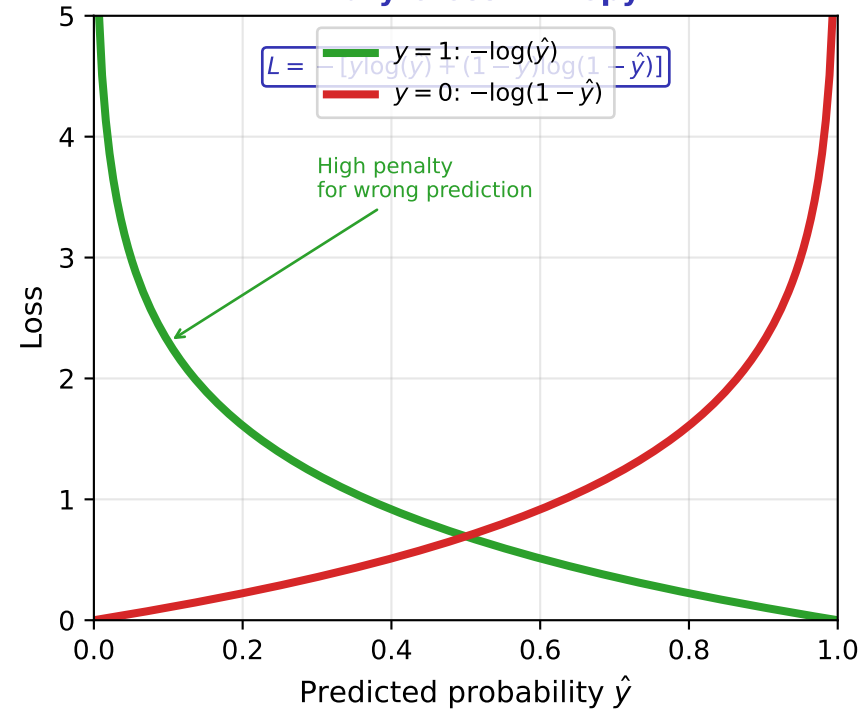
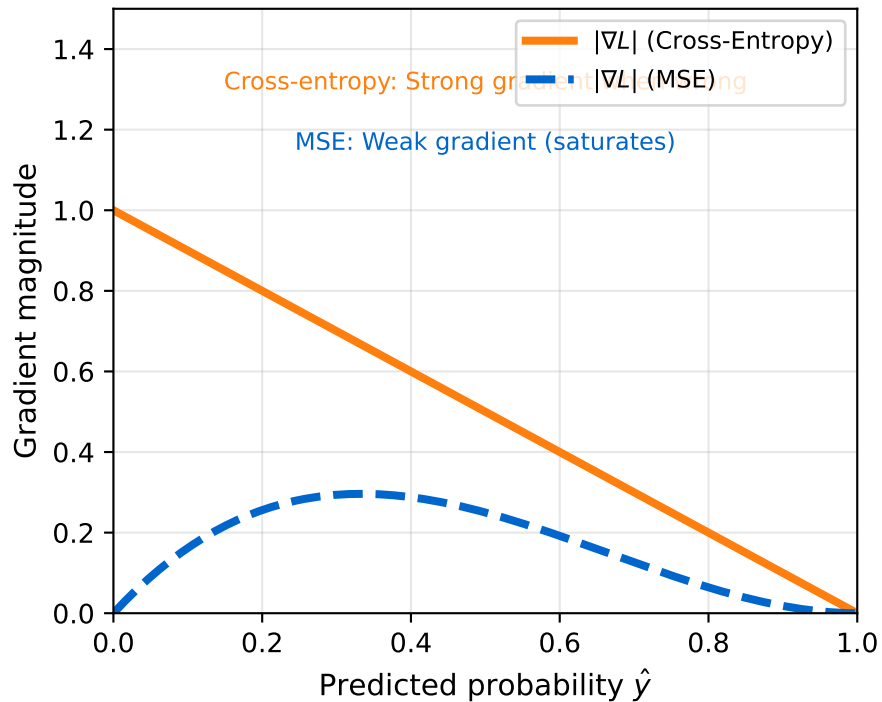


# Cross-Entropy: The Standard Classification Loss

## Binary Cross-Entropy



## Gradient Behavior



## Cross-Entropy Properties

**Probability-based:** Measures information difference

**Strong gradients:** Fast learning from mistakes

**No saturation:** Always learns from errors

**Natural for classification:** Matches softmax output

### Why use Cross-Entropy over MSE?

- Stronger gradients for confident wrong predictions
- Information-theoretic foundation
- Standard for classification tasks