Морфологический подход к синтезу метрических классификаторов и его реализация методом отыскания минимального разреза графа соседства для обучающей выборки

Ю.В. Визильтер, В.С. Горбацевич, viz@gosniias.ru
ФГУП «ГосНИИАС», Москва

ММРО-2011, г. Петрозаводск

Задача синтеза классификатора при обучении с учителем

Компоненты задачи:

пространство объектов $\mathcal{A}$, *множество классов* $C=\{c_1, \ldots, c_l\}$,

разбиение объектов по классам: $c_{\mathcal{A}}(a)$: $a \in \mathcal{A} \mapsto c \in C$,

пространство описаний (признаков) $\mathcal{X}$,

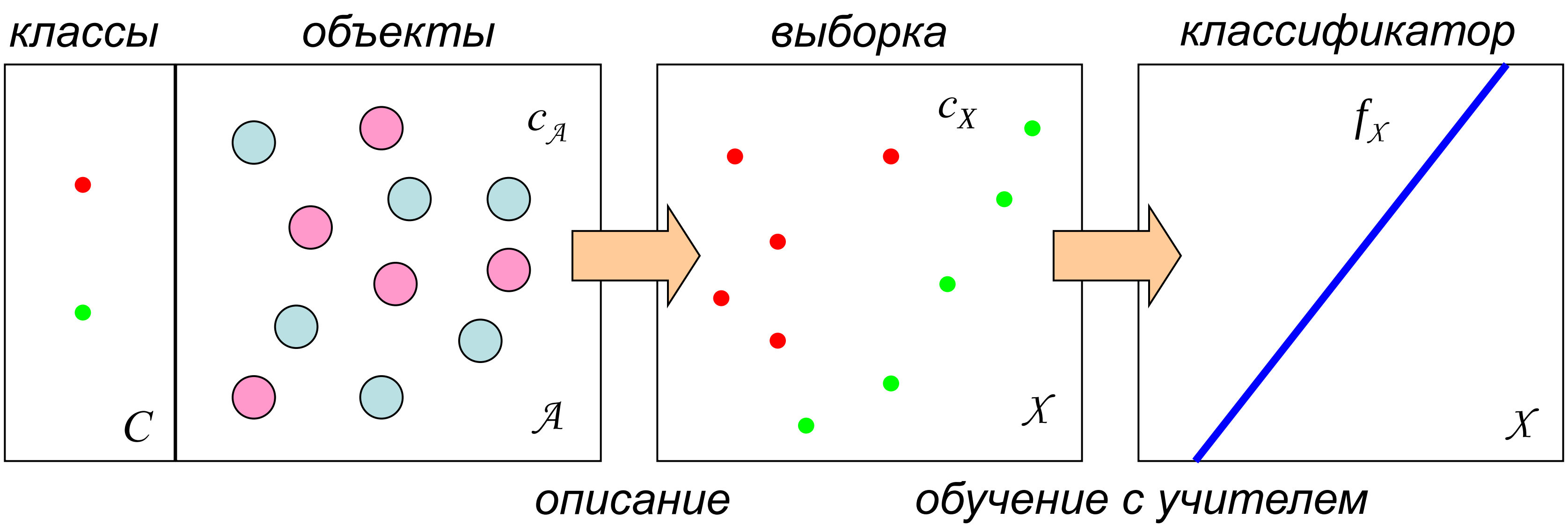
описание объектов признаками: $x_{\mathcal{A}}(a)$: $a \in \mathcal{A} \mapsto x \in \mathcal{X}$.

выборка объектов $A \subseteq \mathcal{A}$, $\|A\| < +\infty$, *выборка описаний* $X \subseteq \mathcal{X}$, $\|X\| < +\infty$.

обучающая выборка: $c_X(x)$: $x_{\mathcal{A}}(a) \in X \mapsto c_{\mathcal{A}}(a) \in C$.

распознающий алгоритм или *классификатор* $f_X(x)$: $x \in \mathcal{X} \mapsto c \in C$.

Требуется: по информации о c_X построить такой $f_X(x)$, который обеспечивает *в некотором смысле наилучшее разбиение* $\mathcal{X}$ на классы из C.



Задача синтеза классификатора при обучении с учителем

Тестовая выборка: $c'_Y(x): x\in Y \mapsto c\in C,\ Y\subseteq \mathrm{X},\ Y\cap \mathcal{X}=\varnothing,\ \|Y\|<+\infty$,

Критерий эмпирического риска на выборке Y: $J_Y(f_{\mathcal{X}}) = d_H(f_Y, c'_Y) / \| Y \|$,

$d_H(f_Y, c'_Y) = \sum_{x\in Y} 1(f(x) \neq c'(x))$ – *число ошибок классификации* на выборке Y.

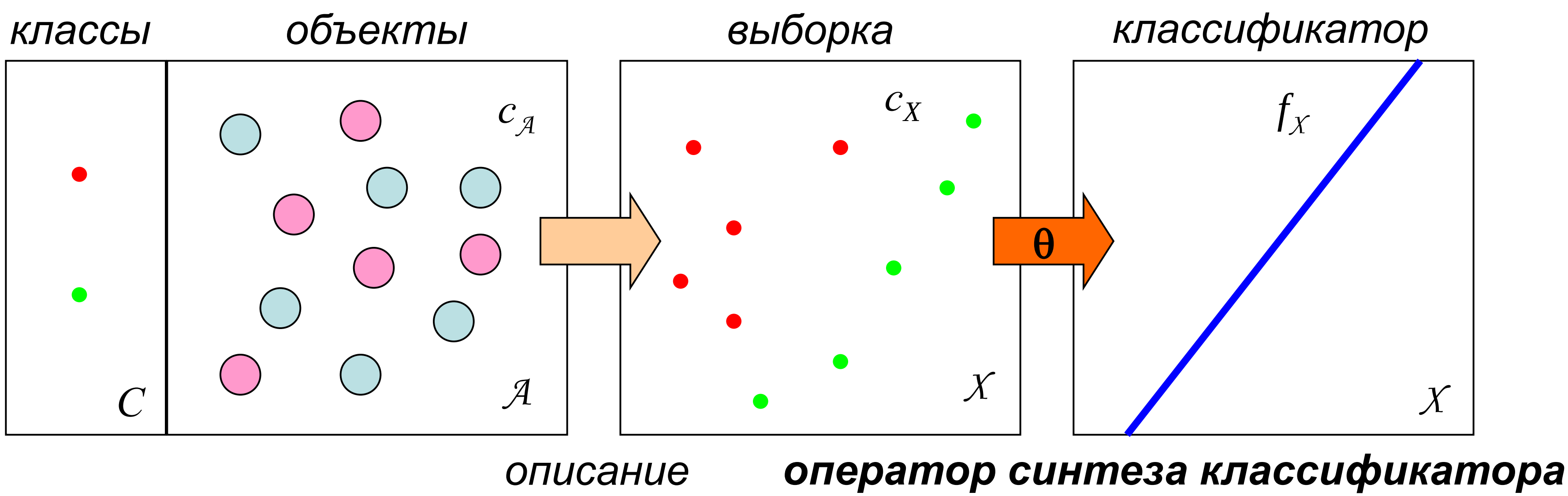
$\| Y \| = \sum_{x\in Y} 1$ – объем выборки Y.

Критерий среднего ожидаемого риска: $J_{\mathcal{X}}(f_{\mathcal{X}}) = E_{Y\subseteq \mathcal{X}}\{J_Y(f_{\mathcal{X}})\}$.

Требуется: Определить ***оператор оптимального синтеза*** θ, доставляющий минимум критерию ожидаемого риска $J_{\mathcal{X}}(f_{\mathcal{X}})$:

$$\theta: c_X\in\Omega_X \mapsto f_{\mathcal{X}}\in\Omega_{\mathcal{X}},\quad \theta = \arg\min_{\theta'} \{J_{\mathcal{X}}(\theta' c_X)\}, \qquad (1)$$

где Ω_X и $\Omega_{\mathcal{X}}$ – *множества всех возможных разбиений* выборки X и пространства $\mathcal{X}$ по классам из C.



Уточнение задачи: обучение классификатора вместо синтеза классификатора

Дополнительные компоненты задачи обучения:

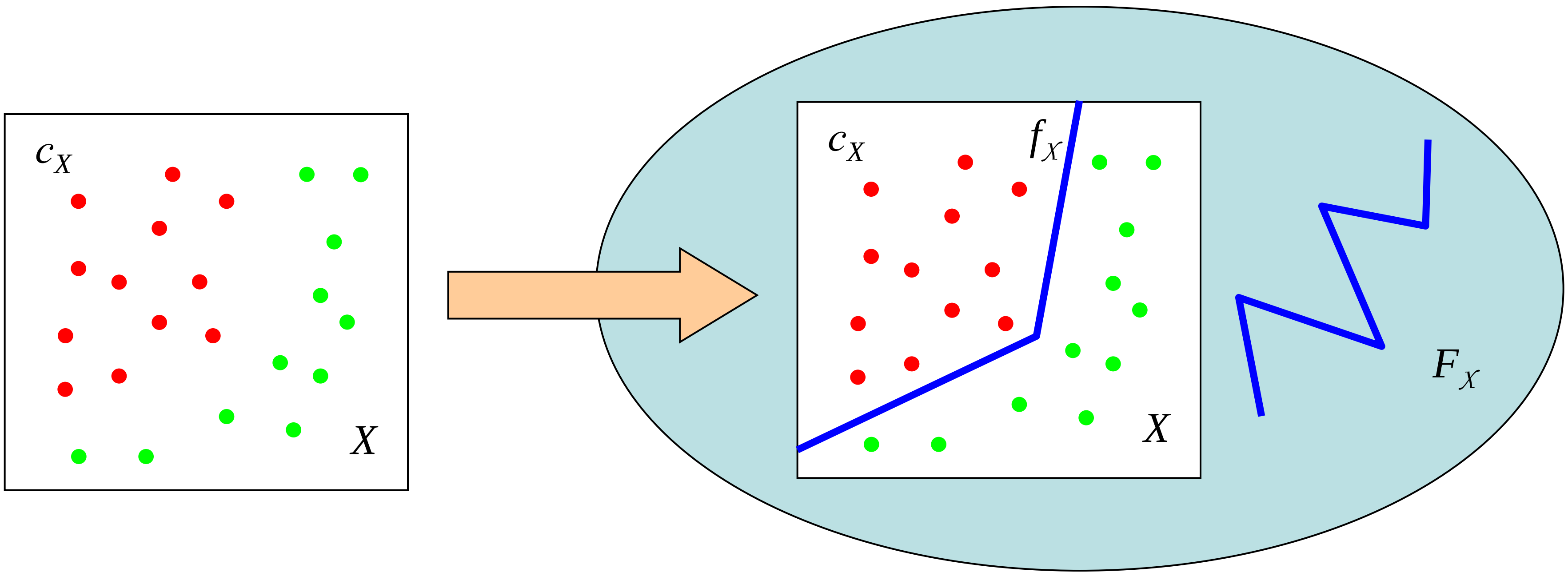
Класс классификаторов $F_{\mathcal{X}}$

Класс алгоритмов обучения Θ для классификаторов из $F_{\mathcal{X}}$ на выборках $X \subseteq \mathcal{X}$.

Оператор обучения:

$$\theta \in \Theta: c_X \in \Omega_X \mapsto f_{\mathcal{X}} \in F_{\mathcal{X}} \subseteq \Omega_{\mathcal{X}},$$

$$\theta = arg\ min_{\theta' \in \Theta} \{J_Y(\theta' c_X)\}, \qquad (2)$$



Уточнение задачи: наблюдаемый риск вместо ожидаемого

Проблема переобучения:

наблюдаемый риск $J_X(\theta c_X)$ имеет глобальный минимум в точке $f_X \equiv c_X$, заведомо непригодный для неизвестной тестовой выборки Y.

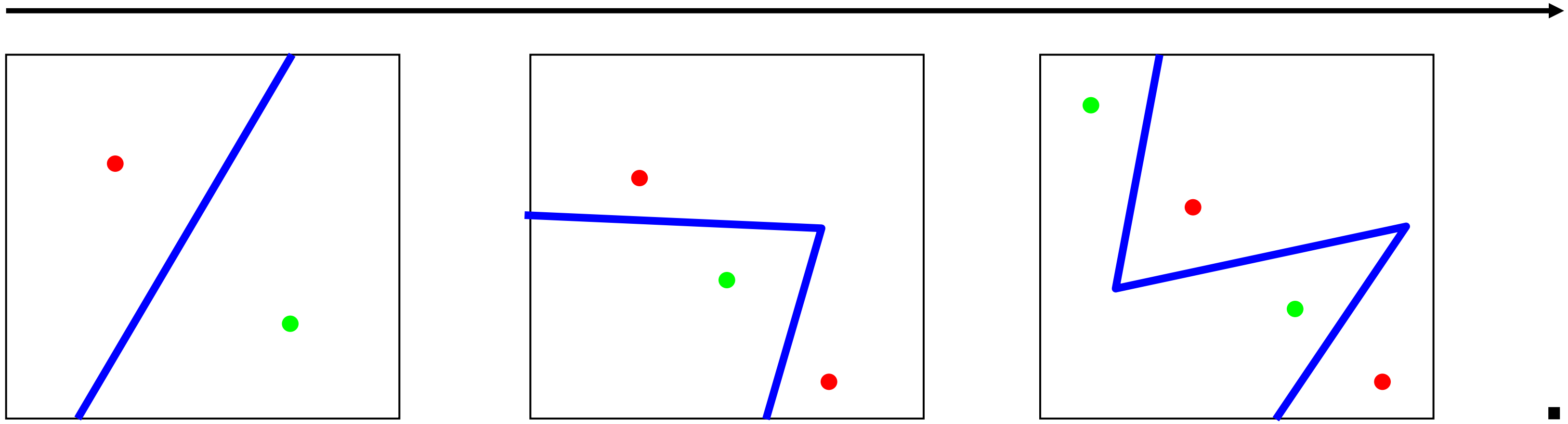
Решение: *теория оценки и контроля переобучения*

Эмпирический риск оценивается по обучающей выборке, но *сложность* решающего правила искусственно ограничивается.

Дополнительные компоненты задачи обучения:

Сложность класса классификаторов $Q(F_X)$.

VC-размерность класса классификаторов



...

Уточнение задачи: обучение классификатора заданного класса с регуляризацией по сложности

Требуется: *минимизировать наблюдаемый риск с регуляризацией по сложности класса обучаемого классификатора*:

$$\theta\in\Theta:\ c_X\in\Omega_X \mapsto f_X\in F_X \subseteq \Omega_X,$$

$$\theta = arg\ min_{\theta'\in\Theta}\ \{J_X(\theta' c_X) + \alpha Q(F_X)\}\ , \tag{3}$$

где $\alpha\geq 0$ – *параметр регуляризации*, определяющий компромисс между *точностью* на X и *сложностью классификатора*, от которой зависит поведение f_X на тестовой выборке Y.

