

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Российский государственный гуманитарный университет»
(ФГБОУ ВО «РГГУ»)

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

СТАТИСТИЧЕСКИЕ МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

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 2024

Статистические методы машинного обучения
Рабочая программа дисциплины (модуля)

Составитель(и):

Кандидат технических наук, доцент Л.О. Шашкин

.....

УТВЕРЖДЕНО

Протокол заседания кафедры МЛИИС

№6 от 08.02.2024

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	8
2. Структура дисциплины	8
3. Содержание дисциплины	9
4. Образовательные технологии	10
5. Оценка планируемых результатов обучения	11
5.1 Система оценивания	11
5.2 Критерии выставления оценки по дисциплине	11
5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине	12
6. Учебно-методическое и информационное обеспечение дисциплины	13
6.1 Список источников и литературы	13
6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».	13
7. Материально-техническое обеспечение дисциплины	14
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов	14
9. Методические материалы	15
9.1 Планы семинарских/ практических/ лабораторных занятий	15
9.2 Методические рекомендации по подготовке письменных работ	19
Приложение 1. Аннотация рабочей программы дисциплины	21

1. Пояснительная записка

1.1 Цель и задачи дисциплины

Цель дисциплины. В процессе обучения машинному обучению преследуются несколько целей. Одна из них – подготовить выпускника, способного внедрить алгоритмы машинного обучения в интеллектуальных системах. Другой целью курса можно считать научение слушателей применять техники статистического машинного обучения для анализа (в том числе, невозможности) алгоритмов интеллектуального анализа данных.

Задачи дисциплины: освоение базовых математических понятий машинного обучения 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
<i>УК-1</i> Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	<i>УК-1.1</i> Знает принципы сбора, отбора и обобщения информации <i>УК-1.2</i> Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности <i>УК-1.3</i> Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт библиографического разыскания, создания научных текстов	Знать: ● теоретические ограничения для алгоритмов машинного обучения; Владеть: ● простейшими навыками применения алгоритмов машинного обучения для переработки больших объемов информации.
<i>ПК-1</i> Способен применять новые информационные технологии в гуманитарных областях знаний с использованием средств интеллектуального анализа данных и машинного обучения,	<i>ПК-1.1</i> Знает области возможного применения новых информационных технологий в гуманитарных областях знаний, включая использование средств интеллектуального анализа данных и машинного обучения, компьютерной	Знать: ● подход к интеллектуальному анализу данных на основе машинного обучения. Уметь: ● применять методы машинного обучения для интеллектуального анализа данных.

<p>компьютерной лингвистики и представления знаний</p>	<p>лингвистики и представления знаний. Знает примеры успешного применения информационных технологий в гуманитарных областях</p> <p><i>ПК-1.2</i> Умеет использовать различные инструментальные средства, платформы для разработки приложений, и прикладные программы в гуманитарных областях</p> <p><i>ПК-1.3</i> Имеет практический опыт использования различных инструментальных средств, платформ для разработки приложений и прикладных программ (включая средства интеллектуального анализа данных, машинного обучения, компьютерной лингвистики и представления знаний) в гуманитарных областях</p>	
<p><i>ПК-3</i> Способен разрабатывать и тестировать новые программы и интерфейсы систем</p>	<p><i>ПК-3.1</i> Знает технологии разработки и тестирования программ, языки программирования и стандарты на представления результатов анализа и проектирования</p> <p><i>ПК-3.2</i> Умеет использовать интегрированные среды разработки, включая средства визуального</p>	<p>Знать:</p> <ul style="list-style-type: none"> ● основные понятия интеллектуального анализа данных и машинного обучения; <p>Уметь:</p> <ul style="list-style-type: none"> ● применять методы машинного обучения для интеллектуального анализа данных <p>Владеть:</p> <ul style="list-style-type: none"> ● простейшими навыками встраивания алгоритмов машинного обучения в новые информационные технологии.

