

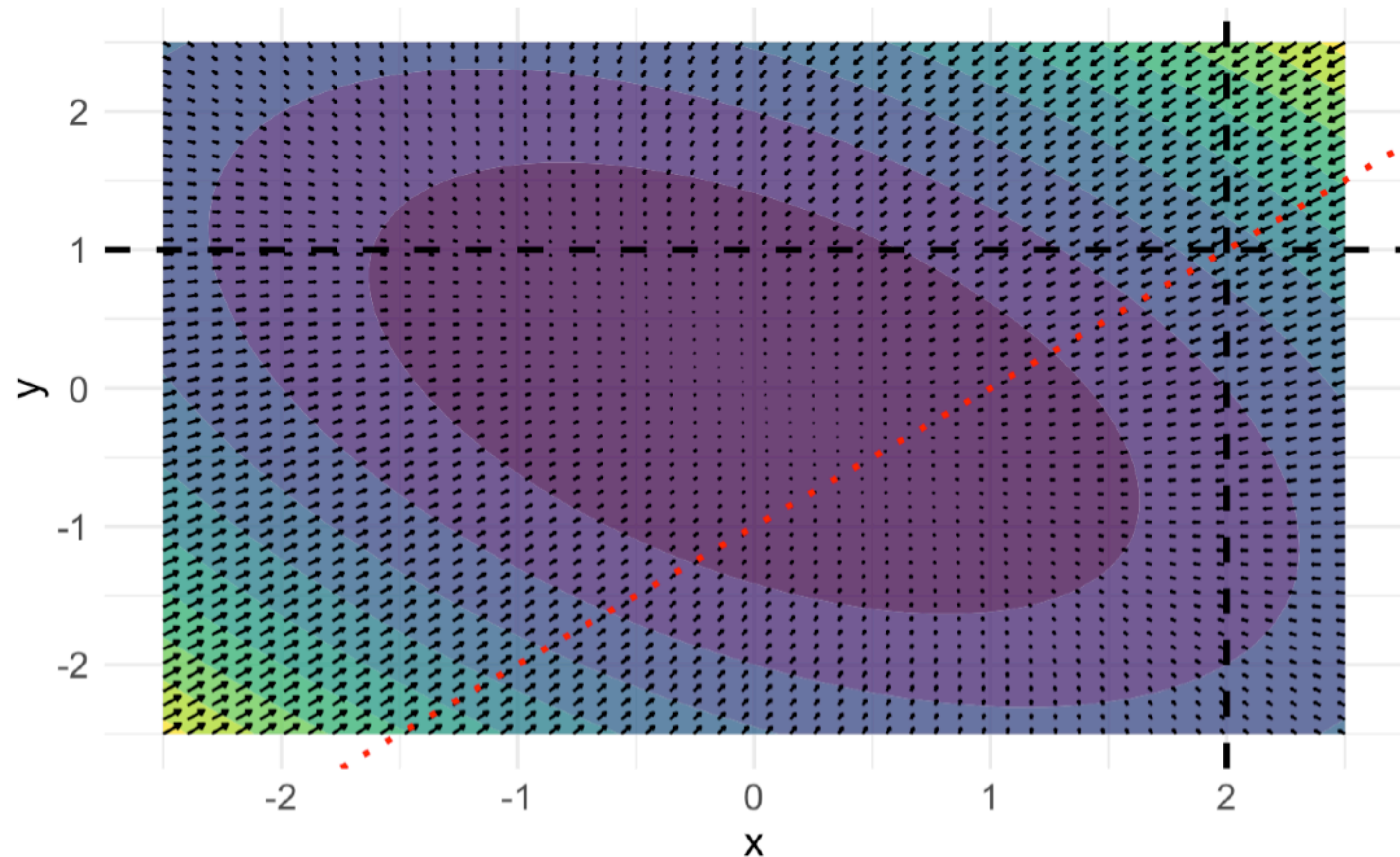
Gradient Descent

$$f(x, y) = x^2 + xy + y^2$$



Step size determined by
Learning Rate ρ :

$$\begin{bmatrix} x_{new} \\ y_{new} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} - \rho \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$



gradient tells us what adjustments we should make to each of our parameters

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

