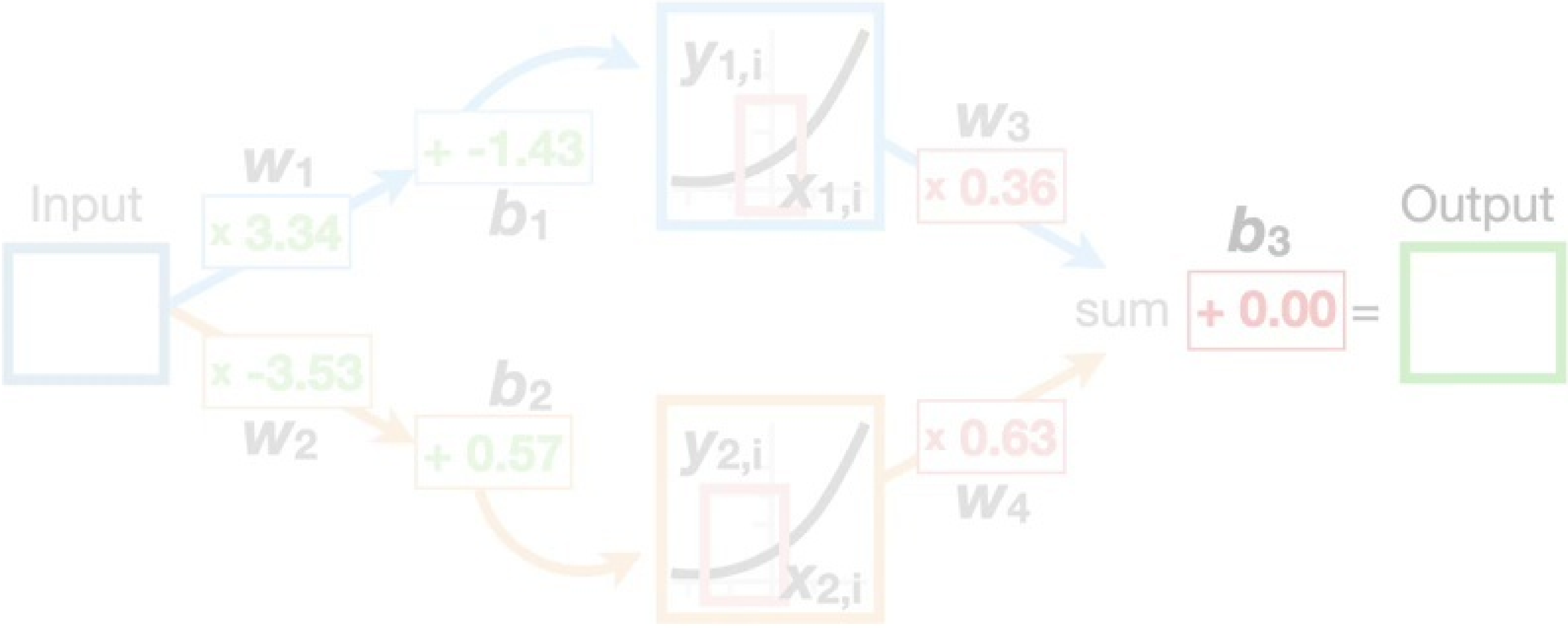
$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted}$$

$$\frac{d\,SSR}{d\,w_4}$$

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

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

$$\text{Predicted}_i = \textbf{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$



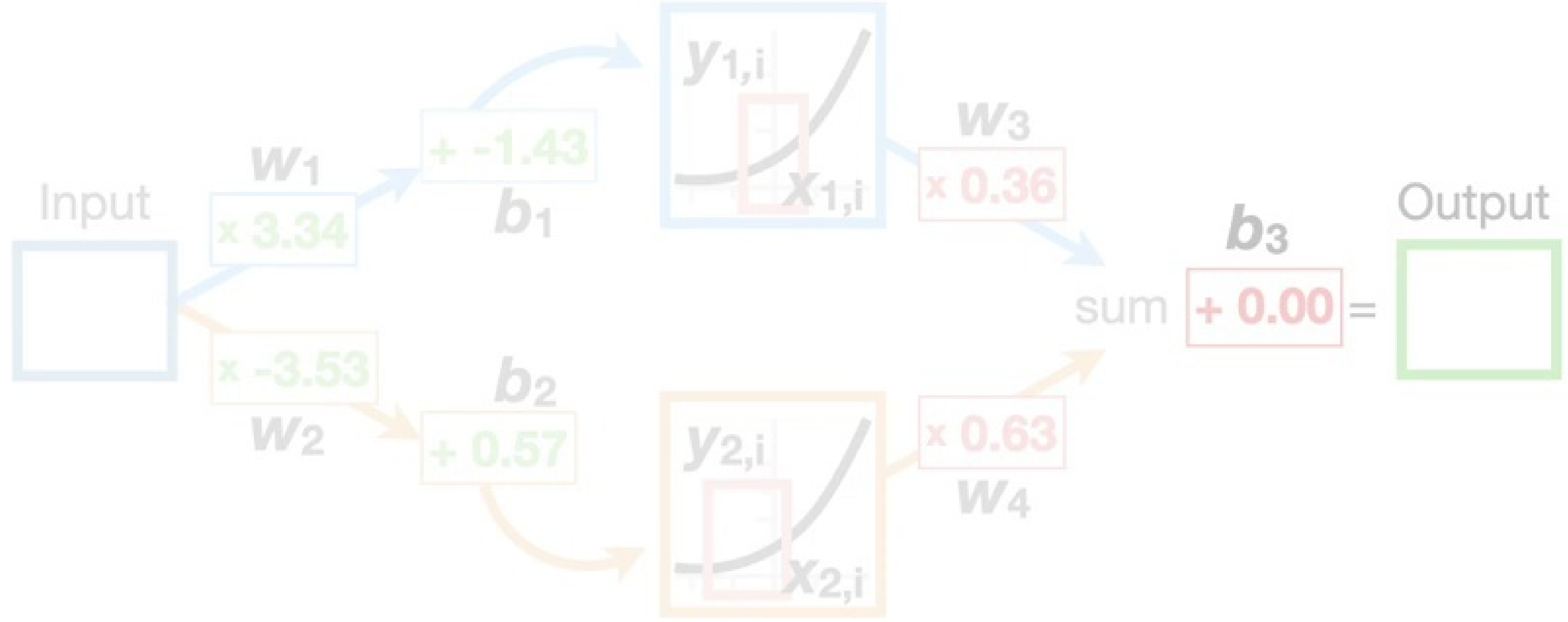
$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3}$$

…times the derivative of the **Predicted** values with respect to ***w*₃**.

$$\frac{d\,SSR}{d\,w_4}$$

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

$$\boxed{\text{Predicted}_i} = \textbf{green squiggle}_i = y_{1,i}\boxed{w_3} + y_{2,i}w_4 + b_3$$



$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3}$$

$$\frac{d\ SSR}{d\ w_4} =$$

Likewise, the derivative with respect to w_4…

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

$$\boxed{Predicted_i} = \textbf{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3}$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted}$$

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

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

$$Predicted_i = \textbf{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3}$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_4}$$

…times the derivative of the **Predicted** values with respect to w_4.

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

$$\boxed{\text{Predicted}_i} = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} \boxed{w_4} + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \boxed{\frac{d\ SSR}{d\ Predicted}} \times \frac{d\ Predicted}{d\ w_3}$$

$$\frac{d\ SSR}{d\ w_4} = \boxed{\frac{d\ SSR}{d\ Predicted}} \times \frac{d\ Predicted}{d\ w_4}$$

NOTE: In both cases, the derivative of the **SSR** with respect to the **Predicted** values…

$$SSR = \sum_{i=1}^{n=3} (Observed_i - Predicted_i)^2 \qquad Predicted_i = \textbf{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3}$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_4}$$

…is the exact same as the derivative used for b_3.

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

$$SSR = \sum_{i=1}^{n=3} (Observed_i - Predicted_i)^2 \qquad Predicted_i = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \boxed{\frac{d\,SSR}{d\,Predicted}} \times \frac{d\,Predicted}{d\,w_3}$$

$$\frac{d\,SSR}{d\,w_4} = \boxed{\frac{d\,SSR}{d\,Predicted}} \times \frac{d\,Predicted}{d\,w_4}$$

Just to remind you, we start by substituting **SSR** with its equation…

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

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

$$\text{Predicted}_i = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3}$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_4}$$

…then we use **The Chain Rule…**

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

$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \qquad \text{Predicted}_i = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3}$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_4}$$

…to move the square to the front…

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

$$SSR = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3}$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_4}$$

…and then we multiply that by the derivative of the stuff **Inside** the parentheses with respect to the **Predicted** values, **-1**.

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

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \textbf{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \boxed{\frac{d\,SSR}{d\,Predicted}} \times \frac{d\,Predicted}{d\,w_3}$$

$$\frac{d\,SSR}{d\,w_4} = \boxed{\frac{d\,SSR}{d\,Predicted}} \times \frac{d\,Predicted}{d\,w_4}$$

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

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

$$SSR = \sum_{i=1}^{n=3}(\text{Observed}_i - \text{Predicted}_i)^2 \qquad \text{Predicted}_i = \textbf{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \boxed{\frac{d\,SSR}{d\,Predicted}} \times \frac{d\,Predicted}{d\,w_3}$$

$$\frac{d\,SSR}{d\,w_4} = \boxed{\frac{d\,SSR}{d\,Predicted}} \times \frac{d\,Predicted}{d\,w_4}$$

So we just plug it in.

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

$$SSR = \sum_{i=1}^{n=3}(\text{Observed}_i - \text{Predicted}_i)^2 \quad \text{Predicted}_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \boxed{\frac{d\,SSR}{d\,Predicted}} \times \frac{d\,Predicted}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\,SSR}{d\,w_4} = \boxed{\frac{d\,SSR}{d\,Predicted}} \times \frac{d\,Predicted}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

So we just plug it in.

$$\frac{d\,SSR}{d\,Predicted} = \frac{d}{d\,Predicted} \sum_{i=1}^{n=3} (Observed_i - Predicted_i)^2 = \boxed{\sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)}$$

$$SSR = \sum_{i=1}^{n=3} (Observed_i - Predicted_i)^2 \quad Predicted_i = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \boxed{\frac{d\ Predicted}{d\ w_3}} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_4} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

Now, to solve for the derivative of the **Predicted** values with respect to w_3…

$$\frac{d\ Predicted}{d\ w_3}$$

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

$$Predicted_i = \text{green squiggle}_i = y_{1,i} w_3 + y_{2,i} w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \boxed{\frac{d\,Predicted}{d\,w_3}} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

…we plug in the equation for the **Predicted** values.

$$\frac{d\,Predicted}{d\,w_3} = \frac{d}{d\,w_3}\text{green squiggle}$$

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

$$\boxed{Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3}$$

$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \boxed{\frac{d\,Predicted}{d\,w_3}} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

…we plug in the equation for the **Predicted** values.

$$\frac{d\,Predicted}{d\,w_3} = \frac{d}{d\,w_3}\text{ green squiggle} = \frac{d}{d\,w_3}(y_{1,i}w_3 + y_{2,i}w_4 + b_3)$$

$$SSR = \sum_{i=1}^{n=3}(Observed_i - Predicted_i)^2 \qquad \boxed{Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3}$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_4} = \sum_{i=1}^{n=3} -2 \times (Obse$$

And the derivative of the first term with respect to $\boldsymbol{w_3}$ is $\boldsymbol{y_{1,i}}$…

$$\frac{d\ Predicted}{d\ w_3} = \frac{d}{d\ w_3}\ \text{green squiggle} = \frac{d}{d\ w_3}(y_{1,i}w_3 + y_{2,i}w_4 + b_3) = y_{1,i}$$

$$SSR = \sum_{i=1}^{n=3}(Observed_i - Predicted_i)^2 \qquad Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_4} = \sum_{i=1}^{n=3} -2 \times (Ob$$

…and the derivatives of the other terms are both **0** since they do not contain $\boldsymbol{w_3}$.

$$\frac{d\ Predicted}{d\ w_3} = \frac{d}{d\ w_3}\ \text{green squiggle} = \frac{d}{d\ w_3}(y_{1,i}w_3 + y_{2,i}w_4 + b_3) = y_{1,i} + 0$$

$$SSR = \sum_{i=1}^{n=3}(Observed_i - Predicted_i)^2 \qquad Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_4} = \sum_{i=1}^{n=3} -2 \times (Ob$$

...and the derivatives of the other terms are both **0** since they do not contain ***w*₃.**

$$\frac{d\ Predicted}{d\ w_3} = \frac{d}{d\ w_3}\ \text{green squiggle} = \frac{d}{d\ w_3}(y_{1,i}w_3 + y_{2,i}w_4 + b_3) = y_{1,i} + 0 + 0$$

$$SSR = \sum_{i=1}^{n=3}(Observed_i - Predicted_i)^2 \quad Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_4} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

And we end up with just $\boldsymbol{y_{1,i}}$.

$$\frac{d\ Predicted}{d\ w_3} = \frac{d}{d\ w_3}\ \text{green squiggle} = \frac{d}{d\ w_3}(y_{1,i}w_3 + y_{2,i}w_4 + b_3)\boxed{= y_{1,i}}$$

$$SSR = \sum_{i=1}^{n=3}(Observed_i - Predicted_i)^2 \qquad Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i) \times y_{1,i}$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

$$\frac{d\,Predicted}{d\,w_3} = \frac{d}{d\,w_3}\,\text{green squiggle} = \frac{d}{d\,w_3}(y_{1,i}w_3 + y_{2,i}w_4 + b_3) \boxed{= y_{1,i}}$$

So we multiply the derivative of the **SSR** with respect to the **Predicted** values by $y_{1,i}$.

$$SSR = \sum_{i=1}^{n=3} (Observed_i - Predicted_i)^2 \qquad Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i) \times y_{1,i}$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \boxed{\frac{d\ Predicted}{d\ w_4}} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

Likewise, the derivative of the **Predicted** values with respect to $\boldsymbol{w_4}$ is…

$$\frac{d\ Predicted}{d\ w_4} = \frac{d}{d\ w_4}\ \text{green squiggle} = \frac{d}{d\ w_4}(y_{1,i}w_3 + y_{2,i}w_4 + b_3)$$

$$SSR = \sum_{i=1}^{n=3}(Observed_i - Predicted_i)^2 \qquad Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{1,i}$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \boxed{\frac{d\,Predicted}{d\,w_4}} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i)$$

…**0** for the first term…

$$\frac{d\,Predicted}{d\,w_4} = \frac{d}{d\,w_4}\text{ green squiggle} = \frac{d}{d\,w_4}(y_{1,i}w_3 + y_{2,i}w_4 + b_3) = 0$$

$$\text{SSR} = \sum_{i=1}^{n=3}(\text{Observed}_i - \text{Predicted}_i)^2 \qquad \text{Predicted}_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i) \times y_{1,i}$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \boxed{\times \frac{d\,Predicted}{d\,w_4}} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

…plus $y_{2,i}$ for the second term…

$$\frac{d\,Predicted}{d\,w_4} = \frac{d}{d\,w_4}\text{ green squiggle} = \frac{d}{d\,w_4}(y_{1,i}w_3 + y_{2,i}w_4 + b_3) = 0 + y_{2,i}$$

$$SSR = \sum_{i=1}^{n=3}(Observed_i - Predicted_i)^2 \qquad Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i) \times y_{1,i}$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \boxed{\frac{d\ Predicted}{d\ w_4}} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

…plus **0** for the third term…

$$\frac{d\ Predicted}{d\ w_4} = \frac{d}{d\ w_4} \text{green squiggle} = \frac{d}{d\ w_4} (y_{1,i}w_3 + y_{2,i}w_4 + b_3) = 0 + y_{2,i} + 0$$

$$SSR = \sum_{i=1}^{n=3} (Observed_i - Predicted_i)^2 \qquad Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i) \times y_{1,i}$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \boxed{\frac{d\ Predicted}{d\ w_4}} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i)$$

…which is just $y_{2,i}$.

$$\frac{d\ Predicted}{d\ w_4} = \frac{d}{d\ w_4} \text{green squiggle} = \frac{d}{d\ w_4}(y_{1,i}w_3 + y_{2,i}w_4 + b_3) = \boxed{y_{2,i}}$$

$$SSR = \sum_{i=1}^{n=3} (Observed_i - Predicted_i)^2 \qquad Predicted_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\ SSR}{d\ w_3} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{1,i}$$

$$\frac{d\ SSR}{d\ w_4} = \frac{d\ SSR}{d\ Predicted} \times \frac{d\ Predicted}{d\ w_4} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{2,i}$$

So we multiply the derivative of the **SSR** with respect to the **Predicted** values by $y_{2,i}$.

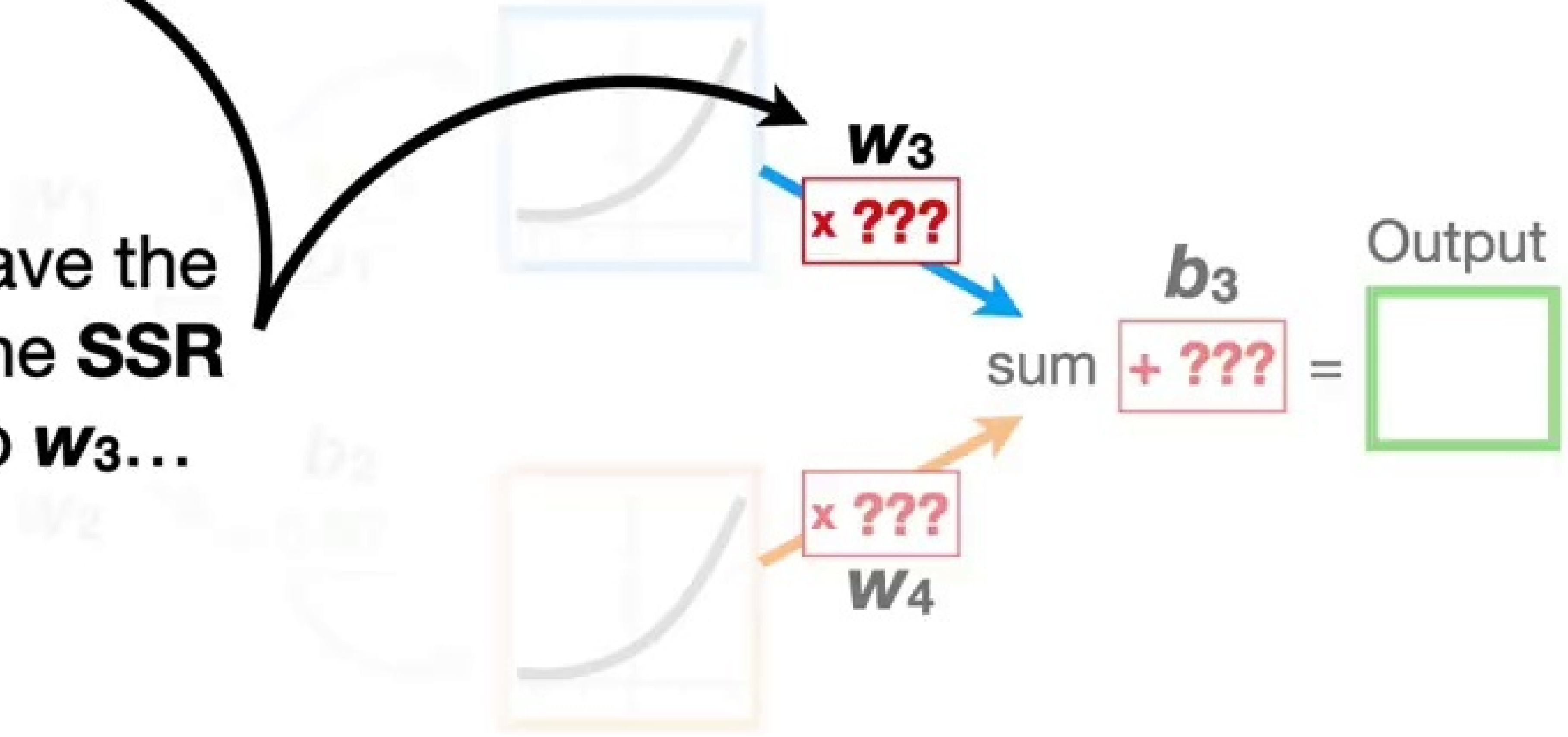
$$\frac{d\ Predicted}{d\ w_4} = \frac{d}{d\ w_4}\ \text{green squiggle} = \frac{d}{d\ w_4}(y_{1,i}w_3 + y_{2,i}w_4 + b_3) \boxed{= y_{2,i}}$$

$$\text{SSR} = \sum_{i=1}^{n=3} (\text{Observed}_i - \text{Predicted}_i)^2 \qquad \text{Predicted}_i = \text{green squiggle}_i = y_{1,i}w_3 + y_{2,i}w_4 + b_3$$

$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i) \times y_{1,i}$$

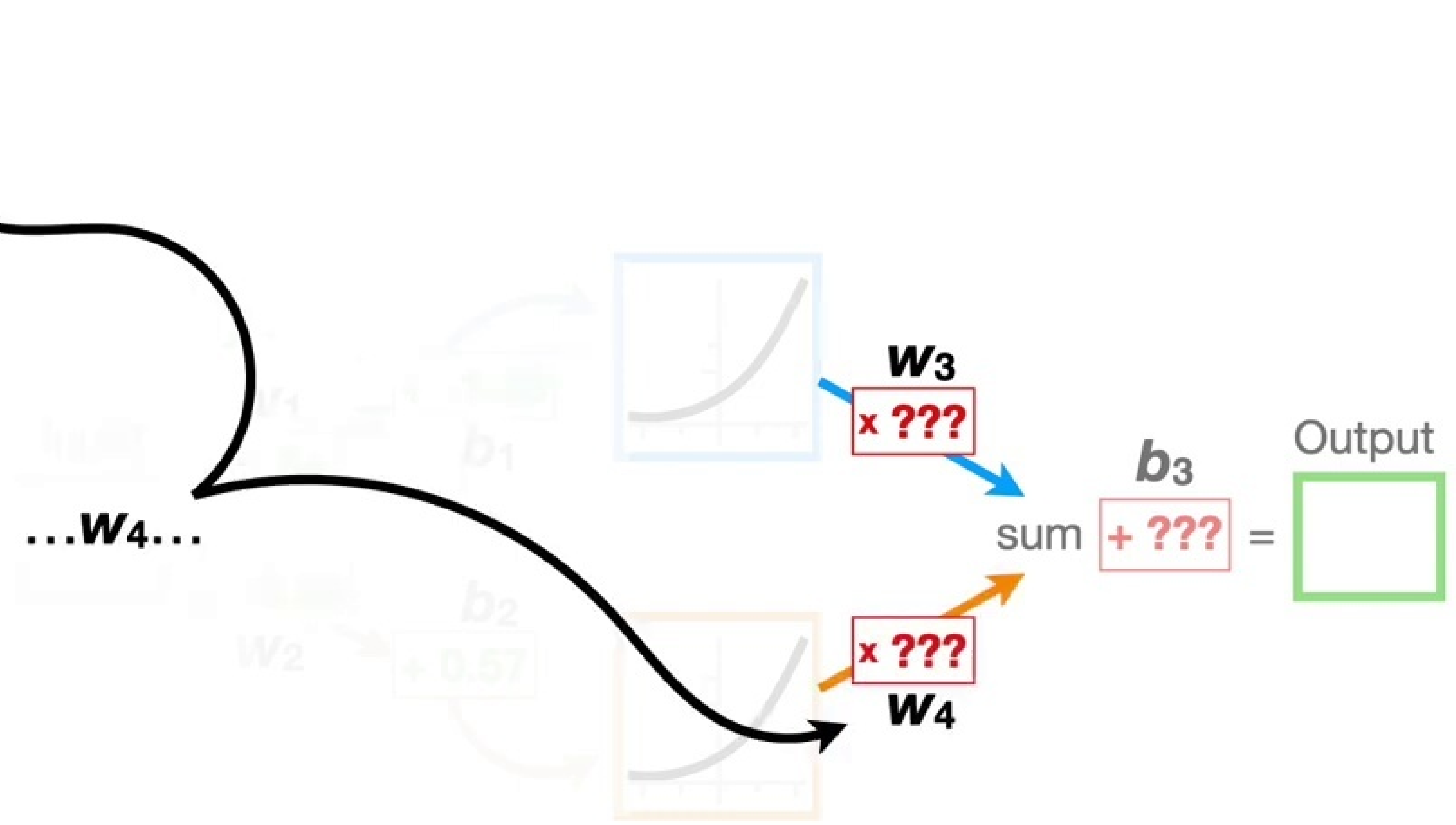
$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i) \times y_{2,i}$$

Now that we have the derivatives of the **SSR** with respect to ***w***$_3$…



$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{1,i}$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{2,i}$$

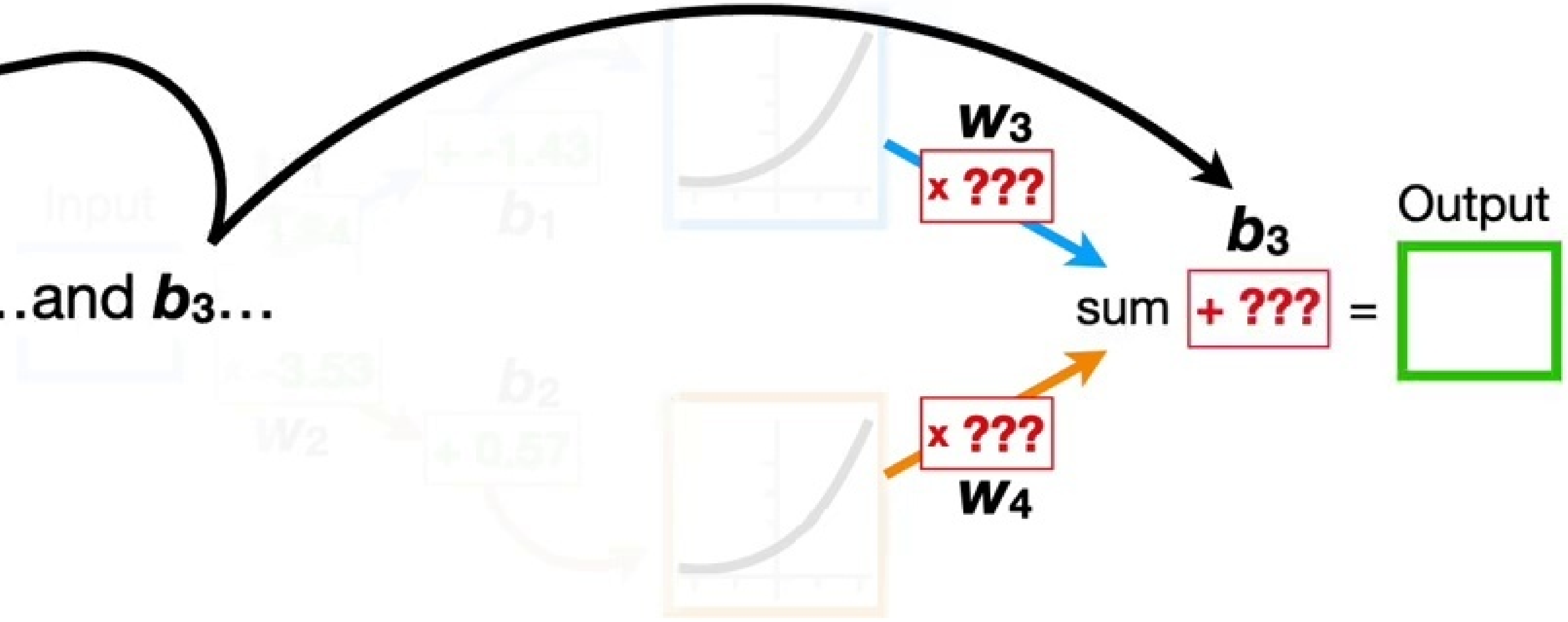


$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i) \times y_{1,i}$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i) \times y_{2,i}$$

$$\frac{d\,SSR}{d\,b_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,b_3} = \sum_{i=1}^{n=3} -2 \times (Observed_i - Predicted_i) \times 1$$

…and b_3…

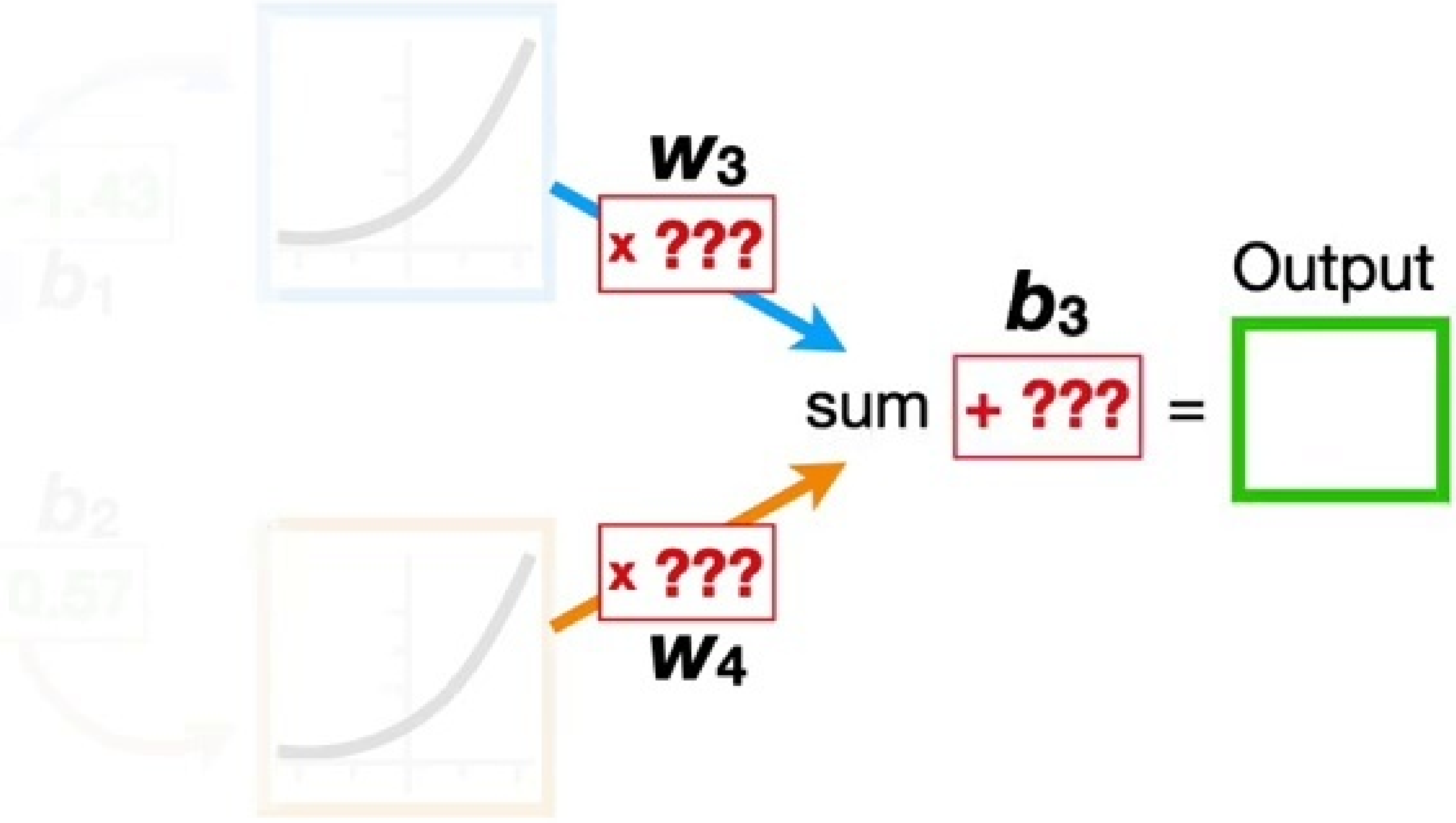


$$\frac{d\,SSR}{d\,w_3} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{1,i}$$

$$\frac{d\,SSR}{d\,w_4} = \frac{d\,SSR}{d\,Predicted} \times \frac{d\,Predicted}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{2,i}$$

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

…we can plug them into **Gradient Descent** to optimize w_3, w_4 and b_3.

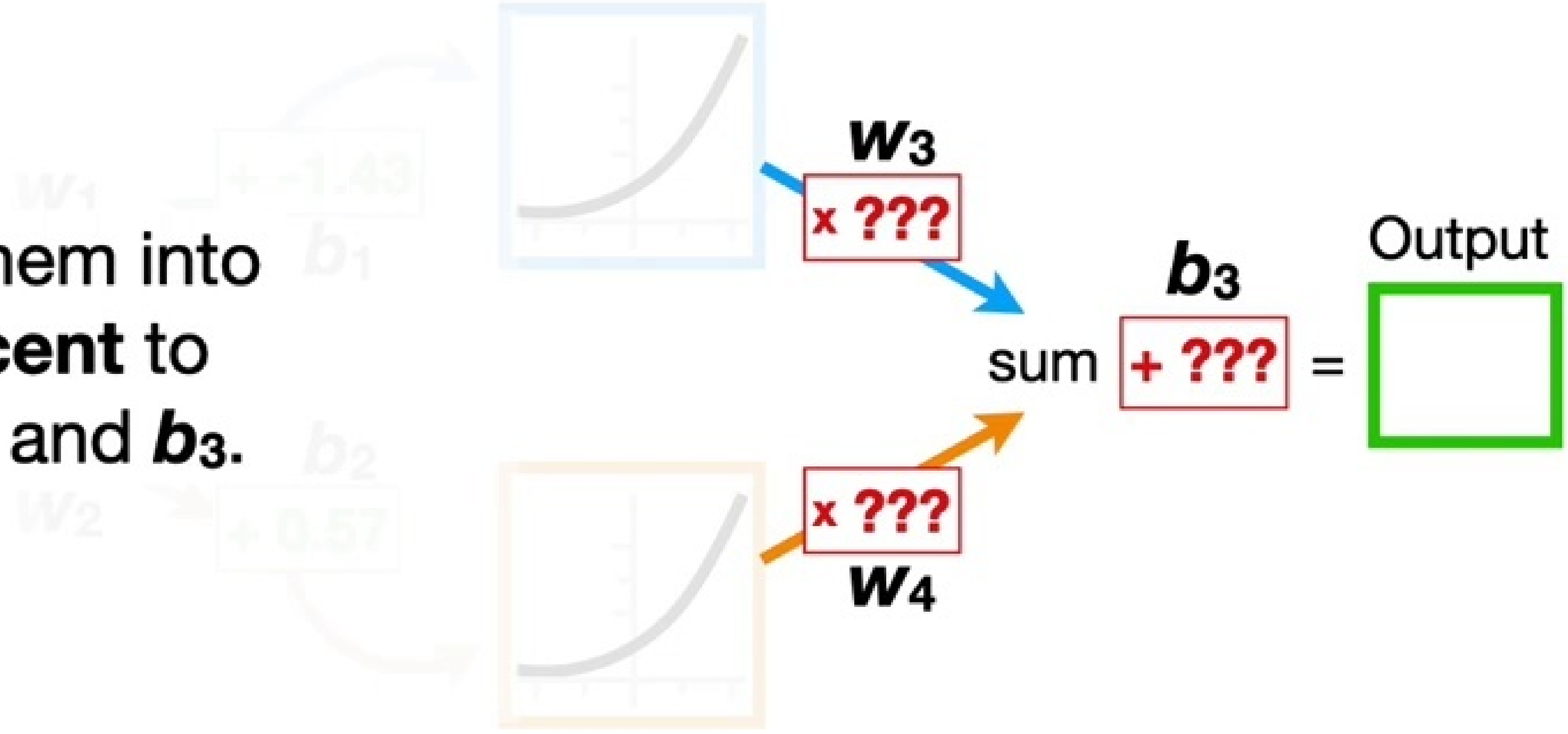


$$\frac{d\,SSR}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{1,i}$$

$$\frac{d\,SSR}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{2,i}$$

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

…we can plug them into **Gradient Descent** to optimize w_3, w_4 and b_3.

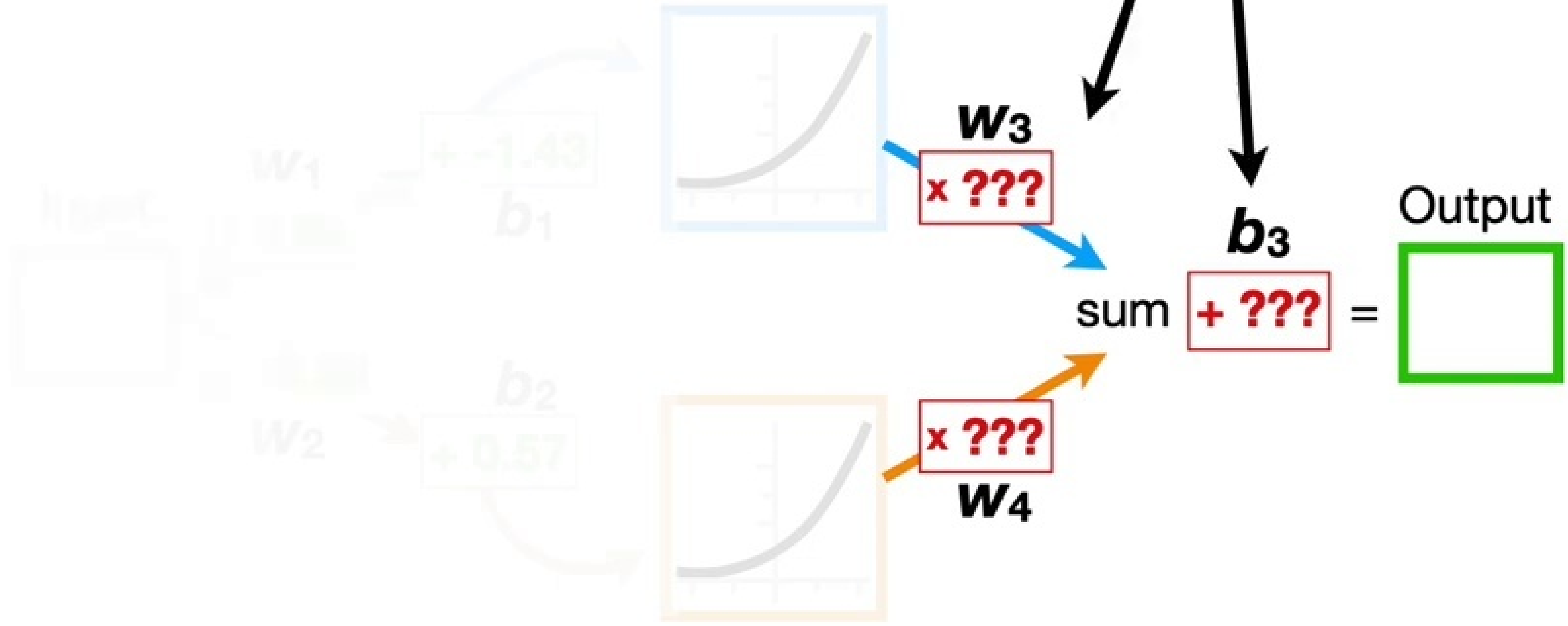


$$\frac{d\,SSR}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{1,i}$$

$$\frac{d\,SSR}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{2,i}$$

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

First, we initialize w_3 and w_4 with random values and set $b_3 = 0$.

