

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В.Ломоносова»

УТВЕРЖДАЮ

Декан факультета ВМК МГУ
имени М.В. Ломоносова



_____ Е.И. Моисеев
академик

«__» _____ 2015 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

«Нестатистические методы анализа данных и классификации»

Уровень высшего образования – подготовка научно-педагогических кадров в аспирантуре

Направление подготовки – 01.06.01 «Математика и механика»

Направленность (профиль) – 01.01.09 «Дискретная математика и математическая кибернетика»

2015

1. НАИМЕНОВАНИЕ ДИСЦИПЛИНЫ

Нестатистические методы анализа данных и классификации

2. УРОВЕНЬ ВЫСШЕГО ОБРАЗОВАНИЯ

Подготовка научно-педагогических кадров в аспирантуре

3. НАПРАВЛЕНИЕ ПОДГОТОВКИ, НАПРАВЛЕННОСТЬ (ПРОФИЛЬ) ПОДГОТОВКИ

Направление подготовки – 01.06.01 «Математика и механика». Направленность (профиль) – 01.01.09 «Дискретная математика и математическая кибернетика»

4. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОСНОВНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина относится к специальным дисциплинам вариативной части образовательной программы.

5. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

Дисциплина участвует в формировании следующих компетенций образовательной программы:

Формируемые компетенции	Планируемые результаты обучения
Способностью формулировать научные задачи в области обеспечения информационной безопасности, применять для их решения методологии теоретических и экспериментальных научных исследований, внедрять полученные результаты в практическую деятельность (ОПК-1);	З1(ОПК-1) ЗНАТЬ научные задачи в области обеспечения информационной безопасности У1(ОПК-1) УМЕТЬ: применять для их решения методологии теоретических и экспериментальных научных исследований, внедрять полученные результаты в практическую деятельность В1(ОПК-1) ВЛАДЕТЬ: Навыками внедрения полученных результатов в практическую деятельность

<p>Владение современными методами построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также методами разработки и реализации алгоритмов их решения на основе фундаментальных знаний в области математики и информатики (ПК-1)</p>	<p>З1 (ПК-1) ЗНАТЬ: современные методы разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения У1 (ПК-1) УМЕТЬ: применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения В1 (ПК-1) ВЛАДЕТЬ: навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методов разработки и реализации алгоритмов их решения</p>
--	--

Оценочные средства для промежуточной аттестации приведены в Приложении.

6. ОБЪЕМ ДИСЦИПЛИНЫ

Объем дисциплины составляет 3 зачетные единицы, всего 108 часов.

24 часа составляет контактная работа с преподавателем – 22 часа занятий лекционного типа, 0 часов занятий семинарского типа (семинары, научно-практические занятия, лабораторные работы и т.п.), 0 часов индивидуальных консультаций, 0 часов мероприятий текущего контроля успеваемости, 0 часов групповых консультаций, 2 часа мероприятий промежуточной аттестации.

84 часов составляет самостоятельная работа аспиранта.

7. ВХОДНЫЕ ТРЕБОВАНИЯ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Учащиеся должны владеть знаниями по дискретной математике и основам кибернетики в объеме, соответствующем основным образовательным программам бакалавриата и магистратуры по укрупненным группам направлений и специальностей 01.00.00 «Математика и механика», 02.00.00 «Компьютерные и информационные науки».

8. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

В процессе обучения технические и программные средства не используются.

9. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

В курсе излагаются основные результаты и методы теории сложности вычислений. Основное внимание уделяется классам сложности и отношениям между ними.

Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе								
		Контактная работа (работа во взаимодействии с преподавателем), часы					Самостоятельная работа обучающегося, часы			
		из них					из них			
		Занятия лекционного типа	Занятия семинарского типа	Групповые консультации	Индивидуальные консультации	Учебные занятия, направленные на проведение текущего контроля успеваемости (коллоквиумы, практические контрольные занятия и др)*	Всего	Выполнение домашних заданий	Подготовка рефератов и т.п.	Всего

<p>Тема 1. Введение в статистические методы анализа данных и классификации Введение в курс. Основные разделы и задачи. Решение задач медицинской и технической диагностики, геологического прогнозирования, распознавания солеобразования в нефтепромысловом оборудовании, распознавание рукописных цифр, понимание текста.</p>	6	2					2		4	4
<p>Тема 2. Тестовый алгоритм классификации. Основные определения. Поиск тупиковых тестов как поиск неприводимых покрытий матрицы сравнения.</p>	6	2					2		4	4

