

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Институт вычислительной математики и информационных технологий



УТВЕРЖДАЮ

Проректор
по образовательной деятельности КФУ
Проф. Таюрский Д.А.

"__" _____ 20__ г.

Программа дисциплины
Машинное обучение Б1.В.ОД.7

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Миссаров М.Д.

Рецензент(ы):

Кашина О.А.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Миссаров М. Д.

Протокол заседания кафедры No ____ от " ____ " _____ 201__ г

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

Протокол заседания УМК No ____ от " ____ " _____ 201__ г

Регистрационный No

Казань
2016

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) заведующий кафедрой, д.н. (доцент) Миссаров М.Д. кафедра анализа данных и исследования операций отделение фундаментальной информатики и информационных технологий, Moukadas.Missarov@kpfu.ru

1. Цели освоения дисциплины

Данная учебная дисциплина реализуется как цикл лекционных и практических занятий, которые знакомят студентов с теоретическими основами и алгоритмами машинного обучения, их возможными практическими реализациями и применением при решении реальных задач. В рамках данного курса студенты должны получить представление о задачах, решаемых с помощью рассматриваемой теории, и принципах построения некоторых основных классификаторов

2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " Б1.В.ОД.7 Дисциплины (модули)" основной образовательной программы 01.04.02 Прикладная математика и информатика и относится к обязательные дисциплины. Осваивается на 1 курсе, 2 семестр.

Данная учебная дисциплина включена в раздел " М2.ДВ.5 Профессиональный" основной образовательной программы 010400.68 Прикладная математика и информатика и относится к дисциплинам по выбору. Осваивается на 1 курсе, 2 семестр.

Цикл: М2.ДВ5.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

| Шифр компетенции | Расшифровка приобретаемой компетенции |
|---|--|
| ОК-6 (общекультурные компетенции) | (общекультурные компетенции) способность использовать нормативные правовые документы в своей деятельности, проявлять настойчивость в достижении цели с учетом моральных и правовых норм и обязанности |
| ОК-7 (общекультурные компетенции) | способность владеть одним из иностранных языков на уровне не ниже разговорного |
| ОК-8 (общекультурные компетенции) | способность самостоятельно, методически правильно использовать методы физического воспитания и укрепления здоровья, готовность к достижению должного уровня физической подготовленности для обеспечения полноценной социальной и профессиональной деятельности |
| ОК-9 (общекультурные компетенции) | способность осознать социальную значимость своей будущей профессии, обладать высокой мотивацией к выполнению профессиональной деятельности |
| ПК-10 (профессиональные компетенции) | способность применять в профессиональной деятельности современные языки программирования и языки баз данных, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии |
| ПК-11 (профессиональные компетенции) | способность приобретать и использовать организационно-управленческие навыки в профессиональной и социальной деятельности |

| Шифр компетенции | Расшифровка приобретаемой компетенции |
|--|---|
| ПК-8 (профессиональные компетенции) | способность формировать суждения о значении и последствиях своей профессиональной деятельности с учетом социальных, профессиональных и этических позиций |
| ПК-9 (профессиональные компетенции) | способность решать задачи производственной и технологической деятельности на профессиональном уровне, включая: разработку алгоритмических и программных решений в области системного и прикладного программирования |

В результате освоения дисциплины студент:

В результате освоения дисциплины обучающийся должен:

Знать:

- принципы построения векторов признаков, решающих правил и классификации;
- основные виды классификаторов;
- принципы построения линейных классификаторов;
- принципы построения нелинейных классификаторов;
- особенности выбора признаков классификации и предварительной обработки данных.

Уметь:

- выбирать подходящий вид классификатора в зависимости от решаемой задачи;
- выбирать набор признаков для классификации и проводить предварительную обработку данных;
- уметь применять алгоритмы построения и обучения классификатора по выборке;
- выполнять вычисления, связанные с обучением и работой классификатора, в среде MATLAB

- .

Владеть:

- навыками выбора, построения, обучения и использования основных классификаторов при решении задач

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 4 зачетных(ые) единиц(ы) 144 часа(ов).

