УДК 004.052'4

А.О. Гурина[1,2], В.Л. Елисеев[1,2]

ИССЛЕДОВАНИЕ МНОГОКЛАССОВОЙ КЛАССИФИКАЦИИ ЗА ПРЕДЕЛАМИ ОБУЧАЮЩЕЙ ВЫБОРКИ

[1] – Национальный исследовательский университет «МЭИ», г. Москва

[2] – АО «ИнфоТеКС», г. Москва

Аннотация. Исследуется распространённая проблема многоклассовой классификации на основе моделей машинного обучения. В виду непредсказуемости классификации объектов за пределами обучающей выборки классификаторы могут работать некорректно на новых данных, а также могут быть уязвимы к состязательным атакам. Выдвигается предположение о том, что при достаточно полной оценке качества классификатора этих проблем можно избежать. Анализируется эффективность применения традиционного подхода к оценке качества классификации. Описываются недостатки традиционных показателей качества, которые не позволяют оценить риск возникновения ошибок и степень подверженности модели машинного обучения состязательным атакам. Предлагается новый критерий качества многоклассовой классификации, включающий четыре показателя. Вычисление показателей основано на соотношении размеров области пространства, занимаемого обучающей выборкой и результатами классификации всех точек дискретизированного пространства признаков в рабочем диапазоне их значений. Проводится экспериментальное исследование для визуальной оценки и сравнения качества двух многоклассовых SVM классификаторов на характерных синтетических наборах данных с помощью традиционных и предлагаемых показателей качества. Демонстрируются эффективность и преимущество введенных показателей по сравнению с традиционными. Подтверждается хорошая интерпретируемость значений показателей качества, а также субъективное соответствие метрик ожидаемым результатам сравнения двух SVM классификаторов. Есть основания полагать, что применение нового подхода к оценке качества позволит строить более надёжные классификаторы на основе машинного обучения.

Ключевые слова: *показатели качества классификации, многоклассовая классификация, количественная оценка качества, критерий качества классификации, пространство признаков, машинное обучение, состязательные атаки, SVM, матрица ошибок, достоверность, точность, полнота.*

Введение

Методы машинного обучения используются в системах принятия решений различных прикладных областей [1]: для биометрической идентификации, в промышленности для управления производством и обнаружения угроз безопасности, в медицине для диагностики заболеваний, в маркетинге для персонализированной рекламы, в сфере информационной безопасности для обнаружения аномалий и кибератак и т. п. Например, в 2020 году 34% компаний в Европе, США и Китае применяли машинное обучение, а по оценкам экспертов, к 2024 году использование машинного обучения вырастет ещё на 42% [1].

Рост спроса на решения на основе машинного обучения связан с растущим внедрением облачных сервисов, увеличением объема неструктурированных данных и с потребностью в глобальной автоматизации процессов. Однако согласно исследованию IBM [2] повсеместному внедрению машинного обучения препятствует недоверие к технологии.

Действительно, множество известных прецедентов некорректной работы машинного обучения [3]- [5] не позволяют полностью полагаться на этот инструмент при решении задач с высокой ценой ошибки. Очевидно, что в некоторых системах ошибки алгоритмов могут нанести серьезный ущерб. Также, справедливо, что для использования технологии, призванной заменить умственный и физический труд человека, должна быть уверенность, что классификатор на основе машинного обучения будет давать предсказуемый и корректный ответ на любых данных, которые ему подаются на вход.

В связи с этим на первый план выходят критерии для оценки качества полученной модели машинного обучения, без которых, в условиях многомерных данных, невозможно оценить эффективность и надежность построенной модели или сравнить между собой два различных алгоритма классификации.

Широко распространенная задача, решаемая алгоритмами машинного обучения, это классификация – отнесение наблюдаемого объекта к тому или другому классу для принятия последующего решения автоматически или человеком. Для оценки качества построенной модели классификатора принято использовать следующие показатели:

1) confusion matrix,
2) accuracy,
3) precision,
4) recall,
5) f-score.

Традиционный метод оценки классификатора путем сравнения значений общепринятых показателей качества с эталонными обладает одним серьезным ограничением – такая оценка в полной мере

справедлива только для проверяемых объектов ограниченного тестового набора. Даже если показатели качества близки к эталонным на тестовом наборе, это не означает, что классификатор будет работать корректно на новых данных. Таким образом, есть основания полагать, что традиционные критерии качества неполны и не позволяют оценить качество классификатора за пределами набора обучающих и тестовых данных. Именно эта неполнота и приводит к неожиданным ошибкам высококачественных, по мнению разработчиков, классификаторов.

Представляется актуальным провести исследование достаточности традиционных показателей качества для создания предсказуемых классификаторов, а также разработать критерий, позволяющий снизить риск ошибок и повысить доверие к использованию методов машинного обучения в прикладных задачах.

Обзор литературы

Для оценки качества моделей машинного обучения и сравнения различных алгоритмов принято использовать стандартные показатели качества, рассчитываемые на основе составляющих матрицы ошибок (confusion matrix) [6]. Матрица ошибок для оценки качества алгоритма двухклассовой классификации (binary classification) показана в таблице 1, причем в ячейках таблицы указывается количество фактов корректной (True) и некорректной (False) классификации для каждого из классов, традиционно называемых положительным (Positive) и отрицательным (Negative).

Табл.1. Матрица ошибок

Класс / Классификация	Positive	Negative
Positive	TP	FP
Negative	FN	TN

Основные показатели, которые необходимо учитывать при оценке качества алгоритма классификации – это рассчитываемые на основе элементов матрицы:

- ошибка первого рода (α), оценивающая долю ложных срабатываний алгоритма классификации (False Positive, FP),

$$\alpha = \frac{FP}{FP + TN}$$

- ошибка второго рода (β), оценивающая долю ложно отвергнутых примеров (False Negative, FN),

$$\beta = \frac{FN}{FN + TP}$$

- мощность критерия - характеристика способности критерия не упустить значимое событие: $1 - \beta$.

