

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Горбунов Алексей Александрович

Должность: Заместитель начальника университета по учебной работе

Дата подписания: 12.07.2024 12:04:44

Уникальный программный ключ:

286e49ee1471d400cc1f45539d51ed7bbf0e9cc7

ФГБОУ ВО «Санкт-Петербургский университет ГПС МЧС России»

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ МАШИННОЕ ОБУЧЕНИЕ И АНАЛИЗ ДАННЫХ

Специалитет по специальности

10.05.03 – Информационная безопасность автоматизированных систем

Специализация «Анализ безопасности информационных систем»

1. Цели и задачи дисциплины

Цели освоения дисциплины:

- формирование у обучающихся необходимой теоретической базы и практических навыков, которые позволят всесторонне и системно понимать проблемы обработки и анализа информации;
- формирование навыков разработки и анализа концептуальных и теоретических моделей при решении научных и прикладных задач в области информационных технологий.

Перечень компетенций, формируемых в процессе изучения дисциплины

Компетенции	Содержание
ПК-5	Способен моделировать и исследовать технологии по автоматизации информационно-аналитической деятельности в сфере безопасности

Задачи дисциплины:

- сформировать целостное представление о современных проблемах машинного обучения, а также анализа и обработки больших данных;
- владеть навыками разработки и анализа концептуальных и теоретических моделей прикладных задач анализа больших данных;
- уметь оценивать время и необходимые аппаратные ресурсы для решения задач анализа и обработки данных.

2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Индикаторы достижения компетенции	Планируемые результаты обучения по дисциплине
ПК-5.1. Использует методологические основы и принципы организации информационно-аналитической деятельности; основные принципы и проблематику теории машинного обучения, методы машинного обучения	Знает методологические основы и принципы организации информационно-аналитической деятельности, основные принципы и проблематику теории машинного обучения, методы машинного обучения Умеет ставить задачи и разрабатывать алгоритмы машинного обучения, осуществлять обоснованный выбор методов машинного обучения при решении профессиональных задач
ПК-5.2. Применяет современные методы и средства автоматизированного сбора, обработки и анализа информации в области технологии автоматизации информационно-аналитической деятельности; реализовывает основные алгоритмы теории машинного	Знает возможности инновационных методов анализа данных с использованием машинного обучения, основные подходы к оценке качества работы методов машинного обучения Умеет применять инновационные методы

обучения оценивает их точность и эффективность	анализа данных с использованием машинного обучения, применять простые метрики качества работы методов машинного обучения
--	--

3. Место дисциплины в структуре основной профессиональной образовательной программы

Дисциплина «Машинное обучение и анализ данных» относится к части, формируемой участниками образовательных отношений, образовательной программы специалитета по специальности **10.05.03 – Информационная безопасность автоматизированных систем**, специализация - **Анализ безопасности информационных систем**.

4. Структура и содержание дисциплины

Дисциплина «Машинное обучение и анализ данных» реализуется:

Для очной формы обучения в рамках обязательной части образовательной программы в объеме 180 академических часов (5 зачетных единицы).

4.1 Распределение трудоемкости дисциплины по видам работ по семестрам для очной формы обучения

Вид учебной работы	Трудоемкость		
	з.е.	час.	по семестрам
			9
Общая трудоемкость дисциплины по учебному плану	5	180	180
Контактная работа, в том числе:		92	92
Аудиторные занятия		90	90
Лекции (Л)		26	26
Практические занятия (ПЗ)		64	64
Консультация		2	2
Самостоятельная работа (СРС)		52	52
Экзамен		36	+

4.2 Тематический план, структурированный по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

