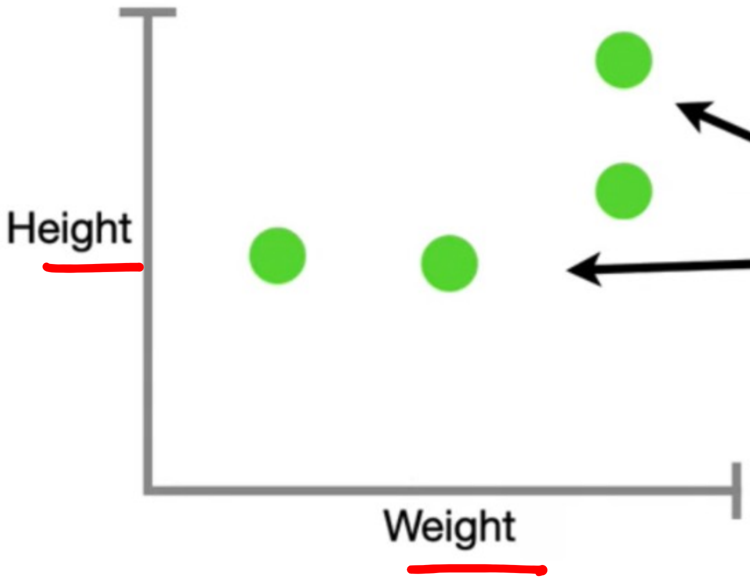
upGrad

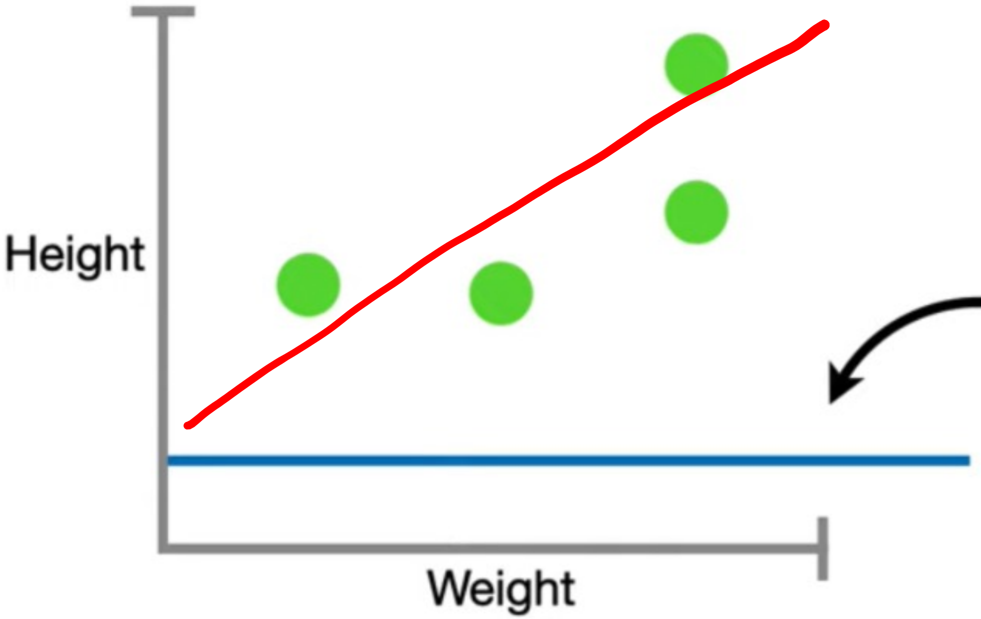
Module : Regularization- Ridge & Lasso

Instructor: Dr. Darshan Ingle

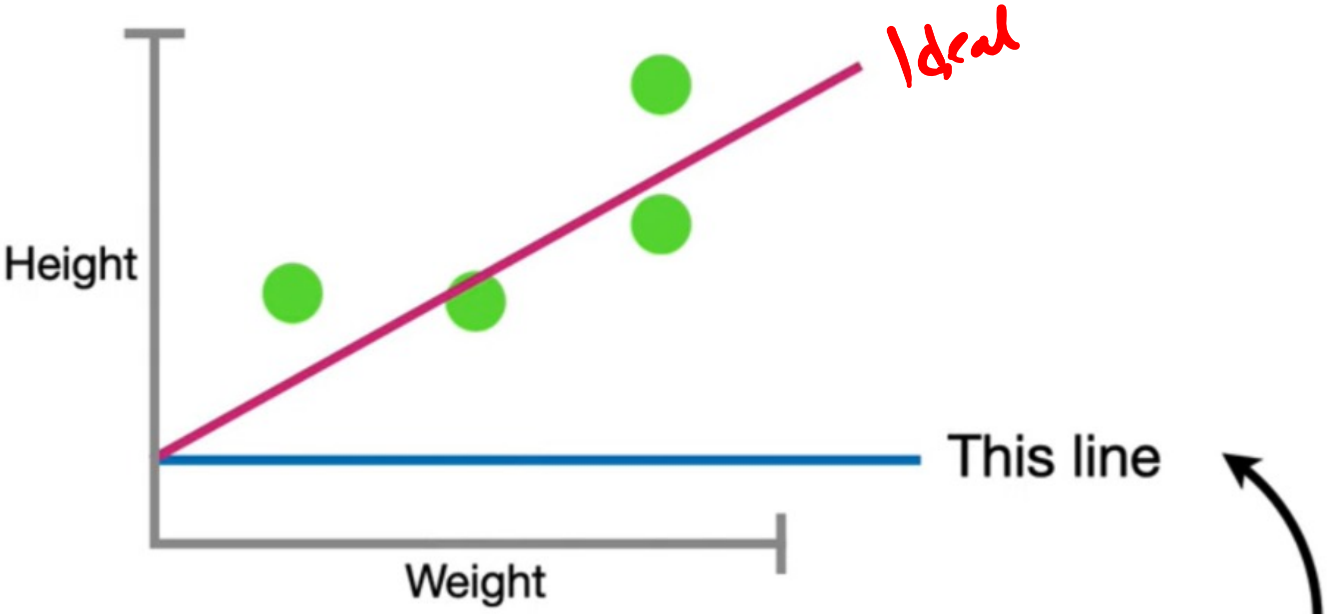


To show you the difference between **Ridge** and **Lasso Regression**, we're going to use a very simple dataset that consists of **Weight** and **Height** measurements.

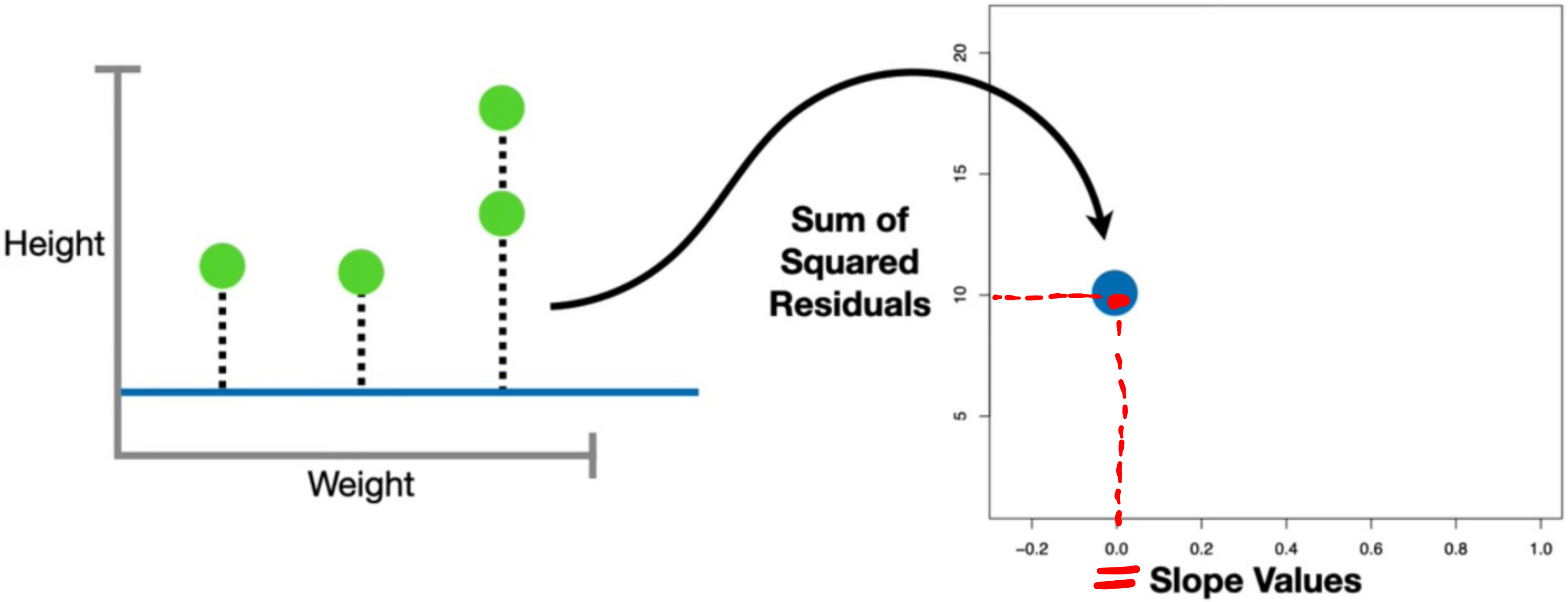
upGrad



And we'll start by fitting this horizontal line to the data.

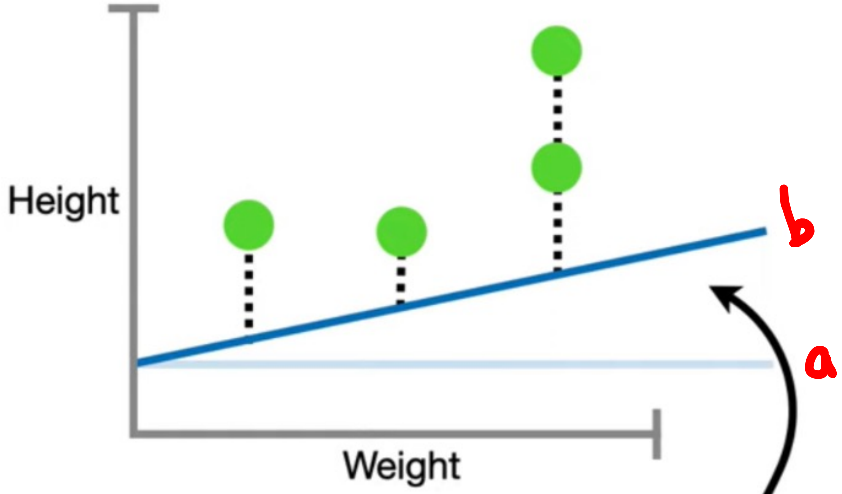


In order to compare the horizontal line to other lines fit to the data...

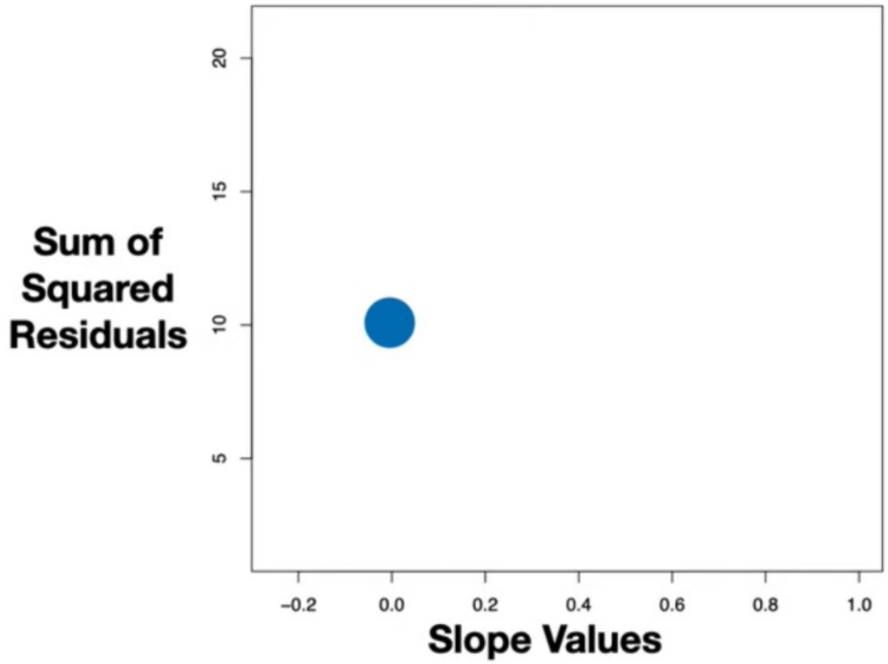


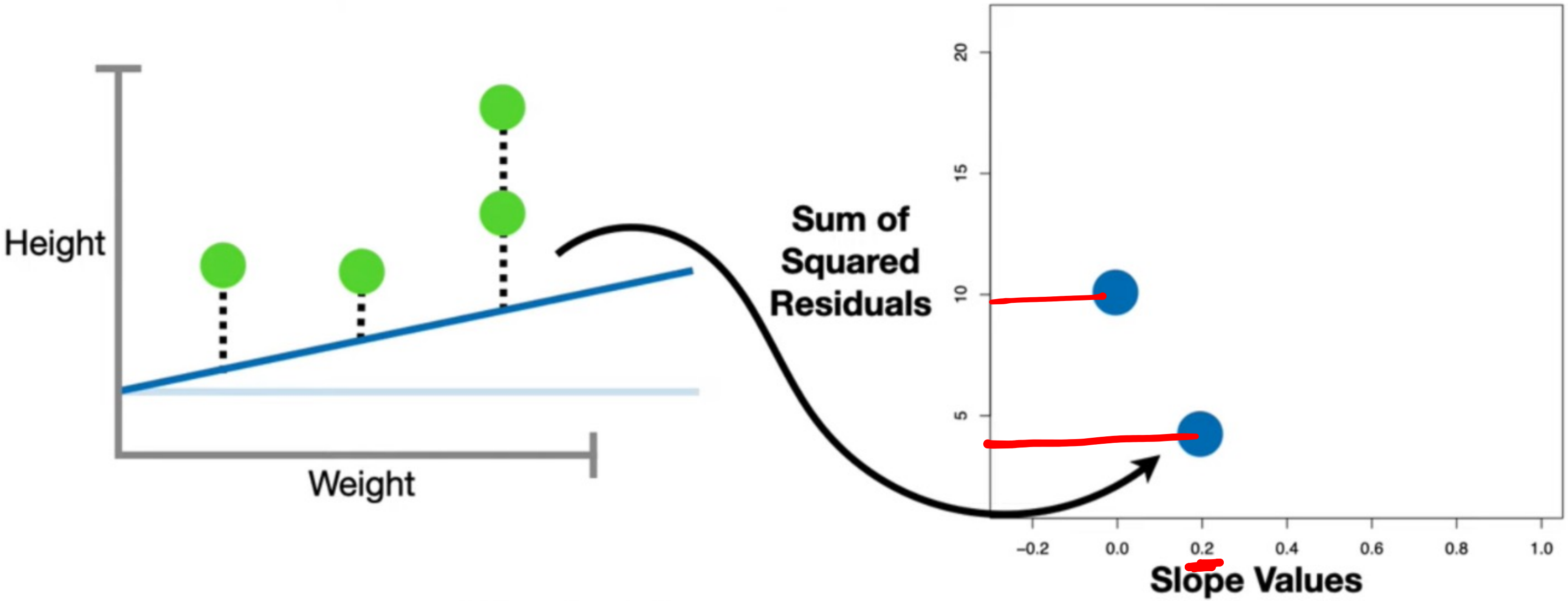
...we will plot the **Sum of the Squared Residuals** on a graph.

Gradient Descent

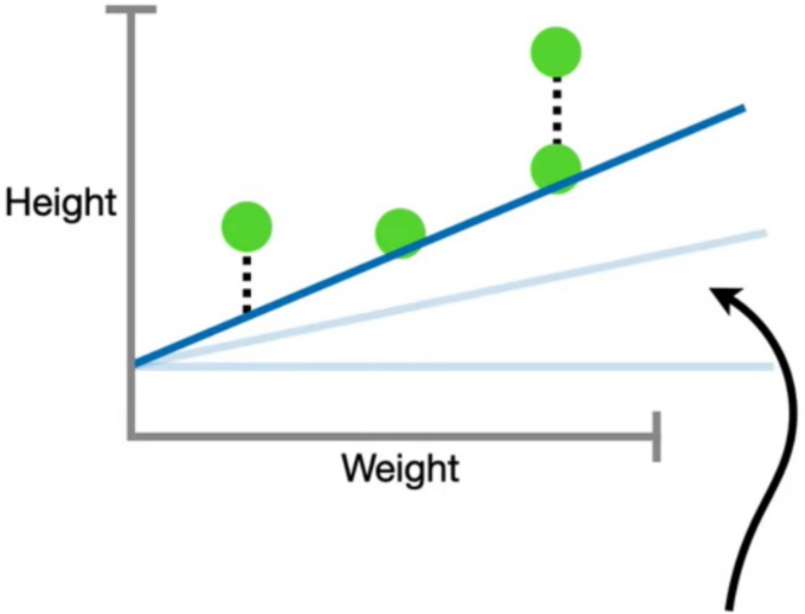


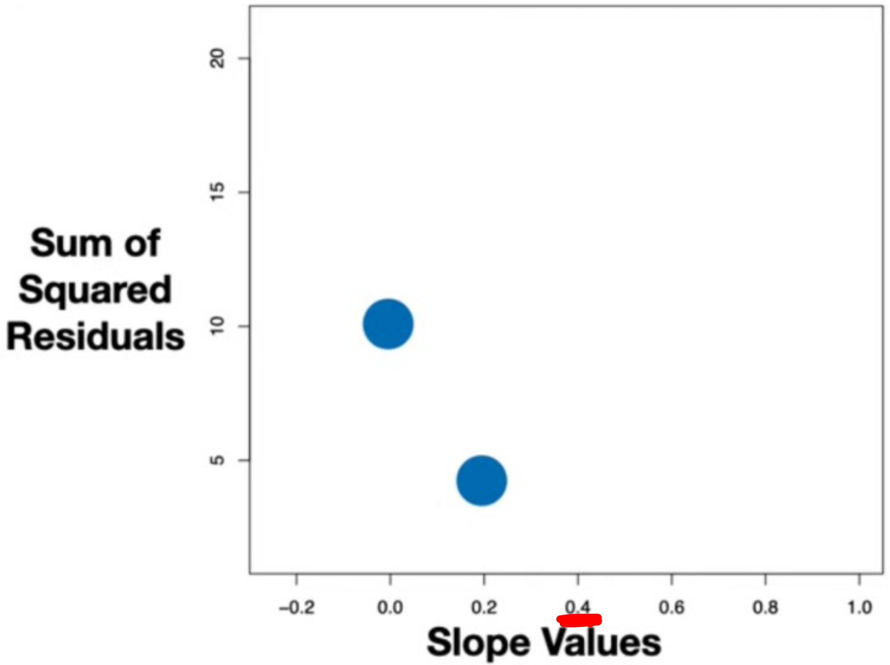
Now let's increase the slope to **0.2**...



