

МИНОБРНАУКИ РОССИИ



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

ИНСТИТУТ ИНФОРМАЦИОННЫХ НАУК И ТЕХНОЛОГИЙ БЕЗОПАСНОСТИ
Факультет информационных систем и безопасности
Кафедра фундаментальной и прикладной математики

**ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ
ДАННЫХ**
РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Направление подготовки 01.04.04 Прикладная математика
Направленность (профиль) Математические методы и модели обработки
и защиты информации в социотехнических системах

Уровень высшего образования: магистратура
Форма обучения: очная, очно-заочная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2022

ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Рабочая программа дисциплины

Составитель:

кандидат физ.-мат. наук, доц., доцент кафедры фундаментальной и прикладной математики
Синицын В.Ю.

УТВЕРЖДЕНО

Протокол заседания кафедры
фундаментальной и прикладной математики
№ 10 от 05.04.2022

ОГЛАВЛЕНИЕ

1. Пояснительная записка	4
1.1. Цель и задачи дисциплины	4
1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций	4
1.3. Место дисциплины в структуре образовательной программы	4
2. Структура дисциплины	5
3. Содержание дисциплины	5
4. Образовательные технологии	7
5. Оценка планируемых результатов обучения	8
5.1 Система оценивания	8
5.2 Критерии выставления оценки по дисциплине	8
5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине	9
6. Учебно-методическое и информационное обеспечение дисциплины	12
6.1 Список источников и литературы	12
6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».	13
6.3 Профессиональные базы данных и информационно-справочные системы	13
7. Материально-техническое обеспечение дисциплины	13
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов	14
9. Методические материалы	15
9.1 Планы практических занятий	15
9.2 Методические рекомендации по подготовке письменных работ	18
Приложение 1. Аннотация рабочей программы дисциплины	19

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Цель дисциплины: формирование у студентов современных представлений об искусственных нейронных сетях и их приложениях, а также об интеллектуальном анализе данных с использованием реальных данных и актуальных прикладных задач.

Задачи дисциплины: познакомить студентов с современными алгоритмами и технологиями машинного обучения и их применением при обработке разнородной информации в социотехнических системах, а также развивать у студентов практические навыки интеллектуального анализа данных и интерпретации результатов исследования для решения прикладных задач.

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

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
ОПК-1. Способен обобщать и критически оценивать опыт и результаты научных исследований в области прикладной математики	ОПК-1.1. Определяет и анализирует источники проблемных ситуаций в экспериментальной и исследовательской деятельности.	<p><i>Знать:</i> основные стандартные типы прикладных задач, решаемых при помощи интеллектуального анализа данных и машинного обучения — классификация, регрессия, кластеризация и их особенности, методы оценивания качества построенных моделей, современные библиотеки для работы с моделями, нейронные сети</p> <p><i>Уметь:</i> работать с большими объемами данных, структурировать их, согласно требованиям заказчика, а также проводить анализ моделей различных типов, применять различные методы анализа данных для решения прикладных задач в социотехнических системах, разрабатывать и исследовать математические модели объектов, систем, процессов и технологий, предназначенных для проведения расчетов, анализа, подготовки решений, проводить научные эксперименты, оценивать результаты исследований</p> <p><i>Владеть:</i> навыками постановки прикладных задач машинного обучения, выбора соответствующих методов для их решения, анализа полученных результатов, а также навыками построения моделей и модификации стандартных методов при решении прикладных задач</p>

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Искусственные нейронные сети и интеллектуальный анализ данных» относится к обязательной части блока дисциплин учебного плана.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин и прохождения практик: «Математические методы исследования социальных систем», «Анализ данных в социотехнических системах», «Иностранный язык в профессиональной деятельности», «Принципы построения

математических моделей в социотехнических системах», «Методология исследовательской деятельности и академическая культура», «Математические модели в истории науки и техники», Учебная практика (Научно-исследовательская работа).

В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: «Программные средства научного исследования», «Интеллектуальные системы», «Современные системы программирования», Производственная практика (Научно-исследовательская работа).

