Управление селективностью аномальных наблюдений в обучении одноклассовому распознаванию образов по методу опорных векторов

Ларин А.О., Середин О.С., Моттль В.В.

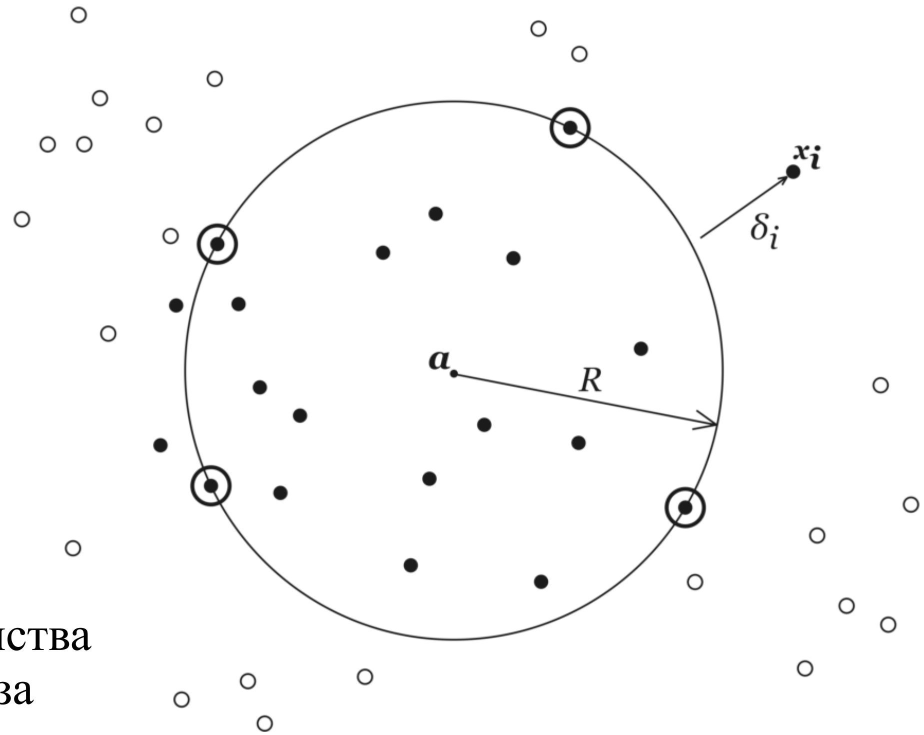
Московский физико-технический институт (ГУ),
Тульский государственный уинверситет

Математические методы распознавания образов – 2015 (ММРО-17)
г. Светлогорск, 2015

Метод описания данных опорными векторами
Support Vector Data Description (SVDD)

Модель описания данных – гиперсфера, заданная центром $\mathbf{a}$ и радиусом R.

Был предложен в диссертации Д. Тэкса в 2001 году



Целевая функция:

$$R^2 + C\sum_{i=1}^{N}\delta_i \to \min_{R^2,\mathbf{a},\delta_1,\ldots,\delta_N},$$

с ограничениями:

$$\rho^2\left(\mathbf{x}_i,\mathbf{a}\right) \le R^2 + \delta_i,$$

$$\delta_i \ge 0, \quad i = 1,\ldots,N,$$

где

N – количество наблюдений

n – размерность признакового пространства

δ_i – величина штрафа объекта за выход за пределы гиперсферы

C – параметр управления селективностью нетипичных объектов

$\rho\left(\mathbf{x}',\mathbf{x}''\right)$ – мера расстояния между двумя объектами в $\mathbb{R}^n$

Метод описания данных опорными векторами
Решение двойственной задачи Лагранжа

Решение двойственной задачи приводит к минимизации функции, с учетом того, что мера между объектами определена как $\rho(\mathbf{x}', \mathbf{x}'') = \left[K(\mathbf{x}', \mathbf{x}') + K(\mathbf{x}'', \mathbf{x}'') - 2K(\mathbf{x}', \mathbf{x}'') \right]^{1/2}$:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j K\left(\mathbf{x}_i \cdot \mathbf{x}_j\right) - \sum_{i=1}^{N} \lambda_i K\left(\mathbf{x}_i \cdot \mathbf{x}_i\right) \to \min_{\lambda_1, \ldots, \lambda_N}$$

со следующими ограничениями:

$$\sum_{i=1}^{N} \lambda_i = 1 \qquad\qquad 0 \le \lambda_i \le C, \quad i = 1, \ldots, N$$

При этом штрафной параметр должен удовлетворять условию:

$$\frac{1}{N} \le C \le 1$$

Метод описания данных опорными векторами
Решающее правило распознавания

Новый объект $\mathbf{z}$ будет отнесен к целевому классу, если расположен на ограничивающей гиперсфере, либо внутри нее

Решающее правило распознавания метода описания данных опорными векторами:

$$d_{SVDD}\left(\mathbf{z}\right) = I\left(\rho^2\left(\mathbf{z},\mathbf{a}\right) \leq R^2\right),$$