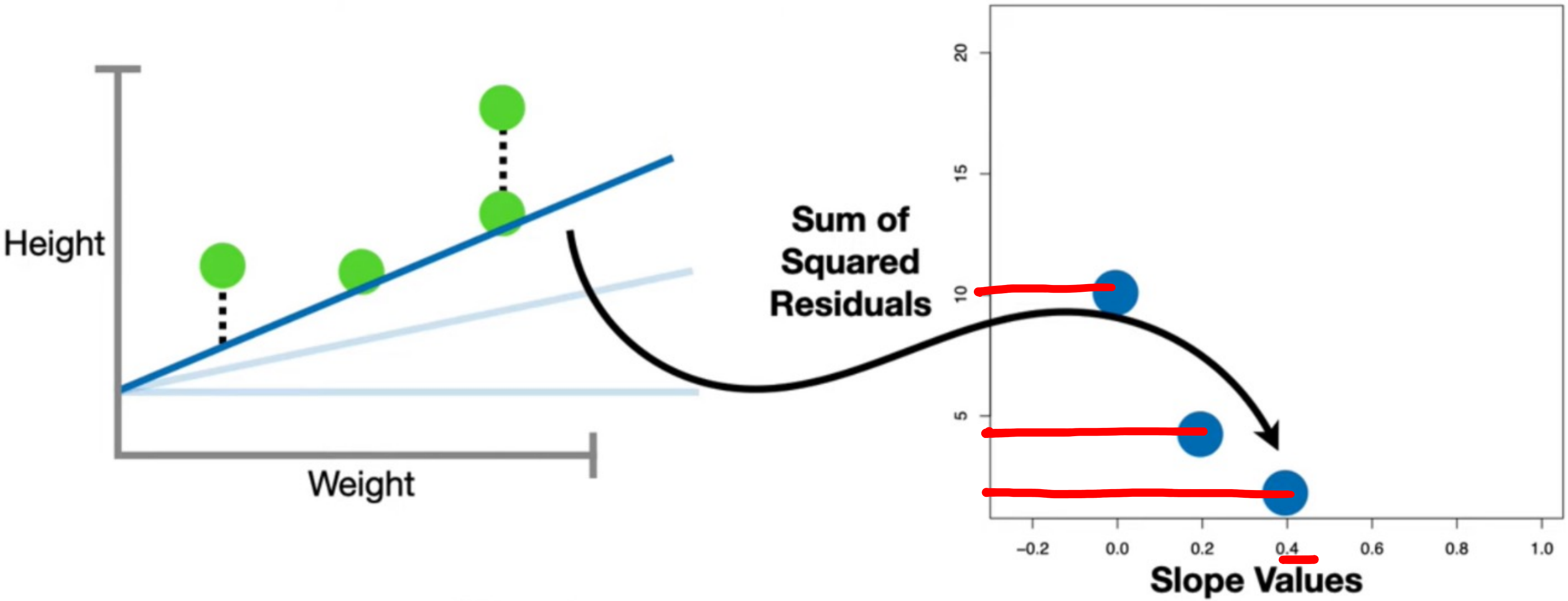


...and calculate a new value for **Sum of the Squared Residuals.**

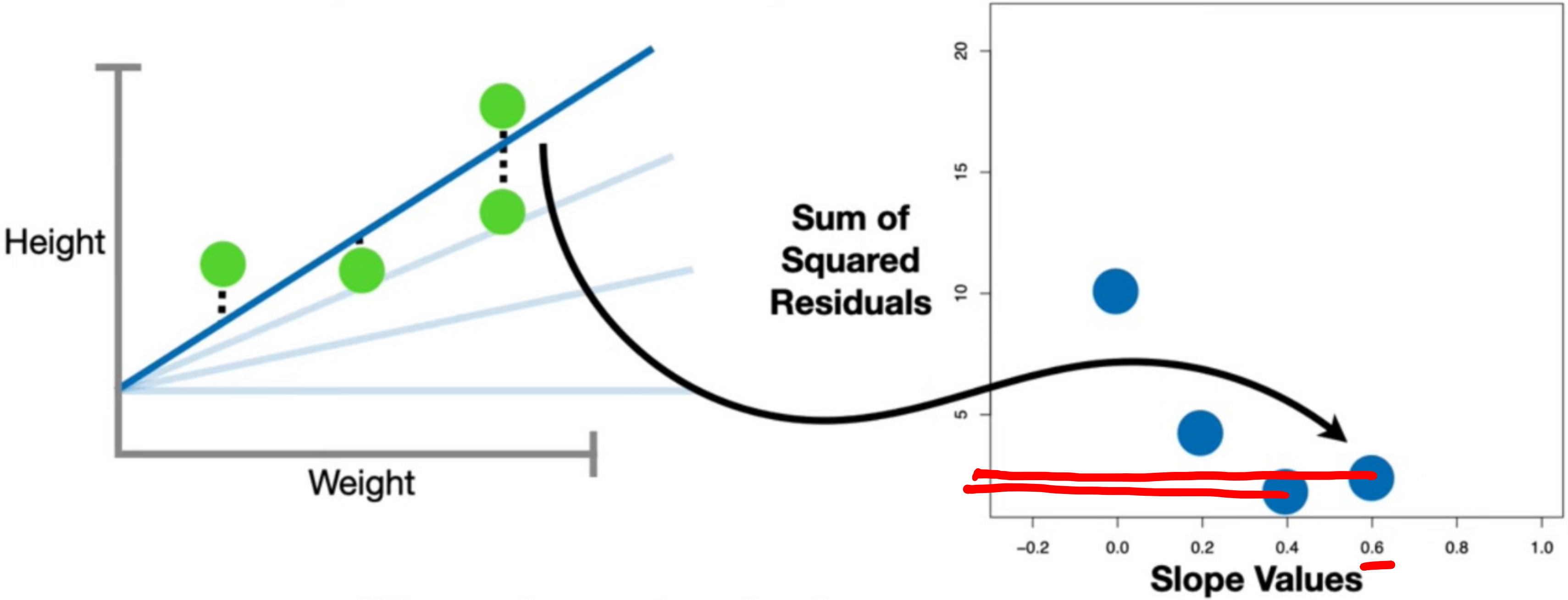




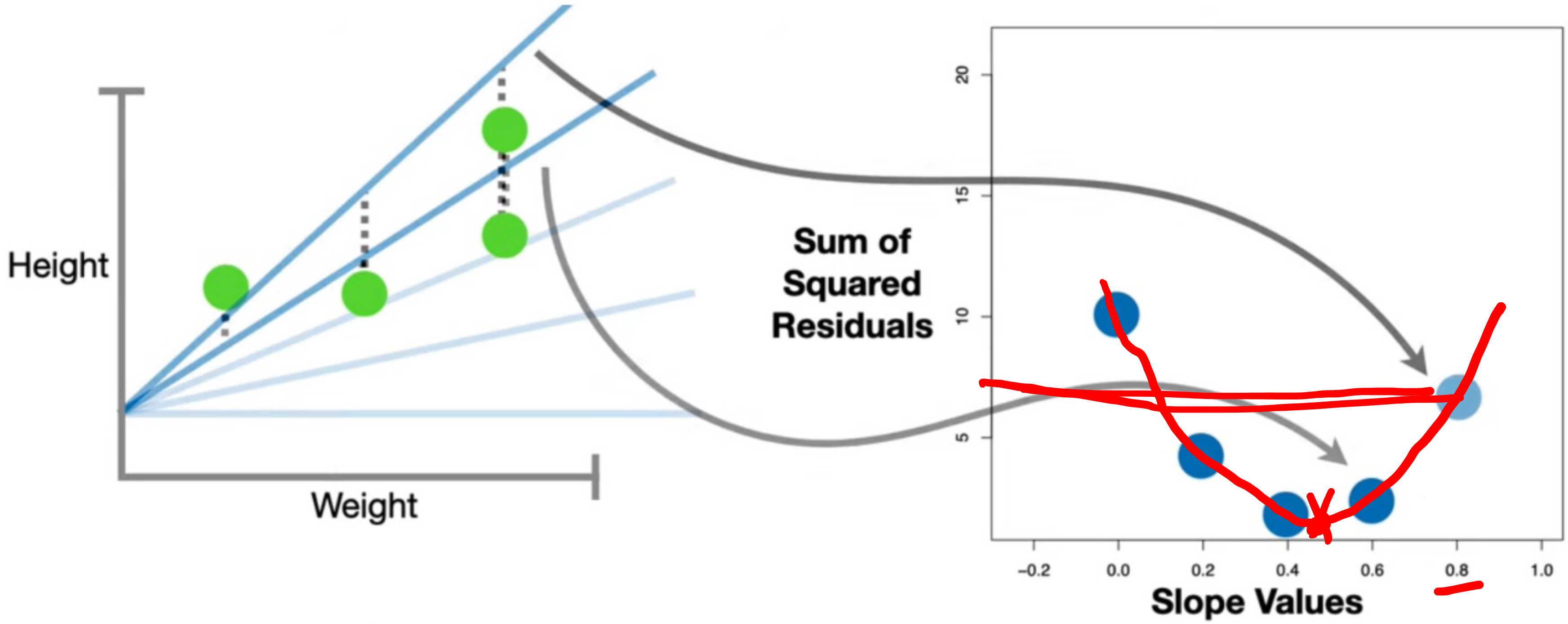
Now let's increase the slope to **0.4**…



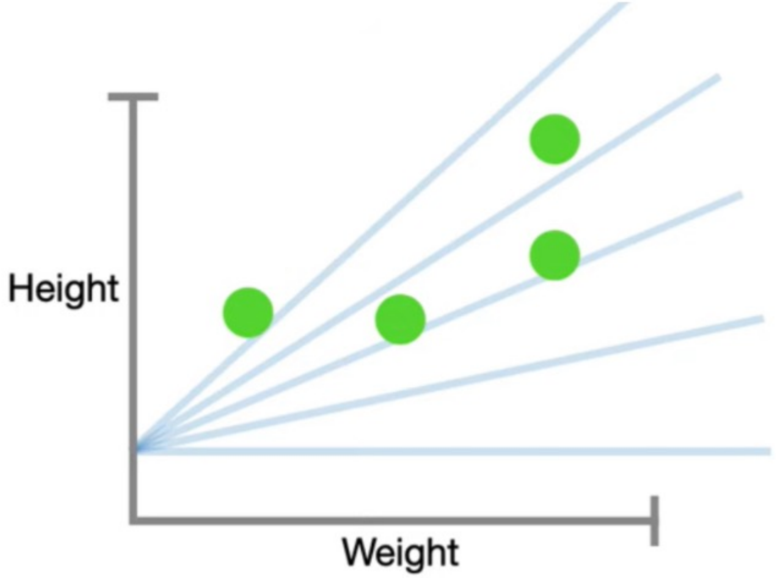
...and calculate a new value for **Sum of the Squared Residuals.**

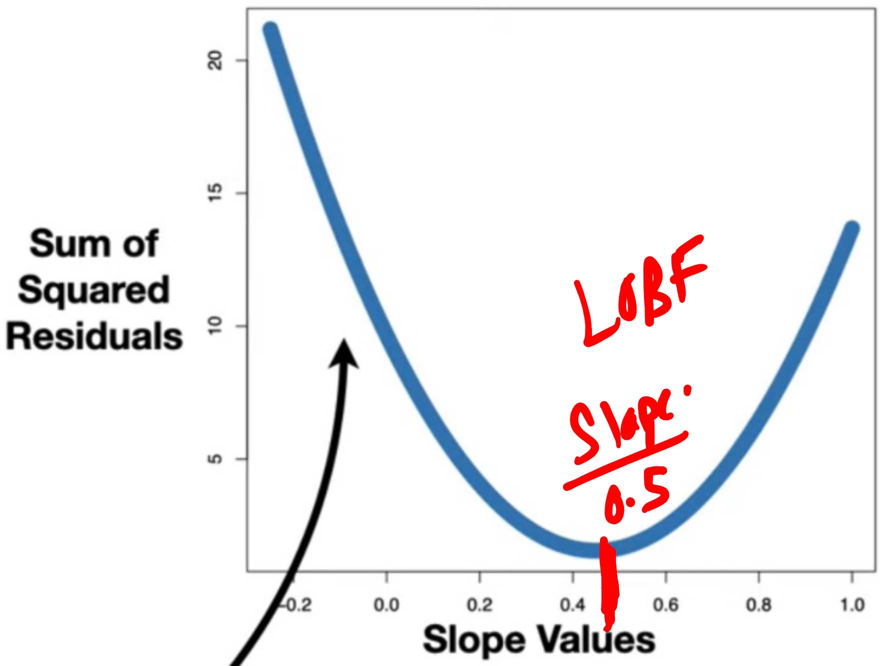


We can keep plugging in new values for the slope and plotting the **Sum of the Squared Residuals**…



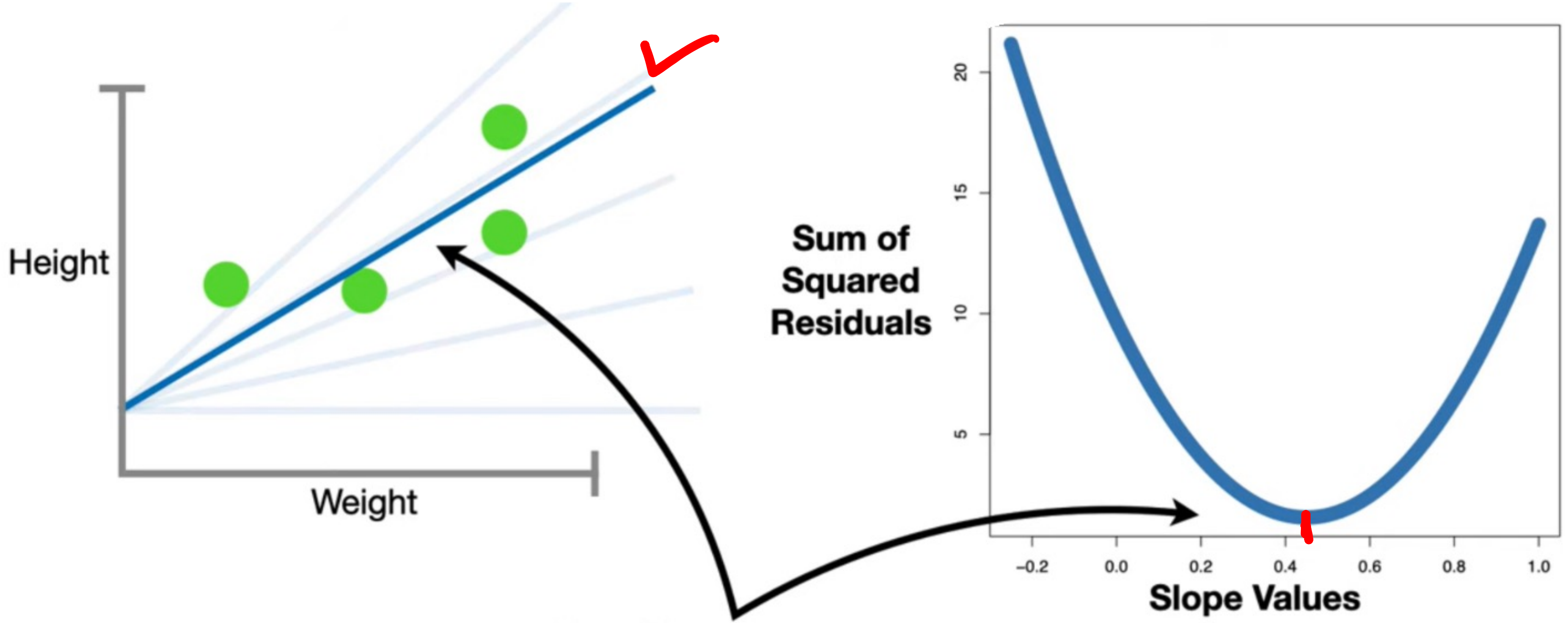
We can keep plugging in new values for the slope and plotting the **Sum of the Squared Residuals**...



