Linear Discriminant Analysis Theory

Classification Methods

Daniel E. Acuna
Associate Professor, University of Colorado Boulder

Contents of This Video

In this video, we will cover:

- LDA approach to student classification
- Gaussian assumptions for each class
- Linear decision boundaries
- Mathematical foundations of LDA
- When LDA works well vs. its limitations
- Comparison with logistic regression

LDA Approach to Classification

Core Idea: Model each class (pass/fail) as a Gaussian distribution, then find the optimal boundary

Student Success Context:

- Pass students: One Gaussian “cloud”
- Fail students: Another Gaussian “cloud”
- Same shape, different centers
- Find line that best separates the clouds

Key Advantage: Optimal when assumptions hold

Gaussian Assumptions: 1D Example

For Pass Students:

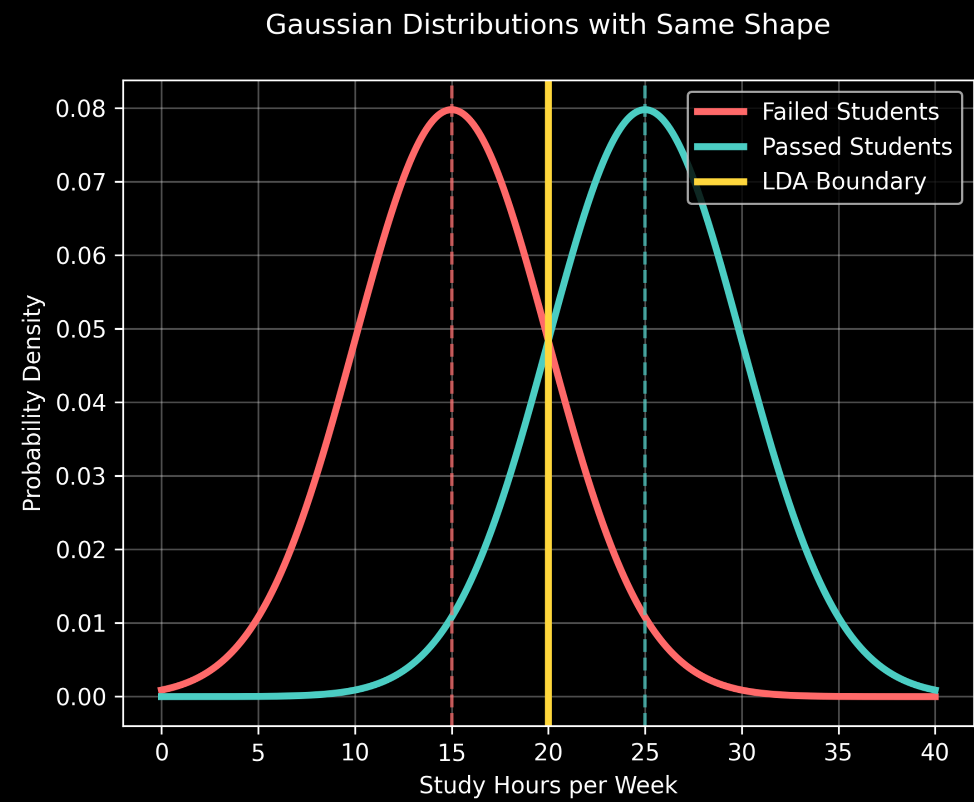
$$\mathbf{x} \sim \square(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$$

For Fail Students:

$$\mathbf{x} \sim \square(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$$

