In ML we often look at functions of vectors \( \omega \in \mathbb{R}^d \).

**Examples:**

1. \( f(\omega) = \omega^T a + b \)
   - \( a \) is a \( d \times 1 \) vector, \( b \) is a scalar

2. \( f(\omega) = \frac{1}{2} \| \omega \|^2 \)
   - \( \omega \) is a \( d \times 1 \) vector

**What do these functions look like?**

1. \( f(\omega) = \omega^T a + b \)
   - Linear function
   - In 1d, slope = \( a \)
   - Intercept = \( b \)

2. \( f(\omega) = \frac{1}{2} \| \omega \|^2 \)
   - In 1d: parabola
   - In higher dimensions, paraboloid

**Gradients:** Derivative for multivariable functions.

\[ f: \mathbb{R}^d \rightarrow \mathbb{R} \quad \text{(f is a function of d variables)} \]

\[ \nabla f(\omega) = \begin{bmatrix} \frac{\partial f}{\partial \omega_1} \\ \vdots \\ \frac{\partial f}{\partial \omega_d} \end{bmatrix} \]

Gradient of \( f \) is a \( d \times 1 \) vector, whose coordinate \( i \) is the partial derivative with respect to \( \omega_i \)

**Examples:**

1. \( f(\omega) = \omega^T a + b = \sum_{i=1}^{d} a_i \omega_i + b \)
   - For each \( i \), \( \frac{\partial f}{\partial \omega_i} = a_i \)

So \[ \nabla f(\omega) = \begin{bmatrix} a_1 \\ \vdots \\ a_d \end{bmatrix} = a \]
2. \[ f(\omega) = \frac{1}{2} \| \omega \|_2^2 = \frac{1}{2} \sum_{i=1}^{d} \omega_i^2 \]

\[ \frac{\partial f}{\partial \omega_i} = \frac{1}{2} 2 \omega_i = \omega_i, \text{ so } \nabla f(\omega) = \omega. \]

Gradient represents the direction along which the function increases the fastest.

Why is this useful? If we want to minimize a function from a starting point, we can follow the direction opposite to the gradient. This is called gradient descent. This continues until we get close to the minimum point where \( \nabla f = 0 \).