2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 5 з.е., 180 академических часов.

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	Лекции	16
3	Практические занятия	34
	Всего:	50

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 130 академических часов.

Структура дисциплины для очно-заочной 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
4	Лекции	12
4	Практические занятия	28
	Всего:	40

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 140 академических часов.

3. Содержание дисциплины

Тема 1. Основные понятия и задачи машинного обучения.

Сущность и цели машинного обучения (machine learning). Интеллектуальный анализ данных (data mining). Обучение с учителем и без учителя. Типы задач: регрессия, классификация, прогнозирование, ранжирование, понижение размерности. Постановка задач обучения по прецедентам. Сбор и подготовка данных для анализа. Качество данных. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные. Методы ресамплинга. Основные понятия: модель зависимости, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность. Оценивание качества моделей. Компромисс между точностью предсказаний и

интерпретируемостью модели. Линейные модели регрессии и классификации. Метод наименьших квадратов. Примеры прикладных задач. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных. Конкурсы по анализу данных. Полигон алгоритмов классификации. Выбор соответствующего задаче метода решения. Построение моделей в вычислительной среде R и с помощью языка программирования Python. Универсальный интерфейс доступа к функциям машинного обучения в пакете caret.

Тема 2. Регрессия.

Роль регрессионного анализа при построении структурных моделей и для решения задач прогнозирования. Линейная модель множественной регрессии. Отбор оптимального набора предикторов линейной модели. Пошаговые методы с добавлением и исключением независимых переменных. Сравнение построенных моделей и оценка информативности предикторов. Обобщенные аддитивные модели. GAM для решения регрессионных задач. Нелинейные регрессионные модели. Полиномиальная регрессия. Ступенчатые функции, регрессионные сплайны. Бинарная и мультиномиальная логистическая регрессия. Регуляризация, частные наименьшие квадраты и kNN-регрессия. Построение деревьев регрессии. Гребневая регрессия. Лассо. Построение регрессионных моделей в вычислительной среде R.

Тема 3. Классификация.

Постановка задач классификации. Использование логистической регрессии для решения задач классификации. Оценивание параметров логистической модели. Ансамбли логистических моделей. Линейный и квадратичный дискриминантный анализ. Деревья классификации. Машины опорных векторов. Метод главных компонент и его применение для понижения размерности и классификации признаков. Методы кластеризации. Кластеризация по методу k-средних. Иерархическая кластеризация. Самоорганизующиеся карты Кохонена. Процедуры сравнения эффективности моделей классификации. Использование языка программирования R для построения и исследования классификаций.

Тема 4. Ресамплинг.

Генерация случайных псевдовыборок. Повторные выборки. Метод складного ножа. Параметрические и непараметрические бутстреп-процедуры. Интервальное оценивание параметров моделей при помощи ресамплинга. Проверка статистических гипотез и разработка статистических критериев на основе рандомизации и бутстреп-процедур. Перестановочные тесты. Исследование устойчивости моделей факторного анализа и тестирование надёжности классификаций при помощи ресамплинга. Перекрёстная проверка. Метод проверочной выборки. Перекрёстная проверка по отдельным наблюдениям. Многократная перекрёстная проверка. Перекрёстная проверка при решении задач классификации. Реализация ресамплинга на языке программирования R.

Тема 5. Отбор и регуляризация линейных моделей.

Отбор оптимального подмножества переменных. Пошаговые методы включения и исключения независимых переменных. Выбор оптимальной модели. Применение генетических алгоритмов для отбора переменных. Методы сжатия. Гребневая регрессия. Лассо. Методы снижения размерности. Регрессия на главные компоненты. Метод частных наименьших квадратов. Особенности работы с данными большой размерности. Нахождение оптимальной модели при помощи методов проверочной выборки и перекрёстной проверки. Использование языка программирования R для отбора и регуляризации линейных моделей.

Тема 6. Деревья решений.

