

Центроид кластера (Cluster centroid)

Синонимы: Центр кластера, Центр масс кластера, Центр тяжести кластера, Center of cluster

Разделы: [Алгоритмы](#)

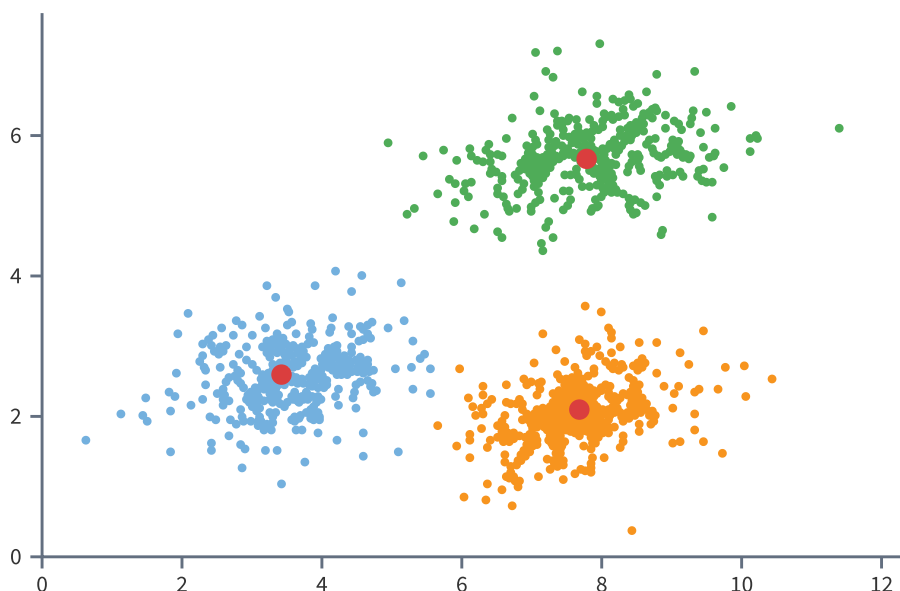
Центроидом или центром кластера называется вектор, каждый элемент которого вычисляется как среднее по соответствующим элементам (координатам) векторов наблюдений, входящих в кластер:

$$X^c = \frac{1}{n} \sum_{i=1}^n x_i,$$

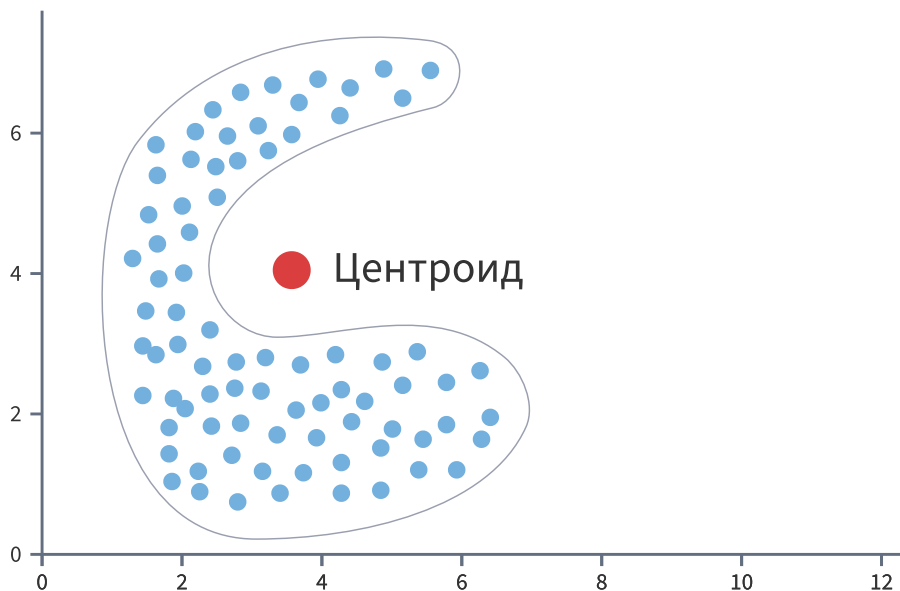
$$Y^c = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$Centroid = (X^c, Y^c).$$

Если кластер имеет правильную сферическую или эллипсоидную форму, то геометрически центроид будет расположен вблизи центра области пространства, занимаемой кластером (для простоты на рисунках представлен 2-мерный случай).



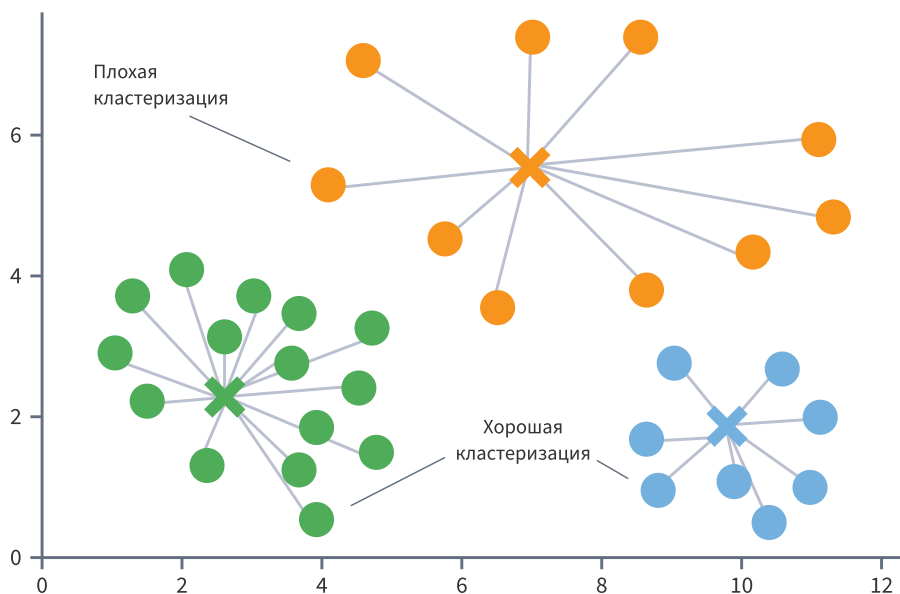
Однако если кластер имеет сложную форму, то его центроид может оказаться за пределами кластера, как показано на рисунке ниже.



Понятие центроида играет очень важную роль в кластерном анализе. Во-первых, расстояние между центроидами кластеров могут служить для оценки качества кластеризации: чем оно больше, тем сильнее различаются кластеры, т.е. лучше кластеризация.

Во-вторых, чем плотнее векторы объектов кластера расположены вокруг его центроида, тем точнее кластеризация и лучше кластерная модель. Обычно, расстояние понимается евклидовым, а в качестве меры используется среднее расстояние от центра до всех объектов:

$$D_c = \frac{1}{m} \sqrt{\sum_{i=1}^n (C - X_i)^2}$$



И, наконец, центроиды используются в алгоритмах кластеризации, основанных на расстоянии, таких, например, как k-means. В нем с помощью случайно выбираемых центроидов производится начальная инициализация кластеров, а в процессе обучения — их перевычисление. Алгоритм работает до тех пор, пока центры кластеров не перестанут смещаться от итерации к итерации.

Алгоритмы кластеризации, использующие центроиды, образуют целый класс (centroid-based algorithms) к которому относятся k-means и нечеткий k-means, k-medoids, EM-алгоритм и другие.