

# START ML

**KARPOV.COURSES**

# МАШИННОЕ ОБУЧЕНИЕ

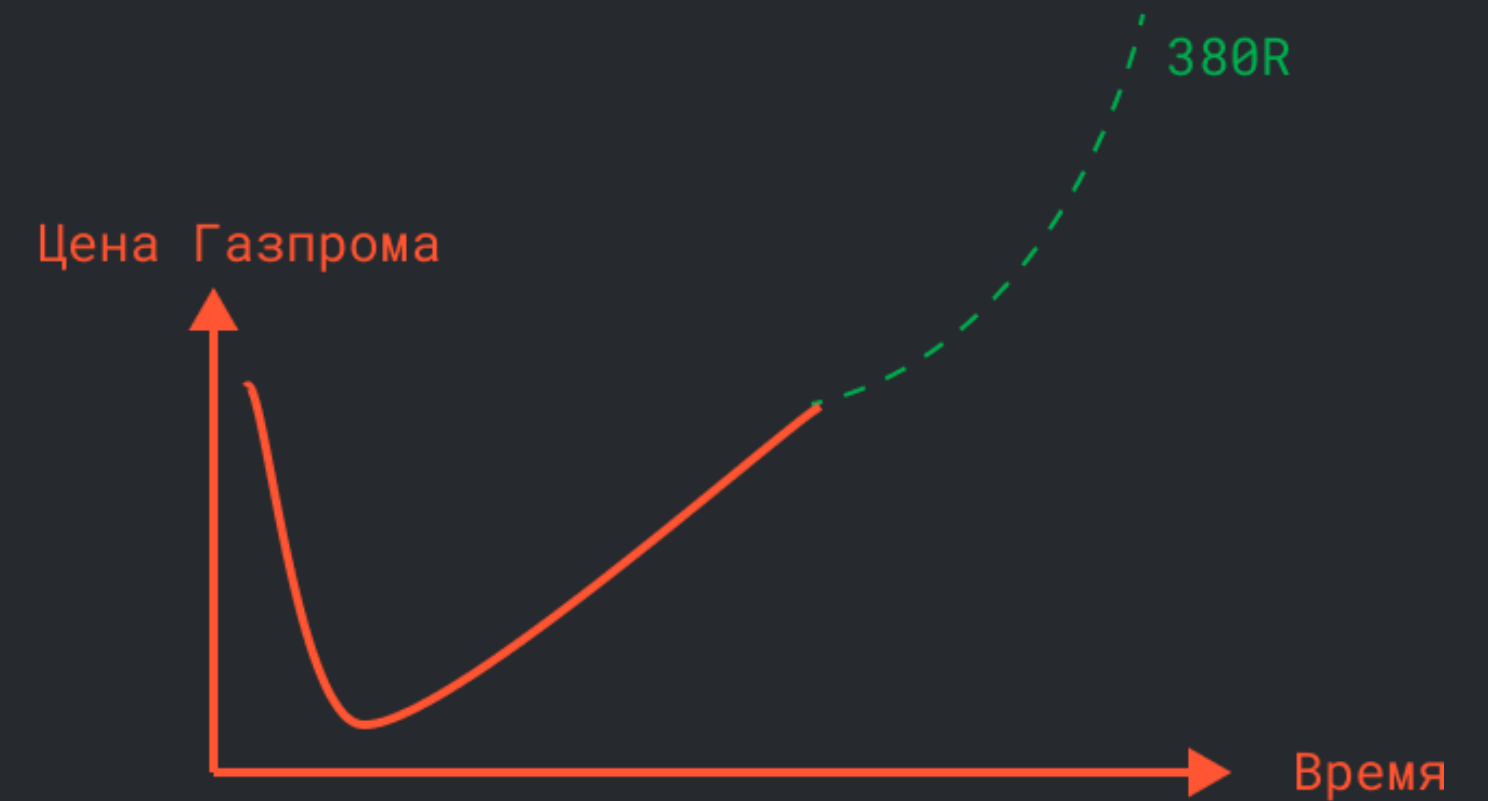
## ПОИСК ЗАКОНОМЕРНОСТЕЙ И РАЗРАБОТКА АЛГОРИТМА

Комплекс методов и инструментов, позволяющий компьютеру строить модели с сильной экспертизой на основе входящих из реального мира данных и накопленных человеком научных знаний.

# Сколько стоят акции через год?

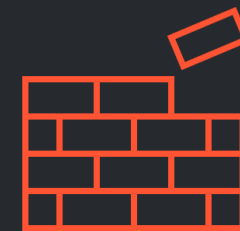
— Инвесторам — максимизация прибыли

— Управляющим менеджерам — стратегия



# Скоринг банковских клиентов

- Мы — финансовая организация
- Кредитуем юридических лиц
- Какие-то из них могут дефолтить
- Круто уметь сортировать клиентов по их кредитному рейтингу!



