

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 квалификации выпускника: магистр

Форма обучения очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2019

«Статистические методы машинного обучения»
Рабочая программа дисциплины
Составитель:
доктор физико-математических наук, профессор
Д.В. Виноградов

УТВЕРЖДЕНО
Протокол заседания кафедры МЛиИС
№__7__ от__10.06.2019

ОГЛАВЛЕНИЕ

1. Пояснительная записка

1.1 Цель и задачи дисциплины (*модуля*)

1.2. Формируемые компетенции, соотнесённые с планируемыми результатами обучения по дисциплине

1.3. Место дисциплины в структуре образовательной программы

2. Структура дисциплины

3. Содержание дисциплины

4. Образовательные технологии

5. Оценка планируемых результатов обучения

5.1. Система оценивания

5.2. Критерии выставления оценок

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

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список источников и литературы

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

7. Материально-техническое обеспечение дисциплины

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

9. Методические материалы

9.1. Планы семинарских занятий

9.2. Методические указания для обучающихся по освоению дисциплины

9.3. Иные материалы

Приложения

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

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. Формируемые компетенции, соотнесённые с планируемыми результатами обучения по дисциплине:

Коды компетенций	Содержание компетенций	Перечень планируемых результатов обучения по дисциплине
ОК-2	способностью выстраивать и реализовывать перспективные линии интеллектуального, культурного, нравственного, физического и профессионального саморазвития и самосовершенствования	Знать: <ul style="list-style-type: none">● подход к интеллектуальному анализу данных на основе машинного обучения. Уметь: <ul style="list-style-type: none">● применять методы машинного обучения для интеллектуального анализа данных.
ОПК-7	готовностью перерабатывать большие объёмы информации и вычленять главное (анализ информации)	Знать: <ul style="list-style-type: none">● теоретические ограничения для алгоритмов машинного обучения; Владеть: <ul style="list-style-type: none">● простейшими навыками применения алгоритмов машинного обучения для переработки больших объёмов информации.
ПК-29	способностью применять новые информационные технологии в гуманитарных областях знаний с использованием средств интеллектуального анализа данных и машинного обучения, компьютерной лингвистики и представления знаний	Знать: <ul style="list-style-type: none">● основные понятия интеллектуального анализа данных и машинного обучения; Уметь: <ul style="list-style-type: none">● применять методы машинного обучения для интеллектуального анализа данных Владеть: <ul style="list-style-type: none">● простейшими навыками встраивания алгоритмов машинного обучения в новые информационные технологии.

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Статистическая теория машинного обучения» относится к вариативной части дисциплин по выбору 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 дисциплины

Структура дисциплины для очной формы обучения

Общая трудоёмкость дисциплины составляет 3 з.е., 114 ч., в том числе контактная работа обучающихся с преподавателем 30 ч., промежуточная аттестация 18 ч., самостоятельная работа обучающихся 84 ч.

№ п/п	Раздел дисциплины/темы	Се ме ст р	Виды учебной работы (в часах)						Формы текущего контроля успеваемости, форма промежуточной аттестации (по семестрам)
			Контактная				Пр оме жу точ ная атт ест аци я	С а м ос то ят ел ь- на я ра б от а	
			Лек ции	Се ми нар	Прак тичес кие занят ия	Лабора торные занят ия			
1	Обучение с ограничением на число ошибок	1	2			2		14	Оценка выполнения практических заданий
2	Снижение размерности и персептрон	1	2			4		14	Оценка выполнения практических заданий
3	Машина опорных векторов	1	2			2		14	Контрольная работа
4	Вероятно-приближенно корректное обучение	1	2			4		14	Оценка выполнения практических заданий
5	Случай невозможности эффективного ВПК-обучения	1	2			2		14	Оценка выполнения практических заданий
6	Размерность Вапника-Червонен	1	2			4		14	Контрольная работа

	киса и ВПК-обучаемость								
	Зачет с оценкой	1					18		Итоговая контрольная работа
	итого:		12			18	18	84	

3. Содержание дисциплины

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 ВЧ-размерности с числом ошибок. Лемма Радона и ВЧ-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 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 ВЧ-размерностью. Неравенство Маркова и метод Бернштейна (напоминание). Оценки больших уклонений. Лемма Вапника-Червоненкиса. Нижняя граница на объем обучающей выборки.

4. Образовательные технологии

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 оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- контрольная работа (темы 1-3)	30 баллов	30 баллов
- контрольная работа (темы 4-6)	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	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

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 (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/ C	«хорошо»/ «зачтено (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	«удовлетвори- тельно»/ «зачтено (удовлетвори- тельно)»/ «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 выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»/ 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 выставляются 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 дисциплине

ФОС для проведения промежуточной аттестации по дисциплине состоит из 4 разделов:

1 раздел. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы;

2 раздел. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания;

3 раздел. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы;

4 раздел. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Для каждого результата обучения по дисциплине (модулю) или практике определяются показатели и критерии оценивания сформированности компетенций на различных этапах их формирования, шкалы и процедуры оценивания.

