

---

AI Ukraine October 2014

# Ритейл с человеческим лицом

Алгоритмы понимания покупателя

---

Роман Захаров, Softcube

---

# Рекомендательная система

---

Автоматический подбор товаров / услуг, основанный на истории поведения покупателя

Хорошая рекомендательная система должна быть неотличима от живого человека, который знает, что нужно покупателю

# Цель рекомендаций

---

- Помочь покупателю найти нужный товар
- Приблизить опыт взаимодействия покупателя с онлайн магазином к живому общению
  
- Продать более дорогой товар
- Продать товар, комплементарный к данному
- Увеличить вовлеченность

# Рекомендации на главной странице



KUN I 72 TIMER

## -20% RABAT\* PÅ SNEAKERS

GÆLDER TIL 26.10.14 - 23.59

MAND KVINDE BØRN

\*FÅ BRANDS UNDTAGET

A promotional banner for sneakers. The background is a dark purple color with a pair of white sneakers. The text is white and centered. It says 'KUN I 72 TIMER' at the top, followed by '-20% RABAT\* PÅ SNEAKERS' in large letters. Below that, it says 'GÆLDER TIL 26.10.14 - 23.59'. At the bottom, there are three buttons labeled 'MAND', 'KVINDE', and 'BØRN'. A small note at the bottom right says '\*FÅ BRANDS UNDTAGET'.



-30% HUMØR  
KØB NU

A promotional banner for clothing. It features a man in a dark jacket sitting in a snowy environment. The text is white and centered. It says '-30% HUMØR' and 'KØB NU'.



MODSTRØM  
KØB NU

A promotional banner for clothing. It features a woman in a light-colored jacket with a fur collar. The text is white and centered. It says 'MODSTRØM' and 'KØB NU'.



MOLO  
KØB NU

A promotional banner for clothing. It features a group of people in winter sports gear, including helmets and goggles. The text is white and centered. It says 'MOLO' and 'KØB NU'.

GRATIS FRAGT

GRATIS FRAGT VED ORDRER OVER 499 DKK

FORTRYDELSESRET

GRATIS OMBYTNING - 30 DAGES BYTTERET

FAKTURA

KØB NU OG BETAL SENERE

# Рекомендации на главной странице



KUN I 72 TIMER

**-20% RABAT\* PÅ SNEAKERS**

GÆLDER TIL 26.10.14 - 23.59

MAND KVINDE BØRN

\*FÅ BRANDS UNDTAGET



-30% HUMØR  
KØB NU



MODSTRØM  
KØB NU



MOLO  
KØB NU

**GRATIS FRAGT**  
GRATIS FRAGT VED ORDRER OVER 499 DKK

**FORTRYDELSESRET**  
GRATIS OMBYTNING - 30 DAGES BYTTERET

**FAKTURA**  
KØB NU OG BETAL SENERE

# Рекомендации на странице категории

Strik & Cardigans  
Undertøj & Strømper  
Jakkesæt  
Shorts & Knickers

Udvid

## TILPAS UDVALG

### PRIS

83 DKK 2000 DKK

### KATEGORIER

- Regular Fit Jeans (456 stk)
- Slim Fit Jeans (401 stk)
- Loose Fit (175 stk)

### BRANDS

- G-Star (91 stk)
- Jack & Jones (82 stk)
- Levi's (65 stk)
- Humør (61 stk)
- Solid (41 stk)
- ID Denim (39 stk)
- Selected (36 stk)
- Blend (35 stk)
- Outfitters Nation (31 stk)
- Tiger of Sweden (31 stk)
- Anerkjendt (29 stk)
- Diesel (28 stk)
- Nudie (24 stk)
- Wrangler (24 stk)
- Shine (23 stk)



Tiger of Sweden 'Igg' jeans  
899,75 DKK



Jack & Jones 'Erik Original' jeans  
~~599,75 DKK~~ 269,75 DKK

-55%



Tiger of Sweden 'Pistolero' jeans  
1.199,75 DKK



CHEAPLOADER 'India' twisted jeans  
~~399,75 DKK~~ 379,75 DKK



CHEAPLOADER 'India' twisted jeans  
~~399,75 DKK~~ 379,75 DKK



CHEAPLOADER 'India' twisted jeans  
~~399,75 DKK~~ 379,75 DKK



# Рекомендации на странице категории

Strik & Cardigans  
Undertøj & Strømper  
Jakkesæt  
Jakkensæt  
Shorts & Knickers

Udvid

## TILPAS UDVALG

### PRIS

83 DKK 2000 DKK

### KATEGORIER

- Regular Fit Jeans (456 stk)
- Slim Fit Jeans (401 stk)
- Loose Fit (175 stk)

### BRANDS

- G-Star (91 stk)
- Jack & Jones (82 stk)
- Levi's (65 stk)
- Humør (61 stk)
- Solid (41 stk)
- ID Denim (39 stk)
- Selected (36 stk)
- Blend (35 stk)
- Outfitters Nation (31 stk)
- Tiger of Sweden (31 stk)
- Anerkjendt (29 stk)
- Diesel (28 stk)
- Nudie (24 stk)
- Wrangler (24 stk)
- Shine (23 stk)



Tiger of Sweden 'lggy' jeans  
899,75 DKK



Jack & Jones 'Erik Original' jeans  
~~599,75 DKK~~ 269,75 DKK

-55%



Tiger of Sweden 'Pistolero' jeans  
1.199,75 DKK



CHEAPLOADER 'India' twisted jeans  
~~399,75 DKK~~ 379,75 DKK



CHEAPLOADER 'India' twisted jeans  
~~399,75 DKK~~ 379,75 DKK



CHEAPLOADER 'India' twisted jeans  
~~399,75 DKK~~ 379,75 DKK



# Рекомендации на странице категории

## THE NORTH FACE M OPEN GATE PU HD – SWEATSHIRT – BLÅ

Pris: 499,75 DKK

Størrelse: Vælg størrelse [Vis størrelsesguiden](#)

Farve: blå

Antal 1 **LÆG I INDKØBSKURV** [Tilføj til ønskeliste](#)

M OPEN GATE PU HD

95% bomuld, 4% Elastane Varenr: 244388\_00

Sweatshirt Hætte kænguru lomme Langærmet

**PRAKTISK INFO** VASKEANVISNINGER

- ✓ **Leveringsid** – afsendelse indenfor 24 timer
- ✓ **FORTRYDELSERET** – gratis ombytning - 30 dages returret
- ✓ **Gratis levering** - Gratis fragt på ordre over 499
- ✓ **Factura** Køb nu og betal senere
- ✓ **Klarna Konto** - Betal 30,00 DKK pr. måned for denne vare

