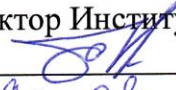


Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Макаренко Елена Владимировна
Должность: Ректор
Дата подписания: 08.11.2023 16:31:03
Уникальный программный ключ:
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ
Директор Института магистратуры
 Иванова Е.А.
« 29 » 08 2022 г.

**Рабочая программа дисциплины
Математические методы анализа больших данных**

Направление 09.04.03 Прикладная информатика
магистерская программа 09.04.03.01 "Информационные системы и технологии в бизнесе"

Для набора 2022 года


Квалификация
магистр


КАФЕДРА Фундаментальная и прикладная математика**Распределение часов дисциплины по курсам**

Курс Вид занятий	1		Итого	
	уп	рп		
Лекции	8	8	8	8
Практические	10	10	10	10
Итого ауд.	18	18	18	18
Контактная работа	18	18	18	18
Сам. работа	225	225	225	225
Часы на контроль	9	9	9	9
Итого	252	252	252	252

ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 22.02.2022 протокол № 7.

Программу составил(и): к.ф.-м.н, доц., Богачев Т.В. 

Зав. кафедрой: к.э.н. Рутта Н.А. 

Методическим советом направления: д.э.н., проф., Щербаков С.М. 

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	формирование у обучающихся способности применять математические методы и модели для обработки и анализа больших данных при решении профессиональных задач
-----	---

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПК-1:Способен адаптировать и применять методы и алгоритмы машинного обучения для решения прикладных задач в различных предметных областях

ПК-3:Способен руководить проектами со стороны заказчика по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов

В результате освоения дисциплины обучающийся должен:

Знать:
- математические модели, методы и алгоритмы для обработки и анализа больших данных (соотнесено с индикатором ПК- 1.1). - принципы построения моделей глубоких нейронных сетей и глубокого машинного обучения (соотнесено с индикатором ПК-3.1).
Уметь:
- выбирать и применять математические модели, методы и алгоритмы для решения прикладных задач анализа больших данных (соотнесено с индикатором ПК-1.2). - руководить выполнением коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов (соотнесено с индикатором ПК-3.2).
Владеть:
- навыками применения методов и алгоритмов машинного обучения, современных инструментальных средств и систем программирования для разработки новых методов и моделей машинного обучения (соотнесено с индикатором ПК-1.3). - навыками проектной деятельности для создания, поддержки и использования систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов (соотнесено с индикатором ПК-3.3).

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература
	Раздел 1. «Математические методы машинного обучения для анализа больших данных»				
1.1	Тема 1. "Метрические алгоритмы. Байесовский подход к обучению". Обобщенный алгоритм. Примеры частных алгоритмов. Метод ближайших соседей. Метод окна Парзена. Понятие выступа объекта. Алгоритм отбора эталонов. Проклятие размерности. Выбор метрики. Применение формула Байеса к задаче машинного обучения. Построение приближенной плотности распределения: одномерный случай (дискретное, кусочно- постоянное, приближение с помощью ядра, приближение нормальным законом распределения), многомерный случай (наивный байесовский классификатор, приближение с помощью ядра, многомерное нормальное распределение). Логистическая регрессия. Бинаризация признаков. Скоринг. Смеси распределений. EM-алгоритм восстановления смеси. /Лек/	1	2	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4

