***Практическое обучение***

***(19.07.04)***

***БФИ1701***

***Лаврентьев.Н.Л.***

**Машинное обучение** – способы решения анализа данных, позволяющие системе обучаться, в ходе обработки данных и решения определённых задач. **Постановка задачи**.

Дано конечное множество (объектов, ситуаций), по каждому из которых есть некоторые данные (описание). Совокупность всех имеющихся описаний представляет собой обучающую выборку. Требуется по этим частным данным выявить общие зависимости, закономерности, взаимосвязи, присущие выборке.

**Два типа задач обучения.**

Обучение с учителем – каждая ситуация представляет собой пару «объект, ответ». Нужно найти зависимость ответов от описаний объектов и построить алгоритм, принимающий на входе описание объекта и выдающий на выходе ответ.

Обучение без учителя – нужно найти зависимость между объектами, в этом случае ответы не задаются.

**Обучение без учителя.**

В рамках обучения без учителя рассмотрим задачу кластеризации.

Задача кластеризации заключается в том, чтобы сгруппировать объекты в кластеры. Функционалы качества могут определяться по-разному, например, как отношение средних межкластерных и внутрикластерных расстояний.

Рассмотрим несколько методов кластеризации:

* K-means (К-средник)

Это итеративный алгоритм кластеризации, основанный на минимизации суммарных квадратичных отклонений точек кластеров от средних координат (центра) этих кластеров. Первоначально выбирается желаемое количество кластеров. P Затем случайным образом из входных данных выбираются выбранное кол-во элементов в соответствии с кол-вом кластеров, в каждый из которых теперь включено по одной точке. Далее ищем ближайшего соседа к текущему центра. Добавляем точку к соответствующему кластеру и пересчитываем положение центра с учетом координат новых точек. Алгоритм заканчивает работу, когда центр перестраёт меняться.

* DBSCAN

Алгоритм группирует тесно расположенные точки, тем самым разделяя их на кластеры. Точки, которые расположены далеко от кластеров отмечаются как выбросы. Для этого алгоритма не нужно заранее знать кол-во кластеров, но необходимо указывать расстояние между точками и минимальное кол-во точек входящих в зону для достижения связи.

* Agglomerative

Алгоритм без фиксированного кол-ва кластеров. С самого начала все точки обозначаем за кластеры, затем наиболее похожие (близкие) кластеры объединяются в новый более крупный кластер. Это происходит до тех пор, пока не образуется один большой корневой кластер.