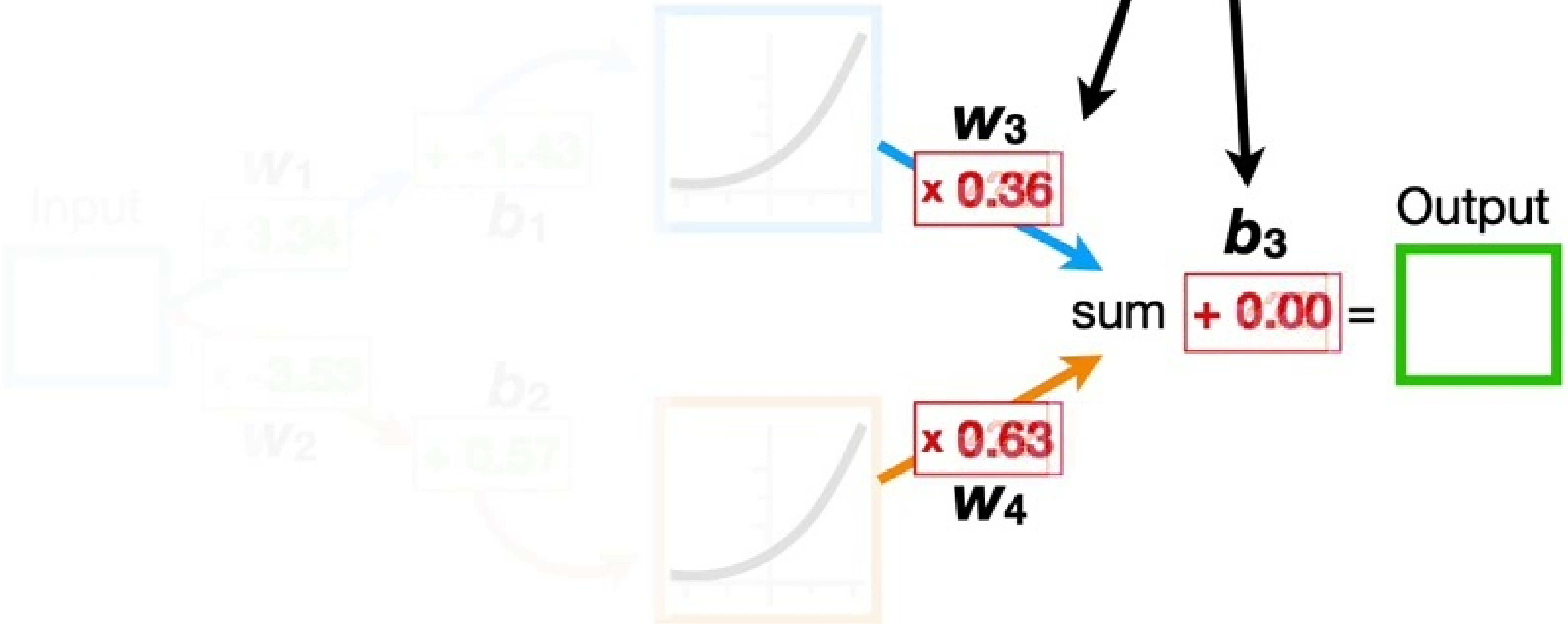


$$\frac{d\,SSR}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{1,i}$$

$$\frac{d\,SSR}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{2,i}$$

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

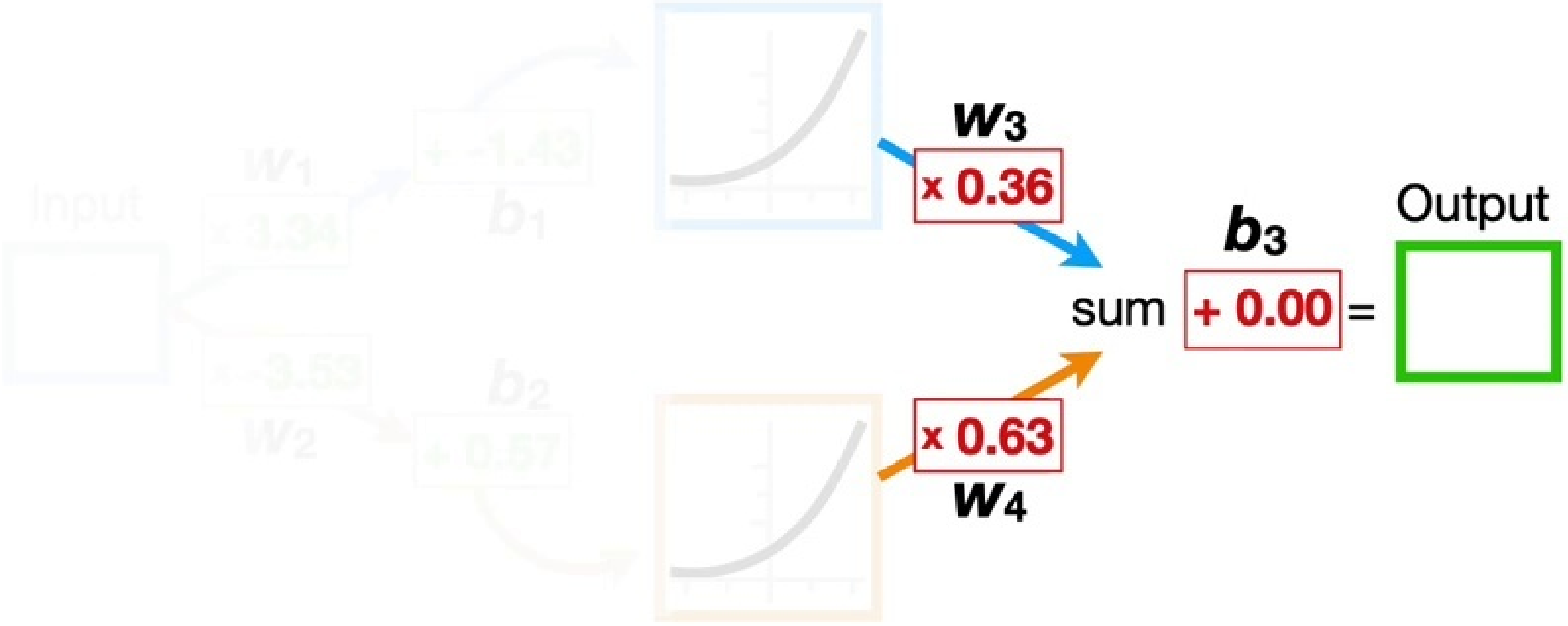
First, we initialize w_3 and w_4 with random values and set $b_3 = 0$.



$$\frac{d\,SSR}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{1,i}$$

Now, starting with the derivative of the **SSR** with respect to ***w*₃**…

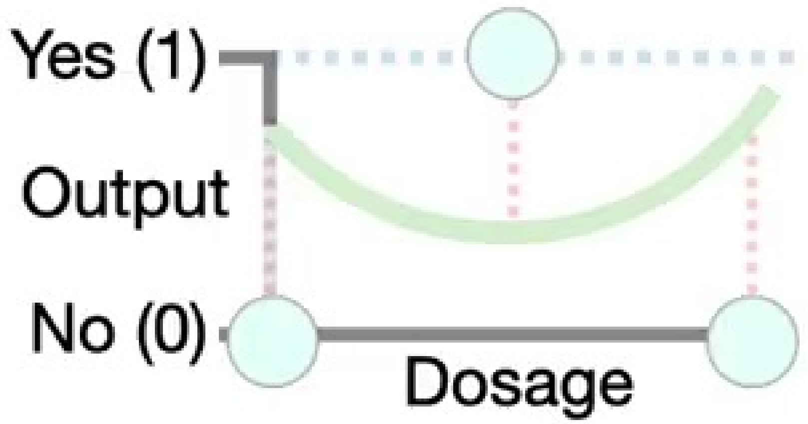
(**NOTE:** It does not matter which derivative we start with.)



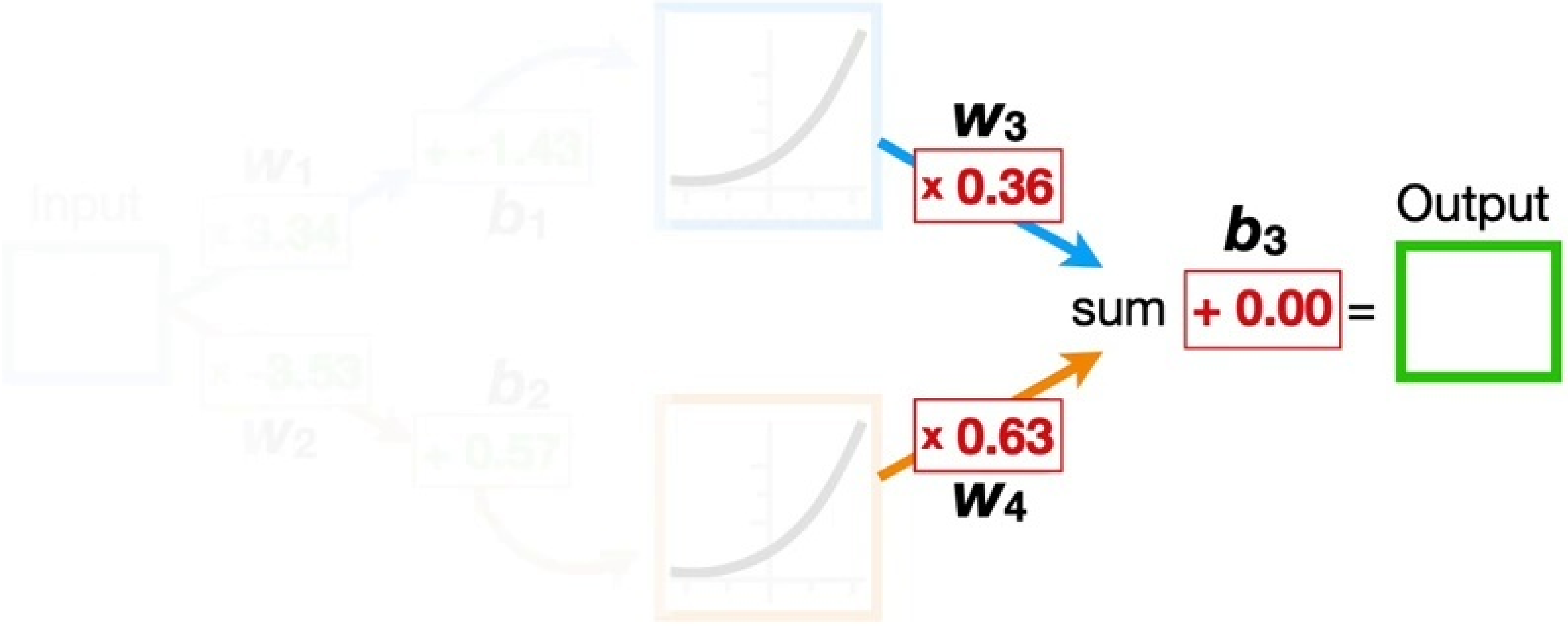
$$\frac{d\,SSR}{d\,w_3} = -2 \times (\text{Observed}_1 - \text{Predicted}_1) \times y_{1,1}$$
$$+ -2 \times (\text{Observed}_2 - \text{Predicted}_2) \times y_{1,2}$$
$$+ -2 \times (\text{Observed}_3 - \text{Predicted}_3) \times y_{1,3}$$

First, we expand the summation.

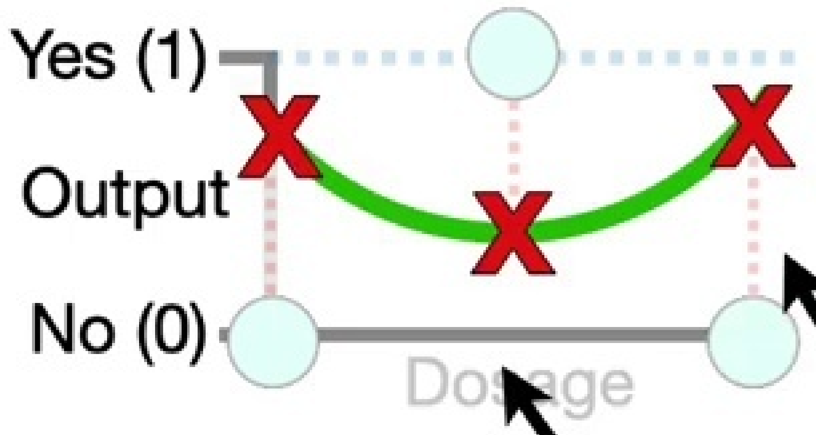
$$\frac{d\,SSR}{d\,w_3} = -2 \times (\text{Observed}_1 - \text{Predicted}_1) \times y_{1,1} + -2 \times (\text{Observed}_2 - \text{Predicted}_2) \times y_{1,2} + -2 \times (\text{Observed}_3 - \text{Predicted}_3) \times y_{1,3}$$



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

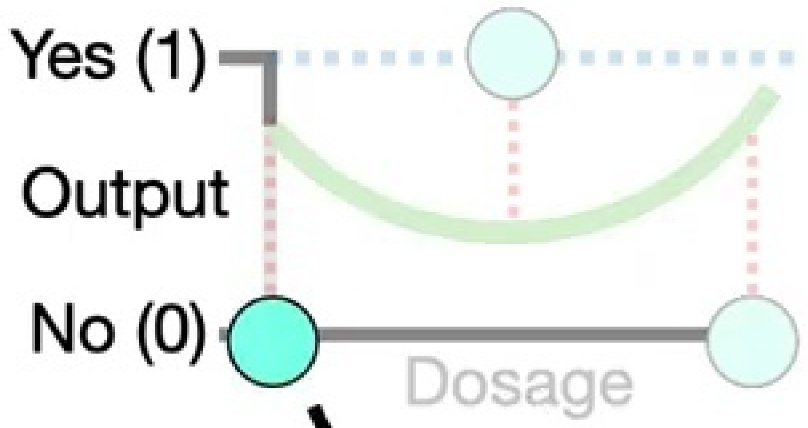


$$\frac{d\,SSR}{d\,w_3} = -2 \times (0 - \text{Predicted}_1) \times y_{1,1} + -2 \times (1 - \text{Predicted}_2) \times y_{1,2} + -2 \times (0 - \text{Predicted}_3) \times y_{1,3}$$

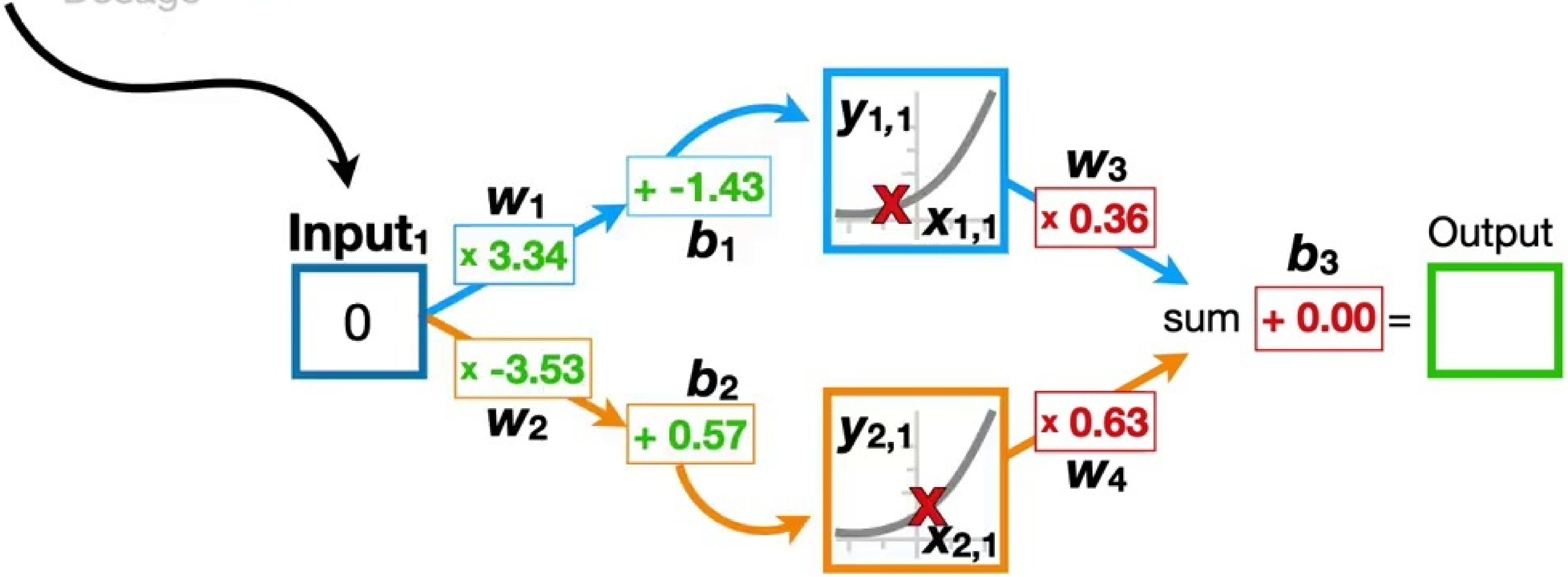


...and plug in **Predicted** values from the **green squiggle**.

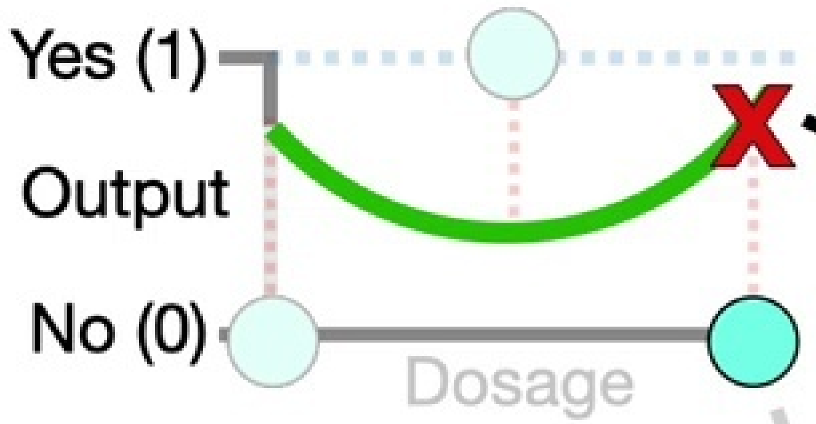
$$\frac{d\ SSR}{d\ w_3} = -2 \times (0 - \text{Predicted}_1) \times y_{1,1} + -2 \times (1 - \text{Predicted}_2) \times y_{1,2} + -2 \times (0 - \text{Predicted}_3) \times y_{1,3}$$



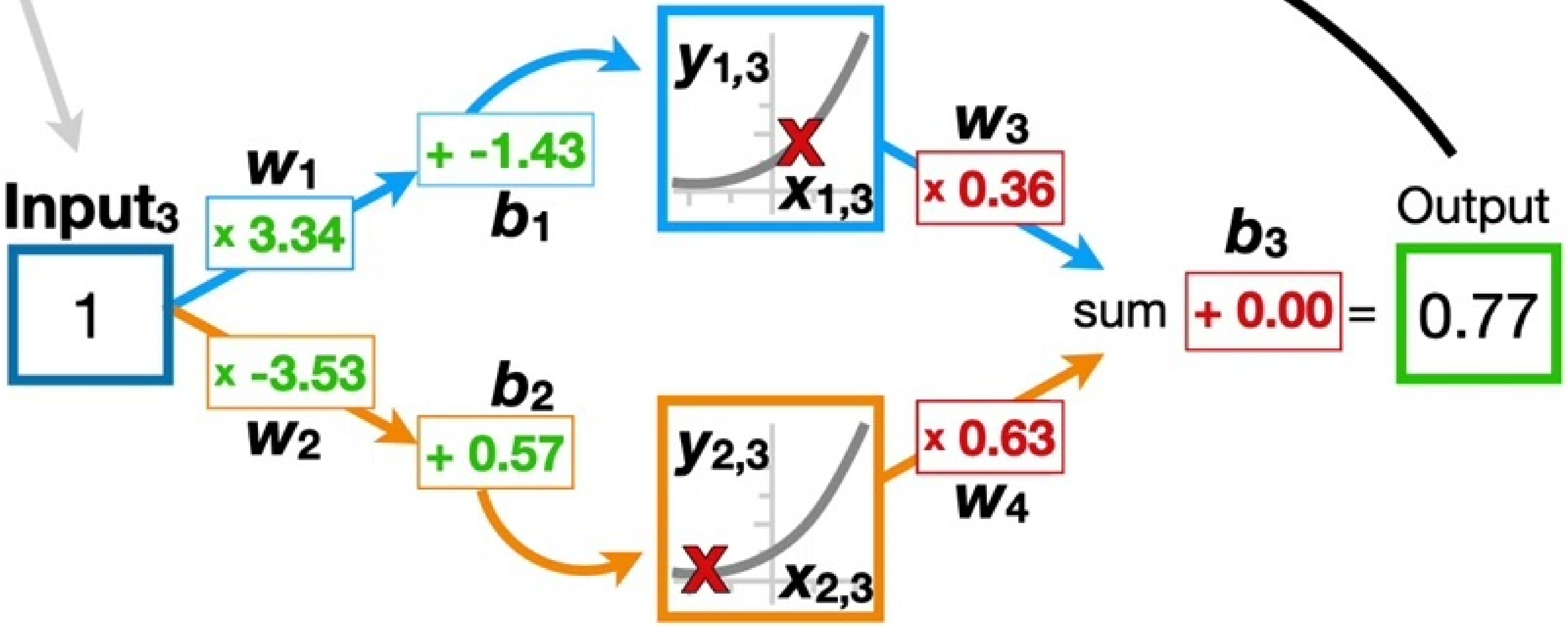
Remember, we get the **Predicted** values on the **green squiggle** by running the **Dosages** through the **Neural Network.**



$$\frac{d\ SSR}{d\ w_3} = -2 \times (0 - 0.72) \times y_{1,1} + -2 \times (1 - 0.46) \times y_{1,2} + -2 \times (0 - \text{Predicted}_3) \times y_{1,3}$$

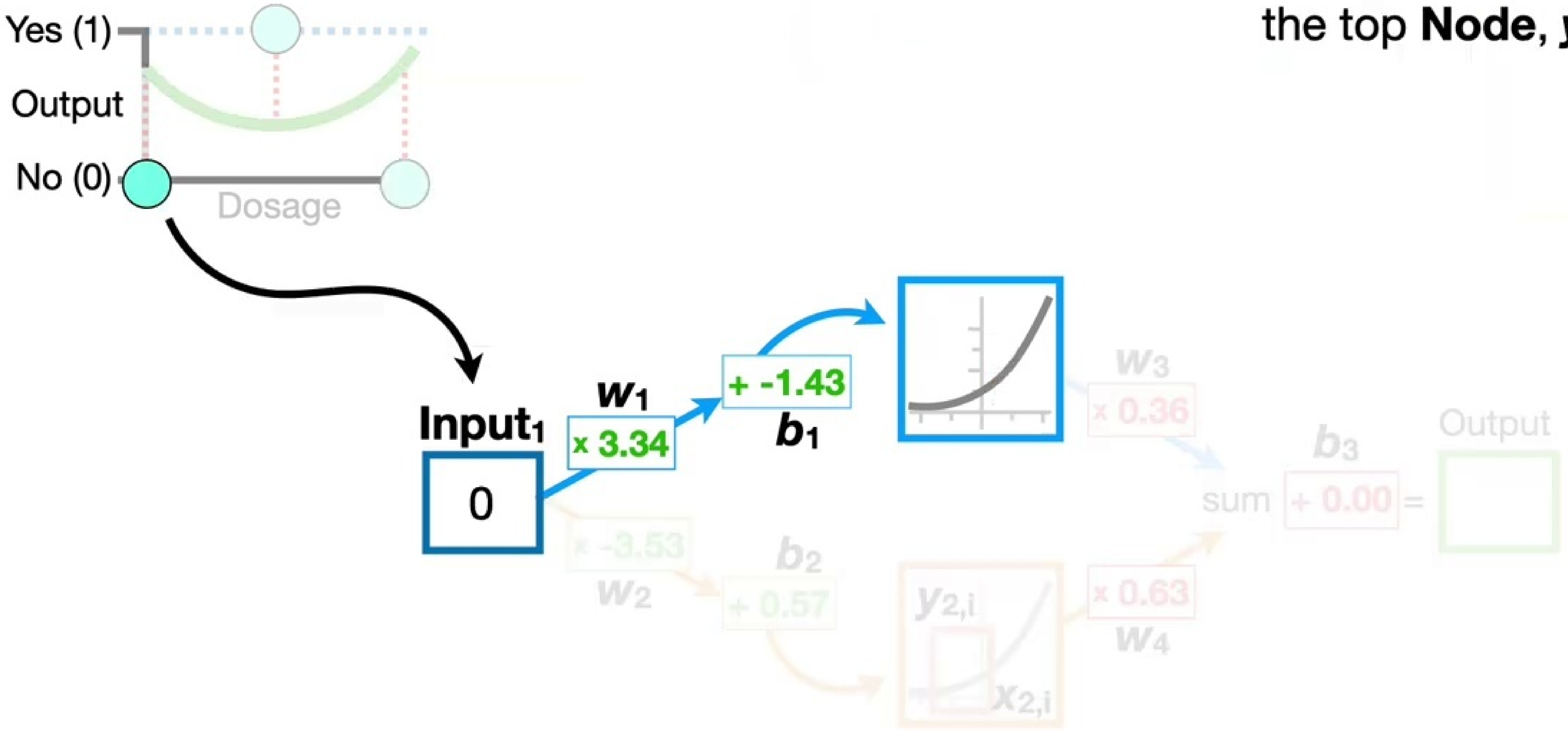


Remember, we get the **Predicted** values on the **green squiggle** by running the **Dosages** through the **Neural Network**.



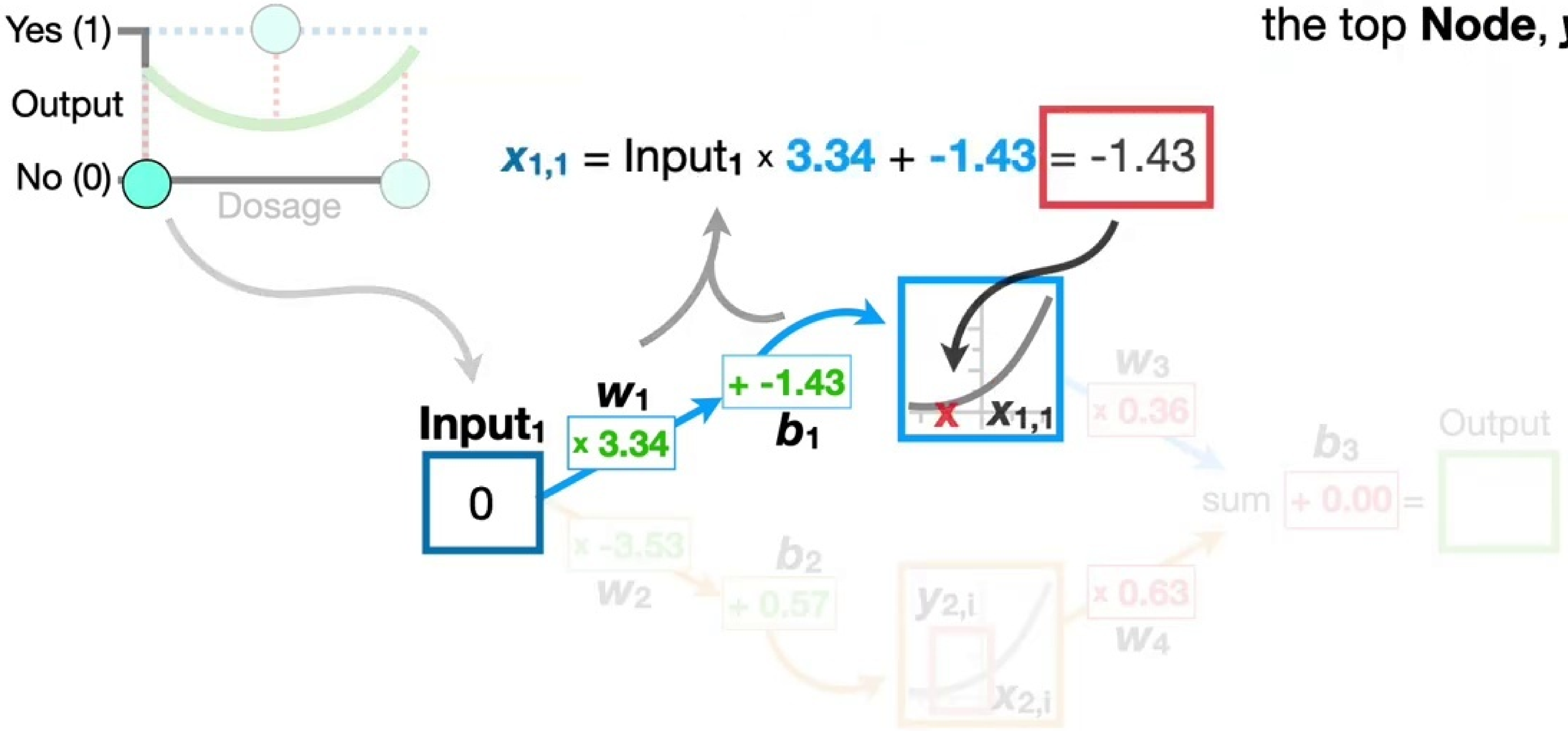
$$\frac{d\ SSR}{d\ w_3} = -2 \times (0 - 0.72) \times y_{1,1} + -2 \times (1 - 0.46) \times y_{1,2} + -2 \times (0 - 0.77) \times y_{1,3}$$

Now we plug in the y-axis coordinates for the **Activation Function** in the top **Node**, $y_{1,i}$.



$$\frac{d\,SSR}{d\,w_3} = -2 \times (0 - 0.72) \times y_{1,1} + -2 \times (1 - 0.46) \times y_{1,2} + -2 \times (0 - 0.77) \times y_{1,3}$$

Now we plug in the y-axis coordinates for the **Activation Function** in the top **Node**, $y_{1,i}$.

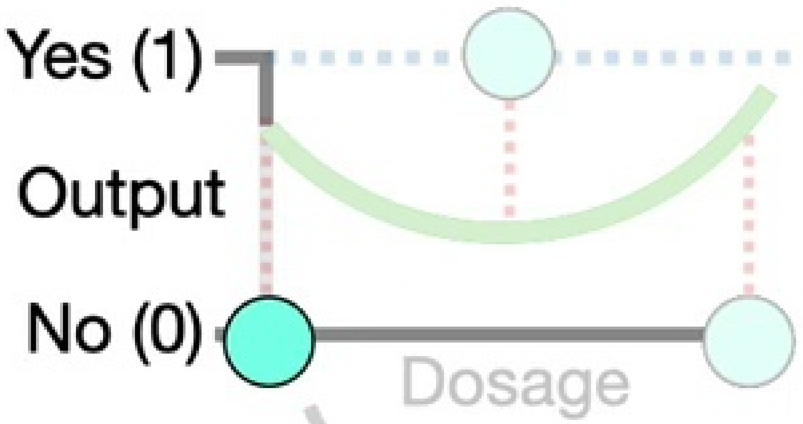


$$\frac{d\,SSR}{d\,w_3} = -2 \times (0 - 0.72) \times 0.21$$

$$+ -2 \times (1 - 0.46) \times y_{1,2}$$

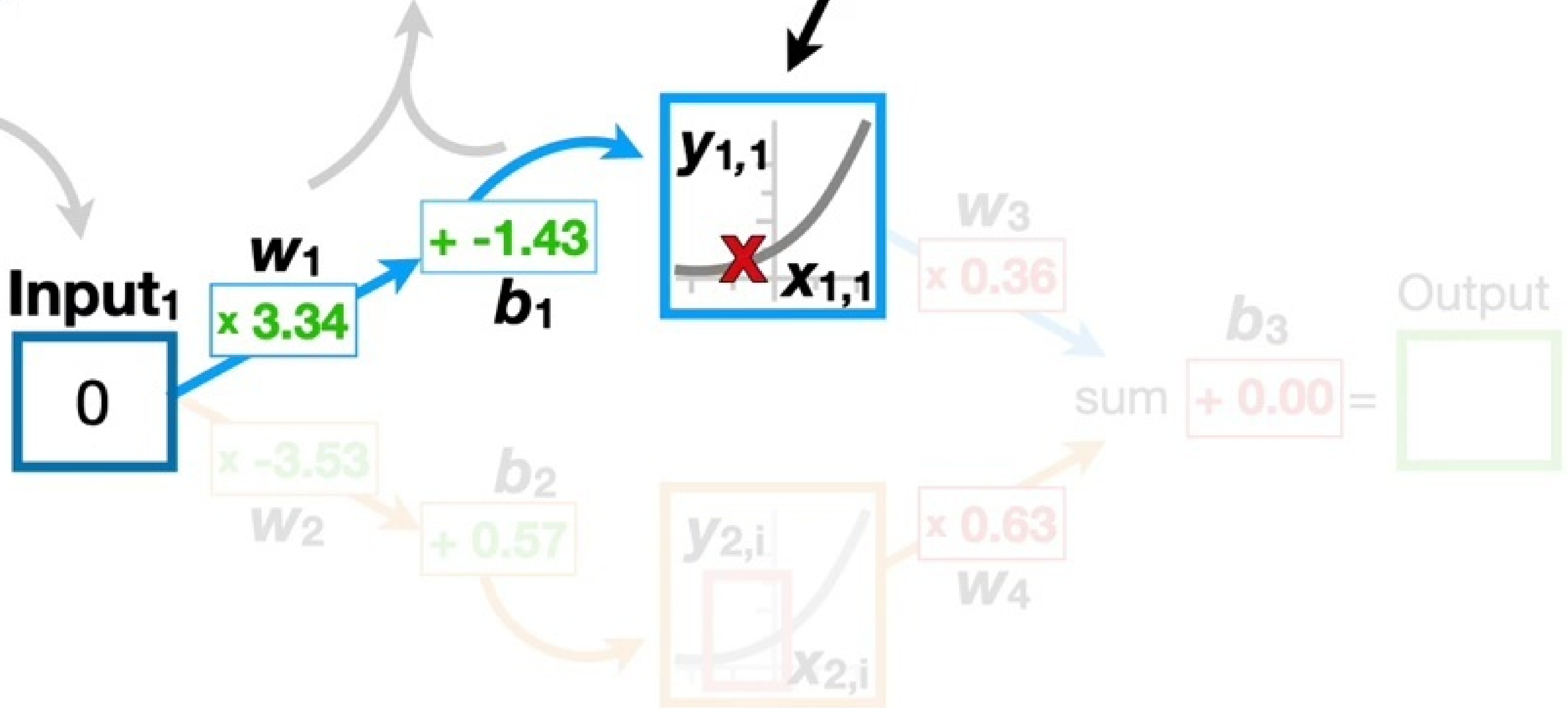
$$+ -2 \times (0 - 0.77) \times y_{1,3}$$

Now we plug in the y-axis coordinates for the **Activation Function** in the top **Node**, $\boldsymbol{y_{1,i}}$.



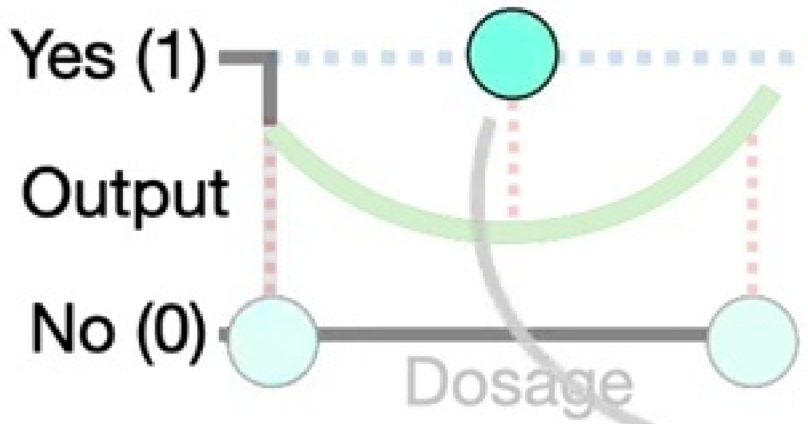
$$y_{1,1} = f(x_{1,1}) = \log(1 + e^x) \boxed{= 0.21}$$

$$x_{1,1} = \text{Input}_1 \times 3.34 + -1.43 = -1.43$$



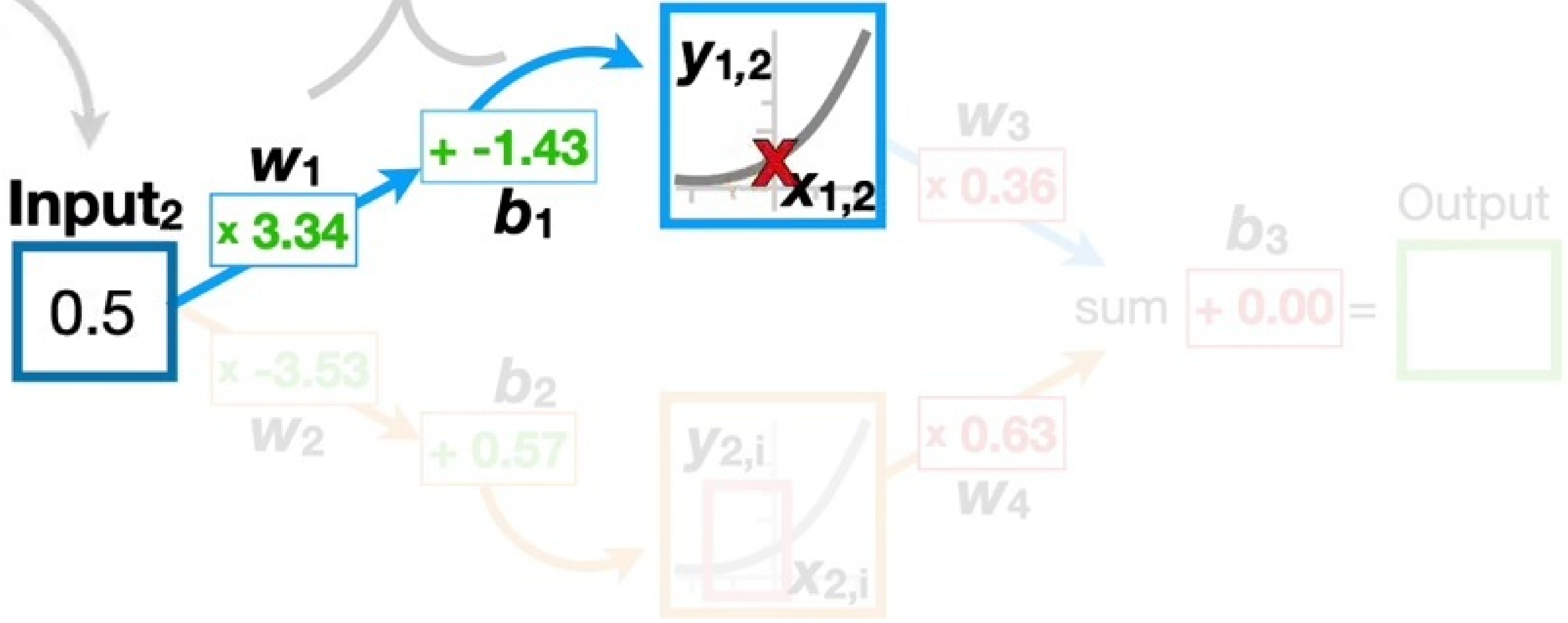
$$\frac{d\,SSR}{d\,w_3} = -2 \times (0 - 0.72) \times 0.21 + -2 \times (1 - 0.46) \times 0.82 + -2 \times (0 - 0.77) \times y_{1,3}$$

Now we plug in the y-axis coordinates for the **Activation Function** in the top **Node**, $y_{1,i}$.



