Литература

1. Лэм Г. Аналоговые и цифровые фильтры. М: Мир, 1982. - 592 с.
2. Нургес Ю. Лагерровы модели в задачах аппроксимации и идентификации//АиТ. - 1987. - 3. - С. 88-96.
3. Прохоров С.А. Аппроксимативный анализ случайных процессов. - 2-е изд., перераб. и доп. - Самара: СНЦ РАН, 2001. - 380 с.
4. King R.E., Paraskevopoulos P.N. Digital Laguerre filters//Int. J. Circuit Theory and Appl. - 1977. - 5(1). - pp. 81-91.
5. Сергиенко А.Б. Цифровая обработка сигналов. - Спб.: Питер, 2002. - 608 с.
6. Clowes G. Choice of the time-scaling factor for linear system approximations using orthonormal Laguerre functions//IEEE Trans. Automatic Control. - 1965. - 10. - pp. 487-489.
7. Волков И.И., Прохоров С.А. Способ повышения точности аппроксимации корреляционных функций ортогональными функциями Лагерра//Приборостроение. - 1974. - 17. - С. 66-72.
8. den Brinker A.C. Meixner-like functions having a rational z-transform//International Journal of Circuit Theory and Applications. - 1995. - 23. - pp. 237-246.
9. den Brinker A.C., Belt H.J. Optimal free parameters in orthonormal approximations//IEEE Trans. Sig. Proc. - 1998. - 46. - pp. 2081-2087.
10. Prokhorov S.A., Kulikovskikh I.M. Unique condition for generalized Laguerre functions to solve pole position problem//Signal Processing. - 2015. - 108. - pp. 25-29.

И.М. Куликовских, К.А. Кузнецова, К.А. Панова, С.А. Прохоров

ПОВЫШЕНИЕ СКОРОСТИ СХОДИМОСТИ ГРАДИЕНТНЫХ МЕТОДОВ С РАСШИРЕННОЙ ЛОГАРИФМИЧЕСКОЙ ФУНКЦИЕЙ ПОТЕРЬ

(Самарский национальный исследовательский университет
им. академика С.П. Королева)

Пусть имеется выборка наблюдений $\{\mathrm{x}_i, y_i\}_{i=1}^{m}$, где $\mathrm{x}_i \in \mathrm{R}^{\mathrm{n}}$ и $y_i \in \{0,1\}$. Поставим задачу бинарной классификации

$$L(\boldsymbol{\theta}) \underset{\boldsymbol{\theta}}{\rightarrow} \min, \tag{1}$$

где функция потерь с вектором весов $\boldsymbol{\theta} \in \mathrm{R}^{\mathrm{n}}$ задана в виде

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{m} l(y_i \boldsymbol{\theta}^{\boldsymbol{T}} \mathrm{x}_i). \tag{2}$$

Примем $\forall i: y_i = 1, \|\mathrm{x}_i\| < 1$ [1].

Задача (1) может быть решена аналитически в виде нормального уравнения, однако наиболее часто на практике прибегают к градиентным методам. Метод градиентного спуска или метод первого порядка предполагает итерационное обновление вектора весов $\boldsymbol{\theta}_t$ в направлении антиградиента $-\nabla L(\boldsymbol{\theta}_t)$ с шагом η [1,2]:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \nabla L(\boldsymbol{\theta}_t), \qquad (3)$$

где градиент функции потерь (2) имеет вид:

$$\nabla L(\boldsymbol{\theta}_t) = \sum_{i=1}^{m} l'(\boldsymbol{\theta}^T \mathrm{x}_i)\mathrm{x}_i.$$

В качестве $L(\boldsymbol{\theta})$ могут выступать различные виды функций потерь, но наиболее часто используется логарифмическая функция потерь с $l(\boldsymbol{\theta}^T \mathrm{x}_i) = \frac{1}{1+e^{-\boldsymbol{\theta}^T \mathrm{x}_i}}$. В условиях почти линейной разделимости классов скорость сходимости метода (3) значительно снижается, что требует применения адаптивных методов оптимизации. Наиболее распространенные адаптивные методы позволяют повысить скорость сходимости процедуры поиска вектора решений, однако не обеспечивают границу погрешности, эквивалентную границе стандартного градиентного метода [3].