• Se alle produkter i kategorien [Sweatshirts](#)

• Flere produkter fra [The North Face](#)

• Se alle [Tøj](#)

ANDRE KUNDER HAR OGSÅ KIGGET PÅ:

-  899,75 DKK 829,75 DKK
-  899,75 DKK 549,75 DKK
-  799,75 DKK

ANBEFALEDE VARER FRA SAMME MÆRKE

-  899,75 DKK 419,83 DKK
-  499,75 DKK 399,75 DKK
-  1.199,75 DKK 999,75 DKK
-  249,75 DKK 127,84 DKK
-  1.999,75 DKK
-  1.499,75 DKK 1.349,75 DKK
-  1.499,75 DKK 299,93 DKK

# Рекомендации на странице категории

**THE NORTH FACE M OPEN GATE PU HD – SWEATSHIRT – BLÅ**

Pris: 499,75 DKK

Størrelse: Vælg størrelse Vis størrelsesguiden

Farve: blå

Antal: 1 **LÆG I INDKØBSKURV** Tilføj til ønskeliste

M OPEN GATE PU HD

95% bomuld, 4% Elastane Varenr: 244388\_00

Sweatshirt Hætte kænguru lomme Langærmet

**PRAKTISK INFO** VASKEANVISNINGER

- ✓ **Leveringsid** – afsendelse indenfor 24 timer
- ✓ **FORTRYDELSERSET** – gratis ombytning - 30 dages returret
- ✓ **Gratis levering** - Gratis fragt på ordre over 499
- ✓ **Factura** Køb nu og betal senere
- ✓ **Klarna Konto** - Betal 30,00 DKK pr. måned for denne vare

• Se alle produkter i kategorien **Sweatshirts**

• Flere produkter fra **The North Face**

• Se alle **Tag**

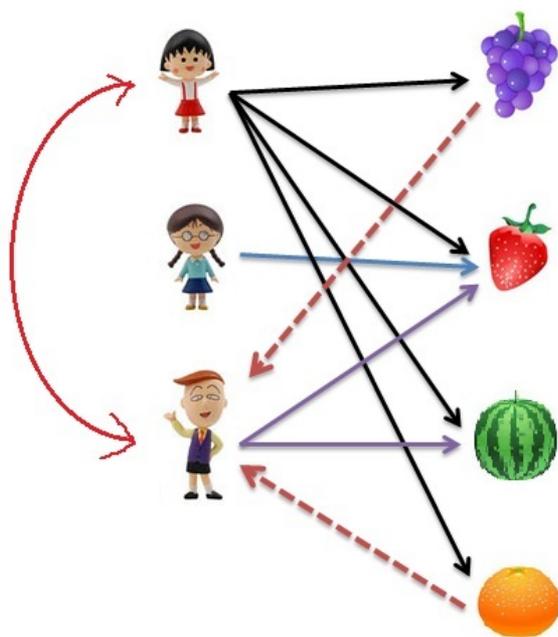
ANBEFALEDE VARER FRA SAMME MÆRKE

Andre kunder har også kigget på:

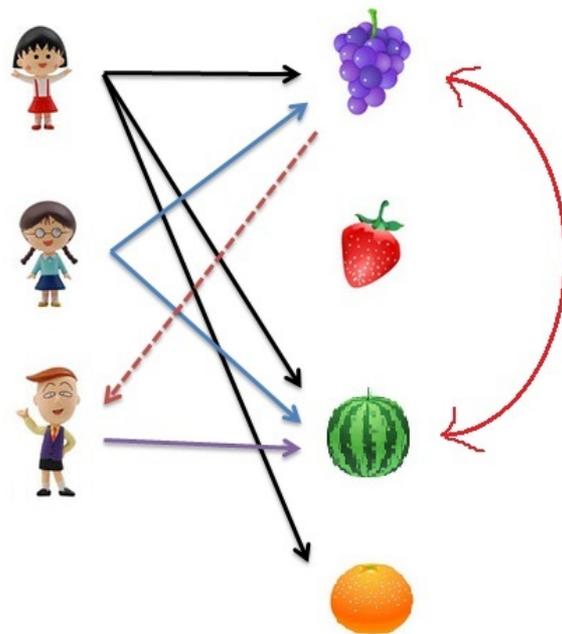
- 899,75 DKK / 629,75 DKK
- 899,75 DKK / 549,75 DKK
- 799,75 DKK
- 899,75 DKK / 419,83 DKK
- 499,75 DKK / 399,75 DKK
- 1.199,75 DKK / 999,75 DKK
- 249,75 DKK / 127,84 DKK
- 1.999,75 DKK
- 1.499,75 DKK / 1.349,75 DKK

