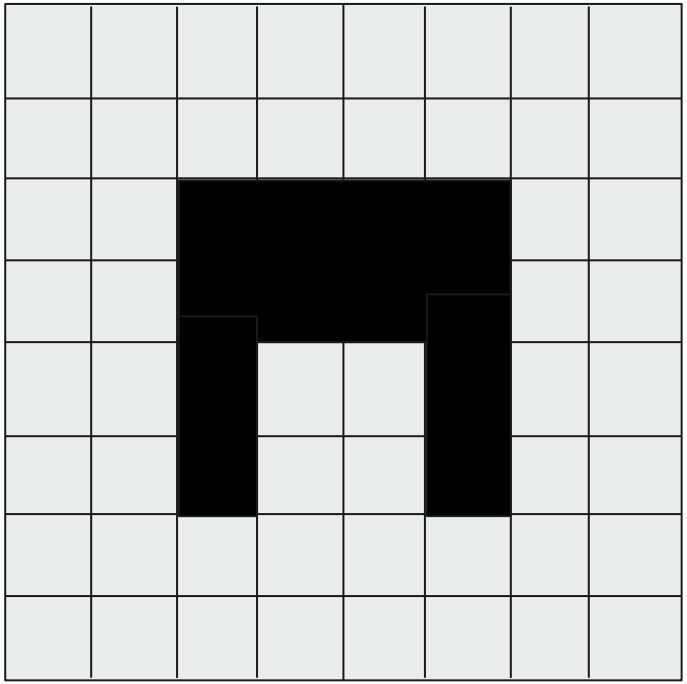
Пример



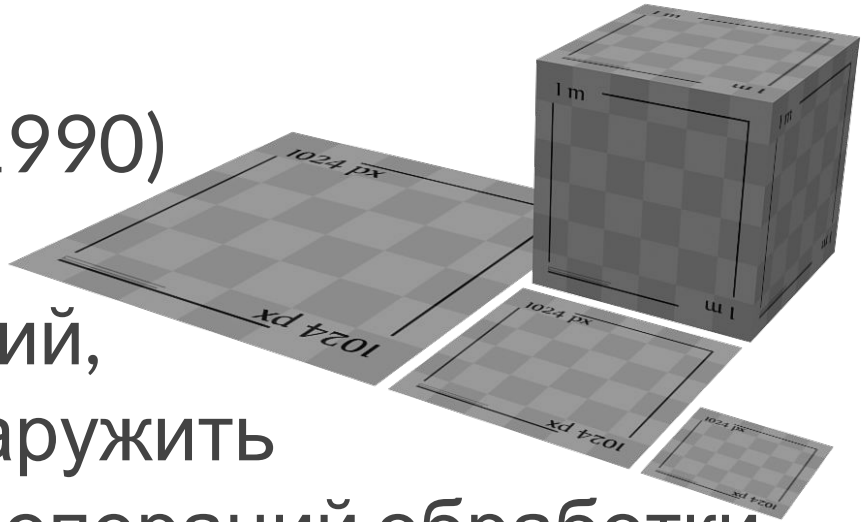
Текстуры

Текстура - это

- **Структурный подход** - множество примитивных *текселов*, расположенных в некотором регулярном или повторяющемся порядке
- **Статистический подход** - количественная характеристика распределения значений интенсивности в области изображения

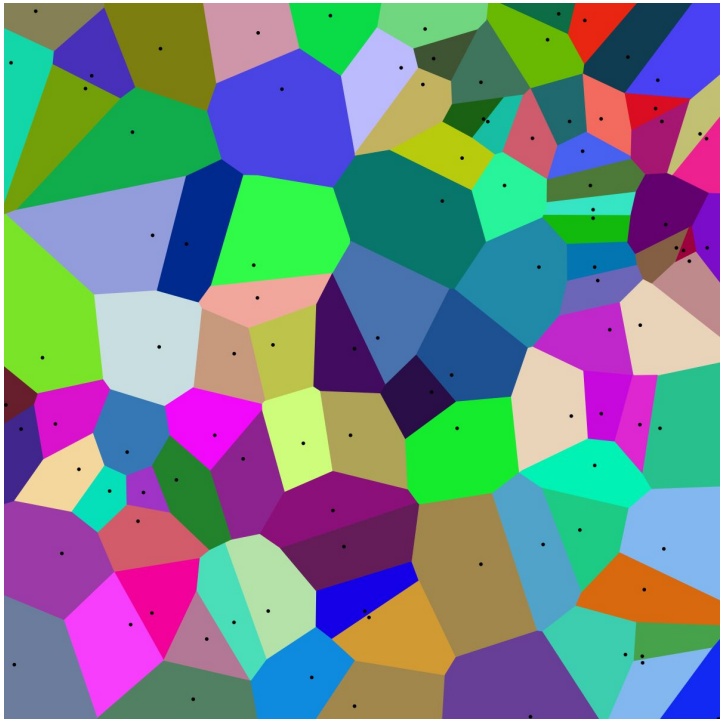
Texel

Тусерьян и Джейн (1990)



Области изображений,
которые можно обнаружить
с помощью простых операций обработки

Диаграмма Воронова



Количественные характеристики текстур

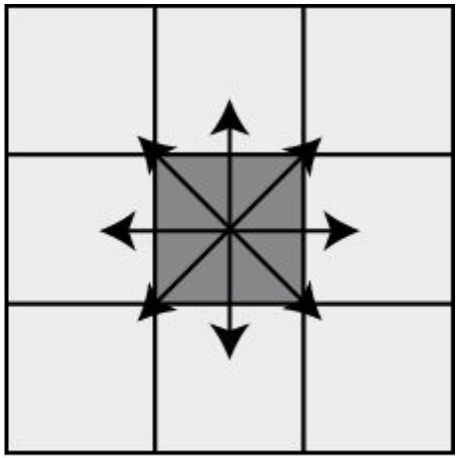
Плотность и направление краёв

$$F_{edgeness} = \frac{|\{p|Mag(p) \geq T\}|}{N}$$

$$F_{magdir} = \Big(H_{mag}(R), H_{dir}(R)\Big)$$

Локальное двоичное разбиение

$$b_i = \begin{cases} 0, p \geq p_i \\ 1, p < p_i \end{cases}$$

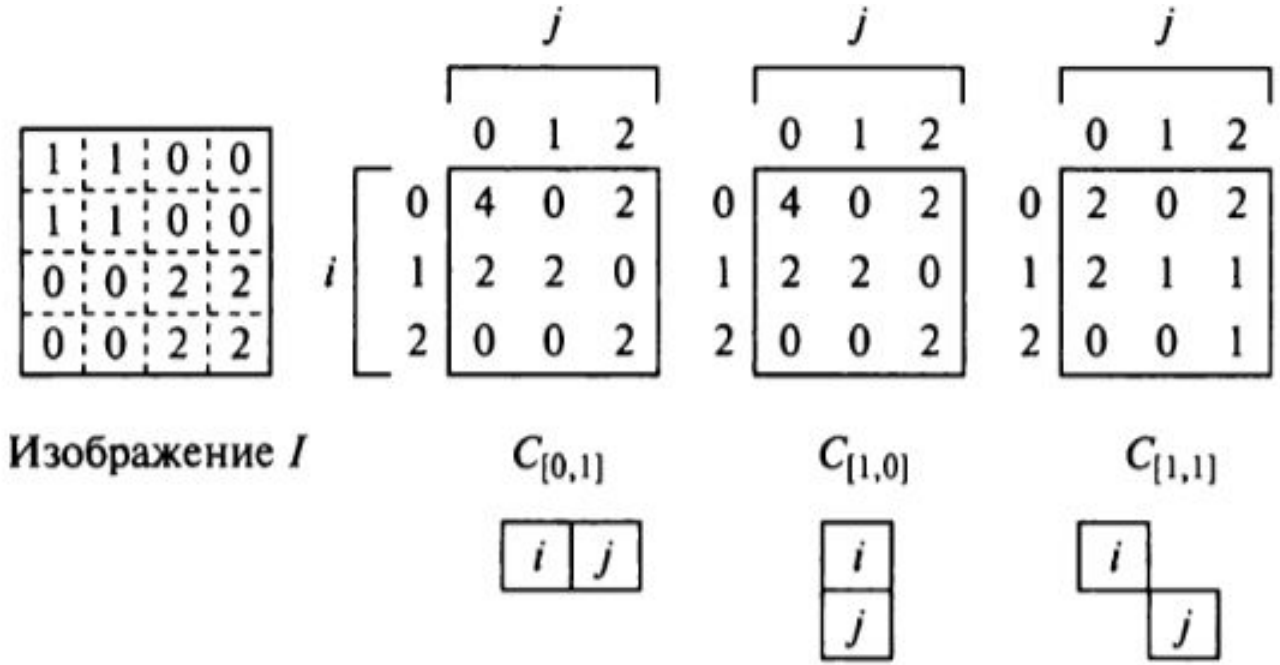


8-Connectivity

Матрица вхождений

Co-occurrence matrix - двумерный массив **C**, в котором индексы строк и столбцов образуют множество **V** допустимых на изображении значений пикселов.

$$C_d[i,j] = |\{[r,c]|I[r,c] = i, I[r+dr, c+dc] = j\}|$$



Матрица вхождений

Нормированная $$N_d[i,j] = \frac{C_d[i,j]}{\sum_i \sum_j C_d[i,j]}$$

Симметричная $$S_d[i,j] = C_d[i,j] + C_{-d}[i,j]$$

$$\textit{Энергия} = \sum_i \sum_j N_d^2[i,j]$$

$$\textit{Энтропия} = -\sum_i \sum_j N_d[i,j] \log_2 N_d[i,j]$$

$$\textit{Контраст} = \sum_i \sum_j (i-j)^2 N_d[i,j]$$

$$\textit{Однородность} = \sum_i \sum_j \frac{N_d[i,j]}{1+|i-j|}$$

$$\textit{Корреляция} = \frac{\sum_i \sum_j (i-\mu_i)(j-\mu_j) N_d[i,j]}{\sigma_i \sigma_j}$$

Энергетические текстурные характеристики Лавса

$$
\begin{array}{lll}
\text{L5} & \text{(Level)} & = [\ 1 \quad 4 \quad 6 \quad 4 \quad 1] \\
\text{E5} & \text{(Edge)} & = [-1 \quad -2 \quad 0 \quad 2 \quad 1] \\
\text{S5} & \text{(Spot)} & = [-1 \quad 0 \quad 2 \quad 0 \quad -1] \\
\text{R5} & \text{(Ripple)} & = [\ 1 \quad -4 \quad 6 \quad -4 \quad 1]
\end{array}
$$

$$\begin{bmatrix} -1 \\ -2 \\ 0 \\ 2 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix} = \begin{bmatrix} -1 & -4 & -6 & -4 & -1 \\ -2 & -8 & -12 & -8 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 8 & 12 & 8 & 2 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}$$

Автокорреляция и спектр мощности

$$\rho(dr, dc) = \frac{\sum_{r=0}^{N} \sum_{c=0}^{N} I[r, c] I[r + dr, c + dc]}{\sum_{r=0}^{N} \sum_{c=0}^{N} I^2[r, c]}$$

$$= \frac{I[r,c] \circ I_d[r,c]}{I[r,c] \circ I[r,c]}$$

Кластеризация

Задача

Группировка пикселей так, чтобы объекты внутри группы (кластера) были похожи (или связаны) друг с другом и отличались от объектов в других группах.