Решение: *Метод структурной минимизации риска*

Пусть даны: *суперкласс* $\boldsymbol{F}_X$ и последовательность *вложенных классов классификаторов нарастающей сложности*:

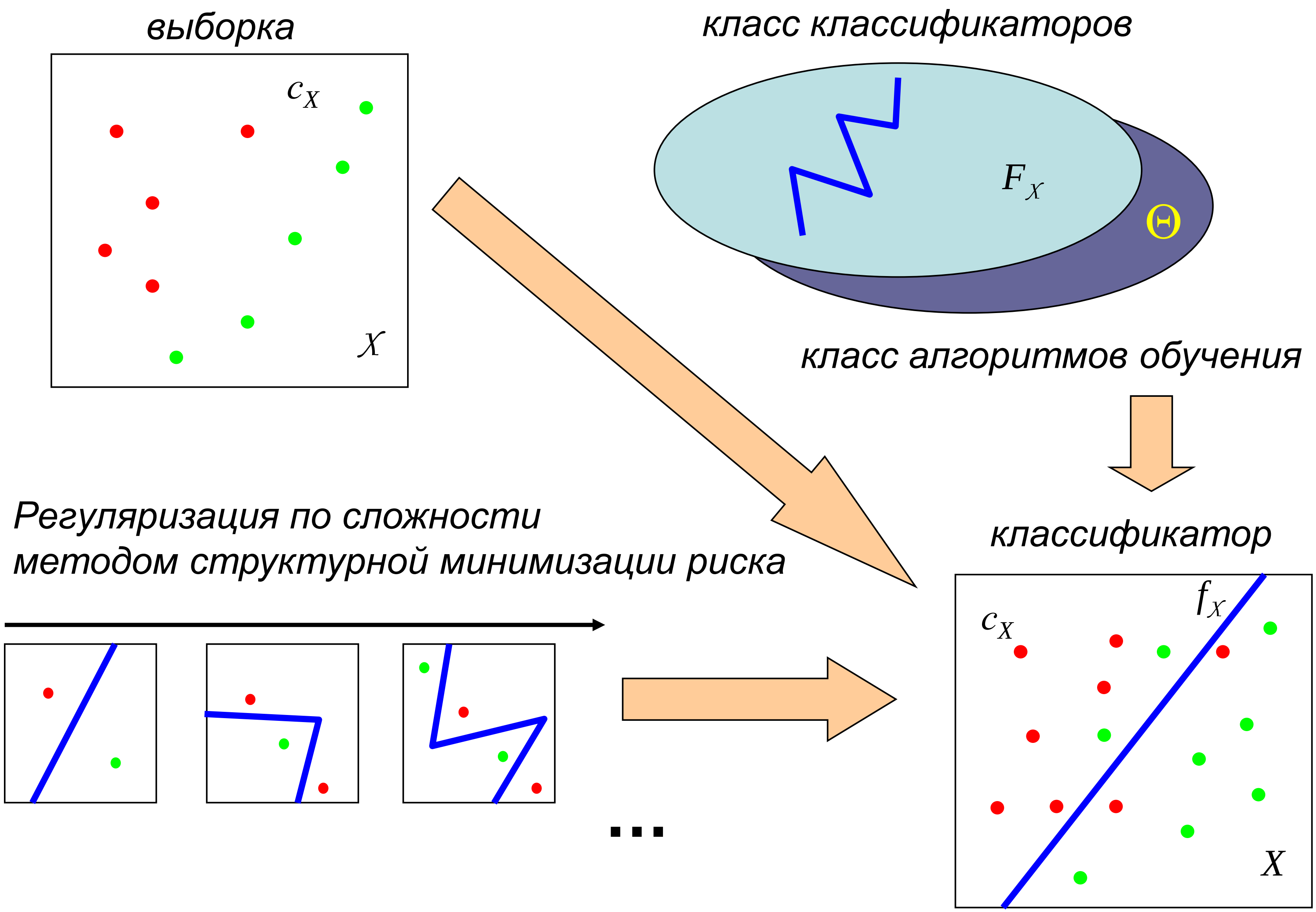
$$F^0{}_X \subseteq F^1{}_X \subseteq \ldots \subseteq F^j{}_X \subseteq \ldots \subseteq \boldsymbol{F}_X \subseteq \Omega_X:$$

$$Q(F^0{}_X) \leq Q(F^1{}_X) \leq \ldots \leq Q(F^j{}_X) \leq \ldots \tag{4}$$

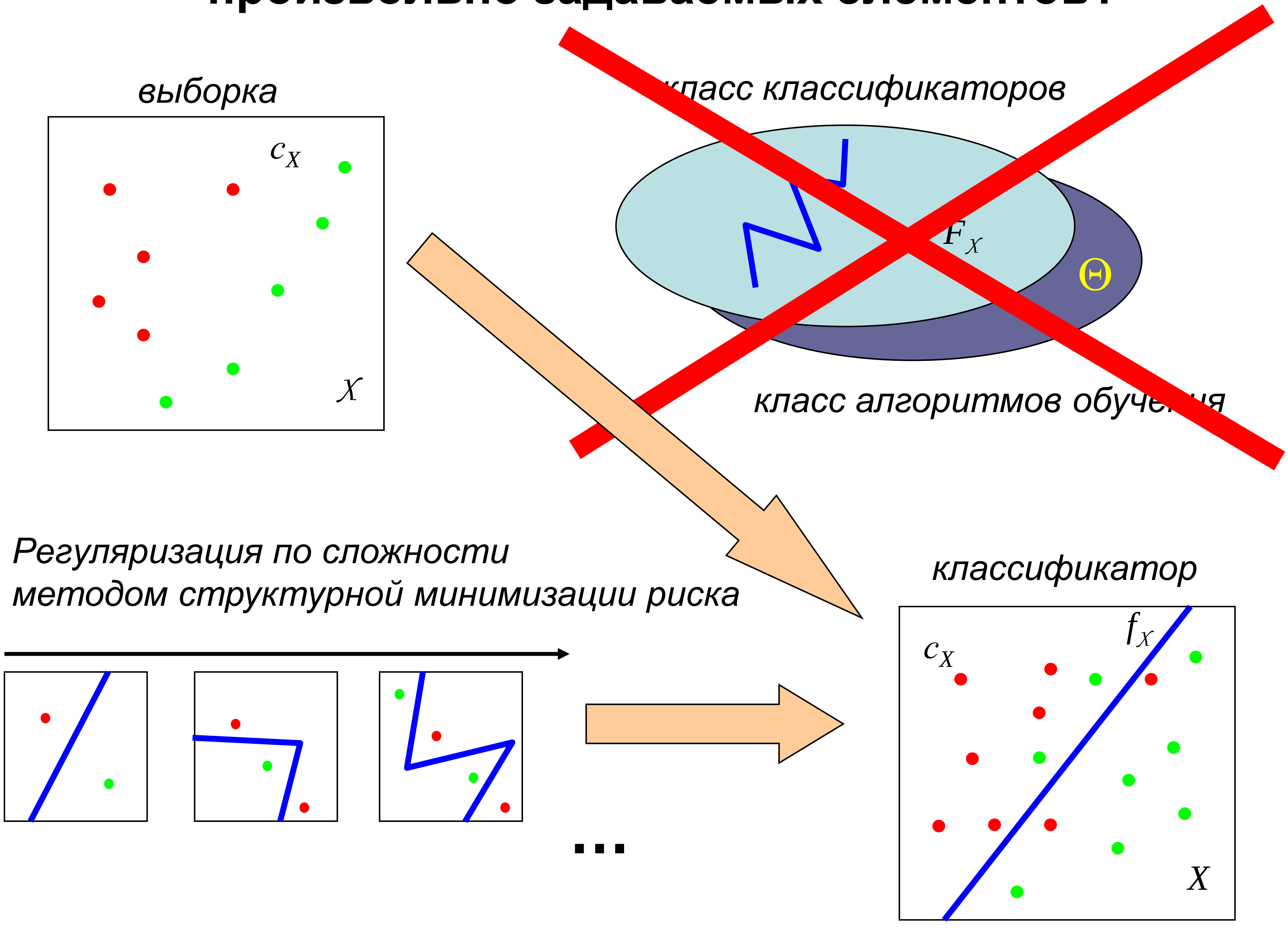
Задача (3) последовательно решается для $F^j{}_X$, j=0,1,2,…, пока значения критерия не перестанут улучшаться.

Значение α подбирается с использованием *валидационной выборки* $Z\subseteq X$, $Z\cap X=\varnothing$, $\|Z\|<+\infty$.

Схема обучения классификатора заданного класса с регуляризацией по сложности

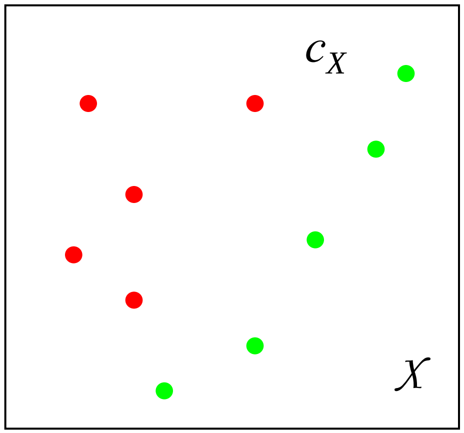


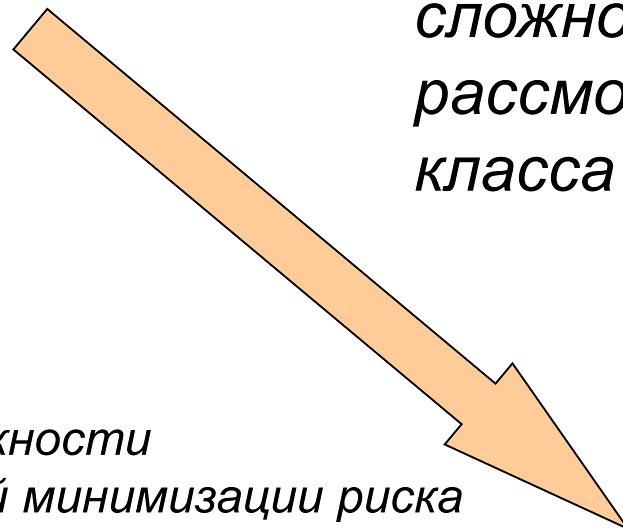
Нельзя ли избавиться в этой схеме от лишних произвольно задаваемых элементов?



Нельзя ли избавиться в этой схеме от лишних произвольно задаваемых элементов?

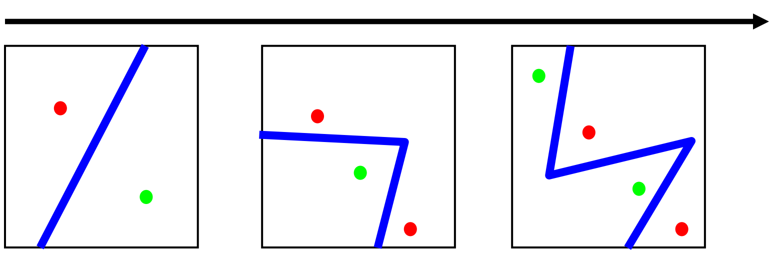
выборка



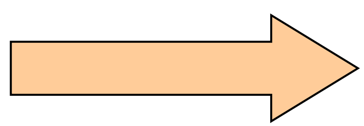


Как осуществлять регуляризацию по сложности, исключив из рассмотрения понятие класса алгоритмов?

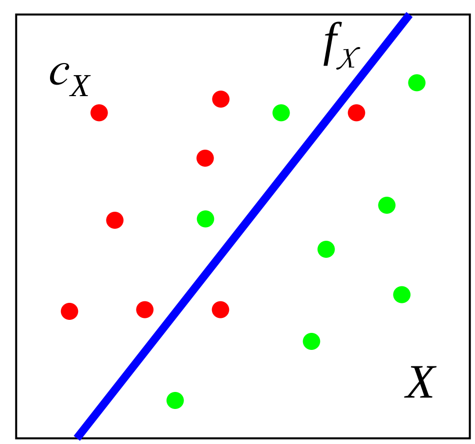
Регуляризация по сложности методом структурной минимизации риска



...



классификатор

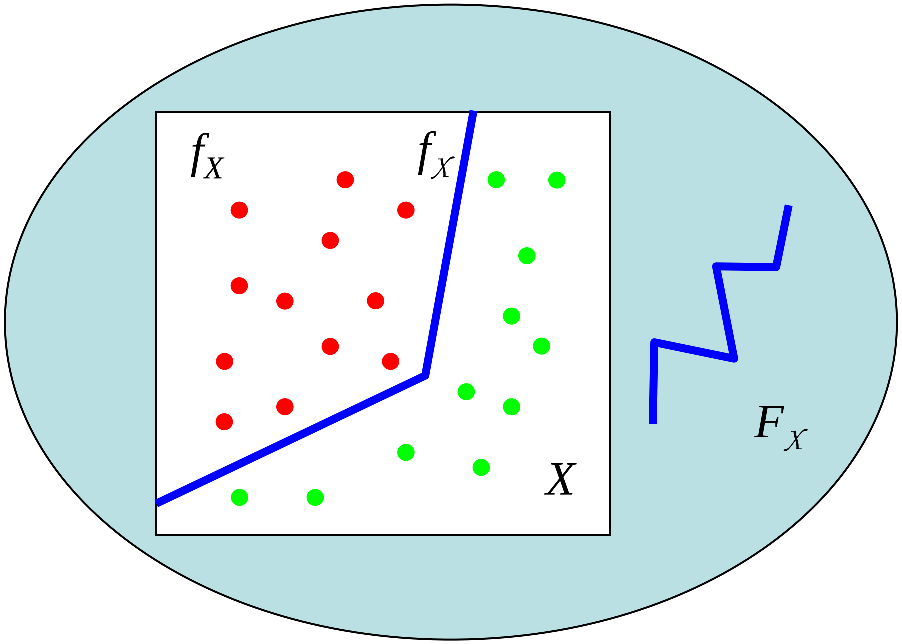


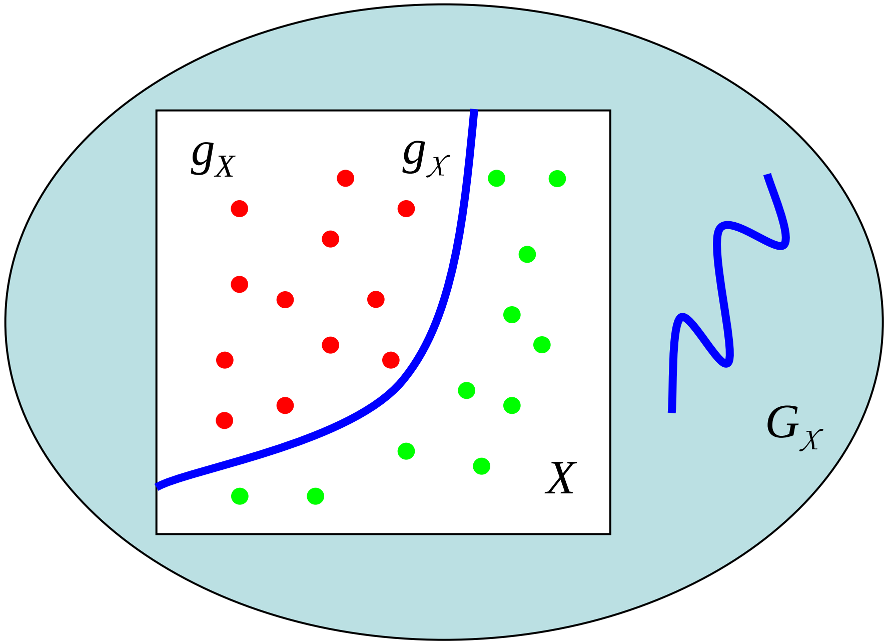
Источники основных идей и предыдущие работы

Принцип эмпирической неразличимости алгоритмов, дающих одинаковые результаты на объектах обучающей выборки.

