ИНСТИТУТ ИНТЕЛЛЕКТУАЛЬНЫХ КИБЕРНЕТИЧЕСКИХ СИСТЕМ

КАФЕДРА КИБЕРНЕТИКИ

ОДОБРЕНО  
  
протокол № 18 / 03   
  
от « 31 » мая 2018 г.

**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ**

МАШИННОЕ ОБУЧЕНИЕ

|  |  |
| --- | --- |
| Направление подготовки (специальность) | 09.04.04 Программная инженерия |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Семестр** | **Интерактив** | **Трудоемкость, кред.** | **Общий объем курса, час.** | **Лекции, час.** | **Практич. занятия, час.** | **Лаборат. работы, час.** | **СРС, час.** | **КСР, час.** | **Форма(ы) контроля, экз./зач./КР/КП** |
| 3 |  | 5 | 180 | 16 | 0 | 32 | 96 | 0 | Э КР |
| ИТОГО | 0 | 5 | 180 | 16 | 0 | 32 | 96 | 0 |  |

Группа: М20-504

АННОТАЦИЯ

Машинное обучение является современным подходом к созданию систем, которые в процессе работы способны вырабатывать решения не по заранее запрограммированным правилам, а на основе предварительного анализа располагаемых данных. Применение методов машинного обучения позволяет создавать интеллектуальные системы управления, распознавания образов, прогнозирования во всех сферах деятельности человека. В курсе студенты знакомятся с основами машинного обучения: постановкой задачи обучения, подготовкой данных, принципами обучения и принятия решения, подходами к организации обучения и валидации результатов, методами и алгоритмами классификации и кластеризации данных. Теоретические основы машинного обучения закрепляются на лабораторных занятиях и используются при выполнении самостоятельного исследования в рамках индивидуальной курсовой работы.

1. ЦЕЛИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Целью освоения учебной дисциплины «Машинное обучение» является овладение будущими специалистами теоретическими знаниями и практическими подходами к построению систем, способных принимать адекватные интеллектуальные решения, сравнимые по качеству и точности с решениями, которые принимает человек. Необходимость создания таких систем возрастает. В то же время растет и возможность их реализации в связи с развитием технологической основы вычислительных систем и средств коммуникаций. Изучение основных задач машинного обучения, методов их решения, алгоритмического и программного обеспечения необходимо будущим специалистам в области программной инженерии при создании современных программных систем, предназначенных для интеллектуальной обработки больших объемов данных в условиях неполноты или отсутствия их формального описания.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Дисциплина «Машинное обучение» относится к общенаучному циклу М1 образовательной программы и является обязательной для магистранта. Дисциплина требует от слушателя общематематической подготовки по дискретной математике, математическому анализу, линейной алгебре, дифференциальным уравнениям, методам оптимизации, а также углубленного знания теории вероятностей и математической статистики. Полученные знания могут быть применены студентами в процессе обучения при выполнении текущей научно-исследовательской и выпускной работ. Дисциплина формирует систему базовых понятий, необходимых для специалиста в области программной инженерии, прикладной математики и информатики, способствует выработке профессиональных решений практических задач.

3. КОМПЕТЕНЦИИ СТУДЕНТА, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ/ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ ОБРАЗОВАНИЯ И КОМПЕТЕНЦИИ СТУДЕНТА ПО ЗАВЕРШЕНИИ ОСВОЕНИЯ ПРОГРАММЫ УЧЕБНОЙ ДИСЦИПЛИНЫ

ОПК-1 – Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте

ОПК-2 – Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач

ОПК-4 – Способен применять на практике новые научные принципы и методы исследований

ОПК-5 – Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем

ПК-14 – Владеет навыками программной реализации систем с параллельной обработкой данных и высокопроизводительных систем

ПК-15 – Владеет навыками создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов

ПК-19 – Владеет навыками создания систем обработки текстов

ПК-4 – Владеет существующими методами и алгоритмами решения задач распознавания и обработки данных

ПК-5 – Владеет существующими методами и алгоритмами решения задач цифровой обработки сигналов

4. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **№ п.п** | **Наименование раздела учебной дисциплины** | **Недели** | **Лекции, час.** | **Практ. занятия / семинары, час.** | **Лабораторные работы, час.** | **Обязат. текущий контроль (форма\*, неделя)** | **Аттестация раздела (форма\*, неделя)** | **Максимальный балл за раздел\*\*** |
|  | *3 Семестр* |  |  |  |  |  |  |  |
| 1 | Постановка задачи машинного обучения. Препроцессинг данных. Методы классификации данных. Композиции методов классификации. | 1-8 | 8 |  | 16 | КР-8 | к.р-8 | 30 |
| 2 | Обучение без учителя. Кластеризация данных. Решение задач машинного обучения на многослойных нейронных сетях. Глубокое обучение. | 9-16 | 8 |  | 16 | КР-16 | к.р-16 | 30 |
|  | *Итого за 3 Семестр* |  | 16 | 0 | 32 |  |  | 60 |
|  | **Контрольные мероприятия за 3 Семестр** |  |  |  |  |  | Э КР | 40 |

