モード探索型クラスタリング

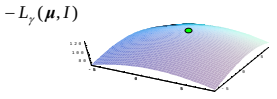
江口 真透　　数理・推論研究系　教授

【 アブストラクト】　非階層クラスタリングの方法を提案する．従来のクラスター法は，例えば k−平均法，ファジィC−平均法，ガウス混合モデル法などはクラスター数を予め決めないと実行できない弱点がある．この発表では，**クラスター数を逐次的に決める方法** を提案する．タリス・エントロピーに関連したパワー・ロス関数を計算して，その局所最小解の個数を k とする．このように k を自動的に決め，k 平均（局所最小解）からの距離によって k 個のクラスターを定めることができる．パワー・エントロピーがデータセットを異なる定常状態をもつ k 個のサブセットに分解していると解釈できる．

【 動機 】　p 次元のデータ $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ に対して正規モデル $N(\boldsymbol{\mu},V)$ に対して次のロス関数を考察していた．

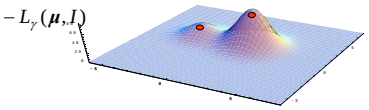
$$L_\gamma(\boldsymbol{\mu},V)=-\det(V)^{-\frac{\gamma}{1+\gamma}}\sum_{i=1}^{n}[\exp\{-\tfrac{\gamma}{2}(\boldsymbol{x}_i-\boldsymbol{\mu})^{\mathrm{T}}V^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\}-1]/\gamma \qquad (1)$$

以下のような2次データに対して平均ベクトル$\boldsymbol{\mu}$の推定をロス関数 (1)の最小化で行うと奇妙なことが．．．



最尤法 ($\gamma=0$)は大域学習

data set



ロバスト推定法 ($\gamma=1$) は局所学習

【 方法 】　予め分散行列V_0を決めて，$L_\gamma(\boldsymbol{\mu},V_0)$ の局所最小解 $\boldsymbol{\mu}_1,\cdots,\ \boldsymbol{\mu}_k$ をk-平均と定義して，k-クラスターを次のように定める．

$$C_j=\{\ \boldsymbol{\mu}:(\boldsymbol{\mu}-\boldsymbol{\mu}_j)^{\mathrm{T}}V_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_j)\le(\boldsymbol{\mu}-\boldsymbol{\mu}_\ell)^{\mathrm{T}}V_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_\ell)\ (\forall\ell\neq j)\ \}\qquad (j=1,\cdots,k) \qquad (2)$$

【アルゴリズム】

(A1)　データ$\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$の中からロス関数 L_γ を最小にするものを初期値として，反復(*)の収束値 $\boldsymbol{\mu}_1$を求める．

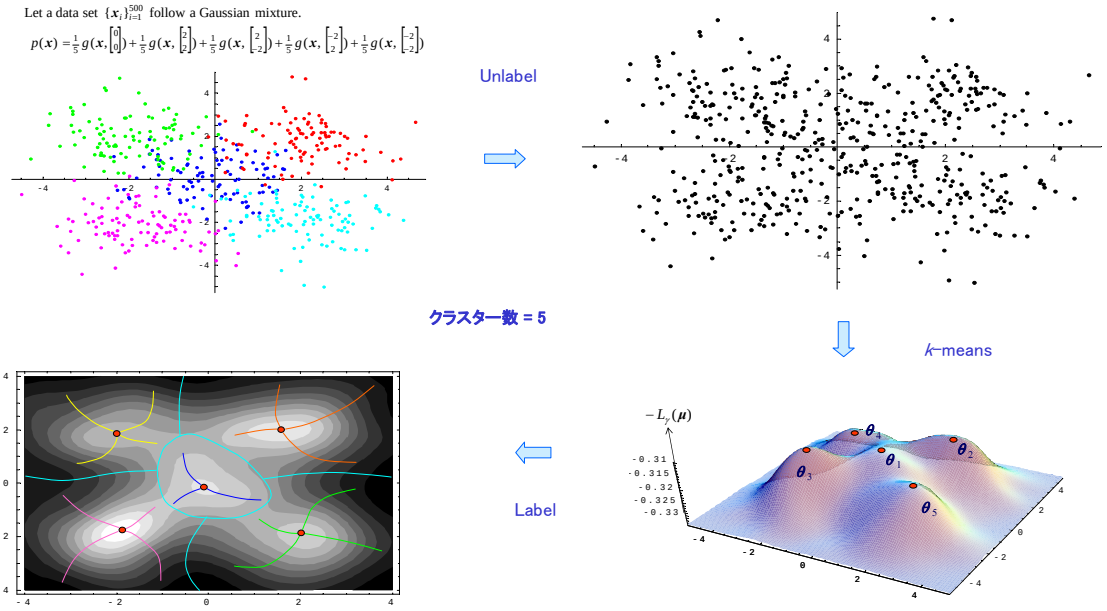
(A2)　$t=2,\cdots$に対して，データ$\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$の中から$\{\boldsymbol{\mu}_1,\cdots,\boldsymbol{\mu}_{t-1}\}$への距離: $\delta(\boldsymbol{x})=\min_{\boldsymbol{\mu}\in\{\boldsymbol{\mu}_1,\cdots,\boldsymbol{\mu}_{t-1}\}}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}V_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$
のα番目に大きいものを初期値として反復(*)の収束値 $\boldsymbol{\mu}_t$ を求める．

$$(*)\qquad \boldsymbol{\mu}\leftarrow\frac{\sum_{i=1}^{n}w_i(\boldsymbol{\mu})\boldsymbol{x}_i}{\sum_{i=1}^{n}w_i(\boldsymbol{\mu})},\qquad w_i(\boldsymbol{\mu})=\exp\{-\tfrac{\gamma}{2}(\boldsymbol{x}_i-\boldsymbol{\mu})V_0^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\}$$

（ 繰り返し重み付け平均）

(A3)　もし $\boldsymbol{\mu}_t\in\{\boldsymbol{\mu}_1,\cdots,\boldsymbol{\mu}_{t-1}\}$ とならば ステップ(A2)を停止し，$k=t-1$と定め，そうでなければステップ(A2)を続ける．

【 簡単な 例題 】



【 結論 と 課題 】　正規分布の平均をγパワー・エントロピー最小化法による推定した．最尤推定値（$\gamma=0$）はデータがどうであれ，標本平均ベクトルで与えられるが，提案方法（$\gamma>0$）では k 平均を推定する．この意味で自発的にクラスター数 k が決まる．提案方法は推定の強いロバスト性が示されていたが，このような大きな **モデル不確定性に対して柔軟な性能** が確認された．提案方法では分散行列はV_0として予め決められていた．クラスターによって共分散が異なる場合はマハラノビスの2乗距離: $(\boldsymbol{\mu}-\boldsymbol{\mu}_j)^{\mathrm{T}}V_j^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_j)$ を使って(2)と同様にしてクラスターを決めることができる．ただし

$$V_j=\arg\min\{\ L_\gamma(\boldsymbol{\mu}_j,V):V>\mathrm{O}\}\qquad (3)$$

今後の課題としてはアルゴリズムの計算量の節約，温度パラメータγのアニーリングについて考察が必要である．

テンソルデータ，関数データ，変数選択 などについても今後の課題となる．