# Коллаборативная фильтрация

---



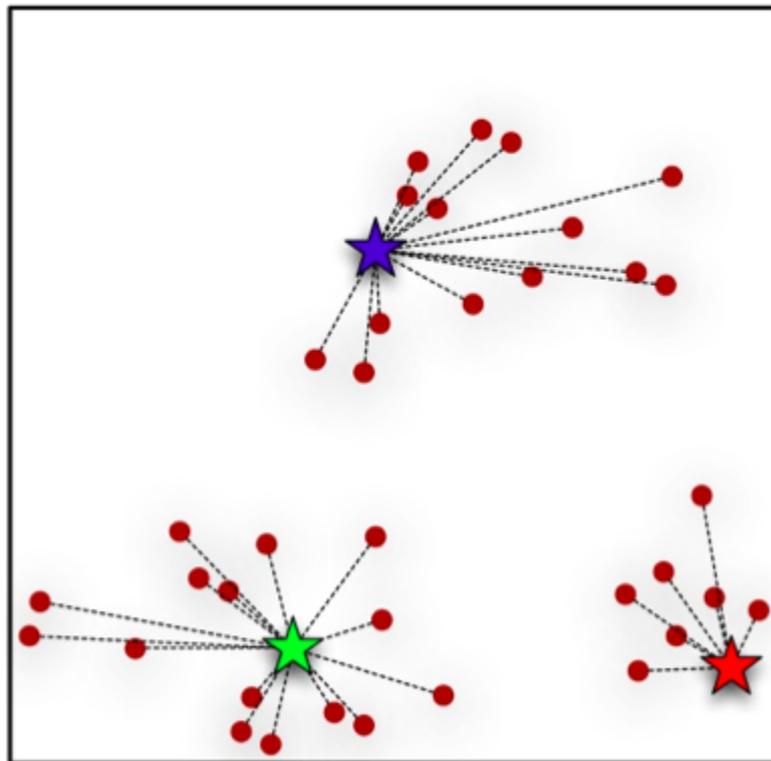
человек-человек



товар-товар

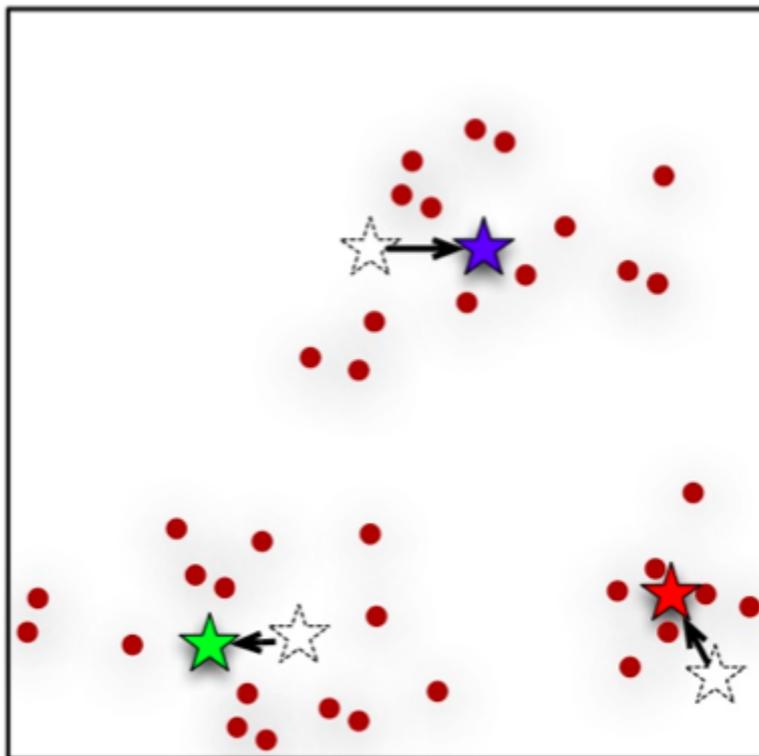
# Метод ближайших соседей

---



# Метод ближайших соседей

---



# Недостатки кластеризации

---

- Огромное количество переменных, как следствие разреженность пространства поиска
- Концентрация на поиске схожести, а не определяющих интересы отличий
- Случайность и сложная оценка достоверности результата

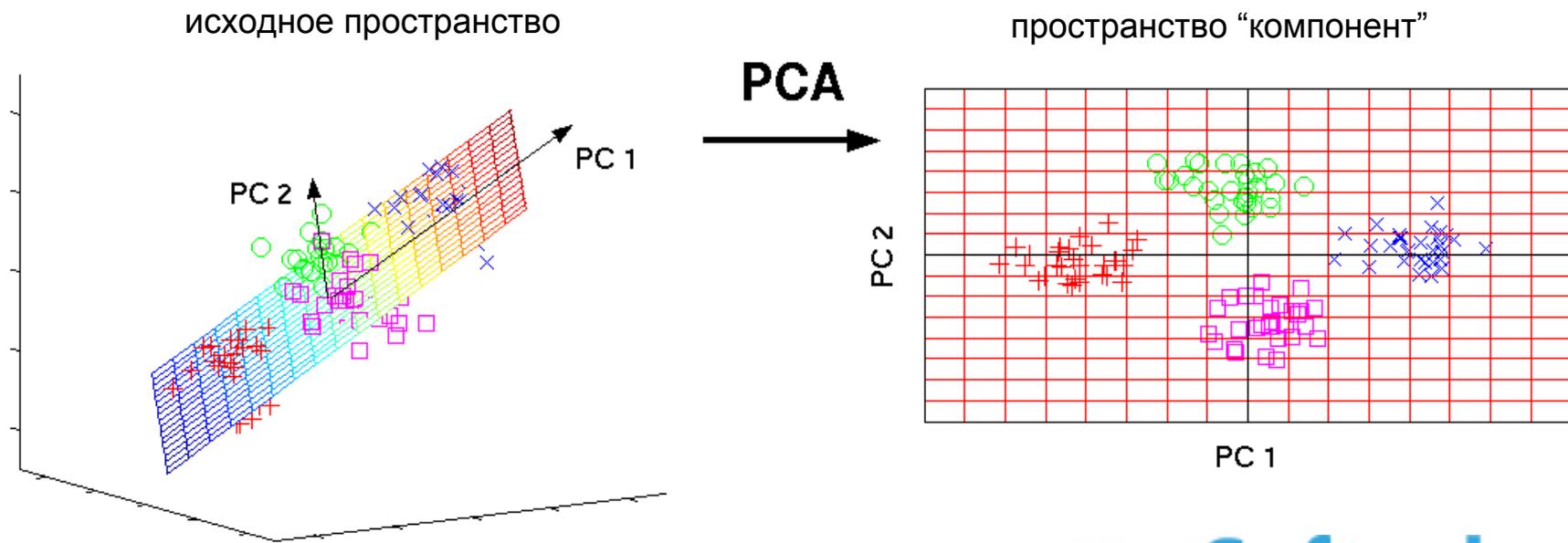
# Формализация предпочтений

---



-2	-1	0	1	2	Мужская одежда
-2	-1	0	1	2	Классическая
-2	-1	0	1	2	Одежда со скидкой
-2	-1	0	1	2	Премиум
-2	-1	0	1	2	Аксессуары
-2	-1	0	1	2	Средний чек

