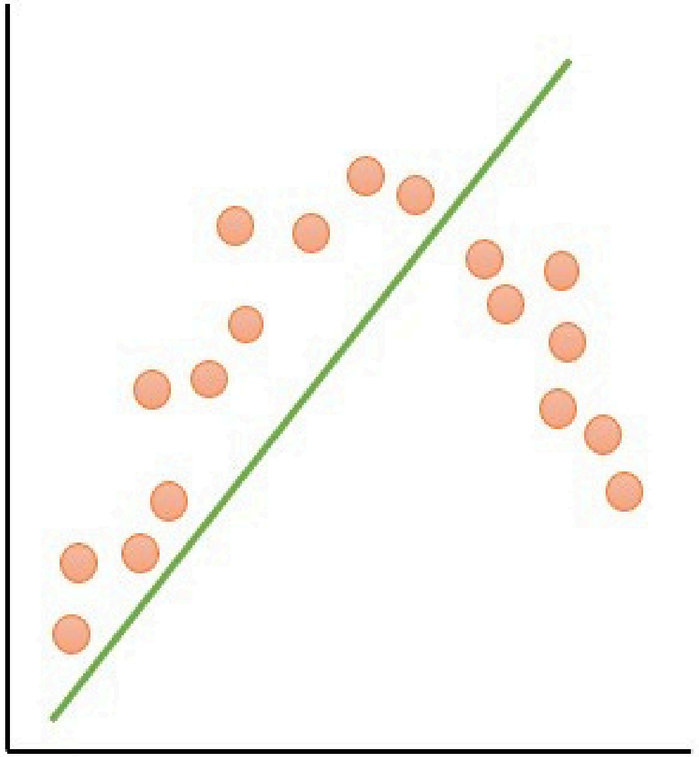
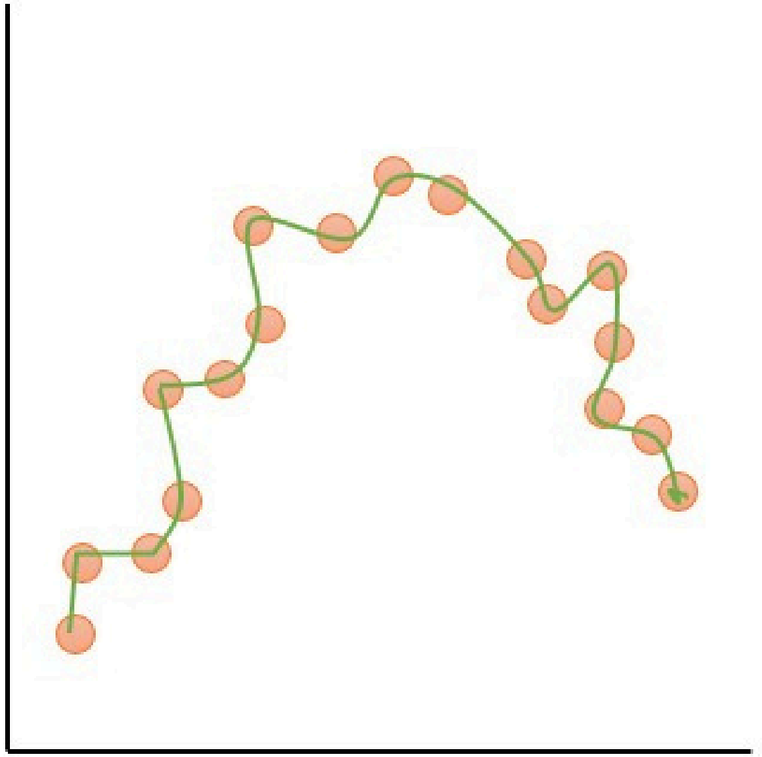
Bias Variance Tradeoff

