K-Nearest Neighbors Intuition

Classification Methods

Daniel E. Acuna
Associate Professor, University of Colorado Boulder

Contents of This Video

In this video, we will cover:

- The KNN classification approach
- Finding similar students with distance measures
- Effect of k on decision boundaries
- Bias-variance tradeoff in KNN
- Curse of dimensionality
- Strengths and limitations for student prediction

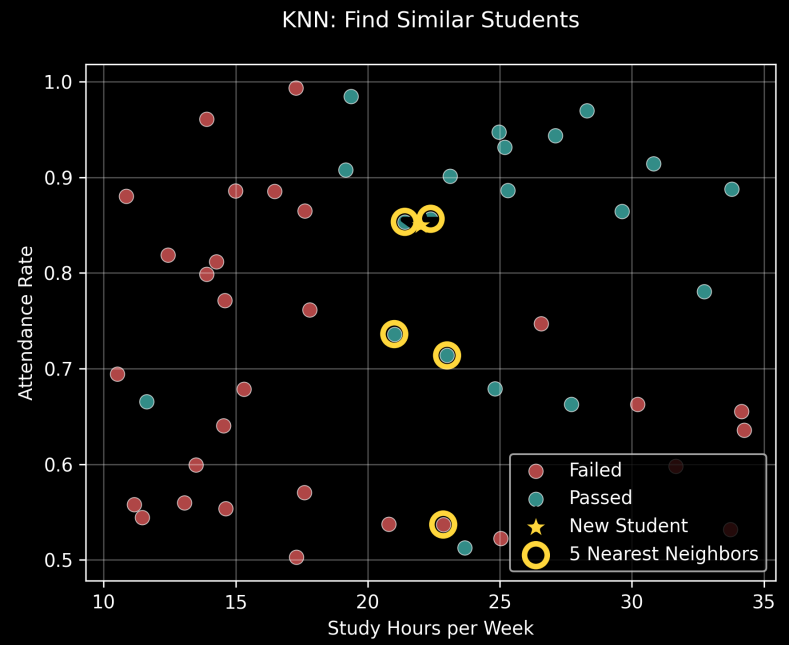
The KNN Approach

Core Idea: Classify a new student based on the outcomes of their k most similar students