Обзор методов, основанных на деревьях решений. Построение и использование регрессионных деревьев. Построение и использование деревьев классификации. Сравнение

деревьев с линейными моделями. Преимущества и недостатки деревьев решений. Ансамбли моделей. Композиции алгоритмов. Бэггинг, случайный лес, бустинг. Бустинг деревьев решений. Реализация методов, основанных на деревьях решений, в вычислительной среде R.

Тема 7. Машины опорных векторов.

Классификация с использованием гиперплоскости. Классификатор с максимальным зазором. Построение классификатора с максимальным зазором. Случай, когда разделяющая гиперплоскость не существует. Общие представления о классификаторах на опорных векторах. Классификация с использованием нелинейных решающих границ. Машина опорных векторов. Пример анализа данных по нарушению сердечной функции. Машины опорных векторов для случаев с несколькими классами. Классификация типа «один против одного». Классификация типа «один против всех». Связь с логистической регрессией. Реализация классификаторов на языке программирования R.

Тема 8. Искусственные нейронные сети.

Биологический прототип. Формальный нейрон. Возможности многослойного перцептрона. Классификация нейронных сетей. Пороговые функции. Обучение нейронных сетей. Сети прямого распространения. Теорема Колмогорова. Алгоритм обратного распространения ошибки. Сети радиально-базисных функций. Обучение без учителя. Правило Хебба. Сети Кохонена. Рекуррентные нейронные сети. Сети Хопфилда. Ассоциативная память. Теория адаптивного резонанса. Эффективные нейронные сети. Оптимизация процесса обучения. Критерии эффективности нейронных сетей. Конструктивный подход к построению нейронных сетей. Бинарные алгоритмы. Древоподобные нейронные сети. Алгоритмы Monoplan, NetLines и NetSphere. Метод динамического добавления узлов. Каскадная корреляция и ее модификации. Эволюционный способ создания нейронных сетей. Генетические алгоритмы. Различные подходы к кодированию нейронных сетей. Прямое кодирование. Порождающее кодирование. Подходы к построению нейронных сетей. Метод мониторинга динамики изменения ошибки. Эволюционное накопление признаков. Организация поиска архитектуры. Искусственные нейронные сети на языках программирования R и Python.

4. Образовательные технологии

Для проведения *занятий лекционного типа* по дисциплине применяются такие образовательные технологии как лекция-визуализация с применением слайд-проектора, лекция с разбором конкретных ситуаций, проблемная лекция.

Для проведения *практических занятий* используются такие образовательные технологии как: решение типовых задач для закрепления и формирования знаний, умений, навыков.

В рамках *самостоятельной работы* студентов проводится консультирование и проверка домашних заданий посредством электронной почты.

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.

5. Оценка планируемых результатов обучения

5.1 Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- опрос	2 балла	10 баллов
- тестирование	10 баллов	10 баллов
- расчётно-графическая работа	10 баллов	30 баллов
- доклад	10 баллов	10 баллов
Промежуточная аттестация - экзамен		
- ответы на вопросы билета		20 баллов
- итоговая контрольная работа		20 баллов
Итого за семестр		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	Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации. 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. Свободно 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 – «высокий».
82-68/ C	хорошо	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

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».
67-50/ D,E	удовлетво- рительно	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 выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».
49-0/ F,FX	неудовлет- ворительно	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 выставляются 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 дисциплине

Текущий контроль

Примерные задания для расчётно-графической работы №1 по теме «Ресамплинг»:

ВАРИАНТ 1

Задачи из книги:

Джеймс Г., Уиттон Д., Хасты Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R. Пер. с англ. С. Э. Мастицкого. – М.: ДМК Пресс, 2016. - 450 с. (основная литература):

Глава 5 стр. 215-220 №№ 5 (b, d), 6 (b, d), 7 (b, d), 8 (b, c), 9 (b, d).

