

Methods and Algorithms

Spring 2026

1 Problem

2 Method

3 Solution

4 Practice

5 Decision Framework

6 Summary

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

- ➊ Explain how logistic regression models binary outcomes
- ➋ Derive the maximum likelihood estimation for logistic regression
- ➌ Interpret classification metrics (precision, recall, AUC)
- ➍ Apply logistic regression for credit scoring decisions

Finance Application: Credit default prediction

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

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

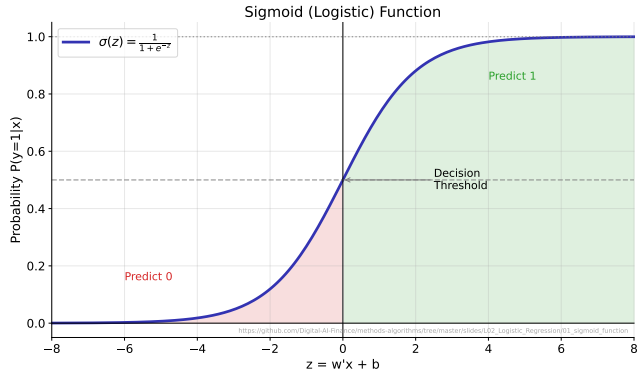
Why Not Linear Regression?

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

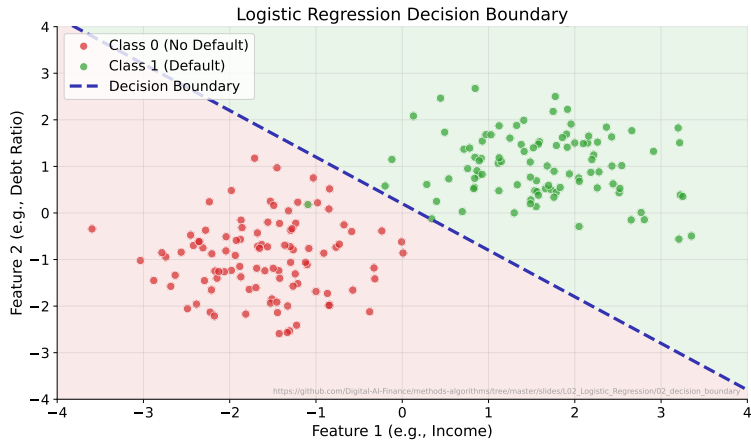
Logistic regression: the industry standard for credit scoring since 1980s

From Linear to Probability

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



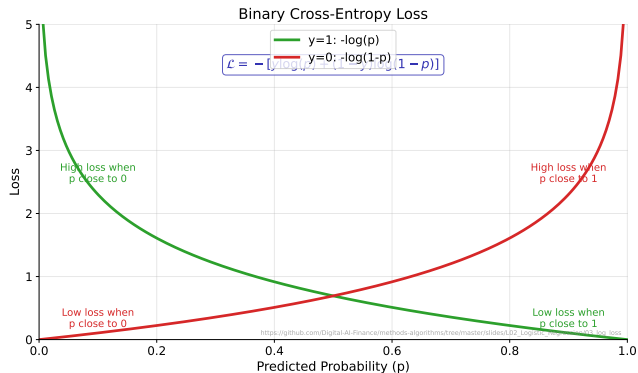
$\sigma(z) = 1/(1 + e^{-z})$ transforms linear combination to probability



The decision boundary is where $P(y = 1|x) = 0.5$, i.e., $w'x + b = 0$

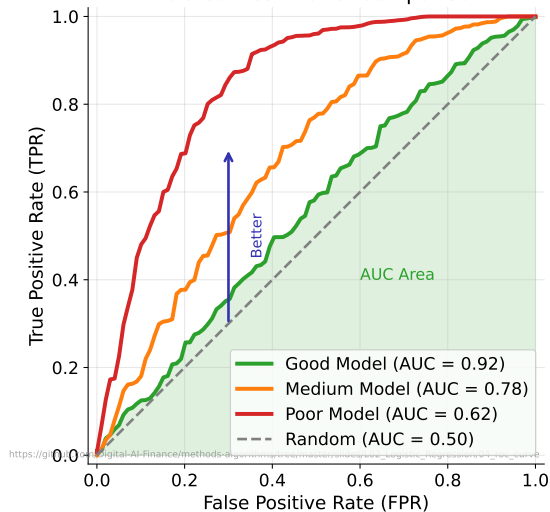
Why Not MSE?

