Proofs, Exercises and Literature - Nonconvex Clustering

1 Proofs

Theorem 1 (Equivalent k-means objectives)**.** *Given a data matrix $D \in \mathbb{R}^{n\times d}$, the following objectives are equivalent k-means objectives:*

$$\min_Y \|D - YX^\top\|^2 \; s.t. \; X = D^\top Y(Y^\top Y)^{-1}, Y \in \mathbb{1}^{n\times r} \quad (1)$$

$$\max_Y \operatorname{tr}(Z^\top DD^\top Z) \qquad s.t. \; Z = Y(Y^\top Y)^{-1/2}, Y \in \mathbb{1}^{n\times r} \quad (2)$$

Proof. For $X = D^\top Y(Y^\top Y)^{-1}$, we have

$$\begin{aligned}
&\|D - YX^\top\|^2 \\
&= \|D\|^2 - 2\operatorname{tr}(D^\top YX^\top) + \|YX^\top\|^2 && \text{(binomial formula)} \\
&= \|D\|^2 - 2\operatorname{tr}(D^\top YX^\top) + \operatorname{tr}(XY^\top YX^\top) && \text{(def. Fro-norm by tr)} \\
&= \|D\|^2 - 2\operatorname{tr}(XY^\top D) + \operatorname{tr}(XY^\top Y(Y^\top Y)^{-1}Y^\top D) && \text{(def. } X\text{)} \\
&= \|D\|^2 - \operatorname{tr}(D^\top Y(Y^\top Y)^{-1}Y^\top D) && (Y^\top Y(Y^\top Y)^{-1} = I) \\
&= \|D\|^2 - \operatorname{tr}((Y^\top Y)^{-1/2}Y^\top DD^\top Y(Y^\top Y)^{-1/2}) && \text{(cycling property tr)}
\end{aligned}$$

As a result, the objective function of k-means is equal to

$$\begin{aligned}
\|D - YX^\top\|^2 &= \|D\|^2 - \operatorname{tr}((Y^\top Y)^{-1/2}Y^\top DD^\top Y(Y^\top Y)^{-1/2}) \\
&= \|D\|^2 - \operatorname{tr}((Z^\top DD^\top Z),
\end{aligned}$$

for $Z = Y(Y^\top Y)^{-1/2}$. Minimizing the term on the left is equivalent to minimizing the negative trace term on the right, which is equivalent to maximizing the trace term on the right. □

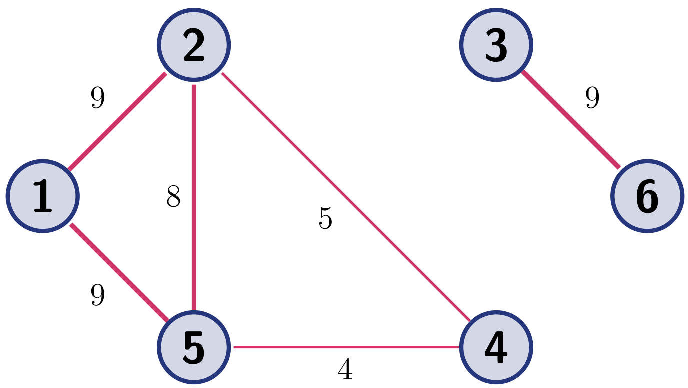
Proposition 2 (Positive Definiteness of Laplacians)**.** *Given a symmetric similarity matrix $W \in \mathbb{R}_+^{n\times n}$, the Laplacian $L = I_W - W$ is positive semi-definite*

Proof: Let $0 \neq v \in \mathbb{R}^n$, then

$$\begin{aligned} v^\top L v = v^\top I_W v - v^\top W v &= \sum_{i=1}^{n} v_i^2 |W_{i\cdot}| - \sum_{1\le i,j\le n} v_i v_j W_{ij} \\ &= \frac{1}{2} \sum_{1\le i,j\le n} (v_i^2 W_{ij} - 2 v_i v_j W_{ij} + v_j^2 W_{ij}) \\ &= \frac{1}{2} \sum_{1\le i,j\le n} W_{ij}(v_i - v_j)^2 \geq 0 \end{aligned}$$

