

Discovery Worksheet

Introduction to Machine Learning & AI Pattern Recognition Through Visual Analysis

Machine Learning for Smarter Innovation Week 0: Pre-Lecture Activity

Instructions

Complete this worksheet before attending the lecture.

Approach:

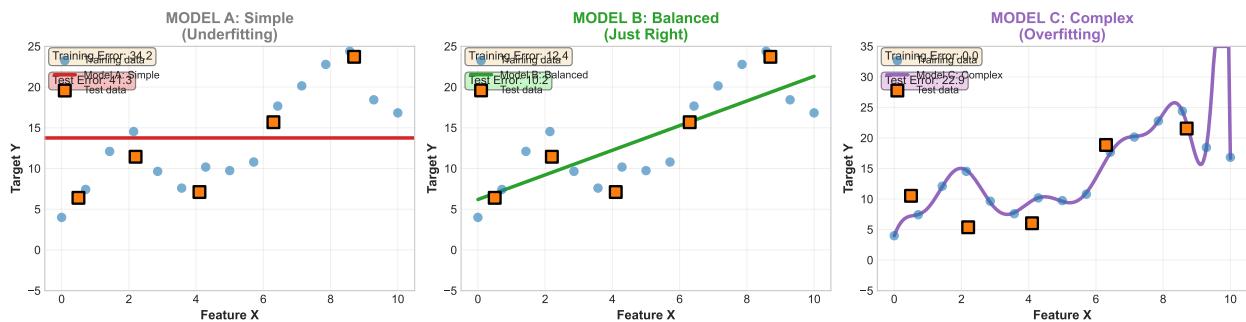
- Examine each chart carefully
- Answer questions based on observed patterns
- Focus on conceptual understanding rather than precise calculations
- Compare observations with a partner when possible

Time needed: 45-55 minutes

Materials: Pencil, calculator (optional)

Learning approach: This worksheet uses visual charts to help identify patterns in machine learning algorithms. The lecture will provide formal terminology and mathematical frameworks for the patterns observed here.

Chart 1: Model Complexity



Observation

Three models predict house prices from square footage. Blue dots represent training data. Orange squares represent new test data the models have not seen.

Tasks

Task 1: Rank model complexity

Rank the three models from simplest to most complex:

Simplest: _____ → _____ → Most complex: _____

Task 2: Identify performance pattern

Which model performs worst on the new orange test points?

Answer: _____

Why does this model struggle with new data?

Task 3: Explain the trade-off

Model C fits the training data perfectly (zero error), yet performs poorly on test data. What is the risk of creating a model that fits training data too closely?