1.2	<p>Тема 1. "Метрические алгоритмы. Байесовский подход к обучению". Обобщенный алгоритм. Примеры частных алгоритмов. Метод ближайших соседей.</p> <p>Метод окна Парзена. Понятие выступа объекта. Алгоритм отбора эталонов.</p> <p>Проклятие размерности. Выбор метрики. Применение формула Байеса к задаче машинного обучения.</p> <p>Построение приближенной плотности распределения: одномерный случай (дискретное, кусочно- постоянное, приближение с помощью ядра, приближение нормальным законом распределения), многомерный случай (наивный байесовский классификатор, приближение с помощью ядра, многомерное нормальное распределение).</p> <p>Логистическая регрессия. Применение Python /Пр/</p>	1	2	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
1.3	<p>Тема 1. "Метрические алгоритмы. Байесовский подход к обучению". Обобщенный алгоритм. Примеры частных алгоритмов. Метод ближайших соседей.</p> <p>Метод окна Парзена. Понятие выступа объекта. Алгоритм отбора эталонов.</p> <p>Проклятие размерности. Выбор метрики. Применение формула Байеса к задаче машинного обучения.</p> <p>Построение приближенной плотности распределения: одномерный случай (дискретное, кусочно- постоянное, приближение с помощью ядра, приближение нормальным законом распределения), многомерный случай (наивный байесовский классификатор, приближение с помощью ядра, многомерное нормальное распределение).</p> <p>Логистическая регрессия. Бинаризация признаков.</p> <p>Скоринг. Смеси распределений. EM-алгоритм восстановления смеси. /Ср/</p>	1	30	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
1.4	<p>Тема 2. "Линейные алгоритмы классификации. Метод опорных векторов".Общая формула линейного классификатора. Метод стохастического градиента.</p> <p>Частные случаи. Обоснование метода СГ. Выступ объекта для линейного классификатора. ROC и AUC. Случаи линейно разделимой и неразделимой выборки.</p> <p>Двойственная задача. Типы объектов. Нелинейное обобщение SVM. SVM-регрессия.</p> <p>L1 регуляризация /Лек/</p>	1	2	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
1.5	<p>Тема 2. "Линейные алгоритмы классификации. Метод опорных векторов".Общая формула линейного классификатора. Метод стохастического градиента.</p> <p>Частные случаи. Обоснование метода СГ. Выступ объекта для линейного классификатора. ROC и AUC. Случаи линейно разделимой и неразделимой выборки.</p> <p>Двойственная задача. Типы объектов. Нелинейное обобщение SVM. SVM-регрессия.</p> <p>L1 регуляризация /Пр/</p>	1	2	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
1.6	<p>Тема 2. "Линейные алгоритмы классификации. Метод опорных векторов".Общая формула линейного классификатора. Метод стохастического градиента.</p> <p>Частные случаи. Обоснование метода СГ. Выступ объекта для линейного классификатора. ROC и AUC. Случаи линейно разделимой и неразделимой выборки.</p> <p>Двойственная задача. Типы объектов. Нелинейное обобщение SVM. SVM-регрессия.</p> <p>L1 регуляризация /Ср/</p>	1	24	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4

1.7	Тема 3. "Методы восстановления регрессии. Логические методы классификации. Композиции алгоритмов". Метод наименьших квадратов. Геометрический смысл. Регуляризация. Сингулярное разложение. Непараметрическая регрессия. Понятие закономерности. Критерий качества закономерностей. Поиск закономерностей. Алгоритмы классификации на основе логических закономерностей. AdaBoost. AnyBoost. Градиентный бустинг. Бэггинг и метод случайных подпространств /Ср/	1	24	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
1.8	Тема 4. "Ранжирование и рекомендательные системы. Тематическое моделирование. Кластеризация."Оценки качества. Подходы к решению задачи: поточечный, попарный, списочный. Латентный семантический анализ (LSA). Вероятностный LSA (PLSA). Латентное размещение Дирихле (LDA). EM-алгоритм. Метод k-средних. DBSCAN /Пр/	1	2	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
1.9	Тема 4. "Ранжирование и рекомендательные системы. Тематическое моделирование. Кластеризация."Оценки качества. Подходы к решению задачи: поточечный, попарный, списочный. Латентный семантический анализ (LSA). Вероятностный LSA (PLSA). Латентное размещение Дирихле (LDA). EM-алгоритм. Метод k-средних. DBSCAN /Ср/	1	16	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
Раздел 2. Искусственные нейронные сети					
2.1	Тема 1. "Искусственные нейронные сети" Типы функций активации нейронов. Представление нейронных сетей с помощью направленных графов. Архитектура сетей. Сети прямого распространения. Рекуррентные сети. Обучение нейронных сетей. Обучение, основанное на коррекции ошибок. Обучение на основе памяти. Обучение Хебба. Математические модели предложенного Хеббом механизма модификации синаптической связи. Конкурентное обучение. Обучение Больцмана. Обучение с учителем. Обучение с подкреплением. Обучение без учителя. /Лек/	1	2	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
2.2	Тема 1. "Искусственные нейронные сети". Обучение нейронных сетей. Обучение, основанное на коррекции ошибок. Обучение на основе памяти. Обучение Хебба. Математические модели предложенного Хеббом механизма модификации синаптической связи. Конкурентное обучение. Обучение Больцмана. Обучение с учителем. Обучение с подкреплением. Обучение без учителя.Применение Phython /Пр/	1	2	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
2.3	Тема 1. "Искусственные нейронные сети" Типы функций активации нейронов. Представление нейронных сетей с помощью направленных графов. Архитектура сетей. Сети прямого распространения. Рекуррентные сети. Обучение нейронных сетей. Обучение, основанное на коррекции ошибок. Обучение на основе памяти. Обучение Хебба. Математические модели предложенного Хеббом механизма модификации синаптической связи. Конкурентное обучение. Обучение Больцмана. Обучение с учителем. Обучение с подкреплением. Обучение без учителя. /Ср/	1	22	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4

