DOI 10.37882/2223-2966.2025.07.37

АНАЛИЗ ЭФФЕКТИВНОСТИ МАШИН ОПОРНЫХ ВЕКТОРОВ В КЛАССИФИКАЦИИ ДАННЫХ И ИХ ПРАКТИЧЕСКИЕ ПРИМЕНЕНИЯ

ANALYSIS OF THE EFFECTIVENESS OF SUPPORT VECTOR MACHINES IN DATA CLASSIFICATION AND THEIR PRACTICAL

Hejjo Muhsen

Summary. The aim of the study is to improve the efficiency of unmanned aerial vehicles (One of the classification methods presented in this work is the support vector machine classifier. The support vector machine (SVM) is a supervised learning method that ensures a correspondence between desired input and output data. This method is based on statistical learning theory and is applied to classification tasks such as disease diagnosis, image classification, and handwritten text recognition. Traditional artificial neural networks face challenges in generalization because they rely on the principle of empirical risk minimization (ERM). Consequently, in 1995, Vapnik developed the support vector machine method to enhance the generalization process based on the principle of structural risk minimization (SRM). This principle surpasses empirical risk minimization, as it focuses on reducing the upper bound of expected risk rather than merely minimizing errors in the training set. This distinction provides the support vector machine with greater generalization capabilities, which is a key goal of statistical learning.

Keywords: Support vectors, support vector machine, Lagrange multiplier, kernel trick, maximum margin.

Хежжо Мухсен
Аспирант, Казанский федеральный университет
muhseen.hejoo@gmail.com

Аннотация. Одним из методов классификации, рассматриваемых в данной работе, является классификатор на основе машин опорных векторов. Машина опорных векторов (SVM) представляет собой метод обучения с учителем, который обеспечивает соответствие между желаемыми входными и выходными данными. Этот метод основан на статистической теории обучения и находит применение в задачах классификации, таких как диагностика заболеваний, классификация изображений и распознавание рукописного текста. Традиционные искусственные нейронные сети сталкиваются с проблемами обобщения, поскольку они основываются на принципе минимизации эмпирического риска (ERM). В связи с этим, в 1995 году Вапник разработал метод машин опорных векторов, направленный на улучшение процесса обобщения на основе принципа минимизации структурного риска (SRM]. Данный принцип превосходит минимизацию эмпирического риска, так как он фокусируется на снижении верхнего предела ожидаемого риска, а не только на уменьшении ошибок в обучающем наборе. Это различие обеспечивает векторной машине повышенные способности к обобщению, что является ключевой целью статистического обучения.

Ключевые слова: опорные векторы, машина опорных векторов, множитель Лагранжа, ядерный трюк, максимальная маржа.

Введение

Современные достижения в области машинного обучения открывают новые возможности для решения сложных задач классификации данных, особенно в контексте автономных систем и медицинской диагностики. Одним из ключевых методов, демонстрирующих высокую эффективность в этих областях, является метод опорных векторов (SVM). Его уникальность заключается в способности находить оптимальные границы решений, минимизируя структурный риск, что обеспечивает устойчивость к переобучению и высокую обобщающую способность. В отличие от традиционных искусственных нейронных сетей, основанных на минимизации эмпирического риска, SVM опирается на принцип структурной минимизации риска, предложенный Вапником, что делает его особенно ценным для работы с ограниченными наборами данных и задачами высокой размерности. Целью данной работы является комплексный анализ метода SVM, включая его теоретические основы, преимущества перед другими подходами, а также практическое применение в реальных сценариях. В статье рассматриваются ключевые аспекты SVM: от математического обоснования максимального зазора и использования множителей Лагранжа до применения ядерных функций для нелинейно разделимых данных. Особое внимание уделено сравнительному анализу SVM с классическими нейронными сетями, что позволяет выделить его сильные стороны в контексте задач классификации. Практическая значимость исследования подтверждается примерами использования SVM в диагностике рака молочной железы и управлении группами беспилотных летательных аппаратов (БЛА). Эти приложения демонстрируют, как теоретические принципы метода трансформируются в эффективные инструменты для решения реальных проблем. Статья структурирована следующим образом: в первом разделе представлены теоретические основы SVM, включая линейные и нелинейные модификации. Далее рассматриваются методы оптимизации и применения ядерных функций. Заключительные разделы посвящены практическим кейсам и выводам, подчеркивающим роль SVM в современных технических и медицинских исследованиях. Работа вносит вклад в развитие методов машинного обучения, предлагая как теоретический анализ, так и практические рекомендации для инженеров и исследователей, работающих в области анализа данных и искусственного интеллекта.

