

МИНОБРНАУКИ РОССИИ



Федеральное государственное автономное образовательное учреждение
высшего образования
«**Российский государственный гуманитарный университет**»
(**ФГАОУ ВО «РГГУ»**)

ОТДЕЛЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ В ГУМАНИТАРНОЙ СФЕРЕ
Кафедра математики, логики и интеллектуальных систем в гуманитарной сфере

ЭФФЕКТИВНЫЕ АЛГОРИТМЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

45.04.04 – Интеллектуальные системы в гуманитарной среде

Когнитивное и программное обеспечение интеллектуальных роботов и программирование
интеллектуальных систем

Уровень высшего образования: магистратура

Форма обучения: очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2025

Эффективные алгоритмы интеллектуального анализа данных
Рабочая программа дисциплины

Составитель(и):

Кандидат технических наук, доцент Л.О. Шашкин

УТВЕРЖДЕНО

Протокол заседания кафедры МЛиИС

№ 3 от 10.12.2024

ОГЛАВЛЕНИЕ

1. Пояснительная записка	5
1.1 Цель и задачи дисциплины	5
1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций	5
1.3. Место дисциплины в структуре образовательной программы	7
2. Структура дисциплины	7
3. Содержание дисциплины	7
4. Образовательные технологии	8
5. Оценка планируемых результатов обучения	9
5.1 Система оценивания	9
5.2 Критерии выставления оценки по дисциплине	10
5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине	11
6. Учебно-методическое и информационное обеспечение дисциплины	12
6.1 Список источников и литературы	12
6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».	12
7. Материально-техническое обеспечение дисциплины	13
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов	13
9. Методические материалы	14
9.1 Планы семинарских/ практических/ лабораторных занятий	14
9.2 Методические рекомендации по подготовке письменных работ	17
Приложение 1. Аннотация рабочей программы дисциплины	19

1. Пояснительная записка

1.1 Цель и задачи дисциплины

Цель дисциплины. В процессе обучения эффективным алгоритмам интеллектуального анализа данных преследуются несколько целей. Одна из них – подготовить выпускника, способного внедрить алгоритмы ИАД в интеллектуальных системах. Другой целью курса можно считать научение слушателей применять эффективные алгоритмы интеллектуального анализа данных к реальным данным.

Задачи дисциплины: освоение способов разработки эффективных алгоритмов ИАД и навыков, необходимых для получения требуемых компетенций в области создания интеллектуальных систем с машинным обучением.

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Знает принципы сбора, отбора и обобщения информации УК-1.2 Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности УК-1.3 Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт библиографического разыскания, создания научных текстов	Знать: <ul style="list-style-type: none">● теоретические ограничения для алгоритмов машинного обучения;● методы создания эффективных алгоритмов ИАД; Уметь: <ul style="list-style-type: none">● применять методы машинного обучения для интеллектуального анализа данных. Владеть: <ul style="list-style-type: none">● навыками применения алгоритмов машинного обучения для переработки больших объемов информации.● навыками встраивания алгоритмов машинного обучения в новые информационные технологии.
ОПК-2 Способен выявлять сущность проблем, возникающих в ходе профессиональной деятельности, привлекать соответствующий математический аппарат и информационные технологии для их решения	ОПК-2.1 Знает примеры решения разных классов задач, возникающих в профессиональной деятельности ОПК-2.2 Умеет использовать инструментальные средства для построения математических моделей, используемых для решения прикладных	

	<p>задач из разных классов</p> <p>ОПК-2.3 Имеет практический опыт участия в решении практических задач, предполагающий использование математического аппарата и информационных технологий</p>	
<p>ОПК-3 Способен использовать фундаментальные знания в области гуманитарных, социальных и лингвистических наук, а также в сфере техники и технологии информатики для совершенствования профессиональной деятельности</p>	<p>ОПК-3.1 Знает основы лингвистики, социологии и математической статистики</p> <p>ОПК-3.2 Умеет пользоваться инструментальными средами, библиотеками и фреймворками для анализа социологических данных и текстов</p> <p>ОПК-3.3 Имеет практический опыт работы с инструментальными средствами для анализа данных, в том числе анализа текстов</p>	
<p>ОПК-5 Способен применять новые информационные технологии в гуманитарных областях знаний с использованием средств интеллектуального анализа данных, машинного обучения, компьютерной лингвистики и представления знаний</p>	<p>ОПК-5.1 Знает теоретические основы машинного обучения, задачи, решаемые с помощью машинного обучения и интеллектуального анализа данных</p> <p>ОПК-5.2 Умеет применять методы интеллектуального анализа данных и машинного обучения для анализа данных в гуманитарных областях</p>	

	ОПК-5.3 Имеет практических опыт применения интеллектуального анализа данных, машинного обучения, компьютерной лингвистики и представления знаний к исследованиям в гуманитарных областях	
--	--	--

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Эффективные алгоритмы интеллектуального анализа данных» относится к обязательной части блока дисциплин учебного плана.