Примерные задания для расчётно-графической работы №2 по теме «Деревья решений»:

ВАРИАНТ 1

Задачи из книги:

Джеймс Г., Уиттон Д., Хасты Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R. Пер. с англ. С. Э. Мастицкого. – М.: ДМК Пресс, 2016. - 450 с. (основная литература):

Глава 8 стр. 359-363 №№ 7, 8 (b, d), 9 (b, d), 10 (b, d), 11 (b), 12.

**Примерные задания для расчётно-графической работы №3
по теме «Машины опорных векторов»:
ВАРИАНТ 1**

Задачи из книги:

Джеймс Г., Уиттон Д., Хастис Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R. Пер. с англ. С. Э. Мастицкого. – М.: ДМК Пресс, 2016. - 450 с. (основная литература):

Глава 9 стр. 397-401 №№ 4, 5 (b, d, f), 6 (b, c), 7 (b, d), 8 (b, d, f, h).

**Примерные задания для тестирования
по теме «Искусственные нейронные сети»:**

Задания из массового открытого онлайн курса:

Нейронные сети. Учебный курс на портале Stepik.org. - Режим доступа: <https://stepik.org/401>

Задание 1. Раздел 2.2 Перцептрон Step 3

Задание 2. Раздел 2.3 Перцептрон: обучение Step 3

Задание 3. Раздел 2.4 Больше искусственных нейронов! Step 10

Задание 4. Раздел 2.5 Градиентный спуск Step 3

Задание 5. Раздел 2.5 Градиентный спуск Step 6

Задание 6. Раздел 2.5 Градиентный спуск Step 15

Задание 7. Раздел 3.1 Многослойный перцептрон Step 5

Задание 8. Раздел 3.2 Алгоритм обратного распространения ошибки Step 2

Задание 9. Раздел 3.2 Алгоритм обратного распространения ошибки Step 11

Задание 10. Раздел 3.4 Целевые функции Step 6

Примерные темы докладов

1. Разработка искусственных нейронных сетей для исследования психологического портрета личности.
2. Сравнительное исследование пятифакторной модели личности для российских и американских респондентов при помощи многомерных статистических методов и численного ресамплинга.
3. Построение и исследование регрессионных моделей для прогнозирования показателей надежности психометрических атласов.
4. Законы Ципфа для массовых открытых онлайн курсов.
5. Применение ансамблей регрессионных моделей.
6. Применение ансамблей логистических моделей.
7. Искусственные нейронные сети как средство проверки статистических гипотез.
8. Разработка статистических критериев с регулируемой мощностью, основанных на численном ресамплинге.
9. Сравнительное исследование мощности некоторых статистических критериев при помощи численного ресамплинга.
10. Критика современной методологии проверки статистических гипотез и усиление статистического вывода при помощи численного ресамплинга.
11. Анализ текстов на естественном языке в системе R.
12. Ресамплинг и его реализация в вычислительной среде R.
13. Бутстреп-оценки параметров моделей и их свойства.
14. Статистическое моделирование в вычислительной среде R.
15. Экспериментальное исследование мощности некоторых статистических критериев.
16. Исследование статистической устойчивости пятифакторной модели личности.
17. Статистические методы построения новых психометрических шкал.

18. Структурное моделирование психологического портрета личности с помощью теста NEO PI-R.
19. Применение комитетов искусственных нейронных сетей.
20. Характеристики качества датчиков псевдослучайных чисел.

Примерные вопросы для опроса см. п.9.1 Планы практических занятий, контрольные вопросы

Промежуточная аттестация (экзамен)
Контрольные вопросы по дисциплине:

1. Постановка задач обучения по прецедентам. Сбор и подготовка данных для анализа. Качество данных.
2. Основные понятия машинного обучения: модель зависимости, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность.
3. Линейные модели регрессии и классификации. Метод наименьших квадратов. Примеры прикладных задач.
4. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных.
5. Построение моделей в вычислительной среде R и с помощью языка программирования Python.
6. Линейная модель множественной регрессии. Отбор оптимального набора предикторов линейной модели. Сравнение моделей.
7. Обобщенные аддитивные модели. GAM для решения регрессионных задач.
8. Нелинейные регрессионные модели. Полиномиальная регрессия. Ступенчатые функции, регрессионные сплайны.
9. Бинарная и мультиномиальная логистическая регрессия.
10. Построение регрессионных моделей в вычислительной среде R.
11. Использование логистической регрессии для решения задач классификации.
12. Линейный и квадратичный дискриминантный анализ.
13. Метод главных компонент и его применение для понижения размерности и классификации признаков.
14. Методы кластеризации. Кластеризация по методу k-средних. Иерархическая кластеризация.
15. Использование языка программирования R для построения и исследования классификаций.
16. Повторные выборки. Параметрические и непараметрические бутстреп-процедуры. Интервальное оценивание параметров моделей при помощи ресамплинга.
17. Проверка статистических гипотез и разработка статистических критериев на основе рандомизации и бутстреп-процедур. Перестановочные тесты.
18. Исследование устойчивости структурных моделей и тестирование надёжности классификаций при помощи ресамплинга.
19. Перекрёстная проверка. Метод проверочной выборки. Перекрёстная проверка при решении задач классификации.
20. Реализация ресамплинга на языке программирования R.
21. Отбор оптимального подмножества переменных. Пошаговые методы включения и исключения независимых переменных. Выбор оптимальной модели.
22. Применение генетических алгоритмов для отбора переменных. Методы сжатия. Гребневая регрессия. Лассо.
23. Методы снижения размерности. Регрессия на главные компоненты. Метод частных наименьших квадратов.

24. Использование языка программирования R для отбора и регуляризации линейных моделей.
25. Построение и использование регрессионных деревьев. Построение и использование деревьев классификации. Сравнение деревьев с линейными моделями.
26. Ансамбли моделей. Композиции алгоритмов. Бэггинг, случайный лес, бустинг. Бустинг деревьев решений.
27. Реализация методов, основанных на деревьях решений, в вычислительной среде R.
28. Построение классификатора с максимальным зазором. Машина опорных векторов.
29. Классификация типа «один против одного». Классификация типа «один против всех». Связь с логистической регрессией.
30. Реализация классификаторов на языке программирования R.
31. Формальный нейрон. Многослойный перцептрон. Классификация нейронных сетей.
32. Пороговые функции. Обучение нейронных сетей. Сети прямого распространения. Теорема Колмогорова.
33. Алгоритм обратного распространения ошибки.
34. Сети радиально-базисных функций. Сети Кохонена. Рекуррентные нейронные сети. Сети Хопфилда.
35. Ассоциативная память. Теория адаптивного резонанса.
36. Критерии эффективности нейронных сетей. Конструктивный подход к построению нейронных сетей. Древоподобные нейронные сети.
37. Алгоритмы Monoplan, NetLines и NetSphere.
38. Метод динамического добавления узлов. Каскадная корреляция и ее модификации.
39. Эволюционный способ создания нейронных сетей. Генетические алгоритмы.
40. Подходы к кодированию нейронных сетей. Прямое кодирование. Порождающее кодирование.

Примерные задания для итоговой контрольной работы:

Вариант 1

Задачи из книги:

Джеймс Г., Уиттон Д., Хасты Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R. Пер. с англ. С. Э. Мастицкого. – М.: ДМК Пресс, 2016. - 450 с. (основная литература):

Глава 5 стр. 215-220 №№ 5 (с), 7 (с), 9 (с);

Глава 8 стр. 359-363 №№ 8 (с), 10 (с), 11 (с);

Глава 9 стр. 397-401 №№ 5 (с), 7 (с), 8 (с, е).

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы

Литература

Основная

1. Введение в статистическое обучение с примерами на языке R / Е. Джеймс, Д. Уитгон, Т. Хасты, Р.В. Тибширани ; пер. с англ. С.Э. Мастицкого. - 2-е изд., испр. - Москва : ДМК Пресс, 2017. - 456 с.
2. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии

машинного обучения / С. Рашка ; пер. с англ. А.В. Логунова. – Москва : ДМК Пресс, 2017. - 418 с.