Большинством разработчиков систем на основе машинного обучения используется следующая методология оценки качества моделей:

- Рассчитывается матрица ошибок;
- На основе элементов матрицы ошибок рассчитываются показатели качества:
  - Доля правильно классифицированных примеров,
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
  - Доля положительно классифицированных примеров, действительно являющихся положительными,
$$Precision = \frac{TP}{TP + FP}$$
  - Доля примеров положительного класса распознанных среди всех примеров положительного класса,
$$Recall = \frac{TP}{TP + FN}$$
- На основе вышеупомянутых показателей качества рассчитывается агрегированный критерий качества:
  - Среднее гармоническое Precision и Recall,
$$F_1 = 2\frac{Precision \cdot Recall}{Precision + Recall}$$
- На основе сравнения значения F_1 и эталонного значения, равного 1, принимается решение о достижении требуемого уровня качества модели.

Кроме этих показателей некоторыми исследователями вводятся и другие, позволяющие более эффективно характеризировать те или иные аспекты алгоритмов машинного обучения, например, [6] и [7]. Также для оценки качества используется так называемые ROC-кривые [8].

Несмотря на принятую методологию, известны случаи некорректной работы введенных в эксплуатацию алгоритмов классификации, высоко оцениваемых с её помощью. Согласно исследованию [4] большой класс классификаторов имеет фундаментальный недостаток – непредсказуемость результата классификации для объектов за пределами обучающей выборки. Досадные ошибки такого рода наблюдаются в системах распознавания объектов [9], компьютерного зрения [10], обработки естественного языка [11] и других системах на основе машинного обучения.

Более того, встал вопрос защиты алгоритмов классификации от состязательных атак злоумышленников, целенаправленно формирующих примеры, вызывающие ошибки классификаторов [5]. Например,

известно, что маленькие наклейки, прикрепленные к стандартному дорожному знаку остановки, заставляют систему компьютерного зрения автономного транспортного средства ошибочно идентифицировать его как знак ограничения скорости [10]. В работе [12] показано, что системы машинного обучения, получающие входные данные с камер и других датчиков, также уязвимы для состязательных атак. А для обмана системы распознавания лиц [13] достаточно одеть специальные очки. Вышеприведенные примеры неудач машинного обучения показывают, что многие даже простые алгоритмы машинного обучения могут вести себя совсем не так, как предполагают разработчики.

Исследователями были предложены некоторые методы борьбы с ошибками классификаторов, среди которых состязательное обучение, автоэнкодеры. Однако ни один способ не может полностью защитить системы машинного обучения от атак этого типа и на текущий момент нет какого-либо общепринятого решения. Как обсуждалось в [14], возможно, состязательные примеры неизбежны, особенно при высокой размерности данных.

В связи с этим разработка способов предотвращения ошибок классификаторов является сейчас важной областью исследований. Необходимо, чтобы модели машинного обучения давали адекватные результаты для всех возможных входных данных.

Справедливо отметить, что возможная причина таких неожиданных ошибок в том, что на этапе оценки качества полученной модели не осуществляется проверка наличия вблизи области целевого класса примеров, не относящихся к классу, но для которых ответ классификатора - положительный. Для такой проверки не обязательно строить ещё одну нейронную сеть для состязательного обучения, достаточно пересмотреть способ оценки качества моделей машинного обучения.

В большинстве случаев модели машинного обучения работают достаточно точно, но только на небольшом количестве входных данных относительно всех возможных, с которыми они могут столкнуться при эксплуатации. Оценка качества многоклассового классификатора в большинстве практических задач осуществляется на основе расчёта показателей для тестового набора примеров, зачастую существенно меньшего ограниченного счетного обучающего множества, что обычно обусловлено труднодоступностью наборов данных для обучения и тестирования. На ограниченной известной части примеров можно добиться высокой точности классификатора согласно традиционным показателям, но значит ли это, что классификатор будет обеспечивать такое же высокое качество и принимать корректные решения на новых данных, которых не было в тестовой или обучающей выборке? Данный вопрос будет исследован в разделе Эксперименты данной статьи.

Очевидно, что традиционный подход к оценке качества классификаторов не может обеспечить полного доверия к показателям качества и результатам применения классификатора в практических задачах, особенно с дрейфом данных, в силу ограниченности набора данных для тестирования и расчета показателей качества на значениях элементов матрицы ошибок. Одним из возможных решений проблемы надежности использования моделей машинного обучения является разработка нового более объективного подхода к оценке качества моделей классификации.

Ранее в [15] была рассмотрена задача оценки качества одноклассового классификатора и предложены показатели для оценки качества, оперирующие не количеством примеров в выборке, а областями, покрывающими точки обучающего множества. Такой подход устраняет недостатки, свойственные общеизвестным критериям качества многоклассовой классификации, и позволяет повысить достоверность оценки качества и доверие к решениям классификаторов. Далее в работе будет рассмотрено обобщение введенных в [15] показателей для оценки алгоритмов многоклассовой классификации.

Методология

Для оценки алгоритмов многоклассовой классификации предлагается обобщить введенный ранее критерий качества одноклассовой классификации [15] на множество классов. Новый критерий качества оперирует дискретными оценками объемов, занимаемыми данными в пространстве признаков для классификации, и включает четыре новых показателя: $Approx$, $Excess$, $Deficit$, $Coating$, которые, в случае оценки многоклассового классификатора, рассчитываются для каждого класса.

Под оценкой объема данных понимается сумма объемов атомарных ячеек дискретного разбиения ограниченной области в пространстве признаков, таких, что в каждой из ячеек есть хотя бы одна из точек множества данных. Будем обозначать объем счетного ограниченного множества данных X^* с разбиением в области Ω на атомарные ячейки размером h как $|X^*|_{\Omega,h}$. В том случае, если все множества данных рассматриваются в одной и той же области разбиения и с одним и тем же размером атомарной ячейки, то индекс Ω, h можно опустить. Область разбиения также будет обозначаться областью сканирования.

Показатели рассчитываются на основе объемов, занимаемых обучающим $|X_T^*|$ и классифицированным $|X_D^*|$ множеством в пространстве признаков, по следующим формулам:

$$Excess = \frac{|X_D^* \backslash X_T^*|}{|X_T^*|} \quad (1)$$

$$Deficit = \frac{|X_T^* \backslash X_D^*|}{|X_T^*|} \quad (2)$$

$$Coating = \frac{|X_T^* \cap X_D^*|}{|X_T^*|} \quad (3)$$

$$Approx = \frac{|X_T^*|}{|X_D^*|} \quad (4)$$

При этом под указанными объемами понимается сумма объемов атомарных элементов разбиения пространства, затронутых имеющимися элементами обучающего $\hat{X}_T$ и классифицированного $\hat{X}_D$ множества. Классифицированное множество $\hat{X}_D$ – это множество сгенерированных точек пространства сканирования, которые были отнесены обученным классификатором к целевому классу, то есть деформированное классификатором целевое множество.