Векторы поддержки и максимальная маржа

Этот метод использует подмножество обучающих примеров, известных как опорные векторы, для представления границы решения путем поиска наибольшего запаса, на который данные расходятся с обоих концов. Это гиперплоскость с максимальными запасами. Цель выбора пределов решения с максимальными запасами заключается в том, что максимальные запасы делают ошибки при обобщении пределов принятия решений минимальными [2].

Машина опорных векторов

Ниже мы показываем машину опорных векторов в случае линейно разделимых и неразделимых данных, а также машину нелинейных опорных векторов [7].

Машина опорных векторов для линейно разделимых данных (SVM)

Если $x_i = (x_1, \ldots\ldots\ldots, x_d)$ представляет собой набор атрибутов линейно разделимых данных двоичной классификации, (представленных кружками и треугольниками), которые предназначены для классификации в набор классов $y_i = \{-1, 1\}$. Взяв обучающий набор N, как показано на рисунке (1).

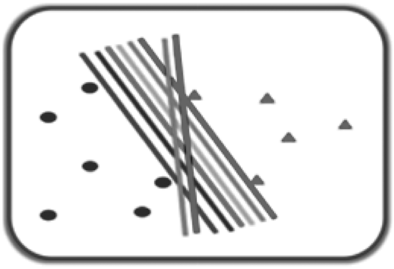

Рис. 1. Уровни разделения данных

Мы замечаем, что существует более одной плоскости для разделения данных, но плоскость, которую ищет машина опорных векторов, это плоскость, к которой приближаются или лежат опорные векторы (обозначенные пунктирными кружками обоих типов), как показано на рисунке (2).

