|  |  |  |  |
| --- | --- | --- | --- |
| **АЛГОРИТМ** | **ПРИНЦИП РАБОТЫ** | **ПРЕИМУЩЕСТВА** | **НЕДОСТАТКИ** |
| **Иерархическая кластеризация** | Строит дерево (иерархию) вложенных кластеров. | * + позволяет построить дендрограмму кластеризации и тем самым лучше проанализировать данные;   + позволяет найти иерархию в данных. | * + необходимо выбирать оптимальное количество кластеров;   + неравномерное распределение кластеров;   + работает на небольшом датасете. |
| **EM-кластеризация (модель гауссовой смеси)** | Предполагает, что любое наблюдение принадлежит ко всем кластерам, но с разной вероятностью. | * + хорошо работает на данных в форме эллипса, окружности, на перекрывающихся данных;   + устойчив к шуму в данных;   + работает быстро при удачном выборе начальных кластеров. | * + необходимо выбирать оптимальное количество кластеров;   + работает долго при неудачном выборе кластеров. |
| **Спектральная кластеризация** | Снижает размерность данных, а затем применяет более простые алгоритмы кластеризации, например *k-means*). | * + подходит для кластеризации данных с большим количеством признаков;   + эффективен для кластеризации разрежённых данных;   + хорошо работает на практически любых распределениях данных. | * + необходимо выбирать оптимальное количество кластеров;   + подходит для задач сбалансированной классификации, то есть количество точек между кластерами не сильно отличается. |
| **Кластеризация на основе плотности** | Находит в датасете области с высокой плотностью и, основываясь на этом, образует кластеры. | * + нет необходимости выбирать оптимальное количество кластеров;   + устойчив к шуму и выбросам. | * + не способен качественно кластеризовать наборы данных с большой разницей в плотности;   + иногда бывает сложно подобрать параметры для моделирования. |