- *Вапник В. Н. Восстановление зависимостей по эмпирическим данным. - М.: Наука, 1979.*

- *Воронцов К.В. Комбинаторная теория надёжности обучения по прецедентам. Диссертация на соискание ученой степени доктора физико-математических наук. ВЦ им. А. А. Дородницына РАН. Москва, 2010.*

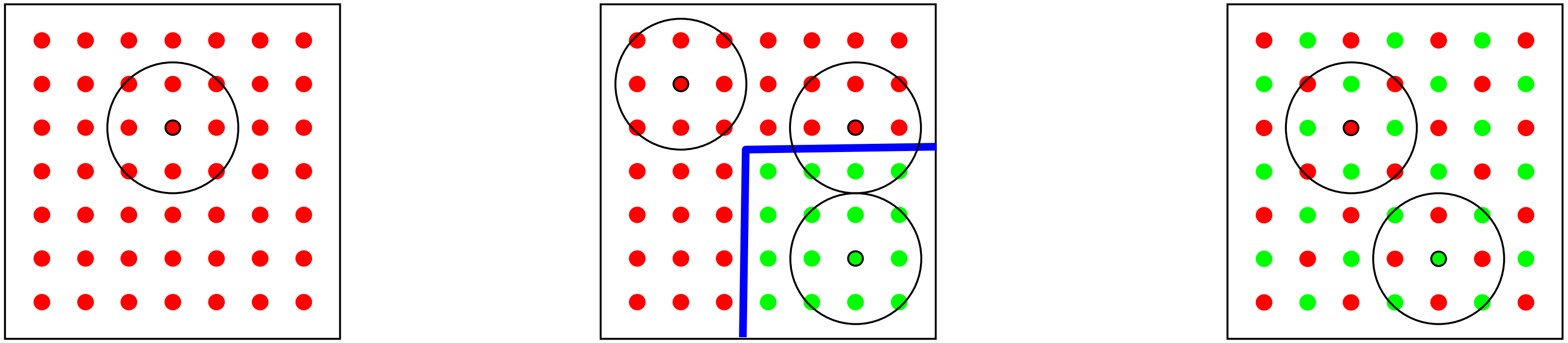




Источники основных идей и предыдущие работы

Принцип компактности: более близкие объекты должны с большей вероятностью принадлежать к одному классу.

- *Айзерман М. А., Браверман Э. М., Розоноэр Л. И. Метод потенциальных функций в теории обучения машин. М.: Наука, 1970. 320 pp.*
- *Хачай М. Ю. Топологический подход к выводу условий равномерной по классу событий сходимости частот к вероятностям. // Интеллектуализация обработки информации: 8-я международная конференция (ИОИ-8), Кипр, г.Пафос, 2010 г.: Сборник докладов. – М.: МАКС Пресс, 2010, с.91-94.*



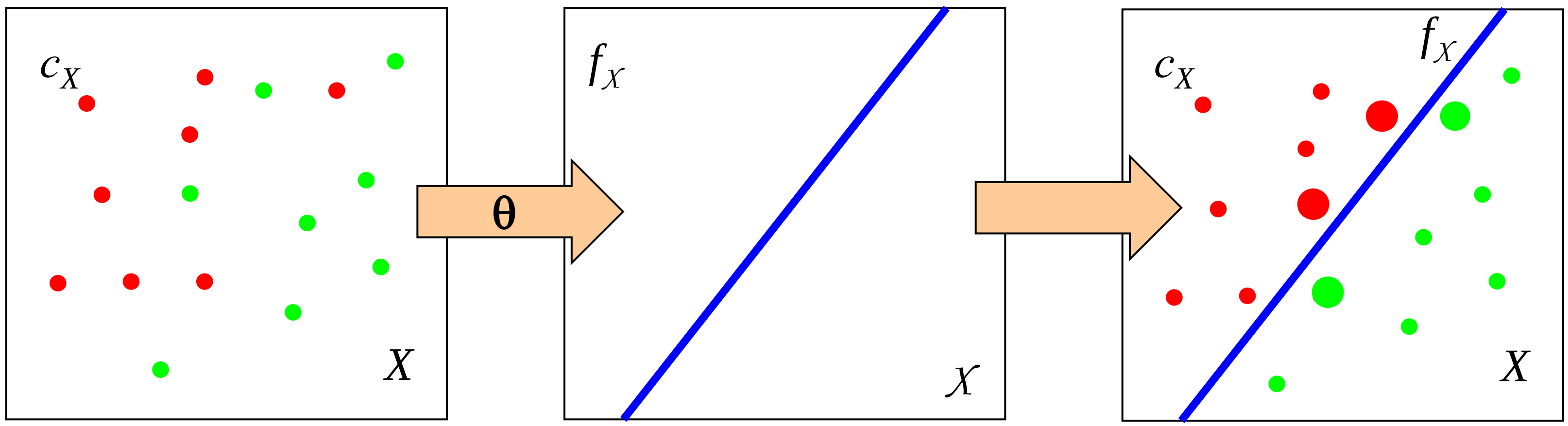
компактный класс *локально компактные классы* *некомпактные классы*

Гипотеза: сложность классификатора это некомпактность на выборке $\Rightarrow$ можно рассматривать задачу синтеза классификатора как задачу *наилучшей разметки* (*optimal labeling*) точек обучающей выборки, связанных отношениями соседства. Таким образом, из области **распознавания образов** или **машинного обучения** мы переходим в область **анализа изображений** и можем применять методы **машинного зрения** и машинной графики.

Предлагаемое решение: Морфологический синтез метрических классификаторов

Традиционная схема:

выборка X *классификатор* ***переразмеченная выборка***



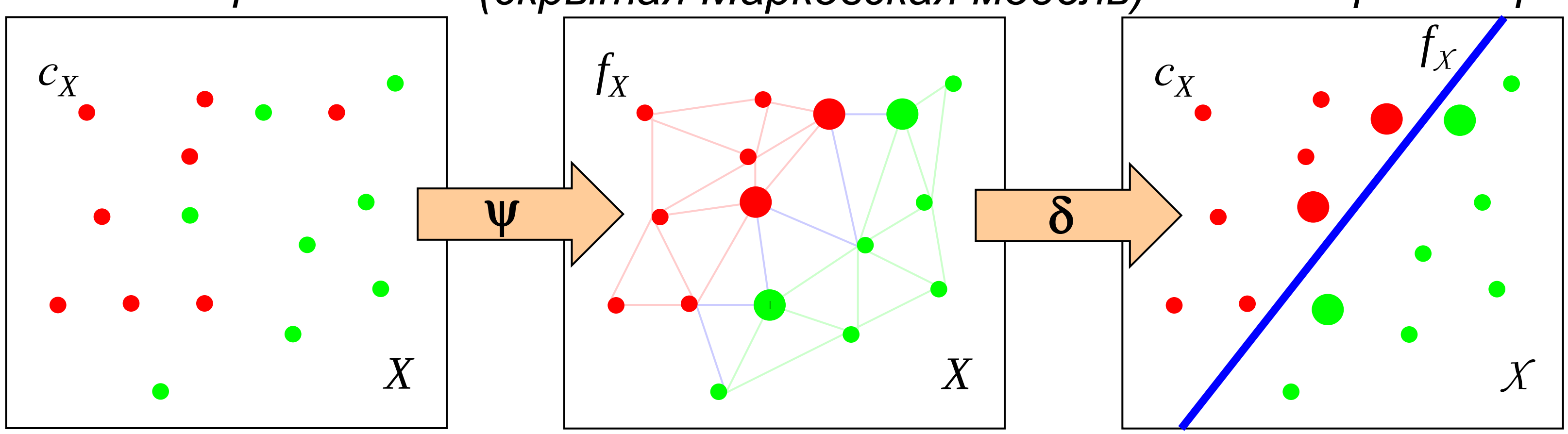
обучение с регуляризацией ***тестирование на выборке***

Предлагаемая схема:

выборка X

переразмеченная выборка X (скрытая Марковская модель)

классификатор на $\mathcal{X}$



переразметка (регуляризация) на X ***интерполяция на $\mathcal{X}$***

Предлагаемое решение: Морфологический синтез метрических классификаторов

Декомпозиция задачи синтеза: Решение (1) предлагается отыскивать в виде композиции решений подзадач:

$$\theta_\alpha = \delta_\alpha \psi_\alpha. \qquad (5)$$

Переразметка обучающей выборки:

ψ_α - ***оператор оптимального синтеза классификатора на обучающей выборке*** X *с учетом его сложности (локальной некомпактности)*

$$\psi_\alpha\colon c_X \in \Omega_X \mapsto f_X \in \Omega_X,\ \psi_\alpha = arg\ min_{\psi'}\ \{J_X(\psi' c_X) + \alpha Q_X(\psi' c_X)\} \qquad (6)$$

Корректное распространение за пределы обучающей выборки:

δ_α - ***оператор оптимальной корректной интерполяции*** *(расширения)* классификатора f_X на $\mathcal{X}$ *с учетом сложности получаемого классификатора* $f_{\mathcal{X}}$:

$$\delta_\alpha\colon f_X \in \Omega_X \mapsto f_{\mathcal{X}} \in \Omega_{\mathcal{X}},\ \delta_\alpha = arg\ min_{\delta'}\ \{J_{NN}(\delta' f_X) + \alpha Q(\delta' f_X)\}, \qquad (7)$$

где $J_{NN}(\delta f_X) = \{+\infty$, если $\exists x \in X$: $\delta f_X(x) \neq f_X(x)$;

$d_H(\delta f_X(x), \delta^{NN} f_X(x))$ – в противном случае$\}$;

d_H – расстояние Хэмминга; δ^{NN} – оператор интерполяции по *ближайшему соседу* (Nearest Neighbor), т.е. разбиение $\mathcal{X}$ по границам *диаграммы Вороного*.

Морфологический подход к синтезу метрических классификаторов

Особенности предложенного решения:

• При решении задачи синтеза (1) вводится единственное дополнительное предположение о *локальной компактности* классов объектов в *метрическом* или *топологическом* пространстве описаний $\mathcal{X}$.

• При регуляризации сложности оценивается не объемлющий класс классификаторов F, а *данный конкретный классификатор* f;

• Сложность $Q_X(f)$ классификатора f оценивается не в общем случае (применительно к «общему положению» *m*-точечного множества в $\mathcal{X}$), а *применительно к данной конкретной выборке* X.

• Сложность классификатора f на выборке X оценивается *степенью локальной некомпактности* $Q_X(f_X)$.

• Задача синтеза классификатора (1) разбивается на две последовательно решаемые подзадачи – *оптимальной переразметки обучающей выборки* X и *оптимальной корректной интерполяции* полученного классификатора на все пространство описаний $\mathcal{X}$.

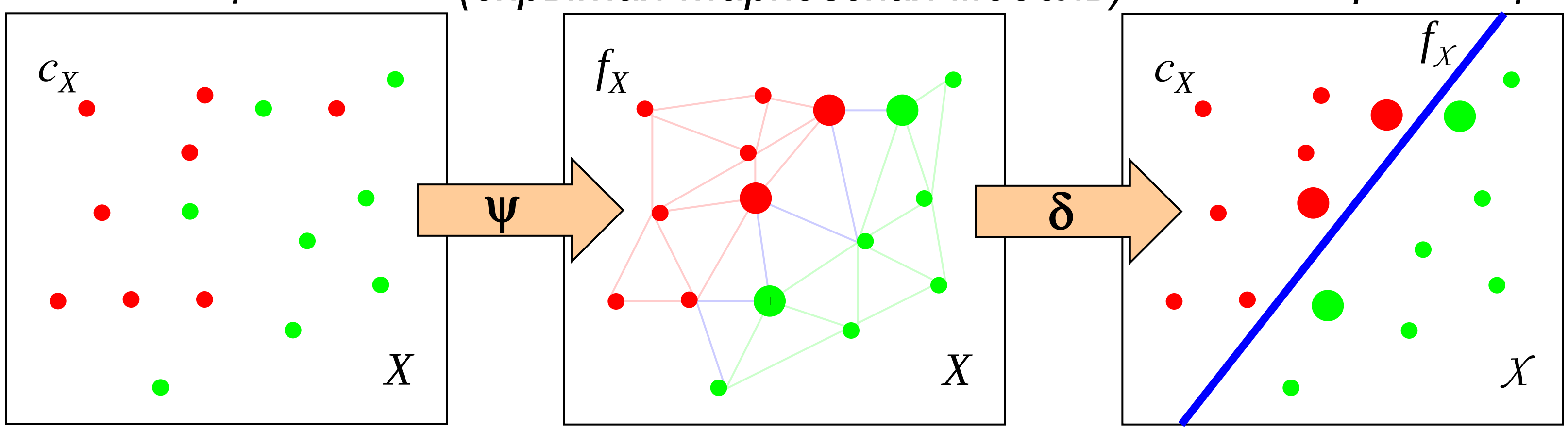
• Операторы обучения являются *критериальными морфологическими проекторами* и образуют систему «форм» возрастающей сложности (в смысле Пытьева), которая используется для структурной минимизации риска.

Содержание следующих слайдов

- *Как оценивается сложность/некомпактность на выборке?*
- *Как находится максимальный поток / минимальный разрез графа?*
- *Кто скрыл Марковскую модель или О чем умолчал Учитель?*
- *Какова связь между морфологией Пытьева и структурной минимизацией риска?*
- *Почему подход морфологический, а классификаторы метрические?*
- *Как мы моделировали, и что получилось*

Предлагаемая схема:

выборка X *переразмеченная выборка X (скрытая Марковская модель)* *классификатор на $\mathcal{X}$*



переразметка (регуляризация) на X ***интерполяция на $\mathcal{X}$***

Оценка сложности как некомпактности на обучающей выборке

Определение *k*-некомпактности:

Система вложенных k-окрестностей $O_k(x) \subseteq X$, $k=1,\ldots,\|X\|-1$, $x \in X \subseteq \mathcal{X}$.

