**プログラミング言語**

**kmeans法**

令和５年８月10日20-416 小柴 颯太

1 .kmeans法について

1.１kmeans法とは

互いに近いデータ同士は同じクラスタであるという考えに基づいたデータ群をk個に分類するクラスタリングの手法。

クラスが3つの場合。

（１）初期値を3つ決める。

（２）３つの初期値から最も近いクラスタに割り振る。

（３）各クラスの重心を測る。

重心が変化しないまで（２）と（３）を繰り返す。

1.２欠点

初期値によってクラスタリング結果が異なる。

例）初期値が(0.1,0.1) (-0.25,1) (0.75,0.25)の場合

　　結果：（0.12630833,0.04005）(-0.03502727,1.00790909) (0.65487143,0.43904286)

初期値が(0,0) (-0.25,1) (0.75,0.25)の場合

結果(0,0) (-0.03502727,1.00790909) (0.61925,0.430475)

1.３kmeans++法

このような結果が初期値に依存してしまうことを改善したkmeans++法がある。

kmeans++法とはクラスタの中心の初期値を遠ざけるように選択することで、kmeans法の初期値問題を改善したアルゴリズムである。

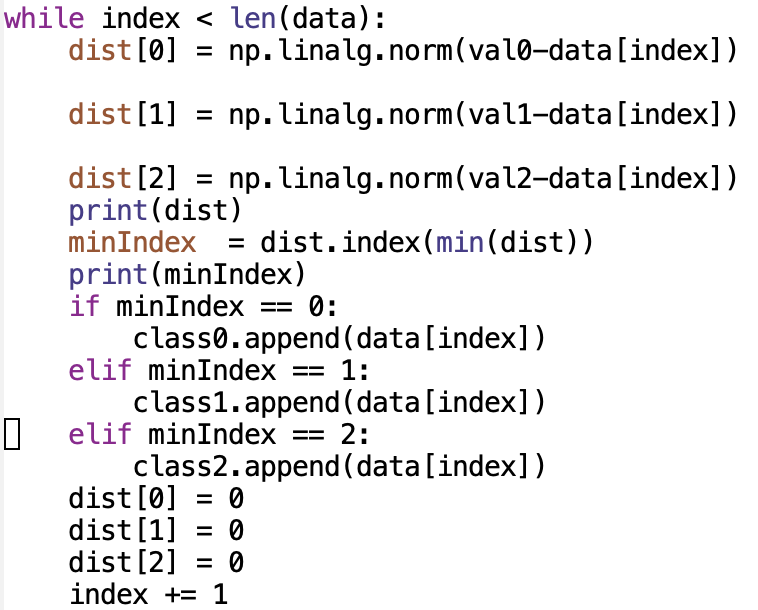
つまり、クラスタの中心同士は離れていた方が良いという考えである。

参考文献：<https://hogetech.info/machine-learning/algorithm/kmeans>

2.pythonによるkmeans法の実装

2.１〜2.４は無限ループの中とする。

2.1クラス分け



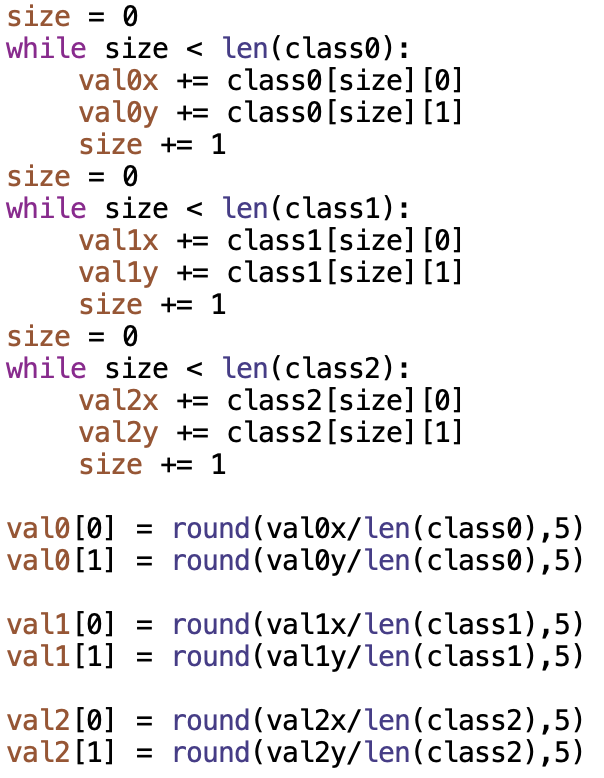
各クラスタの重心との距離を測り最も小さい値をとったものにクラスタリングする。

2.2 比較するために重心を記録、重心を計測するための各クラスの合計を測る変数を生成

テキスト

自動的に生成された説明

2.３ 各クラスの重心を計算



有効桁数を指定しないと重心が更新し続けるので、round()を使い小数点第５位までを出す。

2.４ 重心が移動したか比較をする。

グラフィカル ユーザー インターフェイス, テキスト, アプリケーション

自動的に生成された説明

重心が移動していなかったら、breakでループから抜ける。

2.５ 点をプロット、表示

グラフィカル ユーザー インターフェイス, テキスト

自動的に生成された説明

クラス０を赤、クラス１を青、クラス２を緑で表示し各重心を×とする。

3.実行と結果

makeData.pyを実行

グラフ, 散布図

自動的に生成された説明

結果

グラフ, 散布図

自動的に生成された説明

重心

|  |  |  |
| --- | --- | --- |
| Class0 | Class1 | Class2 |
| 0,0 | 0.3517,0.91866 | 0.60982,0.4975 |

4.最後に

プログラミング言語の授業を通して、最小二乗法、kmeans法を学習し今まで名前は聞いたことあるが深くは理解していなかったデータサイエンスというものを学習することができ興味を持つことができた。データサイエンスを触れることができ自分にとって有意義なものとなった。この授業で機械学習について興味をもったため自身でも学習をして見たいと思った。