Метод второго порядка [2]

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \big(H(\boldsymbol{\theta}_t)\big)^{-1} \nabla L(\boldsymbol{\theta}_t), \qquad (4)$$

где $H(\boldsymbol{\theta}_t)$ задает гессиан функции потерь (2), позволяют достичь более высокой скорости сходимости, что, однако, не учитывает напрямую проблему низкой скорости сходимости, являющуюся результатом линейной разделимости классов [2,4].

В данной работе предлагается подход к повышению скорости сходимости градиентных методов через расширение логарифмической функции потерь с учетом пары гиперапараметров a и b [4]:

$$l(\boldsymbol{\theta}^T \mathrm{x}_i; a, b) = a + \frac{b-a}{1+e^{-\boldsymbol{\theta}^T \mathrm{x}_i}}. \qquad (5)$$

С учетом расширения (5) градиентные методы (3) и (4) примут вид

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \nabla L(\boldsymbol{\theta}_t; a, b),$$

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \big(H(\boldsymbol{\theta}_t; a, b)\big)^{-1} \nabla L(\boldsymbol{\theta}_t; a, b),$$

где $L(\boldsymbol{\theta}_t; a, b)$ определяет *расширенную* логарифмическую функцию потерь.

Результаты исследования

Для проверки предложенного подхода к повышению скорости сходимости градиентных методовбыла проведена серия вычислительных экспериментов на различных наборах данных из репозитория UCI ML. На рис. 1 приведены результаты анализа числа итераций градиентного метода 1-ого порядка в стандартном и расширенном представлениях при варьировании объема выборки $m \in [20,100]$ на наборе данных *Parkinsons*. Количество наблюдений $m = 195$, среди которых $m_{y_i=0} = 48$ и $m_{y_i=1} = 147$. Количество признаков $n = 22$.

На рис. 2 приведены результаты аналогичного анализа для градиентного метода 2-ого порядка $m \in [25,100]$ на наборе данных *Parkinsons*.

Из приведенных результатов следует, что предложенное расширение логарифмической функции потерь оказывает положительное влияние на скорость сходимости как метода 1-ого порядка, так и метода 2-ого порядка.

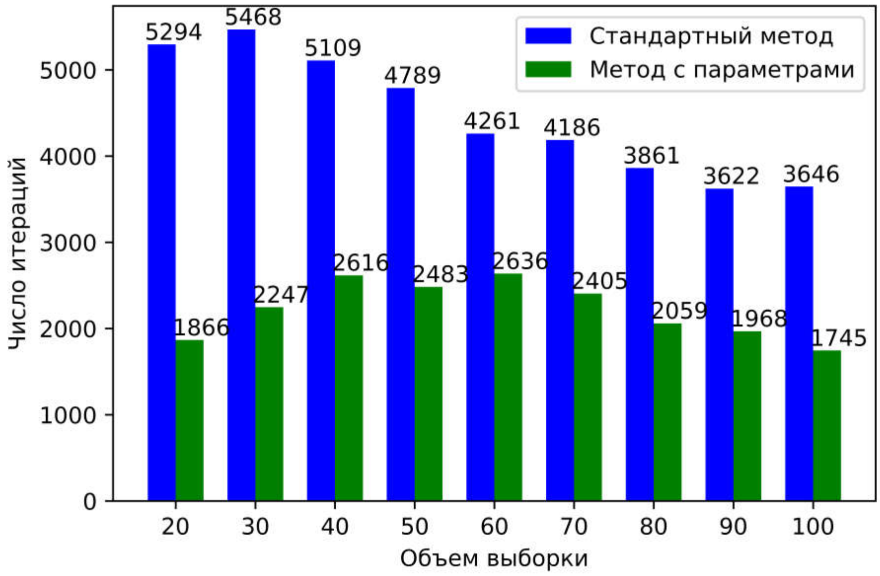

Рисунок 1 – Зависимость числа итераций от объема выборки для градиентного метода 1-ого порядка