Для освоения дисциплины необходимы компетенции, сформированные в ходе изучения следующих дисциплин: «Случайные процессы», «Статистические методы машинного обучения», «Байесовские методы в статистике и машинном обучении».

2. Структура дисциплины¹

Общая трудоёмкость дисциплины составляет 9 з.е., 324 академических часа (ов).

Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
3	Практические занятия	30
4	Практические занятия	60
Всего:		90

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 216 академических часа(ов).

3. Содержание дисциплины²

В курсе изучаются эффективные алгоритмы машинного обучения и их применения в интеллектуальных системах. На практических занятиях студенты приобретают практические навыки обработки небольших массивов данных, навыки анализа алгоритмов машинного обучения встраивания алгоритмов машинного обучения в интеллектуальные системы.

В результате изучения курса студенты должны овладеть методами разработки эффективных алгоритмов машинного обучения, уметь использовать их при решении практических задач интеллектуального анализа данных.

¹ При реализации образовательной программы на очно-заочной и заочной формах обучения, таблица составляется для каждой формы.

² Раздел может быть представлен как в текстовой форме, так и в таблице

Курс должен сочетать современность и строгость изложения материала с его доступностью для слушателей. В основе курса лежит разбор большого числа примеров приложения методов и средств эффективных алгоритмов машинного обучения для решения прикладных задач.

№ п/п	Наименование раздела дисциплины	Содержание
1.	Эффективное представление данных	Операция сходства. Побитовое умножение. Библиотека boost::dynamic_bitset. Обучающая выборка в Анализе формальных понятий. Фундаментальная теорема АФП и ее следствия. Неразложимые элементы в конечных решетках. Алгоритм кодирования признаков с дискретными значениями. Непрерывные признаки. Элементы теории информации. Энтропийная дискретизация непрерывных признаков Выпуклое объединение интервалов.
2.	Обнаружение зависимостей между признаками	Задача регрессии. Формулы для линейной регрессии. Ридж-регрессия. Задача классификации. Байесовский классификатор. Логистическая регрессия. Метод Ньютона-Рафсона. Формулы для логистической регрессии. Критерии значимости регрессии.
3.	Переобучение	Переподгонка полиномиальной регрессии. Фантомные сходства. Модель переобучения. Проблема недообучения. Недостаточность запрета на контр-примеры. Производящие функции для переобучения
4.	ВКФ-метод машинного обучения	Монотонная, немонотонная и спаривающая цепь Маркова. Останавливаемость с вероятностью 1. Спаривающая цепь для случая Булевой алгебры. Ленивый алгоритм спаривающей цепи Маркова. Учет предварительных траекторий Процедуры индукции, абдукции и аналогии. Теорема о достаточном числе ВКФ-гипотез. Вычисления на видеокάρтах.
5.	Цепи Маркова для вероятностных алгоритмов	Конечные цепи Маркова. Неприводимость, апериодичность, эргодичность. Условия баланса. Обратимые цепи Маркова. Цепи Метрополиса-Хастингса. Цепи Гиббса. Метод спаривания для цепей Маркова Цепь Гиббса-Джеррума для раскрашиваний графов Вероятно-приближенное оценивание с помощью цепи Маркова.

4. Образовательные технологии³

³ В разделе указываются образовательные технологии, используемые при реализации различных видов учебных занятий для наиболее эффективного освоения дисциплины. При проведении учебных занятий обеспечивается развитие у обучающихся навыков командной работы, межличностной коммуникации, принятия решений, лидерских качеств (включая при необходимости проведение интерактивных лекций, групповых дискуссий, ролевых игр, тренингов, анализ ситуаций и имитационных моделей, преподавание дисциплин (*модулей*) в форме курсов, составленных на основе результатов научных исследований, в том числе с учётом региональных особенностей профессиональной деятельности выпускников и потребностей работодателей (п.34. Приказ №301).