<p>Тема 3. Тестовый алгоритм классификации. Стохастическая версия тестового алгоритма. Веса тупиковых тестов.</p>	6	2					2		4	4
<p>Тема 3. Модели вычисления оценок (АВО) Этапы моделей АВО, способы их выполнения. Свойства моделей АВО с опорными множествами одной мощности.</p>	6	2					2		4	4
<p>Тема 5. Решающие правила над оценками, свойства оценок. Модели АВО для метрических и порядковых признаков.</p>	6	2					2		4	4
<p>Тема 6. Эффективные формулы вычисления оценок при системе опорных множеств равной мощности по всем опорным множествам и при заданных параметрах «веса признаков» и «веса строк».</p>	6	2					2		4	4

<p>Тема 7. Оптимизация моделей классификации, практические модели АВО в системе РАСПОЗНАВАНИЕ.</p> <p>Стандартный функционал качества распознавания. Релаксационный метод решения совместных систем линейных неравенств. Точный комбинаторный подход. Методы оптимизации функционала (релаксационный, комбинаторный, аппроксимационный). Проблема переобучения и робастные модели.</p>	8	2					2		6	6
---	---	---	--	--	--	--	---	--	---	---

<p>Тема 8. Логические модели анализа данных и классификации. Эвристический подход. Логические закономерности классов (ЛЗК), основные определения. Стандартный функционал качества ЛЗК и частотный критерий. Три алгоритма сокращения границ в ЛЗК. Сведение к задаче ЦЛП.</p>	8	2					2		6	6
---	---	---	--	--	--	--	---	--	---	---

<p>Тема 9. Применения логических закономерностей в анализе данных и классификации. Распознавание на базе ЛЗК, взвешивание и сглаживание в решающих правилах, обобщенные прецеденты, определение выбросов и оптимизация обучающих выборок. Минимизация признакового пространства в задачах распознавания с использованием множеств ЛЗК.</p>	8	2					2		6	6
<p>Тема 10. Бинарные решающие деревья, их свойства. Допустимые разбиения и их поиск. Свойства допустимых разбиений, примеры. Практическое построение БРД.</p>	8	2					2		6	6

Тема 11. Назначение программной системы РАСПОЗНАВАНИЕ, ее структура. Описание графической оболочки, главные окно и меню, методы системы. Ввод и преобразование данных. Визуализация данных.	8	2				2		6	6	
12. Промежуточная аттестация – устный экзамен	32	2					30			
Итого	108	24					84			

10. УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ УЧАЩИХСЯ

Самостоятельная работа учащихся состоит в изучении лекционного материала, учебно-методической литературы, подготовки к экзамену.

11.РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ

Основная учебно-методическая литература

- 1) Журавлев Ю.И. Избранные научные труды. - М.: Магистр, 1998. 420 с.
- 2) Zhuravlev Yu. I., Nazarenko G. I., Vinogradov A. P., Dokukin A. A., Katerinochkina N. N., Kleimenova E. B., Konstantinova M. V., Ryazanov V.V., Sen'ko O. V., and Cherkashov A. M. // Methods for discrete analysis of medical information based on recognition theory and some of their applications. Pattern Recognition and Image Analysis. 2016. Vol. 26, No.3. pp. 643-664.
- 3) Duda, R. O., Hart, P. E., and Stork, D. G. : Pattern Classification. John Wiley and Sons. 2nd edition (2000).

- 4) Журавлев Ю.И., Рязанов В.В., О.В.Сенько, РАСПОЗНАВАНИЕ. Математические методы. Программная система. Практические применения. Изд.во «ФАЗИС», Москва, 2006, 168 с.
- 5) Рязанов В.В. Логические закономерности в задачах распознавания (параметрический подход) // Журнал вычислительной математики и математической физики, Т.47, №10, 2007, с.1793-1808.

Дополнительная учебно-методическая литература

- 6) T. Hastie, R. Tibshirani, J. Friedman The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media. 2013.
- 7) Aggarwal C. Data Mining: The Textbook. IBM T.J. Watson Research Center. Yorktown Heights. New York. 2015. 771 p. DOI: 10.1007/978-3-319-14142-8.
- 8) Katerinochkina N.N. Some approaches to the solution of optimization problems in supervised learning. Computational Mathematics and Mathematical Physics, 2015. Vol. 55, No. 11, pp. 1933–1939.
- 9) I. H. Witten, E. Frank Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). — Morgan Kaufmann, 2005 ISBN 0-12-088407-0
- 10) Biryukov A, Shmakov A, Ryazanov V. Solving the problems of cluster analysis by collectives of algorithms. Journal of Computational Mathematics and Mathematical Physics. 2008;48(1)176-192. DOI: 10.1134/S0965542508010132.