Key Assumption: Same covariance matrix $\boldsymbol{\Sigma}$ for both classes

Simple 1D Case: Same variance σ^2, different means

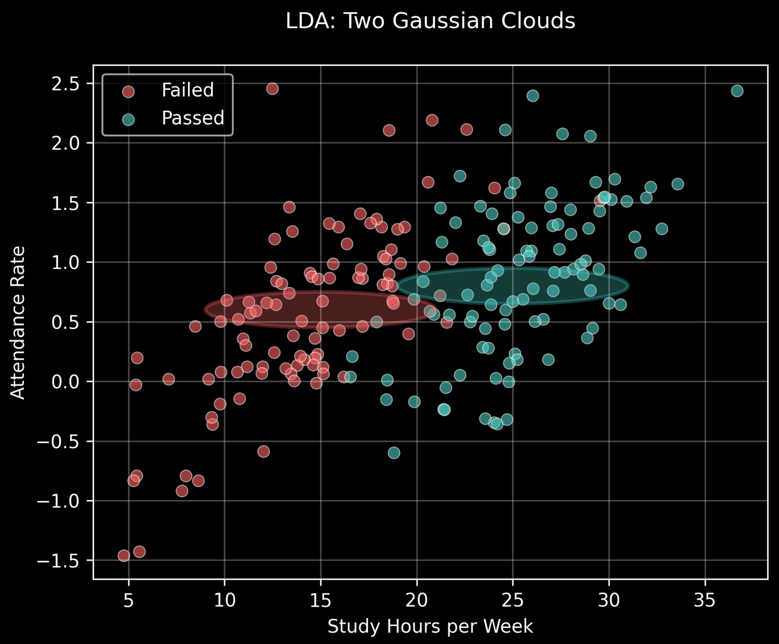


2D Gaussian Clouds

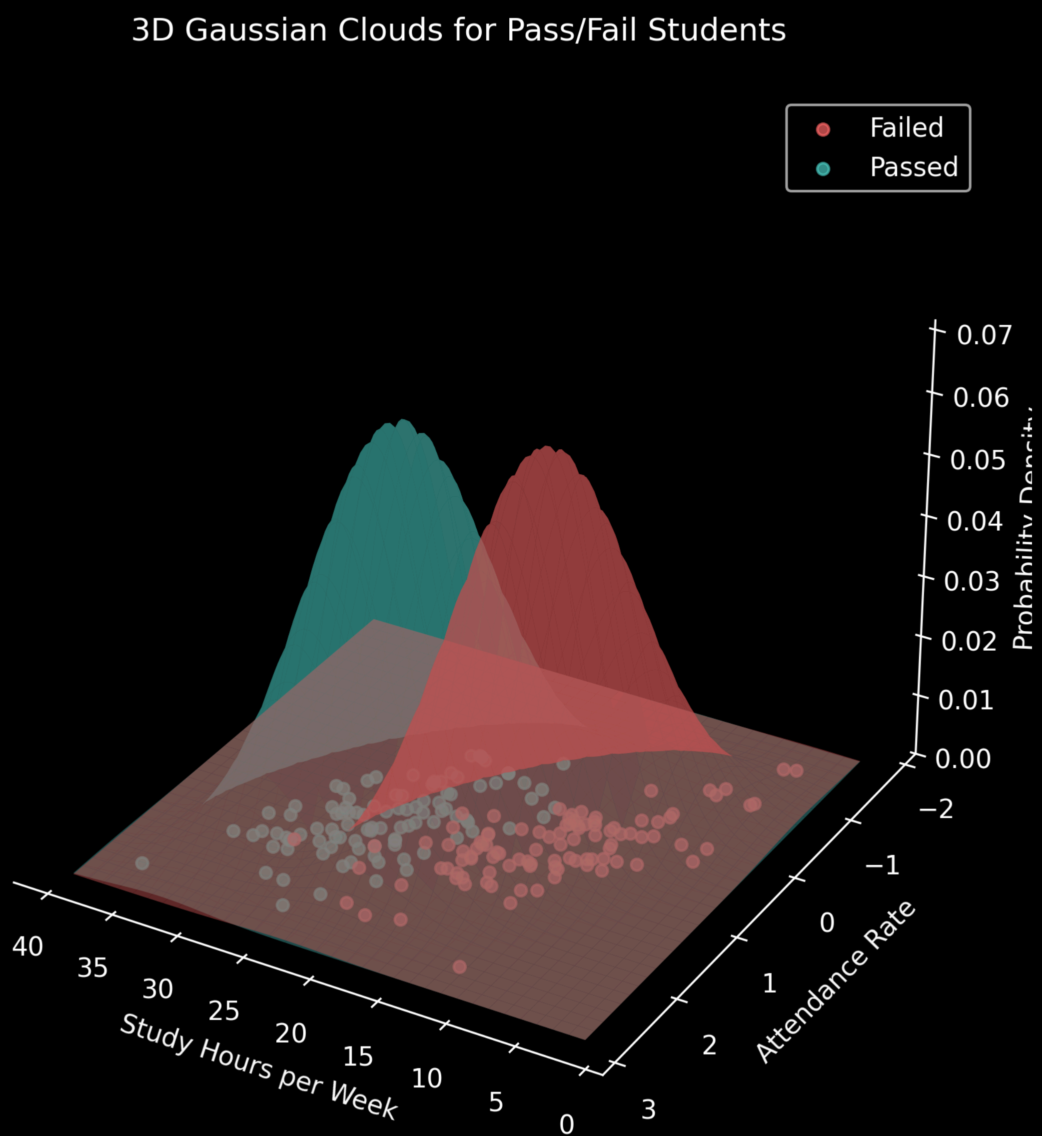
Extending to 2D:

- Pass students: One Gaussian "cloud"
- Fail students: Another Gaussian "cloud"
- Same shape, different centers
- Find line that best separates the clouds