2.4	Тема 2. "Однослойные и многослойные перцептроны" Однослойный перцептрон. Обучение перцептрона. Методы безусловной оптимизации. Метод наискорейшего спуска. Метод Ньютона. Метод Гаусса- Ньютона. Взаимосвязь перцептрона и байесовского классификатора. Многослойный перцептрон. Алгоритм обратного распространения ошибки. Извлечение признаков. Линейный дискриминант Фишера. Сети свертки. /Лек/	1	2	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
2.5	Тема 2. "Однослойные и многослойные перцептроны" Однослойный перцептрон. Обучение перцептрона. Методы безусловной оптимизации. Метод наискорейшего спуска. Метод Ньютона. Метод Гаусса- Ньютона. Взаимосвязь перцептрона и байесовского классификатора. Многослойный перцептрон. Алгоритм обратного распространения ошибки. Извлечение признаков. Линейный дискриминант Фишера. Сети свертки. /Ср/	1	16	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
2.6	Тема 3. "Сети на основе радиальных базисных функций". Теорема Ковера о разделимости множеств. Разделяющая способность поверхности. Задача интерполяции. Теория регуляризации. Функция Грина. Решение задачи регуляризации. Многомерные функции Гаусса. Обобщенные сети на основе радиальных базисных функций. Свойства аппроксимации сетей RBF. Сравнение сетей RBF и многослойных перцептронов. /Ср/	1	30	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
2.7	Тема 4. "Анализ главных компонент". Анализ признаков на основе самоорганизации. Структура анализа главных компонент. Представление данных. Сокращение размерности. Фильтр Хебба для выделения максимальных собственных значений. Анализ главных компонент на основе правила Хебба. /Ср/	1	25	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
2.8	Тема 5. "Карты самоорганизации Кохонена". Модели отображения признаков. Карты самоорганизации. Процессы конкуренции, кооперации и адаптации. Варианты самоорганизующихся карт. Адаптивные тензорные веса. Самоорганизующиеся карты для символьных строк. Самоорганизующиеся карты с эволюционным обучением. Пакеты программ, реализующие самоорганизующиеся карты. Применение Loginom /Пр/	1	2	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
2.9	Тема 5. "Карты самоорганизации Кохонена". Модели отображения признаков. Карты самоорганизации. Процессы конкуренции, кооперации и адаптации. Варианты самоорганизующихся карт. Адаптивные тензорные веса. Самоорганизующиеся карты для символьных строк. Самоорганизующиеся карты с эволюционным обучением. Пакеты программ, реализующие самоорганизующиеся карты. /Ср/	1	12	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
2.10	Тема 6. "Нейродинамические модели". Динамические системы. Пространство состояний. Условие Лившица. Теорема о дивергенции. Устойчивость состояний равновесия. Теоремы Ляпунова. Гиперболические аттракторы. Аддитивная и связанная нейродинамические модели. Модель Хопфилда. Теорема Козна- Гроссберга /Ср/	1	26	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4
2.11	/Экзамен/	1	9	ПК-1 ПК-3	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Л2.4

4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущей и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л1.1	Воронова, Л. И., Воронов, В. И.	Big Data. Методы и средства анализа: учебное пособие	Москва: Московский технический университет связи и информатики, 2016	http://www.iprbookshop.ru/61463.html неограниченный доступ для зарегистрированных пользователей
Л1.2	Келлехер Д., Тирни Б.	Наука о данных: базовый курс: учебное пособие	Москва: Альпина Паблишер, 2020	https://biblioclub.ru/index.php?page=book&id=598235 неограниченный доступ для зарегистрированных пользователей

