

Machine Learning

Part II

Задачи прогнозирования временных
рядов и ML в Retail

Alexey Romanenko (alexromsput@gmail.com)

MIPT, 9 November 2016

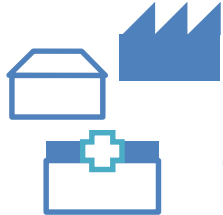
Plan

- Introduction
 - Структура Розничной сети
 - Бизнес KPI компании
- Задачи TS Forecasting и ML в Retail
 - Предобработка данных
 - Прогнозирование при несимметричной функции потерь
 - Предсказание спроса новых товаров
- Contest №3

Структура розничной сети

Объекты

Склады



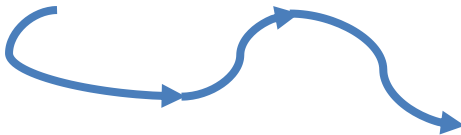
Размещения

Товары

Поставщики и Производители

Физические
ограничения

Пути поставки



Администра
тивные
ограничения

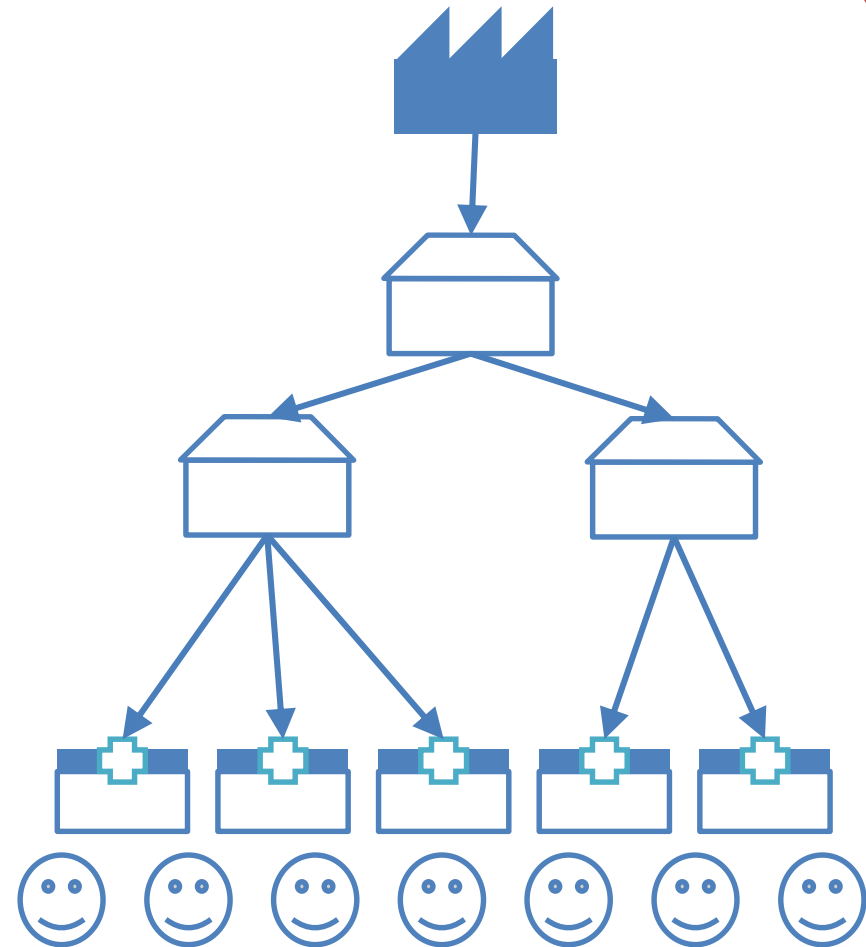
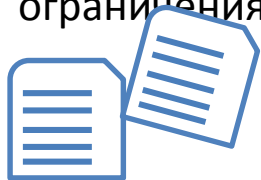
Закупочные цены

