

*Пальмов С.В., к.т.н., доцент
доцент кафедры информационных систем и технологий
ФГБОУ ВО ПГУТИ,
доцент кафедры информационных технологий
ФГБОУ ВО «СамГТУ»
Россия, Самара*

**РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА ОСНОВЕ
МАШИННОГО ОБУЧЕНИЯ ДЛЯ СТУДЕНТОВ ВУЗОВ**

Аннотация: В настоящее время достаточно остро стоит вопрос импортозамещения в ИТ. Вузы, реализующие образовательные программы в указанной сфере, вынуждены искать замену привычным продуктам. Одно из решений данной проблемы – самостоятельная разработка. Целью работы являлось создание программного обеспечения, которое может быть использовано для проведения лабораторных работ по дисциплинам, связанным с изучением алгоритмов машинного обучения. Использовались методы стохастического градиентного спуска, высокоуровневого программирования и сравнительного анализа. Заявленное программное обеспечение было разработано и протестировано; получены положительные результаты, подтверждающие возможность его практического использования.

Ключевые слова: стохастический градиентный спуск, искусственный интеллект, Python, машинное обучение, классификация

*Palmov S.V., candidate of technical sciences, associate professor
Associate Professor of the Department of Information Systems and
Technologies
State University of Telecommunications and Informatics,*

Associate Professor of the Department of Information Technologies

Samara State Technical University

Russia, Samara

SOFTWARE FOR UNIVERSITY STUDENTS

BASED ON MACHINE LEARNING

Abstract: The import substitution issue in IT sector has become highly relevant nowadays, leading universities which offer educational programs in this field to search for the usual products alternatives. One possible solution to the problem is self-development. In this paper, we aimed to create software that can be used for laboratory work in disciplines related to the study of machine learning algorithms. Authors utilized methods of stochastic gradient descent, high-level programming, and comparative analysis to develop the software. It was tested, and positive results have been obtained, confirming the possibility of this solution practical usage.

Keywords: stochastic gradient descent, artificial intelligence, Python, machine learning, classification

Введение

В настоящее время в РФ реализуются мероприятия по импортозамещению. Одним из важных направлений является сфера информационных технологий. Вузы, которые готовят специалистов в данной области, также вынуждены вносить изменения в пул используемых программных средств. Одним из способов выхода из сложившейся ситуации является разработка необходимых решений силами образовательной организации. Следовательно, можно утверждать, что статья, посвященная рассмотрению вопроса самостоятельного создания приложения для обучения студентов, будет актуальной.

Цель работы была сформулирована так: создание программного обеспечения, которое может быть использовано для проведения

лабораторных работ по дисциплинам, связанным с изучением алгоритмов машинного обучения.

Методы и исследование

В работе были использованы следующие методы.

Стохастический градиентный спуск (Stochastic Gradient Descent) [1] – это метод оптимизации, который применяется в ситуациях, когда градиентный спуск может функционировать медленно или неэффективно. В отличие от обычной версии алгоритма, работающей на всем наборе данных, рассматриваемый вариант оперирует случайно выбранными небольшими подмножествами, называемых пакетами (или мини-пакетами). Поэтому, обучение выполняется быстрее, так как каждый пакет содержит только часть данных, что уменьшает количество вычислений и заметно сокращает время, затрачиваемое на построение модели.

Язык высокоуровневого программирования Python. С его помощью было реализовано заявленное программное обеспечение (ПО)

Сравнительный анализ применялся для оценки качества работы ПО.

«Градиентный спуск» предоставляется следующие возможности:

- Построение классификатора (математической модели) на основе стохастического градиентного спуска
- Сохранение созданного / загрузка существующего классификатора
- Классификация новых данных с сохранением результатов в текстовый файл
- Режим тестирования классификатора (оценка качества работы на основе наборов данных с известными значениями целевого показателя)

