

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное учреждение  
высшего профессионального образования  
"Казанский (Приволжский) федеральный университет"  
Институт вычислительной математики и информационных технологий



**УТВЕРЖДАЮ**

Проректор  
по образовательной деятельности КФУ  
Проф. Таюрский Д.А.

"\_\_" \_\_\_\_\_ 20\_\_ г.

**Программа дисциплины**

Методы оптимизации в задачах анализа данных Б1.В.ДВ.2

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

**Автор(ы):**

Габидуллина З.Р.

**Рецензент(ы):**

Миссаров М.Д.

**СОГЛАСОВАНО:**

Заведующий(ая) кафедрой: Миссаров М. Д.

Протокол заседания кафедры No \_\_\_\_ от " \_\_\_\_ " \_\_\_\_\_ 201\_\_ г

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

Протокол заседания УМК No \_\_\_\_ от " \_\_\_\_ " \_\_\_\_\_ 201\_\_ г

Регистрационный No

Казань  
2017

## Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) доцент, к.н. (доцент) Габидуллина З.Р. кафедра анализа данных и исследования операций отделение фундаментальной информатики и информационных технологий , Zulfiya.Gabidullina@kpfu.ru

### 1. Цели освоения дисциплины

Курс "Методы оптимизации в задачах анализа данных" имеет целью усвоение студентами теоретических знаний и приобретение практических навыков к решению задач классификации и идентификации данных, к их анализу, к построению классификаторов различными методами и использованию их для принятия решений.

### 2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " Б1.В.ДВ.2 Дисциплины (модули)" основной образовательной программы 01.04.02 Прикладная математика и информатика и относится к дисциплинам по выбору. Осваивается на 1 курсе, 2 семестр.

Дисциплина Б1.В.ДВ3 "методы оптимизации в задачах анализа данных" входит в учебный план "Прикладная математика и информатика Магистерская программа "Анализ данных и его приложения" (как дисциплина на выбор)). Изучается на 1 курсе во 2 семестре.

### 3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОКП-4 (общекультурные компетенции)	--способность использовать и применять углубленные знания в области прикладной математики и информатики;
ОК-1 (общекультурные компетенции)	--способность к абстрактному мышлению, анализу, синтезу;
ОК-2 (общекультурные компетенции)	-умение логически верно, аргументированно и ясно строить устную и письменную речь;
ОК-3 (общекультурные компетенции)	-готовность к саморазвитию, самореализации, использованию творческого потенциала;
ПК-1 (профессиональные компетенции)	--способность собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям;
ПК-2 (профессиональные компетенции)	--способность приобретать новые научные и профессиональные знания, используя современные образовательные и информационные технологии;

В результате освоения дисциплины студент:

1. должен знать:

роль и место оптимизации и анализа данных в системе фундаментальных и прикладных математических дисциплин.

2. должен уметь:

применять основные результаты методов оптимизации к решению задач анализа данных.

3. должен владеть:

основным математическим аппаратом решения оптимизационных задач.

4. должен демонстрировать способность и готовность:

-формулировать задачи анализа данных в виде оптимизационных математических моделей.

#### 4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 5 зачетных(ые) единиц(ы) 180 часа(ов).

Форма промежуточного контроля дисциплины экзамен во 2 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

#### 4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

##### Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Основные определения.	2	1	1	1	0	Письменное домашнее задание
2.	Тема 2. Одномерная идентификация данных и ранжирование.	2	2	1	1	0	Письменное домашнее задание
3.	Тема 3. Применение метода главного эксперта в задачах идентификации.	2	3-4	1	1	0	Устный опрос
4.	Тема 4. Различные подходы к построению классификаторов (идентификаторов данных).	2	5-6	1	1	0	Письменное домашнее задание

