

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

академик



Е.И. Моисеев

2015 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

«Нестатистический анализ данных»

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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

1. НАИМЕНОВАНИЕ ДИСЦИПЛИНЫ

Нестатистический анализ данных

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 вариативным спецкурсам (по выбору) образовательных программ.

5. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

Дисциплина участвует в формировании следующих компетенций образовательной программы:

Формируемые компетенции	Планируемые результаты обучения
ПК-1 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	З1 (ПК-1) Знать: 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) 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
Способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе междисциплинарных областях. (УК -1)	У1 (УК-1) Уметь: анализировать альтернативные варианты решения исследовательских и практических задач и оценивать потенциальные выигрыши/проигрыши реализации этих вариантов В1(УК-1) Владеть: навыками анализа методологических проблем, возникающих при решении исследовательских и практических задач, в том числе в междисциплинарных областях
Способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий (ОПК-1)	З1(ОПК-1) Знать: современные методы исследования и информационно-коммуникационных технологий в соответствующей профессиональной области У1(ОПК-1) Уметь: уметь самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно коммуникационных технологий

6. ОБЪЕМ ДИСЦИПЛИНЫ

Объем дисциплины составляет 3 зачетные единицы, всего 108 часов.

36 часов составляет контактная работа с преподавателем –занятия лекционного типа.

72 часа составляет самостоятельная работа аспиранта.

7. ВХОДНЫЕ ТРЕБОВАНИЯ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Аспиранты должны владеть знаниями по курсам, связанным с основами математического анализа, аналитической геометрии и дискретной математики, в объеме, соответствующем основным образовательным программам бакалавриата по укрупненным группам направлений и специальностей 01.00.00 «Математика и механика», 02.00.00 «Компьютерные и информационные науки».

8. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Процесс изложения учебного материала сопровождается презентациями и демонстрацией теории и методов решения задач вычислительной геометрии в интерактивном режиме с использованием мультимедийного проектора.

9. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

В курсе дается обзор основных понятий, концепций и методов для построения непрерывных морфологических моделей на основе медиального представления формы объектов. Рассматривается граничное описание формы объектов на основе многоугольной аппроксимации границ бинарного изображения, а также медиальное описание формы объектов на основе скелета и радиальной функции. Для обоих типов моделей предлагаются алгоритмы их построения, использующие в качестве входных данных бинарные растровые изображения.

Наименование и краткое содержание разделов и тем дисциплины, форма промежуточной аттестации по дисциплине	Всего (часы)	В том числе								
		Контактная работа (работа во взаимодействии с преподавателем), часы					Самостоятельная работа учащегося, часы			
		из них					из них			
Занятия лекционного типа	Занятия семинарского типа	Групповые консультации	Индивидуальные консультации	Учебные занятия, направленные на проведение текущего контроля успеваемости: экзамен, практические контрольные занятия и др.	Всего	Выполнение домашних заданий	Подготовка рефератов и т.п..	Всего		
Тема 1. Задача распознавания (классификации) по прецедентам. Тупиковые тесты и их вычисление, случаи вещественных признаков, стохастический тестовый алгоритм.	6	2	-	-	-	-	2	4	-	4
Тема 2. Логические модели распознавания для разнотипных признаков (алгоритмы вычисления)	12	4	-	-	-	-	4	8	-	8

оценок, алгоритм «Кора», модели с представительными наборами, практические реализации).										
Тема 3. Эффективное вычисление оценок для различных типов признаков. Оптимизация моделей распознавания. Релаксационный и комбинаторный алгоритмы поиска максимальной совместной подсистемы системы линейных неравенств. Логические закономерности классов, эвристический критерий качества логических закономерностей.	12	4	-	-	-	-	4	8	-	8
Тема 4. Сведение задачи поиска логических закономерностей классов к задаче целочисленного линейного программирования. Генетические методы поиска, генетический алгоритм поиска логиче-	12	4	-	-	-	-	4	8	-	8

ских закономерностей классов.										
Тема 5. Веса признаков и прецедентов, логические корреляции. Минимизация признакового пространства. Логические описания классов. Обработка множеств логических закономерностей. Минимальные и кратчайшие описания классов. Алгоритмы распознавания, основанные на голосовании по системам логических закономерностей, построение устойчивых на обучении оценок.	12	4	-	-	-	-	4	8	-	8
Тема 6. Построение минимальных по сложности логических закономерностей классов. Бинарные решающие деревья. Допустимые разбиения единичного куба. Алгоритм построения допустимого разбиения, представление допустимого разбиения бинар-	16	4	-	-	-	-	4	12	-	12