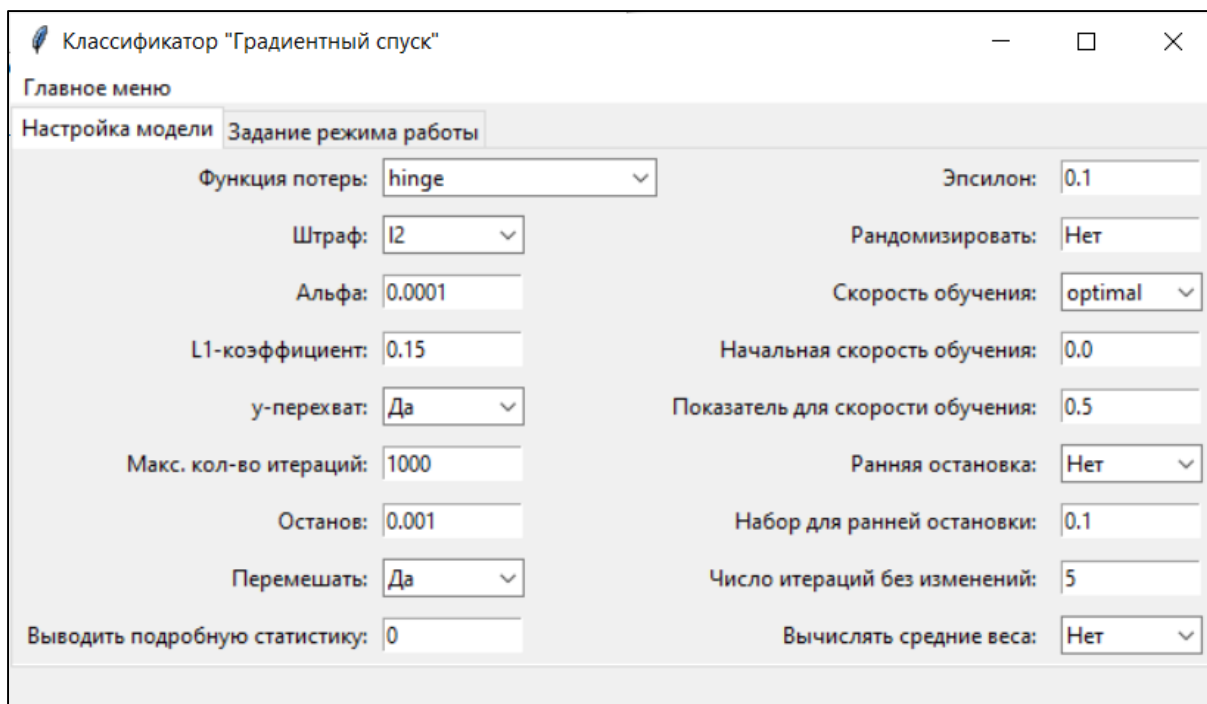


Рис. 1. Интерфейс ПО. Настройка модели



Рис. 2. Интерфейс ПО. Задание режима работы

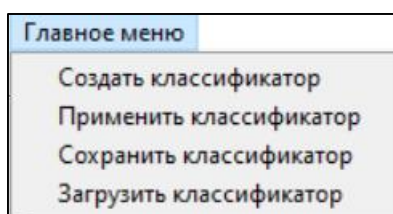


Рис. 3. Интерфейс ПО. Главное меню

Результаты исследования

Для демонстрации возможностей ПО был проведен эксперимент. Исходными данными послужил набор «Голосование депутатов конгресса»

Использовалась перекрестная проверка с k , равным трем; метрика оценки качества – вероятность (P).

Было выполнено девять прогонов ПО (по числу значений параметра Φ)

у

н

к

Таблица 1. Результаты эксперимента

ц	Функция потерь	P
и		
я		
п		
о		
т		
е		
р		
ь		

табл. 1.

Заключение

Как видно из представленных результатов, классификаторы, созданные с использованием первых пяти функций, показали лучшее качество, заметно отличающееся от остальных. Следовательно, последняя четверка значений не является предпочтительной для обработки указанного набора данных. Таким образом, представленное ПО позволяет реализовывать изучение возможностей определенного класса алгоритмов машинного обучения. Это обстоятельство указывает на то, что цель работы достигнута.

Использованные источники:

Никулин, В. Н. Алгоритмы на основе метода стохастического градиентного спуска для рекомендательных систем / В. Н. Никулин, М. Г. Петик // Общество, наука, инновации (НПК - 2015) : Всероссийская ежегодная научно-практическая конференция: Сборник материалов: Общеуниверситетская секция, БФ, ХФ, ФСА, ФАМ, ЭТФ, ФАВТ, ФПМТ, ФЭМ, ФГСН, ЮФ, Киров, 13–24 апреля 2015 года / ФГБОУ ВПО «Вятский государственный университет». – Киров: Вятский государственный университет, 2015. – С. 2033-2034.

2. Пальмов, С. В. Сравнение классификационных возможностей алгоритмов С4.5 и С5.0 / С. В. Пальмов, А. А. Мифтахова // Инфокоммуникационные технологии. – 2015. – Т. 13, № 4. – С. 467-471. – DOI

Сивак, М. А. Построение робастных нейронных сетей с различными функциями потерь / М. А. Сивак, В. С. Тимофеев // Системы анализа и

о

б

р

а

б

о

т

к

и

д

а

н

н

ы

х

.

Ф

О