5.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л2.1	Терновая, Г. Н.	Теория вероятностей и математическая статистика в примерах: электронное учебное пособие	Астрахань: Астраханский государственный архитектурно-строительный университет, ЭБС АСВ, 2019	http://www.iprbookshop.ru/93094.html неограниченный доступ для зарегистрированных пользователей
Л2.2	Плас Дж. Вандер	Python для сложных задач: наука о данных и машинное обучение	Санкт-Петербург: Питер, 2018	https://ibooks.ru/reading.php?short=1&productid=356721 неограниченный доступ для зарегистрированных пользователей
Л2.3	Силен Дэви, Мейсман Арно, Али Мохамед	Основы Data Science и Big Data. Python и наука о данных. — (Серия «Библиотека программиста»)	Санкт-Петербург: Питер, 2019	https://ibooks.ru/reading.php?short=1&productid=376837 неограниченный доступ для зарегистрированных пользователей
Л2.4		Студент. Аспирант. Исследователь: всероссийский научный журнал: журнал	Владивосток: Эксперт-Наука, 2021	https://biblioclub.ru/index.php?page=book&id=685681 неограниченный доступ для зарегистрированных пользователей

5.3 Профессиональные базы данных и информационные справочные системы

Консультант +

Образовательная платформа по Python - <http://pybrain.org/docs/index.html>сайт федеральной государственной службы статистики - <https://rosstat.gov.ru/databases>

5.4. Перечень программного обеспечения

Loginom

Python

5.5. Учебно-методические материалы для студентов с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:

- столы, стулья;

- персональный компьютер / ноутбук (переносной);

- проектор;

- экран / интерактивная доска.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1 Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

1.1 Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
ПК-1:Способен адаптировать и применять методы и алгоритмы машинного обучения для решения прикладных задач в различных предметных областях			
З. математические модели, методы и алгоритмы для обработки и анализа больших данных	изучает основную и дополнительную литературу, содержащую материал об основных понятиях инструментальных средствах и математических методах, используемых при решении профессиональных задач, для подготовки к зачету, и устному опросу	полнота и содержательность ответа на зачете, устном опросе, соответствие ответов материалу, содержащемуся в изученной литературе	УО (Раздел 1 в. 1-17 Раздел 2 в. 1-18) Э (1-36)
У. выбирать и применять математические модели, методы и алгоритмы для решения прикладных задач анализа больших данных	решение практико-ориентированных и практических заданий: применение составление программ на Python с использованием библиотек	правильность решения заданий на составление программ на Python с использованием с использованием библиотек	Раздел 1. ПЗ 1-4 Раздел 2. ПЗ 1-6
В. навыками применения методов и алгоритмов машинного обучения, современных инструментальных средств и систем программирования для разработки новых методов и моделей машинного обучения	решение практико-ориентированных и практических заданий: составление программ на Python с использованием библиотек (NumPy,Pandas, matplotlib, PyBriant)	обоснованность применения методов для: решения заданий на составление программ на Python с использованием библиотек (NumPy,Pandas, matplotlib, PyBriant)	Раздел 1. ПЗ 1-4 Раздел 2. ПЗ 1-6
ПК-3:Способен руководить проектами со стороны заказчика по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов			
З. принципы построения моделей глубоких нейронных сетей и глубокого машинного обучения	изучает основную и дополнительную литературу, содержащую материал об основных понятиях инструментальных средствах и математических методах, используемых при решении профессиональных задач, для подготовки к зачету, и устному опросу	полнота и содержательность ответа на зачете, устном опросе, соответствие ответов материалу, содержащемуся в изученной литературе	УО (Раздел 1 в. 1-17 Раздел 2 в. 1-18) Э (1-36)

У. руководить выполнением коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов	решение практико-ориентированных и практических заданий: составление программ на Python с использованием базовых конструкций, условий, циклов, массивов	правильность решения заданий на составление программ на Python с использованием базовых конструкций, условий, циклов, массивов	Раздел 1. ПЗ 1-4 Раздел 2. ПЗ 1-6
В. навыками проектной деятельности для создания, поддержки и использования систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов	решение практико-ориентированных и практических заданий: применение различных методов с использованием Loginom, составление программ на Python с использованием базовых конструкций, условий, циклов, массивов)	обоснованность применения методов для: решения заданий с использованием Loginom, решения заданий на составление программ на Python с использованием базовых конструкций, условий, циклов, массивов)	Раздел 1. ПЗ 1-4 Раздел 2. ПЗ 1-6

Э – вопросы к экзамену, ПЗ-практическое задание, УО- устный опрос

1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале:

- 84-100 баллов (оценка «отлично»)
- 67-83 баллов (оценка «хорошо»)
- 50-66 баллов (оценка «удовлетворительно»)
- 0-49 баллов (оценка «неудовлетворительно»)