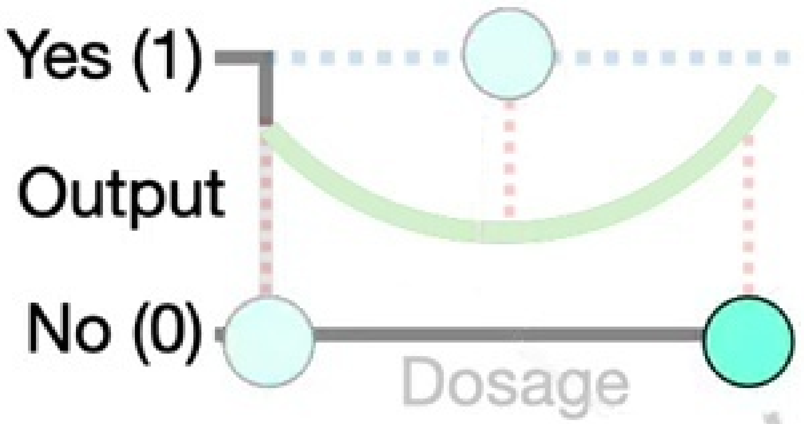
$$y_{1,2} = f(x_{1,2}) = \log(1 + e^x) \boxed{= 0.82}$$

$$x_{1,2} = \text{Input}_2 \times 3.34 + -1.43 = 0.24$$



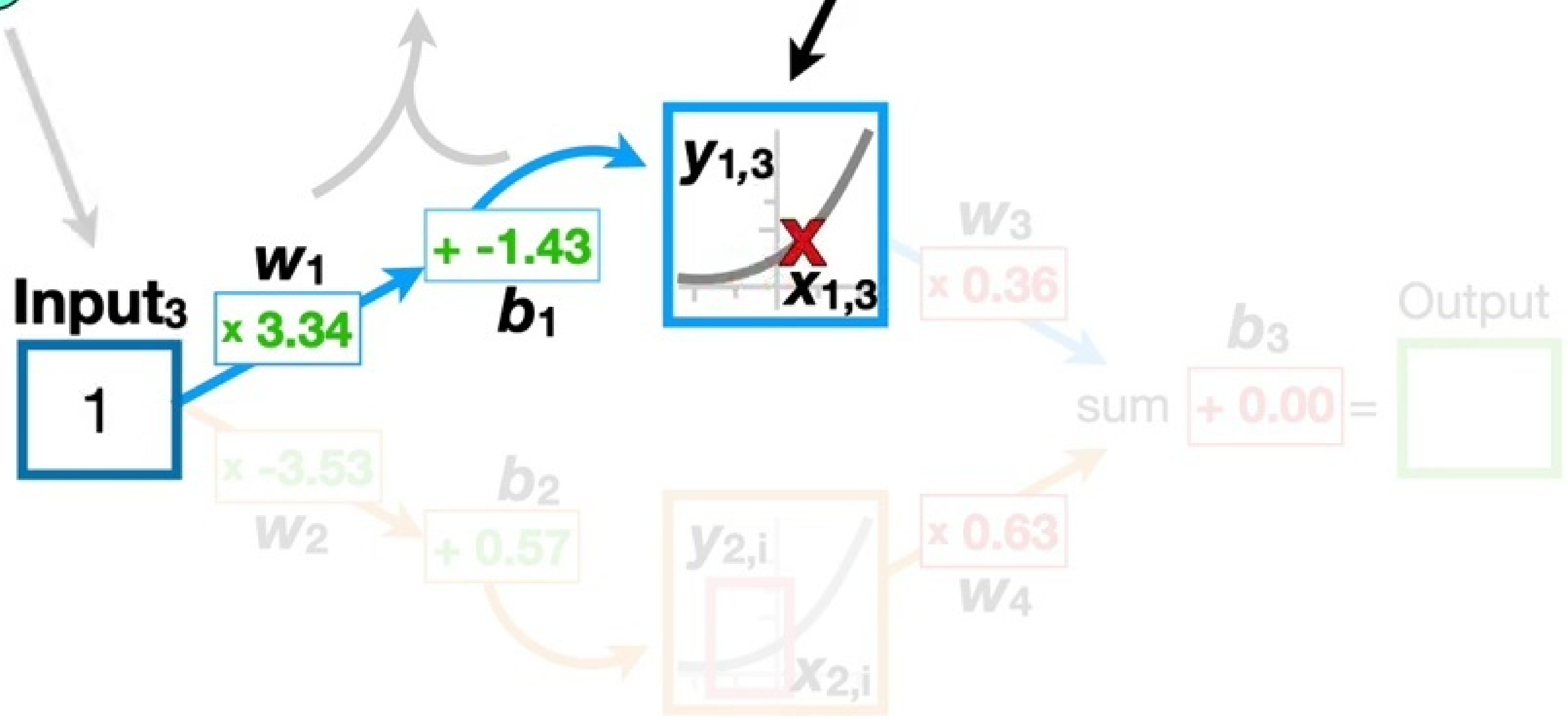
$$\frac{d\,SSR}{d\,w_3} = -2 \times (0 - 0.72) \times 0.21$$
$$+ -2 \times (1 - 0.46) \times 0.82$$
$$+ -2 \times (0 - 0.77) \times 2.04$$

Now we plug in the y-axis coordinates for the **Activation Function** in the top **Node, $y_{1,i}$.**



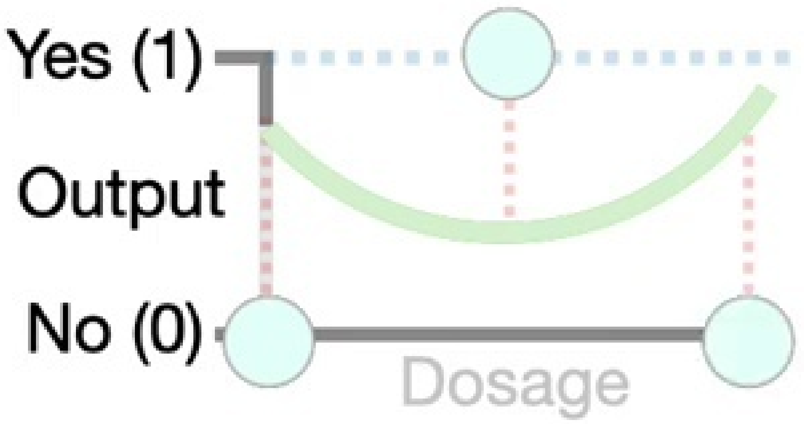
$$y_{1,3} = f(x_{1,3}) = \log(1 + e^x) \boxed{= 2.04}$$

$$x_{1,3} = \text{Input}_3 \times 3.34 + -1.43 = 1.91$$



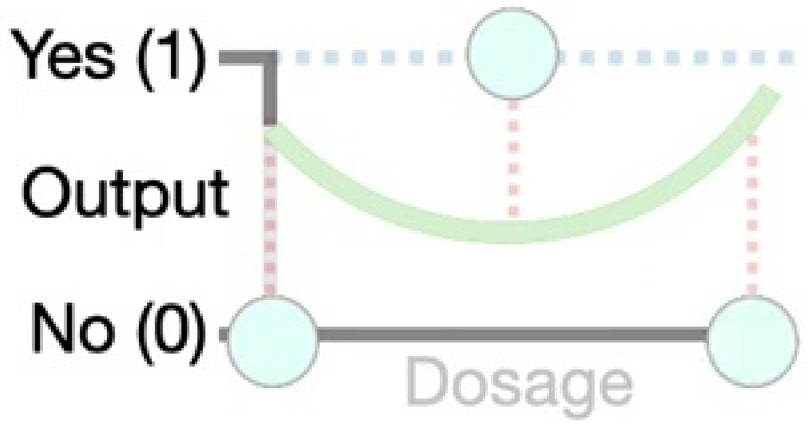
$$\frac{d\ SSR}{d\ w_3} = -2 \times (0 - 0.72) \times \mathbf{0.21} + -2 \times (1 - 0.46) \times \mathbf{0.82} + -2 \times (0 - 0.77) \times \mathbf{2.04}$$

Lastly, we do the math.



$$\frac{d\,SSR}{d\,w_3} = -2 \times (0 - 0.72) \times 0.21 + -2 \times (1 - 0.46) \times 0.82 + -2 \times (0 - 0.77) \times 2.04 \boxed{= 2.58}$$

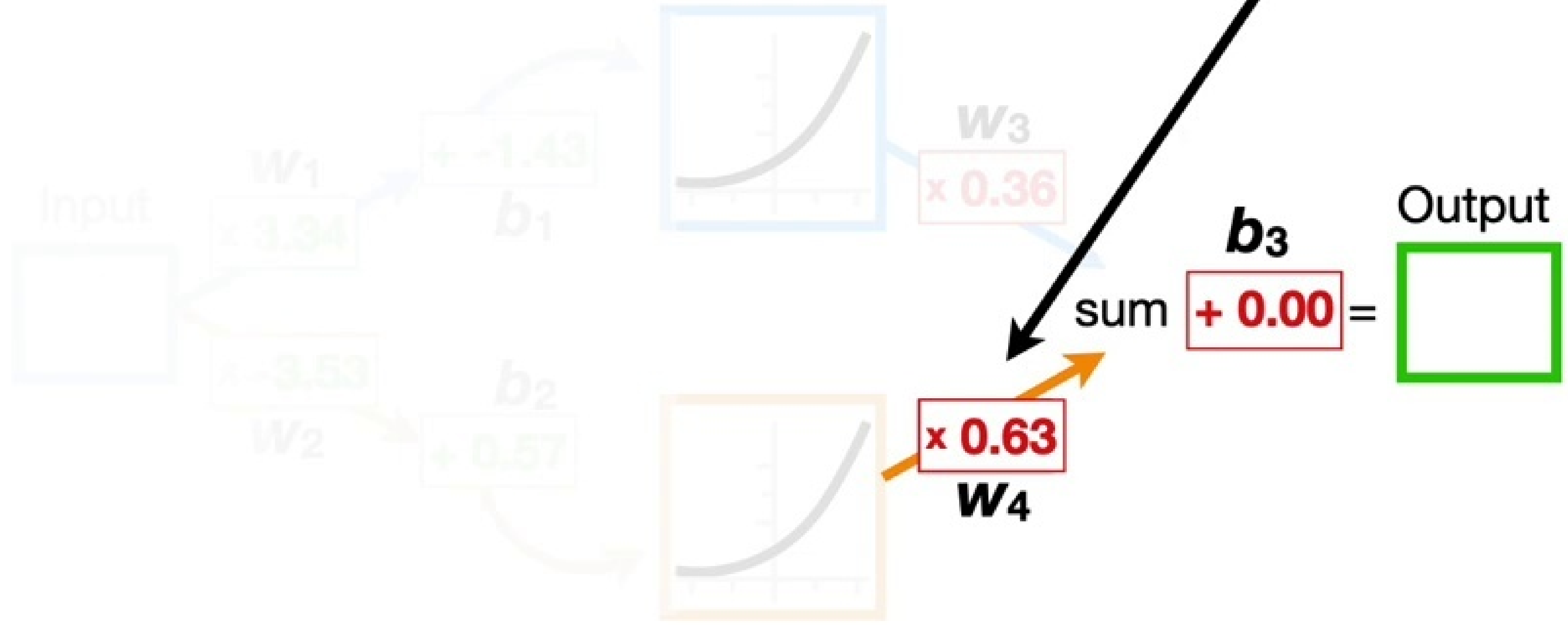
Lastly, we do the math.



$$\frac{d\ SSR}{d\ w_3} = 2.58$$

$$\frac{d\ SSR}{d\ w_4} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{2,i}$$

Likewise, we calculate the derivative of the **SSR** with respect to $\boldsymbol{w_4}$…



$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

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

…and with respect to b_3.

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

Step Size = Derivative × Learning Rate

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

New w_3 = Old w_3 - Step Size

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Now we use the derivatives to calculate the new values for w_3…

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

Step Size = 2.58 × **Learning Rate**

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

New w_3 = Old w_3 - Step Size

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Now we use the derivatives to calculate the new values for $\boldsymbol{w_3}$…

$$\frac{d\ SSR}{d\ w_3} = 2.58$$

$$\frac{d\ SSR}{d\ w_4} = 1.26$$

$$\frac{d\ SSR}{d\ b_3} = 1.90$$

Step Size = 2.58 × **Learning Rate**

New w_3 = Old w_3 - Step Size

NOTE: In this example we've set the **Learning Rate** to **0.1**.

Now we use the derivatives to calculate the new values for w_3…

$$\frac{d\ SSR}{d\ w_3} = 2.58$$

$$\frac{d\ SSR}{d\ w_4} = 1.26$$

$$\frac{d\ SSR}{d\ b_3} = 1.90$$

Step Size = 2.58 × 0.1

New w_3 = Old w_3 - Step Size

NOTE: In this example we've set the **Learning Rate** to **0.1**.

Now we use the derivatives to calculate the new values for $\boldsymbol{w_3}$…

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 2.58 × 0.1 = 0.258

New w_3 = Old w_3 - Step Size

NOTE: In this example we've set the **Learning Rate** to **0.1**.

Now we use the derivatives to calculate the new values for w_3…

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

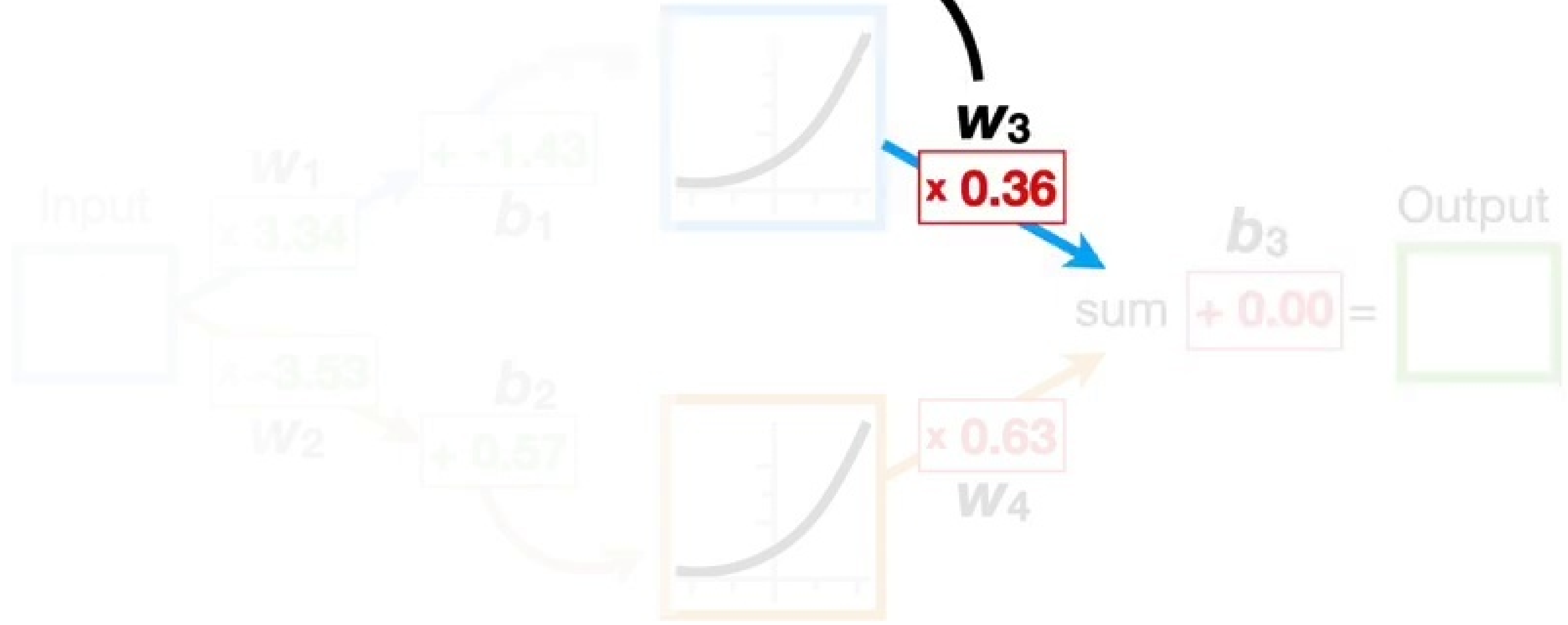
$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 2.58 × 0.1 = 0.258

New w_3 = Old w_3 - Step Size

Now we use the derivatives to calculate the new values for $\boldsymbol{w_3}$…



$$\frac{d\,SSR}{d\,w_3} = 2.58$$

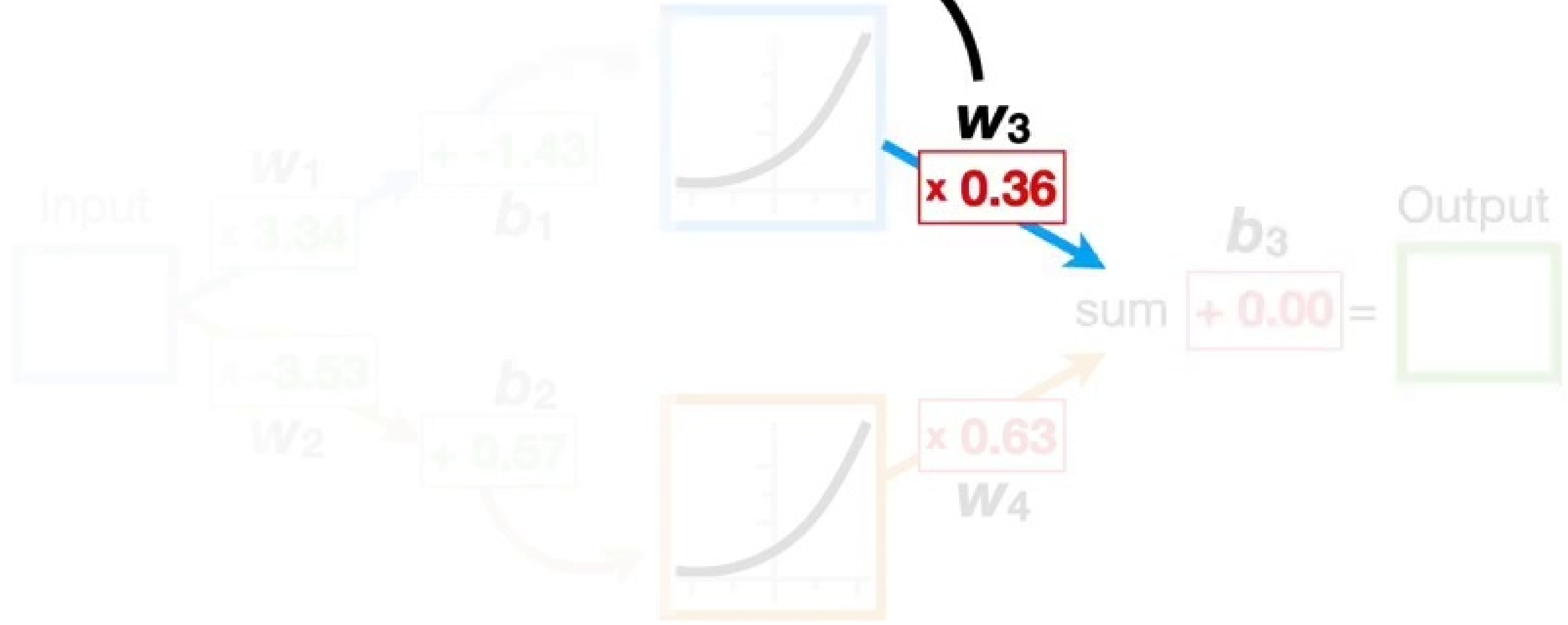
$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 2.58 × 0.1 = 0.258

New w_3 = 0.36 - Step Size

Now we use the derivatives to calculate the new values for w_3…



$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 2.58 × 0.1 $\boxed{= 0.258}$

New w_3 = 0.36 - $\boxed{\textbf{Step Size}}$

Now we use the derivatives to calculate the new values for w_3…

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 2.58 × 0.1 $\boxed{= 0.258}$

New w_3 = **0.36** - $\boxed{0.258}$

Now we use the derivatives to calculate the new values for w_3…

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

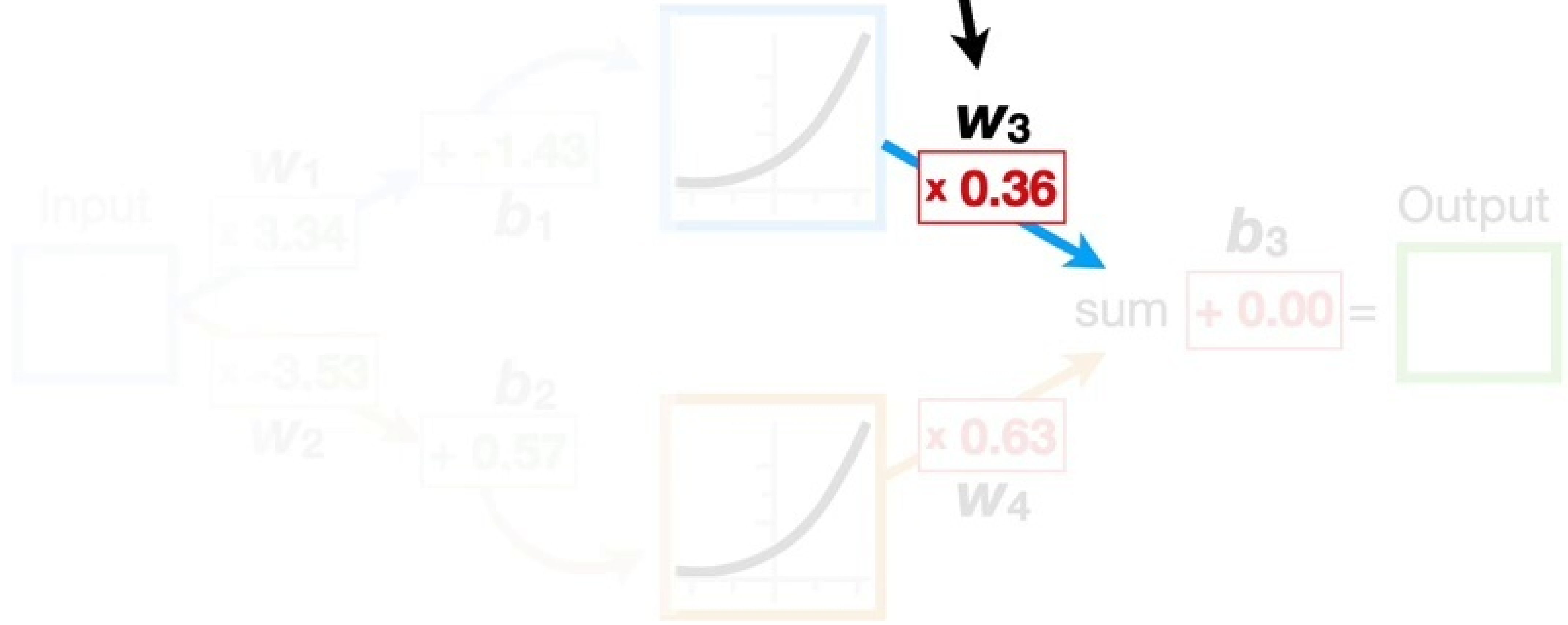
$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 2.58 × 0.1 = 0.258

New w_3 = **0.36** - 0.258 $\boxed{= 0.10}$

Now we use the derivatives to calculate the new values for w_3…



$$\frac{d\ SSR}{d\ w_3} = 2.58$$

$$\frac{d\ SSR}{d\ w_4} = 1.26$$

$$\frac{d\ SSR}{d\ b_3} = 1.90$$

Step Size = 2.58 × 0.1 = 0.258

New w_3 = **0.36** - 0.258 $\boxed{= 0.10}$

Now we use the derivatives to calculate the new values for w_3…

$$\frac{d\ SSR}{d\ w_3} = 2.58$$

$$\frac{d\ SSR}{d\ w_4} = 1.26$$

Step Size = Derivative × Learning Rate

New w_4 = Old w_4 - Step Size

...w_4...

$$\frac{d\ SSR}{d\ b_3} = 1.90$$

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

Step Size = 1.26 × **Learning Rate**

New w_4 = Old w_4 - Step Size

...w_4...

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

Step Size = 1.26 × 0.1

New w_4 = Old w_4 - Step Size

...w_4...

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

$$\frac{d\ SSR}{d\ w_3} = 2.58$$

$$\frac{d\ SSR}{d\ w_4} = 1.26$$

Step Size = 1.26 × 0.1 = 0.126

New w_4 = Old w_4 - Step Size

...w_4...

$$\frac{d\ SSR}{d\ b_3} = 1.90$$

$$\frac{d\ SSR}{d\ w_3} = 2.58$$

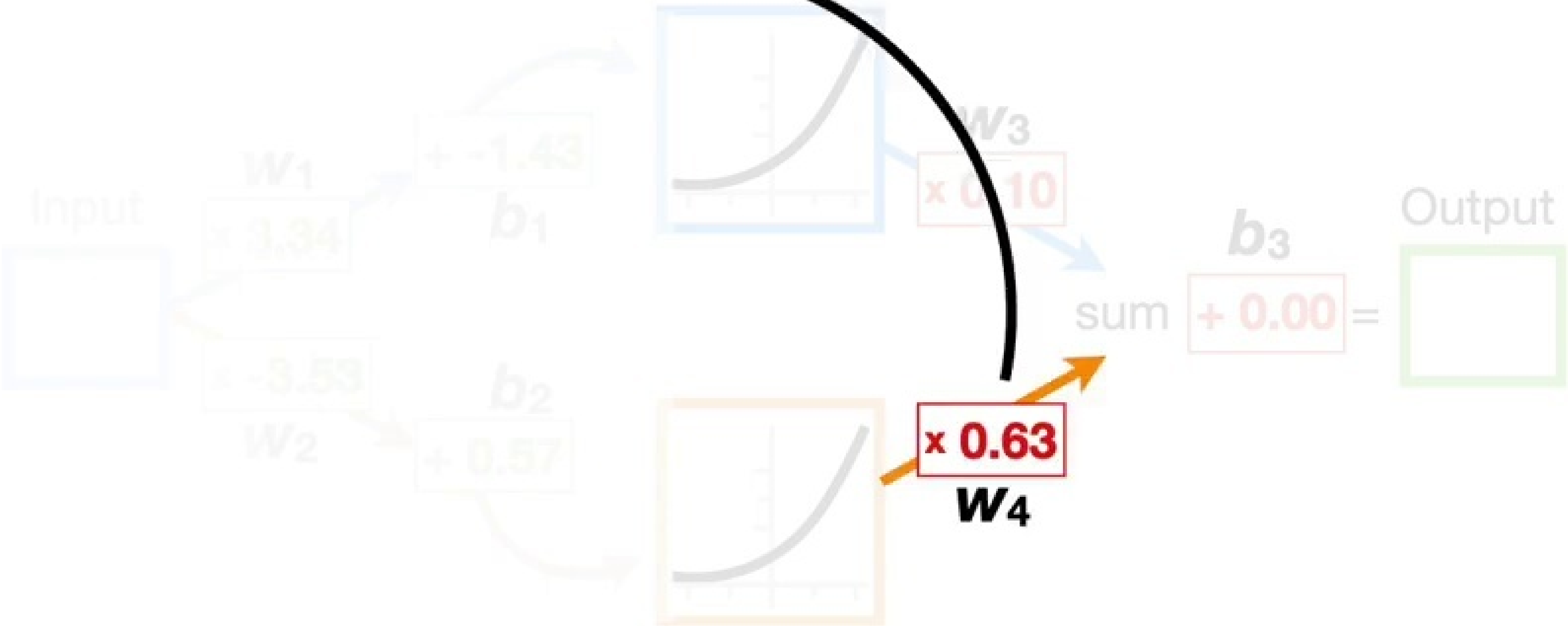
Step Size = 1.26 × 0.1 = 0.126

$$\frac{d\ SSR}{d\ w_4} = 1.26$$

New w_4 = Old w_4 - Step Size

...w_4...

$$\frac{d\ SSR}{d\ b_3} = 1.90$$



$$\frac{d\,SSR}{d\,w_3} = 2.58$$

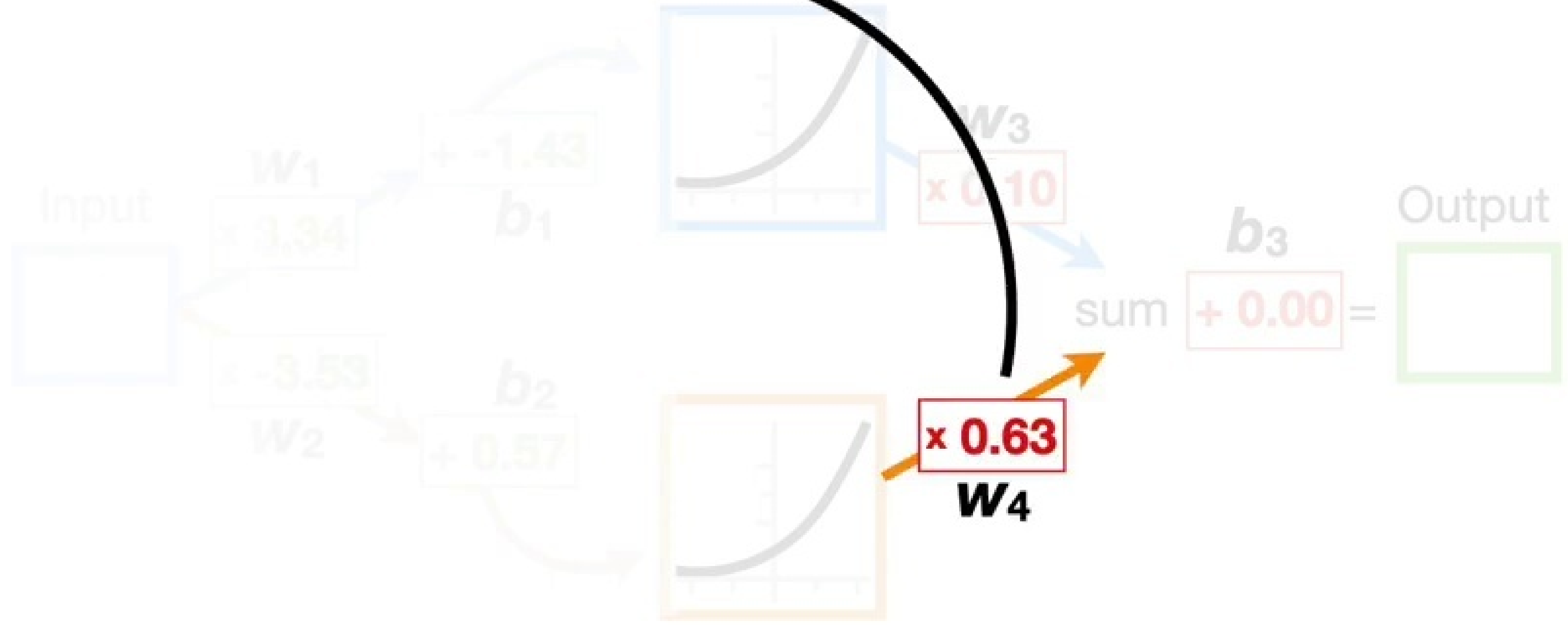
Step Size = 1.26 × 0.1 = 0.126

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

New w_4 = 0.63 - Step Size

...w_4...

$$\frac{d\,SSR}{d\,b_3} = 1.90$$



$$\frac{d\,SSR}{d\,w_3} = 2.58$$

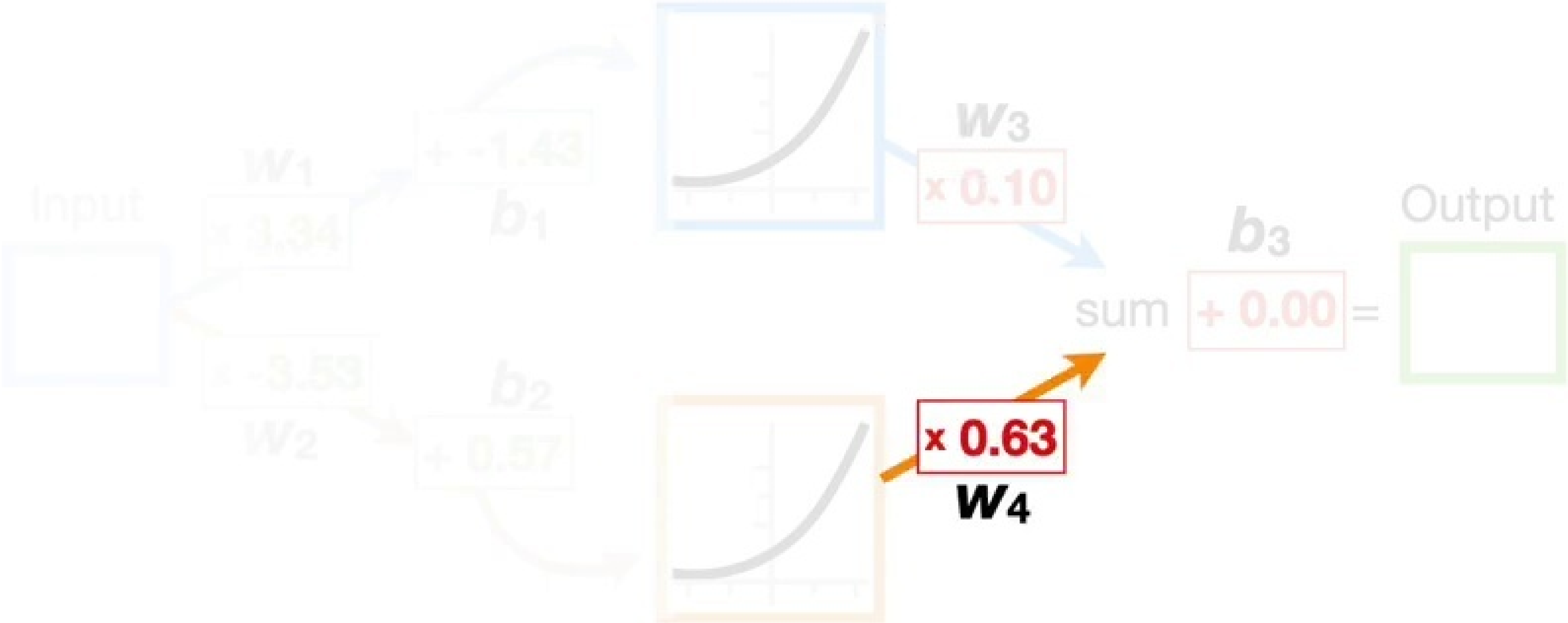
$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\text{Step Size} = 1.26 \times 0.1 = \boxed{0.126}$$

$$\text{New } w_4 = 0.63 - \boxed{\text{Step Size}}$$

$\dots w_4 \dots$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$



$$\frac{d\,SSR}{d\,w_3} = 2.58$$

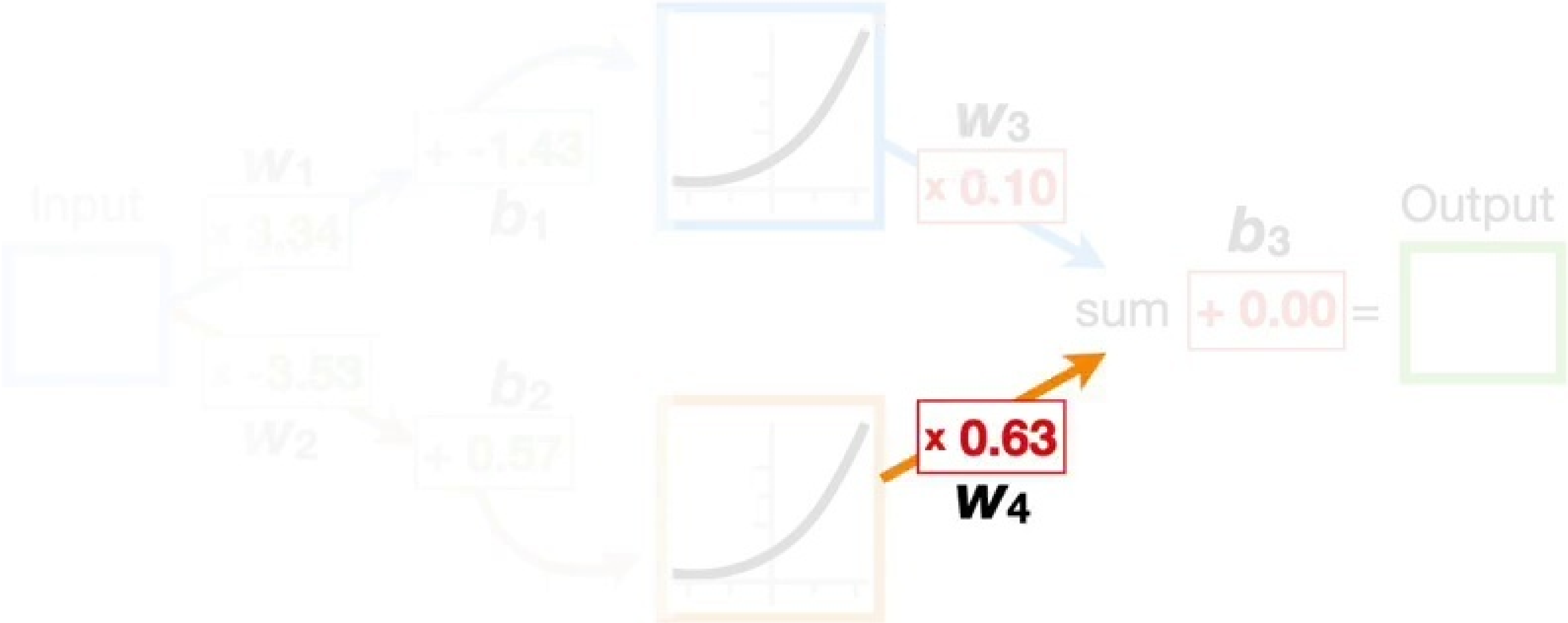
$$\frac{d\,SSR}{d\,w_4} = 1.26$$

Step Size = 1.26 x 0.1 $\boxed{= 0.126}$

New w_4 = **0.63** - $\boxed{0.126}$

...w_4...

