



Кейс: промежуточный ML



С чего все началось

- К вам пришёл владелец некоторой компании
- Они запускают рекламные кампании и хотели бы более «умно» распределить имеющийся бюджет по телеканалам
- Вас просят разработать такую систему
- Важный нюанс: в данной задаче не существует фиксированной рекламы (реклама в конкретное время)

Первые шаги





Первые шаги

- Задавайте произвольные вопросы
- Поймите, что от вас хотят
- Узнайте все нюансы



Первые шаги

- Как оценивается качество рекламной кампании
- Какие доступны данные для обучения
- На основании чего делается оптимизация
- Есть ли дополнительные ограничения

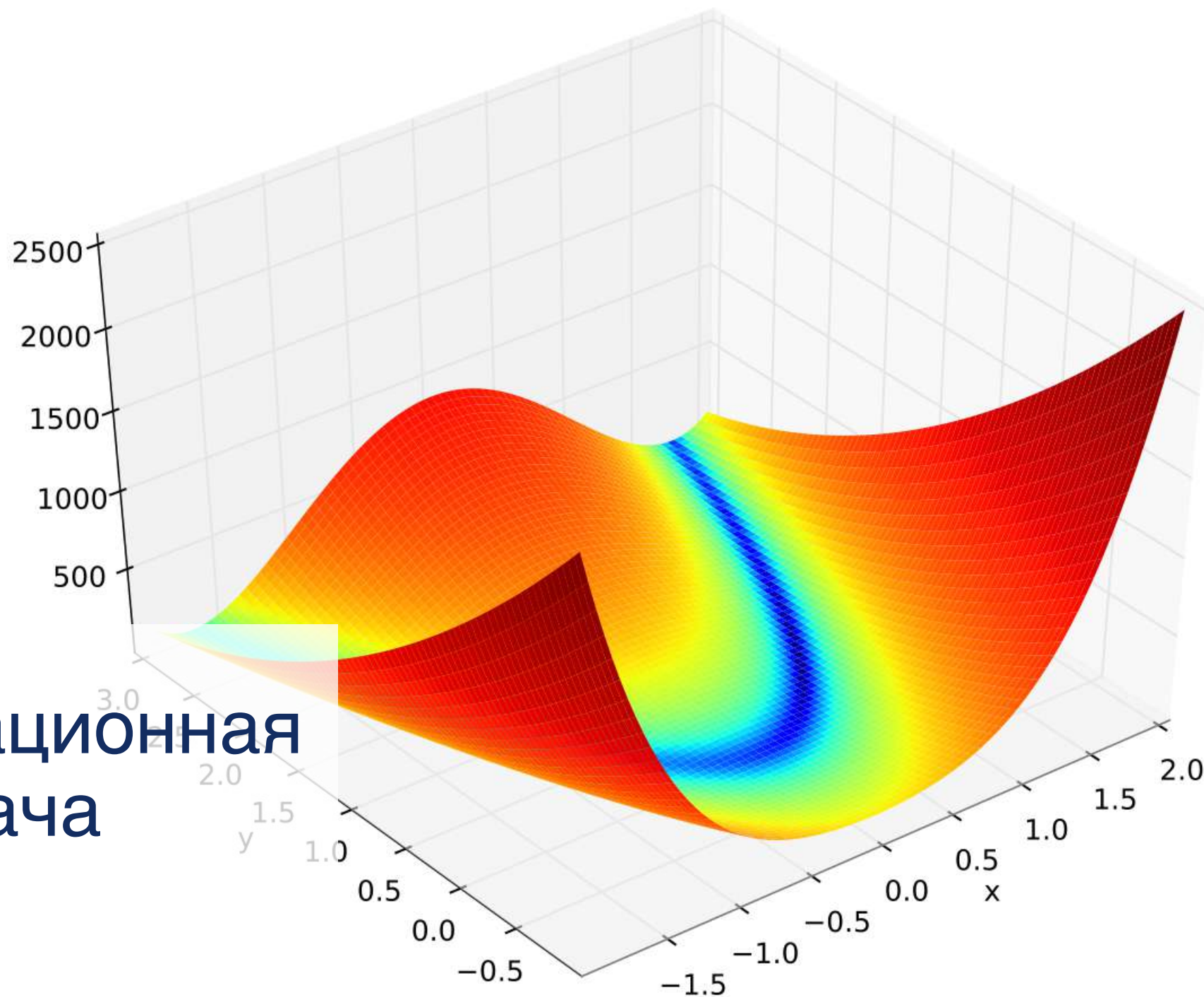


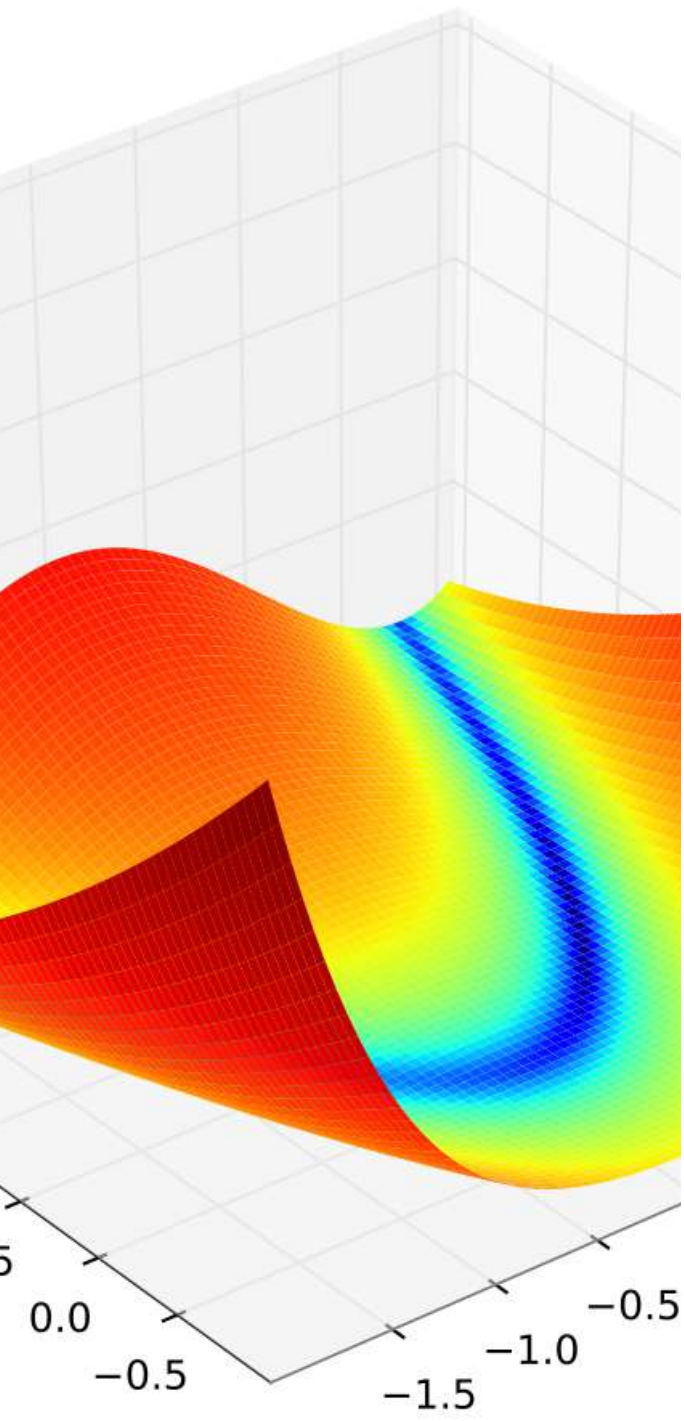
Доступные данные

- Для каждой рекламной кампании есть
 - Количество денег, потраченное на каждый канал
 - Целевая аудитория рекламы
 - Желаемый охват
- Далее ситуация разделяется:
 - У вас есть охваты по всем ЦА внутри каждого каналаЛибо
 - У вас есть суммарные охваты по всем

Как это можно использовать?

Оптимизационная задача

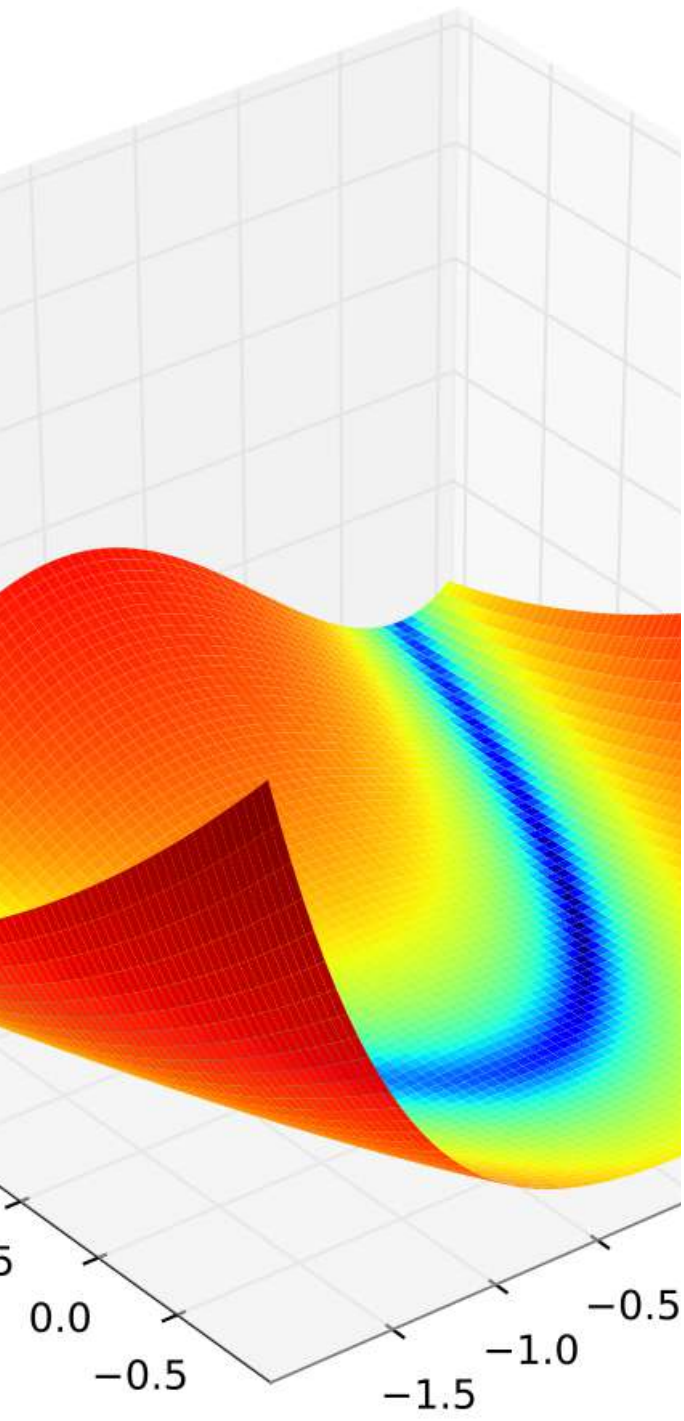




Оптимизационная задача

- Допустим, у вас есть модель $a(x)$, которая предсказывает охват в требуемой ЦА по вектору бюджета x , как её использовать?