Наименование разделов и тем	Всего часов	Количество часов по видам занятий				Самостоятельная работа
		Лекции	Практические занятия	Консультации	Контроль	
9 семестр						
Тема 1. Введение в анализ данных и машинное обучение	18	6	4			8
Тема 2. Линейная регрессия	44	4	10			16
Тема 3. Методы классификации	28	4	16			8
Тема 4. Кластерный анализ и поиск аномалий	28	4	14			10
Тема 5. Нейронные сети	38	8	20			10
Консультация	2			2		
Экзамен	36				+	
Всего за 9 семестр	180	26	64	2	36	52

4.3. Содержание дисциплин для обучающихся

Тема 1. Введение в анализ данных и машинное обучение

Лекции. Задачи машинного обучения. Обучение с учителем и обучение без учителя. Классы задач машинного обучения: регрессия, классификация, кластерный анализ, поиск аномалий. Примеры применения машинного обучения в прикладных задачах.

Практические занятия. Решение типовых задач статистической обработки данных. Применение моделей машинного обучения для решения задач регрессии. Применение моделей машинного обучения для решения задач классификации. Решение задач кластерного анализа.

Самостоятельная работа. Основные понятия теории машинного обучения: проблемы, решаемые методами машинного обучения, модели машинного обучения (геометрические, вероятностные, логические), признаки. Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные. Концептуальное обучение: пространство гипотез, поиск в пространстве гипотез, обучаемость, оценка качества решения задачи.

Рекомендуемая литература:

основная [1];

дополнительная [2].

Тема 2. Линейная регрессия

Лекции. Постановка задачи регрессионного анализа. Парная линейная регрессия. Множественная линейная регрессия. Точечный и интервальный прогноз по модели регрессии. Метрики качества.

Практическая подготовка. Решение задач регрессионного анализа. Парная линейная регрессия. Множественная линейная регрессия. Точечный и интервальный прогноз по модели регрессии.

Самостоятельная работа. Применение моделей машинного обучения для решения задач регрессии. Задача восстановления регрессии, метод наименьших квадратов. Линейный регрессионный анализ. Отбор признаков, коллинеарность, влиятельные наблюдения, анализ остатков. Непараметрическая регрессия (ядерное сглаживание). L1 и L2 регуляризация.

Рекомендуемая литература:

основная [1];

дополнительная [1].

Тема 3. Методы классификации

Лекции. Постановка задачи классификации с обучением. Логистическая регрессия. Понятие о деревьях решений. Метрики качества классификации (точность/специфичность, ROC-кривая, площадь под кривой).

Практические занятия. Применение моделей машинного обучения для решения задач классификации. Бинарная и многоклассовая классификация. Вероятностная постановка задачи классификации. Наивный байесовский классификатор. Метрические методы классификации.

Самостоятельная работа. Основные понятия: априорная вероятность, апостериорная вероятность, функция правдоподобия класса. Статистические (байесовские) методы классификации. Обобщённый метрический классификатор. Линейные методы классификации.

Рекомендуемая литература:

основная [2];

дополнительная [1].

Тема 4. Кластерный анализ и поиск аномалий

Лекции. Постановка задачи кластерного анализа. Метод K-средних. Понятие о методах машинного обучения в задачах поиска аномалий. Применение моделей машинного обучения для решения задач кластеризации. Базовые метрики качества, используемые в задачах кластеризации.

Практические занятия. Решение задач кластерного анализа. Метод K-средних. Применение методов машинного обучения в задачах поиска аномалий. Применение моделей машинного обучения для решения задач кластеризации.

Самостоятельная работа. Операторы Map и Reduce. Лямбда-архитектура. Системы управления потоками данных. Системы хранения больших данных.

Обработка данных в реальном времени. Аналитические платформы. Оборудование для обработки больших данных.

Рекомендуемая литература:

основная [2];

дополнительная [1].

Тема 5. Нейронные сети

Лекции. Основные понятия теории нейронных сетей. Архитектура искусственных нейронных сетей. Методы и алгоритмы обучения искусственных нейронных сетей. Ассоциативные запоминающие нейронные сети. Сети с обратными связями.