Чем больше сходство (или однородность) внутри группы, больше различие между группами, тем более четкая кластеризация.

Проблема

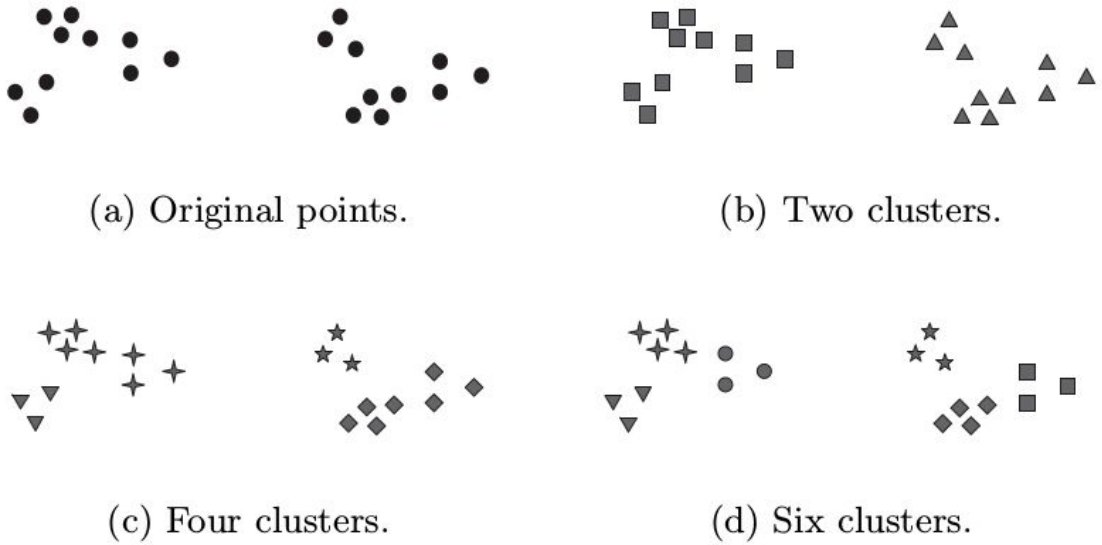
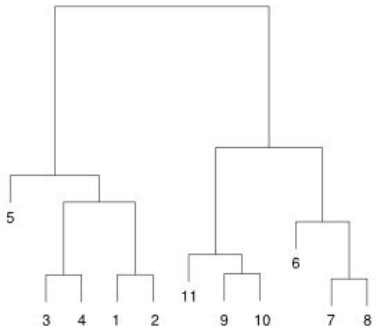
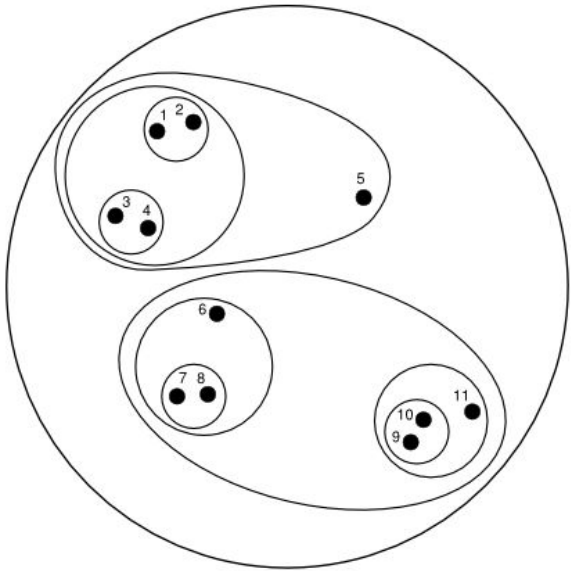


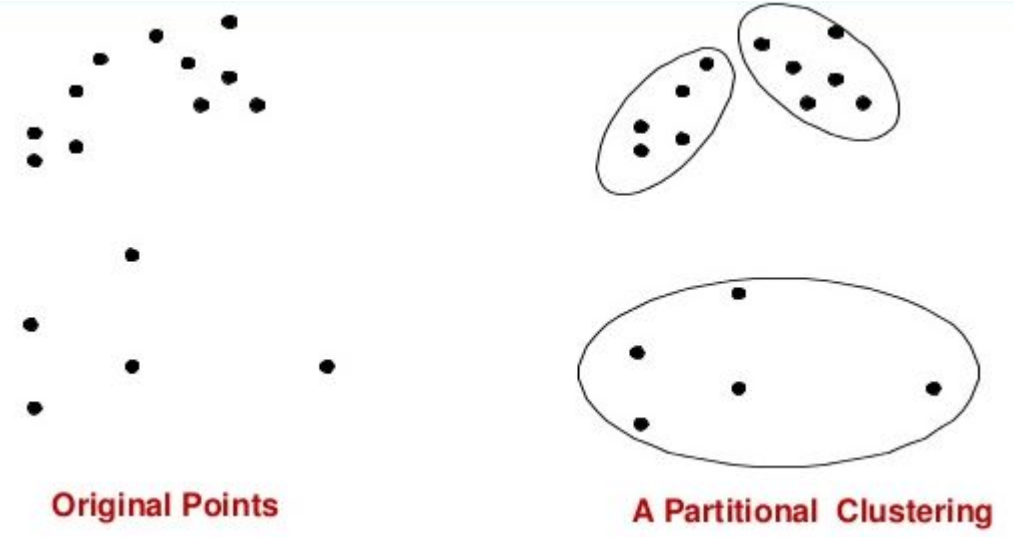
Figure 8.1. Different ways of clustering the same set of points.