где расстояние до центра и радиус вычисляются следующим образом:

$$\rho^2\left(\mathbf{z},\mathbf{a}\right) = K\left(\mathbf{z}\cdot\mathbf{z}\right) - 2\sum_{i=1}^{N_{SV}}\lambda_i K\left(\mathbf{z}\cdot\mathbf{x}_i\right) + \sum_{i=1}^{N_{SV}}\sum_{j=1}^{N_{SV}}\lambda_i\lambda_j K\left(\mathbf{x}_i\cdot\mathbf{x}_j\right),$$

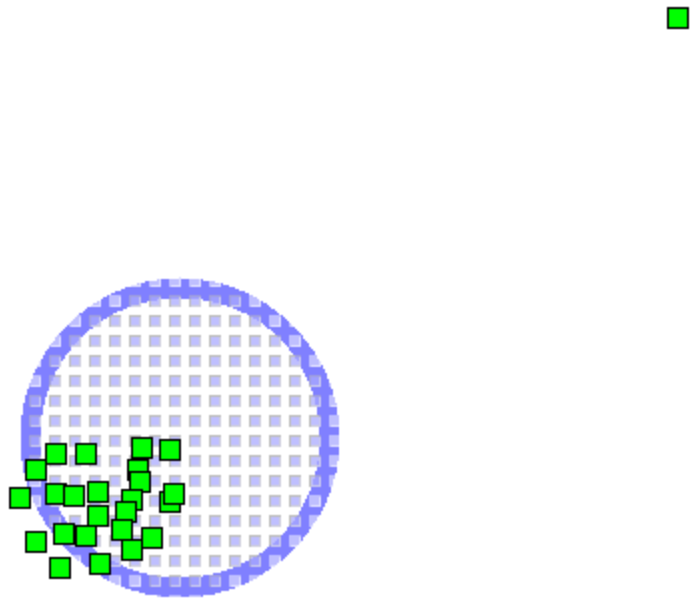
$$R^2 = K\left(\mathbf{x}_k\cdot\mathbf{x}_k\right) - 2\sum_{i=1}^{N_{SV}}\lambda_i K\left(\mathbf{x}_k\cdot\mathbf{x}_i\right) + \sum_{i=1}^{N_{SV}}\sum_{j=1}^{N_{SV}}\lambda_i\lambda_j K\left(\mathbf{x}_i\cdot\mathbf{x}_j\right),$$

где $\mathbf{x}_k$ – некоторый опорный объект

Метод описания данных опорными векторами
Проблема нетипичных объектов в обучающей выборке

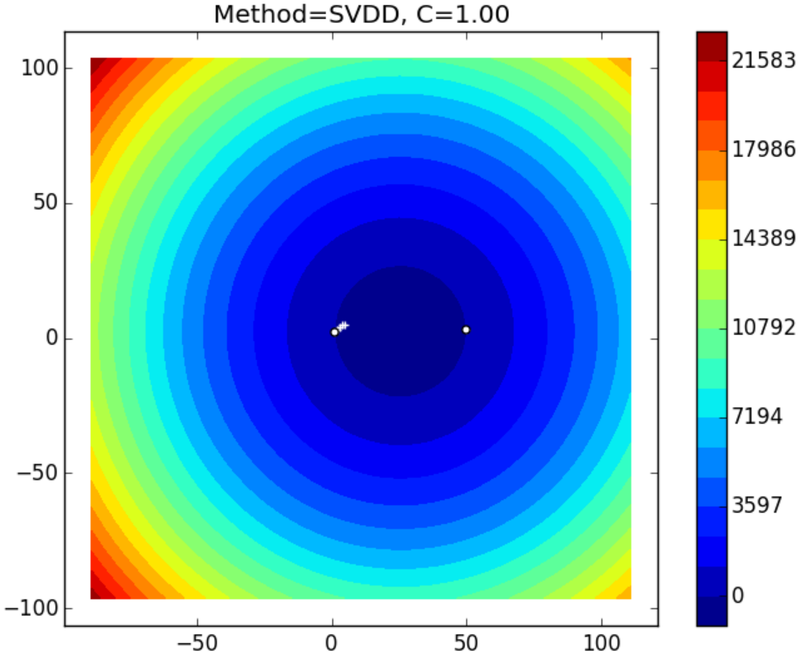
Постановка задачи Д. Тэкса не обеспечивает корректной селекции нетипичных объектов обучающей совокупности – далеких редких выбросов (*outliers*)

$$\begin{cases} R^2 + C\sum_{i=1}^{N}\delta_i \to \min\limits_{R,\mathbf{a},\delta_1,\ldots,\delta_N}, \\ \rho^2\left(\mathbf{x}_i,\mathbf{a}\right) \le R^2 + \delta_i, \quad \delta_i \ge 0, \quad i = 1,\ldots,N \end{cases}$$

