Lesson 28: Classification

Mini-Lecture Version (30 min)

Digital Finance

Learning Objectives: Explain classification problems and contrast with regression — Understand logistic regression and probability estimation — Interpret confusion matrices and classification metrics — Calculate and interpret accuracy, precision, recall, and F1-score

Classification vs. Regression

Regression (Lesson 27):

- Predict continuous values
- $Y \in \{0, 1, \ldots, K\}$ (discrete classes) $\in \mathbb{R}$ (e.g., stock return, price)
- Example: Predict bond yield (3.2%)
- Metrics: R^2, RMSE, MAE

Classification (Today):

- Predict discrete categories
- $Y \in \{0, 1, \ldots, K\}$ (discrete classes)
- Binary: Default (Yes/No)
- Multi-class: Credit rating (AAA, AA, A, ...)
- (See full lecture for details)

Focus Today: Binary classification (most common in finance).

Finance Examples:

Binary Classification:

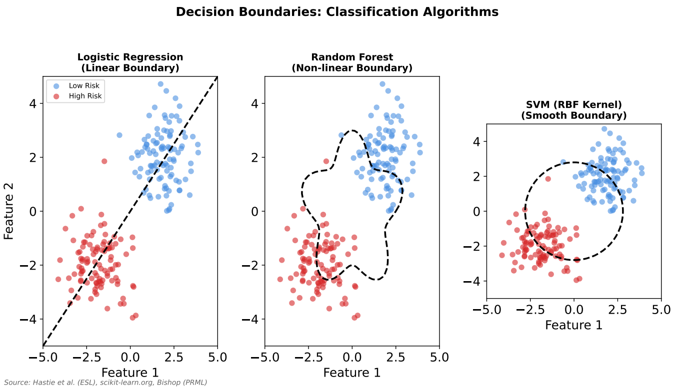
- Loan default (Yes/No)
- Fraud transaction (Fraud/Legitimate)
- Stock direction (Up/Down)
- M&A success (Completed/Failed)

Multi-class:

- Credit ratings (10 grades)
- Customer segments (5 types)
- Market regime (Bull/Bear/Sideways)

This concept is fundamental to understanding Classification.

Decision Boundaries



This concept is fundamental to understanding Classification.

Logistic Regression: The Model

Why Not Linear Regression?

- For binary $Y \in \{0, 1\}$, OLS can predict $\hat{Y} < 0$ or $\hat{Y} > 1$
- Violates probability interpretation
- Residuals not normal

Logistic Regression:

- Predict probability: $P(Y = 1|X)$
- Output constrained to $[0, 1]$
- Use logistic (sigmoid) function

Model:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)}}$$

Logit Form:

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

- Left side: Log-odds (logit)
- Right side: Linear combination (like linear regression)

Interpretation:

- $\beta_j > 0$: Increase in X_j increases probability of $Y = 1$
- $\beta_j < 0$: Decrease probability
- Magnitude: Change in log-odds per unit X_j

Estimation: Maximum Likelihood (not OLS).

This concept is fundamental to understanding Classification.

Example: Loan Default Prediction

Problem:

- Predict default (1) vs. repayment (0)
- Features:
- Credit score
- Debt-to-income ratio
- (See full lecture for details)

Fitted Model:

$$\log\left(\frac{P(\text{Default})}{1-P(\text{Default})}\right) = 2.5 - 0.015 \times \text{CreditScore} + 1.2 \times \text{DebtToIncome}$$

Interpretation:

- **Credit Score** ($\beta = -0.015$):
  - 10-point increase $\rightarrow$ log-odds decrease by 0.15
  - Higher score $\rightarrow$ lower default probability
- **Debt-to-Income** ($\beta = 1.2$):
  - 0.1 increase (10 pp) $\rightarrow$ log-odds increase by 0.12
  - Higher debt burden $\rightarrow$ higher default risk

Example Prediction:
Credit Score $= 650$, D/I $= 0.4$:

$$\log(\text{odds}) = 2.5 - 0.015(650) + 1.2(0.4) = -7.28$$

$$P(\text{Default}) = \frac{1}{1+e^{7.28}} = 0.0007\ (0.07\%)$$

Real-world examples demonstrate Classification applications.