Для расчета показателя $Excess$ объем $|X_D^* \backslash X_T^*|$ определяется как сумма объемов атомарных элементов разбиения пространства (ячеек), в которые входят элементы $\hat{X}_D$, за исключением ячеек, в которые входят также элементы $\hat{X}_T$.

Для расчета показателя $Deficit$ объем $|X_T^* \backslash X_D^*|$ определяется как сумма объемов ячеек, затронутых элементами множества $\hat{X}_T$, но не затронутых элементами множества $\hat{X}_D$.

Для расчета показателя $Coating$ объем $|X_T^* \cap X_D^*|$ рассчитывается как сумма объемов ячеек, затронутых как элементами множества $\hat{X}_T$, так и $\hat{X}_D$.

Особенностью предложенного критерия, которую необходимо учитывать, является влияние выбора размера атомарной ячейки разбиения на значения показателей.

Введенные показатели интерпретируются следующим образом:

1) Если классификатор неправильно классифицирует объекты за пределами целевого класса, показатель $Excess$ (англ. избыток) примет значение больше 0 (аналог α);
2) Если классификатор неправильно классифицирует объекты в пределах целевого класса, показатель $Deficit$ (англ. недостаток) примет значение больше 0 (аналог β);
3) Если классификатор правильно классифицирует все объекты в пределах целевого класса, показатель $Coating$ принимает значение 1 (аналог $1 - \beta$);
4) Точность классификатора с точки зрения аппроксимации целевого множества – $Approx$, оценивается как отношение объема целевого множества к объему множества, отнесенного обученным классификатором к целевому классу.

Идеальный одноклассовый классификатор характеризуется следующими значениями показателей:

$$Excess = 0, \quad Deficit = 0, \quad Coating = 1, \quad Approx = 1 \quad (5)$$

Поскольку предложенный критерий легко обобщается на любое число классов, то методология оценки многоклассового классификатора выглядит следующим образом:

1) Сканирование расширенной области обучающего пространства с некоторым шагом сетки h;
2) Получение ответов классификатора для примеров обучающего множества $\hat{X}_T$;
3) Расчет объема $|X_T^*|$ для каждого класса;
4) Получение ответов классификатора для примеров множества сканирования;
5) Расчет объема $|X_D^*|$ для каждого класса;
6) Расчет показателей $Approx$, $Excess$, $Deficit$, $Coating$ для каждого класса;
7) Принятие решения относительно качества классификатора путем сравнения значений показателей качества для каждого класса со значениями, установленными в (5).

В отличие от традиционного подхода, качество классификатора будет оцениваться не по агрегированным показателям качества классификации объектов всех классов, а по четырем показателям, отражающим точность классификации объектов каждого класса в отдельности. Такой подход представляется более точным и надежным.

Эксперименты

Для наглядного сравнения объективности введенных показателей с традиционными был проведен эксперимент по оценке качества методов SVM с разными параметрами по двум методологиям:

1) Традиционная: Матрица ошибок, $Recall$, $Precision$ и $F-score$ для классификатора;
2) Предлагаемая: $Approx$, $Excess$, $Deficit$, $Coating$ для каждого класса.

В качестве исходных данных был взят сгенерированный набор двумерных данных и разделен на обучающее и тестовое множество. Обучающее множество состоит из трех классов, которые представлены на рис. 1.

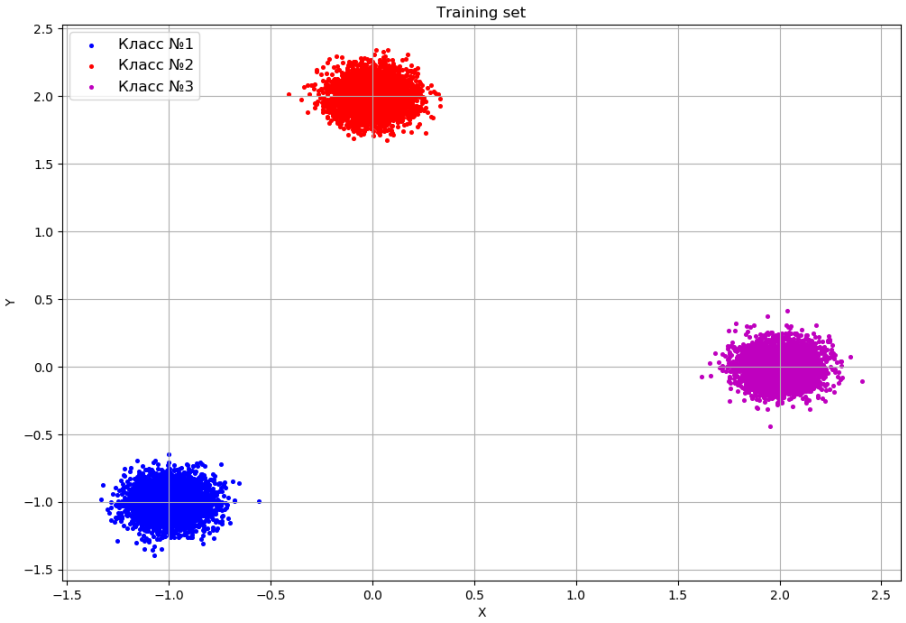


Рис. 1. Обучающие данные

Для классификации объектов было обучено три классификатора SVM с радиально-базисными функциями ядра:

- SVM1 с параметром $gamma = 0{,}0001$ (рис. 2);
- SVM2 с параметром $gamma = 900$ (рис. 3).

Параметр $gamma$ в методе SVM определяет, какое влияние при построении «идеальной» разделяющей линии имеют далеко находящиеся элементы обучающего набора данных. Чем ниже $gamma$, тем больше элементов, которые достаточно далеки от разделяющей линии, принимают участие в процессе определения линии. При высоком значении gamma алгоритм будет опираться только на наиболее близкие к линии элементы. Стоит отметить, что высокие значения gamma могут приводить к переобучению.

