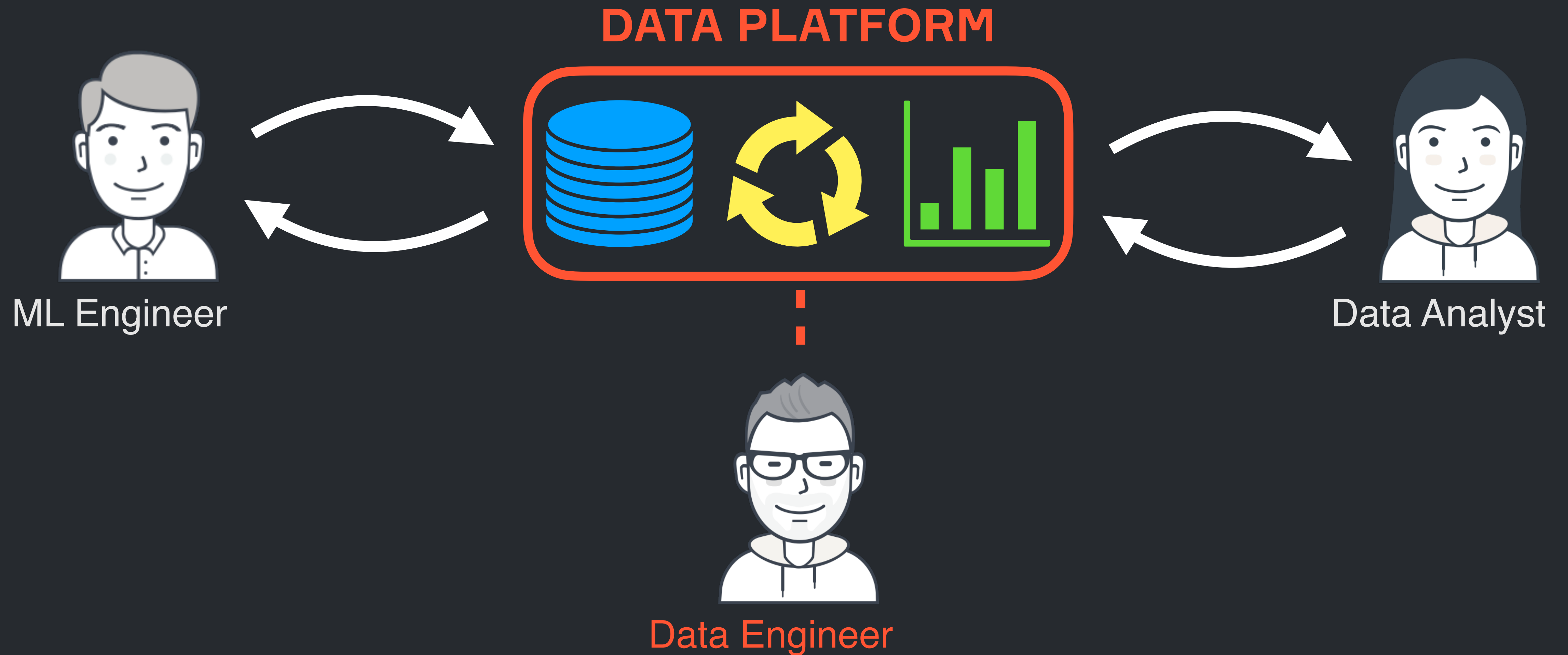


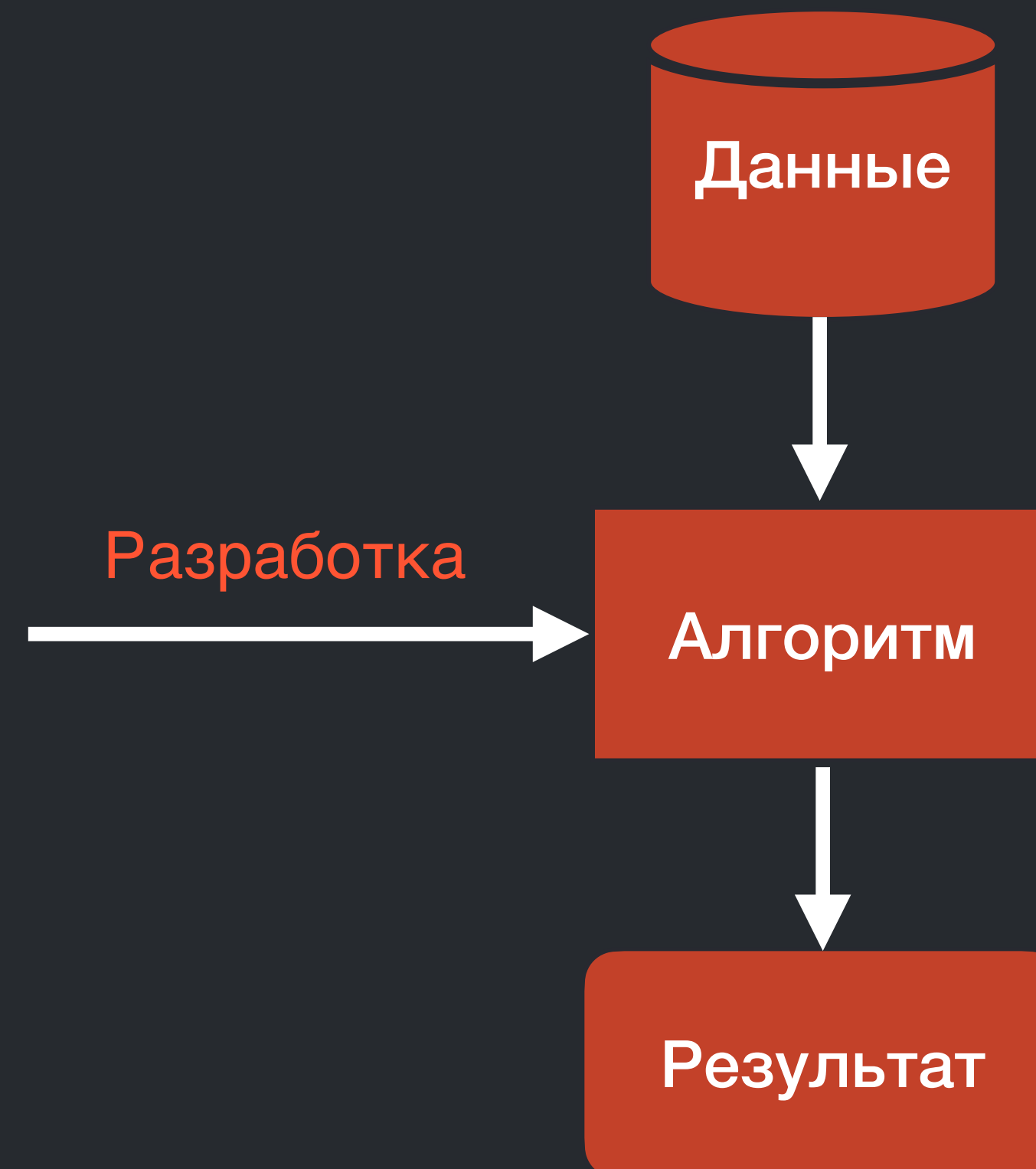
# ВВЕДЕНИЕ В МАШИННОЕ ОБУЧЕНИЕ

KARPOV.COURSES

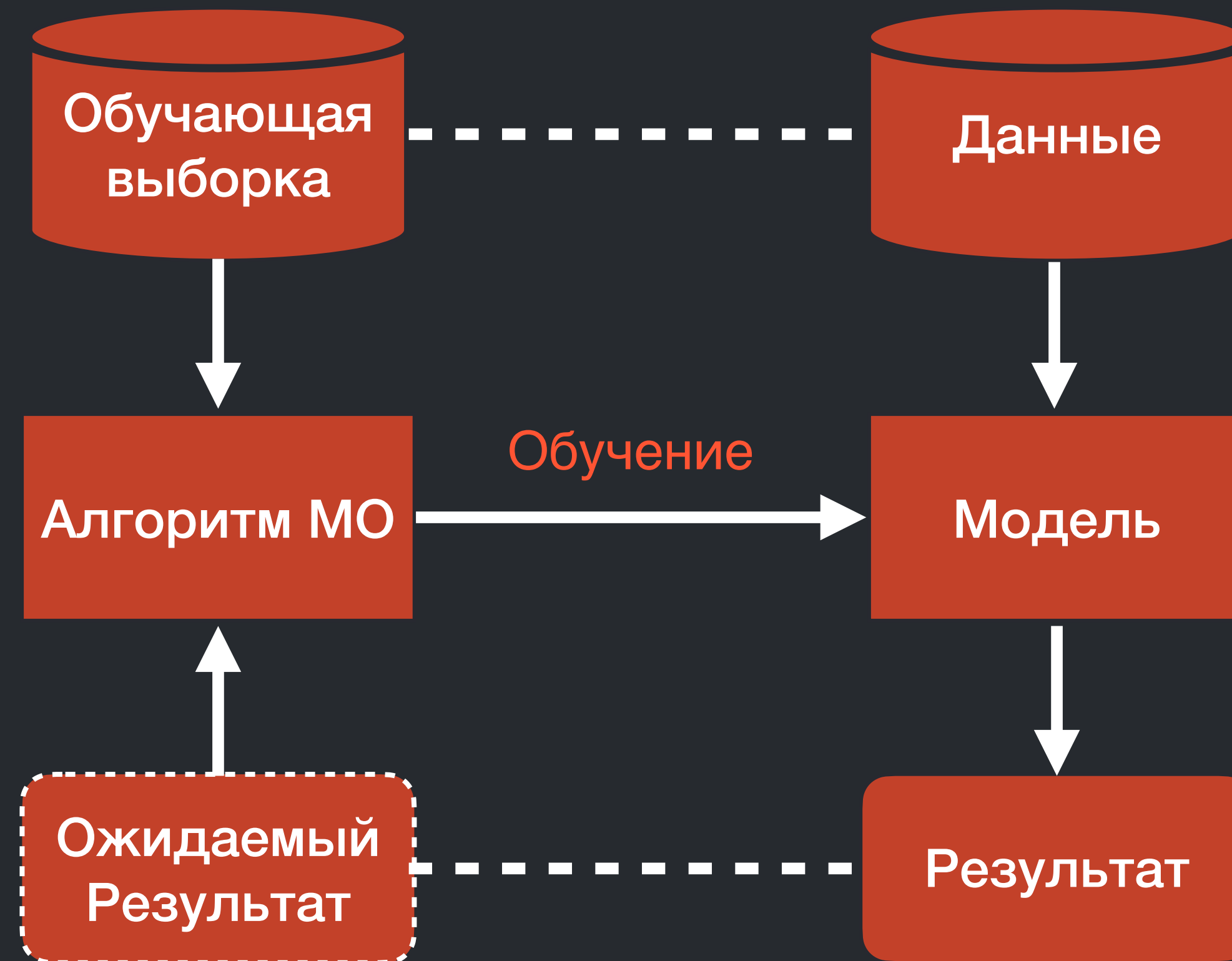
# BIG DATA & ML



# АЛГОРИТМ



# МАШИННОЕ ОБУЧЕНИЕ



# АЛГОРИТМ VS МАШИННОЕ ОБУЧЕНИЕ

	АЛГОРИТМ	МАШИННОЕ ОБУЧЕНИЕ
Результат	Точный	Вероятностный
Проверка	Тестирование	Оценка качества (метрики)
Сложность	Относительно просто реализовать	Требует затрат как для подготовки данных так и в обучении
Интерпретируемость	Прозрачен	Большинство методов трудно объяснимы
Применение	Широкое	Более широкое

# СОСТАВЛЯЮЩИЕ МАШИННОГО ОБУЧЕНИЯ

## ДАННЫЕ

Множество объектов (ситуаций) и их свойств для решения задачи.

