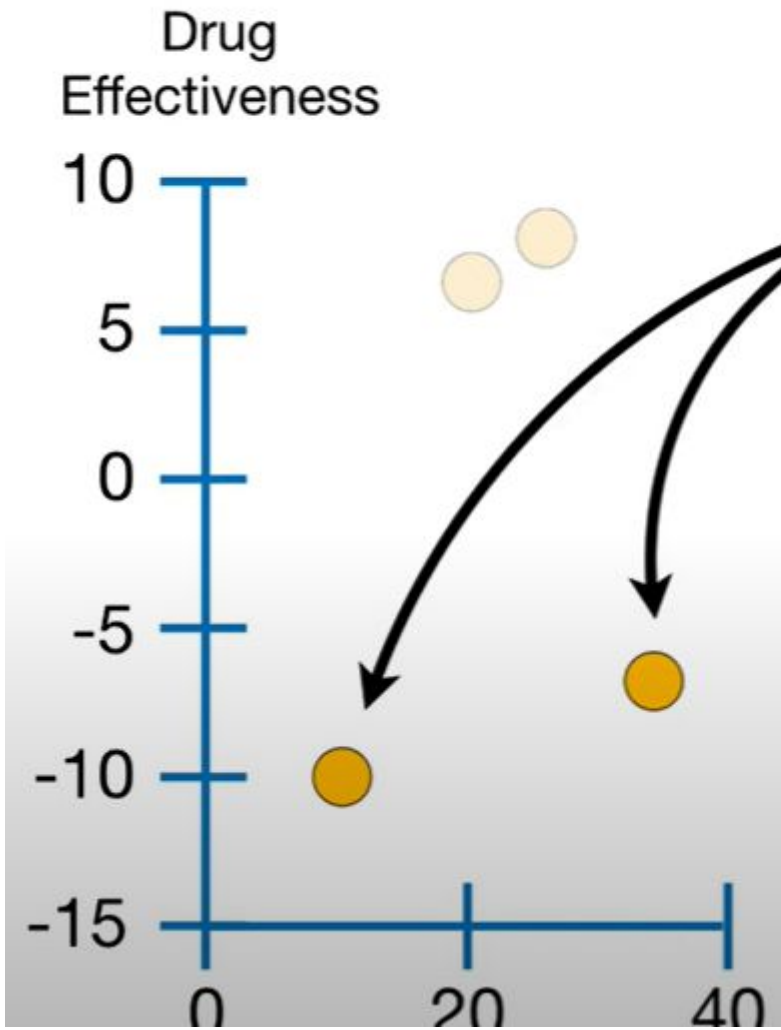
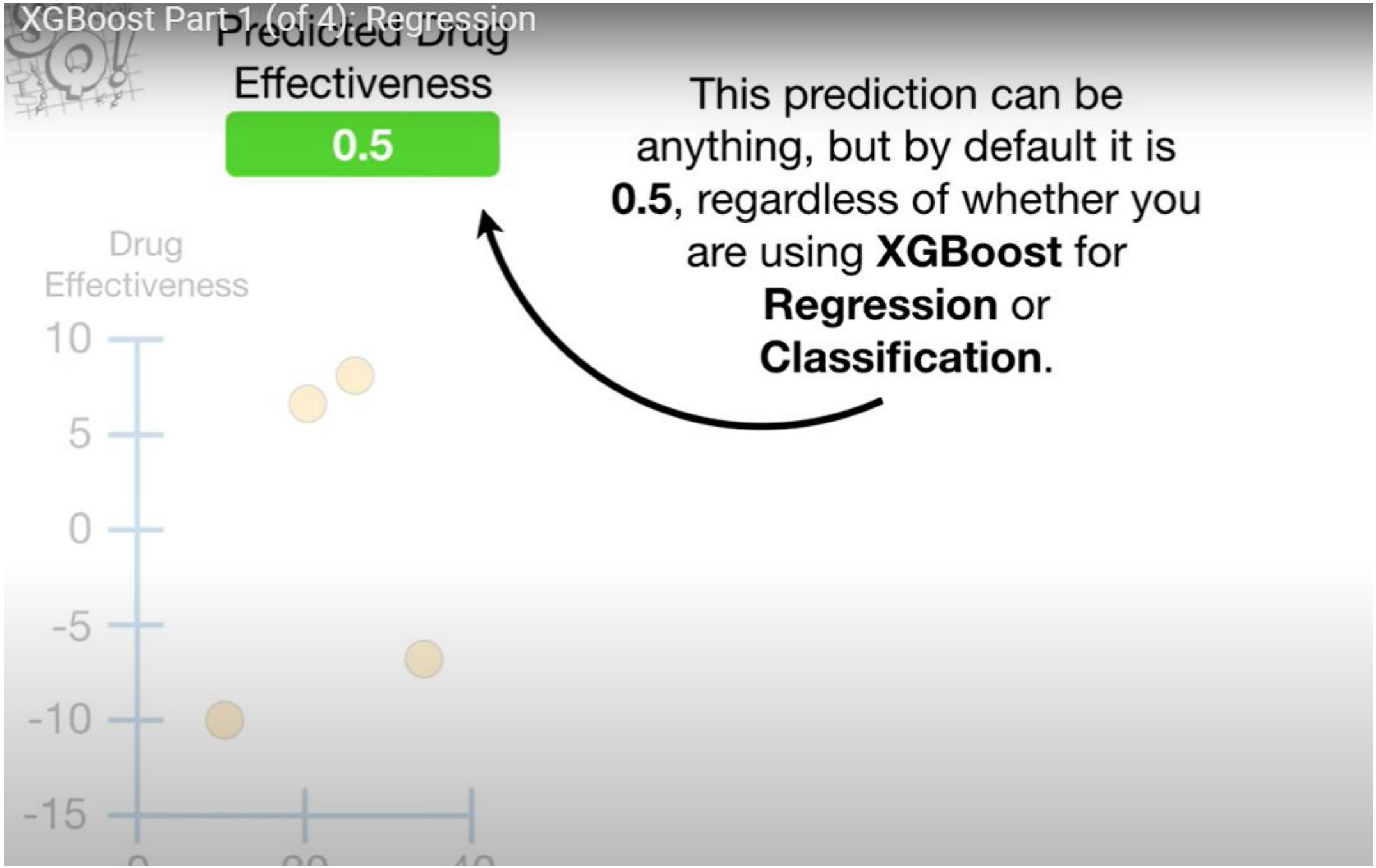


