

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



УТВЕРЖДАЮ

Проректор по образовательной деятельности КФУ

Проф. Д. А. Таюрский

» _____ 20__ г.

подписано электронно-цифровой подписью

Программа дисциплины

Методы оптимизации в задачах анализа данных

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Математическое моделирование

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2020

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

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. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОПК-1	Способен решать актуальные задачи фундаментальной и прикладной математики
ОПК-2	Способен совершенствовать и реализовывать новые математические методы решения прикладных задач

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

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

Должен уметь:

применять основные результаты методов оптимизации к решению задач анализа данных.

Должен владеть:

основным математическим аппаратом решения оптимизационных задач.

Должен демонстрировать способность и готовность:

-формулировать задачи анализа данных в виде оптимизационных математических моделей.

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.О.03 Дисциплины (модули)" основной профессиональной образовательной программы 01.04.02 "Прикладная математика и информатика (Математическое моделирование)" и относится к обязательным дисциплинам.

Осваивается на 1 курсе в 2 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

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

Контактная работа - 32 часа(ов), в том числе лекции - 16 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 16 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 94 часа(ов).

Контроль (зачёт / экзамен) - 54 часа(ов).

Форма промежуточного контроля дисциплины: экзамен во 2 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины / модуля	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)			Самостоятельная работа
			Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Основные определения.	2	1	0	1	10
2.	Тема 2. Одномерная идентификация данных и ранжирование.	2	1	0	1	4
3.	Тема 3. Применение метода главного эксперта в задачах идентификации.	2	2	0	2	4
4.	Тема 4. Различные подходы к построению классификаторов (идентификаторов данных).	2	2	0	2	16
5.	Тема 5. Новый подход (Габидуллиной З.Р.): использование разности Минковского классов данных в постановке оптимизационных задач.	2	2	0	2	16
6.	Тема 6. Математические методы решения оптимизационных задач для построения сепаратора или псевдо-сепаратора классов.	2	2	0	2	12
7.	Тема 7. Метод Гильберта и метод Митчелла-Демьянова-Малоземова для построения классификатора обучающихся классов в отделимом случае.	2	2	0	2	14
8.	Тема 8. Универсальные решатели, реализующие критерии линейной отделимости классов (на примере, минимаксной оптимизационной задачи и критерия отделимости Габидуллиной З.Р.).	2	2	0	2	12
9.	Тема 9. Использование метода главного эксперта в медицинской диагностике.	2	2	0	2	6
	Итого		16	0	16	94

4.2 Содержание дисциплины (модуля)

Тема 1. Основные определения.

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 K- ближайших соседей.

Тема 2. Одномерная идентификация данных и ранжирование.

Идентификация методом разделения. Метод изоляции. Ранжирование параметров. Теоретические вопросы, а также вопросы практического применения методов на одномерных задачах с различным числом идентифицируемых данных. Изучение графической интерпретации результатов практического решения задач одномерной идентификации.

Тема 3. Применение метода главного эксперта в задачах идентификации.

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

Изучение метода с использованием примера выпуклых многогранников, заданных в виде выпуклой комбинации конечного набора векторов. Графическая интерпретация работы метода.

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

Классический подход (Bennet K., Mangasarian O.): оптимизационные задачи, применяемые в SVM (Support Vector Machines). Оптимизационные методы, применимые для решения задач.

Изучение подхода на примере выпуклых многогранников, заданных в виде выпуклой комбинации конечного набора векторов. Графическая интерпретация.

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

Новый подход (Габидуллиной З.Р.): использование разности Минковского классов данных в постановке оптимизационных задач. Разность Минковского для выпуклых многогранников, заданных в виде выпуклой комбинации конечного набора векторов. Построение разности Минковского для различных пар выпуклых многогранников.

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

Мера толщины сепаратора (отделяющего слоя) в отделимом случае. Мера толщины псевдо-сепаратора (слоя неправильно классифицированных точек) в случае неотделимости классов. Построение графической интерпретации для случая выпуклых многогранников, заданных в виде выпуклой комбинации конечного набора векторов.