Hierarchical vs Partitional

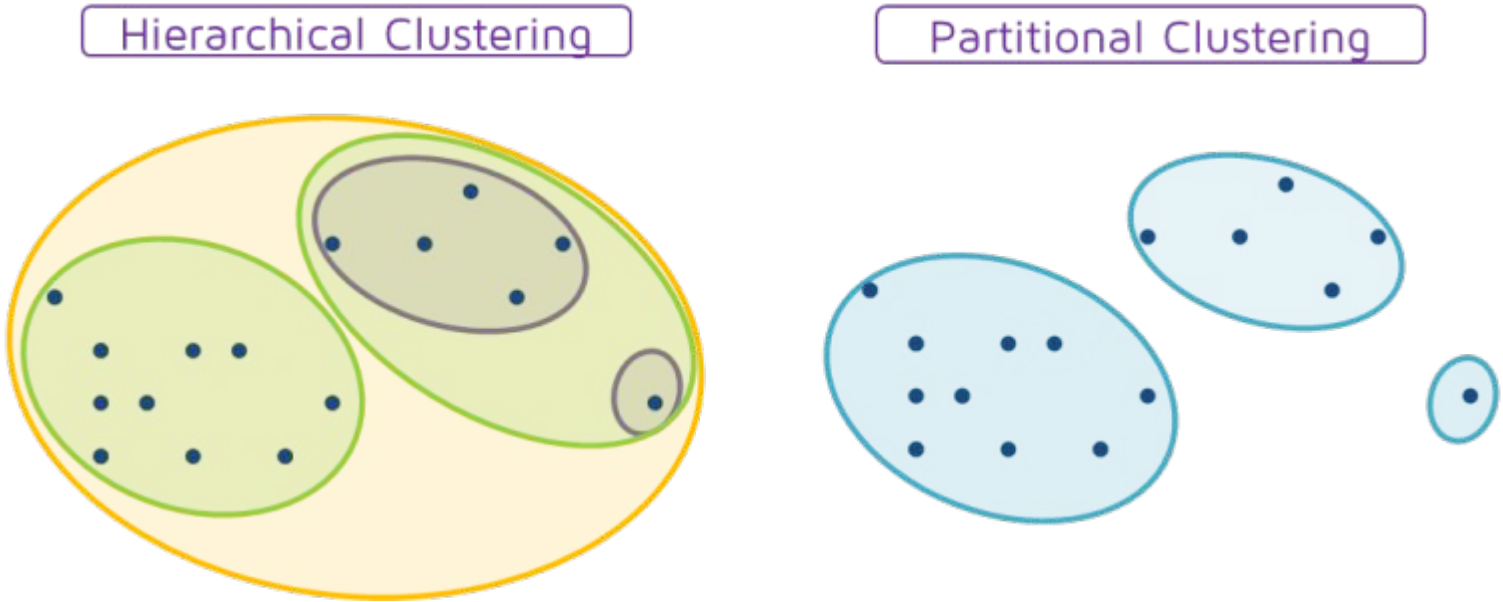




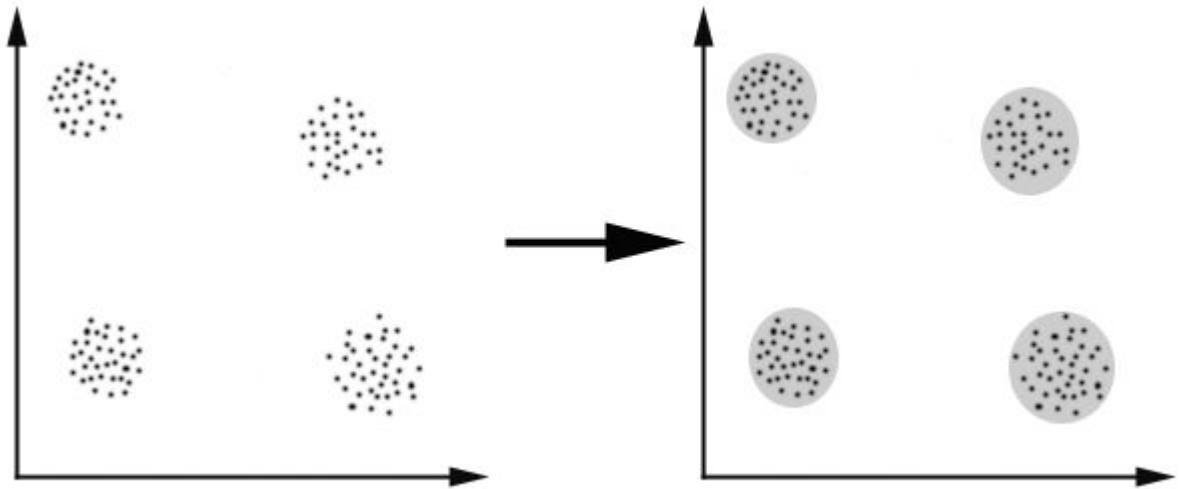
Hierarchical vs Partitional



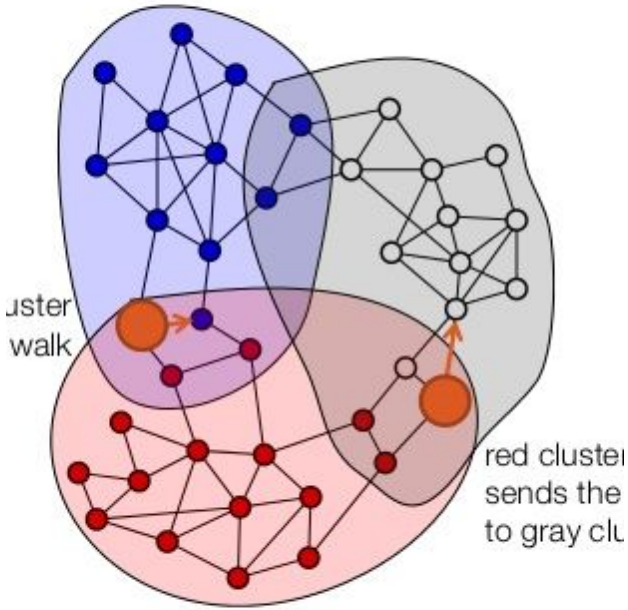
Hierarchical vs Partitional



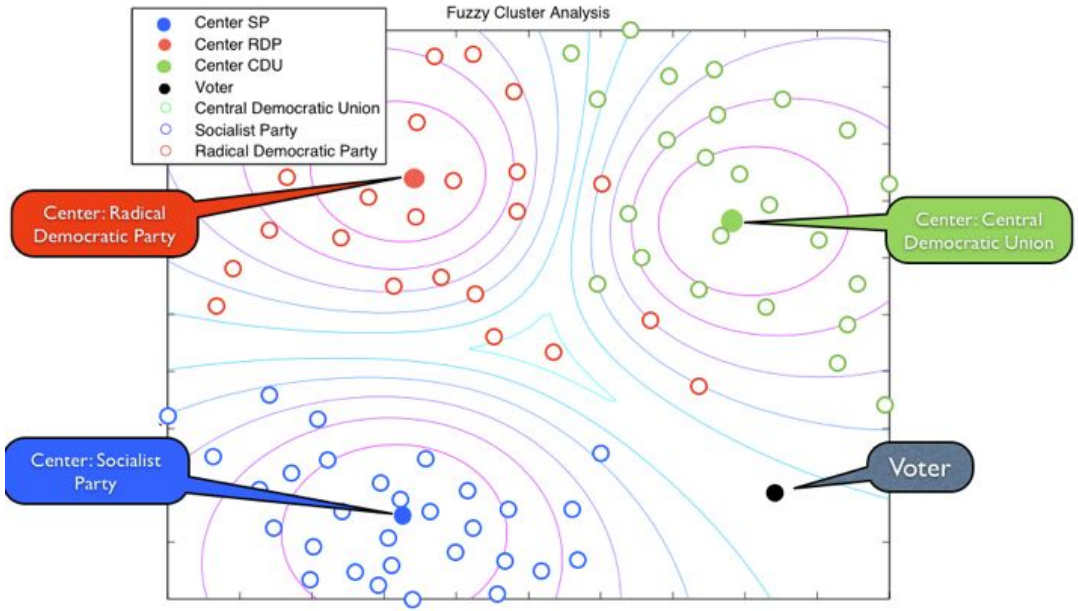
Exclusive vs Overlapping vs Fuzzy



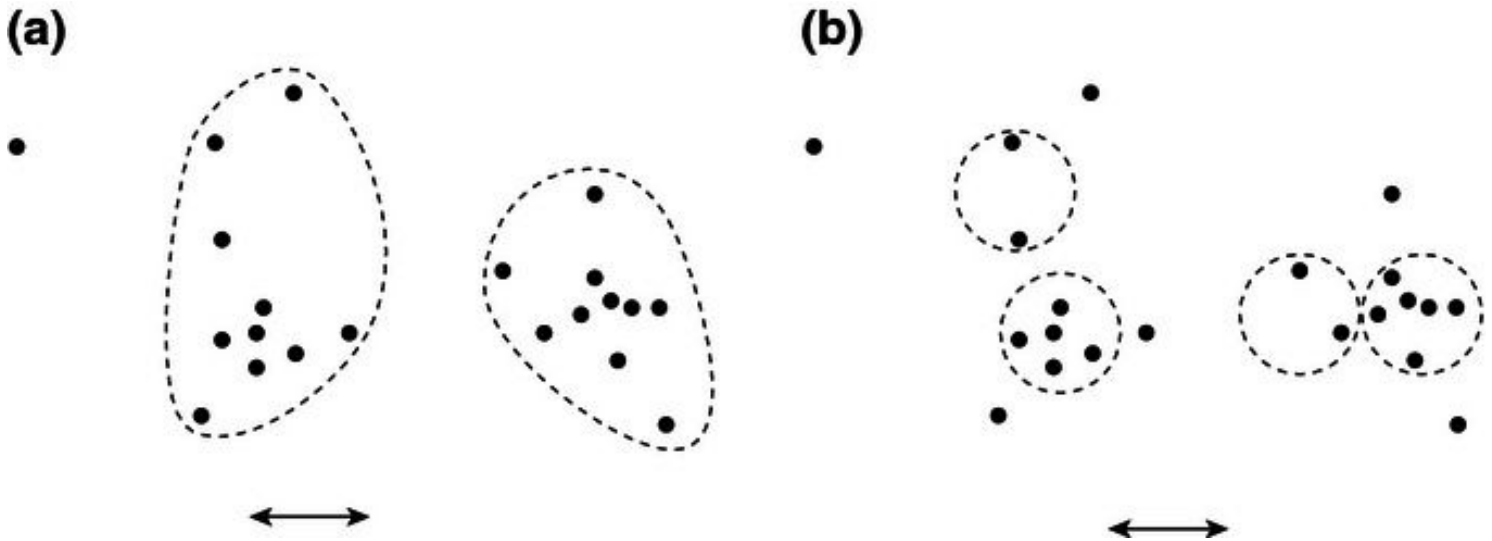
Exclusive vs Overlapping vs Fuzzy



Exclusive vs Overlapping vs Fuzzy



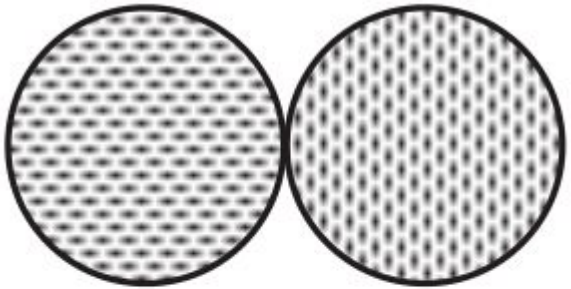
Complete vs Partial



Well-Separated

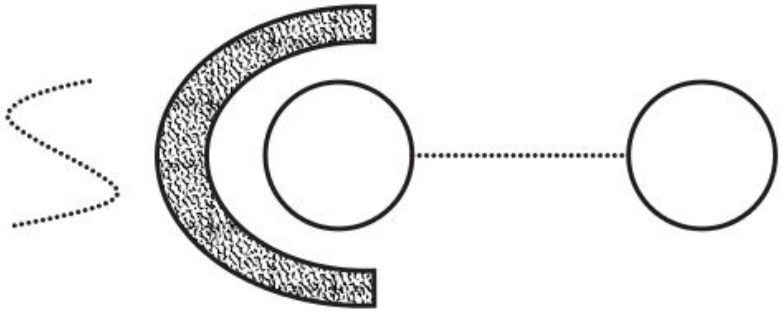
(a) Well-separated clusters. Each point is closer to all of the points in its cluster than to any point in another cluster.

Center-Based



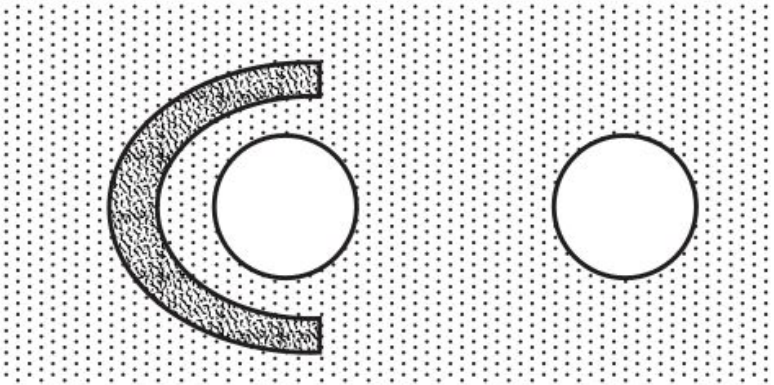
(b) Center-based clusters. Each point is closer to the center of its cluster than to the center of any other cluster.

Contiguity-Based



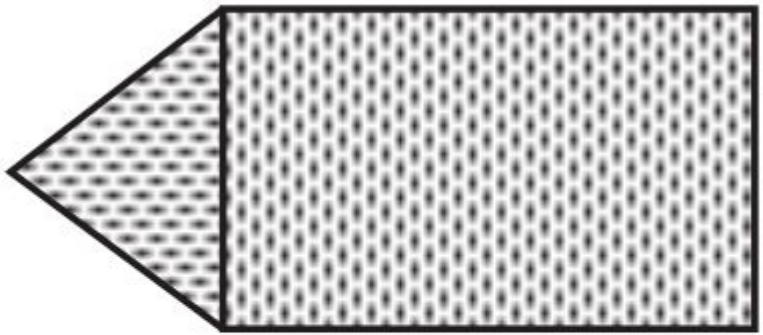
(c) Contiguity-based clusters. Each point is closer to at least one point in its cluster than to any point in another cluster.

