Spontaneous Clustering via Minimum γ-divergence

Akifumi Notsu　**Third-year Ph.D. Student, The Graduate University for Advanced Studies**

【Abstract】

We propose a new method for clustering based on the local minimization of the γ-divergence, which we call the spontaneous clustering. The greatest advantage of the proposed method is that it automatically detects the number of clusters that adequately reflect the data structure. In contrast, exiting methods such as K-means, fuzzy c-means, and model based clustering need to prescribe the number of clusters. We detect all the local minimum points of the γ-divergence, which are defined as the centers of clusters. A simulation study is performed to compare our proposal with existing methods.

【Motivational Example】

Consider the problem of estimating Gaussian mean parameter μ. The maximum likelihood estimator (MLE) of μ is given by the arithmetic mean of the data set as the unique maximum point of the log likelihood function. It is known that the MLE poorly behaves in various situations where Gaussianity assumption is inappropriate. For example, the log likelihood function suggests rather a misleading summary as seen in panel (a) of Figure 1. Alternatively, the γ-loss function properly reflects the data shape. For the same data set in panel (a) of Figure 1, panel (b) shows that the γ-loss function has two local minimum points corresponding to the two normal distributions. We will propose to determine the centers of clusters by such local minimum points.

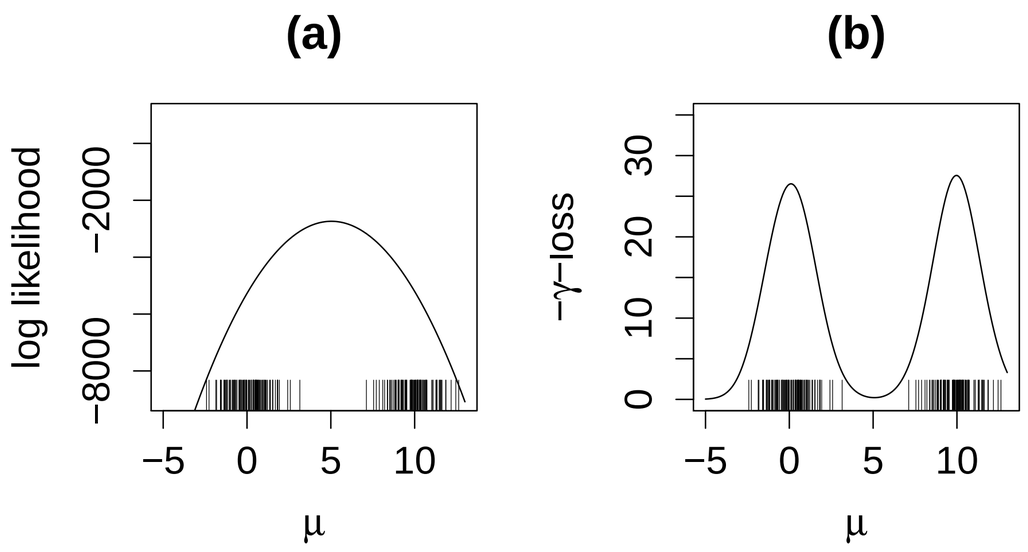

Figure 1. (a) Log likelihood function. (b) Minus γ-loss function ($\gamma = 1$). In panels (a) and (b) the data of size 200 is generated from the mixture of two standard normal distributions centered at 0 and 10, respectively.

【γ-loss Function for Normal Distribution】

We consider the γ-loss function for the normal distribution with mean vector μ and covariance matrix Σ,

$$L_\gamma(\mu, \Sigma) = -\det \Sigma^{-\frac{\gamma}{2(1+\gamma)}} \sum_{i=1}^{n} \exp\left(-\frac{\gamma}{2}(x_i - \mu)^\top \Sigma^{-1}(x_i - \mu)\right).$$

An iteration algorithm to find the local minimum points of $L_\gamma(\mu, \Sigma)$ is proposed in Fujisawa and Eguchi (2008) and Eguchi and Kato (2010). We name it ALGORITHM.