Оценочные средства, соответствующие конкретным этапам формирования компетенций

Код компетенции	Описание этапов формирования компетенции	Наименование оценочных средств
ОК-2, ОПК-7, ПК-29	Знать: - подход к интеллектуальному анализу данных на основе машинного обучения; - основные понятия интеллектуального анализа данных и машинного обучения; - теоретические ограничения для алгоритмов машинного обучения	Выполнение практических заданий Зачет с оценкой
	Уметь: применять методы машинного обучения для интеллектуального анализа данных	Выполнение практических заданий Контрольная работа 1 Контрольная работа 2 Зачет с оценкой
	Владеть: - простейшими навыками применения алгоритмов машинного обучения для переработки больших объемов информации - простейшими навыками встраивания алгоритмов машинного обучения в новые информационные технологии	Выполнение практических заданий Контрольная работа 1 Контрольная работа 2 Зачет с оценкой

Контрольная работа 1

1. Для дизъюнкции $x_1 \vee x_5$ выпишите шаги продвижения и устранения в алгоритме Winnow1 (переменные x_1, \dots, x_5).
2. 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

1. Вычислите размерность Вапника-Червоненкиса для кругов с рациональным центром и рациональным радиусом на плоскости.
2. Для списка решений $<\sim x_3 \ \& \ \sim x_4, 0>, <x_1 \ \& \ \sim x_2, 0>, <\text{true}, 1>$ оракул выдал примеры с номерами 13, 16, 5, 3, 12, 5, 7, 14, 15, 4, 2, 11. Вычислите список-гипотезу и расхождение (распределение равномерное).

Контрольные вопросы к зачету

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 литературы

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 БД и ИСС

№п /п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2019 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2019 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global

	<p>SAGE Journals</p> <p>Журналы Taylor and Francis</p>
3	<p>Профессиональные полнотекстовые БД</p> <p>JSTOR</p> <p>Издания по общественным и гуманитарным наукам</p> <p>Электронная библиотека Grebennikon.ru</p>
4	<p>Компьютерные справочные правовые системы</p> <p>Консультант Плюс,</p> <p>Гарант</p>

7. Материально-техническое обеспечение дисциплины

Занятия проводятся в компьютерном классе ауд. 706, расположенном по адресу 125993, Москва, Миусская пл., д. 6, стр.2,.

Этот компьютерный класс оснащен

- достаточным количеством объединенных в локальную сеть рабочих станций,
- медиапроектором и экраном,
- меловой доской.

В классе имеются возможности

- подключения ноутбука к медиапроектору,
- одновременного доступа в Интернет для преподавателя и студентов.

1. Перечень ПО

№п /п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)
1	Adobe Master Collection CS4	Adobe	лицензионное
2	ОС «Альт Образование» 8	ООО «Базальт СПО	лицензионное
3	Windows 10 Pro	Microsoft	лицензионное
4	Kaspersky Endpoint Security	Kaspersky	лицензионное
5	Microsoft Office 2016	Microsoft	лицензионное

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;
 - зачёт проводится в устной форме или выполняется 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;

- зачёт проводится 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;
 - зачёт проводится в устной форме или выполняется 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 материалы

9.1. Планы лабораторных занятий

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?

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 персептронов.

16. Отступ центральной функции. Опорные вектора.

17. Машина опорных векторов для персептрона.

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 вероятно-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 указания для обучающихся по освоению дисциплины

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 персептрон	2	Снижение размерности. Алгоритм Winnow1. Персептрон. Алгоритм обучения персептрону. 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 персептронов. Отступ центральной функции. Опорные вектора. Машина опорных векторов для персептрона. 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 языки. Преобразование графа в обучающую выборку. 3-членные ДНФ и 3-раскрашивания. Невозможность эффективного	Lecture5.pdf из https://vk.com/club190684626

		ВПК-обучения 3-членным ДНФ. Расширение пространства версий и эффективная ВПК-обучаемость после этого.	
Размерность Вапника-Червонен киса и ВПК-обучаемость	2	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.	Lecture6.pdf из https://vk.com/club190684626

9.3 Иные материалы

Приложение 1
АННОТАЦИЯ 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:

- ОК-2 способностью выстраивать и реализовывать перспективные линии интеллектуального, культурного, нравственного, физического и профессионального саморазвития и самосовершенствования;
- ОПК-7 готовностью перерабатывать большие объемы информации и вычленять главное (анализ информации);
- ПК-29 способностью применять новые информационные технологии в гуманитарных областях знаний с использованием средств интеллектуального анализа данных и машинного обучения, компьютерной лингвистики и представления знаний.

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 **зачета с оценкой**.

Общая трудоемкость освоения дисциплины составляет 3 зачетные единицы.

ЛИСТ ИЗМЕНЕНИЙ

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение №1	08.06.2020	№6

Приложение к листу изменений №1

1. Образовательные технологии (к п.4 на 2020 г.)

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:

- аудио-файлы с лекциями из группы <https://vk.com/club190684626> (Виноградов, Д.В. Статистическое машинное обучение, 2020)
- PDF-презентации лекций из группы <https://vk.com/club190684626> (Виноградов, Д.В. Статистическое машинное обучение, 2020)
- 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.

2. Перечень БД и ИСС (к п. 6.2 на 2020 г.)

№ п/п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru
4	Компьютерные справочные правовые системы Консультант Плюс,

	Гарант
--	--------

3. Состав программного обеспечения (ПО) (к п. 7 на 2020 г.)

№п /п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)
1	Adobe Master Collection CS4	Adobe	лицензионное
2	ОС «Альт Образование» 8	ООО «Базальт СПО	лицензионное
3	Windows 10 Pro	Microsoft	лицензионное
4	Kaspersky Endpoint Security	Kaspersky	лицензионное
5	Microsoft Office 2016	Microsoft	лицензионное
6	Zoom	Zoom	лицензионное