High Bias, Underfitting

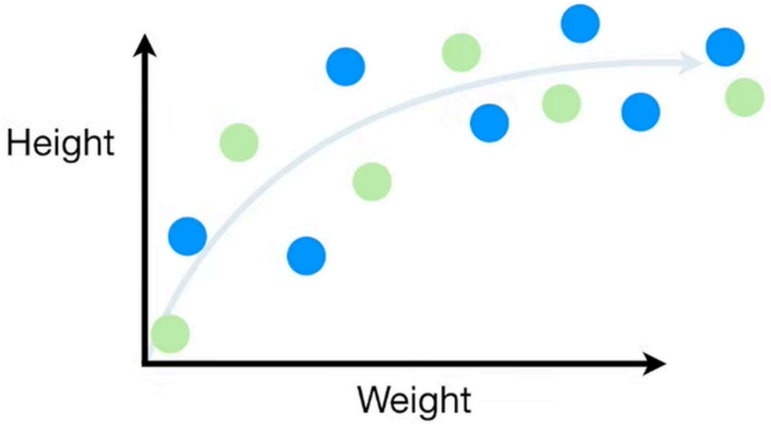


Low Bias, Overfitting

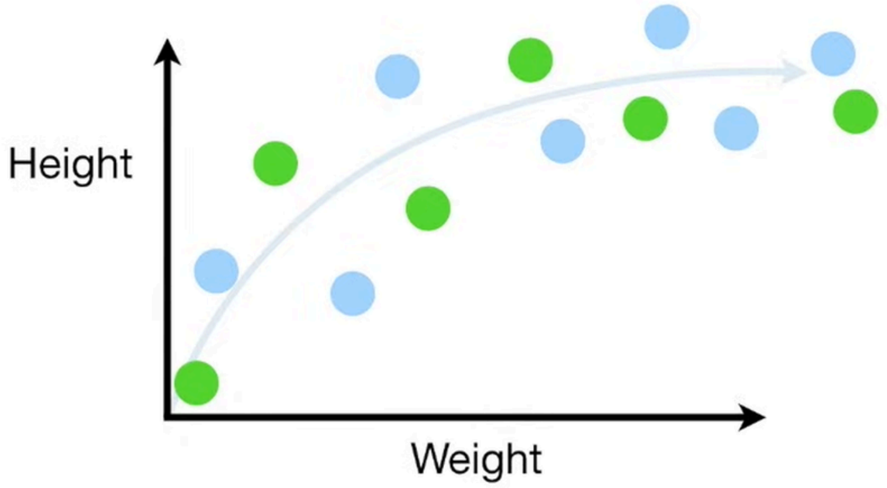


Machine Learning Fundamentals: Bias and Variance, Clearly Explained!!!

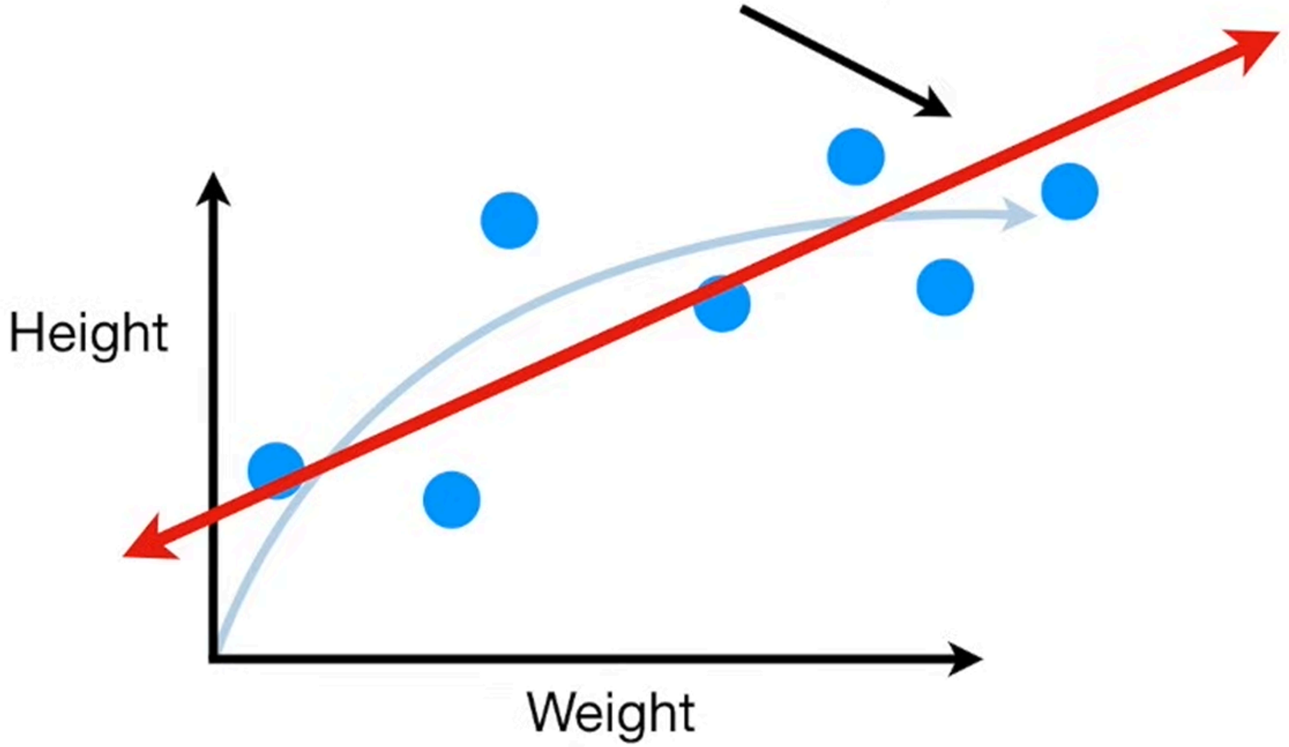
The **Blue Dots** are the **training set**…