# Метод главных компонент



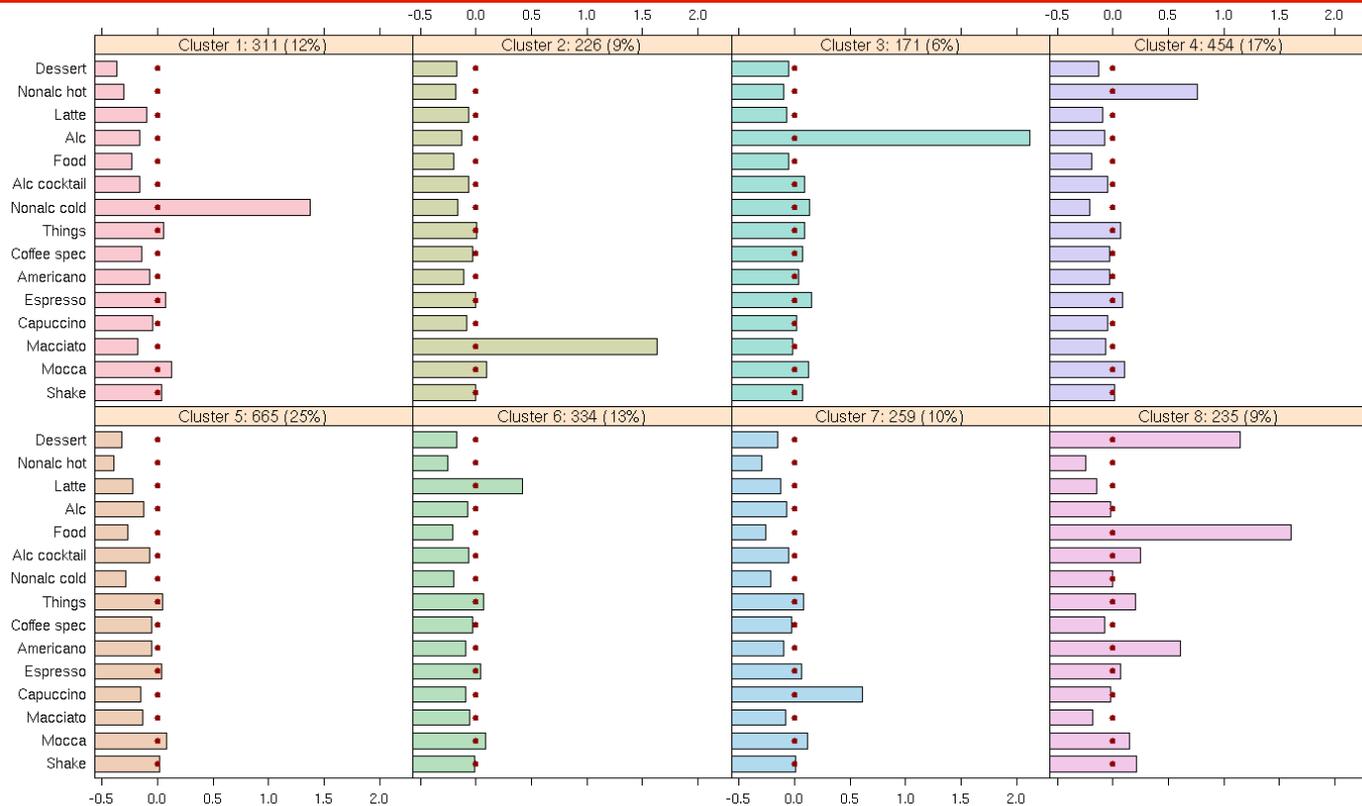
# Портрет покупателя

---



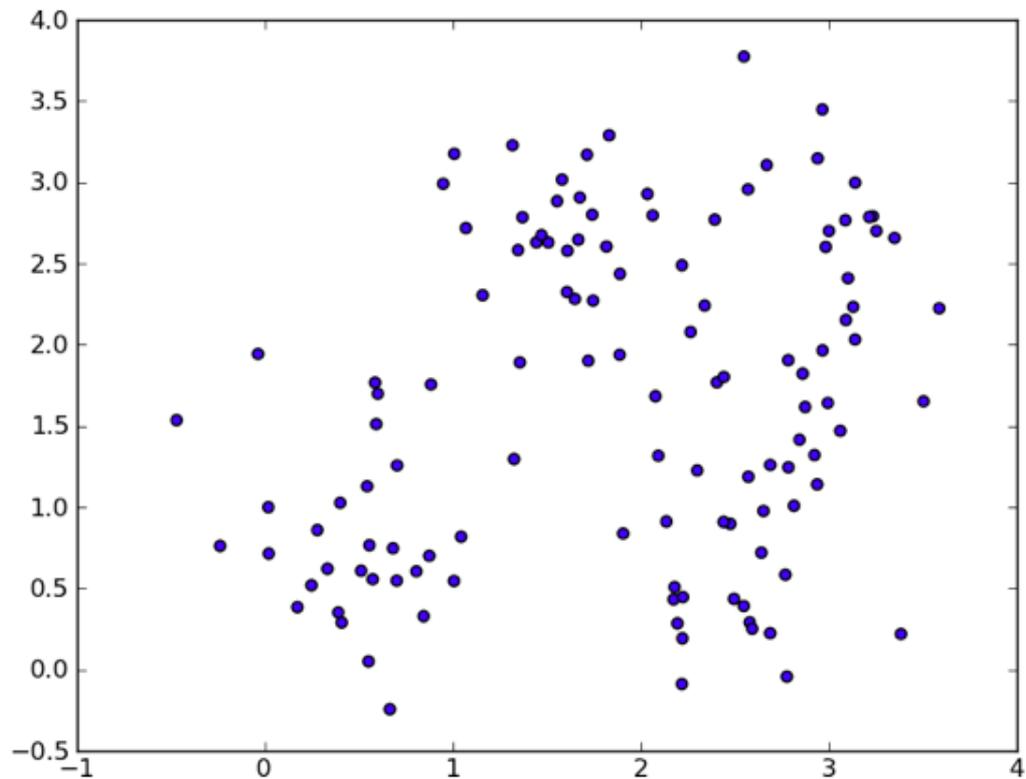


# Отличия определяют сегменты



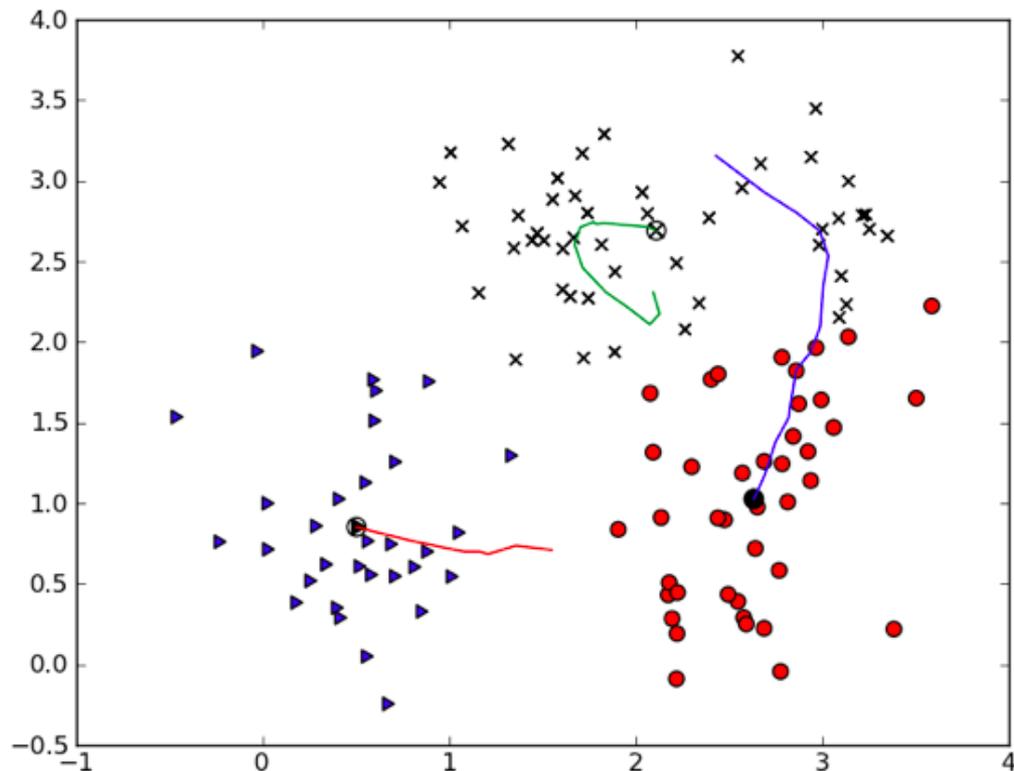
# Метод ближайших соседей

---

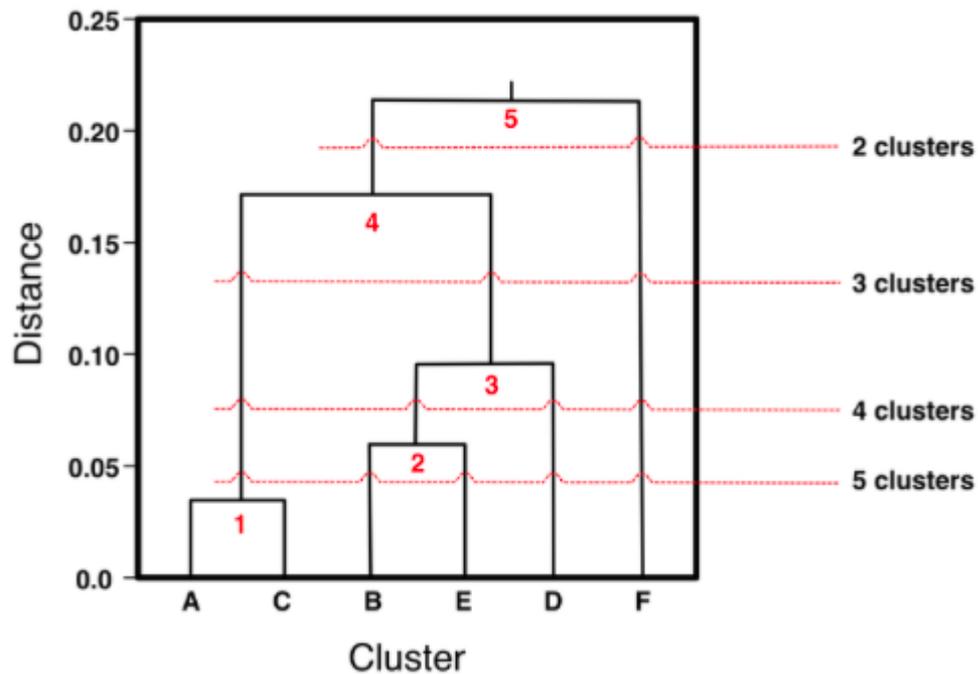
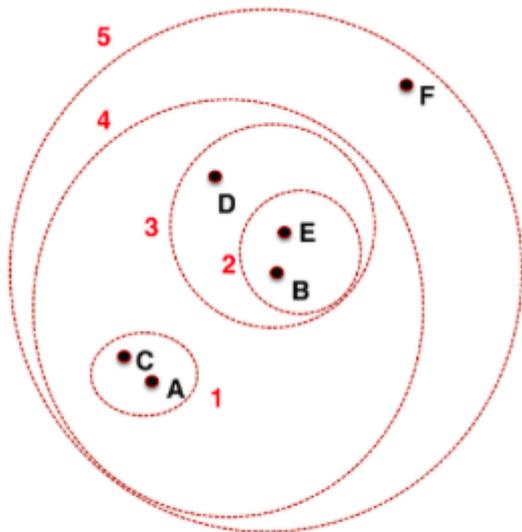


# Метод ближайших соседей

---