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
5.	Тема 5. Новый подход (Габидуллиной З.Р.): использование разности Минковского классов данных в постановке оптимизационных задач.	2	7-8	2	2	0	Письменное домашнее задание
6.	Тема 6. Математические методы решения оптимизационных задач для построения сепаратора или псевдо-сепаратора классов.	2	9-10	2	2	0	Устный опрос
7.	Тема 7. Метод Гильберта и метод Митчелла-Демьянова-Малоземова для построения классификатора обучающих классов в отделимом случае.	2	11-12	2	2	0	Письменное домашнее задание
8.	Тема 8. Универсальные решатели, реализующие критерии линейной отделимости классов (на примере, минимаксной оптимизационной задачи и критерия отделимости Габидуллиной З.Р.).	2	13-14	2	2	0	Письменное домашнее задание
9.	Тема 9. Использование метода главного эксперта в медицинской диагностике.	2	15-18	2	2	0	Письменное домашнее задание Контрольная работа
	Тема . Итоговая форма контроля	2		0	0	0	Экзамен
	Итого			14	14	0	

## 4.2 Содержание дисциплины

### Тема 1. Основные определения.

*лекционное занятие (1 часа(ов)):*

Типы данных. Математическое представление данных. Классы данных. Типы классов (пересекающиеся и непересекающиеся). Задача классификации. Классификация объектов (четкая и нечеткая, бинарная и мультиклассовая). Пространство признаков. Метрики, используемые для отображения матрицы расстояний между объектами. Алгоритм К-ближайших соседей.

**практическое занятие (1 часа(ов)):**

Практическое изучение типов данных и математического представления данных. Разбор на примерах классов данных и типов классов - пересекающихся и непересекающихся. Рассмотрение задачи классификации: классификация объектов (четкая и нечеткая, бинарная и мультиклассовая). Изучение пространства признаков. Разбор метрик, используемых для отображения матрицы расстояний между объектами. Анализ алгоритма К- ближайших соседей.

**Тема 2. Одномерная идентификация данных и ранжирование.**

**лекционное занятие (1 часа(ов)):**

Идентификация методом разделения. Метод изоляции. Ранжирование параметров.

**практическое занятие (1 часа(ов)):**

Практическое рассмотрение одномерной идентификации данных: - идентификация методом разделения. - метод изоляции. Разбор ранжирования параметров.

**Тема 3. Применение метода главного эксперта в задачах идентификации.**

**лекционное занятие (1 часа(ов)):**

Постановка задачи идентификации. Основные идеи метода главного эксперта (Демьяновой В.В) для машинного обучения с учителем.

**практическое занятие (1 часа(ов)):**

Практическое изучение метода главного эксперта в задачах идентификации. Рассмотрение постановки задачи идентификации. Разбор основных идей метода главного эксперта (Демьяновой В.В) для машинного обучения с учителем.

**Тема 4. Различные подходы к построению классификаторов (идентификаторов данных).**

**лекционное занятие (1 часа(ов)):**

Классический подход (Bennet K., Mangasarian O.): оптимизационные задачи, применяемые в SVM (Support Vector Machines).

**практическое занятие (1 часа(ов)):**

Ознакомление с различными подходами к построению классификаторов (идентификаторов) данных. Практическое изучение на примерах классического подхода (Bennet K., Mangasarian O.). Решение оптимизационных задач, применяемых в SVM (Support Vector Machines).

**Тема 5. Новый подход (Габидуллиной З.Р.): использование разности Минковского классов данных в постановке оптимизационных задач.**

**лекционное занятие (2 часа(ов)):**

Новый подход (Габидуллиной З.Р.): использование разности Минковского классов данных в постановке оптимизационных задач.

**практическое занятие (2 часа(ов)):**

Практическое рассмотрение на примерах нового подхода (Габидуллиной З.Р.), основанного на использовании разности Минковского классов данных в постановке оптимизационных задач.

