
HOMEWORK N° 3 : K-means

Please send your homework as a **pdf** file to joseph.salmon@telecom-paristech.fr with email header [CR12-homework] before Wednesday December the 17th.

EXERCICE 1. (Convergence)

1) For a convex set $\mathcal{D} \subset \mathbb{R}^p$, consider $\phi : \mathcal{D} \mapsto \mathbb{R}$ a strictly convex and differentiable function over $\operatorname{Int}(\mathcal{D})$ (the interior $\mathcal{D}$. Define for all $x \in \mathcal{D}$ and all $y \in \operatorname{Int}(\mathcal{D})$:

$$d_\phi(x, y) = \phi(x) - \phi(y) - \langle x - y, \nabla\phi(y)\rangle.$$

Show that $d_\phi(x, y) \geqslant 0$ and that equality holds only when $x = y$.

2) What is d_ϕ for the following ϕ?
— $\phi(x) = \|x\|_2^2/2$ with $\mathcal{D} = \mathbb{R}^p$
— $\phi(x) = \sum_{i=1}^p x_j \log(x_j)$ with $\mathcal{D} = \{x \in \mathbb{R}^p : \forall j \in [p], 0 \leqslant x_j \leqslant 1, \sum_{i=1}^p x_j = 1\}$

3) Suppose we have n observations $x_1, \cdots, x_n$ in $\mathcal{D}$, with weights $\alpha_1, \cdots, \alpha_n$ (*i.e.*, $\forall i \in [n], 1 \leqslant \alpha_i \leqslant 1$ and $\sum_{i=1}^n \alpha_i = 1$). Define the inertia function in the following way : for any $\mu \in \operatorname{Int}(\mathcal{D})$

$$I_n(\mu) = \sum_{i=1}^n \alpha_i d_\phi(x_i, \mu).$$

Show that $\hat{\mu} := \sum_{i=1}^n \alpha_i x_i = \arg\min_{\mu \in \mathcal{D}} I_n(\mu)$ (assuming that $\hat{\mu} \in \operatorname{Int}(\mathcal{D})$).

4) Consider the following variation of the K-means algorithm (with K classes) :
Initialize the centroids as $\mu_1, \cdots, \mu_K$.
Repeat iteratively the following steps :
   - Assignment step : affect x_i to the class k where $k = \arg\min_{k' \in [K]} d_\phi(x_i, \mu_{k'})$. Obtain a partition $\mathcal{X}_1, \cdots, \mathcal{X}_K$ of the set $\{x_1, \cdots, x_n\}$.
   - Estimation step : compute the new centroids

$$\forall k \in [K], \quad \mu_k = \frac{\sum_{x_i \in \mathcal{X}_k} \alpha_i x_i}{\sum_{x_{i'} \in \mathcal{X}_k} \alpha_{i'}}$$

Show that the algorithm convergences (Hint : as for the K-means, show that the inertia is non-increased as the algorithm evolves).

EXERCICE 2. (Practical aspects)

The following exercise can be done in your favorite programming language (though, for exotic ones please add enough comments).

1) Generate 300 (independent) points in $\mathbb{R}^2$ from a mixture of Gaussian with centers $(0, 1), (1, 0), (0, -1), (-1, 0)$, covariance $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$ for $\sigma = 0.8$, and uniform weights (*i.e.*, $1/4$).

2) Code the K-means algorithm introduced in the course. Compare the visual results over the datasets generated before with the following parameters : $K = 3, 4, 5$. Precise how you initialized the algorithm.

3) Compare the inertia value at optimum for $K = 4$, and 5 different centroids initializations.