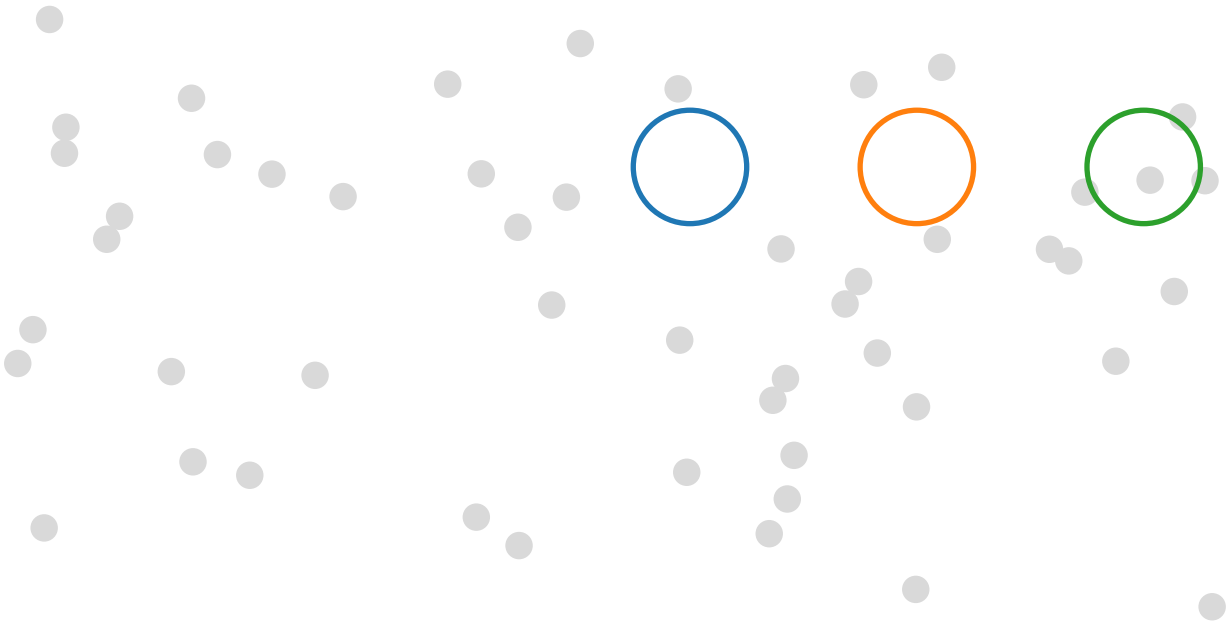
Machine Learning for Innovation

A Discovery Journey

Week 1: Finding Patterns in Chaos



Name: _____

Date: _____

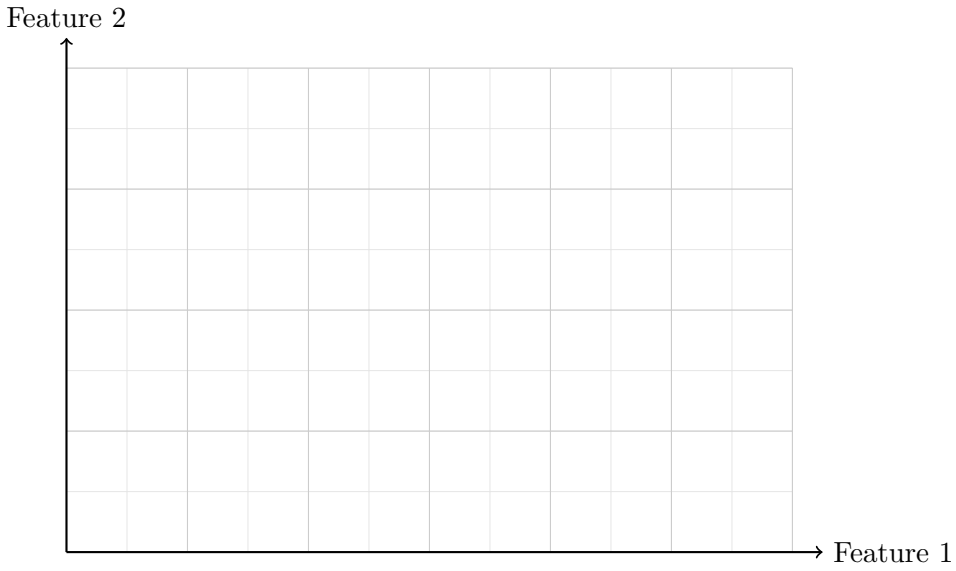
Page 2: The Innovation Challenge

Your Innovation Universe

Think about innovations in your field. They could be products, services, processes, or ideas.

Exercise: Draw Your Innovation Ideas

In the grid below, place 20 different innovation ideas as dots. Don't think too hard - just scatter them naturally.