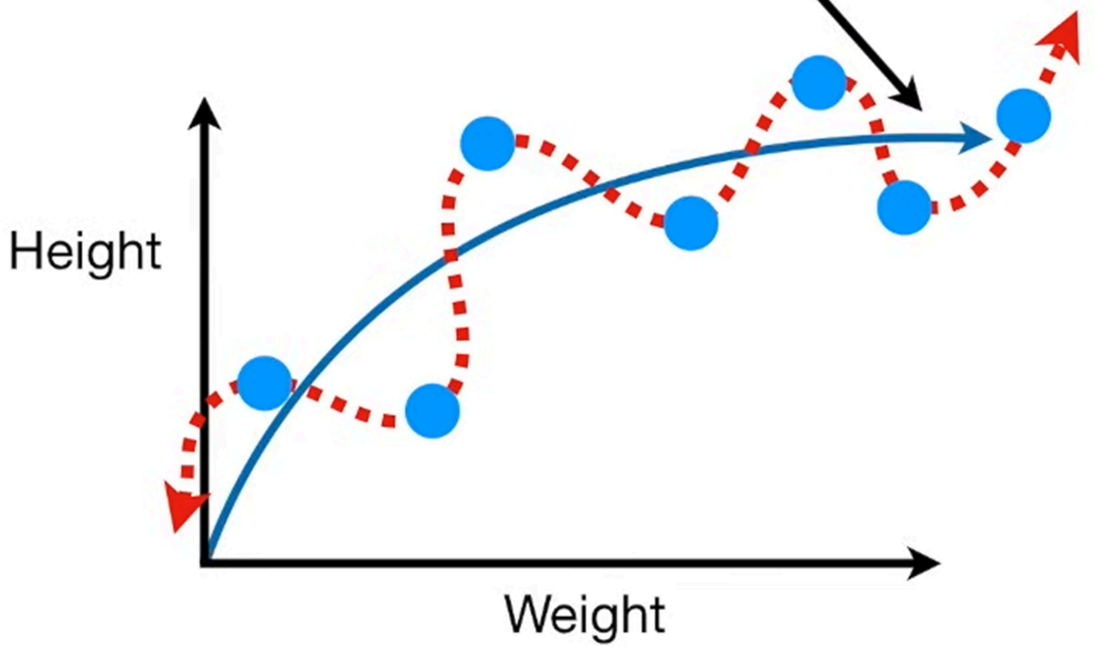
…and the **Green Dots** are the **testing set**.



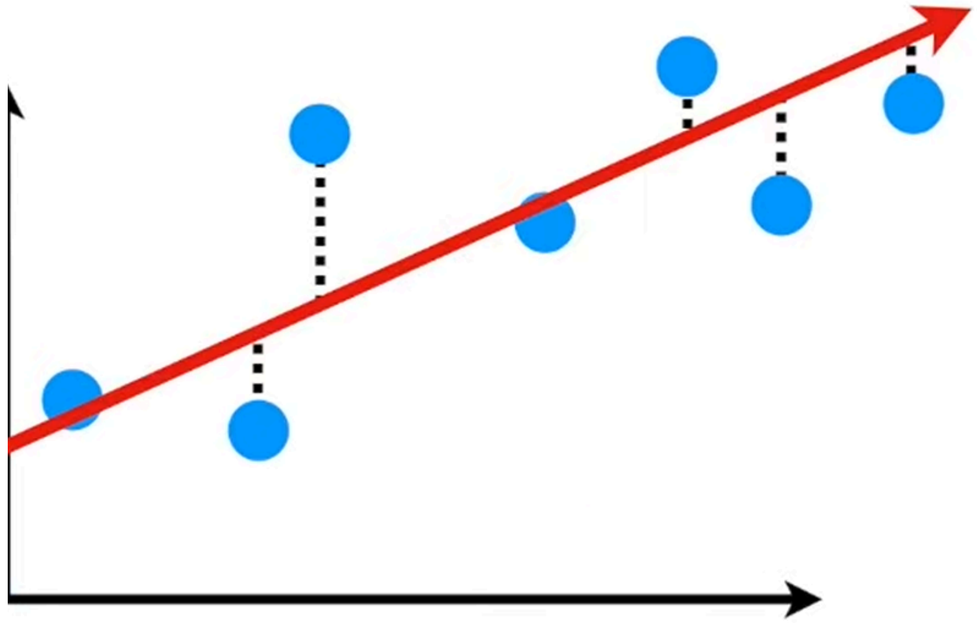
that we will use is Linear Regression
(aka “Least Squares”).