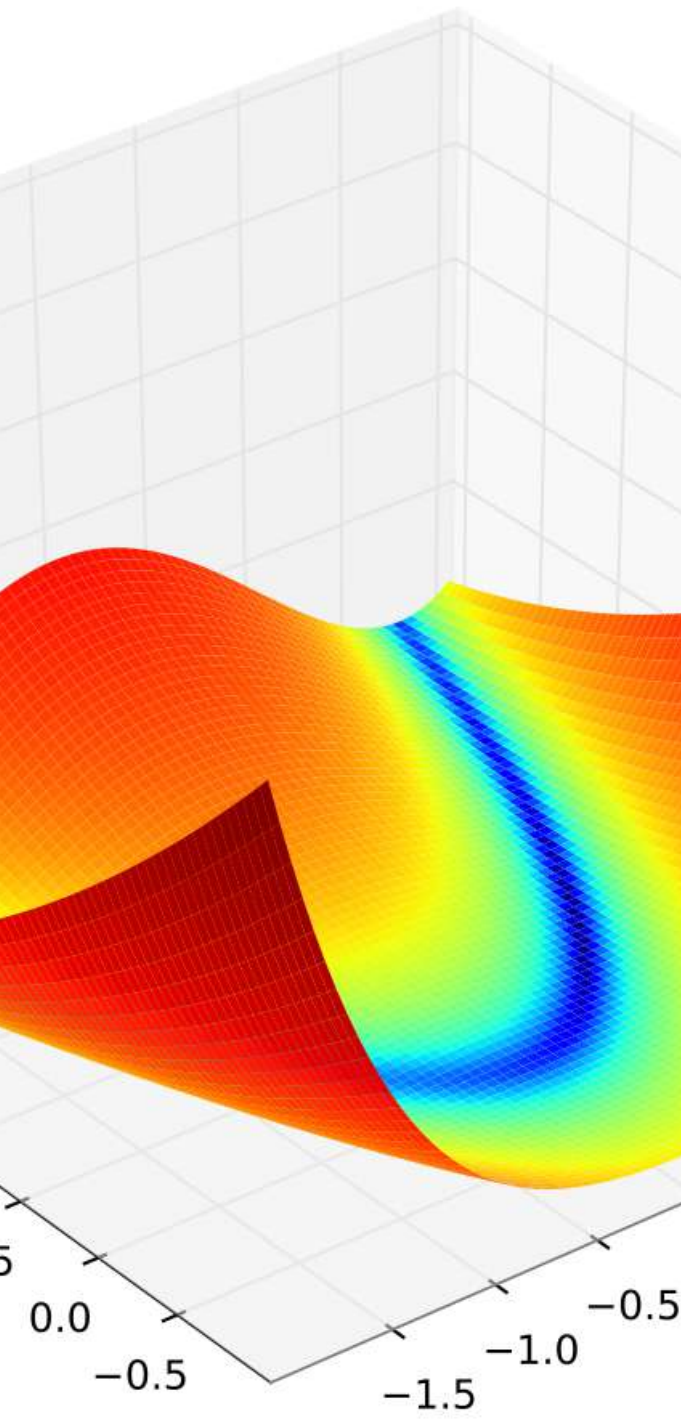
$$\left\{ \max_x a(x) \right.$$



Оптимизационная задача

- Допустим, у вас есть модель $a(x)$, которая предсказывает охват в требуемой ЦА по вектору бюджета x , как её использовать?

$$\left\{ \min_x \sum_{i=1}^n x_i \right.$$

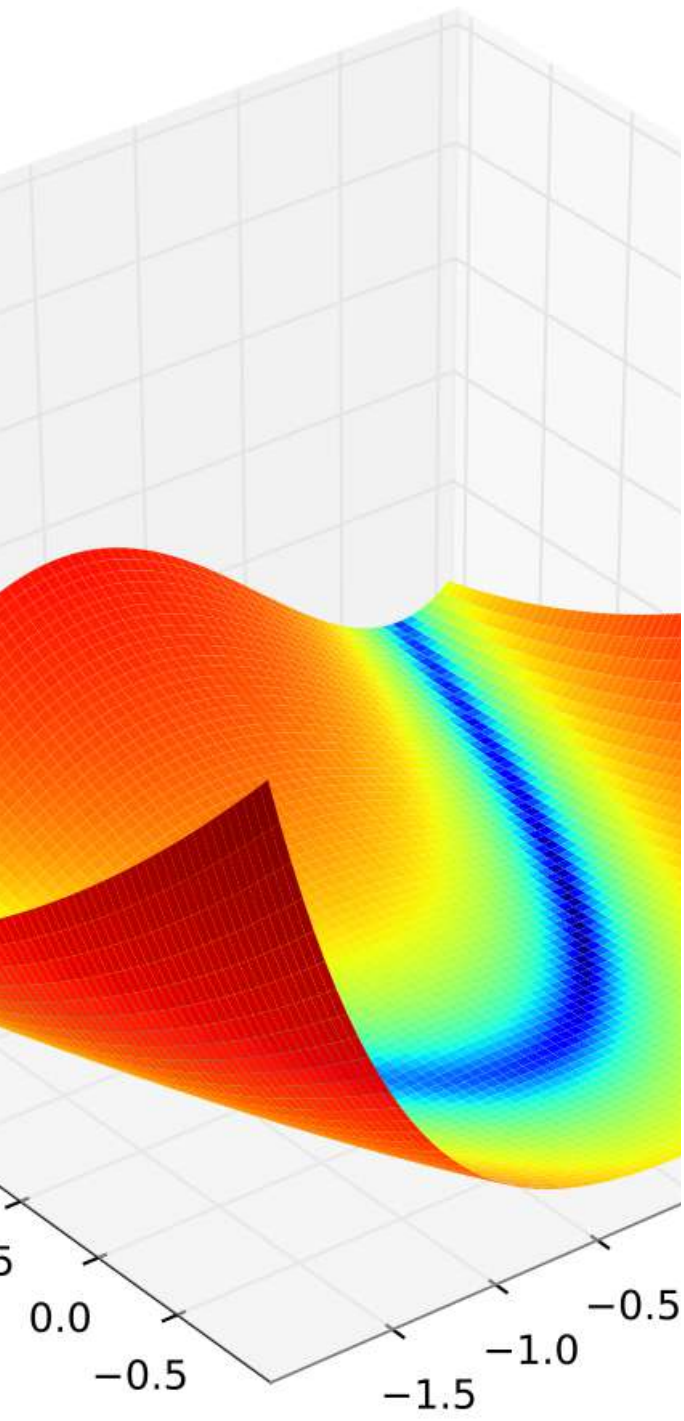


Оптимизационная задача

- Допустим, у вас есть модель $a(x)$, которая предсказывает охват в требуемой ЦА по вектору бюджета x , как её использовать?

