6. Поиск оптимальной цены для многих товара

Повторение одномерного случая

Формальная постановка задачи

Максимизировать прибыль $\boldsymbol{y}(\boldsymbol{p}, \boldsymbol{\xi})$ по цене $\boldsymbol{p} \in [\boldsymbol{p}, +\infty)$.

4 разные постановки:

1. $p^* = \underset{p}{\operatorname{argmax}} E[y(p_n, \xi)]$ за n итераций.
2. $p^* = \underset{p}{\operatorname{argmax}} E[y(p, \xi)]$ за минимальное время.
3. $\max\limits_{p} \sum E[y(p, \xi)]$ Оптимизация траектории.
4. Найти зависимость $y(p)$

$E[\cdot]$ - может быть любым оператором. Можно считать, как мат. ожидание.

Параметрический подход

$\boldsymbol{y}(\boldsymbol{p}, \boldsymbol{\xi}) = \boldsymbol{Q}(\boldsymbol{p})(\boldsymbol{p} - \underline{\boldsymbol{p}})$,

где $\boldsymbol{Q}(\boldsymbol{p})$ - число проданных товаров в зависимости от цены.

Линейная функция: $\boldsymbol{Q}(\boldsymbol{p}) = \boldsymbol{max}(-\boldsymbol{ap} + \boldsymbol{b}, \boldsymbol{0})$

Гиперболическая функция: $\boldsymbol{Q}\ (\boldsymbol{p}) = \boldsymbol{max}(-\boldsymbol{a}/\boldsymbol{p} + \boldsymbol{b}, \boldsymbol{0})$

Экспоненциальная функция: $\boldsymbol{Q}(\boldsymbol{p}) = \boldsymbol{max}(-\boldsymbol{exp}(\boldsymbol{ap} + \boldsymbol{b})\boldsymbol{c} + \boldsymbol{d}, \boldsymbol{0})$

Показательная функция $\boldsymbol{Q}(\boldsymbol{p}) = \boldsymbol{max}(\boldsymbol{ba}^{\boldsymbol{p}} + \boldsymbol{d}, \boldsymbol{0})$

Шум версия 2

$\boldsymbol{y(p,\xi) = Q(p)\xi(p - \underline{p})}$,
где $\boldsymbol{Q(p)}$ - число проданных товаров в зависимости от цены.

Линейная функция: $\mathrm{Q}(\boldsymbol{p}) = \boldsymbol{max(-ap + b, 0)}$
Гиперболическая функция: $\boldsymbol{Q(x) = max(-a/p + b, 0)}$
Экспоненциальная функция: $\boldsymbol{Q(x) = max(-exp(ap + b)c + d, 0)}$

$\boldsymbol{E[\xi] = 1}$ (несмещенный шум)
$\boldsymbol{0 \leq \xi}$ – продажи не могу быть отрицательными

Система координат

Вообще говоря, неудобно работать в пространстве (цена, прибыль).
Удобнее работать в пространстве (процентная наценка, прибыль).
Т.е. сделаем такую замену: $\frac{\boldsymbol{p}-\underline{\boldsymbol{p}}}{\underline{\boldsymbol{p}}} = \boldsymbol{margin}\ (\boldsymbol{m})$

$$\boldsymbol{y}(\boldsymbol{p},\boldsymbol{\xi}) = \boldsymbol{Q}(\boldsymbol{p})(\boldsymbol{p}-\underline{\boldsymbol{p}}),$$

Можно переписать в виде:

$$\boldsymbol{y}(\boldsymbol{m},\boldsymbol{\xi}) = \boldsymbol{\xi}\cdot\underline{\boldsymbol{p}}\cdot\boldsymbol{Q}(\boldsymbol{m})\boldsymbol{m},\ \boldsymbol{m}\in[\boldsymbol{0},\overline{\boldsymbol{m}})$$

Линейная функция: $\mathrm{Q}(\boldsymbol{m}) = \boldsymbol{max}(-\boldsymbol{am}+\boldsymbol{b},\boldsymbol{0})$
Гиперболическая функция: $\boldsymbol{Q}(\boldsymbol{x}) = \boldsymbol{max}(-\boldsymbol{a}/\boldsymbol{m}+\boldsymbol{b},\boldsymbol{0})$
Экспоненциальная функция: $\boldsymbol{Q}(\boldsymbol{x}) = \boldsymbol{max}(-\mathbf{exp}(\boldsymbol{am}+\boldsymbol{b})\boldsymbol{c}+\boldsymbol{d},\boldsymbol{0})$

Как выбрать параметрическую функцию?

Как выбрать лучшую параметризацию?

Очевидно, что от параметризации много что зависит.

Линейная функция: $\boldsymbol{Q(p) = max(-ap + b, 0)}$
Гиперболическая функция: $\boldsymbol{Q\ (p) = max(-a/p + b, 0)}$
Экспоненциальная функция: $\boldsymbol{Q(p) = max(-exp(ap + b)c + d, 0)}$
Показательная функция $\boldsymbol{Q(p) = max(ba^p + d, 0)}$

Как выбрать правильную параметризацию мы обсудим в лекции «непараметрические» методы.
Пока приведем общее рассуждение.

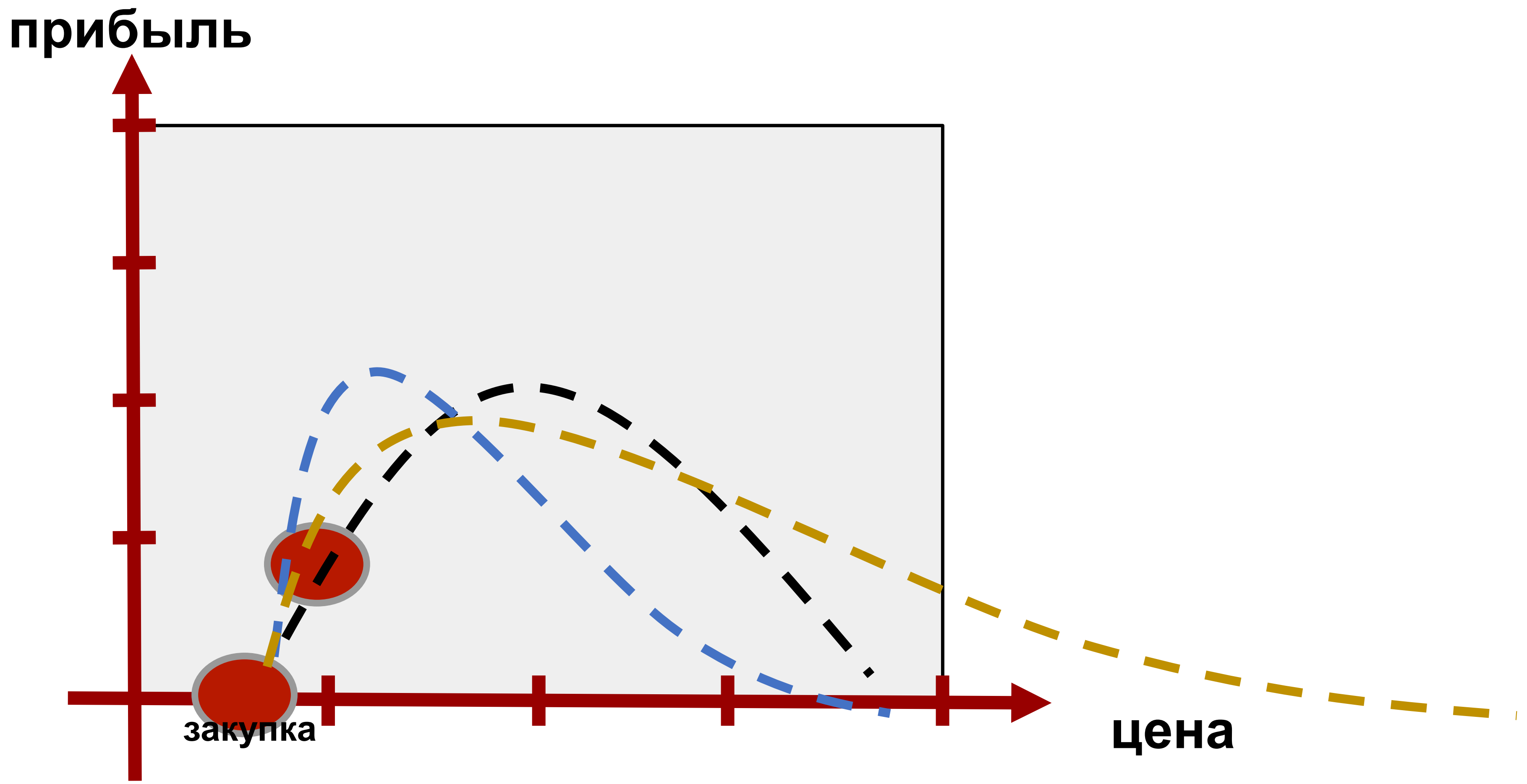
Несогласие в комитете (Query By Committee)

Query By Committee (QBC): у нас есть J алгоритмов (параметрических моделей). Они образуют *комитет*. Выбираем x_i – где модели между собой максимально расходятся, чтобы снизить неопределенность в будущем. Так мы сможем быстрее найти лучшую параметризацию.