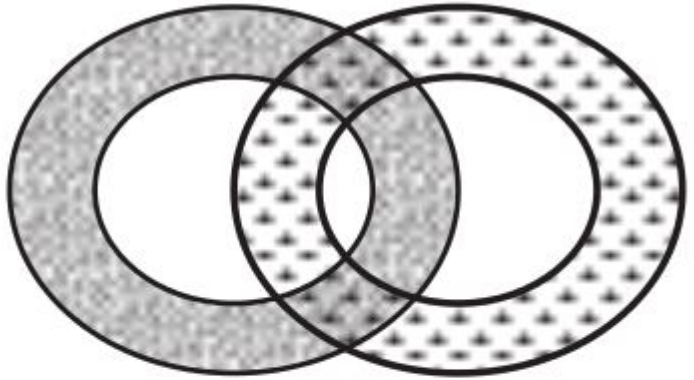
Density-Based



(d) Density-based clusters. Clusters are regions of high density separated by regions of low density.

Shared-Property



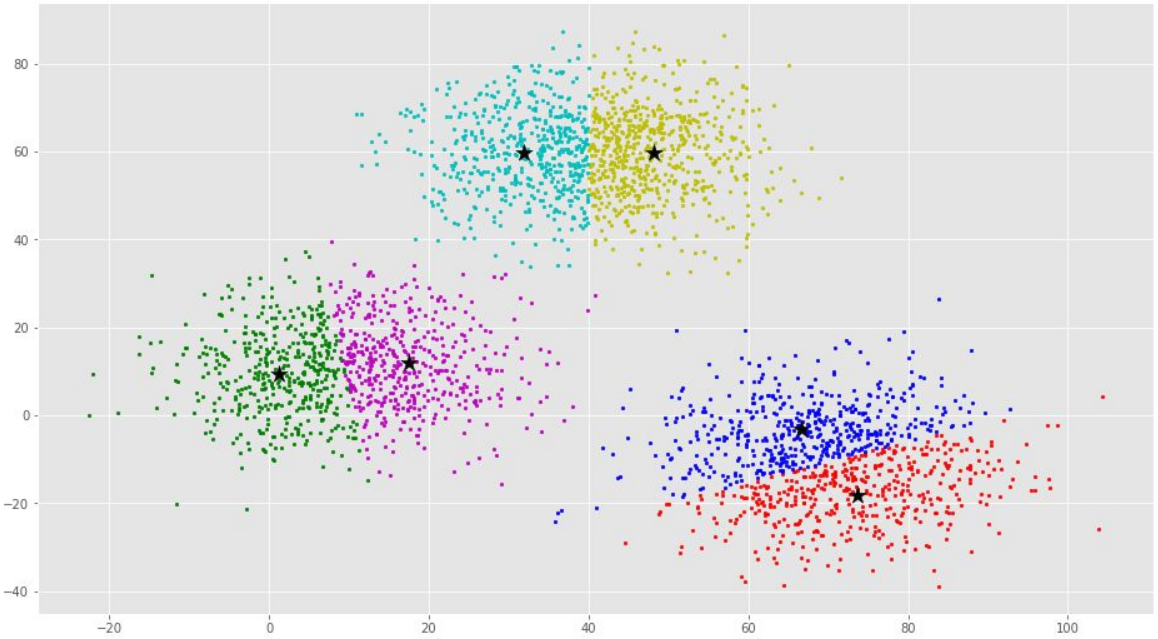


(e) Conceptual clusters. Points in a cluster share some general property that derives from the entire set of points. (Points in the intersection of the circles belong to both.)

Methods

- K-means
- Agglomerative Hierarchical Clustering
- DBSCAN

K-means



Basic K-means algorithm

1. Select *K* points as initial centroids.
2. **repeat**
3. Form *K* clusters by assigning each point to its closest centroid
4. Recompute the centroid of each cluster.
5. **until** Centroids do not change.

Basic K-means algorithm

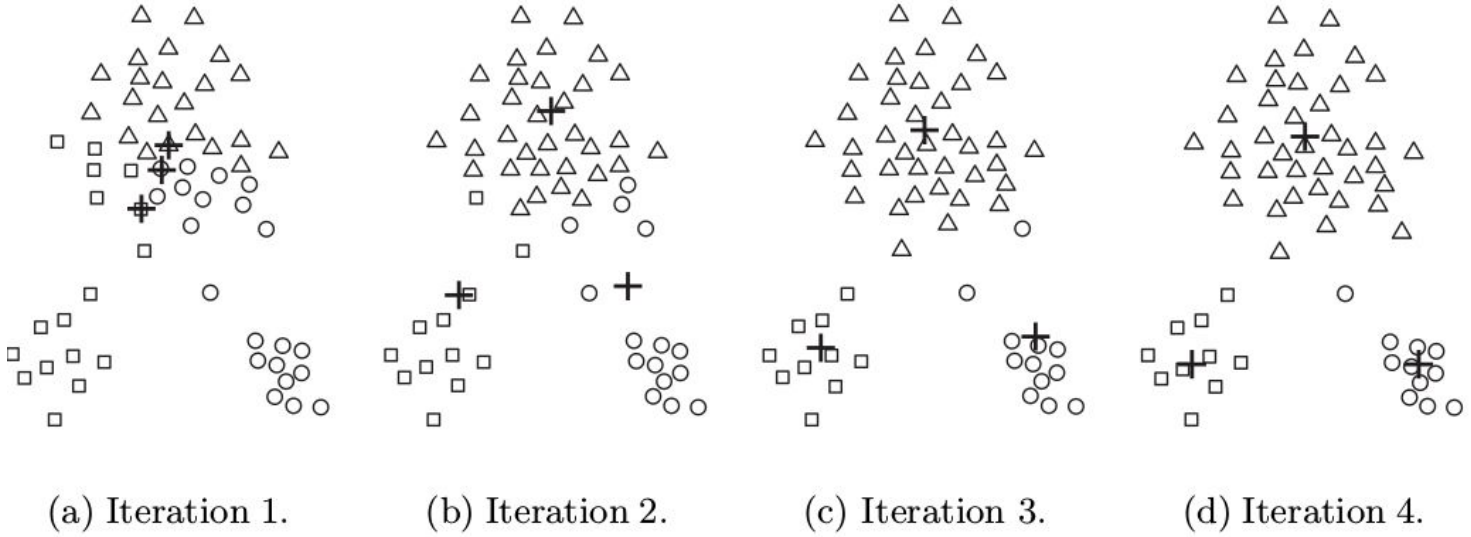


Figure 8.3. Using the K-means algorithm to find three clusters in sample data.

Assigning Points to the Closest Centroid

Зависит от типа объектов, подлежащих кластеризации

Centroids and Objective Function

Table 8.1. Table of notation.

Symbol	Description
$\mathbf{x}$	An object.
C_i	The i^{th} cluster.
$\mathbf{c}_i$	The centroid of cluster C_i.
$\mathbf{c}$	The centroid of all points.
m_i	The number of objects in the i^{th} cluster.
m	The number of objects in the data set.
K	The number of clusters.

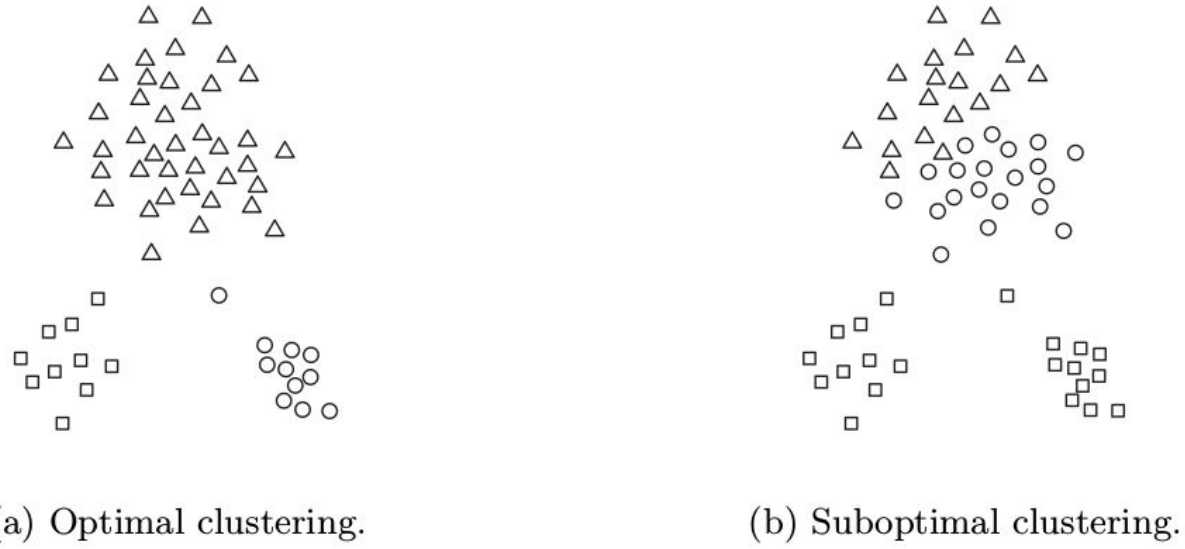
Centroids and Objective Function

$$\mathrm{SSE} = \sum_{i=1}^{K} \sum_{\mathbf{x} \in C_i} dist(\mathbf{c}_i, \mathbf{x})^2$$

Centroids and Objective Function

$$\mathbf{c}_i = \frac{1}{m_i} \sum_{\mathbf{x} \in C_i} \mathbf{x}$$

Выбор начального приближения

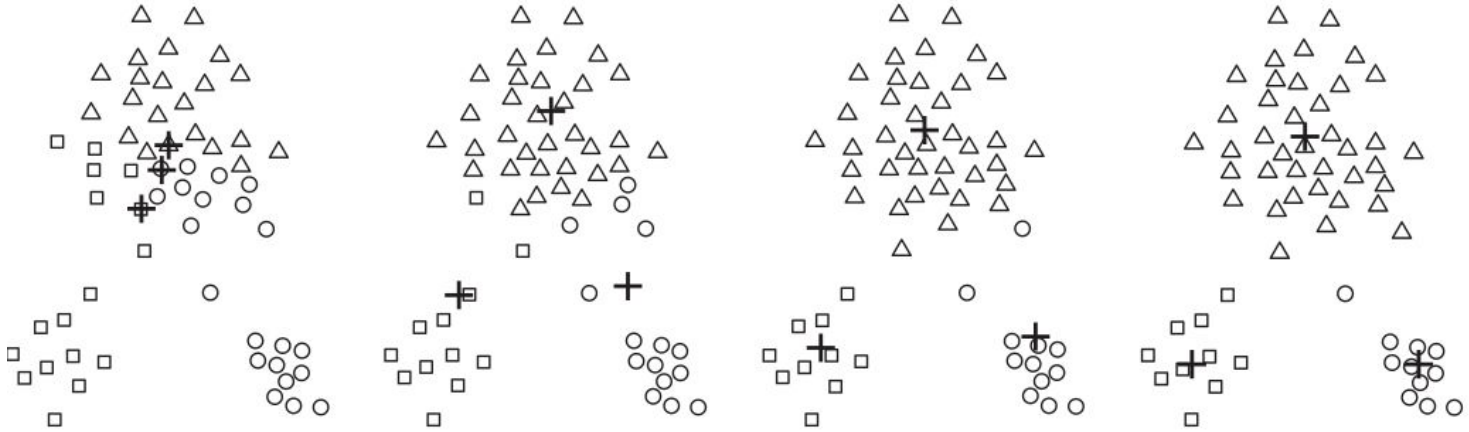


(a) Optimal clustering.