$$\begin{cases} \min_x \sum_{i=1}^n x_i \\ a(x) \geq t \end{cases}$$

- Минимизация расходов при условии выполнения цели по охвату

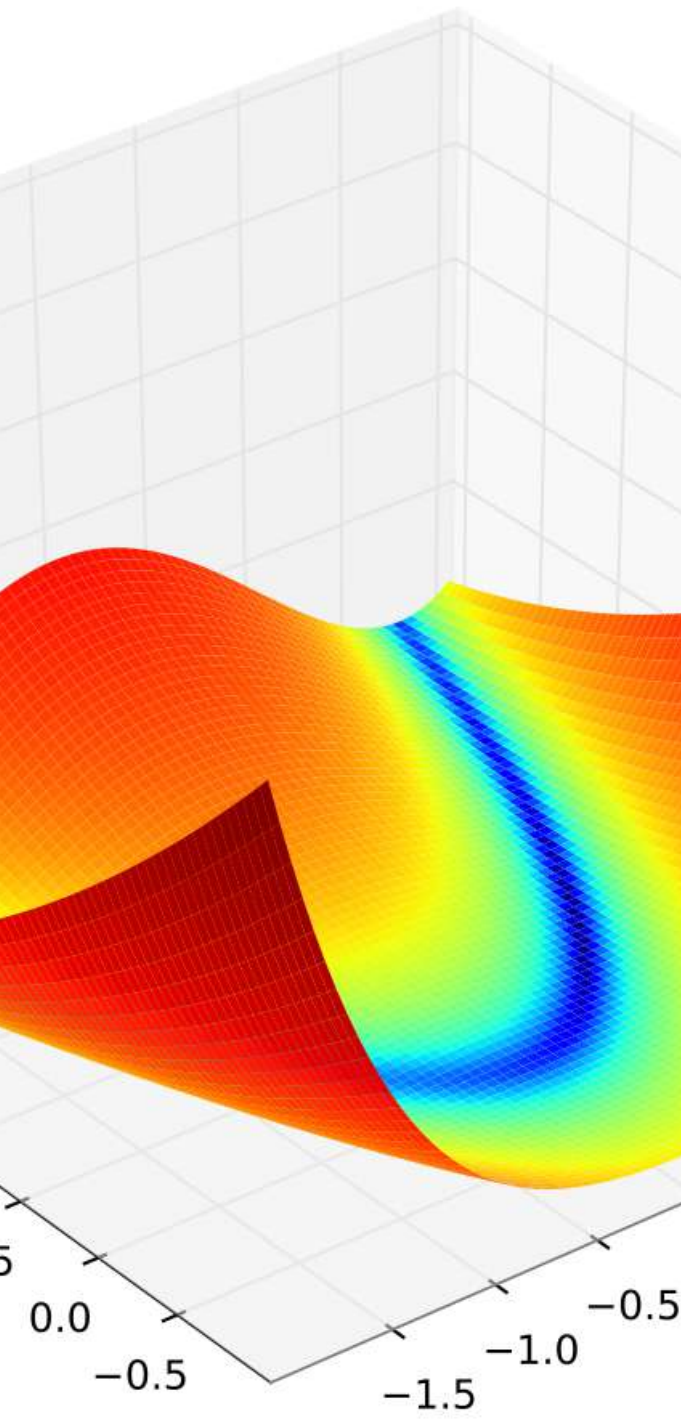


Оптимизационная задача

- Допустим, у вас есть модель $a(x)$, которая предсказывает охват в требуемой ЦА по вектору бюджета x , как её использовать?

$$\begin{cases} \min_x \sum_{i=1}^n x_i \\ a(x) \geq t \end{cases}$$

- Минимизация расходов при условии выполнения цели по охвату
- Но есть ограничение на рассматриваемые вектора x