Применительно к данной дисциплине, следует отметить, что она, по существу, является введением в прикладную математическую теорию, поэтому применение в ней информационных технологий носит вспомогательный характер. В ней используются следующие информационные технологии.

- Преподаватель представляет изучаемый материал посредством демонстрации презентаций.

- Электронная почта и другие средства Интернет используется студентами для общения с преподавателем.

- Сервер учебных материалов отделения используется студентами для создания словаря по курсу, самостоятельной работы и тестирования.

5. Оценка планируемых результатов обучения

5.1 Система оценивания⁴

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль: - контрольная работа (тема 1) - контрольная работа (тема 2)	30 баллов 30 баллов	30 баллов 30 баллов
Промежуточная аттестация зачет		40 баллов
Итого за семестр (дисциплину) зачет		100 баллов
Текущий контроль: - контрольная работа (темы 3-4) - контрольная работа (тема 5)	30 баллов 30 баллов	30 баллов 30 баллов
Промежуточная аттестация зачет с оценкой		40 баллов
Итого за семестр (дисциплину) экзамен		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

⁴ Система оценивания выстраивается в соответствии с учебным планом, где определены формы промежуточной аттестации (зачёт/зачёт с оценкой/экзамен), и структурой дисциплины, где определены формы текущего контроля. Указывается распределение баллов по формам текущего контроля и промежуточной аттестации, сроки отчётности.

5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ А,В	отлично/ зачтено	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	хорошо/ зачтено	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	удовлетво- рительно/ зачтено	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	неудовлет- ворительно/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине⁵

Контрольная работа 1

Реализовать кодирование дискретных признаков массива Mushroom.

Контрольная работа 2

Реализовать кодирование непрерывных признаков массива Wine Quality.

Контрольная работа 3

1. Вычислить ВКФ-гипотезы в массивах Mushroom и Wine Quality.
2. Найти зависимости между признаками в массиве Wine Quality

Контрольная работа 4

Написать алгоритм цепи Метрополиса-Хастингса по заданной матрице смежности графа и стационарному распределению.

1 Контрольные вопросы к зачету

1. Операция сходства. Побитовое умножение. Библиотека `boost::dynamic_bitset`.
2. Обучающая выборка в Анализе формальных понятий. Фундаментальная теорема АФП и ее следствия.
3. Неразложимые элементы в конечных решетках. Алгоритм кодирования признаков с дискретными значениями.
4. Непрерывные признаки. Элементы теории информации. Энтропийная дискретизация непрерывных признаков
5. Выпуклое объединение интервалов.
6. Задача регрессии. Формулы для линейной регрессии. Ридж-регрессия.
7. Задача классификации. Байесовский классификатор.
8. Логистическая регрессия. Метод Ньютона-Рафсона. Формулы для логистической регрессии.
9. Критерии значимости регрессии.

2 Контрольные вопросы к экзамену

10. Переподгонка полиномиальной регрессии.
11. Фантомные сходства. Модель переобучения. Проблема недообучения.
12. Недостаточность запрета на контр-примеры.
13. Производящие функции для переобучения.
14. Монотонная, немонотонная и спаривающая цепь Маркова. Останавливаемость с вероятностью 1.
15. Спаривающая цепь для случая Булевой алгебры.
16. Ленивый алгоритм спаривающей цепи Маркова.
17. Учет предварительных траекторий
18. Процедуры индукции, абдукции и аналогии
19. Теорема о достаточном числе ВКФ-гипотез

⁵ Приводятся примеры оценочных средств в соответствии со структурой дисциплины и системой контроля: варианты тестов, тематика письменных работ, примеры экзаменационных билетов, типовые задачи, кейсы и т.п. Оценочными средствами должны быть обеспечены все формы текущего контроля и промежуточной аттестации. Они должны быть ориентированы не только на проверку сформированности знаний, но также умений и владений.

20. Вычисления на видеокартах.
21. Конечные цепи Маркова. Неприводимость, апериодичность, эргодичность.
22. Условия баланса. Обратимые цепи Маркова.
23. Цепи Метрополиса-Хастингса. Цепи Гиббса.
24. Метод спаривания для цепей Маркова
25. Цепь Гиббса-Джеррума для раскрашивания графов
26. Вероятно-приближенное оценивание с помощью цепи Маркова

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы⁶

а) Основная литература

1. Виноградов Д.В. Вероятностно-комбинаторный формальный метод обучения, основанный на теории решеток. Диссертация на соискание ученой степени д.ф.-м.н. по специальности «Теоретические основы информатики» М.: ФИЦ ИУ РАН, 2019
www.frccsc.ru/diss-council/00207305/diss/list/vinogradov_dv