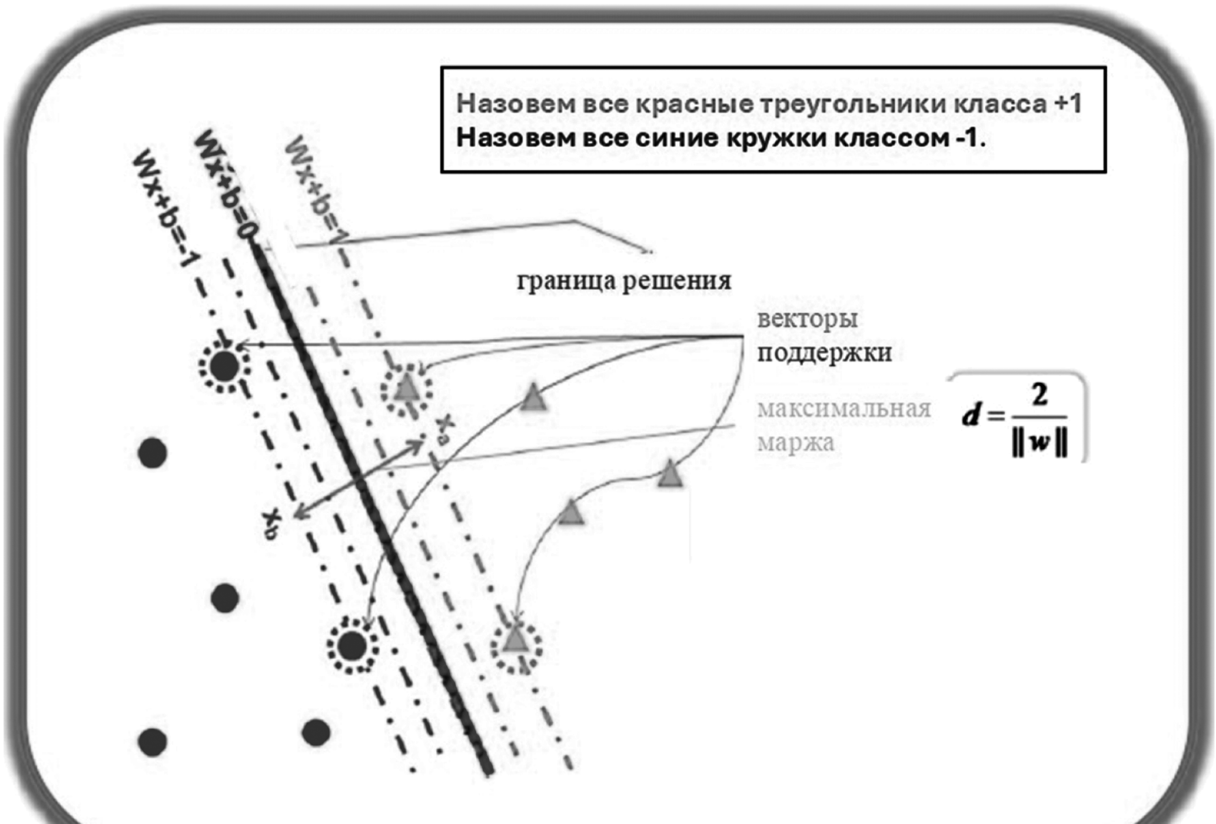

Рис. 2. Граница принятия решения о разделении данных, определяемая опорными векторами для линейно разделимых данных

Если x входной вектор, а w, b посредник модели, то предел решения для линейного классификатора определяется уравнением [8]:

$$wx + b = 0 \tag{1}$$

Точки решения средней прямой удовлетворяют ее уравнению. Что касается точек, лежащих выше предела решения (треугольники), то они достигают $wx + b > 0$ а точки ниже предела решения (кружки) достигают $wx + b < 0$ Если отметить категории Y соответственно любая запись данных классифицируется следующим образом:

$$Y = \begin{cases} 1 \text{ если } wx + b \geq 1 \\ 0 \\ -1 \text{ если } wx + b \leq -1 \end{cases} \tag{2}$$

Что записано в виде:

$$Y_i\left(wx_i + b\right) \geq 1 \tag{3}$$

Рассчитать маржу линейного классификатора

Треугольник, ближайший к пределу решения, достигает $wx_1 + b = 1$ и круг, ближайший к пределу решения, достигает $wx_2 + b = -1$ и расстояние между ними d представляет собой линейную границу книги. Вычитая два уравнения, находим:

$$w(x_1 - x_2) = 2 \quad \Rightarrow \quad wd = 2$$

Если ширина поля d:

$$d = \frac{2}{w} \tag{4}$$

Изучение линейной машины опорных векторов

Линейная машина опорных векторов поставляется с обучающим набором [6]. Чтобы оценить медиаторы предела решения линейного классификатора w, b из обучающего набора. Две медианы оцениваются w, b так, чтобы были выполнены два условия:

$$wx_i + b \geq -1 \text{ если } y_i = -1 \tag{5}$$

$$wx_i + b \leq -1 \text{ если } y_i = -1 \tag{6}$$

Процесс машинного обучения опорных векторов направлен на достижение наибольшего запаса $d = \dfrac{2}{\|w\|}$ от которого данные отходят с обоих концов. Это приводит к задаче нелинейного программирования с ограничениями, представленной целевой функцией

$$f\left(w\right) = \min_{w} \frac{\|w\|^2}{2} \tag{7}$$

С письменными ограничениями

$$y_i\left(wx_i + b\right) \geq 1 \ , i = 1, 2, \ldots\ldots\ldots, N \tag{8}$$

Это задача выпуклой оптимизации, поскольку целевая функция представляет собой уравнение второй степени, а ограничения линейны по параметрам w, b и она решается с использованием метода множителей Лагранжа. Это метод нахождения значений локальных пределов функции нескольких переменных, подчиняющихся одному или нескольким ограничениям. Этот метод сводит задачу с n переменными с k ограничениями к разрешимой задаче с $n+k$ переменными без каких-либо ограничений. Этот метод использует неизвестную постоянную переменную (факториал Лагранжа) для каждого ограничения и формирует линейные структуры, которые включают факториалы Лагранжа в качестве коэффициентов. Сформируем функцию Лагранжа для задачи оптимизации:

$$L_p = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i \left(y_i \left(wx_i + b \right) - 1 \right) \tag{9}$$