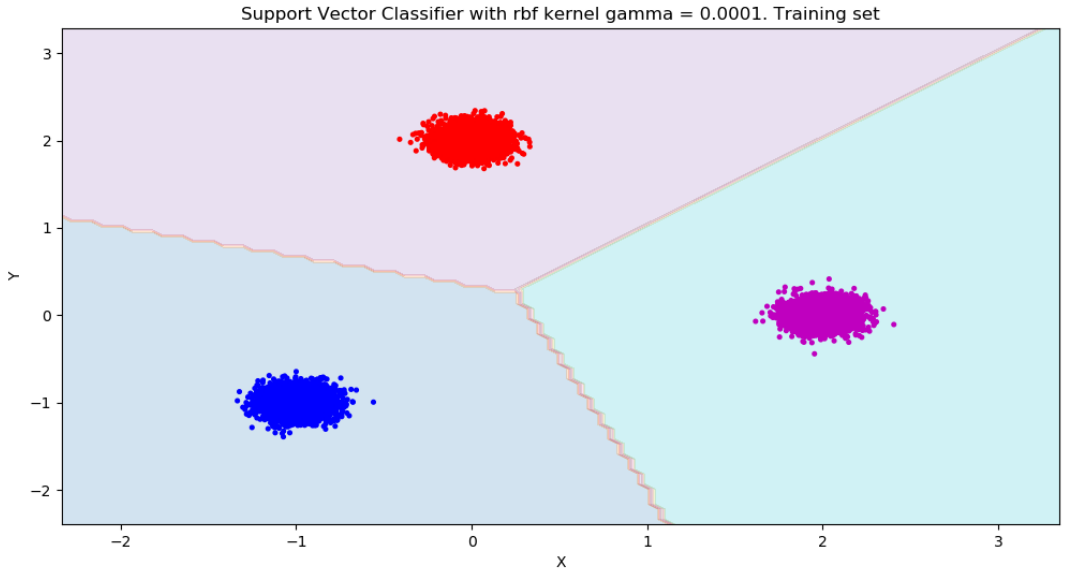


Рис. 2. Разделение пространства признаков обученным классификатором SVM1

Как видно по рис. 2, при низком значении параметра $gamma$ всё пространство признаков разделяется на три открытые области, каждая из которых отдельный класс. Важно отметить, что при использовании такого классификатора объекты, не относящиеся ни к одному из классов, будут ошибочно классифицированы.

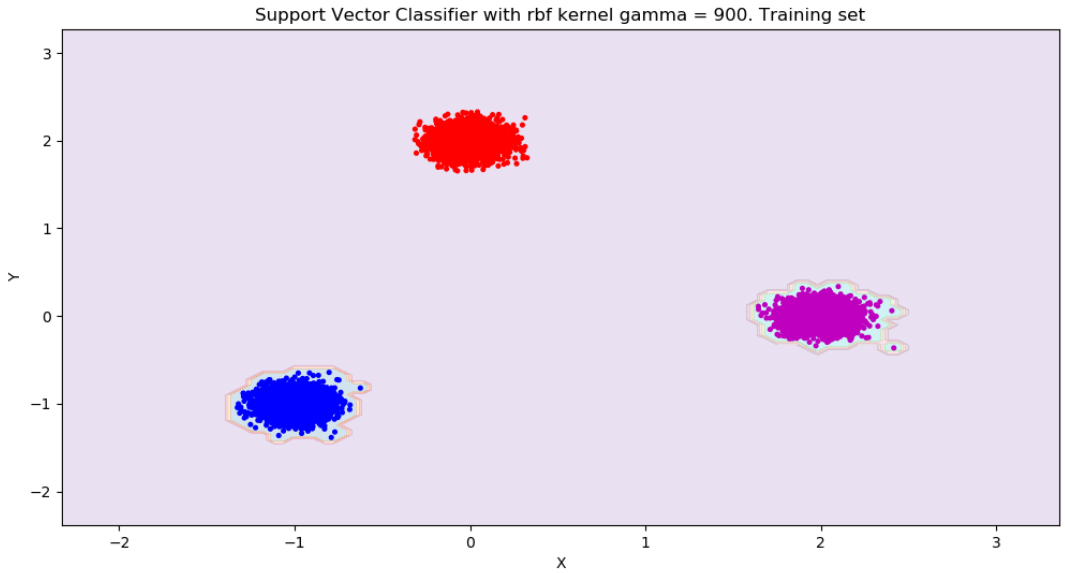


Рис. 3. Разделение пространства признаков обученным классификатором SVM2

При высоком значении gamma для данных двух классов строятся две замкнутые области, при этом точки остального пространства будут отнесены к третьему классу, что также в большинстве случаев будет некорректным решением. При этом очевидно, что произошло

переобучение для двух классов, поэтому точки за пределами выстроенных границ будут классифицированы неверно.

Результаты оценки качества трех классификаторов на тестовом наборе были получены согласно традиционной методологии и сведены в табл. 2.

Таблица 2. Результаты традиционного метода оценки качества

Классификатор		*Традиционные показатели оценки качества*					
SVM	***Gamma***	***F-score***	***Precision***	***Recall***	***Матрица ошибок***		
SVM1	0,0001	1,00	1,00	1,00	1000	0	0
					0	1000	0
					0	0	1000
SVM2	900	0,99	0,99	0,99	999	0	1
					0	998	2
					0	0	1000

Согласно традиционным показателям оценки качества все классификаторы имеют одинаково высокую точность и могут использоваться для решения практической задачи.

Теперь оценим качество трех классификаторов по новой методологии: оценка качества каждого классификатора определяется качеством классификации объектов каждого класса. Рассчитанные значения новых показателей качества представлены в табл. 3.