These two observations have relatively large negative values for **Drug Effectiveness**, and that means that the drug did more harm than good.


XGBoost Part 1 (of 4): Regression
Predicted Drug Effectiveness
0.5
This prediction can be anything, but by default it is **0.5**, regardless of whether you are using **XGBoost** for **Regression** or **Classification**.
Drug Effectiveness
10
5
0
-5
-10
-15

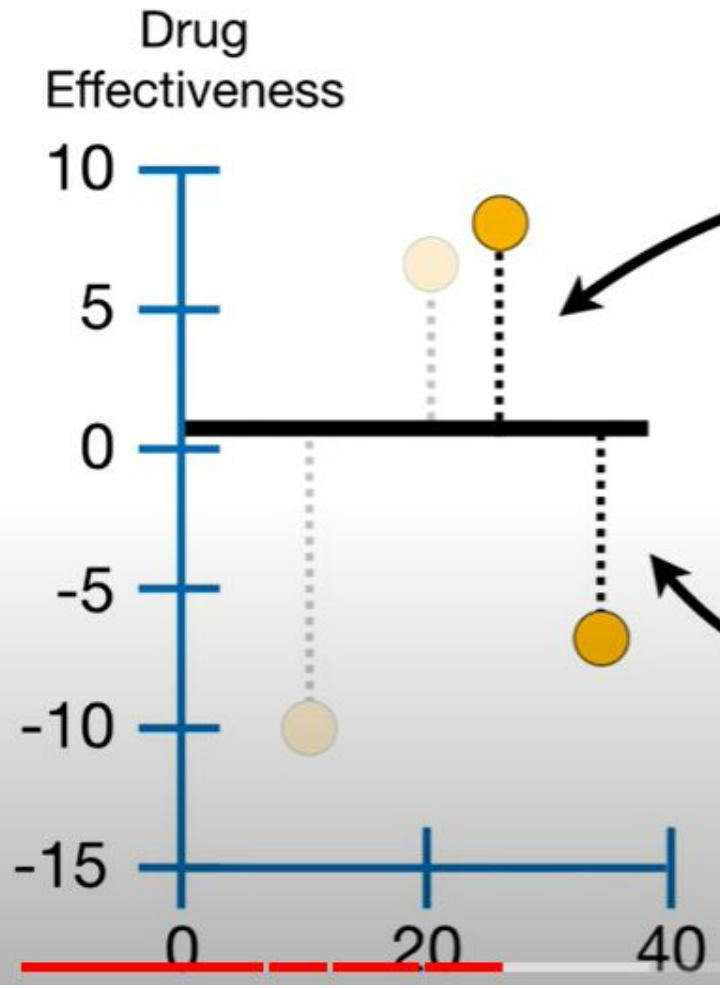
XGBoost Part 1 (of 4): Regression

Predicted Drug Effectiveness

0.5

Press Esc to exit full screen