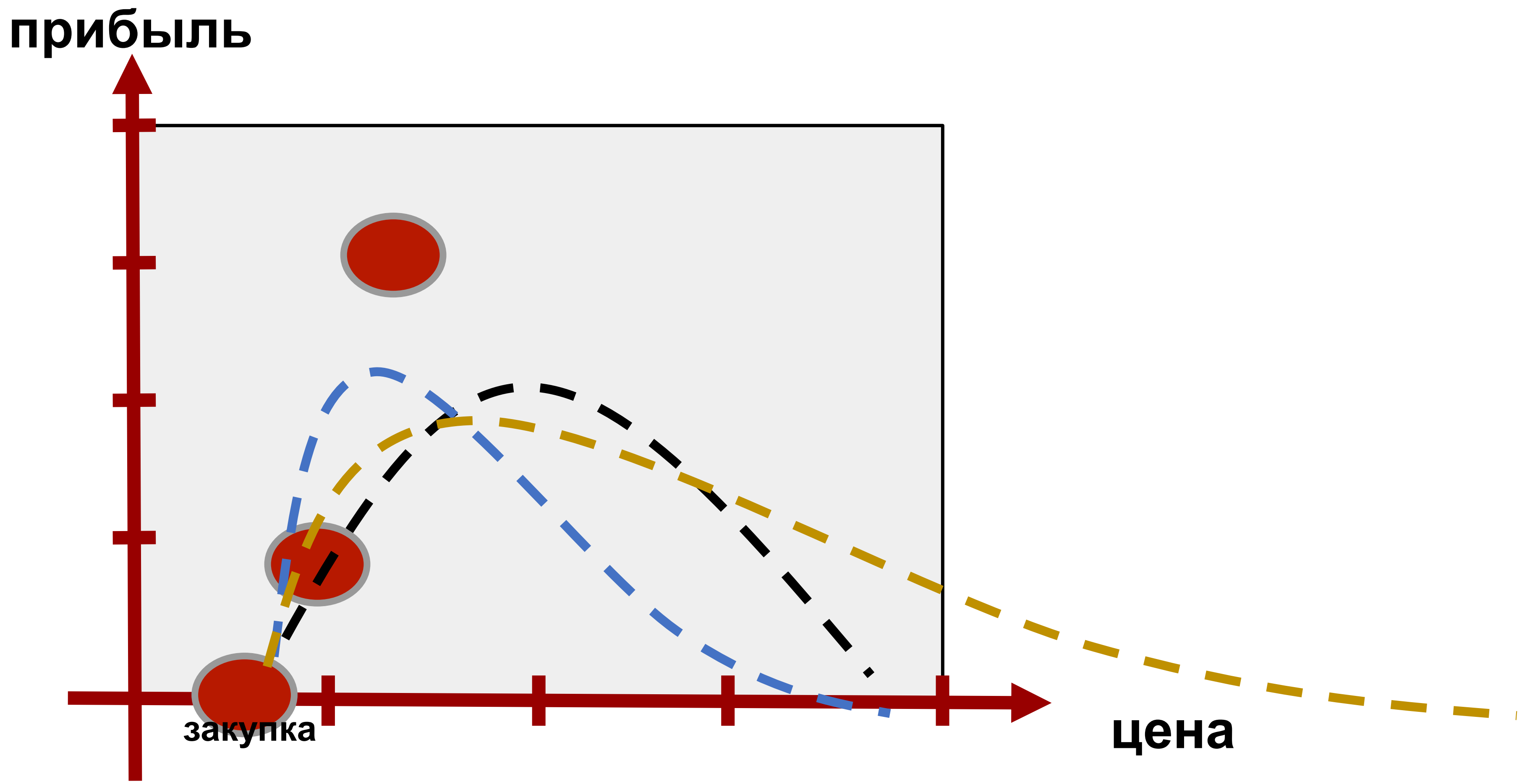
Что значит максимально расходятся?

1) Максимальная выборочная дисперсия
2) Максимальный (интерквантильный) размах
3) Взвешенная оценка (1) и (2)
4) …

