Метод "Fuzzy k-means" - це алгоритм кластеризації, який використовується для розділення множини даних на кілька кластерів. На відміну від стандартного алгоритму "k-means", де кожен об'єкт належить до одного з кластерів, "Fuzzy k-means" дозволяє об'єктам належати до кількох кластерів одночасно з різною ступінню належності.

Основна ідея "Fuzzy k-means" полягає в тому, що кожен об'єкт з множини даних характеризується вектором приналежності, який складається з дійсних чисел від 0 до 1. Кожне число в цьому векторі вказує, наскільки об'єкт належить до кожного з кластерів. Якщо число близьке до 1, то об'єкт майже повністю належить до кластеру, якщо ж число близьке до 0, то об'єкт належить до кластеру з малою ступенем належності.

Параметри та позначення для данного алгоритму:

b - параметр, що контролює ступінь "розмитості" (fuzziness) приналежності об'єкту до кластерів. Зазвичай, значення b обирають в межах від 1 до 2,

– центр кластеру під номером i,

с - кількість кластерів,

N – кількість точок,

координати точки під номером j,

Max\_iter – максимальна кількість ітерацій алгоритму до зупинки,

Eps – мінімальне зміщення центрів кластерів внаслідок ітерації для продовження алгоритму.

Відповідно до параметрів результат алгоритма може відрізнятися.

На вхід до алгоритму подаєтються точки, початкові центри кластерів (також можуть вибиратися випадково) та параметр b. Зазвичай координати точок та центрів центроїдів зберігаються у масивах з розмірностями (N, k) та (c, k) де k кількість координат.

Допоміжні параметри:

- ступінь належності i-го об'єкту до j-го кластеру.

Часто дінні кооефіцієнти також зберігаються у матриці з розмірністю (c, N), адже для кожної точки зберігається 3 ступеня залежності відповідно до кожного центру кластера.

Після роботи алгоритм повертає ступені залежності. Також може повертати і центри кластерів, наприклад для обчислення точних кластерів за допомогою звичайного k-means.

Алгоритм "Fuzzy k-means" складається з таких етапів:

1. Вибрати кількість кластерів, на які будуть розділятися об'єкти множини даних.
2. Ініціалізувати випадковим чином початкові координати центрів кластерів. (можна вибрати певні точки з множини)
3. Для кожного об'єкту вирахувати вектор приналежності, який відображає ступінь належності до кожного з кластерів згідно з формулою:

де - ступінь належності i-го об'єкту до j-го кластеру, – відповідний центр кластеру, c - кількість кластерів, b - параметр, що контролює ступінь "розмитості" (fuzziness) приналежності об'єкту до кластерів. Зазвичай, значення b обирають в межах від 1 до 2.

1. Для кожного кластеру вирахувати нові координати центру, використовуючи формулу:

Де - нові координати центру i-го кластеру

1. Повторювати кроки 3-4 до тих пір, поки вектори приналежності не стабілізуються або досягне максимальної кількості ітерацій. Кластеризацію можна вважати завершеною, коли зміни координат центрів кластерів стають незначними.

Метод "Fuzzy k-means" дозволяє отримувати більш точні результати кластеризації, оскільки дозволяє об'єктам належати до кількох кластерів одночасно з різною ступінню належності. Однак, цей метод може вимагати більшої обчислювальної потужності порівняно зі стандартним "k-means", оскільки він використовує більш складні формули для обчислення векторів приналежності та координат центрів кластерів. Також за його допомогою можна легко помітити точки, що розташовані на перетині кластерів та точки, що явно не належать жодному з них.