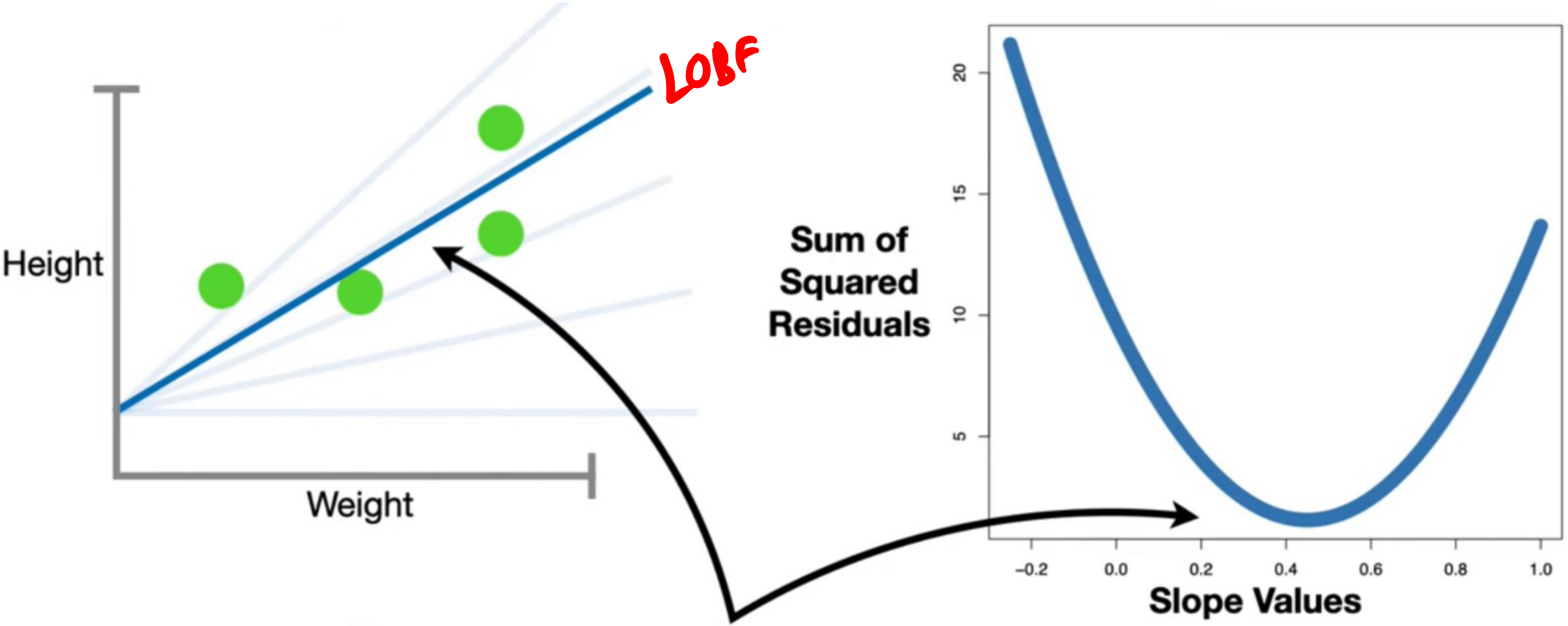


...or we can just plot the curve for this equation.

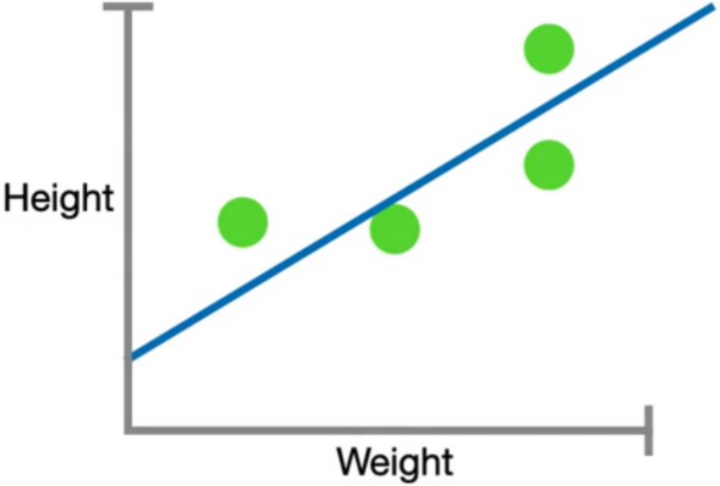
$$\sum (\text{Observed}_i - (\text{intercept} + \text{slope} \times \text{Weight}_i))^2$$



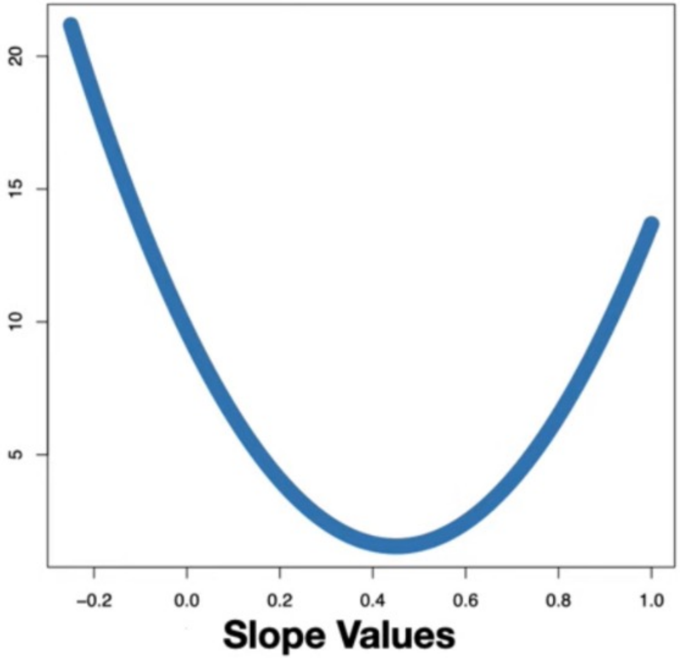
We can see that the best fitting line is at the bottom of the parabola.



In other words, when the slope = **0.45,** we get the lowest **Sum of the Squared Residuals.**



Sum of Squared Residuals + λ × Slope²

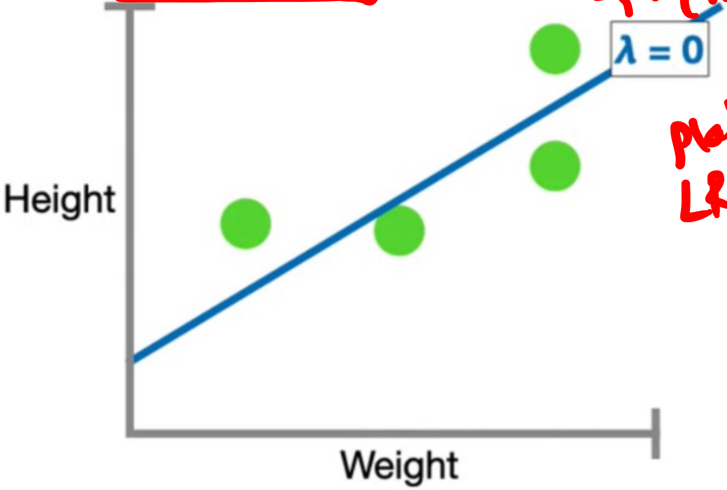


Now let's add the **Ridge Regression Penalty**, aka the **L2 norm**.

$= \text{Loss} + \alpha \cdot (\text{slope})^2$

RR= Loss + Penalty

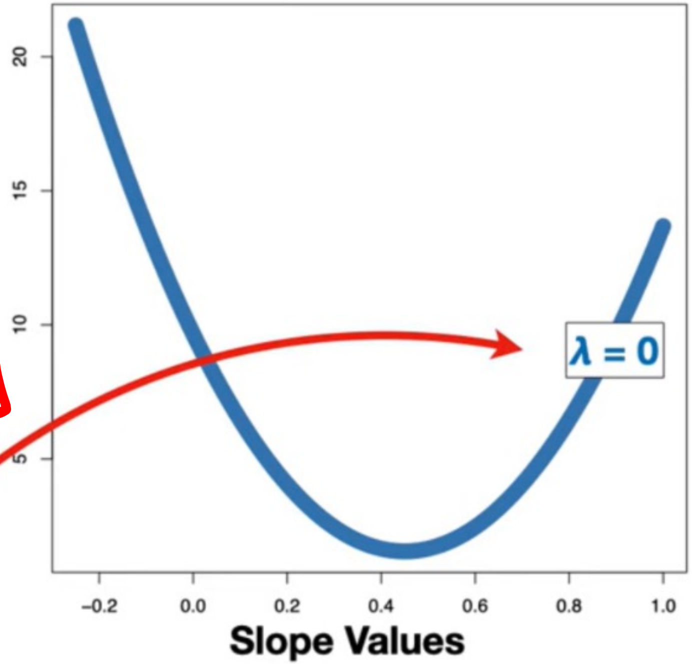
$\alpha \cdot (m)^2$



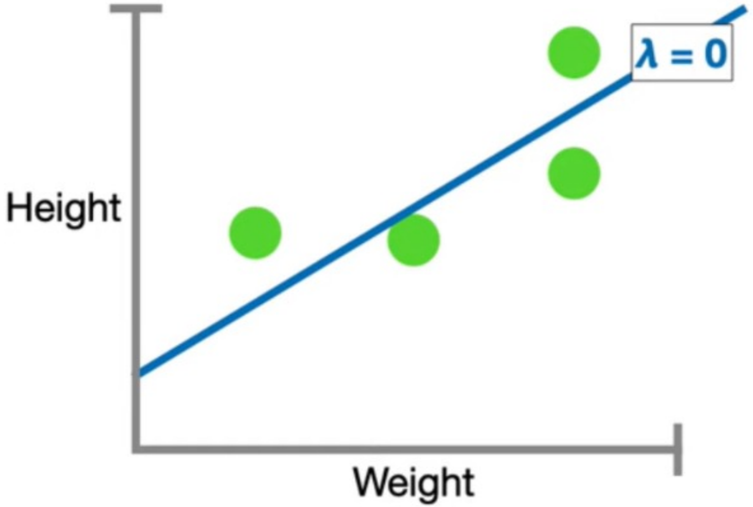
plain LR.

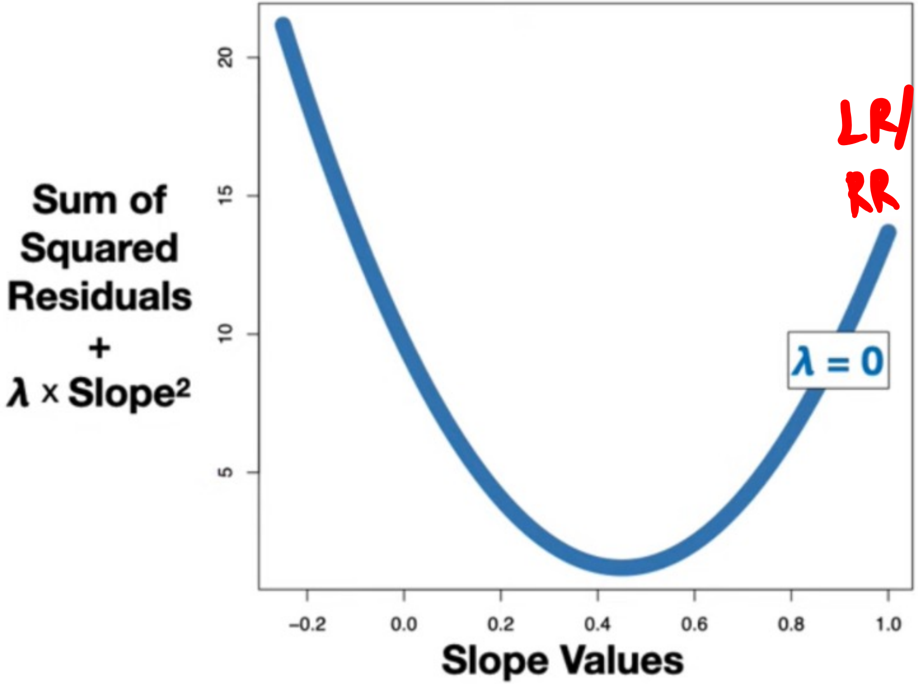
Sum of Squared Residuals
+
$\lambda \times \text{Slope}^2$

0



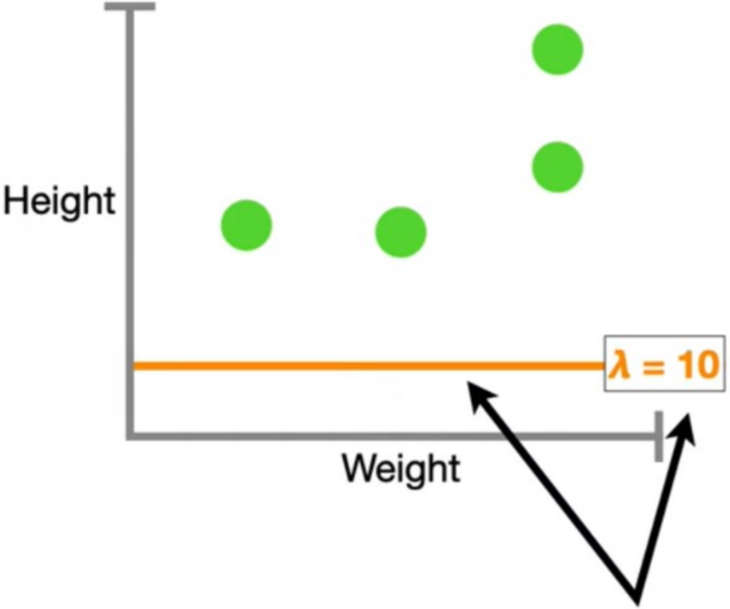
The **thick blue line** that we just drew represents **(lambda) λ = 0.**

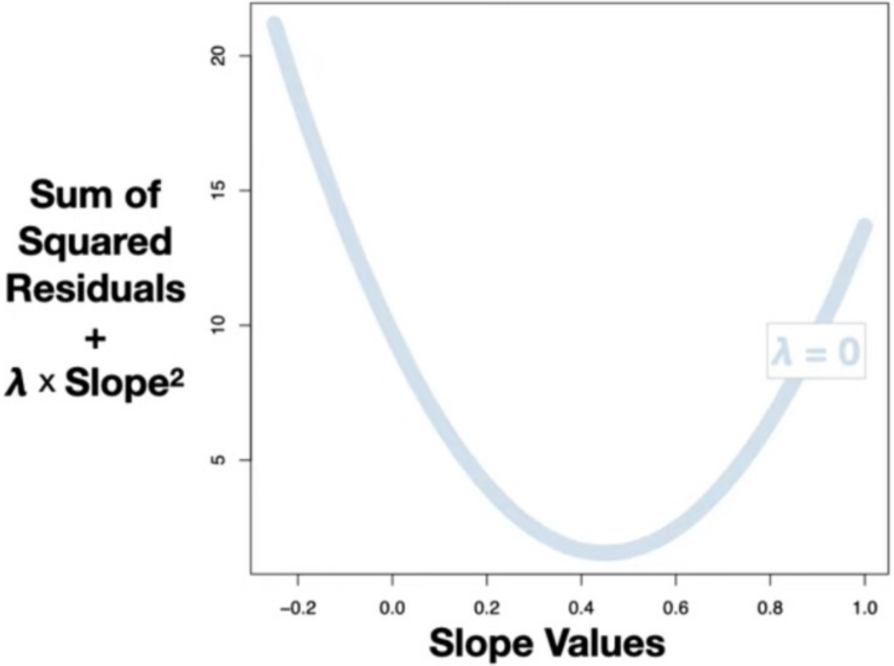




This is because when **(lambda) λ = 0…**

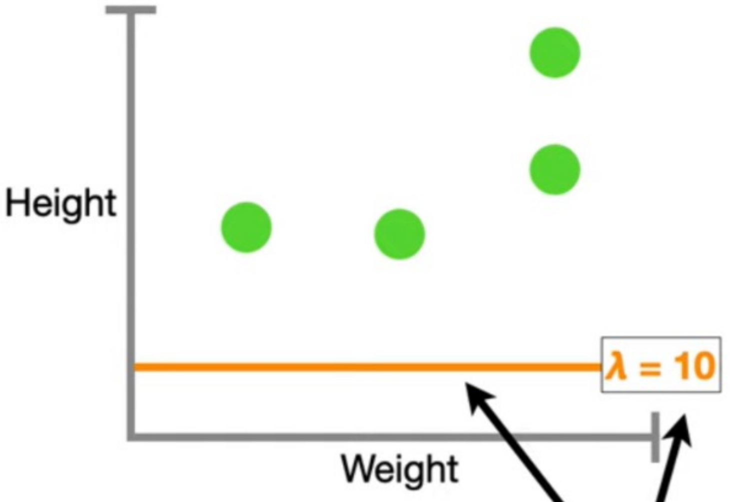
$\left\{ \text{Sum of the Squared Residuals} + (\lambda \times \text{Slope}^2) \right\}$





