

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.ДВ.2

Направление подготовки: 02.04.02 - Фундаментальная информатика и информационные технологии

Профиль подготовки: Математические основы и программное обеспечение информационной безопасности и защиты информации

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Разинков Е.В.

Рецензент(ы):

Столов Е.Л.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Латыпов Р. Х.

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) ассистент, к.н. Разинков Е.В. кафедра системного анализа и информационных технологий отделение фундаментальной информатики и информационных технологий , Evgenij.Razinkov@kpfu.ru

1. Цели освоения дисциплины

Цель освоения дисциплины заключается в ознакомлении с базовыми понятиями машинного обучения, с основными алгоритмами машинного обучения, особенностями их применения.

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

Данная учебная дисциплина включена в раздел " Б1.В.ДВ.2 Дисциплины (модули)" основной образовательной программы 02.04.02 Фундаментальная информатика и информационные технологии и относится к дисциплинам по выбору. Осваивается на 1 курсе, 2 семестр.

Для освоения дисциплины студентам необходимы знания из линейной алгебры, теории вероятностей, математического анализа. Полученные в процессе изучения дисциплины знания могут использоваться при написании курсовых и дипломных работ, при изучении алгоритмов компьютерного зрения и интеллектуального анализа данных.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-1 (общекультурные компетенции)	способностью к абстрактному мышлению, анализу, синтезу
ОК-2 (общекультурные компетенции)	готовностью действовать в нестандартных ситуациях, нести социальную и этическую ответственность за принятые решения
ОК-3 (общекультурные компетенции)	готовностью к саморазвитию, самореализации, использованию творческого потенциала
ОПК-3 (профессиональные компетенции)	способностью использовать и применять углубленные теоретические и практические знания в области фундаментальной информатики и информационных технологий
ОПК-4 (профессиональные компетенции)	способностью самостоятельно приобретать и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять своё научное мировоззрение
ПК-1 (профессиональные компетенции)	способностью проводить научные исследования и получать новые научные и прикладные результаты самостоятельно и в составе научного коллектива
ПК-13 (профессиональные компетенции)	способностью разрабатывать аналитические обзоры состояния области прикладной математики и информационных технологий
ПК-14 (профессиональные компетенции)	способностью выполнять работу экспертов в ведомственных, отраслевых или государственных экспертных группах по экспертизе проектов, тематика которых соответствует направленности (профилю) программы магистратуры

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-16 (профессиональные компетенции)	способностью участвовать в деятельности профессиональных сетевых сообществ по конкретным направлениям
ПК-2 (профессиональные компетенции)	способностью использовать углубленные теоретические и практические знания в области информационных технологий и прикладной математики, фундаментальных концепций и системных методологий, международных и профессиональных стандартов в области информационных технологий
ПК-3 (профессиональные компетенции)	способностью разрабатывать концептуальные и теоретические модели решаемых научных проблем и задач проектной и производственно-технологической деятельности
ПК-7 (профессиональные компетенции)	способностью разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов

В результате освоения дисциплины студент:

1. должен знать:

- возможности алгоритмов машинного обучения;
- классы задач, решаемых с помощью алгоритмов машинного обучения.

2. должен уметь:

- применять на практике алгоритмы машинного обучения;
- обосновать применение того или иного алгоритма машинного обучения для решения конкретной задачи.

3. должен владеть:

- базовым инструментарием машинного обучения;
- программно реализовывать алгоритмы машинного обучения;
- применять алгоритмы машинного обучения на практике;
- анализировать результаты обучения алгоритма, предлагать пути повышения точности алгоритма.

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 4 зачетных(ые) единиц(ы) 144 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. Введение в машинное обучение	2	1	2	0	0	письменное домашнее задание
2.	Тема 2. Линейные модели регрессии	2	2	2	2	0	письменное домашнее задание
3.	Тема 3. Логистическая регрессия	2	3	2	2	0	письменное домашнее задание
4.	Тема 4. Нейронные сети	2	4	2	4	0	письменное домашнее задание
5.	Тема 5. Деревья решений	2	5	2	2	0	письменное домашнее задание
7.	Тема 7. Алгоритм AdaBoost	2	6	2	2	0	письменное домашнее задание
8.	Тема 8. Кластеризация	2	7	2	2	0	контрольная работа
	Тема . Итоговая форма контроля	2		0	0	0	экзамен
	Итого			14	14	0	

4.2 Содержание дисциплины

Тема 1. Введение в машинное обучение

лекционное занятие (2 часа(ов)):

Обзор задач, решаемых алгоритмами машинного обучения. Классификация алгоритмов машинного обучения.

Тема 2. Линейные модели регрессии

лекционное занятие (2 часа(ов)):

Линейная регрессия. Линейные модели регрессии. Базисные функции. Регуляризация.