Локальная мера некомпактности f_X *в окрестности* $O_k(x)$:

$$Q_k(x, f_X) = q_H(O_k(x)) \,/\, \| O_k(x) \|,$$

$$q_H(O_k(x)) = \sum_{y \in O_k(x)} 1(f_X(x) \neq f_X(y)), \| O_k(x) \| = \sum_{y \in O_k(x)} 1. \qquad (10)$$

Глобальная мера k-некомпактности:

$$Q^k{}_X(f_X) = Q_H(X, f_X) \,/\, \| X \|,$$

$$Q_H(X, f_X) = \sum_{x \in X} Q_k(x, f_X) \qquad (11)$$

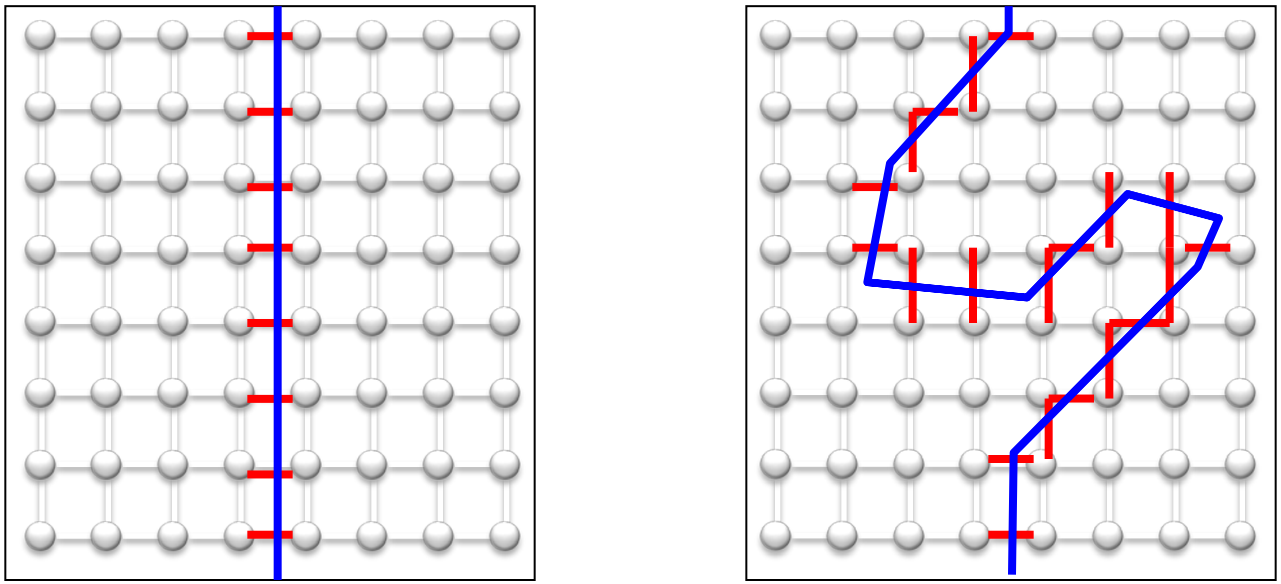
Оценка вероятности того, что один из k ближайших соседей в разбиении $f_X(x)$ будет отнесен к другому классу.

Оценка сложности как некомпактности на обучающей выборке

Свойства *k*-некомпактности:

- VC-усложнению класса классификатора f_X соответствует нарастание меры *k*-некомпактности $Q^k_X(f_X)$ (при фиксированном k).

- увеличение параметра k ведет к предпочтению все более «гладких» разделяющих поверхностей (см. *Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In Proc. IEEE International Conf. Computer Vision (ICCV), pages 26–33, 2003.*).



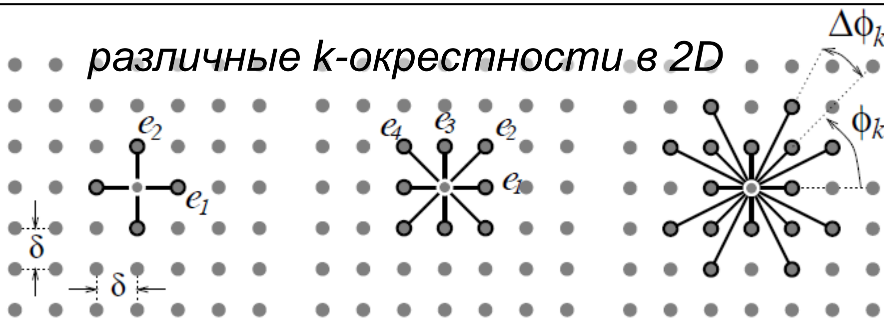
k = 4

Оценка сложности как некомпактности на обучающей выборке

Свойства *k*-некомпактности:

- VC-усложнению класса классификатора f_X соответствует нарастание меры *k*-некомпактности $Q^k_X(f_X)$ (при фиксированном k).

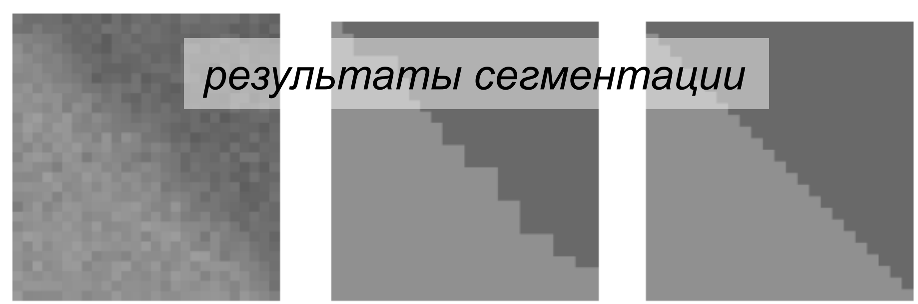
- увеличение параметра k ведет к предпочтению все более «гладких» разделяющих поверхностей (см. *Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In Proc. IEEE International Conf. Computer Vision (ICCV), pages 26–33, 2003.*).



(a) 4 n-system (b) 8 n-system (c) 16 n-system

(a) 4 n-system (b) 8 n-system (c) 128 n-system

(C) Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In Proc. IEEE International Conf. Computer Vision (ICCV), pages 26–33, 2003.

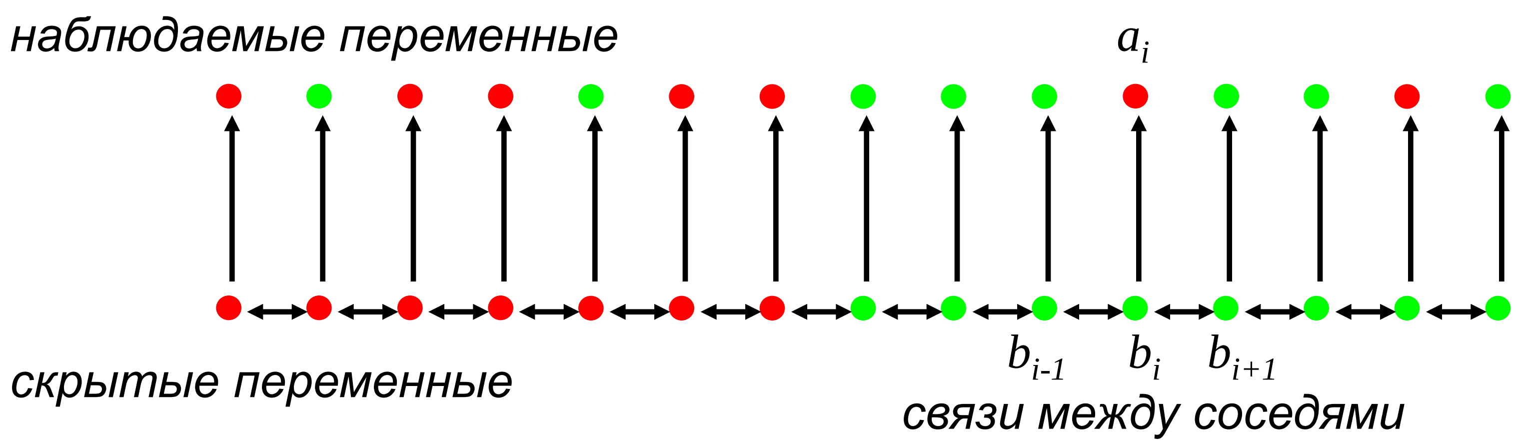


(a) Original data (b) 4 n-system (c) 8 n-system

Источники основных идей и предыдущие работы

Скрытые Марковские модели (СММ) – для «объяснения поведения» наблюдаемых переменных каждой из них ставится в соответствие некоторая новая ненаблюдаемая (скрытая) переменная, причем считается, что вероятность нахождения каждой из скрытых переменных в одном из возможных состояний зависит от состояния конечного числа соседних (по некоторому графу) ненаблюдаемых переменных.

- *Geman S., Geman D. Stochastic relaxation, Gibbs distributions,, the Bayesian restoration of images. – IEEE Trans. Pattern Analysis, Machine Intelligence, 1984, № 6, pp.721-741.*



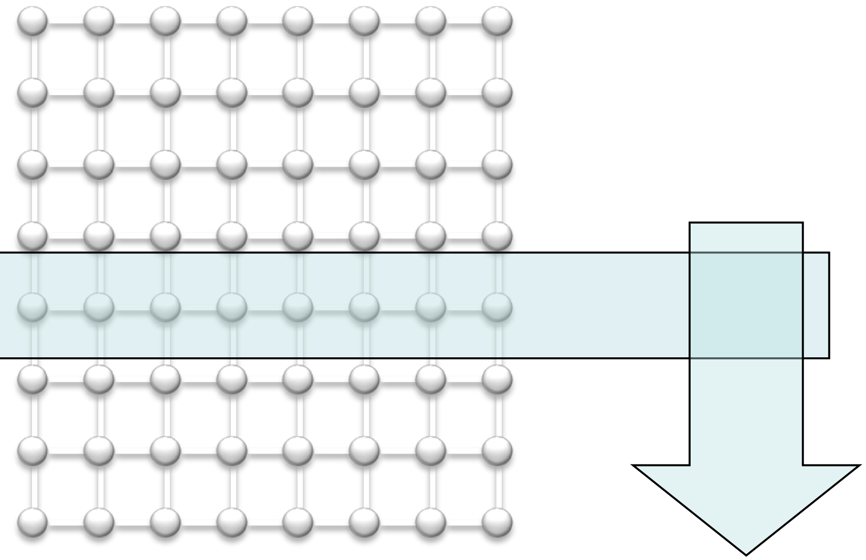
В одномерном случае задачи оптимизации критериев путем реконструкции СММ могут быть точно решены методом динамического программирования

Источники основных идей и предыдущие работы

Методы оптимизации СММ

- **Моделируемый отжиг**
- Методы линейного программирования
- Методы динамического программирования
- **Методы разреза графов**

Граф структурных связей



Граф структурных связей должен быть *ациклическим*

Методы максимального потока / минимального разреза графов

- *L. Ford and D. Fulkerson. Flows in Networks. Princeton University Press,1962.*
- *D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. Journal of the Royal Statistical Society, 51(2):271–279, 1989.*
- *Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In Proc. IEEE International Conf. Computer Vision (ICCV), pages 26–33, 2003.*
- *Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), 26(9):1124–1137, 2004.*
- *V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts?. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), 26(2):147–159, 2004.*

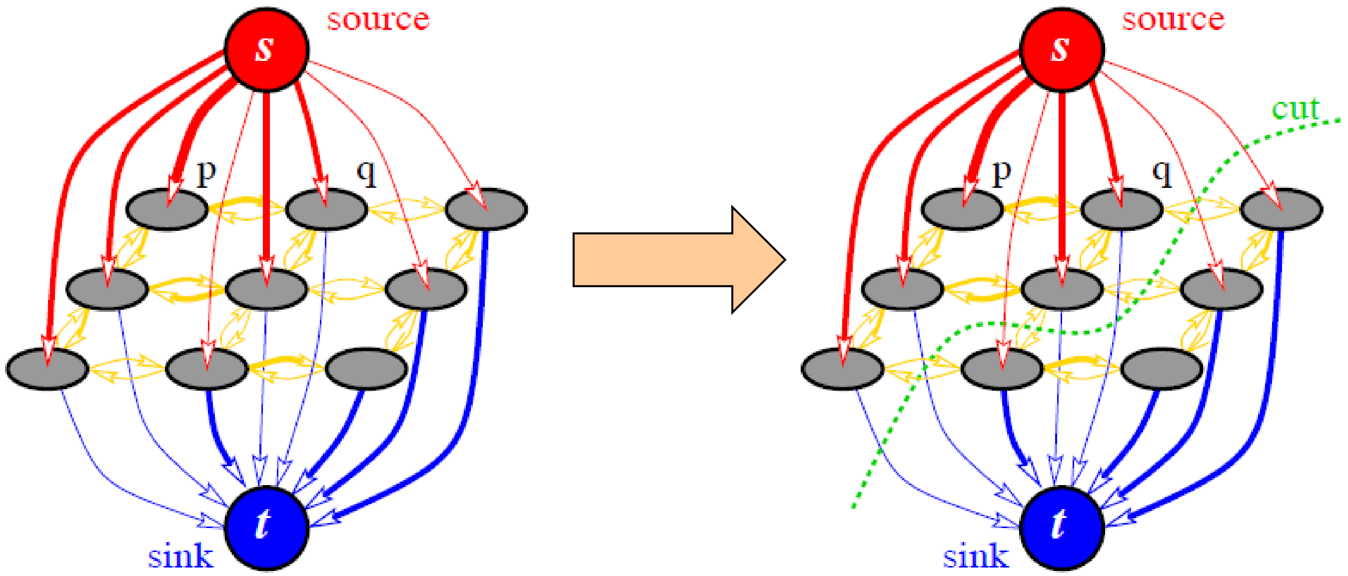
Метод отыскания максимального потока / минимального разреза графов

Минимизация энергии (Energy-based models)

На графе с двумя *терминальными вершинами* позволяет находить минимум *функционала энергии* вида:

$$E(T) = E_0 + \sum_{i=1..N} E_i(t_i) + \sum_{(i,j)\in V} E_{ij}(t_i,t_j), \qquad (12)$$

где N- число нетерминальных вершин графа; $T=\langle t_1,\ldots,t_N\rangle$, $t_1, \ldots,t_N \in\{0,1\}$ – метки ассоциирования каждой нетерминальной вершины с одной из терминальных; $E_i(0)$, $E_i(1)\in\{0,1\}$ – *унарные потенциалы*; $E_{ij}(t_i,t_j)$ – *парные потенциалы*, задаваемые четверкой действительных коэффициентов $E_{ij}(0,0)$, $E_{ij}(0,1)$, $E_{ij}(1,0)$, $E_{ij}(1,1)$; V –подмножество пар индексов переменных, задающее систему соседства на T.



(C) Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), 26(9):1124–1137, 2004.