Таблица 3. Результаты нового метода оценки качества

Классификатор		*Новые показатели оценки качества*				
SVM	***Gamma***	***Номер класса***	***Excess***	***Deficit***	***Coating***	***Approx***
SVM1	0,0001	1	8,50	0	1	**0,10**
		2	13,40	0	1	**0,07**
		3	11,60	0	1	**0,07**
SVM2	900	1	0	0	1	**1**
		2	0,12	0	1	**0,89**
		3	30,90	0	1	**0,03**

Несмотря на то, что традиционные показатели качества показывают, что оба классификатора обеспечивают высокую точность, по значениям показателей $Excess$ и $Approx$ видно, что точность классификаторов разная, как и точность классификации каждого класса. Качество первого классификатора неудовлетворительно, поскольку значения $Excess$ для всех классов велики - классификатор неправильно

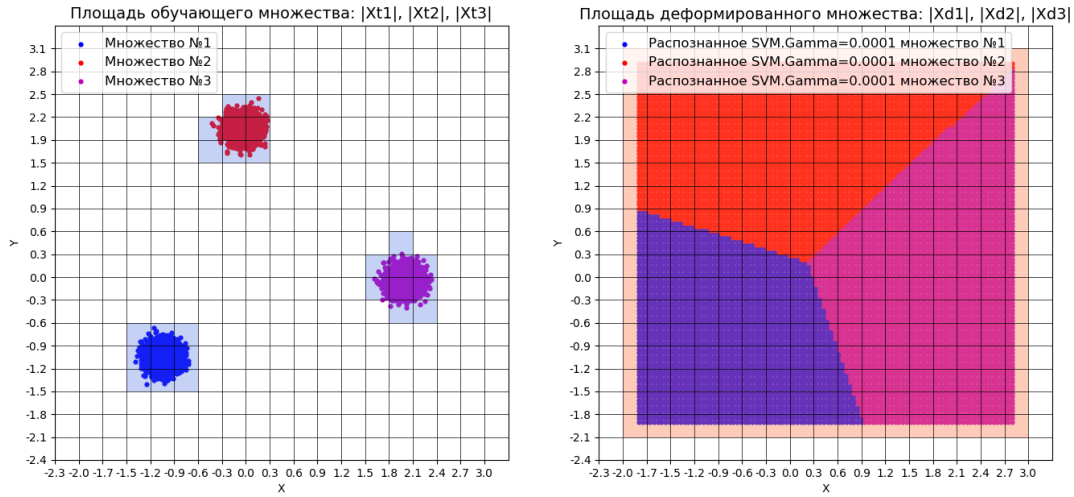
классифицирует объекты за пределами целевого класса (ошибочная классификация точек, не относящихся к целевым классам), а значения показателя $Approx$, наоборот, близки к 0 для всех классов, что означает низкий уровень аппроксимации классификатором целевого множества. Кроме того, высокие значения показателя $Approx$ для 1 и 2 класса подтверждают, что увеличение параметра обучения gamma приводит к высокой степени аппроксимации классификатором области обучающего множества. То есть, только по значениям показателя $Approx$ можно сделать вывод о том, насколько границы класса, выстроенные классификатором, близки к фактическим границам области обучающего множества. Эта информация особенно ценна в многомерном пространстве признаков, когда привычная визуализация обучающего множества и границ затруднительна, а решение о достижении требуемого уровня качества принимается только на основе числовых показателей качества.

Таким образом, выводы, сделанные благодаря анализу значений введённых показателей качества, полностью соответствуют действительному качеству классификаторов, которое можно визуально оценить по рис. 2 и 3. Поскольку традиционные показатели качества не могут предоставить такой точной оценки качества классификаторов, то можно утверждать, что введенные показатели качества более информативны и объективны.

Далее рассмотрим этапы расчета показателей качества более подробно.

Этап 1. Сканирование пространства и определение размера $|X_T^*|$ *и* $|X_D^*|$

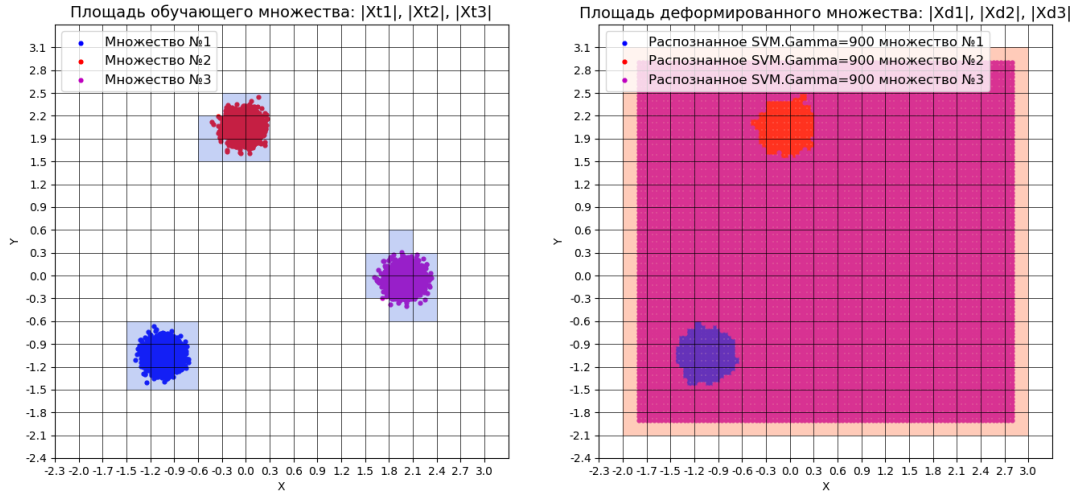
Поскольку размерность пространства признаков - 2, то для сканирования пространства была построена сетка с одинаковым шагом h по двум осям в расширенной области значений признаков (рис. 4 и 5).



а) б)

Рис. 4. Дискретные области обучающего X_T^* и деформированного множества X_D^* (SVM1)

Как видно на рис. 4а обучающие примеры первого класса, например, занимают 9 ячеек сетки, соответственно, объем области $|X_T^*|$ рассчитывается как площадь совокупности этих ячеек: $9h^2$. Примеров пространства сканирования, отнесенных классификатором SVM1 к первому классу, гораздо больше – они занимают 86 ячеек сетки (рис. 4б), и размер области $|X_D^*|$ составляет $86h^2$.



