# Классификация ЛС

1.Классификация – получение категориального ответа на основе набора признаков

2.Класстеризация – распределение данных на группы

3.Регрессия – прогноз на основе выборки объектов с различными признаками

4. Понижение размерности – понижение размерности сведение к большого числа к меньшему

5.Отделение аномалий от стандартных случаев

Категориальный ответ – это набор данных с ограниченным числом уникальных значений

Признак – набор исходных данных, которые поддаются алгоритмам ML

Метка обозначается yn , целевая переменная обозначается y от f – прогнозированный признак дата сета

Дата сет – набор исследуемых данных

Простая запись(таблица):

1. Матрица
2. Разряженная матрица
3. Транзакционные данные

Граф

Упорядоченные данные

Характеристики датасет

1)размерности(количество записей)

2)разрежённость(количество пустых значений)

3)разрешение(диапазон доступных значения)

Процесс обучение – это подача данных на алгоритм ML, которая в результате должна вывести определённый шаблон этих исходных данных