Метод отыскания максимального потока / минимального разреза графов

Минимизация энергии (Energy-based models)

На графе с двумя *терминальными вершинами* позволяет находить минимум *функционала энергии* вида:

$$E(T) = E_0 + \sum_{i=1..N} E_i(t_i) + \sum_{(i,j)\in V} E_{ij}(t_i,t_j), \quad (12)$$

где N- число нетерминальных вершин графа; $T=\langle t_1,\ldots,t_N\rangle$, $t_1, \ldots,t_N \in\{0,1\}$ – метки ассоциирования каждой нетерминальной вершины с одной из терминальных; $E_i(0)$, $E_i(1)\in\{0,1\}$ – *унарные потенциалы*; $E_{ij}(t_i,t_j)$ – *парные потенциалы*, задаваемые четверкой действительных коэффициентов $E_{ij}(0,0)$, $E_{ij}(0,1)$, $E_{ij}(1,0)$, $E_{ij}(1,1)$; V –подмножество пар индексов переменных, задающее систему соседства на T.

Субмодулярность энергии (аналог условия выпуклости функционала)

Энергия (12) считается *субмодулярной*, если:

$$\forall(i,j)\in V\colon E_{ij}(0,0) + E_{ij}(1,1) \le E_{ij}(0,1) + E_{ij}(1, 0). \quad (13)$$

Для субмодулярной энергии (12)-(13) метод построения *минимального разреза графа* гарантирует нахождение точного минимума.

- *V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts?. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), 26(2):147–159, 2004.*

Минимизация сложности метрических классификаторов методом разреза графов

С учетом критерия (11) задача (6) сводится к задаче типа (12). При этом:

$C=\{0,1\}$, $N=\|X\|$,

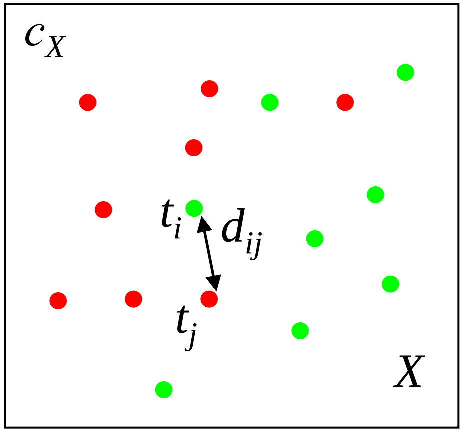
$T=\langle t_1,\ldots,t_N\rangle$, $t_1=f_X(x_1),\ldots,t_N=f_X(x_N)$;

$E_i(x)= 1(f_X(x) \neq c_X(x))$,

$E_{ij}(t_i,t_j)= 1(f_X(x_i) \neq f_X(x_j)) \times d_{ij}$;

$V = \{(i,j)\colon j \in O_k(x_i)\}$.

Энергия $E(T)$ (12) здесь является субмодулярной (13), а значит, метод минимального разреза графа *k*-соседства для выборки c_X действительно порождает α-семейства проекторов из Утверждения 1.



Источники основных идей и предыдущие работы

Математическая морфология и морфологический анализ

Математическая морфология Серра и Матерона – топологические идеи покрытия фигур компактными областями и анализа окрестностей;

- *Serra J. Image Analysis and Mathematical Morphology, Academic Press, London, 1982.*

Теория форм Павель - роль проекторов как распознающих операторов;

- *Pavel M. Fundamentals of Pattern Recognition. Marcel Dekker. Inc., New York, 1989*.

Морфологический анализ Пытьева – формы изображения как разбиения пространства кадра, проекция образа на форму другого образа, операция сравнения форм по сложности, морфологический коэффициент корреляции;

- *Пытьев Ю.П. Морфологический анализ изображений. Доклады АН СССР, 1983. Т. 269. № 5. С. 1061-1064.*
- *Пытьев Ю.П., Чуличков А.И. Методы морфологического анализа изображений // М.: ФИЗМАТЛИТ, 2010. 336с.*

Критериальные проективные морфологии – морфологические проекторы, заданные критериями.

- *Визильтер Ю.В. Обобщенная проективная морфология. «Компьютерная оптика», Том 32, №4. 2008, С.384-399.*
- *J. Darbon and M. Siegelle. Image Restoration with Discrete Constrained Total Variation Part I: Fast and Exact Optimization.Journal of Mathematical Imaging and Vision. Vol. 26, no. 3, pages 261-276, December 2006.*

Морфологический подход к синтезу метрических классификаторов

Проблема переобучения и критериальная морфология

<u>*Утверждение 1.*</u>

Если

ψ_α – критериальный морфологический фильтр

$$\psi_\alpha A = argmin_B \{J(A,B) + \alpha Q(B)\},$$

где A – исходный образ, B – выходной образ, $J(A,B)$ – критерий минимума ошибки аппроксимации, $Q(B)$ – выпуклый критерий сложности образа B, $\alpha \geq 0$ – параметр регуляризации, и критерий $J(A,B)$ *обладает свойствами метрики*,

То

1) оператор ψ_α является проектором: $\psi^2_\alpha = \psi_\alpha$;

2) проектор ψ_α определяет систему вложенных классов, монотонную по α: $\alpha \geq \beta \Rightarrow \psi_\beta \psi_\alpha = \psi_\alpha$. ♦

<u>*Доказательство*</u>*:* см. *Визильтер Ю.В. Обобщенная проективная морфология. «Компьютерная оптика», Том 32, №4. 2008, С.384-399.*

Морфологический подход к синтезу метрических классификаторов

Проблема переобучения и критериальная морфология

Поскольку в задаче (6, 7) количество ошибок классификации (функционал J) имеет вид *расстояния Хэмминга*, из Утверждения 1 следует, что оператор θ_α (5) является *алгебраическим проектором*:

$$\theta_\alpha^2 = \theta_\alpha \Rightarrow \forall x \in X\colon \theta_\alpha f_X(x) = f_X(x). \tag{8}$$

Кроме того, из Утверждения 1 следует, что на основе θ_α образуется система вложенных классов решающих правил, монотонная относительно α:

$$\forall \alpha \geq \beta \Rightarrow F^\alpha_X \subseteq F^\beta_X : Q(F^\alpha_X) \leq Q(F^\beta_X), \tag{9}$$

где $F^\alpha_X = \{f_X\colon \theta_\alpha f_X = f_X\}$ – множество классификаторов (разбиений), стабильное относительно проектора θ_α.

Источники основных идей и предыдущие работы

Кто скрыл Марковскую модель

или О чем умолчал Учитель

Традиционная ситуация использования СММ: *скрытые переменные - «истинные», наблюдаемые – «зашумленные» или «искаженные»*.

Использование СММ в задачах обучения:

Интерпретация 1: *наблюдаемые данные - истинные* (Учитель сообщает истинные метки объектов из обучающей выборки), *скрытые переменные – искусственные и ложные*, поскольку формируемый метрический классификатор заведомо ошибается на обучающей выборке ради компактности формируемых классов.

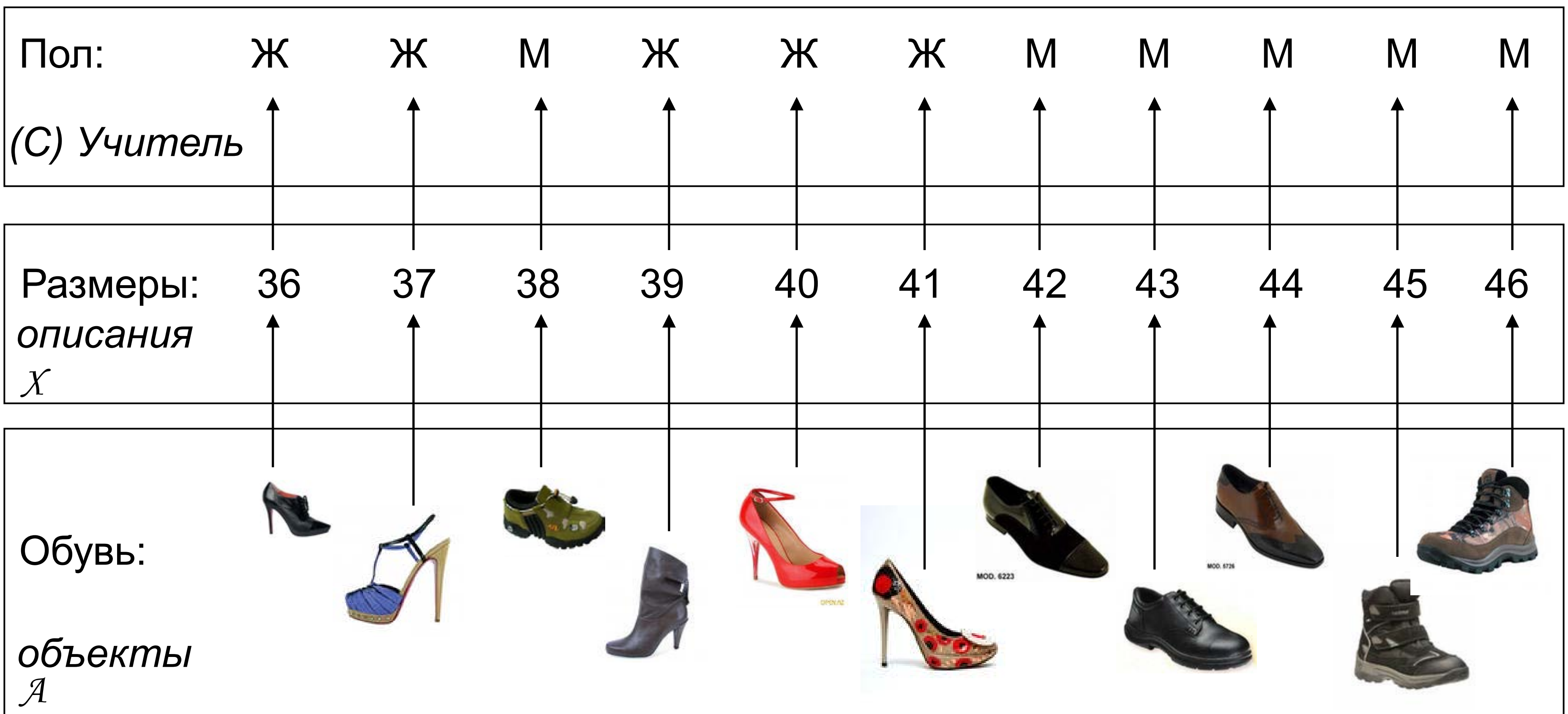
Интерпретация 2: *наблюдаемые данные действительно искажены, а мы пытаемся вскрыть их истинную природу.* Истиной здесь считается разбиение пространства описаний на области преобладания *возможности* меток того или иного класса, определяемой неизвестной Учителю картиной вероятностных распределений.

- *Пытьев Ю.П.*, **«***Возможность как альтернатива вероятности***»,** *М. Наука 2007, 464с.*
- *Пытьев Ю.П., Чуличков А.И. Методы морфологического анализа изображений // М.: ФИЗМАТЛИТ, 2010. 336с.*

Источники основных идей и предыдущие работы

Кто скрыл Марковскую модель
или О чем умолчал Учитель

Интерпретация 2: *наблюдаемые данные действительно искажены, а мы пытаемся вскрыть их истинную природу:* разбиение пространства описаний на области преобладания *возможности* меток того или иного класса.

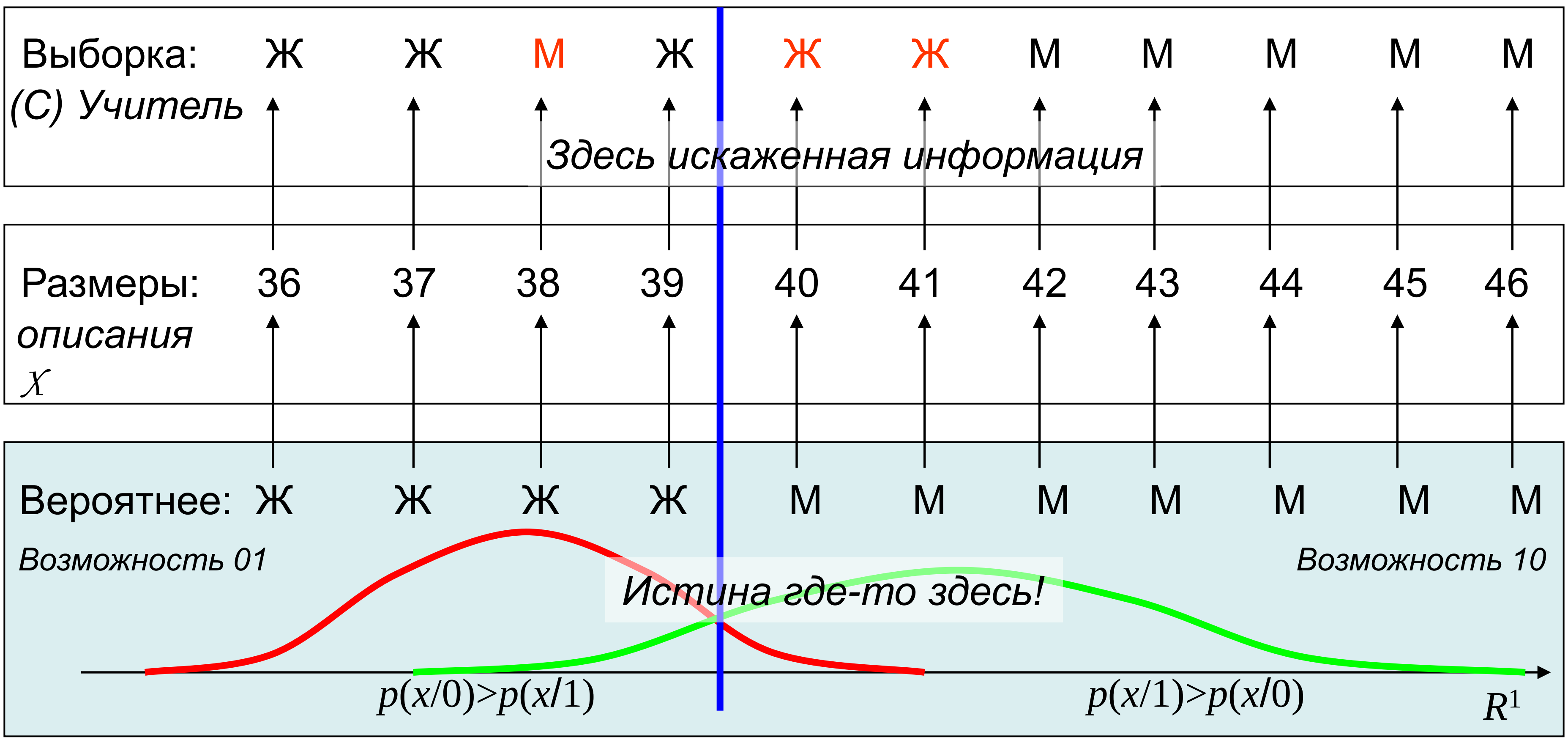


Источники основных идей и предыдущие работы

Кто скрыл Марковскую модель

или О чем умолчал Учитель

Интерпретация 2: *наблюдаемые данные действительно искажены, а мы пытаемся вскрыть их истинную природу:* разбиение пространства описаний на области преобладания *возможности* меток того или иного класса.



