

Министерство образования и науки Российской Федерации  
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

---

В.М. НЕДЕЛЬКО

# ОСНОВЫ СТАТИСТИЧЕСКИХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Утверждено Редакционно-издательским советом университета  
в качестве учебного пособия

НОВОСИБИРСК  
2010

УДК 519.23: 004.852  
Н 421

Рецензенты:

д-р техн. наук, проф. *Г.С. Лбов*,  
канд. физ.-мат. наук *Т.А. Ступина*

Работа подготовлена кафедрой ПС и БД

**Неделько В.М.**

Н 421      Основы статистических методов машинного обучения :  
учеб. пособие / В.М. Неделько. – Новосибирск : Изд-во НГТУ,  
2010. – 72 с.

ISBN 978-5-7782-1385-2

В учебном пособии излагаются основы теории и методов машинного обучения в вероятностной постановке. Под машинным обучением понимается анализ данных, при котором выявляются закономерности или строятся модели, описывающие данные. Дисциплина базируется на методах математической статистики.

Пособие предназначено для студентов ФПМИ НГТУ.

УДК 519.23: 004.852

ISBN 978-5-7782-1385-2

© Неделько В.М., 2010  
© Новосибирский государственный  
технический университет, 2010

## ОГЛАВЛЕНИЕ

Предисловие .....	5
Глава 1. НЕКОТОРЫЕ ЗАДАЧИ И МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ .....	6
§ 1.1. Метод прецедентов .....	6
1.1.1. Задача классификации .....	6
1.1.2. Задача восстановления зависимостей .....	9
§ 1.2. Классификация в пространстве бинарных переменных.....	12
1.2.1. Случай известных распределений .....	12
1.2.2. Выборочная оценка .....	15
1.2.3. Гипотеза независимости переменных .....	18
1.2.4. Ряд Бахадура .....	21
§ 1.3. Дискриминантная функция для нормальных распределений .....	24
1.3.1. Случай известных распределений .....	24
1.3.2. Оценивание параметров .....	30
§ 1.4. Деревья решений.....	33
1.4.1. Задача классификации .....	33
1.4.2. Задача восстановления зависимостей. ....	38
§ 1.5. Прогнозирование бинарного временного ряда .....	42

§ 1.6. Кластерный анализ.....	44
1.6.1. Выделение кластеров.....	45
1.6.2. Иерархическая кластеризация.....	48
§ 1.7. Поиск логических закономерностей .....	50
§ 1.8. Задача поиска глобального экстремума.....	51
§ 1.9. Оценивание достоверности решения .....	53
1.9.1. Использование контрольной выборки .....	53
1.9.2. Оценка скользящего экзамена .....	56
1.9.3. Статистическое моделирование.....	58
Глава 2. ЗАДАЧА МАШИННОГО ОБУЧЕНИЯ В ВЕРОЯТНОСТНОЙ ПОСТАНОВКЕ .....	62
§ 2.1. Статистическая постановка задачи анализа данных.....	62
2.1.1. Задача построения решающей функции .....	62
2.1.2. Общая постановка .....	63
2.1.3. Иллюстративный пример .....	64
2.1.4. Варианты формальных постановок .....	66
§ 2.2. Обзор методов машинного обучения .....	69
2.2.1. Методы, основанные на восстановлении распределений .....	69
2.2.2. Методы, конструирующие решающие правила .....	70
Библиографический список .....	72