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

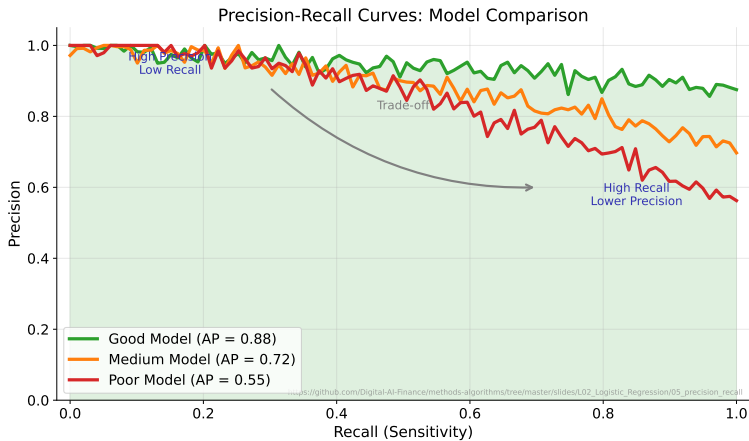


Heavily penalizes confident wrong predictions

ROC Curves: Model Comparison



AUC = probability random positive ranks higher than random negative



Use PR curve when classes are imbalanced (common in fraud detection)

Confusion Matrix: Credit Default Prediction

Actual Label	Predicted Label	
	Predicted: No Default	Predicted: Default
Actual: No Default	TN (True Negative) 850	FP (False Positive) 50
Actual: Default	FN (False Negative) 30	TP (True Positive) 70

Accuracy: 92.0%
Precision: 58.3%
Recall: 70.0%
F1 Score: 0.64

https://github.com/Digital-AI-Finance/methods-algorithms/tree/master/slides/L02_Logistic_Regression/06_confusion_matrix

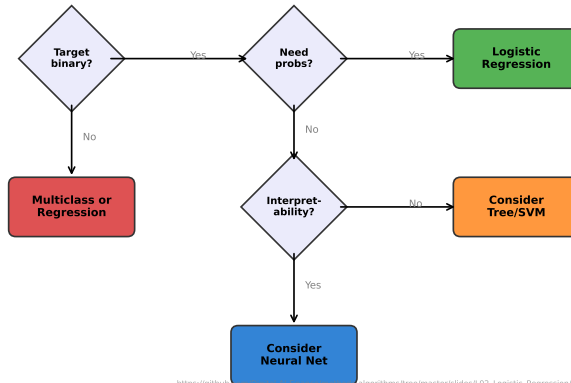
FP = approve bad loans (costly), FN = reject good customers (lost revenue)

Open the Colab Notebook

- Exercise 1: Implement logistic regression from scratch
- Exercise 2: Train model on credit scoring data
- Exercise 3: Evaluate with ROC curve and confusion matrix

Link: <https://colab.research.google.com/> [TBD]

Logistic Regression Decision Guide



https://github.com/DataCamp/algorithms/tree/master/slides/L02_Logistic_Regression/07_decision_flowchart

Key strengths: interpretable coefficients, probability outputs, fast training

- James et al. (2021). *Introduction to Statistical Learning*. <https://www.statlearning.com/>
- Hastie et al. (2009). *Elements of Statistical Learning*. <https://hastie.su.domains/ElemStatLearn/>