б) Дополнительная литература

1. Виноградов, Д.В. Быстрые алгоритмы машинного обучения, 2019
<https://vk.com/club176049295>
2. Виноградов, Д.В. Статистическое машинное обучение, 2019 <https://vk.com/club190684626>
3. Виноградов, Д.В. Байесовские методы в машинном обучении, 2019
<https://vk.com/club174045824>

6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

Национальная электронная библиотека (НЭБ) www.rusneb.ru
ELibrary.ru Научная электронная библиотека www.elibrary.ru
Электронная библиотека Grebennikon.ru www.grebennikon.ru
Cambridge University Press
ProQuest Dissertation & Theses Global
SAGE Journals
Taylor and Francis
JSTOR

7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения: компьютерный класс, компьютер преподавателя, компьютеры студентов, проектор, экран, доступ в интернет.

Состав программного обеспечения:

1. Windows
2. Microsoft Office

⁶ Рекомендуется включать в списки издания из ЭБС и не более 15 печатных изданий.

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением или могут быть заменены устным ответом; обеспечивается индивидуальное равномерное освещение не менее 300 люкс; для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств; письменные задания оформляются увеличенным шрифтом; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

- для глухих и слабослышащих: лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования; письменные задания выполняются на компьютере в письменной форме; экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих: в печатной форме увеличенным шрифтом, в форме электронного документа, в форме аудиофайла.

- для глухих и слабослышащих: в печатной форме, в форме электронного документа.

- для обучающихся с нарушениями опорно-двигательного аппарата: в печатной форме, в форме электронного документа, в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих: устройством для сканирования и чтения с камерой SARA CE; дисплеем Брайля PAC Mate 20; принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих: автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих; акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата: передвижными, регулируемые эргономическими партами СИ-1; компьютерной техникой со специальным программным обеспечением.

9. Методические материалы⁷

9.1 Планы семинарских/ практических/ лабораторных занятий⁸

Тема 1. (14 ч.) Эффективное представление данных

Цель занятий: изучить алгоритмы представления данных для эффективной обработки.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

- Как использовать побитовое умножение?
- Как устроен Анализ формальных понятий?
- Как кодировать дискретные признаки?
- Как работает энтропийная техника для дискретизации непрерывных признаков?

Контрольные вопросы:

1. Операция сходства. Побитовое умножение. Библиотека `boost::dynamic_bitset`.
2. Обучающая выборка в Анализе формальных понятий. Фундаментальная теорема АФП и ее следствия.
3. Неразложимые элементы в конечных решетках. Алгоритм кодирования признаков с дискретными значениями.
4. Непрерывные признаки. Элементы теории информации. Энтропийная дискретизация непрерывных признаков
5. Выпуклое объединение интервалов.

Список источников и литературы:

Основная литература:

2. Виноградов Д.В. Вероятностно-комбинаторный формальный метод обучения, основанный на теории решеток. Диссертация на соискание ученой степени д.ф.м.н. по специальности «Теоретические основы информатики» М.: ФИЦ ИУ РАН, 2019 www.frccsc.ru/diss-council/00207305/diss/list/vinogradov_dv

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

1. Виноградов, Д.В. Быстрые алгоритмы машинного обучения, 2019 <https://vk.com/club176049295>

⁷ Методические материалы по дисциплине могут входить в состав рабочей программы, либо разрабатываться отдельным документом.

⁸ План занятий строится в соответствии со структурой дисциплины (п.2). Разделы плана включают: название темы, количество часов, форму проведения занятия, его содержание (вопросы для обсуждения, задания, контрольные вопросы, кейсы и т.п.), список литературы. При необходимости, планы практических и лабораторных занятий могут содержать указания по выполнению заданий и требования к материально-техническому обеспечению занятия.

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

Тема 2. (16 ч.) Обнаружение зависимостей между признаками

Цель занятий: изучить применение регрессионной техники для обнаружения зависимостей между признаками.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

Что такое задача регрессии?

Почему для задачи классификации используют логистическую регрессию?

Как метод Ньютона-Рафсона дает явные формулы для логистической регрессии ?