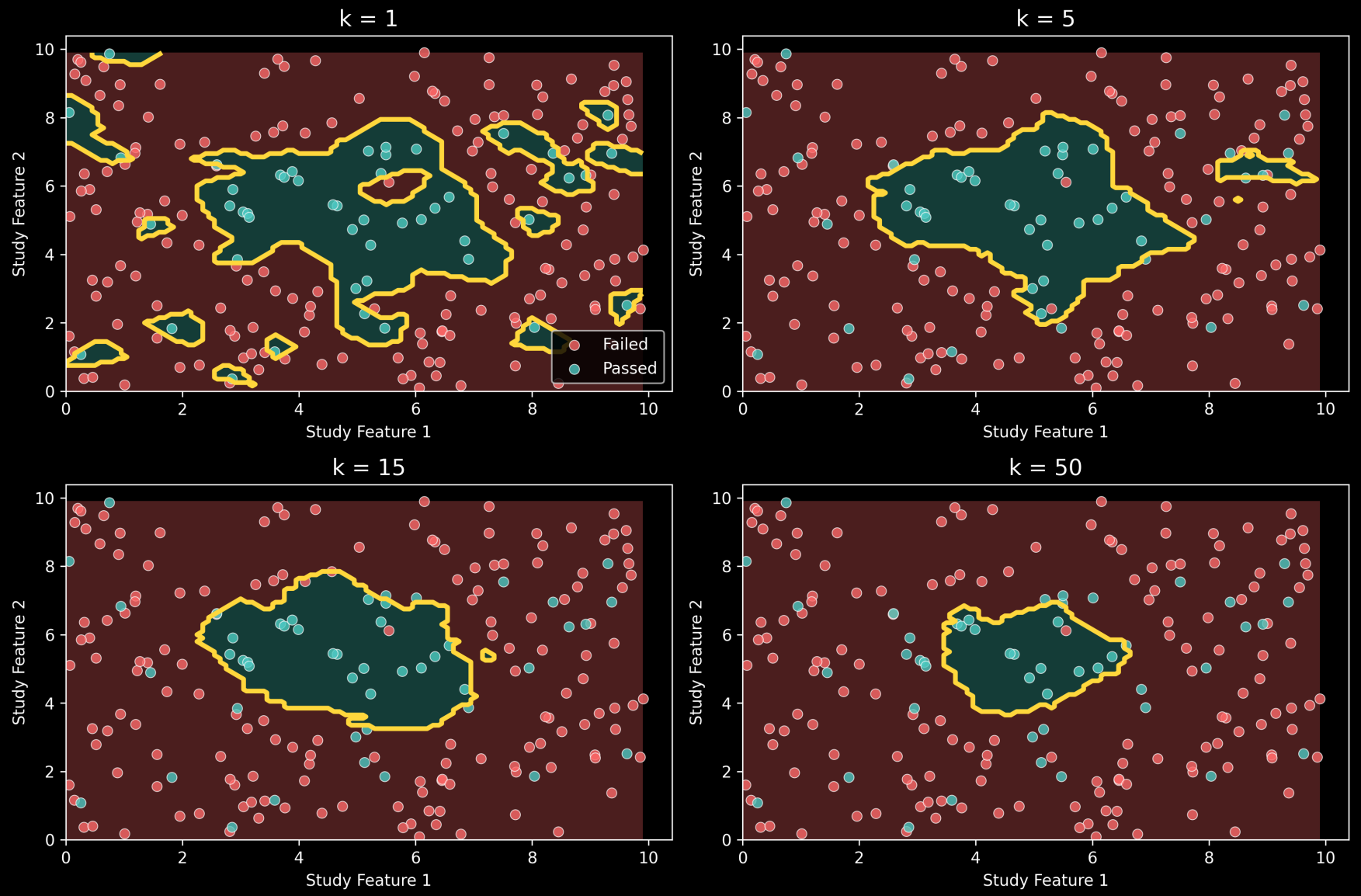
Student Success Example:

- New student: 22 study hours, 85% attendance, 8 assignments completed
- Find k=5 most similar students in database
- Check how many passed vs failed
- Predict based on majority vote

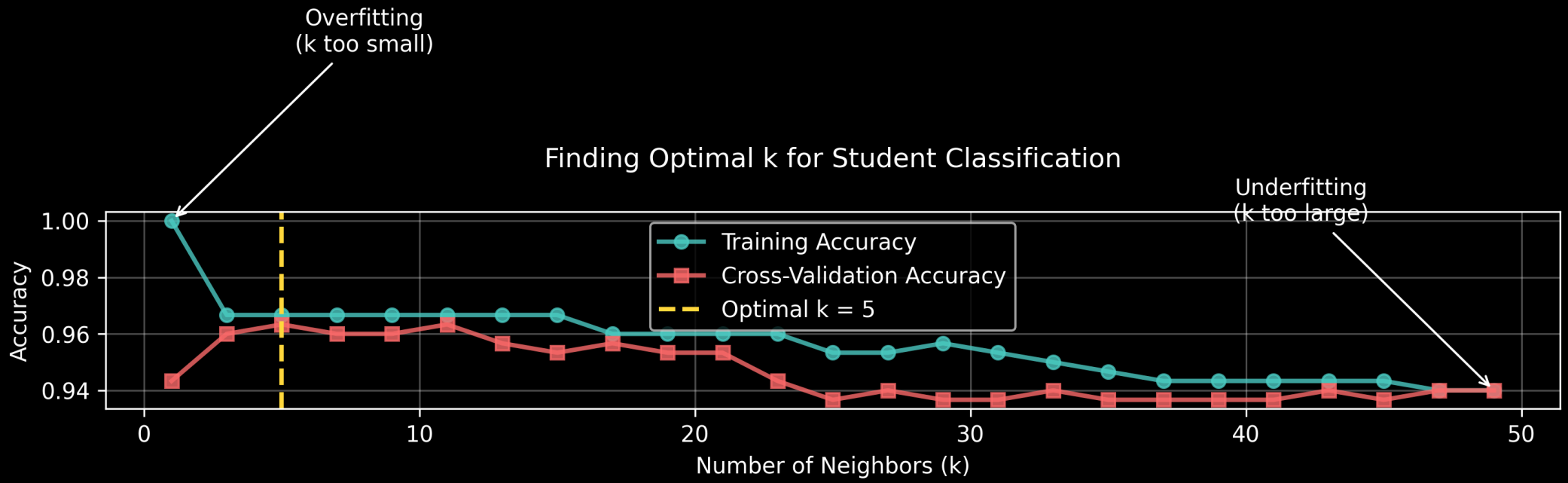
No Training Required! Just store data and compute at prediction time