а) б)

Рис. 5. Дискретные области обучающего X_T^* и деформированного множества X_D^* (SVM2)

Для сравнения, площадь классифицированного SVM2 множества $|X_D^*|$ для первого класса полностью совпадает с площадью обучающего $|X_T^*|$, а вот для третьего класса при $|X_T^*| = 9h^2$, $|X_D^*| = 272h^2$ (рис. 5). Такая разница в размерах области целевого класса и области данных,

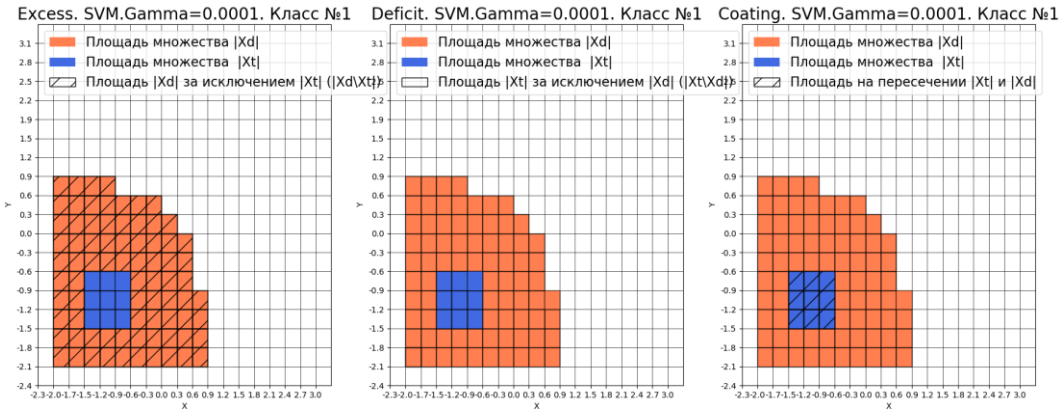
относимых классификатором к целевому классу, говорит об избыточном обобщении и риске ошибок классификатора.

Этап 2. Расчет и сравнение значений показателей качества

Для расчета величины показателя $Excess$, согласно формуле (1), необходимо определить область, включающую X_D^* за исключением области X_T^*, а затем найти её отношение к области обучающего множества X_T^*. На рис. 6-11 показаны области X_T^* и X_D^*, а искомая область заштрихована. На рис. 6а область $X_D^* \backslash X_T^*$ заштрихована.

Для расчета показателя $Deficit$ необходим размер области, включающей X_T^* и не включающей X_D^*. Поскольку область, построенная классификатором, полностью охватывает область обучающего множества, то такой области нет (рис. 6б), а показатель, согласно формуле (2), принимает значение 0.

Поскольку области X_T^* и X_D^* пересекаются только в области X_T^*, то, согласно формуле (3), величина $Coating$ достигает максимального значения – 1.



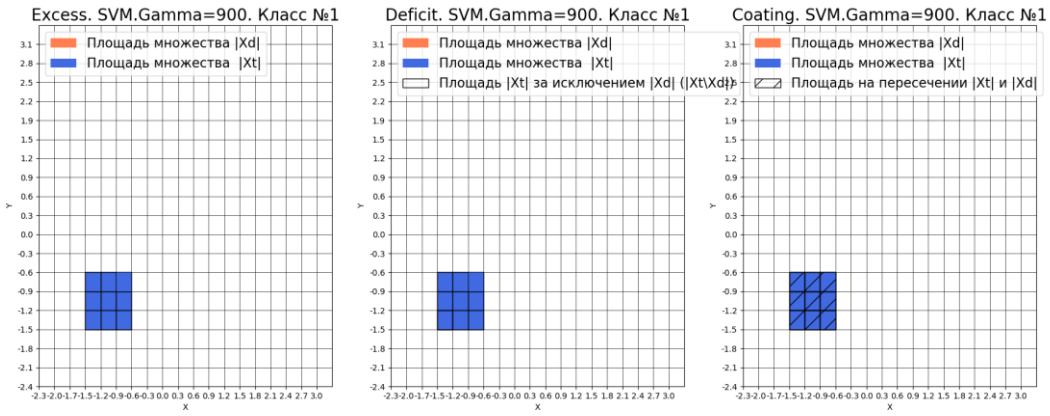
а) б) в)

Рис. 6. Визуализация дискретных областей, необходимых для расчета показателей Excess, Deficit и Coating (SVM1. Класс 1)

По рис. 6а можно оценить отношение размера области $|X_D^* \backslash X_T^*|$ к размеру области $|X_T^*|$ величину показателя $Excess$. Величина показывает насколько область, построенная классификатором, больше, чем область целевого класса и насколько велик риск ошибок первого рода (α). Чем больше значение показателя $Excess$, тем больше данных, для которых велик риск ложного срабатывания классификатора.

Как видно из рисунка 6б, величина $Deficit = 0$, что означает нулевую вероятность ошибки второго рода (β) внутри обучающего множества.

На рис. 7 показаны аналогичные данные для расчета показателей качества классификации объектов первого класса классификатором SVM2.

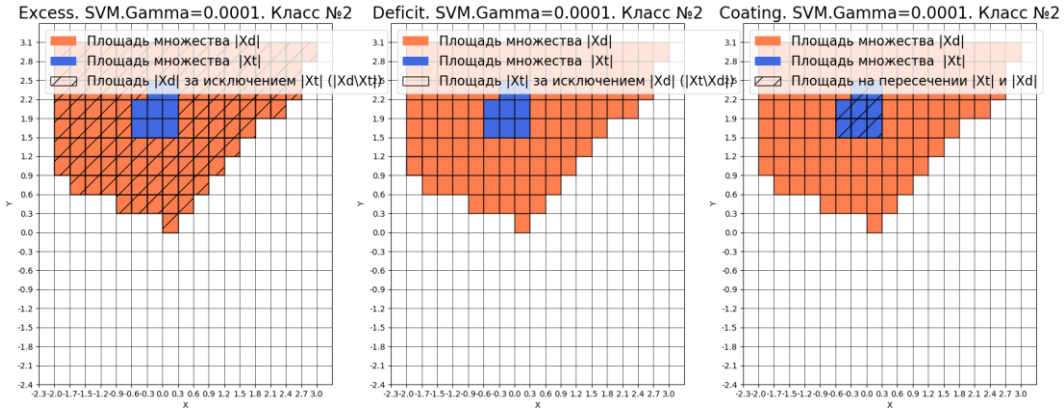


а) б) в)