# Поиск аномалий в процессах

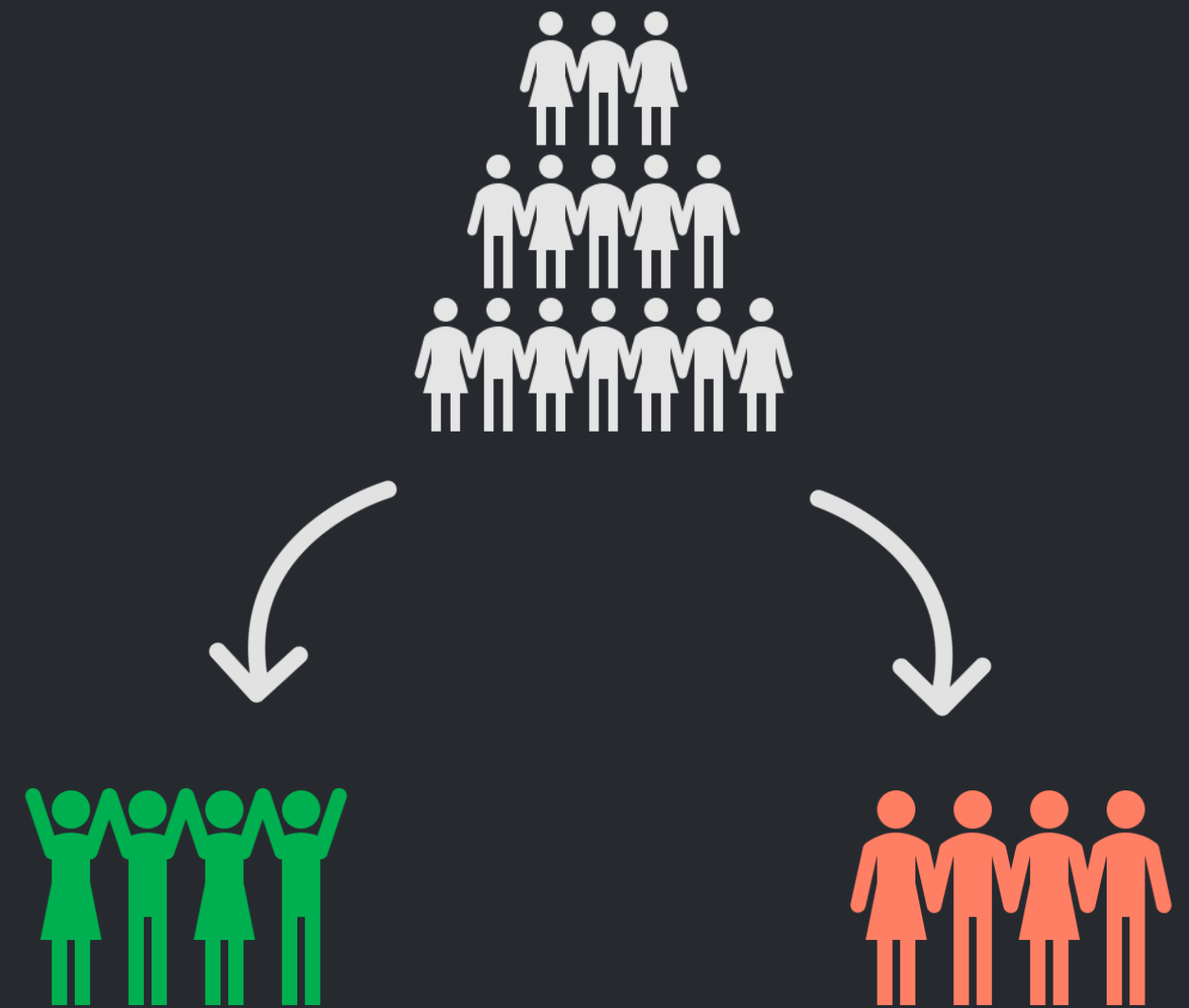
- Занимаемся опасным производством
- Есть ряд ежедневных процессов и процедур
- Хотим фиксировать и сообщать об аномалиях
- Экономим не только убытки, но и жизни



Все в норме?

# Существуют ли явные группы клиентов, и если да, то какие?

- Допустим, мы маркетинговое агентство
- Хотим понять, кому давать рекламу, кому нет
- То есть выделить клиентов с откликом и без
- Минимизируем расходы маркетинга



# ПОЧЕМУ ЧЕЛОВЕКА НЕДОСТАТОЧНО?

## ВОПРОСЫ И ПРОБЛЕМЫ

- Сколько стоят акции через год?
- Скоринг банковских клиентов
- Поиск аномалий в процессах
- Существуют ли явные группы клиентов, и если да, то какие?