Где λ_i факториалы Лагранжа. Чтобы сделать функцию Лагранжа минимальной, придаем частным производным L по w, b нулевое значение, находим:

$$\frac{\partial L_p}{\partial w} = 0 \;\Rightarrow\; \boldsymbol{w} = \sum_{i=1}^{N} \lambda_i y_i \chi_i \tag{10}$$

$$\frac{\partial L_p}{\partial b} = 0 \;\Rightarrow\; \sum_{i=1}^{N} \lambda_i y_i = 0 \tag{11}$$

Факториалы Лагранжа λ_i неизвестны, поэтому решение невозможно для w, b Поэтому мы преобразуем ограничения-неравенства в набор ограничений-равенств и получаем факториальные ограничения Лагранжа, известные как Каруша-Куна. — Условия Такера (ККТ).

$$\lambda_i \geq 0 \tag{12}$$

$$\lambda_i \left[y_i \left(wx_i + b \right) - 1 \right] = 0 \tag{13}$$

Число медиаторов по-прежнему велико, и для простоты преобразуем функцию Лагранжа в функцию для факториалов Лагранжа, только взяв задачу, сопровождающую задачу оптимизации:

$$L_D = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \lambda_i \lambda_j y_i y_j x_i x_j \tag{14}$$

Решая сопутствующую задачу (с использованием численных методов, таких как квадратичное программирование), получаем значения λ_i Расчет w, b Используем два уравнения ...:

Это будет предел решения [3]

$$\left(\sum_{i=1}^{N}\lambda_i y_i \chi_i\right)x + b = 0 \qquad (15)$$

Машина опорных векторов для линейно неразделимых данных (Нелинейно разделимая SVM)

В случае неразделимых линейных данных, как показано на рисунке (3)

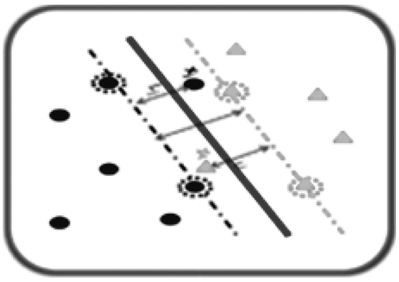

Рис. 3. Граница принятия решения о разделении данных, определяемая опорными векторами для данных, которые не являются линейно разделимыми

В этом случае машина опорных векторов использует мягкий запас, который позволяет пределу решения допускать небольшие ошибки обучения, поскольку данные не являются линейно разделимыми. Когда машина опорных векторов находит компромисс между шириной поля и количеством ошибок обучения, допущенных пределом решения, путем уменьшения линейных ограничений для соответствия данным путем введения формальных переменных L_i, пределы решения принимают форму:

$$Y = \begin{cases} 1 \text{ если } wx + b \geq 1 - L_i \\ 0 \\ -1 \text{ если } wx + b \leq -1 + L_i ; L_i > 0 \end{cases} \qquad (16)$$

Чтобы алгоритм обучения не увеличивал предел решения и некорректно классифицировал множество обучающих примеров, предел решения модифицируется путем наложения штрафа на целевую функцию путем присвоения больших значений ее формальным переменным, так что

$$f(w) = \frac{\|w\|^2}{2} + C(\sum_{i=1}^{N} L_i)^k \qquad (17)$$

Где C, k представляет собой параметр штрафа за классификацию и определяется пользователем. Для $k = 1$ это функция Лагранжа для задачи оптимизации с ограничениями.

$$L_p = \frac{\|w\|^2}{2} + C\sum_{i}^{N} L_i - \\ -\sum_{i}^{N}\lambda_i\left(y_i\left(wx_i + b\right) - 1 + L_i\right) - \sum_{i}^{N}\mu_i L_i \qquad (18)$$

