

Министерство науки и высшего образования Российской Федерации

—  
САНКТ-ПЕТЕРБУРГСКИЙ  
ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО

---

Физико-механический институт  
Высшая школа прикладной математики и вычислительной физики

*Н. О. Кадырова Л. В. Павлова*

СТАТИСТИЧЕСКИЙ АНАЛИЗ  
БОЛЬШИХ ДАННЫХ:  
ПОДХОД НА ОСНОВЕ МАШИН  
ОПОРНЫХ ВЕКТОРОВ

Учебное пособие



**ПОЛИТЕХ-ПРЕСС**

Санкт-Петербургский  
политехнический университет  
Петра Великого

Санкт-Петербург

2022

ББК 22.172я73

К13

**Рецензенты:**

Кандидат физико-математических наук, старший научный сотрудник  
лаборатории лазерной диагностики плазмы и взаимодействия плазмы  
с поверхностью ФТИ им. А. Ф. Иоффе РАН *А. Н. Баженов*

Кандидат физико-математических наук, доцент ВШ ПМиВФ СПбПУ  
*О. И. Заяц*

*Кадырова Н. О. Статистический анализ больших данных: подход на основе машин опорных векторов* : учеб. пособие / Н. О. Кадырова, Л. В. Павлова. – СПб. : ПОЛИТЕХ-ПРЕСС, 2022. – 60 с.

Соответствует содержанию федеральной дисциплины «Машины опорных векторов» государственного образовательного стандарта по направлению подготовки бакалавров 01.03.02 «Прикладная математика и информатика», специальность 01.03.02\_02 «Системное программирование».

Рассмотрены основные принципы и идеи современного подхода к решению задачи восстановления зависимостей по эмпирическим данным. Приведены основные базовые идеи реализации SVM-методов. Сделан обзор наиболее эффективных алгоритмов построения машин опорных векторов для задач бинарной классификации, кластеризации и восстановления регрессии.

Предназначено для студентов, обучающихся по бакалаврским и магистерским программам, а также для аспирантов, изучающих методы и алгоритмы машинного обучения.

Ил. 16. Библиогр.: 43 назв.

Печатается по решению

Совета по издательской деятельности Ученого совета

Санкт-Петербургского политехнического университета Петра Великого.

ISBN 978-5-7422-7813-9

doi:10.18720/SPBPU/2/id22-81

© Кадырова Н. О., Павлова Л. В., 2022

© Санкт-Петербургский политехнический университет Петра Великого, 2022

## ОГЛАВЛЕНИЕ

Введение .....	6
1. Общая постановка задачи обучения с учителем .....	10
2. Структурная минимизация риска .....	11
3. Линейный SVM-классификатор .....	13
3.1. Линейно разделимые данные .....	13
3.2. Линейно неразделимые данные .....	17
4. Нелинейное SVM-обучение .....	20
5. Ядерные методы .....	23
5.1. Представление данных .....	24
5.2. Положительно-определенные ядра .....	24
5.3. Ядро как скалярное произведение .....	24
5.4. Ядро как мера подобия объектов .....	25
5.5. Ядро как мера гладкости функций .....	25
6. SVM-подход к решению задачи восстановления регрессии .....	27
6.1. Постановка задачи .....	27
6.2. Функции потерь .....	28
6.3. Нелинейное SVR-обучение .....	32
7. Качество алгоритмов обучения .....	34
8. Основные этапы построения машин опорных векторов .....	37
8.1. Характеристики качества обучения .....	38
8.2. Стратегии формирования выборок при настройке машины опорных векторов .....	38
8.3. Настройка машины опорных векторов .....	40
9. Основные подходы к построению SV-машин на практике .....	42
9.1. Решение двойственной задачи SVM-обучения .....	42
9.2. Декомпозиция двойственной задачи .....	45
9.3. Базовые идеи SV-алгоритмов .....	45
9.4. Преимущества SVM-подхода .....	47
10. SV-алгоритм решения задачи кластеризации .....	49
10.1. SV-кластеризация .....	49
10.2. Определение границ кластеров .....	50
10.3. Маркировка кластеров .....	53
10.4. Подбор значений гиперпараметров SVC-алгоритма .....	54
БИБЛИОГРАФИЧЕСКИЙ СПИСОК .....	57