## НЕДОСТАТКИ ЭКСПЕРТИЗЫ ЧЕЛОВЕКА

- Нужна скорость ответа
- Экспертов может быть мало
- Сложные задачи
- Большое количество данных

# РЕЗЮМЕ

- Машинное обучение – не магия!
- Умеет решать целый комплекс различных задач
- Делает работу человека быстрой и эффективной
- Справляется со сложными и нагруженными вопросами



# ОСНОВНЫЕ ТИПЫ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

## ОБУЧЕНИЕ С УЧИТЕЛЕМ

- Регрессия
- Классификация



# Обучение с учителем

## РЕГРЕССИЯ

- Имеем прецеденты
- Отвечаем неограниченным числом
- Пример: какая капитализация?

100M



125M



?? M



## КЛАССИФИКАЦИЯ

- Имеем прецеденты
- Ответы из ограниченного множества
- Пример: какой фрукт на картинке?

Клубник

а



Вишня



??



# ОСНОВНЫЕ ТИПЫ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

## ОБУЧЕНИЕ С УЧИТЕЛЕМ

- Регрессия
- Классификация



## ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ

- Кластеризация
- Рекомендательные  
системы



# Обучение с учителем

## КЛАСТЕРИЗАЦИЯ

- Есть только объекты
- Хотим сформировать группы
- Сколько их? Какие они будут по сути?



## РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ

- Пользователи, предметы, оценки
- Никогда наверняка не знаем лучшую рекомендацию



# РЕЗЮМЕ

- Выделяют обучение с учителем и без него
- Главное отличие – наличие или отсутствие явных ответов для объектов
- Но есть и другие направления МО
- Их затрагивать в курсе явно не будем

# КОМПОНЕНТЫ КЛАССИЧЕСКОЙ ML-ЗАДАЧИ

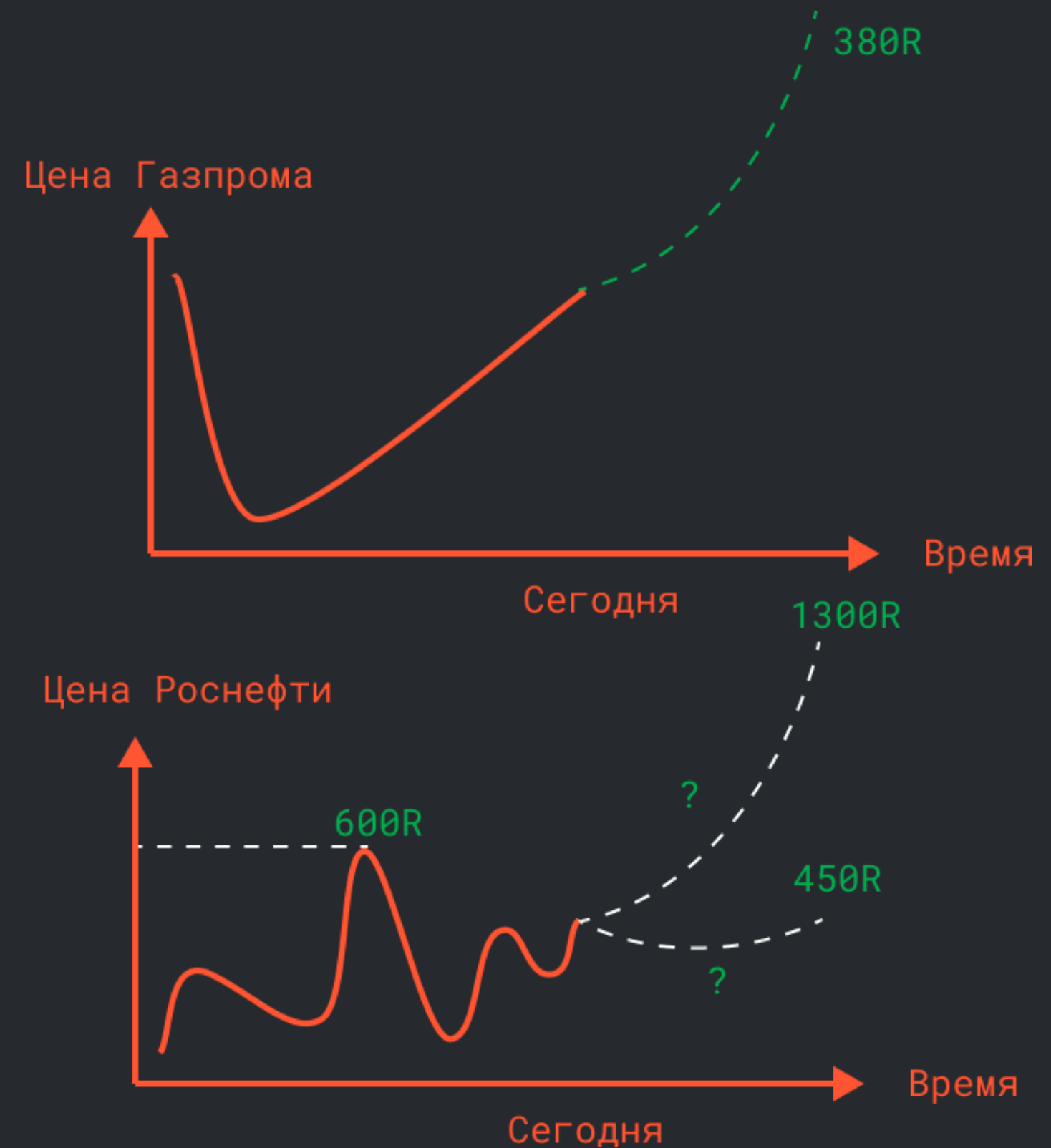
- Выборка: объект и признаки (object and features)
- Ответ (target)
- Функция потерь и функционал качества
- Метрики качества
- Алгоритм / семейство моделей
- Оценка модели