Форма промежуточного контроля дисциплины зачет во 2 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

Тематический план дисциплины/модуля

| N | Раздел Дисциплины/ Модуля | Семестр | Неделя семестра | Виды и часы аудиторной работы, их трудоемкость (в часах) | | | Текущие формы контроля |
|----|--|---------|--------------------|---|-------------------------|------------------------|---------------------------|
| | | | | Лекции | Практические занятия | Лабораторные работы | |
| 1. | Тема 1. Тема 1. Основные понятия и определения. Примеры прикладных задач | 2 | 1-2 | 0 | 0 | 6 | домашнее задание |
| 2. | Тема 2. Тема 2. Линейные классификаторы | 2 | 3-4 | 0 | 0 | 6 | домашнее задание |
| 3. | Тема 3. Тема 3. Метод опорных векторов | 2 | 5-6 | 0 | 0 | 6 | контрольная работа |
| 4. | Тема 4. Тема 4. Методы восстановления регрессии | 2 | 7-8 | 0 | 0 | 6 | домашнее задание |
| 5. | Тема 5. Тема 5. Искусственные нейронные сети | 2 | 9-11 | 0 | 0 | 8 | контрольная работа |
| 6. | Тема 6. Тема 6. Выбор признаков и подготовка данных | 2 | 12-13 | 0 | 0 | 6 | домашнее задание |
| 7. | Тема 7. Тема 7. Контекстно-зависимая классификация | 2 | 14-15 | 0 | 0 | 4 | домашнее задание |
| | Тема . Итоговая форма контроля | 2 | | 0 | 0 | 0 | зачет |
| | Итого | | | 0 | 0 | 42 | |

4.2 Содержание дисциплины

Тема 1. Тема 1. Основные понятия и определения. Примеры прикладных задач

лабораторная работа (6 часа(ов)):

Признаки, вектора признаков. Объекты, классы. Классификация. Классификатор. Обучение, виды обучения "с учителем" и "без учителя". Разбор примеров прикладных задач.

Тема 2. Тема 2. Линейные классификаторы

лабораторная работа (6 часа(ов)):

Разбор примеров и решение задач по темам: линейная модель классификации, метод стохастического градиента, алгоритм Персептрона.

Тема 3. Тема 3. Метод опорных векторов

лабораторная работа (6 часа(ов)):

Основы метода опорных векторов. Случай линейно разделимой выборки. Случай линейно неразделимой выборки. Ядра и спрямляющие пространства. Разбор примеров и решение задач.

Тема 4. Тема 4. Методы восстановления регрессии

лабораторная работа (6 часа(ов)):

Метод наименьших квадратов. Непараметрическая регрессия: ядерное сглаживание. Линейная регрессия. Метод главных компонент. Разбор примеров и решение задач по этим темам.

Тема 5. Искусственные нейронные сети

лабораторная работа (8 часа(ов)):

Проблема полноты. Задача исключаящего "или". Вычислительные возможности двух- и трехслойных сетей. Метод обратного распространения ошибки. Изучение на лабораторном занятии алгоритма построения нейронных сетей.

Тема 6. Выбор признаков и подготовка данных

лабораторная работа (6 часа(ов)):

Влияние выбора набора признаков на результаты классификации. Предварительная обработка данных. Недостающие значения. Выбор признаков на основе проверки гипотез. Выбор подмножества признаков.

Тема 7. Контекстно-зависимая классификация

лабораторная работа (4 часа(ов)):

Марковские цепи. Алгоритм Витерби. Скрытые марковские модели. Применение в задачах распознавания голоса. Решение задач по теории марковских моделей в машинном обучении.

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

| N | Раздел Дисциплины | Семестр | Неделя семестра | Виды самостоятельной работы студентов | Трудоемкость (в часах) | Формы контроля самостоятельной работы |
|----|--|---------|-----------------|---------------------------------------|------------------------|---------------------------------------|
| 1. | Тема 1. Основные понятия и определения. Примеры прикладных задач | 2 | 1-2 | подготовка домашнего задания | 14 | домашнее задание |
| 2. | Тема 2. Линейные классификаторы | 2 | 3-4 | подготовка домашнего задания | 14 | домашнее задание |
| 3. | Тема 3. Метод опорных векторов | 2 | 5-6 | подготовка к контрольной работе | 20 | контрольная работа |
| 4. | Тема 4. Методы восстановления регрессии | 2 | 7-8 | подготовка домашнего задания | 12 | домашнее задание |
| 5. | Тема 5. Искусственные нейронные сети | 2 | 9-11 | подготовка к контрольной работе | 14 | контрольная работа |
| 6. | Тема 6. Выбор признаков и подготовка данных | 2 | 12-13 | подготовка домашнего задания | 14 | домашнее задание |
| 7. | Тема 7. Контекстно-зависимая классификация | 2 | 14-15 | подготовка домашнего задания | 14 | домашнее задание |
| | Итого | | | | 102 | |

5. Образовательные технологии, включая интерактивные формы обучения

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

Тема 1. Тема 1. Основные понятия и определения. Примеры прикладных задач

домашнее задание , примерные вопросы:

Выполнение домашнего задания. Подготовка к лабораторным занятиям. Доработка заданий, выполняемых на лабораторных занятиях.

