$X = (x_i)_{i=1}^{\ell}$ $a : X \to \{1, \ldots, K\}$

K-means

DBSCAN

Иерархическая кластеризация

$C^1 = \{\{x_1\}, \ldots \{x_\ell\}\}$

$\underline{d(X_m, X_n)}$ – мера близости между кластером m и n

1. Расстояние между самыми дальними
2. Между самыми ближними
3. Между центрами

$C^j = \{X_1, \ldots, X_{\ell-j+1}\}$.

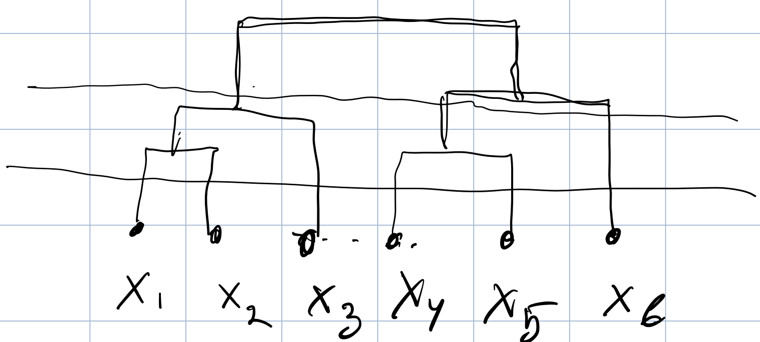
$(m, n) = \operatorname{argmin} d(X_n, X_m)$.

$1 \le m \le n \le \ell - j + 1$

$$C^{j+1} = \left(C^j \setminus \{X_m, X_n\} \right) \cup \{X_m \cup X_n\}.$$

$\vdots$

$$C^\ell = \{\{x_1, \dots x_\ell\}\}$$



– дендрограмма.
иерархия объектов.

Спектральная кластеризация

$G = (X, E)$ X – мн-во объектов.

проведем ребра между вершинами
если $\rho(x_i, x_j) < \varepsilon$.

$\mathcal{D} = \mathrm{diag.}(d_1, \dots d_\ell)$

d_i - степень вершины x_i $\quad d_i = \sum_{j=1} w_{ij}$ ← вес ребра

$W = (w_{ij})_{ij=1}^{\ell}$ - матрица смежности

$L = D - W$ - лапласиан графа

---

1) $f \in \mathbb{R}^{\ell}$ $\quad f^T L f = \frac{1}{2} \sum_{ij=1}^{\ell} w_{ij} (f_i - f_j)^2$

2) L - симметрическая неотрицательно определенная

Теорема

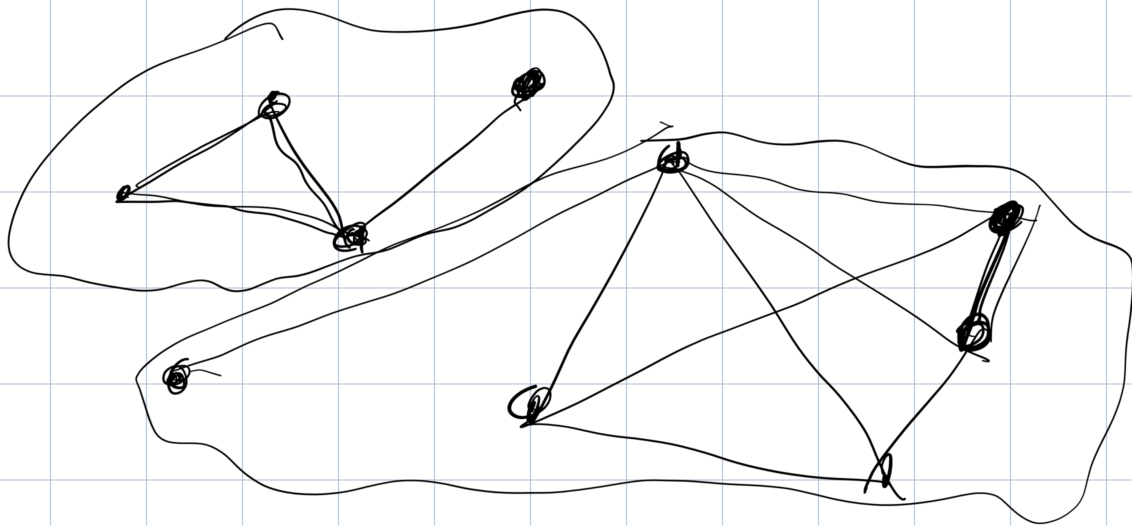
L - лапласиан G то

(1) $\lambda = 0$ в L то оно имеет кратность равную числу компонент в G

(2) Пусть $A_1, \dots A_k$

$f_1, \dots f_k$ $\quad f_i = ([x_j \in A_i])_{j=1}^{\ell}$

являются с. в. соответствующие $\lambda = 0$



$$\begin{pmatrix}1\\1\\1\\0\\0\\0\\0\end{pmatrix} \quad \begin{pmatrix}0\\0\\0\\1\\1\\0\\0\end{pmatrix} \quad \begin{pmatrix}0\\0\\0\\0\\0\\1\\1\end{pmatrix}$$

Гипотеза Если x_i и x_j похожие объекты $\rho(x_i, x_j) \approx 0$
Тогда у собств. вектора f_j соотв. маленькому $\lambda \approx 0$ $f_{ji} \approx f_{jk}$.

Алгоритм спектральной кластеризации

1) $X \to G \to \underset{l\times l}{L} = D - W$
2) $L \to m$ наименьших с.з.

и соотв им собств. векторы

$u_1, \dots, u_m$

3) $U = (u_1 | \dots | u_m) \in \mathbb{R}^{\ell \times m}$

новая матрица объекты × признаки

4) Запускаем K-means над U