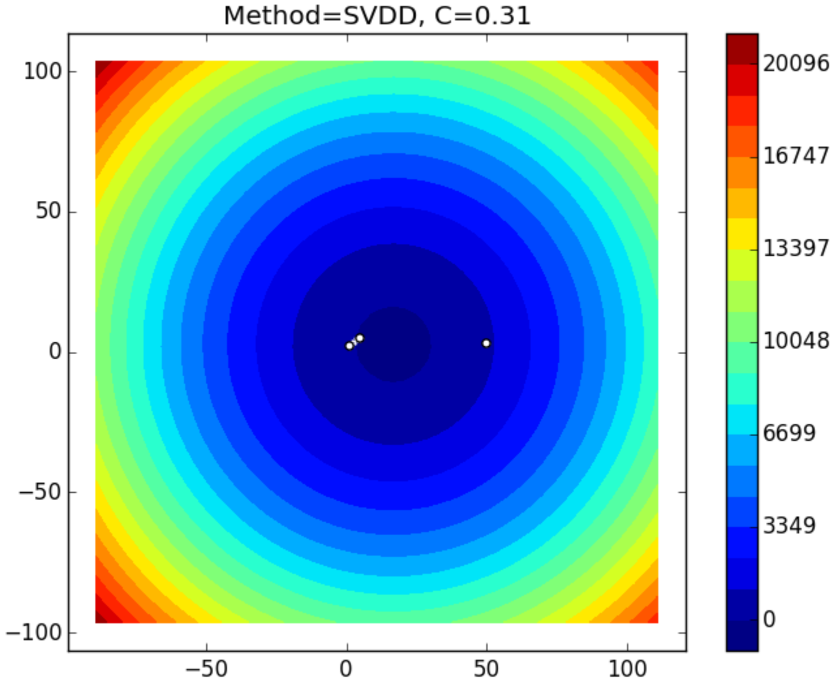


Величина штрафа допустимого выхода объектов обучающей выборки за пределы описывающей гиперсферы <u>является несоизмеримой</u> с расстоянием до её центра в оптимизационной задаче

Метод описания данных опорными векторами
Проблема нетипичных объектов в обучающей выборке



Штрафной параметр C = 1.0
Гиперсфера охватывает всю обучающую выборку



Штрафной параметр C = 0.3
Часть данных из обучающей выборки оказывается вне границы описывающей гиперсферы

Метод описания данных опорными векторами
Нетипичные объекты (outliers) в обучающей выборке

Существующая постановка Тэкса:

Предлагаемая постановка:

Целевая функция:

$$\begin{cases} R^2 + C\sum_{i=1}^{N}\delta_i \to \min\limits_{R,\mathbf{a},\delta_1,\ldots,\delta_N}, \\ \rho^2(\mathbf{x}_i,\mathbf{a}) \le R^2 + \delta_i, \quad \delta_i \ge 0, \quad i = 1,\ldots,N \end{cases}$$

$$\begin{cases} R + C\sum_{i=1}^{N}\delta_i \to \min\limits_{R,\mathbf{a},\delta_1,\ldots,\delta_N}, \\ \rho(\mathbf{x}_i,\mathbf{a}) \le R + \delta_i, \quad \delta_i \ge 0, \quad i = 1,\ldots,N \end{cases}$$

Решающее правило:

$$d_{SVDD}(\mathbf{z}) = I\left(\rho^2(\mathbf{z},\mathbf{a}) \le R^2\right)$$

$$d(\mathbf{z}) = I\left(\rho(\mathbf{z}-\mathbf{a}) \le R\right)$$

где

$\rho(\mathbf{x}',\mathbf{x}'')$ – мера расстояния между двумя объектами в $\mathbb{R}^n$

Метод описания данных опорными векторами
Эквивалентная постановка без ограничений

<u>Теорема</u>. Критерий:

$$\begin{cases} R + C\sum\limits_{i=1}^{N}\delta_i \to \min\limits_{R,\mathbf{a},\delta_1,\ldots,\delta_N}, \\ \rho\left(\mathbf{x}_i,\mathbf{a}\right) \le R + \delta_i, \quad \delta_i \ge 0, \quad i = 1,\ldots,N, \end{cases} \qquad C > 0$$

эквивалентен задаче <u>без ограничений</u>:

$$\boxed{R + C\sum_{i=1}^{N}\max\left(0,\, \rho\left(\mathbf{x}_i,\mathbf{a}\right) - R\right) \to \min_{R,\mathbf{a}}}$$

<u>Теорема</u>. Если $C \ge 1$, то в точке решения $\rho\left(\mathbf{x}_i,\mathbf{a}\right) \le R, \quad \delta_i = 0, \quad i = 1,\ldots,N$

Таким образом, для поиска нетипичных объектов (*outliers*), необходимо задавать параметр $0 < C < 1$