# Иерархическая сегментация



# Достоверные сегменты

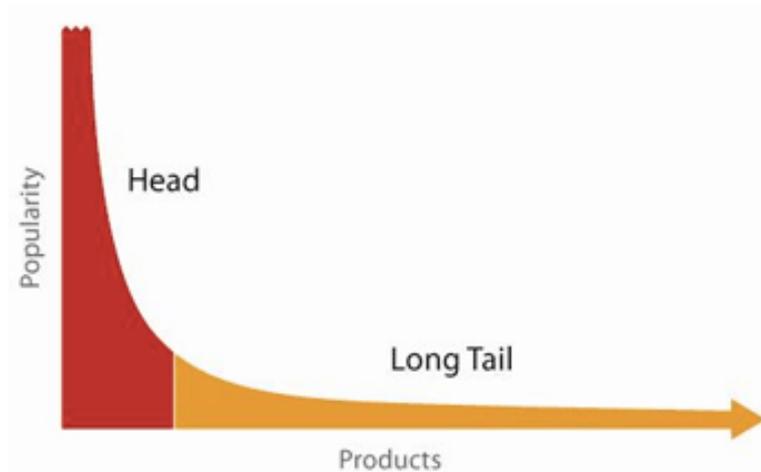
---

- тщательно подбирать количество сегментов
- внутренняя оценка качества сегментации:
  - “сплоченность” точек вокруг центра
  - расстояние между сегментами
- внешняя оценка качества сегментации:
  - обучение на известных примерах
  - экспертная оценка