$$\frac{d\,SSR}{d\,b_3} = 1.90$$



$$\frac{d\ SSR}{d\ w_3} = 2.58$$

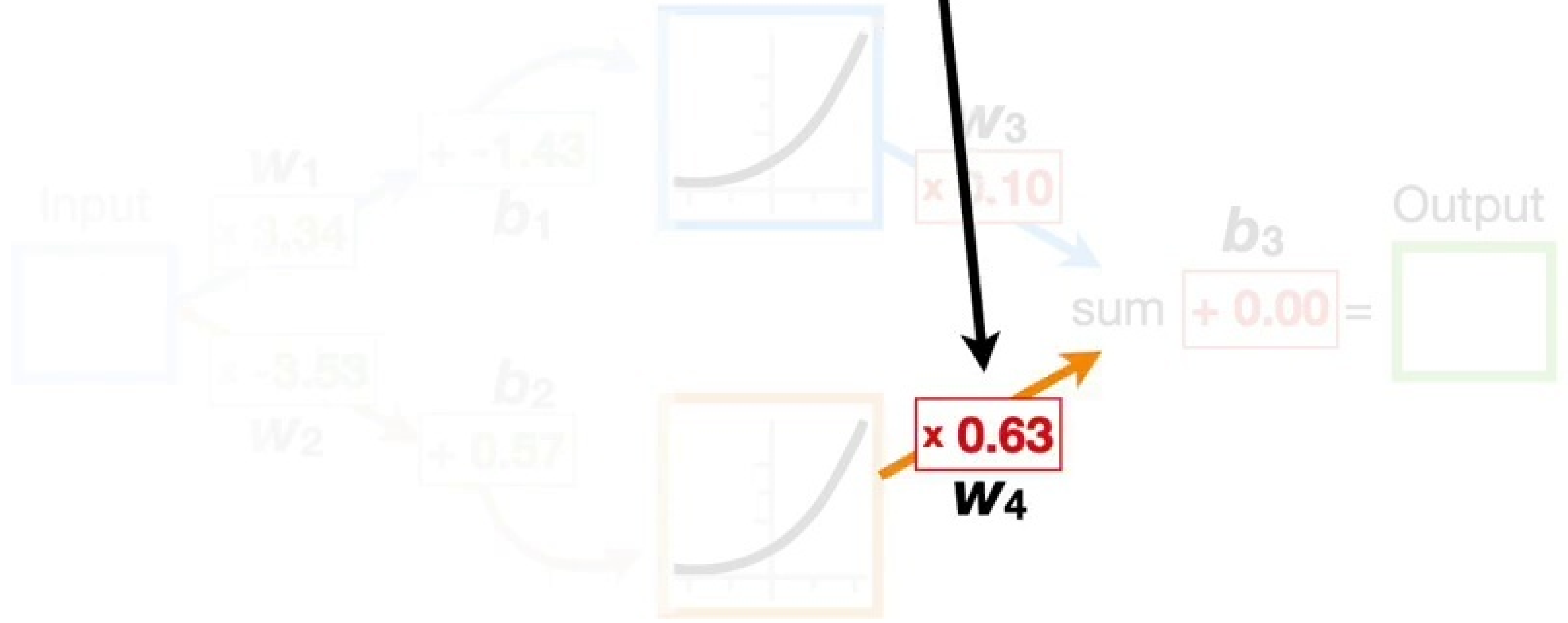
$$\frac{d\ SSR}{d\ w_4} = 1.26$$

$$\frac{d\ SSR}{d\ b_3} = 1.90$$

Step Size = 1.26 × 0.1 = 0.126

New w_4 = 0.63 - 0.126 $\boxed{= 0.50}$

…w_4…



$$\frac{d\,SSR}{d\,w_3} = 2.58$$

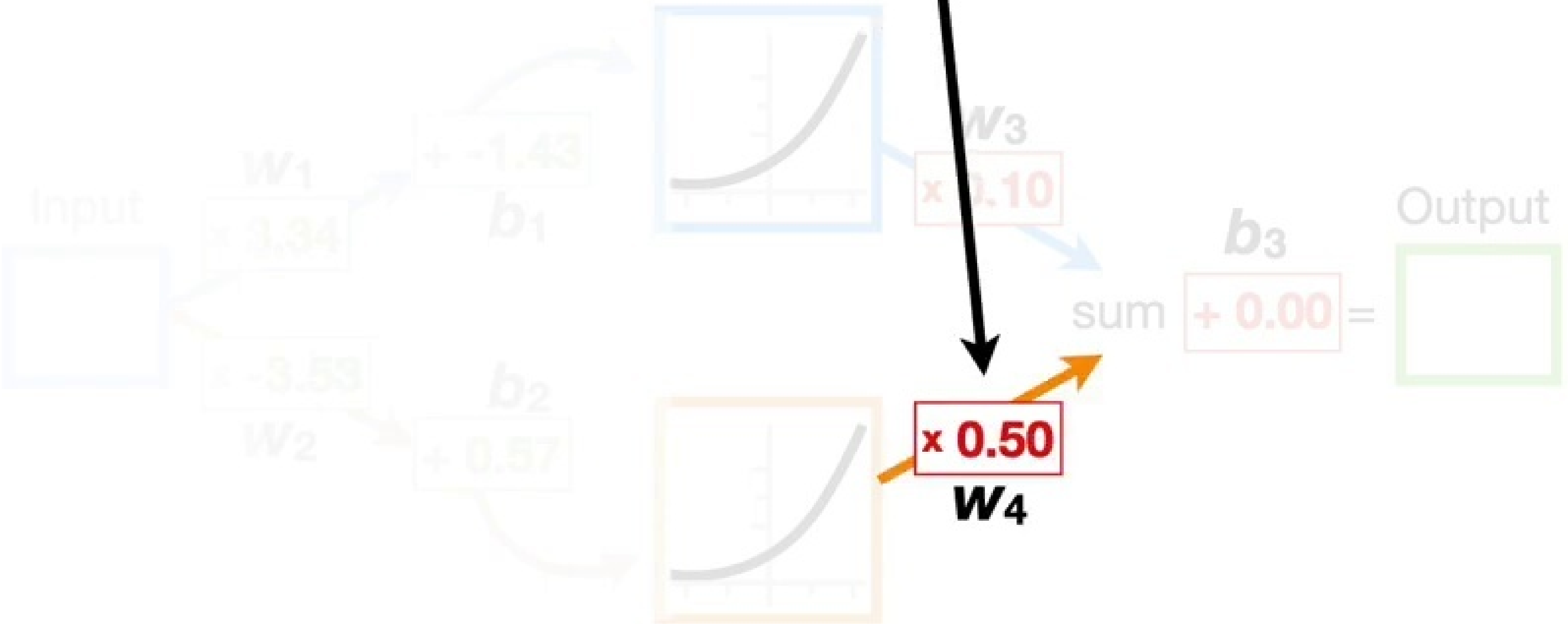
Step Size = 1.26 × 0.1 = 0.126

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

New w_4 = 0.63 - 0.126 $\boxed{= 0.50}$

...w_4...

$$\frac{d\,SSR}{d\,b_3} = 1.90$$



$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = Derivative × Learning Rate

New b_3 = Old b_3 - Step Size

…and b_3.

$$\frac{d\ SSR}{d\ w_3} = 2.58$$

$$\frac{d\ SSR}{d\ w_4} = 1.26$$

$$\frac{d\ SSR}{d\ b_3} = 1.90$$

Step Size = 1.90 × **Learning Rate**

New b_3 = Old b_3 - Step Size

…and b_3.

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 1.90 × 0.1

New b_3 = Old b_3 - Step Size

…and b_3.

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 1.90 x 0.1 = 0.19

New b_3 = Old b_3 - Step Size

…and b_3.

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 1.90 × 0.1 = 0.19

New b_3 = Old b_3 - Step Size

…and b_3.

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 1.90 × 0.1 = 0.19

New b_3 = 0.00 - Step Size

…and b_3.

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size $= 1.90 \times 0.1 \boxed{= 0.19}$

New $\boldsymbol{b_3}$ $=$ **0.00** $-$ $\boxed{\textbf{Step Size}}$

…and $\boldsymbol{b_3}$.

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 1.90 × 0.1 $\boxed{= 0.19}$

New b_3 = **0.00** − $\boxed{0.19}$

…and b_3.

$$\frac{d\,SSR}{d\,w_3} = 2.58$$

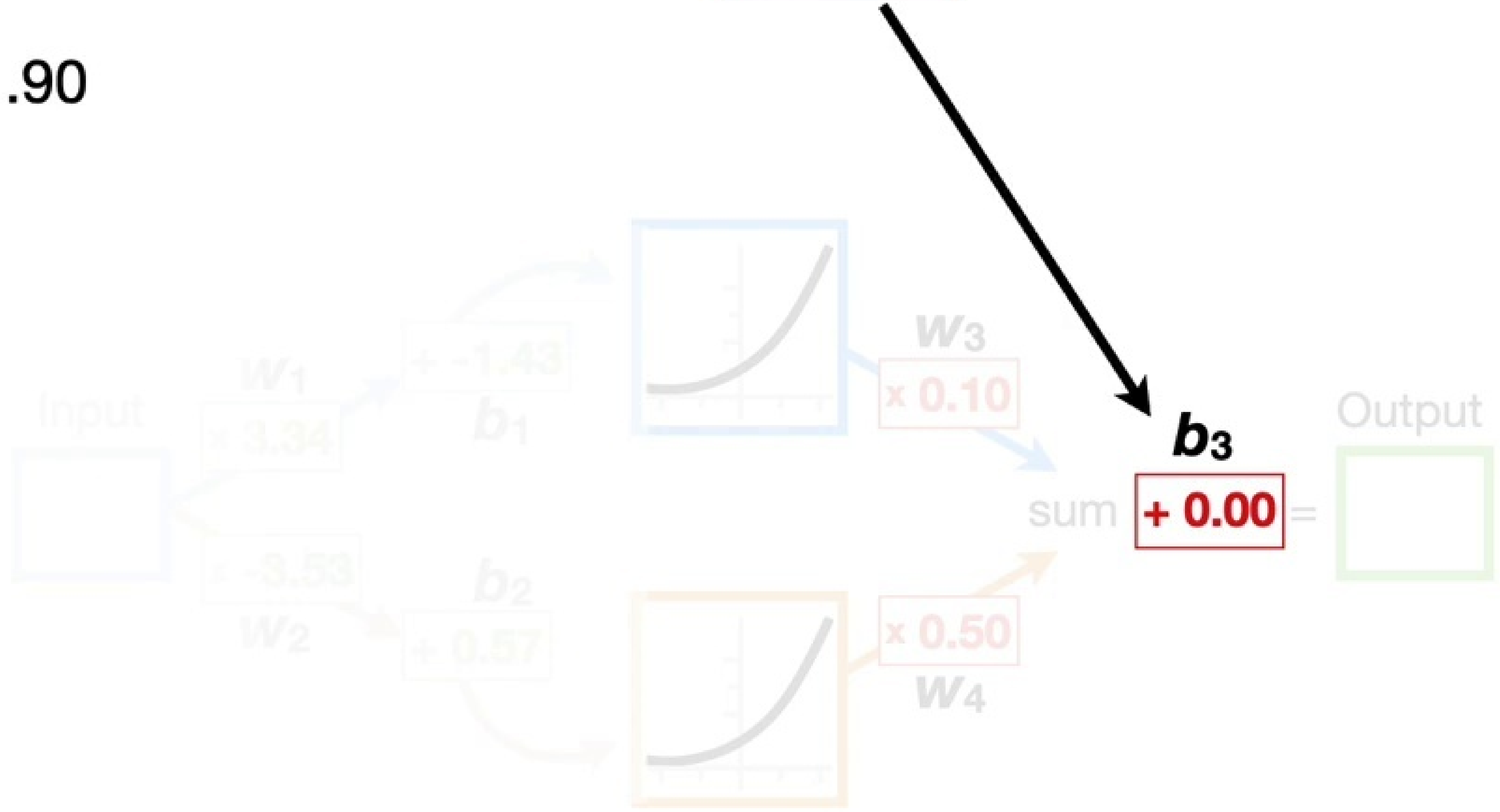
$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 1.90 × 0.1 = 0.19

New b_3 = 0.00 - 0.19 $\boxed{= -0.19}$

…and b_3.



$$\frac{d\,SSR}{d\,w_3} = 2.58$$

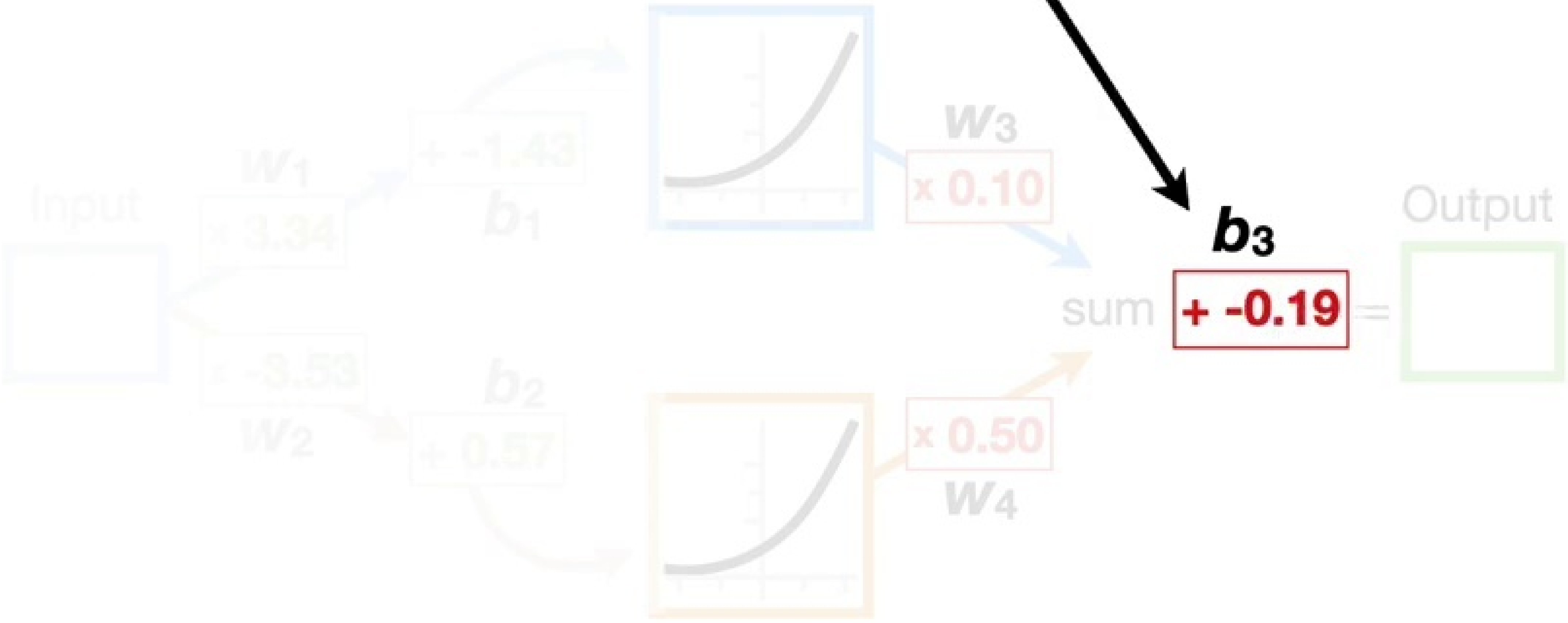
$$\frac{d\,SSR}{d\,w_4} = 1.26$$

$$\frac{d\,SSR}{d\,b_3} = 1.90$$

Step Size = 1.90 × 0.1 = 0.19

New b_3 = 0.00 - 0.19 $\boxed{= -0.19}$

…and b_3.

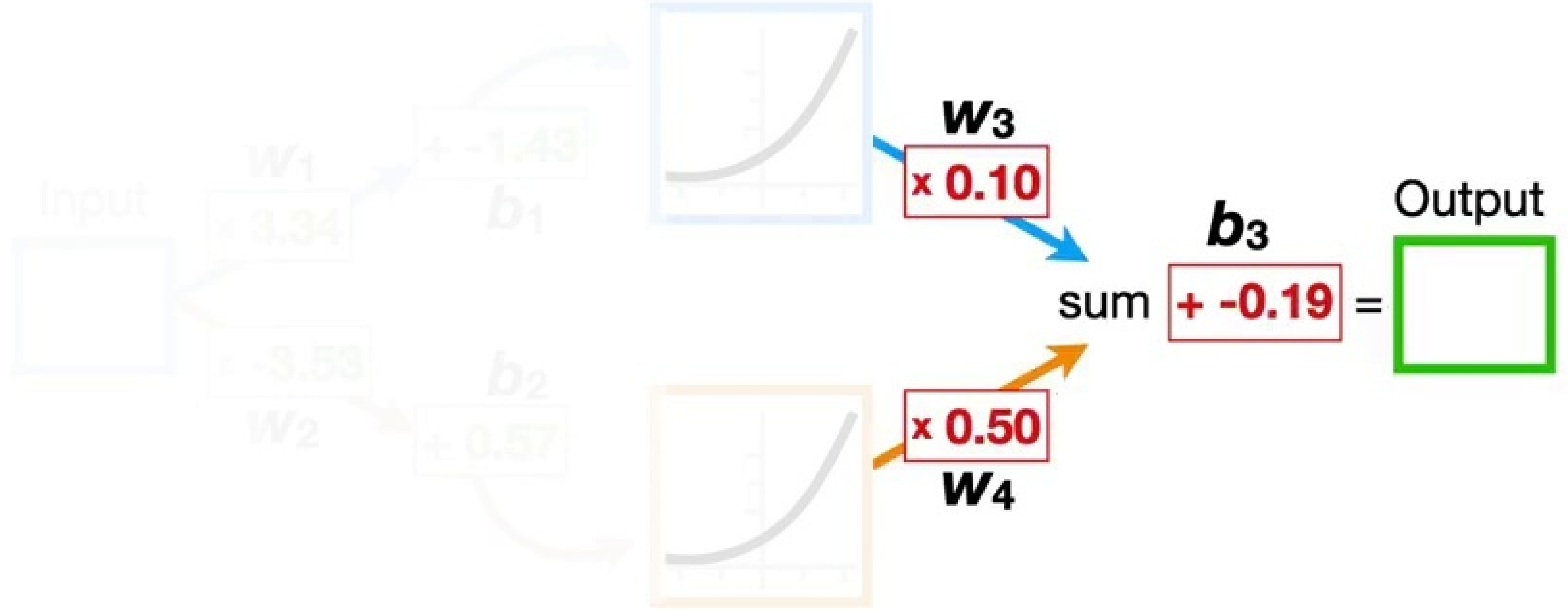


$$\frac{d\ SSR}{d\ w_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{1,i}$$

$$\frac{d\ SSR}{d\ w_4} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{2,i}$$

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

Now we repeat that process until the **Predictions** no longer improve very much, or we reach a maximum number of steps or we meet some other criteria.

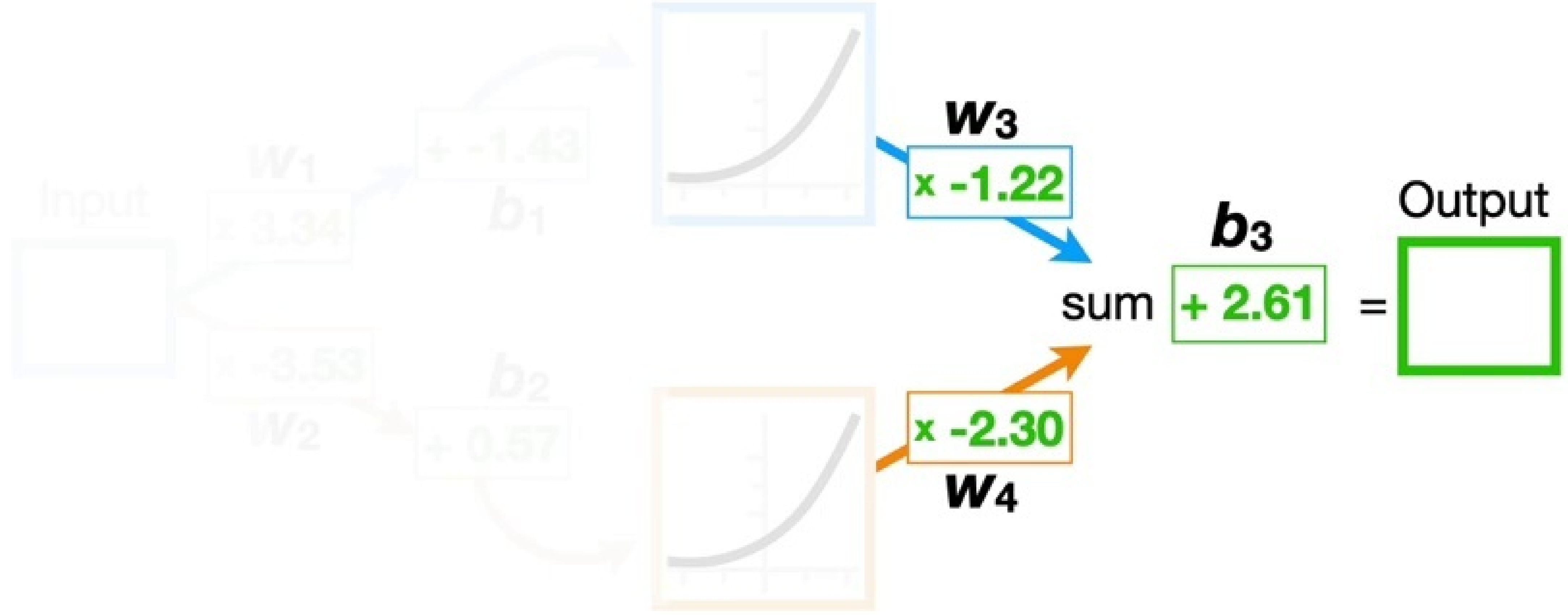


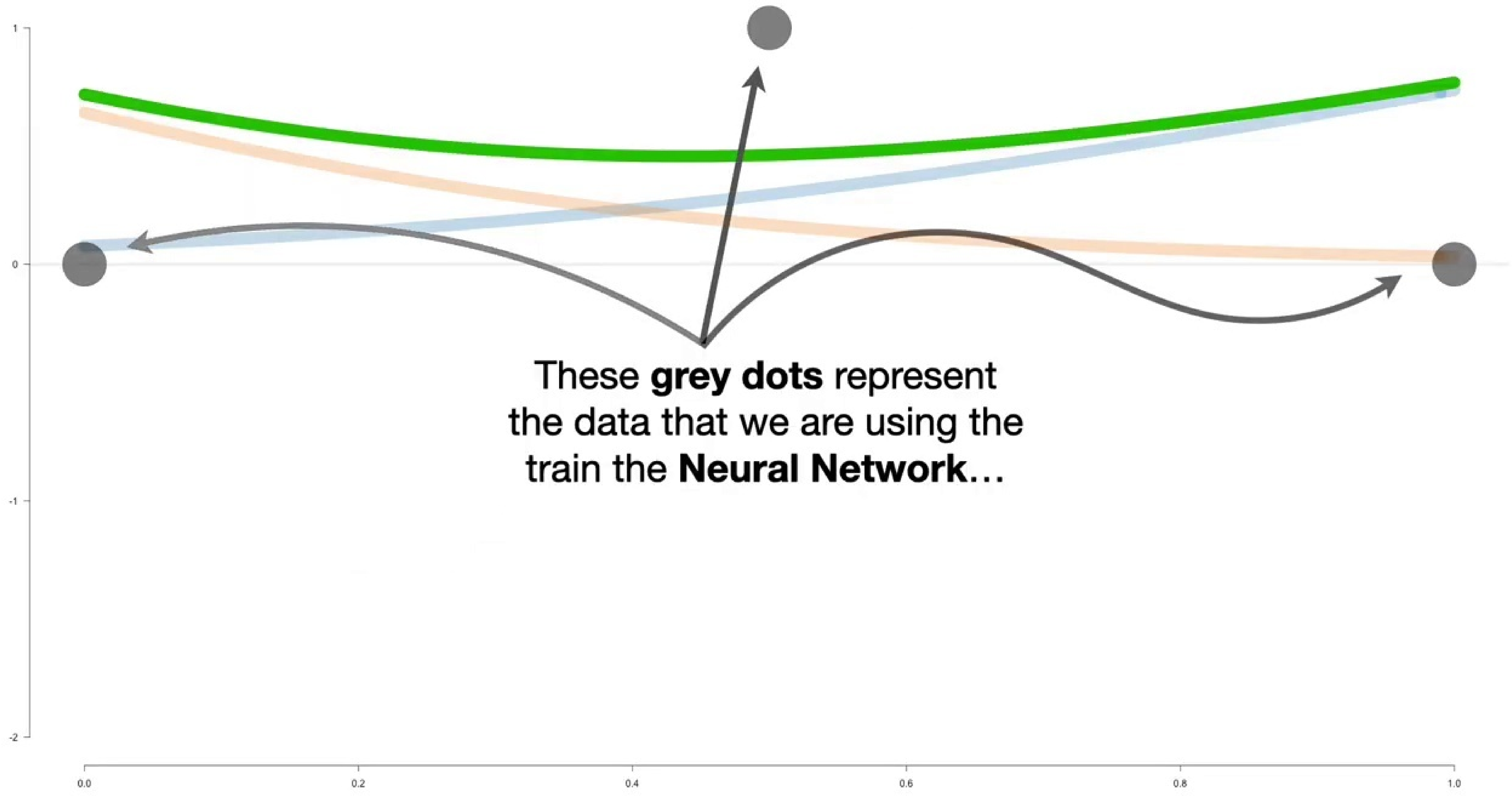
$$\frac{d\,SSR}{d\,w_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{1,i}$$

$$\frac{d\,SSR}{d\,w_4} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times y_{2,i}$$

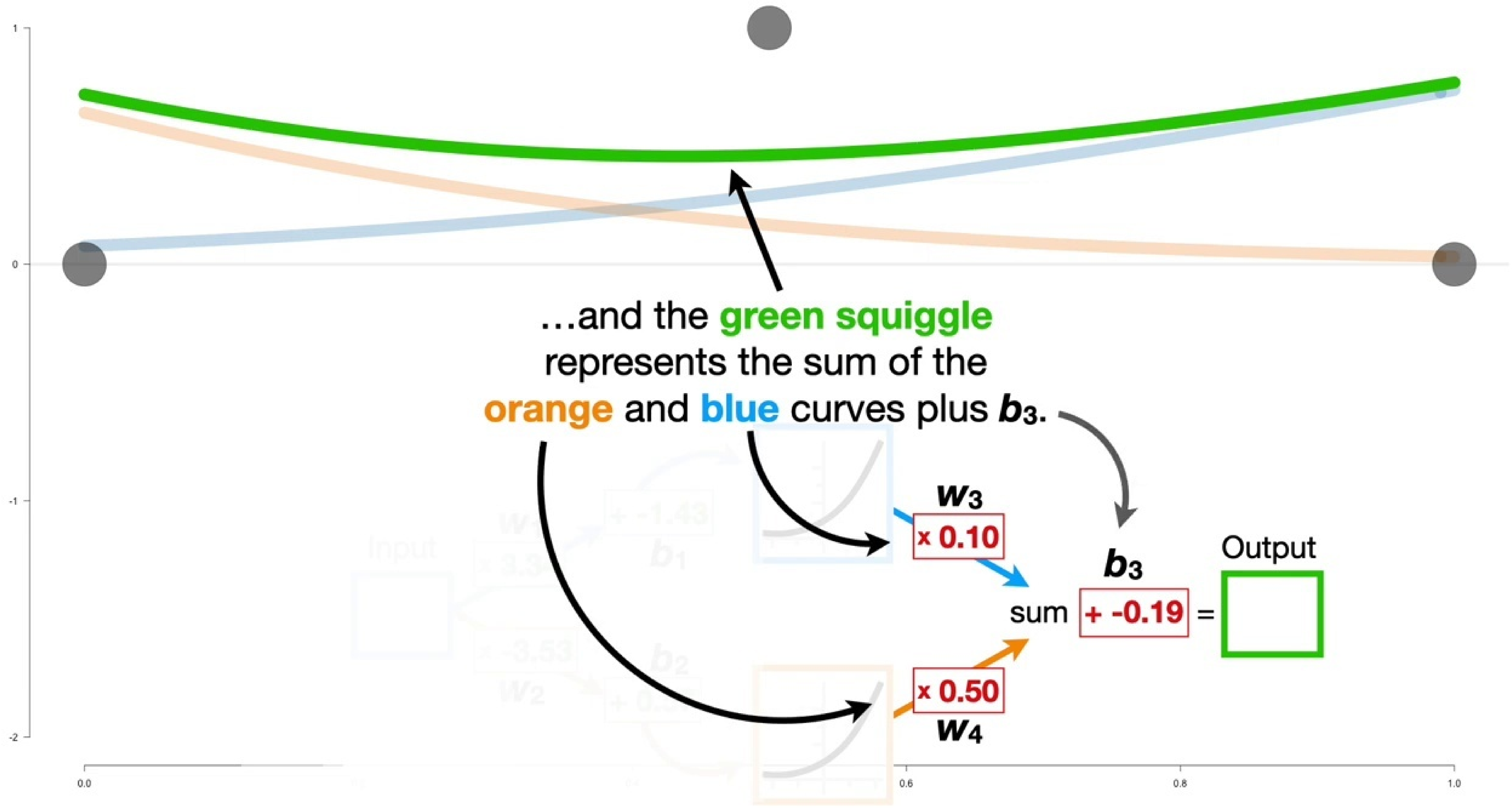
$$\frac{d\,SSR}{d\,b_3} = \sum_{i=1}^{n=3} -2 \times (\text{Observed}_i - \text{Predicted}_i) \times 1$$

Now we repeat that process until the **Predictions** no longer improve very much, or we reach a maximum number of steps or we meet some other criteria.

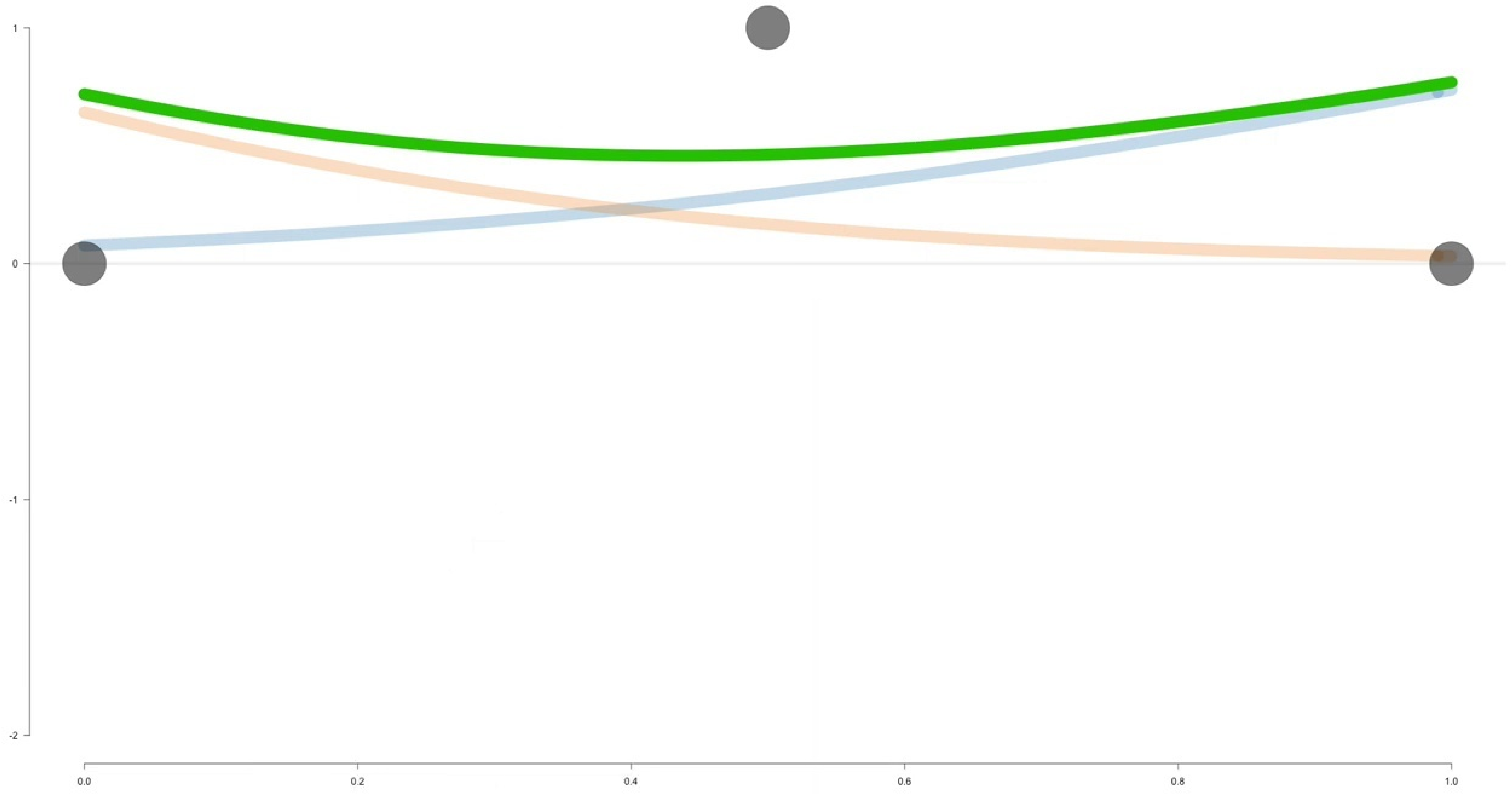


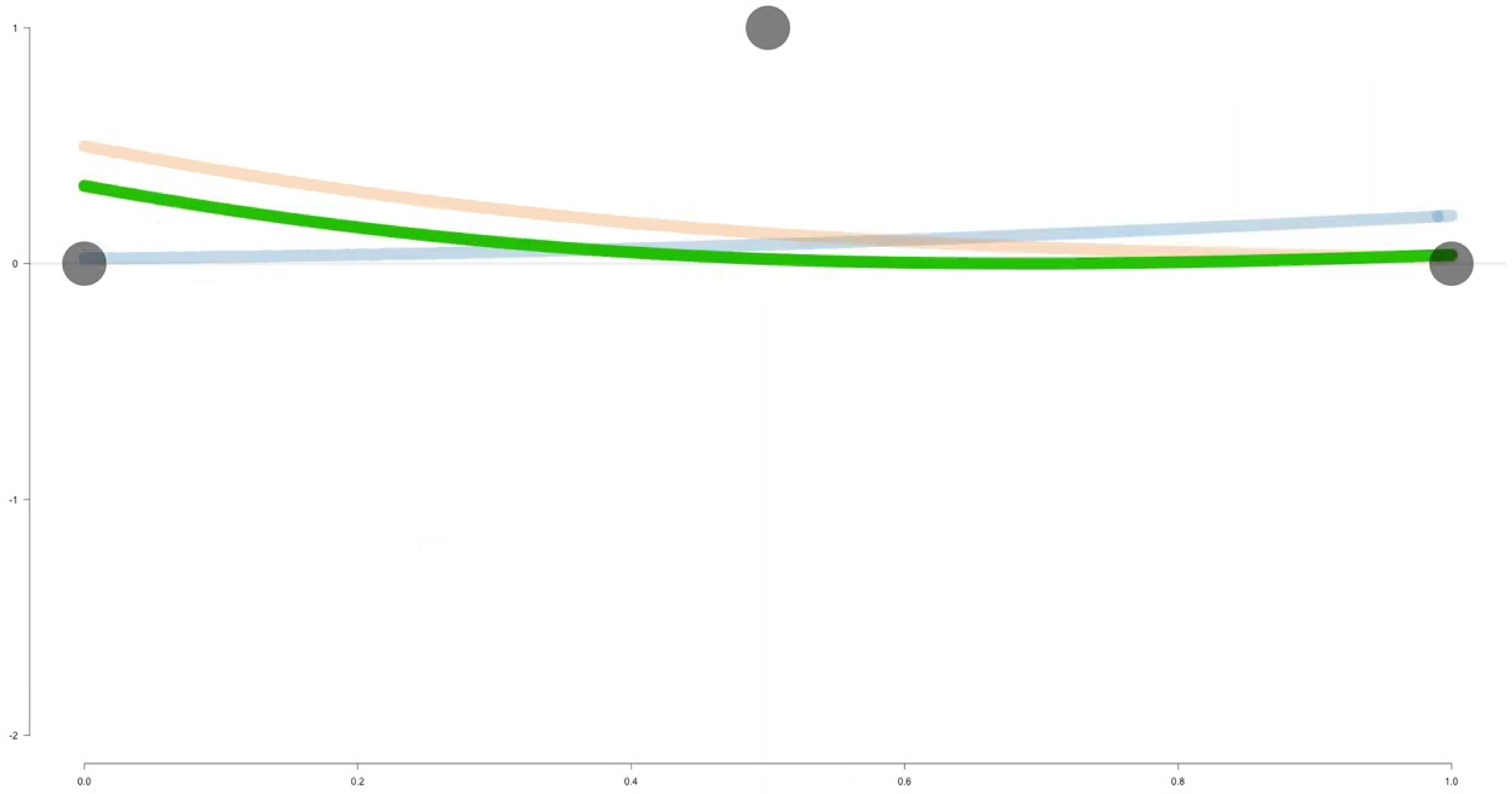


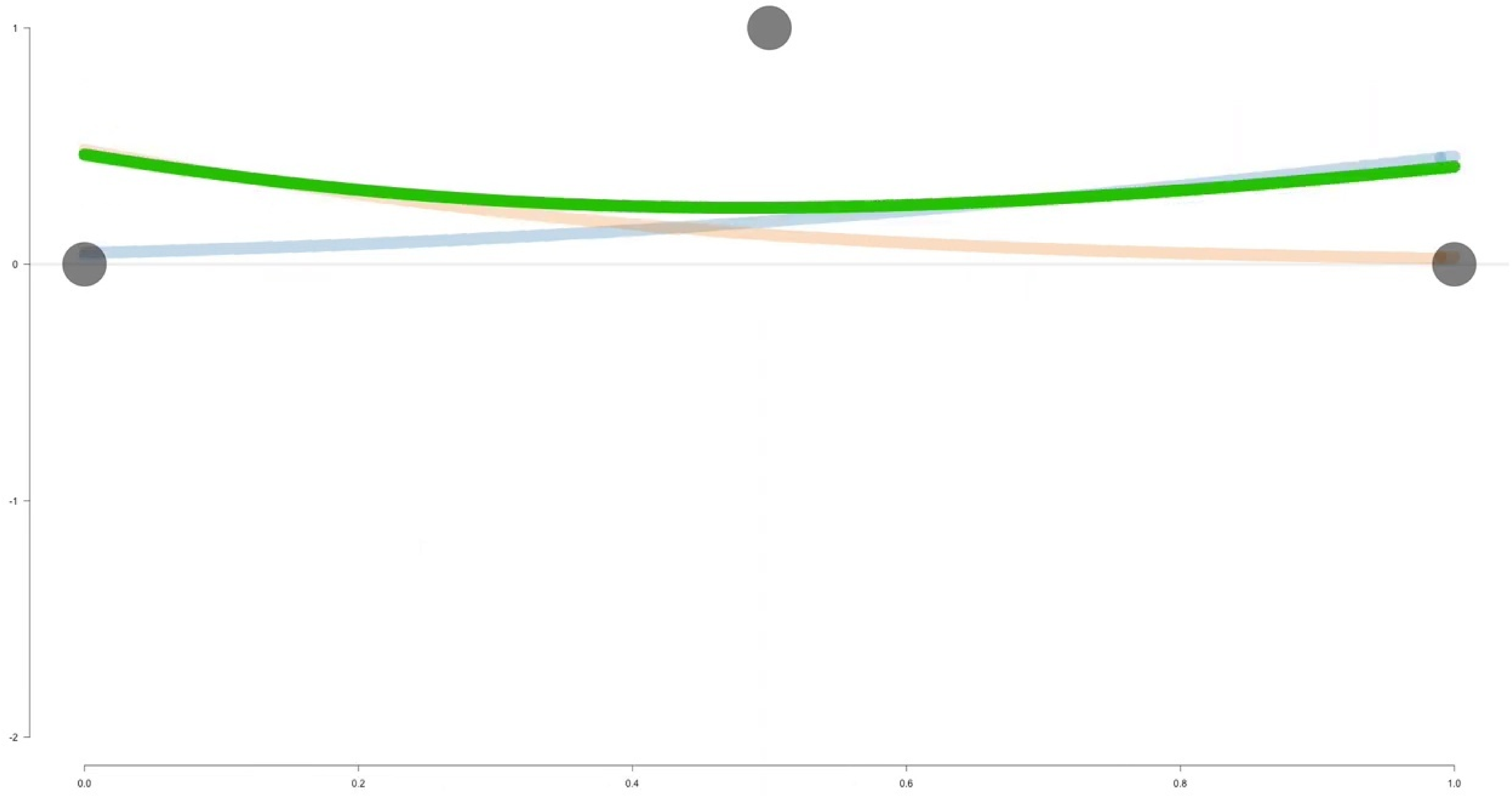
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0
These **grey dots** represent the data that we are using the train the **Neural Network**…



