

9. ファジィクラスタリング

7月15日

田中洸一

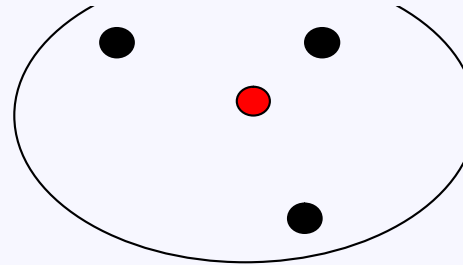
内容

- 問題提起
- ファジィクラスタリング
- 更新式
- 数値例

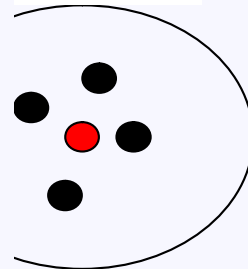
問題提起(1)

= データ

= 重心



クラスタ 1



クラスタ 2

クラスタ数を2として
クラスタリングする

図 : k-means によるクラスタリング (イメージ図)

- 重心は見た目妥当な位置にあり、問題はなさそう

問題提起(2)

- ちなみにk-meansについて

評価関数
$$\sum_{k=1}^K \sum_{x \in C_k} \|x - c_k\|^2$$

※ x はデータ

※ K はクラスタ数

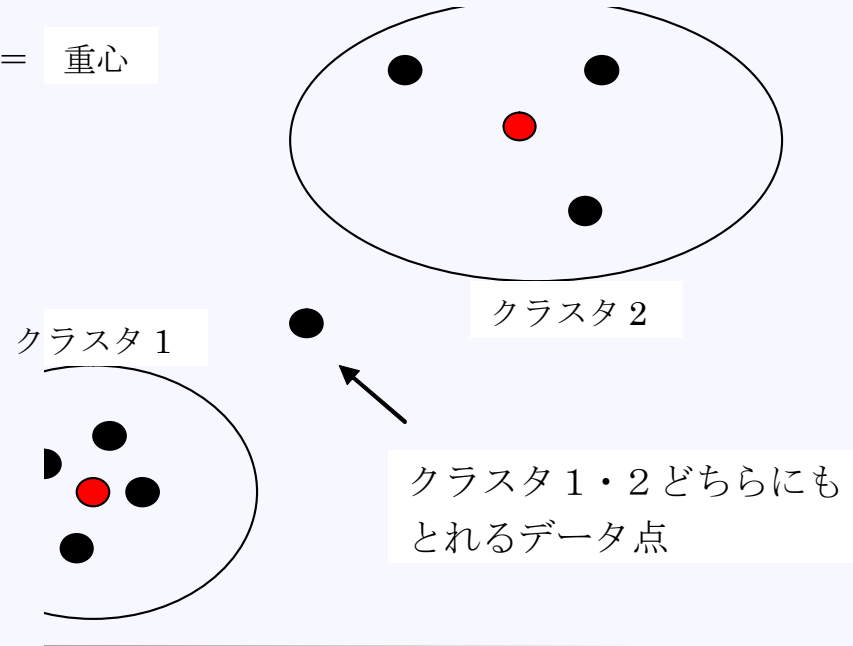
※ c_k はクラスタ C_k の重心

評価関数を最小化となるようなデータのクラスタへ割り当てる

問題提起(3)

= データ点

= 重心



図：k-means によるクラスタリング (イメージ図)

仮にこのデータ点をクラスタ2に割り当てるとすると、

問題提起(4)

= データ点

= 重心

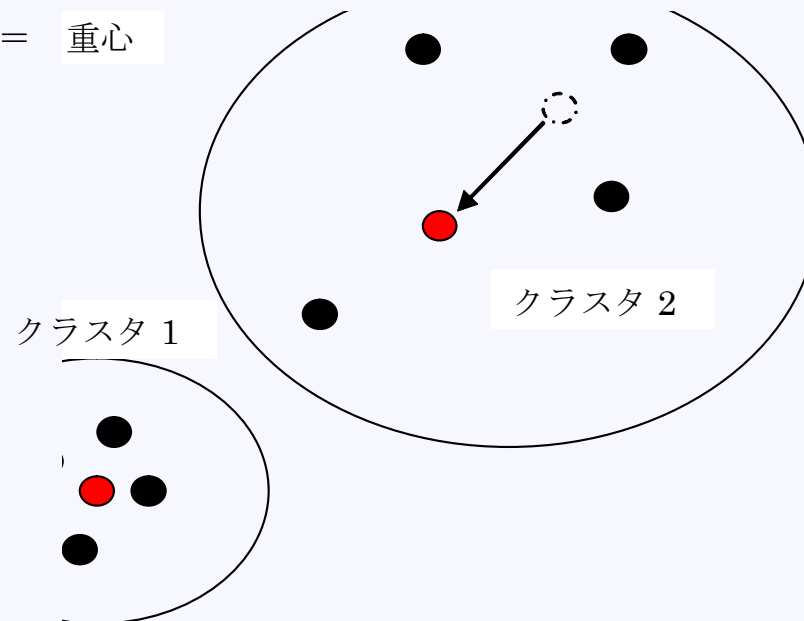


図 : k-means によるクラスタリング (イメージ図)