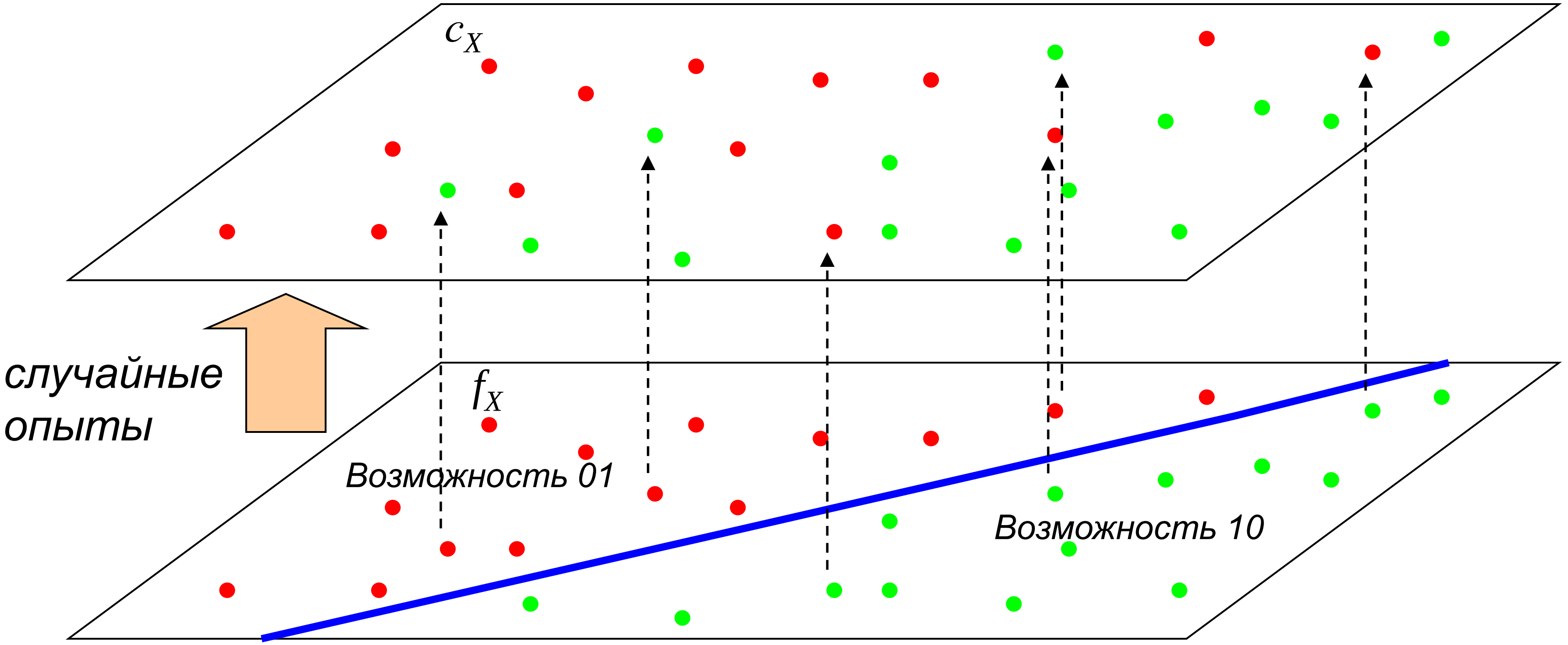
Источники основных идей и предыдущие работы

Кто скрыл Марковскую модель

или О чем умолчал Учитель

Интерпретация 2: *наблюдаемые данные действительно искажены, а мы пытаемся вскрыть их истинную природу:* разбиение пространства описаний на области преобладания *возможности* меток того или иного класса.

случайные наблюдения классов на выборке



детерминированная возможность классов на выборке

Минимизация сложности метрических классификаторов методом разреза графов

Методика моделирования и реализация алгоритмов

Количество классов – 2

Размерность пространства описаний – 2

Модель распределения – смеси двумерных гауссовых распределений.

Граф соседства строился на основе триангуляции Делоне с динамическим кэшированием треугольников

- *Скворцов А.В. Обзор алгоритмов построения триангуляции Делоне. Вычислительные методы и программирование. 2002. Т3. с.14-39.*)

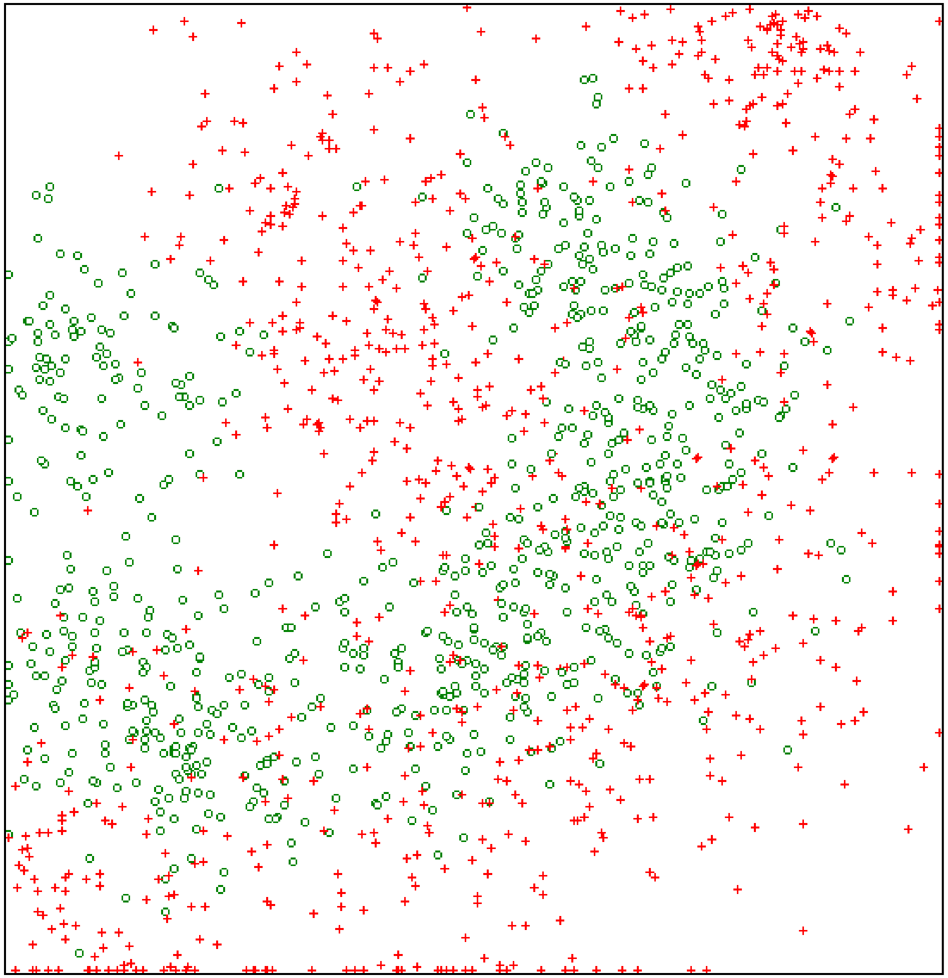
Нахождение разрезов графов - с использованием библиотеки

- *Yuri Boykov, Vladimir Kolmogorov. MAXFLOW - software for computing mincut/maxflow in a graph.V.3.01: http://www.cs.ucl.ac.uk/staff/V.Kolmogorov/software.html*

Минимизация сложности метрических классификаторов методом разреза графов

Пример обучения с учителем (supervised learning)

Обучающая выборка - 1000 объектов (по 500 для каждого класса).

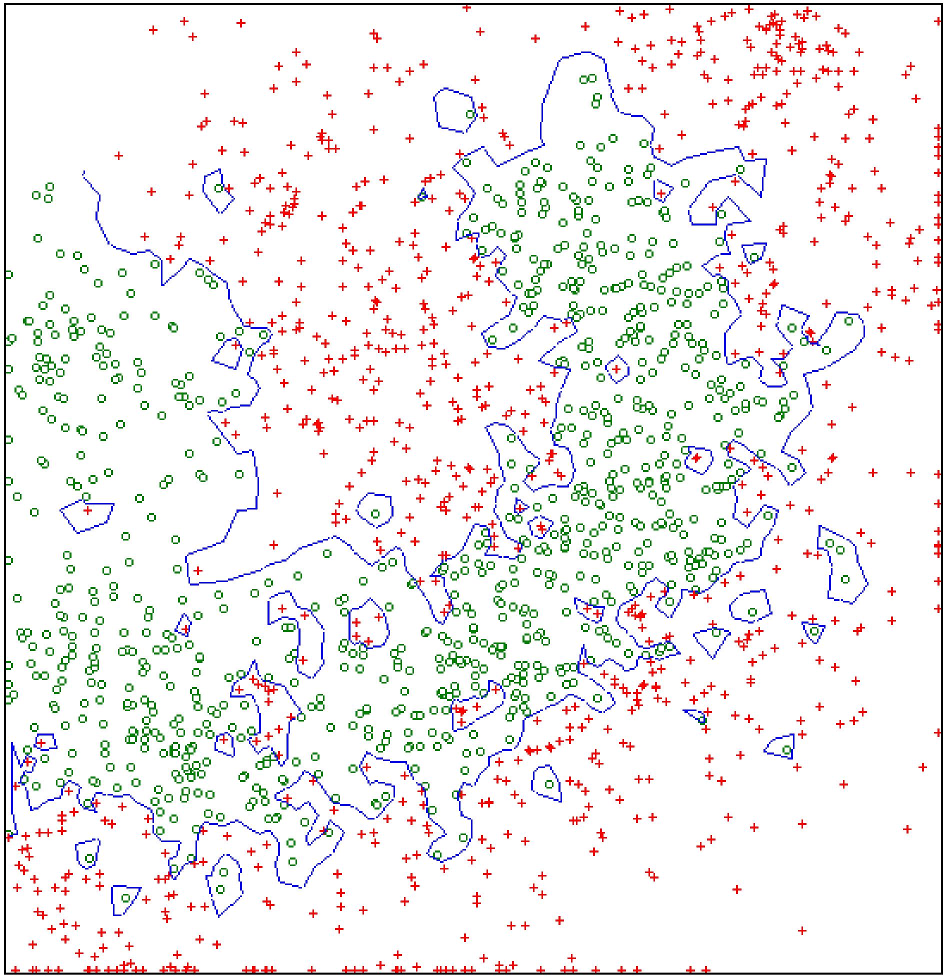


α=0 (обучающая выборка)

Минимизация сложности метрических классификаторов методом разреза графов

Пример обучения с учителем (supervised learning)

Обучающая выборка - 1000 объектов (по 500 для каждого класса).

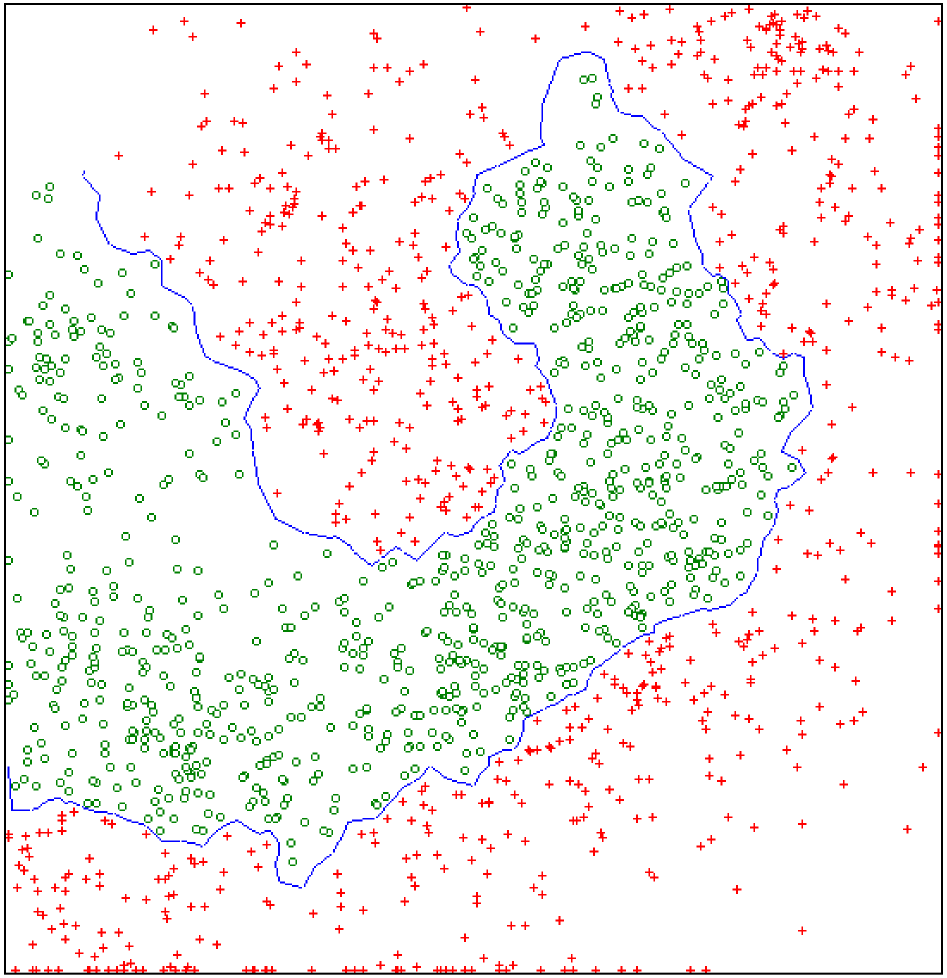


α=1000 (переобучение)

Минимизация сложности метрических классификаторов методом разреза графов

Пример обучения с учителем (supervised learning)

Обучающая выборка - 1000 объектов (по 500 для каждого класса).

