

Pattern Recognition and Machine Learning

Lecture 10 Clustering

Chizhi Chris ZHANG

zhangchizhi@ciomp.ac.cn

Advanced Computing and Digital Technology Research Center

University of Chinese Academy of Sciences



Autumn 2026

Lecture plan and time allocation

Topic	Minutes
Unsupervised learning	22.5
Similarity and distance	22.5
Clustering evaluation	22.5
Clustering algorithms	67.5

The six scores 0, 1, 2, 7, 8, 9

Define distance; check a grouping; follow each algorithm on the same scores.

The numerical steps are shown on the slides.

The six scores do not arrive with true group names

Clustering starts with observations and a chosen similarity rule. There are no verified class labels in the fitting objective. A useful grouping must serve a stated purpose.

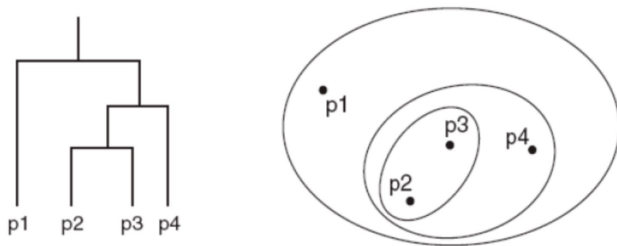
For customer grouping, one row must represent one customer over a stated time window. Aggregate transactions into behavior features; identifiers are keys, not measures of behavioral proximity.

transactions $\rightarrow x_i = (\text{Visits}, \text{Spend}, \text{recency})$ for customer i

These are six activity measurements. No row supplies a verified target class.

$$X = (0, 1, 2, 7, 8, 9).$$

The same points admit multiple useful partitions



The reference places a dendrogram beside its nested groups. It contains the same four points at different resolutions: p_2, p_3 join first, then p_4 , and finally p_1 .

Check. Different resolutions answer different questions. A hierarchy preserves these alternatives; it does not establish a universally correct customer taxonomy.

Hard, soft, partial, and nested membership answer different questions

A hard partition assigns each row to one cluster. Soft probabilistic membership assigns nonnegative weights summing to one. Density clustering may leave noise unassigned; a hierarchy returns nested groups.

$$\mathbf{z}_i \in \{1, \dots, K\} \quad \text{versus} \quad r_i = (0.7, 0.3), \quad \sum_k r_{ik} = 1$$

Check. A responsibility is model-based uncertainty, not the percentage of a person belonging to a social category. Cluster IDs can be permuted without changing the partition.



Changing group numbers preserves the same partition

Every pair of observations stays together or stays apart in exactly the same way. Only the arbitrary group names changed.

Score	0	1	2	7	8	9
Result A	1	1	1	2	2	2
Result B	2	2	2	1	1	1

A new object needs an explicit assignment calculation

An analysis can describe the current groups $\{0, 1, 2\}$ and $\{7, 8, 9\}$ without yet specifying what happens to a future measurement.

Nearest centre assigns the new object to the low-score group. The boundary solves $|x - 1| = |x - 8|$, so $x = 4.5$. A tie at 4.5 needs a declared policy.

$$c_L = 1 \quad c_H = 8$$

$x = 1.5$

$$|1.5 - 1| = .5, \quad |1.5 - 8| = 6.5.$$

$$|x - 1| = |x - 8| \quad x = 4.5$$

Check. The rule preserves this fitted grouping convention. Its relevance and stability still need evaluation.

Compute how scaling changes the nearest neighbor

Compute absolute distances in this one-dimensional example and retain the full matrix for later methods.

