**Вопросы к лабораторной работе №6**

1. Что решают задачи кластеризации в машинном обучении?

Группировка данных: Кластеризация помогает разделить данные на группы, где объекты в каждой группе похожи друг на друга.

Поиск скрытых структур: Кластеризация может выявить внутренние закономерности и структуры в данных, которые могут быть невидимы на первый взгляд.

Упрощение данных: Группировка данных в кластеры позволяет упростить сложные данные и сделать их более понятными для анализа.

1. Расскажите принцип работы метода K-means.

Принцип работы метода K-means:

1. Начинается с инициализации центров кластеров.
2. Каждый объект данных относится к кластеру, чей центр находится ближе всего.
3. Пересчитываются центры кластеров на основе объектов, отнесенных к этим кластерам.
4. Процесс повторяется до сходимости, когда центры кластеров больше не меняются или меняются незначительно.
5. Как можно выбрать оптимальное количество кластеров в K-means?

Метод локтя: построение графика зависимости инерции от количества кластеров и выбор точки, где изменение инерции замедляется (локоть).

Метод силуэта: рассчитывается силуэт для различных значений количества кластеров и выбирается значение с наибольшим силуэтом.

1. Расскажите принцип работы метода иерархической кластеризации.
2. Каждый объект начинает как отдельный кластер.
3. Два самых близких кластера объединяются в один на каждом шаге, пока все объекты не будут объединены в один кластер.
4. Результатом является дерево (дендрограмма), показывающее иерархию объединения кластеров.
5. Для чего можно использовать дендрограмму в методе иерархической кластеризации?

По высоте срезов на дендрограмме можно определить оптимальное количество кластеров: точку, где происходит значительное увеличение расстояния между узлами дерева.

1. Какие метрики используют для оценки качества кластеризации?

Силуэт: Оценка того, насколько объекты внутри одного кластера похожи друг на друга и насколько они отличаются от объектов в других кластерах.

Компактность кластеров (англ. Cluster Cohesion): идея данного метода в том, что чем ближе друг к другу находятся объекты внутри кластеров, тем лучше разделение.

Разделимость кластеров (англ. Cluster Separation): в данном случае идея противоположная — чем дальше друг от друга находятся объекты разных кластеров, тем лучше.