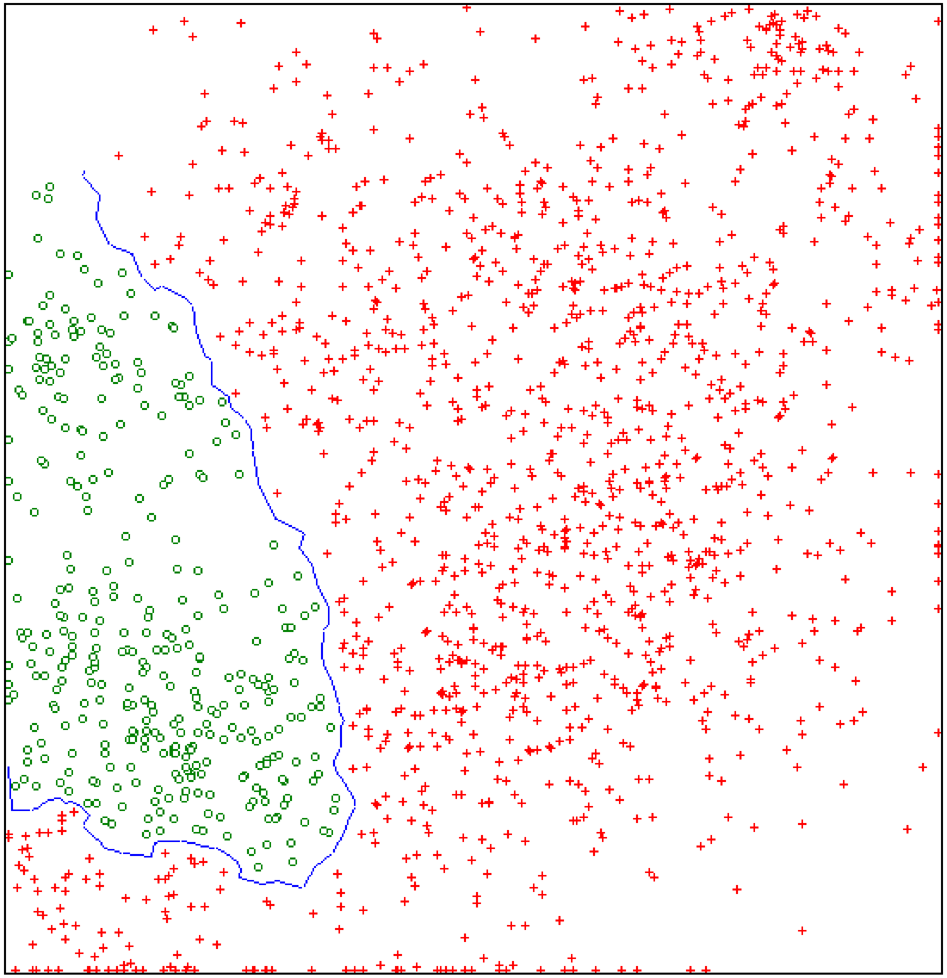


α=4500 (оптимум)

Минимизация сложности метрических классификаторов методом разреза графов

Пример обучения с учителем (supervised learning)

Обучающая выборка - 1000 объектов (по 500 для каждого класса).

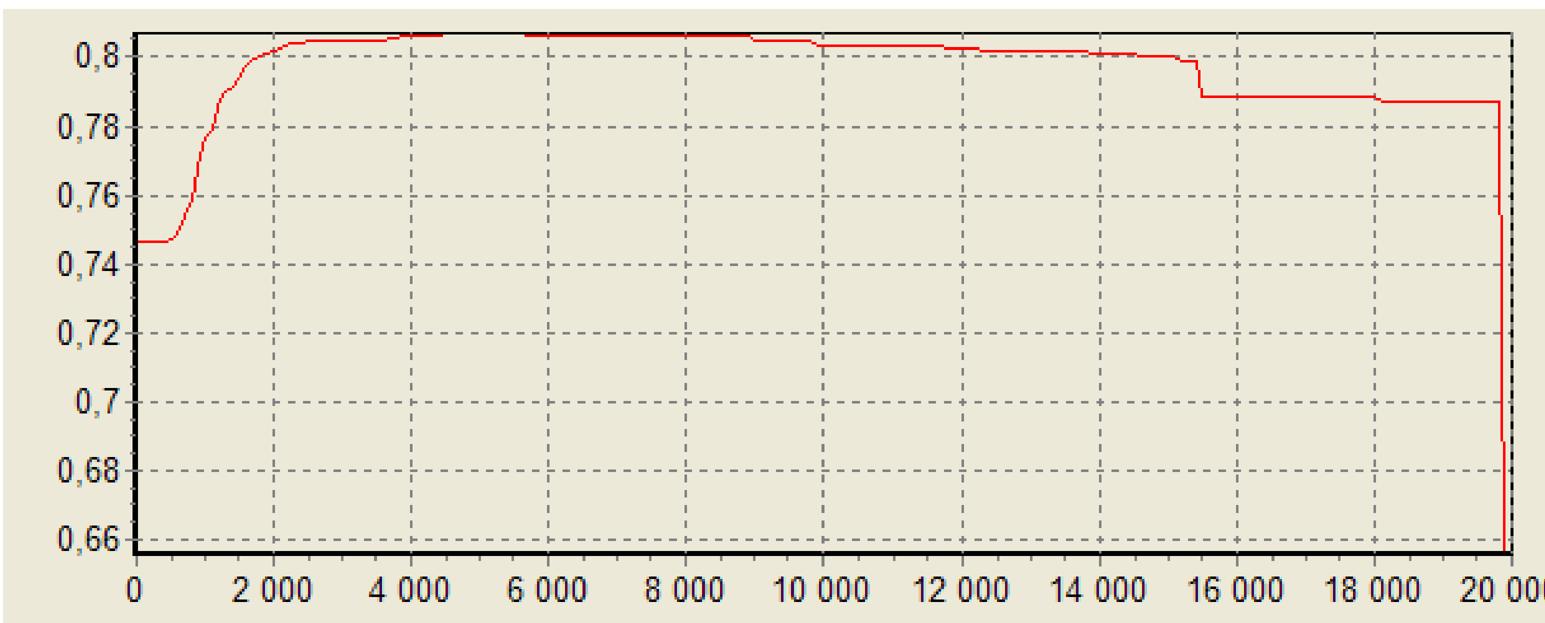


α=20000 (перебор)

Минимизация сложности метрических классификаторов методом разреза графов

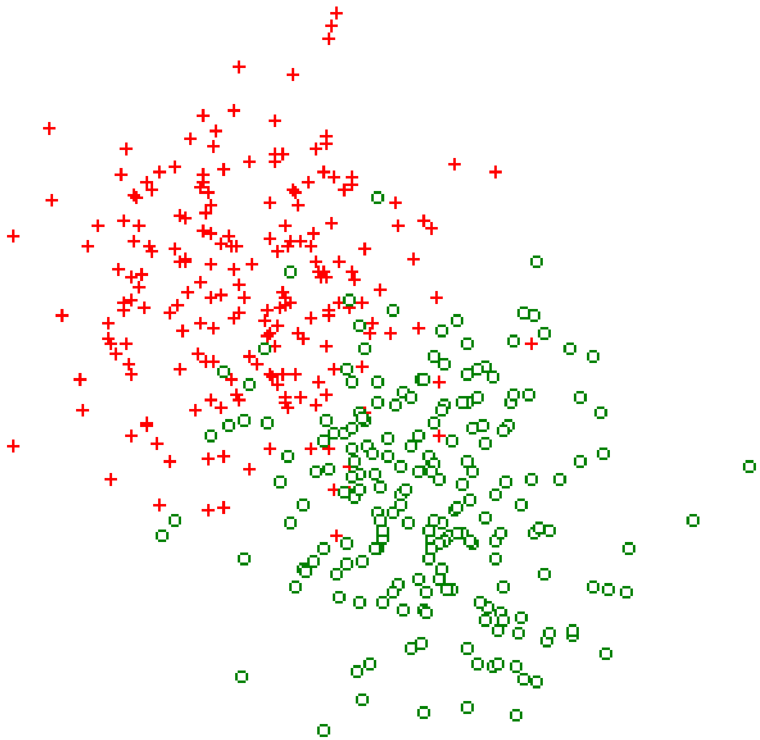
Пример обучения с учителем (supervised learning)
Обучающая выборка - 1000 объектов (по 500 для каждого класса).

Зависимость вероятности правильного распознавания от параметра α, (получена методом Монте-Карло по 100000 независимых выборок).



Минимизация сложности метрических классификаторов методом разреза графов

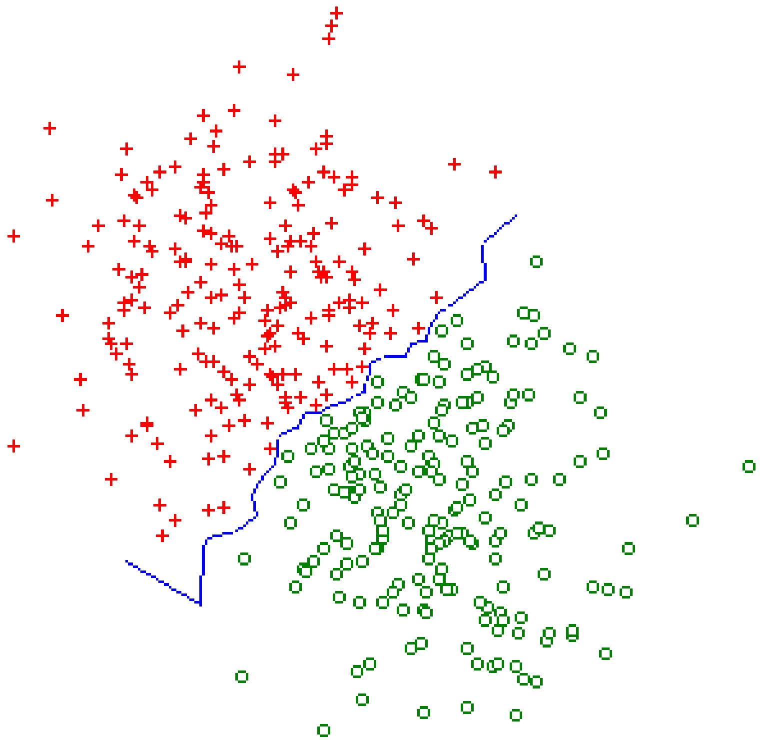
Пример обучения с учителем (supervised learning)



обучающая выборка

Минимизация сложности метрических классификаторов методом разреза графов

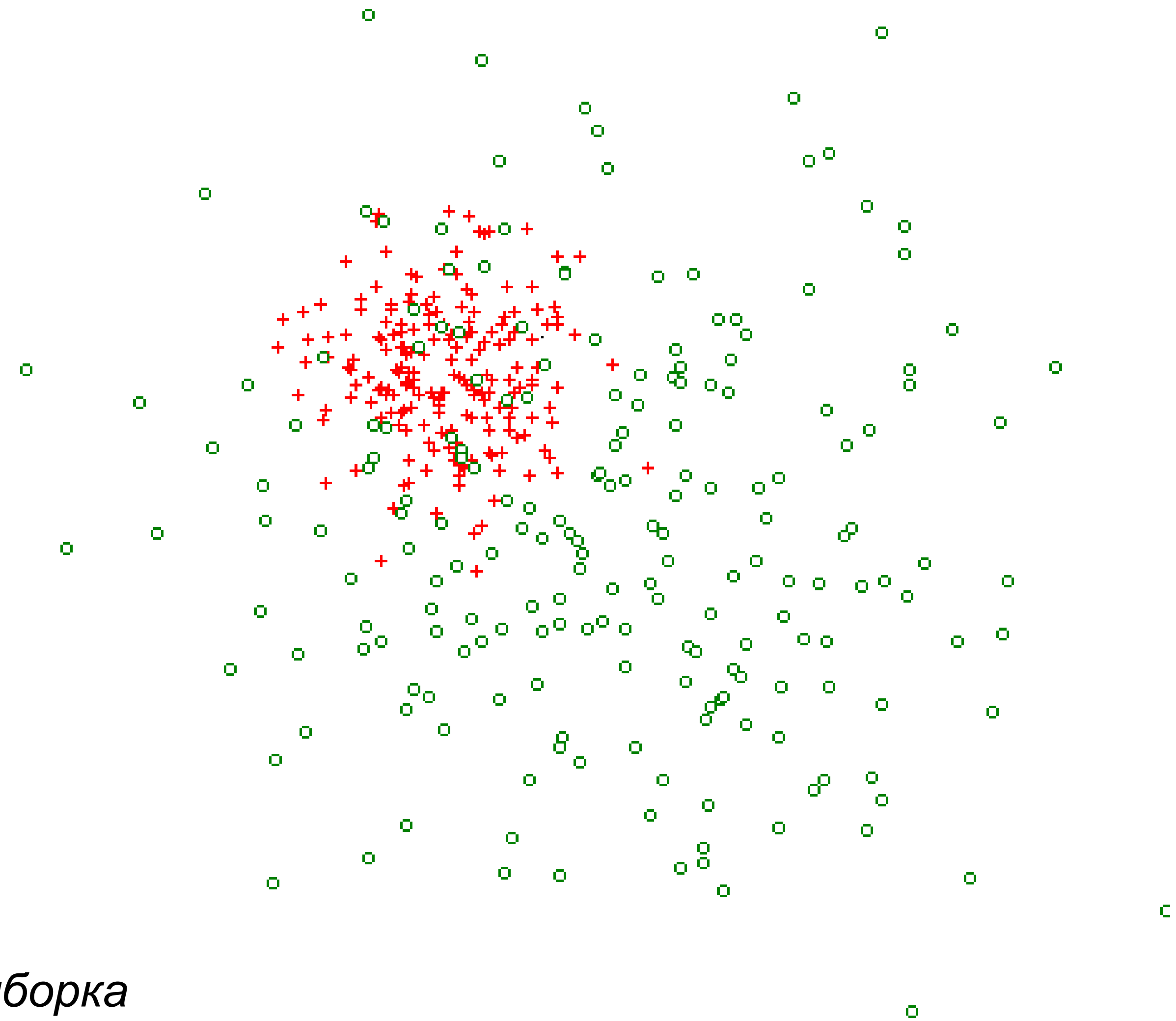
Пример обучения с учителем (supervised learning)



результат переразметки

Минимизация сложности метрических классификаторов методом разреза графов

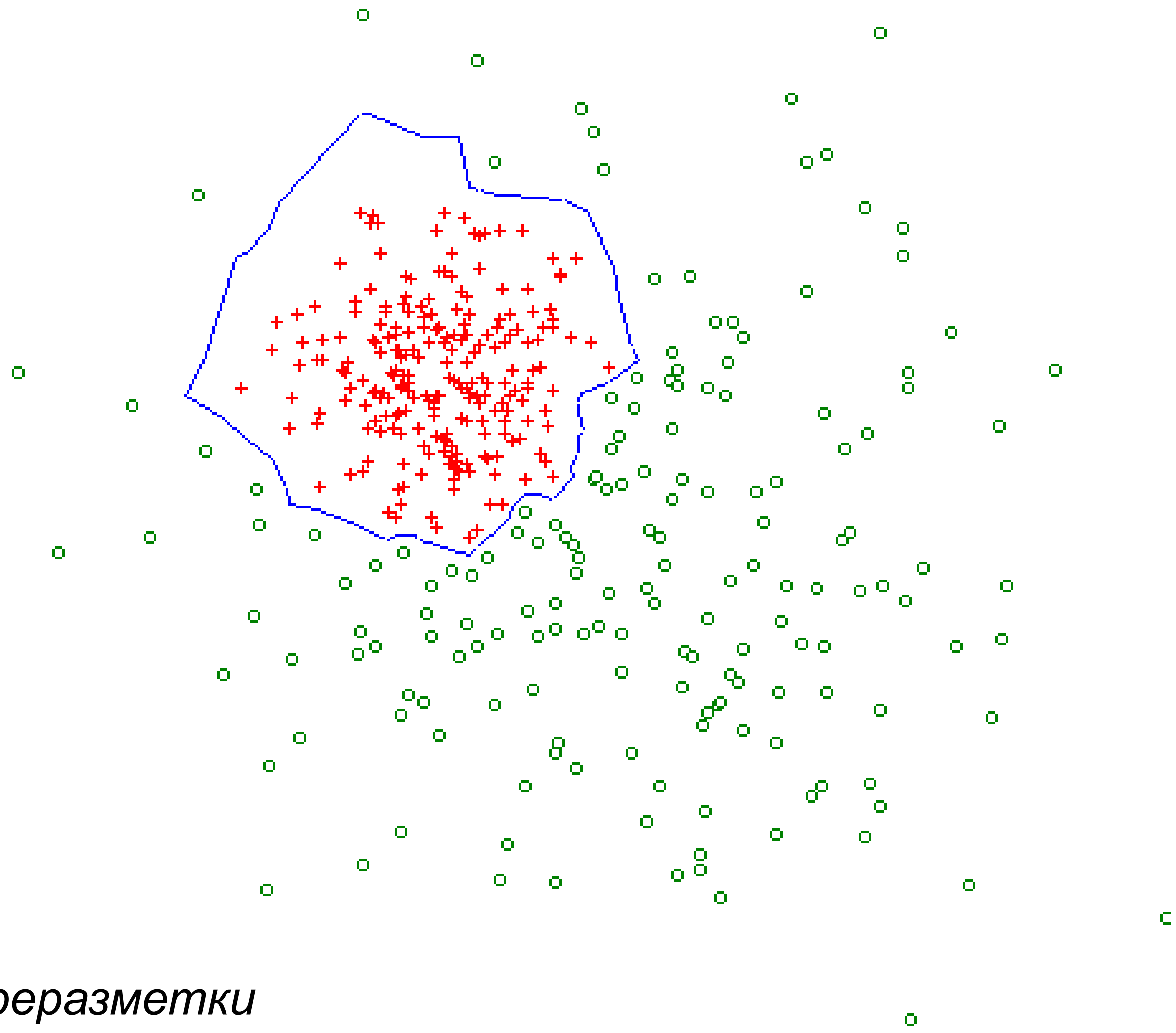
Пример обучения с учителем (supervised learning)