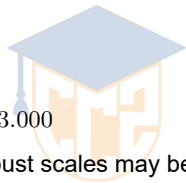
Euclidean distance weights coordinate changes by their numerical units. Standardization uses training or analysis-reference statistics to express changes relative to typical feature variation.

$$\mathbf{A} = (1, 100), \mathbf{B} = (2, 110), \mathbf{C} = (4, 101)$$

$$d(A, B) = \sqrt{1 + 100} = 10.05; \quad d(A, C) = \sqrt{9 + 1} = 3.16$$

$$\mathbf{s} = (1, 100) : \quad d_z(\mathbf{A}, \mathbf{B}) = \sqrt{1 + 0.01} = 1.005; \quad d_z(\mathbf{A}, \mathbf{C}) = \sqrt{9 + 0.0001} = 3.000$$

Check. Changing currency units should not accidentally redefine customer similarity. Robust scales may be preferable with extreme observations.



Calculate cosine and Jaccard rather than mixing their meanings

Quantity	Definition	Worked example
Euclidean distance	$\sqrt{\sum_j (x_j - z_j)^2}$	(1, 2) versus (2, 4): $\sqrt{5}$
Cosine similarity	$x^T z / (\ x\ \ z\)$	For the same two vectors: 1
Jaccard similarity	$ A \cap B / A \cup B $	$\{a, b\}, \{b, c\}$: $1/3$

Smaller distances mean closer rows; larger similarities mean closer rows. The feature meaning determines which comparison is useful.

Cosine ignores overall positive length scaling. Jaccard ignores jointly absent items. Neither is the same question as Euclidean distance.

$$\mathbf{x} = (1, 2) \quad \mathbf{z} = (2, 4)$$

$$\mathbf{x}^T \mathbf{z} = 10, \quad \|\mathbf{x}\| = \sqrt{5}, \quad \|\mathbf{z}\| = \sqrt{20},$$

$$\cos(\mathbf{x}, \mathbf{z}) = \frac{10}{\sqrt{5}\sqrt{20}} = 1.$$

$$A = \{a, b\}$$

$$B = \{b, c\}$$

$$A \cap B = \{b\}$$

$$A \cup B = \{a, b, c\}$$

$$J(A, B) = 1/3.$$

Build a distance matrix for the shared six-row dataset

From score 0 to the six scores, distances are $(0, 1, 2, 7, 8, 9)$.

$$d(a, b) = |a - b|.$$

$$d(2, 0) = 2, \quad d(2, 1) = 1, \quad d(2, 2) = 0,$$

$$d(2, 7) = 5, \quad d(2, 8) = 6, \quad d(2, 9) = 7.$$

$$d(a, b) = d(b, a)$$

One-dimensional Euclidean distance is absolute difference. The matrix is symmetric and its diagonal is zero.

$$D = \begin{pmatrix} 0 & 1 & 2 & 7 & 8 & 9 \\ 1 & 0 & 1 & 6 & 7 & 8 \\ 2 & 1 & 0 & 5 & 6 & 7 \\ 7 & 6 & 5 & 0 & 1 & 2 \\ 8 & 7 & 6 & 1 & 0 & 1 \\ 9 & 8 & 7 & 2 & 1 & 0 \end{pmatrix}$$

Check. Another meaningful feature can change proximity. A gap is evidence for this geometry, not a business label.

Ask three different questions about clustering quality

Consider the proposed groups $\{0, 1, 2\}$ and $\{7, 8, 9\}$ before choosing an algorithm.

Agreement with reference labels, separation in the chosen geometry, and stability under perturbation are different claims.

Evidence	Question	Required evidence
Agreement with reference	Agreement with independent reference labels	Reference labels excluded from fitting
Geometric Separation	Within-group closeness and between-group separation	Task-relevant distance
Stability	Membership under resampling or changed settings	Repeated fits and membership comparisons

Check. Geometric separation alone cannot establish practical value. Reference labels may describe a different objective from the intended grouping.

Compute a second silhouette to see why one average is incomplete

For a row, a is its mean distance to other members of its own group. b is the smallest mean distance to another group. Compare both using the same metric.

$$s = \frac{b - a}{\max(a, b)}$$

$$x = 1: \quad \mathbf{a} = (1 + 1)/2 = 1,$$

$$b = (6 + 7 + 8)/3 = 7 \quad \Rightarrow \quad s = 6/7 \approx 0.857$$

Check. Silhouette favors structures well separated in the chosen distance. It can penalize meaningful curved or variable-density groups.