Термин $\sum_{i}^{N}\mu_i L_i$ был добавлен для того, чтобы значения не были отрицательными L_i , $L_p = \frac{\|w\|^2}{2} + C\sum_{i}^{N} L_i$ представляет собой целевую функцию, которую необходимо минимизировать до наименьшее значение. В то время как $\sum_{i}^{N}\lambda_i\left(y_i\left(wx_i + b\right) - 1 + L_i\right)$ представляет собой арбитражные ограничения, которые преобразуются с помощью условий Каруша-Куна-Такера (ККТ) в ограничения равенства

$$\begin{cases} L_i \geq 0, \lambda_i \geq 0, \mu_i \geq 0 \\ \lambda_i\left(y_i\left(wx_i + b\right) - 1 + L_i\right) = 0 \\ \mu_i L_i = 0 \end{cases} \qquad (19)$$

Коэффициенты Лагранжа μ_i в последнем уравнении равны нулю для $L_i \geq 0$ что соответствует неправильно классифицированным примерам обучения.

Чтобы сделать функцию Лагранжа минимальной, приведем частные производные L по L_i, w, b нулевому значению. Находим:

$$\frac{\partial L}{\partial w_j} = w_j - \sum_{i=1}^{N}\lambda_i y_i \chi_{ij} = 0 \Rightarrow w_j = \sum_{i=1}^{N}\lambda_i y_i \chi_{ij} \qquad (20)$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N}\lambda_i y_i = 0 \Rightarrow \sum_{i=1}^{N}\lambda_i y_i = 0 \qquad (21)$$

$$\frac{\partial L}{\partial L_i} = C - \lambda_i - \mu_i = 0 \Rightarrow \lambda_i + \mu_i = C \qquad (22)$$

Подставив его в функцию Лагранжа, найдем функцию Лагранжа для задачи, сопровождающей задачу оптимизации:

$$L_D = \frac{1}{2}\sum_{i,j}^{N}\lambda_i\lambda_j y_i y_j x_i x_j + C\sum_{i} L_i - \\ -\sum_{i}\lambda_i\left\{y_i\left(\sum_{j}\lambda_j y_j x_i x_j + b\right) - 1 + L_i\right\} - \sum_{i}\left(C - \lambda_i\right)L_i \qquad (23)$$

Отметим, что функции Лагранжа совпадают в случае линейно разделимых и нелинейно разделимых данных. Ограничения на функцию Лагранжа в двух случаях различны. В случае разделимых данных оно было $\lambda_i = 0$, но в случае нелинейно разделимых данных λ_i оно должно достигать $0 \leq \lambda_i \leq$.

Вычисление медиаторов предела решения осуществляется путем решения задачи, сопровождающей задачу оптимизации по нахождению факториалов Лагранжа λ_i и подстановки их в (17) и условия ККТ[2].

Машина опорных векторов для SVM нелинейных данных

Если набор данных имеет нелинейную границу решения (рис. 4), он преобразуется так, что линейную границу решения можно использовать для разделения набора обучающих данных в пространстве, в которое он был преобразован...[4]

То есть признаки x в исходном пространстве заменяются преобразованными признаками $\Phi(x)$. Таким образом, задачей обучения в нелинейной машине опорных векторов является следующая задача оптимизации:

$$f(w) = \min_{w} \frac{\|w\|^2}{2}$$

Итак, это:

$$y_i\left(w.\Phi(x_i) + b\right) \geq 1 \quad i = 1,2,\ldots\ldots,N \tag{24}$$

Функция Лагранжа, связанная с задачей оптимизации:

$$L_D = \sum_{i=1}^{N}\lambda_i - \frac{1}{2}\sum_{i,j}^{N} L_i\lambda_i\lambda_j y_i y_j \Phi(x_i)\Phi(x_j) \tag{25}$$

Решая сопутствующую задачу (с использованием численных методов, таких как квадратичное программирование), мы получаем значения λ_i Для расчета w, b мы используем два уравнения [4]

$$\boldsymbol{w} = \sum_{i,j}^{N}\lambda_i y_i \Phi(x_i) \tag{26}$$

$$\lambda_i\left\{y_i\left(\sum_j \lambda_j \Phi(x_j).\Phi(x_j) + b\right) - 1\right\} = 0 \tag{27}$$

