УДК 004.8:517

Ж.В. Иванова, Ян Юйпин
Витебский государственный университет имени П.М. Машерова, Витебск, Беларусь

ГЛАДКАЯ ФУНКЦИЯ ПОТЕРЬ ДЛЯ УСКОРЕННОГО АДММ

Ключевые слова: метод опорных векторов (SVM), метод множителей с переменным направлением (ADMM), функция потерь.

В эпоху быстрого развития цифровой экономики и технологий искусственного интеллекта большие массивы данных с множеством признаков стали ключевой движущей силой в научных исследованиях и промышленности.

Все больше задач от распознавания образов и классификации текстов в машинном обучении, до скрининга заболеваний в медицинской диагностике и прогнозирования рисков в экономике требуют эффективных и надежных алгоритмов классификации. Одним из таких алгоритмов является метод опорных векторов (SVM), который был впервые предложен К. Кортес и В. Вапником [1]. Основная цель этого метода – разделить образцы двух классов, определяя оптимальную гиперплоскость $\mathbf{w}^T\mathbf{x} + b = 0$, что приводит к моделям SVM с жесткой и мягкой границей для линейно-разделимых и линейно-неразделимых случаев. В модели SVM с мягкой границей требуется найти:

$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 + \sigma \sum_{i=1}^{m} \ell\,(d),$$

где $\sigma > 0, d = 1 - y_i \cdot (\mathbf{w}^T\mathbf{x}_i + b)$; $\mathbf{w} \in R^n$ – вектор весов; $\mathbf{x}_i \in R^n$ – вектор признаков i-го объекта, $y_i = \{-1, +1\}$ – метка класса, которому принадлежит объект**;** $\ell\,(d)$ – функция потерь.

В данной модели функция потерь напрямую определяет ее разреженность и устойчивость. Стандартная функция 0/1-loss обеспечивает идеальный баланс: она присваивает постоянную потерю, равную 1, неправильно классифицированным или критическим образцам ($d \geq 0$) для достижения усечения и нулевую потерю – правильно классифицированным образцам ($d < 0$) для обеспечения разреженности. Однако ее невыпуклый и разрывный характер делает задачу оптимизации NP-трудной, требующей на практике приближенных функций потерь, которые можно разделить на два класса: выпуклые и невыпуклые.

Выпуклые потери (hinge loss, квадратичная функция потерь и т. д.) упрощают процесс оптимизации, но имеют недостаток неограниченности, что делает их чувствительными к выбросам в обучающих данных и недостаточно устойчивыми. Невыпуклые потери (Ramp Loss, усеченная функция потерь Хубера, и т. д.) улучшают способность противостоять выбросам за счет ограниченности и могут обеспечивать разреженность модели, что делает их предпочтительным выбором для крупномасштабных задач классификации. Среди них усеченная вогнутая функция потерь (ℓ_{tri}), предложенная в [2]: при $d \leq 0$ потери равны 0, для обеспечения разреженности; при $d \geq 1/4$ фиксированы на уровне 1, для повышения устойчивости. При $0 < d < 1/4$ функция задается переходным квадратичным членом $8d - 16d^2$. Соответствующие алгоритмы L_{tri}-SVM и L_{tri}-ADMM дополнительно обеспечивают эффективность вычислений на больших данных и надёжную глобальную сходимость. Однако ℓ_{tri} является кусочно-гладкой, что может привести к колебаниям сходимости в процессе численной оптимизации.

Мы предлагаем новую функцию потерь (ℓ_s)

$$\ell_s(d) = \begin{cases} 0, & d \leq 0, \\ \frac{1}{2}\sin\left(\omega d - \frac{\pi}{2}\right) + \frac{1}{2}, & 0 < d < \frac{\pi}{\omega}, \\ 1, & d \geq \frac{\pi}{\omega}, \end{cases}$$

где $\omega \geq \pi$ регулирует крутизну синусоидального перехода.

Эта функция обеспечивает разреженность модели и устойчивость к выбросам благодаря своей структуре: промежуточный переходный сегмент использует плавную синусоидальную кривую для достижения глобальной непрерывности в C^1, устраняет резкие изменения производных при $d = 0$ и $d = \pi/\omega$, а также предоставляет более стабильную информацию о градиенте для численной оптимизации.

Основываясь на $\ell_s(d)$, разработана модель ℓ_s-SVM, для которой определен проксимальный оператор целевой функции и P-стационарная точка, установлено эквивалентное соотношение между P-стационарной точкой и локальным минимумом модели, охарактеризованы свойства опорных векторов ℓ_s-SVM. Кроме того, разработан ускоренный алгоритм ADMM с интегрированной стратегией рабочего набора, даны формулы обновления всех переменных в замкнутой форме и доказана глобальная сходимость алгоритма.

Литература

1. Cortes C., Vapnik V. Support-vector networks //Machine Learning. 1995. №20(3). С. 273-297.

2. Wang H., Li W. A novel highly efficient alternating direction method of multipliers for large-scale trimmed concave SVM. //Applied Soft Computing. 2024. №167.

***Сведения об авторах*:**

Иванова Жанна Викторовна – доцент кафедры математики Витебского государственного университета им. П.М. Машерова, кандидат физико-математических наук, доцент, e-mail: IvanovaZhV@gmail.com, SPIN-код: 9115-9219.

Ян Юйпин – магистрант Витебского государственного университета им. П.М. Машерова, e-mail: yupingyang325@gmail.com.