## ПРИЗНАКИ

Свойства или характеристики используемые для обучения.

## АЛГОРИТМ

Метод применяемый для создания модели.

## МЕТРИКА

Функция для оценки качества модели.

# ОСНОВНЫЕ ПРОБЛЕМЫ В ML

## ДАННЫЕ

- Где взять?
- Как хранить?
- Чем больше тем лучше!

## РЕЗУЛЬТАТЫ ПО ДАННЫМ

- Ручная разметка
- Сбор результатов внутри системы
- Готовые датасеты

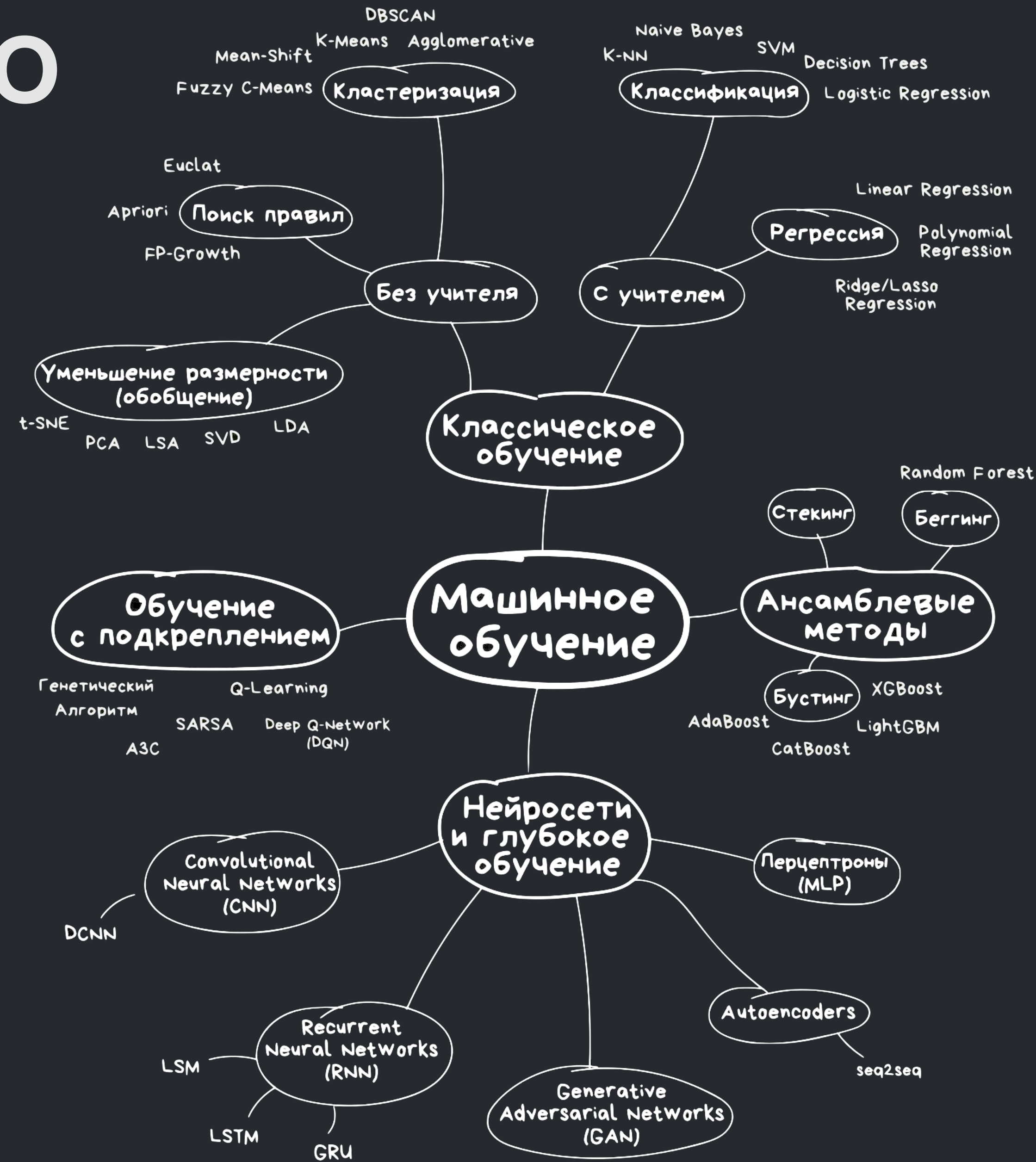
## ВЫЧИСЛИТЕЛЬНЫЕ МОЩНОСТИ

- Где взять CPU/GPU?

## ПРОИЗВОДИТЕЛЬНОСТЬ (NRT)

- Оптимизация
- Масштабирование

# МЕТОДЫ МО





# ОСНОВНЫЕ ВИДЫ МО

## КЛАССИЧЕСКОЕ ОБУЧЕНИЕ

Просты данные и признаки.

## ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ

Возможно взаимодействие со средой исполнения для обучения.

## АНСАМБЛИ

Комбинирование моделей для улучшения качества.

## ГЛУБОКОЕ ОБУЧЕНИЕ

Применение нейросетей любого типа.