**Тема 6. Математические методы решения оптимизационных задач для построения сепаратора или псевдо-сепаратора классов.**

**лекционное занятие (2 часа(ов)):**

Мера толщины сепаратора (отделяющего слоя) в отделимом случае. Мера толщины псевдо-сепаратора (слоя неправильно классифицированных точек) в случае неотделимости классов

**практическое занятие (2 часа(ов)):**

Практическое рассмотрение на примерах меры ошцины сепаратора (отделяющего слоя) в отделимом случае. Расчет меры толщины псевдо-сепаратора (слоя неправильно классифицированных точек) в случае неотделимости классов

### **Тема 7. Метод Гильберта и метод Митчелла-Демьянова-Малоземова для построения классификатора обучающих классов в отделимом случае.**

**лекционное занятие (2 часа(ов)):**

Метод Гильберта и метод Митчелла-Демьянова-Малоземова для построения классификатора обучающих классов в отделимом случае.

**практическое занятие (2 часа(ов)):**

Практическое изучение на примерах метода Гильберта и метода Митчелла-Демьянова-Малоземова для построения классификатора обучающих классов в отделимом случае.

### **Тема 8. Универсальные решатели, реализующие критерии линейной отделимости классов (на примере, минимаксной оптимизационной задачи и критерия отделимости Габидуллиной З.Р.).**

**лекционное занятие (2 часа(ов)):**

Универсальные решатели, реализующие критерии линейной отделимости классов (на примере, минимаксной оптимизационной задачи и критерия отделимости Габидуллиной З.Р.) . Реализация построения классификатора и псевдо-классификатора в пакете Matlab.

**практическое занятие (2 часа(ов)):**

Рассмотрение на примере минимаксной оптимизационной задачи и критерия отделимости Габидуллиной З.Р. универсальных решателей, реализующих критерии линейной отделимости классов . Выполнение реализации построения классификатора и псевдоклассификатора в пакете Matlab.

### **Тема 9. Использование метода главного эксперта в медицинской диагностике.**

**лекционное занятие (2 часа(ов)):**

Использование метода главного эксперта для прогнозирования эффективности различных типов терапии для лечения заболеваний. Использование базы данных СНМО-253 (разбиение базы данных на подбазы).

**практическое занятие (2 часа(ов)):**

Подробное рассмотрение использования метода главного эксперта для прогнозирования эффективности различных типов терапии для лечения заболеваний. Разбор использования базы данных СНМО-253 (разбиение базы данных на подбазы).

## **4.3 Структура и содержание самостоятельной работы дисциплины (модуля)**

<b>N</b>	<b>Раздел Дисциплины</b>	<b>Семестр</b>	<b>Неделя семестра</b>	<b>Виды самостоятельной работы студентов</b>	<b>Трудоемкость (в часах)</b>	<b>Формы контроля самостоятельной работы</b>
1.	Тема 1. Основные определения.	2	1	подготовка домашнего задания	10	домашнее задание
2.	Тема 2. Одномерная идентификация данных и ранжирование.	2	2	подготовка домашнего задания	10	домашнее задание
3.	Тема 3. Применение метода главного эксперта в задачах идентификации.	2	3-4	подготовка к устному опросу	12	устный опрос



N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
4.	Тема 4. Различные подходы к построению классификаторов (идентификаторов данных).	2	5-6	подготовка домашнего задания	12	домашнее задание
5.	Тема 5. Новый подход (Габидуллиной З.Р.): использование разности Минковского классов данных в постановке оптимизационных задач.	2	7-8	подготовка домашнего задания	12	домашнее задание
6.	Тема 6. Математические методы решения оптимизационных задач для построения сепаратора или псевдо-сепаратора классов.	2	9-10	подготовка к устному опросу	12	устный опрос
7.	Тема 7. Метод Гильберта и метод Митчелла-Демьянова-Малоземова для построения классификатора обучающих классов в отделимом случае.	2	11-12	подготовка домашнего задания	12	домашнее задание
8.	Тема 8. Универсальные решатели, реализующие критерии линейной отделимости классов (на примере, минимаксной оптимизационной задачи и критерия отделимости Габидуллиной З.Р.).	2	13-14	подготовка домашнего задания	12	домашнее задание
9.	Тема 9. Использование метода главного эксперта в медицинской диагностике.	2	15-18	подготовка домашнего задания	12	домашнее задание
				подготовка к контрольной работе	12	контрольная работа
Итого					116	

## 5. Образовательные технологии, включая интерактивные формы обучения

Лекции и практические занятия проводятся в аудитории, оснащенной доской и мелом.