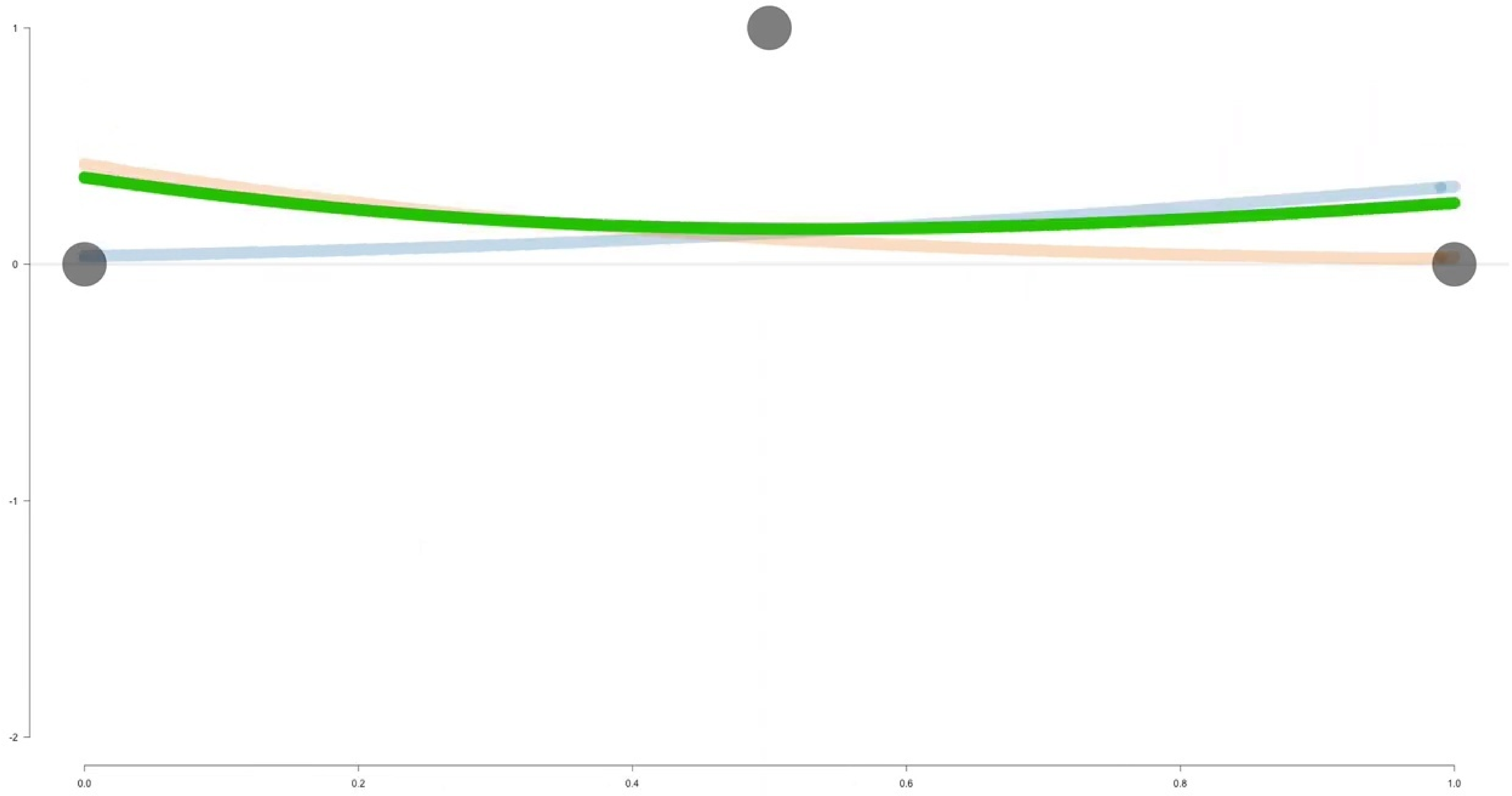
…and the green squiggle
represents the sum of the
orange and blue curves plus b3.
w3
x 0.10
b3
Output
sum
+ -0.19
=
x 0.50
w4


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

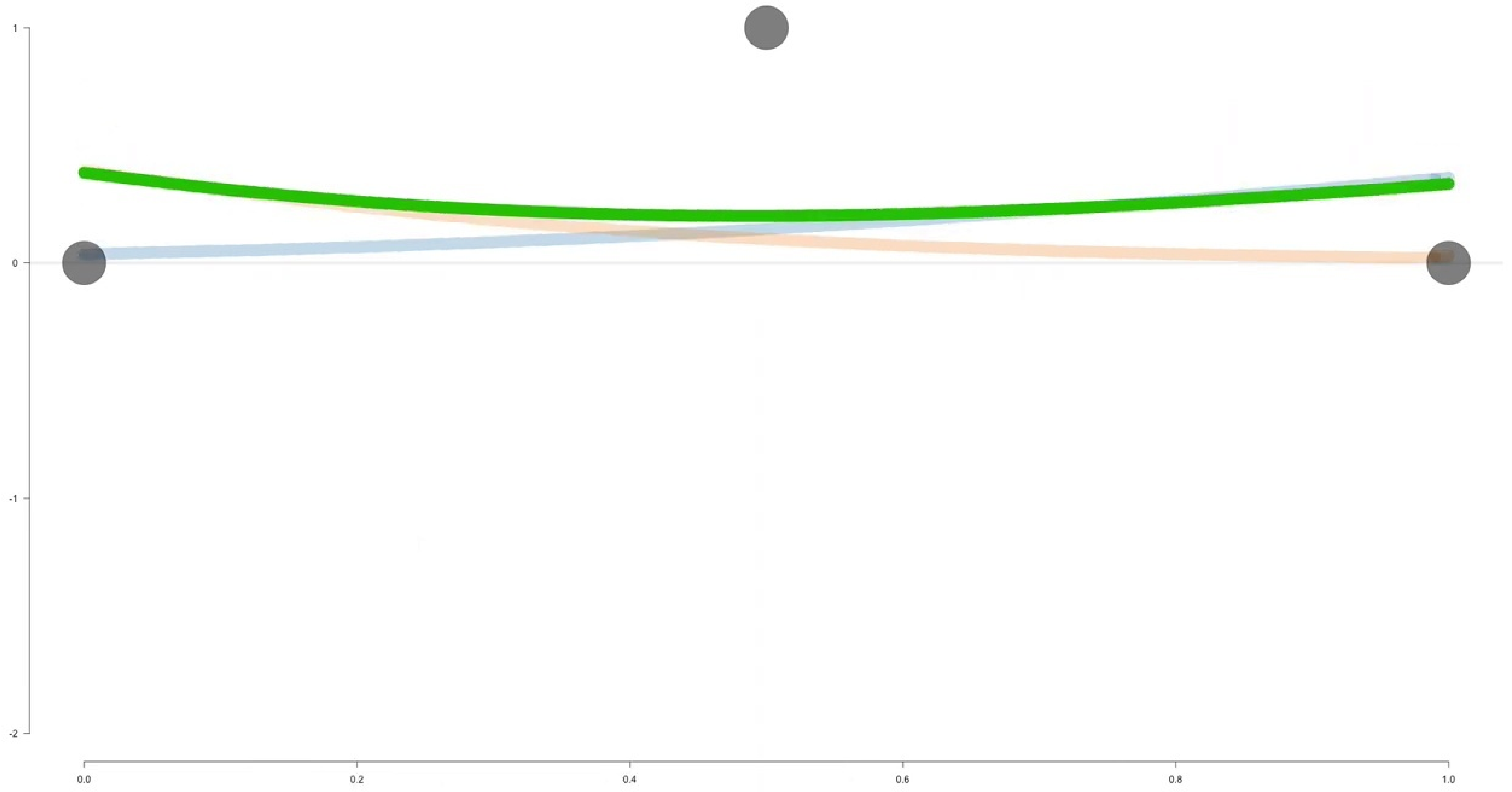

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

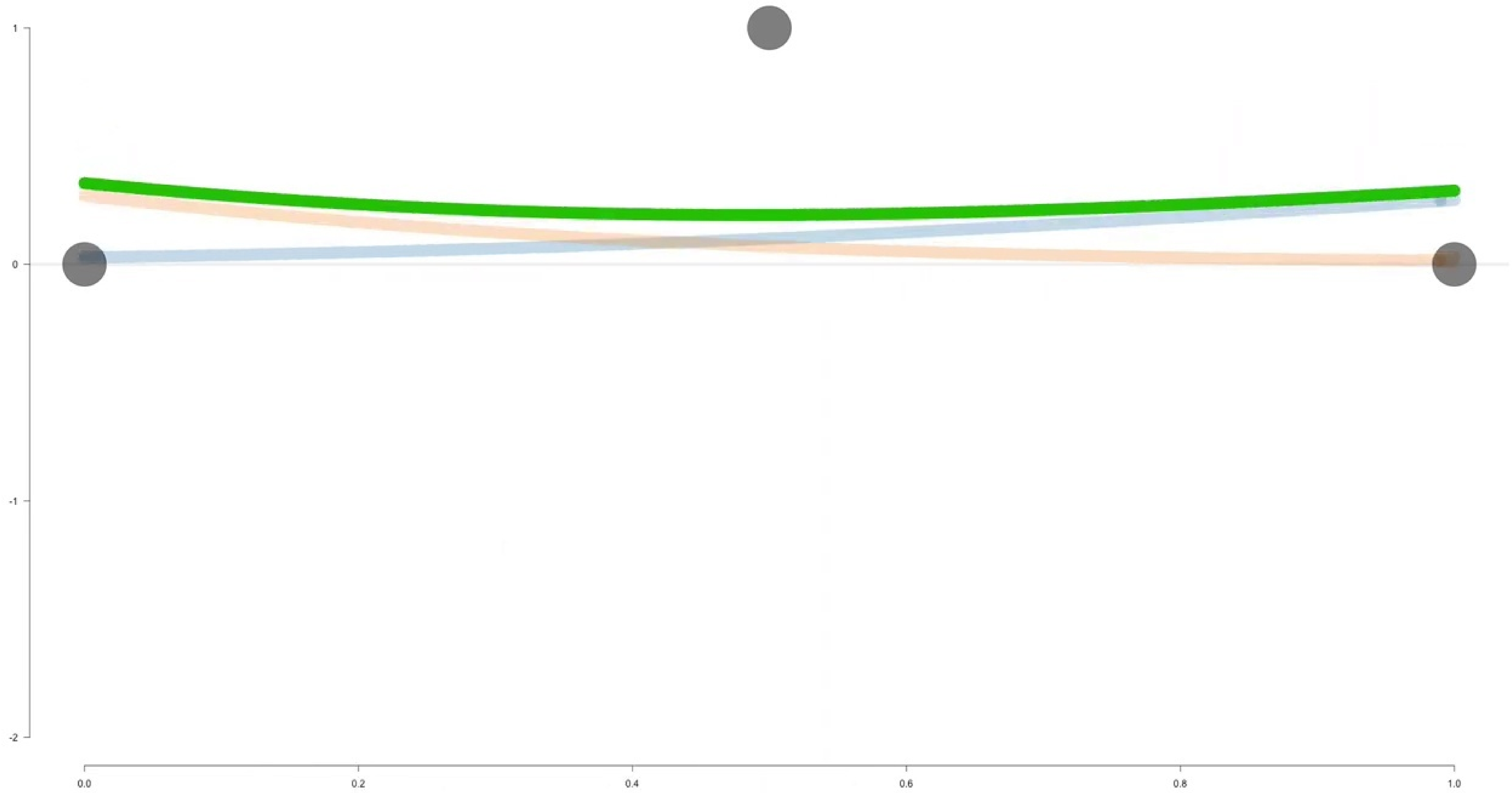


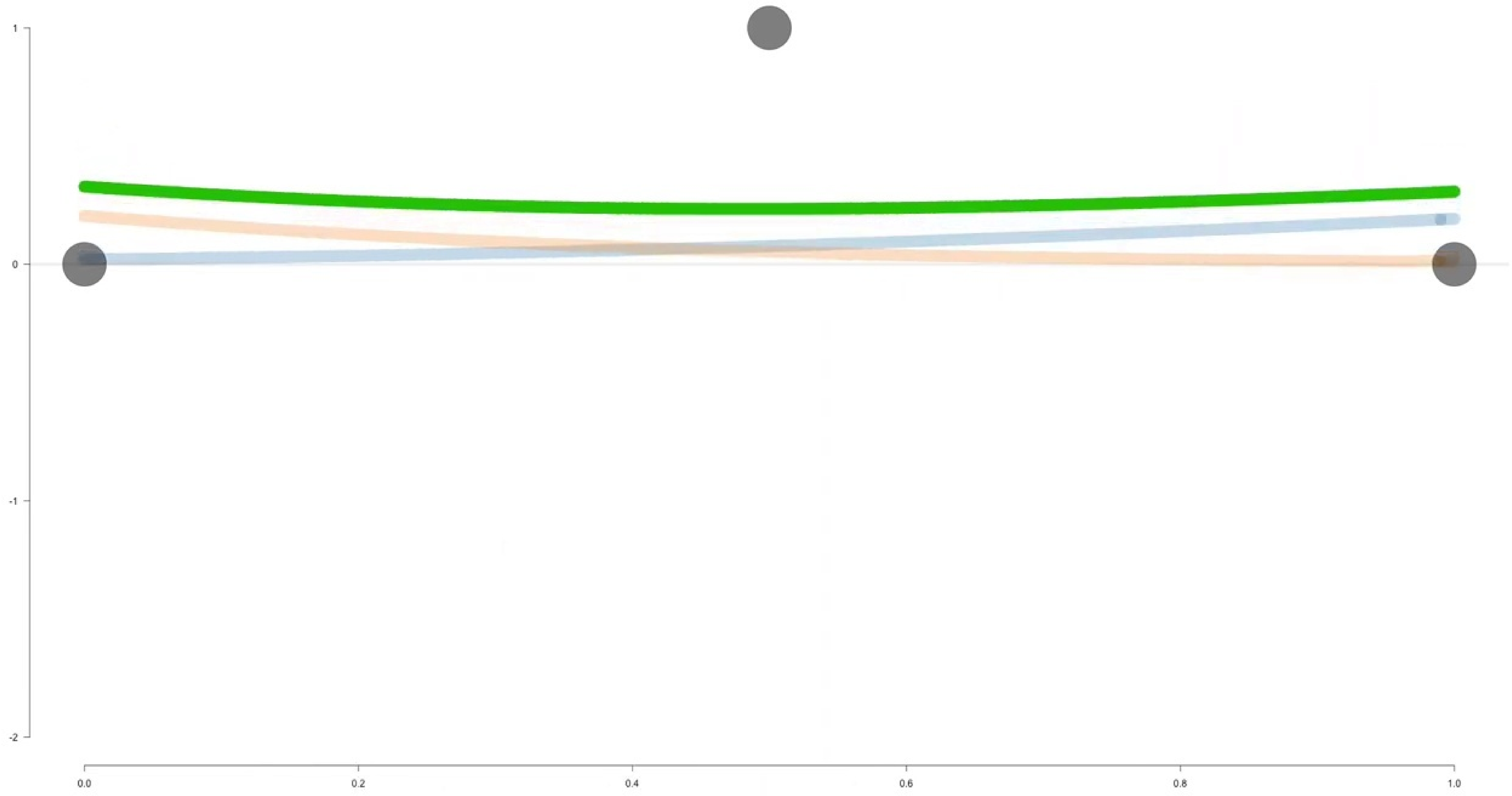
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



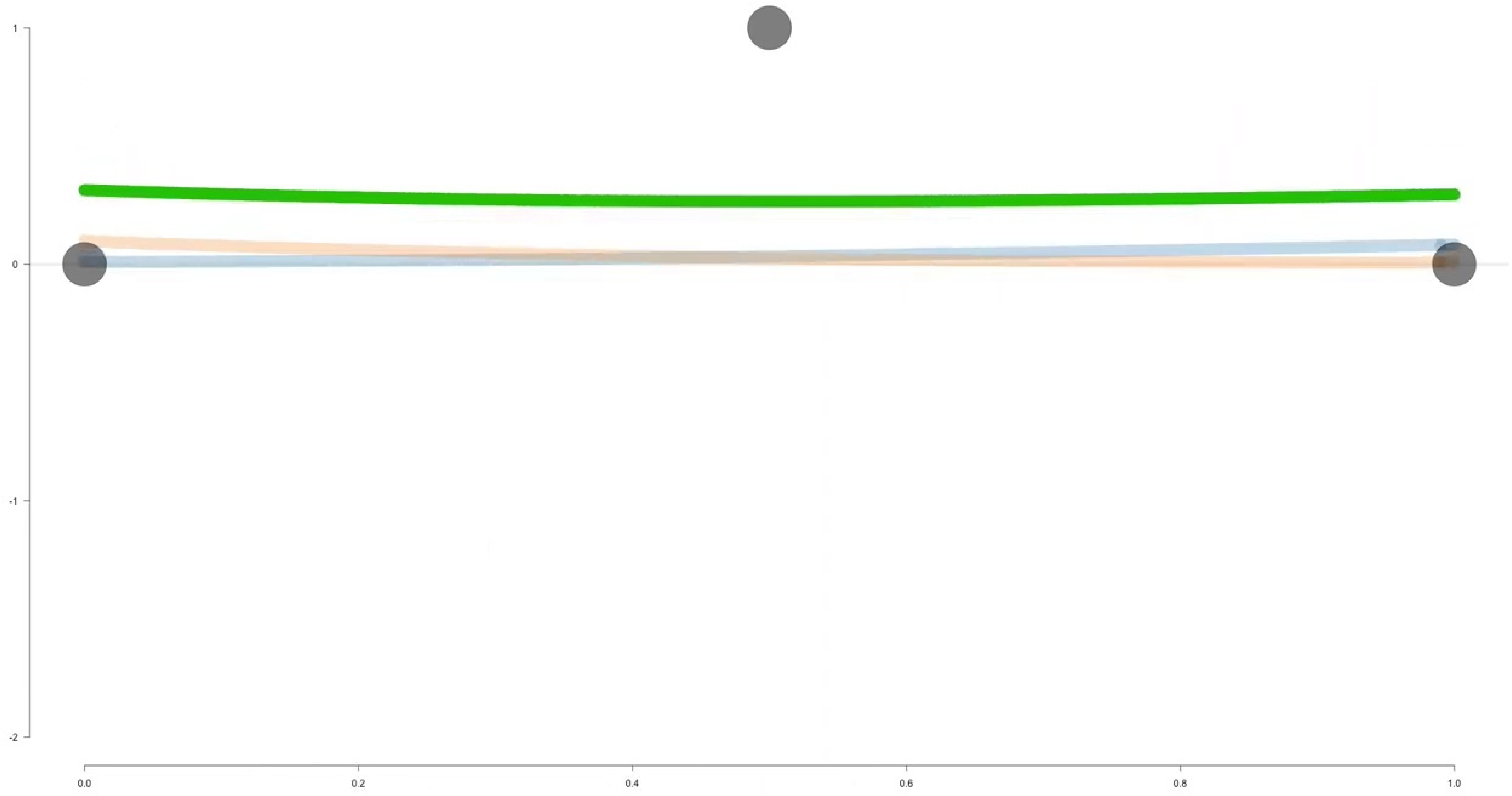
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

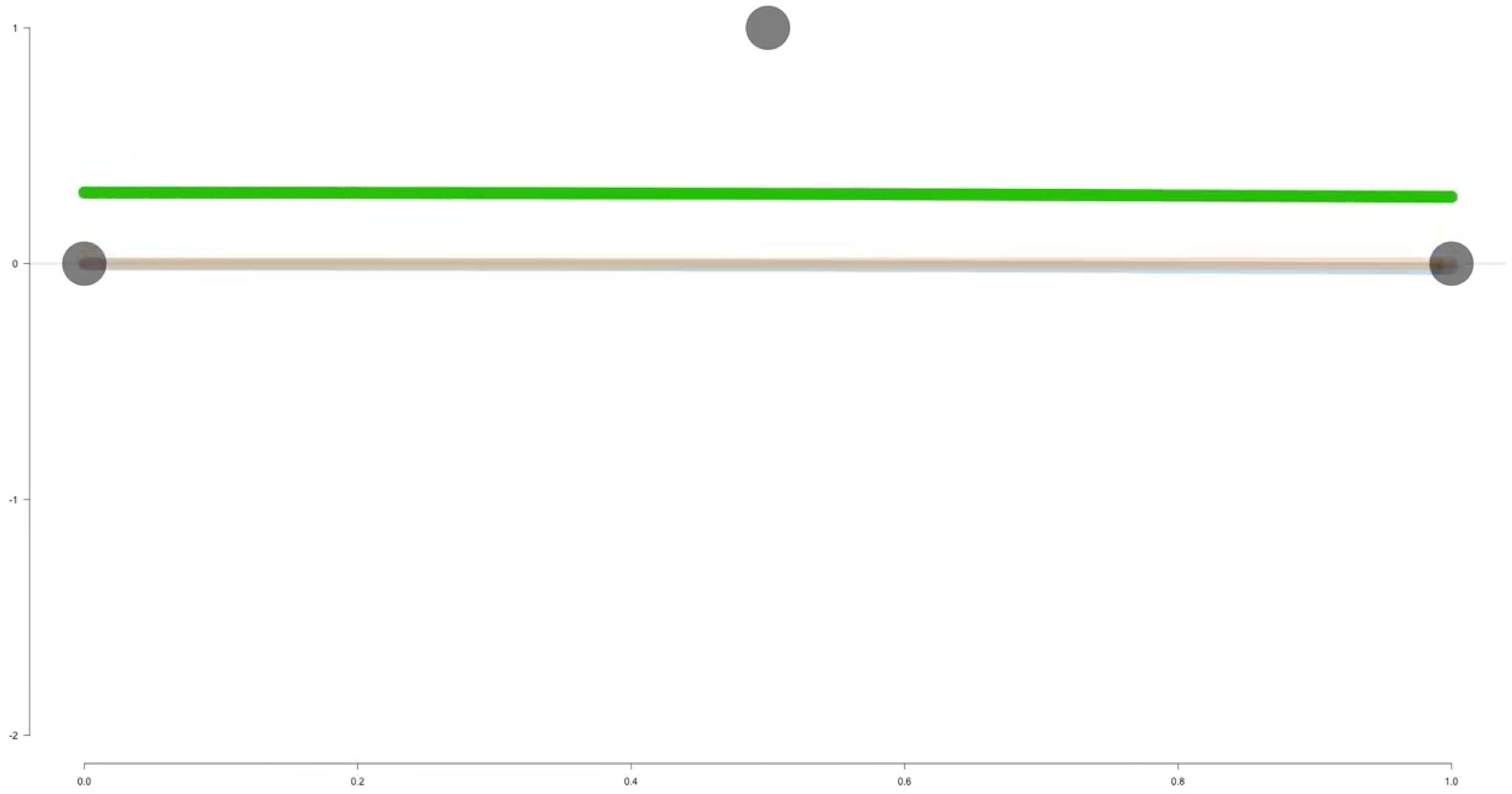

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

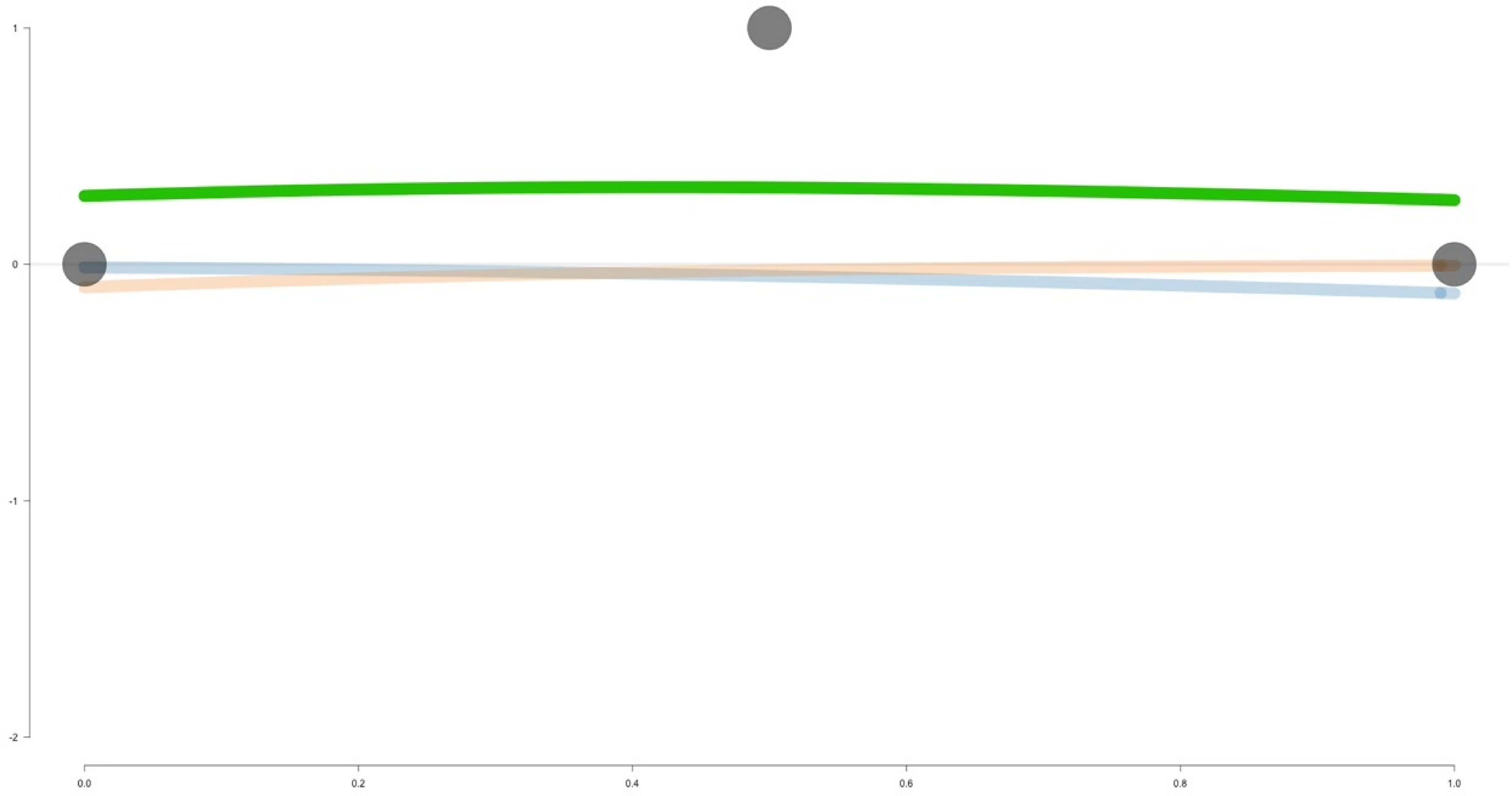


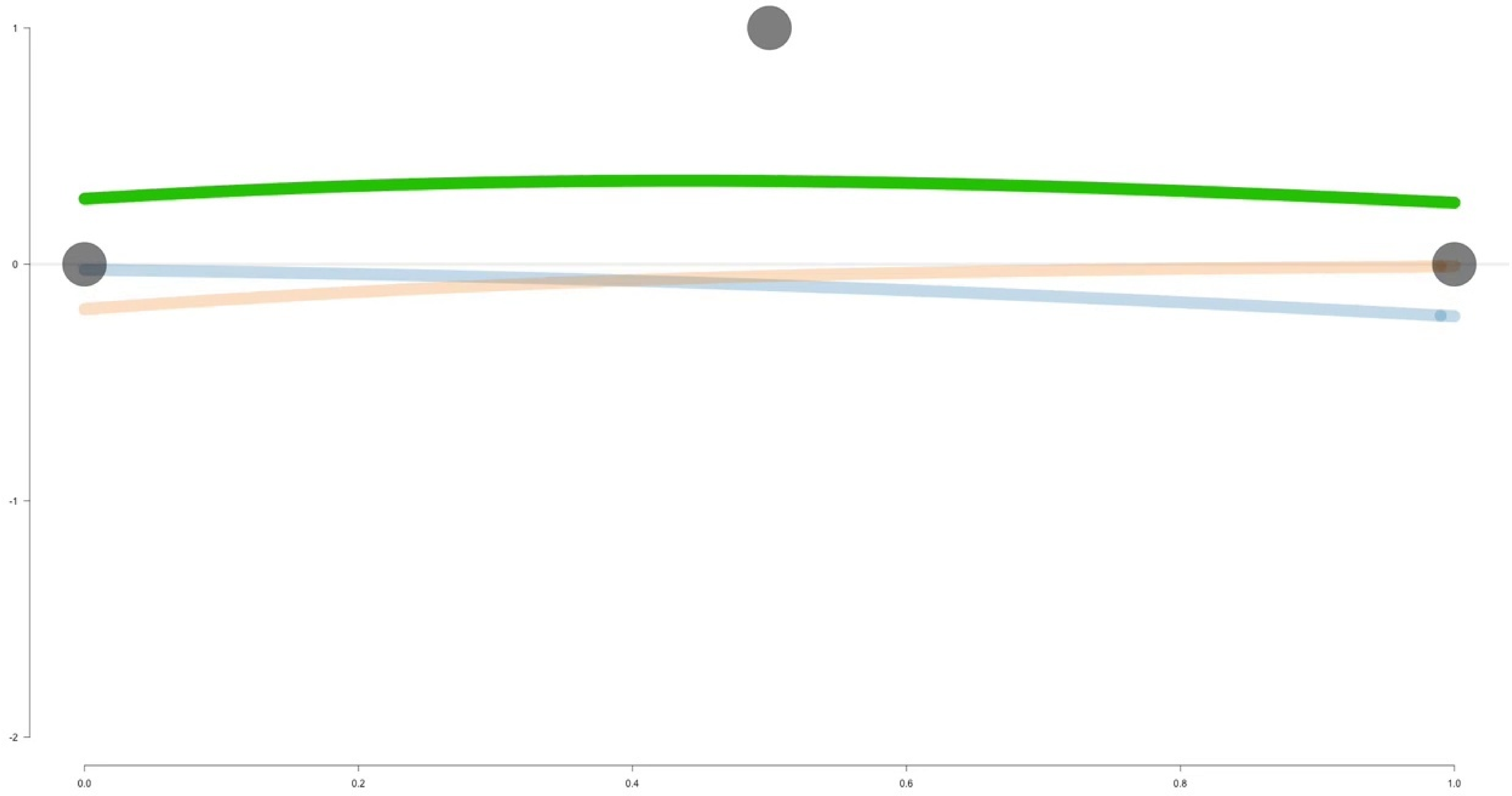
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



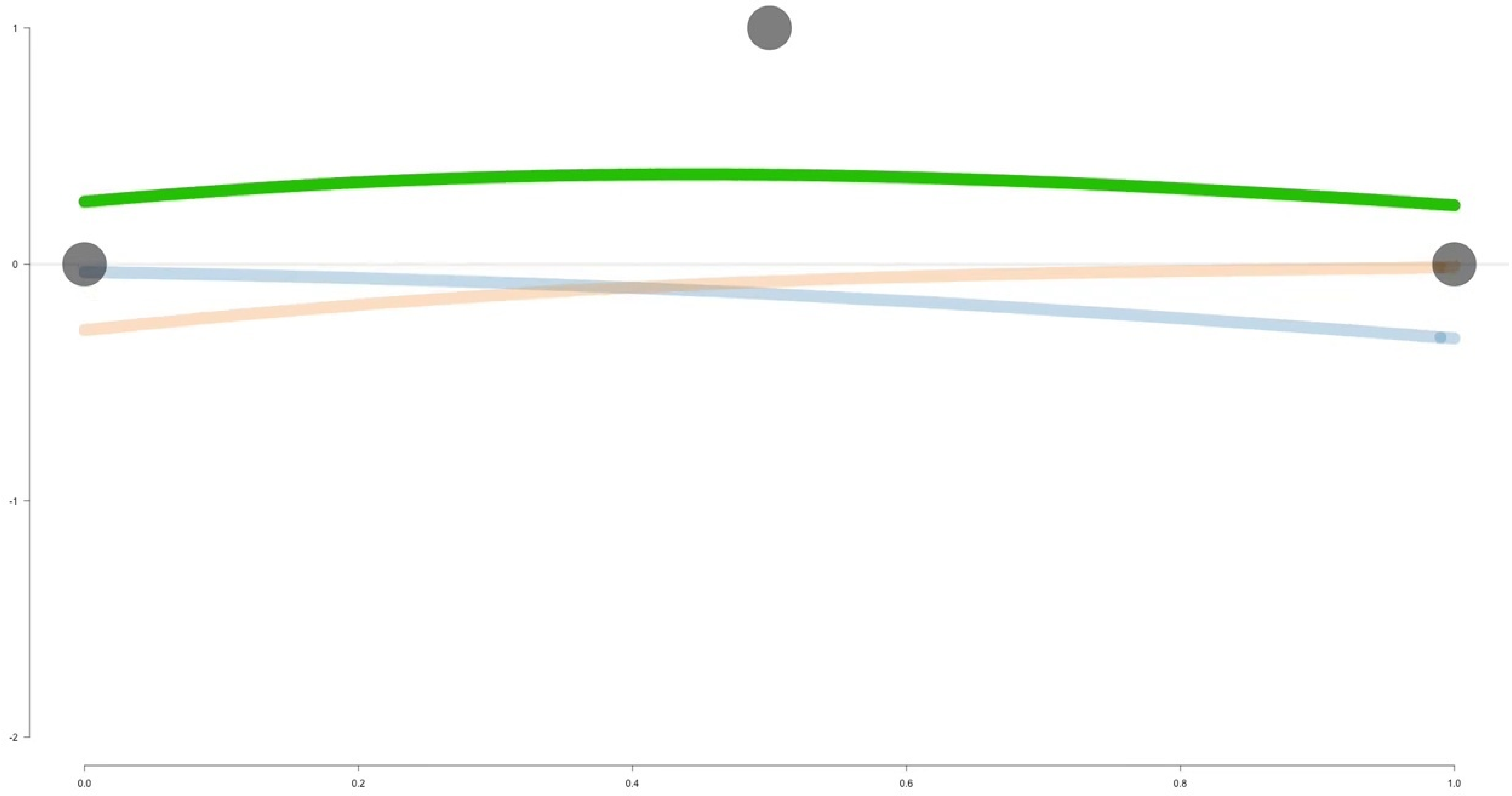
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