Сроки поставки

Ограничения по объемы заказов

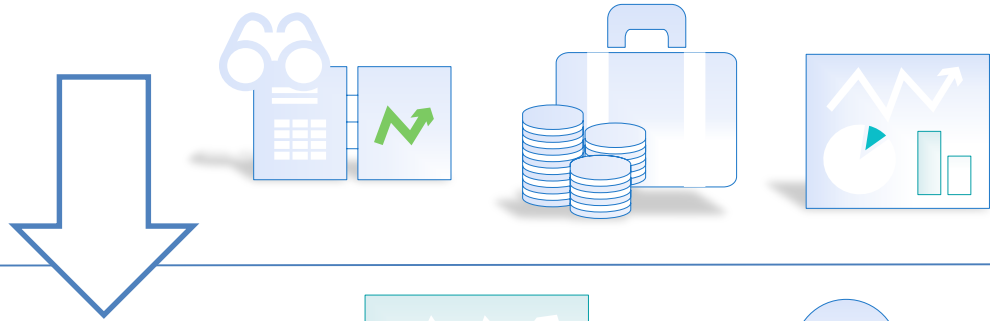
Целевой уровень сервиса

И другие 20-25 типов ограничений

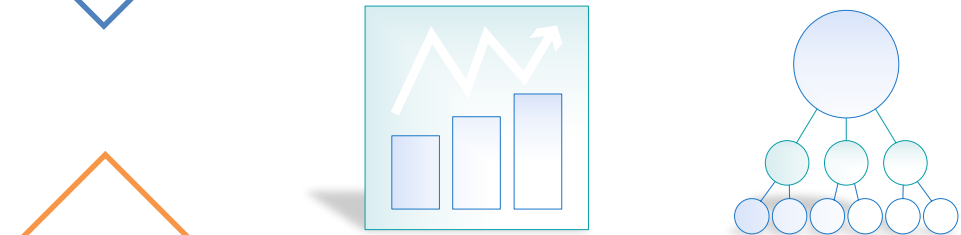


Описание процессов

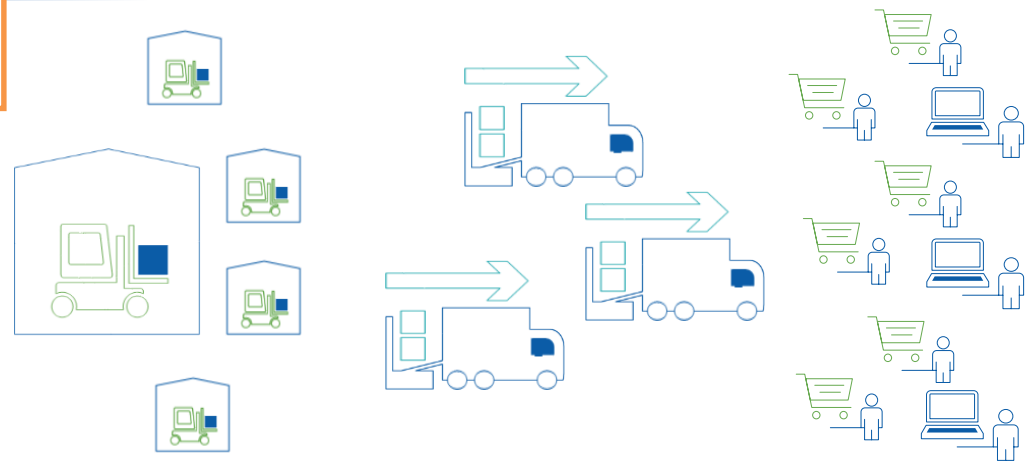
Планирование
Ценообразование
Маркетинг



Прогнозирование спроса
Товародвижение
Оптимизация остатков



Складская
логистика
Товародвижение
Оптимизация
остатков



Ключевая задача

- Ключевая задача в Retail – оптимизация прибыли

$$Profit = q \cdot (p_{sale} - p_{buy}) - cost(q, \delta, \dots)$$

– q – объём продаж;

– δ – стоимость денег (ставка дисконтирования);

– $p_{sale} - p_{buy}$ - наценка на товар;

- $Profit$ - сложно вычисляемая функция
- KPI (отслеживается), пример: оборачиваемость

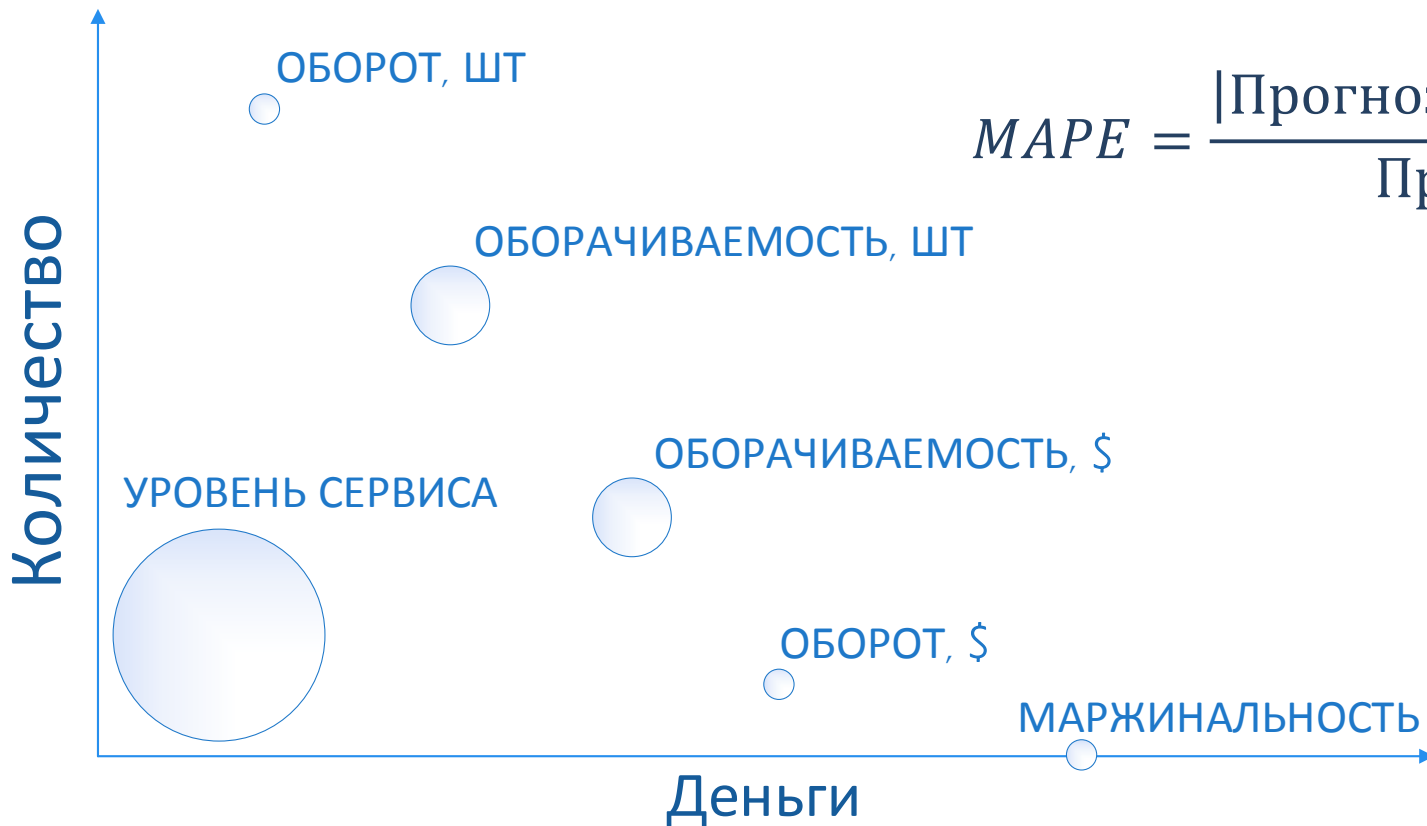
$$\text{Оборачиваемость} = \frac{\text{Продажи}}{\text{Запасы}}$$

