



輪講用資料4/22

Fuzzy Collaborative Forecasting and Clustering

FCM-Type fuzzy Clustering (sec. 4.1, 4.2)

Koki Kitamori (B4)

4.1節 Introduction

- クラスタリングは大きく分けて2種類
 - 階層的アルゴリズム
 - 距離を基準に樹形図を作る
 - 計算量が多く小さいデータにしか使えない
 - 非階層的アルゴリズム
 - k-Means : 計算コストが小さく人気
 - FCM : k-Meansの拡張
 - FCCM : 目的関数を修正したFCMのようなもの

4.2.1 k-Means

- n個のobject→C個のクラスタに分割
- クラスタC→セントロイド b_c に代表される
- 最近隣割り当てとセントロイドの更新を反復

- 目的関数
$$J_{km} = \sum_{c=1}^C \underbrace{\sum_{i \in G_c} ||x_i - b_c||^2}_{\text{クラスタ内のユークリッド距離の2乗和}}$$

- u_{ci} を導入 $J_{km} = \sum_{c=1}^C \sum_{i=1}^n u_{ci} ||x_i - b_c||^2, u_{ci} = \{0,1\}$

4.2.3 Fuzzy C-Means

- クラスタへの所属度 $u_{ci} \in [0,1], \sum_{c=1}^C = 1$ を導入
- $\theta(> 1)$ メンバーシップのファジィ度
- 目的関数 $J_{km} = \sum_{c=1}^C \sum_{i=1}^n u_{ci}^{\theta} \|x_i - b_c\|^2$ (u_{ci} を $\theta(> 1)$ 乗して非線形化)
- 更新則
$$u_{ci} = \left(\sum_{\ell=1}^C \left(\frac{\|x_i - b_c\|^2}{\|x_i - b_{\ell}\|^2} \right)^{\frac{1}{\theta-1}} \right)^{-1}$$
$$b_c = \frac{\sum_{i=1}^n u_{ci}^{\theta} x_i}{\sum_{i=1}^n u_{ci}^{\theta}}$$
- K-Meansと同様に初期値の問題は発生

4.2.4 FCM Variants1

- エントロピー項を付加し非線形化
- $\lambda(> 1)$ メンバーシップのファジィ度

- 目的関数 $J_{efcm} = \sum_{c=1}^C \sum_{i=1}^n u_{ci} \|x_i - b_c\|^2 + \lambda \sum_{c=1}^C \sum_{i=1}^n u_{ci} \log u_{ci}$

- 更新則

$$u_{ci} = \frac{\exp\left(-\frac{1}{\lambda} \|x_i - b_c\|^2\right)}{\sum_{\ell=1}^C \exp\left(-\frac{1}{\lambda} \|x_i - b_{\ell}\|^2\right)}$$

$$b_c = \frac{\sum_{i=1}^n u_{ci} x_i}{\sum_{i=1}^n u_{ci}}$$

- FCMは統計モデルと関連

4.2.5 FCM Variants2

- FCL(Fuzzy c-Lines) プロトタイプ中心を直線に拡張
- FCV(Fuzzy c-Varieties) プロトタイプ中心を超平面に拡張

これらはローカルPCAと同一視される

- FCRM(Fuzzy c-Regression Models) 回帰誤差→回帰モデル推定

クラスタリング基準