

# Мульти моделирование SVM

Сергей Иванович, Александр Адуенко

Московский физико-технический институт  
Факультет управления и прикладной математики  
Кафедра «Интеллектуальные системы»

## Цель исследования

Отобрать оптимальный набор ядер и построить на них композицию SVM. Описать отличия множеств опорных объектов, генерируемые разными ядрами.

### Проблемы

Существующие методы комбинирования алгоритмов плохо работают с малым количеством сильных классификаторов  
Различные ядра могут в реальности быть “похожими” и давать схожие результаты, их нецелесообразно использовать в композиции.

### Предположение

Можно использовать вектора отступов как новые объекты и построить классификатор над ними. Схожесть ядер можно описать с помощью новой метрики, основанной на множествах опорных объектов.

- 1 Rauf Izmailov, Vladimir Vapnik and Akshay Vashist. Multidimensional Splines with Infinite Number of Knots as SVM Kernels. 2013
- 2 Alex J Smola et al. A Tutorial on Support Vector Regression. 2004
- 3 Corinna Cortes and Vladimir Vapnik. Support-Vector Networks. 1995

# Постановка задачи

$X^l = (\mathbf{x}_i, y_i)_{i=1}^l$  — обучающая выборка,  $\mathcal{K} = \{K_i\}_{i=1}^m$  — множество ядер.

## Определение

**Модель** с порядковым номером  $s$  — SVM с ядром  $K_s \in \mathcal{K}$

Совокупность моделей генерирует для обучающей выборки матрицу  $M \in \mathbb{R}^{l \times m}$ , где  $(i, j)$ -й элемент — отступ  $i$ -го объекта на  $j$ -й модели.  $M$  — новая матрица “объект-признак”

Пусть  $\mathcal{A}$  — множество алгоритмов вида

$$\mathcal{A} = \{a(\vec{x}) = g(\vec{x}, \theta) | \theta \in \Theta\} \quad g : \mathbb{R}^m \rightarrow Y$$

# Постановка задачи

## Определение

Мультимодель — пара  $(g, \mathcal{K}')$ , где  $\mathcal{K}' \subset \mathcal{K}$

Тогда перед нами стоит задача отбора ядер из  $\mathcal{K}$  (то есть задача отбора признаков из  $M$ ) и а также выбор оптимального алгоритма для агрегации множества моделей в мультимодель с лучшим качеством классификации и регрессии.

$$L(y, g(M, \theta)) \rightarrow \min_{\mathcal{A}, \Theta}$$

# Цели эксперимента

Необходимо найти способ определять схожие модели, то есть дающие схожие результаты, чтобы не включать таковые в мультимодель.

## Гипотеза

Если множества опорных объектов пары классификаторов похожи, то и векторы отступов похожи.

## Цель эксперимента

Сформулировать понятие “похожести” векторов отступов и различных моделей и проверить гипотезу.

## Эксперимент. Ядра и данные.

В качестве исходных данных взяты датасеты German Credits, Wine и Heart disease из UCI.

Ядра:

- Линейное
- Полиномиальное (степени 3, 4, 5)
- RBF-ядро ( $\gamma \in \{0.0001, 0.001, 0.01, 0.1, 1\}$ )
- INK-spline ядро

## Эксперимент. Метрики

Расстояние между ядрами представим в виде нормализованной симметрической разности:

$$\rho_{X'}(K_i, K_j) = \frac{\# [SV_i \Delta SV_j]}{\# [SV_i \cup SV_j]}$$

В качестве меры сходства классификаторов возьмем отнормированную корреляцию Пирсона векторов отступов.

$$\rho_{X'}(M_i, M_j) = 1 - \text{corr}(M_i, M_j)$$

Проанализируем эволюцию распределения пар расстояний в зависимости от параметра регуляризации.



