Research Paper Assignment - Week 7

Applied AI I - Interpretable Clustering Study

IMPORTANT: This is a LEARNING EXERCISE!

Don't panic! This assignment is designed to help you:

- Apply clustering techniques thoughtfully
- Connect unsupervised learning to real meaning
- Practice interpreting ML results critically
- Build skills in explainable AI
- Understand when clusters are meaningful vs artificial

Remember: The journey of critical analysis is more important than "perfect" clusters. Focus on understanding WHAT clusters mean and WHETHER they're interpretable!

Assignment Overview

Research Question: "Do discovered clusters in dataset X align with interpretable structure? Can we explain and validate the meaning of unsupervised clusters?" **Goal:** Apply clustering methods and rigorously assess whether discovered groups have meaningful, interpretable structure.

Deliverables:

1. Research paper (4-6 pages PDF)
2. Jupyter notebook with all experiments
3. All code must be reproducible

Due: [3 weeks from assignment date - To be announced by instructor]

Value: This demonstrates your ability to apply unsupervised learning critically and explain complex ML results!

Learning Objectives

By completing this assignment, you will: 1. **Apply** clustering algorithms (K-Means, DBSCAN, Hierarchical) 2. **Interpret** discovered clusters using domain knowledge 3. **Validate** clusters using both internal and external metrics 4. **Explain** cluster membership using explainability techniques 5. **Assess** whether clusters reflect real structure vs algorithmic artifacts 6. **Communicate** interpretability findings clearly

Key Skill: Learning to question and interpret unsupervised learning results!

Part 1: Dataset Selection

Requirements

Your dataset MUST be:

- ✓ **Structured**: Tabular data with interpretable features
- ✓ **Real-world**: Actual application domain
- ✓ **Sufficient size**: Minimum 500 samples
- ✓ **Unlabeled or labels-aside**: You'll cluster without labels, then validate
- ✓ **Interpretable features**: You can explain what features mean
- ✓ **Publicly available**: With proper citation

Ideal characteristics: - Features have clear domain meaning - Expect 2-10 natural groups

Dataset Sources

Where to Find Datasets:

- UCI Machine Learning Repository
- Kaggle Datasets (search "clustering" or "segmentation")

Datasets NOT Allowed

Do NOT use:

- ✗ Time series (unless aggregated to features)
- ✗ < 500 samples
- ✗ Datasets without interpretable feature
- ✗ Datasets from previous assignments

Why? Need interpretable features to assess meaningful structure!

Dataset Requirements

You MUST:

1. **Understand your domain**

- What do features represent?
- What groups might naturally exist?
- Why would clustering be useful here?

1. **Explore the data**

- Feature distributions
- Correlations
- Outliers
- Missing values

1. **Prepare features**

- Handle missing values
- Scale/normalize features
- Select relevant features (don't use IDs!)
- Consider dimensionality reduction if > 15 features

4. **Set aside ground truth (if available)**

- Don't use labels during clustering!
- Use only for validation afterward

Part 2: Methodology

Experimental Design

You will apply **3 clustering algorithms:**

1. **K-Means**

- Try k = 2 to 8 (or appropriate range)
- Find optimal k using Elbow method + Silhouette
- Centroid-based, assumes spherical clusters

1. **Hierarchical Clustering**

- Agglomerative (bottom-up)
- Try different linkages (ward, complete, average)
- Create dendrogram
- Cut at different heights

1. **DBSCAN**

- Density-based
- Find optimal eps using k-distance plot
- Try different min_samples values
- Handles noise as outliers

Validation Strategy

Internal Validation (No ground truth):

- Silhouette Score
- Davies-Bouldin Index
- Calinski-Harabasz Index
- Within-cluster sum of squares (WCSS)

External Validation (If ground truth available):

- Adjusted Rand Index (ARI)
- Normalized Mutual Information (NMI)

- Homogeneity, Completeness, V-measure

Interpretability Assessment:

- Feature importance per cluster
- Cluster profiles (mean features)
- Domain expert evaluation (or your own domain understanding)
- Visual interpretability

Required Experiments

Experiment 1: K-Means with Optimal K

Steps:

1. Scale features (StandardScaler)
2. Apply K-Means for k = 2 to 8
3. Plot Elbow curve (WCSS vs k)
4. Plot Silhouette scores vs k
5. Select optimal k
6. Train final K-Means model
7. Analyze cluster centers

Questions:

- What is optimal k?
- Do metrics agree?
- Are cluster centers interpretable?