# Недостатки коллаборативной фильтрации

---

- Проблема «холодного старта»
- Разреженность данных
- Однообразие



**Softcube**  
Making important visible



# Текстовый анализ: TF-IDF

$$\text{TF-IDF}_{(n,d)} = \text{TF}_{(n,d)} \times \text{IDF}_{(n)}$$

Значимость термина

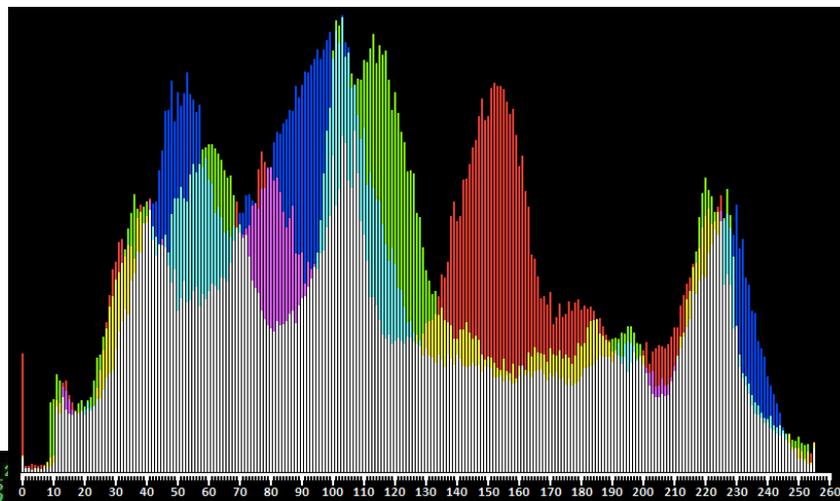
Частота термина n  
в документе d

Уникальность  
термина среди  
документов



Softcube  
Making important visible

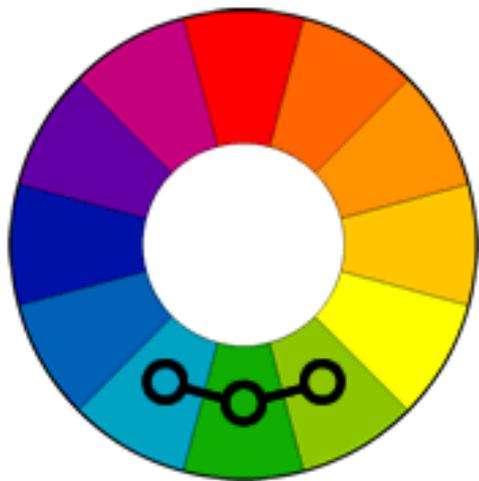
# Анализ изображений



```
5217 499 14552 19481 ;  
8086 19347 1313 10965  
8405 16619 23985 2999  
  
27537 10424 25389 1537 20441 9864 12744 7763 1277 1677 29158 28076 22299 506  
4995 4457 1385 4292 26448 11718 8258 19109 2985 30466 7421 22887 9671 21950  
38916 24671 15159 10919 32163 32056 16266 30036 24032 17906 10831 25201 6539 266  
65  
10558 1310 25173 19867 17078 16274 29094 27109 31050 7940 1257 5517 23127 11979  
29739 6725 18937 1732 5860 21614 19902 6388 27412 3548 31931 622 12946 30169  
10909 25463 31393 23148 19427 745 16837 1755 27211 16597 12245 27502 15754 7276  
5200 29302 15424 20768 7924 9070 14635 20230 1677 3617 24731 4735 4620 24620  
10790 7959 14296 30277 4764 8621 18281 8157 12862 32213 20754 15042 810 19066  
16414 13543 27674 456 23327 23400 25422 26696 7011 6857 15815 28493 9576 5247  
32438 5776 5298 11697 15156 4375 15450 31551 26163 12635 25352 12036 6325 24745  
28645 27224 10812 9673 26282 6617 11649 32478 27134 15980 9576 4749 14334 18302  
1588 31841 21789 25232 22263 14446 22830 25586 1869 29237 4799 11470 19786 15048  
  
29865 12905 8113 27183 485 2726 8295 12297 11169 8652 2855 6006 21907 32136  
31862 565 29838 12590 9434 24330 31571 26460 15431 8035 2473 9335 18906 12346  
4529 9779 5866 31293 26402 2177 19355 11565 28347 1203 19092 23172 23339 2110  
20300 9050 13961 28358 1633 32137 8538 4215 4129 21520 434 1516 9612 1717  
5195 17609 8769 3311 5076 31181 10649 16386 27275 17423 1241 4204 8718 13703
```