практическое занятие (2 часа(ов)):

Программная реализация алгоритма линейной регрессии.

Тема 3. Логистическая регрессия

лекционное занятие (2 часа(ов)):

Целевая функция логистической регрессии. Регуляризация логистической регрессии.

практическое занятие (2 часа(ов)):

Программная реализация алгоритма логистической регрессии.

Тема 4. Нейронные сети

лекционное занятие (2 часа(ов)):

Структура нейрона. Структура нейронной сети. Обучения нейронной сети с помощью алгоритма обратного распространения ошибки.

практическое занятие (4 часа(ов)):

Программная реализация алгоритма обратного распространения ошибки.

Тема 5. Деревья решений

лекционное занятие (2 часа(ов)):

Структура деревьев решений. Виды разделяющих функций. Обучения дерева решений. Алгоритм Random Forest.

практическое занятие (2 часа(ов)):

Программная реализация дерева решений.

Тема 7. Алгоритм AdaBoost

лекционное занятие (2 часа(ов)):

Описание алгоритма AdaBoost. Математическое обоснование алгоритма. Каскад классификаторов.

практическое занятие (2 часа(ов)):

Программная реализация алгоритма AdaBoost.

Тема 8. Кластеризация

лекционное занятие (2 часа(ов)):

Обзор существующих алгоритмов классификации. Алгоритм k-means.

практическое занятие (2 часа(ов)):

Программная реализация алгоритма k-means.

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. Введение в машинное обучение	2	1	Изучение литературы по приложениям машинного обучения.	8	домашнее задание
2.	Тема 2. Линейные модели регрессии	2	2	Проведение экспериментов.	8	домашнее задание
3.	Тема 3. Логистическая регрессия	2	3	Проведение экспериментов.	8	домашнее задание
4.	Тема 4. Нейронные сети	2	4	Проведение экспериментов.	16	домашнее задание
5.	Тема 5. Деревья решений	2	5	Проведение экспериментов.	8	домашнее задание
7.	Тема 7. Алгоритм AdaBoost	2	6	Проведение экспериментов.	16	домашнее задание
8.	Тема 8. Кластеризация	2	7	подготовка к контрольной работе	4	контрольная работа
				Проведение экспериментов.	12	домашнее задание
Итого					80	

5. Образовательные технологии, включая интерактивные формы обучения

Обучение происходит в форме лекционных и лабораторных занятий, а также самостоятельной работы студентов.

Теоретический материал излагается на лекциях. Причем конспект лекций, который остается у студента в результате прослушивания лекции не может заменить учебник. Его цель - формулировка основных утверждений и определений. Прослушав лекцию, полезно ознакомиться с более подробным изложением материала в учебнике. Список литературы разделен на две категории: необходимый для сдачи зачета минимум и дополнительная литература.

Изучение курса подразумевает не только овладение теоретическим материалом, но и получение практических навыков для более глубокого понимания разделов на основе примеров и упражнений, иллюстрирующих рассматриваемые теоретические положения. Аудиторные занятия подразумевают интерактивные обсуждения различных примеров или в активной форме самостоятельного выполнения заданий студентами.

Самостоятельная работа предполагает выполнение домашних заданий. Практические задания, выполненные в аудитории, предназначены для указания общих методов решения задач определенного типа. Закрепить навыки можно лишь в результате самостоятельной работы.

Кроме того, самостоятельная работа включает подготовку к зачету. При подготовке к сдаче зачета весь объем работы рекомендуется распределять равномерно по дням, отведенным для подготовки к зачету, контролировать каждый день выполнения работы. Лучше, если можно перевыполнить план. Тогда будет резерв времени.

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

Тема 1. Введение в машинное обучение

домашнее задание , примерные вопросы:

Разбор возможных приложений алгоритмов машинного обучения.

Тема 2. Линейные модели регрессии

домашнее задание , примерные вопросы:

Проведение экспериментов с программной реализацией линейной регрессии. Исследование влияния параметров алгоритма на значение целевой функции на обучающей и тестовой выборке.

Тема 3. Логистическая регрессия

домашнее задание , примерные вопросы:

Проведение экспериментов с программной реализацией логистической регрессии. Исследование влияния параметров алгоритма на точность алгоритма на обучающей и тестовой выборке.

Тема 4. Нейронные сети

домашнее задание , примерные вопросы:

Проведение экспериментов с программной реализацией нейронной сети. Исследование влияния параметров алгоритма на точность алгоритма на обучающей и тестовой выборке.

Тема 5. Деревья решений

домашнее задание , примерные вопросы:

Проведение экспериментов с программной реализацией дерева решений. Исследование влияния параметров алгоритма на точность алгоритма на обучающей и тестовой выборке.