Experiment 2: Hierarchical Clustering

Steps:

1. Apply hierarchical clustering
2. Create dendrogram
3. Try ward, complete, average linkage
4. Cut tree at different heights
5. Compare to K-Means results

Questions:

- Does dendrogram suggest natural groupings?
- Which linkage works best?
- Agree with K-Means on number of clusters?

Experiment 3: DBSCAN

Steps:

1. Create k-distance plot to find eps
2. Apply DBSCAN with different eps and min_samples
3. Identify outliers (label = -1)
4. Compare to K-Means/Hierarchical

Questions:

- How many outliers detected?
- Different cluster structure than K-Means?
- More or fewer clusters?

Experiment 4: Dimensionality Reduction Visualization

Steps:

1. Apply PCA to 2D
2. Plot clusters in 2D space
3. Apply UMAP to 2D (if time permits)
4. Visualize for all 3 algorithms

Questions:

- Do clusters separate visually?
- Same structure across algorithms?
- Any overlap or noise? #### Experiment 5: Cluster Interpretation

Steps:

1. For best algorithm, create cluster profiles
2. Calculate mean feature values per cluster
3. Identify distinguishing features
4. Name clusters based on characteristics
5. (Optional) Apply SHAP to explain cluster assignment

Questions:

- What defines each cluster?
- Can you name clusters meaningfully?
- Do clusters align with domain knowledge?

Part 3: Interpretability Analysis

Core Question: Are Clusters Meaningful?

You MUST assess:

1. Feature-based Interpretation

- Which features distinguish clusters?
- Are differences meaningful or trivial?
- Do feature profiles make domain sense?

2. Domain Knowledge Validation

- Do clusters correspond to known categories?
- Can you explain clusters to a domain expert?
- Would these groups be useful in practice?

3. Stability

- Do different algorithms find similar structure?
- Are results stable across random initializations?
- Sensitive to outliers or parameter choices?

4. Actionability

- Can you act on these clusters?
- Are recommendations different per cluster?
- Do clusters provide business value?

Part 5: Interpretability Requirements

Critical Assessment

You MUST critically assess interpretability:

Ask yourself:

1. **Can I explain each cluster?**
- If yes: How? What defines it?
- If no: Why not? Missing domain knowledge? No real structure?

2. **Do clusters make sense?**
- Align with expectations?
- Surprising but plausible?
- Or arbitrary?

3. **Are clusters actionable?**
- Different strategy per cluster?
- Or same treatment for all?

4. **Are clusters stable?**

- Different algorithms agree?
- Robust to parameters?
- Or fragile?

5. **Are clusters meaningful or artifacts?**

- Real structure in data?
- Or imposed by algorithm?
- Evidence for either?

Be honest! If clusters aren't interpretable, that's a valid finding!

When Clusters Are NOT Interpretable

Possible reasons:

1. **No real structure**

- Data is uniformly distributed
- Algorithms force groupings
- Validation metrics are poor

2. **Wrong number of clusters**

- Over-segmentation (too many)
- Under-segmentation (too few)
- Try different k values

3. **Wrong features**

- Irrelevant features dominate
- Important features missing
- Need feature engineering

4. **Wrong algorithm**

- K-Means assumes spherical
- Data has non-spherical structure
- Try DBSCAN or hierarchical

5. **Insufficient domain knowledge**

- Can’t interpret features
- Don’t know what’s meaningful
- Need domain expert

If clusters aren’t interpretable, discuss WHY in your paper!

