kNN - k Nearest Neighbor - k Ближайших Соседей — это один из алгоритмов классификации. Задача классификации— это отнесение объекта к одному из заранее определенных классов на основании его формализованных признаков

алгоритм будет использоваться для построения модели из обучающих данных, а также для прогнозирования новых точек данных. Он также будет содержать информацию, которую алгоритм извлек из обучающих данных. В случае KNeighborsClassifier он будет просто хранить учебный набор.

**Стандартизация** набора данных – это корректировка каждого значения x так, чтобы они находились примерно в одном диапазоне.

Теперь мы можем применить метод transform для стандартизации всех признаков, чтобы они имели примерно одинаковый масштаб. Мы сохраним преобразованные семплы в переменной scaled\_features

Шаг 2: Назначение объектов к ближайшим центроидам

Каждый объект назначается к ближайшему центроиду на основе некоторой метрики расстояния (обычно евклидово расстояние).

Объекты, находящиеся ближе к одному центроиду, считаются более похожими друг на друга, чем на объекты, находящиеся ближе к другим центроидам.

Шаг 3: Перерасчет центроидов

Вычисляются новые центроиды путем вычисления среднего значения всех объектов, назначенных к каждому кластеру.

Это обновление центроидов на основе текущего разделения объектов на кластеры.

Коэффициент силуэта ( sklearn.metrics.silhouette\_score) является примером такой оценки, где более высокий показатель коэффициента силуэта относится к модели с лучше определенными кластерами. Коэффициент силуэта определяется для каждого образца и состоит из двух баллов:

a : Среднее расстояние между образцом и всеми другими точками того же класса.

b : Среднее расстояние между образцом и всеми другими точками в следующем ближайшем кластере .