\* – сокращенное наименование формы контроля

\*\* – сумма максимальных баллов должна быть равна 100 за семестр, включая зачет и (или) экзамен

Сокращение наименований форм текущего контроля и аттестации разделов:

|  |  |
| --- | --- |
| **Обозначение** | **Полное наименование** |
| к.р | Контрольная работа |
| Э | Экзамен |
| КР | Курсовая работа |

КАЛЕНДАРНЫЙ ПЛАН

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Недели** | **Темы занятий / Содержание** | **Лек., час.** | **Пр./сем., час.** | **Лаб., час.** |
|  | *3 Семестр* | 16 | 0 | 32 |
| **1-8** | **Постановка задачи машинного обучения. Препроцессинг данных. Методы классификации данных. Композиции методов классификации.** | 8 |  | 16 |
| 1 - 2 | **Постановка задачи машинного обучения. Препроцессинг данных. Метод главных компонент.** Типы задач машинного обучения. Примеры прикладных задач. Основные понятия и определения. Функционал качества. Объекты и признаки. Методы предобработки исходных данных. Метод главных компонент. Проблема переобучения и понятие обобщающей способности. Методика тестирования обучаемых алгоритмов. Кросс-валидация. Приёмы генерации модельных данных. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 2 |  | 4 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 3 - 6 | **Задачи классификации данных. Логические и метрические методы. Метод опорных векторов.** Постановка задачи классификации. Байесовский подход к классификации. Наивный байесовский классификатор. Нормальный дискриминантный анализ. Линейный дискриминант Фишера. Метод опорных векторов. Метод ближайших соседей. Деревья решений. Бинарная классификация. Оценка точности классификации. Специфичность и чувствительность. ROC-анализ. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 4 |  | 8 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 7 - 8 | **Композиции методов классификации. Повышение точности решения.** Подходы к построению ансамблей классификаторов. Базовый классификатор и решающее правило. Бустинг, стекинг, бэггинг. Бутстреп-метод. Алгоритм AdaBoost. Случайный лес. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 2 |  | 4 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| **9-16** | **Обучение без учителя. Кластеризация данных. Решение задач машинного обучения на многослойных нейронных сетях. Глубокое обучение.** | 8 |  | 16 |
| 9 - 11 | **Обучение без учителя. Кластеризация данных. Оценка валидности кластеризации.** Постановка задачи кластеризации данных. Агломеративные методы кластеризации. Метод Уорда. Представление результатов иерархической кластеризации. Дендрограмма. Метод k-средних. Статистический анализ качества кластеризации данных. Силуэты кластеров, индексы Данна и ДэвисаБолдина. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 3 |  | 6 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 12 - 14 | **Решение задач машинного обучения на многослойных нейронных сетях.** Математическая модель многослойной нейронной сети. Многослойный персептрон. Постановка задачи обучения. Метод обратного распространения ошибки. Применение многослойного персептрона для решения задач машинного обучения с учителем. Самообучающиеся нейронные сети. Сеть Кохонена. Карта Кохонена. Особенности применения самообучающихся нейронных сетей для решения задач кластеризации данных. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 3 |  | 6 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 15 - 16 | **Новый взгляд на задачи машинного обучения. Глубокое обучение.** Проблема формирования вектора признаков. Сети глубокого доверия. Метод обучения. Предобучение слоёв глубокой сети. Автоэнкодеры. Ограниченная машина Больцмана. Свёрточные нейронные сети. Математическая модель сети. Ядро свёртки. Особенности архитектуры свёрточных сетей. Области практического применения. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 2 |  | 4 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |

Сокращенные наименования онлайн опций:

|  |  |
| --- | --- |
| **Обозначение** | **Полное наименование** |
| ЭК | Электронный курс |
| ПМ | Полнотекстовый материал |
| ПЛ | Полнотекстовые лекции |
| ВМ | Видео-материалы |
| АМ | Аудио-материалы |
| Прз | Презентации |
| Т | Тесты |
| ЭСМ | Электронные справочные материалы |
| ИС | Интерактивный сайт |