2 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

Вопросы к экзамену

1. Основные определения: прецедент, обучающая выборка, признаки объектов, виды признаков, матрица объектов-признаков. Модель алгоритмов, метод обучения, функционал качества алгоритма.
2. Вероятностная постановка задачи обучения. Принцип максимума правдоподобия. Связь максимизации правдоподобия и минимизации эмпирического риска.
3. Проблема переобучения и обобщающая способность алгоритма. Состоятельные методы обучения. Эмпирические оценки обобщающей способности.
4. Выбор алгоритма для вероятностной постановки задачи. Функционал среднего риска.
5. Метрические алгоритмы классификации. Обобщенный метрический классификатор. Виды и

особенности частных случаев: методы ближайшего соседа, к ближайших соседей, взвешенных соседей, парзеновского окна постоянной и переменной ширины.

6. Классификация объектов по значению отступа. Алгоритм STOLP отбора эталонных объектов. Выбор метрики и проклятие размерности.
7. Приближенное вычисление плотности распределения. Наивный байесовский классификатор. Одномерный случай. Многомерный случай. Смеси распределений. EM- алгоритм разделения смеси. Смеси многомерных нормальных распределений.
8. Линейные алгоритмы классификации. Модель Мак Каллока-Питтса, алгоритм стохастического градиента для минимизации функционала среднего риска. Эвристики для улучшения сходимости и обобщающей способности.
9. Логистическая регрессия. Случайные величины с экспонентным законом распределения. Теорема о линейности байесовского классификатора (с доказательством).
10. Метод опорных векторов (SVM). Случай линейно разделимой выборки. Случай линейно неразделимой выборки. Функция Лагранжа. Классификация объектов в зависимости от значений множителей Лагранжа. Двойственная задача. Обучение SVM.
11. Кривая ошибок и AUC. Формула вычисления AUC. Примеры.
12. Алгоритмы восстановления регрессии. Метод наименьших квадратов. Непараметрическая регрессия. Многомерная линейная регрессия. Подход с использованием SVD-разложения матрицы. Гребневая регрессия. Метод главных компонент PCA.
13. Логические методы классификации. Понятие информативности предиката: эвристическое, вероятностное, энтропийное. Поиск информативных закономерностей. Алгоритмы для номинальных и порядковых признаков.
14. Градиентный алгоритм синтеза конъюнкций. Построение решающего списка и решающего дерева. Редукция деревьев.
15. Композиции алгоритмов. AdaBoost. AnyBoost. Градиентный бустинг. Бэггинг, метод случайных подпространств.
16. Ранжирование и рекомендательные системы. Оценки качества. Алгоритмы построения ранжирующих систем: поточечный, попарный и списочный.
17. Тематическое моделирование. Векторная модель, LSA, PLSA, LDA.
18. Кластеризация. EM-алгоритм, метод k-средних.
19. Обучение нейронных сетей. Обучение, основанное на коррекции ошибок. Обучение на основе памяти.
20. Обучение Хебба. Математические модели предложенного Хеббом механизма модификации синаптической связи. Конкуренционное обучение.
21. Обучение Больцмана.
22. Обучение с учителем. Обучение с подкреплением. Обучение без учителя.
23. Методы безусловной оптимизации. Метод наискорейшего спуска. Метод Ньютона. Метод Гаусса-Ньютона.
24. Взаимосвязь персептрона и байесовского классификатора. Многослойный персептрон.
25. Алгоритм обратного распространения ошибки. Извлечение признаков.
26. Линейный дискриминант Фишера. Сети свертки.
27. Теорема Ковера о разделимости множеств. Разделяющая способность поверхности.
28. Задача интерполяции. Теория регуляризации. Функция Грина. Решение задачи регуляризации.
29. Многомерные функции Гаусса.
30. Обобщенные сети на основе радиальных базисных функций. Свойства аппроксимации сетей RBF. Сравнение сетей RBF и многослойных персептронов
31. Анализ признаков на основе самоорганизации. Структура анализа главных компонент. Представление данных. Сокращение размерности.
32. Фильтр Хебба для выделения максимальных собственных значений. Анализ главных компонент на основе правила Хебба
33. Модели отображения признаков. Карты самоорганизации. Процессы конкуренции, кооперации и адаптации. Варианты самоорганизующихся карт.
34. Адаптивные тензорные веса. Самоорганизующиеся карты для символьных строк. Самоорганизующиеся карты с эволюционным обучением.
35. Динамические системы. Пространство состояний. Условие Лившица. Теорема о дивергенции.

