Discovery Worksheet

Introduction to Machine Learning & AI

Discover the Patterns Before the Formulas

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

How to Use This Worksheet

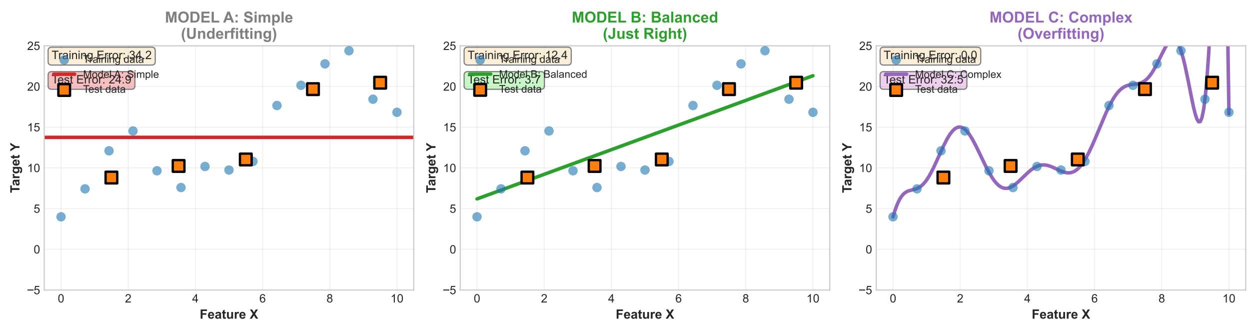
Important: Complete this **BEFORE** attending the lecture!
Your mission:

- Look at the 6 charts (one per section)
- Answer the discovery questions using calculations and observations
- Find the patterns YOURSELF before hearing the technical terms
- Compare answers with a partner (different answers are OK!)

Time needed: 75-90 minutes
Materials: Calculator, ruler, pencil

Learning Philosophy: You already understand these concepts from everyday life. This worksheet helps you discover the mathematical patterns behind what you already know intuitively. The lecture will then give you the formal language to describe your discoveries.

Discovery 1: The Overfitting Paradox



Observe the Chart

Three models are trying to predict house prices from square footage. All three were trained on the same blue dots (training data). Orange squares are NEW houses the models have never seen (test data).

Discovery Tasks

Task 1: Calculate Training Errors

For each model, measure how far predictions are from the blue training points.

- Model A (red horizontal line): Approx. training error = __________
- Model B (green curve): Approx. training error = __________
- Model C (purple wiggly): Approx. training error = __________

Which model has the LOWEST training error? ____________________

Task 2: Calculate Test Errors

Now measure how well each model predicts the NEW orange test points.

- Model A test error: __________
- Model B test error: __________
- Model C test error: __________

Which model has the LOWEST test error? ____________________

Task 3: The Paradox

Discovery Question:
Model C has the BEST training performance but the WORST test performance!
Why does perfect training lead to poor testing?

____

____

____

Task 4: Plot the Trade-off

On the grid below, plot training error (x-axis) vs test error (y-axis) for all three models:

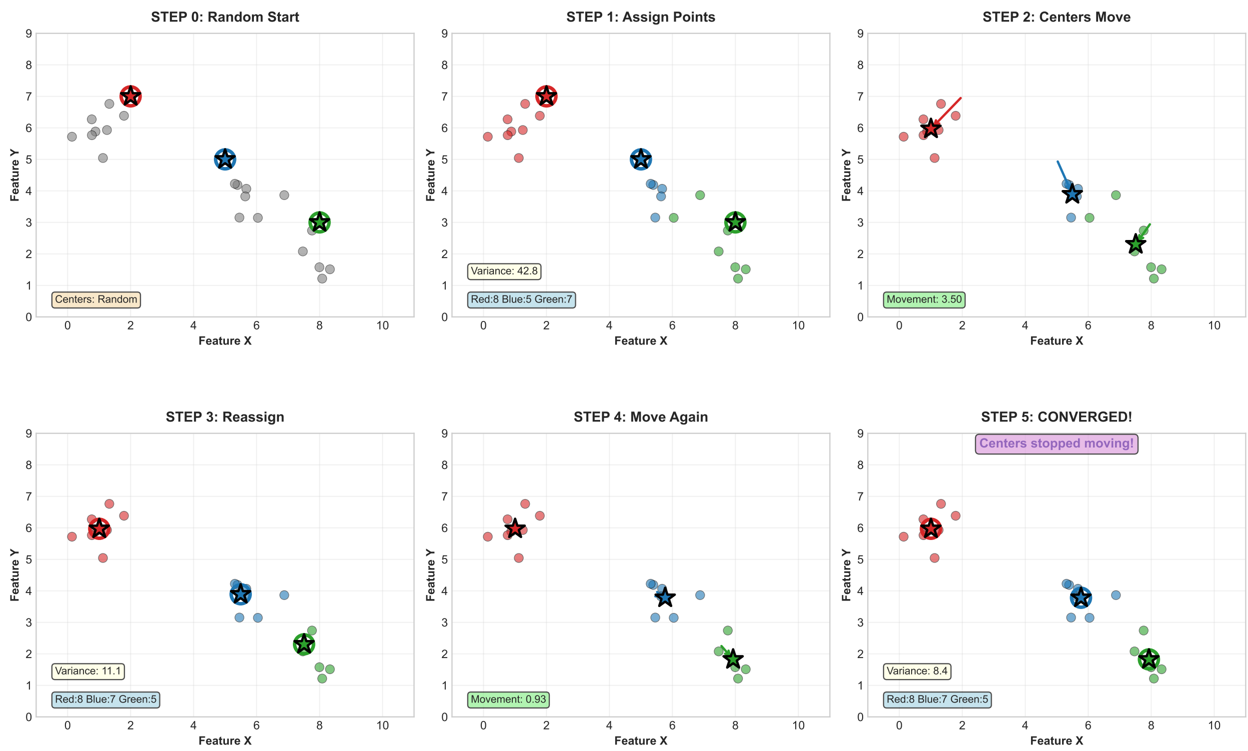
What pattern do you see? ____

Task 5: Predict Model D
Imagine Model D is even MORE complex than Model C (degree 20 polynomial).

- Predicted training error: __________ (Why? _________________________)
- Predicted test error: __________ (Why? _________________________)

Key Insight:
Model A problem: ______________________________ (too simple)
Model C problem: ______________________________ (too complex)
Model B achieves: ______________________________ (balanced)
In the lecture, you'll learn these have names: Bias and Variance

Discovery 2: The Moving Centers Algorithm



Observe the Chart

Watch how three centers (marked with stars) move over 5 steps to find natural groups in the data.

Discovery Tasks

Task 1: Calculate Center Movements

In Step 2, centers move to the average position of their assigned points.
Calculate the Red center position in Step 2:

- Red cluster points visible in Step 1: Count = _____
- Approximate center coordinates: (_____, _____)
- Movement distance from Step 0 to Step 2: __________ units

Task 2: Calculate Within-Cluster Variance

Variance measures how spread out points are from their center.
For the Red cluster in Step 1:

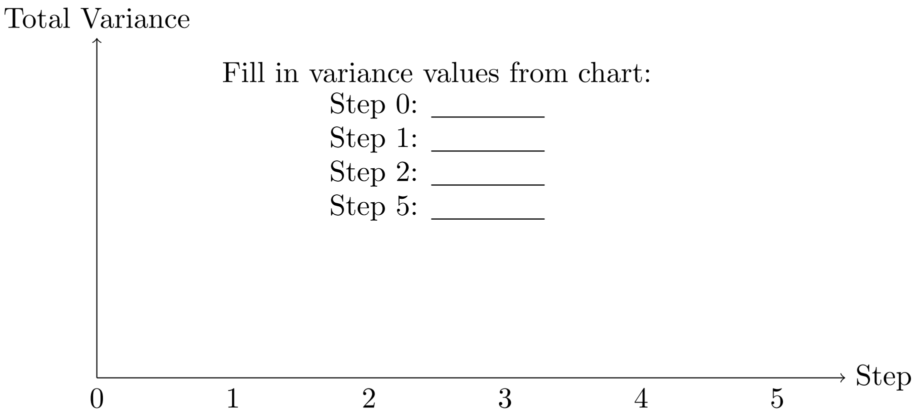
$$\text{Variance} = \frac{1}{n}\sum(\text{distance from point to center})^2$$