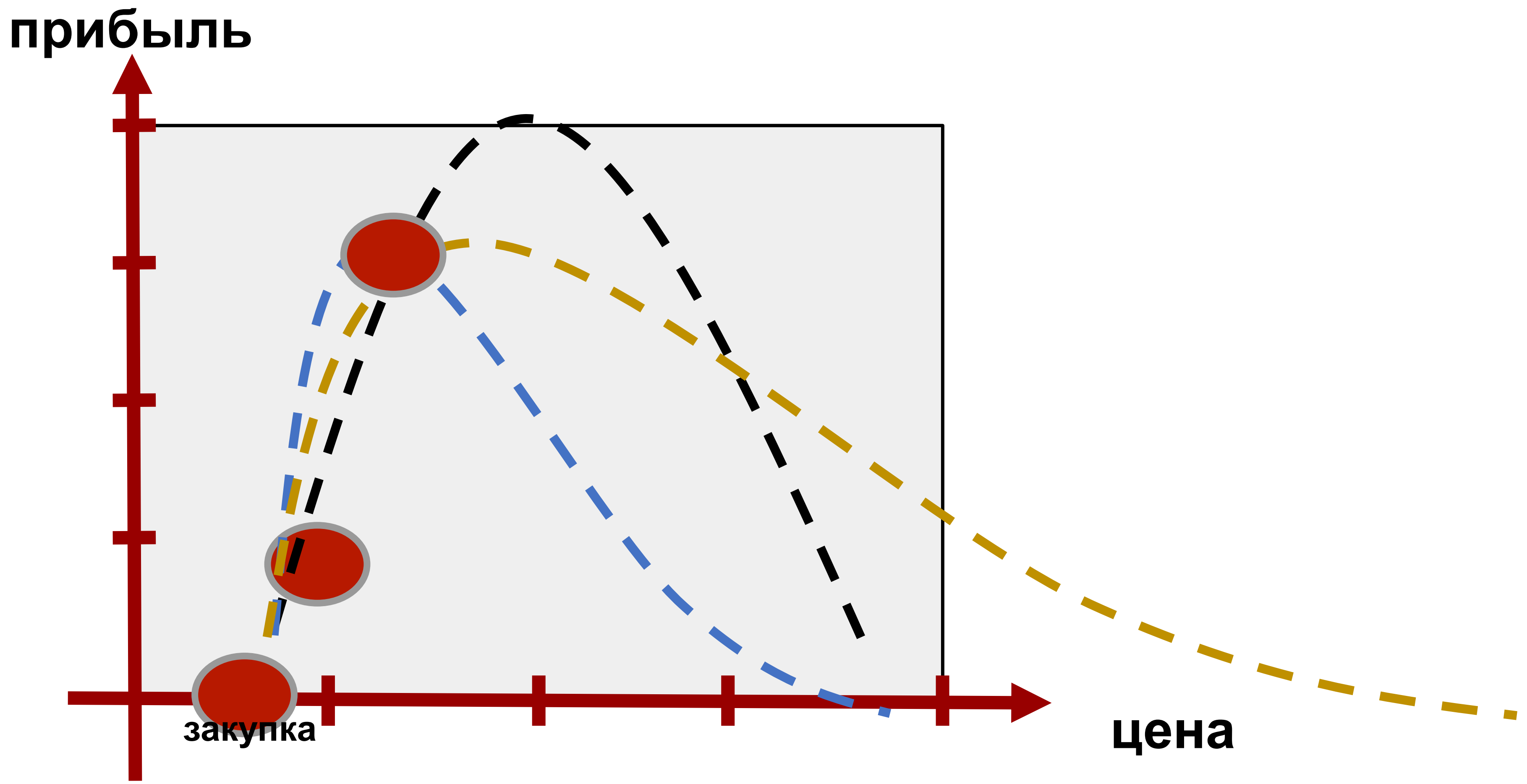
Алгоритм. Выбираем точку



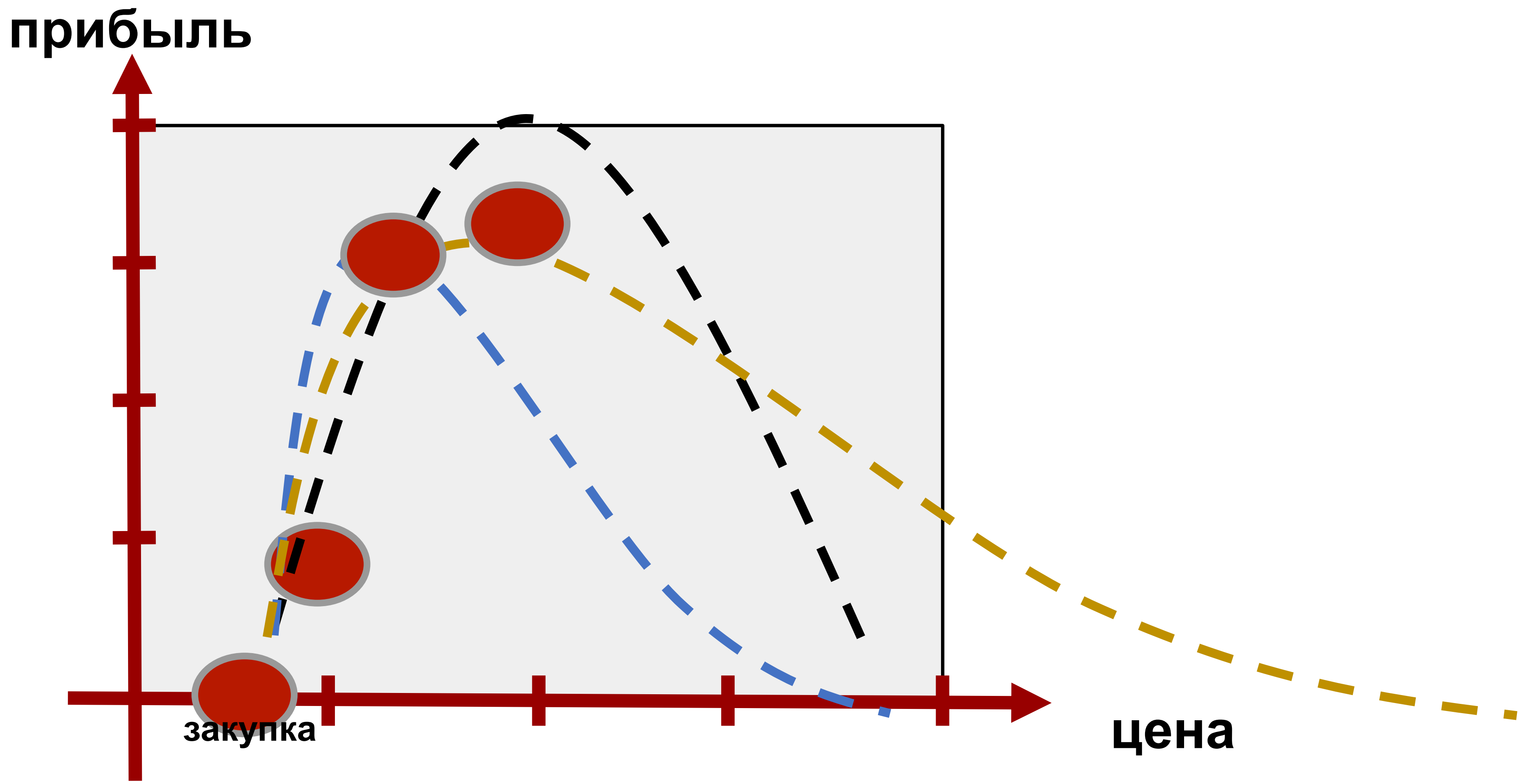
Алгоритм. Получаем оценку



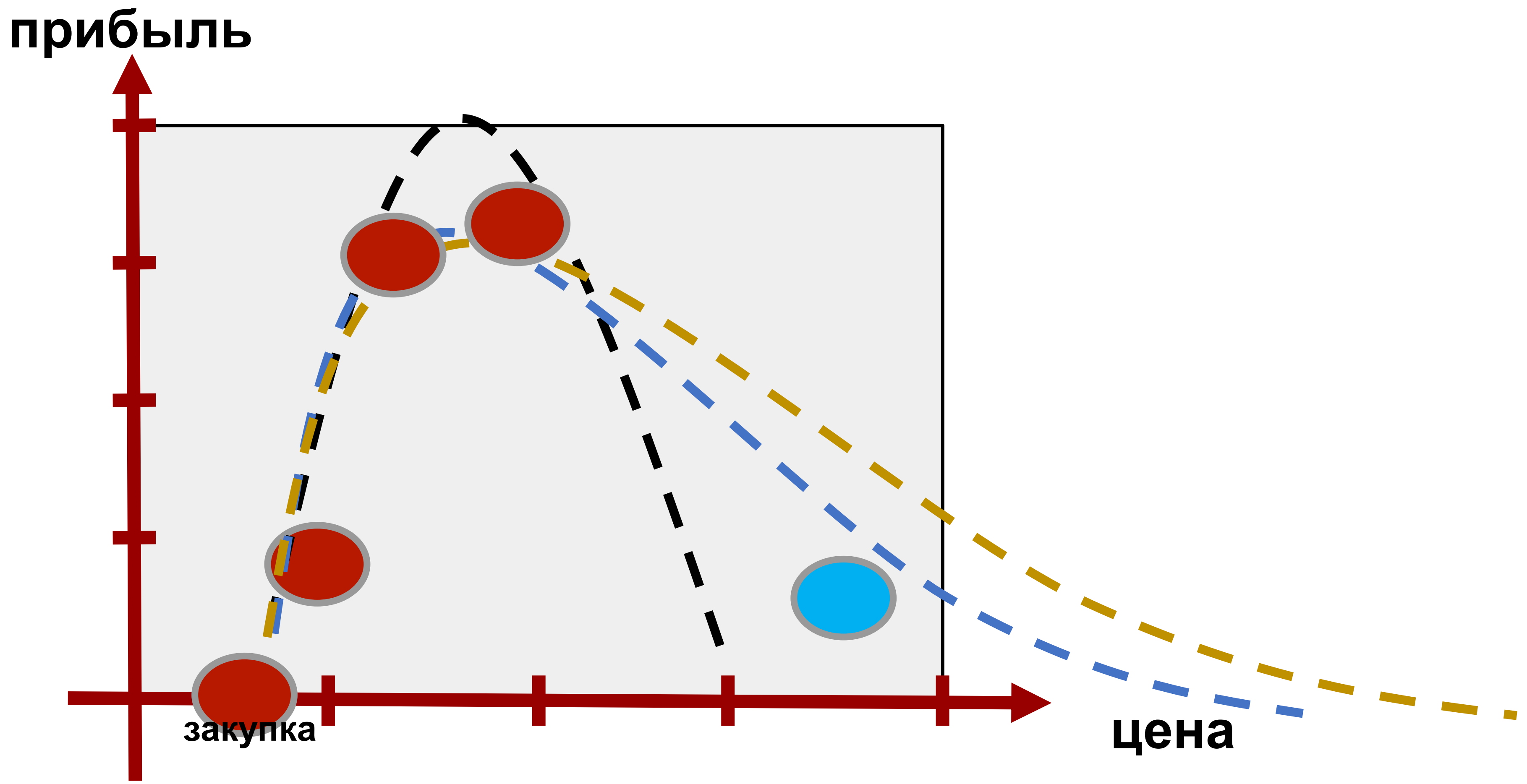
Алгоритм. Пересчитываем параметры



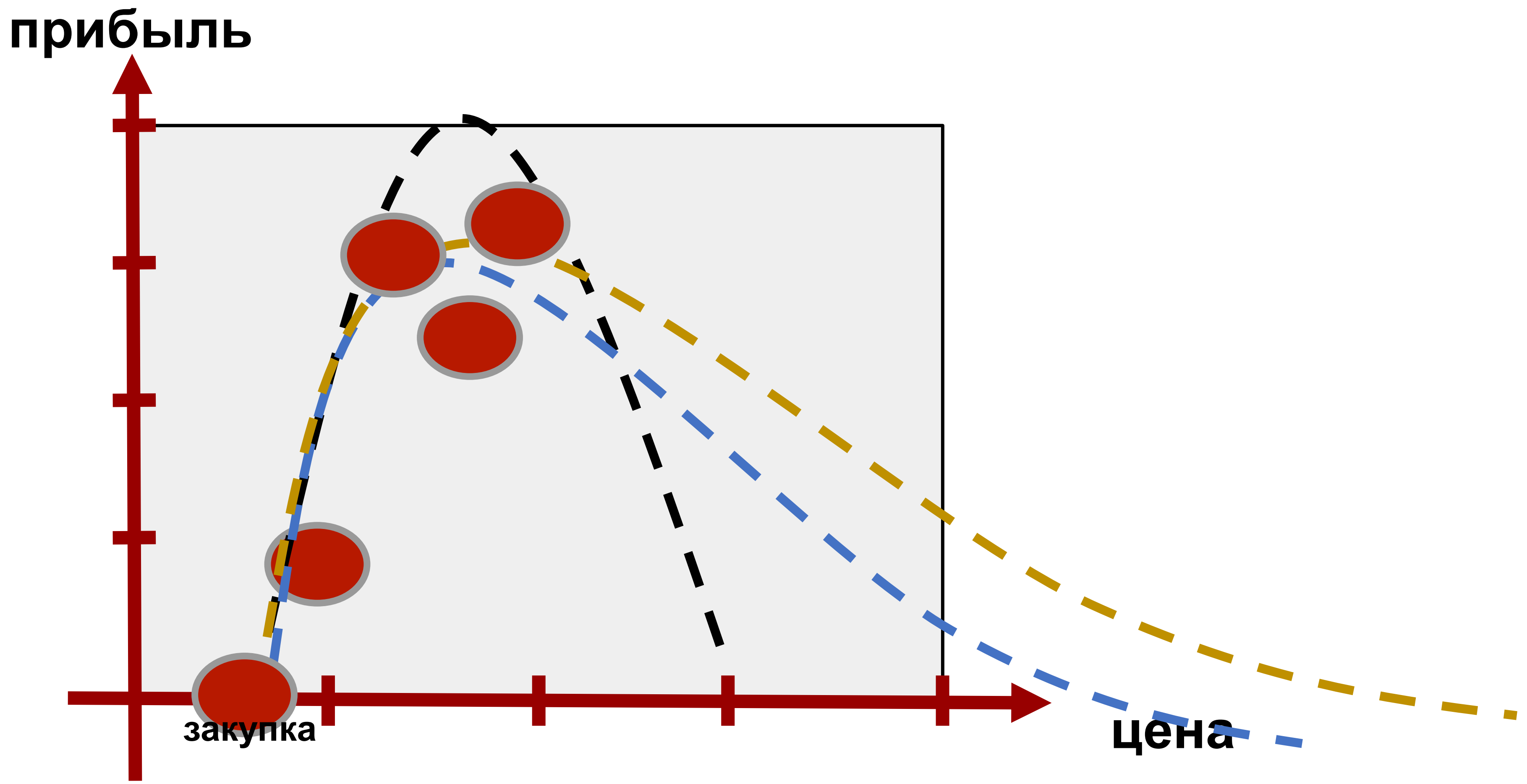
Алгоритм. Выбираем точку



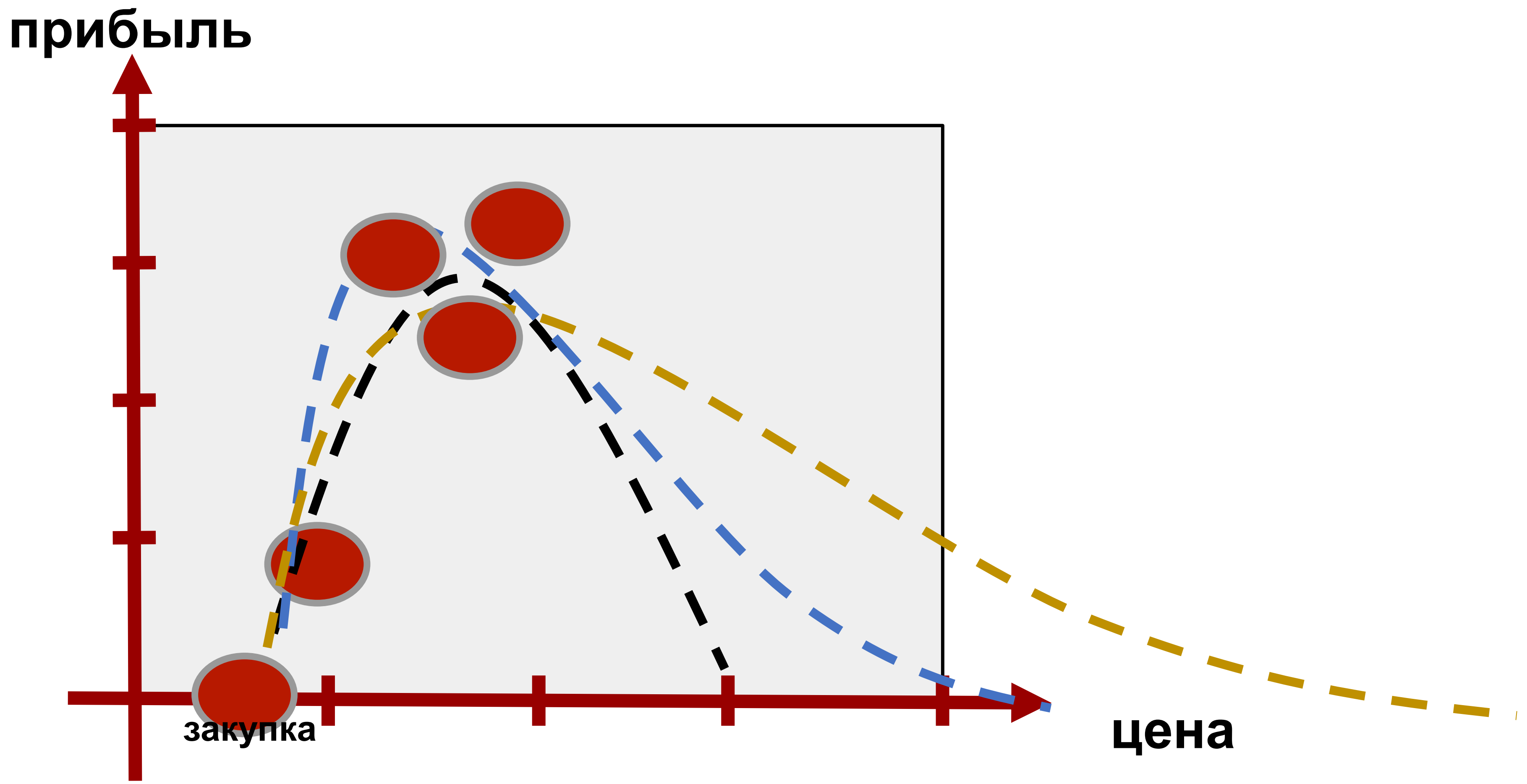
Алгоритм. Пересчитываем параметры



Алгоритм. Выбираем точку



Алгоритм. Пересчитываем параметры



Как выбрать лучшую параметризацию?