	<p>программирования, умеет использовать средства компьютерной поддержки этапов анализа и проектирования</p> <p><i>ПК-3.3</i> Имеет практический опыт разработки и тестирования прикладных программ</p>	
<p><i>ПК-5</i> Способен использовать технические, программные средства и языки программирования для разработки алгоритмов и программ в области интеллектуального анализа данных, интеллектуальных и информационных систем</p>	<p><i>ПК-5.1</i> Знает синтаксис, семантику, возможности и ограничения языков программирования, применяемых для разработки программных средств интеллектуального анализа данных, интеллектуальных и информационных систем</p> <p><i>ПК-5.2</i> Умеет применять современные интегрированные среды разработки для создания систем интеллектуального анализа данных и интеллектуальных информационных систем</p> <p><i>ПК-5.3</i> Имеет практический опыт участия в разработке систем интеллектуального анализа данных, интеллектуальных и информационных систем</p>	<p>Знать:</p> <ul style="list-style-type: none"> ● основные понятия интеллектуального анализа данных и машинного обучения; <p>Уметь:</p> <ul style="list-style-type: none"> ● применять методы машинного обучения для интеллектуального анализа данных <p>Владеть:</p> <ul style="list-style-type: none"> ● простейшими навыками встраивания алгоритмов машинного обучения в новые информационные технологии.

<i>ПК-5</i> Способен использовать	<i>ПК-5.1</i> Знает синтаксис, семантику, возможности	Знать: основные понятия теории нечетких
-----------------------------------	---	---

<p>технические, программные средства и языки программирования для разработки алгоритмов и программ в области интеллектуального анализа данных, интеллектуальных и информационных систем</p>	<p>и ограничения языков программирования, применяемых для разработки программных средств интеллектуального анализа данных, интеллектуальных и информационных систем</p>	<p>множеств, операции над нечеткими множествами; нечеткие отношения и их свойства; основные понятия теории генетических алгоритмов; методы построения формальных моделей для применения эволюционных алгоритмов; методы повышения эффективности эволюционных алгоритмов; основные понятия теории нейронных сетей; основные типы нейронных сетей и методы их обучения.</p> <p>Уметь: разрабатывать и тестировать алгоритмы, моделирующие эволюцию; создавать программы, реализующие нейронные сети различных типов.</p> <p>Владеть навыками: навыками решения оптимизационных задач с помощью программ, использующих эволюционные модели; навыками обучения и применения нейронных сетей.</p>
	<p><i>ПК-5.2</i> Умеет применять современные интегрированные среды разработки для создания систем интеллектуального анализа данных и интеллектуальных информационных систем</p>	
	<p><i>ПК-5.3</i> Имеет практический опыт участия в разработке систем интеллектуального анализа данных, интеллектуальных и информационных систем</p>	

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Статистическая теория машинного обучения» относится к части, формируемой участниками образовательных отношений блока дисциплин учебного плана.

Для освоения дисциплины необходимы компетенции, сформированные (в рамках бакалавриата) в ходе изучения следующих дисциплин: «Дискретная математика», «Теория алгоритмов», «Теория вероятностей и математическая статистика».

2. Структура дисциплины¹

Общая трудоёмкость дисциплины составляет 3 з.е., 108 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:

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

Семестр	Тип учебных занятий	Количество часов
1	Лекции	12
	Лабораторные работы	18
	Всего:	30

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 60 академических часа(ов).

3. Содержание дисциплины²

В курсе изучаются понятия статистического машинного обучения и его применения в интеллектуальных системах. На практических занятиях студенты приобретают практические навыки обработки небольших массивов данных, навыки анализа алгоритмов машинного обучения и вычисления размерности Вапника-Червоненкиса как средства оценки качества обучения.

