

パターン情報学 プログラミング課題 2

03-170312
航空宇宙工学科
新幡 駿

2017 年 10 月 28 日

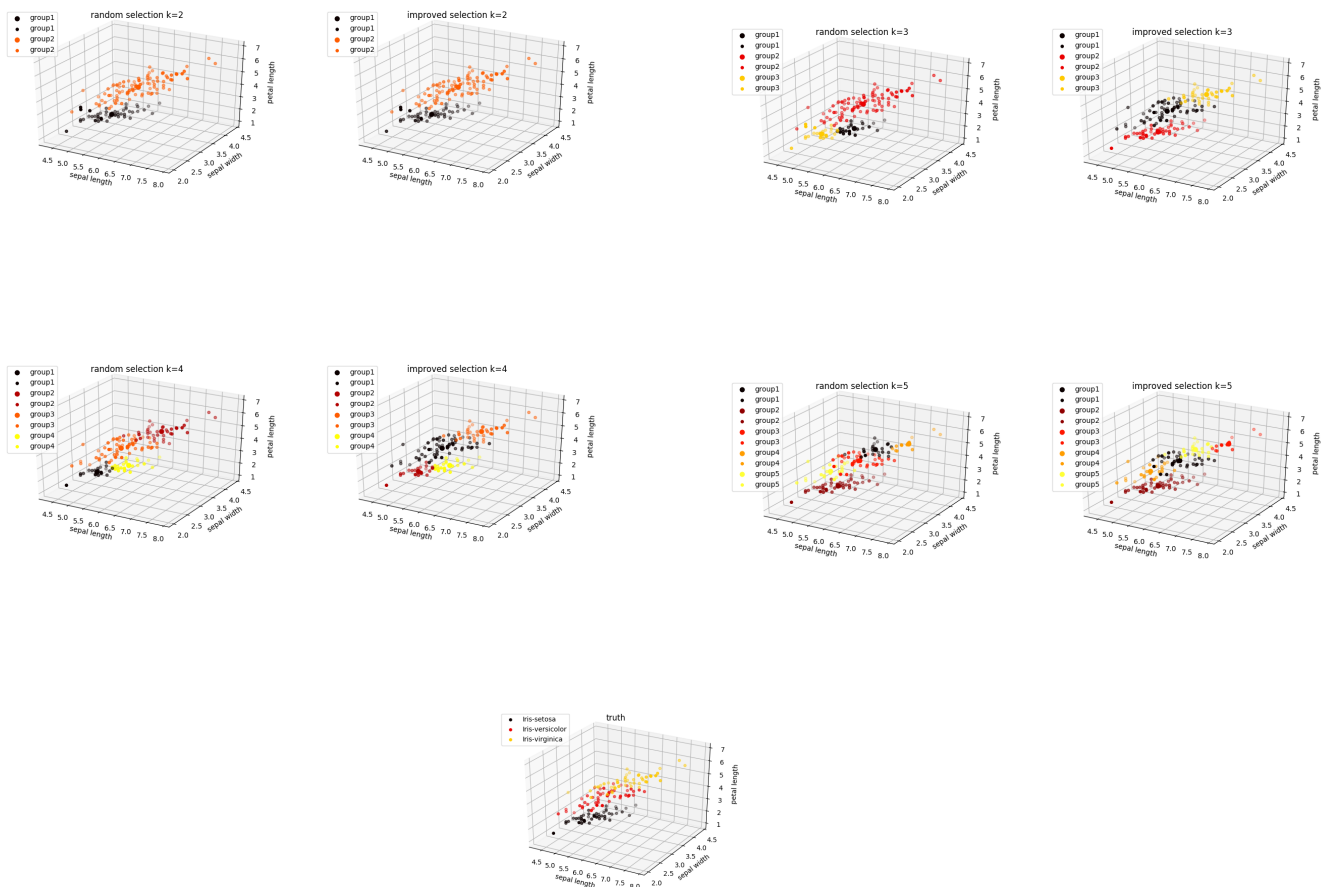
1 アルゴリズムについて

初期クラスタの代表点をランダムに選ぶ方法と、離れるように選ぶ方法で実装した。離れた初期クラスタの選び方は

1. ランダムに初めの一個の点を選ぶ
2. 全ての点について既に選んだ点からの距離を計算し、和が最大になるものを新しい点として選ぶことを $k-1$ 回繰り返す

以上のようなアルゴリズムである。クラスタの更新の収束判定はクラスタの代表点の距離が一回前のクラスタの代表点から閾値 (0.0001) 以下であることとした。また可視化には [1] を参考にした。

2 結果



3 考察

二つの初期クラスタの比較からわかるように、初期状態の影響を受ける。また $k=3$ の時はなるべく離れるようにクラスタの代表点を選んだほうが正解に近かった。初期状態の影響による差は k が大きほうが多かった。また収束するまでの速度は二つの初期クラスタで大きく変わらなかった。

参考文献

- [1] stack overflow <https://stackoverflow.com/questions/28033046/matplotlib-scatter-color-by-categorical-factors/28033497>