Мы можем адаптировать параметризацию под получаемые результаты. То есть по полученным прецедентам выбирать лучшую из известных функций. Параметризации желательно выбрать заранее, чтобы на них не переобучиться.

Важно при этом не переобучиться.

Двумерный случай

Пример зависимости

Пример 1

Пусть есть 2 товара: A и B.
Функция принятия решения покупателей – покупать товар с минимальной ценой пропорционально его цене. В случае равенства (с точностью до 3ех рублей) покупать случайным образом один из товаров с вероятностью 0.5.

Пример 1 (итерация 1)

Допустим стартовая точка $p_1, p_2 = 100, 100$

При этом продажи $q_1, q_2 =$ 99,101(из-за шума).
Решаем параметрическую задачу оптимизации.
Прогнозный оптимум **находится в разных точках**.
Вообще говоря не важно в каких. Пусть это будут
$p_1, p_2 = 105, 95$.
В итоге получаем что в точке $q_1, q_2 = 0, 210$

Что делать?

Пример 1 (итерация 1)

Что будет дальше?
На практике: вы должны выкатить alert и расследовать инцидент, т.к. есть экстремальные скачки.

Если вы этого не сделали, что будет дальше делать алгоритм?
Первый товар нашел «правый край»
Второй товар экстремально увеличивается в прибыли с уменьшением цены.

Пример 1 (итерация 2)

Что будет дальше?

Первый товар сильно опуститься в цене.
Второй в зависимости от параметризации тоже опуститься.
$p_1, p_2 = 70, 82.$
В итоге получаем что в точке $q_1, q_2 = 250{,}0$

На практике: ВОТ тут уже точно должны выкатить alert и расследовать инцидент.

Пример 1 (итерация 2)

Давайте посмотрим на динамику:

$p_1, p_2 = 100, 100 \quad q_1, q_2 = 99,101$

$p_1, p_2 = 105, 95 \quad q_1, q_2 = 0, 210$

$p_1, p_2 = 70, 82 \quad q_1, q_2 = 250,0$

На практике: ВОТ тут уже точно должны выкатить alert и расследовать инцидент. По крайней мере по второму товару.

Пример 1 (итерация 3)

Давайте посмотрим на динамику:

$p_1, p_2 = 100, 100 \quad q_1, q_2 = 99{,}101$

$p_1, p_2 = 105, 95 \quad q_1, q_2 = 0, 210$

$p_1, p_2 = 70, 82 \quad q_1, q_2 = 250{,}0$

$p_1, p_2 = 60, 95 \quad q_1, q_2 = 260{,}0$

Пример 1 (итерация 4)

Давайте посмотрим на динамику:

$p_1, p_2 = 100, 100 \quad q_1, q_2 = 99{,}101$

$p_1, p_2 = 105, 95 \quad q_1, q_2 = 0, 210$

$p_1, p_2 = 70, 82 \quad q_1, q_2 = 250{,}0$

$p_1, p_2 = 60, 95 \quad q_1, q_2 = 260{,}0$

Пусть в оптимум расчетный по 1-ому товару. в ~70.

$p_1, p_2 = 73, 75 \quad q_1, q_2 = 120{,}120$ (плохо)

Пример 1 (итерация 5)

Давайте посмотрим на динамику:

$p_1, p_2 = 100, 100 \quad q_1, q_2 = 99{,}101$

$p_1, p_2 = 105, 95 \quad q_1, q_2 = 0, 210$

$p_1, p_2 = 70, 82 \quad q_1, q_2 = 250{,}0$

$p_1, p_2 = 60, 95 \quad q_1, q_2 = 260{,}0$

Пусть в оптимум расчетный по 1-ому товару. в ~70.

$p_1, p_2 = 73, 75 \quad q_1, q_2 = 120{,}120$ (плохо)

$p_1, p_2 = 65, 75 \quad q_1, q_2 = 260{,}0$ (очень плохо)

Зависимость и независимость

Двумерный случай

Какие могут быть варианты:

1. 2 независимых товара:
   - Сведение к двум одномерным случаям
   - Поиск общей «эластичности»
2. 2 зависимых товара:
   - Частные случаи
   - Двумерная параметризация

Два независимых товара

Два независимых товара

Общая постановка:

$$\boldsymbol{y}\ \ (\boldsymbol{m_1}, \boldsymbol{m_2}, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot (\underline{\boldsymbol{p}}_1 \cdot \boldsymbol{Q_1}(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_1} + \underline{\boldsymbol{p}}_2 \cdot \boldsymbol{Q_2}(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_2})$$

Если у нас 2 независимых товара, то:

$$\boldsymbol{y}\ \ (\boldsymbol{m_1}, \boldsymbol{m_2}, \boldsymbol{\xi}) = \boldsymbol{y_1}(\boldsymbol{m_1}, \boldsymbol{\xi}) + \boldsymbol{y_2}(\boldsymbol{m_2}, \boldsymbol{\xi})$$

$$\boldsymbol{y_1}(\boldsymbol{m_1}, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot \underline{\boldsymbol{p}}_1 \cdot \boldsymbol{Q_1}(\boldsymbol{m_1})\boldsymbol{m_1}$$

$$\boldsymbol{y_2}(\boldsymbol{m_2}, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot \underline{\boldsymbol{p}}_2 \cdot \boldsymbol{Q_2}(\boldsymbol{m_2})\boldsymbol{m_2}$$

Применяем те же самые методы, что для одномерной оптимизации для каждого из товаров.

Как проверить независимость? См. кластеризацию.

Два независимых товара

Есть 2 независимые задачи:

$$y_1(m_1, \xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1) m_1$$

$$y_2(m_1, \xi) = \xi \cdot \underline{p}_2 \cdot Q(m_2) m_2$$

Применяем те же самые методы, что для одномерной оптимизации для каждого из товаров.

Как проверить независимость? См. кластеризацию.
Проверка на KVI, якоря, товары-заменители.

Но! Мы проверяли только независимость в продажах, а что если есть какие-то другие зависимости?!

Важное замечание

Полностью независимых товаров не бывает! Если кто-то завтра начнет продавать золото по 1 рублю за кг, он может обрушить продажи всех остальных товаров.

$$\boldsymbol{y_1(m_1, \xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1) m_1}$$
$$\boldsymbol{y_2(m_1, \xi) = \xi \cdot \underline{p}_2 \cdot Q(m_2) m_2}$$

Поэтому мы подразумеваем, что $\boldsymbol{m_1, m_2 > 0}$. Или покупатель не знает о себестоимости товара.
И в целом $\boldsymbol{m_1, m_2}$ находятся в «адекватной» области – нет экстремально низких или высоких цен.

Как проверить независимость?

Общий смысл проверки:
$\boldsymbol{Q_1(m_1, m_2)}$ не зависит от $\boldsymbol{m_2}$
$\boldsymbol{Q_2(m_1, m_2)}$ не зависит от $\boldsymbol{m_1}$

Или $\boldsymbol{Q_1(m_1, x) = Q_1(m_1, z)}$ (на уровне шума), где $\boldsymbol{x \neq z}$
То есть число продаж одного товара не зависит от изменения цены (маржи) другого.

Как проверить независимость?

Общий смысл проверки:
$\boldsymbol{Q_1(m_1, m_2)}$ не зависит от $\boldsymbol{m_2}$
$\boldsymbol{Q_2(m_1, m_2)}$ не зависит от $\boldsymbol{m_1}$

Или $\boldsymbol{Q_1(m_1, x) = Q_1(m_1, z)}$ (на уровне шума), где $\boldsymbol{x \neq z}$
То есть число продаж одного товара не зависит от изменение цены (маржи) другого.

Проблема: если это 2 новых товара (или на эти товары никогда не менялась цена), как убедиться, что действительно нет зависимости?
Ведь если мы итерируем по $\boldsymbol{m_1}$ и $\boldsymbol{m_2}$ одновременно, то у нас не будет ситуации, когда $\boldsymbol{Q_1(m_1, x) = Q_1(m_1, z)}$, поскольку $\boldsymbol{m_1}$ всегда разное.

Как проверить независимость?

Проблема: если это 2 новых товара (или на эти товары никогда не менялась цена), как убедиться, что действительно нет зависимости?
Ведь если мы итерируем по $\boldsymbol{m_1}$ и $\boldsymbol{m_2}$ одновременно, то у нас не будет ситуации, когда $\boldsymbol{Q_1(m_1, x) = Q_1(m_1, z)}$, поскольку $\boldsymbol{m_1}$ всегда разное.

На практике:
1) Случаи зависимости редки и ими изначально можно пренебречь
2) Нужно тестировать ☺

Общая эластичность

Поиск общей эластичности

На самом деле похожие товары обладают похожей эластичностью.
В рамках нашей постановки «похожей» параметризацией.

Как выбрать «правильную» параметрическую модель (вернее переключаться между ними) мы обсудим в следующих лекциях.
Но как это можно использовать сейчас?

2 варианта:

- Мы предполагаем, что оптимум по марже находится в одной и той же точке.
- Мы считаем $\boldsymbol{Q_1(m_1) \sim Q_2(m_2)}$ функции спроса одинаковые, но с разными параметрами.