Хитрость ядра

Трюк с ядром — это метод вычисления сходства между векторами в пространстве, в которое они были преобразованы, с использованием исходных признаков, то есть нахождение произведения $\Phi(x_j).\Phi(x_j)$ чтобы узнать сходство между двумя векторами $x_j x_j$ Это делается с помощью функции сходства K в исходном пространстве, известной как функция ядра. Существует множество функций ядра, таких как функция Гаусса.

$$K(x,y) = e^{\frac{-\|x-y\|}{2\sigma^2}} \tag{28}$$

Функция логистического ядра

$$K(x,y) = tanh(kx.y - \delta) \tag{29}$$

Полиномиальная функция

$$K(x,y) = \Phi(x).\Phi(y) = (x.y + 1)^2 \tag{30}$$

Некоторые функции ядра показаны в Приложении 3.

Преимущества и недостатки машины опорных векторов Адв. И недостатки SVM

1. Легкость обучения (обучения).
2. Он характеризуется своей силой для небольшого и большого количества переменных.
3. Он использует передовые математические принципы, чтобы избежать переобучения и хранения большого количества примеров.

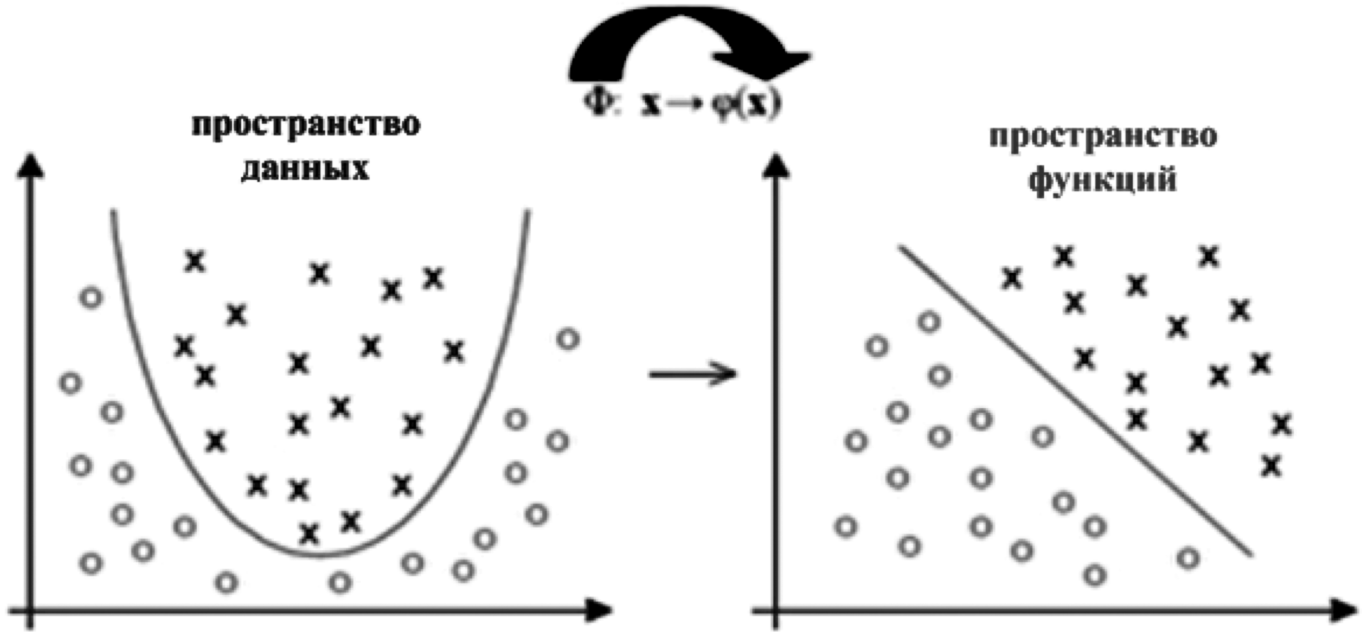

Рис. 4. Преобразование нелинейных данных в линейно разделимые данные

4. В отличие от сетей нет локальных минимумов.
5. Возможность преобразования данных в более высокие измерения.