Pick 3 visible red points and calculate:

- Point 1 distance to red center: __________
- Point 2 distance to red center: __________
- Point 3 distance to red center: __________
- Average squared distance: __________

Task 3: Track Variance Reduction

The chart shows total variance values at each step.



Is variance increasing or decreasing? ____________________________

Task 4: Convergence Detection

Discovery Question:
How do you know the algorithm has finished? What stops changing?

____

____

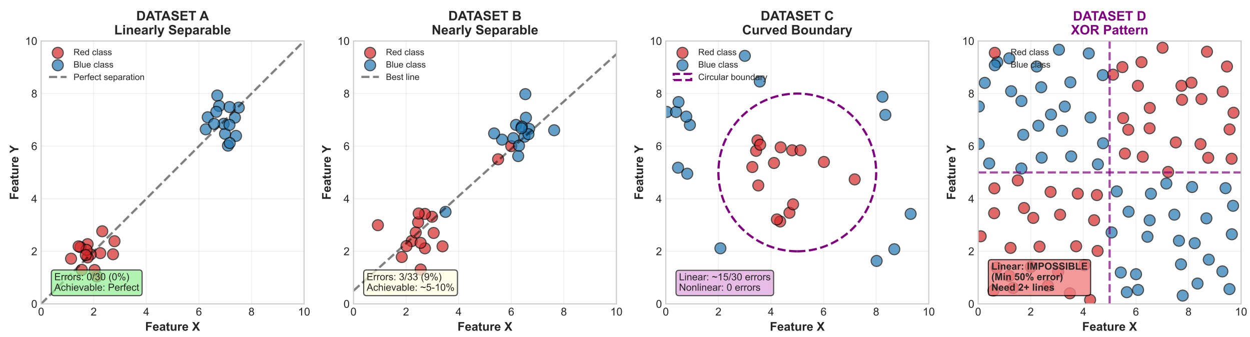
Task 5: Discover the Rules
By watching the 6-step sequence, you can figure out the algorithm!
The algorithm repeats two rules:
Rule 1 (Assignment): Each point ____

Rule 2 (Update): Each center moves to _______________________________________

Task 6: What Gets Optimized?
The algorithm is trying to minimize something. What?
The objective is to minimize: ____

Key Insight:
This "dance" has a name: **K-Means Clustering**
K = number of ______________________
Means = _________________________________ position
It minimizes ____________________________ within clusters

Discovery 3: The Impossible Separation



Observe the Chart

Four datasets show red and blue points. Your task: separate them with the simplest possible rule.

Discovery Tasks

Task 1: Draw Linear Boundaries

Using a ruler, draw the best straight line to separate red from blue in each dataset.
Count classification errors for each:

- Dataset A errors: ______/______ = __________%
- Dataset B errors: ______/______ = __________%
- Dataset C errors: ______/______ = __________%
- Dataset D errors: ______/______ = __________%

Which dataset(s) CAN be perfectly separated with a straight line? ____________________________

Task 2: Mathematical Proof (Dataset D - XOR Pattern)

Any straight line has equation: $ax + by + c = 0$
Points are classified as:

- Red if $ax + by + c > 0$
- Blue if $ax + by + c < 0$

Test the four corners of Dataset D:

Point	Actual Color	$ax + by + c$	Predicted Color
(1,1)	Red	__________	__________
(1,9)	Blue	__________	__________
(9,1)	Blue	__________	__________
(9,9)	Red	__________	__________

Proof by Contradiction:
Assume (1,1) is Red: $a + b + c > 0$
Then (9,9) is Red: $9a + 9b + c > 0$
But (1,9) is Blue: $a + 9b + c < 0$
And (9,1) is Blue: $9a + b + c < 0$
Add last two inequalities: $(a + 9b + c) + (9a + b + c) < 0$

Simplifies to: $10a + 10b + 2c < 0$
But we need: $9a + 9b + c > 0$ (for point (9,9) to be Red)
Contradiction! No single line can separate XOR pattern.

