Theory Discovery Tools

Building Your Own Clustering Theory - Week 1

BSc Machine Learning for Innovation
Discovery Cards & Worksheets

How to Use These Tools

These cards and worksheets guide you through discovering clustering theory yourself. Cut out the cards, use them during exercises, and document your discoveries in the worksheets.

Part 1: Discovery Cards

Observation Cards

Cut along dotted lines. Use these to guide your observations.

OBSERVATION CARD 1
the spacing between points?
What do you notice about...
Look for: gaps, clusters, density
Document: Where are points close? Far?
Count: How many groups do you see?

OBSERVATION CARD 2
the shapes of groups?
What do you notice about...
Look for: circles, lines, blobs
Document: Are groups round? Stretched?
Question: Why these shapes?

OBSERVATION CARD 3
points that don't fit?
What do you notice about...
Look for: isolated points, bridges
Document: Which points are alone?
Consider: Are they noise or important?

OBSERVATION CARD 4
the center of groups?
What do you notice about...
Look for: dense areas, empty centers
Document: Where is the "middle"?
Measure: Distance from center to edges

Hypothesis Cards

Use these to form theories about clustering.

HYPOTHESIS CARD 1

If... points are within distance D of each other

Then... _____

Because...

_____

Test this with: Different D values

HYPOTHESIS CARD 2

If... we use the average position

Then... _____

Because...

_____

Test this with: Calculate group averages

HYPOTHESIS CARD 3

If... groups have similar sizes

Then... _____

Because...

_____

Test this with: Count points per group

HYPOTHESIS CARD 4

If... we know the number of groups

Then... _____

Because...

_____

Test this with: Try different K values

Test Cards

Validate your theories with these experiments.

TEST CARD 1

Try your theory on...

A dataset with clear, separated groups

Does it work? YES / NO
What happened? _____
Adjust theory: _____

TEST CARD 2

Try your theory on...

A dataset with overlapping groups

Does it work? YES / NO
What happened? _____
Adjust theory: _____

TEST CARD 3

Try your theory on...

A dataset with different densities

Does it work? YES / NO
What happened? _____
Adjust theory: _____

TEST CARD 4

Try your theory on...

A dataset with outliers

Does it work? YES / NO
What happened? _____
Adjust theory: _____

Reflection Cards

Understand why your discoveries work.

REFLECTION CARD 1

My clustering rule succeeds when:

_____

It fails when:

_____

The key insight is:

_____

This works because...

REFLECTION CARD 2

Distance matters because:

_____

Different distances give:

_____

The best distance is:

_____

This works because...

REFLECTION CARD 3

Groups form naturally when:

_____

The pattern I see is:

_____

This relates to innovation by:

_____

This works because...

REFLECTION CARD 4

Outliers are important because:

_____

They represent:

_____

In innovation, outliers are:

_____

This works because...

Part 2: Theory Building Worksheets

Worksheet 1: Rule Creator

Template for writing clustering rules

Create Your Clustering Algorithm

Step 1: Define "Belongs Together"

Two points belong in the same group if:

_____

_____

Step 2: Write Your Algorithm

1. Start with: _____
2. For each point: _____
3. Calculate: _____
4. Assign to group if: _____
5. Repeat until: _____
6. Stop when: _____

Step 3: Handle Special Cases

If a point doesn't fit any group: _____
If groups overlap: _____
If I don't know K: _____

Step 4: Name Your Algorithm

My clustering algorithm is called: _____
Because it: _____

Worksheet 2: Pattern Finder

Guide for identifying regularities

Finding Patterns in Data

Visual Pattern Recognition

Pattern Type	I See This	It Means	Algorithm Needed
Circular groups	YES / NO		
Elongated groups	YES / NO		
Nested groups	YES / NO		
Chain connections	YES / NO		
Varying density	YES / NO		
Clear outliers	YES / NO		

Numerical Pattern Recognition

When I count groups at different scales:
- Fine scale (small distance): _____ groups
- Medium scale: _____ groups
- Coarse scale (large distance): _____ groups

The pattern is: _____

Innovation Pattern Recognition

In my field, natural groups form around:

- _____
- _____
- _____

These patterns suggest the clustering approach: _____

Worksheet 3: Theory Tester

Framework for validation

Testing Your Clustering Theory

Theory Statement
My theory: _____
_____

Prediction Table

If my theory is correct...	I expect to see...	Actually saw...
On dataset A:		
On dataset B:		
On dataset C:		

Success Metrics
My clustering is good when:

- Within-cluster distance is: _____
- Between-cluster distance is: _____
- Number of outliers is: _____
- Groups are balanced: YES / NO / SOMETIMES

Theory Refinement
What worked: _____
What failed: _____
Revised theory: _____
_____

Worksheet 4: Concept Connector

Linking discoveries to formal terms

From Your Words to Formal Theory

Translation Table
Connect what you discovered to the formal terminology:

What I Called It	Formal Term	Why It's Important
"Group middle point"	Centroid	Represents the cluster
"How spread out"		
"Doesn't fit anywhere"		
"How tight the group is"		
"Space between groups"		
"Best number of groups"		

Algorithm Matching
My approach is most similar to:

- ☐ K-means (uses centers, equal groups)
- ☐ DBSCAN (uses density, finds shapes)
- ☐ Hierarchical (builds tree, multiple scales)
- ☐ GMM (uses probability, overlapping groups)
- ☐ Something new: _____

Innovation Applications
My clustering discovery applies to innovation by:

1. _____
2. _____
3. _____

Key Insight
The most important thing I learned about clustering is:

_____
_____