

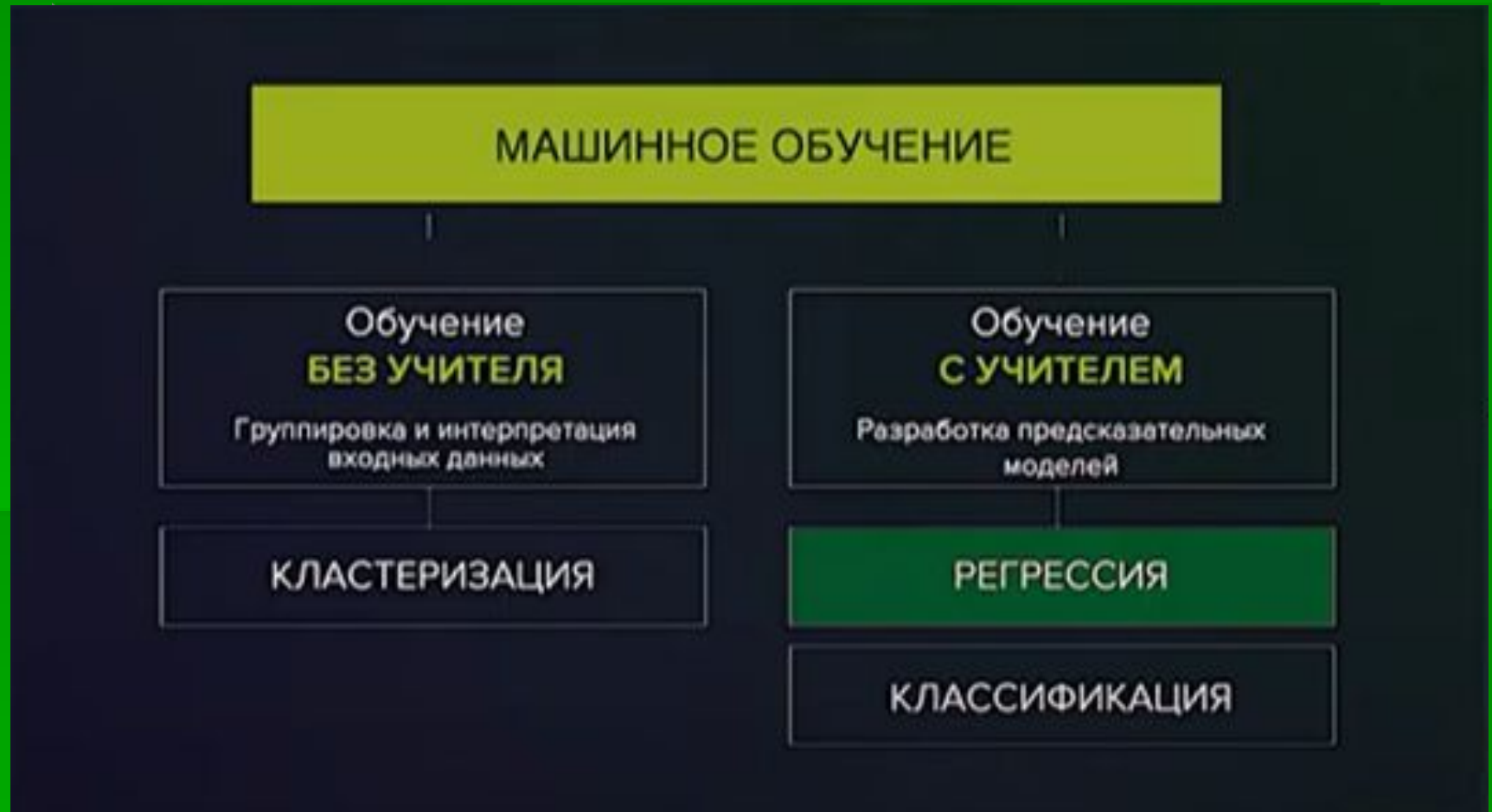
Описательная статистика и регрессия

Занятие 5

Подходы к МО



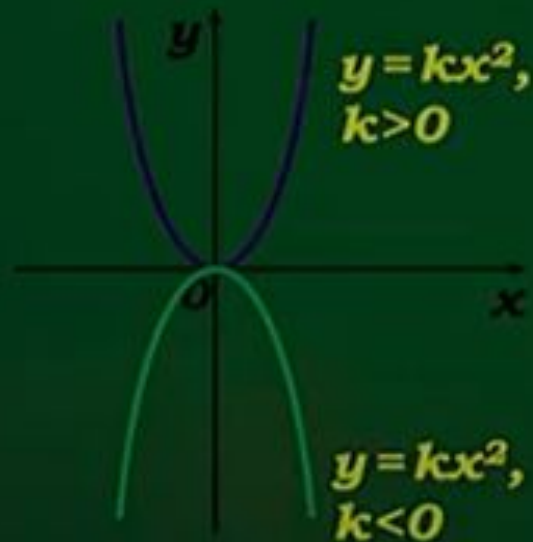