## **6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов**

### **Тема 1. Основные определения.**

домашнее задание , примерные вопросы:

Изучение лекционного материала. Выполнение домашнего задания на тему: -Задача классификации. Классификация объектов (четкая и нечеткая, бинарная и мультиклассовая).

### **Тема 2. Одномерная идентификация данных и ранжирование.**

домашнее задание , примерные вопросы:

Изучение лекционного материала. Выполнение домашнего задания на тему: -Одномерная идентификация данных и ранжирование.

### **Тема 3. Применение метода главного эксперта в задачах идентификации.**

устный опрос , примерные вопросы:

Изучение лекционного материала. Подготовка к устному опросу по пройденным темам.

### **Тема 4. Различные подходы к построению классификаторов (идентификаторов данных).**

домашнее задание , примерные вопросы:

Изучение лекционного материала. Выполнение домашнего задания на тему: -Различные подходы к построению классификаторов (идентификаторов данных).

### **Тема 5. Новый подход (Габидуллиной З.Р.): использование разности Минковского классов данных в постановке оптимизационных задач.**

домашнее задание , примерные вопросы:

Изучение лекционного материала. Выполнение домашнего задания на тему: -Новый подход (Габидуллиной З.Р.): использование разности Минковского классов данных в постановке оптимизационных задач.

### **Тема 6. Математические методы решения оптимизационных задач для построения сепаратора или псевдо-сепаратора классов.**

устный опрос , примерные вопросы:

Изучение лекционного материала. Подготовка к устному опросу по пройденным темам.

### **Тема 7. Метод Гильберта и метод Митчелла-Демьянова-Малоземова для построения классификатора обучающих классов в отделимом случае.**

домашнее задание , примерные вопросы:

Изучение лекционного материала. Выполнение домашнего задания на тему: -Метод Гильберта и метод Митчелла-Демьянова-Малоземова для построения классификатора обучающих классов в отделимом случае.

### **Тема 8. Универсальные решатели, реализующие критерии линейной отделимости классов (на примере, минимаксной оптимизационной задачи и критерия отделимости Габидуллиной З.Р.).**

домашнее задание , примерные вопросы:

Изучение лекционного материала. Выполнение домашнего задания на тему: -Универсальные решатели, реализующие критерии линейной отделимости классов (на примере, минимаксной оптимизационной задачи и критерия отделимости Габидуллиной З.Р.).

### **Тема 9. Использование метода главного эксперта в медицинской диагностике.**

домашнее задание , примерные вопросы:

Изучение лекционного материала. Выполнение домашнего задания на тему: -Использование метода главного эксперта в медицинской диагностике.

контрольная работа , примерные вопросы:

Подготовка к контрольной работе (выполнению индивидуальных заданий) по всем пройденным темам.

## Тема . Итоговая форма контроля

Примерные вопросы к экзамену:

Примерные вопросы к экзамену:

1. Типы данных. Математическое представление данных.
2. Классы данных. Типы классов (пересекающиеся и непересекающиеся).
3. Задача классификации. Классификация объектов (четкая и нечеткая, бинарная и мультиклассовая).
4. Пространство признаков. Метрики, используемые для отображения матрицы расстояний между объектами.
5. Алгоритм К- ближайших соседей.
6. Одномерная идентификация данных.
7. Ранжирование.
8. Применение метода главного эксперта в задачах идентификации.
9. Различные подходы к построению классификаторов (идентификаторов данных)
10. Постановка задачи идентификации.
11. Основные идеи метода главного эксперта (Демьяновой В.В) для машинного обучения с учителем.
12. Классический подход (Bennet K., Mangasarian O.): оптимизационные задачи, применяемые в SVM (Support Vector Machines)
13. Новый подход (Габидуллиной З.Р.): использование разности Минковского классов данных в постановке оптимизационных задач.
14. Математические методы решения оптимизационных задач для построения сепаратора классов. или псевдо-сепаратора классов.
15. Математические методы решения оптимизационных задач для построения псевдо-сепаратора классов.
16. Метод Гильберта для построения классификатора обучающих классов в отделимом случае.
17. Метод Митчелла-Демьянова-Малоземова для построения классификатора обучающих классов в отделимом случае.
18. Универсальные решатели, реализующие критерии линейной отделимости классов (на примере, минимаксной оптимизационной задачи и критерия отделимости Габидуллиной З.Р.).
19. Использование метода главного эксперта в медицинской диагностике.