Now let's see what happens when we set **(lambda) λ = 10.**

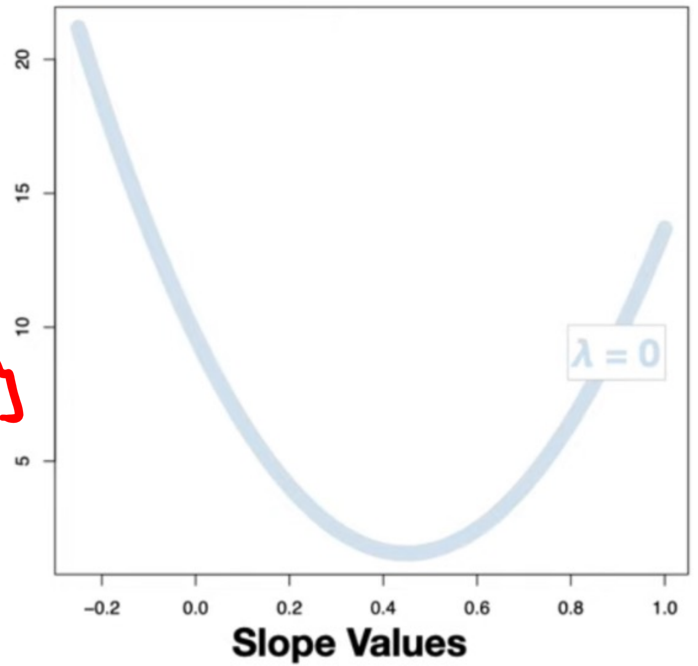
α=0.1
α=1
α=5

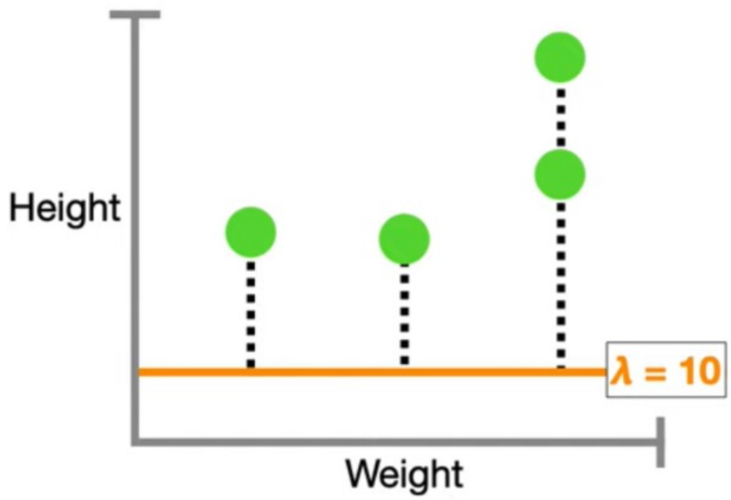


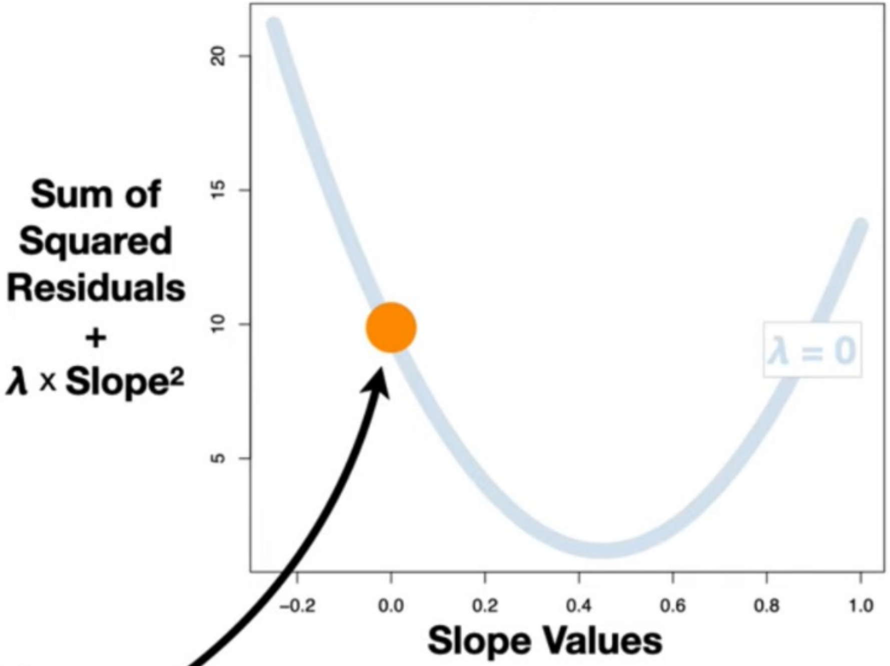
Sum of Squared Residuals + $\lambda \times \text{Slope}^2$

10×0^2

And just like before, we'll start with a horizontal line, only this time the **line is orange**.

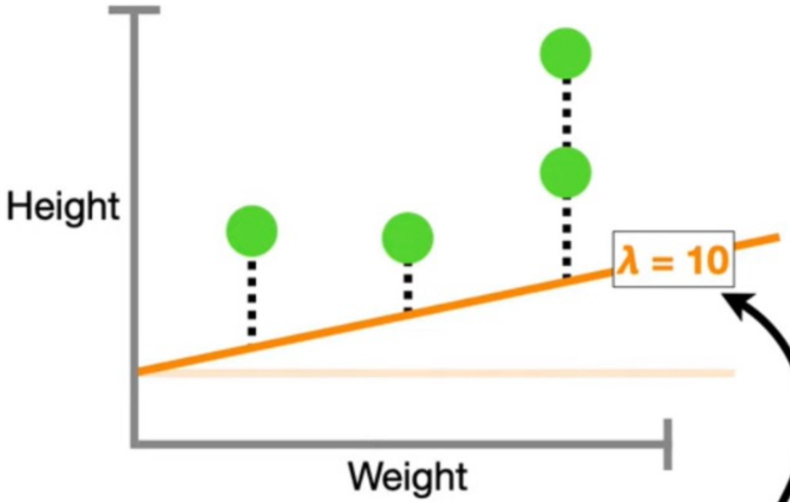




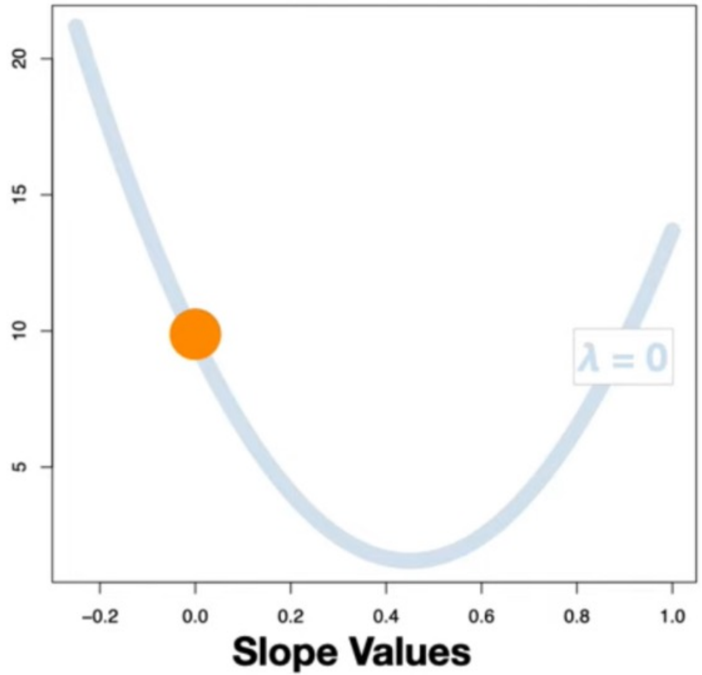


So we plot the **Sum of the Squared Residuals** here.

Sum of the Squared Residuals + 0

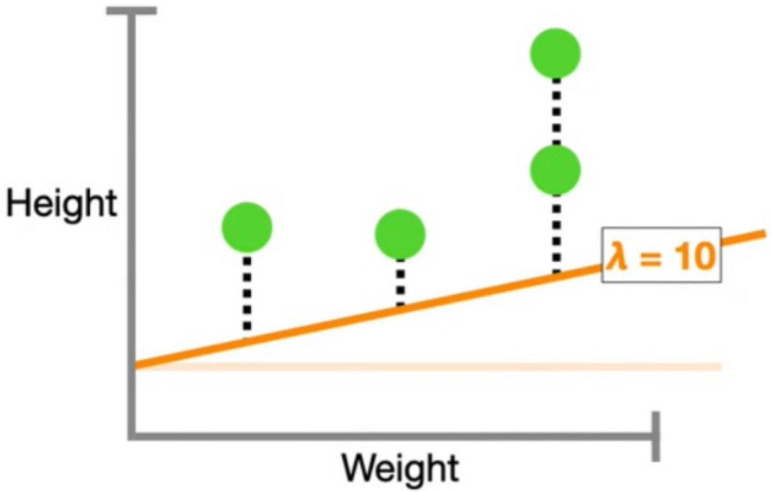


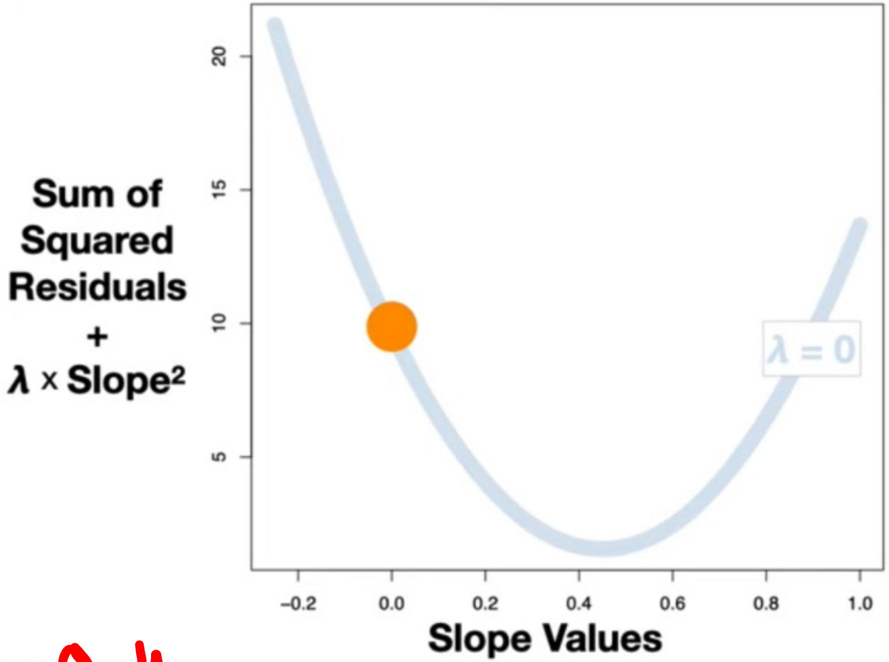
Sum of Squared Residuals + λ × Slope²



Now let's increase the slope to **0.2**.

Sum of the Squared Residuals + (λ × Slope²)





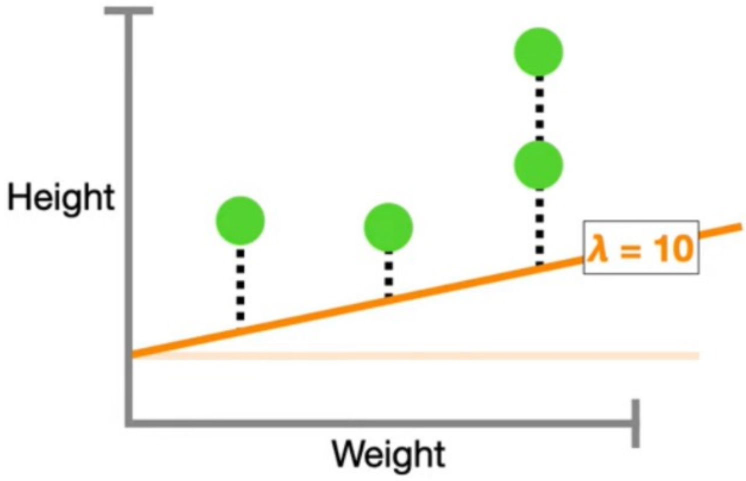
...but now the 0.4 penalty is...

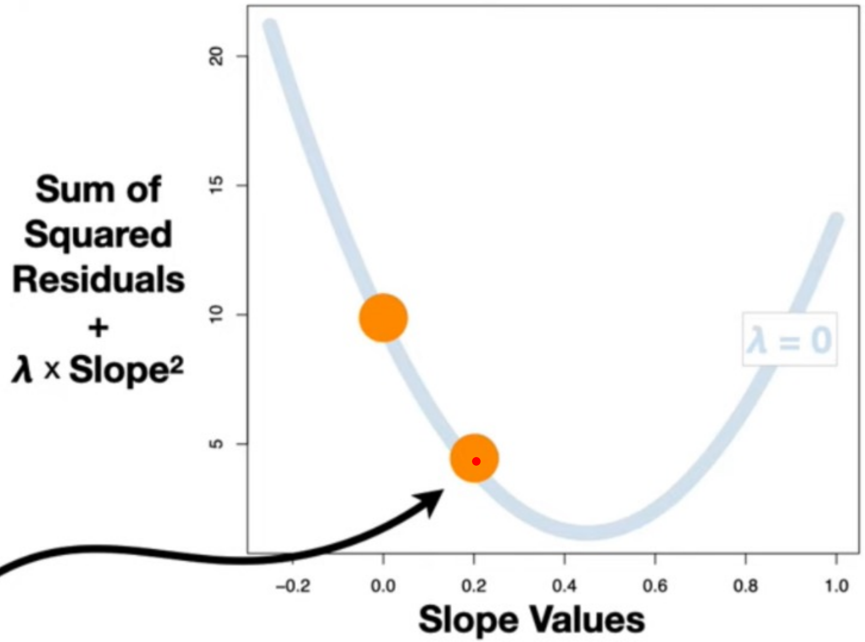
10 × 0.04

10 × $(0.2)^2$

Sum of the Squared Residuals + $\boxed{(\lambda \times \text{Slope}^2)}$

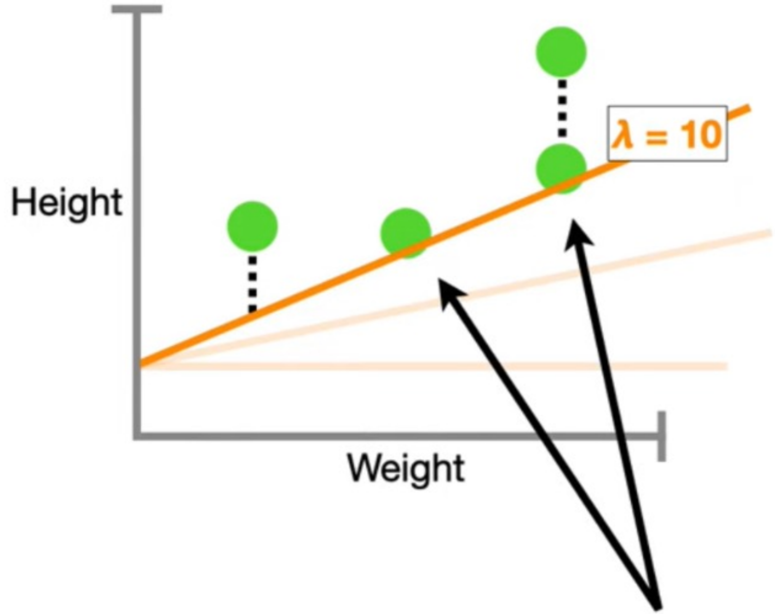
0.4

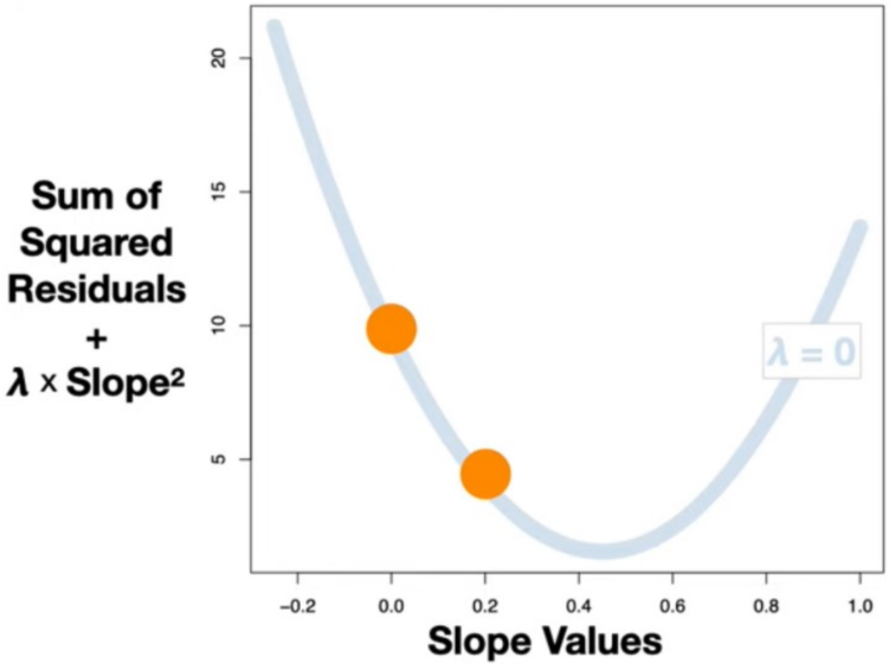




...and that gives us this point on the graph.