Task 3: Nonlinear Solutions

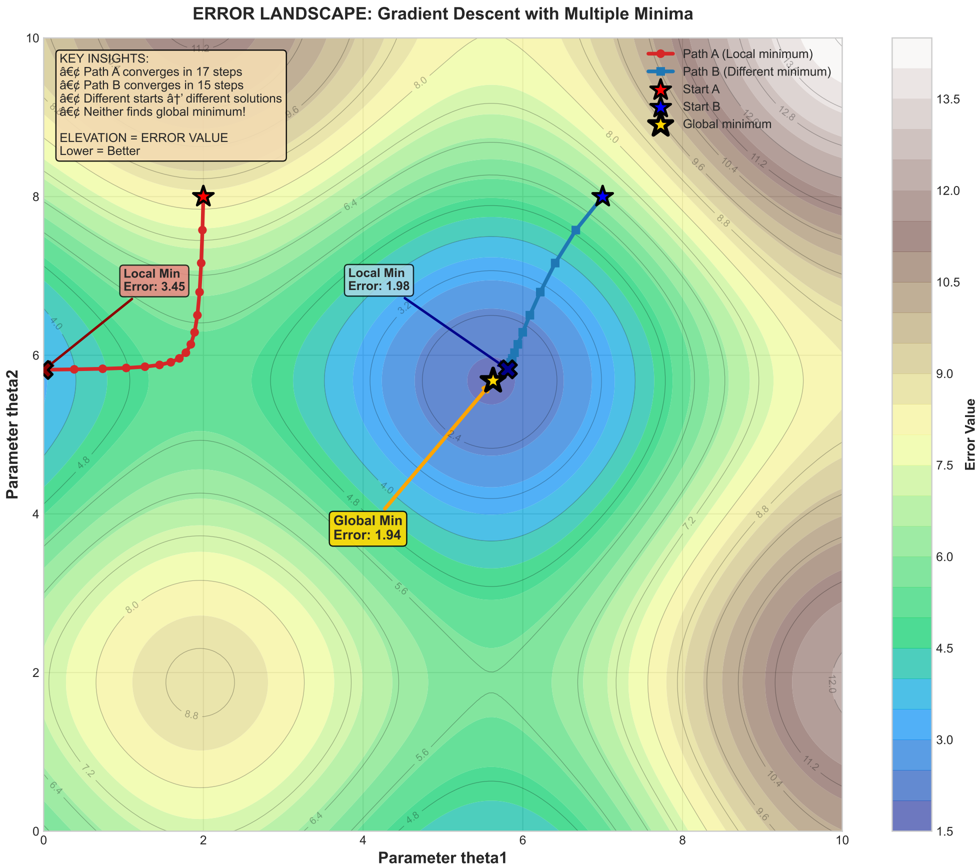
For Dataset C (circular pattern), what shape boundary works?
Boundary type: ____________________
Equation: $x^2 + y^2 =$ __________

For Dataset D (XOR pattern), propose a solution using TWO lines:
Line 1: $x =$ __________ (vertical)
Line 2: $y =$ __________ (horizontal)
Combined rule: Point is Red if ____

Task 4: When Linearity Fails

Key Discoveries:
Linear models work when: ____
Linear models fail when: ____
Solution for complex patterns: ___________________________________
In the lecture, you'll learn how neural networks combine multiple lines to solve XOR

Discovery 4: The Optimization Landscape



Observe the Chart

This topographic map shows "error elevation." Two paths (Red and Blue) descend from different starting points. Contour lines show equal error values.