Task 4: Optimal complexity

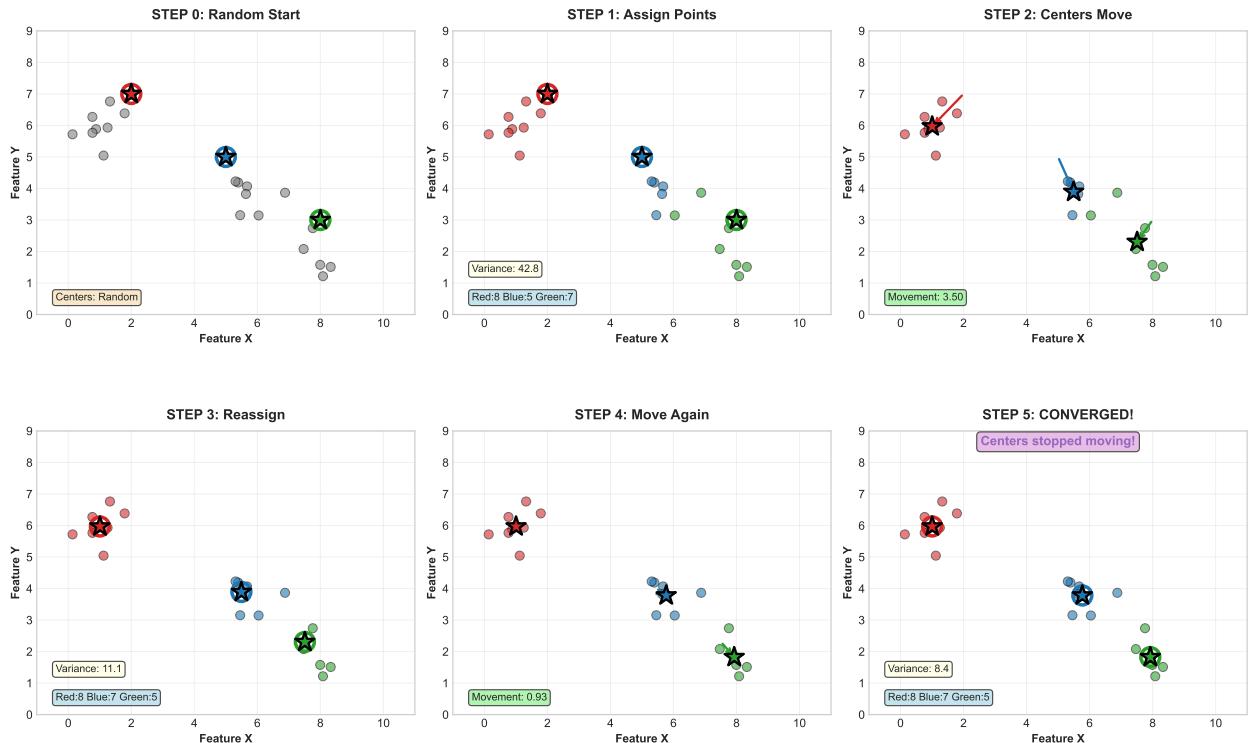
Which model achieves the best balance between simplicity and accuracy?

Answer: _____

What makes this model better than the other two?

Summary: A model that fits training data perfectly may fail on new data. The optimal model balances simplicity with predictive accuracy. This trade-off is fundamental to machine learning.

Chart 2: Clustering Algorithm



Observation

The chart shows six steps of an algorithm that groups data points into clusters. Centers are marked with stars.

Tasks

Task 1: Describe center movement

How do centers move from Step 1 to Step 2?

Task 2: Observe point-to-center distances

From Step 0 to Step 5, what happens to the distance between points and their assigned centers?

Distances: _____ (increasing/decreasing/staying same)

What does this suggest about the algorithm's objective?

Task 3: Identify assignment rule

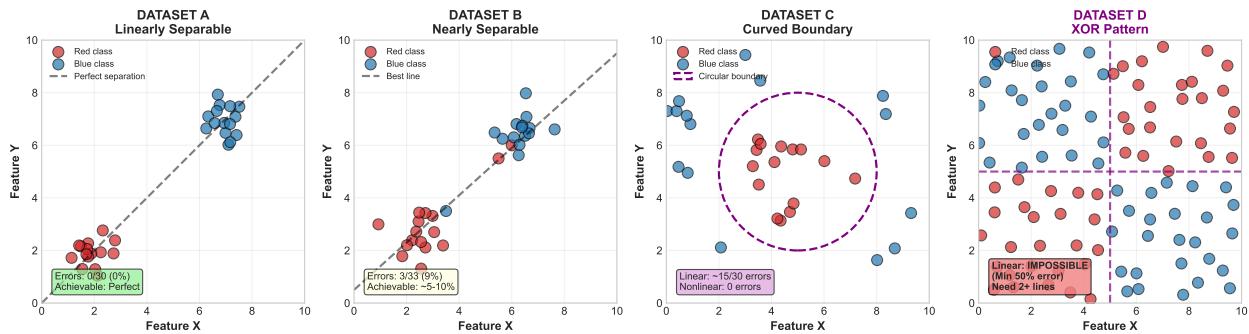
How does the algorithm decide which cluster a point belongs to?

Task 4: Detect convergence

How can the algorithm determine when to stop iterating?

Summary: The algorithm alternates between two steps: assigning points to the nearest center, then moving centers to the average position of their assigned points. This process minimizes the total distance from points to their cluster centers.

Chart 3: Classification Boundaries



Observation

Four datasets show red and blue points. The goal is to separate the two colors.

