DP-means におけるファジィ所属度の推定

○西田 豊

(大阪大学 大学院基礎工学研究科)

キーワード: DP-means, K -means, ファジィクラスタリング

Estimating fuzzy membership in DP-means

Yutaka NISHIDA

(Graduate School of Engineering Science, Osaka Univ.)

Key words: DP-means, K -means, fuzzy clustering

目 的

K -means はある仮定を置いたとき, EM アルゴリズムのある極限として導くことができる (Bishop, 2006). すなわち混合分布の要素である正規分布の共分散を固定し, 分散共分散を $\sigma^2 \mathbf{I}$ であると仮定したとき, $\sigma \rightarrow 0$ の極限では, EM アルゴリズムが K -means と同様にハードな割り当てとなる.

Kulis and Jordan (2012) は生成モデルとして記述される正規混合モデルをベイズに拡張し, EM アルゴリズムの場合と同様に $\sigma \rightarrow 0$ の極限を考えることにより, デリクレ過程 (DP) 混合モデルに基づくハードなクラスタリングアルゴリズムを導いた. これを DP-means と呼び, その目的関数は以下のように定義される.

$$\sum_{i=1}^n \sum_{j=1}^k u_{ij} \| \mathbf{x}_i - \boldsymbol{\mu}_j \|^2 + \lambda k \quad (1)$$

$\mathbf{x}_i$ は個体 i のデータベクトル, u_{ij} は個体 i がクラスター j に所属するか否かを示すパラメータ, $\boldsymbol{\mu}_j$ はクラスター j の平均を示すパラメータ, λ はクラスター数を調整するパラメータ, k はクラスター数である. この関数の最適化アルゴリズムを以下に示す.

- (1) クラスター数パラメータ λ をセットする.
- (2) クラスター数 k を 1, クラスター平均 $\boldsymbol{\mu}_1$ を全平均とする.
- (3) 収束するまで以下を繰り返す.
 - すべてのデータポイント $\mathbf{x}_i$ に対し,
 - $j = 1, \dots, k$ について $d_{ij} = \| \mathbf{x}_i - \boldsymbol{\mu}_j \|^2$ を計算する.
 - もし $\min_j d_{ij} > \lambda$ であれば, $k = k + 1$, データ i の所属をクラスター k , クラスター k の平均 $\boldsymbol{\mu}_k$ を $\mathbf{x}_i$ とする.
 - そうでなければ, データ i の所属を $\arg \min_j d_{ij}$ とする.
 - 各クラスターの平均 $\boldsymbol{\mu}_j$ を求める.

K -means と異なり, DP-means では事前にクラスター数を指定する必要はなく (代わりにクラスター数パラメータを設定する), 全平均をクラスター平均とするシングルクラスターを初期値としてスタートする. データポイントと既存のクラスター中心との距離が λ よりも遠い場合は新しいクラスターを生成する.

DP-means は DP 混合モデルのクラスター内分散を 0 に近づけることによってハードクラスタリングを達成しており K -means と似たアルゴリズムを持つ. ノンパラメトリックベイズの特性を持ち, 計算も早い DP-means のアルゴリズムを使いたいが, ソフトな割り当てによるクラスタリングを行いたいという需要も考えられる. そこで本研究ではエントロピー正則化の枠組み (Miyamoto, Ichihashi and Honda, 2008) を用い

て, ファジィ所属度を推定できるソフトな DP-means を提案する.

方 法

DP-means の目的関数にエントロピー正則化項を加え, 目的関数を修正する. ファジィ DP-means の目的関数を以下のように定義する.

$$\sum_{j=1}^k \sum_{i=1}^n u_{ij} \| \mathbf{x}_i - \boldsymbol{\mu}_j \|^2 + \lambda k + \eta^{-1} \sum_{j=1}^k \sum_{i=1}^n u_{ij} \log u_{ij} \quad (2)$$

η はファジィの程度を調整するパラメータである. アルゴリズムは DP-means とほぼ同じであるが, 所属度の推定にはエントロピー正則化ファジィ K -means の方法を用いる.

データ解析

iris データを用いて, 提案したファジィ DP-means がうまく働くかを検討する. ファジィパラメータ η を 0.2, クラスター数パラメータ λ を 10 としたとき, クラスター数 $k=3$ となった (Figure 1). DP-means とは異なり, 所属度は連続的となりソフトな割り当てを可能としている.

Figure 1. ファジィ DP-means による分類結果

引用文献

- Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Kulis, B. and Jordan, M.I. (2012). Revisiting k-means: New algorithms via Bayesian nonparametrics. *Proceedings of the 29th International Conference on Machine Learning*.
- Miyamoto, S., Ichihashi, H. and Honda, K. (2008). *Algorithms for Fuzzy Clustering*. Springer-Verlag.