Discovery Tasks

Task 1: Read the Terrain

- Path A (Red) starts at coordinates: (_____, _____)
- Path A ends in valley with error: __________
- Path B (Blue) starts at coordinates: (_____, _____)
- Path B ends in valley with error: __________

Did both paths find the SAME minimum? _______________
Global minimum (gold star) has error: __________

Task 2: Calculate Gradients
Gradient = rate of change = slope
Pick a point on Path A around coordinates (3, 7):
Read contour values:

- Error at (3.0, 7): $E_1 \approx$ __________
- Error at (3.5, 7): $E_2 \approx$ __________
- Error at (3, 7.5): $E_3 \approx$ __________

Calculate gradients:

$$\frac{\partial E}{\partial x} \approx \frac{E_2 - E_1}{0.5} = \underline{\hspace{3cm}}$$

$$\frac{\partial E}{\partial y} \approx \frac{E_3 - E_1}{0.5} = \underline{\hspace{3cm}}$$

Descent direction (negative of gradient): $(-\frac{\partial E}{\partial x}, -\frac{\partial E}{\partial y}) = (\_\_\_\_\_, \_\_\_\_\_)$

Task 3: Step Size Experiments

If learning rate (step size) is TOO LARGE:

- Problem: ____
- Risk: ____

If learning rate is TOO SMALL:

- Problem: ____
- Risk: ____

Optimal strategy: ____

Task 4: Local vs Global Minima

Discovery Question:
Why did Path A get stuck in a local minimum instead of finding the global minimum?

____

____

How could we help Path A escape its valley and find the better solution?

____

____

Task 5: Optimization Strategy

Update rule for gradient descent:

$$\theta_{\text{new}} = \theta_{\text{old}} - \alpha \cdot \nabla E(\theta)$$

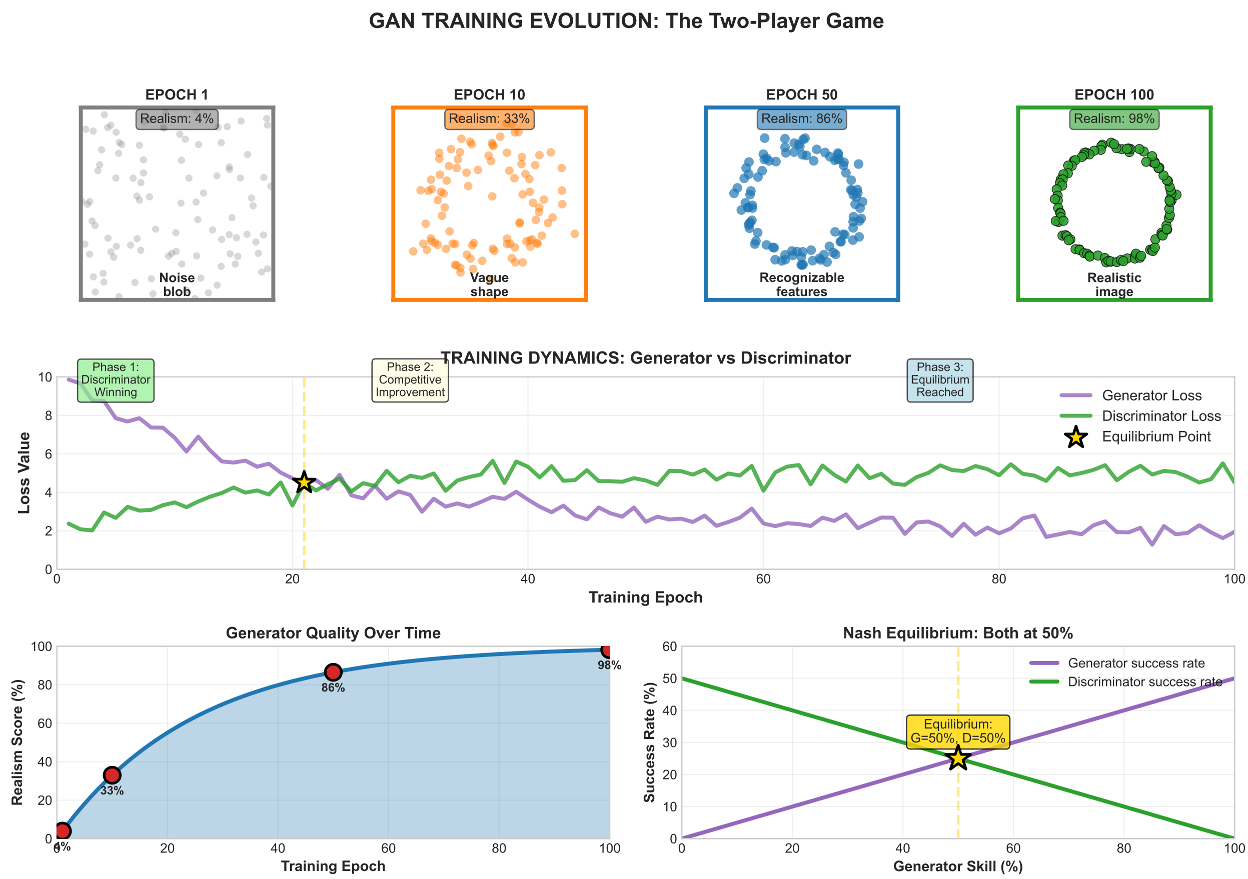
where:

- θ = parameters (model weights) = ______________________
- α = learning rate (step size) = ______________________
- ∇E = gradient (direction) = ______________________

If gradient is positive, we move _____________ (left/right)
If gradient is negative, we move _____________ (left/right)

Key Insights:
Gradient descent follows the ____________________________ direction
Different starting points lead to ____________________________ solutions
Solution: Random restarts, momentum, or _______________________

Discovery 5: The Two-Player Game



Observe the Chart

Top row shows generated samples improving from random noise to realistic. Middle shows loss curves for two competing players. Bottom shows quality metrics.

Discovery Tasks

Task 1: Track Quality Improvement

From the top row, read realism scores:

- Epoch 1 realism: __________% (looks like: _________________________)
- Epoch 10 realism: __________% (improvement: __________%)
- Epoch 50 realism: __________% (improvement: __________%)
- Epoch 100 realism: __________% (improvement: __________%)

Total improvement from Epoch 1 to 100: _______________%

Task 2: Analyze Loss Dynamics

From the middle training curve:
When is the Generator (purple) winning? Epochs: _______________
When is the Discriminator (green) winning? Epochs: _______________
When do they reach equilibrium? Around epoch: __________
At equilibrium:

- Generator loss ≈ __________
- Discriminator loss ≈ __________

- Are they equal? ___________

Task 3: Game Theory Calculations
From the bottom-right Nash equilibrium chart:
The formula for success rates:

- Generator success $= G \times (1 - D)$ where $G =$ generator skill, $D =$ discriminator skill
- Discriminator success $= D \times (1 - G)$

Fill in the table:

G skill	D skill	G success	D success	Who wins?
0.2	0.8	________	________	___________
0.5	0.5	________	________	___________
0.8	0.2	________	________	___________
0.9	0.1	________	________	___________

At what skill levels are both players equally successful? $G =$ _____, $D =$ _____
This is called **Nash Equilibrium**.

Task 4: Training Evolution
From the quality curve (bottom-left):

- Steepest improvement happens between epochs: ___________ and ___________
- Improvement slows down after epoch: ___________
- Will realism ever reach 100%? ___________ (Why/why not? ___________________________)

Task 5: The Adversarial Insight

Discovery Questions:
Why do BOTH players improve even though only one can win each round?

____

____

If we trained only the Generator (no Discriminator feedback), what would happen?