Tasks

Task 1: Attempt linear separation

Using a ruler or mentally, try to draw a straight line that separates red from blue in each dataset.
Which datasets allow perfect or near-perfect separation with a straight line?

Possible: _____

Task 2: Identify impossible cases

Which dataset(s) cannot be separated by any straight line?

Answer: _____

What pattern in the data makes linear separation impossible?

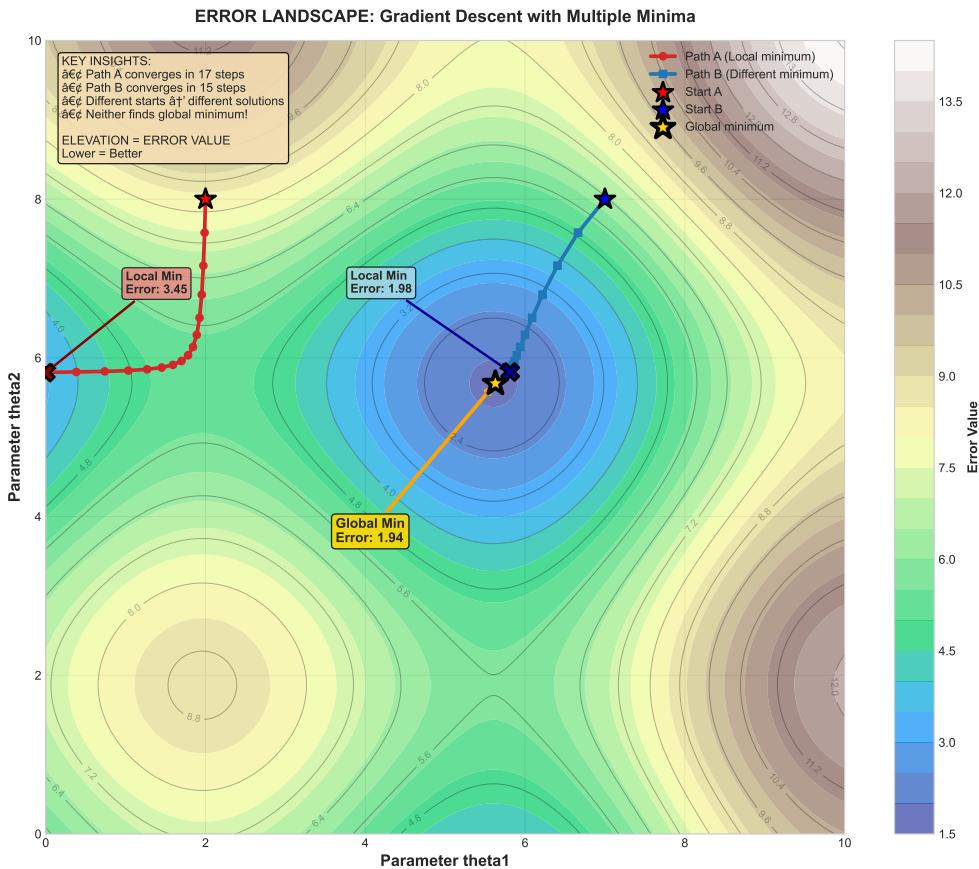
Task 3: Propose alternative solutions

For Dataset D (the impossible case), what alternative approach could work?

Optional challenge: Can you prove mathematically that Dataset D is impossible for a single line?

Summary: Linear models work when data classes can be separated by a straight line or plane. Some patterns require more complex decision boundaries. This limitation motivates the development of nonlinear models such as neural networks.

Chart 4: Optimization Paths



Observation

This topographic map shows an error landscape where elevation represents error value. Two paths descend from different starting points.

Tasks

Task 1: Compare final destinations

Do Path A (red) and Path B (blue) end in the same valley?

Answer: _____

What does this tell us about the importance of starting position?

Task 2: Identify the better solution

Which path finds the valley with lower error?

Answer: _____

Is this the global minimum (lowest point on the entire landscape)?

Answer: _____

Task 3: Analyze step size effects

What would happen if steps were too large?