В результате изучения курса студенты должны овладеть основными идеями и методами статистического машинного обучения, моделями (вероятно-приближенно корректного и с ограничением на число ошибок) машинного обучения, уметь использовать их при решении практических задач интеллектуального анализа данных.

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 статистического машинного для решения прикладных задач.

№ п/п	Наименование раздела дисциплины	Содержание
1.	Обучение с ограничением на число ошибок	Методы интеллектуального анализа данных. Типы машинного обучения. Обучение с ограничением на число ошибок. Алгоритм пересечения для обучения конъюнктивным понятиям. Пространство версий и алгоритм голосования. Алгоритм случайного выбора.
2.	Снижение размерности и персептрон	Снижение размерности. Алгоритм Winnow1. Персептрон. Алгоритм обучения персептрону. Зазор в обучающей выборке. Теорема о числе ошибок.
3.	Машина опорных векторов	Сколы и расщепленные множества. Размерность Вапника-Червоненкиса. Связь VC-размерности с числом ошибок. Лемма Радона и VC-размерность персептронов. Отступ центральной функции. Опорные вектора. Машина опорных векторов для персептрона. Двойственная задача. Теорема Каруша-Куна-Таккера и опорные вектора. Скользящий контроль.
4.	Вероятно-приближенно корректное обучение	Вероятно-приближенно корректное обучение. Эффективная обучаемость. Алгоритм пересечения для

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

		обучения конъюнктивным понятиям. Бритва Оккама. Выборки большого объема для конечного класса гипотез. Списки решений. Алгоритм Ривеста обучения спискам решений.
5.	Случай невозможности эффективного ВПК-обучения	Элементы теории сложности вычислений (напоминание). NP-трудность задачи 3-раскрашивания графа. Вероятностно-полиномиальные языки. Преобразование графа в обучающую выборку. 3-членные ДНФ и 3-раскрашивания. Невозможность эффективного ВПК-обучения 3-членным ДНФ. Расширение пространства версий и эффективная ВПК-обучаемость после этого.
6.	Размерность Вапника-Червоненкиса и ВПК-обучаемость	Размерность Вапника-Червоненкиса (напоминание). Функция расщепления и ее мажоранта. Рост мажоранты. Выборки большого объема для класса с конечной VC-размерностью. Неравенство Маркова и метод Бернштейна (напоминание). Оценки больших уклонений. Лемма Вапника-Червоненкиса. Нижняя граница на объем обучающей выборки.

4. Образовательные технологии³

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.

В период временного приостановления посещения обучающимися помещений и территории РГГУ для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

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).

5. Оценка планируемых результатов обучения

5.1 Система оценивания⁴

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль: - контрольная работа (темы 1-3) - контрольная работа (темы 4-6)	30 баллов 30 баллов	30 баллов 30 баллов
Промежуточная аттестация экзамен		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	A	
83 – 94		B	
68 – 82	хорошо	зачтено	
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	FX	
0 – 19		не зачтено	F

5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	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>

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.

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
82-68/ С	хорошо/ зачтено	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. Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами. Достаточно хорошо ориентируется в учебной и профессиональной литературе. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».
67-50/ D,E	удовлетворительно/ зачтено	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. 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. Демонстрирует достаточный уровень знания учебной литературы по дисциплине. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».
49-0/ F,FX	неудовлетворительно/ не зачтено	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. 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. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

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

Контрольная работа 1

- Для дизъюнкции $x_1 \vee x_5$ выпишите шаги продвижения и устранения в алгоритме Winnow1 (переменные x_1, \dots, x_5).
- Найдите опорные вектора и проведите разделяющую прямую для обучающей выборки $\{(0,0), -1\}, \{-3,3\}, +1\}, \{(3,5), +1\}, \{(0,5), +1\}, \{(1,-2), -1\}, \{-2,0\}, +1\}$.

Контрольная работа 2

- Вычислите размерность Вапника-Червоненкиса для кругов с рациональным центром и рациональным радиусом на плоскости.

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.