Информационные технологии, используемые в процессе обучения

В процессе обучения используются пакет прикладных программ РАСПОЗНАВАНИЕ.

Материально-техническая база

Медиапроектор и экран для проведения лекций-презентаций.

12. ЯЗЫК ПРЕПОДАВАНИЯ

Русский

13. РАЗРАБОТЧИК ПРОГРАММЫ

Разработчик – д.ф.-м.н. В.В.Рязанов

14. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

РЕЗУЛЬТАТ ОБУЧЕНИЯ	КРИТЕРИИ и ПОКАЗАТЕЛИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТА ОБУЧЕНИЯ из соответствующих карт компетенций					ОЦЕНОЧНЫЕ СРЕДСТВА
	1	2	3	4	5	
	Неудовлетворительно	Неудовлетворительно	Удовлетворительно	Хорошо	Отлично	
<p>УМЕТЬ: самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий У1 (ОПК-1)</p>	Отсутствие умений	Частично освоенное умение самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий	В целом успешное, но не систематическое умение самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий	В целом успешное, но содержащее отдельные пробелы умение самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий	Успешное и систематическое умение самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий	доклад на научном семинаре

<p>ЗНАТЬ: современные методы исследования и информационно-коммуникационных технологий в соответствующей профессиональной области З1(ОПК-1)</p>	<p>Отсутствие знаний</p>	<p>Фрагментарные представления современных методов исследования и информационно-коммуникационных технологий в соответствующей профессиональной области</p>	<p>В целом сформированные, но неполные знания о современных методах исследования и информационно-коммуникационных технологий в соответствующей профессиональной области</p>	<p>Сформированные, но содержащие отдельные пробелы знания о современных методах исследования и информационно-коммуникационных технологий в соответствующей профессиональной области</p>	<p>Сформированные систематические знания о современных методах исследования и информационно-коммуникационных технологий в соответствующей профессиональной области</p>	<p>Устный экзамен</p>
<p>ЗНАТЬ: современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения З1 (ПК-1)</p>	<p>Отсутствие знаний</p>	<p>Фрагментарные представления о современных методах построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методах разработки и реализации алгоритмов их решения</p>	<p>В целом сформированные, но неполные знания о современных методах построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методах разработки и реализации алгоритмов их решения</p>	<p>Сформированные, но содержащие отдельные пробелы знания о современных методах построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методах разработки и реализации алгоритмов их решения</p>	<p>Сформированные систематические знания о современных методах построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методах разработки и реализации алгоритмов их решения</p>	<p>Устный экзамен</p>

<p>УМЕТЬ: применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения У1 (ПК-1)</p>	<p>Отсутствие умений</p>	<p>Фрагментарные умения применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения</p>	<p>В целом успешное, но не систематическое умение применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения</p>	<p>Успешное, но содержащее отдельные пробелы умение применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения</p>	<p>Сформированное умение применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения</p>	<p>Устный экзамен</p>
<p>ВЛАДЕТЬ: навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методов разработки и реализации алгоритмов их решения В1 (ПК-1)</p>	<p>Отсутствие навыков</p>	<p>Фрагментарное владение навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методов разработки и реализации алгоритмов их решения</p>	<p>В целом успешное, но не полное владение навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методов разработки и реализации алгоритмов их решения</p>	<p>Успешное, но содержащее отдельные пробелы владение навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методов разработки и реализации алгоритмов их решения</p>	<p>Сформированное владение навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методов разработки и реализации алгоритмов их решения</p>	<p>Устный экзамен, контрольные работы</p>

Темы коллоквиума.

- 1) Основные подходы нестатистического анализа данных и классификации.
- 2) Модели голосования (частичной прецедентности)
- 3) Методы обучения при классификации данных.
- 4) Частотный функционал построения ЛЗК.
- 5) Задача поиска логических закономерностей данных.
- 6) Практические подходы к построению бинарных решающих деревьев.
- 7) Леса из бинарных решающих деревьев.
- 8) Архитектура системы РАСПОЗНАВАНИЕ.

Групповая творческая работа.

Учащиеся создают новое приложение по своему выбору. Работа должна продемонстрировать наличие знаний, умений и навыков по следующим темам.

- 1) Методы частичной прецедентности в больших задачах.
- 2) Задачи обучения в нестатистическом анализе данных и классификации
- 3) Реализация частотного критерия построения ЛЗК.
- 4) Построение ЛЗК для порядковых признаков.
- 5) Построение ЛЗК для больших задач.
- 6) Создание оптимального леса бинарных решающих деревьев.
- 7) Архитектура системы РАСПОЗНАВАНИЕ