$s(1) = 6/7 \approx .857$ while $s(0) = .8125$. Both are well separated here, but group position changes the ratio.

$$\mathbf{a} = (|0 - 1| + |0 - 2|)/2 = 1.5,$$

$$b = (7 + 8 + 9)/3 = 8,$$

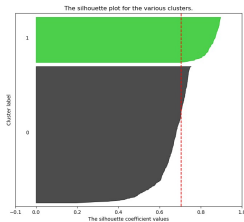
$$\mathbf{s}(0) = (8 - 1.5)/8 = .8125.$$

$$\mathbf{s}(1) = 6/7 \approx .857 \quad \mathbf{s}(0) = .8125$$

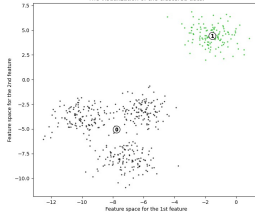
Check. A high overall average can hide a small weak group. Use each row's own group and nearest alternative when computing its value.

Read the distribution of silhouette values

Silhouette analysis for KMeans clustering on sample data with n_clusters = 2



The visualization of the clustered data



Each horizontal bar is one observation. A silhouette near 1 means it is closer to its own cluster; near 0 suggests a boundary case; a negative value suggests another cluster may fit it better.



Pair agreement checks all six pairs of four IDs

Three pairs agree out of six: $R/I = 3/6 = .5$. Group names never enter the calculation.

Pair	Together in A?	Together in B?	Agree?
(1, 2)	Yes	Yes	Yes
(1, 3)	No	Yes	No
(1, 4)	No	No	Yes
(2, 3)	No	Yes	No
(2, 4)	No	No	Yes
(3, 4)	Yes	No	No

$$RI = 3/6 = .5$$

Chance adjustment on the same four objects

Partitions are $A = \{12 \mid 34\}$ and $B = \{123 \mid 4\}$. The contingency counts are:

	123	4	Sum
12	2	0	2
34	1	1	2
Sum	3	1	4

Let $\mathbf{a} = \sum_{ij} \binom{n_{ij}}{2} = 1$, $\mathbf{r} = \sum_i \binom{n_{i\cdot}}{2} = 2$, $\mathbf{c} = \sum_j \binom{n_{\cdot j}}{2} = 3$, and $T = \binom{4}{2} = 6$.

$$E = rc/T = 2(3)/6 = 1,$$

$$ARI = \frac{a - E}{(r + c)/2 - E} = \frac{1 - 1}{2.5 - 1} = 0.$$

RI was .5. ARI is zero because same-cluster pair overlap equals the fixed-size chance expectation here. They are different scales.

Describe groups in original units and test stability

For our score example, report both membership and centre. In real data, retain group size, spread, visits and recency.

Group	Scores	Centre
Low activity	0, 1, 2	1
High-activity group	7, 8, 9	8

Check. Stable groups need not justify different actions. A behavior group is not a causal or value judgment.



K-means searches for a compact partition around K centers

Different algorithms assume different group structures: compact centers, nested similarity, connected dense regions or probability mixtures. Examine those assumptions before choosing an algorithm.

Key idea

Choose K centers and assign every row to exactly one. The squared-error objective measures how far rows are from the center of their assigned group.

$$J = \sum_{i=1}^N \sum_{k=1}^K r_{ik} \|x_i - \mu_k\|^2, \quad r_{ik} \in \{0, 1\}, \quad \sum_k r_{ik} = 1$$

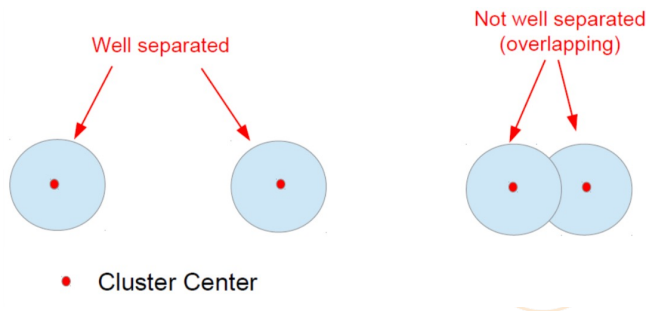
$$X = \{0, 1, 2, 7, 8, 9\} \quad K = 2$$

Check. K-means discovers the best compact groups for this objective, not all possible kinds of structure. The center need not be an observed customer.