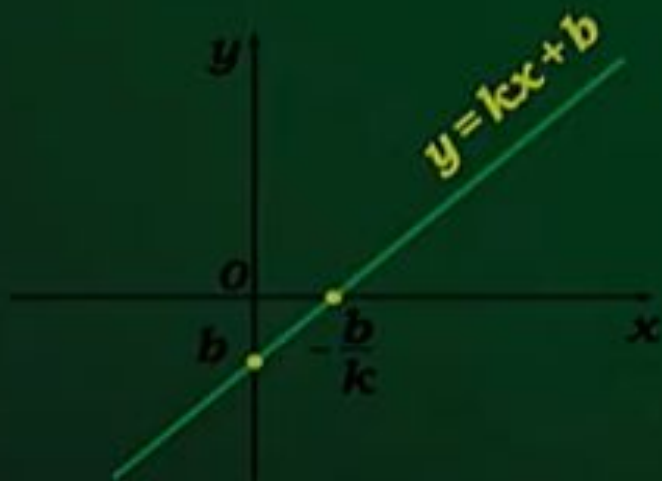
Регрессия



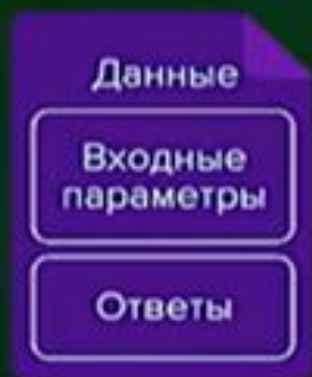
Математические основы машинного обучения