KPI Retail

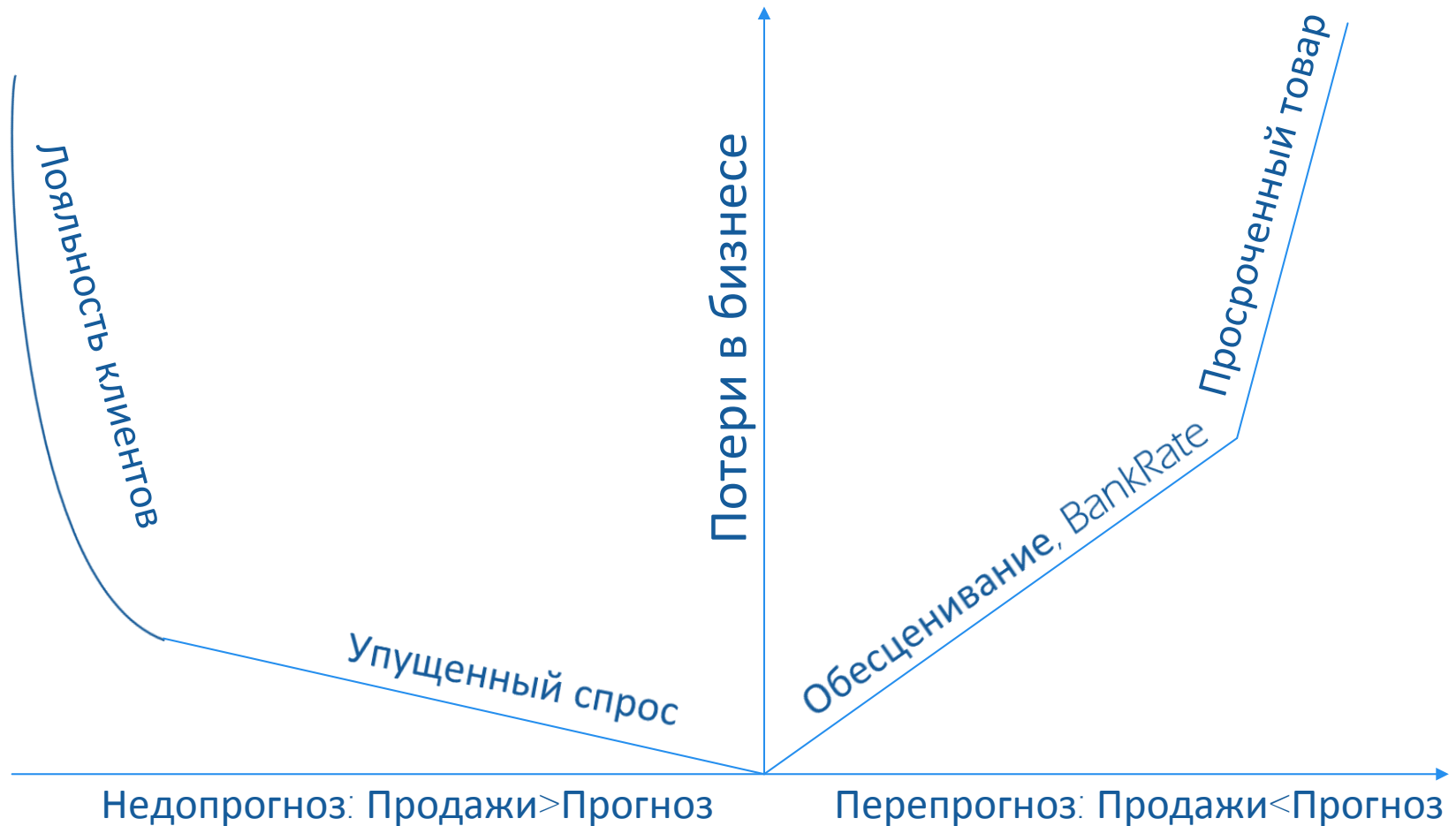
$$MSE = (\text{Прогноз} - \text{Продажи})^2$$

