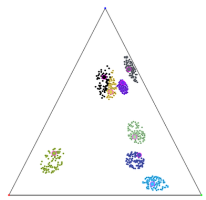
On Clustering Histograms with k-Means by Using Mixed α-Divergences
Entropy 16(6): 3273-3301 (2014). BibTeX:`J2014-ClusteringMixedDivergence` [1]

The *mixed divergence* $M_\lambda(p:q:r) = \lambda D(p:q) + (1-\lambda) D(q:r)$ for $\lambda \in [0,1]$ includes the *sided* ($\lambda \in \{0,1\}$) and the *symmetrized* divergences ($\lambda = \frac{1}{2}$). In particular, the *mixed α-divergences* are defined by $M_{\lambda,\alpha}(p:x:q) = \lambda D_\alpha(p:x) + (1-\lambda) D_\alpha(x:q) = M_{1-\lambda,-\alpha}(q:x:p)$. The *$\alpha$-Jeffreys symmetrized divergence* ($\lambda = \frac{1}{2}$) is $S_\alpha(p,q) = M_{\frac{1}{2},\alpha}(q:p:q)$ and the *skew symmetrized α-divergence* is defined by $S_{\lambda,\alpha}(p:q) = \lambda D_\alpha(p:q) + (1-\lambda) D_\alpha(q:p) = M_{\lambda,\alpha}(q:p:q)$. We describe hard k-means type and soft EM type clustering methods for mixed and symmetrized divergences. For mixed divergences, we define *coupled k-means* where each cluster has two dual centroids, and show how to extend the k-means++ seeding to the case of mixed divergences.



In particular, we report a guaranteed probabilistic bound of mixed k-means++ α-seeding, and show that the dual centroids in clusters are $\pm\alpha$-means. When symmetrized centroids are not available in closed form, we use *variational k-means* clustering with one centroid per cluster. We show that the symmetrized Jeffreys J_α-centroid of a set of n weighted histograms $\mathcal{H}$ amount to computing the symmetrized α-centroid for the weighted α-mean and $-\alpha$-mean: $\min J_\alpha(x,\mathcal{H}) = \min_x \left(D_\alpha(x:r_\alpha) + D_\alpha(l_\alpha:x)\right)$, where $r_\alpha^i = \begin{cases} \left(\sum_{j=1}^n w_j (h_j^i)^{\frac{1-\alpha}{2}}\right)^{\frac{2}{1-\alpha}} & \alpha \neq 1 \\ r_1^i = \prod_{j=1}^n (h_j^i)^{w_j} & \alpha = 1 \end{cases}$, $\tilde{r}_\alpha^i = \frac{r_\alpha^i}{w(\tilde{r}_\alpha)}$ and $l_\alpha^i = r_{-\alpha}^i$ ($\tilde{l}_\alpha^i = \tilde{r}_{-\alpha}^i$). We consider mixed/symmetrized α-divergences and their centroids defined either on *positive arrays* or on *frequency histograms*. Finally, we report a *soft mixed α-clustering* where each histogram belongs to all clusters according to some weight distribution. This latter algorithm also learns the α and λ parameters (provided that $\lambda_{\text{init}} \not\in \{0,1\}$).

References

[1] Frank Nielsen, Richard Nock, and Shun-ichi Amari. On clustering histograms with k-means by using mixed α-divergences. *Entropy*, 16(6):3273–3301, 2014.