# КЛАССИЧЕСКОЕ ОБУЧЕНИЕ

## ОБУЧЕНИЕ С УЧИТЕЛЕМ

- **КЛАССИФИКАЦИЯ**

Предсказать класс объекта

- **РЕГРЕССИЯ**

Предсказать значение

## ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ

- **КЛАСТЕРИЗАЦИЯ**

Группировка объектов по схожести

- **АССОЦИАЦИЯ**

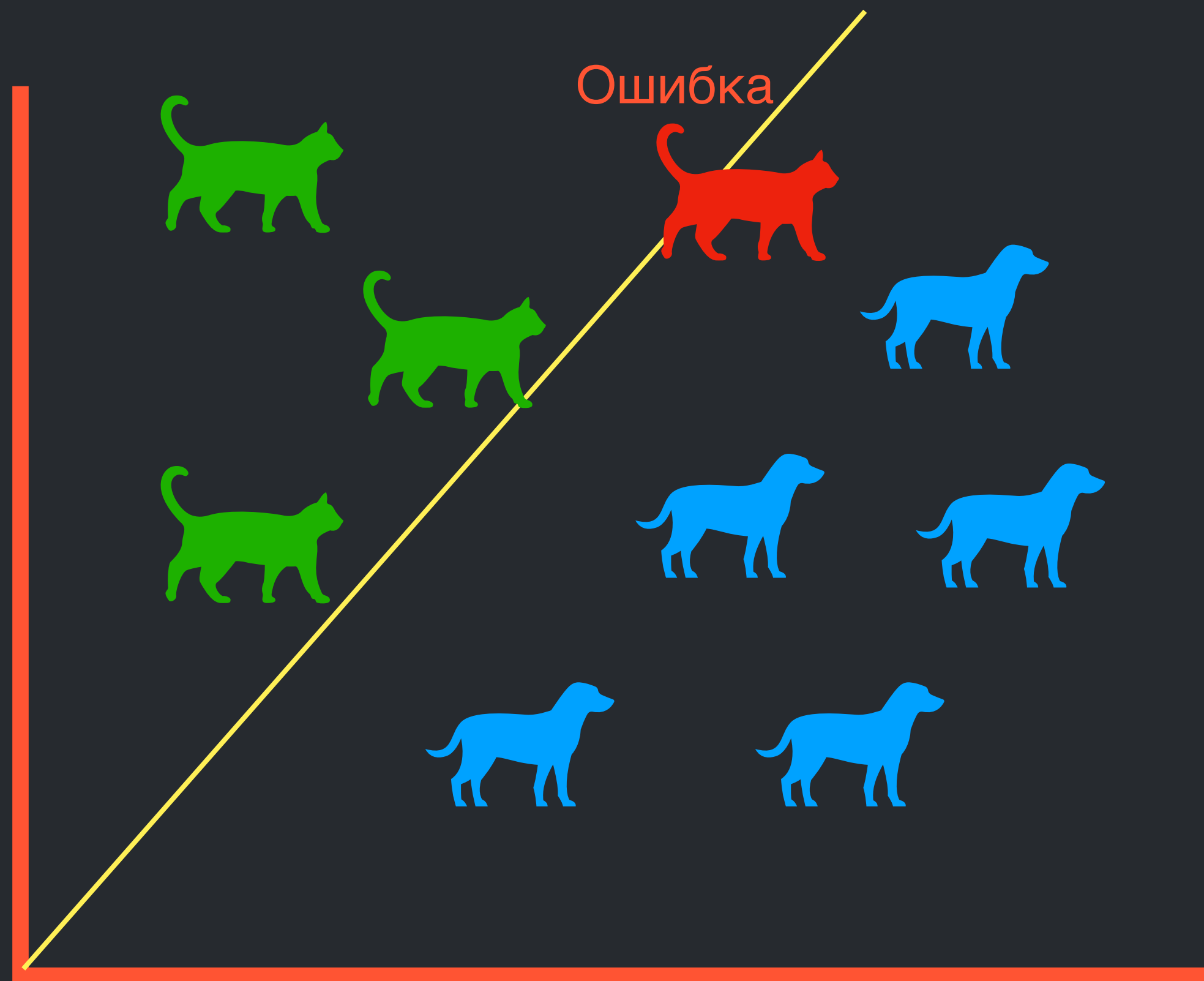
Выявление последовательностей

- **УМЕНЬШЕНИЕ РАЗМЕРНОСТИ**

Выявление зависимостей

# ЗАДАЧА КЛАССИФИКАЦИИ

Задача классификации - получение категориального ответа на основе набора признаков.



## Метрики

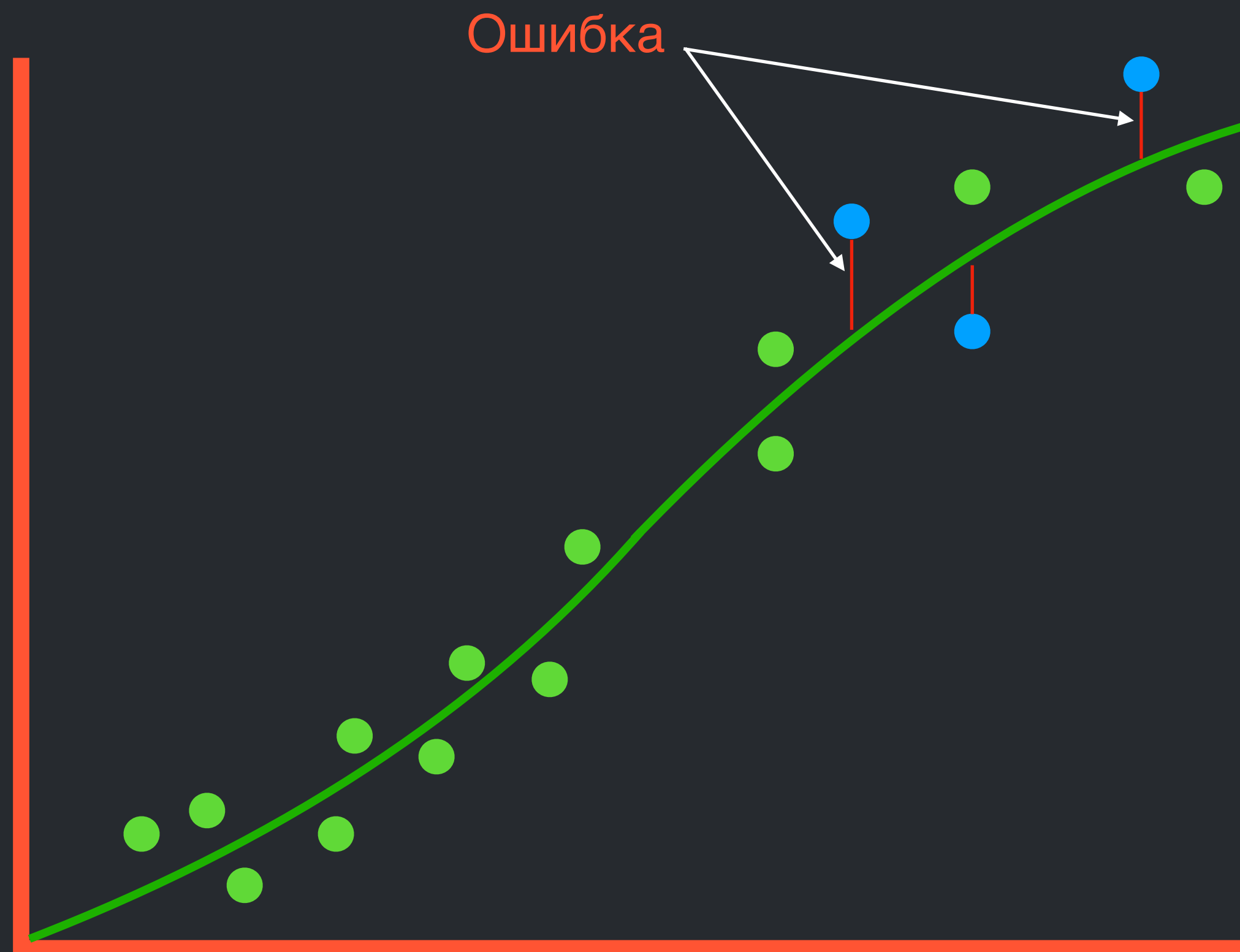
**Accuracy** - доля правильных ответов в наборе данных