$$MAE = |\text{Прогноз} - \text{Продажи}|$$

$$MAPE = \frac{|\text{Прогноз} - \text{Продажи}|}{\text{Продажи}}$$



Функция потерь при прогнозировании

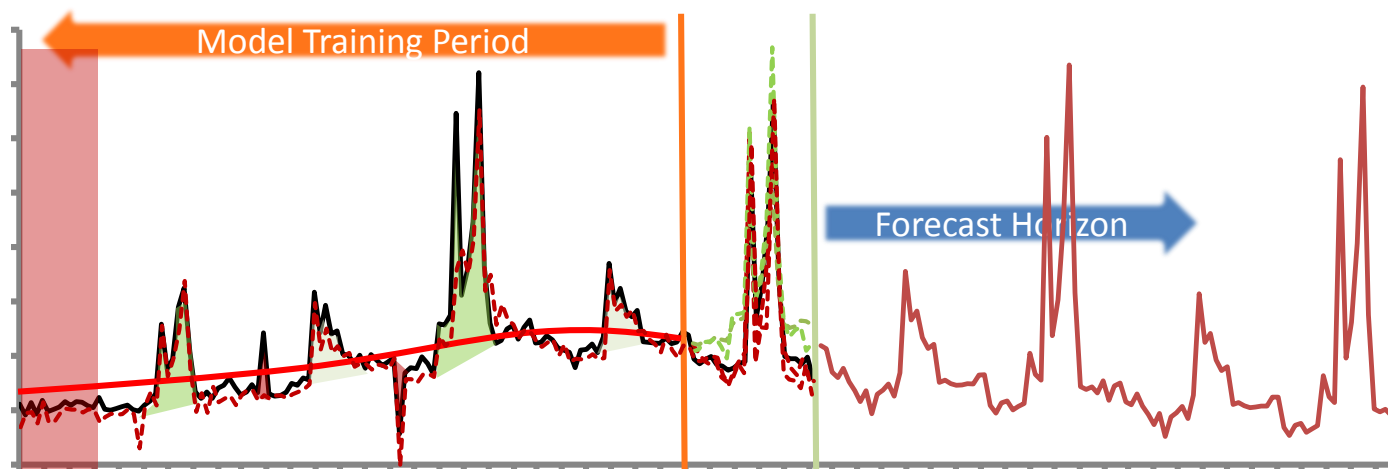


Plan

- Introduction
 - Структура Розничной сети
 - Бизнес KPI компании
- Задачи TS Forecasting и ML в Retail
 - Предобработка данных
 - Прогнозирование при несимметричной функции потерь
 - Предсказание спроса новых товаров
- Contest №3

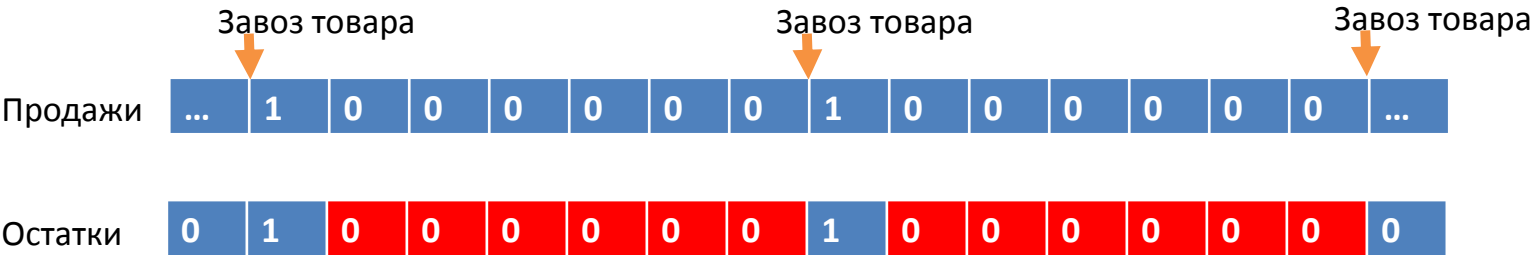
Прогнозирование в ритейл

90% временных рядов прогнозируются по следующей схеме

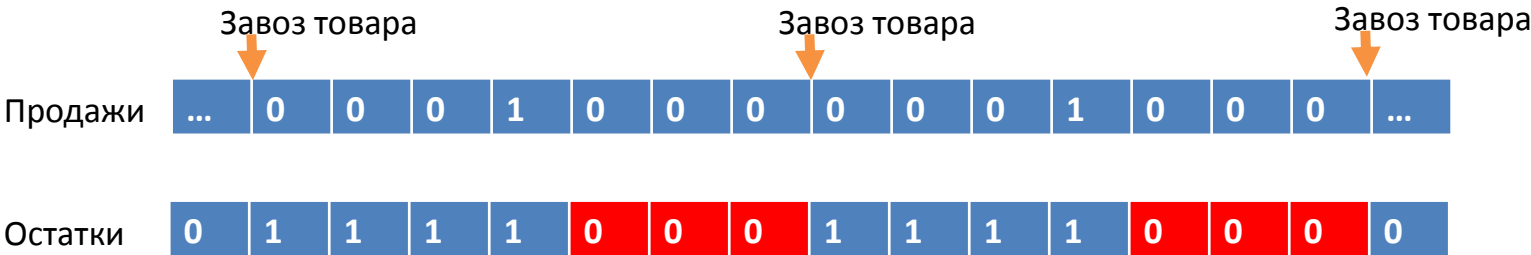


ВОССТАНОВЛЕНИЕ СПРОСА: УЧЁТ ОСТАТКОВ

Товар 1



Товар 2



Пример

ДВА ТОВАРА В ОДНОМ МАГАЗИНЕ

Товар 1

		Завоз товара							Завоз товара								Завоз товара	
Продажи	...	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	...	
Остатки	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	

Товар 2

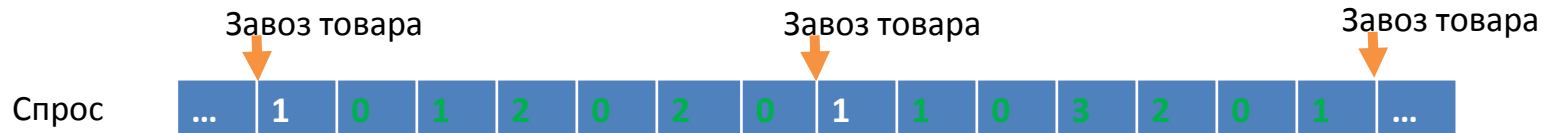
		Завоз товара							Завоз товара								Завоз товара	
Продажи	...	0	0	0	1	0	0	0	0	0	0	1	0	0	0	...		
Остатки	0	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0		