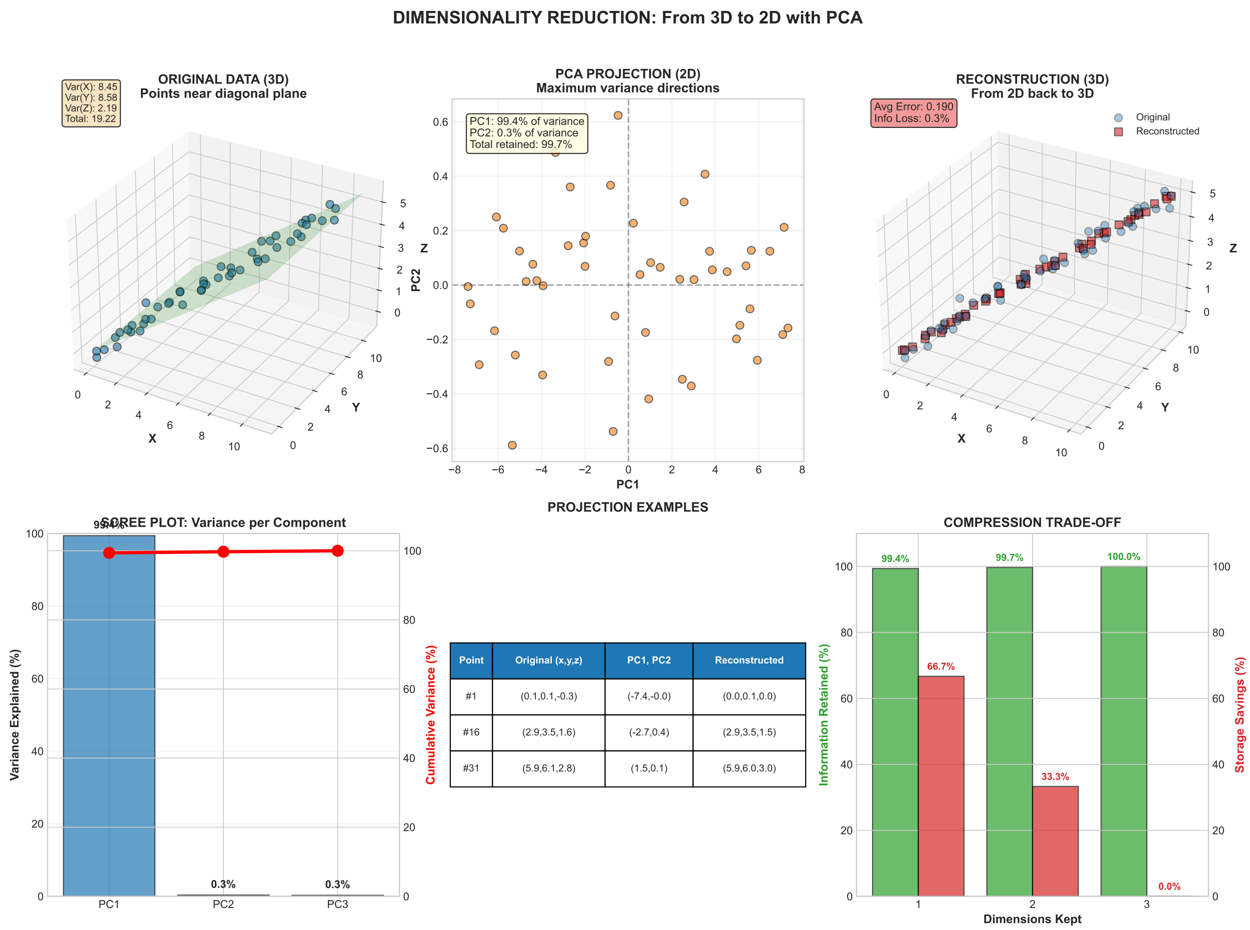
____

____

The art student analogy: Generator = ___________________________, Discriminator = ___________________________

Key Insights:
Competition drives ________________________________
Equilibrium occurs at ________________________________
This approach is called **Generative Adversarial Networks (GANs)**

Discovery 6: The Dimensionality Revelation



Observe the Chart

Top-left shows 3D data. Top-middle shows 2D projection. Top-right shows reconstruction error. Bottom shows variance accounting.