# Цветовой круг Иттена

---

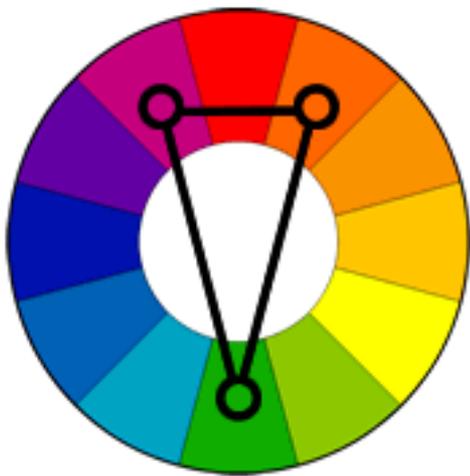


Аналоговая триада



# Цветовой круг Иттена

---



Контрастная триада



# Анализ изображений

Product preview



Vila dress  
54.95 € 16.49 €

Content Design

Product for preview VILA GRINNING – DRESS – GREEN

Number of recommendation - 4 +

Recommendations Preview



Vila dress  
54.95 € 16.49 €



Vila dress  
54.95 € 16.49 €



Vila dress  
54.95 € 16.49 €



Vila dress  
54.95 € 16.49 €



Vila dress  
54.95 € 16.49 €



Vila dress  
54.95 € 16.49 €

Text similarity

Images

# Big Data: чем больше тем лучше

---

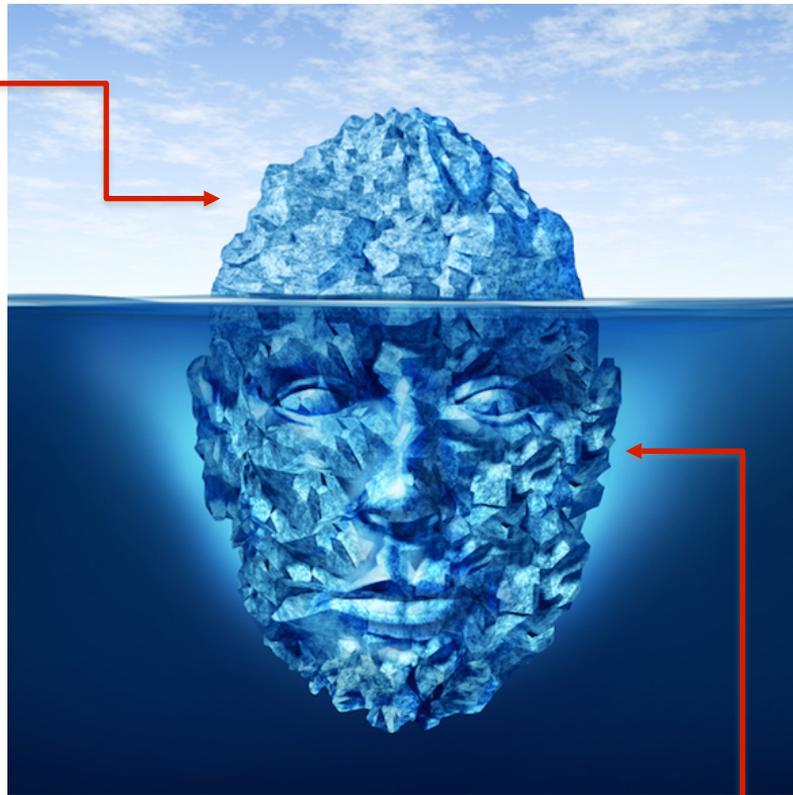
- важен сам объем данных:
- в ширину (немного данных о миллионе пользователей)
- в глубину (много данных о нескольких тысячах пользователей)



# Больше данных

---

- имя
- email
- история заказов



- клики
- переходы
- просмотры
- open rate & CTR в письмах
- история поиска
- etc.

# Товарные рекомендации основанные на кликах via ensemble learning

---

	<i>товар 1</i>	<i>товар 2</i>	<i>...</i>	<i>товар p</i>	<i>факт покупки y</i>
<i>покупатель 1</i>	$x_{11}$	$x_{12}$	<i>...</i>	$x_{1p}$	<i>1</i>
<i>покупатель 2</i>	$x_{21}$	$x_{22}$	<i>...</i>	$x_{2p}$	<i>0</i>
<i>...</i>	<i>...</i>	<i>...</i>	<i>...</i>	<i>...</i>	<i>0</i>
<i>покупатель n</i>	$x_{n1}$	$x_{n2}$	<i>...</i>	$x_{np}$	<i>1</i>

- для каждого товара мало покупателей
  - количество товаров, которые просматривал пользователь измеряется сотнями
  - задача: определить паттерн, ведущий к покупке
-

# Задача классификации

---

Необходимо отобрать небольшое количество товаров, определяющих факт покупки конкретного товара с большой вероятностью

- выдача рекомендаций в риал-тайм требует небольшого количества фильтров
  - можно определять товары, которые понравятся пользователю вне контекста
-

# Задача классификации: критерии

---

- Точность
- Стабильность
- Длина сигнатуры

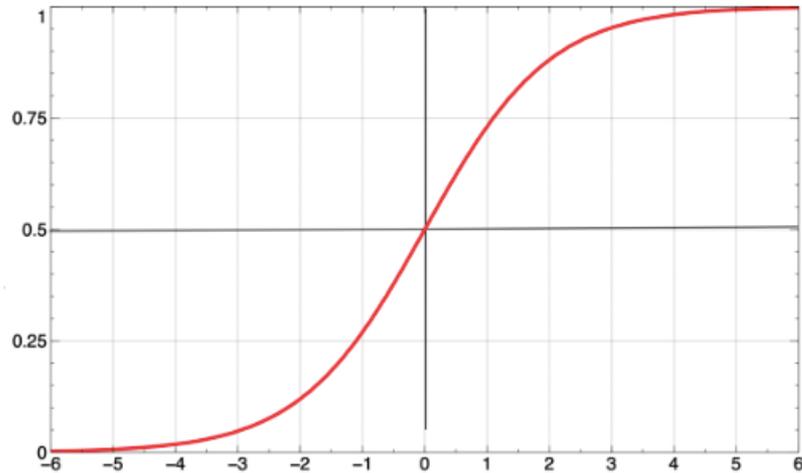
# Логистическая регрессия

---

$$\text{Prob}(y|\mathbf{x}) = \frac{1}{1 + \exp(-y(\mathbf{w}^T \mathbf{x} + v))}$$

$$\min_{\mathbf{w}, v} \frac{1}{n} \sum_{i=1}^n f(y_i(\mathbf{w}^T \mathbf{x}_i + v)) + \lambda \sum_{j=1}^p w_j^2$$