Effect of k on Predictions



Choosing the Right k

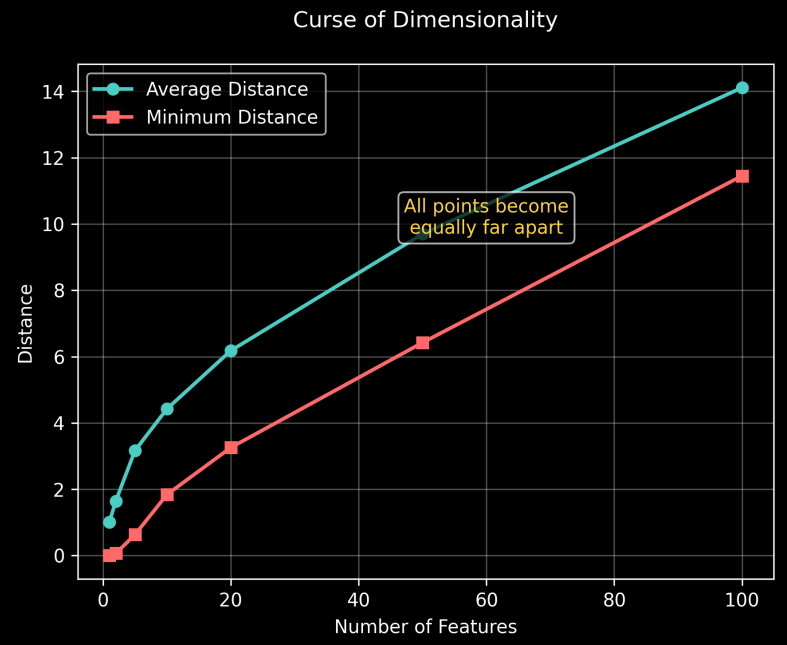


Curse of Dimensionality

The Problem: As the number of features increases, all students become equally “far” from each other

Student Features Example: - Study hours, attendance, assignments, GPA, sleep, job, major, library time, tutoring, forum participation, study space, study time preference, practice exams…

Result: - Distances become meaningless - No clear “nearest” neighbors - KNN performance degrades