Discovery Tasks

Task 1: Variance Calculations

From the top-left 3D plot, read variance values:

- Var(X) = __________
- Var(Y) = __________
- Var(Z) = __________
- Total variance = __________

From the top-middle 2D projection:

- Var(PC1) = __________ (__________% of total)
- Var(PC2) = __________ (__________% of total)
- Total retained = __________%

Information lost by using only 2D: __________%

Task 2: Reconstruction Error

From the table in bottom-middle, pick one sample point:

- Original 3D: (_____, _____, _____)
- Projected to 2D: (_____, _____)
- Reconstructed 3D: (_____, _____, _____)

Calculate error:

$$\text{Error} = \sqrt{(x_{orig} - x_{recon})^2 + (y_{orig} - y_{recon})^2 + (z_{orig} - z_{recon})^2} = \_\_\_\_\_\_\_\_\_\_$$

Average error from chart: __________
Is reconstruction good or bad? _________________________

Task 3: Compression Analysis
From bottom-right chart:

Dimensions	Info Retained	Storage Saved	Good Trade-off?
3 (original)	100%	0%	N/A
2	__________%	__________%	__________
1	__________%	__________%	__________

Calculate compression ratio for 2D:
Original storage: 50 points × 3 dimensions = 150 numbers
Compressed storage: 50 points × 2 dimensions = __________ numbers
Compression ratio: __________ (__________% reduction)
With only __________% information loss!

Task 4: When Does PCA Work?

Discovery Questions:
Why does this dataset compress so well from 3D to 2D?
Hint: Look at the 3D visualization - where do most points lie?

____

____

If points were randomly scattered in all 3 dimensions, would PCA work as well?

____

Task 5: Principal Component Direction
From the scree plot (bottom-left):

- PC1 captures: __________% of variance (MOST important direction)
- PC2 captures: __________% of variance (SECOND most important)
- PC3 captures: __________% of variance (LEAST important)

Sum of all three: __________% (should be 100%)
The "elbow" in the scree plot suggests keeping __________ components.

Key Insights:
PCA finds directions of ______________________________ variance
Data lying near a lower-dimensional _________________________ compresses well
Trade-off: Storage savings vs _________________________
This technique is called **Principal Component Analysis (PCA)**

Final Reflection: Connecting Your Discoveries

Cross-Discovery Patterns

1. Optimization Appears Everywhere
You discovered optimization in multiple places:

- Discovery 1: Finding the model that balances training and test error
- Discovery 2: K-means minimizes within-cluster ____________________
- Discovery 4: Gradient descent minimizes ____________________
- Discovery 5: GANs optimize a competitive ____________________

What do all these have in common? ____

2. The Complexity Trade-off

- Discovery 1: Too simple (bias) vs too complex (____________________)
- Discovery 3: Linear (simple) vs nonlinear (____________________)
- Discovery 6: More dimensions (complex) vs fewer dimensions (____________________)

General principle: ____

3. When Simple Rules Fail

- Discovery 3 Dataset D: Single line cannot solve ____________________
- Solution shown: Combine multiple ____________________
- This motivates: Neural networks with multiple ____________________

Prepare for Lecture

You've Discovered the Core Ideas!
In the lecture, you'll learn the formal names:

- Discovery 1 → **Bias-Variance Tradeoff**
- Discovery 2 → **K-Means Clustering**
- Discovery 3 → **Linear Separability & Neural Networks**
- Discovery 4 → **Gradient Descent Optimization**
- Discovery 5 → **Generative Adversarial Networks (GANs)**
- Discovery 6 → **Principal Component Analysis (PCA)**

The formulas will make sense because you've already discovered the patterns!

Three Most Important Insights

What are the three most surprising or important patterns you discovered?

1. ____
____

2. ____

____

3. ____

____

Questions for Lecture

What questions do you still have? Write 2-3 questions to ask during the lecture:

1. ____
2. ____
3. ____

Excellent work! Bring this completed worksheet to class.