36. Устойчивость состояний равновесия. Теоремы Ляпунова. Гиперболические аттракторы. Аддитивная и связанная нейродинамические модели. Модель Хопфилда. Теорема Коэна-Гроссберга.

Критерии оценивания:

Основой для определения баллов, набранных при промежуточной аттестации, служит объем и уровень усвоения материала, предусмотренного рабочей программой дисциплины. При этом необходимо руководствоваться следующим:

- 84-100 баллов (оценка «**отлично**») - изложенный материал фактически верен, наличие глубоких исчерпывающих знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; правильные, уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе, усвоение основной и знакомство с дополнительной литературой;

- 67-83 баллов (оценка «**хорошо**») - наличие твердых и достаточно полных знаний в объеме пройденной программы дисциплины в соответствии с целями обучения, правильные действия по применению знаний на практике, четкое изложение материала, допускаются отдельные логические и стилистические погрешности, обучающийся усвоил основную литературу, рекомендованную в рабочей программе дисциплины;

- 50-66 баллов (оценка «**удовлетворительно**») - наличие твердых знаний в объеме пройденного курса в соответствии с целями обучения, изложение ответов с отдельными ошибками, уверенно исправленными после дополнительных вопросов; правильные в целом действия по применению знаний на практике;

- 0-49 баллов (оценка «**неудовлетворительно**») - ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и «наводящие» вопросы.

Практические задания

Раздел 1. «Математические методы машинного обучения для анализа больших данных»

Тема 1. "Метрические алгоритмы. Байесовский подход к обучению". Обобщенный алгоритм.

Примеры частных алгоритмов. Метод ближайших соседей.

Метод окна Парзена. Понятие выступа объекта. Алгоритм отбора эталонов.

Проклятие размерности. Выбор метрики. Применение формула Байеса к задаче машинного обучения. Построение приближенной плотности распределения: одномерный случай (дискретное, кусочно- постоянное, приближение с помощью ядра, приближение нормальным законом распределения), многомерный случай (наивный байесовский классификатор, приближение с помощью ядра, многомерное нормальное распределение). Логистическая регрессия. Применение Python

Тема 2. "Линейные алгоритмы классификации. Метод опорных векторов". Общая формула линейного классификатора. Метод стохастического градиента. Частные случаи. Обоснование метода СТ. Выступ объекта для линейного классификатора. ROC и AUC. Случаи линейно разделимой и неразделимой выборки. Двойственная задача. Типы объектов. Нелинейное обобщение SVM. SVM-регрессия.

L1 регуляризация

Тема 3. "Методы восстановления регрессии. Логические методы классификации. Композиции алгоритмов". Метод наименьших квадратов. Геометрический смысл. Регуляризация.

Сингулярное разложение. Непараметрическая регрессия. Понятие закономерности. Критерий качества закономерностей. Поиск закономерностей.

Алгоритмы классификации на основе логических закономерностей. AdaBoost. AnyBoost. Градиентный бустинг. Бэггинг и метод случайных подпространств

Тема 4. "Ранжирование и рекомендательные системы. Тематическое моделирование. Кластеризация." Оценки качества. Подходы к решению задачи: поточечный, попарный, списочный. Латентный семантический анализ (LSA). Вероятностный LSA (PLSA). Латентное размещение Дирихле (LDA). EM-алгоритм. Метод k- средних.

DBSCAN

Каждое задание оценивается в 7 баллов

Критерии оценивания:

7 б. – задание выполнено верно;

6 б. – при выполнении задания были допущены неточности, не влияющие на результат;

4-5 б. – при выполнении задания были допущены ошибки;

3-1 б. – при выполнении задания были допущены существенные ошибки.

0 б. – задание не выполнено.

Максимальное количество баллов по практическим заданиям раздела 1 – 28 б.

Раздел 2. Искусственные нейронные сети

Тема 1. "Искусственные нейронные сети". Обучение нейронных сетей. Обучение, основанное на коррекции ошибок. Обучение на основе памяти. Обучение Хебба. Математические модели предложенного Хеббом механизма модификации синаптической связи. Конкурентное обучение. Обучение Больцмана. Обучение с учителем. Обучение с подкреплением. Обучение без учителя. Применение Python

Тема 2. "Однослойные и многослойные перцептроны". Методы безусловной оптимизации. Метод наискорейшего спуска. Метод Ньютона. Метод Гаусса- Ньютона. Взаимосвязь перцептрона и байесовского классификатора. Многослойный перцептрон. Алгоритм обратного распространения ошибки. Извлечение признаков. Линейный дискриминант Фишера. Сети свертки.