обучающая выборка

Минимизация сложности метрических классификаторов методом разреза графов

Пример обучения с учителем (supervised learning)



результат переразметки

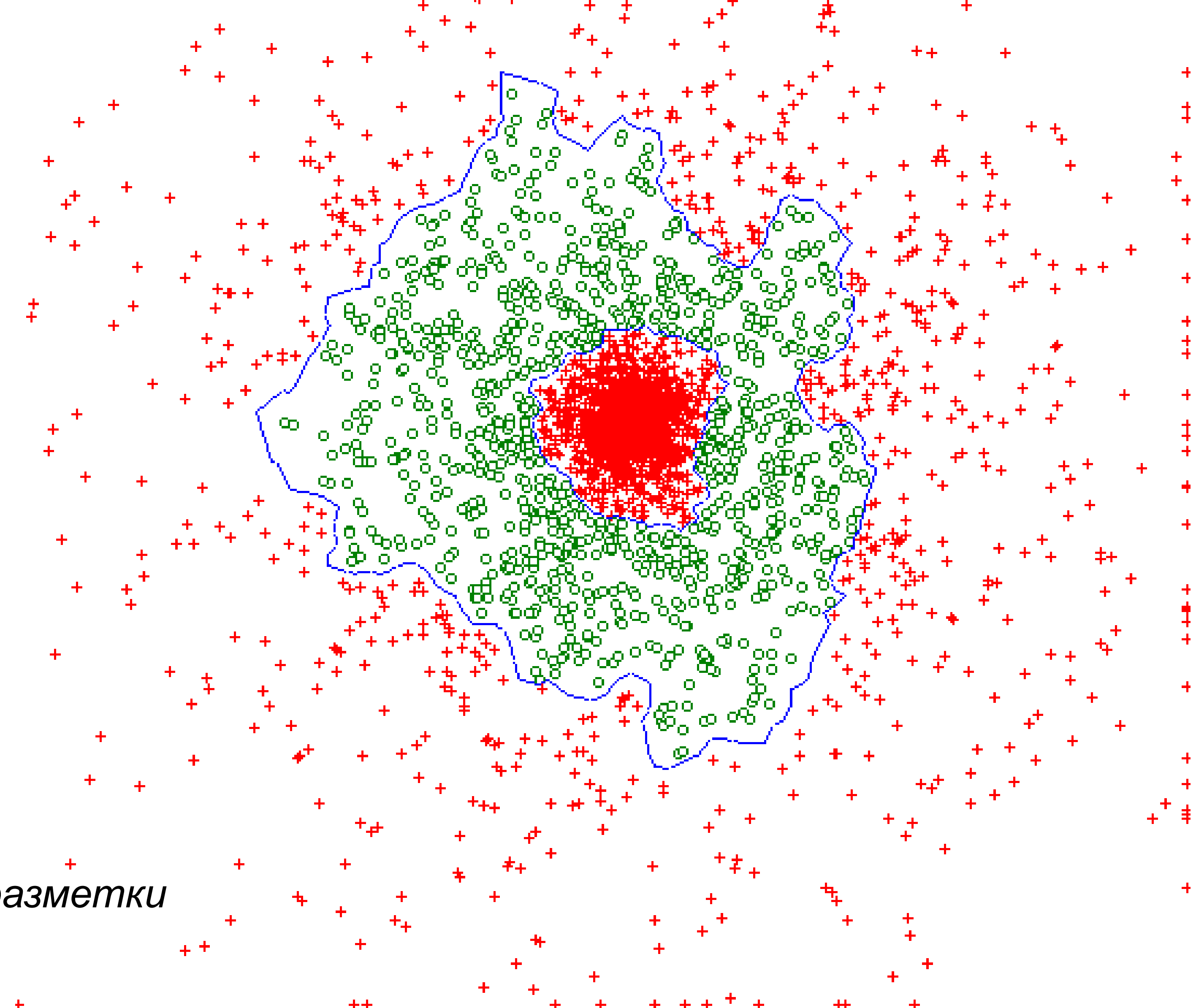
Минимизация сложности метрических классификаторов методом разреза графов

Пример обучения с учителем (supervised learning)

обучающая выборка

Минимизация сложности метрических классификаторов методом разреза графов

Пример обучения с учителем (supervised learning)

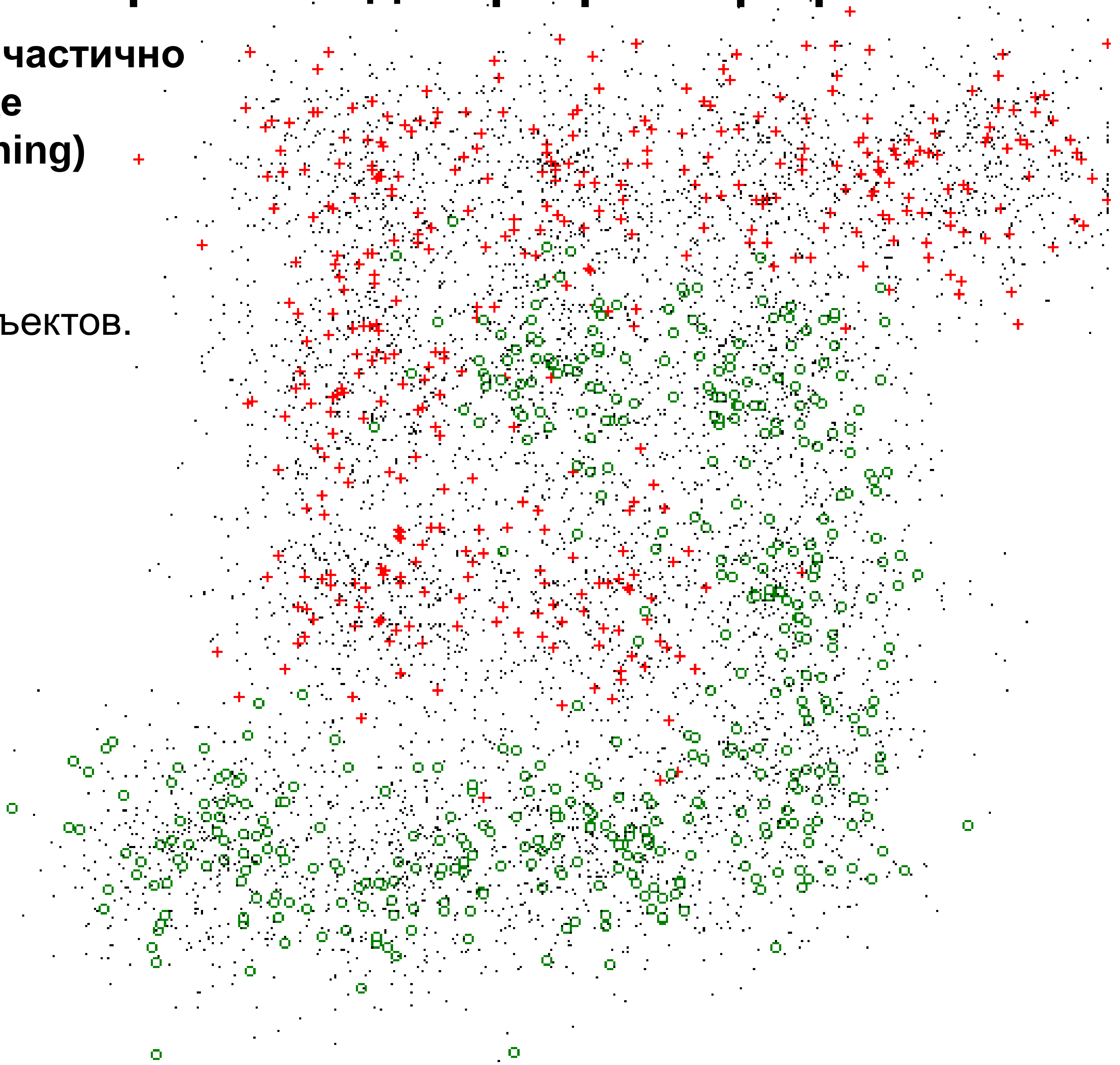


результат переразметки

Минимизация сложности метрических классификаторов методом разреза графов

Пример обучения по частично размеченной выборке (semi-supervised learning)

Размечено 10% точек
из выборки в 10000 объектов.

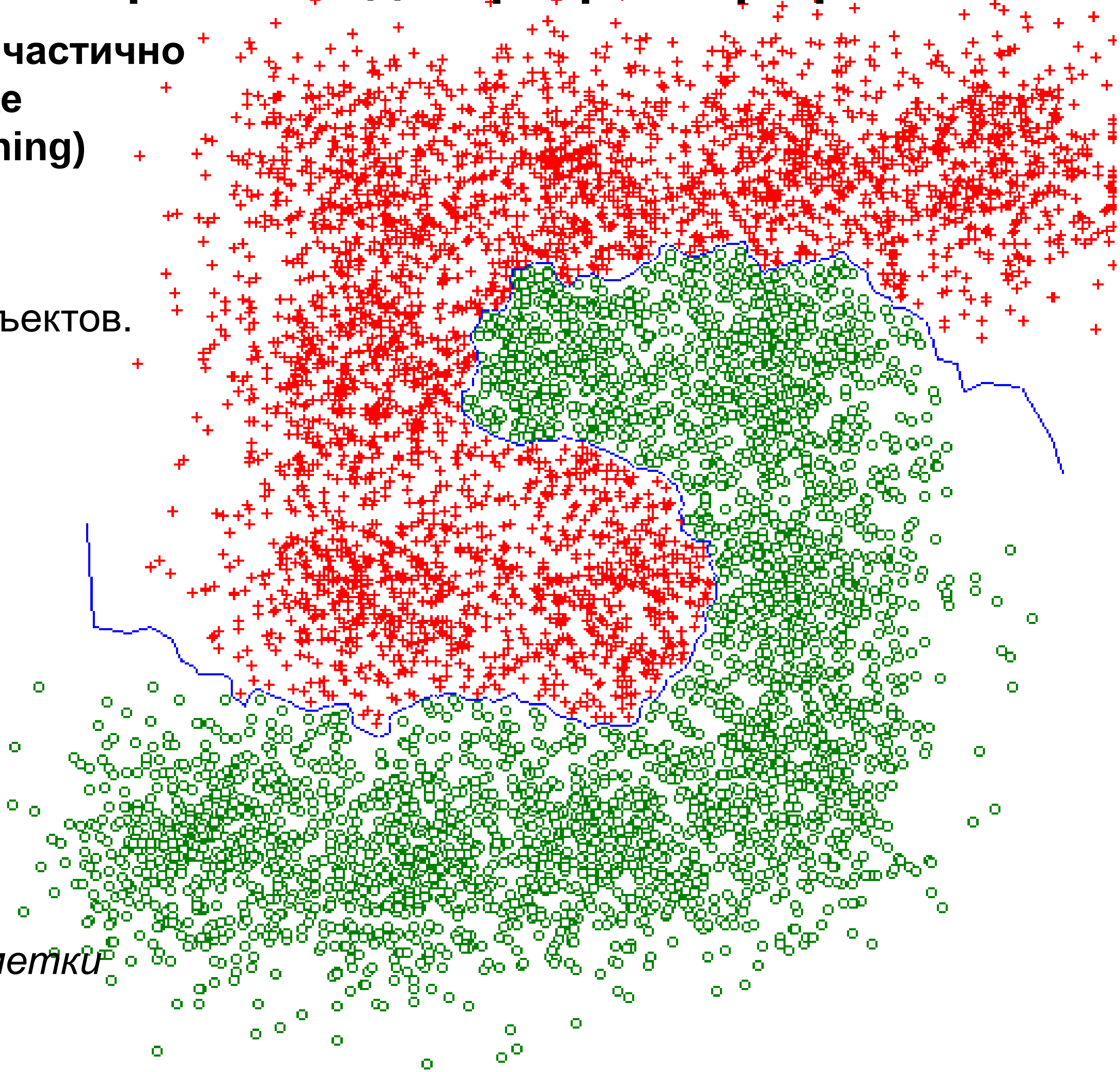


обучающая выборка

Минимизация сложности метрических классификаторов методом разреза графов

Пример обучения по частично размеченной выборке (semi-supervised learning)

Размечено 10% точек
из выборки в 10000 объектов.



результат переразметки

Основные результаты

1. Предложен и обоснован морфологический подход к синтезу классификаторов в задаче обучения с учителем, предполагающий решение двух последовательных подзадач – переразметки обучающей выборки и оптимальной корректной интерполяции (расширения) решающего правила.
2. Доказана проективность морфологических операторов машинного обучения. Показано, что классы морфологических классификаторов и алгоритмов обучения образуют систему вложенных классов возрастающей сложности в смысле Пытьева и могут быть использованы для структурной минимизации риска переобучения.
3. Показано, что для случая двухклассовой классификации метод минимального разреза графов позволяет отыскивать глобальный оптимум введенного критерия качества классификации.
4. Проведено первичное экспериментальное исследование для случая двумерных модельных данных.

Направления дальнейших исследований

- исследование поведения предложенных процедур обучения на многомерных модельных и реальных данных, связанных с задачами распознавания сложных образов;
- исследование подходов к оптимизации метрических свойств пространств описаний на обучающих выборках в задачах распознавания сложных образов.

Спасибо за внимание!