距離が離れたデータのため、重心が大きく移動してしまう。

問題提起(5)

重心が大きく移動してしまう



他の分類にも悪影響を与えてしまう



データの距離に応じて重心の移動を調整する
必要がある



ファジィクラスタリング

ファジィクラスタリング(1)

- ファジィクラスタリングとはk-means評価関数にクラスタに属する重み g_{ik} を考慮したもの

$$\text{k-means} \quad \sum_{k=1}^K \sum_{x \in C_k} \|x - c_k\|^2 \quad (9.1)$$

$$\text{Fuzzy c-means} \quad J = \sum_{i=1}^N \sum_{k=1}^K (g_{ik})^m \|x_i - c_k\|^2 \quad (9.2)$$

※データ $N(x_1, x_2, \dots, x_N)$

※ $m > 1$ とし、値が大きいとよりファジィな分割となる

ファジィクラスタリング (2)

- ファジィクラスタリングをするには(9.2)の g_{ik} と c_k を求めていく。具体的には、

[step1] 初期値 $c_k^{(0)}$ を設定。

[step2] 式(9.2)において、 c_k を $c_k^{(t)}$ に固定したときに、 J を最小化する $g_{ik}^{(t)}$ を求める。

[step3] 式(9.2)において、 g_{ik} を $g_{ik}^{(t)}$ に固定したときに、 J を最小化する $c_k^{(t+1)}$ を求める。

step2,3を収束するまで繰り返す

更新式 (step2 ($g_{ik}^{(t)}$) について)

- $x_i = c_k$ となるような c_k が存在する場合
 - $x_i = c_k$ となる k については $g_{ik} = 1$
 - それ以外の k については $g_{ik} = 0$
- $x_i = c_k$ となるような c_k が存在しない場合

$$\text{更新式 } g_{ik}^{(t)} = \frac{1}{\sum_{j=1}^K \left(\frac{\|x_i - c_k^{(t)}\|^2}{\|x_i - c_j^{(t)}\|^2} \right)^{2/(m-1)}} \quad (9.5)$$

更新式 (step3 ($c_k^{(t+1)}$) について)

$$\text{更新式 } c_k^{(t+1)} = \frac{\sum_{i=1}^N \left(g_{ik}^{(t)} \right)^m x_{ij}}{\sum_{i=1}^N \left(g_{ik}^{(t)} \right)^m}$$

Fuzzy c-meansはk-meansと同様、クラスタリング結果が初期値 $c_k^{(0)}$ に依存する

数値例(1)

データ 座標

1 (2.0 ,1)

2 (1.0 ,2)

3 (2.0 ,4)

4 (4.0 ,4)

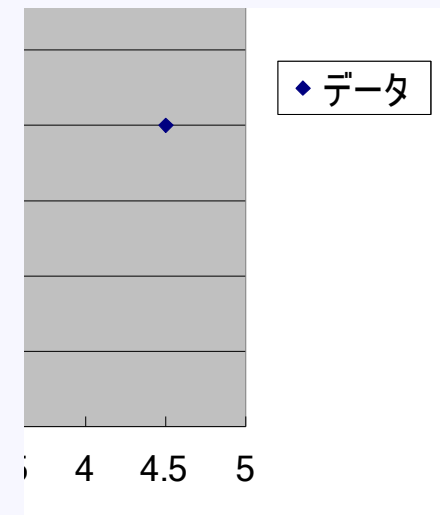
5 (4.5 , 2)

6 (1.5 ,2)

7 (1.0 ,1)

8 (4.0 ,3)

9 (4.2 ,3)



数値例(2)

- ・ファジィクラスタリングでクラスタリングする
 - ・クラスタ数=2
 - ・繰り返しの最大回数=50
 - ・他はデフォルト

g_{ic}	1	2
[1,]	0.08473883	0.915261172
[2,]	0.02765818	0.972341823
[3,]	0.55261063	0.447389367
[4,]	0.93853187	0.061468128
[5,]	0.86656817	0.133431834
[6,]	0.01700489	0.982995106
[7,]	0.04273702	0.957262975
[8,]	0.99865974	0.001340262
[9,]	0.99403469	0.005965306

数値例(3)

重心

座標

クラスタ1 (3.984629 3.105817)

クラスタ2 (1.410223 1.653752)

よって、クラスタリング結果は

データ

2 2 1 1 1 2 2 1 1

数値例(4)

- 同様のデータをk-meansでクラスタリングする

重心 座標

クラスタ1 (1.375 1.5)

クラスタ2 (3.740 3.2)

クラスタ番号は逆だが、cmeansと同じクラスタリング結果となった。

データ

1 1 2 2 2 1 1 2 2

数値例(5)