Discovery Questions

1. Look at your dots. Can you see any natural groups? Circle them.

2. What makes grouping difficult? List three challenges:
   - _____
   - _____
   - _____

3. Why might a computer be better at finding these patterns? _____

   _____

Think About This

Key Insight: Human brains can only visualize 2-3 dimensions at once. But innovations have hundreds of features!

Page 3: The Hidden Dimensions

Exercise: Feature Discovery

Pick ONE innovation idea (e.g., a smart home device, an app, a new service). Write it here:
My Innovation: _____
Now, list ALL the factors that could affect its success:

Technical Features

- _____
- _____
- _____
- _____
- _____

User Features

- _____
- _____
- _____
- _____
- _____

Market Features

- _____
- _____
- _____
- _____
- _____

Financial Features

- _____
- _____
- _____
- _____
- _____

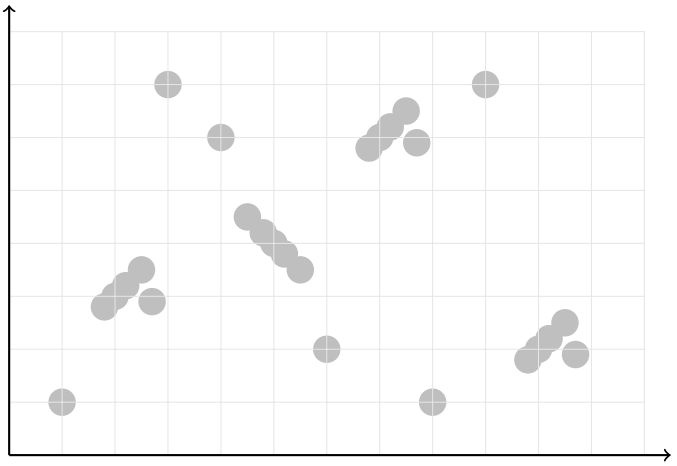
Exercise: Dimension Realization

1. How many total features did you identify? _____
2. If each innovation has this many features, how would you compare 1000 innovations?
   _____
3. This is why we need Machine Learning! It can handle hundreds of dimensions that we can't visualize.

Page 4: Discovering K-Means Clustering

Exercise: Manual Clustering Challenge

You'll now experience what the K-Means algorithm does!
Step 1: Place 3 center points (marked with X) anywhere on the grid with dots:



Exercise: K-Means Steps

Step 2: For each dot, draw a line to its nearest center (X).
Step 3: Count how many dots belong to each center:

- Center 1: ____ dots
- Center 2: ____ dots
- Center 3: ____ dots

Step 4: Now move each X to the middle of its assigned dots (this is the "mean" in K-means!).
Step 5: Would you need to repeat? Why? _____

What Happens When K Changes?

1. What if you used K=2 centers instead? Draw how the groups would change.
2. What if K=5? Would that be better or worse? Why? _____

Page 5: How Do We Measure "Close"?

Exercise: Distance Discovery

Consider two innovation ideas with features:

- Innovation A: (x=2, y=3)
- Innovation B: (x=5, y=7)

Euclidean Distance

"As the crow flies" - straight line
Formula: $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$
Calculate:

$$\begin{aligned} d &= \sqrt{(5-2)^2 + (7-3)^2} \\ &= \sqrt{\text{----} + \text{----}} \\ &= \sqrt{\text{----}} \\ &= \text{----} \end{aligned}$$

When to use: Physical measurements

Manhattan Distance

"City blocks" - grid walking
Formula: $d = |x_2 - x_1| + |y_2 - y_1|$
Calculate:

$$\begin{aligned} d &= |5-2| + |7-3| \\ &= \text{----} + \text{----} \\ &= \text{----} \end{aligned}$$

When to use: Features are independent

Exercise: Distance in Innovation Context

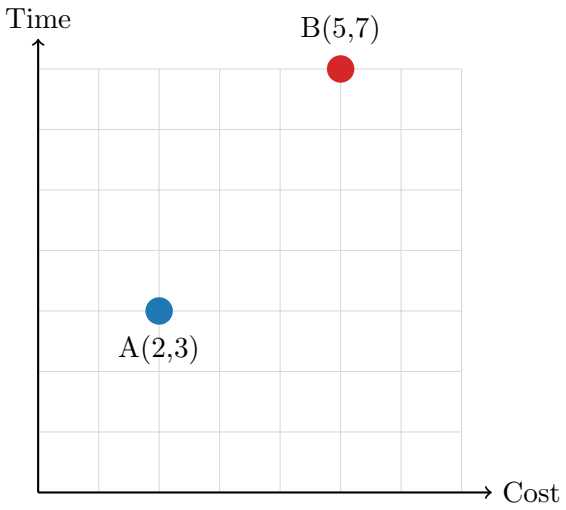
1. If x = "development cost" and y = "time to market", which distance makes more sense?