Метод описания данных опорными векторами
Нетипичные объекты (outliers) в обучающей выборке

Рассмотрим следующую задачу: $$R + C\sum_{i=1}^{N}\max\left(0,\, \rho\left(\mathbf{x}_i, \mathbf{a}\right) - R\right) \to \min_{R,\mathbf{a}}$$

Будем искать центр гиперсферы как выпуклую комбинацию объектов обучающей совокупности:

$$\mathbf{a} = \sum_{i=1}^{N}\lambda_i \mathbf{x}_i, \quad \sum_{i=1}^{N}\lambda_i = 1, \quad \lambda_i \geq 0,\ i = 1,\ldots,N.$$

С учетом этого выражения и факта, что функция расстояния между объектами была определена как $\rho(\mathbf{x}', \mathbf{x}'') = \left[K(\mathbf{x}', \mathbf{x}') + K(\mathbf{x}'', \mathbf{x}'') - 2K(\mathbf{x}', \mathbf{x}'')\right]^{1/2}$, эквивалентная постановка задачи примет вид:

$$R + C\sum_{k=1}^{N}\max\left[0,\ \left(K\left(\mathbf{x}_k, \mathbf{x}_k\right) - 2\sum_{i=1}^{N}\lambda_i K\left(\mathbf{x}_k, \mathbf{x}_i\right) + \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j K\left(\mathbf{x}_i, \mathbf{x}_j\right)\right)^{\frac{1}{2}} - R\right] \to \min_{R,\lambda_1\ldots\lambda_N}$$

при условии, что: $$\sum_{i=1}^{N}\lambda_i = 1, \quad \lambda_i \geq 0,\ i = 1,\ldots,N$$

Решающее правило: $$d\left(\mathbf{z}\right) = I\left[\left(K\left(\mathbf{z}, \mathbf{z}\right) - 2\sum_{i=1}^{N}\lambda_i K\left(\mathbf{z}, \mathbf{x}_i\right) + \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j K\left(\mathbf{x}_i, \mathbf{x}_j\right)\right)^{\frac{1}{2}} \leq R\right]$$

Метод описания данных опорными векторами
Нетипичные объекты (outliers) в обучающей выборке

Еще раз критерий обучения (без ограничений):

$$R + C\sum_{k=1}^{N}\max\left[0,\ \left(K(\mathbf{x}_k,\mathbf{x}_k) - 2\sum_{i=1}^{N}\lambda_i K(\mathbf{x}_k,\mathbf{x}_i) + \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j K(\mathbf{x}_i,\mathbf{x}_j)\right)^{1/2} - R\right] \to \min_{R,\lambda_1 \ldots \lambda_N}$$

Этот критерий является выпуклым, но не всюду дифференцируемым, и для его минимизации нельзя применять обычные оптимизационные методы

Прежде, чем решать задачу оптимизации применим метод сглаживающей функции

$$P\left(f(\mathbf{v}),\varepsilon\right) = \begin{cases} \dfrac{1}{2}\varepsilon e^{\frac{f(\mathbf{v})}{\varepsilon}}, & f(\mathbf{v}) \le 0, \\ f(\mathbf{v}) + \dfrac{1}{2}\varepsilon e^{-\frac{f(\mathbf{v})}{\varepsilon}}, & f(\mathbf{v}) > 0 \end{cases}$$

Получим следующую оптимизационную задачу:

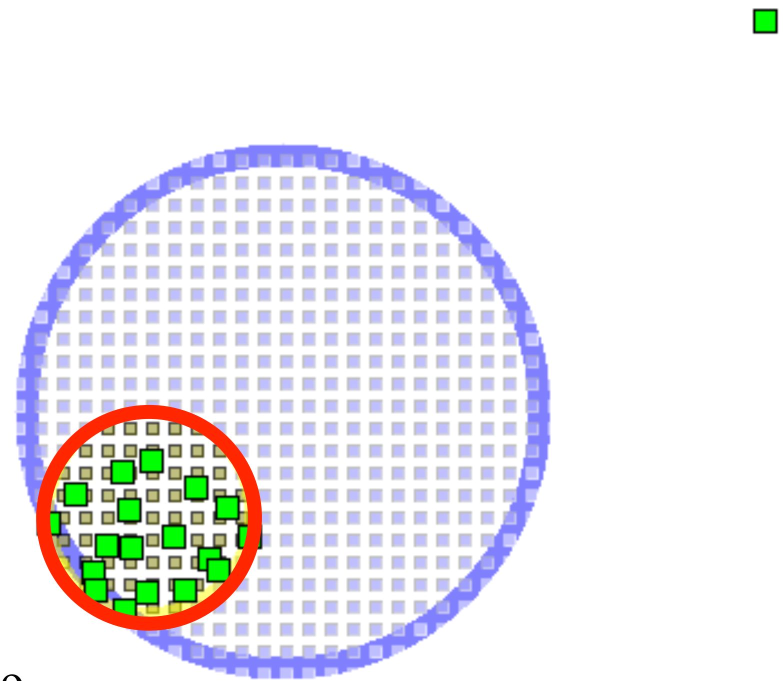
$$\boxed{R + C\sum_{k=1}^{N} P\left(\left[\left(K(\mathbf{x}_k,\mathbf{x}_k) - 2\sum_{i=1}^{N}\lambda_i K(\mathbf{x}_i,\mathbf{x}_k) + \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j K(\mathbf{x}_i,\mathbf{x}_j)\right)^{1/2} - R\right],\varepsilon\right) \to \min_{R,\lambda_1 \ldots \lambda_N \mid \varepsilon}}$$