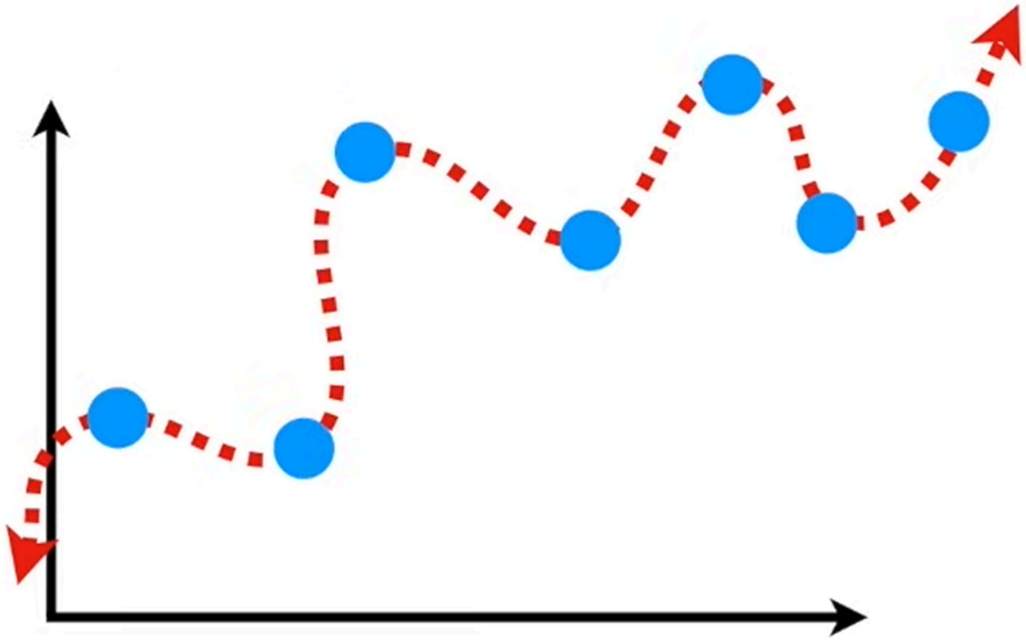


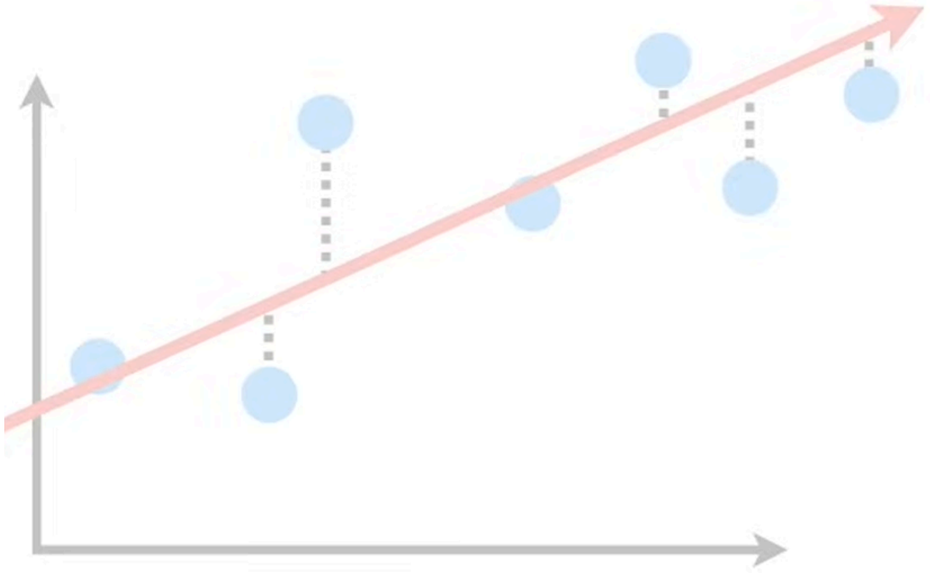
The **Squiggly Line** is super flexible and hugs the **training set** along the arc of the true relationship.