-10.5, 6.5, 7.5, -7.5



$$\text{Similarity Score} = \frac{(-10.5 + 6.5 + 0)^2}{4 + 0}$$

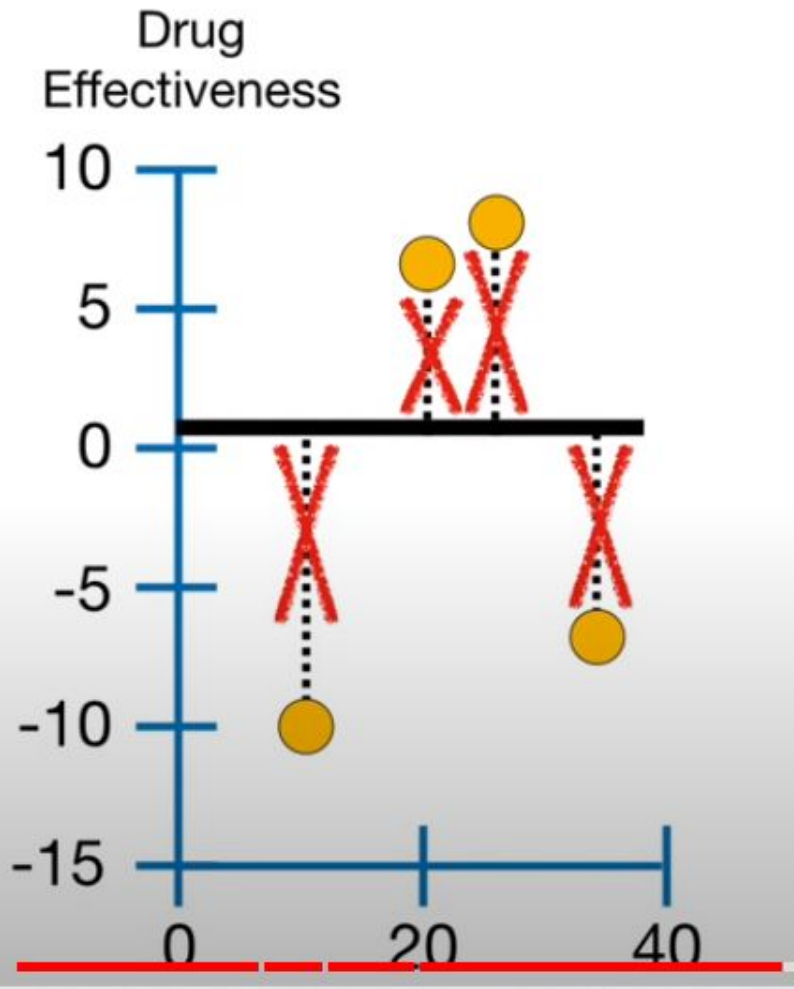
In other words, when we add this **Residual**…

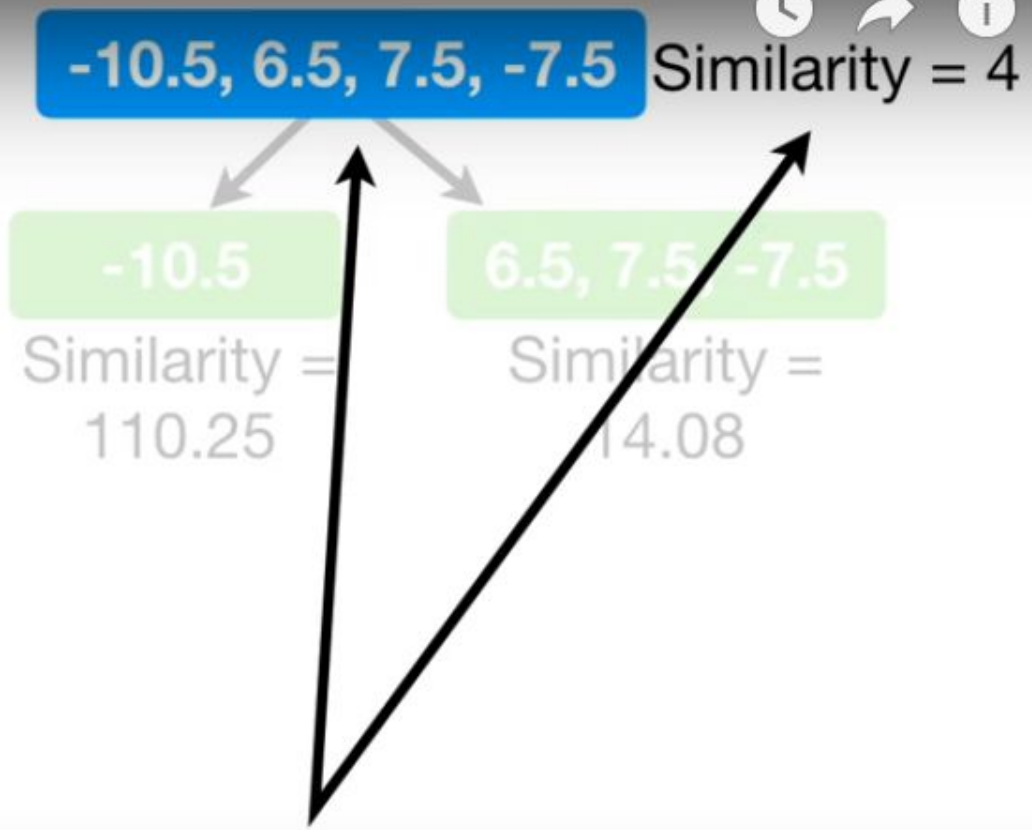
…to this **Residual**…

XGBoost Part 1 (of 4): Regression

Predicted Drug Effectiveness

0.5





…we see that when the **Residuals** in a node are very different, they cancel each other out and the **Similarity Score** is relatively small.

Link

- https://medium.com/@prathameshsonawane/xgboost-how-does-this-work-e1cae7c5b6cb