ОСНОВНЫЕ ПРЕДПОЛОЖЕНИЯ, ЗАКЛАДЫВАЕМЫЕ В МОДЕЛЬ

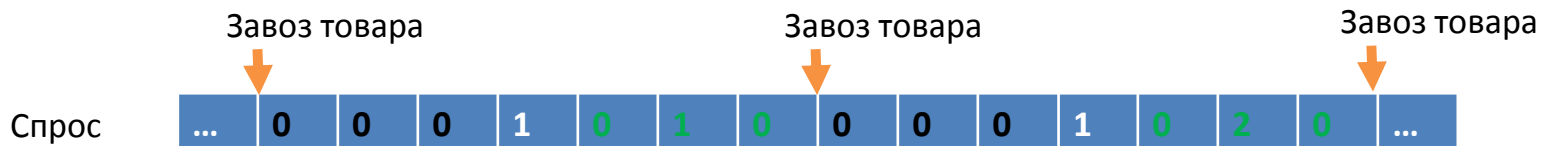
- Количество продаж на разных отрезках времени – НОРСВ
- Вероятность продать k штук за t дней (при бесконечном остатке) равна

$$P_{\infty}(x = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Товар 1: $\lambda = 1$



Товар 2: $\lambda = 0.4231$



ВОССТАНОВЛЕНИЕ СПРОСА: МОДЕЛЬ ПОРОЖДЕНИЯ И ГИБЕЛИ

- m – остаток товара в магазине
- k – количество проданных штук
- $P_m(x = k)$ - вероятность продать k штук при остатке m

$$P_m(x = k) = \begin{cases} \frac{(\lambda)^k}{k!} e^{-\lambda}, & 0 \leq k < m, \\ 0, & k > m \text{ или } k < 0, \\ 1 - \sum_{l=0}^{m-1} \frac{(\lambda)^l}{l!} e^{-\lambda}, & k = m. \end{cases}$$

Нужно найти неизвестный параметр λ :

- требуется большая история данных (так как сочетаний различных m и k – большое)
- найти оптимальный λ можно только приближённо
- необходимо адаптировать λ для каждого участка временного ряда

Процесс порождения и гибели http://sci.alnam.ru/book_otsp.php?id=58

ОЦЕНКА ИНТЕНСИВНОСТИ СПРОСА ММП

- $\sum_{i:k_i>0} k_i$ - суммарные продажи товара в данном магазине
- $n_{0 \leq k < m}$ - количество дней, когда были продажи были меньше остатка
- $n_{k=m>0}$ - количество дней, когда были продажи были равны остатку (все распродано)
- m_i - остаток в день i
- m - максимальный остаток на всю историю

$$\ln L = \ln \lambda \cdot \sum_{i:k_i>0} k_i - \lambda \cdot n_{0 \leq k < m} + \sum_{i: k_i=m_i>0} \left(1 - \sum_{l=1}^{m_i} \frac{(\lambda)^l}{l!} e^{-\lambda} \right) \rightarrow \max_{\lambda>0}$$

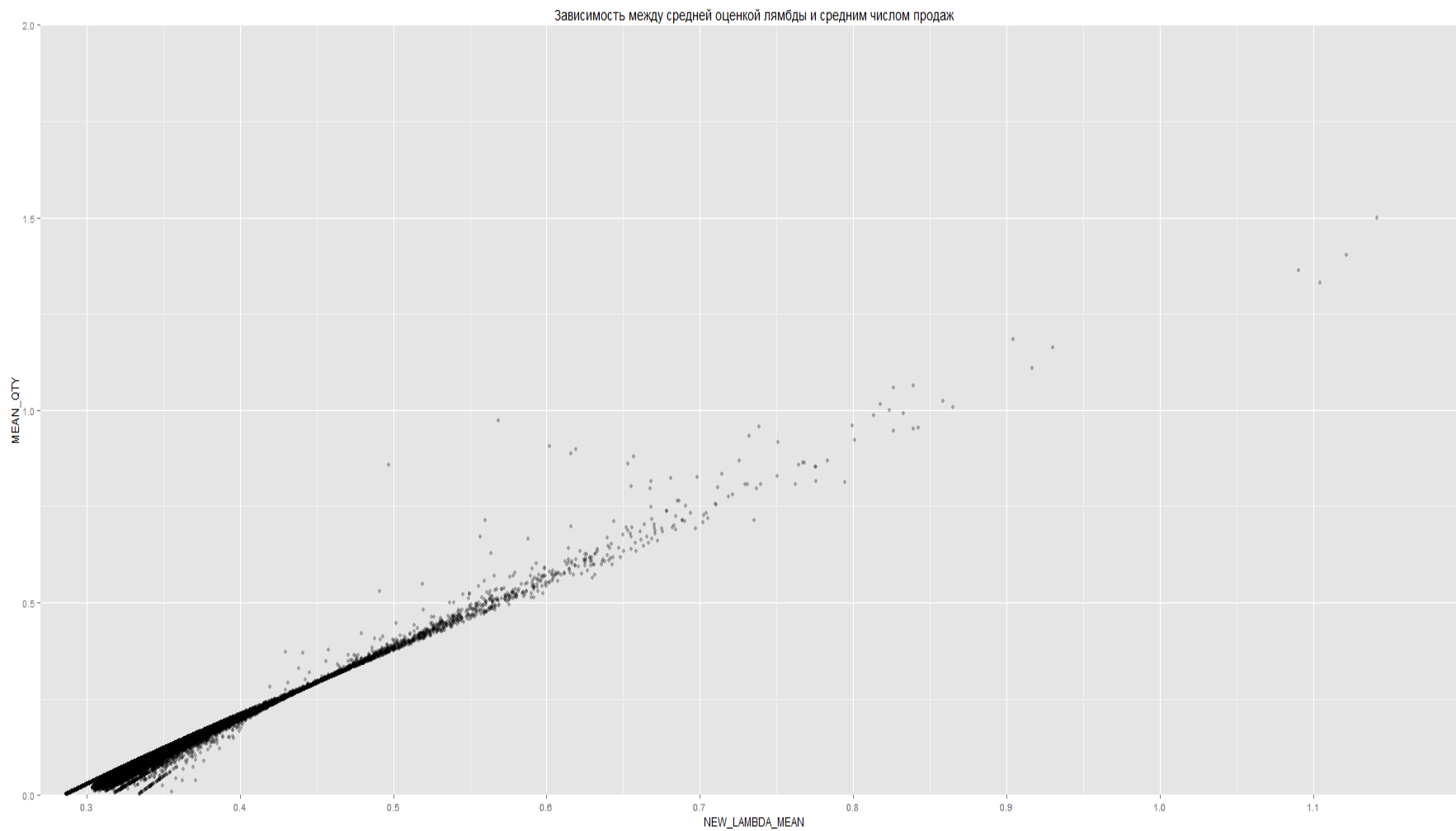
Решение:

$$\lambda = \frac{\sum k_i}{n_{0 \leq k < m} + \alpha \cdot n_{k=m>0}},$$

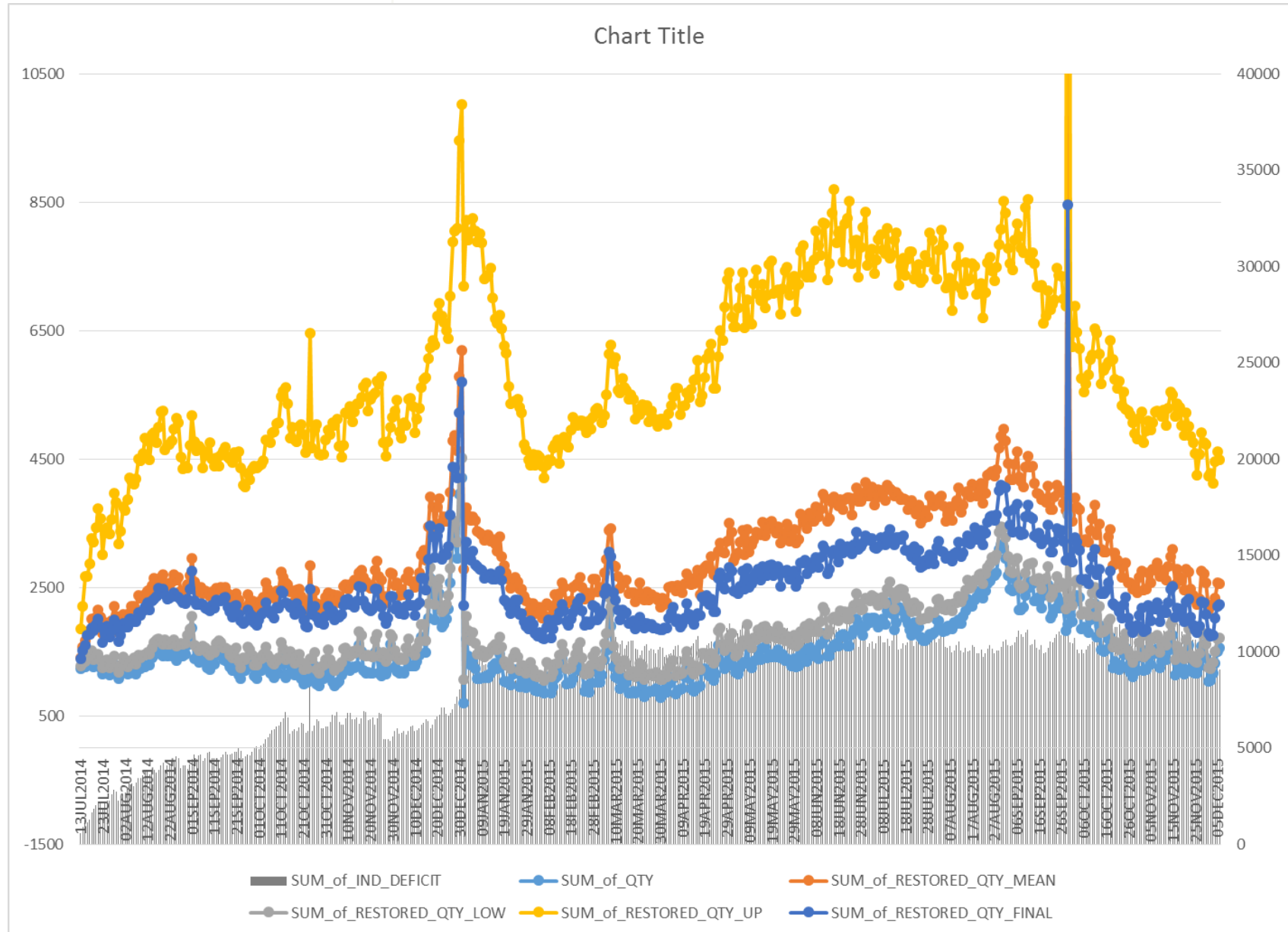
где

$$\alpha \in [0,1]$$

ОЦЕНКА ИНТЕНСИВНОСТИ СПРОСА



РЕЗУЛЬТАТ ВОССТАНОВЛЕНИЯ СПРОСА



Резюме по предобработке данных

- Без восстановления спроса прогноз окажется существенно заниженным
- Разметка дефицитных дней – непростая задача (учёт сроков годности, количества на полке и т.д.)
- Разметка и учёт других событий (промо, события на макро уровне) – задача ML

Plan

- Introduction
 - Структура Розничной сети
 - Бизнес KPI компании
- Задачи TS Forecasting и ML в Retail
 - Предобработка данных
 - Прогнозирование при несимметричной функции потерь
 - Предсказание спроса новых товаров
- Contest №3