Compare separated groups with overlapping groups

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Separated means are insufficient if within-group spread is large. Compare separation with variation inside each group, using a distance meaningful for the measured features.



The mean is the exact update for squared distance

With centers fixed, every row minimizes its own contribution independently. Select the center with the smallest squared distance; comparing ordinary distances gives the same winner.

$$\mathbf{z}_i = \arg \min_k \|\mathbf{x}_i - \mu_k\|^2$$

$$\mu_1 = 0, \mu_2 = 7: \quad x = 2 \Rightarrow (4, 25),$$

$$x = 8 \Rightarrow (64, 1)$$

$$\mathbf{C}_1 = \{0, 1, 2\} \quad \mathbf{C}_2 = \{7, 8, 9\}$$

Check. Assignment uses the current centers, not the final centers. Fix a tie rule so repeated equal-distance rows are handled consistently.

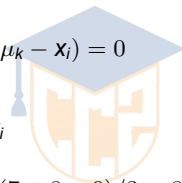
With assignments fixed, each center solves a separate quadratic problem. Differentiating the sum of squared distances sets the center to the arithmetic mean of its assigned rows.

$$\frac{\partial}{\partial \mu_k} \sum_{i \in \mathcal{C}_k} (x_i - \mu_k)^2 = 2 \sum_{i \in \mathcal{C}_k} (\mu_k - x_i) = 0$$

$$\mu_k = \frac{1}{|\mathbf{C}_k|} \sum_{i \in \mathbf{C}_k} x_i$$

$$\mu_1 = (0 + 1 + 2)/3 = 1 \quad \mu_2 = (7 + 8 + 9)/3 = 8$$

Check. The mean follows from the squared-error objective. Changing the distance or loss can require a different prototype update.

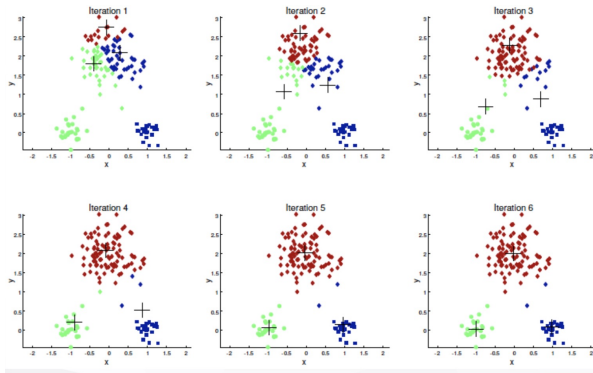


Read the original iteration figure as alternating decisions

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Read the six panels in temporal order. Colored points indicate assignments and crosses indicate current centers. First identify moved centers, then inspect rows whose nearest center changes.

Check. Each exact assignment or mean update cannot increase SSE. Different starting points can still reach different local minima.



Spread seeds with K-means++ and still repeat the run

After choosing the first seed, sample the next row with probability proportional to its squared distance from the nearest chosen seed. This favors uncovered regions without always choosing the farthest outlier.

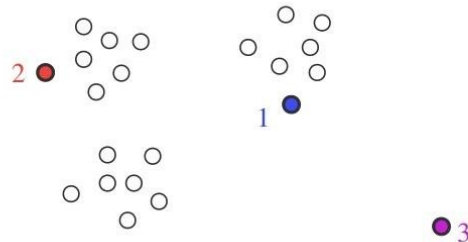
$$P(x_i \text{ next}) = \frac{D(x_i)^2}{\sum_j D(x_j)^2}$$

Initial centers 0 : $D^2 = (0, 1, 4, 49, 64, 81)$, $\sum D^2 = 199$



Distant seeds can also be outliers

K-means++ favors points far from existing centers. A distant outlier can therefore become a seed. Repeated initializations and objective comparisons reduce dependence on one starting choice.