**F-мера** - гармоническое среднее между точностью и полнотой

...

# ЗАДАЧА РЕГРЕССИИ

Задача регрессии - прогноз значения на основе выборки объектов с различными признаками.



## Метрики

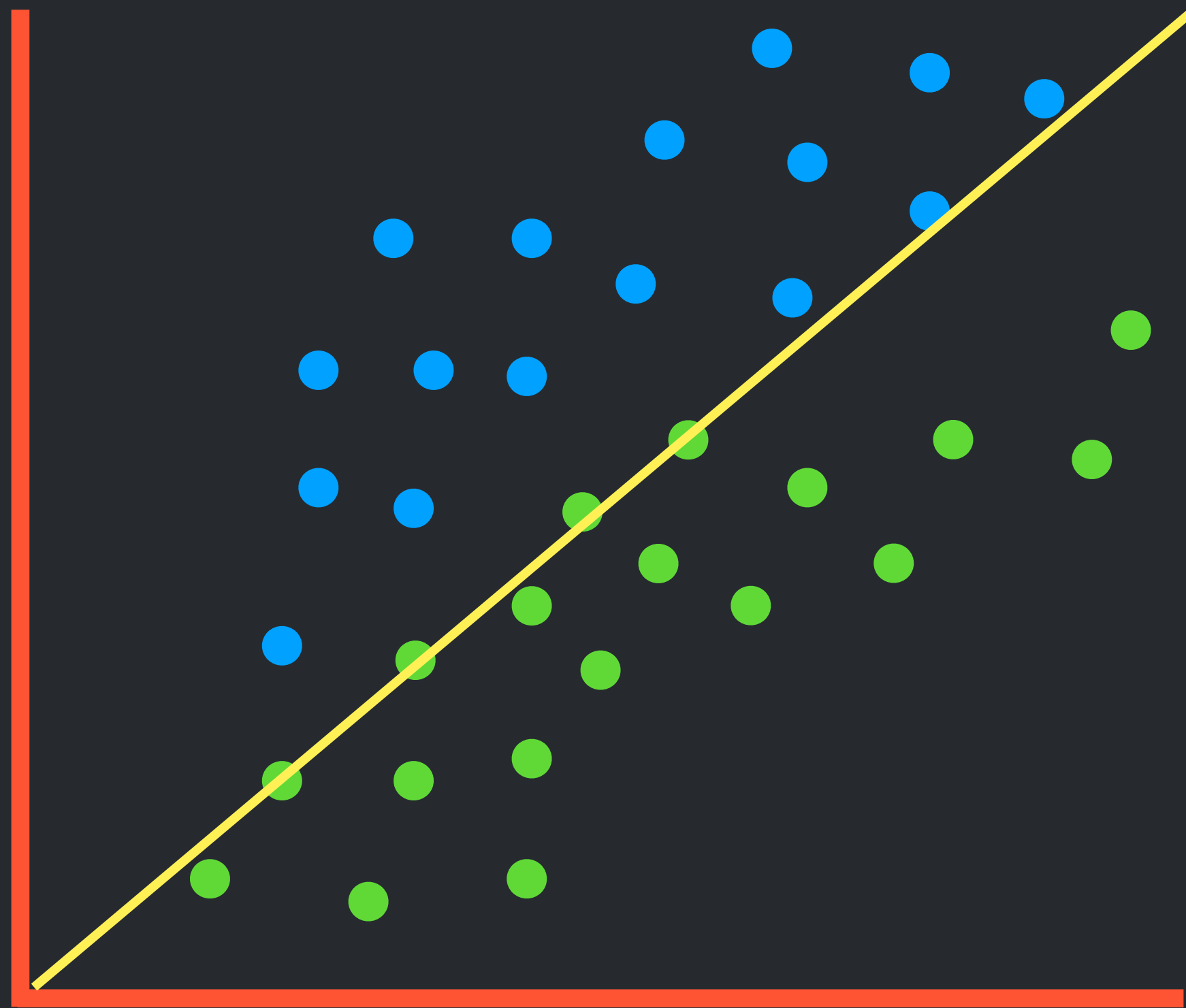
**MSE** - средний квадрат отклонения

**RMSE** - корень среднего квадрата отклонения

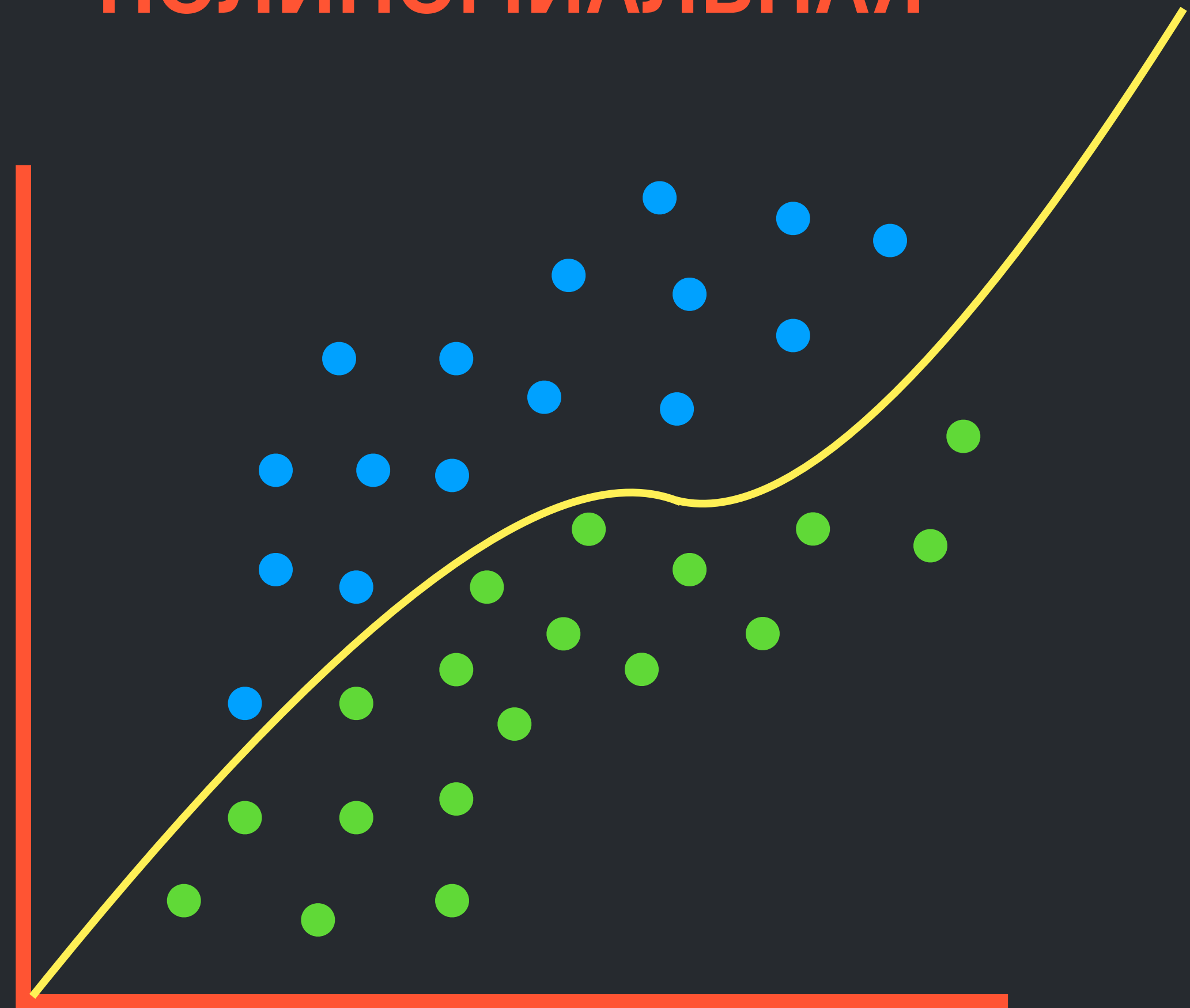
...