Это недостаток машины опорных векторов.
1. Необходимо выбрать функцию Кемеля.
2. Это отнимает время.

Приложения для машин опорных векторов

Ниже мы представляем два применения машины опорных векторов. В первом показан механизм расчета параметров машины опорных векторов для установки границы решения двумерного набора данных, а во втором показано применение машины опорных векторов к образцу. пациентов с раком молочной железы, использующих пакет программного обеспечения Matlab.

Приложение (1): В этом примере мы показываем применение машины опорных векторов к двумерному набору данных.

Таблица 1.

Машины опорных векторов
к двумерному набору данных

Входные данные		Целевой результат
x_1	x_2	y
0.3858	0.4687	1
0.4871	0.611	–1
0.9218	0.4103	–1
0.7382	0.8936	–1
0.1763	0.579	1
0.4057	0.3529	1
0.9355	0.8132	–1
0.8146	0.0093	1

Сначала необходимо рассчитать коэффициенты Лагранжа λ_i соответствующие каждому обучающему примеру.

Это делается путем решения задачи оптимизации

$$L = \sum_{i=1}^{N}\lambda_i - \frac{1}{2}\sum_{i,j}^{N}\lambda_i\lambda_j y_i y_j x_i x_j$$

Используя квадратичное программирование, находим:

$$\lambda_1 = 65.5261,\ \lambda_2 = 65.5261,\ \lambda_3 = 0,\ \lambda_4 = 0,$$
$$\lambda_5 = 0,\ \lambda_6 = 0,\ \lambda_7 = 0,\ \lambda_8 = 0$$

$\lambda_1 \neq 0$ означает, что первый обучающий пример представляет собой опорный вектор для обучающего набора.

$\lambda_2 \neq 0$ означает, что второй обучающий пример представляет собой опорный вектор для обучающего набора.

Уравнение предела решения:

$$wx + b = w_1x_1 + w_2x_2 + b = 0.$$

Вычислим два посредника w, b

Матрица весов

$$w = \{w_1, w_2\} \Rightarrow w_i = \sum_{i=1}^{N}\lambda_i y_i x_i$$
$$w_1 = \lambda_1 y_1 x_{11} + \lambda_2 y_2 x_{21}$$
$$w_1 = 65.5261 * 1 * 0.3858 + 65.5261 * (-1) * 0.4871 = -6.64$$
$$w_2 = \lambda_1 y_1 x_{12} + \lambda_2 y_2 x_{22}$$
$$w_2 = 65.5261 * 1 * 0.4687 + 65.5261 * (-1) * 0.611 = -9.32$$

Предвзятость

$$b = \{b_1, b_2\} = \{1 - w.x, -1 - w.x\}$$
$$b_1 = 1 - w_1x_{11} - w_2x_{12} =$$
$$= 1 - (-6.64)(0.3858) - (-9.32)(0.4687) = 7.93$$
$$b_2 = -1 - w_1x_{21} - w_2x_{22} =$$
$$= -1 - (-6.64)(0.4871) - (-9.32)(0.611) = 7.9289$$

b представляет собой среднее из двух значений b_1, b_2 т.е.

$$b = \frac{b_1 + b_2}{2} = \frac{7.93 + 7.9289}{2} = 7.93$$

Уравнение становится пределом решения

$$-6.64x_1 - 9.32x_2 + 7.93$$

Приложение (2): В этом приложении мы показываем диагноз рака молочной железы для выборки из 80 записей из набора данных о пациентах с раком молочной железы с использованием машины опорных векторов (SVM).

При применении машины опорных векторов к набору данных о пациентах с раком молочной железы машина опорных векторов выполняет этап обучения для оценки параметров классификатора, затем обобщает свои знания для записей, на которых она не обучалась, и выполняет процесс классификации. Машина опорных векторов была обучена на 57 записях и протестирована на 23 записях. В таблице показано 2 количество обучающих и тестовых выборок для доброкачественных и злокачественных случаев при применении машины опорных векторов.

