

УДК 004.93(075)

ББК 16.633я7

М994

Рецензенты: д-р техн. наук, доц. С. Б. П о п о в,
д-р техн. наук, проф. В. А. Ф у р с о в

Мясников, Владислав Валерьевич

М994 **Основы статистической теории распознавания образов и машинного обучения:** практикум / *В.В. Мясников*. – Самара: Издательство Самарского университета, 2023. – 124 с.

ISBN 978-5-7883-1932-2

Издание представляет собой сборник указаний к лабораторным работам по статистической теории распознавания образов, для каждой из которых приводятся описание теоретических основ, задание и список контрольных вопросов. Разделы теоретических основ содержат описания наиболее распространенных методов и алгоритмов статистической теории распознавания образов: методов оптимальной классификации (классификатор Байеса, минимаксный классификатор и классификатор Неймана-Пирсона), методов построения и настройки линейных классификаторов и классификаторов, основанных на оценивании плотностей вероятностей, а также методов автоматической классификации – кластеризации и таксономии.

Настоящий практикум предназначен для обучающихся факультета информатики по направлениям 01.03.02 Прикладная математика и информатика и 10.05.03 Информационная безопасность автоматизированных систем.

УДК 004.93(075)

ББК 16.633я7

ISBN 978-5-7883-1932-2

© Самарский университет, 2023

ОГЛАВЛЕНИЕ

Введение	7
1 Моделирование экспериментальных данных	
для решения задач распознавания образов	8
1.1 Теоретические основы лабораторной работы.....	8
1.1.1 Моделирование случайного вектора с нормальным законом распределения.....	8
1.1.2 Оценивание параметров нормального закона распределения.....	11
1.1.3 Меры близости нормальных распределений	11
1.1.4 Моделирование бинарных случайных векторов с независимыми координатами	14
1.2 Порядок выполнения лабораторной работы.....	15
1.2.1 Исходные данные	15
1.2.2 Общий план выполнения работы.....	15
1.2.3 Содержание отчета.....	16
2 Оптимальные стратегии теории статистического	
расознавания образов	17
2.1 Теоретические основы лабораторной работы.....	17
2.1.1 Постановка задачи классификации	17
2.1.2 Качество классификатора	20
2.1.3 Оптимальные стратегии классификации: классификатор Байеса.....	21
2.1.4 Оптимальные стратегии классификации: минимаксный классификатор	24
2.1.5 Оптимальные стратегии классификации: классификатор Неймана-Пирсона	27
2.1.6 Типовые решения оптимальных стратегий классификации: нормально распределенные вектора признаков	28
2.1.7 Типовые решения оптимальных стратегий классификации: бинарные вектора признаков	33

2.1.8 Вычисление вероятностей ошибочной классификации: общий случай.....	35
2.1.9 Вычисление вероятностей ошибочной классификации: нормально распределенные вектора признаков	37
2.1.10 Вычисление вероятностей ошибочной классификации: бинарные вектора признаков.....	39
2.1.11 Экспериментальная оценка вероятностей ошибочной классификации.....	40
2.2 Порядок выполнения лабораторной работы.....	42
2.2.1 Исходные данные	42
2.2.2 Общий план выполнения работы.....	42
2.2.3 Содержание отчета.....	43
3 Линейные классификаторы	44
3.1 Теоретические основы лабораторной работы.....	44
3.1.1 Постановка задачи построения линейного классификатора	44
3.1.2 Линейный классификатор, минимизирующий суммарную вероятность ошибочной классификации	45
3.1.3 Обобщенная формула построения линейных классификаторов для различных критериев. Классификатор Фишера.....	49
3.1.4 Линейный классификатор, минимизирующий СКО решения	51
3.1.5 Последовательная корректировка линейного классификатора: алгоритм перцептрона	55
3.1.6 Последовательная корректировка линейного классификатора: стохастическая аппроксимация и процедура Роббинса-Монро.....	57
3.1.7 Общая схема построения линейных классификаторов, основанная на методе стохастической аппроксимации.....	60
3.2 Порядок выполнения лабораторной работы.....	63
3.2.1 Исходные данные	63
3.2.2 Общий план выполнения работы.....	64
3.2.3 Содержание отчета.....	65

4 Линейные классификаторы. Метод опорных векторов	66
4.1 Теоретические основы лабораторной работы.....	66
4.1.1. Метод опорных векторов для случая линейно разделимых классов	66
4.1.2. Метод опорных векторов для случая линейно неразделимых классов	70
4.1.3 Использование ядер в методе опорных векторов и переход в пространства высокой размерности.....	71
4.1.4 Достоинства и недостатки метода опорных векторов ..	73
4.2 Порядок выполнения лабораторной работы.....	74
4.2.1 Исходные данные	74
4.2.2 Общий план выполнения работы.....	74
4.2.3 Содержание отчета.....	75
5 Классификаторы, основанные на непараметрических оценках плотности вероятностей	76
5.1 Теоретические основы лабораторной работы.....	76
5.1.1. Метод Парзена оценивания плотности вероятностей..	77
5.1.2. Метод К ближайших соседей.....	81
5.1.3 Оценка априорных вероятностей.....	84
5.2 Порядок выполнения лабораторной работы.....	84
5.2.1 Исходные данные	84
5.2.2 Общий план выполнения работы.....	85
5.2.3 Содержание отчета.....	85
6 Автоматическая классификация	86
6.1 Теоретические основы лабораторной работы.....	86
6.1.1 Постановка задачи автоматической классификации	86
6.1.2 Меры сходства.....	87
6.1.3 Критерии кластеризации	88
6.1.4 Простой алгоритм выделения кластеров	89
6.1.5 Алгоритм максиминного расстояния	90
6.1.6 Алгоритм К внутригрупповых средних	92
6.3 Порядок выполнения лабораторной работы.....	96
6.3.1 Исходные данные	96

6.3.2 Общий план выполнения работы.....	96
6.3.3 Содержание отчета.....	97
7 Справочные сведения о языке программирования PYTHON и средствах машинного обучения.....	98
7.1 Язык Python.....	98
7.1.1 Общие сведения.....	98
7.1.2 Установка интерпретатора языка и пакетов расширений.....	100
7.2 PyCharm – интегрированная среда разработки для языка программирования Python	105
7.3 Библиотека машинного обучения scikit-learn (machine learning in Python)	108
7.4 Прикладная программа синтеза и классификации данных....	109
7.4.1 Синтеза данных для классификации	109
7.4.2 Классификации данных с использованием метода опорных векторов на scikit-learn.....	110
8 Контрольные вопросы.....	114
8.1 Контрольные вопросы к лабораторной работе № 1	114
8.2 Контрольные вопросы к лабораторной работе № 2	114
8.3 Контрольные вопросы к лабораторной работе № 3	115
8.4 Контрольные вопросы к лабораторной работе № 4	115
8.5 Контрольные вопросы к лабораторной работе № 5	116
8.6 Контрольные вопросы к лабораторной работе № 6	117
9 Варианты заданий	118
9.1 Варианты заданий к лабораторным работам № 1–5	118
9.2 Варианты заданий к лабораторной работе 6.....	119
Заключение	121
Библиографический список	122