2. Для списка решений $\langle \sim x_3 \ \& \ \sim x_4, 0 \rangle$, $\langle x_1 \ \& \ \sim x_2, 0 \rangle$, $\langle \text{true}, 1 \rangle$ оракул выдал примеры с номерами 13, 16, 5, 3, 12, 5, 7, 14, 15, 4, 2, 11. Вычислите список-гипотезу и расхождение (распределение равномерное).

Список теоретических вопросов, выносимых на промежуточную аттестацию

1. Методы интеллектуального анализа данных. Типы машинного обучения. Обучение с ограничением на число ошибок. Алгоритм пересечения для обучения конъюнктивным понятиям.
2. Пространство версий и алгоритм голосования.
3. Алгоритм случайного выбора.
4. Снижение размерности. Алгоритм Winnow1.
5. Персептрон. Алгоритм обучения персептрону. Зазор в обучающей выборке. Теорема о числе ошибок.
6. Сколы и расщепленные множества. Размерность Вапника-Червоненкиса. Связь ВЧ-размерности с числом ошибок.
7. Лемма Радона и ВЧ-размерность персептронов.
8. Отступ центральной функции. Опорные вектора.
9. Машина опорных векторов для персептрона. Двойственная задача. Теорема Каруша-Куна-Таккера и опорные вектора.
10. Скользящий контроль для машины опорных векторов.
11. Вероятно-приближенно корректное обучение. Эффективная обучаемость. Алгоритм пересечения для обучения конъюнктивным понятиям.
12. Бритва Оккама. Выборки большого объема для конечного класса гипотез.
13. Списки решений. Алгоритм Ривеста обучения спискам решений.
14. NP-трудность задачи 3-раскрашивания графа.
15. Вероятностно-полиномиальные языки.
16. Преобразование графа в обучающую выборку. 3-членные ДНФ и 3-раскрашивания.
17. Невозможность эффективного ВПК-обучения 3-членным ДНФ.
18. Расширение пространства версий и эффективная ВПК-обучаемость после этого.
19. Функция расщепления и ее мажоранта. Рост мажоранты.
20. Выборки большого объема для класса с конечной ВЧ-размерностью.
21. Лемма Вапника-Червоненкиса (без доказательства).
22. Нижняя граница на объем обучающей выборки.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы⁶

а) Основная литература

1. Вьюгин, В. В. Математические основы машинного обучения и прогнозирования : учебное пособие / В. В. Вьюгин. — Москва : МЦНМО, 2014. — 304 с. — ISBN 978-5-4439-2014-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/56397>.

б) Дополнительная литература

1. Виноградов, Д.В. Статистическое машинное обучение, 2020 <https://vk.com/club190684626>

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

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

учётom 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. (2 ч.) Обучение с ограничением на число ошибок

Цель занятий: усвоить методологию обучения с ограничением на число ошибок.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

Какие бывают типы машинного обучения?

Как устроено обучение с ограничением на число ошибок?

Что такое пространство версий?

В чем проблемы алгоритма голосования?

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.

Контрольные вопросы:

1. Методы интеллектуального анализа данных.
2. Типы машинного обучения.
3. Обучение с ограничением на число ошибок.
4. Алгоритм пересечения для обучения конъюнктивным понятиям.
5. Пространство версий и алгоритм голосования.
6. Алгоритм случайного выбора.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

1. Виноградов, Д.В. Статистическое машинное обучение, 2020
<https://vk.com/club190684626>

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

Тема 2. (4 ч.) Снижение размерности и персептрон

Цель занятий: изучить идеологию снижения размерности и алгоритм обучения персептрону.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

- Чем полезно снижение размерности в данных?
- Как достигается снижение размерности в бинарных данных?
- Что такое персептрон?
- Каково максимальное число ошибок при обучении персептрону?

