

Описание параметров метода
многомерной автоматической
итерации квантования (MASC)

1. Основные параметры алгоритма:

k - число кластеров (целое число int)

Определяется через метод поиска чисел
смысла (автоматический) или
вручную, если известна предполагаемая
структура данных (в данном случае
нет)

max iter - максимальное число
итераций (целое число int)

по умолчанию $\text{max iter} = 100$

Ограничивает количество итераций
для предотвращения бесконечных
вычислений. Если алгоритм выходит
за рамки, то итерации прекраща-
ются автоматически.

tol - критерий остановки (число с
плавающей точкой float) по умол-
чанию $1e-4$. Возвращает минималь-
ное изменение целевой функции для остано-
вки алгоритма. Но есть, если

центры кластеров итеративно итеративно
тест на 0.0001, алгоритм завершает
работу

2. Параметры управления итерационным
и управлением кластерами
max_noise - пороговый уровень шума
(значение с плавающей точкой float)
по умолчанию 0.2 (диапазон от 0 до 1)
radius_multiplier - коэффициент
расширения радиуса (значение с плава-
ющей точкой float) Увеличивает
исходные радиусы кластеров для
более точной кластеризации в случае
если в данных много шума и следует
снизить уровень кластеризации.
k_neighbors (значение int) по умолчанию 5.
Определяет количество "соседей"
"разбросанных" вокруг кластера
3. Метод оптимизации параметров
Автоматически подбирает лучший
кластер: find_optimal_k
Также доступен метод

полюса и симметричный метод.

Возвращает $optimal_k$ - наилучшее измерение между центрами, если они дают в результате разное значение.

Адаптивные радиусы ($adaptive_optimal_radius$) - Для каждого кластера вычисляется среднее расстояние до k ближайших соседей. Радиусы корректируются с учетом $radius_multiplier$.

3. Параметр оценки качества кластеризации
В данной задаче присутствуют две целевые метрики:

1. Индекс Каллини - Харабаса оценивает разделение кластеров на основе переноса. Чем выше значение тем лучше кластеры разделены друг от друга. Значение 80, 2 говорит о том, что кластеры достаточно хорошо разделены, но есть недочеты в зонировании.
2. Симметричный подтверждение оценивает насколько каждая точка похожа на

есть мастер по сравнению с другими
мастерами. Значение 0,0270 в
мизе, что означает: количество
различных, имеющих такие находки
на графиках.

Таким образом, авторские работы
высшего ~~уровня~~ мастера, но с их
мелкими различиями.