Part 6: Code Submission

Jupyter Notebook Requirements

Your notebook MUST include:

1. **Clear structure**

- Markdown headers for each section
- Explanatory text between code blocks
- Analysis and interpretation

2. **Reproducible code**

- All imports at top
- Random seeds set
- Clear variable names
- Comments on non-obvious code

3. **All experiments**

- Each algorithm in separate section
- Parameter selection process
- Final models with analysis
- All visualizations displayed

4. **Interpretability analysis**

- Cluster profiling code
- Feature importance calculations
- Visualization code
- Domain knowledge notes

Code Structure Template

```python
# ============================================

# Week 7: Interpretable Clustering Study

# Student: [Your Name]

# Dataset: [Dataset Name]

# ============================================


# -------- IMPORTS --------

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns


from sklearn.cluster import KMeans, DBSCAN, AgglomerativeClustering

from sklearn.preprocessing import StandardScaler

from sklearn.decomposition import PCA

from sklearn.metrics import silhouette_score, davies_bouldin_score, adjusted_rand_score

from scipy.cluster.hierarchy import dendrogram, linkage


import warnings

warnings.filterwarnings('ignore')


# Random seed
```

```python
RANDOM_STATE = 42

np.random.seed(RANDOM_STATE)


# -------- 1. LOAD DATA --------

# [Your code here]


# -------- 2. EXPLORATORY ANALYSIS --------

# [Your code here]


# -------- 3. PREPROCESSING --------

# [Your code here]


# -------- 4. OPTIMAL K SELECTION --------

# [Your code here - Elbow, Silhouette]


# -------- 5. K-MEANS EXPERIMENT --------

# [Your code here]


# -------- 6. HIERARCHICAL CLUSTERING --------

# [Your code here]


# -------- 7. DBSCAN EXPERIMENT --------

# [Your code here]
```

```python
# -------- 8. COMPARISON & VISUALIZATION --------

# [Your code here - PCA/UMAP plots]


# -------- 9. CLUSTER INTERPRETATION --------

# [Your code here - profiles, feature importance]


# -------- 10. EXTERNAL VALIDATION (if ground truth) --------

# [Your code here]


# -------- 11. EXPLAINABILITY (OPTIONAL) --------

# [Your code here - SHAP, surrogate model]
```

Part 7: Evaluation Rubric

Total: 100 points

Dataset Selection & Preprocessing (10 points)

Dataset Choice (5 points):

- 5: Excellent domain, interpretable features, appropriate for clustering
- 3: Adequate dataset but minor issues
- 1: Poor choice (too simple, uninterpretable features)

Preprocessing (10 points):

- 8-10: Thorough, well-documented, appropriate techniques
- 5-7: Adequate preprocessing
- 2-4: Minimal preprocessing
- 0-1: Inadequate or incorrect

Methodology & Experimental Design (20 points)

Experimental Design (10 points):

- 9-10: Rigorous methodology, all algorithms correctly applied
- 7-8: Good methodology, minor issues
- 5-6: Adequate but some flaws
- 0-4: Poor experimental design

Implementation (10 points):

- 9-10: All 3 algorithms correctly implemented, parameter selection justified
- 7-8: Most correct, minor bugs
- 5-6: Some implementation issues
- 0-4: Major errors

Results & Validation (20 points)

Completeness (8 points):

- 7-8: All required experiments and metrics
- 5-6: Most experiments, minor omissions
- 3-4: Some experiments missing
- 0-2: Incomplete results

Visualizations (7 points):

- 6-7: All required figures, professional quality
- 4-5: Most figures, adequate quality
- 2-3: Some figures, poor quality
- 0-1: Missing or very poor

Validation Rigor (5 points):

- 5: Multiple metrics, internal + external (if available)

- 3-4: Adequate validation
- 1-2: Minimal validation
- 0: No validation

Interpretability Analysis (25 points)

THIS IS THE MOST IMPORTANT SECTION!

Cluster Interpretation (10 points):

- 9-10: Deep, domain-grounded interpretation of all clusters
- 7-8: Good interpretation, adequate depth
- 5-6: Superficial interpretation
- 0-4: Minimal or incorrect interpretation

Critical Assessment (10 points):

- 9-10: Rigorous assessment of whether clusters are meaningful
- 7-8: Good critical thinking
- 5-6: Some critical analysis
- 0-4: No critical assessment

Explainability Methods (5 points):

- 5: Multiple explainability techniques applied well
- 3-4: Basic explainability
- 1-2: Minimal explainability
- 0: No explainability analysis

Discussion & Critical Thinking (15 points)

Depth of Analysis (8 points):