Оптимум в одной точке по m

На практике это очень частая история. На многие группы товаров оптимум по марже (процентной наценке) находится в одной и той же точке.
Например:

- Конфеты
- Кофе
- Крупы
- ...
- Электроника в рамках одного бренда
- Спорт инвентарь
- …

Важно: это может работать когда m>>0.

Оптимум в одной точке по m

Тогда мы можем итерировать «общую» функцию:

$$\boldsymbol{y}(\boldsymbol{m},\boldsymbol{\xi}) = \boldsymbol{\xi} \cdot \underline{\boldsymbol{p}} \cdot \boldsymbol{Q}(\boldsymbol{m})\boldsymbol{m}$$

$$m^* = \underset{m}{\operatorname{argmax}}\, E[y(\boldsymbol{m},\boldsymbol{\xi})] = \underset{m}{\operatorname{argmax}}\, \boldsymbol{Q}(\boldsymbol{m})\boldsymbol{m}$$

Обратите внимание, что от $\underline{\boldsymbol{p}}$ не зависит, т.к. $\underline{\boldsymbol{p}} > 0$. Это влияет только на значение и на установку цены.

Только за шаг можем тестировать 2 точки. Если у нас N товаров, то тестировать сразу N точек.

Функции спроса одинаковые

Такое предположение на практике удобно использовать для новых товаров. Или если на один товары уже были итерации.

Как это можно использовать?

- Это экономит на выборе параметрической функции, что часто бывает очень ценно. То есть не нужно тестировать дополнительные точки в поисках оптимальной параметризации.
- Можно тестировать на «лучшую» параметрическую функцию для менее рискованного товара.

Независимые товары

В худшем случае можно просто искать оптимум независимо. А это мы уже умеем делать.

В лучшем случае можно ускорить нахождение оптимума:

- Через «общий» оптимум.
- Через известную функцию Q

Два зависимых товара

Два зависимых товара

Общая постановка:

$$\boldsymbol{y}\quad(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot (\underline{\boldsymbol{p}}_1 \cdot \boldsymbol{Q}_1(\boldsymbol{m}_1, \boldsymbol{m}_2)\boldsymbol{m}_1 + \underline{\boldsymbol{p}}_2 \cdot \boldsymbol{Q}_2(\boldsymbol{m}_1, \boldsymbol{m}_2)\boldsymbol{m}_2)$$

1. Частные случаи:
   - Якорь + сопутствующий
   - Два взаимозаменяемых
2. Двумерная параметризация
3. Зависимость через модель потребления

Якорь + сопутствующий

Якорь + сопутствующий

Вариант 1: суммарная цена, когда покупатель покупает корзину целиком.
Здесь удобно вернуться в пространство цен.

$$y\ \ (p_1, p_2, \xi) = \xi(Q_{1,2}(p_1 + p_2)(p_1 + p_2 - \underline{p}_1 - \underline{p}_2)$$

Тогда мы вернулись к одномерной задаче с заменой переменных $p = p_1 + p_2$.

Якорь + сопутствующий

Общая постановка:

$$\boldsymbol{y}\ \ (\boldsymbol{m_1}, \boldsymbol{m_2}, \xi) = \boldsymbol{\xi} \cdot (\underline{\boldsymbol{p}}_{\boldsymbol{1}} \cdot \boldsymbol{Q_1}(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_1} + \underline{\boldsymbol{p}}_{\boldsymbol{2}} \cdot \boldsymbol{Q_2}(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_2})$$

Пусть товар 1 – это якорь.

Вариант 2:

$$\boldsymbol{y}\ \ (\boldsymbol{m_1}, \boldsymbol{m_2}, \xi) = \boldsymbol{\xi} \cdot (\underline{\boldsymbol{p}}_{\boldsymbol{1}} \cdot \boldsymbol{Q_1}(\boldsymbol{m_1})\boldsymbol{m_1} + \underline{\boldsymbol{p}}_{\boldsymbol{2}} \cdot \boldsymbol{Q_2}(\boldsymbol{q_1}, \boldsymbol{m_2})\boldsymbol{m_2})$$

$$\boldsymbol{q_1} = \boldsymbol{\xi} \cdot \boldsymbol{Q_1}(\boldsymbol{m_1})$$

Где $\boldsymbol{Q_2}$ - зависит от числа проданных единиц первого товара и маржи (цены) второго товара.

Якорь + сопутствующий

$$\boldsymbol{y}\ \ (\boldsymbol{m_1}, \boldsymbol{m_2}, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot (\underline{\boldsymbol{p}}_{\boldsymbol{1}} \cdot \boldsymbol{Q_1}(\boldsymbol{m_1})\boldsymbol{m_1} + \underline{\boldsymbol{p}}_{\boldsymbol{2}} \cdot \boldsymbol{Q_2}(\boldsymbol{q_1}, \boldsymbol{m_2})\boldsymbol{m_2})$$

$$\boldsymbol{q_1} = \boldsymbol{\xi} \cdot \boldsymbol{Q_1}(\boldsymbol{m_1})$$

Где $\boldsymbol{Q_2}$ - зависит от числа проданных единиц первого товара и маржи (цены) второго товара.

Примеры:

$$\boldsymbol{Q}(\boldsymbol{q_1}, \boldsymbol{m_2}) = \boldsymbol{max}((\boldsymbol{q_1} + \boldsymbol{a})(\boldsymbol{b} - \boldsymbol{c}\boldsymbol{m_2}), \boldsymbol{0})$$

$$\boldsymbol{Q}(\boldsymbol{q_1}, \boldsymbol{m_2}) = \boldsymbol{max}((\boldsymbol{q_1}+\boldsymbol{a})((\boldsymbol{b} - \boldsymbol{c} \log \boldsymbol{m_2}), \boldsymbol{0})$$

Т.е., если $\boldsymbol{a}$ – это продажи без якоря, как правило, можно вычислить заранее, как среднее число продаж без якоря. Часто $\boldsymbol{a} = 0$.

Т.е. если $\boldsymbol{q_1} = \boldsymbol{0}$ сопутствующий не продается.

Якорь + сопутствующий

Пример. Рассмотрим случай, когда $a=0$.

$$y\ \ (m_1,m_2) = (\underline{p}_1 \cdot Q_1(m_1)m_1 + \underline{p}_2 \cdot Q_2(q_1,m_2)m_2)=$$
$$= \underline{p}_1 \cdot Q_1(m_1)m_1 + \underline{p}_2 \cdot Q_1(m_1)(b - cm_2)m_2 =$$
$$= Q_1(m_1)(\underline{p}_1 \cdot m_1 + \underline{p}_2 \cdot (b - cm_2)m_2)$$

$$\frac{\partial y\ \ (m_1,m_2)}{\partial m_2} = -2c\underline{p}_2 \cdot Q_1(m_1)m_2 + b\underline{p}_2 \cdot Q_1(m_1) = 0$$
$$\frac{\partial y\ \ (m_1,m_2)}{\partial m_1} = Q_1'(m_1)\left(\underline{p}_1 \cdot m_1 + \underline{p}_2 \cdot (b - cm_2)m_2\right) + Q_1(m_1)\underline{p}_1 = 0$$

$$Q_1(m_1) > 0 \Rightarrow\ -2c\underline{p}_2 \cdot m_2 + b\underline{p}_2 = 0 \Rightarrow m_2 = \frac{b\underline{p}_2}{2c\underline{p}_2}$$

В качестве упражнения можно дорешать.

Якорь + сопутствующий

Вариант 3: Когда у нас есть 2 типа покупателей в какой-то пропорции (неизвестной) из варианта 1 и варианта 2.

$$\boldsymbol{y}\quad(\boldsymbol{m_1},\boldsymbol{m_2}) = \boldsymbol{\alpha}(\boldsymbol{Q_{1,2}}(\boldsymbol{p_1}+\boldsymbol{p_2})\left(\boldsymbol{p_1}+\boldsymbol{p_2}-\underline{\boldsymbol{p}}_{\boldsymbol{1}}-\underline{\boldsymbol{p}}_{\boldsymbol{2}}\right)+$$
$$+\boldsymbol{\beta}(\underline{\boldsymbol{p}}_{\boldsymbol{1}}\cdot\boldsymbol{Q_1}(\boldsymbol{m_1})\boldsymbol{m_1}+\underline{\boldsymbol{p}}_{\boldsymbol{2}}\cdot\boldsymbol{Q_2}(\boldsymbol{q_1},\boldsymbol{m_2})\boldsymbol{m_2})$$

При этом у мы не можем их разделить.
Получаем «сложную» параметрическую модель: $\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{Q_{1,2}},\boldsymbol{Q_1},\boldsymbol{Q_2}$.
Решаем из общих соображений.

Товары с общим спросом

Товары с общим спросом

Вариант всего один ☺.

$$\boldsymbol{y}\ (\boldsymbol{m_1}, \boldsymbol{m_2}, \boldsymbol{\xi}) = \color{red}{\boldsymbol{\xi}} \cdot (\underline{\boldsymbol{p}}_1 \cdot \boldsymbol{Q}_1(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_1} + \underline{\boldsymbol{p}}_2 \cdot \boldsymbol{Q}_2(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_2})$$

$$\boldsymbol{Q_1}(\boldsymbol{m_1}, \boldsymbol{m_2}) + \boldsymbol{Q_2}(\boldsymbol{m_1}, \boldsymbol{m_2}) = \boldsymbol{Q}(\boldsymbol{m_1}, \boldsymbol{m_2})$$

На практике удобная модель:

$$\boldsymbol{Q_1}(\boldsymbol{m_1}, \boldsymbol{m_2}) = \begin{cases} \boldsymbol{Q}(\boldsymbol{p_1}), & \boldsymbol{p_1} < \boldsymbol{p_2} \\ \dfrac{\boldsymbol{Q}}{\boldsymbol{2}} & \boldsymbol{p_1} = \boldsymbol{p_2} \\ \boldsymbol{0}, & \boldsymbol{p_1} > \boldsymbol{p_2} \end{cases}$$

Она на самом деле означает, что товары надо ходить одновременно!