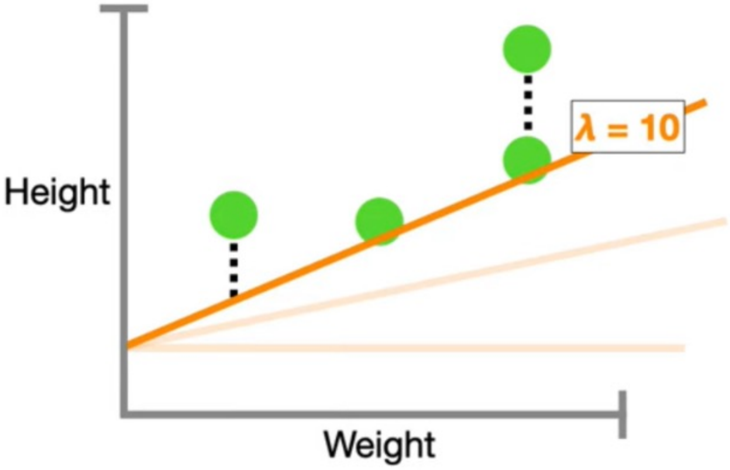
Sum of the Squared Residuals + $\boxed{0.4}$

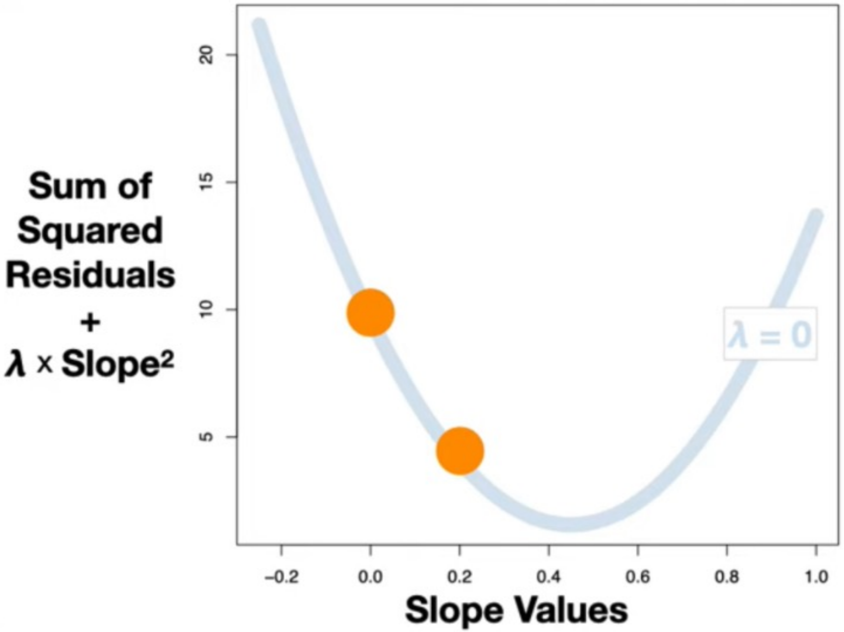




And the **Residuals** are
even smaller…

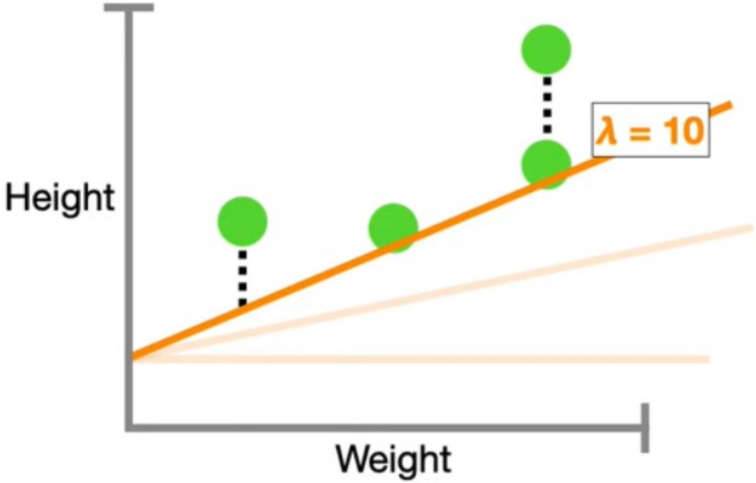
Sum of the Squared Residuals + (λ × Slope^2)

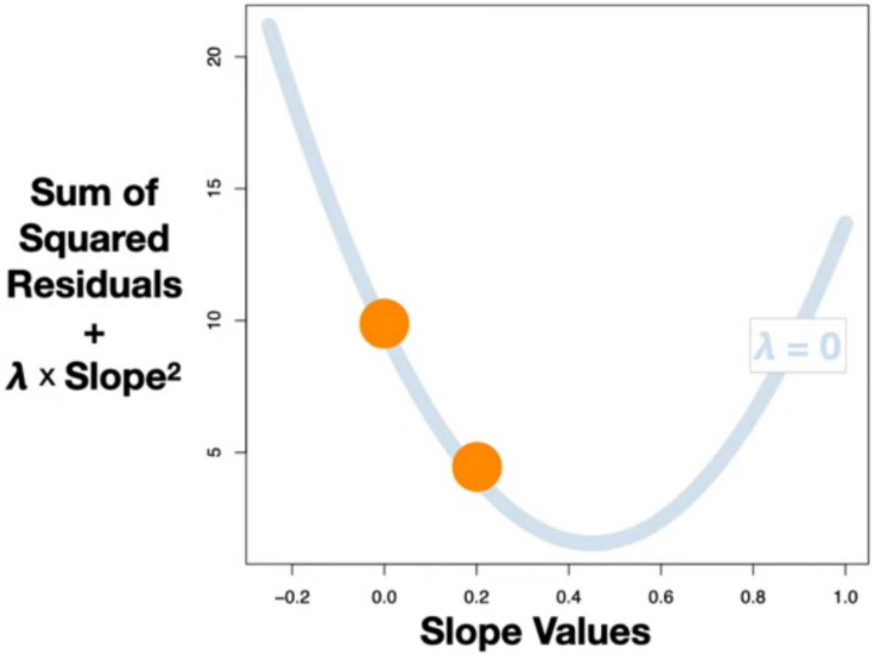




…but now the penalty is…

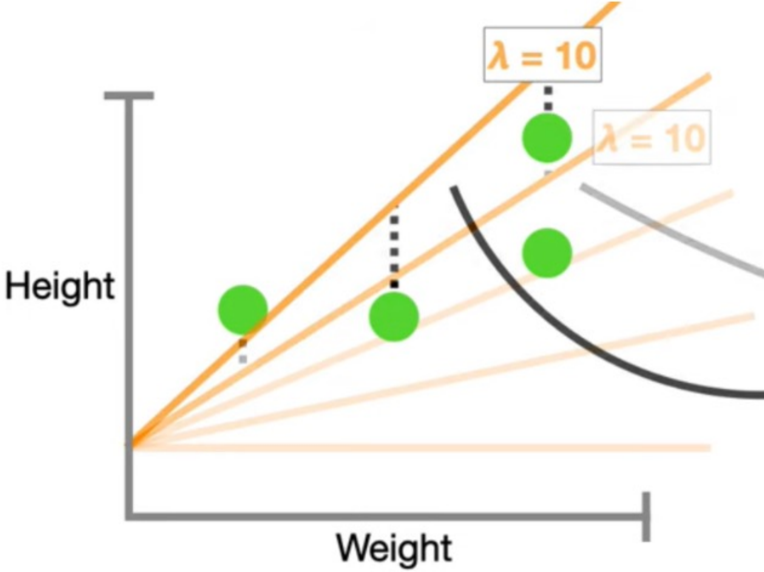
Sum of the Squared Residuals + $\boxed{(\lambda \times \text{Slope}^2)}$

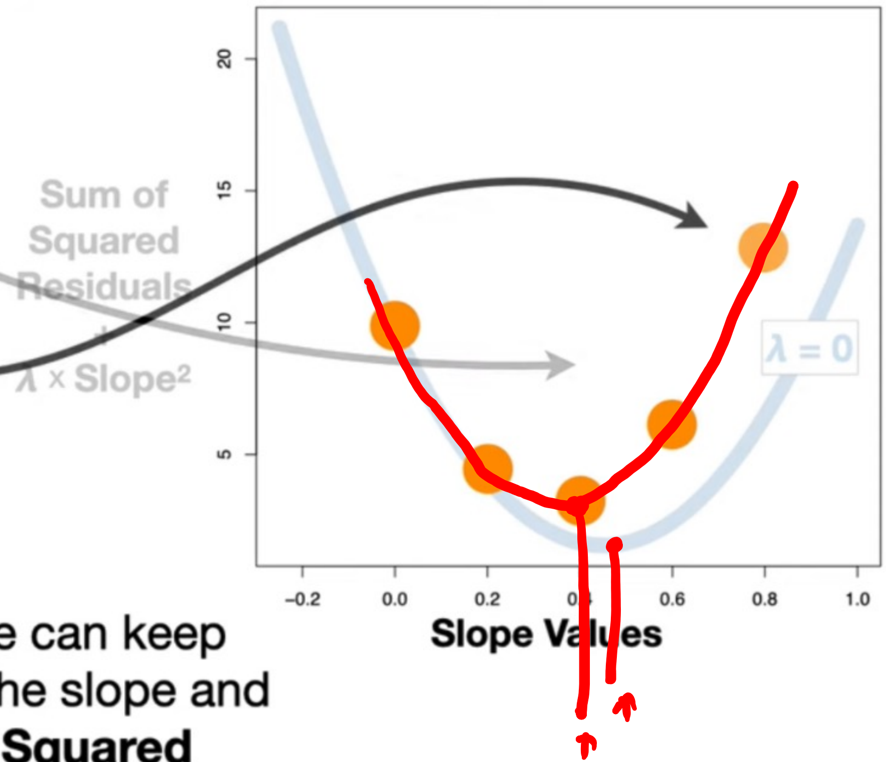




...but now the penalty is...

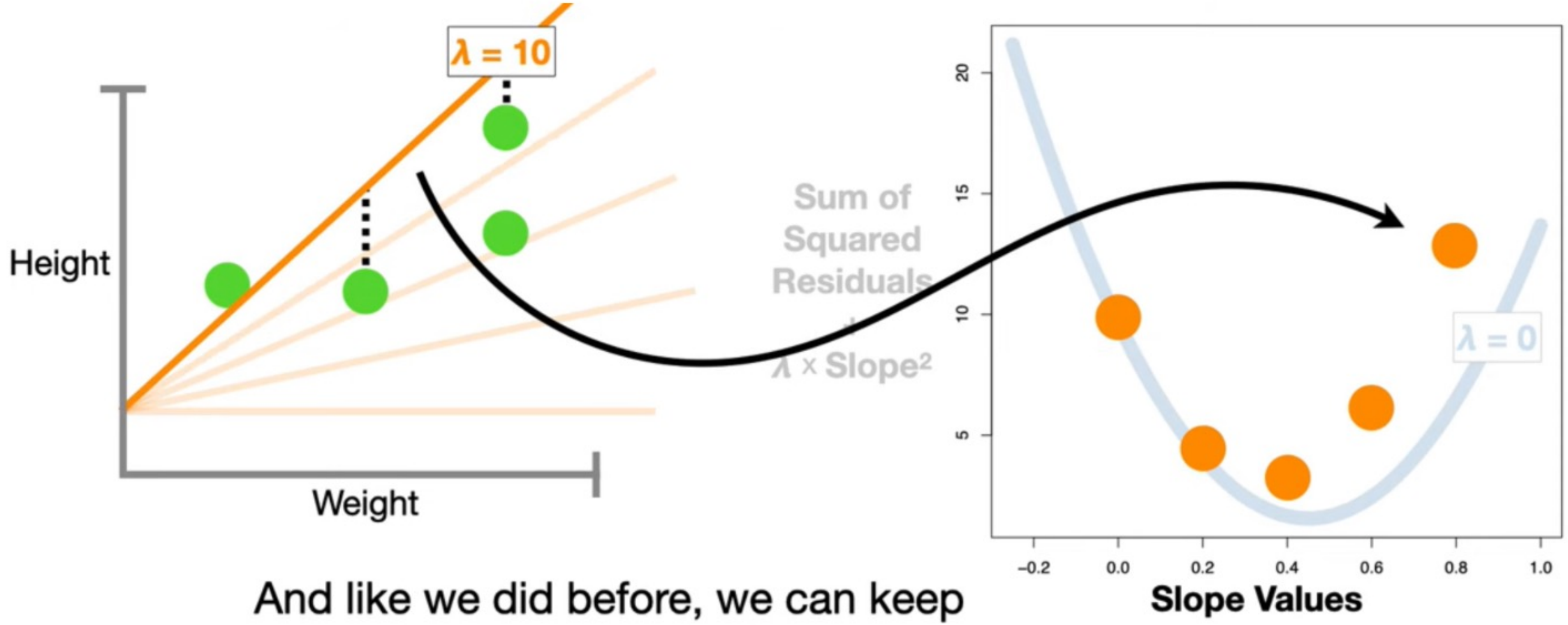
Sum of the Squared Residuals + $\boxed{1.6}$





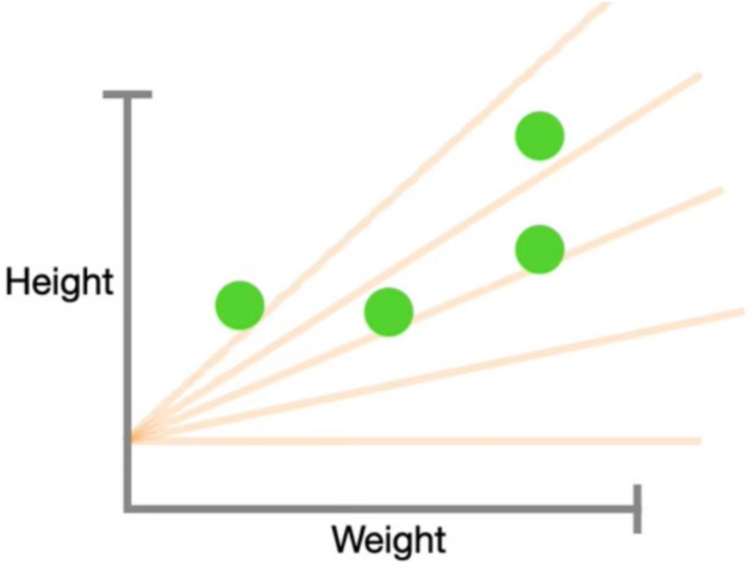
And like we did before, we can keep plugging in new values for the slope and plotting the **Sum of the Squared Residuals plus the Penalty**…

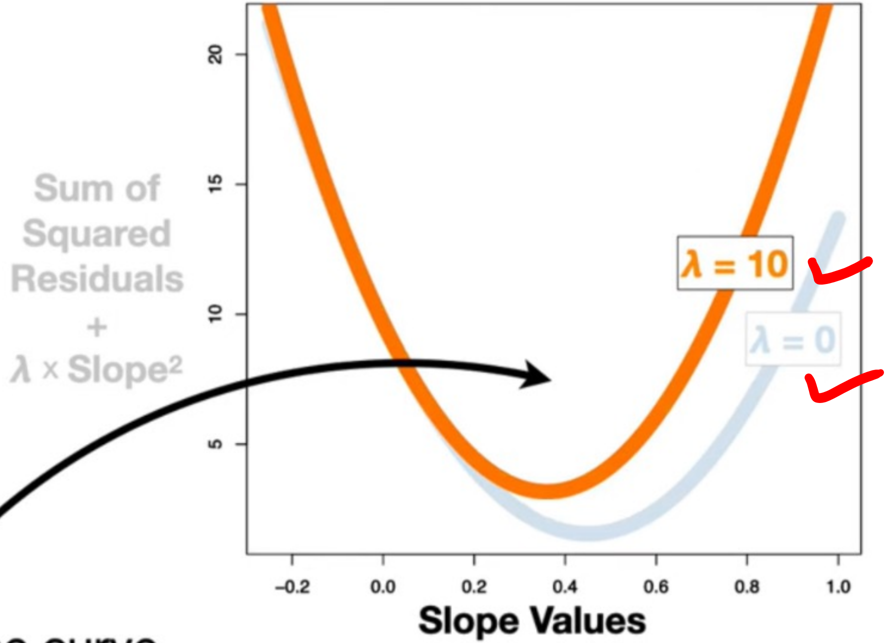
Sum of the Squared Residuals + ($\lambda \times \text{Slope}^2$)



And like we did before, we can keep plugging in new values for the slope and plotting the **Sum of the Squared Residuals plus the Penalty**…

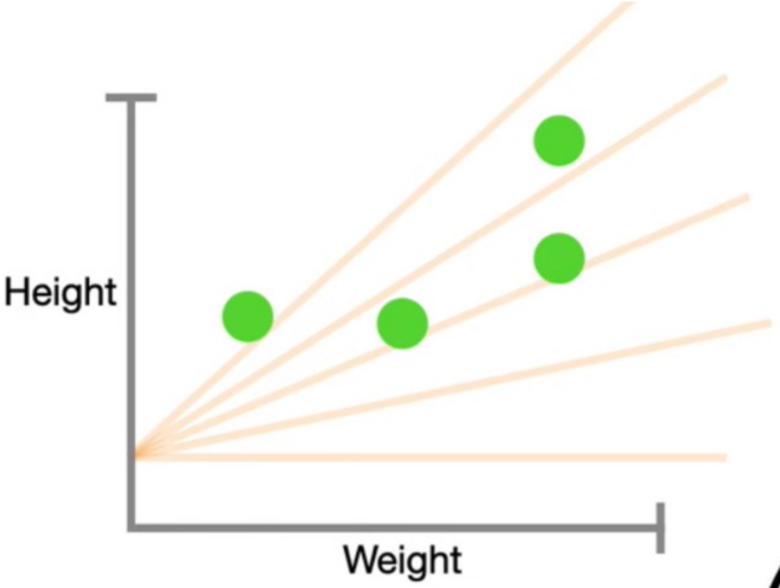
Sum of the Squared Residuals + ($\lambda \times \text{Slope}^2$)

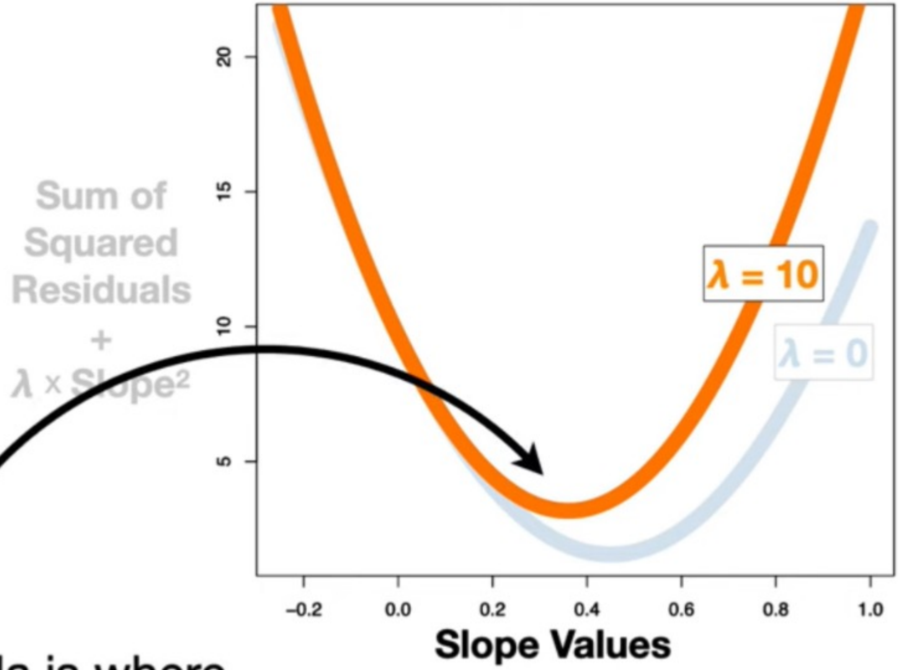




…or we can just plot the curve with a **thick orange line** that represents **(lambda) λ = 10**