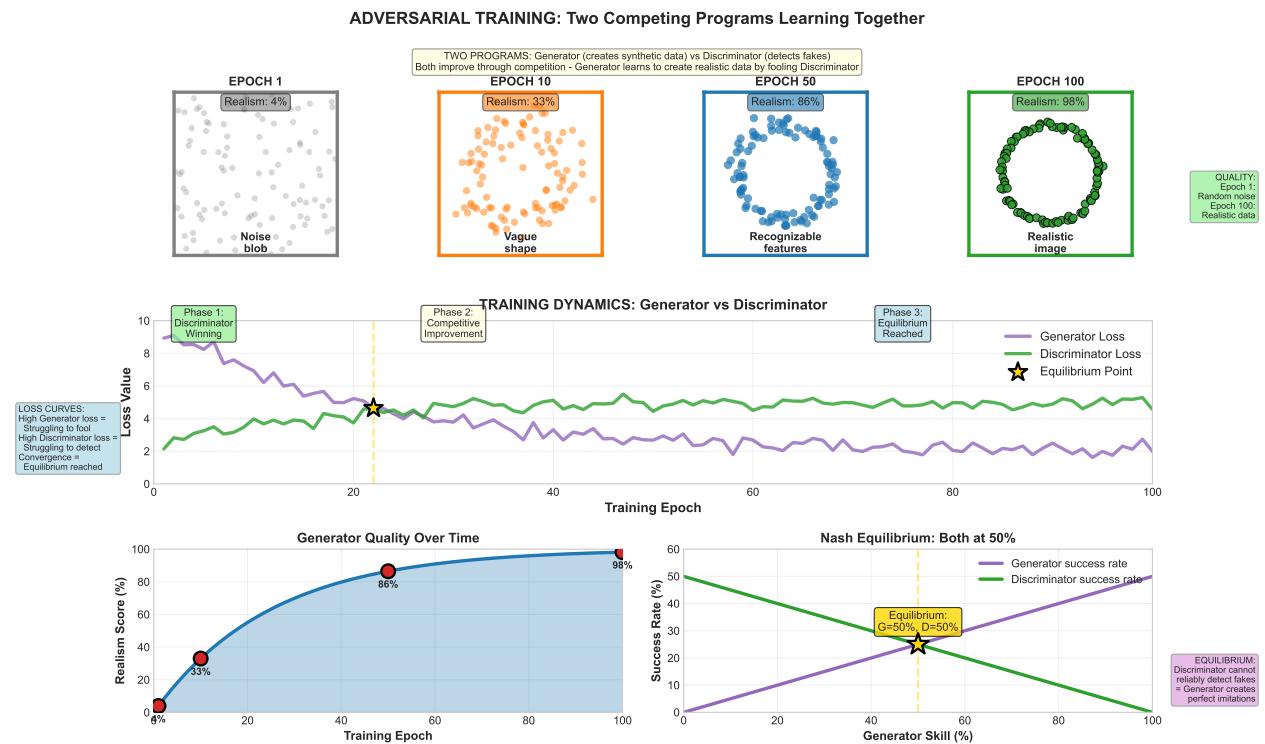
What would happen if steps were too small?

Task 4: Propose improvement strategies

How might the algorithm escape from a suboptimal local minimum?

Summary: Optimization algorithms follow the gradient downhill to find error minima. Different starting points can lead to different solutions. The learning rate (step size) must balance speed with stability. Complex landscapes may contain multiple local minima.

Chart 5: Adversarial Training



Observation

The chart shows two programs competing over 100 training epochs. The Generator creates synthetic data while the Discriminator attempts to detect fakes.

Tasks

Task 1: Track quality progression

Describe how generated samples change from Epoch 1 to Epoch 100.

Task 2: Analyze competitive dynamics

Examine the middle training dynamics panel showing loss curves.

In Phase 1 (Epochs 1-10), which program has the advantage?

Answer: _____

In Phase 3 (Epochs 50-100), what happens to both loss curves?

