

Алгоритми кластеризації даних великих об'ємів

Задача кластеризації

Нехай D — множина точок n -вимірного простору.

Означення

Кластеризацією $C = \{C \mid C \subseteq D\}$ називається таке розбиття D на підмножини, для якого виконується $\cup_{C_i \in C} C_i = D$ і $\forall C_i, C_j \in C : C_i \cap C_{j \neq i} = \emptyset$. Множини C_i називаються кластерами.

Застосування

- ▶ розпізнавання зображень, мови
- ▶ соціологія
- ▶ медицина
- ▶ маркетингові дослідження

Алгоритми

- ▶ K-means
- ▶ DBSCAN
- ▶ UPGMA
- ▶ Neighbor-joining