Exercises for the course
Machine Learning for Data Science
Winter Semester 2024/25

G. Montavon
Institute of Computer Science
Department of Mathematics and Computer Science
Freie Universität Berlin

Exercise Sheet 5 (theory part)

Exercise 1: K-Means Clustering (15 + 10 P)

The K-means optimization problem is given by $\arg\min_{\boldsymbol{\mu},\boldsymbol{c}} \sum_{i=1}^N \|\boldsymbol{x}_i - \boldsymbol{\mu}_{c_i}\|^2$ where $\boldsymbol{c} \in \{1,\dots,K\}^N$ is the cluster assignment function. When considering the latter to be fixed, and only letting the centroids $\boldsymbol{\mu}$ vary, the optimization problem can be restated as:

$$\arg\min_{\boldsymbol{\mu}} \sum_{k=1}^K \sum_{i \in \mathcal{C}_k} \|\boldsymbol{x}_i - \boldsymbol{\mu}_k\|^2$$

where $\mathcal{C}_k$ is the set of instances that are assigned to cluster k.

(a) Show that the solution of this optimization problem is given by:

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_k)_{k=1}^K \quad \text{where} \quad \boldsymbol{\mu}_k = \frac{\sum_{i \in \mathcal{C}_k} \boldsymbol{x}_i}{\sum_{i \in \mathcal{C}_k} 1}$$

After having shown convexity, we get the minimum of the objective where $\nabla J(\boldsymbol{\mu}) = 0$. That is, for all k:

$$\frac{\partial}{\partial \boldsymbol{\mu}_k} \sum_k \sum_{i \in \mathcal{C}_k} \|\boldsymbol{x}_i - \boldsymbol{\mu}_k\|^2 = \sum_{i \in \mathcal{C}_k} 2(\boldsymbol{\mu}_k - \boldsymbol{x}_i) \stackrel{\text{def}}{=} \mathbf{0}$$

Hence, we get for all k:

$$\sum_{i \in \mathcal{C}_k} \boldsymbol{\mu}_k = \sum_{i \in \mathcal{C}_k} \boldsymbol{x}_i$$

or

$$\boldsymbol{\mu}_k \sum_{i \in \mathcal{C}_k} 1 = \sum_{i \in \mathcal{C}_k} \boldsymbol{x}_i$$

and therefore

$$\boldsymbol{\mu}_k = \frac{\sum_{i \in \mathcal{C}_k} \boldsymbol{x}_i}{\sum_{i \in \mathcal{C}_k} 1}$$

(b) A data point $\boldsymbol{x}$ is assigned onto cluster c if

$$\forall_{k\,:\,k \neq c}: \ \|\boldsymbol{x} - \boldsymbol{\mu}_c\| < \|\boldsymbol{x} - \boldsymbol{\mu}_k\|.$$

Show that this condition for assignment onto cluster c can be equivalently formulated as a min-pooling over affine functions, specifically, we assign to cluster c if

$$\min_{k\,:\,k \neq c} \left\{ \boldsymbol{w}_k^\top \boldsymbol{x} + b_k \right\} > 0$$

where $\boldsymbol{w}_k = (\boldsymbol{\mu}_c - \boldsymbol{\mu}_k)$ and $b_k = \frac{1}{2}(\|\boldsymbol{\mu}_k\|^2 - \|\boldsymbol{\mu}_c\|^2)$.