Учёт несимметричности потерь

- $L(y, \hat{y}) = \begin{cases} k_1 |y - \hat{y}|, & y \geq \hat{y}; \\ k_2 |y - \hat{y}|, & y < \hat{y} \end{cases}$ - штраф
- $p(y)$ - априорная плотность распределения спроса
- Оптимальный прогноз (заказ):

$$\hat{y}^* = \operatorname{argmax}_{\hat{y}} \int L(y, \hat{y}) \cdot p(y) dy$$

- Для функции потерь выше

$$\hat{y}^* = \Phi^{-1} \left(\frac{k_1}{k_1 + k_2} \right)$$

(квантиль $\frac{k_1}{k_1 + k_2}$ распределения плотности прогноза)

Оценка плотности распределения

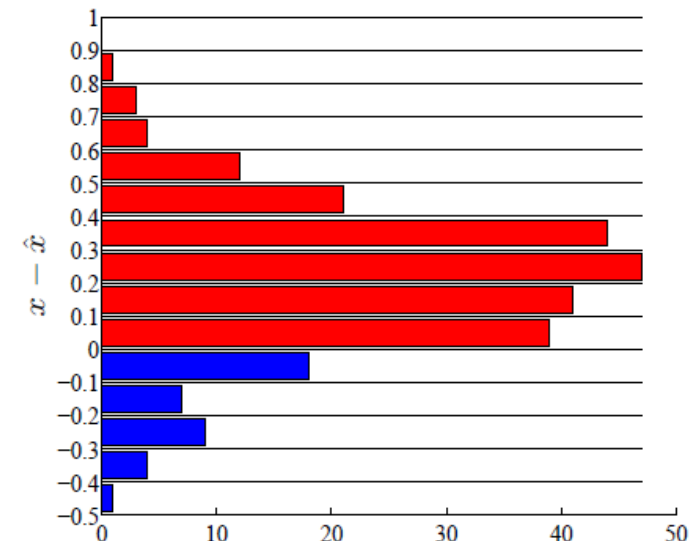
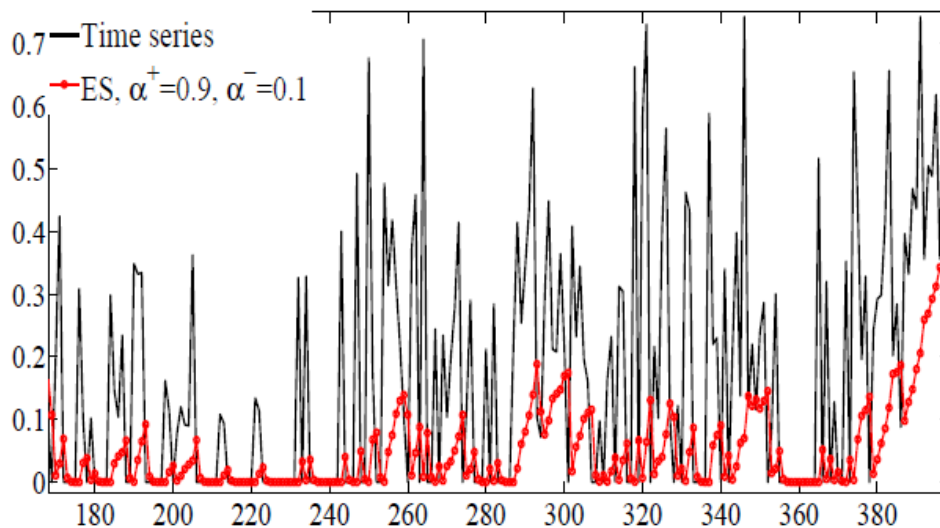
- Непараметрический
 - Прогнозирование эмпирической плотности распределения
 - [Bootstrapping](#)
- Параметрический подход
 - оценка параметров известного распределения

[Density forecasting a Survey, A. S. Tay, \(2000\)](#)

Построение несимметричного прогноза

- Несимметричное ES:

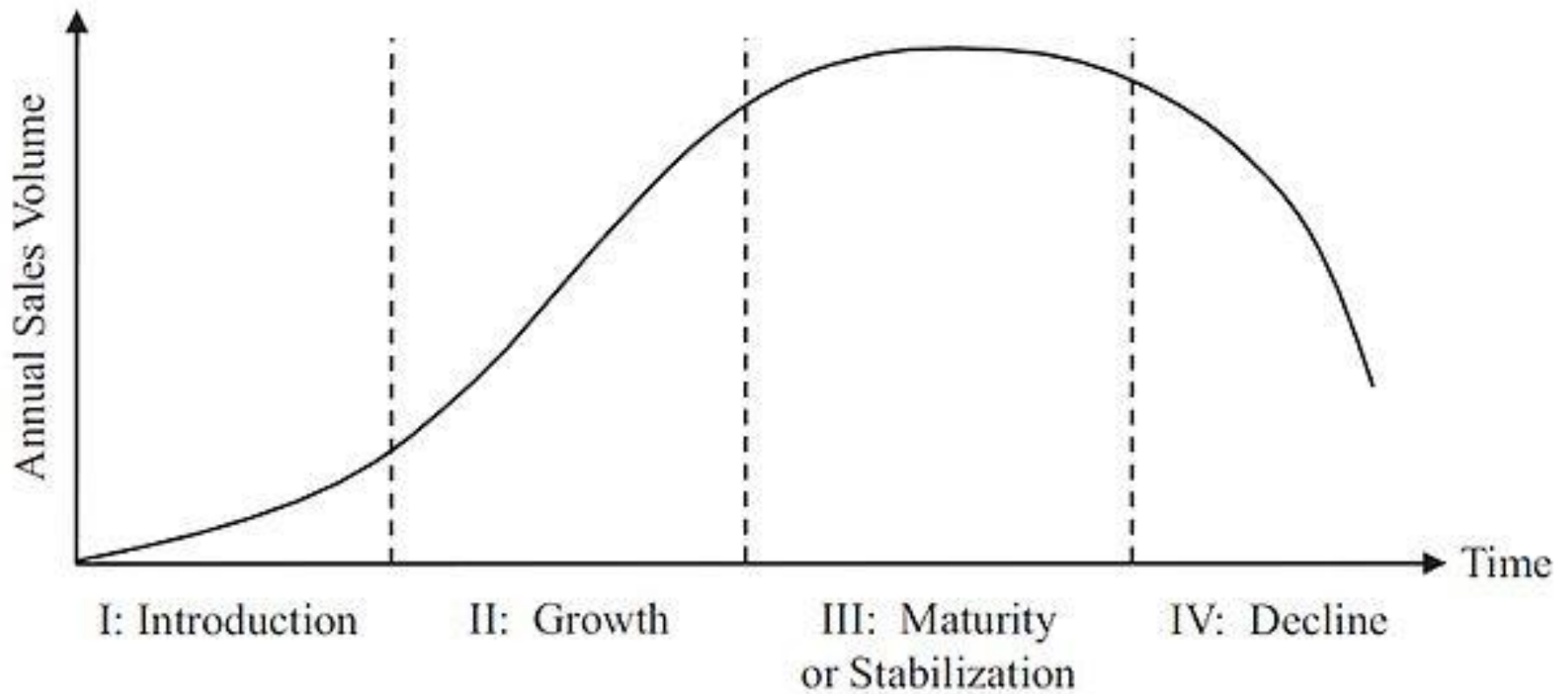
$$\hat{y}_{t+1} = \hat{y}_t + \begin{cases} \alpha^+ \cdot e_t, & e_t \geq 0 \\ \alpha^- \cdot e_t, & e_t < 0 \end{cases}$$