Контрольные вопросы:

7. Снижение размерности.
8. Алгоритм Winnow1.
9. Персептрон.
10. Алгоритм обучения персептрону.
11. Зазор в обучающей выборке.
12. Теорема о числе ошибок.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

2. Виноградов, Д.В. Статистическое машинное обучение, 2020
<https://vk.com/club190684626>

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

Тема 3. (2 ч.) Машина опорных векторов

Цель занятий: усвоить алгоритм машины опорных векторов.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

- Как нерасщепленность ограничивает число ошибок ?
- Какова размерность Вапника-Червоненкиса для персептронов?

Что такое опорные вектора?

Как провести разделяющую гиперплоскость по опорным векторам?

Контрольные вопросы:

13. Сколы и расщепленные множества. Размерность Вапника-Червоненкиса.
14. Связь ВЧ-размерности с числом ошибок.
15. Лемма Радона и ВЧ-размерность перцептронов.
16. Отступ центральной функции. Опорные вектора.
17. Машина опорных векторов для перцептрона.
18. Двойственная задача. Теорема Каруша-Куна-Таккера и опорные вектора.
19. Скользящий контроль для машины опорных векторов.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

3. Виноградов, Д.В. Статистическое машинное обучение, 2020
<https://vk.com/club190684626>

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

Тема 4. (4 ч.) Вероятно-приближенно корректное обучение

Цель занятий: усвоить методологию ВПК-обучения.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

- Как обучаться из вероятностно порождаемых данных?
- Как устроено ВПК-обучение?
- Что такое бритва Оккама для ВПК-обучения?
- Как обучаться спискам решений?

Контрольные вопросы:

20. Вероятно-приближенно корректное обучение.
21. Эффективная обучаемость.
22. Алгоритм пересечения для обучения конъюнктивным понятиям.
23. Бритва Оккама.
24. Выборки большого объема для конечного класса гипотез.
25. Списки решений.
26. Алгоритм Ривеста обучения спискам решений.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

4. Виноградов, Д.В. Статистическое машинное обучение, 2020
<https://vk.com/club190684626>

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

Тема 5. (2 ч.) Случай невозможности эффективного ВПК-обучения

Цель занятий: понять причины невозможности обучения.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

Почему алгоритма обучения может не существовать?

Вычислительная сложность: чем она может помешать?

Что такое вероятностно-полиномиальный язык?

Контрольные вопросы:

27. NP-трудность задачи 3-раскрашивания графа.
28. Вероятностно-полиномиальные языки.
29. Преобразование графа в обучающую выборку.
30. 3-членные ДНФ и 3-раскрашивания.
31. Невозможность эффективного ВПК-обучения 3-членным ДНФ.
32. Расширение пространства версий и эффективная ВПК-обучаемость после этого.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

5. Виноградов, Д.В. Статистическое машинное обучение, 2020
<https://vk.com/club190684626>

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

Тема 6. (4 ч.) Размерность Вапника-Червоненкиса и ВПК-обучаемость

Цель занятий: усвоить связь между конечной ВЧ-размерностью и ВПК-обучаемостью.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

- Как растет функция расщепления?
- Как получаются оценки больших уклонений?
- Почему классы с конечной ВЧ-размерностью ВПК-обучаемы?
- Почему классы с бесконечной ВЧ-размерностью не являются ВПК-обучаемыми?

Контрольные вопросы:

33. Размерность Вапника-Червоненкиса (напоминание).
34. Функция расщепления и ее мажоранта.
35. Рост мажоранты.
36. Выборки большого объема для класса с конечной ВЧ-размерностью.
37. Неравенство Маркова и метод Бернштейна (напоминание).
38. Оценки больших уклонений.
39. Лемма Вапника-Червоненкиса.
40. Нижняя граница на объем обучающей выборки.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