Choose K using diminishing returns and usefulness

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

For the best achievable K-means objective, increasing K cannot make SSE larger: extra centers can reproduce a coarser solution. Training SSE therefore rewards ever finer partitions.

$$X = \{0, 1, 2, 7, 8, 9\} : \quad J_1 = 77.5, \quad J_2 = 4, \quad J_3 = 2.5, \quad J_6 = 0$$

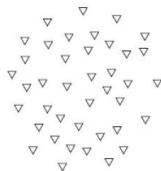
Check. Combine an elbow with silhouette, repeated-start stability, and interpretable profiles. There may be no clear elbow in real data.

A compact-center objective has geometric biases

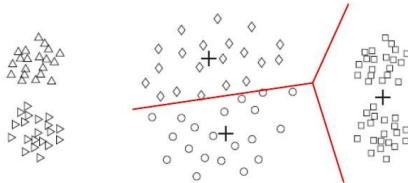
ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Nearest-center assignments create convex Voronoi regions. K-means therefore struggles with curved groups, strongly unequal sizes, and strongly unequal spreads or densities.

Check. More starts can improve optimization but cannot change the model geometry. Consider another representation or method when the desired group is nonconvex.



(a) Original points.



(b) Three K-means clusters.

An outlier separates mean, median, and medoid

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

A mean minimizes squared error; a coordinate median minimizes one-dimensional absolute error; a medoid is an observed row minimizing total chosen dissimilarity to its group.

$$C = \{0, 1, 2, 30\} : \quad \text{Mean} = 8.25, \quad \text{Median} = 1.5$$

$$\sum_{x \in C} |x - 1| = 31, \quad \sum_{x \in C} |x - 2| = 31$$



Check. K-medoids can use broader distances and be more robust, but it is not simply the median and often costs more computation.

A Gaussian mixture models density before assigning groups

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Assume a latent component z generates each row: first select component k with weight π_k , then draw the row from its Gaussian distribution.

$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k), \quad \pi_k \geq 0, \quad \sum_k \pi_k = 1$$



Check. A component is a model of a distribution, not necessarily a one-to-one real customer type. Multiple components can approximate one complicated density.

Mean, covariance, and weight play distinct roles

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

The mean gives a component location. Covariance gives within-component spread and orientation. The mixing weight gives how often that component is selected before seeing a row.

$$\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 4 & 1.5 \\ 1.5 & 1 \end{pmatrix}, \quad \pi = 0.3$$



Check. A broad component can explain a farther row better than a narrow nearby component. Nearest mean alone is not the GMM assignment rule.

The E step converts component densities into responsibilities

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Key idea

Given current parameters, Bayes rule assigns normalized posterior weights to each row. These responsibilities sum to one across components for that row.

$$r_{ik} = P(\mathbf{z}_i = k \mid \mathbf{x}_i) = \frac{\pi_k \mathcal{N}(\mathbf{x}_i \mid \mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_i \mid \mu_j, \Sigma_j)}$$

Check. E means expectation of hidden membership indicators. It estimates missing assignments under the current model; it does not observe true cluster labels.



One responsibility follows from a density ratio

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Means are 1, 8, variances are both 9 and mixture weights are both 1/2. Common Gaussian constants cancel in the ratio.

$$\frac{u_2(x)}{u_1(x)} = \exp\left(-\frac{(x-8)^2 - (x-1)^2}{18}\right) = \exp\left(\frac{7(x-4.5)}{9}\right).$$

$$x = 2$$

$$r_1(2) = \frac{u_1}{u_1 + u_2} = \frac{1}{1 + u_2/u_1} = \frac{1}{1 + e^{-1.9444}} \approx .875.$$

$$r_2(2) = 1 - .875 = .125$$



Compute an E step for the same six activity scores

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Use two equal-weight one-dimensional Gaussians with initial means 1, 8 and variances 9, 9. The first responsibility has a simple logistic expression after canceling common constants.