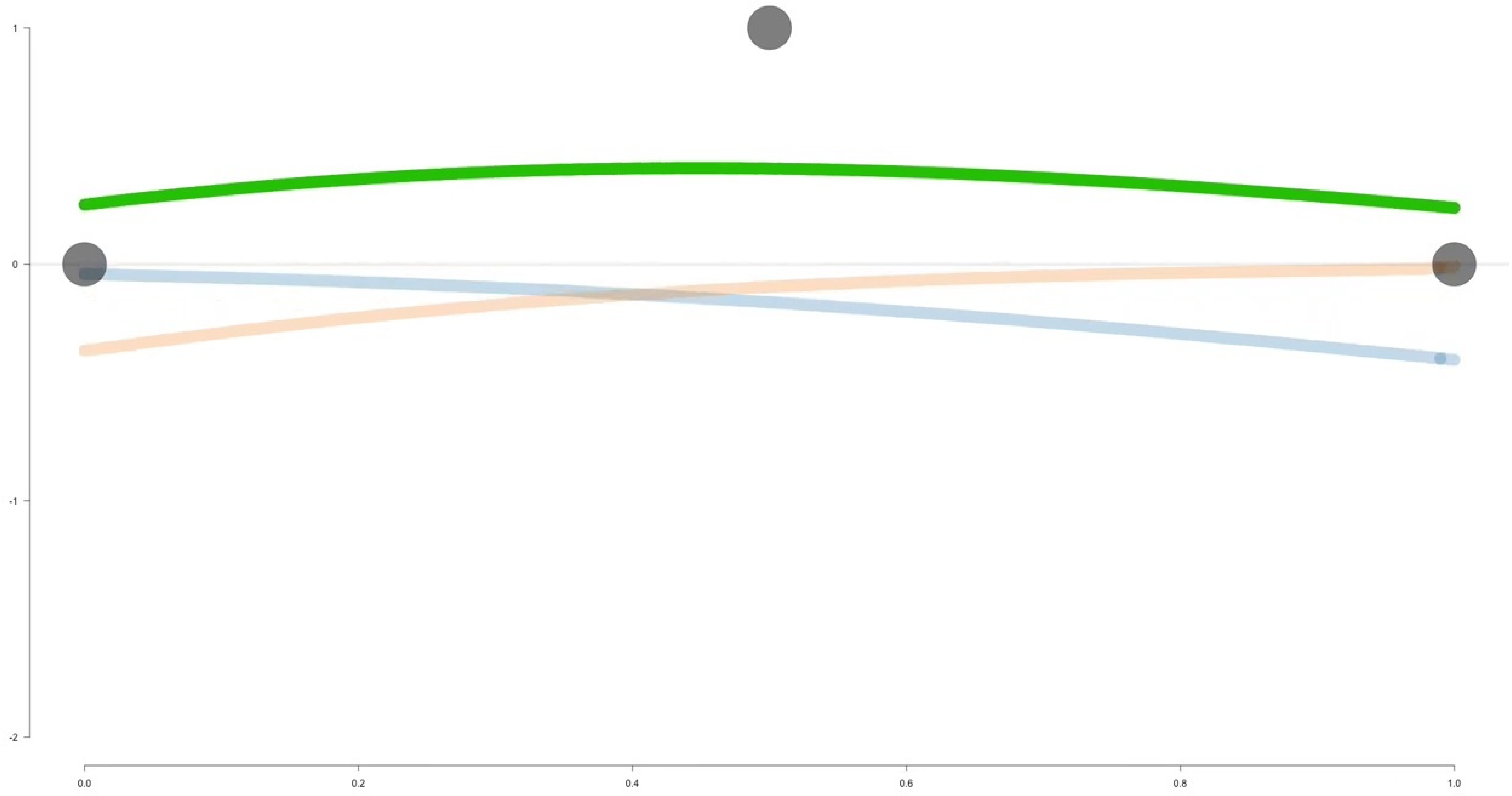

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

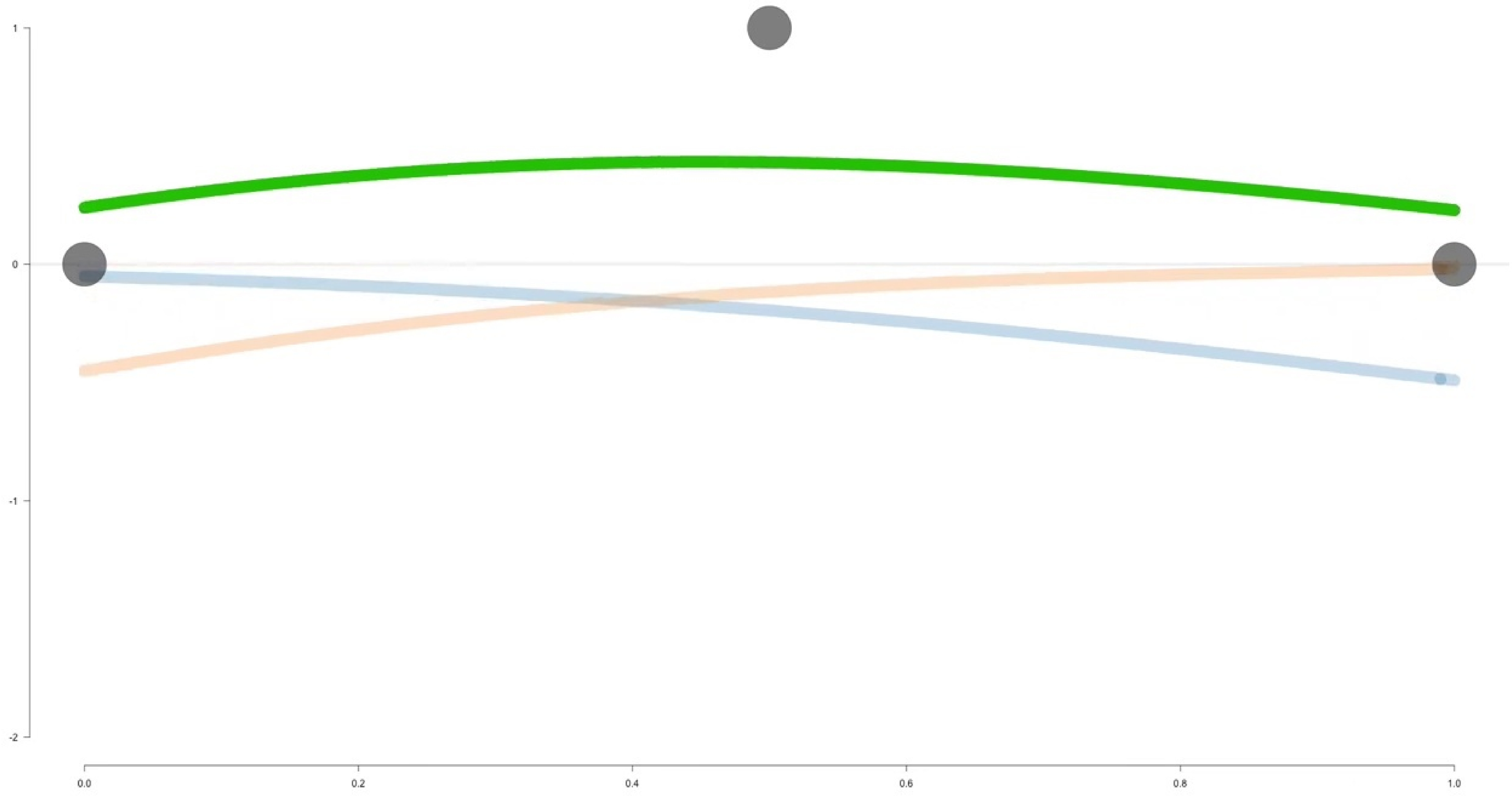


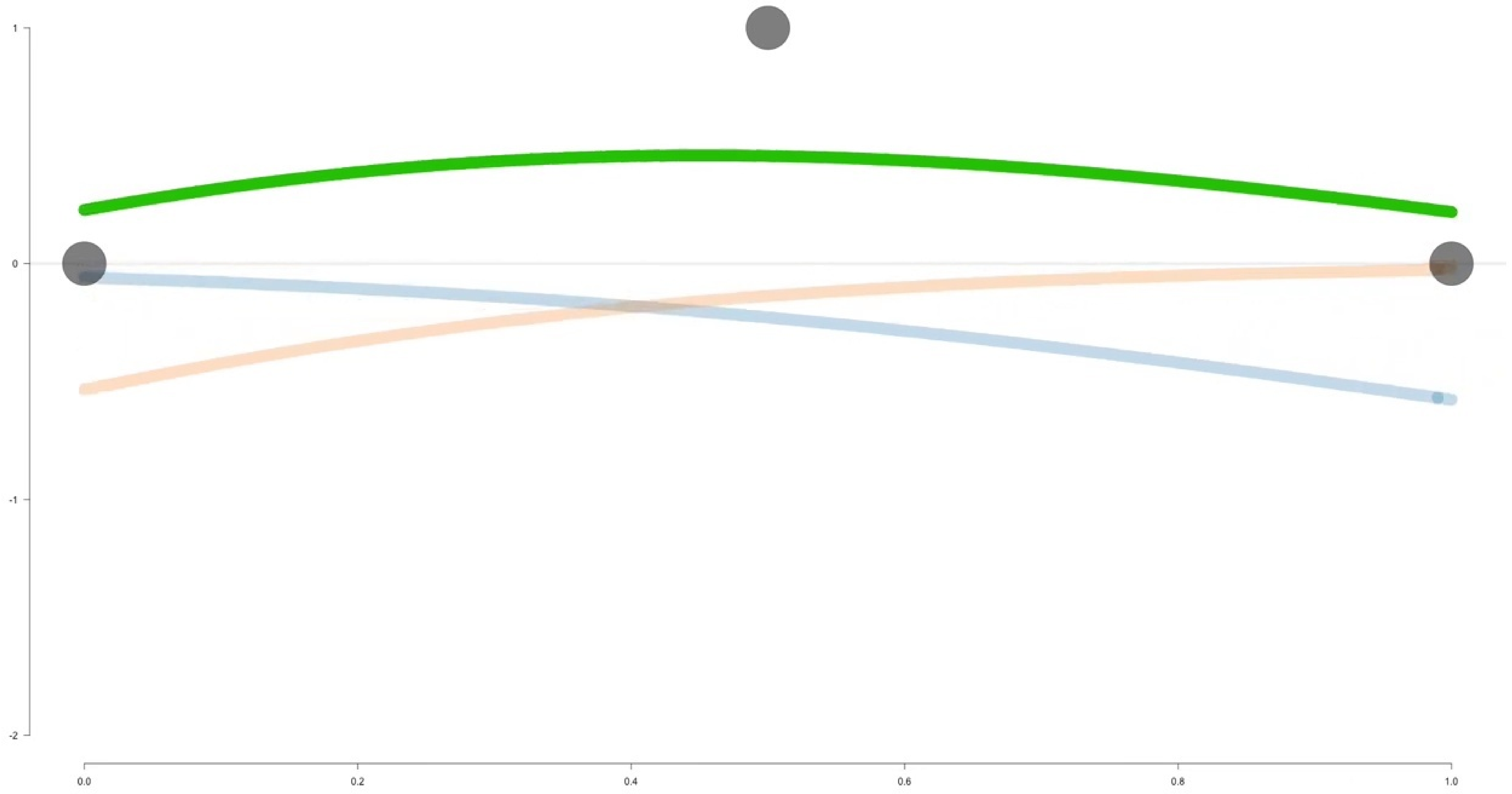
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

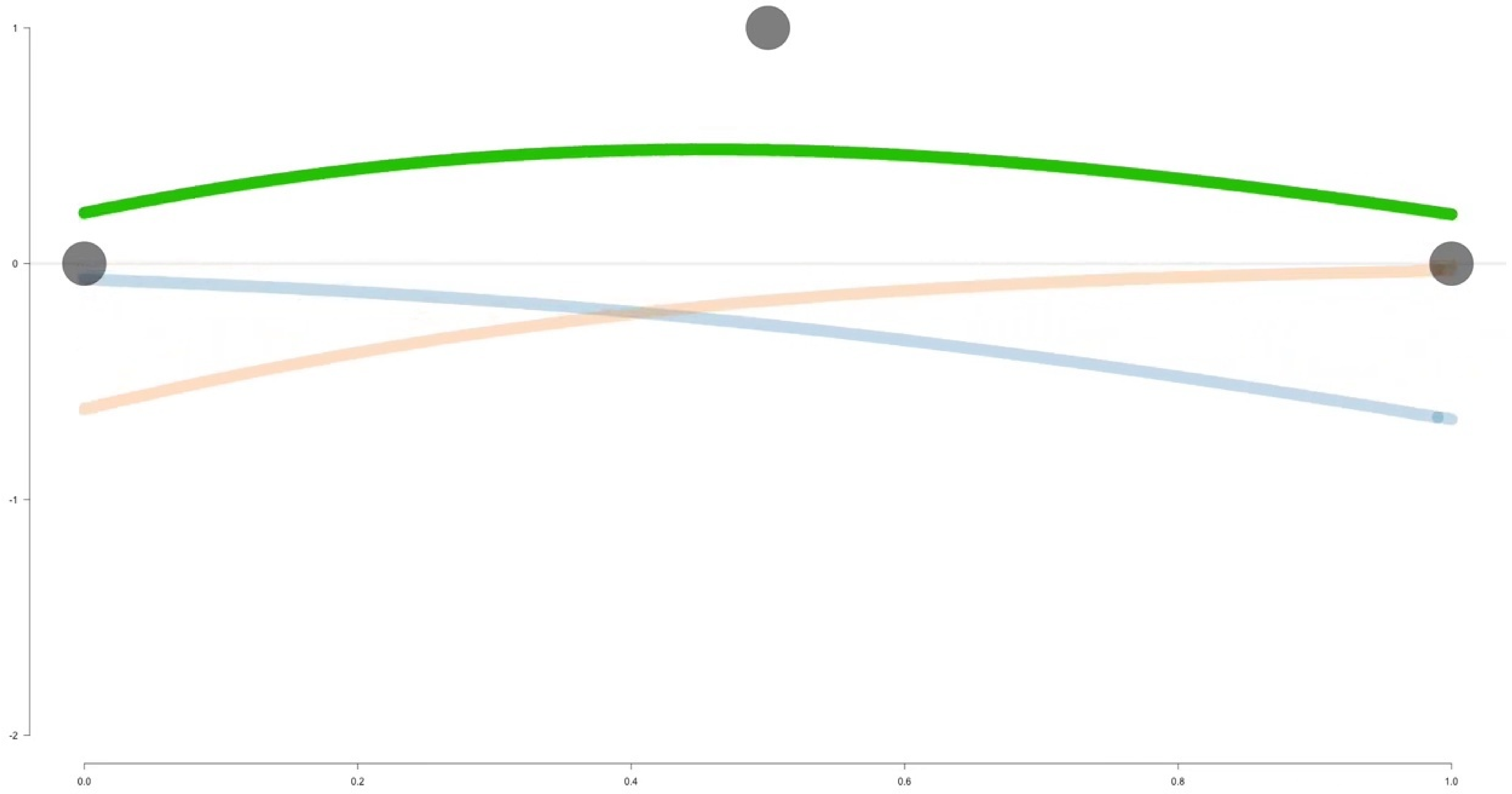

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

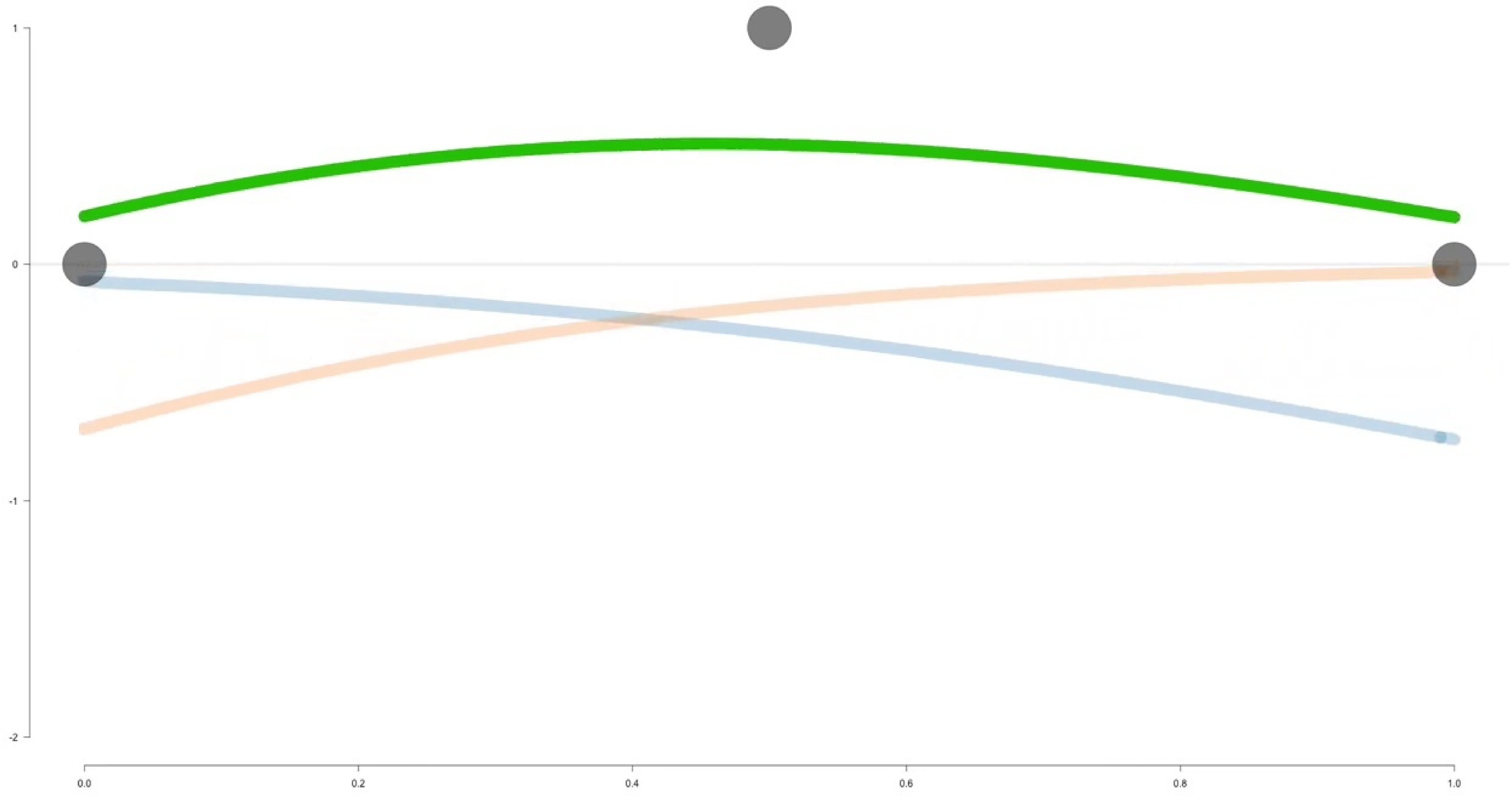


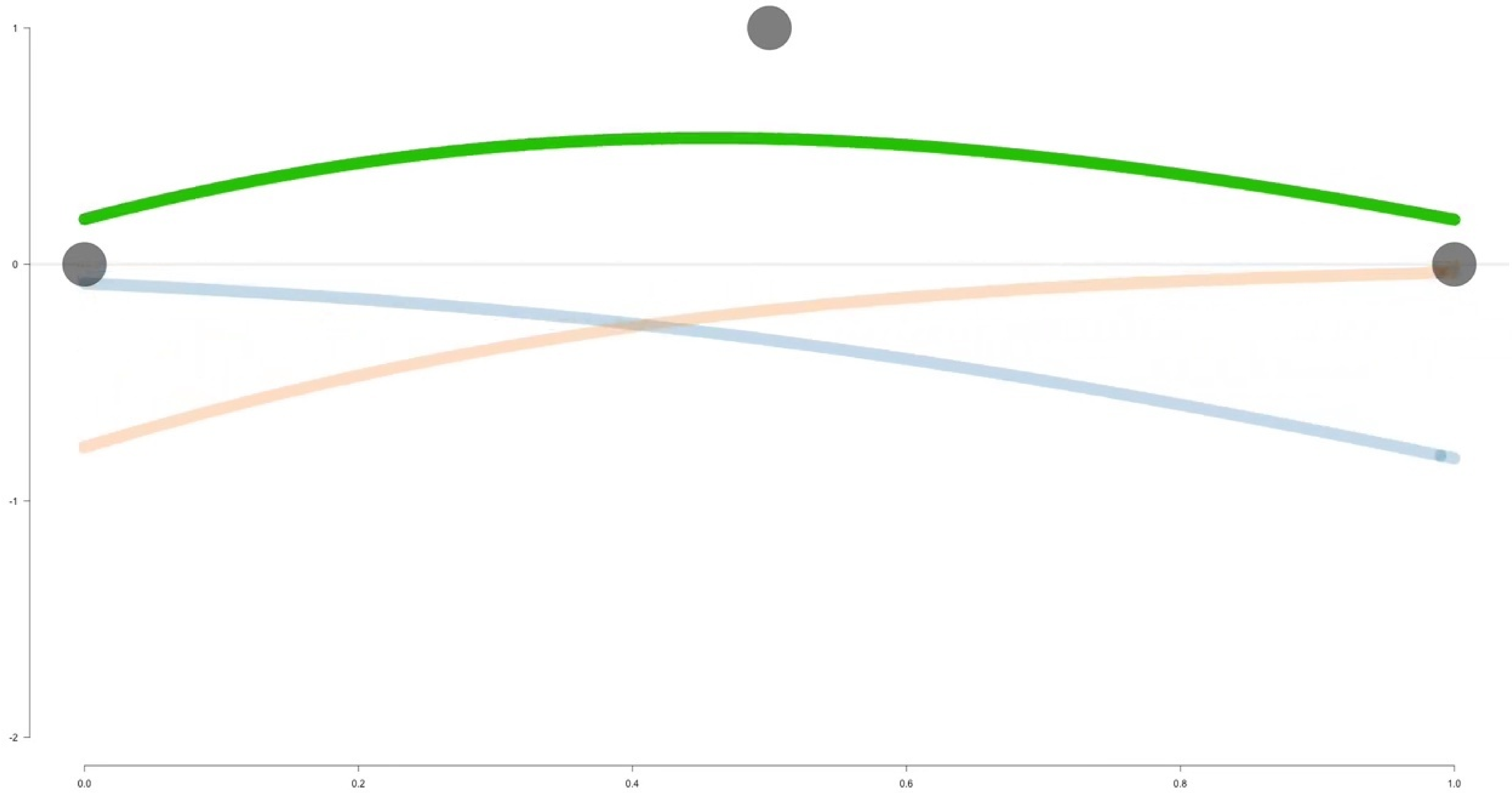
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

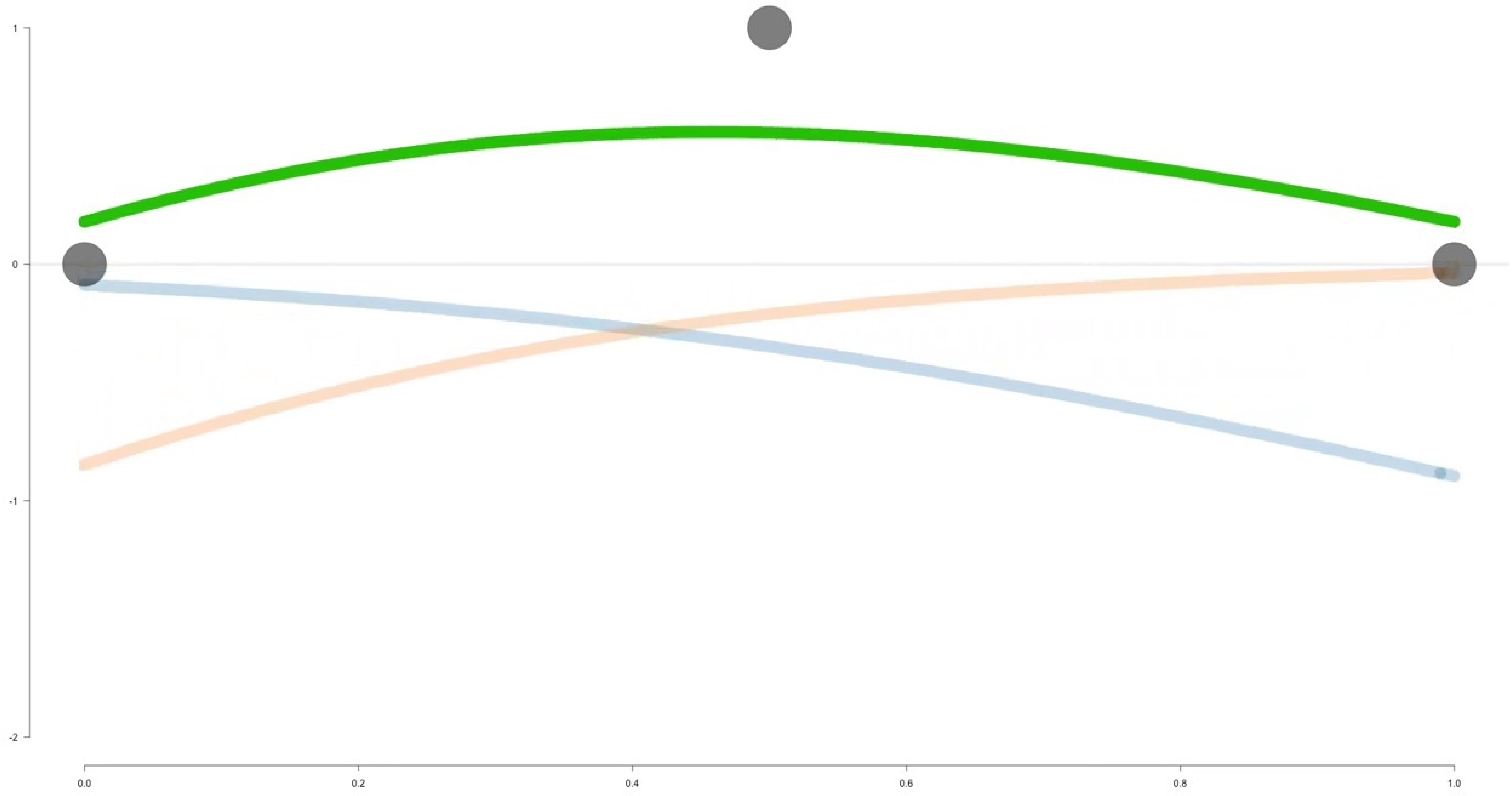

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

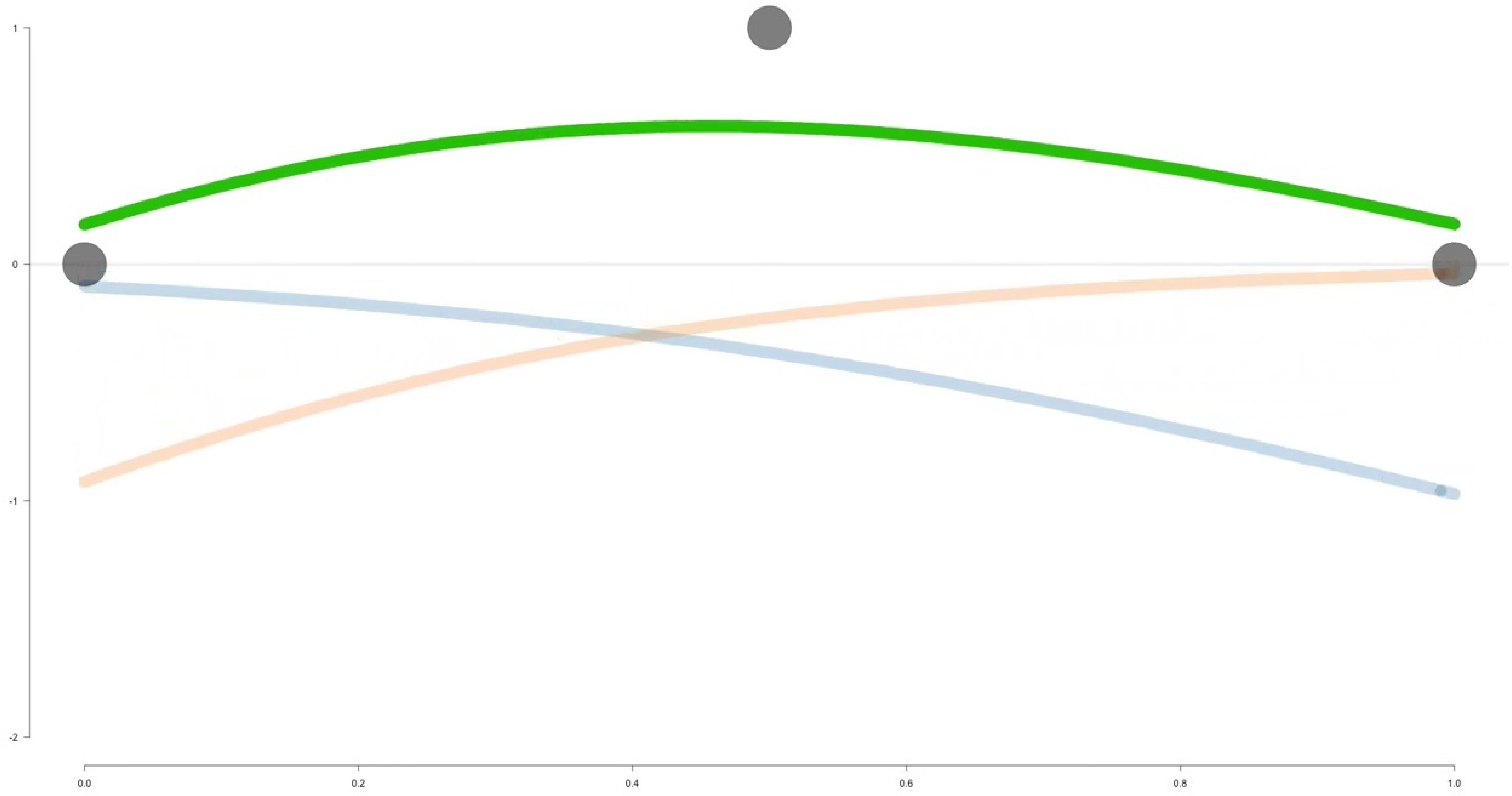

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

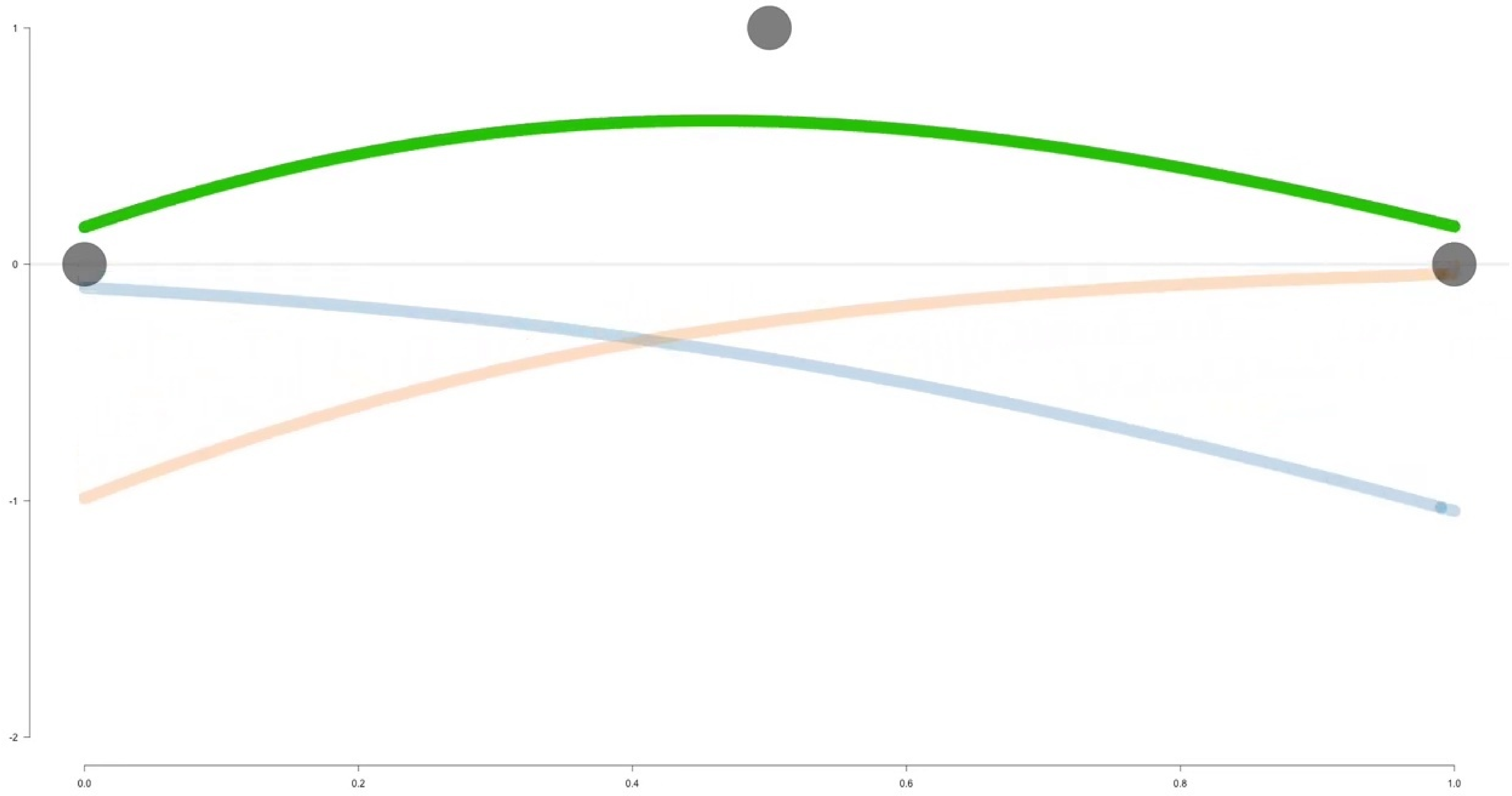


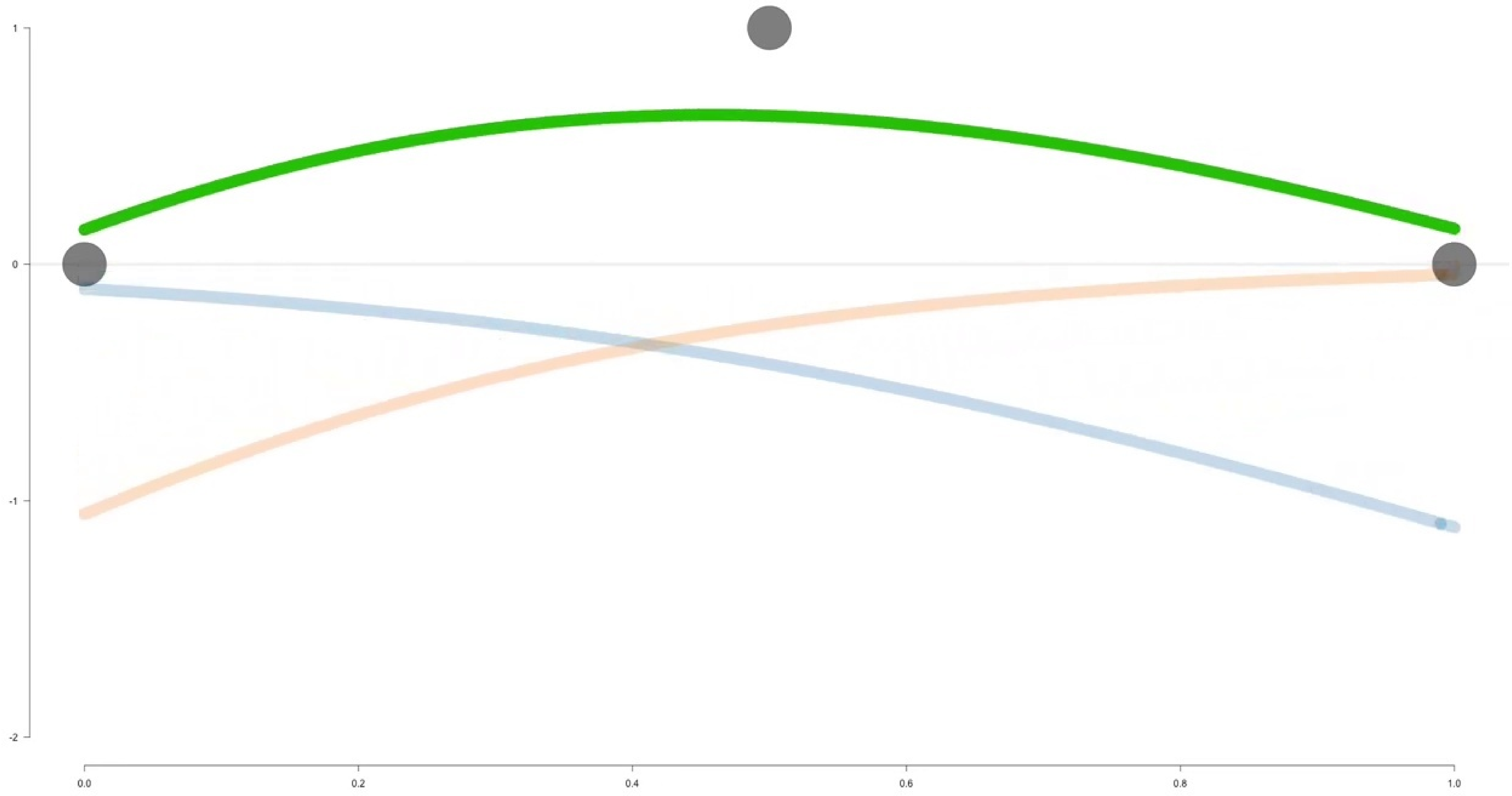
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



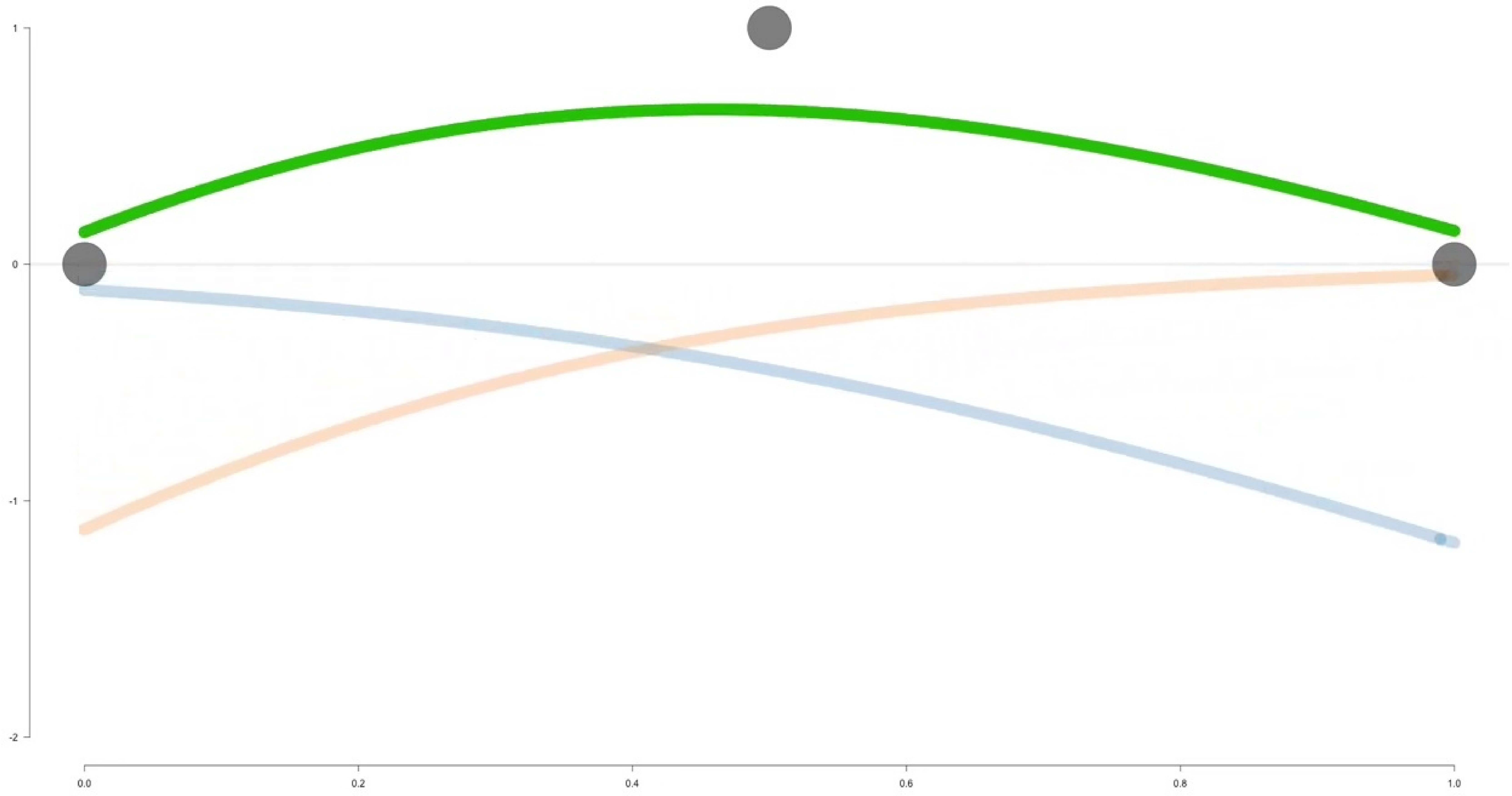
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

