Local Spectral Clustering of Density Upper Level Sets

Anonymous Authors[1]

1. Empirical Performance of PPR on Gaussian Mixture Models

The assumptions and theory of Section 2 are tailored towards density functions with sharp transitions in gradient around the perimeter of the density cluster. These type of functions will satisfy A3 with a small ratio of Λ_σ to λ_σ, while still satisfying A4, potentially for $\gamma << 1$.

As mentioned previously, one line of work on spectral algorithms assesses their performance on mixture models. Particularly well-developed is the characterization of clustering performance on Gaussian Mixture Models (GMMs). Gaussian Mixture Models have smooth derivatives, and are therefore not good candidates for our work (at least, for reasonable values of the parameters in (A1)-(A4)). Since they are a classical choice in the clustering literature, we investigate PPR performance on them empirically.

Figure 1 tells a story with roughly the same narrative as our theoretical work. The left two plots demonstrate instances, in one and two-dimensions, respectively, where the PPR cut has small error relative to a density cut. The right two plots, by contrast, show PPR failing to recover, even approximately, a density cut.

In the bottom right panel in particular, we see the effect of a high diameter on the performance of PPR . Although any one vertex on the boundary of a density level set may have few cut edges, the length of the boundary is sufficiently long to make the entire cut too large for PPR to select it. Instead, the PPR cut defaults to a more geometrically compact shape, which has much smaller (unweighted) perimeter.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

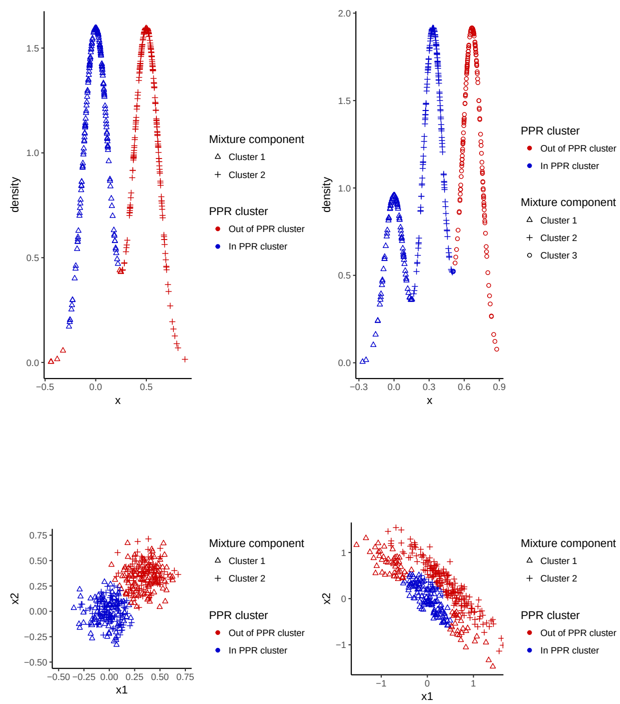

Figure 1. Algorithm **??** run over 4 example datasets. Parameters α and π_0 were tuned to minimize normalized cut of output set. Colors represent PPR cuts; shapes correspond to the maximum density among mixture components. Sample size $n = 400$.