ФУНКЦИЯ

Способы задания функции



Модель обучения

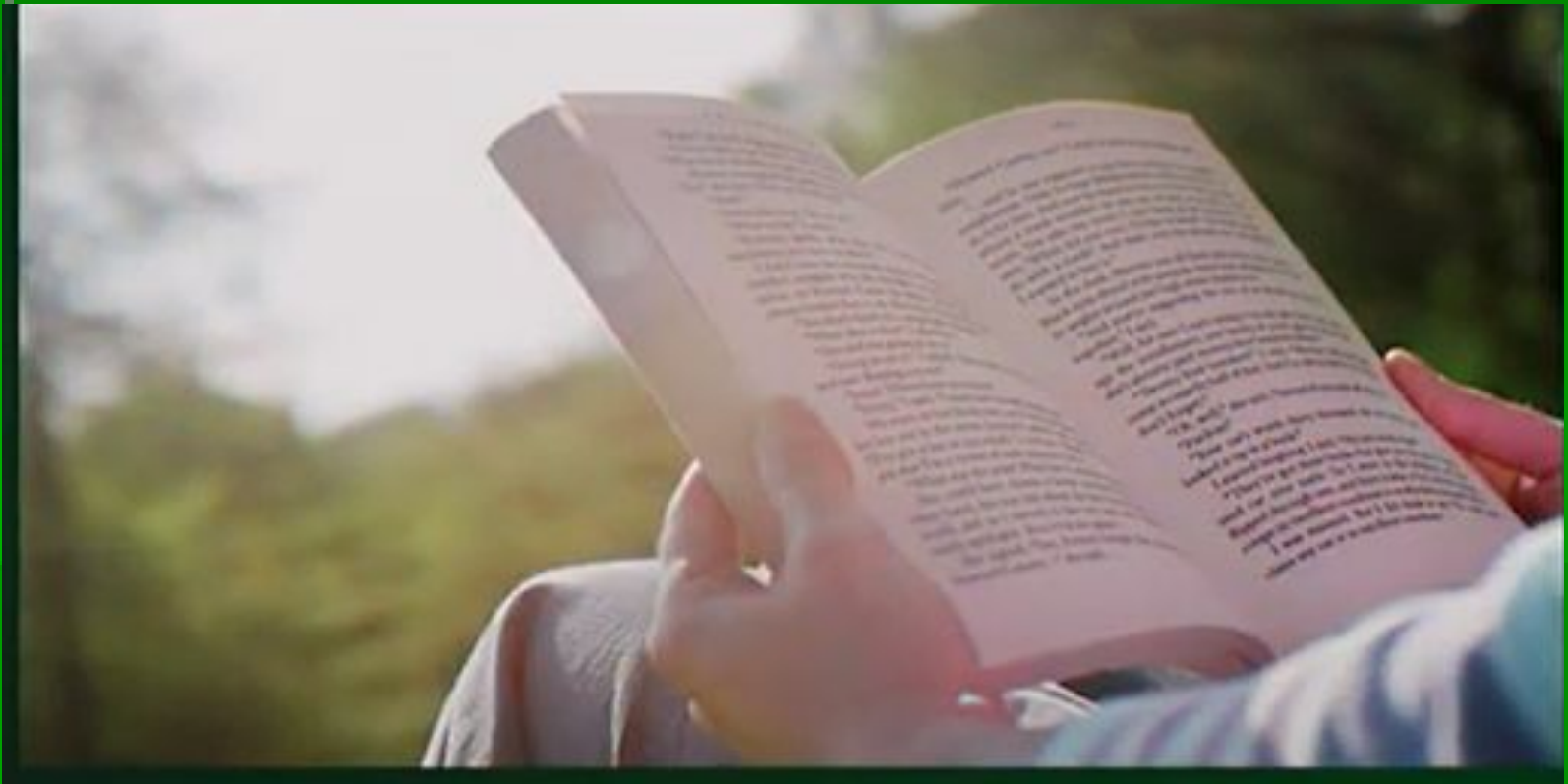


Алгоритм обучения
с учителем



Модель МО

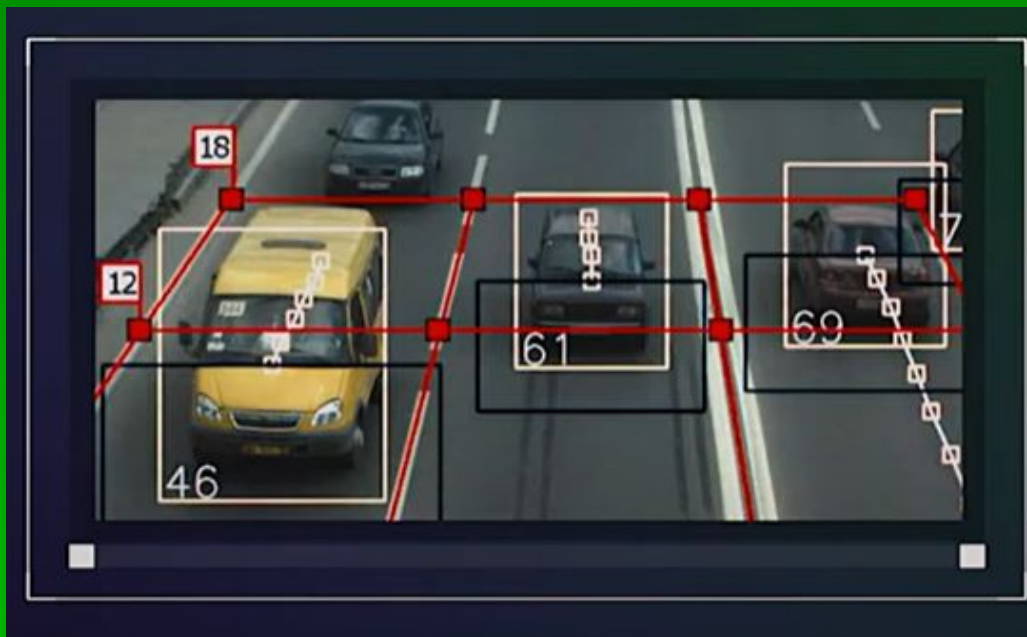
Откуда берутся данные



Что? Где? Когда?



Сенсоры



Запросы и соц.сети



Математическая статистика

Математическая статистика – наука о данных и методах их обработки.