Real Student Features: - Study hours per week - Attendance rate - Same covariance structure for both groups



3D Gaussian Clouds for Pass/Fail Students



Gaussian Assumptions

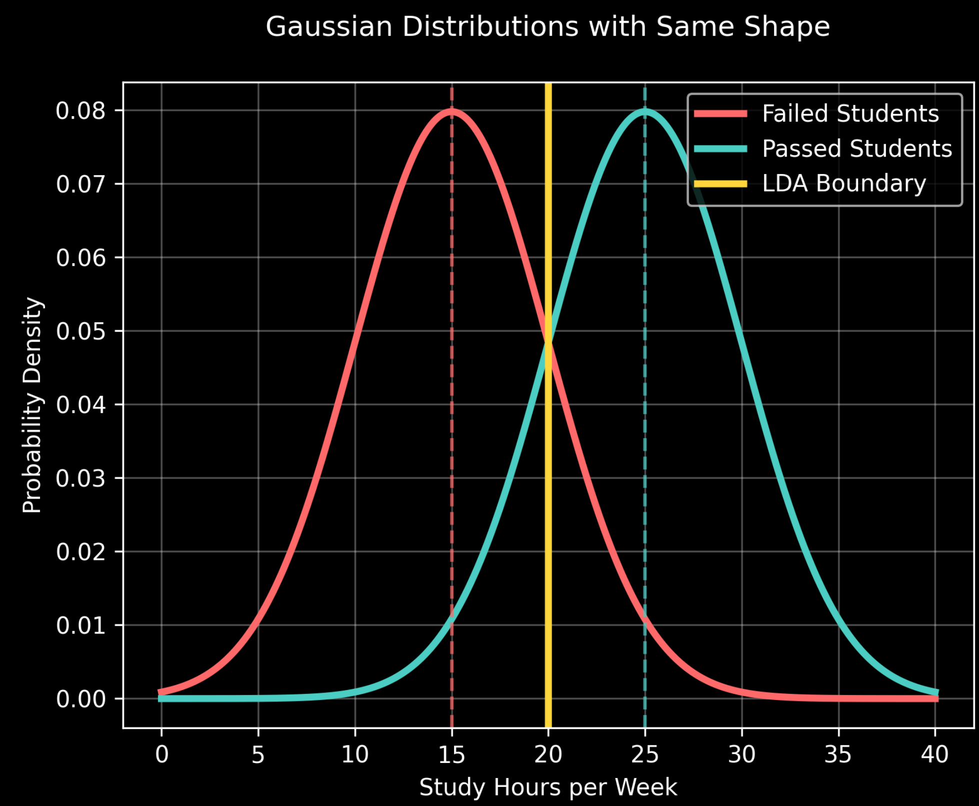
For Pass Students:

$$\mathbf{x} \sim \square(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$$