6. Виноградов, Д.В. Статистическое машинное обучение, 2020
<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
Обучение с ограничением на число ошибок	2	Методы интеллектуального анализа данных. Типы машинного обучения. Обучение с ограничением на число ошибок. Алгоритм пересечения для обучения конъюнктивным понятиям. Пространство версий и алгоритм голосования. Алгоритм случайного выбора.	Lecture1.pdf из https://vk.com/club190684626
Снижение размерности и перцептрон	2	Снижение размерности. Алгоритм Winnow1. Перцептрон. Алгоритм обучения перцептрону. Зазор в обучающей выборке. Теорема о числе ошибок.	Lecture2.pdf из https://vk.com/club190684626
Машина опорных векторов	2	Сколы и расщепленные множества. Размерность Вапника-Червоненкиса. Связь ВЧ-размерности с числом ошибок. Лемма Радона и ВЧ-размерность перцептронов. Отступ центральной функции. Опорные вектора. Машина опорных векторов для перцептрона. Двойственная задача. Теорема Каруша-Куна-Таккера и опорные вектора. Скользящий контроль.	Lecture3.pdf из https://vk.com/club190684626
Вероятно-приближенно корректное обучение	2	Вероятно-приближенно корректное обучение. Эффективная обучаемость. Алгоритм пересечения для обучения конъюнктивным понятиям. Бритва Оккама. Выборки большого объема для конечного класса гипотез. Списки решений. Алгоритм Ривеста обучения спискам решений.	Lecture4.pdf из https://vk.com/club190684626
Случай невозможности эффективного ВПК-обучения	2	Элементы теории сложности вычислений (напоминание). NP-трудность задачи 3-раскрашивания графа. Вероятностно-полиномиальные	Lecture5.pdf из https://vk.com/club190684626

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>языки. Преобразование графа в обучающую выборку. 3-членные ДНФ и 3-раскрашивания. Невозможность эффективного ВПК-обучения 3-членным ДНФ. Расширение пространства версий и эффективная ВПК-обучаемость после этого.</p>	
<p>Размерность Вапника-Червоненкиса и ВПК-обучаемость</p>	2	<p>Размерность Вапника-Червоненкиса (напоминание). Функция расщепления и ее мажоранта. Рост мажоранты. Выборки большого объема для класса с конечной ВЧ-размерностью. Неравенство Маркова и метод Бернштейна (напоминание). Оценки больших уклонений. Лемма Вапника-Червоненкиса. Нижняя граница на объем обучающей выборки.</p>	<p>Lecture6.pdf из https://vk.com/club190684626</p>

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина «Статистические методы машинного обучения» реализуется на Отделении интеллектуальных систем в гуманитарной сфере кафедрой математики, логики и интеллектуальных систем.

В процессе обучения машинному обучению преследуются несколько целей. Одна из них – подготовить выпускника, способного внедрить алгоритмы машинного обучения в интеллектуальных системах. Другой целью курса можно считать научение слушателей применять техники статистического машинного обучения для анализа (в том числе, невозможности) алгоритмов интеллектуального анализа данных.

Задачи дисциплины: освоение базовых математических понятий машинного обучения 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
- ПК-1 Способен применять новые информационные технологии в гуманитарных областях знаний с использованием средств интеллектуального анализа данных и машинного обучения, компьютерной лингвистики и представления знаний
- ПК-3 Способен разрабатывать и тестировать новые программы и интерфейсы систем
- ПК-5 Способен использовать технические, программные средства и языки программирования для разработки алгоритмов и программ в области интеллектуального анализа данных, интеллектуальных и информационных систем.

В результате освоения дисциплины обучающийся должен:

Знать:

- подход к интеллектуальному анализу данных на основе машинного обучения;
- основные понятия интеллектуального анализа данных и машинного обучения;
- теоретические ограничения для алгоритмов машинного обучения

Уметь: применять методы машинного обучения для интеллектуального анализа данных

Владеть:

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

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

Общая трудоемкость освоения дисциплины составляет 3 зачетные единицы.