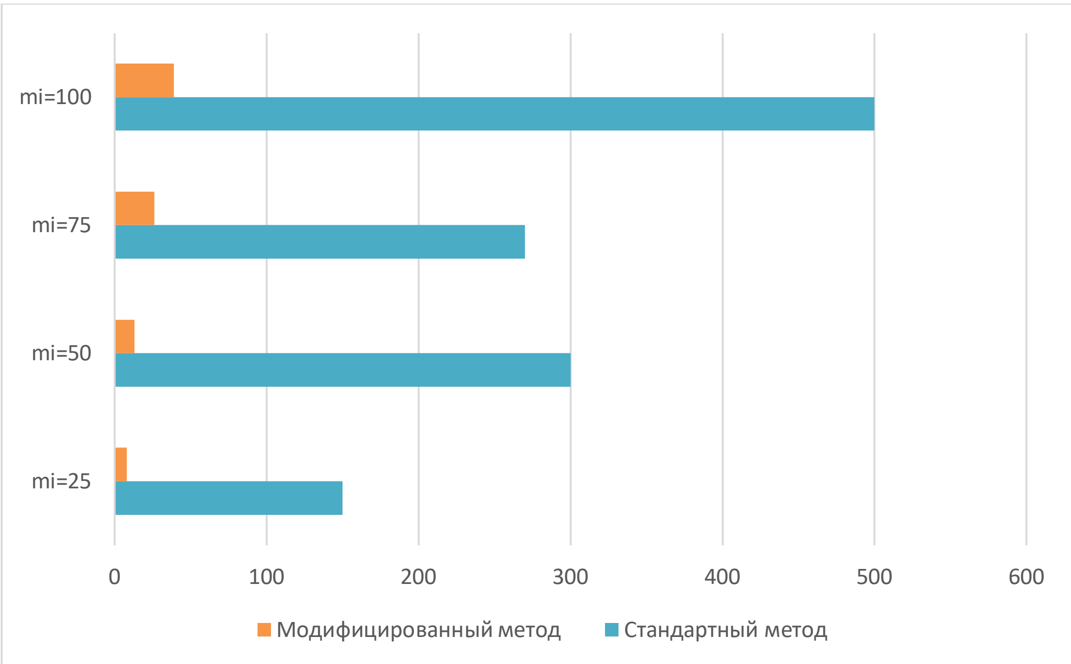

Рисунок 2 – Зависимость числа итераций от объема выборки для градиентного метода 2-ого порядка

Благодарности

Работа выполнена при поддержке гранта Президента РФ (проект № МК-6218.2018.9) и Российского фонда фундаментальных исследований (проект № 18-37-00219).

Литература

1. Soudry D., Hoffer E., Nacson M.S., Gunasekar S., Srebro N. The implicit bias of gradient descent on separable data//JMLR. – 2018. – 19. – pp. 1–57.
2. Воронцов К.В. Математические методы обучения по прецедентам (теория обучения машин). http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf (дата обращения: 28.05.2019)
3. Kim H.S., Kang J.H., Park W.M., Ko S.H., Cho Y.H., Yu D.S., Song Y.S., Choi J.W. Convergence analysis of optimization algorithms. arXiv:1707.01647 (дата обращения: 28.05.2019)
4. Кузнецова К.А., Наумова Ю.А., Панова К.А. Исследования влияния верхней и нижней асимптот на скорость сходимости логарифмической функции потерь//Новые информационные технологии в научных исследованиях: материалы XXIII Всероссийской научно-технической конференции. Рязанский государственный радиотехнический университет. – С. 67-68.

Д.А. Лебедев, В.Г. Литвинов

АВТОМАТИЗИРОВАННАЯ СИСТЕМА КЛАССИФИКАЦИИ МУЗЫКАЛЬНЫХ ЗАПИСЕЙ ПО ЖАНРАМ

(Самарский университет)

Аннотация: в данной статье рассматривается способ классификации музыкальных записей по жанрам. Определены основные принципы, которым должно соответствовать приложение. Исследование показало, что классификация музыкальных записей по жанрам на основе распознанного текста может дать приемлемый по точности результат.

Ключевые слова: выборка, звук, частота звука, распознавание аудио, факторы точности распознавания.

Выборка или выборочная совокупность — часть генеральной совокупности элементов, которая охватывается экспериментом (наблюдением, опросом).

Характеристики выборки:

Качественная характеристика выборки — что именно мы выбираем и какие способы построения выборки мы для этого используем.

Количественная характеристика выборки — сколько случаев выбираем, другими словами объём выборки.

Существует необходимость в сборе вторичной информации. [1].

Звуком называется колебательное движение частиц упругой среды, распространяющееся в виде волн в газообразной, жидкой или твердой среде, которые, воздействуя на слуховой анализатор человека, вызывают слуховые ощущения. Источником звука является колеблющееся тело, например: колебания струны, вибрация камертона, движение диффузора громкоговорителя и др. [2].