Практические занятия. Изучение модели нейрона персептрона. Изучение архитектуры персептронной однослойной нейронной сети. Изучение применения многослойной нейронной сети для распознавания изображений. Изучение применения многослойной нейронной сети для предсказания временных процессов.

Самостоятельная работа. Построения сетей различной архитектуры с помощью инструментального программного пакета Neural Network Toolbox системы MATLAB. Полносвязные нейронные сети. Сверточные нейронные сети. Рекуррентные нейронные сети. Самоорганизующихся слои Кохонена. Структура многослойной нейронной сети. Функции активации. Проблема полноты. Задача исключаящего или. Полнота двухслойных сетей в пространстве булевских функций. Алгоритм обратного распространения ошибок. Формирование начального приближения.

Рекомендуемая литература:

основная [3];

дополнительная [3].

5. Методические рекомендации по организации изучения дисциплины

При реализации программы дисциплины «Машинное обучение и анализ данных» используются лекционные и практические занятия.

Общими целями занятий являются:

- обобщение, систематизация, углубление, закрепление теоретических знаний по конкретным темам дисциплины;
- формирование умений применять полученные знания на практике, реализация единства интеллектуальной и практической деятельности;
- выработка при решении поставленных задач профессионально значимых качеств: самостоятельности, ответственности, точности, творческой инициативы.

Целями лекции являются:

– систематизированные основы научных знаний по дисциплине, раскрывать состояние и перспективы развития соответствующей области науки и техники;

– концентрировать внимание обучающихся на наиболее сложных и узловых вопросах;

– стимулировать их активную познавательную деятельность и способствовать формированию творческого мышления.

В ходе практического занятия обеспечивается процесс активного взаимодействия обучающихся с преподавателем; приобретаются практические навыки и умения. Цели практического занятия: выработка практических умений и приобретение навыков, закрепление пройденного материала по соответствующей теме дисциплины.

Самостоятельная работа обучающихся направлена на углубление и закрепление знаний, полученных на лекциях и других занятиях, выработку навыков самостоятельного активного приобретения новых, дополнительных знаний, подготовку к предстоящим занятиям.

6. Оценочные материалы по дисциплине «Машинное обучение и анализ данных»

Текущий контроль успеваемости обеспечивает оценивание хода освоения дисциплины, проводится в соответствии с содержанием дисциплины по видам занятий в форме опроса и тестирования.

Промежуточная аттестация обеспечивает оценивание промежуточных и окончательных результатов обучения по дисциплине, проводится в форме экзамена.

6.1. Примерные оценочные материалы:

6.1.1. Текущего контроля

Типовые вопросы для опроса:

1. В чем смысл кросс-валидации?
2. Что такое наивный байесовский классификатор?
3. Что такое ROC-кривая, как она определяется?
4. Понятие (искусственного) нейрона. Понятие (искусственной) нейронной сети. Понятие функции активации.
5. Требования к функции активации. Виды функций активации.
6. Понятие обучения (настройки) нейронной сети. Понятие обучающего и тестового множеств.

Типовые задания для тестирования:

1. Выберите способ борьбы с переобучением
А) упростить модель (например, уменьшить глубину решающего дерева)

- Б) усложнить модель (например, увеличить глубину решающего дерева)
- В) сменить язык программирования
- Г) сменить модель компьютера

2. Зачем при построении и обучении модели машинного обучения от алгоритма скрывают часть данных?

- А) на этих объектах измеряется качество
- Б) из-за экономической выгоды
- В) позволяет уменьшить отчетность

3. С чего стоит начать решение задачи машинного обучения?

- А) с создания базовой модели
- Б) с придумывания хороших признаков
- В) с создания модели, которая использует все самые современные алгоритмы машинного обучения

4. Определение ранга пожара на основе анализа первичных данных является задачей ...

- А) регрессии
- Б) классификации
- В) кластеризации

5. Кросс-валидация - это ...