Making Predictions: Decision Threshold

From Probabilities to Classes:

- Logistic regression outputs $P(Y = 1|X)$
- Need decision rule to classify
- Default threshold: 0.5
- If $P > 0.5$, predict class 1
- (See full lecture for details)

Custom Thresholds:

- Threshold is tunable hyperparameter
- Lower threshold (0.3): More sensitive (more positives)
- Higher threshold (0.7): More specific (fewer false positives)

Example (Fraud Detection):

- Transaction 1: $P(\text{Fraud}) = 0.8 \rightarrow$ Flag as fraud
- Transaction 2: $P(\text{Fraud}) = 0.3 \rightarrow$ Depends on threshold
  - Threshold $= 0.5 \rightarrow$ Legitimate
  - Threshold $= 0.2 \rightarrow$ Fraud (more cautious)

Threshold Selection:

- Depends on cost of errors
- Fraud: False negative costly (miss fraud) $\rightarrow$ low threshold
- Spam: False positive costly (block good email) $\rightarrow$ high threshold
- Use ROC curve for tuning (later)

This concept is fundamental to understanding Classification.

Confusion Matrix

2x2 Table of Predictions vs. Actuals:

	Actual 0	Actual 1
Predict 0	TN (True Neg)	FN (False Neg)
Predict 1	FP (False Pos)	TP (True Pos)

Definitions:

- **TP:** Correctly predicted positive (fraud caught)
- **TN:** Correctly predicted negative (legit confirmed)
- **FP:** False alarm (Type I error, legit flagged as fraud)
- **FN:** Missed positive (Type II error, fraud missed)

Example (Fraud Detection):
Out of 1000 transactions:

	Not Fraud	Fraud
Predict Not	920 (TN)	10 (FN)
Predict Fraud	30 (FP)	40 (TP)

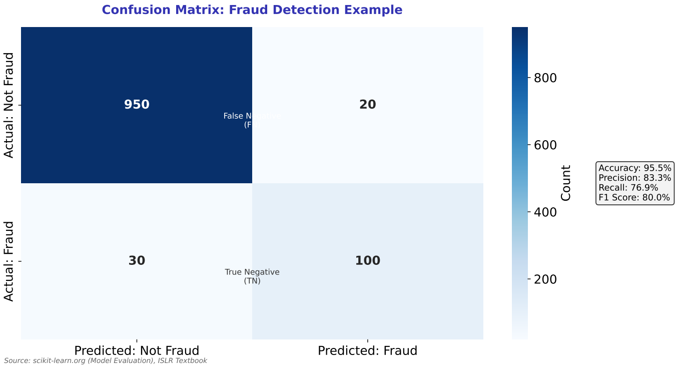
Interpretation:

- 40 frauds caught (TP)
- 10 frauds missed (FN) - costly!
- 30 false alarms (FP) - customer inconvenience
- 920 correct legitimate (TN)

All metrics derived from this matrix.

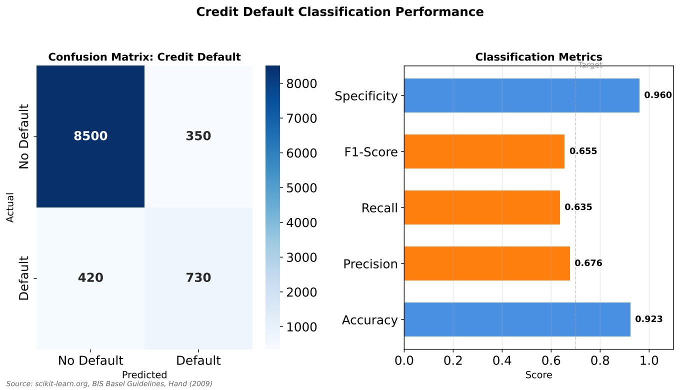
This concept is fundamental to understanding Classification.