Таблица 2.

Количество обучающих и тестовых выборок для доброкачественных и злокачественных случаев

Данные/Классы	Обучающая выборка	Тестовая выборка	Итого
Доброкачественный	38	16	54
Злокачественный	19	7	26
Итого	57	23	80
% от набора данных	71 %	29 %	100 %

Функция логистического ядра также использовалась с посредниками.

$$\sigma = 0.3,\ C = 1$$

Его производительность была превосходной, поскольку он правильно классифицировал 100 % тестового набора. Где выходные и целевые векторы совпадают в тестовом примере, как показано в таблице (3).

На рисунке (5) показана гистограмма выходных данных SVM в тестовом примере.

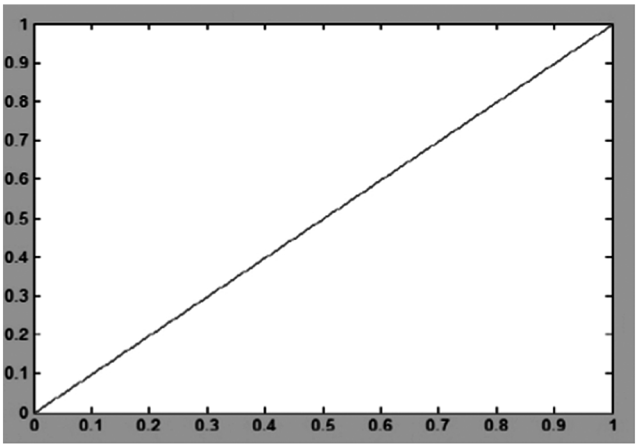

Рис. 5. Гистограмма вывода SVM в тестовом примере

Заключение

В данной статье представлены различные формы машины опорных векторов, что способствует улучшению процесса обобщения, основанного на принципе снижения структурного риска. Кроме того, данный метод был применен к выборке пациентов с раком молочной железы.

Таблица 3.

Целевые и выходные векторы после применения машины опорных векторов

Целевой выход	1	0	1	1	0	1	1	0	0	1	1	0	1	1	1		0	1	1	1	1	0	1
Фактический выход	1	0	1	1	0	1	1	0	0	1	1	0	1	1	1		0	1	1	1	1	0	1

ЛИТЕРАТУРА

1. Gunn, S., Support Vector Machines for Classification and Regression, Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science,1998.
2. Izenman, Alan J., Modern Multivariate Statistical Technique, Department of Statistics Temple University, 2008.
3. Hastie, Tibshirani, Friedman, The Elements of Statistical Learning, Data Mining, Inference, and Prediction, 2009.
4. Steinwart, Christmann, Support Vector Machines., Information Science and Statistics, 2008.
5. Анисимова И.В. Вычислительные и компьютерные технологии определения коэффициентов переноса в моделях многокомпонентных смесей / И.В. Анисимова, В.Н. Игнатьев; Федеральное Государственное Бюджетное Образовательное Учреждение Высшего Образования «Казанский Национальный Исследовательский Технический Университет им. А.Н. Туполева-Каи». Казанский государственный технический университет им. А.Н. Туполева, 2019. — 256 с. — ISBN 978-5-7579-2394-9. — EDN JQWPEO.
6. H.S.Hota, Diagnosis of Breast Cancer Using Intelligent Techniques, International Journal of Emerging Science and Engineering (IJESE), ISSN: 2319-6378, Volume-1, Issue-3, January 2013.
7. S. Aruna, Dr S.P. Rajagopalan, L.V. Nandakishore, An Empirical Comparison Of Supervisied Learning Algorithms In Disease Detection, 2011.
8. Madhavi Pradhan Ketki Kohale Parag Naikad Ajinkya Pachore Eknath Palwe, Design of Classifier for Detection of Diabetes using Neural Network and Fuzzy k- Nearest Neighbor Algorithm, 2012.

© Хежжо Мухсен (muhseen.hejoo@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»