с ограничениями: $\boxed{\sum_{i=1}^{N}\lambda_i = 1,\quad \lambda_i \ge 0,\ i = 1,\ldots,N}$

где $\varepsilon > 0$ – малый параметр сглаживания

Метод описания данных опорными векторами
Результаты описания данных новым методом

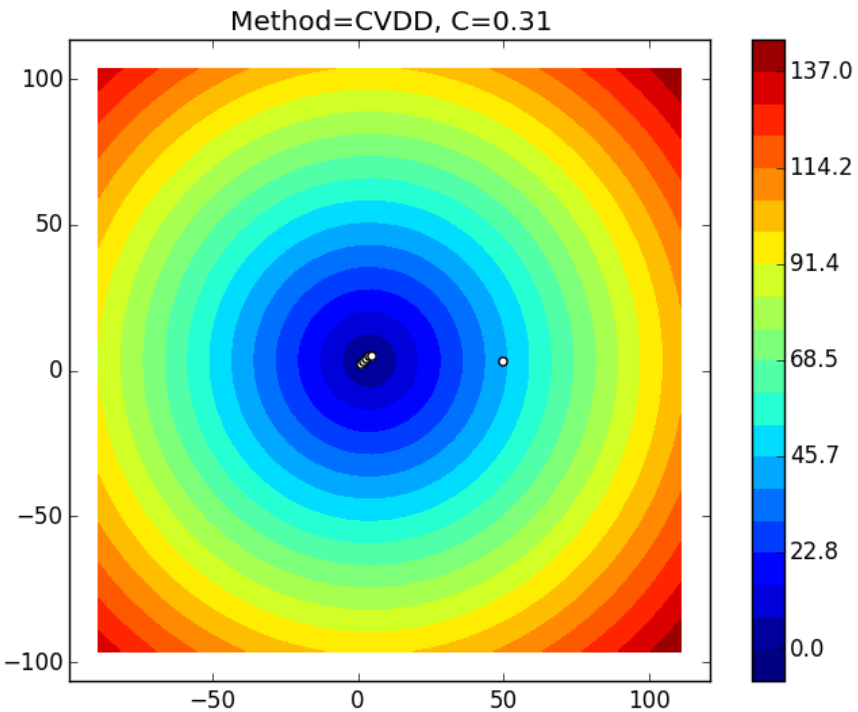
Модельный эксперимент работы классификатора с нетипичными объектами в обучающей выборке в сравнении с оригинальной постановкой задачи Тэкса

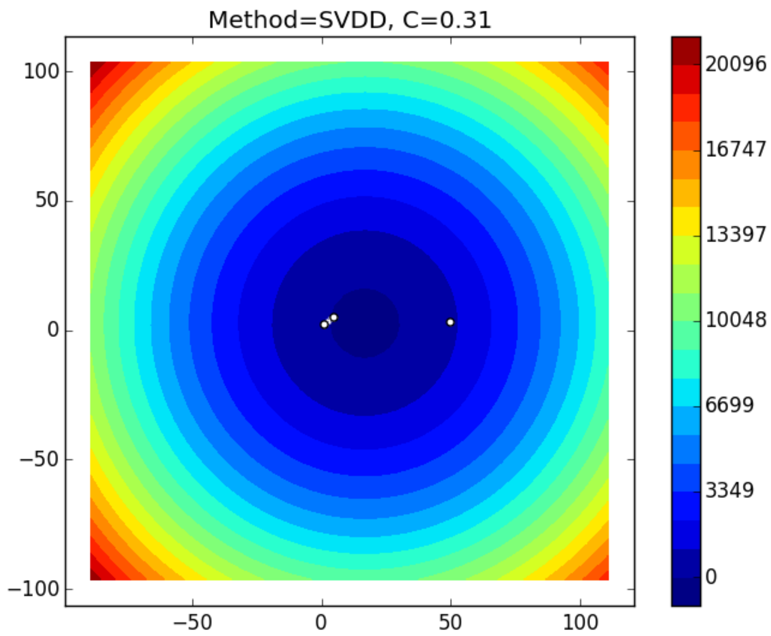


Исходный метод Тэкса SVDD. Недостаточно избирателен относительно аномалий (*синий*)
Предложенный метод учета аномалий (*желтый*)