Answer: _____

Task 3: Explain mutual improvement

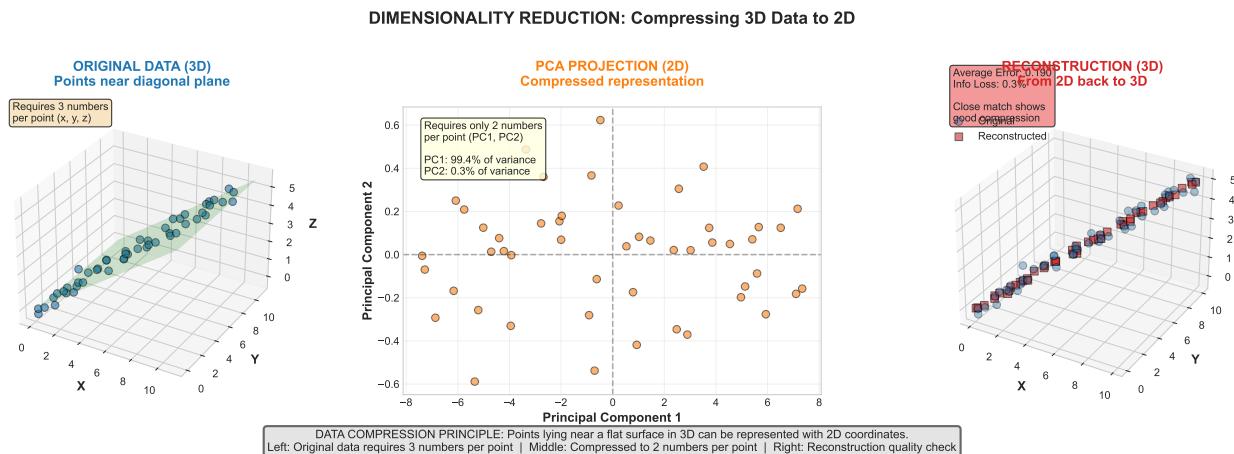
Both programs are competing, yet both improve. How is this possible?

Task 4: Predict equilibrium outcome

When both programs reach equilibrium, the Discriminator cannot reliably distinguish real from fake data. What does this imply about the Generator's performance?

Summary: Adversarial training uses competition between two programs to drive improvement. The Generator learns to create realistic data by attempting to fool the Discriminator. The Discriminator improves its detection ability, which forces the Generator to improve further. At equilibrium, the Generator produces data indistinguishable from real examples.

Chart 6: Dimensionality Reduction



Observation

The chart shows data transformation from three dimensions, then reconstruction back to three dimensions.

Tasks

Task 1: Assess data structure

Examine the left panel showing 3D data.

Do the points fill the entire 3D space, or do they lie near a flat surface?

Answer: _____

Task 2: Evaluate compression feasibility

The middle panel shows the same data represented with only 2 numbers per point instead of 3.

Is this compression possible without major information loss?

Answer: _____

Why or why not?

Task 3: Analyze reconstruction quality

The right panel overlays original points (blue) with reconstructed points (red).

Are the reconstructed points close to the original positions?

Answer: _____

What does this indicate about the quality of the 2D representation?

Task 4: Generalize the principle

Under what conditions can high-dimensional data be compressed to fewer dimensions?

Summary: Data lying near a lower-dimensional structure can be compressed by projecting onto principal directions of variance. This reduces storage requirements and computational complexity while preserving most information. The technique is particularly effective when original dimensions are correlated.

Connections Across Charts

Recurring Patterns

Examine the six charts and identify common themes:

1. Optimization objectives

Multiple charts show algorithms minimizing some quantity:

- Chart 1: Minimizing _____
- Chart 2: Minimizing _____
- Chart 4: Minimizing _____

2. Complexity trade-offs

Several charts illustrate balancing competing objectives:

- Chart 1: Simple models vs complex models
- Chart 3: Linear boundaries vs nonlinear boundaries
- Chart 6: Full dimensions vs compressed dimensions

What general principle connects these trade-offs?

3. Iterative improvement

Which charts show algorithms that improve through repeated steps?

Charts: _____

Key Insights

What are the three most important patterns observed across all charts?

1. _____

2. _____

3. _____

Questions for Lecture

List 2-3 questions to ask during the lecture:

1. _____
2. _____
3. _____

Bring this completed worksheet to class.