$$r_{i1} = \frac{1}{1 + \exp(7(x_i - 4.5)/9)}, \quad r_{i2} = 1 - r_{i1}$$

x_i	0	1	2	7	8	9
r_{i1}	.971	.938	.875	.125	.062	.029



Weighted observations produce the updated mean

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Hold the E-step responsibilities fixed. Treat each row as a fractional contribution to each component, then maximize the expected complete-data log likelihood.

$$N_k = \sum_i r_{ik}, \quad \pi_k^{\text{new}} = N_k/N, \quad \mu_k^{\text{new}} = \frac{\sum_i r_{ik} x_i}{N_k}$$

$$\Sigma_k^{\text{new}} = \frac{\sum_i r_{ik} (x_i - \mu_k^{\text{new}})(x_i - \mu_k^{\text{new}})^T}{N_k}$$

These rounded terms explain the arithmetic. The final variance and likelihood use unrounded responsibilities.



$$\mathbf{x} = (0, 1, 2, 7, 8, 9), \quad \mathbf{r}_1 \approx (.971, .938, .875, .125, .062, .029).$$

$$N_1 \approx .971 + .938 + .875 + .125 + .062 + .029 = 3,$$

$$\sum_i r_{i1} x_i \approx 0(.971) + 1(.938) + 2(.875) + 7(.125) + 8(.062) + 9(.029) = 4.320.$$

$$\mu_1^{\text{new}} \approx 4.320/3 = 1.440.$$

EM update: calculate the variance

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Use all six responsibility-weighted squared deviations from $\mu_1 = 1.440445$. Their sum divided by $N_1 = 3$ gives the new variance.

Scores	Responsibility	Squared deviation	Weighted contribution
0	0.9707	2.075	2.014
1	0.9383	0.194	0.182
2	0.8748	0.313	0.274
7	0.1252	30.909	3.869
8	0.0617	43.028	2.654
9	0.0293	57.147	1.675

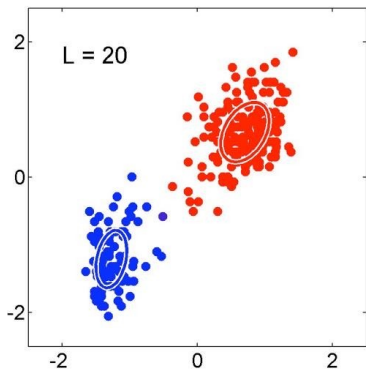


$$\sigma_1^2 = \frac{10.667371}{3} \approx 3.556.$$

By symmetry, $\mu_2 = 7.559555$, $\sigma_2^2 = 3.556$, with equal mixture weights. These new parameters are used in the next E-step.

Read fitted mixture ellipses as probability models

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER



The original plot shows a mixture after 20 iterations. Locate component means, then read ellipse direction, spread and overlap.



Prevent tiny covariance from producing a false victory

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

A component can concentrate around one observation with variance approaching zero, driving Gaussian density at that row upward. Unregularized mixture likelihood can therefore be unbounded.

$$\mathcal{N}(\mu \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \rightarrow \infty \quad \text{as } \sigma^2 \rightarrow 0$$

Condition to check

A very high training likelihood may mean collapse rather than a meaningful new group. Check component size and covariance eigenvalues.

Choose mixture complexity and connect it back to K-means

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

For these univariate mixtures, $p = K$ means $+K$ variances $+(K - 1)$ free mixture weights $= 3K - 1$.

Let p be the number of fitted parameters, N the number of observations, and $\log \hat{L}$ the maximized log likelihood. Smaller criteria are preferred under their assumptions.

$$\text{AIC} = 2p - 2 \log \hat{L}, \quad \text{BIC} = p \log N - 2 \log \hat{L}.$$

$N = 100$	p	$\log \hat{L}$	AIC	BIC
2 components	5	-120	250	263.03
3 components	8	-116	248	268.84



The larger fit improves likelihood. AIC prefers it here; BIC charges a larger complexity cost and prefers the simpler fit.