Товары с общим спросом

То есть просто итерируем в одних и тех же ценах.

$$y\ \ (m_1, m_2, \xi) = \xi \cdot \left(\underline{p}_1 \cdot Q_1(m_1, m_2)m_1 + \underline{p}_2 \cdot Q_2(m_1, m_2)m_2\right),$$

$$p_1 === p_2$$

$$y\ \ (m_1, *, \xi) = \xi \cdot \left(\underline{p}_1 \cdot Q_1(m_1, m_2)m_1\right)$$

Товары с общим спросом

То есть просто итерируем в одних и тех же ценах.

$$\boldsymbol{y}\ \ (\boldsymbol{m_1}, \boldsymbol{m_2}, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot \left(\underline{\boldsymbol{p}}_{\mathbf{1}} \cdot \boldsymbol{Q}_{\mathbf{1}}(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_1} + \underline{\boldsymbol{p}}_{\mathbf{2}} \cdot \boldsymbol{Q}_{\mathbf{2}}(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_2}\right),$$

Другая функция:

$$\boldsymbol{Q_1}(\boldsymbol{m_1}, \boldsymbol{m_2}) = \boldsymbol{Q}\{\boldsymbol{p_2} - \boldsymbol{p_1}, \boldsymbol{x}\}$$

$\{\boldsymbol{p_2} - \boldsymbol{p_1}, \boldsymbol{x}\}$ - положительная срезка до уровня x.

$$\{\boldsymbol{p_2} - \boldsymbol{p_1}, \boldsymbol{x}\} = \begin{cases} \mathbf{1}, \boldsymbol{p_2} - \boldsymbol{p_1} \geq \boldsymbol{x} \\ \text{линейная функция на интервале } (-\text{x}, \text{x}) \\ \mathbf{0}, \boldsymbol{p_2} - \boldsymbol{p_1} \leq \boldsymbol{x} \end{cases}$$

Товары с общим спросом

Легко показать, что все тоже самое:

$$y\ \ (m_1, m_2, \xi) = \xi \cdot \left(\underline{p}_1 \cdot Q_1(m_1, m_2)m_1 + \underline{p}_2 \cdot Q_2(m_1, m_2)m_2\right),$$

$$p_1 === p_2$$

$$y\ \ (m_1, *, \xi) = \xi \cdot \left(\underline{p}_1 \cdot Q_1(m_1, m_2)m_1\right)$$

Двумерная параметризация

Двумерная параметризация

Теперь давайте представим, что есть какая-то более сложная зависимость.

$$\boldsymbol{y}\ \ (\boldsymbol{m_1}, \boldsymbol{m_2}, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot \left(\underline{\boldsymbol{p}}_1 \cdot \boldsymbol{Q_1}(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_1} + \underline{\boldsymbol{p}}_2 \cdot \boldsymbol{Q_2}(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_2}\right),$$

Один пример мы уже видели:

$$\boldsymbol{y}\ \ (\boldsymbol{m_1}, \boldsymbol{m_2}) = \boldsymbol{\alpha}(\boldsymbol{Q_{1,2}}(\boldsymbol{p_1} + \boldsymbol{p_2})\left(\boldsymbol{p_1} + \boldsymbol{p_2} - \underline{\boldsymbol{p}}_1 - \underline{\boldsymbol{p}}_2\right) +$$
$$+\boldsymbol{\beta}(\underline{\boldsymbol{p}}_1 \cdot \boldsymbol{Q_1}(\boldsymbol{m_1})\boldsymbol{m_1} + \underline{\boldsymbol{p}}_2 \cdot \boldsymbol{Q_2}(\boldsymbol{q_1}, \boldsymbol{m_2})\boldsymbol{m_2})$$

Двумерная параметризация

Теперь давайте представим, что есть какая-то более сложная зависимость.

$$\boldsymbol{y}\ \ (\boldsymbol{m_1}, \boldsymbol{m_2}, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot \left(\underline{\boldsymbol{p}}_{\boldsymbol{1}} \cdot \boldsymbol{Q_1}(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_1} + \underline{\boldsymbol{p}}_{\boldsymbol{2}} \cdot \boldsymbol{Q_2}(\boldsymbol{m_1}, \boldsymbol{m_2})\boldsymbol{m_2}\right),$$

Какие могут быть общие предположения на функцию $\boldsymbol{Q_1(m_1, m_2)}$?
$\boldsymbol{Q_1(m_1, *)}$ - убывает (не возрастаем) по $\boldsymbol{m_1}$ но не факт ☺
$\boldsymbol{Q_1(m_1, m_2)}$ - убывает (не возрастаем) по $\boldsymbol{m_1}$ при $\boldsymbol{p_1 < p_2}$ (скорее всего).

Многомерный случай

Многомерный случай

Хорошая новость, что в целом зависимостей не так много.
Теперь у нас есть много товаров, хотим решать многомерную задачу.

$$y\ (m_1, m_2, \dots) = \xi \cdot \left(\sum_i \underline{p}_i \cdot Q_i(m_1, m_2, \dots) m_i\right)$$

Но у нас же есть кластеризация, можем перейти к ним. Если они независимые.

$$y_1(m_1, \xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1) m_1$$
$$y_2(m_2, \xi) = \xi \cdot \underline{p}_2 \cdot Q(m_2) m_2$$
$$y_3(m_3, \xi) = \xi \cdot \underline{p}_3 \cdot Q(m_3) m_3$$

Многомерный случай

Довольно очевидно, что очень тяжело построить двумерную параметризацию. Это на практике довольно редко удается.

Многомерную параметризацию в общем случае тем более не построить.

Нужны другие методы для общего случая!

Итерация в кластере

Будем сразу считать, что мы работаем не с товарами, а с кластерами.

$$y \quad (m \quad , \xi) = \xi \cdot \underline{p} \quad \cdot Q(m \quad)m$$

Что значит в таком случае итерация?
Для всех товаров устанавливается наценка m. Считается их совокупная прибыль $y(m)$. После этого строится новая аппроксимация $y(m)$.

Оптимум кластера

Давайте рассмотрим отдельную задачу.

$$y_1(m_1, \xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1)m_1$$

Вообще говоря y_1 - это функция кластера. Т.е. мы ищем оптимум целого кластера.

Что значит, например, $\underset{m_1}{\operatorname{argmax}} E[y_1(m_1, \xi)] = m_1^* = 11\%$?
Это означает, что оптимальная «средняя» процентная наценка = 11%.
Цены при этом могут быть совершенно разными, т.к. $\underline{p}_i$ разные.

Система координат

Вообще говоря, неудобно работать в пространстве (цена, прибыль).
Удобнее работать в пространстве (процентная наценка, прибыль).
Т.е. сделаем такую замену:

$$\frac{\boldsymbol{p} - \underline{\boldsymbol{p}}}{\underline{\boldsymbol{p}}} = \boldsymbol{margin}\ (\boldsymbol{m})$$

$$\boldsymbol{p} = \boldsymbol{m}\underline{\boldsymbol{p}} + \underline{\boldsymbol{p}} = \underline{\boldsymbol{p}}(\mathbf{1} + \boldsymbol{m})$$

Оптимум кластера

Понятно, что если кластер плохо сбалансирован, погрешность в оптимуме может быть очень существенной.

Откуда она возникает?
На практике, по переменной m также же осуществляется дискретизация.
Допустим m = [9,2%, 10%,11%,12%,13,6%].
Оптимум оказался в 11%.
В кластере есть 10 товаров:

$$\underline{p}_1 = 100, \underline{p}_2 = 1000, \underline{p}_3 = 2000, \ldots \underline{p}_{10} = 10000$$

$$p_1 = 100(1 + 11\%) = 111$$

$$p_{10} = 10000(1 + 11\%) = 11100$$

Оптимум кластера

Допустим m= [9,2%, 10%,11%,12%,13,6%].
Оптимум оказался в 11%.
В кластере есть 10 товаров:

$$\underline{p}_1 = 100, \underline{p}_2 = 1000, \underline{p}_3 = 2000, \dots \underline{p}_{10} = 10000$$

$$p_1 = 100(1 + 11\%) = 111$$

$$p_{10} = 10000(1 + 11\%) = 11100$$

С одной стороны, вроде бы все в порядке. «Точность» наших бинов = 1%.
Оптимум для 1-ого товара с точностью до 1 рубля. (1%)
Оптимум для 10-го товара с точностью до 100 рублей. (1%)

Но что делать, если, например, конкуренты установят цену на 10-ый товар в 10050? Или еще какие-то внешние факторы требуют большей дискретизации?

Оптимум кластера

Какие варианты решения?

- Большая дискретизация по m (очевидно!)
- Для каждого товара на каждой итерации устанавливать случайную цену внутри бина.
- Дополнительная кластеризация.

Случайная наценка (цена)

Делать бинеризация следующим образом.
Например, 6 бинов с границами:
[9,2%,9,6%), [9.6%,10.4%), [10.4%,11,5%),
[11.5%,12.3%),[12,3,13%), [13%,13.6%)

В случае реализации бина, брать для каждого товара значение m случайно из, например, нормального распределения с мат. ожиданием, соответствующим центру отрезка и небольшой дисперсией.

Равномерное распределение на практике работает чуть хуже.

Случайная наценка (цена)

Плюсы:

- Дополнительная устойчивость кластера. Т.е. если на оценку сильно влияет один товар, то за счет рандомизации можно тестировать разные точки на другие товары.
- Дополнительная статистика.
- Возможный дополнительный (но случайный) учет дискретизации

Минусы:

- В одной и той же точке – всегда разные цены.
- Неудобно для покупателя.
- Дополнительные затраты и итерации.

Зачем вообще кластеризация?

Для динамического ценообразования нужно:

1. Похожие товары (чтобы лучше учитывать статистику)
2. Товары, на которые общий спрос (они канибализируют друг друга)
3. Якорный товар и его кластер
4. И заодно научиться выделять KVI ☺

Для каждой задачи – своя кластеризация.

Зачем вообще кластеризация?

Для динамического ценообразования нужно:
1. Похожие товары (чтобы лучше учитывать статистику)
2. Товары, на которые общий спрос (они канибализируют друг друга)
3. Якорный товар и его кластер
4. И заодно научиться выделять KVI ☺

Для каждой задачи – своя кластеризация.

Т.е. получается, что один товар может лежать сразу в нескольких кластерах.

Один товар и 2 кластера.

Допустим, есть какой-то товар который находится как минимум в двух кластерах.

Например, есть общая кластеризация.
И отдельная кластеризация, которая объединяет это товар как «якорь» с сопутствующими.