Метод описания данных опорными векторами
Сравнение результатов работы алгоритмов

Существующая постановка Тэкса





Предлагаемая постановка

Метод описания данных опорными векторами в евклидовых метрических пространствах с аффинными операциями

Пусть Ω – некоторое множество объектов реального мира с заданной на нем метрикой $\rho(\omega', \omega'')$

Метрику будем называть евклидовой, если для любого конечного подмножества объектов $\{\omega_i, i = 1, \ldots, N\} \subseteq \Omega$ матрица $\{-\rho^2(\omega_i, \omega_j), i, j = 1, \ldots, N\}$ условно положительно определена

Абрамов В.И., Середин О.С., Моттль В.В. Обучение распознаванию образов по методу опорных объектов в евклидовых метрических пространствах с аффинными операциями // Известия ТулГУ, Естественные науки, – Тула: Изд-во ТулГУ, Вып. 2, Ч. 1, 2013, С.119-136.

Метод описания данных опорными векторами в евклидовых метрических пространствах с аффинными операциями

<u>Определение.</u> Элемент $\omega_c \subset \tilde{\Omega}$ называют аффинной комбинацией элементов $\left\{\omega_1,\ldots,\omega_N\right\} \subset \tilde{\Omega}$, $\rho\left(\omega_i,\omega_j\right) > 0$, с коэффициентами $\mathbf{c} = (c_1,\ldots c_N) \in \mathbb{R}^N$, $\sum_{i=1}^{N} c_i = 1, \quad c_i \geq 0, i = 1,\ldots,N$ и обозначается $\omega_c = \mathrm{A}_{i=1}^{N} c_i \omega_i$, если

$$\omega_c = \arg\min \sum_{i=1}^{N} c_i \rho^2\left(\omega,\omega_i\right)$$

<u>Теорема.</u> Для любой совокупности $\left\{\omega_1,\ldots,\omega_N\right\} \subset \tilde{\Omega}$ и коэффициентов $\mathbf{c} = (c_1,\ldots c_N) \in \mathbb{R}^N$, $\sum_{i=1}^{N} c_i = 1, \quad c_i \geq 0, i = 1,\ldots,N$ существует единственный элемент $\omega_c \in \tilde{\Omega}$, удовлетворяющий условию выше

<u>Теорема.</u> Пусть $\left\{\omega,\omega_1,\ldots,\omega_N\right\} \subset \tilde{\Omega}$ – конечная совокупность евклидова метрического пространства. Тогда расстояние между ω и заданной аффинной комбинацией $\omega_c = \mathrm{A}_{i=1}^{N} c_i \omega_i, \quad \sum_{i=1}^{N} c_i = 1, \quad c_i \geq 0, i = 1,\ldots,N$ полностью определяется выражением:

$$\boxed{\rho^2\left(\omega_c,\omega\right) = \sum_{i=1}^{N} c_i \rho^2\left(\omega_i,\omega\right) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \rho^2\left(\omega_i,\omega_j\right)}$$

Метод описания данных опорными векторами в евклидовых метрических пространствах с аффинными операциями

По аналогии с признаковым случаем будем решать следующую задачу:

$$R + C\sum_{i=1}^{N}\max\left(0,\,\rho\left(\omega_i,\omega_c\right)-R\right)\to\min_{R,\omega_c}$$

Будем искать центр гиперсферы как аффинную комбинацию объектов обучающей совокупности:

$$\omega_c = \mathrm{A}_{i=1}^{N} c_i\omega_i,\quad \sum\nolimits_{i=1}^{N} c_i = 1,\quad \mathrm{c}_i \ge 0,\, i = 1,\mathrm{K}\;\,, N$$

С учетом этого выражения и факта, что функция расстояния между объектом и центром гиперсферы была определена ранее, эквивалентная постановка задачи примет вид:

$$R + C\sum_{k=1}^{N}\max\left[0,\;\left(\sum_{i=1}^{N} c_i\rho^2\left(\omega_i,\omega_k\right)-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j\rho^2\left(\omega_i,\omega_j\right)\right)^{1/2} - R\right]\to\min_{R,\mathrm{c}_1\mathrm{K}\; c_N}$$

при условии, что: $\displaystyle\sum_{i=1}^{N} c_i = 1,\quad c_i \ge 0,\; i = 1,\mathrm{K}\;\,, N$

Решающее правило: $$d\left(\omega\right) = I\left[\left(\sum_{i=1}^{N} c_i\rho^2\left(\omega_i,\omega\right)-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j\rho^2\left(\omega_i,\omega_j\right)\right)^{1/2} \le R\right]$$

Метод описания данных опорными векторами в евклидовых метрических пространствах с аффинными операциями

