

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

Введение в машинное обучение Б1.В.ДВ.1

Направление подготовки: 38.04.05 - Бизнес-информатика

Профиль подготовки: Математические методы и информационные технологии в бизнесе

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.ДВ.1 Дисциплины (модули)" основной образовательной программы 38.04.05 Бизнес-информатика и относится к дисциплинам по выбору. 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 7 зачетных(ые) единиц(ы) 252 часа(ов).

Форма промежуточного контроля дисциплины экзамен 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	3	6	0	домашнее задание
2.	Тема 2. Тема 2. Линейные классификаторы	2	3-4	3	6	0	домашнее задание
3.	Тема 3. Тема 3. Метод опорных векторов	2	5-6	3	6	0	контрольная работа
4.	Тема 4. Тема 4. Методы восстановления регрессии	2	7-8	2	4	0	домашнее задание
5.	Тема 5. Тема 5. Искусственные нейронные сети	2	9-11	3	6	0	контрольная работа
6.	Тема 6. Тема 6. Выбор признаков и подготовка данных	2	12-13	2	4	0	домашнее задание
7.	Тема 7. Тема 7. Контекстно-зависимая классификация	2	14-15	2	4	0	домашнее задание
	Тема . Итоговая форма контроля	2		0	0	0	экзамен
	Итого			18	36	0	

4.2 Содержание дисциплины

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

лекционное занятие (3 часа(ов)):

Признаки, вектора признаков. Объекты, классы. Классификация. Классификатор. Обучение, виды обучения ?с учителем? и ?без учителя?.

практическое занятие (6 часа(ов)):

Разбор примеров прикладных задач.

Тема 2. Тема 2. Линейные классификаторы

лекционное занятие (3 часа(ов)):

Линейная модель классификации. Метод стохастического градиента. Алгоритм Персептрона.

практическое занятие (6 часа(ов)):

Разбор примеров и решение задач по теме "Линейные классификаторы".

Тема 3. Тема 3. Метод опорных векторов

лекционное занятие (3 часа(ов)):

Основы метода опорных векторов. Случай линейно разделимой выборки. Случай линейно неразделимой выборки. Ядра и спрямляющие пространства.

практическое занятие (6 часа(ов)):

Разбор примеров и решение задач по темам: Случай линейно разделимой выборки. Случай линейно неразделимой выборки. Ядра

Тема 4. Методы восстановления регрессии

лекционное занятие (2 часа(ов)):

Метод наименьших квадратов. Непараметрическая регрессия: ядерное сглаживание. Линейная регрессия. Метод главных компонент.

практическое занятие (4 часа(ов)):

Решение задач по теме "Методы восстановления регрессии".

Тема 5. Искусственные нейронные сети

лекционное занятие (3 часа(ов)):

Проблема полноты. Задача исключаящего "или". Вычислительные возможности двух- и трехслойных сетей. Метод обратного распространения ошибки.

практическое занятие (6 часа(ов)):

Изучение алгоритмов построения нейронных сетей.

Тема 6. Выбор признаков и подготовка данных

лекционное занятие (2 часа(ов)):

Влияние выбора набора признаков на результаты классификации. Предварительная обработка данных. Недостающие значения.

практическое занятие (4 часа(ов)):

Выбор признаков на основе проверки гипотез. Выбор подмножества признаков.

Тема 7. Контекстно-зависимая классификация

лекционное занятие (2 часа(ов)):

Марковские цепи. Алгоритм Витерби. Скрытые марковские модели. Применение в задачах распознавания голоса.

практическое занятие (4 часа(ов)):

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	18	домашнее задание
2.	Тема 2. Линейные классификаторы	2	3-4	подготовка домашнего задания	20	домашнее задание
3.	Тема 3. Метод опорных векторов	2	5-6	подготовка к контрольной работе	24	контрольная работа
4.	Тема 4. Методы восстановления регрессии	2	7-8	подготовка домашнего задания	18	домашнее задание
5.	Тема 5. Искусственные нейронные сети	2	9-11	подготовка к контрольной работе	24	контрольная работа

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
6.	Тема 6. Тема 6. Выбор признаков и подготовка данных	2	12-13	подготовка домашнего задания	20	домашнее задание
7.	Тема 7. Тема 7. Контекстно-зависимая классификация	2	14-15	подготовка домашнего задания	20	домашнее задание
	Итого				144	

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 экзамену:

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:

1. Степанов, Роман Григорьевич. Технология Data Mining: Интеллектуальный анализ данных: учебное пособие / Р. Г. Степанов; Казан. гос. ун-т. Казань: Казанский государственный университет, 2009. 110 с
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.znanium.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://znanium.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 "Введение в машинное обучение" предполагает использование следующего материально-технического обеспечения:

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.

Практические занятия проводятся в аудитории, оборудованной доской и мелом (маркером).

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 38.04.05 "Бизнес-информатика" и магистерской программе Математические методы и информационные технологии в бизнесе .

Автор(ы):

Миссаров М.Д. _____

"__" _____ 201__ г.

Рецензент(ы):

Кашина О.А. _____

"__" _____ 201__ г.