$$
\begin{aligned}
&\forall_{k\,:\,k\neq c}:\ \ \|\boldsymbol{x}-\boldsymbol{\mu}_c\| < \|\boldsymbol{x}-\boldsymbol{\mu}_k\| \\
&\Leftrightarrow \forall_{k\,:\,k\neq c}:\ \ \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{\mu}_c\|^2 < \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{\mu}_k\|^2 \\
&\Leftrightarrow \forall_{k\,:\,k\neq c}:\ \ \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{\mu}_k\|^2 - \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{\mu}_c\|^2 > 0 \\
&\Leftrightarrow \min_{k:\,k\neq c}\left\{\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{\mu}_k\|^2 - \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{\mu}_c\|^2\right\} > 0 \\
&\Leftrightarrow \min_{k:\,k\neq c}\left\{\frac{1}{2}\|\boldsymbol{x}\|^2 - \boldsymbol{x}^\top\boldsymbol{\mu}_k + \frac{1}{2}\|\boldsymbol{\mu}_k\|^2 - \frac{1}{2}\|\boldsymbol{x}\|^2 + \boldsymbol{x}^\top\boldsymbol{\mu}_c - \frac{1}{2}\|\boldsymbol{\mu}_c\|^2\right\} > 0 \\
&\Leftrightarrow \min_{k:\,k\neq c}\Big\{\underbrace{(\boldsymbol{\mu}_c-\boldsymbol{\mu}_k)^\top}_{\boldsymbol{w}}\boldsymbol{x} + \underbrace{\frac{1}{2}\Big(\|\boldsymbol{\mu}_k\|^2 - \|\boldsymbol{\mu}_c\|^2\Big)}_{b}\Big\} > 0
\end{aligned}
$$

Exercise 2: Spectral Clustering (15 + 10 P)

In the lecture, it was mentioned that the eigenvalues λ of the Laplacian matrix $L = D - A$ (where D and A are the degree and adjacency matrices respectively) can be related to the corresponding eigenvector u as:

$$
\lambda = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij}\big(u_i - u_j\big)^2.
$$

(a) *Prove* the equation above.

$$
\begin{aligned}
\lambda &= u^\top L u \\
&= \sum_{ij} u_i L_{ij} u_j \\
&= \sum_{ij} u_i D_{ij} u_j - \sum_{ij} u_i A_{ij} u_j \\
&= \sum_{i} u_i^2 D_{ii} - \sum_{ij} u_i A_{ij} u_j \\
&= \sum_{i} u_i^2 \sum_j A_{ij} - \sum_{ij} u_i A_{ij} u_j \\
&= \sum_{ij} \big(u_i^2 A_{ij} - u_i A_{ij} u_j\big) \\
&= \sum_{ij} \big(\frac{1}{2}u_i^2 A_{ij} + \frac{1}{2}u_j^2 A_{ij} - u_i A_{ij} u_j\big) \\
&= \sum_{ij} \frac{1}{2}(u_i - u_j)^2 A_{ij}
\end{aligned}
$$

(b) From the equation above, we can see that the eigenvalue λ influences the extent by which the associated eigenvector u can vary between connected nodes.

Show that eigenvectors associated to the eigenvalue $\lambda = 0$, cannot vary within a connected component, that is, denoting by $\boldsymbol{u}$ the eigenvector, show that if i and j are part of the same connected component (i.e. if there is a sequence of edges connecting i and j), then $u_i = u_j$.

We start the equation above

$$\lambda = \sum_{ij} \frac{1}{2}(u_i - u_j)^2 A_{ij}$$

If the eigenvalue satisfies $\lambda = 0$, this implies that all summands (necessarily positive) must also be zero, i.e.

$$\forall_{ij} : \ \frac{1}{2}(u_i - u_j)^2 A_{ij} = 0.$$

If a pair of points (i, j) is connected, then we have an adjacency term $A_{ij} > 0$ and then it is necessary that $(u_i - u_j)^2 = 0$, in other words

$$u_i = u_j \tag{1}$$

We now consider the case where two nodes m and n are not necessarily connected but part of the same connected component. If that is the case, there must be a sequence of connected nodes,

$$k_1, \dots, k_T$$

where $m = k_1$ and $n = k_T$, and where k_t is connected to k_{t+1} for all t. In that case, relying on the property of connected nodes stated above, specifically Eq. (1), we can construct the chain of equalities

$$u_{k_1} = u_{k_2} = \cdots = u_{k_{T-1}} = u_{k_T},$$

hence $u_m = u_n$.