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

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

Реализация построения классификатора в пакете Matlab, графическая интерпретация результатов решения. Изучение теоретического материала и особенностей применения, вопросы сходимости.

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

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, графическая интерпретация результатов решения. Изучение математического аппарата.

Тема 9. Использование метода главного эксперта в медицинской диагностике.

Использование метода главного эксперта для прогнозирования эффективности различных типов терапии для лечения заболеваний. Использование базы данных СНЕМО-253 (разбиение базы данных на подбазы). Изучение импорта и экспорта данных в пакете прикладных программ, средств графической интерпретации результатов.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Вид работ	Методические рекомендации
лекции	Реализация данной дисциплины предполагает как очную, так и дистанционную форму обучения. Студентам рекомендуется посещать лекционные занятия. На лекционных занятиях, каждому студенту рекомендуется вести подробные записи лекционных материалов, задавать преподавателю возникающие вопросы, перед каждой лекцией просматривать теоретический материал предыдущей лекции для лучшего освоения следующей темы.
лабораторные работы	Реализация данной дисциплины предполагает как очную, так и дистанционную форму обучения. Для выполнения лабораторной работы требуются навыки работы в пакете прикладных программ Matlab для решения оптимизационных задач, позволяющих строить линейный классификатор отделимых данных, псевдо-классификатор неотделимых данных, вычисления толщины сепаратора и псевдо-сепаратора, построения графической интерпретации полученных результатов. Примеры генерации отделяемых классов данных: 1. $z_1=(0,2), z_2=(3,3), z_3=(2,0); p_1=(-2,0), p_2=(-3,-3), p_3=(0,-2)$. 2. $z_1=(-3,3), z_2=(3,3), z_3=(3,-3); p_1=(-1,1), p_2=(-2,-2), p_3=(3,-3)$. 3. $z_1=(-2,3), z_2=(3,2), z_3=(3,-3); p_1=(-1,-1), p_2=(0,2), p_3=(2,0)$.
самостоятельная работа	Каждому студенту необходимо самостоятельно проработать сценарии решения каждого типа задач, разобранных на занятиях. Следуя рекомендациям преподавателя, выполнять все аудиторские и домашние задания (в том числе, выполнять все лабораторные задания по решению задач отделения данных и готовиться к активному участию в обсуждении теоретических вопросов).
экзамен	Используя лекционный материал, литературу из основного и дополнительного списка подготовиться по следующим вопросам к экзамену: 1. Одномерная идентификация данных и ранжирование. Постановка задачи и метод решения. 2. Постановка задачи идентификации. Применение метода главного эксперта в задачах идентификации. 3. Задача классификации данных. Различные подходы к построению классификаторов (идентификаторов данных). 4. Новый подход (Габидуллиной З.Р.): использование разности Минковского классов данных в постановке оптимизационных задач. 5. Математические методы решения оптимизационных задач для построения сепаратора или псевдо-сепаратора классов. 6. Построение классификатора путем сведения к задаче о дополнении. 7. Метод последовательных проектирований Брегмана. Постановка задачи отделения классов данных. 8. Метод Гильберта для построения классификатора обучающих классов в отделимом случае. 9. Метод Митчелла-Демьянова-Малоземова для построения классификатора обучающих классов в отделимом случае. 10. Универсальные решатели, реализующие критерии линейной отделимости классов (на примере, минимаксной оптимизационной задачи и критерия отделимости Габидуллиной З.Р.). 11. Постановка задачи медицинской диагностики. 12. Использование метода главного эксперта в медицинской диагностике. 14. Визуализация элементов баз данных, результатов классификации. 15. Оптимизационный подход Мангазариана-Беннет к классификации данных.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

Мультимедийная аудитория.

Компьютерный класс.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 01.04.02 "Прикладная математика и информатика" и магистерской программе "Математическое моделирование".