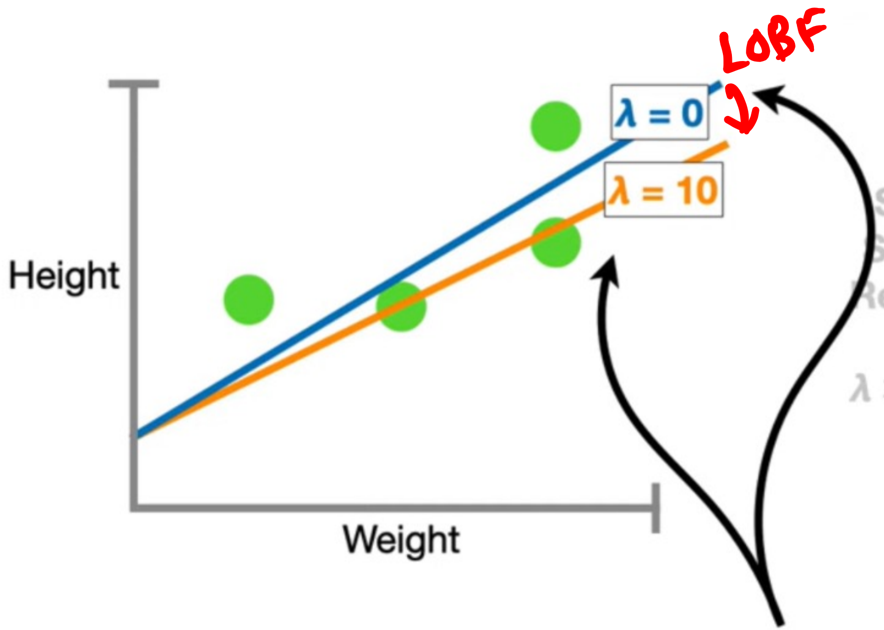
Sum of the Squared Residuals + (λ × Slope^2)

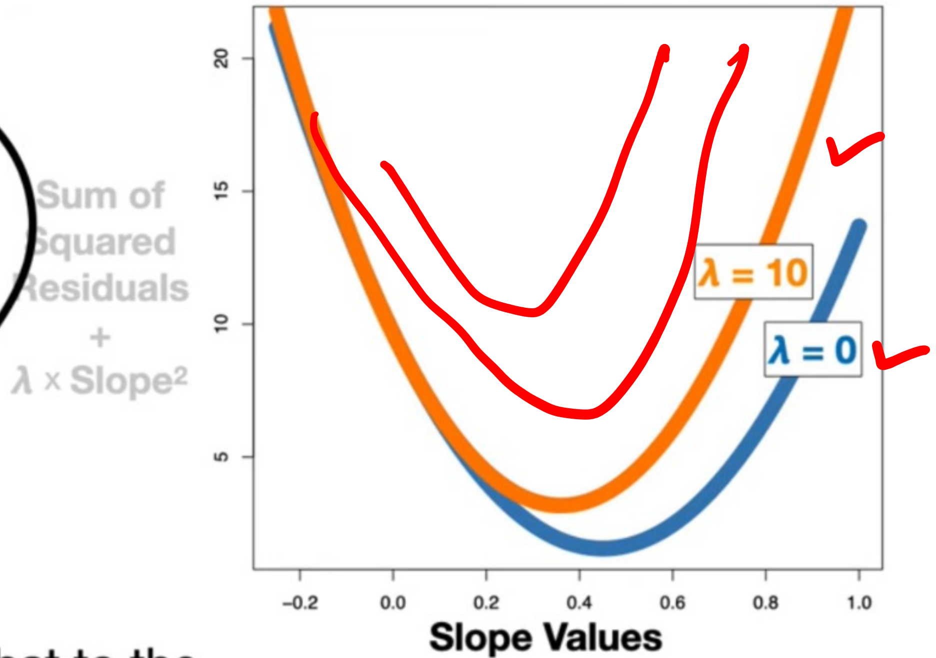




The bottom of the parabola is where the slope gives us the lowest **Sum of Squared Residuals plus Penalty**…

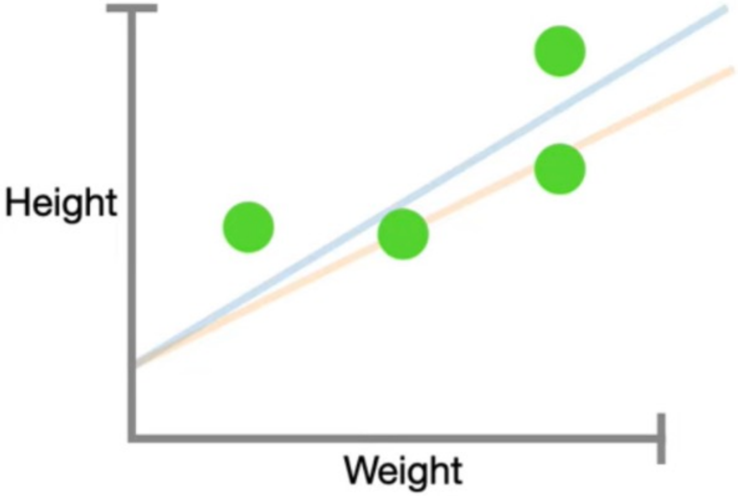
Sum of the Squared Residuals + (λ × Slope^2)

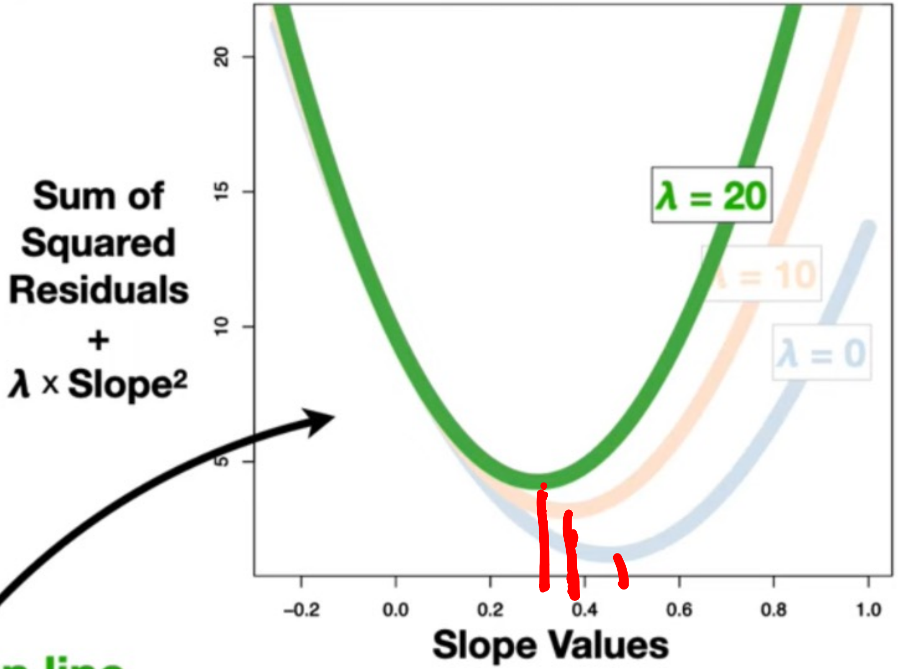




And when we compare that to the optimal slope when **(lambda) $\lambda = 0$**, we see that setting **$\lambda = 10$** results in a smaller optimal slope.

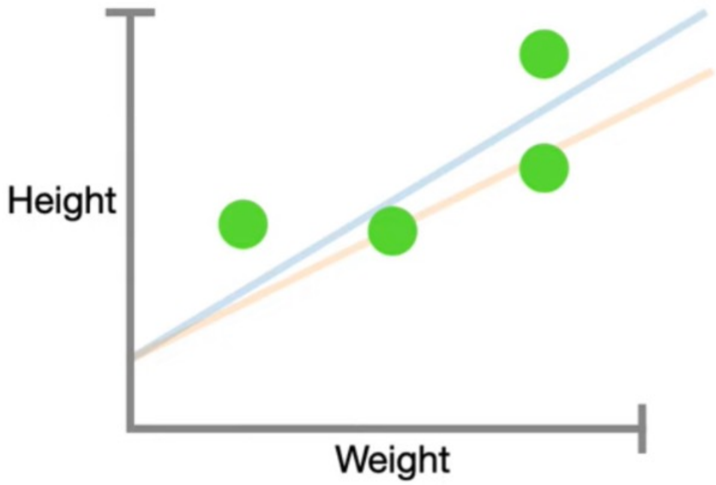
Sum of the Squared Residuals + (λ × Slope^2)

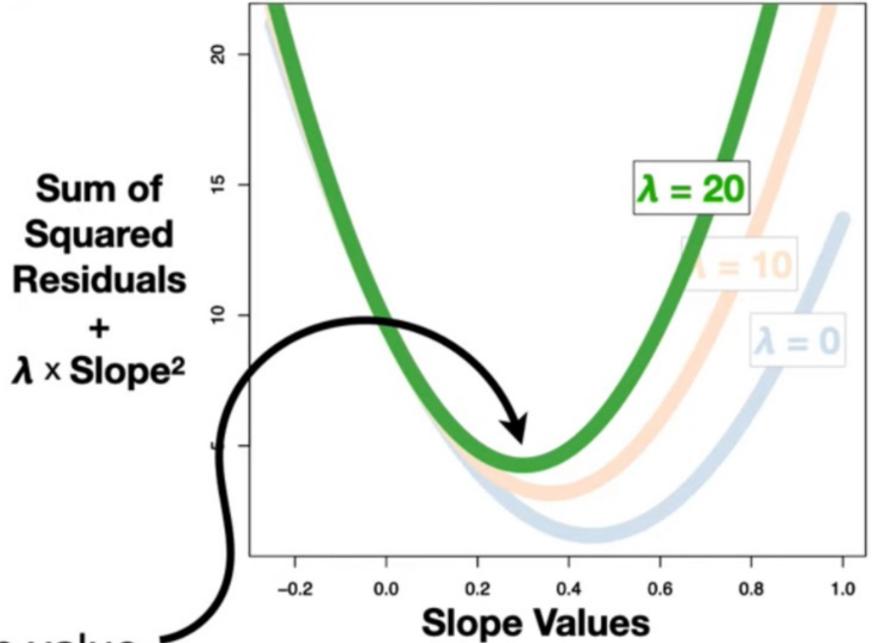




Likewise, the **thick green line** represents **(lambda) λ = 20**.

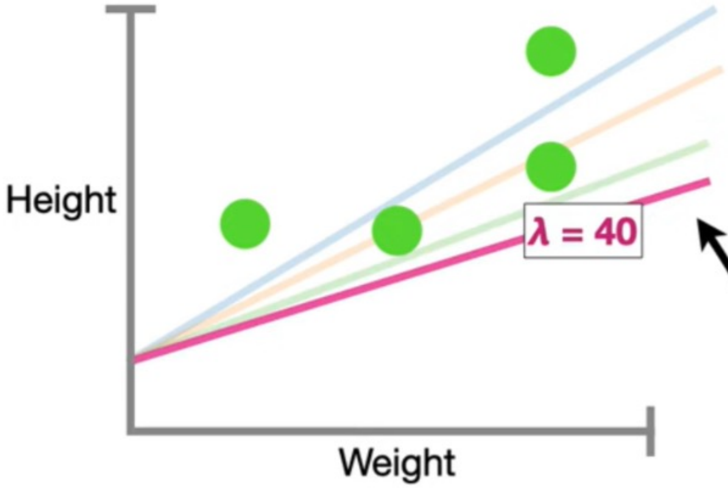
Sum of the Squared Residuals + (λ × Slope²)



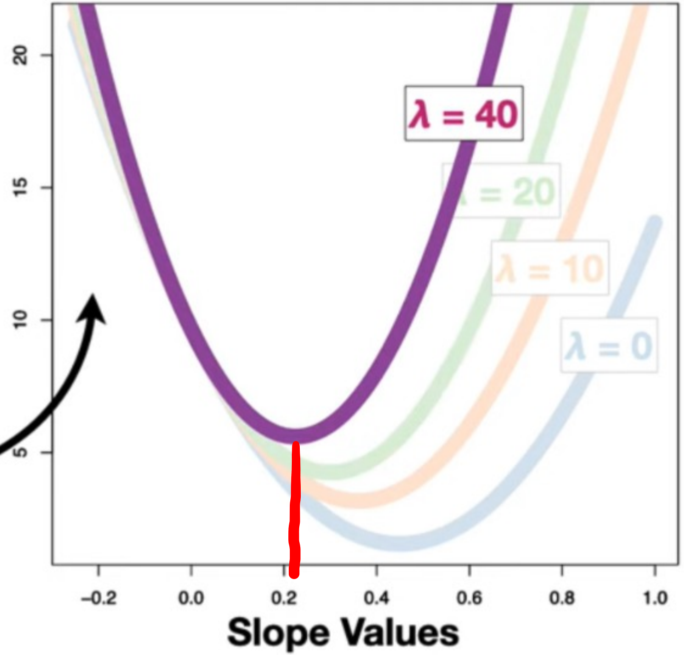


We see that the minimum value
is closer to zero…

Sum of the Squared Residuals + (λ × Slope^2)



Sum of Squared Residuals + $\lambda \times \text{Slope}^2$

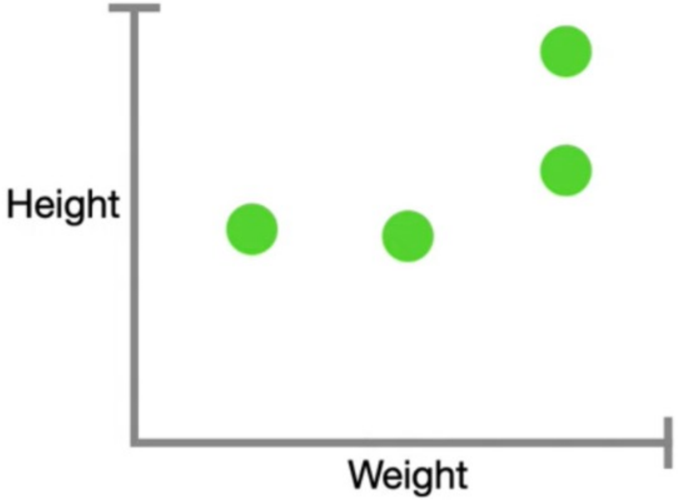


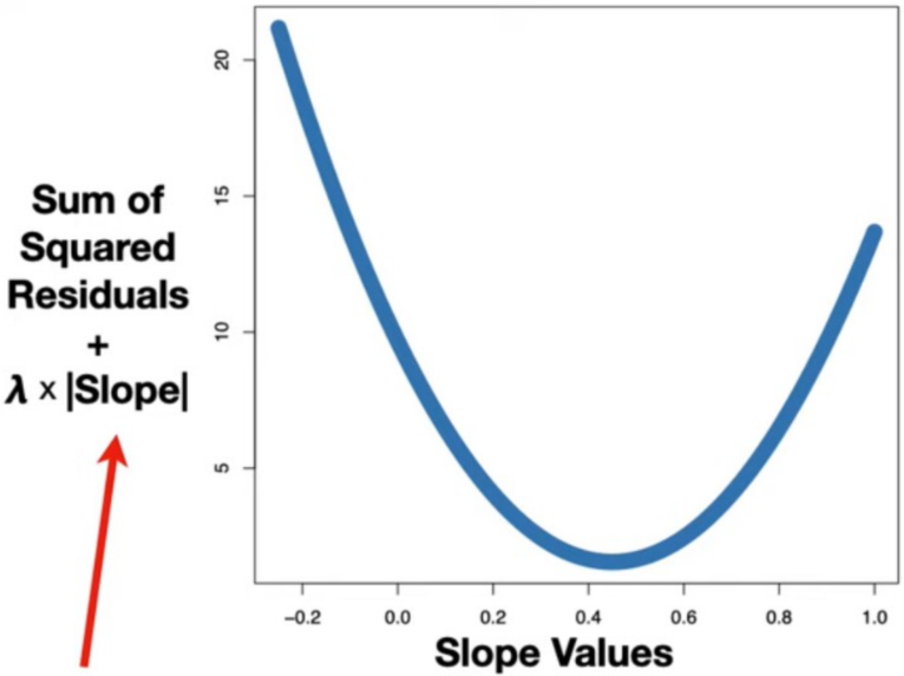
The **purple lines** represent **(lambda) λ = 40**, and it shrinks the slope even more.