Рис. 7. Визуализация дискретных областей, необходимых для расчета показателей Excess, Deficit и Coating (SVM2. Класс 1)

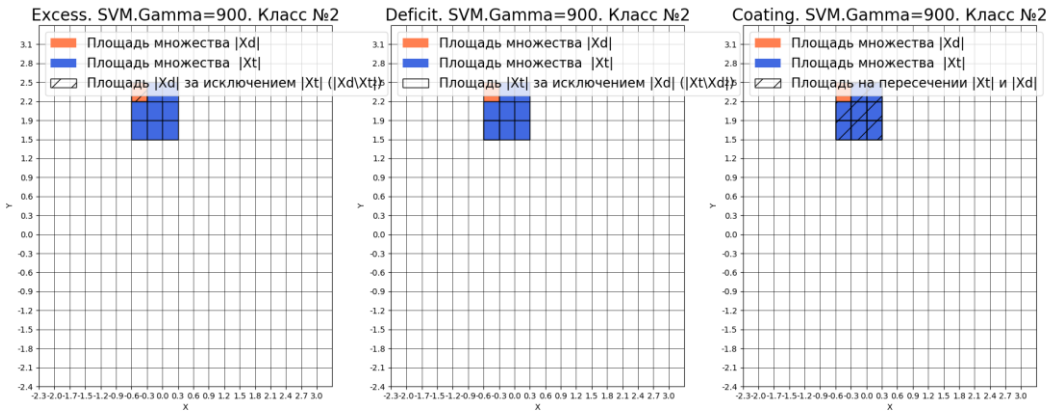
Как можно заметить, классификатор построил границы класса в полном соответствии с целевым множеством, а все величины принимают идеальные значения (см. (5)).

Соответствующим образом по рис. 8-11 можно сравнить показатели качества классификации оставшихся двух классов, обеспечиваемые классификаторами SVM1 и SVM2.



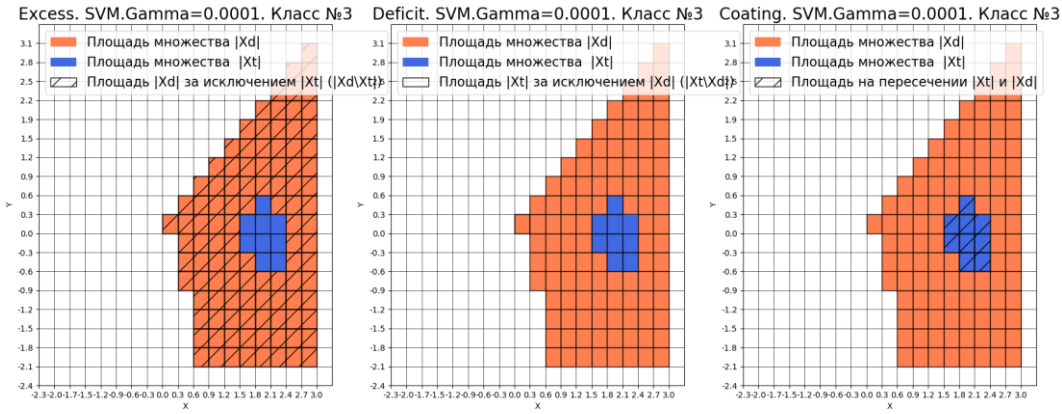
а) б) в)

Рис. 8. Визуализация дискретных областей, необходимых для расчета показателей Excess, Deficit и Coating (SVM1. Класс 2)



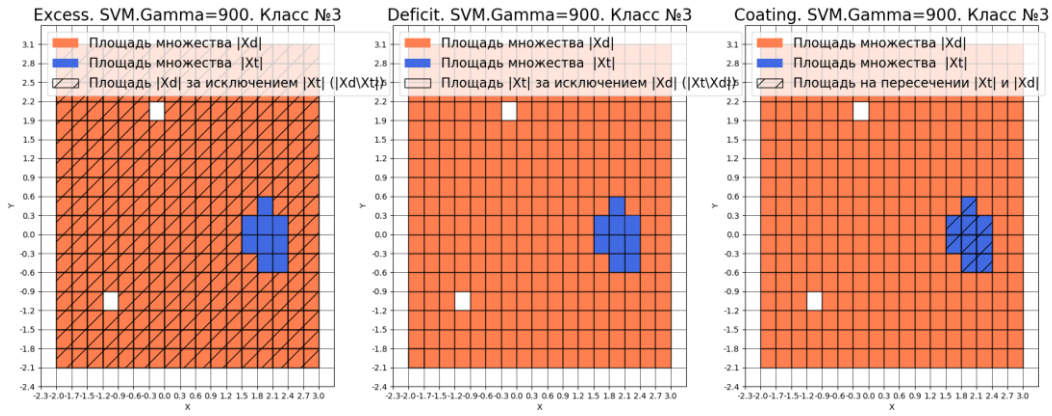
а) б) в)

Рис. 9. Визуализация дискретных областей, необходимых для расчета показателей Excess, Deficit и Coating (SVM2. Класс 2)



а) б) в)

Рис. 10. Визуализация дискретных областей, необходимых для расчета показателей Excess, Deficit и Coating (SVM1. Класс 3)



а) б) в)

Рис. 11. Визуализация дискретных областей, необходимых для расчета показателей Excess, Deficit и Coating (SVM2. Класс 3)

Таким образом, с помощью введённых показателей можно рассчитать объем дискретных областей и оценить насколько деформированная область каждого класса, построенная классификатором после обучения, не соответствует идеалу – области, занимаемой точками представительной обучающей выборки. Ценность такой оценки особенно возрастает в пространствах высокой размерности.

Анализ полученных результатов эксперимента подтверждает полезность введенных показателей для оценки качества многоклассовых классификаторов, а также преимущество по сравнению с традиционным подходом к оценке качества классификации.

Обсуждение

Подход к оценке качества классификации, предлагаемый в данной статье, представляется более объективным и информативным по сравнению с общеизвестными критериями качества многоклассовой классификации, основной недостаток которых в неполноте оценки – они позволяют оценить качество только внутри обучающего и тестового множества и не позволяют оценить риски ошибок классификаторов на новых данных.

