

第 VII 部

市場の分析

8 K-means 演習

ポイント

- K-mean の手順
- クラスタ数

- 計算手順

8.1 はじめに

- 具体的な値を使ってクラスタリングを行います。
- クラスタとは『塊』です。
- 特別な意味はありません。

命題

 $a, b > 0$ とする。

$$a^2 > b^2 \implies a > b \quad (1)$$

証明

$$a^2 > b^2 \quad (2)$$

$$a^2 - b^2 > 0 \quad (3)$$

$$(a + b)(a - b) > 0 \quad (4)$$

仮定より、 $a, b > 0$ なので左辺第一因数は正

である。

したがって第二因数も正でなくてはならない。

$$a - b > 0 \quad (5)$$

$$a > b \quad (6)$$

(終)

8.2 *K-means* の手順

1. 予めクラスタ数を決定する。
2. ランダムに初期値となる各クラスタの中心点を布置。
3. 各データを最も近い中心を持つクラスタに分類。
4. 所属するデータの算術平均を新たな中心とする。
5. 3.4. を中心点が移動しなくなるまで繰り返す。

8.2.1 クラスタ数の設定

- 2 クラスタだと、特定の値の優劣のみで判断されやすい。
- 3 クラスタ以上だと、それぞれの特徴を表現できる。
- 多すぎるクラスタ数だと特長を把握する難易度があがる。

8.2.2 初期値の設定

- それぞれの軸における最大値と最小値の間の値が好ましい。
- 初期値により分類結果は異なるので恣意的な値は避ける。
- 初期値は乱数で発生させる。

8.3 使用予定の Excel365 の関数

- SUMSQ 関数
- SQRT 関数
- AVERAGEIFS 関数
- COUNTIFS 関数
- MIN 関数
- IFS 関数
- IF 関数

SUMSQ(数値 1, [数値 2], ...)

引数の平方和を返します

表 1 SUMSQ 関数の引数

引数	摘要
数値 1	平均を求める 1 つ目の数値、セル参照、またはセル範囲を指定します。
数値 2,...	平均を求める追加の数値、セル参照、または範囲 (最大 255)。

AVERAGEIFS(**平均範囲**, **条件範囲 1**, **条件 1**, [**条件範囲 2**, **条件 2**], ...)

複数の検索条件に一致するすべてのセルの平均値 (算術平均) を返します。

表 2 AVERAGEIFS 関数の引数

引数	摘要
平均範囲	平均する範囲を指定します。
条件範囲 1	「条件 1」 による評価の対象となる範囲を指定します。
条件 1	「条件範囲 1」 の条件を指定します。

MIN(数値 1, [数値 2], ...)

引数の最小値を返します

表 3 MIN 関数の引数

引数	摘要
数値 1	最小値を求める 1 つ目の数値、セル参照、またはセル範囲を指定します。
数値 2,...	最小値を求める追加の数値、セル参照、または範囲 (最大 255)。

IFS(論理式 1, 値が真の場合 1, 論理式 2, 値が真の場合 2)

論理式の結果に応じて、定められた値を返します。

表 4 IFS 関数の引数

引数	摘要
論理式 1	条件 1
値が真の場合 1	論理式 1 の結果が「真」の場合に返す値。
論理式 2	条件 2
値が真の場合 2	論理式 2 の結果が「真」の場合に返す値。

COUNTIFS(条件範囲 1, 条件 1, [条件範囲 2, 条件 2], ...)

複数の検索条件に一致するすべてのセルの個数を返します。

表 5 COUNTIFS 関数の引数

引数	摘要
条件範囲 1	「条件 1」による評価の対象となる範囲を指定します。
条件 1	「条件範囲 1」の条件を指定します。

8.4 具体的な計算手順

1. クラスター中心を 3 組作る。
2. 3 つのクラスター中心とそれぞれの点の差を求める。
3. クラスター中心からの差の平方和を求める。
4. もっとも近いクラスター中心に所属させる。
5. それぞれのクラスターの中心を求めクラスター中心を更新。
6. クラスター中心と更新されたクラスター中心を比較。
 - (a) 変化していたらクラスター中心を更新し 2. に戻る。
 - (b) 変化していなかったら終了。