(b) Suboptimal clustering.

Figure 8.4. Three optimal and non-optimal clusters.

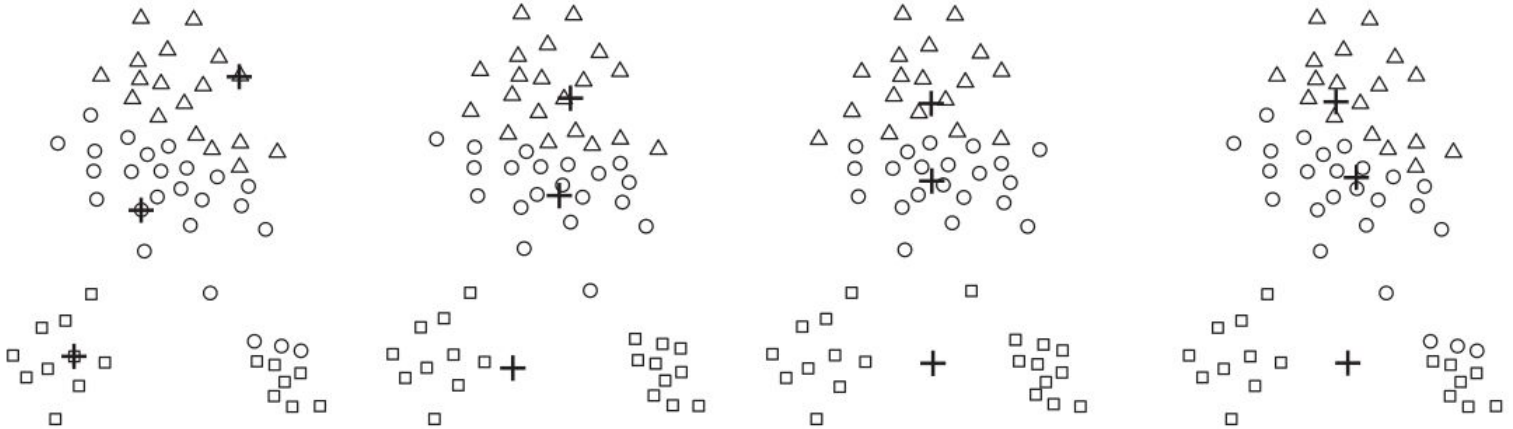
Выбор начального приближения



(a) Iteration 1. (b) Iteration 2. (c) Iteration 3. (d) Iteration 4.

Figure 8.3. Using the K-means algorithm to find three clusters in sample data.

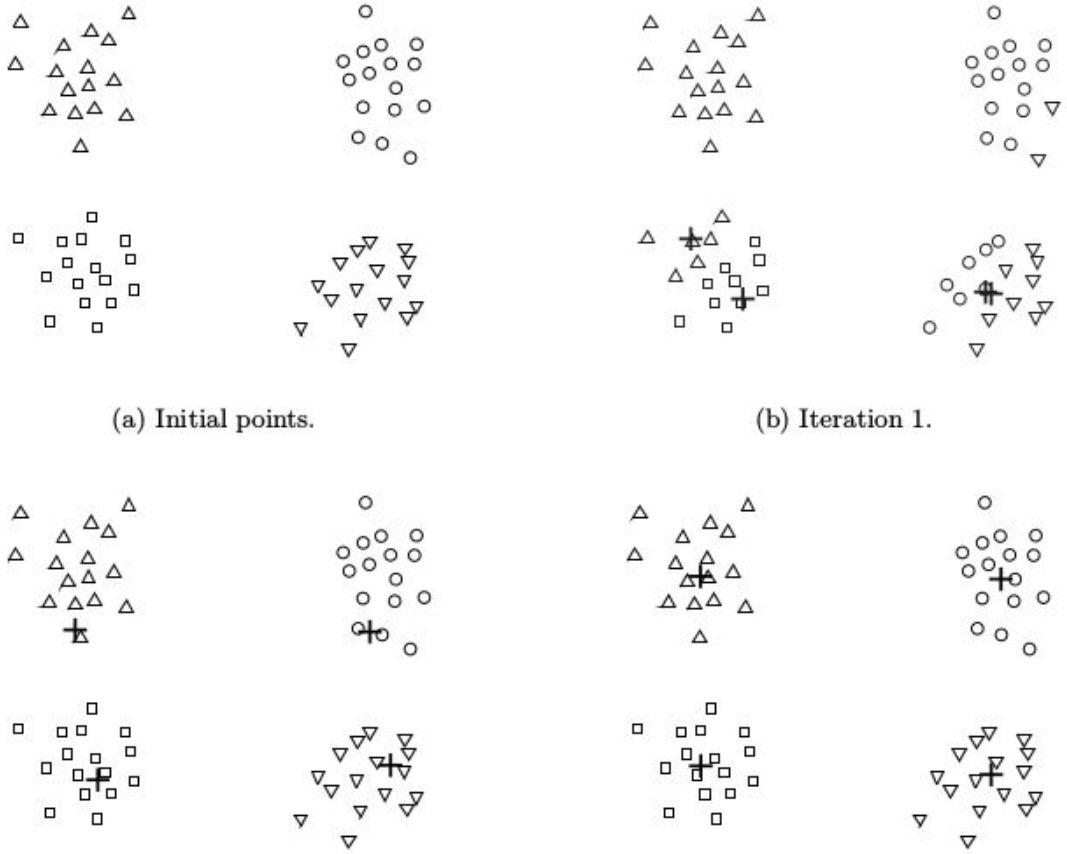
Выбор начального приближения



(a) Iteration 1. (b) Iteration 2. (c) Iteration 3. (d) Iteration 4.

Figure 8.5. Poor starting centroids for K-means.

Limit of Random Initialization



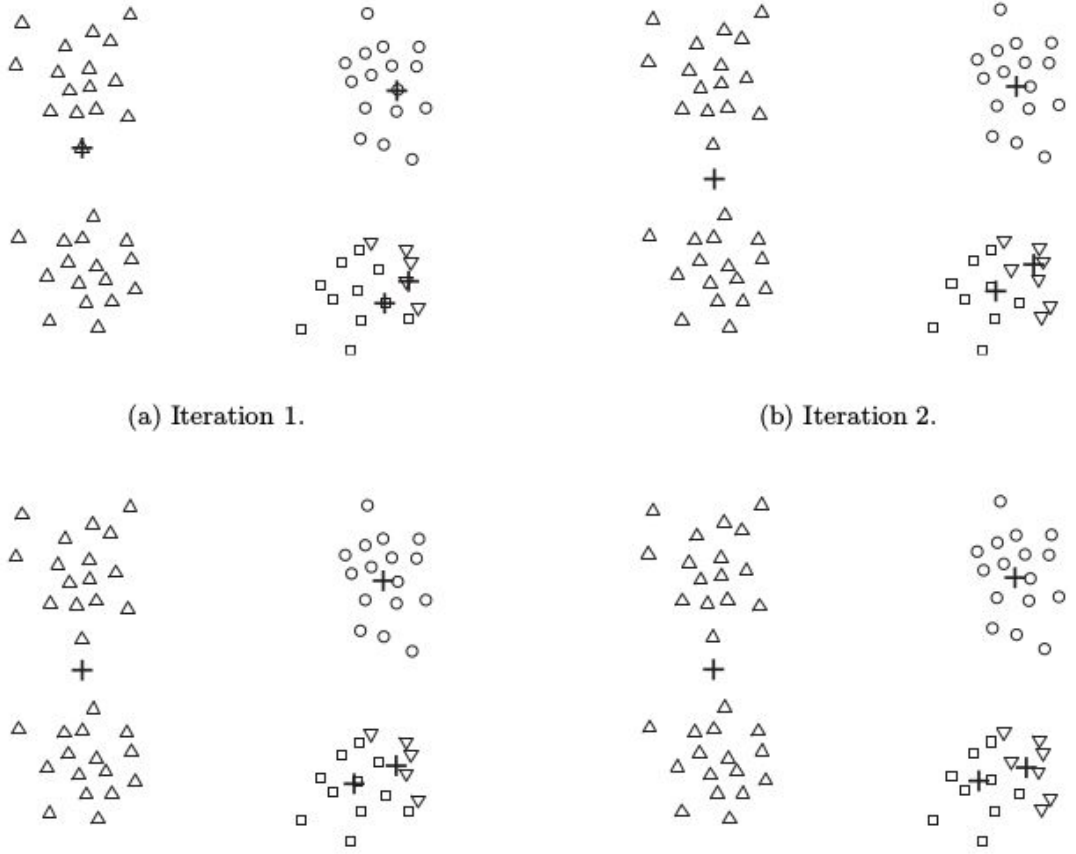
(a) Initial points.

(b) Iteration 1.

(c) Iteration 2.

(d) Iteration 3.

Figure 8.6. Two pairs of clusters with a pair of initial centroids within each pair of clusters.



(a) Iteration 1.

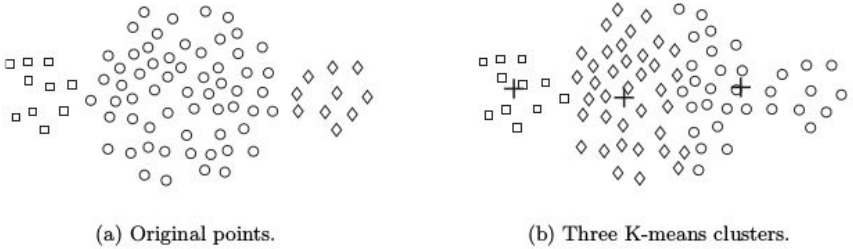
(b) Iteration 2.