Plan

- Introduction
 - Структура Розничной сети
 - Бизнес KPI компании
- **Задачи TS Forecasting и ML в Retail**
 - Предобработка данных
 - Прогнозирование при несимметричной функции потерь
 - Предсказание спроса новых товаров
- Contest №3

Product Life Cycle Stages



New Product Forecasting

- Demand Shaping
 - [Time series clustering](#)
 - [Time series warping](#)

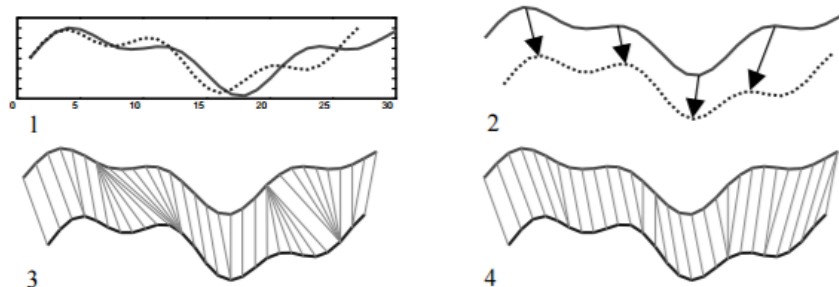


Figure 6: 1) Two artificial signals. 2) The intuitive feature to feature warping alignment. 3) The alignment produced by classic DTW. 4) The alignment produced by DDTW.

- Volume prediction (HW)
 - regression with categorical features

Dynamic Time Warping

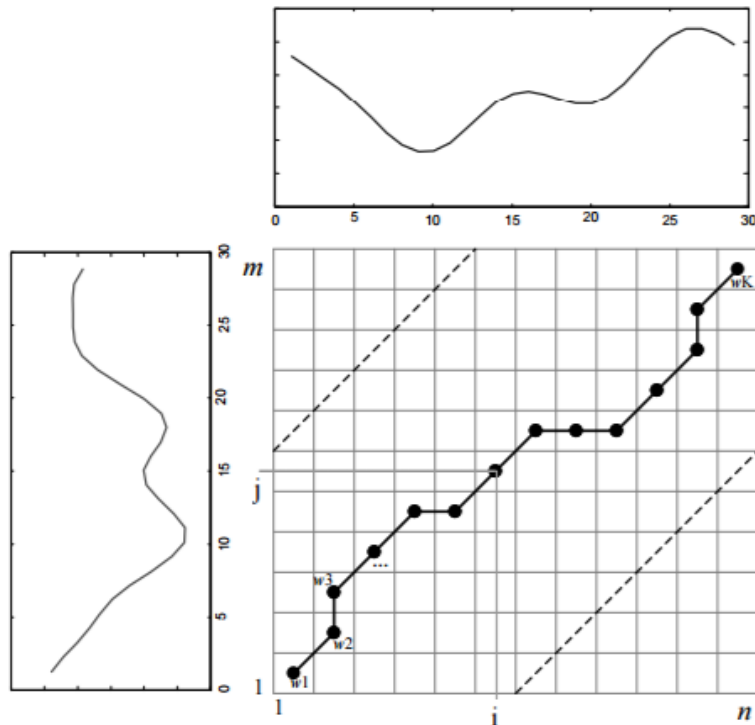


Figure 3: An example warping path.

Derivative Dynamic Time Warping



Figure 4: A pictorial representation of two alternative step-patterns:

- A) The pattern corresponding to $\gamma(i,j) = d(i,j) + \min\{ \gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1) \}$
- B) The pattern corresponding to $\gamma(i,j) = d(i,j) + \min\{ \gamma(i-1,j-1), \gamma(i-1,j-2), \gamma(i-2,j-1) \}$

Заключение

- Прогнозирование занимает ключевую роль в retail
- Для точного прогнозирования требуется тщательная предобработка данных
- Приходится считаться с несимметричными потерями
- Прогнозирование спроса по новым товарам – наименее формализованная задача