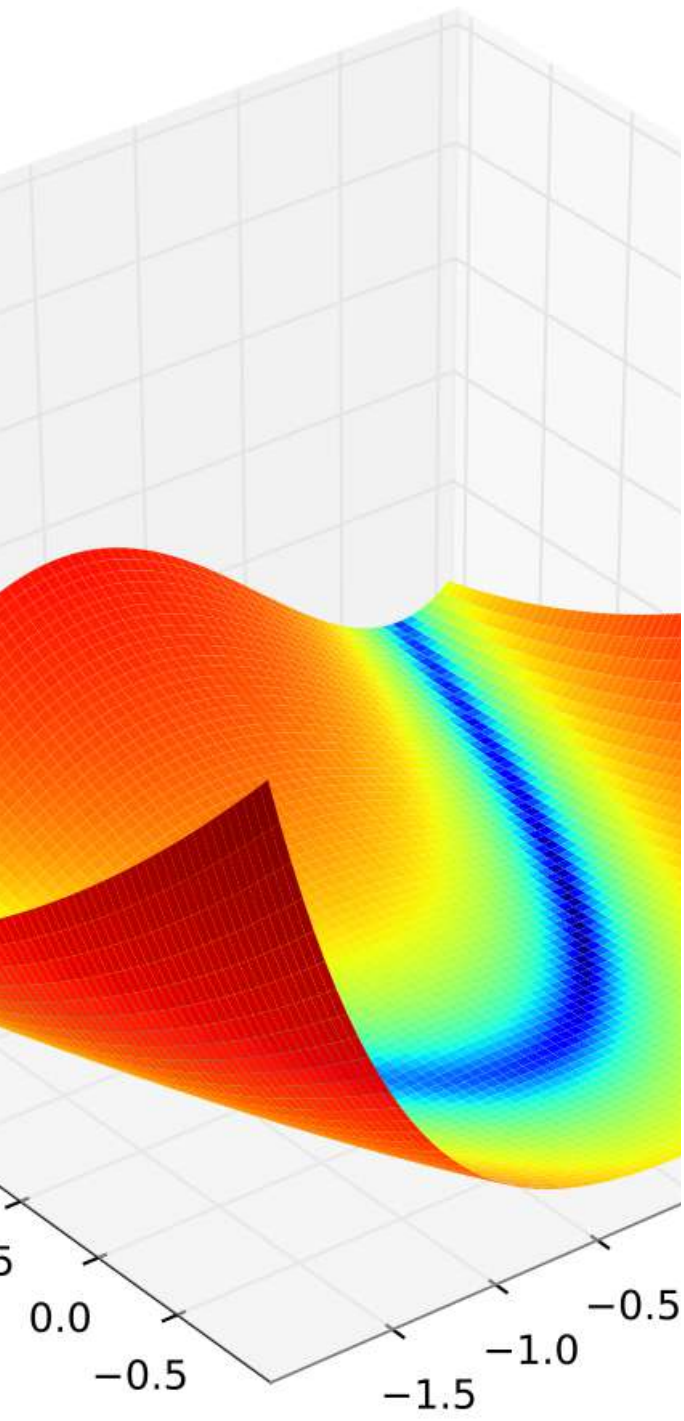


Оптимизационная задача

- Допустим, у вас есть модель $a(x)$, которая предсказывает охват в требуемой ЦА по вектору бюджета x , как её использовать?

$$\begin{cases} \min_x \sum_{i=1}^n x_i \\ a(x) \geq t \\ l_i \leq x_i \leq r_i \end{cases}$$

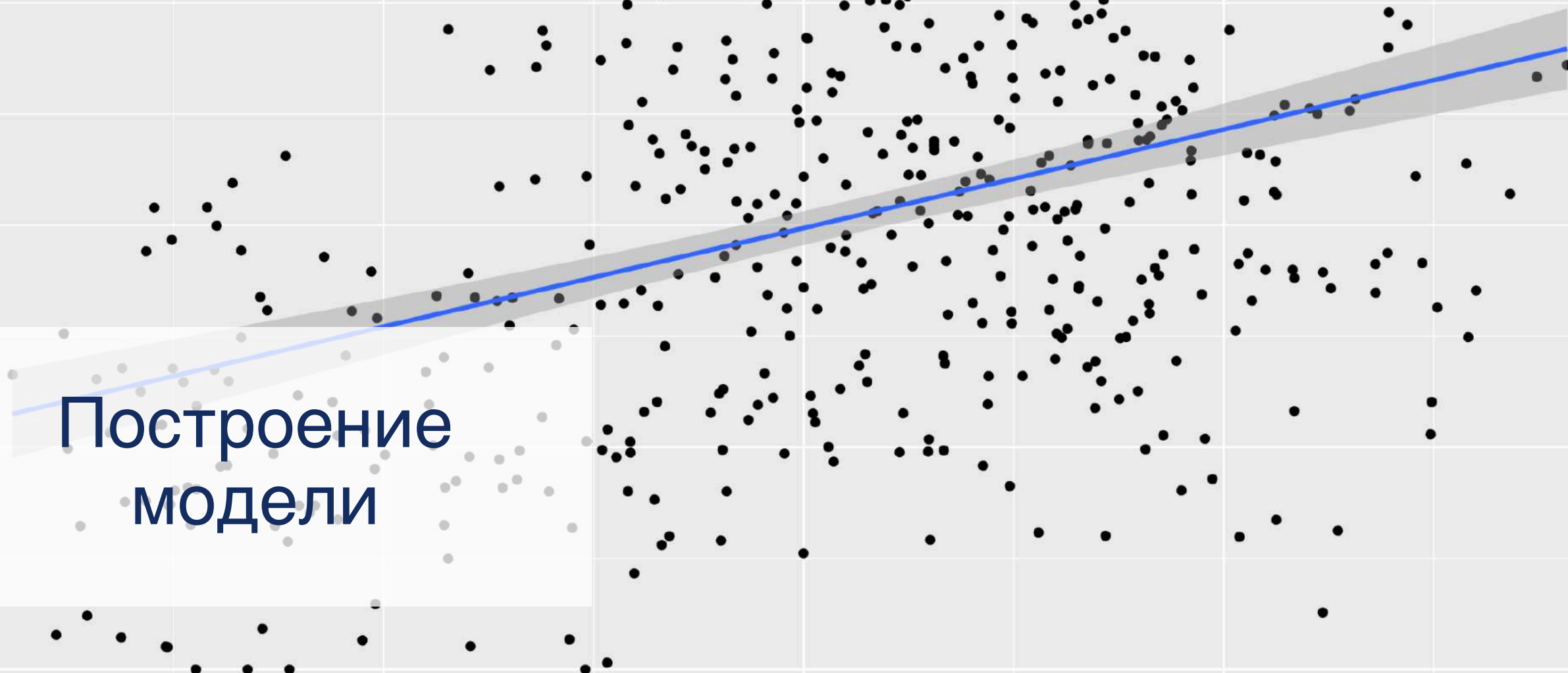
- Минимизация расходов при условии выполнения цели по охвату в окрестности какой-то точки



Оптимизационная задача

$$\begin{cases} \min_x \sum_{i=1}^n x_i \\ a(x) \geq t \\ l_i \leq x_i \leq r_i \end{cases}$$