- 7-8: Insightful, considers validity, limitations, domain context
- 5-6: Good analysis

- 3-4: Superficial
- 0-2: Minimal

Answering Research Question (7 points):

- 6-7: Clear answer with strong evidence
- 4-5: Adequate answer
- 2-3: Weak answer
- 0-1: Doesn't answer question

Writing Quality (10 points)

Organization & Clarity (5 points):

- 5: Well-structured, clear, logical flow
- 3-4: Adequate structure
- 1-2: Poor organization
- 0: Very poorly organized

Technical Communication (5 points):

- 5: Clear explanation of technical concepts
- 3-4: Mostly clear
- 1-2: Unclear or overly technical
- 0: Poor communication

BONUS POINTS (up to +10)

Extra Credit:

- +3: SHAP for cluster membership explanation
- +3: Decision tree surrogate model for interpretability
- +2: UMAP in addition to PCA

- +2: Stability analysis across multiple runs
- +2: Particularly insightful domain analysis
- +2: Exceptional visualizations

Maximum: 110 (capped at 100 for final grade)

Part 8: Common Pitfalls

Clustering Pitfalls

Pitfall 1: Forcing structure where none exists

- Some datasets have no natural clusters!
- If metrics are poor and clusters unininterpretable, that's valid!
- Don't force interpretation

Pitfall 2: Using wrong number of clusters

- Elbow and Silhouette may disagree
- Try range of k values
- Consider domain knowledge

Pitfall 3: Not scaling features

- K-Means sensitive to scale
- Always use StandardScaler or similar
- Exception: Already same scale and meaningful

Pitfall 4: Interpreting clusters as causal

- Clusters show correlation, not causation
- Don't overinterpret
- Be cautious with recommendations

Interpretability Pitfalls

Pitfall 5: Cherry-picking interpretations

- If clusters don't make sense, say so!
- Don't force meaningless interpretations
- Honest negative results are valuable

Pitfall 6: Ignoring domain knowledge

- Can't interpret without understanding domain
- Research your dataset's context
- Consult domain literature

Pitfall 7: Trusting metrics blindly

- High Silhouette ≠ interpretable clusters
- Metrics are necessary but not sufficient
- Always validate with domain sense

Pitfall 8: Comparing apples and oranges

- Different algorithms find different structures
- No single "correct" clustering
- Context determines which is "best"

Analysis Pitfalls

Pitfall 9: Not visualizing in reduced dimensions

- Hard to understand high-D clusters
- Always use PCA/UMAP visualization
- Helps spot problems

Pitfall 10: Ignoring outliers

- Outliers affect K-Means strongly
- DBSCAN handles outliers explicitly
- Decide: Remove or keep?

Pitfall 11: Overfitting to training data

- Clusters should generalize
- If using train/test, cluster on train, validate on test
- Or use cross-validation with clustering

Part 9: Tips for Success

DO ✓

- ✅ Choose dataset you understand
- ✅ Start with EDA (always!)
- ✅ Scale features before clustering
- ✅ Try multiple k values
- ✅ Use multiple validation metrics
- ✅ Visualize clusters in 2D
- ✅ Create detailed cluster profiles
- ✅ Think critically about interpretability
- ✅ Be honest about limitations
- ✅ Write for domain experts, not just ML experts

DON'T ✗

- ❌ Use datasets with uninterpretable features
- ❌ Skip scaling
- ❌ Use only one validation metric
- ❌ Force interpretations that don't make sense
- ❌ Ignore when algorithms disagree
- ❌ Treat clusters as ground truth
- ❌ Forget to set random seeds
- ❌ Skip dimensionality reduction visualization

Part 10: Final Encouragement

You Can Do This!

This assignment is different:

- Not about “best” clustering
- About understanding what clusters MEAN
- About critical thinking
- About honest analysis

Success means:

- Applying methods correctly
- Thinking critically about results
- Explaining findings clearly
- Being honest about limitations

Remember:

- Clusters aren’t always meaningful (valid finding!)
- Interpretability matters more than perfect metrics
- Domain knowledge is essential
- Critical thinking > algorithmic perfection

Start early. Explore thoroughly. Think critically. Write clearly.

You’ve got this!

Happy clustering!