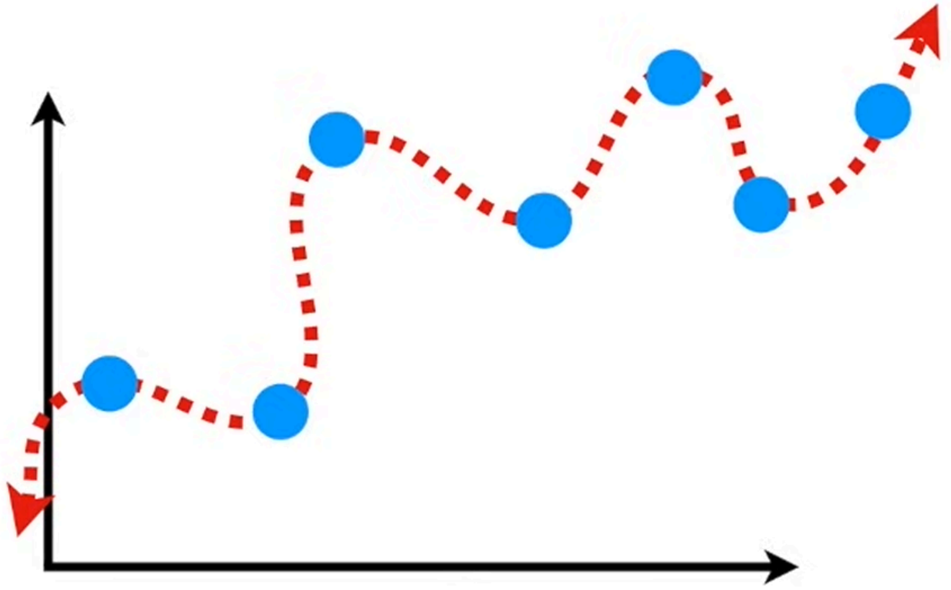
We can compare how well the **Straight Line** and the **Squiggly Line** fit the **training set** by calculating their sums of squares.



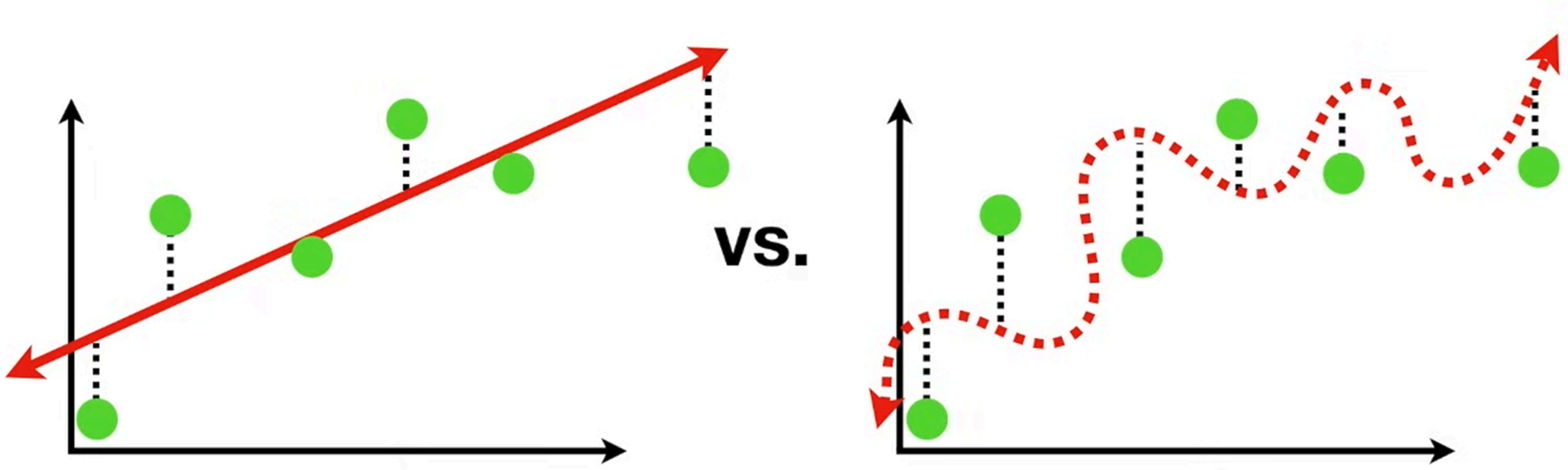




Notice how the **Squiggly Line** fits the data so well that the distances between the line and the data are all 0.



In the contest to see whether the **Straight Line** fits the **testing set** better than the **Squiggly Line**…



Underfitting
overfitting