Еще раз критерий обучения (без ограничений):

$$R + C\sum_{k=1}^{N}\max\left[0,\quad \left(\sum_{i=1}^{N} c_i\rho^2\left(\omega_i,\omega_k\right) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \rho^2\left(\omega_i,\omega_j\right)\right)^{\frac{1}{2}} - R\right] \to \min_{R,\lambda_1 \ldots \lambda_N}$$

Этот критерий является выпуклым, но не всюду дифференцируемым, и для его минимизации нельзя применять обычные оптимизационные методы

Прежде, чем решать задачу оптимизации применим метод сглаживающей функции

$$P\left(f(\mathbf{v}),\varepsilon\right) = \begin{cases} \frac{1}{2}\varepsilon e^{\frac{f(\mathbf{v})}{\varepsilon}}, & f(\mathbf{v}) \le 0, \\ f(\mathbf{v}) + \frac{1}{2}\varepsilon e^{-\frac{f(\mathbf{v})}{\varepsilon}}, & f(\mathbf{v}) > 0 \end{cases}$$

Получим следующую оптимизационную задачу:

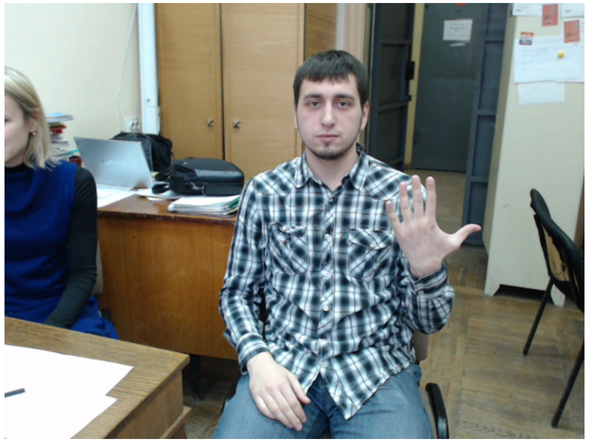
$$\boxed{R + C\sum_{k=1}^{N} P\left(\left[\left(\sum_{i=1}^{N} c_i\rho^2\left(\omega_i,\omega_k\right) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \rho^2\left(\omega_i,\omega_j\right)\right)^{\frac{1}{2}} - R\right],\varepsilon\right) \to \min_{R,c_1 \ldots c_N|\varepsilon}}$$

с ограничениями:

$$\boxed{\sum_{i=1}^{N} c_i = 1,\quad c_i \ge 0,\ i = 1,\ldots,N}$$

где $\varepsilon > 0$ – малый параметр сглаживания

Экспериментальное исследование



Экспериментальное исследование

Таблица расстояний между объектами

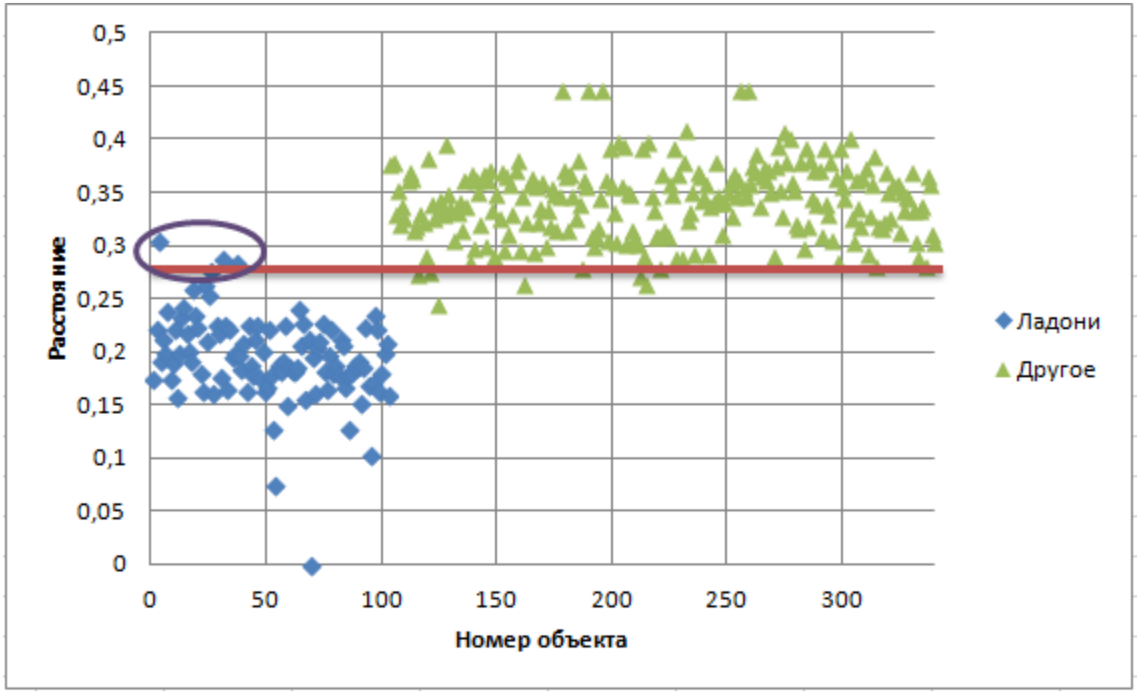
	1	2	3	…	102	103	104	105	…	339	340
1	0	0.10	0.15		0.21	0.22	0.54	0.86		0.51	0.46
2	0.10	0	0.2		0.14	0.26	0.6	0.46		0.52	0.81
3	0.15	0.2	0		0.24	0.16	0.34	0.48		0.36	0.52
…											
102	0.21	0.14	0.24		0	0.23	0.46	0.77		0.51	0.62
103	0.22	0.26	0.16		0.23	0	0.63	0.48		0.54	0.67