$\pi_k = 1/K$, $\Sigma_k = \sigma^2 I$, $\sigma^2 \rightarrow 0 \Rightarrow$ nearest-mean responsibilities become hard.

A density criterion is not a direct measure of the practical usefulness of a customer grouping.

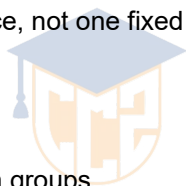
A hierarchy keeps several resolutions in one tree

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Agglomerative clustering begins with singleton groups and repeatedly merges the closest two. Divisive clustering starts with one group and splits it. The tree records a sequence, not one fixed K.

$$\{0|1|2|7|8|9\} \longrightarrow \{01|2|78|9\} \longrightarrow \{012|789\} \longrightarrow \{012789\}$$

Check. Specify linkage before merging. Different definitions of distance between groups generate different trees from the same pairwise distances.



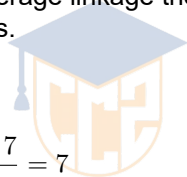
Linkage defines the distance between two multi-row groups

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Single linkage uses the nearest cross-pair, complete linkage the farthest, and average linkage the mean of all cross-pair distances. These definitions emphasize different structures.

$$A = \{0, 1, 2\}, B = \{7, 8, 9\}$$

$$d_{\text{single}} = 5, \quad d_{\text{complete}} = 9, \quad d_{\text{average}} = \frac{7 + 8 + 9 + 6 + 7 + 8 + 5 + 6 + 7}{9} = 7$$



Work through agglomeration using the shared distance matrix

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Use single linkage and a stated tie rule. The smallest original distance is 1, occurring between several adjacent scores. Merge one tied pair, then update group distances using the minimum cross-pair.

$$\{0\}, \{1\} \rightarrow \{0, 1\} \quad \text{at height 1}$$

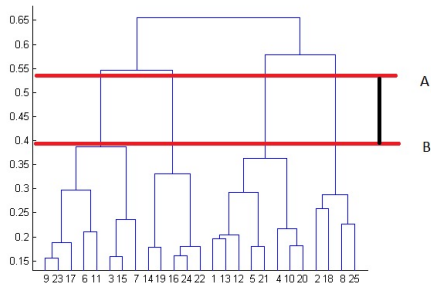
$$d(\{0, 1\}, \{2\}) = \min(2, 1) = 1$$

$$\{0, 1\}, \{2\} \rightarrow \{0, 1, 2\}; \quad \{7, 8, 9\} \text{ forms similarly}$$



Read a dendrogram through merge heights and cuts

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER



Leaves represent observations, branches record merges, and vertical heights encode the chosen linkage. A horizontal cut intersects branches to produce a partition at that resolution.

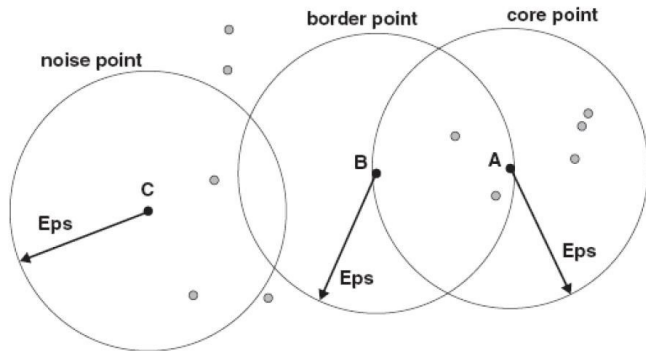
Check. Horizontal spacing and branch rotation do not define sample distances. A large visual gap is a clue, not a universal K-selection theorem.

DBSCAN separates core points, border points, and noise

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Specify radius ϵ and minimum neighbor count m . A core has enough rows in its neighborhood; a border is near a core but cannot expand; noise belongs to no accepted core-connected region.