(c) Iteration 3.

(d) Iteration 4.

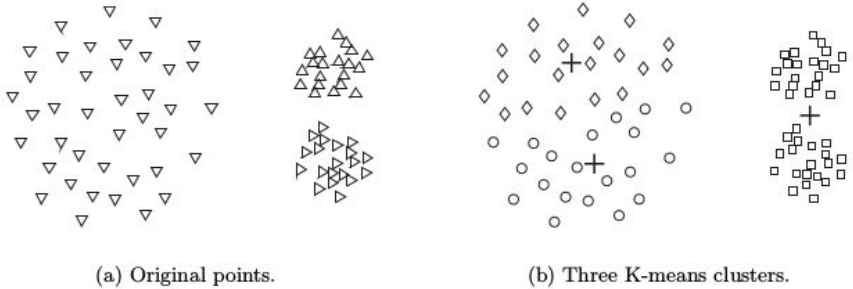
Figure 8.7. Two pairs of clusters with more or fewer than two initial centroids within a pair of clusters.

Плюсы и минусы



(a) Original points. (b) Three K-means clusters.

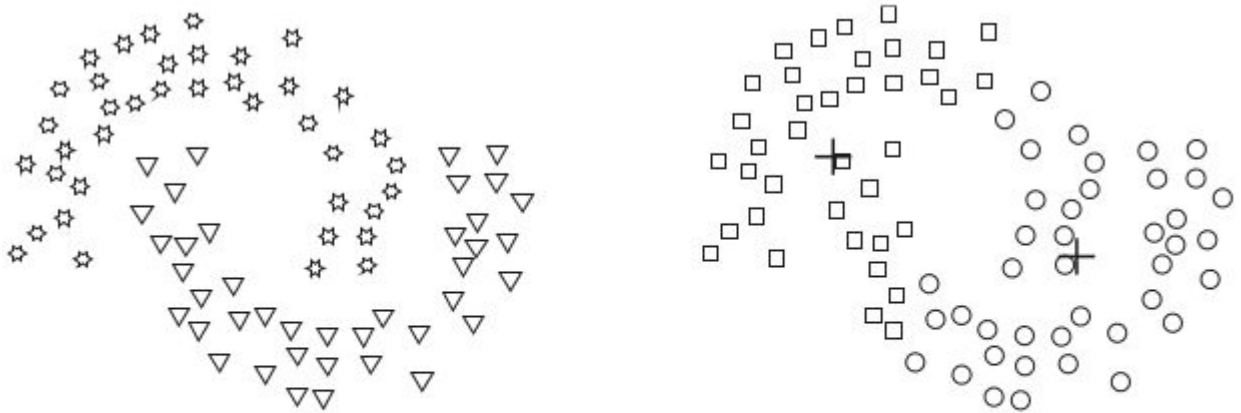
Figure 8.9. K-means with clusters of different size.



(a) Original points. (b) Three K-means clusters.

Figure 8.10. K-means with clusters of different density.

Плюсы и минусы



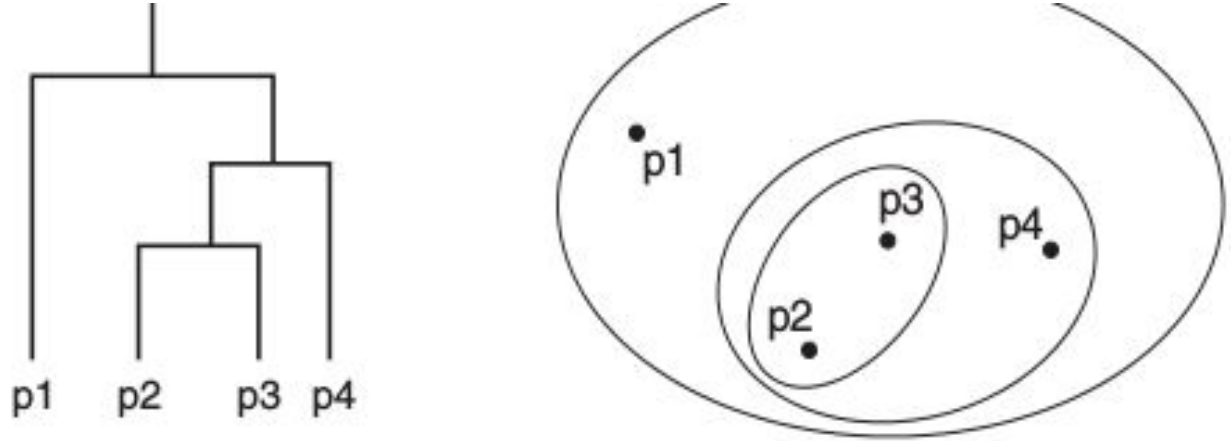
(a) Original points.

(b) Two K-means clusters.

Figure 8.11. K-means with non-globular clusters.

Иерархическая кластеризация

- Agglomerative
- Divisive



(a) Dendrogram. (b) Nested cluster diagram.

Figure 8.13. A hierarchical clustering of four points shown as a dendrogram and as nested clusters.

Basic Agglomerative Hierarchical Clustering Algorithm

1. Compute the proximity matrix, if necessary.
2. **repeat**
3. Merge the closest two clusters.
4. Update the proximity matrix to reflect the proximity between the new cluster and the original clusters.
5. **until** Only one cluster remains

Определение близости между кластерами

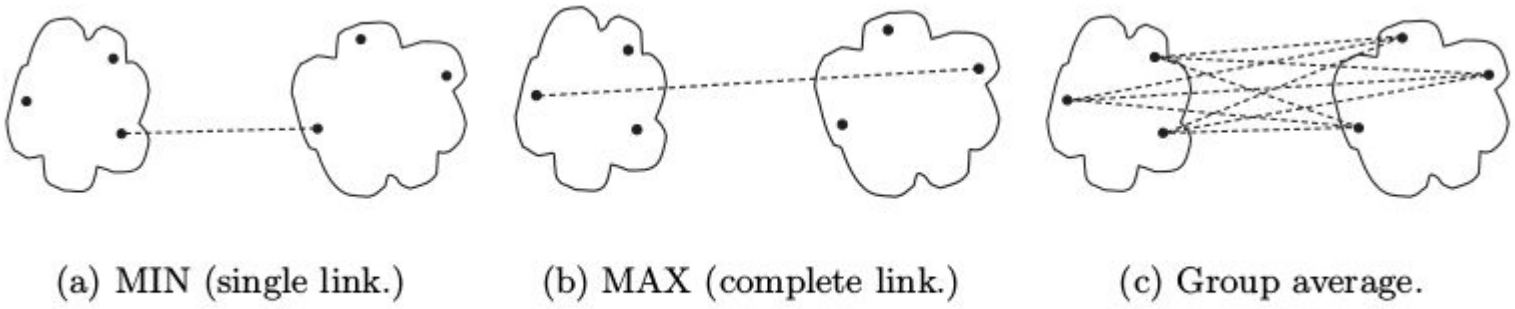
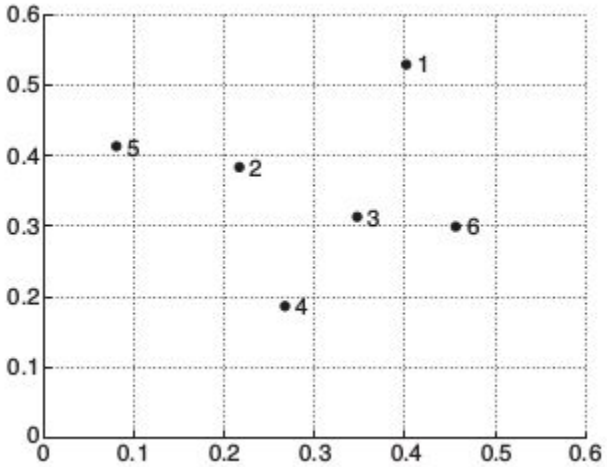


Figure 8.14. Graph-based definitions of cluster proximity

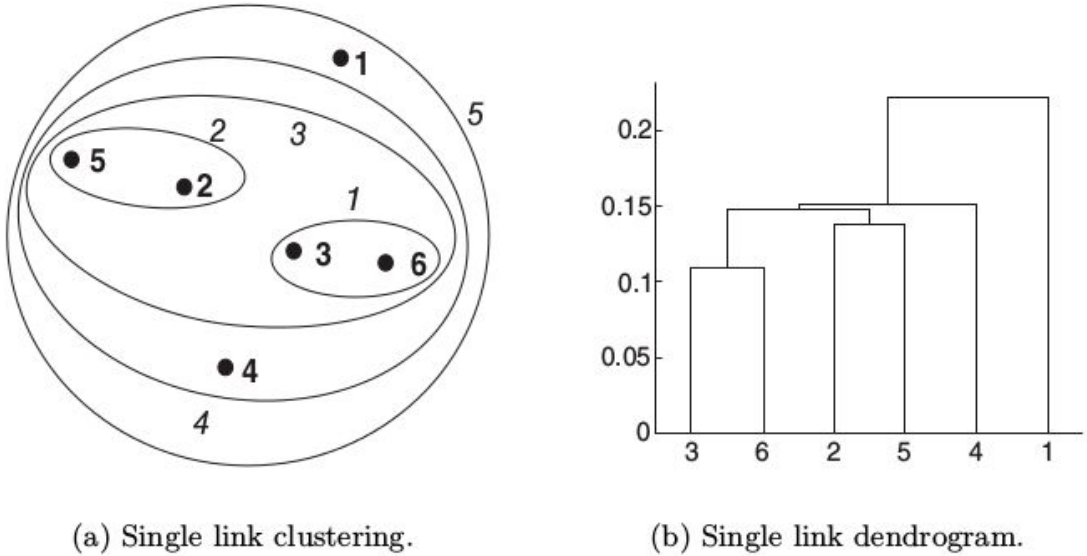
Пример

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00



Point	x Coordinate	y Coordinate
p1	0.40	0.53
p2	0.22	0.38
p3	0.35	0.32
p4	0.26	0.19
p5	0.08	0.41
p6	0.45	0.30

MIN

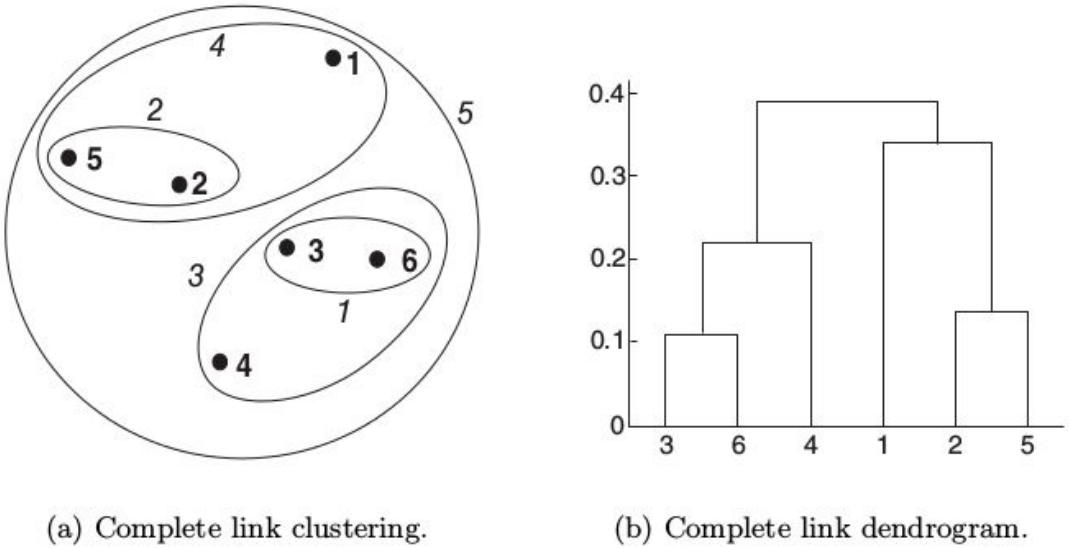


(a) Single link clustering.

