

## Loss Function Definition

A loss function  $L(\hat{y}, y)$  measures the discrepancy between the predicted output  $\hat{y}$  and the true target  $y$ , guiding the network to minimize prediction errors during training.

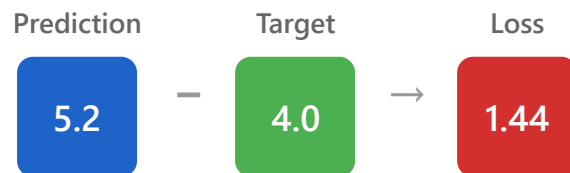
### Regression Tasks

#### Mean Squared Error (MSE)

$$L = (1/n) \sum (\hat{y}_i - y_i)^2$$

- **Penalizes large errors** quadratically
- **Differentiable** everywhere
- **Sensitive to outliers**

#### Example Calculation



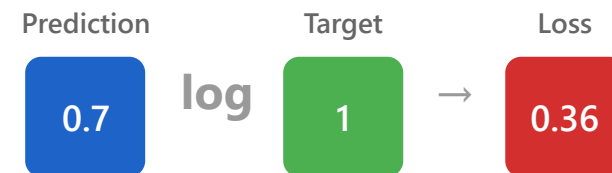
### Classification Tasks

#### Cross-Entropy Loss

$$L = -\sum y_i \log(\hat{y}_i)$$

- **Measures probability distribution** difference
- **Works with softmax** output
- **Penalizes confident mistakes** heavily

#### Example Calculation



## Common Loss Functions

### MSE

**Task:** Regression

**Formula:**  $\sum (\hat{y} - y)^2$

**Use Case:** Continuous values

### Cross-Entropy

**Task:** Classification

**Formula:**  $-\sum y \log(\hat{y})$

**Use Case:** Class probabilities

### MAE

**Task:** Regression

**Formula:**  $\sum |\hat{y} - y|$

**Use Case:** Robust to outliers

### Training Objective

The goal is to find parameters  $\theta$  that minimize the average loss:  $\theta^* = \operatorname{argmin} (1/n) \sum L(f(x_i; \theta), y_i)$