# МОДЕЛИ КЛАССИЧЕСКОГО МО

ЛИНЕЙНАЯ



ПОЛИНОМИАЛЬНАЯ

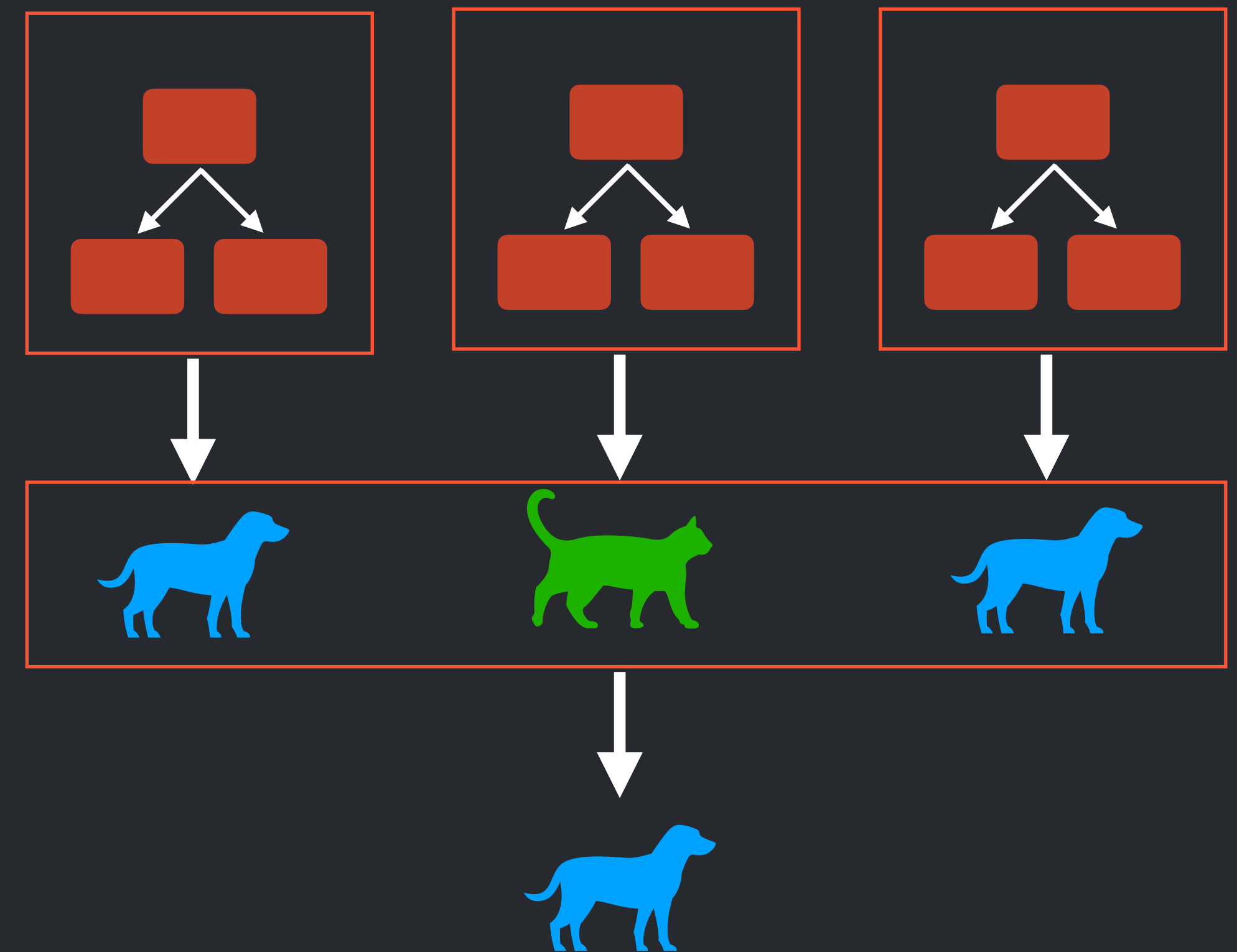


# МОДЕЛИ КЛАССИЧЕСКОГО МО

## ДЕРЕВО РЕШЕНИЙ



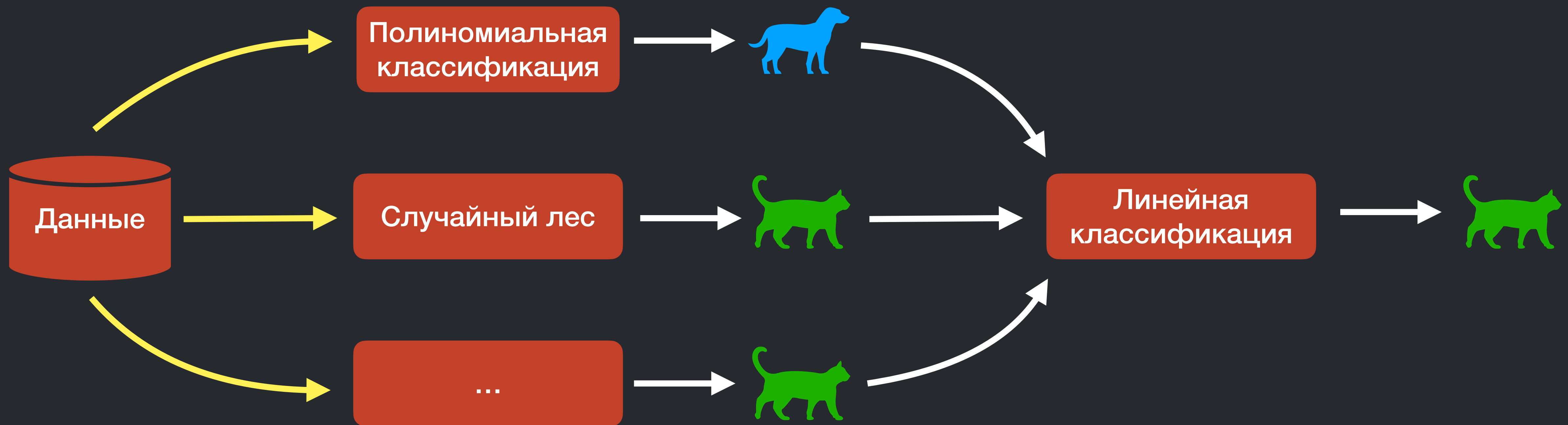
## СЛУЧАЙНЫЙ ЛЕС



# АНСАМБЛИ

## СТЕКИНГ

Обучение **набора