(b) Single link dendrogram.

Figure 8.16. Single link clustering of the six points shown in Figure 8.15.

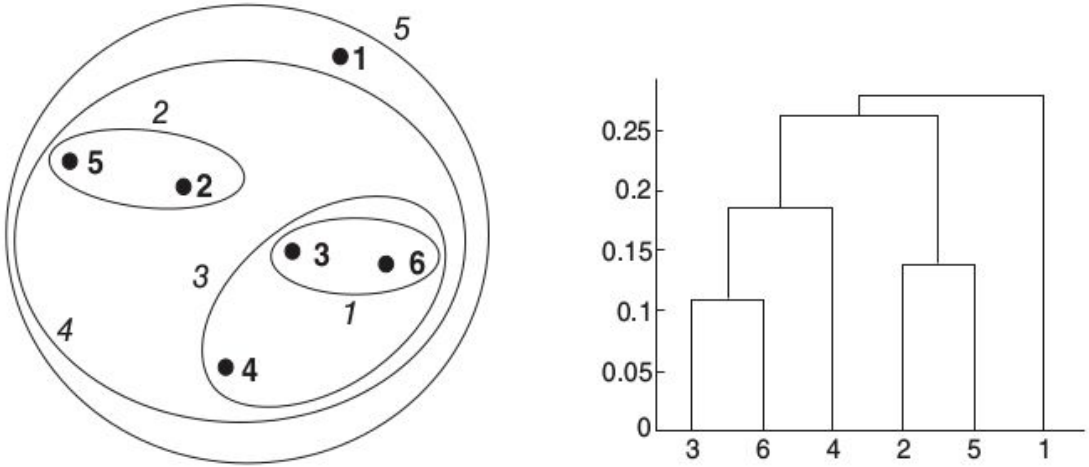
MAX



(a) Complete link clustering. (b) Complete link dendrogram.

Figure 8.17. Complete link clustering of the six points shown in Figure 8.15.

Group Average



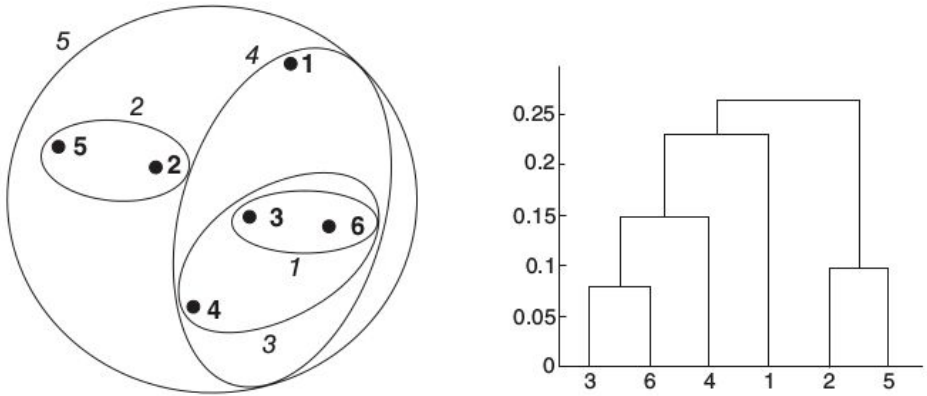
(a) Group average clustering.

(b) Group average dendrogram.

Figure 8.18. Group average clustering of the six points shown in Figure 8.15.

Table 8.5. Table of Lance-Williams coefficients for common hierarchical clustering approaches.

Clustering Method	α_A	α_B	β	γ
Single Link	$1/2$	$1/2$	0	$-1/2$
Complete Link	$1/2$	$1/2$	0	$1/2$
Group Average	$\frac{m_A}{m_A+m_B}$	$\frac{m_B}{m_A+m_B}$	0	0
Centroid	$\frac{m_A}{m_A+m_B}$	$\frac{m_B}{m_A+m_B}$	$\frac{-m_A m_B}{(m_A+m_B)^2}$	0
Ward's	$\frac{m_A+m_Q}{m_A+m_B+m_Q}$	$\frac{m_B+m_Q}{m_A+m_B+m_Q}$	$\frac{-m_Q}{m_A+m_B+m_Q}$	0



(a) Ward's clustering. (b) Ward's dendrogram.

Figure 8.19. Ward's clustering of the six points shown in Figure 8.15.

DBSCAN

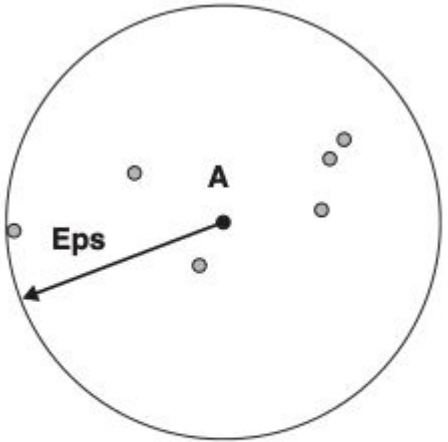


Figure 8.20. Center-based density.

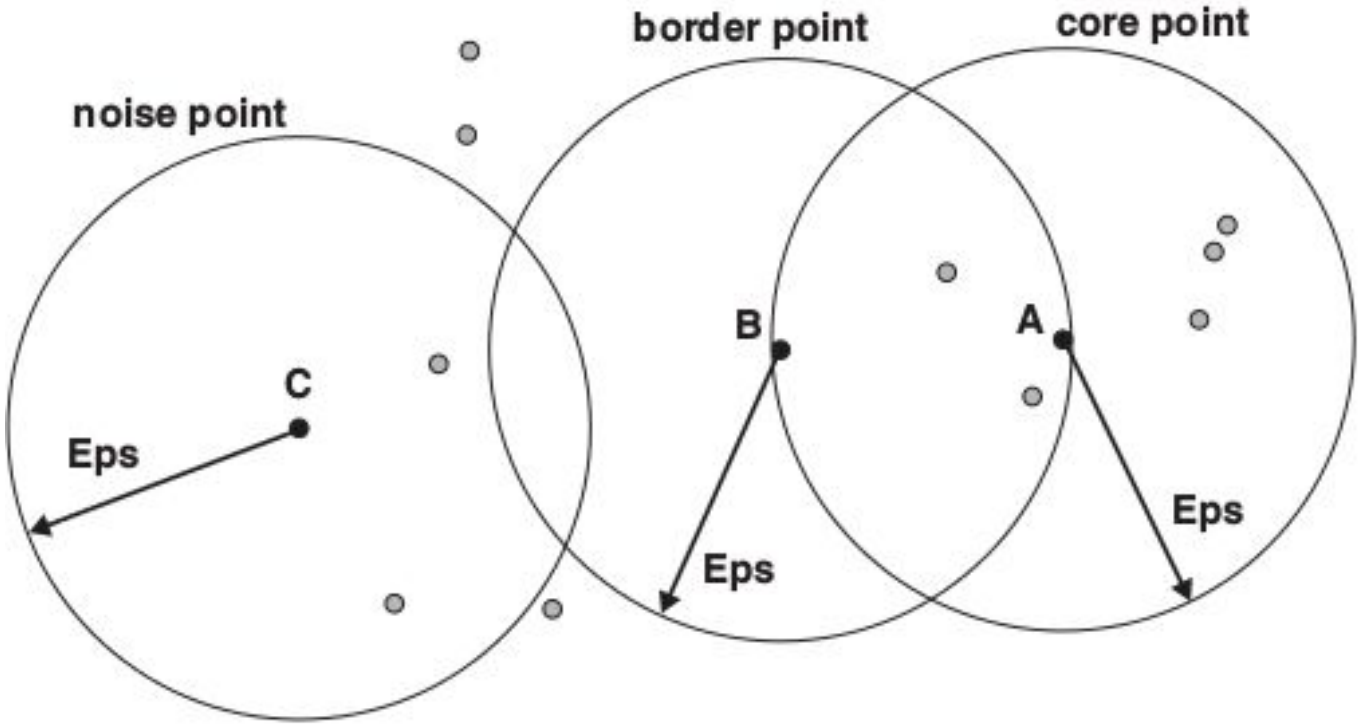


Figure 8.21. Core, border, and noise points.

The DBSCAN Algorithm

1. Label all points as core, border, or noise points.
2. Eliminate noise points.
3. Put an edge between all core points that are within Eps of each other.
4. Make each group of connected core points into a separate cluster.
5. Assign each border point to one of the clusters of its associated core points.

Определение параметров

Figure 8.22. Sample data.