Мы решаем эти две задачи независимо:

$\boldsymbol{y_1}(\boldsymbol{m_1}, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot \underline{\boldsymbol{p}}_{\boldsymbol{1}} \cdot \boldsymbol{Q}(\boldsymbol{m_1})\boldsymbol{m_1}$ - задача по первой кластеризации

$\boldsymbol{y_2}(\boldsymbol{m_1}, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot \underline{\boldsymbol{p}}_{\boldsymbol{1}} \cdot \boldsymbol{Q}(\boldsymbol{m_1})\boldsymbol{m_1}$ - задача по второй кластеризации

$$\underset{m_1}{\operatorname{argmax}} E[y_1(m_1, \xi)] = m_1^{*1} = 11\%$$

$$\underset{m_2}{\operatorname{argmax}} E[y_2(m_1, \xi)] = m_1^{*2} = 16\%$$

Один товар и 2 кластера.

Мы решаем эти две задачи независимо:

$$\boldsymbol{y_1(m_1, \xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1) m_1}$$
$$\boldsymbol{y_2(m_1, \xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1) m_1}$$

$$\underset{m_1}{\operatorname{argmax}} E[y_1(m_1, \xi)] = m_1^* = 11\%$$
$$\underset{m_2}{\operatorname{argmax}} E[y_2(m_1, \xi)] = m_1^* = 16\%$$

И что с этим делать? ☺

Один товар и 2 кластера.

Мы решаем эти две задачи независимо:

$$\boldsymbol{y_1(m_1,\xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1)m_1}$$
$$\boldsymbol{y_2(m_1,\xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1)m_1}$$

$$\underset{m_1}{\operatorname{argmax}}\, E[y_1(m_1,\xi)] = m_1^{*1} = 11\%$$
$$\underset{m_2}{\operatorname{argmax}}\, E[y_2(m_1,\xi)] = m_1^{*2} = 16\%$$

И что с этим делать? ☺

Нужно как-то «усреднить» результаты.

Логично это делать учитывая «вес» кластера.

И снова кластеризация

Характеристики кластера

Какие могут быть характеристики кластера?

- Число товаров
- Диапазон по закупочной цене
- Число продаж (шт.) - (скользящее) среднее
- Выручка - (скользящее) среднее
- Прибыль - (скользящее) среднее
- Устойчивость по числу продаж
- Устойчивость по выручке

Устойчивость кластера

Под устойчивостью можно понимать:

- Максимальную амплитуду (или интерквантильный размах) или дисперсия в какой-то точке или бине. Желательно, разумеется, в оптимальном, бине.
- Изменение выручки (прибыли) при изменение цены. Другими, словами эластичность по цене. [см. лекцию №1]

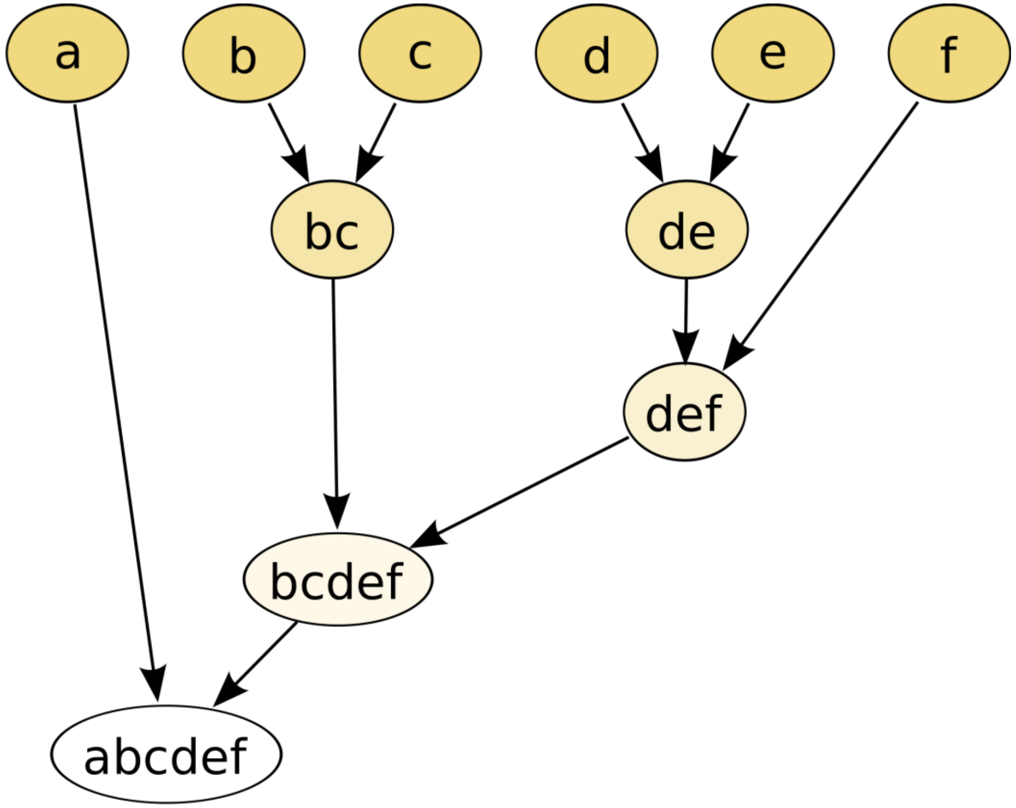
Чуть позже мы обсудим, что все это мы нормируем на спрос и внешние факторы.

Характеристики кластера

С точки зрения практики:
Чем больше выручка (или прибыль) – тем важнее кластер для бизнеса.
Чем устойчивее он – тем больше «доверяем» оценке (которая является следствием параметрической модели).

Иерархическая кластеризация

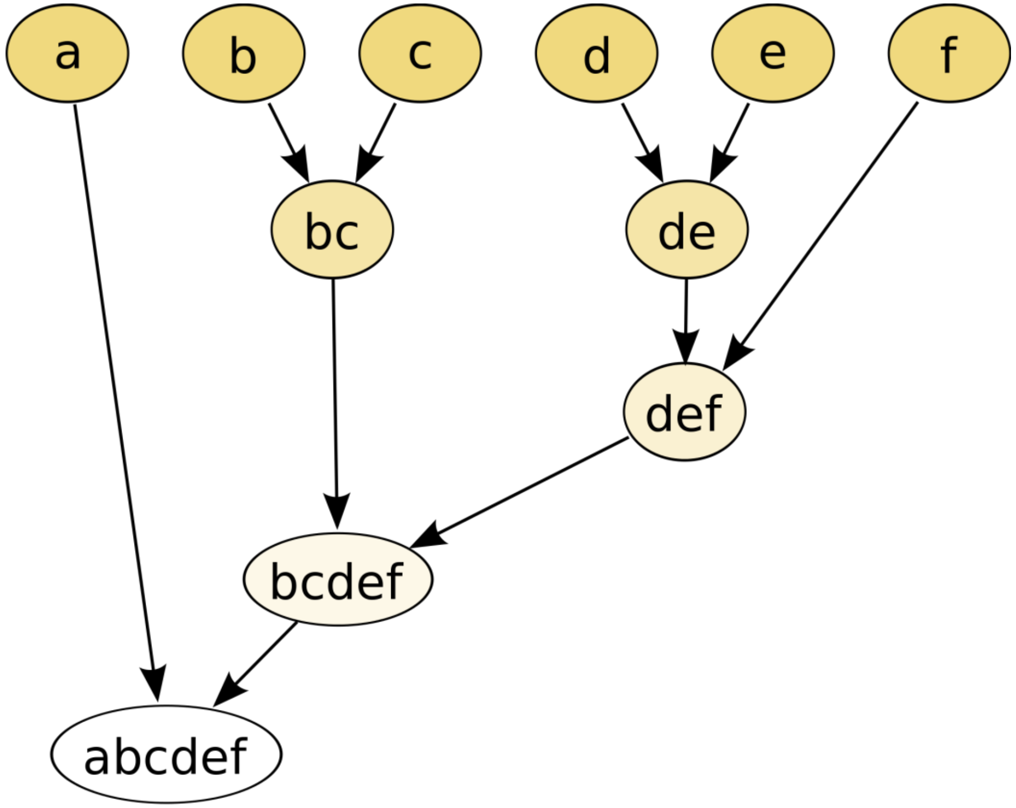
Иерархическая кластеризация— совокупность алгоритмов упорядочивания данных, направленных на создание иерархии (дерева) вложенных кластеров.



Иерархическая кластеризация

Если мы строим иерархическую кластеризацию, то на верхнем (на картинке нижем) уровне иерархии у нас все товары.

У этого кластера, очевидно максимальная выручка (и прибыль), но не максимальная устойчивость!



Как «взвесить» кластера?

Нужны какие-то эвристики.
Варианты:

- Прибыль кластера делить на эластичность $\alpha \frac{y(m)}{(\Delta y(m)/\Delta m)}$
- Логарифм прибыли делить на размах кластера $\alpha \frac{ln(y(m))}{max(y(m))-min(y(m))}$
- $\frac{ln(Ny(m))}{max(y(m))-min(y(m))}$, где N число товаров.

Дополнительные приоритеты

Кластера, где товары с общим спросом – должны получать больший приоритет. Т.к. смотри пример 1.

Если с кластером, который агрегирует общий спрос «все хорошо» - то не нужно брать товары из него для доп. кластеризации.

Кластера, которые учитывают сопутствующие товары должны получать дополнительный приоритет.

Что мы можем посчитать заранее

Мы заранее можем оценить кластера по:

- Число товаров
- Диапазон по закупочной цене
- Число продаж (шт.) - (скользящее) среднее
- Выручка - (скользящее) среднее
- Прибыль - (скользящее) среднее

По историческим данным.

Как «оценить» качество кластера?

Цель кластеризации сделать такое разбиение (или несколько), чтобы кластеризация была:

- Устойчивый
- На ней хорошо работала хотя бы одна параметрическая модель (как следствие находим оптимум)
- Участвовал при взвешивании

Кластеризировали - кластеризировали, да не выкластеризировали!

Надо перекластеризовать!

Один товар и 2 кластера.

Возвращаемся к нашему примеру:

$$\boldsymbol{y_1(m_1,\xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1)m_1}$$
$$\boldsymbol{y_2(m_1,\xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1)m_1}$$

$$C_1 \to \underset{m_1}{\operatorname{argmax}} E[y_1(m_1,\xi)] = m_1^{*1} = 11\%$$
$$C_2 \to \underset{m_1}{\operatorname{argmax}} E[y_2(m_1,\xi)] = m_1^{*2} = 16\%$$

Взвешиваем.

$$\frac{W[C_1]m_1^{*1} + W[C_2]m_1^{*2}}{W[C_1] + W[C_2]} = m_1^{*} = 12\%$$

Перекластеризация

На практике мы можем сделать бесконечно много кластеризации. С приемлемым качеством с точки зрения постановки задачи кластеризации.
И один товар может находится в десятках кластеров (или даже в сотнях с учетом иерархий).