$y = 15 + 1.9x_1 + \ldots$

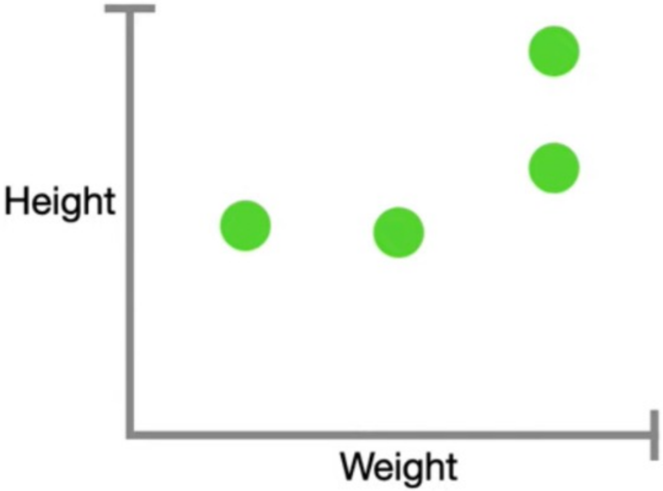
reduce

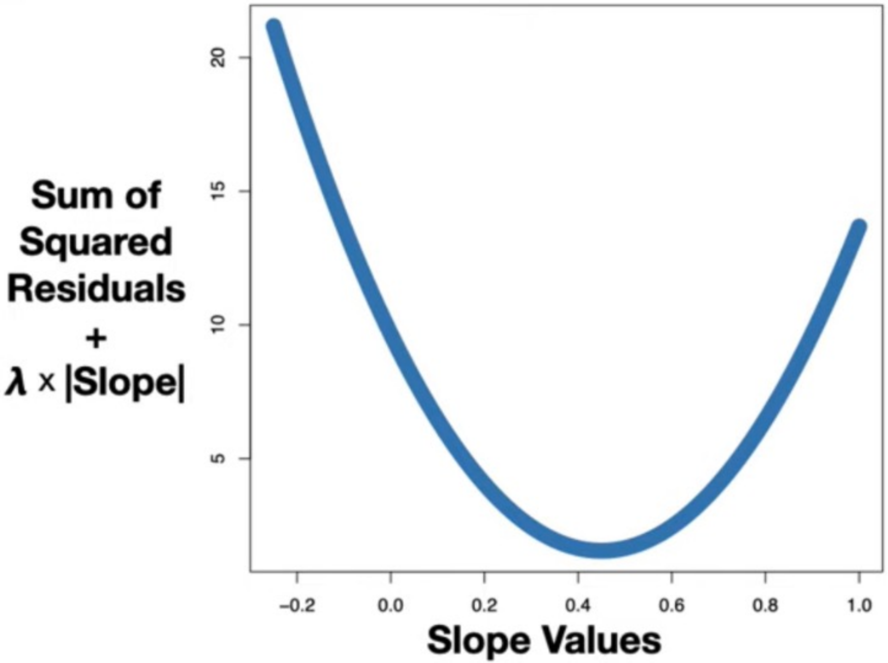
Sum of the Squared Residuals + ($\lambda \times \text{Slope}^2$)





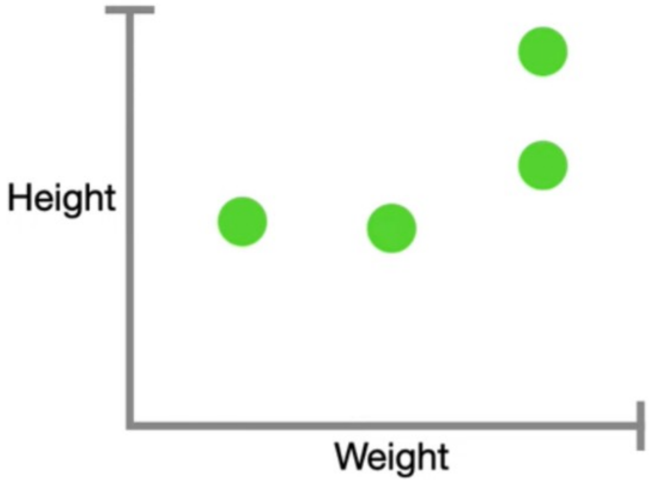
Now let's see what happens if we use the **Lasso Penalty**, aka, the **L1 norm**, or, if you asked me, I'd call it the **Absolute Value Penalty**.

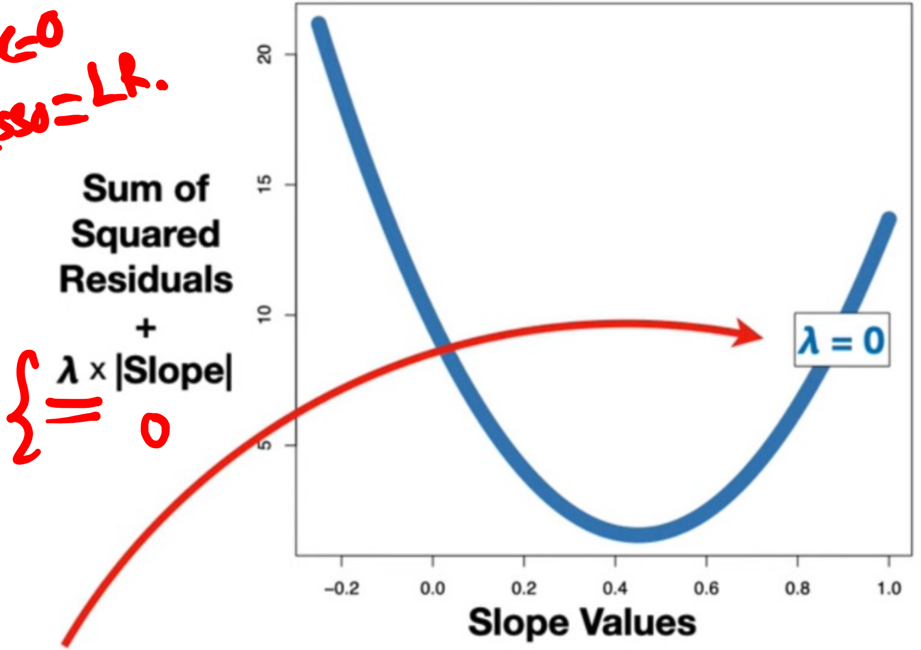




Unfortunately, no one asked me.

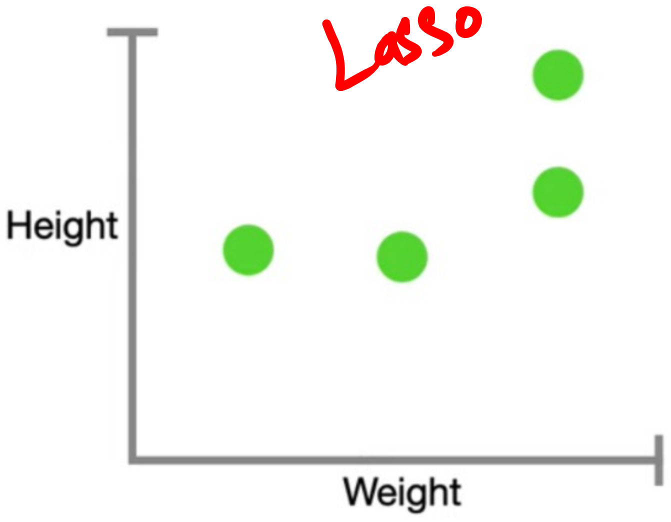
Slope=0
Lasso=LR.

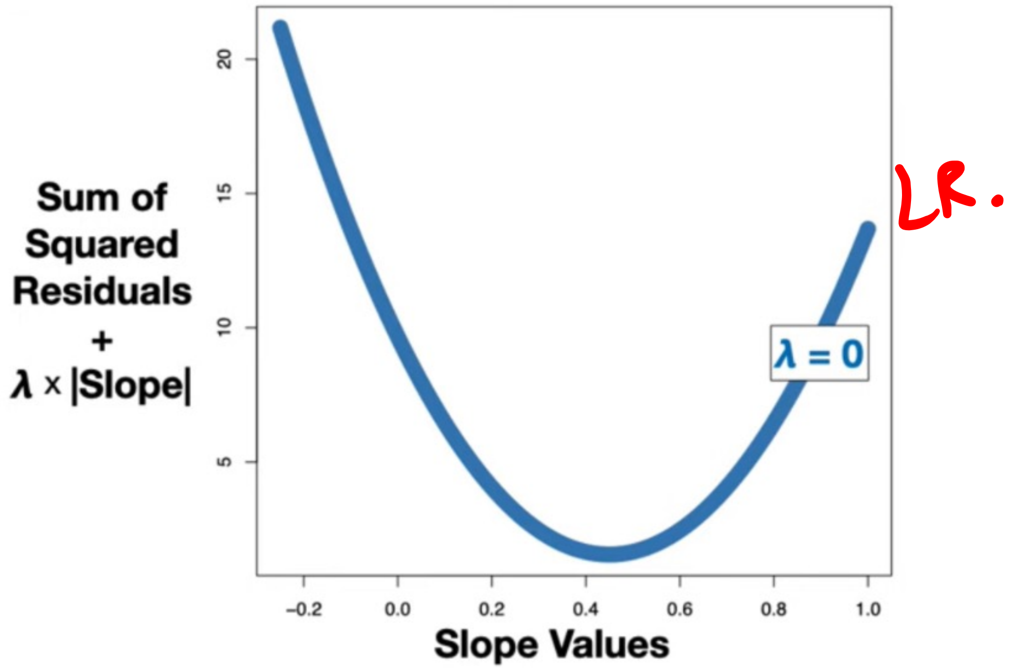




Again, the **thick blue line** represents **(lambda) λ = 0**, so there is no extra penalty.

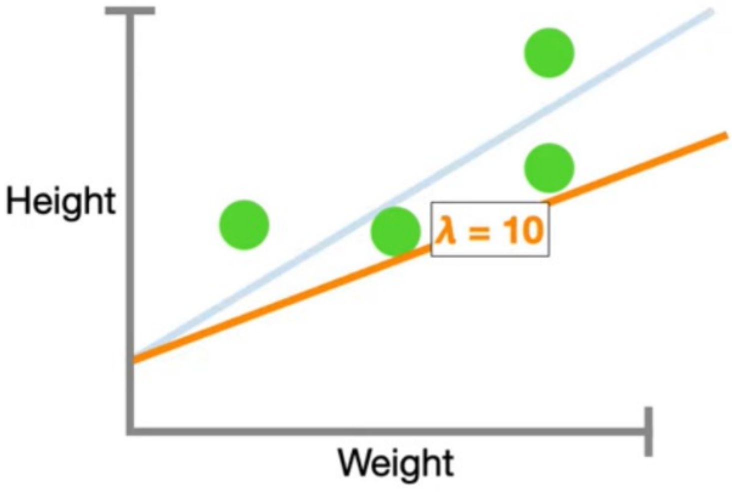
Sum of the Squared Residuals + (λ × |Slope|)

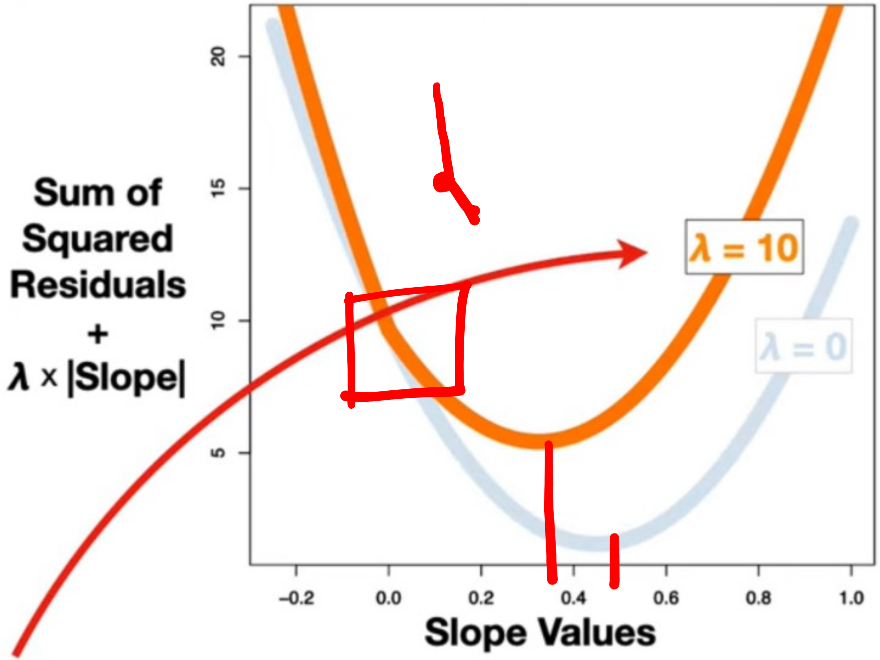




...and we are left with the original
Sum of the Squared Residuals.

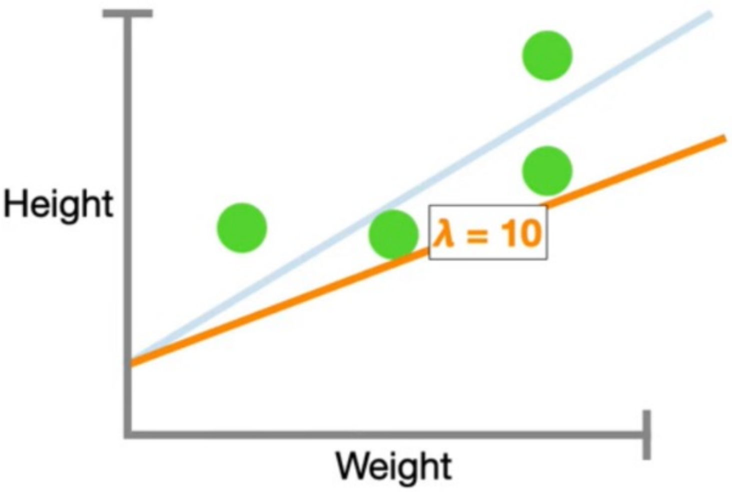
$\boxed{\text{Sum of the Squared Residuals}}$ + 0 =

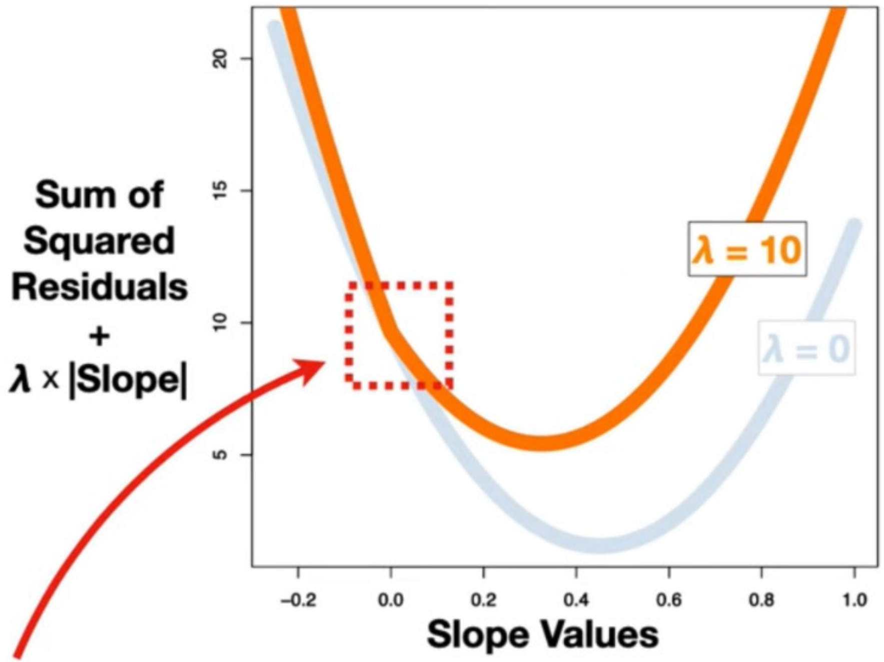




The **thick orange line** represents **(lambda) λ = 10**, so now we are turning on the penalty and shrinking the slope.

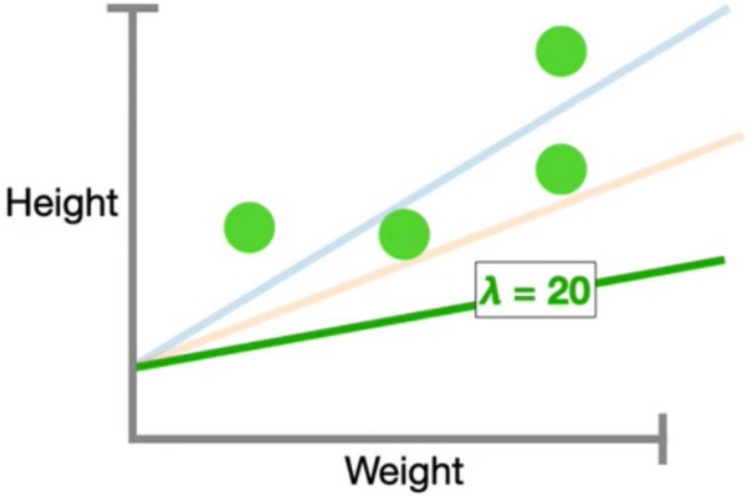
Sum of the Squared Residuals + (λ × |Slope|)

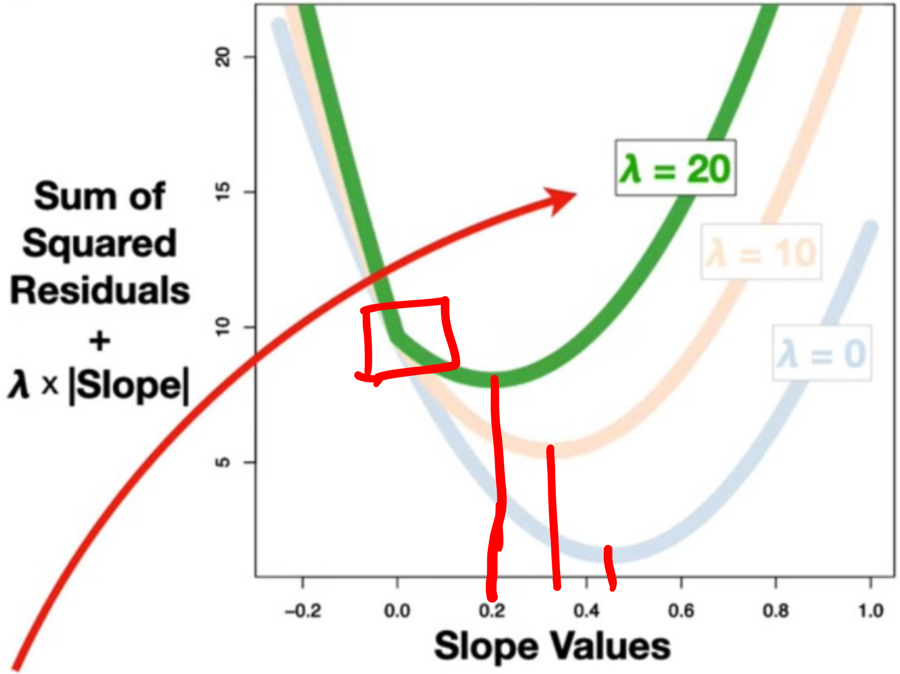




NOTE: When **(lambda) λ = 10** we start to see a kink in the curve where the slope is **0**.

