L2 Regularization in Logistic Regression

Classification Methods

Daniel E. Acuna
Associate Professor, University of Colorado Boulder

Contents of This Video

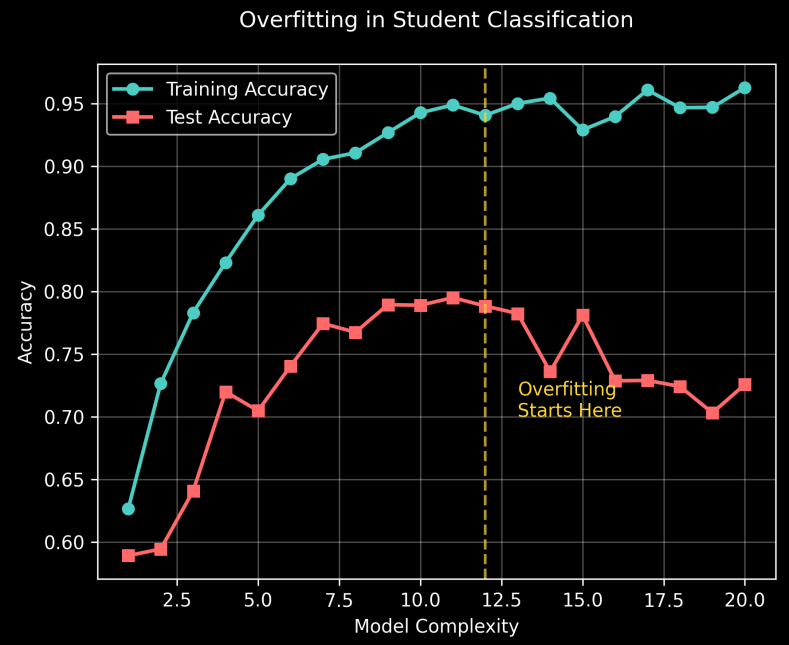
In this video, we will cover:

- The overfitting problem in student classification
- What is L2 regularization?
- Mathematical formulation of regularized loss
- Effect of regularization parameter λ (lambda)
- Visualizing smooth vs. complex decision boundaries

The Overfitting Problem

Student Success Prediction Challenge:

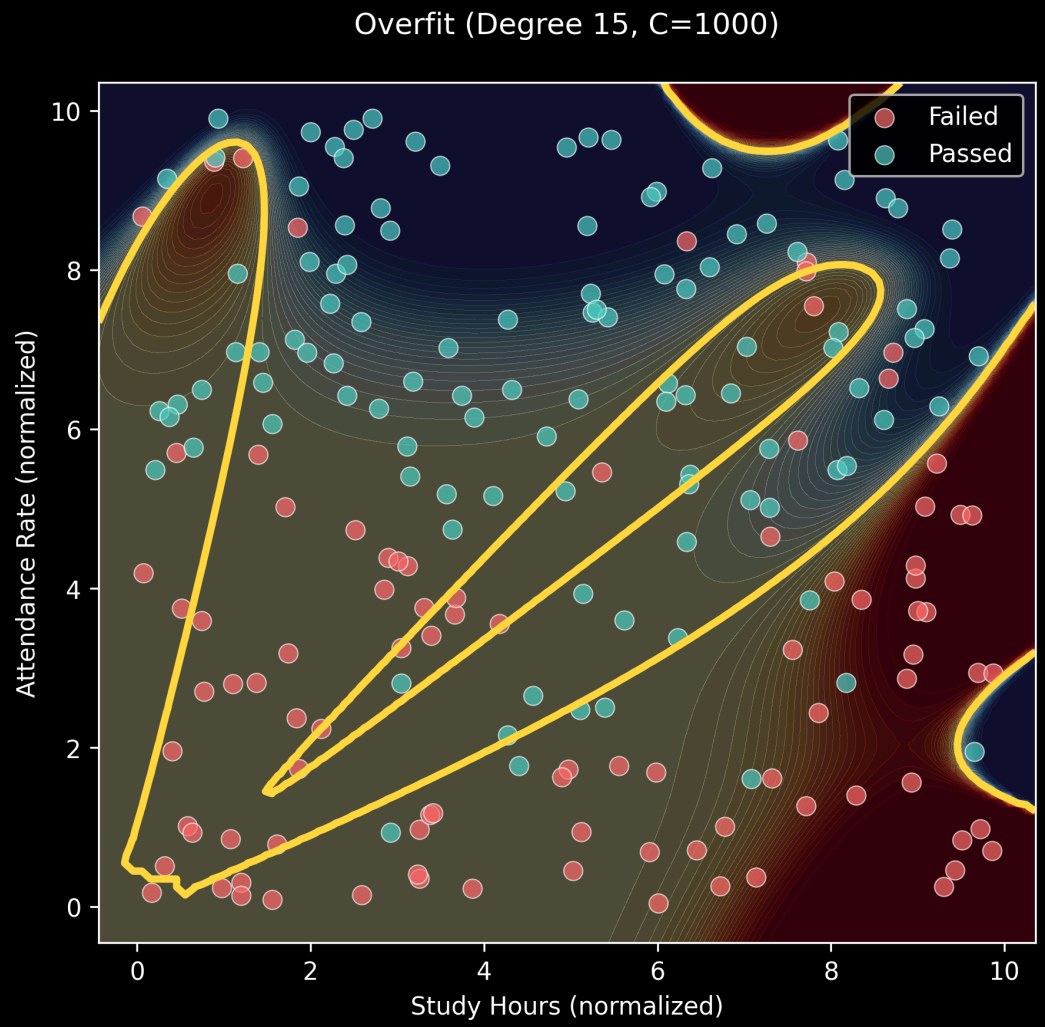
- Many features: study hours, attendance, assignments, GPA, sleep, job status, etc.
- Complex patterns in training data
- Model might memorize noise instead of true patterns