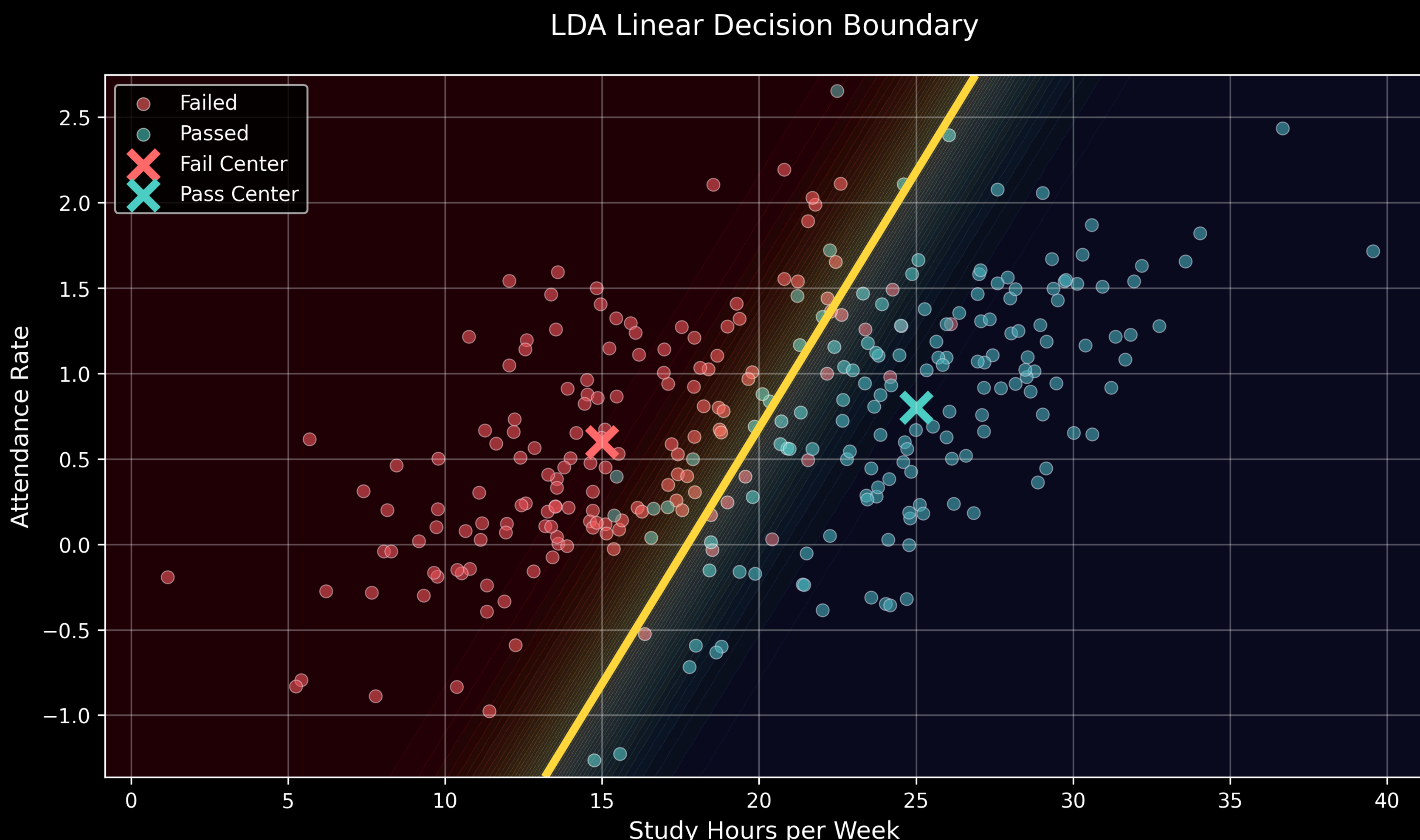
For Fail Students:

$$\mathbf{x} \sim \square(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$$

Key Assumption: Same covariance matrix $\boldsymbol{\Sigma}$ for both classes



Linear Decision Boundary



Mathematical Foundation

LDA Decision Rule

For a new student with features $\mathbf{x}$, predict:

- **Pass** if: $\delta_1(\mathbf{x}) > \delta_2(\mathbf{x})$
- **Fail** if: $\delta_2(\mathbf{x}) > \delta_1(\mathbf{x})$

Where:

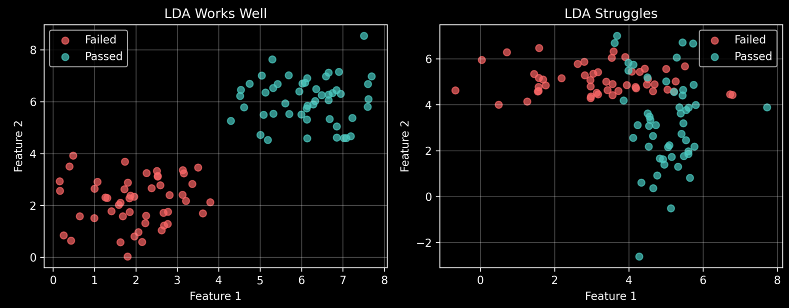
$$\delta_k(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k - \frac{1}{2} \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + \log(\pi_k)$$

- $\boldsymbol{\mu}_k$ = mean of class k
- $\boldsymbol{\Sigma}$ = shared covariance matrix
- π_k = prior probability of class k

When LDA Works Well

Ideal Conditions:

- Classes roughly follow Gaussian distributions
- Similar variability between pass/fail students
- Well-separated class centers
- Small to moderate number of features
- Limited training data (LDA is efficient)



Student Success Examples:

- Standardized test scores and grades
- Consistent measurement scales
- Similar student populations

Limitations of LDA

LDA vs. Logistic Regression

Aspect	LDA	Logistic Regression
Approach	Generative (models classes)	Discriminative (models boundary)
Assumptions	Gaussian classes, equal covariance	Minimal assumptions
Decision Boundary	Always linear	Linear (with regularization)
Training Data	Efficient with small datasets	Needs more data typically
Robustness	Sensitive to assumptions	More robust to violations
Interpretability	Class centers meaningful	Coefficients show feature importance

What We've Covered

In this video, we've explored:

- LDA's Gaussian modeling approach for each class
- Equal covariance assumption and its implications
- Mathematical foundation of linear decision boundaries
- When LDA works well vs. its limitations
- Comparison with logistic regression approaches
- Real-world considerations for student classification