2 Exercises

1. Consider the following graph:



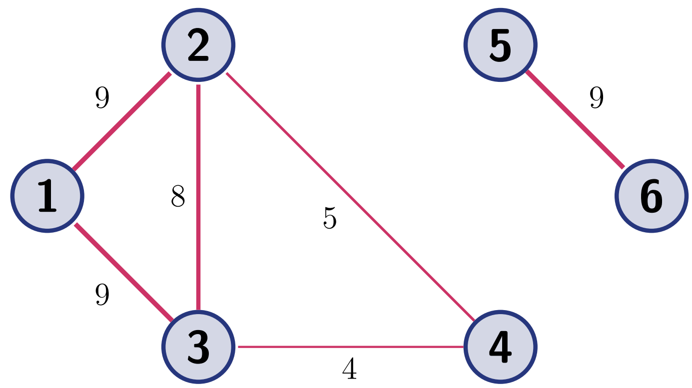
   (a) Denote the weighted adjacency matrix of this graph, where column i corresponds to node i.

   Solution: The weighted adjacency matrix is

$$W = \begin{pmatrix} 0 & 9 & 0 & 0 & 9 & 0 \\ 9 & 0 & 0 & 5 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 \\ 0 & 5 & 0 & 0 & 4 & 0 \\ 9 & 8 & 0 & 4 & 0 & 0 \\ 0 & 0 & 9 & 0 & 0 & 0 \end{pmatrix}$$

   (b) The graph has two connected components. Find a reordering of the nodes, such that the corresponding weighted adjacency matrix has a block-diagonal structure.

   Solution: We numerate the nodes such that the nodes within the connected components have subsequent numbers.



The corresponding weighted adjacency matrix has now a block-diagonal structure:

$$W_c = \begin{pmatrix} 0 & 9 & 9 & 0 & 0 & 0 \\ 9 & 0 & 8 & 5 & 0 & 0 \\ 9 & 8 & 0 & 4 & 0 & 0 \\ 0 & 5 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 & 9 & 0 \end{pmatrix}$$

(c) Compute the within cluster similarity and the cut, when the clusters are equal to the connected components:

$$Sim(Y; W) = \mathrm{tr}(Y^\top W Y (Y^\top Y)^{-1})$$
$$Cut(Y; W) = \mathrm{tr}(Y^\top W (\mathbf{1} - Y)(Y^\top Y)^{-1})$$

Solution: We use the notation of the previous question, where the adjacency matrix has a block-diagonal structure. The cluster indicator matrix is then given by

$$Y = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

Then we have

$$
\begin{aligned}
Sim(Y;W) &= \frac{Y_{\cdot 1}^\top W_c Y_{\cdot 1}}{|Y_{\cdot 1}|} + \frac{Y_{\cdot 2}^\top W_c Y_{\cdot 2}}{|Y_{\cdot 2}|} \\
&= \frac{2(9+9+8+5+4)}{4} + \frac{2\cdot 9}{2} \\
&= 26.5 \\
Cut(Y;W) &= \frac{Y_{\cdot 1}^\top W_c(\mathbf{1} - Y_{\cdot 1})}{|Y_{\cdot 1}|} + \frac{Y_{\cdot 2}^\top W_c(\mathbf{1} - Y_{\cdot 2})}{|Y_{\cdot 2}|} \\
&= 0
\end{aligned}
$$

3 Recommended Literature

As always, the best exercise is to go through the lecture and see if you can follow the steps with pen and paper. If you feel like reading though, you can have a look at the following material:

Friedman, Hastie, and Tibshirani. The elements of statistical learning. 2001.

The book of Friedman, Hastie, and Tibshirani has a section on Spectral clustering, with a plot of the eigenvalues of the Laplacian, which might be helpful to get some intuition about the Laplacian.

14.5.3 Spectral Clustering