【Spontaneous Clustering Based on the Normal Distribution】

The spontaneous clustering based on the normal distribution is defined as follows. We set Θ_μ and $\Theta_{(\mu,\Sigma)}$ are the empty sets at the start of the algorithm. ALGORITHM is employed in the spontaneous clustering below.

Step 1-1 If Θ_μ is the empty set, choose M initial values $x_{(1)}, \ldots, x_{(M)}$ in the data set $\{x_1, \ldots, x_n\}$ at random. Otherwise, choose initial values in $\{x_1, \ldots, x_n\}$ as follows: $x_{(1)}, \ldots, x_{(M)}$ are M maximum points of $d(\cdot, \Theta_\mu)$, where

$$d(x, \Theta_\mu) = \min_{\hat{\mu} \in \Theta_\mu} \|x - \hat{\mu}\|.$$

Step 1-2 Apply ALGORITHM to the data set M times with each initial value $x_{(i)}, i = 1, \ldots, M$ to find the local minimum points of $L_\gamma(\mu)$. Then add the obtained local minimum points to Θ_μ.

Step 1-3 Repeat Step 1-1 and 1-2 until the number of elements in Θ_μ does not increase.

Step 1-4 For each local minimum point $\hat{\mu} \in \Theta_\mu$, obtain a minimum point of $L_\gamma(\hat{\mu}, \Sigma)$ with respect to Σ, denoted by $\hat{\Sigma}$, with ALGORITHM. Then add $(\hat{\mu}, \hat{\Sigma})$ to $\Theta_{(\mu,\Sigma)}$.

Step 2 Write $\Theta_{(\mu,\Sigma)}$ by $\{(\hat{\mu}_k, \hat{\Sigma}_k)\}_{k=1}^K$ and assign each observation x_i to the $\hat{k}$-th cluster with

$$\hat{k} = \underset{k=1,\ldots,K}{\operatorname{argmin}} (x_i - \hat{\mu}_k)^\top \hat{\Sigma}_k^{-1} (x_i - \hat{\mu}_k).$$

【Selection Procedure for γ】

We propose two methods to select the value of γ. One is a heuristic choice of γ that depends on the range of the data. Our proposal is $\hat{\gamma} = 72/R^2$, where R is defined by the maximum range:

$$R = \max_{j=1,\ldots,p} \left\{ \left(\max_{i=1,\ldots,n} x_{ij} \right) - \left(\min_{i=1,\ldots,n} x_{ij} \right) \right\},$$

where $x_i = (x_{i1}, \ldots, x_{ip})^\top$. We also propose a more sophisticated method based on AIC. The value of γ which minimizes AIC is recommended as the optimal selection of γ.

【Simulation】

We demonstrate the performance of the spontaneous clustering in comparison with the K-means algorithm. The value of γ for the spontaneous clustering is determined by the two methods described above. The number of clusters for the K-means algorithm is determined by two criteria, CH and the gap statistic. The performance of clustering is measured by BHI.

The sample of size 200 is generated from the mixture of five standard normal distributions centered at $(0,0)^\top$, $(3,3)^\top$, $(-3,3)^\top$, $(-3,-3)^\top$, $(3,-3)^\top$ with equal mixing proportion. We simulated 100 runs, and compared clustering results from the spontaneous clustering with those from the K-means algorithm.

Table 1. Frequencies of Choosing K Clusters.

K	1	2	3	4	5
SC with the range	0	0	0	9	91
SC with AIC	0	0	0	1	99
K-means with CH	0	0	0	0	100
K-means with Gap	91	7	0	0	2

Table 2. Mean Value of BHI and DM1-DM5.

	BHI	DM1	DM2	DM3	DM4	DM5
SC with the range	0.93	0.38	0.38	0.37	0.33	0.34
SC with AIC	0.94	0.34	0.32	0.28	0.27	0.26
K-means with CH	0.95	0.25	0.23	0.21	0.21	0.21
K-means with Gap	0.22	0.16	0.49	0.23	0.41	0.21

References

Fujisawa, H. & Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis*, 99(9):2053–2081.

Eguchi, S. & Kato, S. (2010). Entropy and divergence associated with power function and the statistical application. *Entropy*, 12:262–274.