Контрольные вопросы:

6. Задача регрессии. Формулы для линейной регрессии. Ридж-регрессия.
7. Задача классификации. Байесовский классификатор.
8. Логистическая регрессия. Метод Ньютона-Рафсона. Формулы для логистической регрессии.
9. Критерии значимости регрессии.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

1. Виноградов, Д.В. Байесовские методы в машинном обучении, 2019
<https://vk.com/club174045824>
2. Виноградов, Д.В. Быстрые алгоритмы машинного обучения, 2019
<https://vk.com/club176049295>

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

Тема 3. (20 ч.) Переобучение

Цель занятий: осознать проблему переобучения.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

Что такое переобучение?

Что такое фантомные сходства?

Почему запрет на контр-примеры недостаточен?

Контрольные вопросы:

10. Переподгонка полиномиальной регрессии.
11. Фантомные сходства. Модель переобучения. Проблема недообучения.
12. Недостаточность запрета на контр-примеры.
13. Производящие функции для переобучения

Список источников и литературы:

Основная литература:

3. Виноградов Д.В. Вероятностно-комбинаторный формальный метод обучения, основанный на теории решеток. Диссертация на соискание ученой степени д.ф.-м.н. по специальности «Теоретические основы информатики» М.: ФИЦ ИУ РАН,

2019 www.frccsc.ru/diss-council/00207305/diss/list/vinogradov_dy

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

2. Виноградов, Д.В. Быстрые алгоритмы машинного обучения, 2019
<https://vk.com/club176049295>

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

Тема 4. (20 ч.) ВКФ-метод машинного обучения

Цель занятий: усвоить методологию ВКФ-метода машинного обучения.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

- Как использовать вероятностные алгоритмы поиска сходств?
- Почему ленивая схема ускоряет вычисления?
- Как влияет ранняя остановка?
- Сколько ВКФ-гипотез породить для минимизации ошибки первого рода?
- Как вычислять ВКФ-сходства на видеокартах?

Контрольные вопросы:

14. Монотонная, немонотонная и спаривающая цепь Маркова. Останавливаемость с вероятностью 1.
15. Спаривающая цепь для случая Булевой алгебры.
16. Ленивый алгоритм спаривающей цепи Маркова.
17. Учет предварительных траекторий
18. Процедуры индукции, абдукции и аналогии.
19. Теорема о достаточном числе ВКФ-гипотез.
20. Вычисления на видеокартах.

Список источников и литературы:

Основная литература:

4. Виноградов Д.В. Вероятностно-комбинаторный формальный метод обучения, основанный на теории решеток. Диссертация на соискание ученой степени д.ф.-м.н. по специальности «Теоретические основы информатики» М.: ФИЦ ИУ РАН, 2019 www.frccsc.ru/diss-council/00207305/diss/list/vinogradov_dy

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

3. Виноградов, Д.В. Быстрые алгоритмы машинного обучения, 2019
<https://vk.com/club176049295>

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

Тема 5. (20 ч.) Цепи Маркова для вероятностных алгоритмов

Цель занятий: усвоить методологию применения цепей Маркова для создания вероятностных алгоритмов

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

- Как сделать алгоритм Метрополиса-Хастингса?
- Как сделать алгоритм типа Гиббса?

Что такое цепь Гиббса-Джеррума?

Как делать вероятно-приближенное оценивание с помощью цепи Маркова?

Контрольные вопросы:

21. Конечные цепи Маркова. Неприводимость, апериодичность, эргодичность.
22. Условия баланса. Обратимые цепи Маркова.
23. Цепи Метрополиса-Хастингса. Цепи Гиббса.
24. Метод спаривания для цепей Маркова
25. Цепь Гиббса-Джеррума для раскрашиваний графов
26. Вероятно-приближенное оценивание с помощью цепи Маркова.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

3. Виноградов, Д.В. Статистическое машинное обучение, 2019
<https://vk.com/club190684626>

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

9.2 Методические рекомендации по подготовке письменных работ⁹

Наименование раздела дисциплины	Кол-во часов	Вопросы для изучения	Литература
Эффективное представление данных	14	1. Операция схождения. Побитовое умножение. Библиотека boost::dynamic_bitset. 2. Обучающая выборка в Анализе формальных понятий. Фундаментальная теорема АФП и ее следствия. 3. Неразложимые элементы в конечных решетках. Алгоритм кодирования признаков с дискретными значениями. 4. Непрерывные признаки. Элементы теории информации. Энтропийная дискретизация непрерывных признаков 5. Выпуклое объединение интервалов.	Lecture1.pdf и Lecture2.pdf из https://vk.com/club176049295
Обнаружение зависимостей между признаками	16	1. Задача регрессии. Формулы для линейной регрессии. Ридж-регрессия. 2. Задача классификации. Байесовский классификатор.	Lecture2.pdf из https://vk.com/club176049295 Lecture7.pdf из https://vk.com/club1740458

⁹ В раздел включаются требования к подготовке, содержанию, и оформлению письменных работ предусмотренных учебным планом или рабочей программой (курсовая работа, эссе, реферат, доклад и т.п.). При наличии кафедральных или факультетских рекомендаций по подготовке письменных работ раздел включается в РПД по усмотрению преподавателя - составителя. Если письменная работа не предусмотрена, раздел не включается.