разных алгоритмов** и передача их результатов на вход последнему, который и принимает итоговое решение.

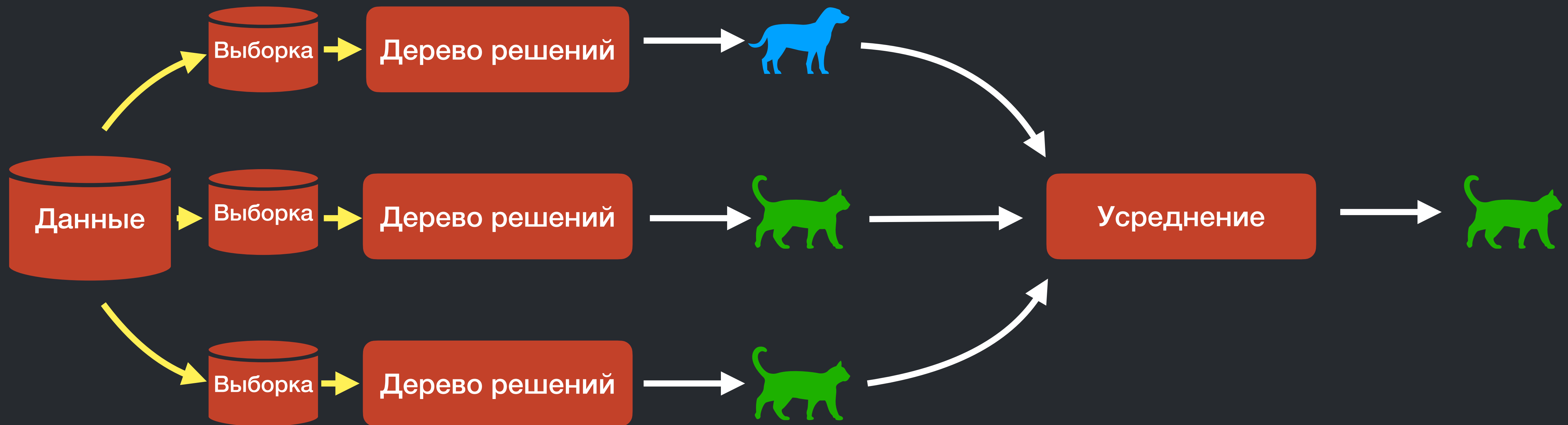


# АНСАМБЛИ

## БЭГГИНГ

Обучаем один алгоритм много раз на случайных выборках из исходных данных.

