Lecture 29

Kernel Clustering with Quantum Circuits. Part 2.

of the course "Fundamentals of Quantum Computing"
(by IAPS and QUANTERALL)

Stoyan Mishev, Vesselin Gueorguiev and Vladimir Gerdjikov

INSTITUTE *for* ADVANCED PHYSICAL STUDIES

NEW BULGARIAN UNIVERSITY

March 03, 2023

Plan

Part 1. Kernel K-means.

Part 2. Graphs and clustering

Part 1. Kernel K-means.

$i=5 \to z_i = 2$

$i=105, \quad z_i = 1$

K-means

Cost function

data points

cluster centroids

$$C(z, \mu) = \sum_i \| x_i - \mu_{z_i} \|^2$$

index of the cluster x_i belongs to

The task is to find z_i and μ_{z_i}: $C(z,\mu)$ is minimal

Iterative Algorithm:

$$z_i = \arg\min_k \| x_i - \mu_k \|^2 \quad (\text{for } \forall x_i \to z_i)$$

$$\mu_k = \frac{1}{N_k} \sum_{i \in C_k} x_i$$

# of points belonging to the cluster k

$\langle \varphi(x_j) | \varphi(x_i) \rangle$ | Feature map $\varphi_k(x_i)$

$= K_{ij}$

if $x_i \to X_{F \times N}$ then $\varphi(x_i) \to \Phi_{K \times N}$ | def

Cost : $C(z, \mu) = \sum \| \varphi(x_i) - \mu_{z_i} \|^2$

Define $Z = \underbrace{\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & & \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & & & & 0 \\ \vdots & & \vdots & & \vdots & & \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \end{pmatrix}}_{K} \Big\} N$ — assignment matrix

— $\forall$ row shows which point belongs to which cluster

$N \times K$ — # of cluster

# of point

$\sum_k z_{nk} = 1$

$\sum_n z_{nk} = N_k$

$\to L = \mathrm{diag}(1 / N_k) \equiv \mathrm{diag}\left(1 / \sum_n z_{nk}\right) \to K \times K$ matrix

---

Construct $\Phi_{K \times N} \, Z_{N \times K} \, L_{K \times K} \, Z^T_{K \times N} = M_{K \times N}$ → $\forall$ column contains a copy of the cluster mean μ_k assigned to corresp. data point

$K \Big\{ \underbrace{\begin{pmatrix} 5 & 5 & 2 & & \\ 5 & 5 & 3 & & \\ 1 & 1 & 1 & & \\ 1 & 1 & 5 & & \\ 2 & 2 & 4 & & \end{pmatrix}}_{N}$

The objective can be rewritten as: (cost function)

$\mu = \Phi Z L Z^T$

$$\Rightarrow C = tr\left[(\Phi_{K\times N} - \mu_{K\times N})(\Phi_{K\times N} - \mu_{K\times N})^T\right]$$

Proof that $(Z L Z^T)^2 = Z L Z^T$ (projector) $\Rightarrow (I - ZLZ^T)^2 = I - ZLZ^T$

$$\Rightarrow C = tr(\underbrace{\Phi\Phi^T}_{K\text{-kernel matrix}}) - tr(\Phi Z L Z^T \Phi^T) =$$

$$= tr(K) - tr(L^{1/2} Z^T K Z L^{1/2})$$

The task is to binary matrix Z: $\max tr(\underbrace{L^{1/2} Z^T}_{H^T} K \underbrace{Z L^{1/2}}_{H})$

(this is hard)

$\Rightarrow H = Z \cdot L^{1/2}$ is not binary anymore and

the task now is to find $\max_H tr[H^T K H]$: $\boxed{H^T \cdot H = I}$

If we interprete the columns of H as a collection of K mutually $\perp$ vectors then the second term of the objective can be rewritten as:

$$\sum_{k=1}^{K} h_k^T K h_k$$

h_k can be expressed derived from the eigenvectors of K:

$$K = U \Lambda U^T \Rightarrow H = U_{[1:k]} \cdot R \qquad (R^T R = I)$$

where R is a rotation inside the eigenvalue space.

Then $tr[H^T K H] = \sum_{k=1}^{K} \lambda_k$, where λ_k are the largest K eigenvalues.

(– Kernel PCA is similarly solved by the K largest eigenvalues)

So after we obtain H and then $Z = H \cdot L^{-1/2}$

We treshold the values of Z $\nearrow 1$, $\searrow 0$ and obtain

approx. clustering solution.

It works better if we first normalize H:

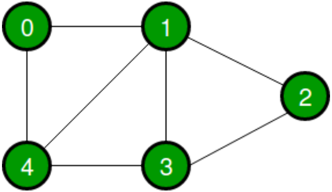
$$H_{nk}' = \frac{H_{nk}}{\sqrt{\sum_k H_{nk}^2}}$$

Part 2. Graphs and clustering

Finite (undirected / with weights) graph

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{A}\}$$

with nodes $\mathcal{V}$, edges $\mathcal{E}$ and adjacency matrix $\mathbf{A}$



	0	1	2	3	4
0	0	1	0	0	1
1	1	0	1	1	1
2	0	1	0	1	0
3	0	1	1	0	1
4	1	1	0	1	0

Unnormalized:

$$\Delta = D - A$$

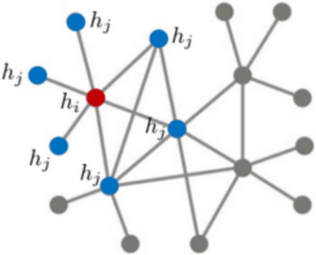
Normalized:

$$\Delta = I - D^{-1/2} A D^{-1/2}$$

$$D = \mathrm{diag}(\sum_{j \neq i} A_{ij})$$

Мярка за гладкост:

$$(\Delta h)_i = h_i - \sum_{j \in \mathcal{N}_i} \frac{1}{\sqrt{d_i d_j}} A_{ij} h_j$$



Proposition 1 (Properties of L) *The matrix L satisfies the following properties:*

1. *For every vector $f \in \mathbb{R}^n$ we have*

$$f'Lf = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2.$$

2. *L is symmetric and positive semi-definite.*
3. *The smallest eigenvalue of L is 0, the corresponding eigenvector is the constant one vector $\mathbb{1}$.*
4. *L has n non-negative, real-valued eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$.*

Algorithm

Unnormalized spectral clustering

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.

- Construct a similarity graph by one of the ways described in Section 2. Let W be its weighted adjacency matrix.
- Compute the unnormalized Laplacian L.
- **Compute the first k eigenvectors $u_1, \dots, u_k$ of L.**
- Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \dots, u_k$ as columns.
- For $i = 1, \dots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of U.
- Cluster the points $(y_i)_{i=1,\dots,n}$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, \dots, C_k$.

Output: Clusters $A_1, \dots, A_k$ with $A_i = \{j \mid y_j \in C_i\}$.

https://colab.research.google.com/drive/
1fKcrCyAbhZEMxbkNYchpkf49VTT86bfZ?usp=sharing

https://qiskit.org/documentation/machine-learning/tutorials/03_quantum_kernel.html#Clustering

THANK YOU FOR YOUR ATTENTION!

БЛАГОДАРЯ ЗА ВНИМАНИЕТО!