А) подход для оценки обобщающей способности алгоритма машинного обучения, при котором все известные данные делят на несколько частей, а потом скрывают по очереди каждую из этих частей, обучая алгоритм на открытых данных и оценивая качество алгоритма на скрытых данных.

Б) функция, измеряющая качество модели машинного обучения.

В) открытая платформа для проведения конкурсов по машинному обучению и предиктивной аналитике.

Г) функция, измеряющая отрицательный эффект от различия между фактическим и заданным курсом.

6. Кластеризация - это ...

А) задача машинного обучения, в которой метки объектов принимают ограниченное число значений, например, город проживания, пол клиента.

Б) задача машинного обучения, в которой метки объектов принимают любое численное значение, например, стоимость квартиры, сумма кредита.

В) задача машинного обучения, заключающаяся в объединении похожих объектов в однородные группы.

7. Процедура обучения применяется:

- А) К конфигурациям из большого числа ИН (нейронным ансамблям)
- Б) К нейронным сетям

- В) К нейроЭВМ
- Г) Ко всем конструкциям, использующих нейроны

8. Среди основных компонентов искусственного нейрона можно выделить следующие:

- А) Весовые коэффициенты
- Б) Сборочные конструкции
- В) Погрешность
- Г) Функция обучения
- Д) Пороговый элемент

9. Обучающие множества должны

- А) Быть достаточно большими
- Б) Содержать необходимую информацию для выявления важных особенностей и зависимостей
- В) Многократно повторяться как в обучающей, так и в тестовой выборке
- Г) Быть представлены в бинарном виде

10. Нейронные сети прямого распространения относят к

- А) Статическим
- Б) Динамическим
- В) Биологическим
- Г) Синаптическим
- Д) Рекуррентным

6.1.2. Промежуточной аттестации

Примерный перечень вопросов, выносимых на экзамен

1. Основные понятия машинного обучения.
2. Обучение с учителем и обучение без учителя.
3. Классы задач машинного обучения: регрессия, классификация, кластерный анализ.
4. Постановка задачи регрессионного анализа.
5. Парная линейная регрессия.
6. Множественная линейная регрессия.
7. Точечный и интервальный прогноз по модели регрессии.
8. Постановка задачи классификации с обучением.
9. Линейные алгоритмы классификации.
10. Логистическая регрессия.
11. Понятие о деревьях решений.
12. Метрики качества классификации
13. Постановка задачи кластерного анализа.

14. Метод К-средних
15. Понятие о методах машинного обучения в задачах поиска аномалий.
16. Метрики качества, используемые в задачах кластеризации.
17. Понятие нейронной сети.
18. Виды нейронных сетей.
19. Многослойный персептрон.
20. Методы обучения нейронной сети.
21. Метод градиентного спуска
22. Метод обратного распространения ошибки.
23. Сеть Кохонена
24. Нейронные сети с обратными связями, сеть Хопфилда.

6.2. Шкала оценивания результатов промежуточной аттестации и критерии выставления оценок

Система оценивания включает:

Форма контроля	Показатели оценивания	Критерии выставления оценок	Шкала оценивания
экзамен	правильность и полнота ответа	дан правильный, полный ответ на поставленный вопрос, показана совокупность осознанных знаний по дисциплине, доказательно раскрыты основные положения вопросов; могут быть допущены недочеты, исправленные самостоятельно в процессе ответа.	отлично
		дан правильный, недостаточно полный ответ на поставленный вопрос, показано умение выделить существенные и несущественные признаки, причинно-следственные связи; могут быть допущены недочеты, исправленные с помощью преподавателя.	хорошо
		дан недостаточно правильный и полный ответ; логика и последовательность изложения имеют нарушения; в ответе отсутствуют выводы.	удовлетворительно
		ответ представляет собой разрозненные знания с существенными ошибками по вопросу; присутствуют фрагментарность, нелогичность изложения; дополнительные и уточняющие вопросы не приводят к	неудовлетворительно

