L02: Logistic Regression

Classification with Probability Estimates

Methods and Algorithms – MSc Data Science

Learning Objectives

By the end of this lecture, you will be able to:

1. Explain how logistic regression models binary outcomes
2. Derive the maximum likelihood estimation for logistic regression
3. Interpret classification metrics (precision, recall, AUC)
4. Apply logistic regression for credit scoring decisions

Finance Application: Credit default prediction

These objectives span Bloom's levels: Understand, Apply, Analyze

Why Logistic Regression?

The Business Problem

- Banks must decide: approve or reject loan applications
- Need probability of default, not just yes/no prediction
- Regulatory requirement: interpretable, auditable models

Why Not Linear Regression?

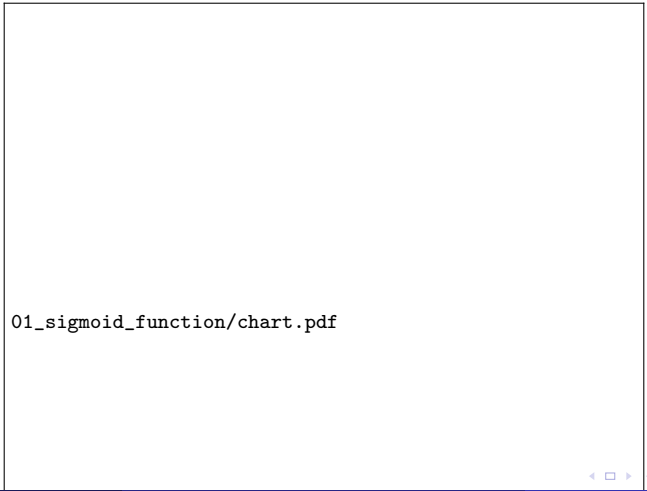
- Linear regression can predict values outside [0,1]
- Binary outcomes need probability-based approach
- Logistic regression outputs calibrated probabilities

Logistic regression: the industry standard for credit scoring since 1980s

The Sigmoid Function

From Linear to Probability

- Maps any real number to (0, 1) range
- Smooth, differentiable, interpretable

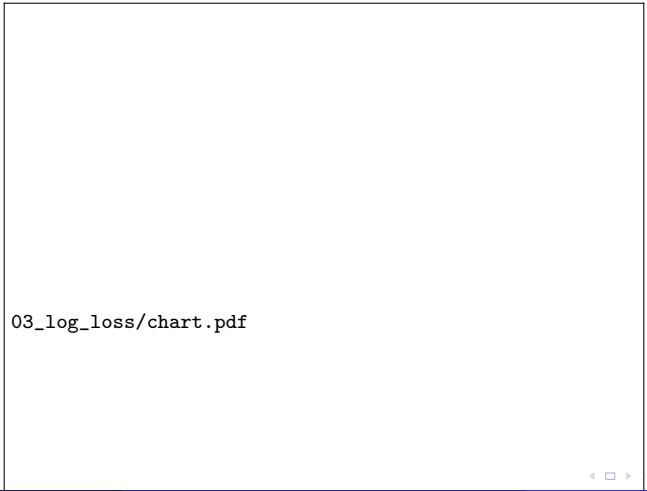


Decision Boundary

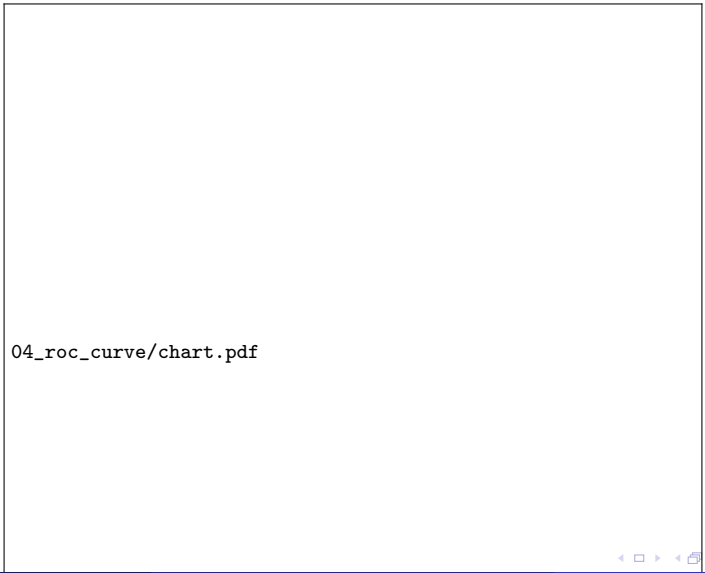
Binary Cross-Entropy Loss

Why Not MSE?

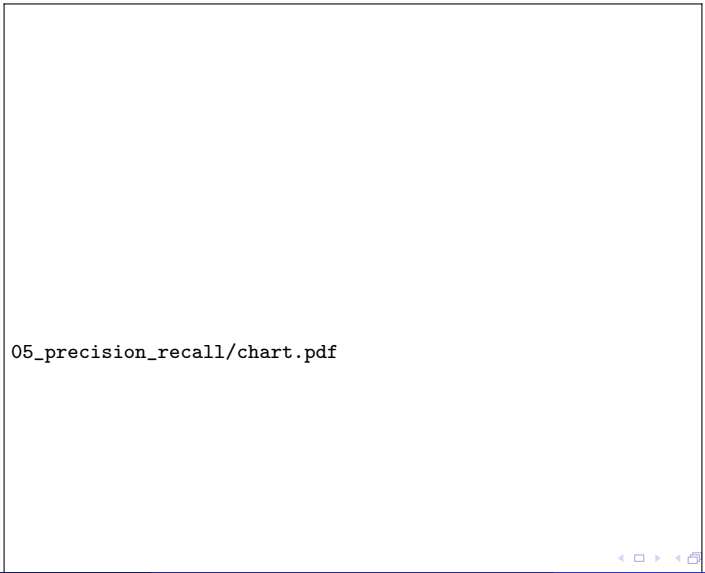
- MSE with sigmoid creates non-convex loss landscape
- Cross-entropy is convex, guarantees global optimum



ROC Curve and AUC



Precision-Recall Trade-off



Confusion Matrix: Reading the Results

When to Use Logistic Regression