Дополнительная

1. Груздев, А.В. Прогнозное моделирование в IBM SPSS Statistics и R: Метод деревьев решений / А.В. Груздев. - Москва : ДМК Пресс, 2016. - 278 с.

6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

1. Аникин Ю. и др. Введение в аналитику больших массивов данных. Дистанционный учебный курс на портале Intuit.ru [Электронный ресурс]. - Режим доступа: <http://www.intuit.ru/studies/courses/12385/1181/info>
2. Воронцов К.В. Машинное обучение. Курс лекций [Электронный ресурс]. - Режим доступа: <http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%BE>
3. Нейронные сети. Учебный курс на портале Stepik.org. - Режим доступа: <https://stepik.org/401>
4. Программирование на Python. Учебный курс на портале Stepik.org. - Режим_доступа: <https://stepik.org/67>
5. Шитиков В.К., Мастицкий С.Э. (2017) Классификация, регрессия и другие алгоритмы Data Mining с использованием R. 351 с. – Электронная книга, адрес доступа: <https://github.com/ranalytics/data-mining>

Национальная электронная библиотека (НЭБ) www.rusneb.ru

ELibrary.ru Научная электронная библиотека www.elibrary.ru

6.3 Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.ruh.ru/ru/bases>

Информационные справочные системы:

1. Консультант Плюс
2. Гарант

7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения:

- *для лекций*: учебные аудитории, оснащённые доской, компьютером или ноутбуком, проектором (стационарным или переносным) для демонстрации учебных материалов.

Состав программного обеспечения:

1. Windows
2. Microsoft Office
3. Kaspersky Endpoint Security

- *для практических занятий*: компьютерный класс или лаборатория, оснащённые доской, компьютером или ноутбуком для преподавателя, компьютерами для обучающихся, проектором (стационарным или переносным) для демонстрации учебных материалов.

Состав программного обеспечения:

1. Windows
2. Microsoft Office
3. Mozilla Firefox
4. Язык программирования R

5. Язык программирования Python
6. Kaspersky Endpoint Security

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 материалы

9.1 Планы практических занятий

Тема 1. Основные понятия и задачи машинного обучения.

Форма проведения – решение типовых задач для закрепления и формирования знаний, умений, навыков

Примерные задачи для решения в аудитории:

Из книги Джеймс Г. и др. Введение в статистическое обучение с примерами на языке R. (основная литература):

Задачи для решения в аудитории:	Глава 2 стр. 65-70 №№ 2, 4, 6, 8 (а, с), 9 (а, с), 10(а, с)
Домашнее задание:	Глава 2 стр. 65-70 №№ 1, 3, 5, 8 (b), 9 (b, d), 10(b, d, e)
Дополнительно:	Глава 2 стр. 65-70 все остальные задачи

Контрольные вопросы:

1. Постановка задач обучения по прецедентам. Сбор и подготовка данных.
2. Модель зависимости, метод обучения.
3. Функция потерь и функционал качества.
4. Принцип минимизации эмпирического риска, обобщающая способность.
5. Линейные модели регрессии и классификации.
6. Метод наименьших квадратов. Примеры прикладных задач.
7. Методика экспериментального исследования и сравнения алгоритмов.
8. Построение моделей в вычислительной среде R.

Тема 2. Регрессия.

Форма проведения – решение типовых задач для закрепления и формирования знаний, умений, навыков

Примерные задачи для решения в аудитории:

Из книги Джеймс Г. и др. Введение в статистическое обучение с примерами на языке R. (основная литература):

Задачи для решения в аудитории:	Глава 3 стр. 135-142 №№ 2, 4, 6, 8 (а), 10 (с), 12(а, с)
Домашнее задание:	Глава 3 стр. 135-142 №№ 1, 3, 5, 8 (b), 9 (а), 11(b)
Дополнительно:	Глава 3 стр. 135-142 все остальные задачи

Контрольные вопросы:

1. Линейная модель множественной регрессии.
2. Отбор оптимального набора предикторов линейной модели.
3. Сравнение моделей.
4. Обобщенные аддитивные модели. GAM для решения регрессионных задач.
5. Нелинейные регрессионные модели.
6. Полиномиальная регрессия. Ступенчатые функции, регрессионные сплайны.
7. Бинарная и мультиномиальная логистическая регрессия.
8. Построение регрессионных моделей в вычислительной среде R.