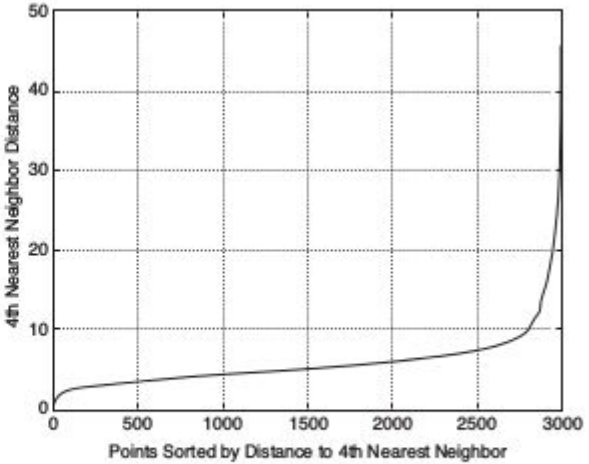
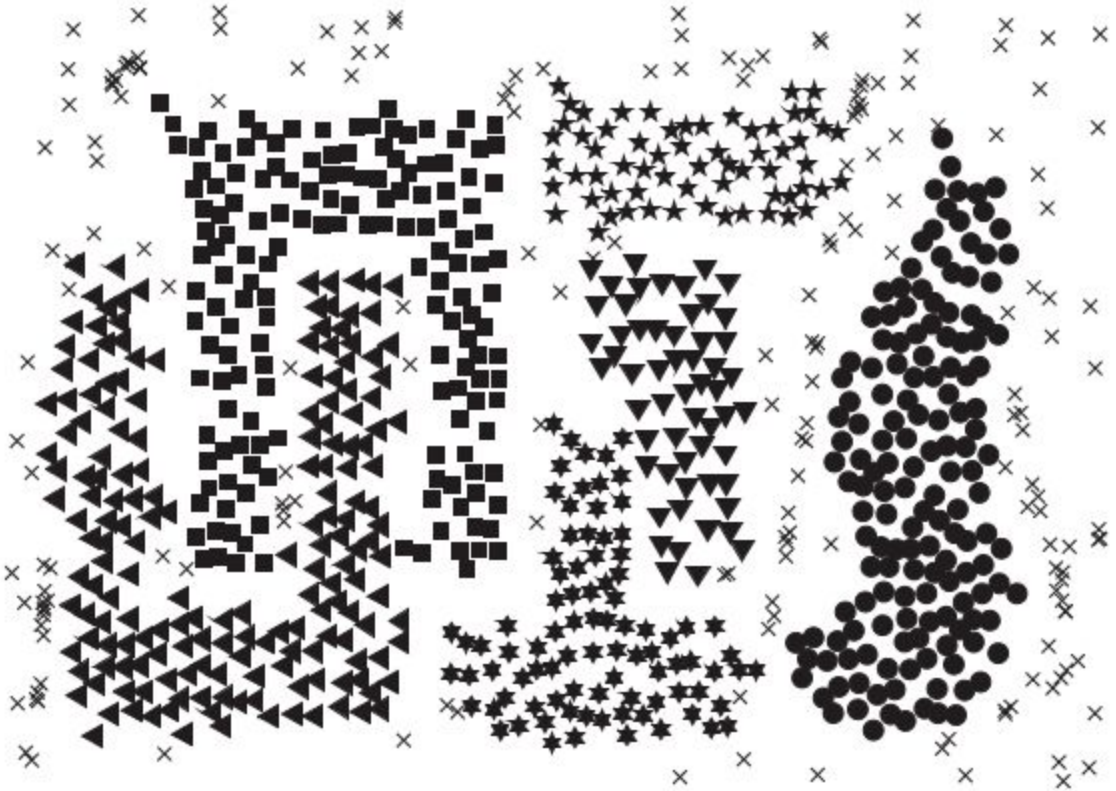
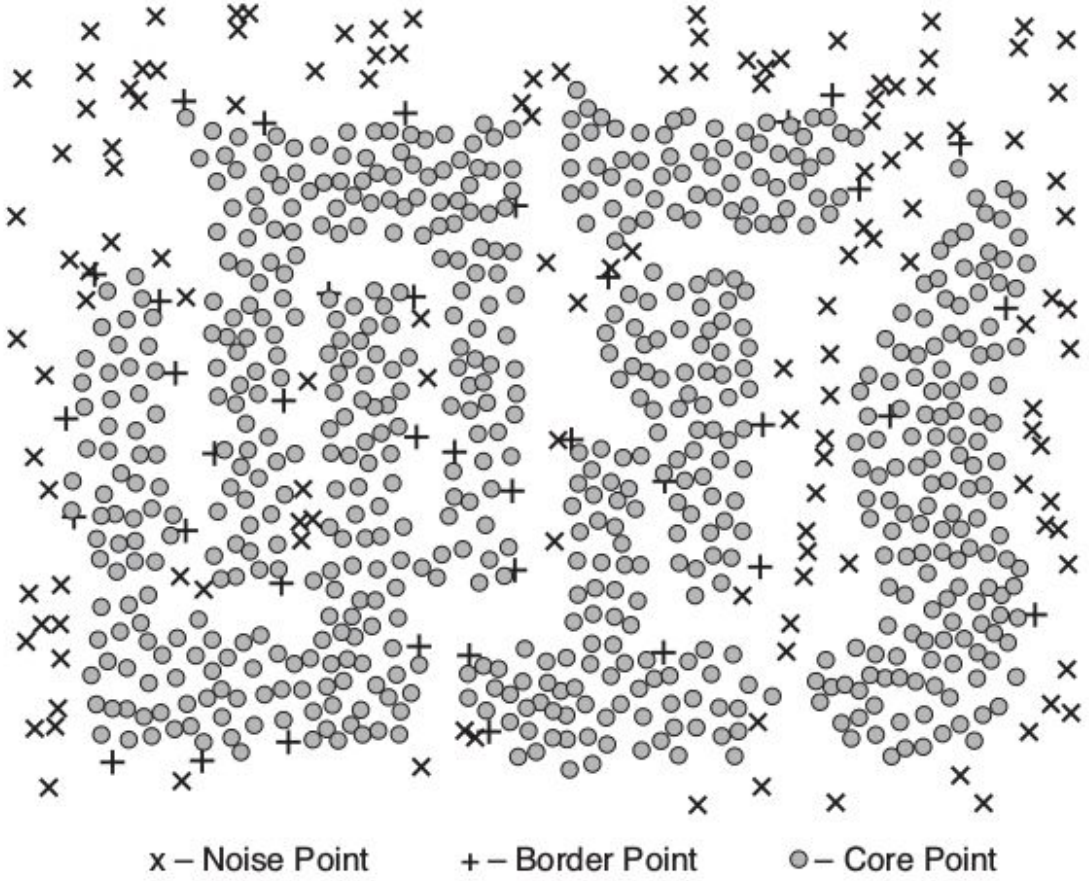


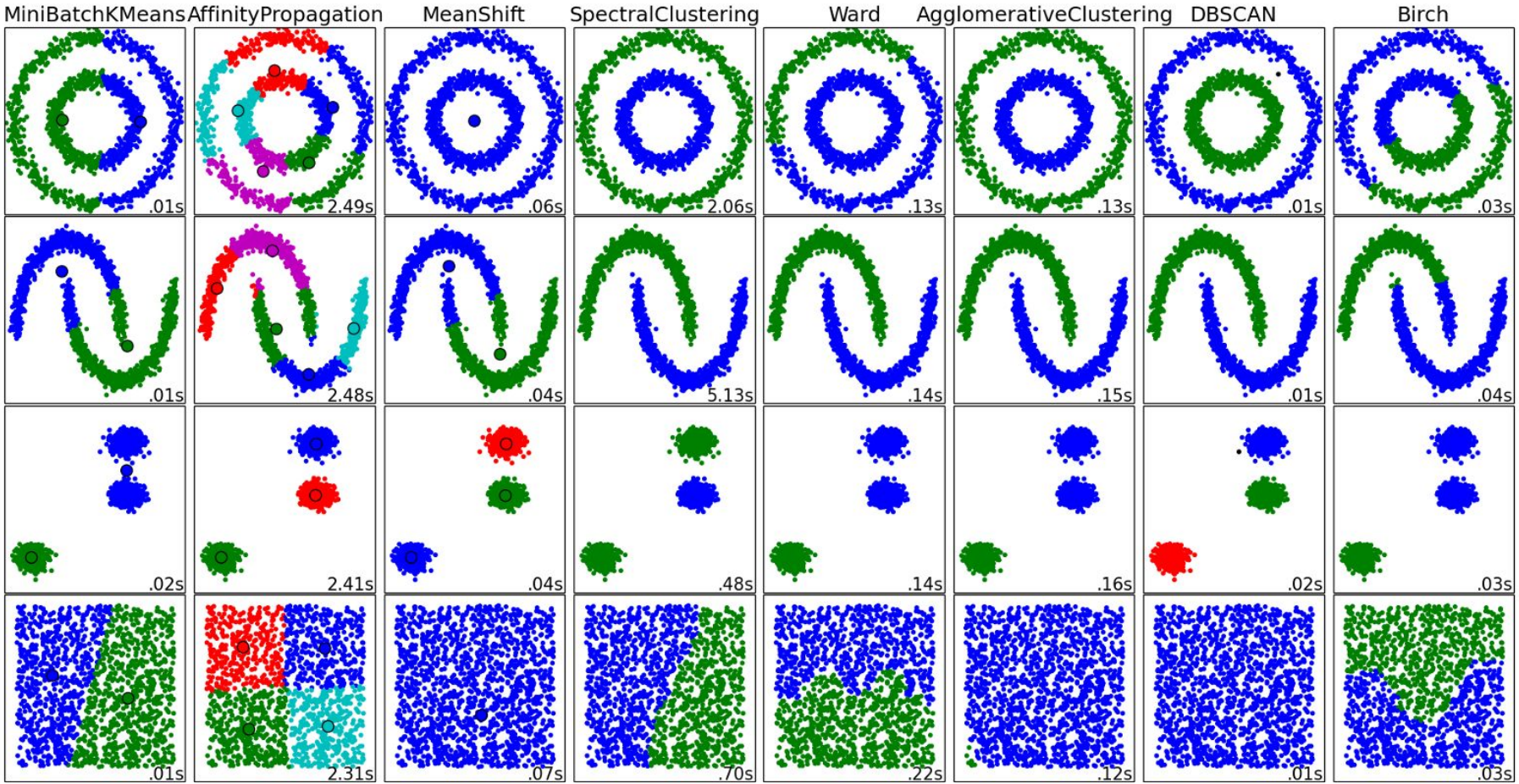
Figure 8.23. K-dist plot for sample data.



(a) Clusters found by DBSCAN.



(b) Core, border, and noise points.


MiniBatchKMeans
AffinityPropagation
MeanShift
SpectralClustering
Ward
AgglomerativeClustering
DBSCAN
Birch
.01s
2.49s
.06s
2.06s
.13s
.13s
.01s
.03s
.01s
2.48s
.04s
5.13s
.14s
.15s
.01s
.04s
.02s
2.41s
.04s
.48s
.14s
.16s
.02s
.03s
.01s
2.31s
.07s
.70s
.22s
.12s
.01s
.03s