		<p>3. Логистическая регрессия. Метод Ньютона-Рафсона. Формулы для логистической регрессии.</p> <p>4. Критерии значимости регрессии.</p>	24
Переобучение	20	<p>1. Переподгонка полиномиальной регрессии.</p> <p>2. Фантомные сходства. Модель переобучения. Проблема недообучения.</p> <p>3. Недостаточность запрета на контр-примеры.</p> <p>4. Производящие функции для переобучения</p>	Lecture3.pdf и Lecture4.pdf из https://vk.com/club1760492 95
ВКФ-метод машинного обучения	20	<p>1. Монотонная, немонотонная и спаривающая цепь Маркова. Останавливаемость с вероятностью 1.</p> <p>2. Спаривающая цепь для случая Булевой алгебры.</p> <p>3. Ленивый алгоритм спаривающей цепи Маркова.</p> <p>4. Учет предварительных траекторий</p> <p>5. Процедуры индукции, абдукции и аналогии.</p> <p>6. Теорема о достаточном числе ВКФ-гипотез.</p> <p>7. Вычисления на видеокартах.</p>	Lecture5.pdf, Lecture6.pdf, Lecture7.pdf, Lecture8.pdf и Lecture9.pdf из https://vk.com/club1760492 95
Цепи Маркова для вероятностных алгоритмов	20	<p>1. Конечные цепи Маркова. Неприводимость, апериодичность, эргодичность.</p> <p>2. Условия баланса. Обратимые цепи Маркова.</p> <p>3. Цепи Метрополиса-Хастингса. Цепи Гиббса.</p> <p>4. Метод спаривания для цепей Маркова</p> <p>5. Цепь Гиббса-Джеррума для раскрашивания графов</p> <p>6. Вероятно-приближенное оценивание с помощью цепи Маркова.</p>	Lecture7.pdf, Lecture8.pdf, Lecture9.pdf и Lecture10.pdf из https://vk.com/club1906846 26

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина «Эффективные алгоритмы интеллектуального анализа данных» реализуется на Отделении интеллектуальных систем в гуманитарной сфере кафедрой математики, логики и интеллектуальных систем.

В процессе обучения эффективным алгоритмам интеллектуального анализа данных преследуются несколько целей. Одна из них – подготовить выпускника, способного внедрить алгоритмы ИАД в интеллектуальных системах. Другой целью курса можно считать научение слушателей применять эффективные алгоритмы интеллектуального анализа данных к реальным данным.

Задачи дисциплины: освоение способов разработки эффективных алгоритмов ИАД и навыков, необходимых для получения требуемых компетенций в области создания интеллектуальных систем с машинным обучением.

Дисциплина направлена на формирование следующих компетенций:

- *УК-1* Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
- *ОПК-2* Способен выявлять сущность проблем, возникающих в ходе профессиональной деятельности, привлекать соответствующий математический аппарат и информационные технологии для их решения
- *ОПК-3* Способен использовать фундаментальные знания в области гуманитарных, социальных и лингвистических наук, а также в сфере техники и технологии информатики для совершенствования профессиональной деятельности
- *ОПК-5* Способен применять новые информационные технологии в гуманитарных областях знаний с использованием средств интеллектуального анализа данных, машинного обучения, компьютерной лингвистики и представления знаний

В результате освоения дисциплины обучающийся должен:

Знать:

- методы создания эффективных алгоритмов ИАД;
- теоретические ограничения для алгоритмов машинного обучения;

Уметь: применять методы машинного обучения для интеллектуального анализа данных

Владеть:

- навыками применения алгоритмов машинного обучения для переработки больших объемов информации;
- навыками встраивания алгоритмов машинного обучения в новые информационные технологии.

По дисциплине предусмотрена промежуточная аттестация в форме **зачета** и **экзамена**.

Общая трудоемкость освоения дисциплины составляет 9 зачетных единиц.