ТЕМЫ ЛАБОРАТОРНЫХ РАБОТ

|  |  |
| --- | --- |
| **Недели** | **Темы занятий / Содержание** |
|  | *3 Семестр* |
| 1 - 2 | **Методы предобработки исходных данных. Метод главных компонент.**  Методы предобработки исходных данных. Метод главных компонент. |
| 3 - 6 | **Метрические и логические методы классификации.**  Метрические и логические методы классификации. |
| 7 - 8 | **Композиции методов классификации.**  Композиции методов классификации. |
| 9 - 11 | **Методы кластеризации данных.**  Методы кластеризации данных. |
| 12 - 14 | **Решение задач машинного обучения на многослойных нейронных сетях.** Решение задач машинного обучения на многослойных нейронных сетях. |
| 15 - 16 | **Методы глубокого обучения.** Методы глубокого обучения. |

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Занятия проводятся по классической схеме чтения лекций и проведения лабораторных занятий. Контроль самостоятельной работы и освоения изученной темы проводится на лабораторных занятиях. В процессе контрольного опроса (КО) каждый слушатель должен ответить на вопросы преподавателя или провести краткие расчеты по предложенным задачам. Результаты контрольного опроса оцениваются в баллах. Для усвоения методов и алгоритмов машинного обучения студенты выполняют исследовательскую курсовую работу по индивидуальному заданию. В процессе выполнения курсовой работы студенты знакомятся с дополнительной литературой по теме проекта. Преподаватель осуществляет текущий контроль результатов исследовательской работы студента и учитывает их в течение семестра в оценках за разделы курса. Курсовая работа студента, представленная в форме Пояснительной записки, содержит описание всех этапов выполнения задания и анализ полученных результатов. Самостоятельная работа студента включает: 1) ознакомление с литературой по темам курса, изучение электронных учебников или курсов по машинному обучению и анализу данных; 2) подготовку к лабораторным занятиям; 3) изучение прикладных программ пакетов, используемых в лабораторном практикуме.

6. ТРЕБОВАНИЯ К ФОНДУ ОЦЕНОЧНЫХ СРЕДСТВ В РАМКАХ РЕАЛИЗУЕМОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

В качестве оценочного средства используется 100 балльная семестровая система, учитывающая, кроме уровня подготовки студента и результатов выполнения курсовой работы, посещаемость лекций и лабораторных занятий, активность (ответы на вопросы при проведении контрольного опроса).

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

ОСНОВНАЯ ЛИТЕРАТУРА:

1. ЭИ Ф 70 Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : , Москва: ДМК Пресс, 2015

2. ЭИ Ч-45 Применение искусственных нейронных сетей и системы остаточных классов в криптографии : учебное пособие, Москва: Физматлит, 2012

3. ЭИ К 75 Самоорганизующиеся карты : , Москва: Бином. Лаборатория знаний, 2014

4. ЭИ М71 Лабораторный практикум по курсу "Введение в теорию нейронных сетей" : , О. А. Мишулина, А. Г. Трофимов, М. В. Щербинина, Москва: МИФИ, 2007

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ:

1. MATLAB & Simulink Student Version ()

LMS И ИНТЕРНЕТ-РЕСУРСЫ:

1. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. Third Edition. Elsevier, 2012. (http://www.cse.hcmut.edu.vn/~chauvtn/data\_mining/Texts/%5B1%5D%20Data%20Mining%20%20Concepts%20and%20Techniques%20%283rd%20Ed%29.pdf)

2. Yoshua Bengio. Learning Deep Architectures for AI // Foundations and Trends in Machine Learning: Vol (http://www.nowpublishers.com/article/Details/MAL-006)

3. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, (http://www.springer.com/us/book/9780387848570 )

4. К. В. Воронцов. Математические методы обучения по прецедентам (теория обучения машин). (http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf)

5. Bishop C.M. Neural Networks for Pattern Recognition. Oxford, 1994. (http://cs.du.edu/~mitchell/mario\_books/Neural\_Networks\_for\_Pattern\_Recognition\_\_Christopher\_Bishop.pdf)

6. Хайкин С. Нейронные сети: Полный курс. – М.: Издательский дом «Вильямс», 2006. (http://www.twirpx.com/file/57977/)

7. Aggarwal, C.C. Data Mining. The Textbook. Springer, XXIX, 2015. (http://www.springer.com/978-3-319-14141-1)

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Специальное материально-техническое обеспечение не требуется

9. ##Definition not found: 'static\_section\_edu\_stud'##

В качестве оценочного средства используется 100 балльная семестровая система, учитывающая, кроме уровня подготовки студента и результатов выполнения курсовой работы, посещаемость лекций и лабораторных занятий, активность (ответы на вопросы при проведении контрольного опроса).

Рабочая программа дисциплины составлена в соответствии с ОС НИЯУ МИФИ (ФГОС) и учебным планом основной образовательной программы (программ).

Автор(ы):

|  |  |  |
| --- | --- | --- |
|  | Трофимов Александр Геннадьевич, к.т.н. |  |