Приложение 2
к рабочей программе дисциплины (модуля)
Б1.О.03 Методы оптимизации в задачах анализа данных

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Математическое моделирование

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2020

Основная литература:

1. Пантелеев, А. В. Методы оптимизации. Практический курс: учебное пособие с мультимедиа сопровождением / А. В. Пантелеев, Т. А. Летова. - Москва : Логос, 2011. - 424 с: ил. (Новая университетская библиотека). - ISBN 978-5-98704-540-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/469213> (дата обращения: 12.05.2020). - Режим доступа: по подписке.
2. Аттетков, А. В. Методы оптимизации: учебное пособие / А.В. Аттетков, В.С. Зарубин, А.Н. Канатников. - Москва: ИЦ РИОР: НИЦ Инфра-М, 2019. - 270 с. ил.; - (Высшее образование: Бакалавриат). - ISBN 978-5-16-103309-8. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1002733> (дата обращения: 12.05.2020). - Режим доступа: по подписке.
3. Дайитбегов, Д. М. Компьютерные технологии анализа данных в эконометрике: монография / Д.М. Дайитбегов. - 3-е изд., испр. и доп. - Москва : Вузовский учебник: НИЦ Инфра-М, 2013. - XIV, 587 с. (Научная книга). ISBN 978-5-9558-0275-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/365692> (дата обращения: 12.05.2020). - Режим доступа: по подписке.
4. Ашманов, С. А. Теория оптимизации в задачах и упражнениях : учебное пособие / С. А. Ашманов, А. В. Тимохов. - 2-е изд., стер. - Санкт-Петербург : Лань, 2012. - 448 с. - ISBN 978-5-8114-1366-9. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/3799> (дата обращения: 12.05.2020). - Режим доступа: для авториз. пользователей.
5. Лесин, В. В. Основы методов оптимизации : учебное пособие / В. В. Лесин, Ю. П. Лисовец. - 4-е изд., стер. - Санкт-Петербург : Лань, 2016. - 344 с. - ISBN 978-5-8114-1217-4. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/86017> (дата обращения: 12.05.2020). - Режим доступа: для авториз. пользователей.
6. Ярушкина, Н. Г. Интеллектуальный анализ временных рядов: учебное пособие / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. - Москва : ИД ФОРУМ: ИНФРА-М, 2012. - 160 с. (Высшее образование). ISBN 978-5-8199-0496-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/249314> (дата обращения: 12.05.2020). - Режим доступа: по подписке.

Дополнительная литература:

1. Степанов, Р. Г. Технология Data Mining: интеллектуальный анализ данных: учебное пособие / Р. Г. Степанов; Казан. гос. ун-т. - Казань: Казанский государственный университет, 2009. - 110 с.
2. Анализ данных и процессов: учебное пособие / Барсегян А.А., Куприянов М.С., Холод И.И. - Санкт-Петербург: БХВ-Петербург, 2009. - 512 с. - ISBN . - Текст : электронный. - URL: <https://znanium.com/catalog/product/350638> (дата обращения: 12.05.2020). - Режим доступа: по подписке.
3. Заботин, И. Я. Методы и вычислительные приемы в линейном программировании : учебное пособие / И. Я. Заботин, Я. И. Заботин. - Казань : КФУ, 2014. - 116 с. - ISBN 978-5-00019-272-6. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/72810> (дата обращения: 12.05.2020). - Режим доступа: для авториз. пользователей.
4. Алексеева, Т. В. Информационные аналитические системы: учебник / Т. В. Алексеева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - Москва : МФПУ Синергия, 2013. - 384 с. - (Университетская серия). - ISBN 978-5-4257-0092-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/451186> (дата обращения: 12.05.2020). - Режим доступа: по подписке.

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.О.03 Методы оптимизации в задачах анализа данных

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Математическое моделирование

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2020

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. 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, учебно-методические комплексы, 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 образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен обучающимся. 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.