<p>ным решающим деревом. Прямые методы построения бинарных решающих деревьев, критерии ветвления.</p>									
<p>Тема 7. Задачи кластерного анализа, меры близости, функции подобия. Критерии качества кластеризации при заданном числе кластеров: дисперсионный и родственные критерии, основанные на матрицах рассеяния, след в качестве критерия, основанные на матрицах рассеяния, определитель матрицы внутригруппового рассеяния. Критерии кластеризации при неизвестном числе кластеров, меры концентрации.</p>	18	6				6	12		12
<p>Тема 8. Итеративная оптимизация критериев кластеризации на примере дисперсионного критерия. Алгоритмы иерархической группировки. Критерии устойчи-</p>	12	4				4	8		8

ности кластеризаций и их вычисление.											
Тема 8. Алгоритмы кластеризации, основанные на поиске центров сгущений (алгоритмы k-средних, нечетких k-средних, ФОРЕЛЬ, построения оптимальных покрытий). Эвристические алгоритмы кластеризации, алгоритмы «к-эталонов» и «взаимного поглощения». Восстановление компонент смеси по заданной обучающей выборке.	6	2					2	4		4	
Промежуточная аттестация –устный экзамен	2	-	-	-	-		2		-		
Итого	108						36	72			

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.
2. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации. Проблемы кибернетики. М.: Наука. 1978. Вып. 33. С.5-68.
3. Донской В.И., Башта А.И. Дискретные модели принятия решений при неполной информации. Симферополь: Таврия. 1992.
4. Рязанов В.В. Логические закономерности в задачах распознавания (параметрический подход) // Журнал вычислительной математики и математической физики. Т.47, № 10. 2007. С. 1793-1808.
5. Ковшов Н.В., Моисеев В.Л., Рязанов В.В. Алгоритмы поиска логических закономерностей в задачах распознавания // Журнал вычислительной математики и математической физики. 2008. Т.48, № 2. С. 329-344.

Дополнительная учебно-методическая литература

- 1) T. Hastie, R. Tibshirani, J. Friedman The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media. 2013.
- 2) H. Witten, E. Frank. Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). — Morgan Kaufmann, 2005 ISBN 0-12-088407-0.
- 3) Ковшов Н.В., Моисеев В.Л., Рязанов В.В. Алгоритмы поиска логических закономерностей в задачах распознавания. Журнал вычислительной математики и математической физики, Т.48, 2008, N 2, стр. 329-344.
- 4) Загоруйко Н. Г. Прикладные методы анализа данных и знаний. — Новосибирск: ИМ СО РАН, 1999. ISBN 5-86134-060-9
- 5) Рязанов В.В. Логические закономерности в задачах распознавания (параметрический подход) // Журнал вычислительной математики и математической физики, Т.47, №10, 2007, с.1793-1808.
- 6) Zhuravlev Yu. I., Nazarenko G. I., Vinogradov A. P., Dokukin A. A., Katerinotchkina 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.

Ресурсы информационно-телекоммуникационной сети «Интернет»

Информационные технологии, используемые в процессе обучения

В процессе обучения используется пакет прикладных программ РАСПОЗНАВАНИЕ. Основная цель спецкурса состоит в изложении основанных на оптимизационных, дискретных и эвристических подходах методов анализа данных. Рассматриваются логические модели распознавания (классификации с учителем) и анализа разнотипных многомерных данных, методы оптимизации моделей распознавания, алгоритмы поиска скрытых логических закономерностей и связей по признаковым описаниям, методы создания качественных моделей объектов, ситуаций, явлений или процессов. Рассматриваются практические численные методы решения данных задач, и их применения в медицине, бизнесе, химии, технике и других областях.

Материально-техническая база

Медиапроектор и экран для проведения лекций-презентаций.

12. ЯЗЫК ПРЕПОДАВАНИЯ

Русский

13. РАЗРАБОТЧИК ПРОГРАММЫ

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	
УМЕТЬ: самостоятельно осуществлять научно-исследовательскую деятельность в соответ-	Отсутствие умений	Частично освоенное умение самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с ис-	В целом успешное, но не систематическое умение самостоятельно осуществлять научно-исследовательскую деятельность в соответ-	В целом успешное, но содержащее отдельные пробелы умение самостоятельно осуществлять научно-исследовательскую деятельность в соот-	Успешное и систематическое умение самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профес-	доклад на научном семинаре

<p>ствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий</p> <p>У1 (ОПК-1)</p>		<p>пользованием современных методов исследования и информационно-коммуникационных технологий</p>	<p>нальной области с использованием современных методов исследования и информационно-коммуникационных технологий</p>	<p>ветствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий</p>	<p>сиональной области с использованием современных методов исследования и информационно-коммуникационных технологий</p>	
<p>ЗНАТЬ: современные методы исследования и информационно-коммуникационных технологий в соответствующей профессиональной области</p> <p>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</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>

реализации алгоритмов их решения 31 (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 У1 (ПК-1)	Отсутствие умений	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	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	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	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	Устный экзамен
ВЛАДЕТЬ: навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методов разработки и реализации	Отсутствие навыков	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	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	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	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	Устный экзамен, контрольные работы