Тема 3. Классификация.

Форма проведения – решение типовых задач для закрепления и формирования знаний, умений, навыков

Примерные задачи для решения в аудитории:

Из книги Джеймс Г. и др. Введение в статистическое обучение с примерами на языке R. (основная литература):

Задачи для решения в аудитории: Глава 4 стр. 186-191 №№ 2, 4, 6, 8, 10 (a), 11 (c), 12(a)
 Домашнее задание: Глава 4 стр. 186-191 №№ 1, 3, 5, 9, 10 (b), 11 (a), 12(b)
 Дополнительно: Глава 4 стр. 186-191 все остальные задачи

Контрольные вопросы:

1. Использование логистической регрессии для решения задач классификации.
2. Линейный и квадратичный дискриминантный анализ.
3. Метод главных компонент и его применение для понижения размерности и классификации признаков.
4. Кластеризация по методу k-средних.
5. Иерархическая кластеризация.
6. Использование языка программирования R для построения и исследования классификаций.

Тема 4. Ресамплинг.

Форма проведения – решение типовых задач для закрепления и формирования знаний, умений, навыков

Примерные задачи для решения в аудитории:

Из книги Джеймс Г. и др. Введение в статистическое обучение с примерами на языке R. (основная литература):

Задачи для решения в аудитории: Глава 5 стр. 215-220 №№ 2, 4, 5 (a), 6 (a), 7 (c), 8 (a), 9 (c)
 Домашнее задание: Глава 5 стр. 215-220 №№ 1, 3, 5 (c), 6 (c), 7 (a), 8 (c), 9 (a)
 Дополнительно: Глава 5 стр. 215-220 все остальные задачи

Контрольные вопросы:

1. Параметрические и непараметрические бутстреп-процедуры.
2. Интервальное оценивание параметров моделей при помощи ресамплинга.
3. Проверка статистических гипотез и разработка статистических критериев на основе рандомизации и бутстреп-процедур. Перестановочные тесты.
4. Исследование устойчивости структурных моделей при помощи ресамплинга.
5. Тестирование надёжности классификаций при помощи ресамплинга.
6. Перекрёстная проверка. Метод проверочной выборки.
7. Перекрёстная проверка при решении задач классификации.
8. Реализация ресамплинга на языке программирования R.

Тема 5. Отбор и регуляризация линейных моделей.

Форма проведения – решение типовых задач для закрепления и формирования знаний, умений, навыков

Примерные задачи для решения в аудитории:

Из книги Джеймс Г. и др. Введение в статистическое обучение с примерами на языке R.
(основная литература):

Задачи для решения в аудитории: Глава 6 стр. 281-287 №№ 2, 4, 6, 8 (с), 9 (а), 10 (с), 11 (а)
 Домашнее задание: Глава 6 стр. 281-287 №№ 1, 3, 5, 8 (b), 9 (с), 10 (а), 11 (с)
 Дополнительно: Глава 6 стр. 281-287 все остальные задачи

Контрольные вопросы:

1. Пошаговые методы включения и исключения независимых переменных. Выбор оптимальной модели.
2. Применение генетических алгоритмов для отбора переменных.
3. Методы сжатия. Гребневая регрессия. Лассо.
4. Методы снижения размерности.
5. Регрессия на главные компоненты. Метод частных наименьших квадратов.
6. Использование языка программирования R для отбора и регуляризации линейных моделей.

Тема 6. Деревья решений.