# Эксперимент. German credit

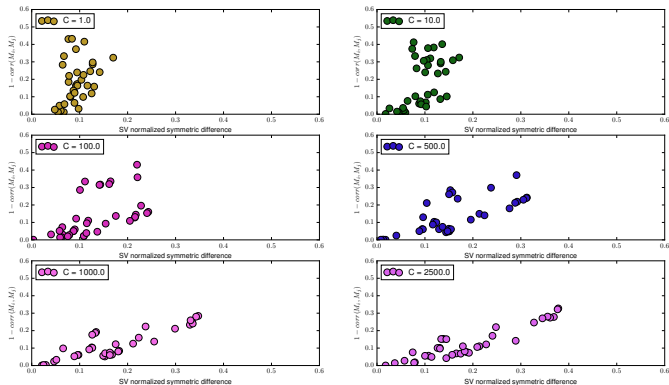


Рис.: German credit

## Эксперимент. German credit

Таблица: German info

	$\langle \#SV \rangle$	$\langle \rho(M_i, M_j) \rangle$	$\langle \rho(K_i, K_j) \rangle$	Correlation
$C = 1.0$	603.4	0.184	0.094	0.376
$C = 10.0$	603.6	0.187	0.097	0.537
$C = 100.0$	594.7	0.134	0.131	0.556
$C = 500.0$	584.3	0.133	0.161	0.717
$C = 1000.0$	581.6	0.120	0.172	0.870
$C = 2500.0$	577.9	0.126	0.189	0.918

# Эксперимент. Wine

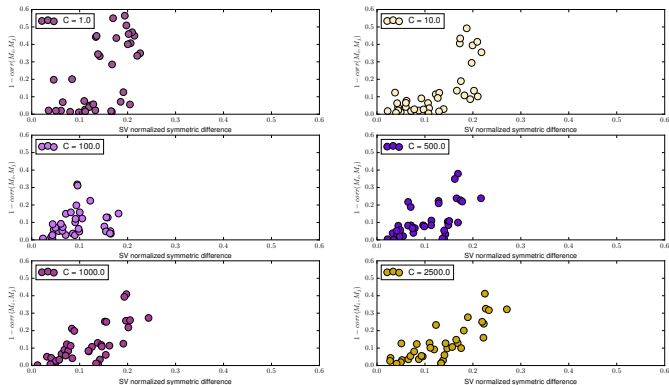


Рис.: Wine

# Эксперимент. Wine

Таблица: Wine info

	$\langle \#SV \rangle$	$\langle \rho(M_i, M_j) \rangle$	$\langle \rho(K_i, K_j) \rangle$	Correlation
$C = 1.0$	3284.1	0.220	0.144	0.600
$C = 10.0$	3284.9	0.130	0.121	0.687
$C = 100.0$	3275.0	0.091	0.091	0.270
$C = 500.0$	3252.6	0.110	0.105	0.591
$C = 1000.0$	3235.2	0.124	0.118	0.694
$C = 2500.0$	3208.6	0.127	0.133	0.795

# Эксперимент. Heart disease

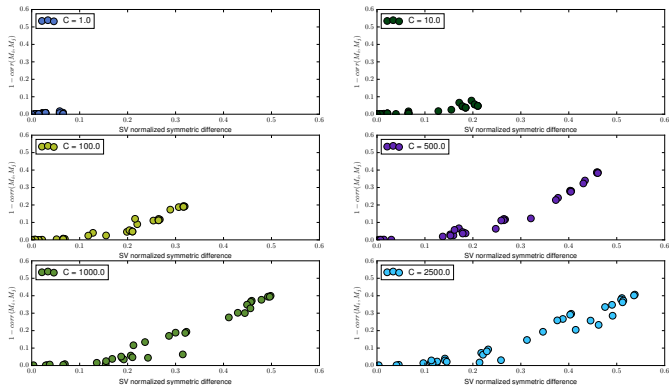


Рис.: Heart disease

## Эксперимент. Heart disease

Таблица: Heart info

	$\langle \#SV \rangle$	$\langle \rho(M_i, M_j) \rangle$	$\langle \rho(K_i, K_j) \rangle$	Correlation
$C = 1.0$	272.0	0.003	0.027	0.608
$C = 10.0$	260.8	0.020	0.088	0.929
$C = 100.0$	249.1	0.063	0.152	0.927
$C = 500.0$	231.9	0.135	0.238	0.940
$C = 1000.0$	223.1	0.157	0.268	0.953
$C = 2500.0$	211.4	0.166	0.297	0.962

# Результаты

- С ростом константы регуляризации расстояние между ядрами и расстояние между их отступами лучше коррелируют между собой.
- При высоких параметре регуляризации коэффициент корреляции Пирсона достигает более 0.8, то есть расстояния практически линейно зависят друг от друга.
- Вектора средних ядерных и отступных расстояний коррелируют по-разному на различных датасетах (на Wine и Heart корреляции Пирсона 0.85 и 0.99 соответственно, на German —  $-0.92$ ).

Мы показали, что на примере довольно разных задач выполняется поставленная гипотеза.

## Результат

Если множества опорных объектов пары классификаторов похожи, то и векторы отступов похожи