Тема 3. "Сети на основе радиальных базисных функций". Теорема Ковера о разделимости множеств. Разделяющая способность поверхности. Задача интерполяции. Теория регуляризации. Функция Грина. Решение задачи регуляризации. Многомерные функции Гаусса. Обобщенные сети на основе радиальных базисных функций. Свойства аппроксимации сетей RBF. Сравнение сетей RBF и многослойных перцептронов

Тема 4. "Анализ главных компонент". Анализ признаков на основе самоорганизации. Структура анализа главных компонент. Представление данных. Сокращение размерности. Фильтр Хебба для выделения максимальных собственных значений. Анализ главных компонент на основе правила Хебба

Тема 5. "Карты самоорганизации Кохонена". Модели отображения признаков. Карты самоорганизации. Процессы конкуренции, кооперации и адаптации. Варианты самоорганизующихся карт. Адаптивные тензорные веса. Самоорганизующиеся карты для символьных строк. Самоорганизующиеся карты с эволюционным обучением. Пакеты программ, реализующие самоорганизующиеся карты. Применение Logitom

Тема 6. "Нейродинамические модели". Динамические системы. Пространство состояний. Условие Лившица. Теорема о дивергенции. Устойчивость состояний равновесия. Теоремы Ляпунова. Гиперболические аттракторы. Аддитивная и связанная нейродинамические модели. Модель Хопфилда. Теорема Коэна- Гроссберга.

Каждое задание оценивается в 7 баллов

Критерии оценивания:

7 б. – задание выполнено верно;

6 б. – при выполнении задания были допущены неточности, не влияющие на результат;

4-5 б. – при выполнении задания были допущены ошибки;

3-1 б. – при выполнении задания были допущены существенные ошибки.

0 б. – задание не выполнено.

Максимальное количество баллов по практическим заданиям раздела 2 – 42 б.

Максимальное количество баллов по практическим заданиям - 70

Перечень вопросов для устного опроса

Раздел 1. Математические методы машинного обучения для анализа больших данных

1. Основные определения: прецедент, обучающая выборка, признаки объектов, виды признаков, матрица объектов-признаков. Модель алгоритмов, метод обучения, функционал качества алгоритма.
2. Вероятностная постановка задачи обучения. Принцип максимума правдоподобия. Связь максимизации правдоподобия и минимизации эмпирического риска.
3. Проблема переобучения и обобщающая способность алгоритма. Состоятельные методы обучения. Эмпирические оценки обобщающей способности.
4. Выбор алгоритма для вероятностной постановки задачи. Функционал среднего риска.
5. Метрические алгоритмы классификации. Обобщенный метрический классификатор. Виды и особенности частных случаев: методы ближайшего соседа, k ближайших соседей, взвешенных соседей, парзеновского окна постоянной и переменной ширины.
6. Классификация объектов по значению отступа. Алгоритм STOLP отбора эталонных объектов. Выбор метрики и проклятие размерности.
7. Приближенное вычисление плотности распределения. Наивный байесовский классификатор. Одномерный случай. Многомерный случай. Смеси распределений. EM- алгоритм разделения смеси. Смеси многомерных нормальных распределений.
8. Линейные алгоритмы классификации. Модель Мак Каллока-Питтса, алгоритм стохастического градиента для минимизации функционала среднего риска. Эвристики для улучшения сходимости и обобщающей способности.
9. Логистическая регрессия. Случайные величины с экспонентным законом распределения. Теорема о линейности байесовского классификатора (с доказательством).
10. Метод опорных векторов (SVM). Случай линейно разделимой выборки. Случай линейно неразделимой выборки. Функция Лагранжа. Классификация объектов в зависимости от значений множителей Лагранжа. Двойственная задача. Обучение SVM.
11. Кривая ошибок и AUC. Формула вычисления AUC. Примеры.
12. Алгоритмы восстановления регрессии. Метод наименьших квадратов. Непараметрическая регрессия. Многомерная линейная регрессия. Подход с использованием SVD-разложения матрицы. Гребневая регрессия. Метод главных компонент PCA.
13. Логические методы классификации. Понятие информативности предиката: эвристическое, вероятностное, энтропийное. Поиск информативных закономерностей. Алгоритмы для номинальных и порядковых признаков. Градиентный алгоритм синтеза конъюнкций. Построение решающего списка и решающего дерева. Редукция деревьев.
14. Композиции алгоритмов. AdaBoost. AnyBoost. Градиентный бустинг. Бэггинг, метод случайных подпространств.
15. Ранжирование и рекомендательные системы. Оценки качества. Алгоритмы построения ранжирующих систем: поточечный, попарный и списочный.
16. Тематическое моделирование. Векторная модель, LSA, PLSA, LDA.
17. Кластеризация. EM-алгоритм, метод k-средних.

