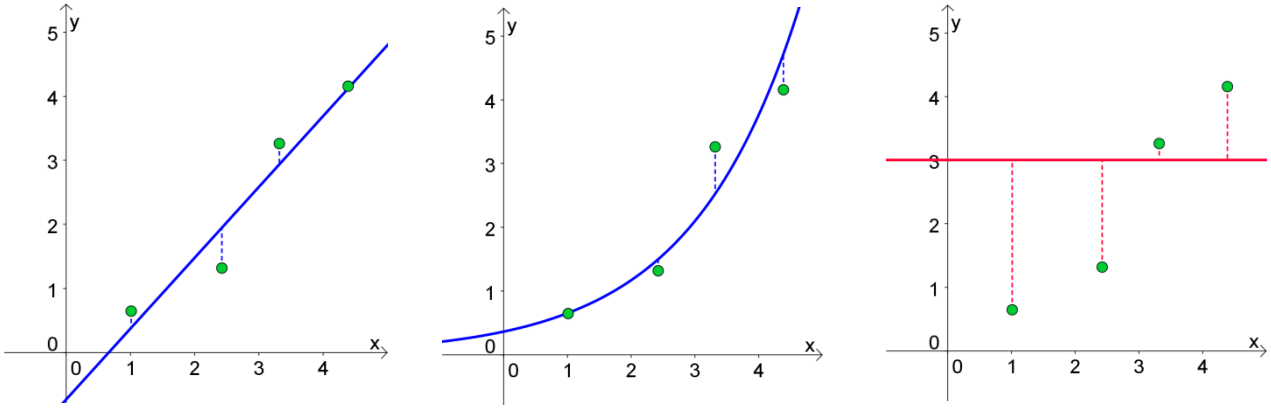
The Method of Least-Squares

Q: Given a scatterplot of data, how do we find a function that "best fits" the data?



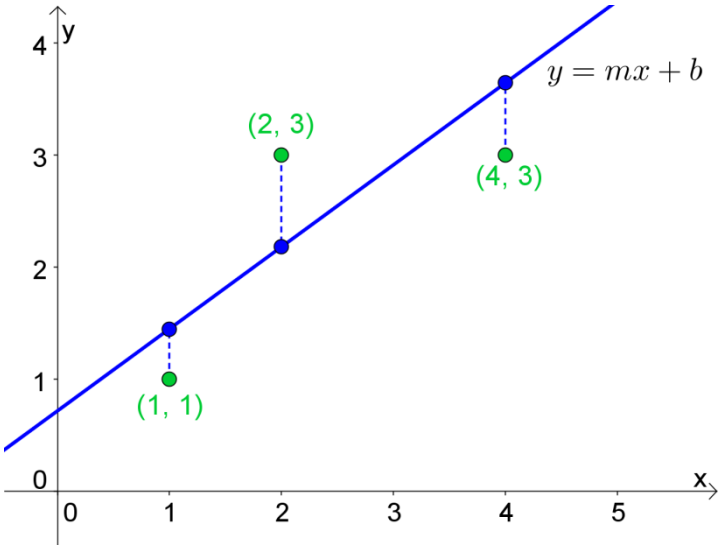
If we can *minimize* the *vertical distances* from the points to the function, then we'll have a good fit. It turns out to be more convenient to use the squares of the vertical distances.

A function is a "best fit" according to the _____ if the sum of the squares of the vertical distances is minimized.

We'll focus on using a linear model here ($y = mx + b$), and so we'll get a two-variable function (the two variables are m and b) that we can minimize using partial derivatives.

Ex 1.

Use the least-squares criterion to find the equation of the line that is closest to the three points $(1,1)$, $(2,3)$, and $(4,3)$.



Note: The above method can be generalized to give two "nice" formulas for finding m and b:

$$m = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \qquad \text{and} \qquad b = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x)^2}$$

This result is used in statistics classes.