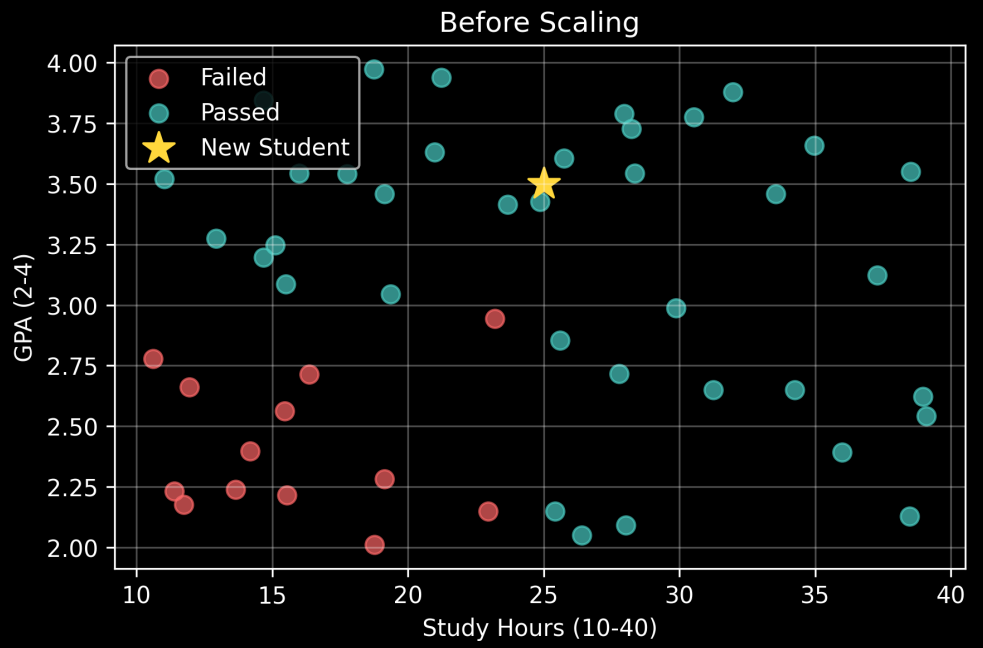
Strengths of KNN

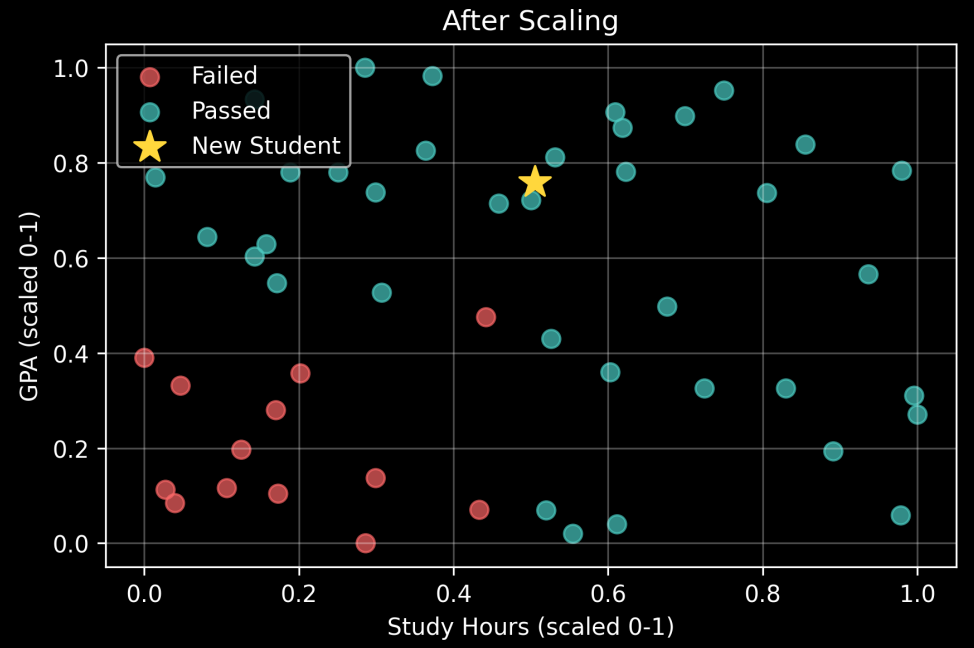
Limitations of KNN

Key Disadvantages:

- **Computational cost:** Slow for large datasets
- **Storage requirements:** Must keep all training data
- **Sensitive to irrelevant features:** All features treated equally
- **Curse of dimensionality:** Poor with many features
- **Imbalanced data:** Majority class dominates

Feature Scaling Critical: Hours (0-40) vs GPA (0-4) vs Attendance (0-1)





What We've Covered

In this video, we've explored:

- KNN's similarity-based classification approach
- Distance measures for finding nearest neighbors
- Effect of k parameter on model complexity
- Bias-variance tradeoff in neighbor selection
- Curse of dimensionality challenges
- Strengths and limitations for student classification