Бинарное изображение образца ладони

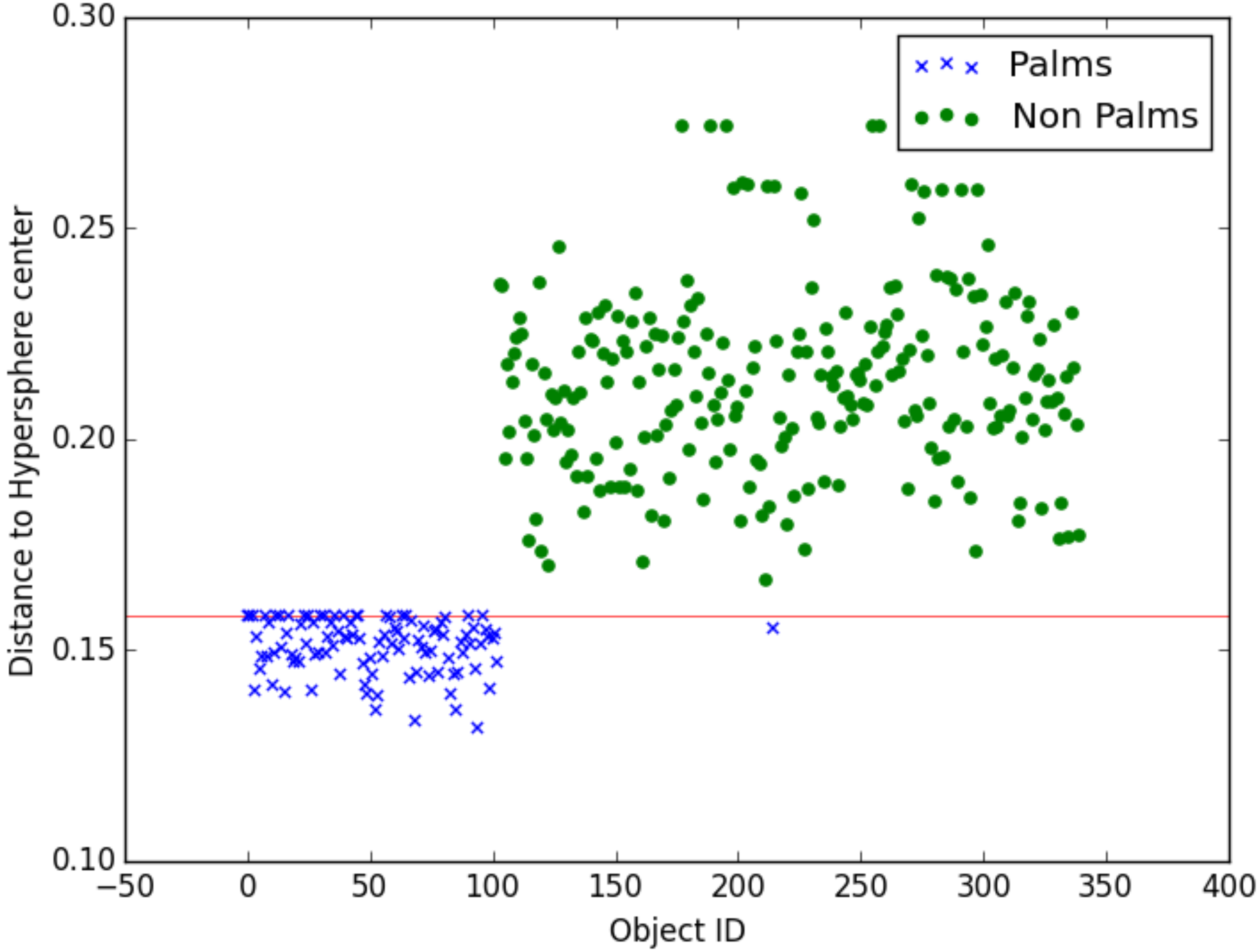
Экспериментальное исследование

Неверно классифицированные объекты.
Ошибка первого рода.

Неверно классифицированные объекты.
Ошибка второго рода.



Экспериментальное исследование
Беспризнаковый одноклассовый классификатор



Спасибо за внимание