Описательная статистика позволяет обобщить первичные результаты, полученные при сборе данных.

Помогает компактно описать данные, понять их структуру, провести классификацию.

Задача описательной статистики

С использованием математических методов, свести сотни значений выборки к нескольким итоговым показателям, которые дают представление о выборке.

Визуализация данных

Позволяет увидеть, как входные признаки X связаны с целевой переменной Y .

Служит ориентиром при построении модели.

Регрессия

Цель: прогнозирование непрерывных параметров какого-либо объекта.

Данный тип задач лежит в основе решения следующих проблем:

- Прогнозирование стоимости ценных бумаг
- Анализ спроса или объёма продаж
- Установление медицинских диагнозов
- Выявление любых зависимостей числа от времени
- Определение стоимости автомобиля по его пробегу
- Прогнозирование количества пробок на дорогах в зависимости от времени суток.

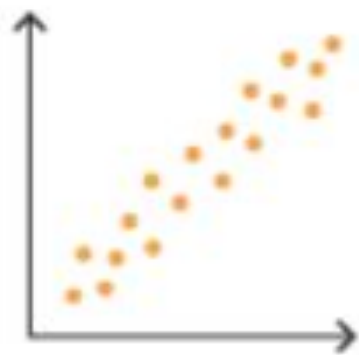
Решении регрессии

Чтобы решить задачу регрессии, требуется построить алгоритм, так называемый **регрессор**.

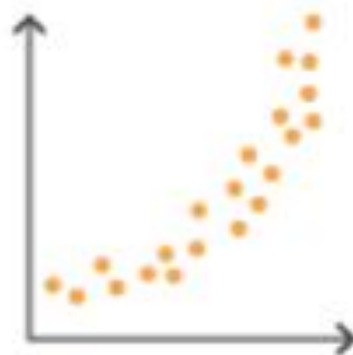
Этот алгоритм сможет спрогнозировать значение интересующей переменной. Это и будет результат работы машинного обучения — предсказание или, как обычно говорят, **прогноз**.

Обработав набор данных, алгоритм вернет число, максимально близкое к настоящему ответу.

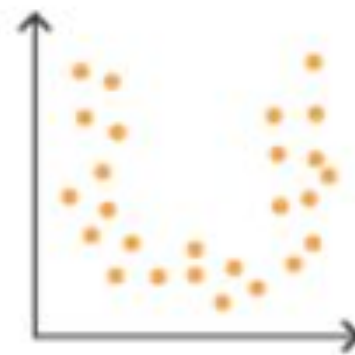
Характер зависимости



Линейная



Экспонента



Нелинейная

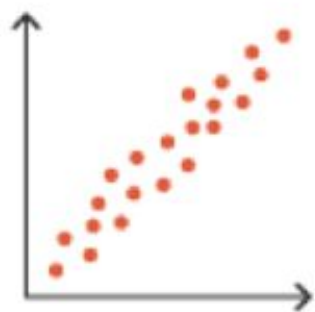
Коэффициент корреляции



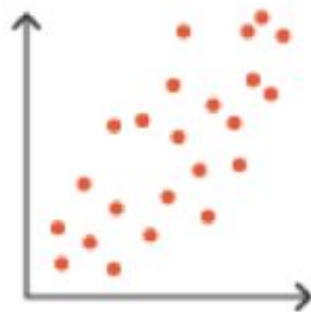
Положительная Отрицательная Отсутствие

Степень корреляции

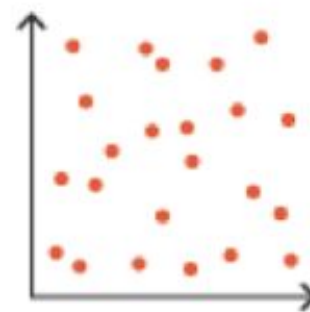
Степень корреляции



Сильная



Слабая



Отсутствие

Коэффициент детерминированности

Величина, полученная в результате построения трендов, R^2 называется **коэффициентом детерминированности**. Он определяет, насколько удачной является полученная регрессионная модель.

Коэффициент детерминированности всегда заключен в диапазоне от 0 до 1. Если он равен 1, то функция точно проходит через табличные значения, если 0, то выбранный вид регрессионной модели предельно неудачен.

Чем R^2 ближе к 1, тем удачнее регрессионная модель.