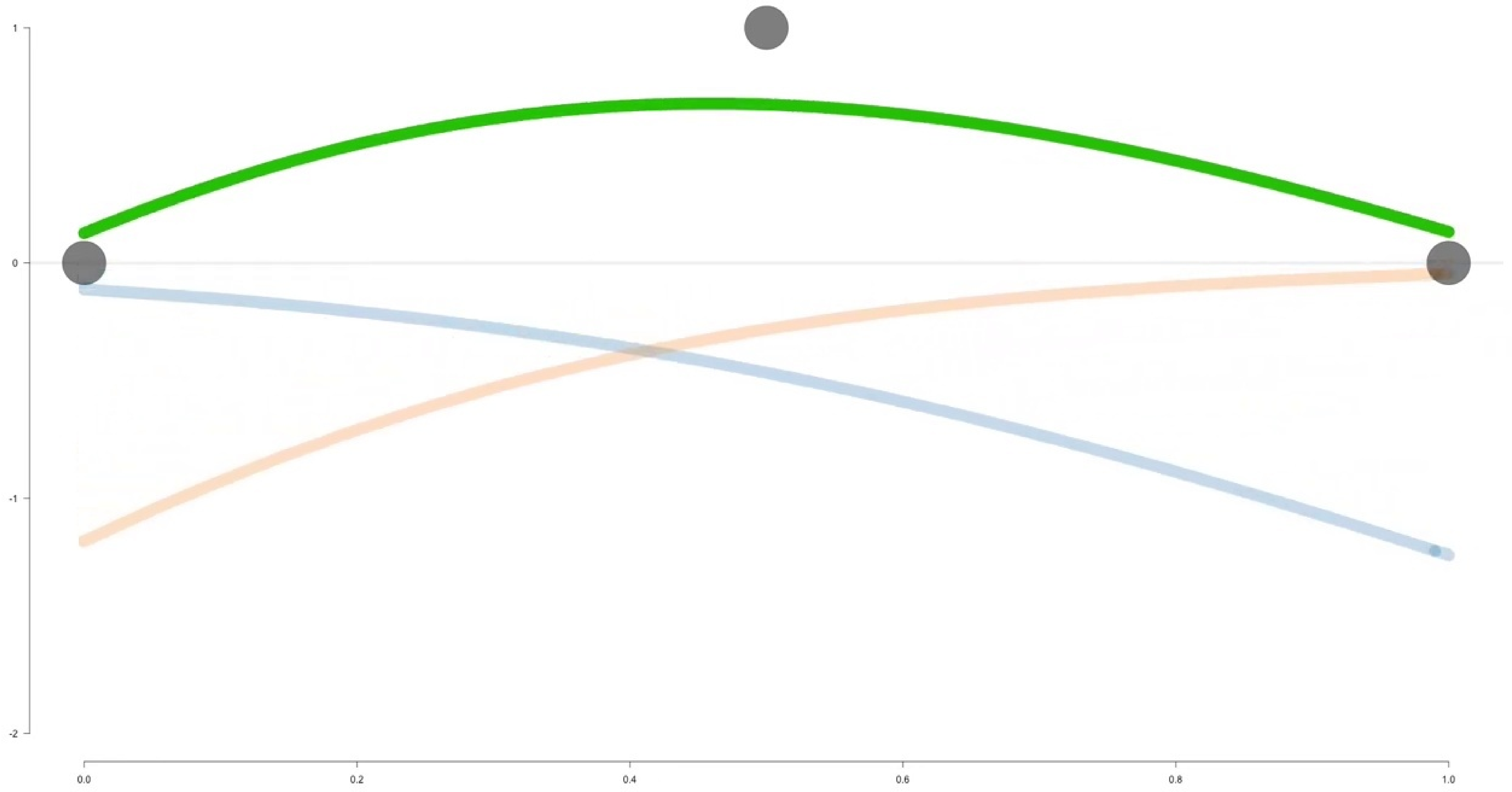

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

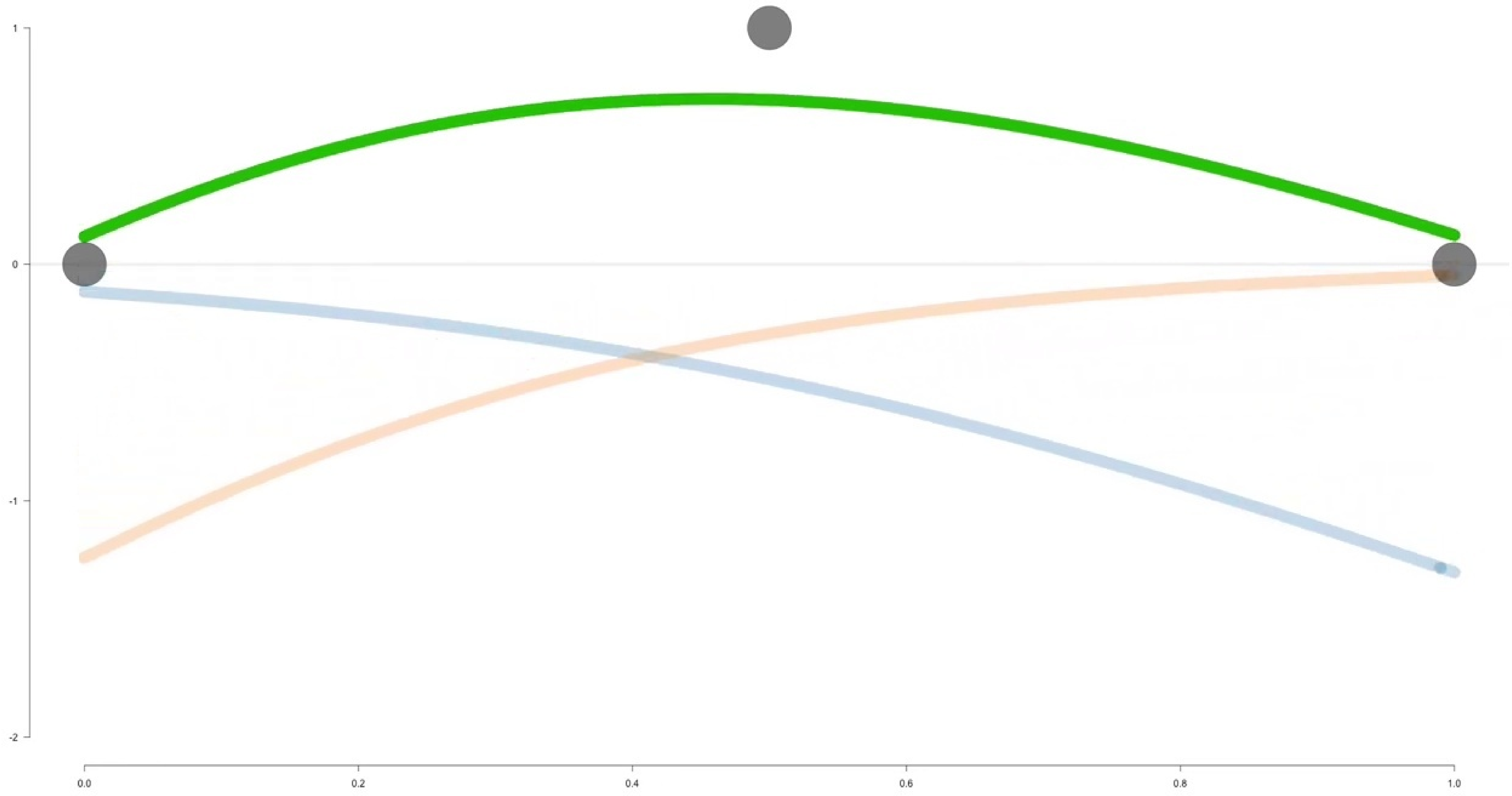


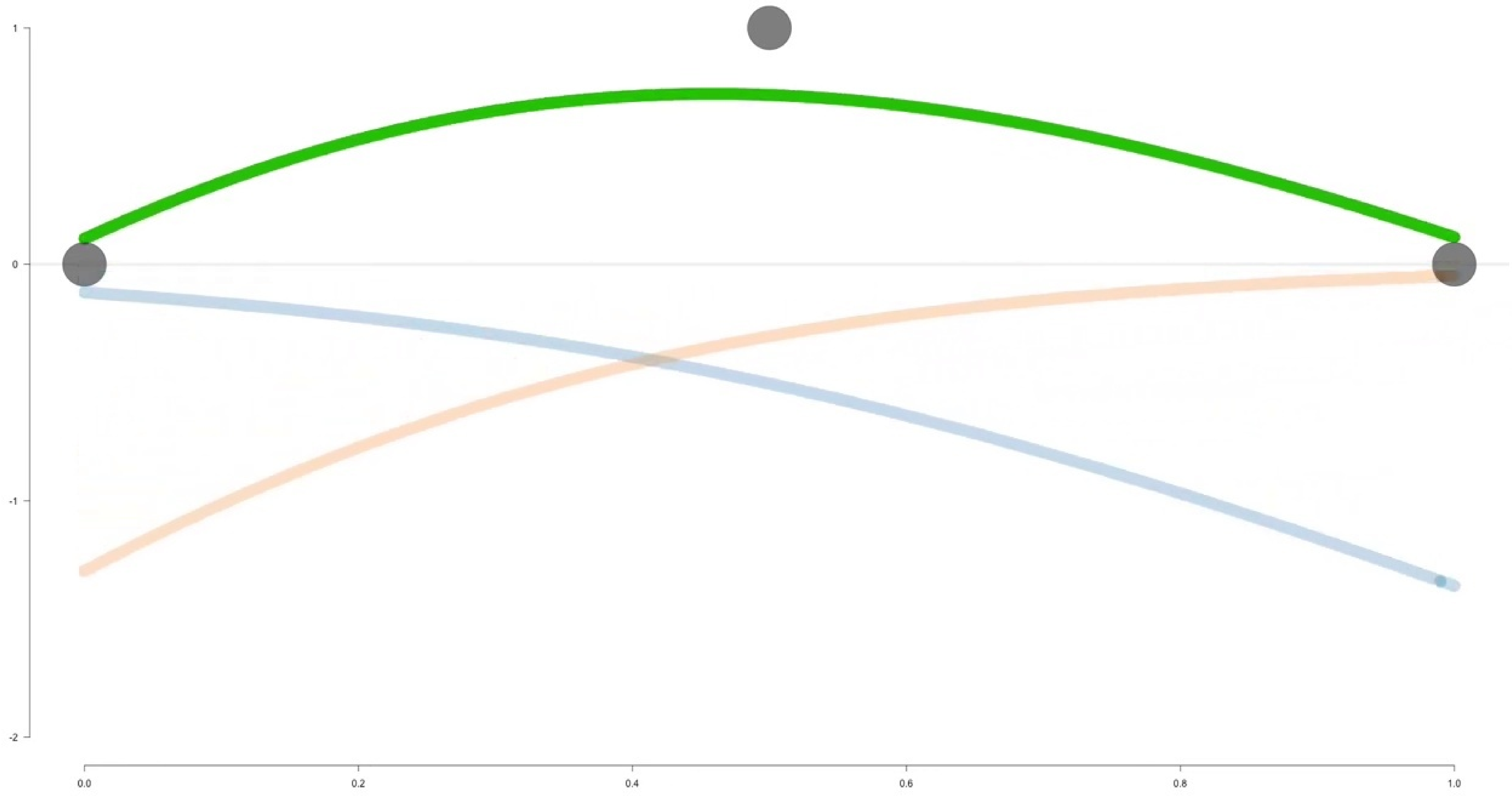
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

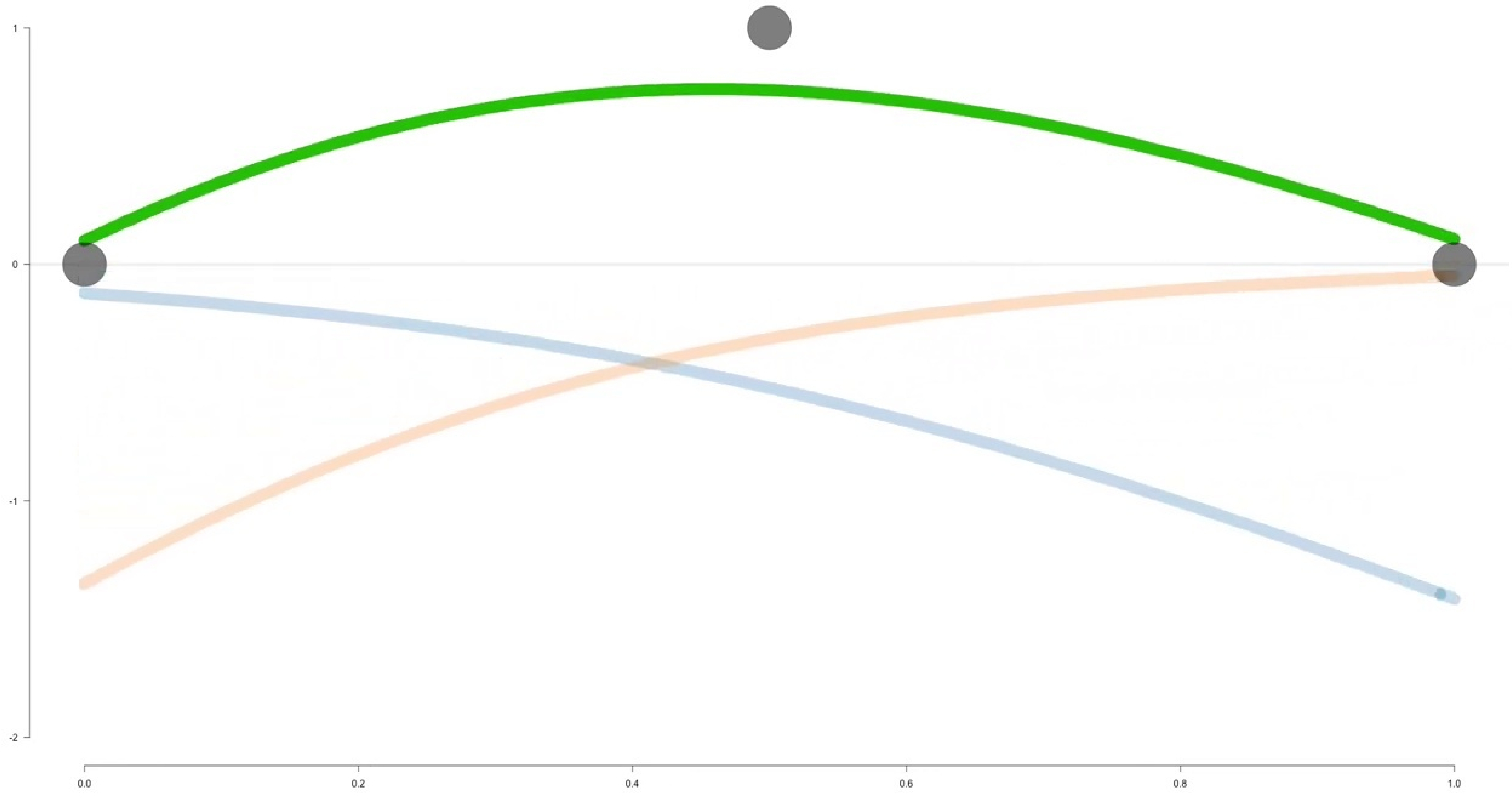

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



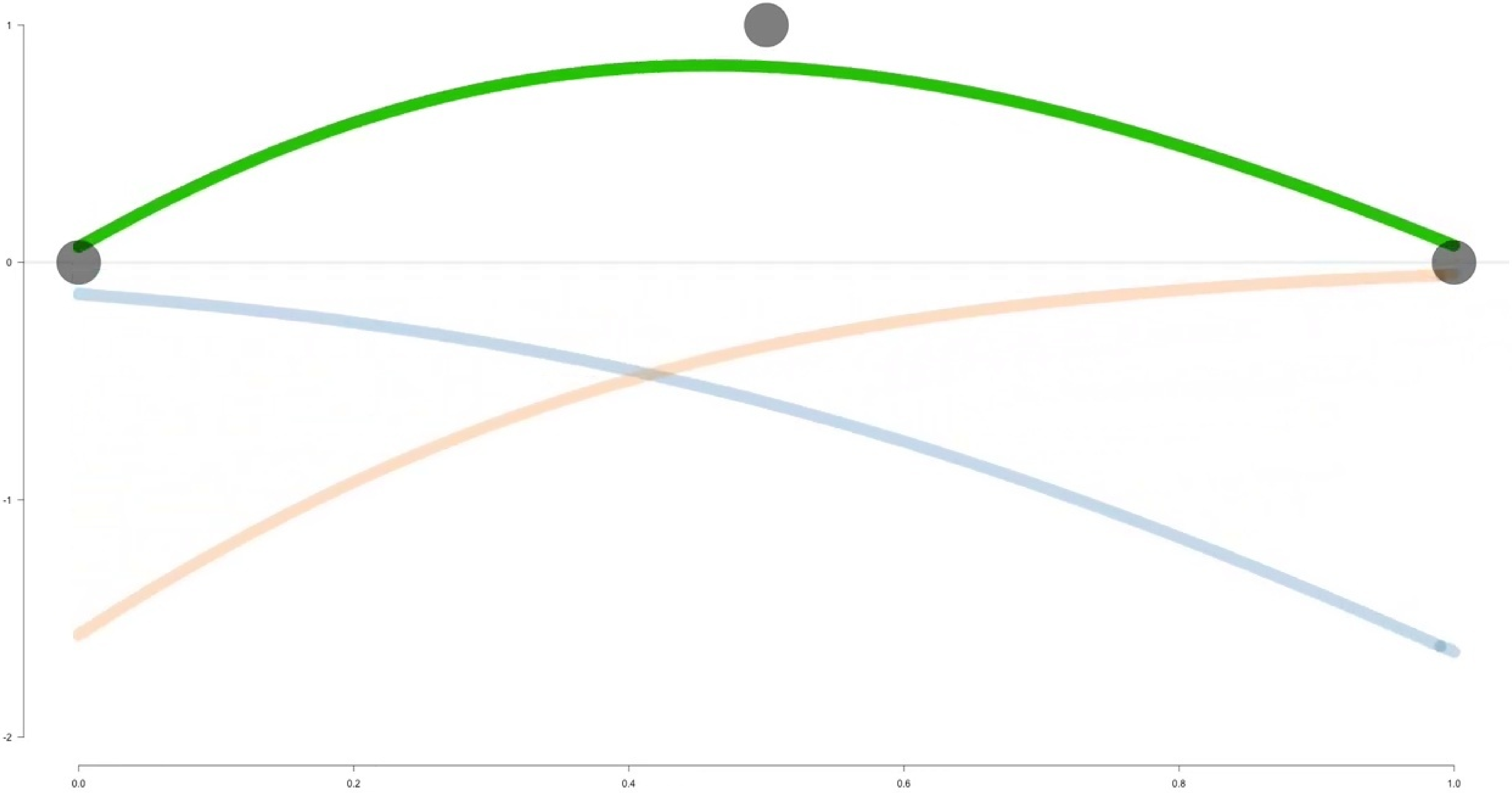
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

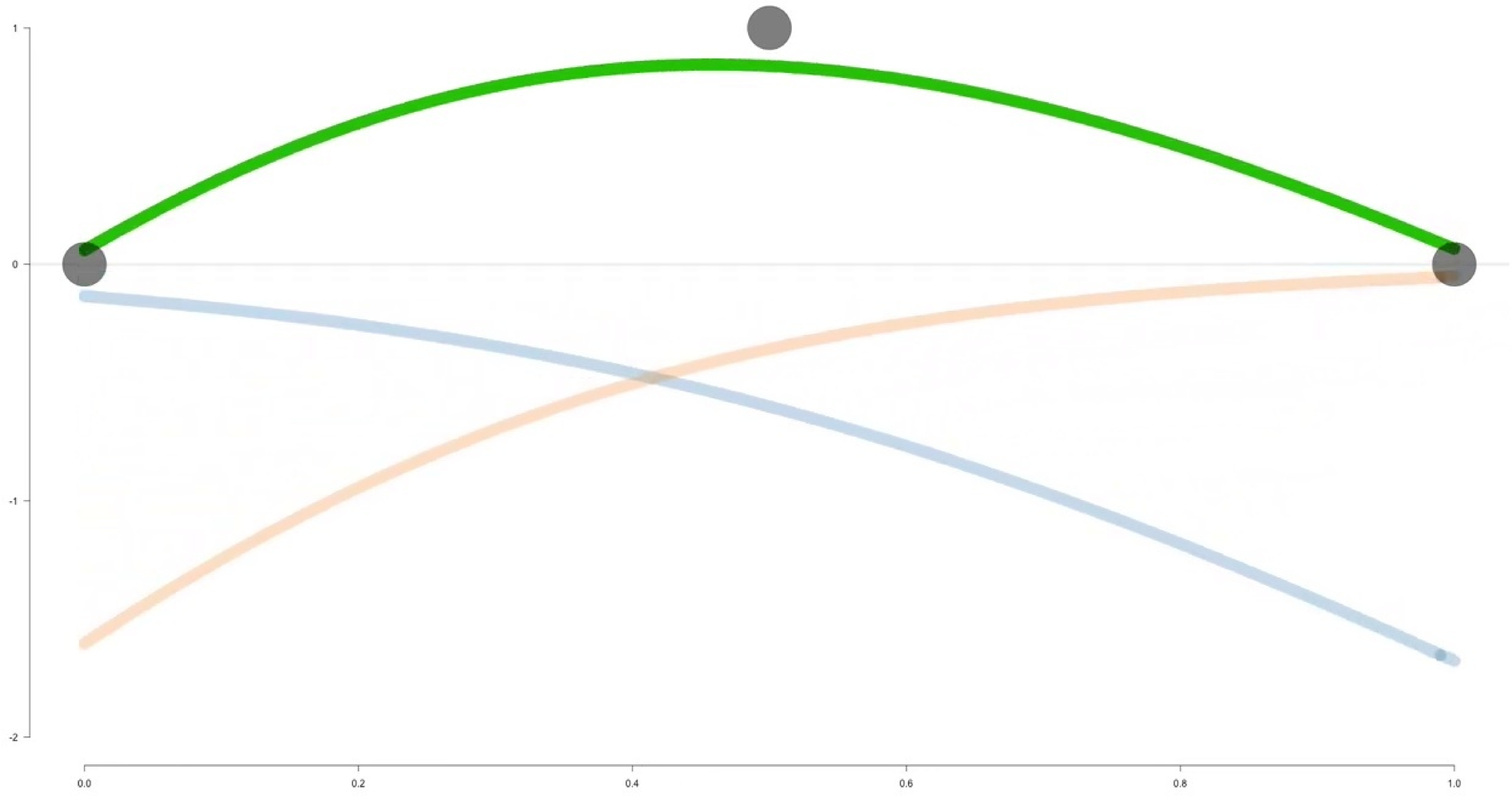

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



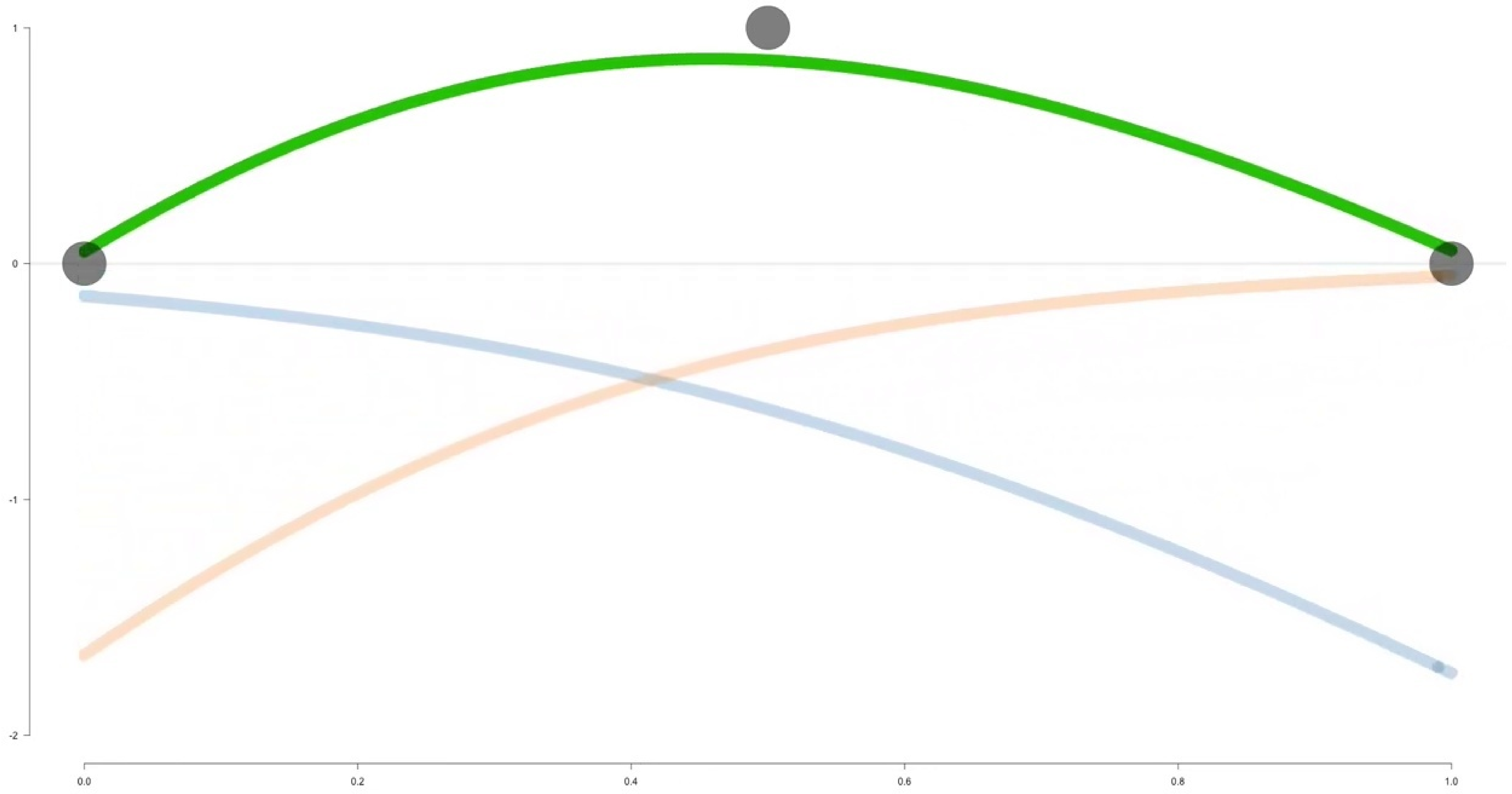
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

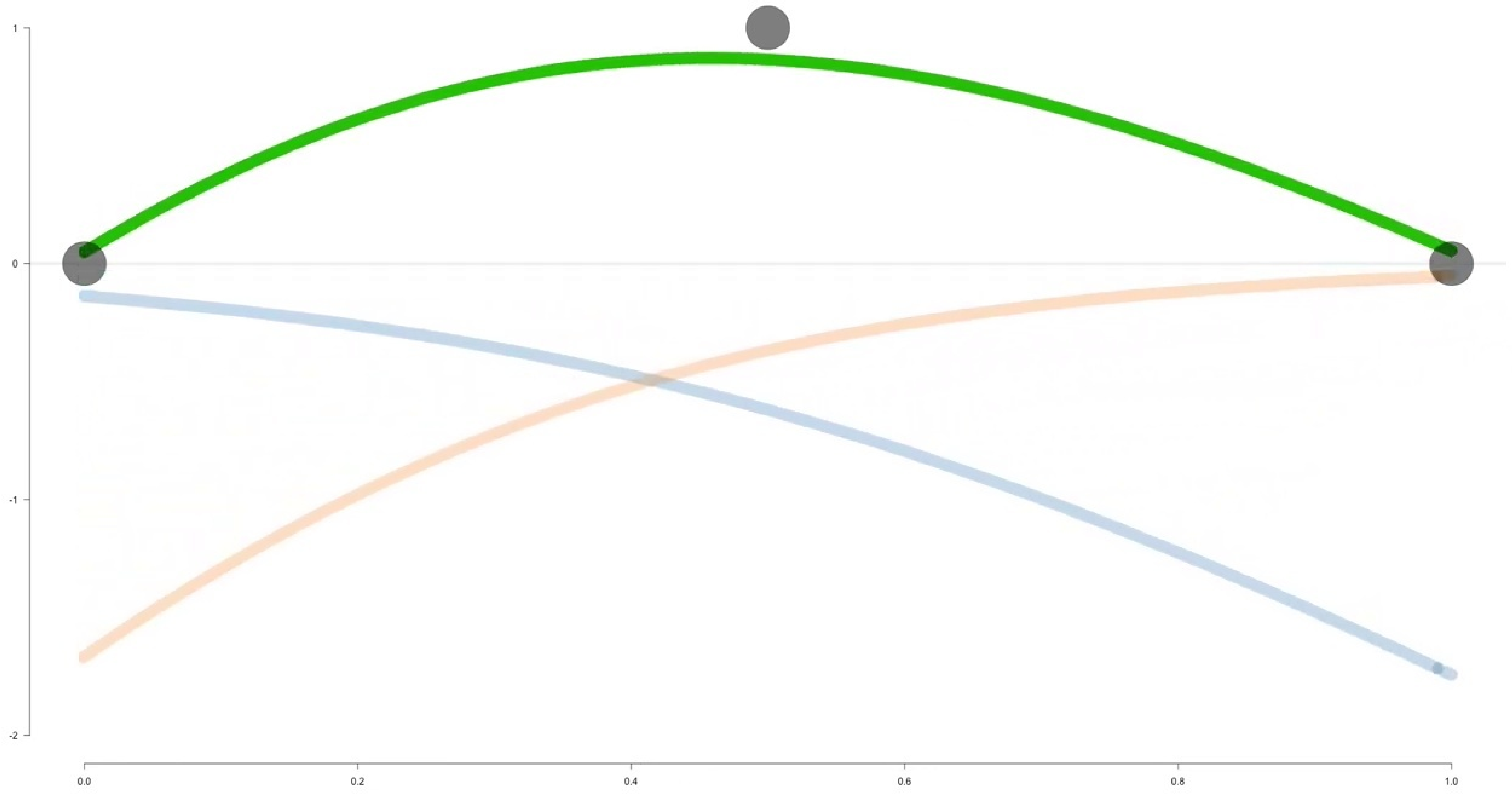


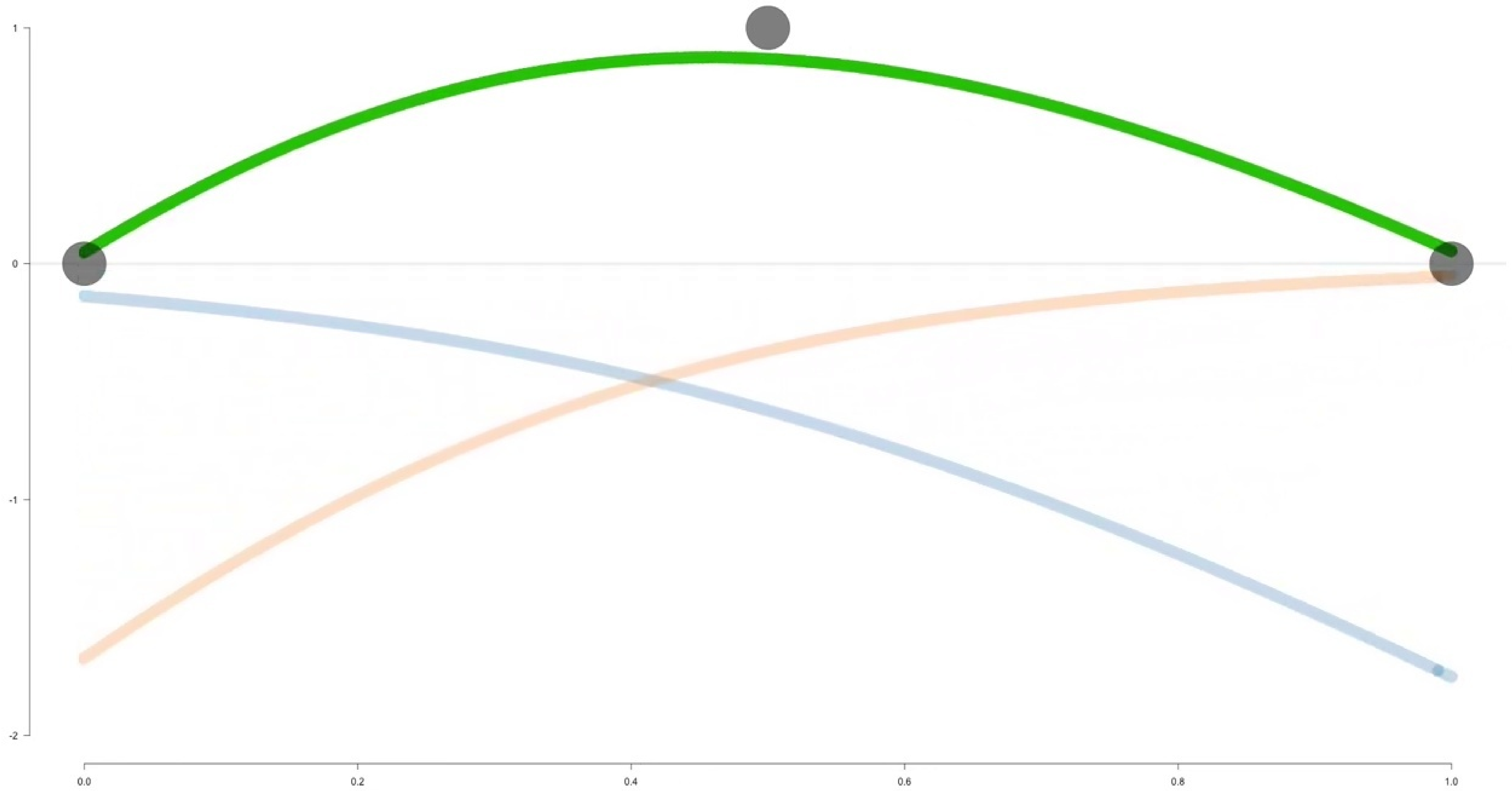
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



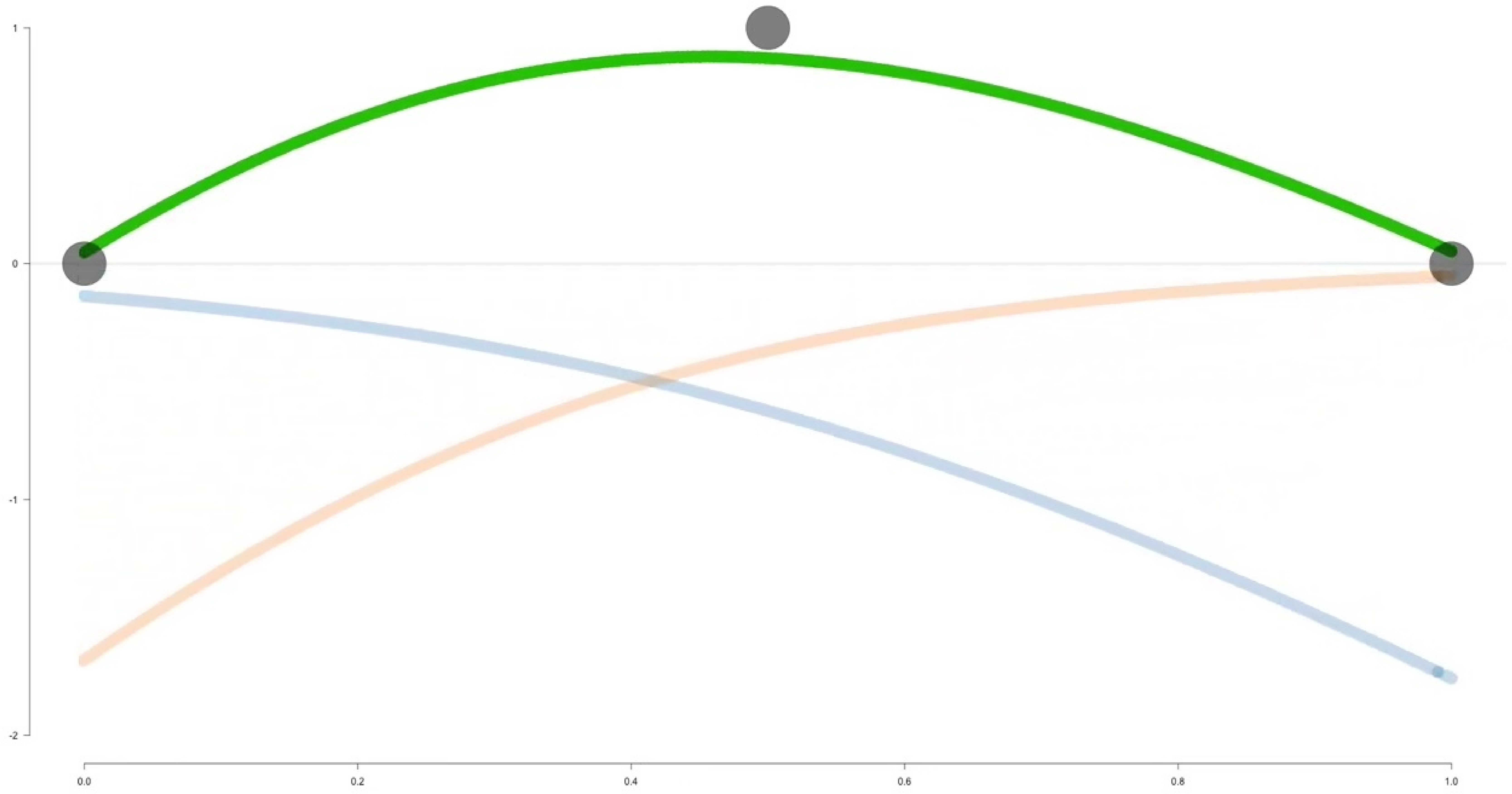
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