В конце усредняем ответы.

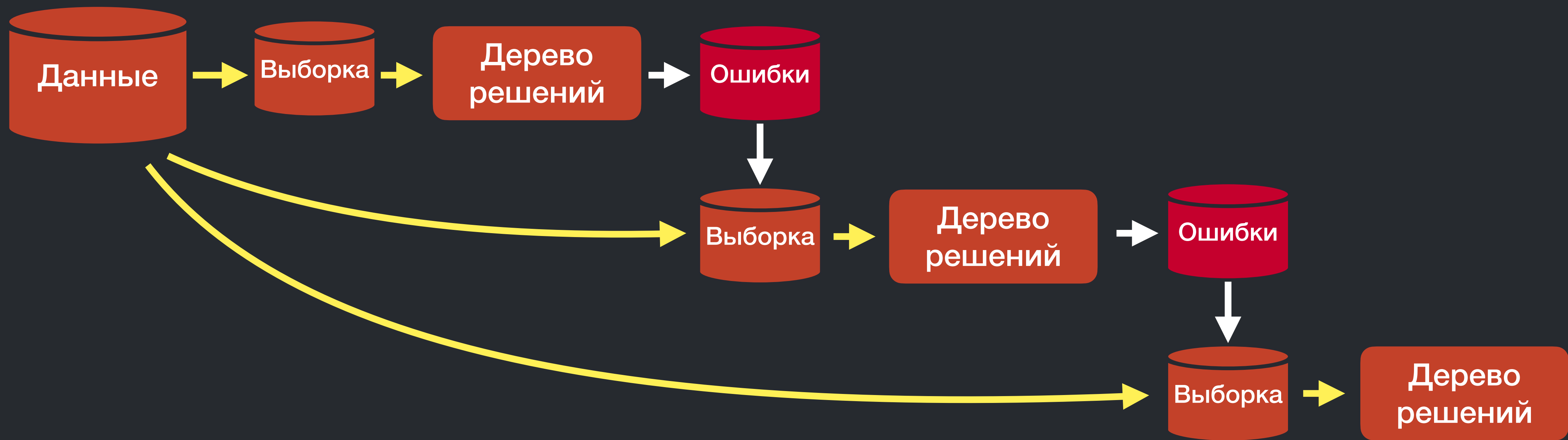




# АНСАМБЛИ

## БУСТИНГ

Обучаем **алгоритм последовательно**, каждый следующий уделяет особое внимание тем случаям, на которых ошибся предыдущий.



# ВИДЫ ПАРАМЕТРОВ МОДЕЛИ

**Параметры модели** — параметры, которые изменяются и оптимизируются в процессе обучения модели и итоговые значения этих параметров являются результатом обучения модели.

**Гиперпараметры модели** — параметры, значения которых задается до начала обучения модели и не изменяется в процессе обучения.  
При этом у модели может не быть гиперпараметров.

**Оптимизация гиперпараметров** — процесс поиска набора оптимальных гиперпараметров для алгоритма обучения.

# ОПТИМИЗАЦИЯ ГИПЕРПАРАМЕТРОВ

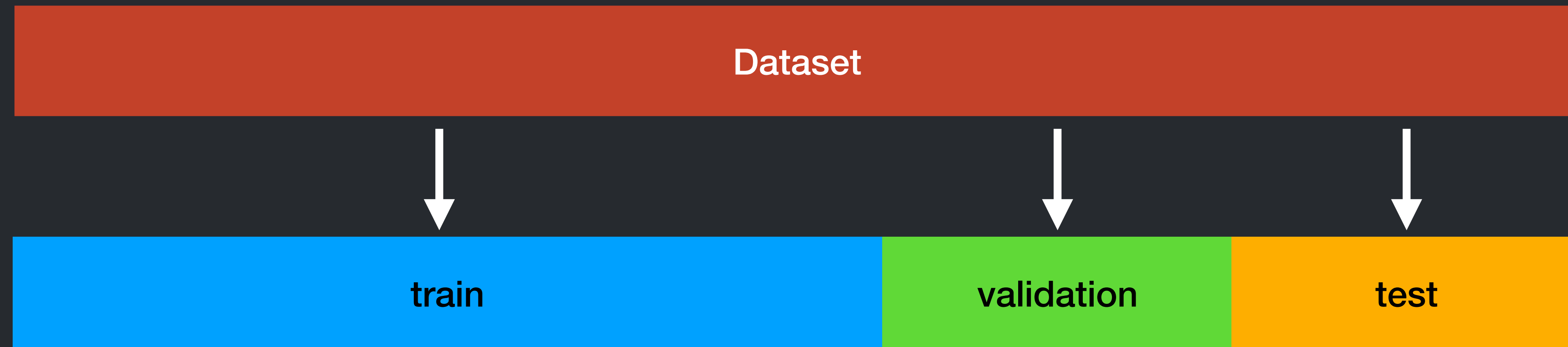
## ПРИМЕНЯЕМЫЕ АЛГОРИТМЫ ОПТИМИЗАЦИИ

**Grid Search** — алгоритм поиска гиперпараметров на основе перебора комбинаций гиперпараметров из заданного множества.

**Random Search** — алгоритм поиска гиперпараметров на основе перебора случайно выбранных комбинаций гиперпараметров из заданного диапазона.

**Gradient-based** — алгоритмы поиска гиперпараметров на основе градиентного спуска.

# ПРОЦЕСС ОБУЧЕНИЯ



**train** - применяется для обучения модели.

**validation** - применяется для оптимизации параметров модели.

**test** - применяется для итоговой оценки качества модели.

# ПРОЦЕСС ОБУЧЕНИЯ