Тема 2. Тема 2. Линейные классификаторы

домашнее задание , примерные вопросы:

Выполнение домашнего задания. Подготовка к лабораторным занятиям. Доработка заданий, выполняемых на лабораторных занятиях.

Тема 3. Тема 3. Метод опорных векторов

контрольная работа , примерные вопросы:

Подготовка к контрольной работе.

Тема 4. Тема 4. Методы восстановления регрессии

домашнее задание , примерные вопросы:

Выполнение домашнего задания. Подготовка к лабораторным занятиям. Доработка заданий, выполняемых на лабораторных занятиях.

Тема 5. Тема 5. Искусственные нейронные сети

контрольная работа , примерные вопросы:

Подготовка к контрольной работе.

Тема 6. Тема 6. Выбор признаков и подготовка данных

домашнее задание , примерные вопросы:

Выполнение домашнего задания. Подготовка к лабораторным занятиям. Доработка заданий, выполняемых на лабораторных занятиях.

Тема 7. Тема 7. Контекстно-зависимая классификация

домашнее задание , примерные вопросы:

Выполнение домашнего задания. Подготовка к лабораторным занятиям. Доработка заданий, выполняемых на лабораторных занятиях.

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

Примерные вопросы для зачета:

7.1. Основная литература:

1. Ярушкина Н. Г. Интеллектуальный анализ временных рядов: Учебное пособие / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. - М.: ИД ФОРУМ: ИНФРА-М, 2012. - 160 с.: <http://znanium.com/bookread.php?book=249314>
2. Компьютерные технологии анализа данных в эконометрике / Д.М. Дайитбегов. - 2-е изд., испр. и доп. - М.: Вузовский учебник: ИНФРА-М, 2010. - 578 с.: 70x100 1/16. - (Научная книга). (переплет) ISBN 978-5-9558-0191-9 <http://www.znanium.com/bookread.php?book=251791>
3. Информационные аналитические системы [Электронный ресурс] : учебник / Т. В. Алексеева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - М.: МФПУ Синергия, 2013. - 384 с. - (Университетская серия). - ISBN 978-5-4257-0092-6. <http://www.znanium.com/bookread.php?book=451186>
4. Кашина О.А., Миссаров М.Д. Электронный курс "Анализ данных в среде R", 2013

<http://zilant.kpfu.ru/course/view.php?id=17341>

5. Кашина О.А., Миссаров М.Д. Электронный курс

"Статистический анализ данных", 2013

<http://zilant.kpfu.ru/course/view.php?id=17260>

7.2. Дополнительная литература:

Технология Data Mining: Интеллектуальный анализ данных, Степанов, Роман Григорьевич, 2009г.

2. Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С.

Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. ? 3-е изд., перераб. и доп. ? СПб.:

БХВ-Петербург, 2009. ? 512 с.: ил. + CD-ROM ? (Учебная литература для вузов).

<http://www.znaniyum.com/bookread.php?book=350638>

3. Математические методы анализа дискретных структур генетического кода/Гупал В.М. - М.: ИЦ РИОР, НИЦ ИНФРА-М, 2015. - 334 с.: 60x90 1/16. - (Научная мысль) ISBN

978-5-369-01462-2

<http://znaniyum.com/bookread2.php?book=516085>

7.3. Интернет-ресурсы:

Data Mining Labs - <http://dmlabs.org/> al Computing

The R Project for Statistical Computing - <http://www.r-project.org/>

курс - <http://www.intuit.ru/studies/courses/2312/612/info>

курс - <http://www.intuit.ru/studies/courses/6/6/info>

программный комплекс RStudio для интеллектуального анализа данных - <http://www.rstudio.com/>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Машинное обучение" предполагает использование следующего материально-технического обеспечения:

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен студентам. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, УМК, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

Лабораторные занятия проводятся в компьютерном классе.

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 01.04.02 "Прикладная математика и информатика" и магистерской программе Анализ данных и его приложения .

Автор(ы):

Миссаров М.Д. _____

"__" _____ 201__ г.

Рецензент(ы):

Кашина О.А. _____

"__" _____ 201__ г.