алгоритмов их решения В1 (ПК-1)						
УМЕТЬ: анализировать альтернативные варианты решения исследовательских и практических задач и оценивать потенциальные выигрыши/проигрыши реализации этих вариантов Код У1 (УК-1)	Отсутствие умений	Частично освоенное умение анализировать альтернативные варианты решения исследовательских и практических задач и оценивать потенциальные выигрыши/проигрыши реализации этих вариантов	В целом успешно, но не систематически осуществляемые анализ альтернативных вариантов решения исследовательских и практических задач и оценка потенциальных выигрышей/проигрышей реализации этих вариантов	В целом успешно, но содержащие отдельные пробелы анализ альтернативных вариантов решения исследовательских задач и оценка потенциальных выигрышей/проигрышей реализации этих вариантов	Сформированное умение анализировать альтернативные варианты решения исследовательских и практических задач и оценивать потенциальные выигрыши/проигрыши реализации этих вариантов	Устный экзамен
ВЛАДЕТЬ: навыками анализа методологических проблем, возникающих при решении исследовательских и практических задач, в том числе в междисциплинарных областях Код В1 (УК-1)	Отсутствие навыков	Фрагментарное применение навыков анализа методологических проблем, возникающих при решении исследовательских и практических задач	В целом успешное, но не систематическое применение навыков анализа методологических проблем, возникающих при решении исследовательских и практических задач	В целом успешное, но содержащее отдельные пробелы применение навыков анализа методологических проблем, возникающих при решении исследовательских и практических задач	Успешное и систематическое применение навыков анализа методологических проблем, возникающих при решении исследовательских и практических задач, в том числе в междисциплинарных областях	Устный экзамен

Темы экзамена.

1. Задача распознавания (классификации) по прецедентам. Тупиковые тесты и их вычисление, случаи вещественных признаков, стохастический тестовый алгоритм.
2. Логические модели распознавания для разнотипных признаков (алгоритмы вычисления оценок, алгоритм «Кора», модели с представительными наборами, практические реализации).
3. Эффективное вычисление оценок для различных типов признаков.
4. Оптимизация моделей распознавания. Релаксационный и комбинаторный алгоритмы поиска максимальной совместной подсистемы системы линейных неравенств.
5. Логические закономерности классов, эвристический критерий качества логических закономерностей.
6. Сведение задачи поиска логических закономерностей классов к задаче целочисленного линейного программирования.
7. Логические закономерности классов, стандартный критерий качества логических закономерностей и его оптимизация.
8. Генетические методы поиска, генетический алгоритм поиска логических закономерностей классов.
9. Веса признаков и прецедентов, логические корреляции. Минимизация признакового пространства.
10. Логические описания классов. Обработка множеств логических закономерностей. Минимальные и кратчайшие описания классов.
11. Алгоритмы распознавания, основанные на голосовании по системам логических закономерностей, построение устойчивых на обучении оценок.
12. Построение минимальных по сложности логических закономерностей классов.
13. Бинарные решающие деревья. Допустимые разбиения единичного куба.
14. Алгоритм построения допустимого разбиения, представление допустимого разбиения бинарным решающим деревом.
15. Прямые методы построения бинарных решающих деревьев, критерии ветвления.
16. Задачи кластерного анализа, меры близости, функции подобия.
17. Критерии качества кластеризации при заданном числе кластеров: дисперсионный и родственные критерии, основанные на матрицах рассеяния, след в качестве критерия, основанные на матрицах рассеяния, определитель матрицы внутригруппового рассеяния.
18. Критерии кластеризации при неизвестном числе кластеров, меры концентрации.
19. Итеративная оптимизация критериев кластеризации на примере дисперсионного критерия.
20. Алгоритмы иерархической группировки.
21. Критерии устойчивости кластеризаций и их вычисление
22. Алгоритмы кластеризации, основанные на поиске центров сгущений (алгоритмы к-средних, нечетких к-средних, ФОРЕЛЬ, построения оптимальных покрытий).
23. Эвристические алгоритмы кластеризации, алгоритмы «к-эталонов» и «взаимного поглощения».
24. Восстановление компонент смеси по заданной обучающей выборке.
25. Нейросетевые алгоритмы обучения и самообучения. Метод встречного распространения, сеть Хопфилда, алгоритмы Хэбба.
26. Логические корректоры для решения задач классификации.
27. Решение задач кластерного анализа коллективами алгоритмов, метод коллективных к-средних.
28. Комитетный синтез коллективных кластеризаций, критерии качества коллективных кластеризаций.

29. Нахождение оптимальных коллективных решений задачи кластерного анализа.
30. Эвристические методы построения коллективных кластеризаций и практические алгоритмы
31. Методы восстановления регрессионных зависимостей по прецедентам, основанные на решении задач распознавания и дискретной оптимизации