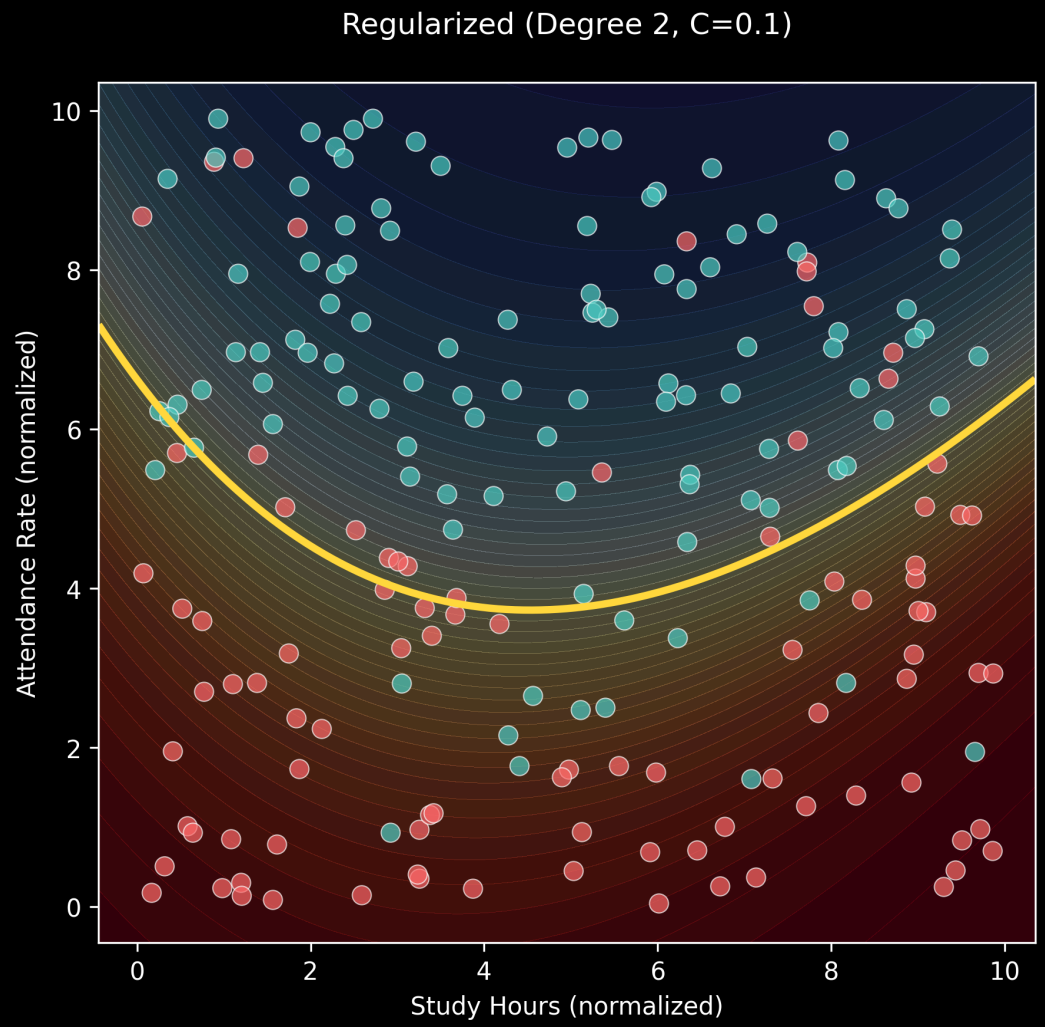


Overfitting Signs:

- Perfect accuracy on training students
- Poor performance on new students
- Overly complex decision boundaries

Complex vs. Simple Decision Boundaries





L2 Regularization Mathematical Formulation

Original Loss Function:

$$J(\mathbf{w}, b) = -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log(y^{(\hat{i})}) + (1 - y^{(i)}) \log(1 - y^{(\hat{i})}) \right]$$

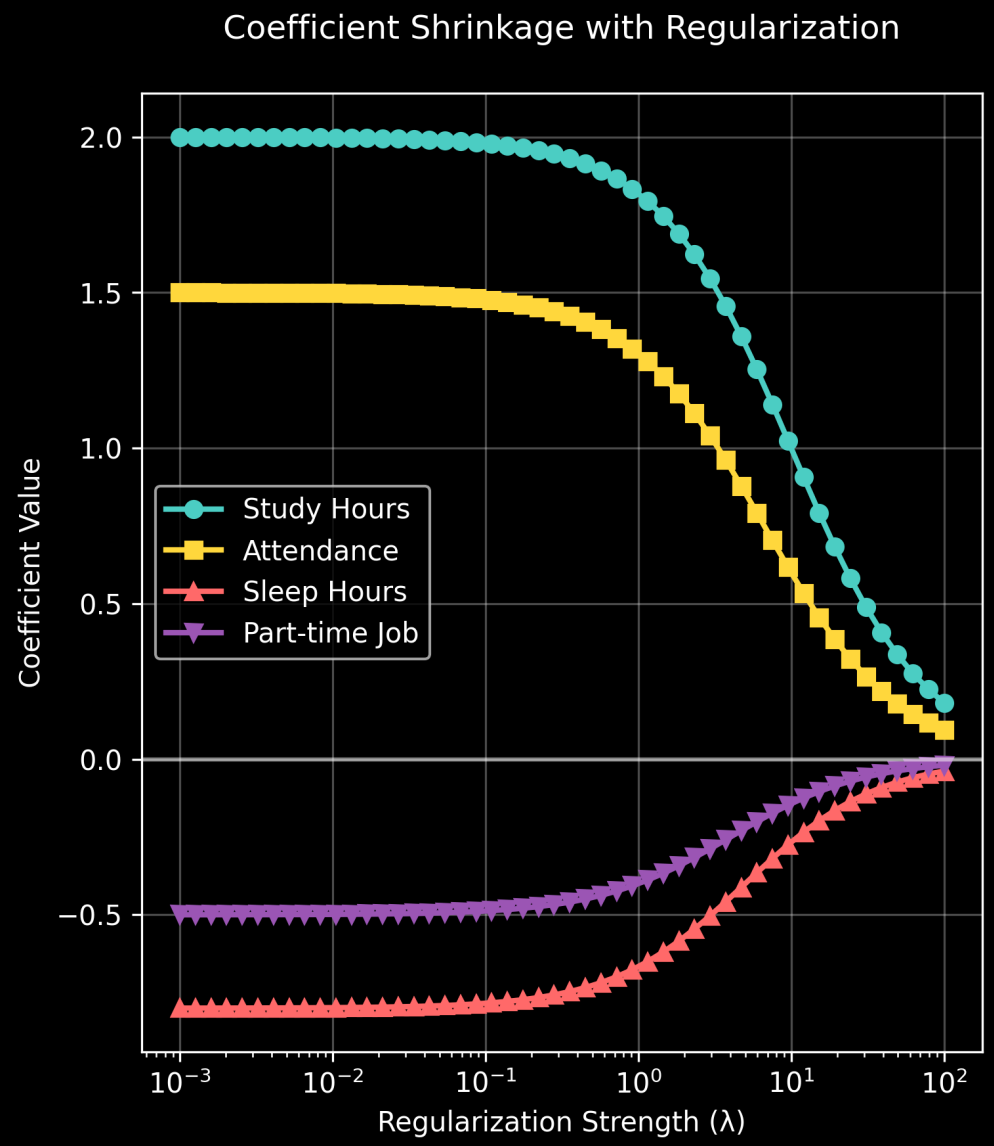
Regularized Loss Function:

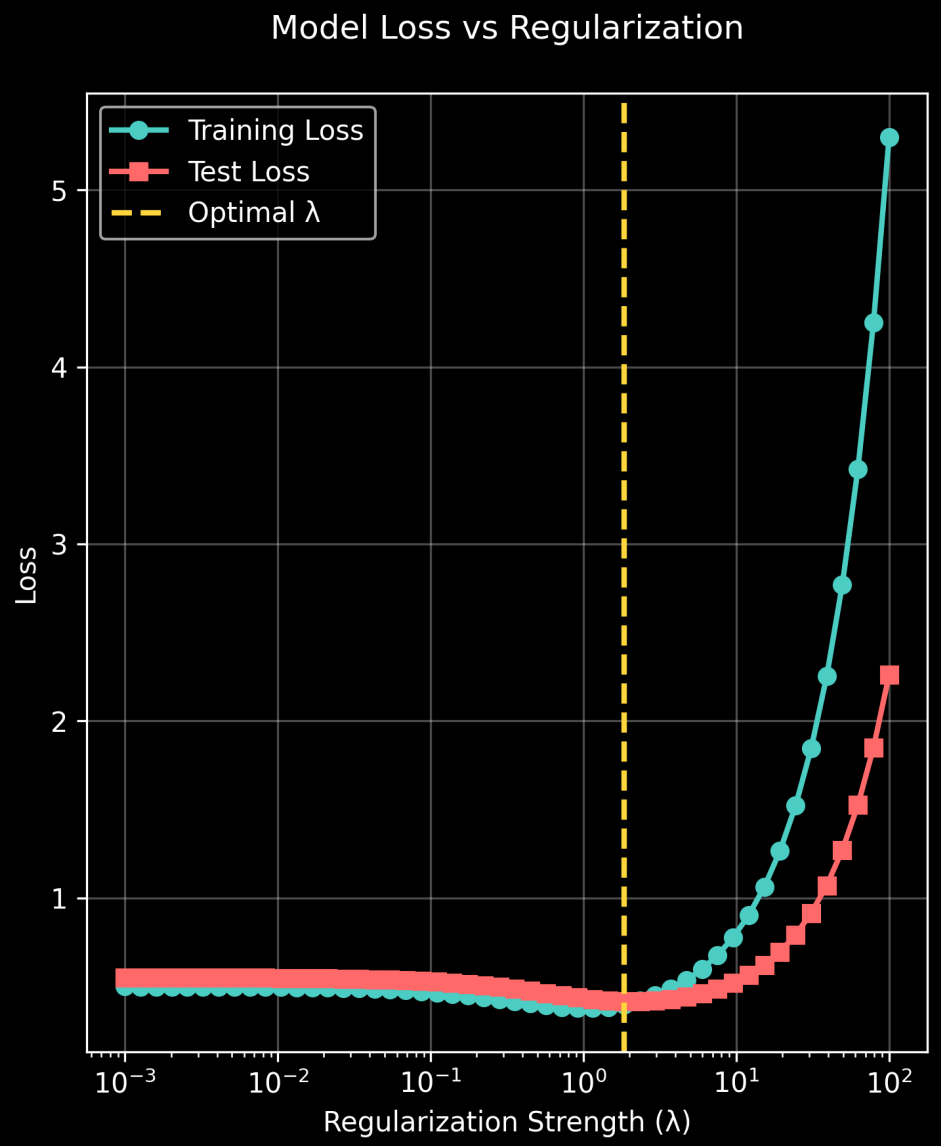
$$J_{reg}(\mathbf{w}, b) = J(\mathbf{w}, b) + \lambda \sum_{j=1}^{p} w_j^2$$

Where:

- λ (lambda) = regularization strength

Effect of Regularization Parameter λ





Regularization Parameter Values

Common Parameter Notation:

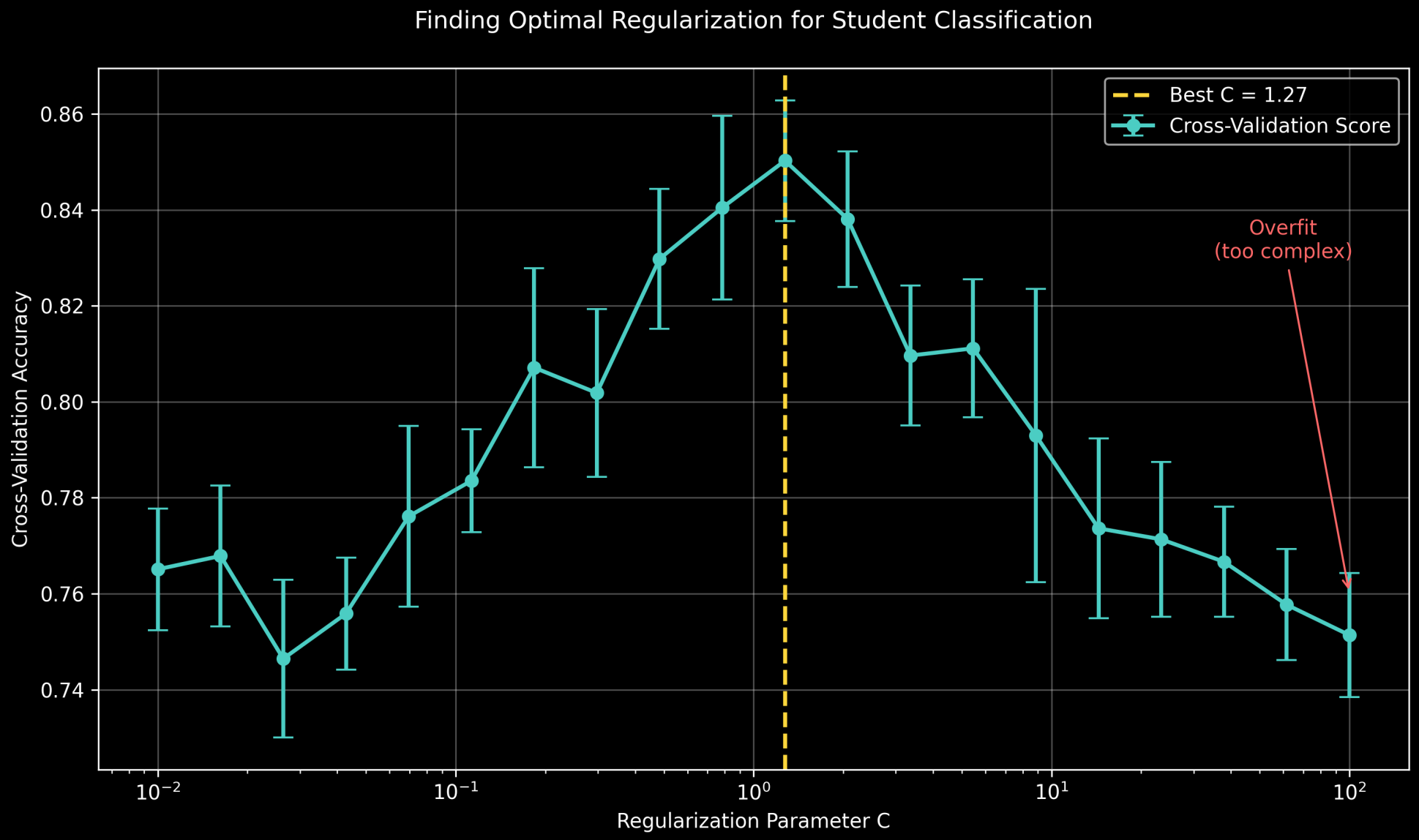
- **λ (lambda)**: Direct regularization strength
- **C**: Inverse regularization (C = 1/λ)

Effect on Student Classification:

- **Large λ (small C)**: Simple model, may underfit
- **Small λ (large C)**: Complex model, may overfit
- **Optimal λ**: Best balance for new students

λ Value	C Value	Model Behavior
0.001	1000	Very complex
0.1	10	Moderate
1.0	1	Balanced
10	0.1	Simple
100	0.01	Very simple

Cross-Validation for Parameter Selection



What We've Covered

In this video, we've explored:

- Overfitting problems with complex student data
- L2 regularization as a solution to overfitting
- Mathematical formulation of regularized loss function
- Effect of regularization parameter λ on model complexity
- Cross-validation for optimal parameter selection