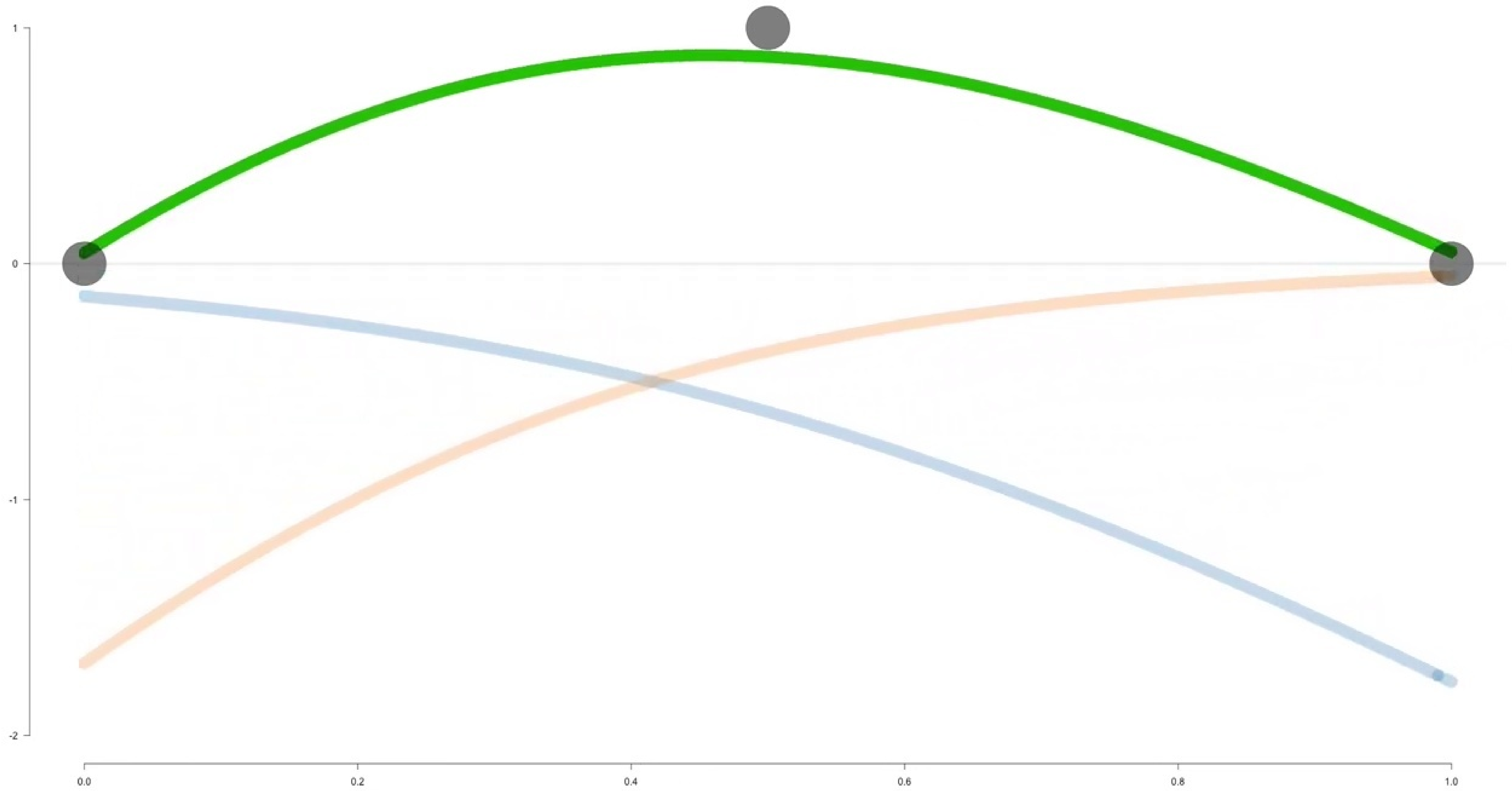

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

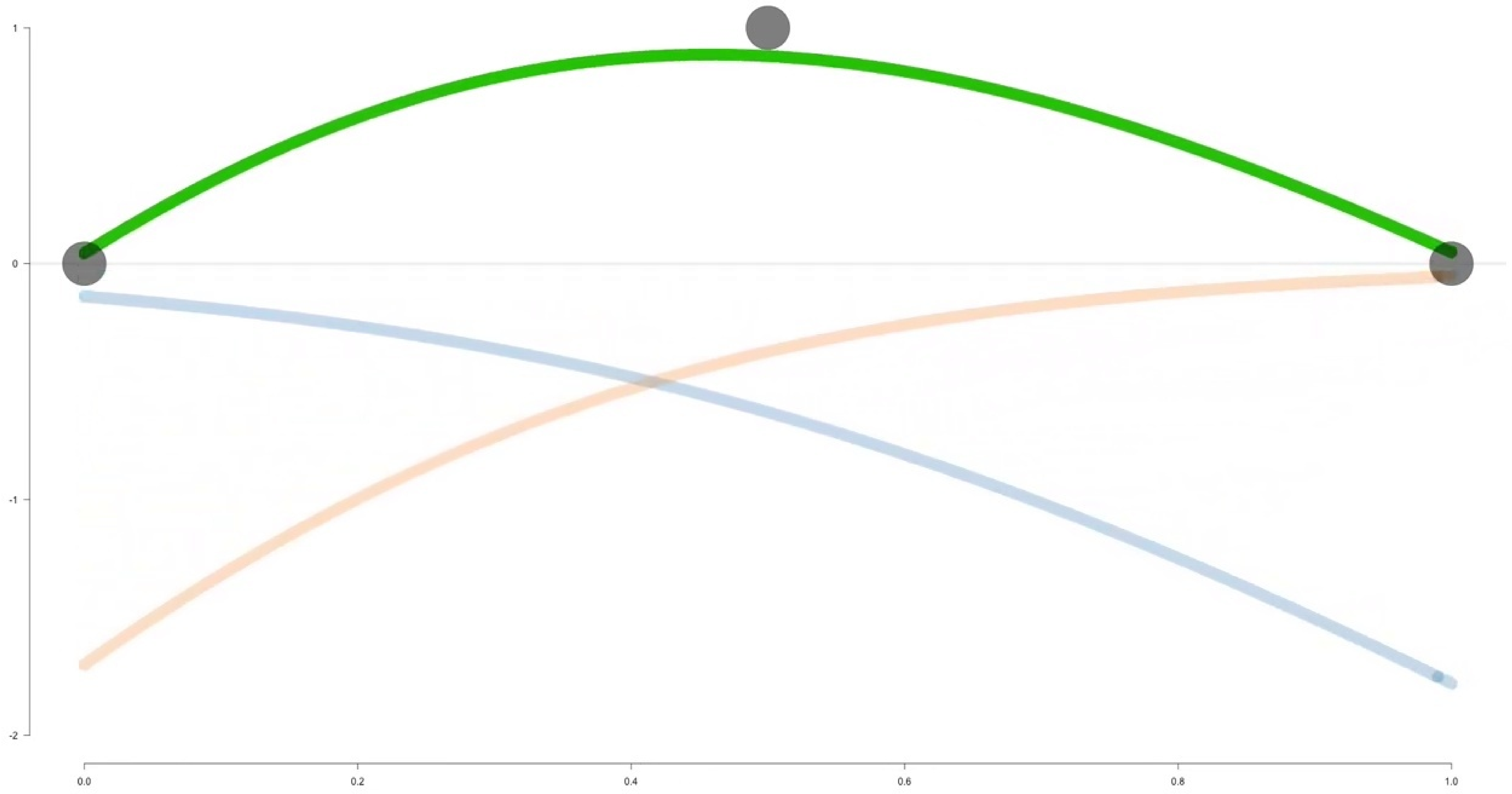


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

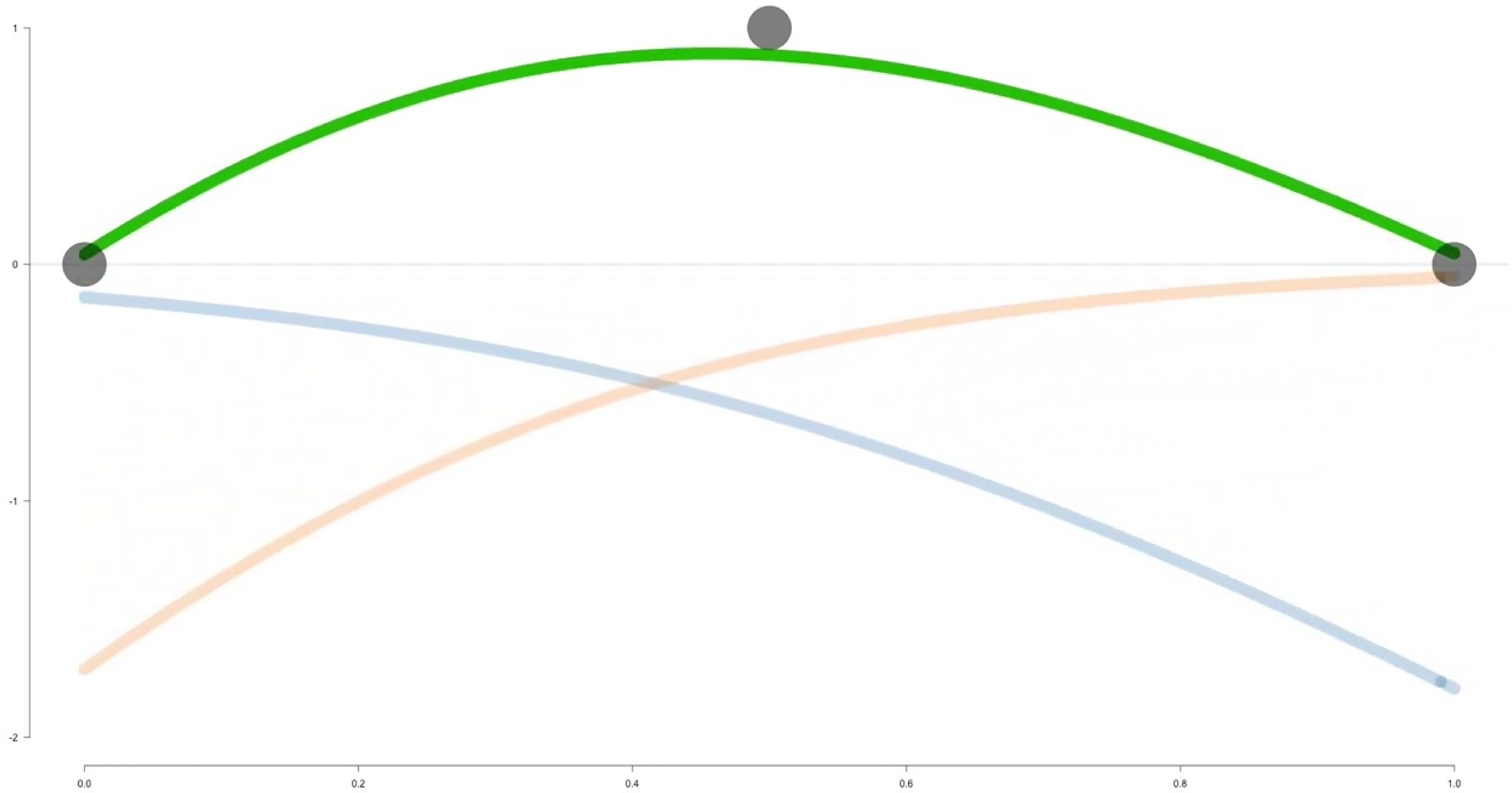

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

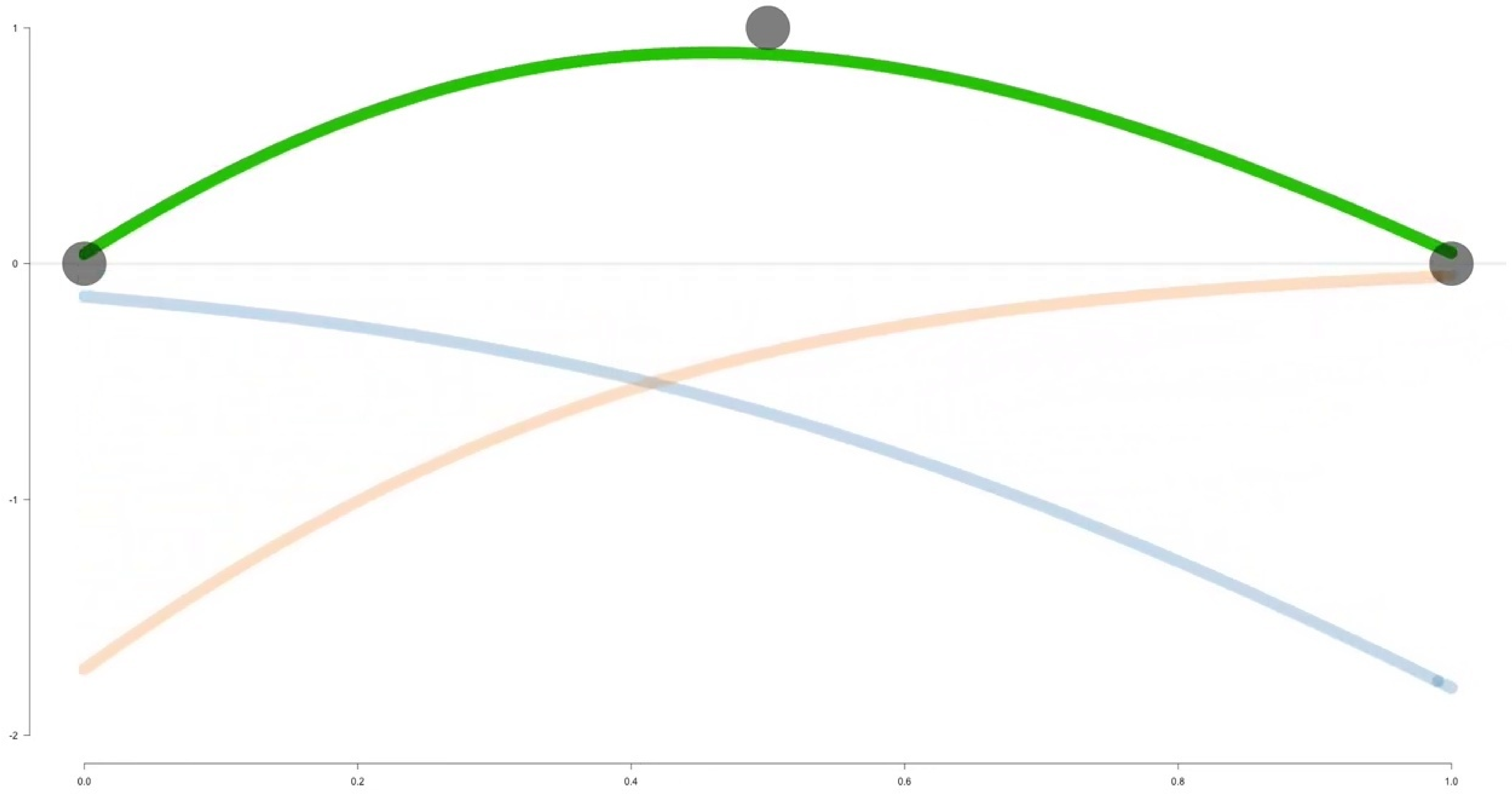


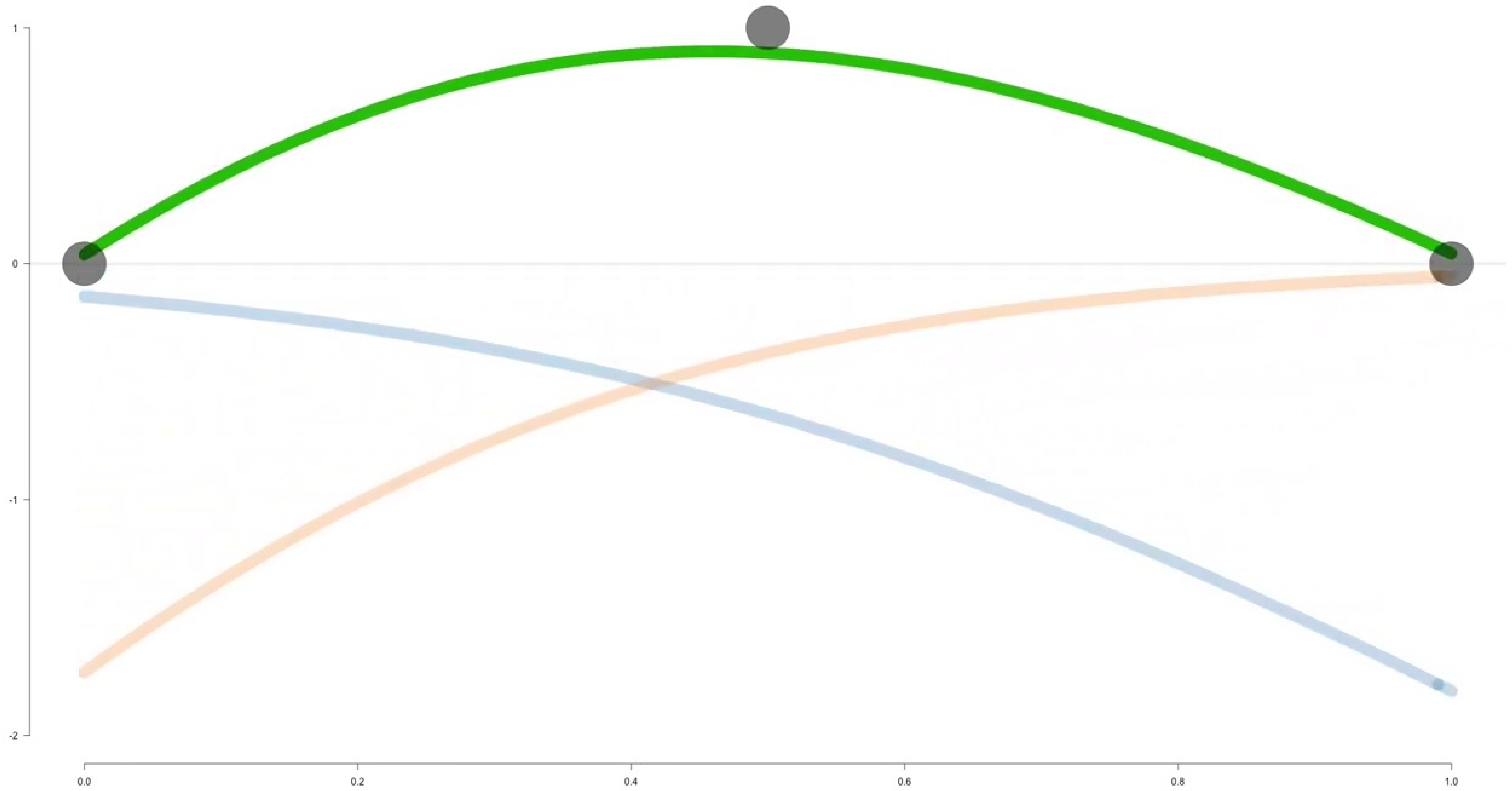
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



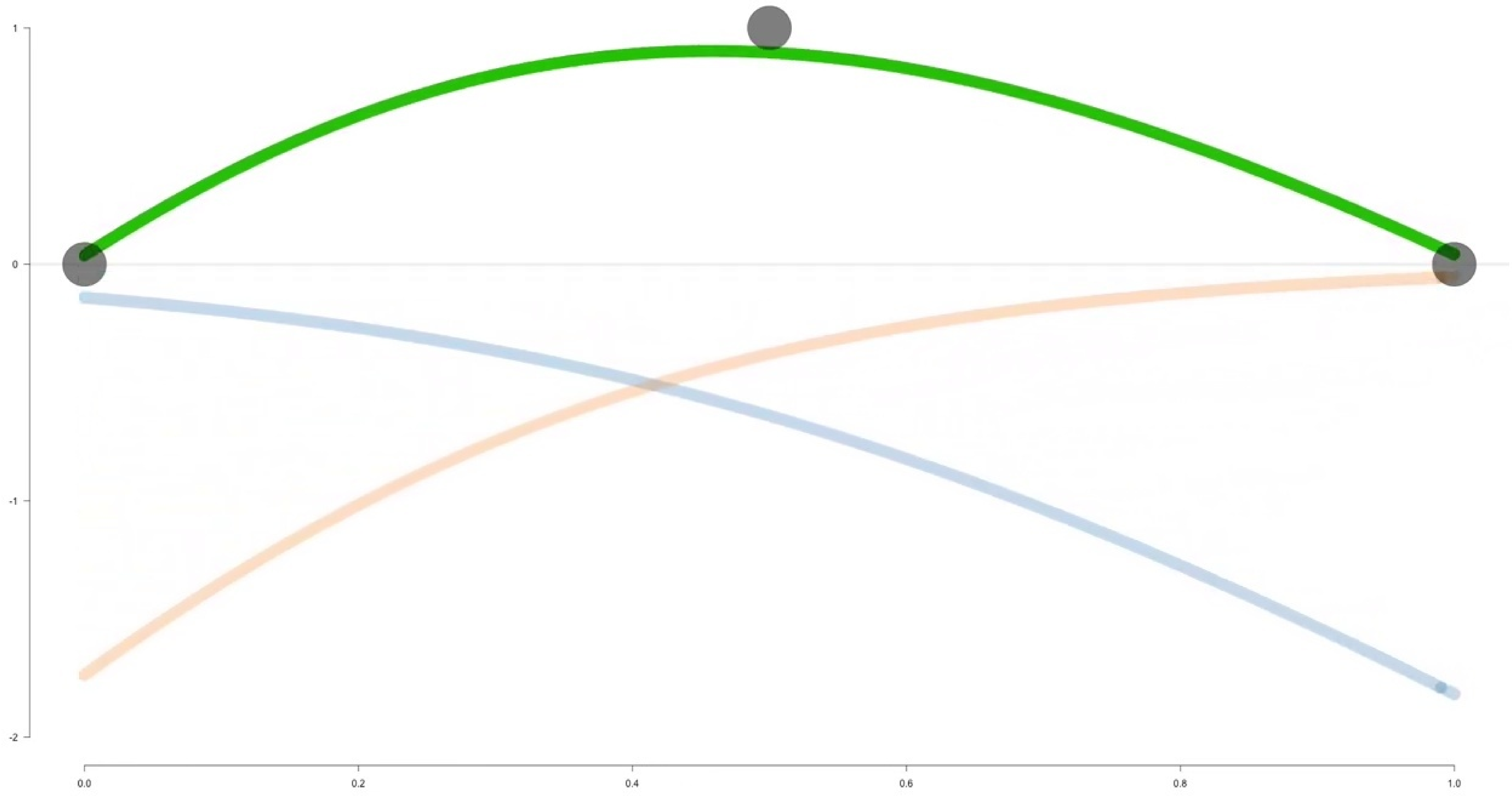
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

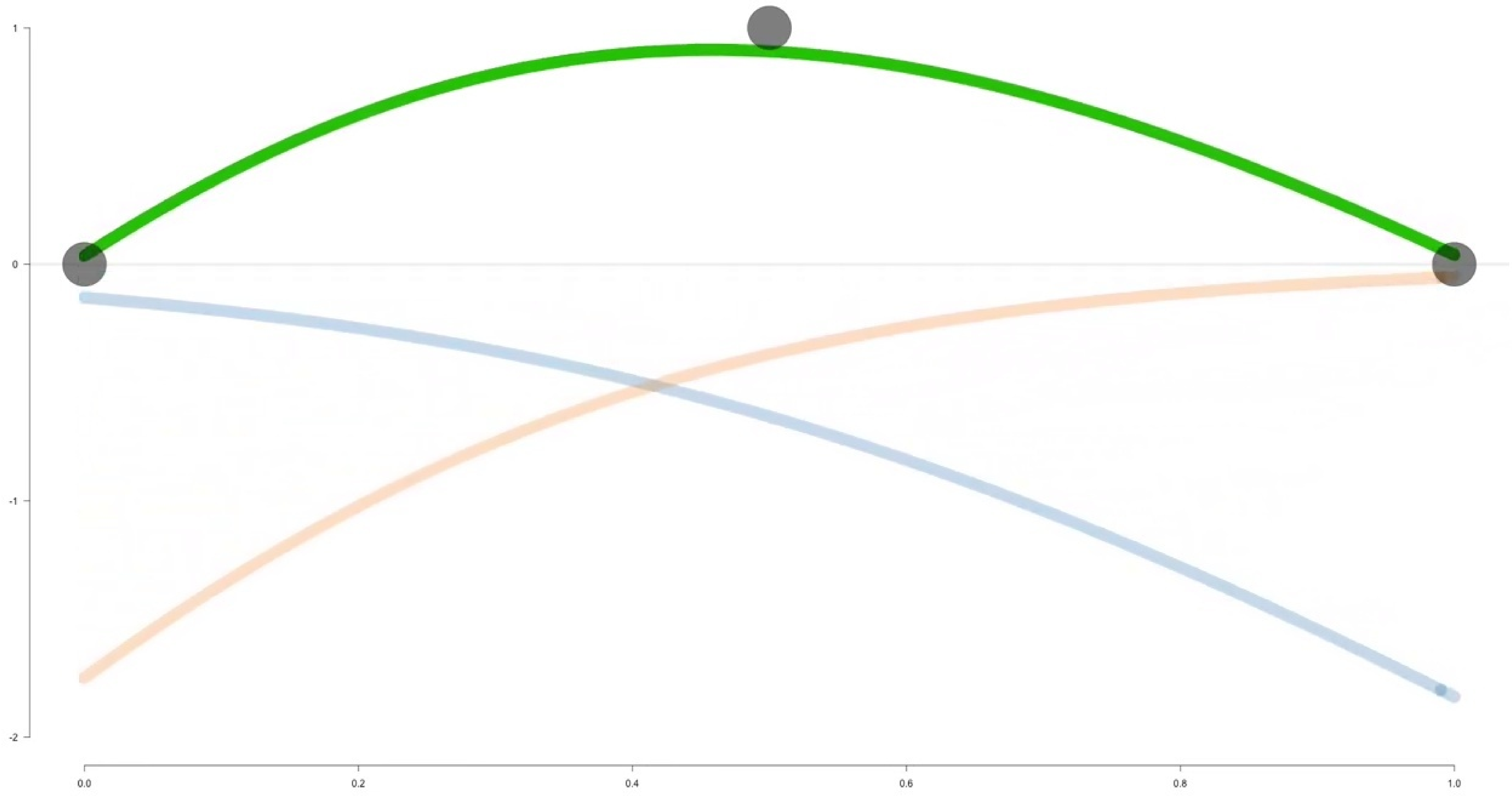

1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



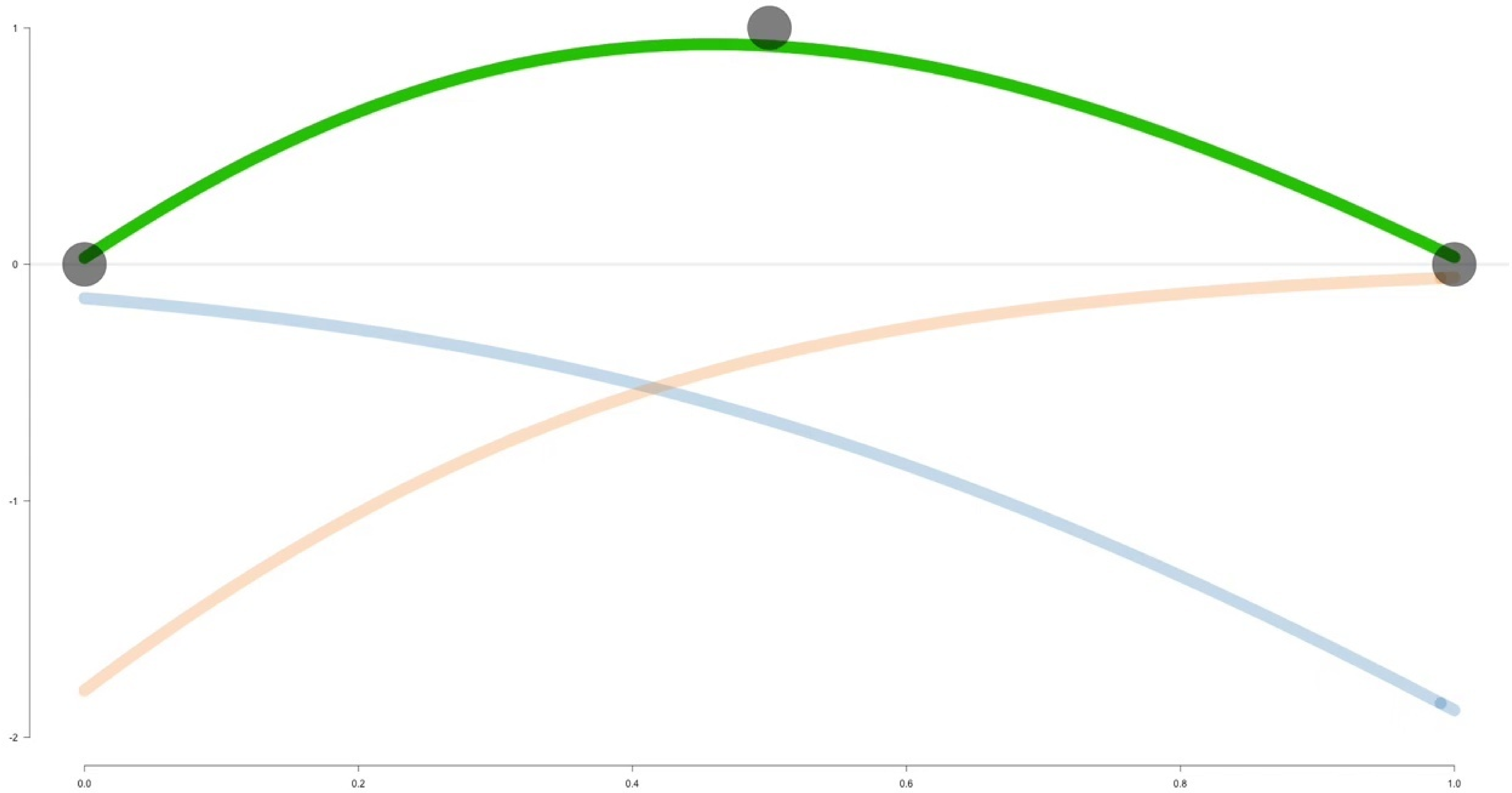
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



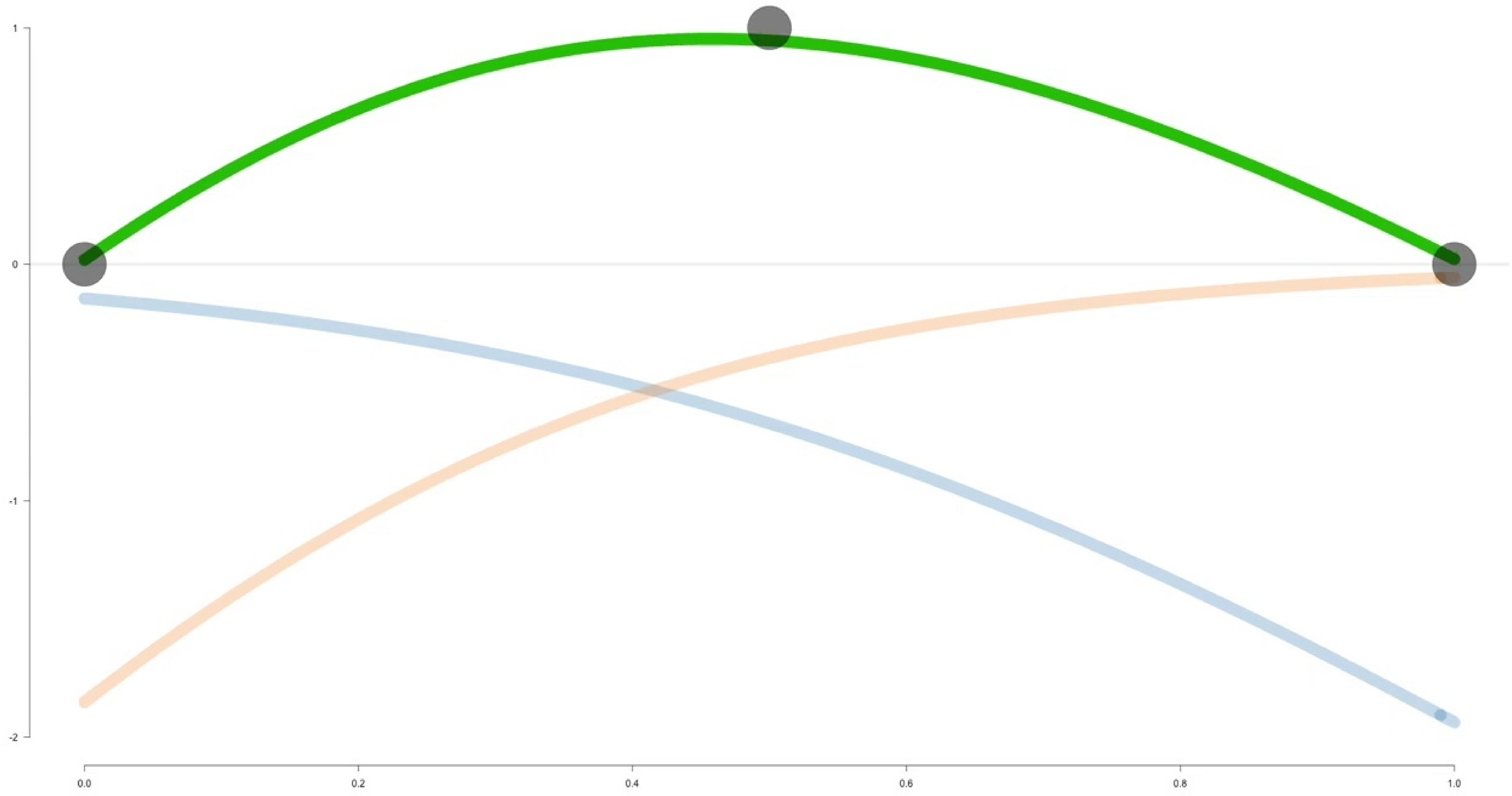
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



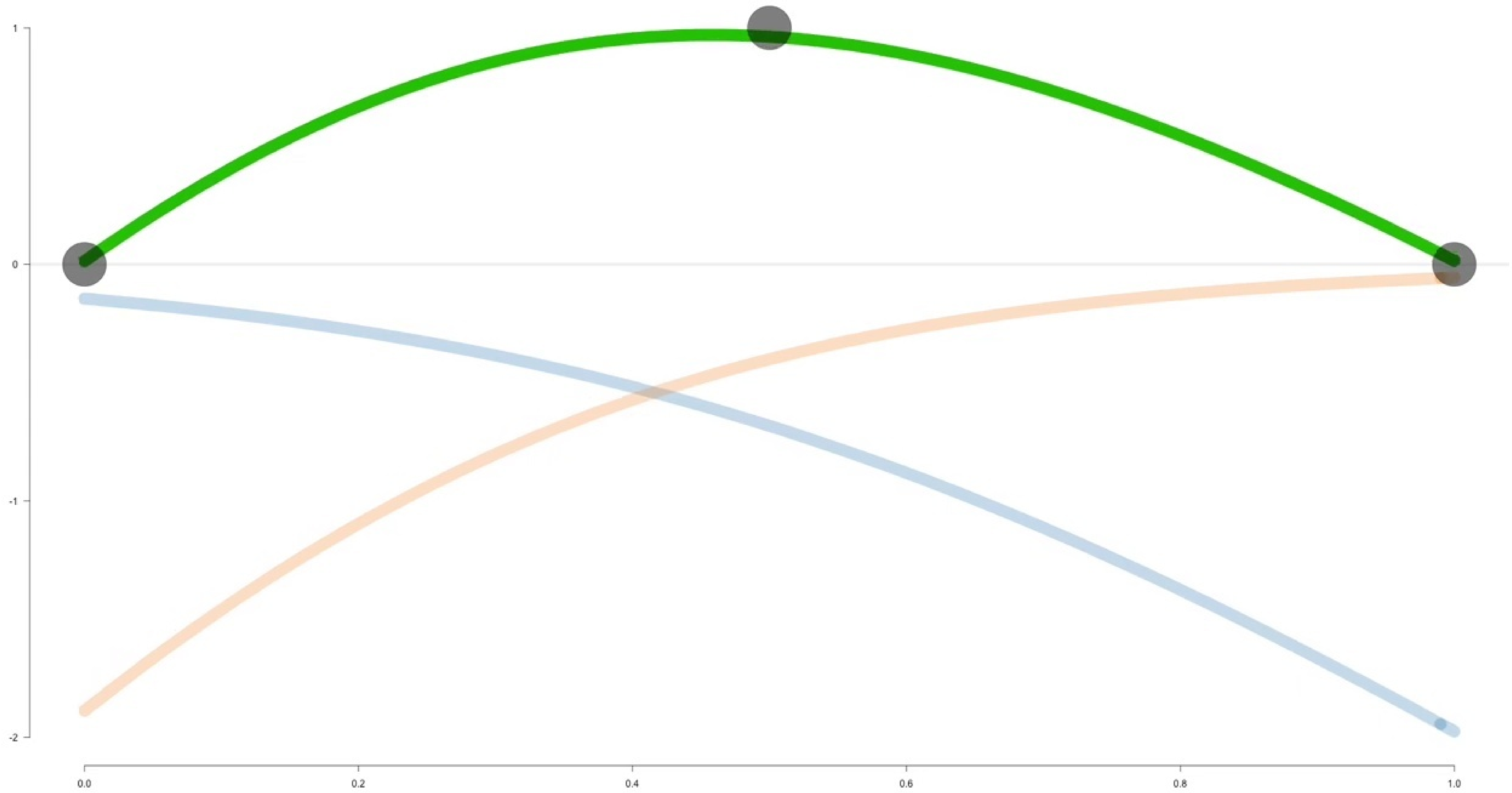
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0

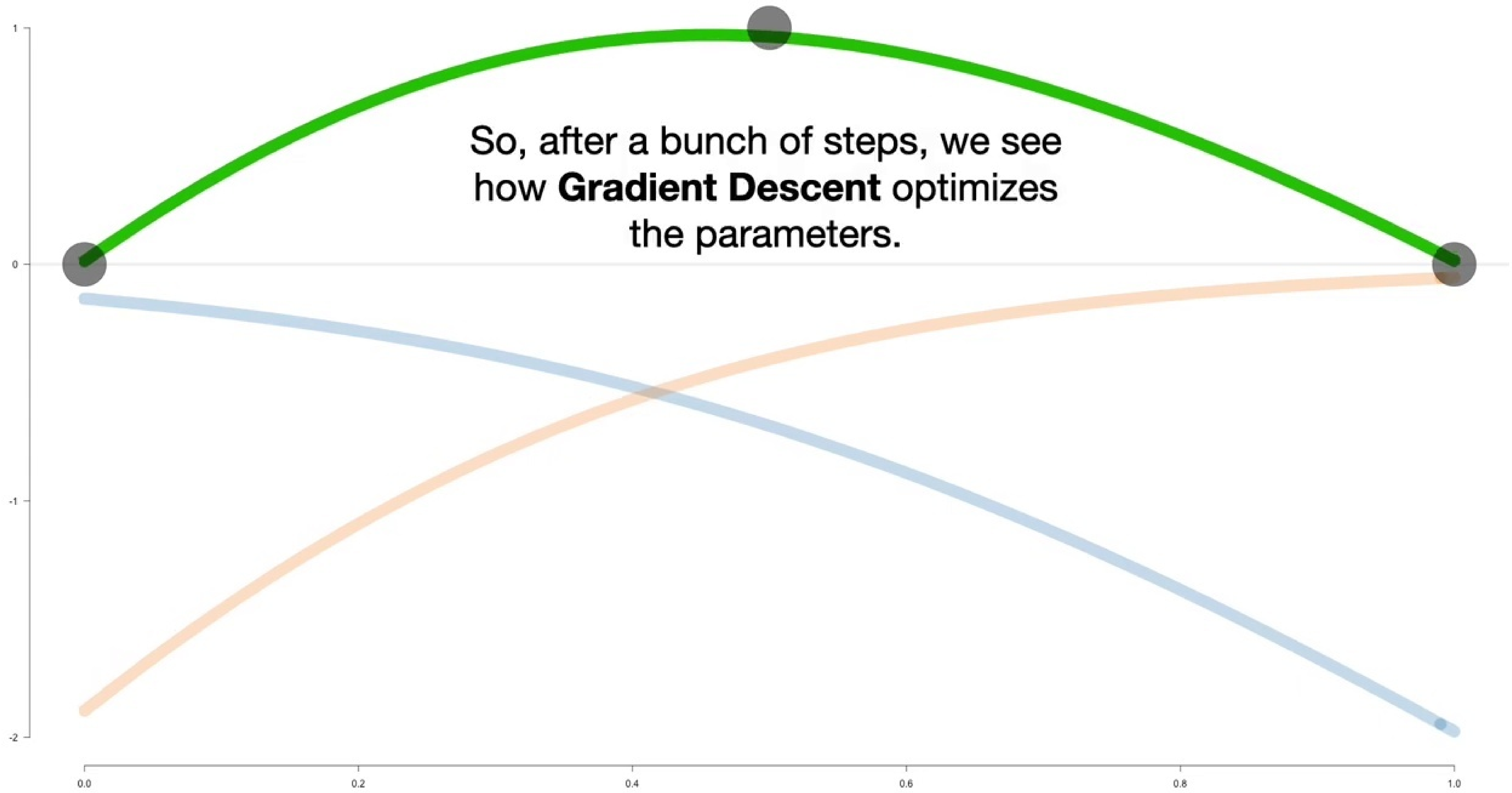


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0


1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0



So, after a bunch of steps, we see how **Gradient Descent** optimizes the parameters.
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0