Максимальное количество баллов по разделу 1 – 22 б

Раздел 2. Искусственные нейронные сети

1. Обучение нейронных сетей. Обучение, основанное на коррекции ошибок. Обучение на основе памяти.
2. Обучение Хебба. Математические модели предложенного Хеббом механизма модификации синаптической связи. Конкурентное обучение.
3. Обучение Больцмана.
4. Обучение с учителем. Обучение с подкреплением. Обучение без учителя.

5. Методы безусловной оптимизации. Метод наискорейшего спуска. Метод Ньютона. Метод Гаусса- Ньютона.
6. Взаимосвязь персептрона и байесовского классификатора. Многослойный персептрон.
7. Алгоритм обратного распространения ошибки. Извлечение признаков.
8. Линейный дискриминант Фишера. Сети свертки.

9. Теорема Ковера о разделимости множеств. Разделяющая способность поверхности.
10. Задача интерполяции. Теория регуляризации. Функция Грина. Решение задачи регуляризации.
11. Многомерные функции Гаусса.
12. Обобщенные сети на основе радиальных базисных функций. Свойства аппроксимации сетей RBF. Сравнение сетей RBF и многослойных персептронов
13. Анализ признаков на основе самоорганизации. Структура анализа главных компонент. Представление данных. Сокращение размерности.
14. Фильтр Хебба для выделения максимальных собственных значений. Анализ главных компонент на основе правила Хебба
15. Модели отображения признаков. Карты самоорганизации. Процессы конкуренции, кооперации и адаптации. Варианты самоорганизующихся карт.
16. Адаптивные тензорные веса. Самоорганизующиеся карты для символьных строк. Самоорганизующиеся карты с эволюционным обучением.
17. Динамические системы. Пространство состояний. Условие Лившица. Теорема о дивергенции.
18. Устойчивость состояний равновесия. Теоремы Ляпунова. Гиперболические аттракторы. Аддитивная и связанная нейродинамические модели. Модель Хопфилда. Теорема Коэна-Гроссберга.

Максимальное количество баллов по разделу 2 – 8 б.

Критерии оценивания:

Для каждого вопроса:

2 балла - дан полный, развёрнутый ответ на поставленный вопрос, изложение материала при ответе – грамотное и логически стройное;

1 балл - дан неполный ответ на поставленный вопрос

0 баллов - обучающийся не владеет материалом по заданному вопросу.

Максимальное количество баллов по устному опросу – 30

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

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

Текущий контроль успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

Промежуточная аттестация проводится в форме экзамена 2 семестр

Экзамен проводится по окончании теоретического обучения до начала экзаменационной сессии в письменном виде. Количество вопросов в задании – 2. Проверка ответов и объявление результатов производится в день экзамена. Результаты аттестации заносятся в зачетную ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать задолженность в установленном порядке.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

- лекции;
- практические занятия.

В ходе практических работ развиваются навыки применения математических методов, выбора инструментальных средств для обработки и анализа данных в профессиональной деятельности

При подготовке к практическим занятиям каждый студент должен:

- изучить рекомендованную учебную литературу;

В процессе подготовки к практическим занятиям студенты могут воспользоваться консультациями преподавателя.

Теоретические вопросы должны быть изучены студентами в ходе самостоятельной работы. Контроль самостоятельной работы студентов над учебной программой курса осуществляется методом устного опроса и выполнения практических заданий. В ходе самостоятельной работы каждый студент обязан прочитать основную и по возможности дополнительную литературу по изучаемой теме. Выделить непонятные термины, найти их значение в литературе.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами. Также обучающиеся могут взять на дом необходимую литературу на абонементе университетской библиотеки или воспользоваться читальными залами.