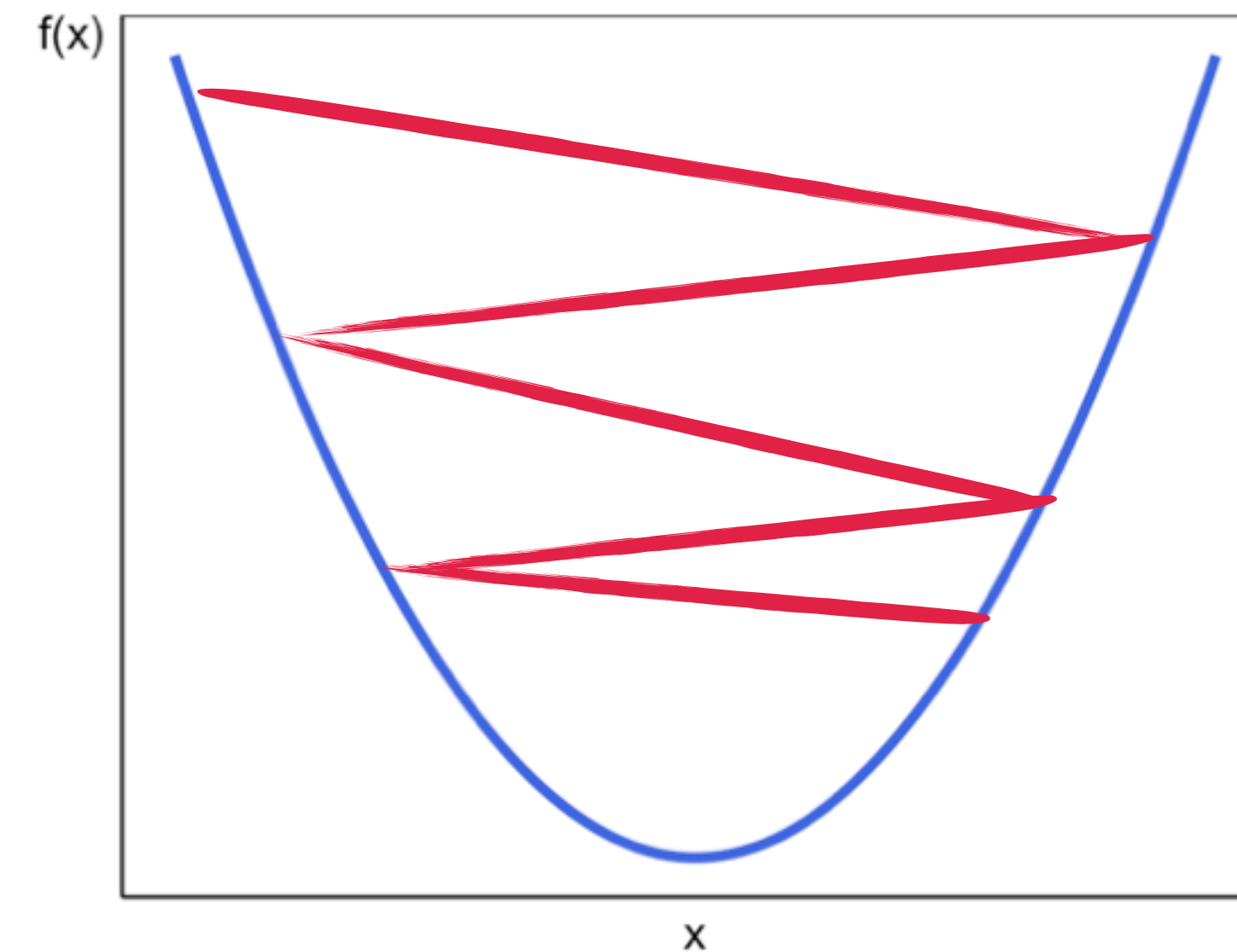
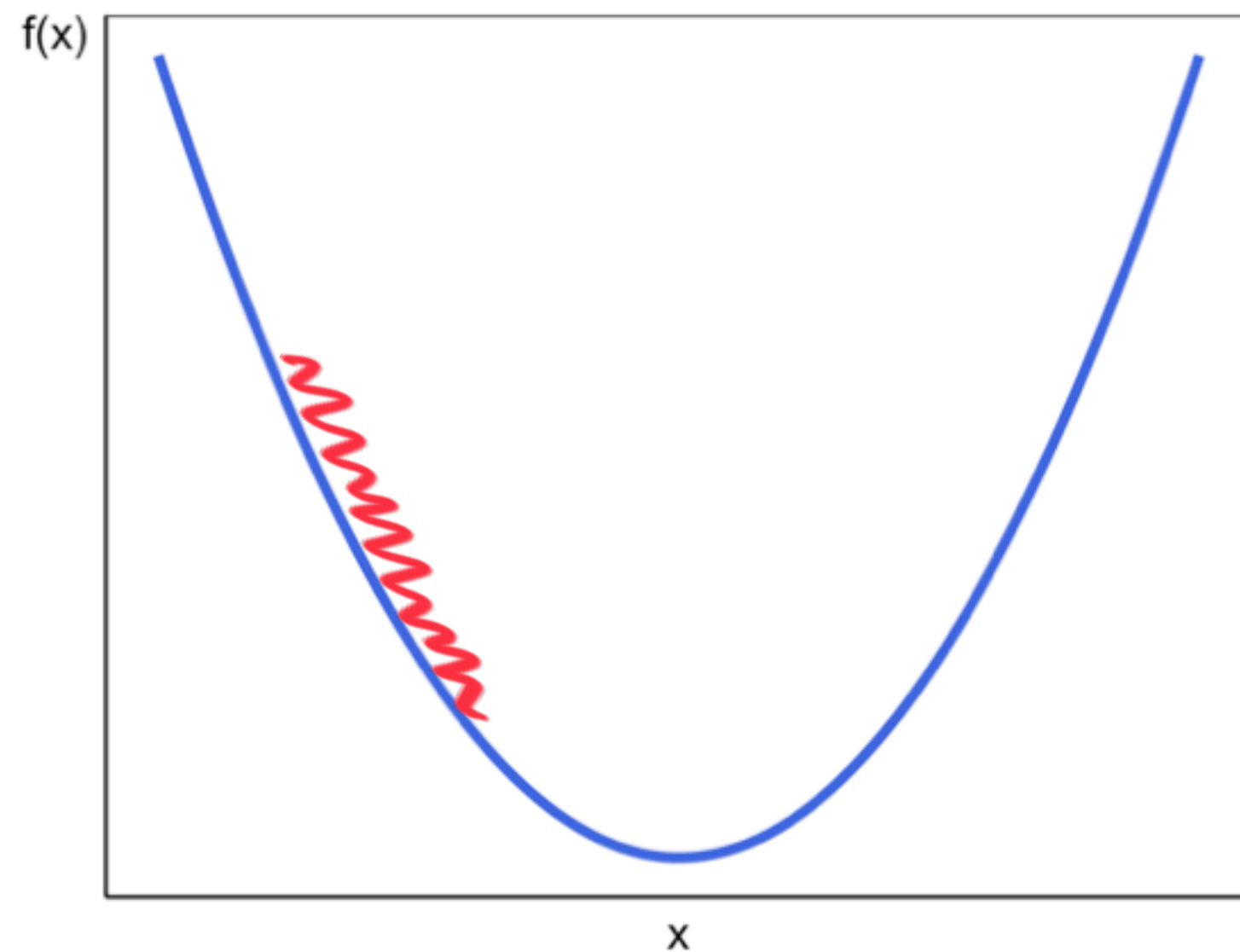