Преимущество и уникальность предлагаемого подхода в том, что он оперирует не количеством примеров в обучающей и тестовой выборке, а областями, покрываемыми точками обучающего и деформированного множества. Введенные показатели позволяют оценивать качество классификации объектов каждого класса даже в многомерных пространствах, когда невозможно визуализировать области обучающих и классифицируемых точек.

Можно предположить, что новый критерий качества классификации в виде совокупности показателей $Approx$, $Excess$, $Deficit$, $Coating$ заслуживает большего доверия на этапе тестирования и

позволит соответственно повысить доверие к результатам многоклассовой классификации. Данные показатели можно использовать для сравнения, целенаправленной оптимизации многоклассовых классификаторов и для управления границами, формируемыми классификаторами на этапе обучения.

Стоит отметить, что метод SVM, как и многие многоклассовые классификаторы, после обучения делит пространство на открытые области классов, что делает возможным формирование состязательных примеров и некорректный результат классификации примеров за пределами целевых классов. Введенные показатели качества позволяют оценивать величину и положение областей, формируемых классификатором после обучения, относительно целевых, а затем оценивать и минимизировать риск неправильной классификации за пределами обучающей выборки.

Ранее был предложен способ многоклассовой классификации на основе автокодировщиков [16], позволяющий получать замкнутую область деформированного множества для каждого класса, а также управлять его величиной для точной классификации и выявления аномалий. Такой способ совместно с новым критерием качества может позволить строить классификаторы, близкие к идеальным в многомерном пространстве.

В то же время, нельзя не отметить объективные недостатки подхода:

- необходимость расчета показателей для каждого класса;
- экспоненциальный рост вычислительной сложности и потребляемой памяти с ростом размерности пространства признаков X;
- высокая вычислительная сложность сканирования пространства для определения деформированного множества X_D^*, растущая экспоненциально с уменьшением шага сетки разбиения;
- экспоненциально растущий с уменьшением шага сетки объём памяти для выполнения операций над дискретизированными множествами;
- зависимость величин рассчитываемых показателей от шага сетки.

При этом вычислительная сложность определения X_T^* линейно зависит от размерности обучающего множества $\hat{X}_T$.

Заключение

Была рассмотрена проблема классификации за пределами обучающей выборки и недостатки традиционных показателей оценки качества многоклассовой классификации. Сформулирован новый критерий качества многоклассовых классификаторов, позволяющий оценивать качество классификации объектов каждого класса в

отдельности на основе теоретико-множественных операций в предположении о непрерывности целевых и классифицированных множеств. Качество классификации объектов каждого класса характеризуется показателями $Approx$, $Excess$, $Deficit$, $Coating$, основанными на оценке объемов множеств. Подтверждена хорошая интерпретируемость значений показателей и продемонстрированы преимущества и большая точность введенных показателей относительно общепринятых критериев качества.

Примечание

Работа проведена при финансовой поддержке РФФИ, проект № 20-37-90073.

Библиографический список

1. Шпрингер Е. 17 примеров применения машинного обучения в 5 отраслях бизнеса. Журнал Mail.ru Cloud Solutions. [Электрон. ресурс] https://mcs.mail.ru/blog/17-primerov-mashinnogo-obucheniya. 2020. (дата обращения 03.06.2021).

2. From Roadblock to Scale: The Global Sprint Towards AI. New research commissioned by IBM in partnership with Morning Consult. [Электрон. ресурс] http://filecache.mediaroom.com/mr5mr_ibmnews/183710/Roadblock-to-Scale-exec-summary.pdf. 2020. (дата обращения 03.06.2021).

3. Sarah Pike. Почему одного только машинного обучения недостаточно. [Электрон. ресурс] https://www.kaspersky.ru/blog/ai-fails/18678/. 2017. (дата обращения 03.06.2021).

4. J. Goodfellow I., Shlens J., Sze Ch. Explaining and harnessing adversarial examples. Google Inc., Mountain View, CA : s.n. ICLR 2015. pp. 1-11.

5. Hern A. Want to beat facial recognition? Get some funky tortoiseshell glasses. [Электрон. ресурс] https://www.theguardian.com/technology/2016/nov/03/how-funky-tortoiseshell-glasses-can-beat-facial-recognition. 2016. (дата обращения 03.06.2021).

6. Forman G. An extensive empirical study of feature selection metrics for text classification // Journal of Machine Learning Research (JMLR). 2003. pp. 1-27.

7. Powers D. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. s.l. : Technical Report SIE-07-001, December 2007. pp. 1-24.

8. Fawcett T. An Introduction to ROC Analysis // Pattern Recognition Letters. 2006. Vol. 27, pp. 861–874.

9. Szegedy Ch., Zaremba W., Sutskever I. et al. Intriguing properties of neural networks. Computer Vision and Pattern Recognition. ICLR. 2014. http://arxiv.org/abs/1312.6199.abs/1312.6199.

10. Harris M. Researchers Find a Malicious Way to Meddle with Autonomous Cars. [Электрон. ресурс] https://www.caranddriver.com/news/a15340148/researchers-find-a-malicious-way-to-meddle-with-autonomous-cars/. 2017. (дата обращения 03.06.2021).

11. Robin J., Liang P. Adversarial Examples for Evaluating Reading Comprehension Systems. Computation and Language. 2017. https://arxiv.org/pdf/1707.07328.pdf.

12. Kurakin A., Goodfellow I., Bengio S. Adversarial examples in the physical world. ICLR 2017. https://arxiv.org/abs/1607.02533.

13. Mahmood S., et al. Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition // The 2016 ACM SIGSAC Conference. 2016. pp. 1528-1540. 10.1145/2976749.2978392.

14. Shafahi A., et al. Are adversarial examples inevitable? 2020. https://arxiv.org/pdf/1809.02104.pdf.

15. Гурина А.О., Елисеев В.Л. Эмпирический критерий качества одноклассовой классификации // 27-я Международная научно-технической конференции Информационные системы и технологии (ИСТ-2021). Нижний Новгород: Нижегородский государственный технический университет им. Р.Е. Алексеева, 2021.

16. Гурина А.О., Елисеев В.Л. Нейросетевой метод классификации в условиях нестационарного множества классов // XXVI международная научно-техническая конференция «Информационные системы и технологии» (ИСТ-2020). Н. Новгород : НГТУ им. Р.Е. Алексеева, 2020.