# ПРИМЕР

## ВЫБОРКА И ОБЪЕКТЫ

- То, на чем мы обучаем модели
- В случае обучения с учителем:
- Имеем объект и ответ к нему
- Объекты описываются набором признаков

$$x_i = (d^1, d^2, \dots, d^k)$$
$$X = \{(x_i, y_i)\}$$



# ПРИМЕР

## ПРИЗНАКИ

— Вещественные

$$d^j \in R$$

— Бинарные

$$d^j \in \{0, 1\}$$

— Категориальные

$$d^j \in D$$

— Порядковые

$$d^j \in D \subset Z$$

— Множественные (в т.ч.  
картинки, текст)

$$d^j \in 2^D$$

— Сколько активов на счете компании?

— Дефолтила ли компания последние  
5 лет?

1 – да; 0 – нет

— Отрасль фирмы

  
Добыча, IT, Красота...

— Какой у компании кредитный  
рейтинг?

1 – BBB, 2 – A,

3 – AA...

— Кто крупнейшие конкуренты?

{Газпром, Лукойл} или {X5  
Retail}





$d^1$ Сколько активов на счете компании?	$d^2$ Дефолтила ли компания последние 5 лет?	$d^3$ Отрасль фирмы?	$d^4$ Какой у компании кредитный рейтинг?	$d^5$ Кто крупнейшие конкуренты?	$y$ Таргетное значение
500M \$	Нет	Металлургия	АА	ИП Кузьминов Алексей	?
100M \$	Да	Ремонт	BBB	∅	?
250M\$	Нет	Телеком	AAA	Билайн, Мегафон	?

# ПРИМЕР

**TARGET:**  $Y = \{y_i\}$

— Вещественный

$$y_i \in R$$

— Бинарный

$$y_i \in \{0, 1\}$$

— Множественный

$$y_i \in \{0, 1, 2, \dots, d\}$$

— Задачи без явного таргета

$$y_i \in \emptyset$$

— Какая будет цена акции через год?



— Вырастет ли цена акций?

1 – да; 0 – нет

— Какая стратегия будет лучшей?

1 – hold, 2 – buy, 3 – sell

— Какие акции добавить в имеющийся портфель?

{Microsoft, Apple, Pfizer}



$d^1$ Сколько активов на счете компании?	$d^2$ Дефолтила ли компания последние 5 лет?	$d^3$ Отрасль фирмы?	$d^4$ Какой у компании кредитный рейтинг?	$d^5$ Кто крупнейшие конкуренты?	$y$ Сколько будет стоить одна акция через год?
500M \$	Нет	Металлургия	АА	ИП Кузьминов Алексей	300\$
100M \$	Да	Ремонт	BBB	∅	100\$
250M\$	Нет	Телеком	AAA	Билайн, Мегафон	Не листингуется

# РЕЗЮМЕ

- Узнали, что такое МО, и на какие вопросы оно умеет отвечать
- Узнали о главных компонентах: объекты, признаки, ответы
- На примере разобрали, какие типы этих компонент встречаются
- Пора к практике!

**СПАСИБО**

**ТАБАКАЕВ НИКИТА**