Тема 7. Алгоритм AdaBoost

домашнее задание , примерные вопросы:

Проведение экспериментов с программной реализацией алгоритма AdaBoost. Исследование влияния параметров слабых классификаторов на точность алгоритма на обучающей и тестовой выборке.

Тема 8. Кластеризация

домашнее задание , примерные вопросы:

Проведение экспериментов с программной реализацией алгоритма k-means. Исследование влияния инициализации алгоритма на значение целевой функции.

контрольная работа , примерные вопросы:

Программная реализация каскада с деревьями решений в качестве слабых классификаторов.

Тема . Итоговая форма контроля

Примерные вопросы к экзамену:

Вопросы к зачету:

- Классификация алгоритмов машинного обучения.
- Регрессия, линейная регрессия, линейные модели регрессии.
- Целевая функция линейной регрессии.
- Классификация. Логистическая регрессия.
- Целевая функция логистической регрессии.
- Регуляризация.
- Нейронные сети. Структура нейронной сети.
- Алгоритм обратного распространения ошибки.
- Деревья решений. Обучение дерева решений.
- Алгоритм Random Forest.
- Алгоритм AdaBoost.
- Каскад классификаторов. Обучение каскада.
- Кластеризация. Алгоритм k-means.

7.1. Основная литература:

1. Джонс, М. Т. Программирование искусственного интеллекта в приложениях [Электронный ресурс] / М. Тим Джонс ; Пер. с англ. А. И. Осипов. - 2-е изд. - М. : ДМК Пресс, 2011. - 312 с.: ил. - ISBN 978-5-94074-746-8.

<http://znanium.com/bookread2.php?book=410211>

2. Саттон, Р. С. Обучение с подкреплением [Электронный ресурс] / Р. С. Саттон, Э. Г. Барто ; пер. с англ. - Эл. изд. - М.: БИНОМ. Лаборатория знаний, 2012. - 399 с.: ил. - (Адаптивные и интеллектуальные системы). - ISBN 978-5-9963-1349-5.

<http://e.lanbook.com/view/book/4405/>

3. Шапиро, Л. Компьютерное зрение [Электронный ресурс] / Л. Шапиро, Дж. Стокман; пер. с англ. - 2-е изд. (эл.). - М.: БИНОМ. Лаборатория знаний, 2013. - 752 с.: ил. - (Лучший зарубежный учебник). - ISBN 978-5-9963-1312-9.

<http://e.lanbook.com/view/book/42631/>

7.2. Дополнительная литература:

1. 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. Режим доступа: <http://znanium.com/bookread2.php?book=451186>

2. Куприянов М. С., Барсегян, А. А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP [Электронный ресурс] / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко. - 2-е изд., перераб. и доп. - СПб.: БХВ-Петербург, 2007. - 384 с. Режим доступа:<http://znanium.com/bookread2.php?book=489445>

7.3. Интернет-ресурсы:

Видео-лекции курса "Машинное обучение" от Yandex -

<https://yandexdataschool.ru/edu-process/courses/machine-learning>

Научный портал по математическим наукам - <http://www.mathnet.ru>

Портал образовательных ресурсов по ИТ - <http://www.intuit.ru>

Портал статей по применению ИТ и машинному обучению -

http://habrahabr.ru/hub/machine_learning/

Профессиональный интернет-ресурс по машинному обучению - <http://www.machinelearning.ru/>

8. Материально-техническое обеспечение дисциплины(модуля)

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

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов. Типовая комплектация мультимедийной аудитории состоит из: мультимедийного проектора, автоматизированного проекционного экрана, акустической системы, а также интерактивной трибуны преподавателя, включающей тач-скрин монитор с диагональю не менее 22 дюймов, персональный компьютер (с техническими характеристиками не ниже Intel Core i3-2100, DDR3 4096Mb, 500Gb), конференц-микрофон, беспроводной микрофон, блок управления оборудованием, интерфейсы подключения: USB, audio, HDMI. Интерактивная трибуна преподавателя является ключевым элементом управления, объединяющим все устройства в единую систему, и служит полноценным рабочим местом преподавателя. Преподаватель имеет возможность легко управлять всей системой, не отходя от трибуны, что позволяет проводить лекции, практические занятия, презентации, вебинары, конференции и другие виды аудиторной нагрузки обучающихся в удобной и доступной для них форме с применением современных интерактивных средств обучения, в том числе с использованием в процессе обучения всех корпоративных ресурсов. Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

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.

Аудитория для проведения лекций с мультимедийным оборудованием, компьютерный класс.

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 02.04.02 "Фундаментальная информатика и информационные технологии" и магистерской программе Математические основы и программное обеспечение информационной безопасности и защиты информации .

Автор(ы):

Разинков Е.В. _____

"__" _____ 201__ г.

Рецензент(ы):

Столов Е.Л. _____

"__" _____ 201__ г.