7. Ресурсное обеспечение дисциплины

7.1. Лицензионное и свободно распространяемое программное обеспечение

Перечень лицензионного и свободно распространяемого программного обеспечения:

- Статистическая диалоговая система STADIA [ПО-6FF-561] - Статистическая диалоговая система [Лицензионное. Номер в Едином реестре российских программ для электронных вычислительных машин и баз данных - 9064]

- SMath Studio [ПО-А68-516] - Программное обеспечение для вычисления математических выражений и построения графиков функций [Свободно распространяемое. Номер в Едином реестре российских программ для электронных вычислительных машин и баз данных - 12849]

- МойОфис Образование [ПО-41В-124] - Полный комплект редакторов текстовых документов и электронных таблиц, а также инструментарий для работы с графическими презентациями [Свободно распространяемое. Номер в Едином реестре российских программ для электронных вычислительных машин и баз данных - 4557]

- Astra Linux Common Edition релиз Орел [ПО-25В-603] - Операционная система общего назначения "Astra Linux Common Edition" [Коммерческая (Full Package Product). Номер в Едином реестре российских программ для электронных вычислительных машин и баз данных - 4433]

7.2. Профессиональные базы данных и информационные справочные системы

Информационная справочная система — Сервер органов государственной власти Российской Федерации <http://россия.рф/> (свободный доступ); профессиональные базы данных — Портал открытых данных Российской Федерации <https://data.gov.ru/> (свободный доступ); федеральный портал «Российское образование» <http://www.edu.ru> (свободный доступ); система официального опубликования правовых актов в электронном виде <http://publication.pravo.gov.ru/> (свободный доступ); федеральный портал «Совершенствование государственного управления» <https://ar.gov.ru> (свободный доступ); электронная библиотека университета <http://elib.igps.ru> (авторизованный доступ); электронно-библиотечная система «ЭБС IPR BOOKS» <http://www.iprbookshop.ru> (авторизованный доступ).

7.3. Литература

Основная литература:

1. 1. Ракитский А.А. Методы машинного обучения : учебно-методическое пособие / Ракитский А.А.. — Новосибирск : Сибирский государственный уни-

верситет телекоммуникаций и информатики, 2018. — 32 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/90591.html>

2. Воронина, В. В. Теория и практика машинного обучения : учебное пособие / В. В. Воронина. — Ульяновск : УлГТУ, 2017. — 290 с. — ISBN 978-5-9795-1712-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165053>

3. Барский А.Б. Введение в нейронные сети : учебное пособие / Барский А.Б.. — Москва, Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 357 с. — ISBN 978-5-4497-0309-5. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/89426.html>

Дополнительная литература:

1. Сараев П.В. Методы машинного обучения : методические указания и задания к лабораторным работам по курсу / Сараев П.В.. — Липецк : Липецкий государственный технический университет, ЭБС АСВ, 2017. — 48 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/83183.html>

2. Мельниченко А.С. Математическая статистика и анализ данных : учебное пособие / Мельниченко А.С.. — Москва : Издательский Дом МИСиС, 2018. — 45 с. — ISBN 978-5-906953-62-9. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/78563.html>

3. Вакуленко С.А. Нейронные сети : учебное пособие / Вакуленко С.А., Жихарева А.А.. — Санкт-Петербург : Санкт-Петербургский государственный университет промышленных технологий и дизайна, 2019. — 110 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/102447.html>

7.4 Материально-техническое обеспечение дисциплины

Для проведения и обеспечения занятий используются помещения, которые представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой специалитета, оснащенные оборудованием и техническими средствами обучения: автоматизированное рабочее место преподавателя, маркерная доска, мультимедийный проектор, документ-камера, посадочные места обучающихся.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде университета.

Авторы: канд. техн. наук, доцент Матвеев А.В., канд. техн. наук, доцент
Максимов А.В.