Форма проведения – решение типовых задач для закрепления и формирования знаний, умений, навыков

Примерные задачи для решения в аудитории:

Из книги Джеймс Г. и др. Введение в статистическое обучение с примерами на языке R.
(основная литература):

Задачи для решения в аудитории: Глава 8 стр. 359-363 №№ 2, 4, 6, 8 (а), 9 (а), 10 (а), 11 (а)
 Домашнее задание: Глава 8 стр. 359-363 №№ 1, 3, 5, 8 (с), 9 (с), 10 (с), 11 (с)
 Дополнительно: Глава 8 стр. 359-363 все остальные задачи

Контрольные вопросы:

1. Построение и использование регрессионных деревьев.
2. Построение и использование деревьев классификации.
3. Ансамбли моделей. Композиции алгоритмов.
4. Бэггинг, случайный лес, бустинг. Бустинг деревьев решений.

Тема 7. Машины опорных векторов.

Форма проведения – решение типовых задач для закрепления и формирования знаний, умений, навыков

Примерные задачи для решения в аудитории:

Из книги Джеймс Г. и др. Введение в статистическое обучение с примерами на языке R.
(основная литература):

Задачи для решения в аудитории: Глава 9 стр. 397-401 №№ 2, 5 (а), 6 (а), 7 (а), 8 (с)
 Домашнее задание: Глава 9 стр. 397-401 №№ 1, 3, 5 (с), 6 (с), 7 (с), 8 (а)
 Дополнительно: Глава 9 стр. 397-401 все остальные задачи

Контрольные вопросы:

1. Построение классификатора с максимальным зазором.
2. Машина опорных векторов.
3. Классификация типа «один против одного».
4. Классификация типа «один против всех».
5. Реализация классификаторов на языке программирования R.

9.2 Методические рекомендации по подготовке письменных работ

Отчет по выполнению расчетно-графических работ по дисциплине «Искусственные нейронные сети и интеллектуальный анализ данных» объемом 15-20 страниц выполняется студентом по каждой работе отдельно. Правила оформления отчета по выполнению расчетно-графических работ совпадают с правилами оформления курсовой работы, которые содержатся в «Методических рекомендациях по подготовке и оформлению курсовой работы» (официальный сайт кафедры ФПМ ИИНТБ РГГУ).

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина «Искусственные нейронные сети и интеллектуальный анализ данных» реализуется на факультете информационных систем и безопасности кафедрой фундаментальной и прикладной математики.

Цель дисциплины: формирование у студентов современных представлений об искусственных нейронных сетях и их приложениях, а также об интеллектуальном анализе данных с использованием реальных данных и актуальных прикладных задач.

Задачи: познакомить студентов с современными алгоритмами и технологиями машинного обучения и их применением при обработке разнородной информации в социотехнических системах, а также развивать у студентов практические навыки интеллектуального анализа данных и интерпретации результатов исследования для решения прикладных задач.

Дисциплина направлена на формирование следующих компетенций:

- ОПК-1. Способен обобщать и критически оценивать опыт и результаты научных исследований в области прикладной математики.

В результате освоения дисциплины обучающийся должен:

Знать: основные стандартные типы прикладных задач, решаемых при помощи интеллектуального анализа данных и машинного обучения — классификация, регрессия, кластеризация и их особенности, методы оценивания качества построенных моделей, современные библиотеки для работы с моделями, нейронные сети

Уметь: работать с большими объемами данных, структурировать их, согласно требованиям заказчика, а также проводить анализ моделей различных типов, применять различные методы анализа данных для решения прикладных задач в социотехнических системах, разрабатывать и исследовать математические модели объектов, систем, процессов и технологий, предназначенных для проведения расчетов, анализа, подготовки решений, проводить научные эксперименты, оценивать результаты исследований

Владеть: навыками постановки прикладных задач машинного обучения, выбора соответствующих методов для их решения, анализа полученных результатов, а также навыками построения моделей и модификации стандартных методов при решении прикладных задач.

По дисциплине предусмотрена промежуточная аттестация в форме экзамена.

Общая трудоемкость освоения дисциплины составляет 5 зачетных единиц.