### 7.1. Основная литература:

1. Методы оптимизации: Учебное пособие / А.В. Аттетков, В.С. Зарубин, А.Н. Канатников. - М.: ИЦ РИОР: НИЦ Инфра-М, 2013. - 270 с.: ил.; 60x90 1/16. - (Высшее образование: Бакалавриат). (переплет) ISBN 978-5-369-01037-2, 700 экз.  
<http://www.znaniyum.com/bookread.php?book=350985>
2. Ашманов С.А. Тимохов А.В. Теория оптимизации в задачах и упражнениях. - Санкт-Петербург: Лань, 2012. - 448с.  
[http://e.lanbook.com/books/element.php?pl1\\_id=3799](http://e.lanbook.com/books/element.php?pl1_id=3799)
3. Лесин, В.В. Основы методов оптимизации. [Электронный ресурс] / В.В. Лесин, Ю.П. Лисовец. ? Электрон. дан. ? СПб. : Лань, 2016. ? 344 с. ? Режим доступа: <http://e.lanbook.com/book/86017> ? Загл. с экрана.
4. Интеллектуальный анализ временных рядов [Электронный ресурс]: Учебное пособие / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. - М.: ИД ФОРУМ: ИНФРА-М, 2012. - 160 с. . - Режим доступа: <http://www.znaniyum.com/bookread.php?book=249314>

## 7.2. Дополнительная литература:

1. Степанов, Роман Григорьевич. Технология Data Mining: Интеллектуальный анализ данных: учебное пособие / Р. Г. Степанов; Казан. гос. ун-т. Казань: Казанский государственный университет, 2009. 110 с
2. Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. 3-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2009. 512 с.: ил. + CD-ROM (Учебная литература для вузов).  
<http://www.znanium.com/bookread.php?book=350638>
3. Заботин, И.Я. Методы и вычислительные приемы в линейном программировании [Электронный ресурс] : учебное пособие / И.Я. Заботин, Я.И. Заботин. Электрон. дан. Казань : КФУ (Казанский (Приволжский) федеральный университет), 2014. 116 с. Режим доступа: [http://e.lanbook.com/books/element.php?pl1\\_id=72810](http://e.lanbook.com/books/element.php?pl1_id=72810) Загл. с экрана.
4. Информационные аналитические системы [Электронный ресурс] : учебник / Т. В. Алексеева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - М.: МФПУ Синергия, 2013. - 384 с. - (Университетская серия). - ISBN 978-5-4257-0092-6.  
<http://www.znanium.com/bookread.php?book=451186>

## 7.3. Интернет-ресурсы:

Data Mining Labs - <http://dmlabs.org/>

The R Project - <http://www.r-project.org/>

курс - <http://www.intuit.ru/studies/courses/2312/612/info>

курс - <http://www.intuit.ru/studies/courses/6/6/info>

Программный комплекс RStudio для интеллектуального анализа данных - <http://www.rstudio.com/>

## 8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Методы оптимизации в задачах анализа данных" предполагает использование следующего материально-технического обеспечения:

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен студентам. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, УМК, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен студентам. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

Лекции и практические занятия проводятся в аудитории, оснащенной доской и мелом (маркером).

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 01.04.02 "Прикладная математика и информатика" и магистерской программе Анализ данных и его приложения .

Автор(ы):

Габидуллина З.Р. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.

Рецензент(ы):

Миссаров М.Д. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.