Check. Noise is an algorithmic role under chosen parameters, not proof of a faulty measurement. Border assignment can be ambiguous near two groups.



Count neighborhoods on the same six-row example

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Let $\varepsilon = 1.1$ and $m = 3$, including the row itself. Neighbor counts follow directly from our distance matrix and determine the point roles.

$$\varepsilon = 1.1 \quad m = 3$$

x	0	1	2	7	8	9
count	2	3	2	2	3	2
Role	border	core	border	border	core	border

Check. Add an isolated score 20: it has one neighbor (itself), is near no core, and is therefore noise under these parameters.



Expand through core connections, not through every nearby point

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

Start from an unvisited core, join its neighborhood, and continue through newly found cores. Border rows are accepted but do not launch further expansion.

$$N_{\varepsilon}(x) = \{z : d(x, z) \leq \varepsilon\} \quad |N_{\varepsilon}(x)| \geq m \Rightarrow x \text{ is core}$$

Condition to check

Dense reachability is not ordinary transitive closeness through arbitrary sparse bridges. Distinguish the core graph from attached border rows.

Radius and neighbor count control different thresholds

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

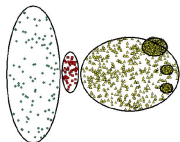
Radius creates neighborhood links; minimum count asks how much local evidence is enough.
Neither directly specifies the number of groups.

Parameters	Six scores, counting each row itself
$\varepsilon = 0.5, m = 3$	No core: every row is noise.
$\varepsilon = 1.1, m = 3$	Two groups, each with one core and two borders.
$\varepsilon = 5.1, m = 3$	Cross 2 Present. 7 connections define one group

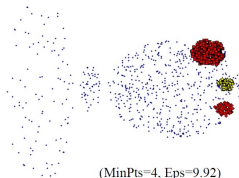


One radius may not fit both sparse and dense groups

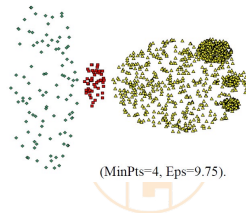
One neighborhood radius may not suit groups with very different densities. Inspect which points become noise and which groups join as the radius changes.



Original Points



(MinPts=4, Eps=9.92)



(MinPts=4, Eps=9.75).

A similarity graph turns grouping into a connectivity question

The graph links 0–1–2 and 7–8–9. Its degrees are (1, 2, 1, 1, 2, 1) and $L = D - W$. Each connected component supplies a zero-eigenvalue direction.

$$L = \begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

$$u_1 = \frac{1}{\sqrt{3}}(1, 1, 1, 0, 0, 0)^T, \quad u_2 = \frac{1}{\sqrt{3}}(0, 0, 0, 1, 1, 1)^T, \quad Lu_1 = Lu_2 = 0.$$

Use the two vectors as coordinate columns. The first three rows all equal $(1/\sqrt{3}, 0)$; the last three rows all equal $(0, 1/\sqrt{3})$. Clustering these rows recovers the two connected components. Spectral clustering uses eigenvectors associated with small eigenvalues as row representations.



Select the structural question before selecting the package

Choose a clustering method by its intended use, representation and group assumptions. Internal compactness can support a comparison, while external evidence is needed to interpret the groups.

Intended structure	Method and checks
Compact groups around centers	K-means: scale and outliers
Overlapping components with different spreads	GMM: covariance structure and component count
Retain coarse and fine groupings	Hierarchical: linkage and cut height
Dense regions with isolated noise	DBSCAN: radius and varying densities
Meaningful similarity graph	Spectral: construction of edges

Several methods agree on these six points because the structure is simple. Overlapping or curved groups can lead to different results.

Does a lower clustering objective reveal the true groups?

Question

Increasing K lowers the optimized within-cluster squared error. Does that prove the customer grouping is more useful?

Evidence and reasoning

No. More centers can fit the same points more closely. For outreach, define customer measurements and a time window, inspect group stability and interpretability, then check the intended response on independent evidence.

Check. Compactness supports one geometric objective. It cannot supply class names or establish practical value by itself.