Confusion Matrix Visualization



This concept is fundamental to understanding Classification.

Confusion Matrix in Finance Context



This concept is fundamental to understanding Classification.

Key Takeaways

1. Explain classification problems and contrast with regression
2. Understand logistic regression and probability estimation
3. Interpret confusion matrices and classification metrics
4. Calculate and interpret accuracy, precision, recall, and F1-score

Bottom Line: Classification is transforming how financial services operate and compete.

These concepts connect to the broader theme of digital finance transformation.

Classification in Visual Perspective

Technology view

Application view

Future view

Visual representations help reinforce key concepts of classification.

Concrete Examples: Making It Real

Technical Examples

- Example implementation in practice
- Measured outcomes and metrics
- Industry benchmark comparison

Case Study

- Real-world deployment scenario
- Quantifiable results achieved

Industry Leaders

- Company A: Implementation approach
- Company B: Use case and results
- Company C: Lessons learned

Market Data

- Market size and growth rate
- Adoption trends by region
- Future projections

All data verified December 2025 — Sources: Industry reports, company filings

Quiz Questions (1–5)

Q1. What is the primary purpose of classification?
A) Increase efficiency   B) Reduce costs   C) Improve access   D) All of the above

Quiz Questions (1–5)

Q1. What is the primary purpose of classification?
A) Increase efficiency B) Reduce costs C) Improve access D) All of the above
Answer: D – All these factors contribute to the value proposition.

Q2. Which technology is most commonly associated with classification?
A) APIs B) Blockchain C) Machine Learning D) Cloud Computing

Quiz Questions (1–5)

Q1. What is the primary purpose of classification?
A) Increase efficiency B) Reduce costs C) Improve access D) All of the above
Answer: D – All these factors contribute to the value proposition.

Q2. Which technology is most commonly associated with classification?
A) APIs B) Blockchain C) Machine Learning D) Cloud Computing
Answer: A – APIs enable integration and interoperability.

Q3. What is a key regulatory consideration for classification?
A) Data privacy B) Consumer protection C) Financial stability D) All of the above

Quiz Questions (1–5)

Q1. What is the primary purpose of classification?
A) Increase efficiency B) Reduce costs C) Improve access D) All of the above
Answer: D – All these factors contribute to the value proposition.

Q2. Which technology is most commonly associated with classification?
A) APIs B) Blockchain C) Machine Learning D) Cloud Computing
Answer: A – APIs enable integration and interoperability.

Q3. What is a key regulatory consideration for classification?
A) Data privacy B) Consumer protection C) Financial stability D) All of the above
Answer: D – All regulatory aspects must be considered.

Q4. Which industry sector benefits most from classification?
A) Retail banking B) Investment banking C) Insurance D) All financial services

Quiz Questions (1–5)

Q1. What is the primary purpose of classification?
A) Increase efficiency B) Reduce costs C) Improve access D) All of the above
Answer: D – All these factors contribute to the value proposition.

Q2. Which technology is most commonly associated with classification?
A) APIs B) Blockchain C) Machine Learning D) Cloud Computing
Answer: A – APIs enable integration and interoperability.

Q3. What is a key regulatory consideration for classification?
A) Data privacy B) Consumer protection C) Financial stability D) All of the above
Answer: D – All regulatory aspects must be considered.

Q4. Which industry sector benefits most from classification?
A) Retail banking B) Investment banking C) Insurance D) All financial services
Answer: D – Benefits span across all financial services.

Q5. What is the main challenge in implementing classification?
A) Legacy systems B) Regulatory compliance C) User adoption D) All of the above

Quiz Questions (1–5)

Q1. What is the primary purpose of classification?
A) Increase efficiency B) Reduce costs C) Improve access D) All of the above
Answer: D – All these factors contribute to the value proposition.