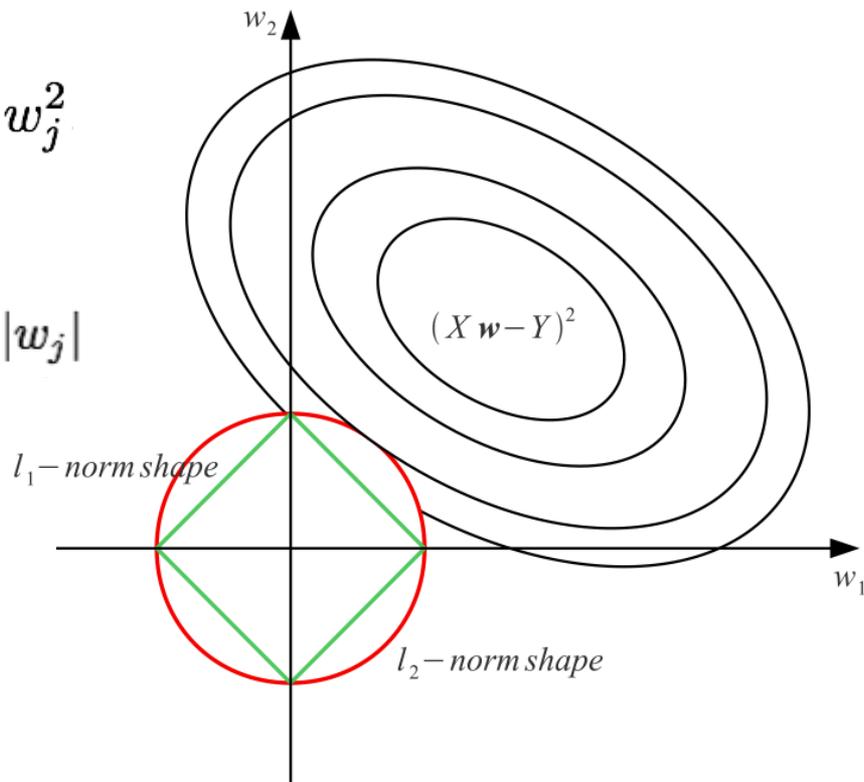
$$f(z) = \log(1 + \exp(-z))$$



# Логистическая регрессия: L1 vs L2

$$\min_{\mathbf{w}, v} \frac{1}{n} \sum_{i=1}^n f(y_i(\mathbf{w}^T \mathbf{x}_i + v)) + \lambda \sum_{j=1}^p w_j^2$$

$$\min_{\mathbf{w}, v} \frac{1}{n} \sum_{i=1}^n f(y_i(\mathbf{w}^T \mathbf{x}_i + v)) + \lambda \sum_{j=1}^p |w_j|$$



# Формализация критериев

---

Точность  $BCR = \frac{1}{2} \left( \frac{TP}{P} + \frac{TN}{N} \right)$

Стабильность  $K(\{\mathbf{S}_1, \dots, \mathbf{S}_k\}) = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^k \frac{|\mathbf{S}_i \cap \mathbf{S}_j| - \frac{s^2}{p}}{s - \frac{s^2}{p}}$

# Много логистических регрессий

---

## Algorithm ELR

**Input:** A learning sample  $\mathbf{X} \in \mathbf{R}^{n \times p}$  and class labels  $\mathbf{y} \in \{-1, 1\}^n$

**Input:** A regularization parameter  $\lambda$  for estimating a  $l_2$ -LR model (3)

**Output:** A vector  $\mathbf{prob} \in [0, 1]^p$  of feature relevance

Initialize  $\mathbf{prob}$

$\overline{BCR} \leftarrow 0.5$  // Default initialization of the average BCR

repeat

    Randomly split  $\mathbf{X}$  into TRAINING (80%) and VALIDATION (20%)

    Draw  $n$  out of  $p$  features at random according to  $\mathbf{prob}$

$(\mathbf{w}, v) \leftarrow$  a  $l_2$ -LR model  $M$  learned on TRAINING restricted to  $n$  features

    Compute BCR of  $M$  on VALIDATION

$quality \leftarrow \log(1 + BCR - \overline{BCR})$

    // Update the feature relevance vector

    foreach  $j$  among the  $n$  sampled features do

$\mathbf{prob}_j \leftarrow \frac{1}{Z} \left( \mathbf{prob}_j + quality \cdot \mathbf{w}_j^{2 \cdot \text{sign}(quality)} \right)$

        //  $Z$  is the normalization constant to define a distribution

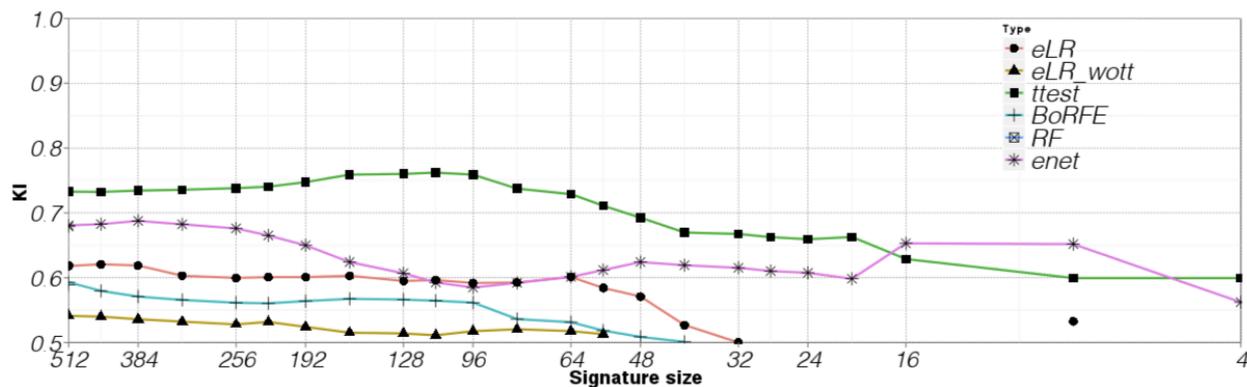
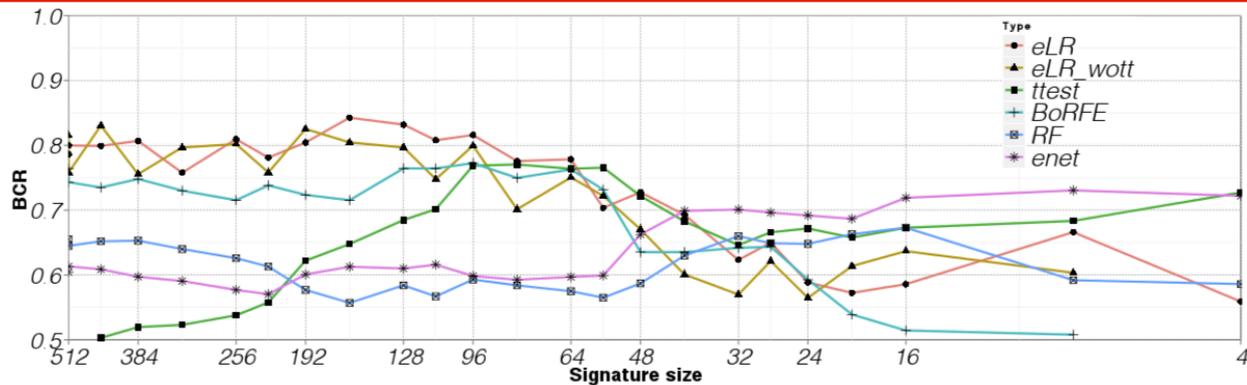
    Update average  $\overline{BCR}$

until no significant change of  $\overline{BCR}$  between consecutive iterations;

return  $\mathbf{prob}$

---

# Результаты



# Спасибо за внимание!

---

[roman.zakharov@softcube.com](mailto:roman.zakharov@softcube.com)  
softcube.com