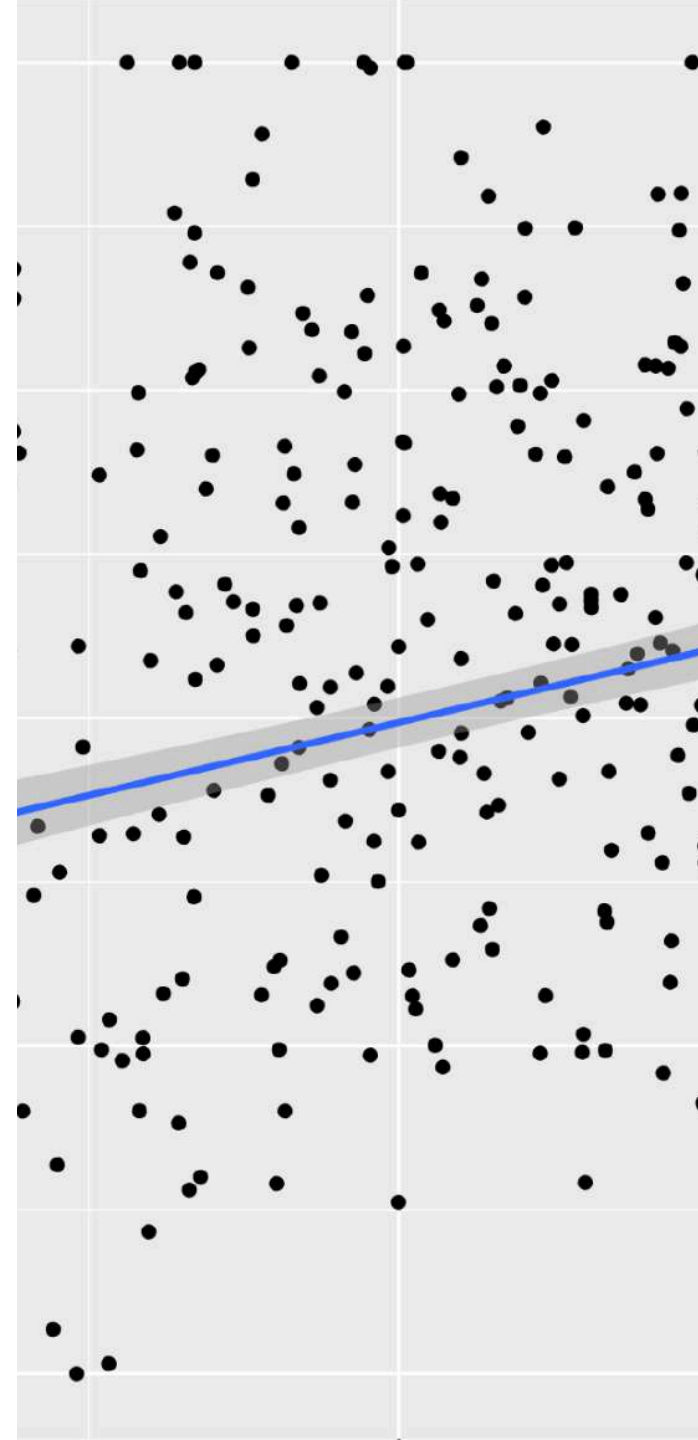
- Решение может не существовать, если значение t неадекватное
- Как проверить, что решение системы существует?



Построение
модели

Построение модели

- Какую модель мы будем обучать, если
 - Есть охваты для каждого канала
 - Есть только суммарные охваты
- Какие дополнительные данные мы могли бы учесть, если у нас были десятки тысяч рекламных кампаний?
- Как выкрутиться если данных очень мало?



Если данных очень мало

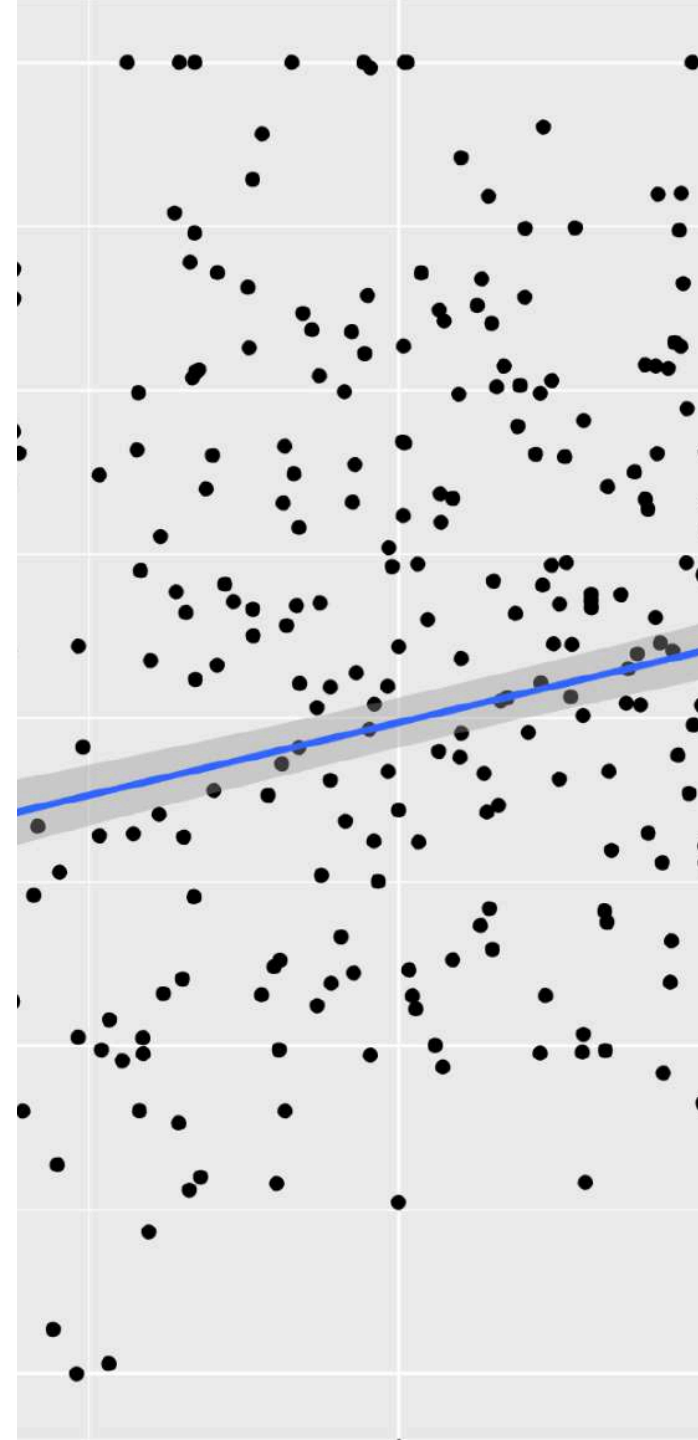
Для линейной модели с квадратичной функцией потерь:

$$\operatorname{argmin}_w \|Xw - y\|^2 = (X^T X)^{-1} X^T y$$

Если X – квадратная обратимая матрица, то:

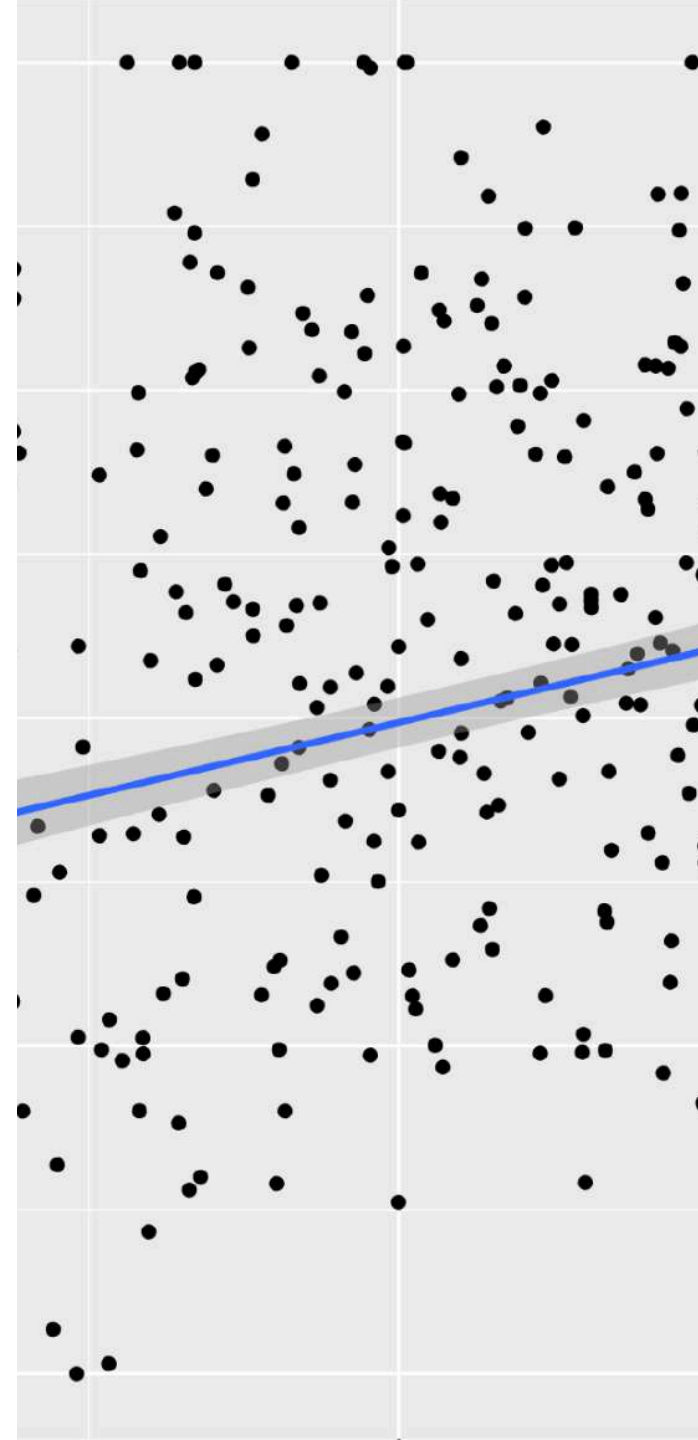
$$(X^T X)^{-1} X^T y = X^{-1} y,$$

Т.е. модель сможет «объяснить» любые наблюдения y . Что с этим делать?



Если данных очень мало

Т.е. модель сможет «объяснить» любые наблюдения у. Что с этим делать?