Перекластеризация

Что происходит с кластером, если у него низкий вес?

Товары, которые были в этом кластере за счет того, что на них «подействовали» другие кластеры оказались не в той точке, где прогнозируется оптимум по этому кластеру. И более того у всех товаров может оказаться разная наценка. Как в таком случае оценить следующую итерацию по этому кластеру?

Оценка итерации

Когда мы говорим про кластер, мы можем взять итоговые значения p и посчитать среднюю m.

$$\frac{\boldsymbol{p} - \underline{\boldsymbol{p}}}{\underline{\boldsymbol{p}}} = \boldsymbol{margin\ (m)}$$

Но лучше взвесить на продажи:

$$\frac{\mathbf{1}}{\boldsymbol{n}} \sum_{i} \frac{\boldsymbol{n_i}(\boldsymbol{p_i} - \underline{\boldsymbol{p_i}})}{\underline{\boldsymbol{p_i}}} = \boldsymbol{m}$$

где $\boldsymbol{n_i}$ - это продажи i-ого товара в кластере, n – число всех продаж.

Перекластеризация

Таким образом у нас появляются «хорошие» и «плохие» кластера.

Хорошие:
- Часто используются
- Параметризация работает хорошо
- Большая прибыль (выручка)

Плохие:
- Редко используются
- Параметризация не устойчивая

А зачем нам «плохие» кластера?

Таким образом у нас есть тройки:

- Кластер, который определяет перечнем товаров
- Параметризация
- «Вес», который является следствием первых двух пунктов и итераций (возможно в прошлом)

Тогда задачу можно переформулировать. А давайте искать сразу хорошие кластера!

Переобучение

Может показаться, что можно просто подобрать набор товаров и идеальную параметризацию к нему. По историческим данным.

Но это, к сожалению, переобучение!

Переобучение

Как бороться с переобучением.

- Заранее выбрать большое кол-во вариантов разбиения.
- Заранее выбрать несколько параметрических функций.

Случайным образом взять несколько разбиений, присвоить каждому кластеру параметрическую функцию.

Один товар и 2 кластера.

Возвращаемся к нашему примеру:

$$\boldsymbol{y_1(m_1,\xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1)m_1}$$
$$\boldsymbol{y_2(m_1,\xi) = \xi \cdot \underline{p}_1 \cdot Q(m_1)m_1}$$

$$C_1 \to \underset{m_1}{\operatorname{argmax}} E[y_1(m_1,\xi)] = m_1^{*1} = 11\%$$
$$C_2 \to \underset{m_2}{\operatorname{argmax}} E[y_2(m_1,\xi)] = m_1^{*2} = 16\%$$

$$\frac{W[C_1]m_1^{*1} + W[C_2]m_1^{*2}}{W[C_1] + W[C_2]} = m_1^*$$

Итоги

1. Один товар может находится сразу во многих кластерах.
Кластеризации могут быть разные. Помимо «смысловой» близости, можно отдельно использовать сегментную близость, например: «эконом», «стандарт», «бизнес», «люкс».

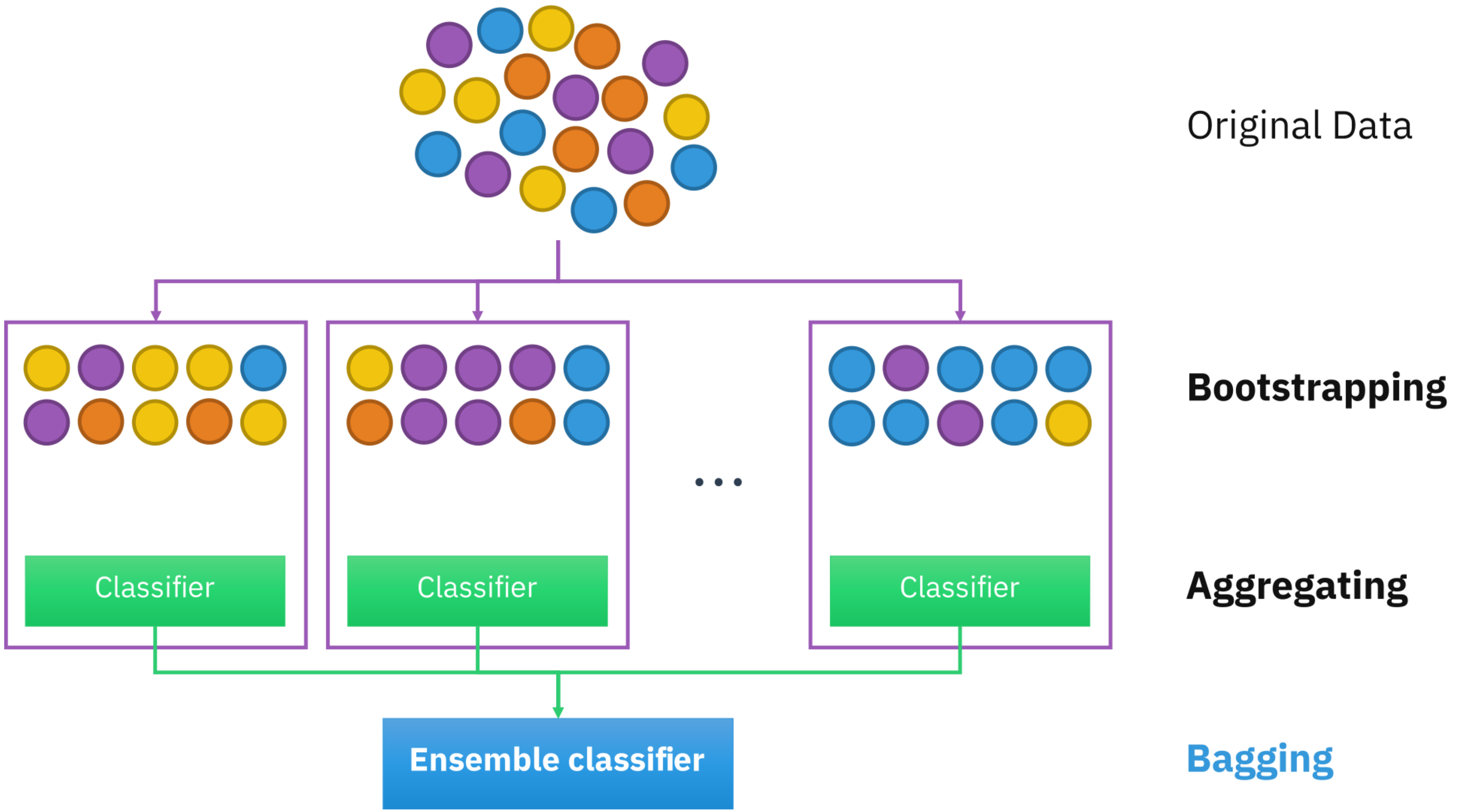
2. Чтобы найти цену для товара на следующую итерацию – нужно взвешивать по кластерам.

3. «Плохие» кластера нужно убирать из расчета и заменять на новые.

4. Наличие большого числа кластеров позволяет решить 2 важные задачи: повысить качество, решить возможную проблему дискретизации.

Bootstrap aggregating

Bootstrap aggregating



Bootstrap aggregating

Идея: разобьём выборку на несколько подвыборок,
для каждой применим алгоритм – усредним результат.

https://habr.com/ru/company/ods/blog/324402/#1-begging

Bootstrap aggregating

Идея: разобьём выборку на несколько подвыборок для каждой применим алгоритм – усредним результат.

Эту идею эксплуатируют:

- RandomFortest
- GradientBoosting

https://habr.com/ru/company/ods/blog/324402/#1-begging

Здесь тоже самое на уровне идеи

Каждый кластер – это «подвыборка».
Для каждого кластера решается своя задача регрессии (в рамках параметрической задачи).

Рассчитывается усредненный (взвешенный) оптимум и происходит итерация в нем.

https://habr.com/ru/company/ods/blog/324402/#1-begging

Общий алгоритм (семинар)

Задание

Описать общую архитектуру решения.

- Алгоритм
  - Кластеризация
  - Параметризация
  - Поиск оптимума
- Какие нужно предусмотреть модули?
- Какие алерты (предупреждения) сделать?
- Как компоненты будут отличать для разных постановок задач?

Какие проблемы могут быть?

Формальная постановка задачи

Максимизировать прибыль $\boldsymbol{y}(\boldsymbol{p}, \boldsymbol{\xi})$ по цене $\boldsymbol{p} \in [\boldsymbol{p}, +\infty)$.

4 разные постановки:

1. $p^* = \underset{p}{\operatorname{argmax}} E[y(p_n, \xi)]$ за n итераций.
2. $p^* = \underset{p}{\operatorname{argmax}} E[y(p, \xi)]$ за минимальное время.
3. $\max_p \sum E[y(p, \xi)]$ Оптимизация траектории.
4. Найти зависимость $y(p)$

$E[\cdot]$ - может быть любым оператором. Можно считать, как мат. ожидание.

Решение

1. Эмбедер
Вход: товары, выход: эмбединги
2. Кластеризация
Вход:
- получает эмбединги
- ограничения на кластера

Выход: кластер
3. Оптимизатор
Вход:
- Параметрическую функцию, шум
- Тип задачи
- Данные: (y, m) по всему кластеру

Выход: m* (или последовательность)

4. Выбор m «взвешивание».
Вход:
- Функция веса
- Кластера
- История по кластеру (y, m)
- m по каждому кластеру
- Справочник

Выход: m* для итерации

Справочник:
Минимально и максимально допустимые цены

Решение

5. Мониторинги

- Кол-во продаж (выручка) упала на X%
- Кол-во продаж (выручка) выросло на Y%
- Скорость продаж резко выросла
- Сток близится к 0.
- Товар (кластер) в прогнозном оптимуме показывает существенно худший результат, чем в известной точке. Выводы:
    - Параметрическая плохая
    - Товар зависимый

Решение

Кластеризацию (много разных)

6. Большой цикл

Оптимизацию -> точку для итерации

Оценку качества кластеров -> заменили «плохие» кластеры на новые

Оценку качества параметризации –> меняем модель параметризации

Смотрим мониторинги –> выводим из ДЦ товары которые попали в мониторги