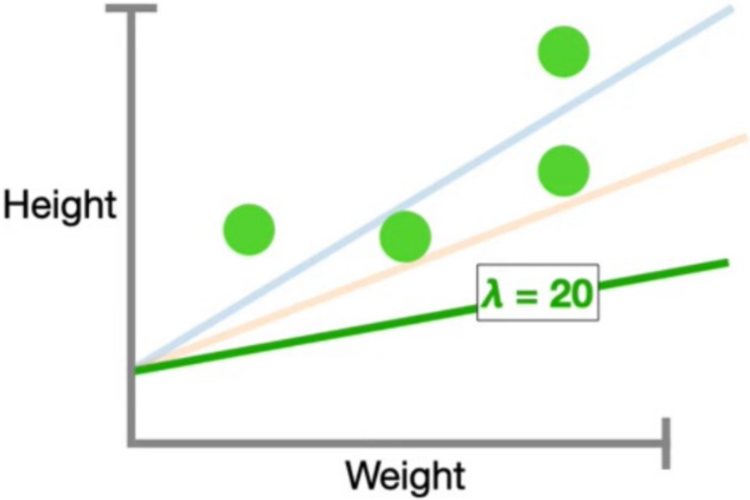
Sum of the Squared Residuals + ($\lambda \times$ |Slope|)

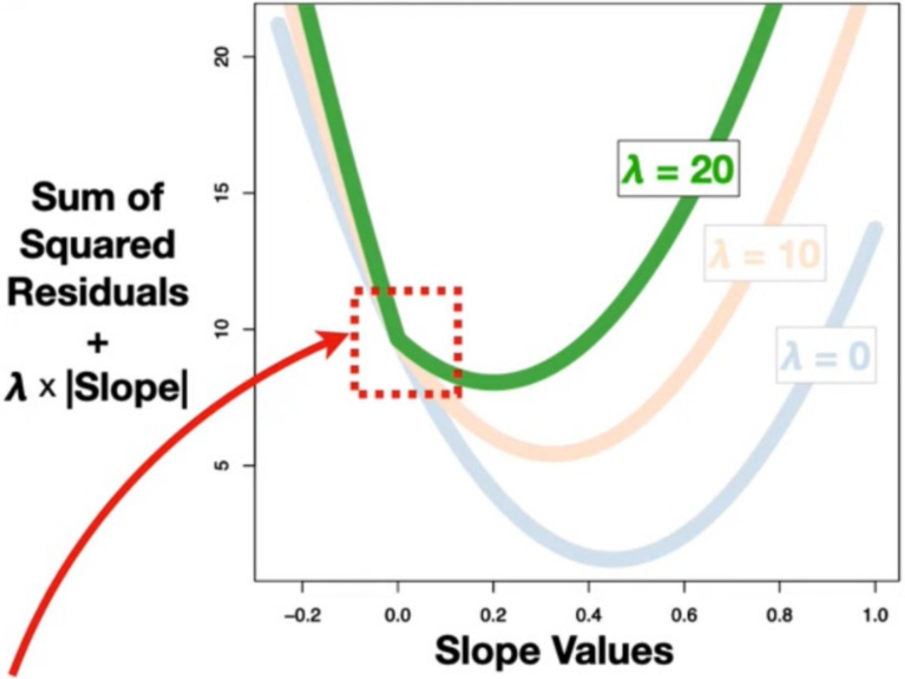




The **thick green line** represents
(lambda) λ = 20…

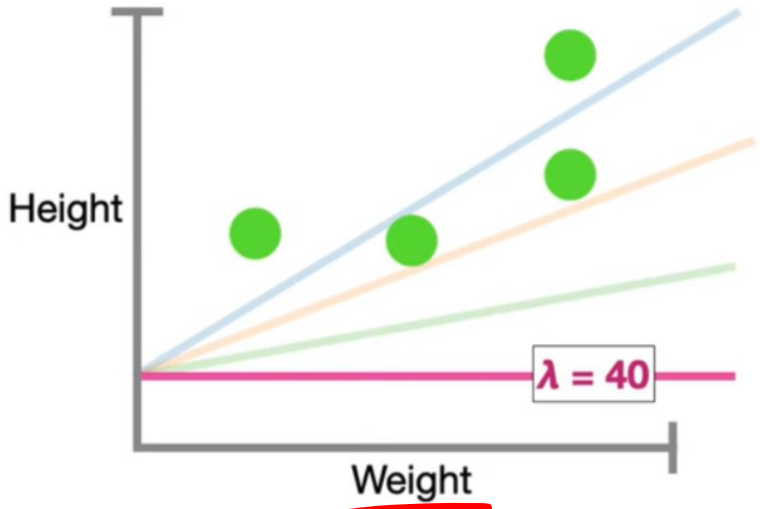
Sum of the Squared Residuals + (λ × |Slope|)

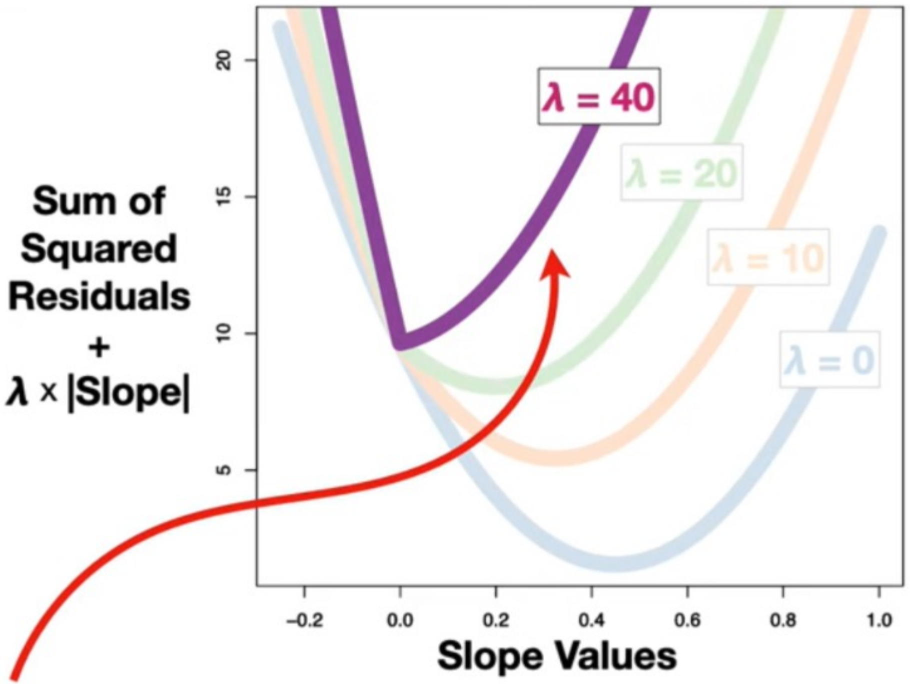




…and this kink at **0** is becoming more **prominent**.

Sum of the Squared Residuals + (**λ** × |Slope|)



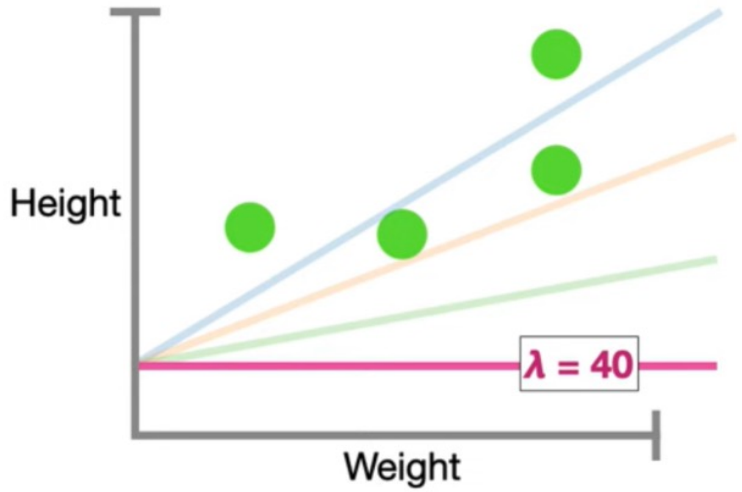


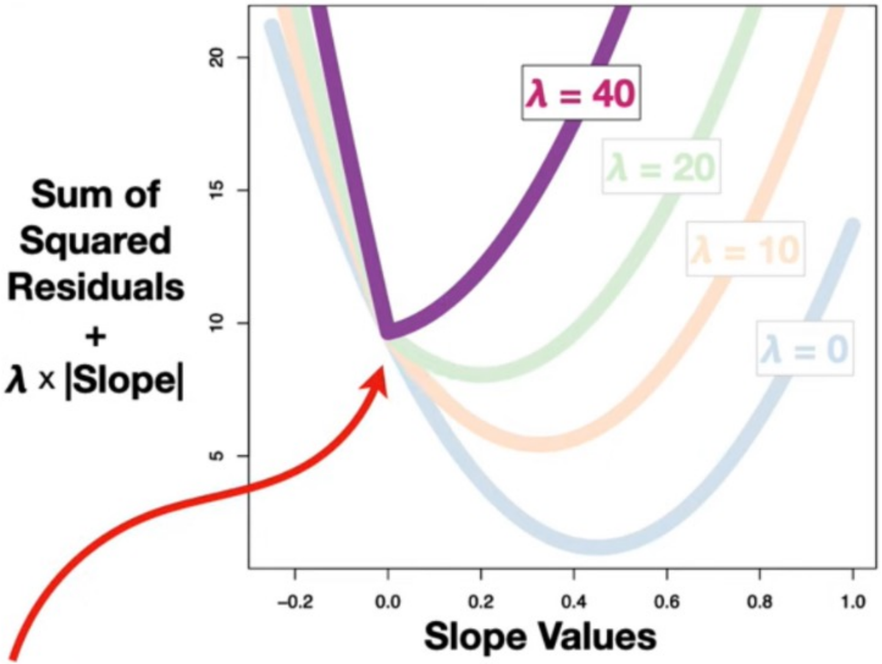
Lastly, the **thick purple line** represents **(lambda) λ = 40**...

Lowest point GD Curve

$\boxed{m = 0}$

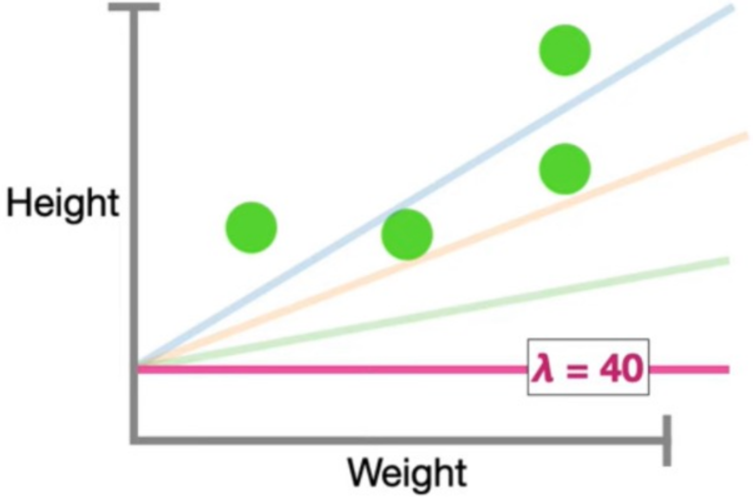
Sum of the Squared Residuals + (λ × |Slope|)

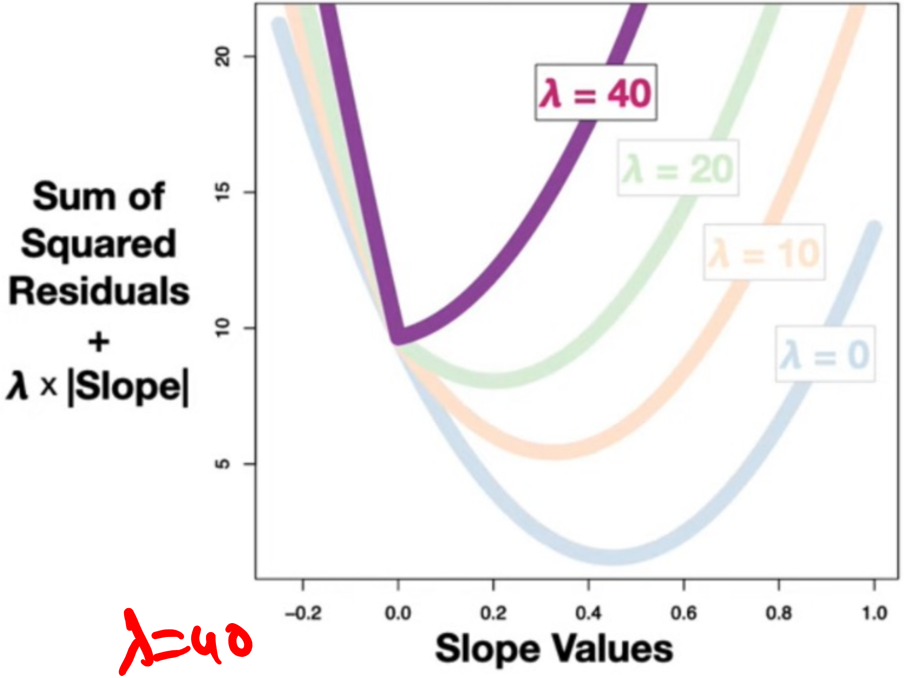




Now the lowest point in the **purple curve**, aka, the optimal slope given the **Absolute Value Penalty** when **λ = 40**, is **0**.

Sum of the Squared Residuals + (λ × |Slope|)





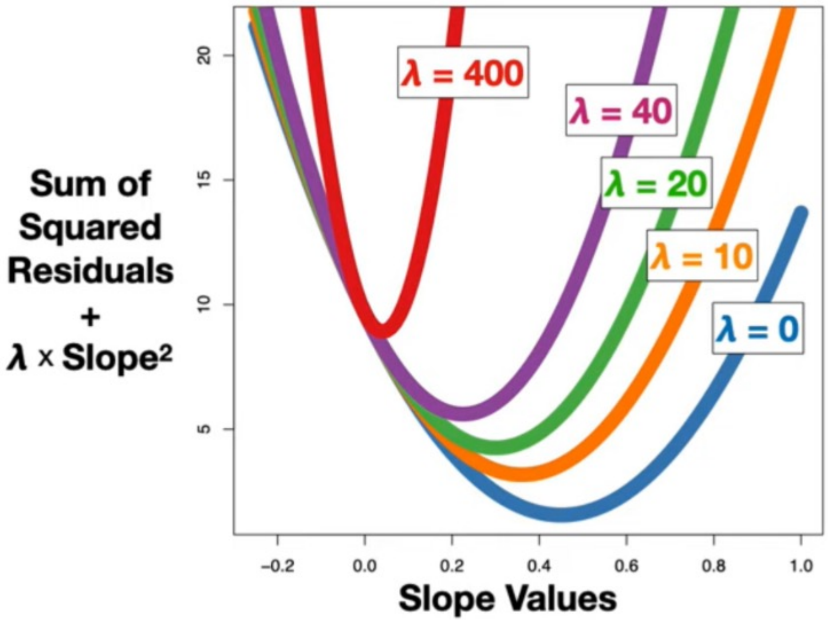
λ=40

And that means that when (lambda) $\lambda = 40$, we ignore **Weight** as a variable when predicting **Height**.

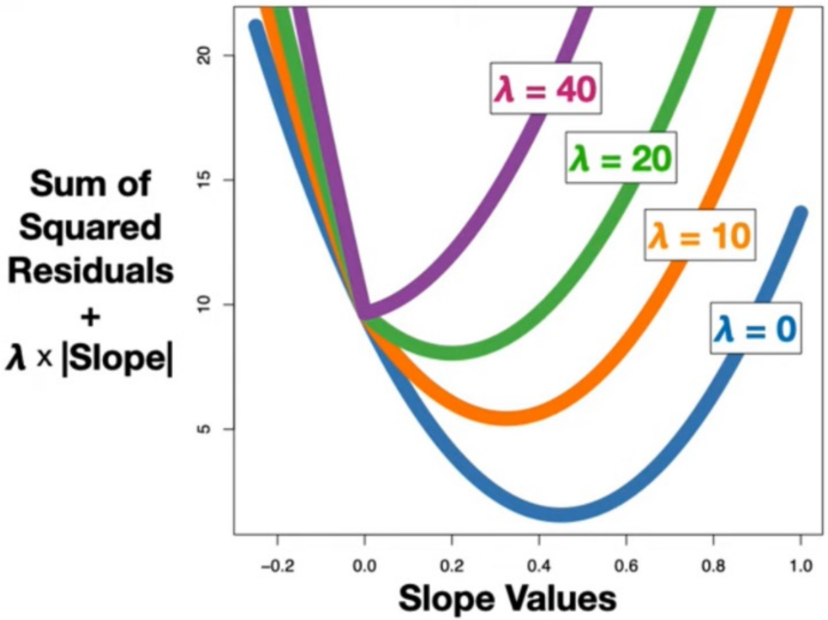
Sum of the Squared Residuals + (λ × |Slope|)

In summary…

upGrad



Sum of Squared Residuals + λ x Slope²
20
15
10
5
λ = 400
λ = 40
λ = 20
λ = 10
λ = 0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
Slope Values



Sum of Squared Residuals + λ x |Slope|
20
15
10
5
λ = 40
λ = 20
λ = 10
λ = 0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
Slope Values

upGrad

#LifeKoKaroLift

Thank You!