Estimation: Gradient Descent

- The goal is to minimize the loss function
- The gradient tells us which direction to move
- The **Learning Rate** controls how big of a step we take
 - Small steps mean slower convergence
 - Large steps mean we might step over minima

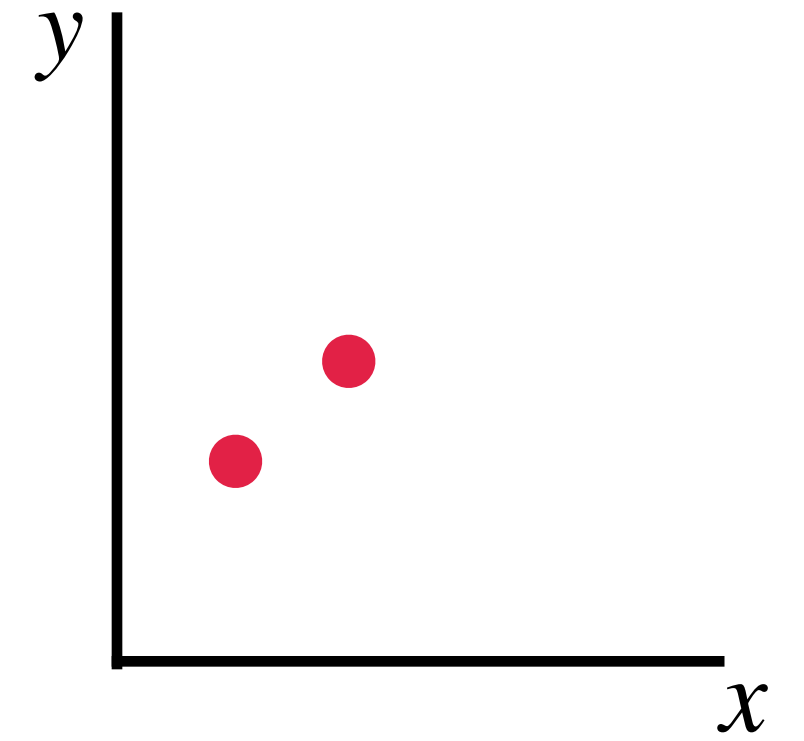


Example: Very Simple Linear Regression

$$\hat{y} = b_0 + b_1x$$

Loss function:

$$\begin{aligned} RSS &= \sum_i^N (\text{actual} - \text{predicted})^2 = \sum_i^N (y_i - \hat{y}_i)^2 \\ &= \sum_i^N (y_i - (b_0 + b_1x))^2 \\ &= \sum_i^N (y_i - b_0 - b_1x)^2 \end{aligned}$$



Assume only 2 data points: $(x_1, y_1) = (1, 2)$, $(x_2, y_2) = (2, 3)$