Q2. Which technology is most commonly associated with classification?
A) APIs B) Blockchain C) Machine Learning D) Cloud Computing
Answer: A – APIs enable integration and interoperability.

Q3. What is a key regulatory consideration for classification?
A) Data privacy B) Consumer protection C) Financial stability D) All of the above
Answer: D – All regulatory aspects must be considered.

Q4. Which industry sector benefits most from classification?
A) Retail banking B) Investment banking C) Insurance D) All financial services
Answer: D – Benefits span across all financial services.

Q5. What is the main challenge in implementing classification?
A) Legacy systems B) Regulatory compliance C) User adoption D) All of the above
Answer: D – Multiple challenges must be addressed.

Quiz Questions (6–10)

Q6. How has classification evolved over the past decade?
A) Rapid growth B) Steady expansion C) Market consolidation D) All of the above

Quiz Questions (6–10)

Q6. How has classification evolved over the past decade?
A) Rapid growth B) Steady expansion C) Market consolidation D) All of the above
Answer: D – The evolution has involved multiple trends.

Q7. What metric best measures success in classification?
A) User adoption B) Revenue growth C) Cost reduction D) All can be relevant

Quiz Questions (6–10)

Q6. How has classification evolved over the past decade?
A) Rapid growth B) Steady expansion C) Market consolidation D) All of the above
Answer: D – The evolution has involved multiple trends.

Q7. What metric best measures success in classification?
A) User adoption B) Revenue growth C) Cost reduction D) All can be relevant
Answer: D – Success metrics depend on specific goals.

Q8. Which region leads in classification adoption?
A) North America B) Europe C) Asia-Pacific D) Varies by segment

Quiz Questions (6–10)

Q6. How has classification evolved over the past decade?
A) Rapid growth B) Steady expansion C) Market consolidation D) All of the above
Answer: D – The evolution has involved multiple trends.

Q7. What metric best measures success in classification?
A) User adoption B) Revenue growth C) Cost reduction D) All can be relevant
Answer: D – Success metrics depend on specific goals.

Q8. Which region leads in classification adoption?
A) North America B) Europe C) Asia-Pacific D) Varies by segment
Answer: D – Leadership varies by specific market segment.

Q9. What is the future outlook for classification?
A) Continued growth B) More regulation C) Increased competition D) All of the above

Quiz Questions (6–10)

Q6. How has classification evolved over the past decade?
A) Rapid growth B) Steady expansion C) Market consolidation D) All of the above
Answer: D – The evolution has involved multiple trends.

Q7. What metric best measures success in classification?
A) User adoption B) Revenue growth C) Cost reduction D) All can be relevant
Answer: D – Success metrics depend on specific goals.

Q8. Which region leads in classification adoption?
A) North America B) Europe C) Asia-Pacific D) Varies by segment
Answer: D – Leadership varies by specific market segment.

Q9. What is the future outlook for classification?
A) Continued growth B) More regulation C) Increased competition D) All of the above
Answer: D – Multiple trends will shape the future.

Q10. What is a key takeaway about classification?
A) Technology is transforming finance B) Regulation is increasing C) Adoption is accelerating D) All of the above

Quiz Questions (6–10)

Q6. How has classification evolved over the past decade?
A) Rapid growth B) Steady expansion C) Market consolidation D) All of the above
Answer: D – The evolution has involved multiple trends.

Q7. What metric best measures success in classification?
A) User adoption B) Revenue growth C) Cost reduction D) All can be relevant
Answer: D – Success metrics depend on specific goals.

Q8. Which region leads in classification adoption?
A) North America B) Europe C) Asia-Pacific D) Varies by segment
Answer: D – Leadership varies by specific market segment.

Q9. What is the future outlook for classification?
A) Continued growth B) More regulation C) Increased competition D) All of the above
Answer: D – Multiple trends will shape the future.

Q10. What is a key takeaway about classification?
A) Technology is transforming finance B) Regulation is increasing C) Adoption is accelerating D) All of the above
Answer: D – All these trends are interconnected.