   _____

2. If you had 100 features instead of 2, could you still calculate distance?

   _____

3. Draw the two different distance paths on this grid:



Page 6: How Good Are Your Clusters?

The Silhouette Score

Measures how well-separated your clusters are (ranges from -1 to 1).

Exercise: Calculate Your Own Silhouette

Look at this simple example with 3 clusters:

For point X:

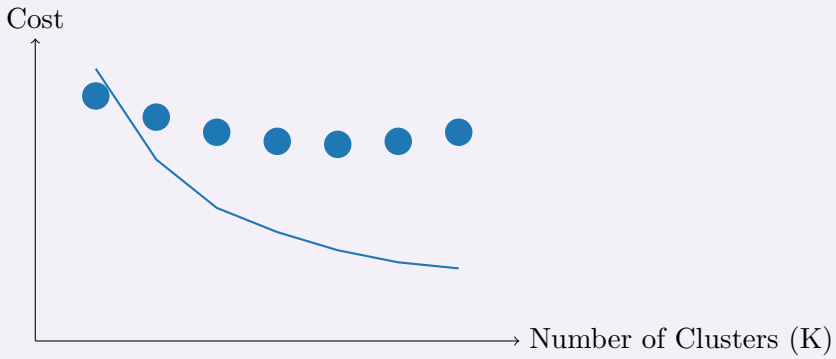
- a(X) = average distance to points in same cluster = ____
- b(X) = average distance to points in nearest other cluster = ____
- Silhouette(X) = $\frac{b(X)-a(X)}{\max(a(X),b(X))}$ = ____

Exercise: Interpret the Score

- Score close to 1: _____(Well/Poorly) clustered
- Score close to 0: _____(On the border/In the center)
- Score close to -1: _____(Correct/Wrong) cluster

The Elbow Method

Plot the "cost" (sum of distances) vs. number of clusters:

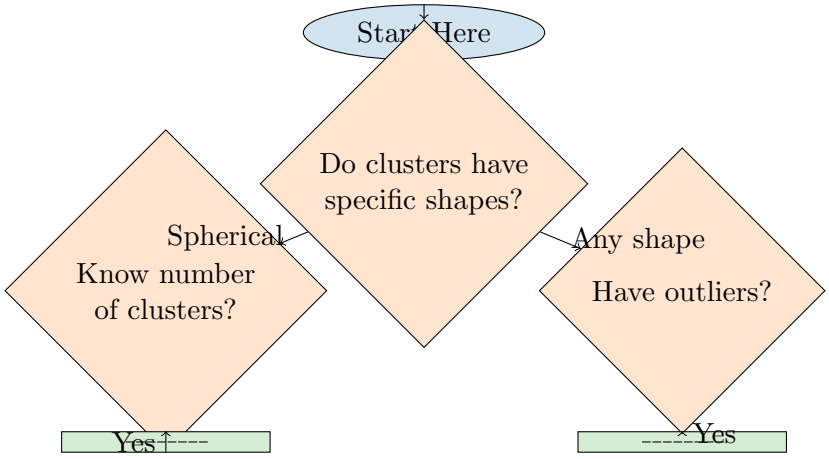


Circle where you think the "elbow" is. This is your optimal K!

Page 7: Choosing the Right Algorithm

Exercise: Build Your Decision Tree

Fill in the blanks to create your algorithm selection guide:



Exercise: Match the Scenario

Draw lines connecting scenarios to the best algorithm:

Scenarios:

1. Customer segments (equal size)
2. Finding fraud (rare events)
3. Product categories (hierarchy)
4. Geographic regions (density)
5. User behaviors (overlapping)

Algorithms:

- K-Means
- DBSCAN
- Hierarchical
- Gaussian Mixture
- Mean Shift

Your Innovation Context

Which algorithm would work best for YOUR innovation challenge? Why?

Page 8: Your Innovation Clustering Project

Exercise: Design Your Project

Now apply everything you've learned to a real innovation challenge!

1. Define Your Innovation Dataset

What innovations will you analyze? (products, services, ideas, etc.)

_____

How many items? _____
List 5 key features you'll measure:

1. _____
2. _____
3. _____
4. _____
5. _____

2. Choose Your Method

Which clustering algorithm? _____
Why this one?

_____

_____

How many clusters (K) do you expect? _____
Why? _____

3. Define Success

What silhouette score would be "good enough"? _____
How will you validate your clusters make business sense?

_____

_____

4. Reflection

Three things I discovered about clustering today:

1. _____
2. _____
3. _____

One question I still have:

_____

How I'll use this in my innovation work:

_____

_____

Congratulations!

You've discovered the fundamentals of clustering for innovation.
Next week: Advanced clustering techniques!