

Piecewise elliptical classification based on bregman matrix divergence

陳宥任

國立臺灣大學數學研究所

Abstract

If a crowd of data (interior data) is encompassed by another set of data (exterior data) and we are to find a closed surface centralized at the mean of interior data to wrap and separate it from exterior data, the first candidate for the closed surface may be an ellipse. (Kulis, 2006) applies (Bregman, 1967) trying to find a suitable ellipse to separate this kind of data. To be more specific, the crowd of data (interior data) can be properly described by its mean and covariance matrix, while the other set of data (exterior data) can be thought of as an unknown number of crowds of data (e.g., a mixed multivariate skew normal). However, from the conventional PCA, ellipses (i.e. semi positive definite matrices) exhibit rigid structures, that is, symmetry and orthogonality. These two properties are too restrictive for the following situations, assuming that interior data is like a unit ball (A) and exterior data is like two unit balls (B, C). First, B is 4 units far from A on the left while C is 40 units far from A on the right. The best cuts should be 2 units far from A on the left and 20 units far from A on the right. Unfortunately, an ellipse centralized at the center of A can only have cuts equally far from the left and right of A. Second, B is 4 units far from A on the left while B is 8 units far from A on the top right. In this situation, AB and AC are not orthogonal, which makes ellipses unsuitable separators. Therefore, we offer a method- Nonorthogonal Ray Decomposition (NRD)- to improve this situation.