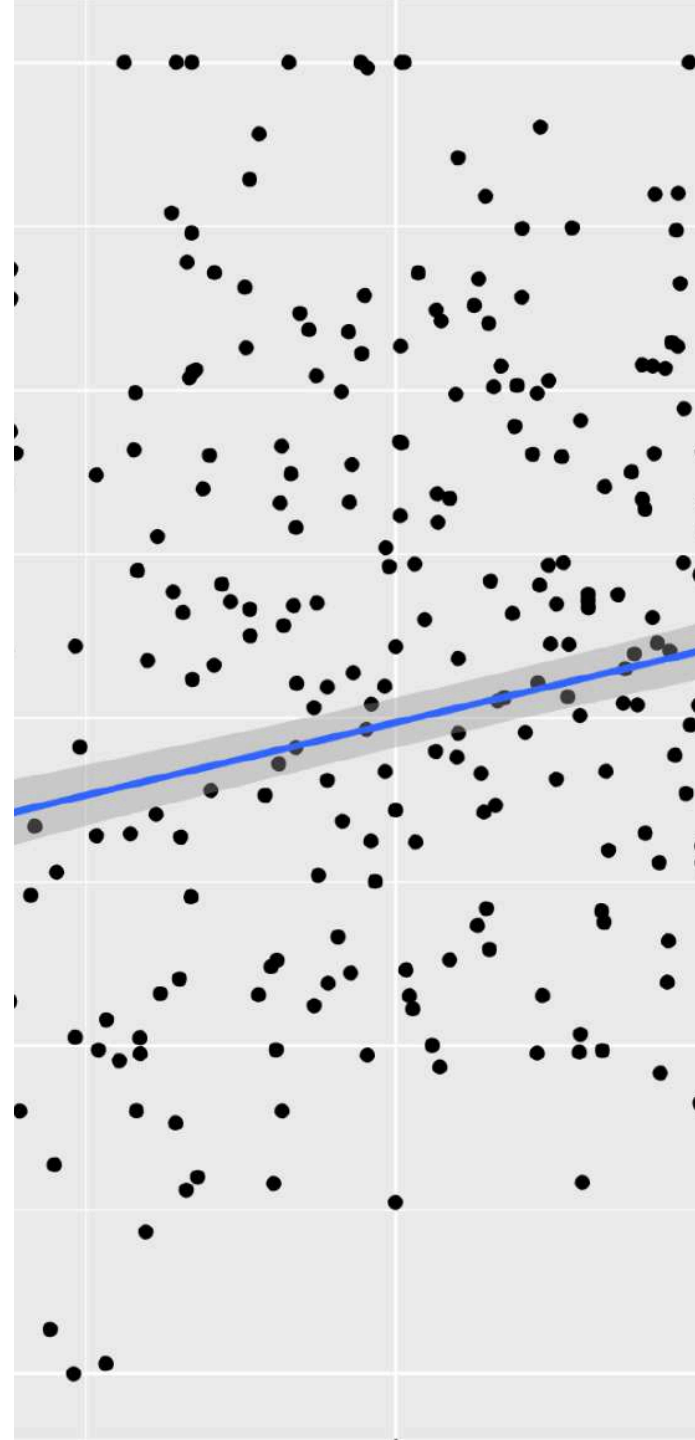


Если данных очень мало

Т.е. модель сможет «объяснить» любые наблюдения у. Что с этим делать?

Подсказки:

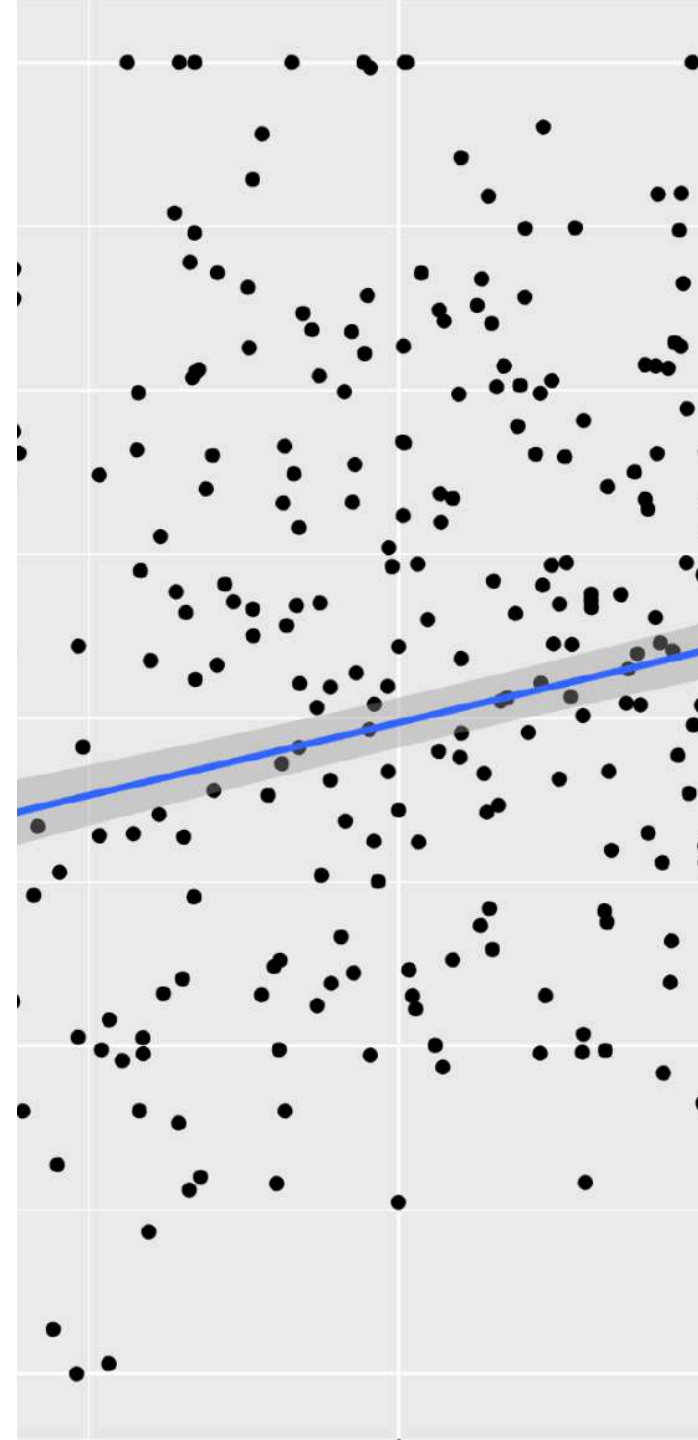
1. Что здесь Bias, а что Variance?
2. Как бороться с переобучением?



Итоговая оптимизация

$$\begin{cases} \min_x \sum_{i=1}^n x_i \\ \sum_{i=1}^n w_i x_i \geq t \\ l_i \leq x_i \leq r_i \end{cases}$$

Это задача линейного программирования, есть много библиотек для эффективного решения



The background is a light teal color with a repeating pattern of stylized speech bubbles and exclamation marks. The speech bubbles are in dark blue, orange, and light orange colors, some with horizontal lines inside. The exclamation marks are in orange and dark blue. Some elements have small black lines radiating from them, suggesting motion or emphasis.

Discussion

Занимательные факты

- Как часто будет выполняться условие?

$$\sum_{i=1}^n w_i x_i \geq t$$



Занимательные факты

- Как часто будет выполняться условие?

$$\sum_{i=1}^n w_i x_i \geq t$$

Зависит от обучения



Занимательные факты

- Как часто будет выполняться условие?

$$\sum_{i=1}^n w_i x_i \geq t$$

Зависит от обучения:

- MSE - ?
- MAE – 0.5
- Quantile loss - α



Где может проявляться

- Купи на X рублей и получи скидку
- Среднее время подачи
- Фин. Модели для проса



А что в итеративном подходе?



А что в итеративном подходе?

- Редко нужны эвристики, часто легче сразу сделать модель
- Связь точности и бизнеса – большой вопрос. Можно делать ухудшающие
- Полезно иметь «стандартный флоу» для таких задач, чтобы минимизировать время на создание mvp

