

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего 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

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Методы прикладной математической статистики

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2017

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

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. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-1	способность к абстрактному мышлению, анализу, синтезу
ОК-3	готовность к саморазвитию, самореализации, использованию творческого потенциала
ОПК-3	способностью самостоятельно приобретать с помощью информационных технологий 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
ОПК-4	способностью использовать и применять углубленные знания в области прикладной математики и информатики
ПК-3	способностью углубленного анализа проблем, постановки и обоснования задач научной и проектно-технологической деятельности
ПК-4	способностью разрабатывать концептуальные и теоретические модели решаемых задач проектной и производственно-технологической деятельности

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- основные постановки и классификацию задач машинного обучения;
- основы регрессионного и кластерного анализа ;
- методы байесовской и линейной классификации, деревьев решений, нейронных сетей;

Должен уметь:

- формализовать задачи в различных прикладных областях на основе математических моделей регрессионного и кластерного анализа, теории классификации;

- применять методы и алгоритмы машинного обучения при решении прикладных задач анализа данных

Должен владеть:

- современным математическим аппаратом анализа данных;
- навыками постановки научно-исследовательских задач.

Должен демонстрировать способность и готовность:

- осуществлять поиск информации и предварительную обработку данных по полученному заданию, подбор адекватных методов машинного обучения, необходимых для решения поставленных задач
- применять в профессиональной деятельности знания, умения, навыки, полученные в ходе освоения дисциплины

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.Б.9 Дисциплины (модули)" основной профессиональной образовательной программы 01.04.02 "Прикладная математика и информатика (Методы прикладной математической статистики)" и относится к базовой (общепрофессиональной) части.

Осваивается на 1 курсе в 2 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

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

Контактная работа - 42 часа(ов), в том числе лекции - 26 часа(ов), практические занятия - 16 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 102 часа(ов).

Контроль (зачёт / экзамен) - 0 часа(ов).

Форма промежуточного контроля дисциплины: зачет во 2 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины / модуля	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)			Самостоятельная работа
			Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Классификация задач машинного обучения. Распределения и расстояния на данных. Программное обеспечение решения задач машинного обучения.	2	2	2	0	12
2.	Тема 2. Регрессионный анализ. Основы работы в среде R.	2	6	4	0	21
3.	Тема 3. Кластерный анализ. Ввод и визуализация данных в R.	2	4	2	0	15
4.	Тема 4. Методы классификации. Применение пакета R для анализа регрессионных моделей.	2	4	4	0	15
5.	Тема 5. Деревья решений. Развитие методов регрессионного анализа в среде R.	2	4	2	0	15
6.	Тема 6. Нейронные сети. Применение пакета R для реализации алгоритмов нейронных сетей.	2	6	2	0	24
	Итого		26	16	0	102

4.2 Содержание дисциплины (модуля)

Тема 1. Классификация задач машинного обучения. Распределения и расстояния на данных. Программное обеспечение решения задач машинного обучения.

Примеры и классификация задач машинного обучения. Обучение с учителем и обучение без учителя. Проблема переобучения. Задачи регрессии и классификации. Вероятностные распределения на данных. Функции потерь. Эмпирический риск. Типы данных, обработка данных. Меры сходства, метрики, ультраметрики. Расстояния на евклидовых пространствах. Расстояние Хэмминга и редакционное расстояние. Обзор современного программного обеспечения, используемого для решения задач машинного обучения. Плюсы и минусы наиболее распространённых пакетов и сервисов: R, Python, MICE, rpart, PARTY, CARET, randomForest, nnet, e1071, kernLab.

Тема 2. Регрессионный анализ. Основы работы в среде R.

Гауссовское распределение и распределения, с ним связанные. Свойства гауссовского распределения. Метод максимального правдоподобия. Однофакторная линейная регрессия. Проверка гипотез о коэффициентах регрессии. Коэффициент корреляции. Коэффициент детерминации. Множественная линейная регрессия. Нелинейная регрессия. Гребневая регрессия. Метод лассо. Источники получения пакета R и интерфейса RStudio. Установка, запуск, настройка пакета. Структура пакета. Меню. Подключение внешних пакетов.

Тема 3. Кластерный анализ. Ввод и визуализация данных в R.

Эвристические методы кластеризации: метод связанных компонент, метод кратчайшего дерева, алгоритм Форель. Алгоритм k-средних. Алгоритмы иерархической кластеризации. Методы "ближнего соседа", "дальнего соседа", средней связи, центров масс, метод Уорда. Формула Ланса-Уильямса. Функционалы качества кластерного разбиения. Типы данных в R. Ввод данных. Инструменты предварительной обработки данных: масштабирование, агрегирование, панорамирование, фильтрация. Визуализация количественных и качественных данных. Основные виды графических объектов: графики функций, гистограммы, мозаичные диаграммы, скрипичные диаграммы, облако точек, "ящик с усами". Интерактивная визуализация. Визуализация сетей в виде графов, построения с использованием географических карт. Многомерная визуализация. Визуализация сетевых, иерархических и географических данных. Тепловые карты.

Тема 4. Методы классификации. Применение пакета R для анализа регрессионных моделей.

Метод k ближайших соседей и его модификации. Формула Байеса. Байесовская классификация. Потери и риски. Наивный байесовский классификатор. Примеры байесовской классификации в случае дискретного и непрерывного распределений. Обобщения линейной регрессионной модели: полиномиальная и экспоненциальная модели. Снижение размерности в задачах регрессионного анализа.

Тема 5. Деревья решений. Развитие методов регрессионного анализа в среде R.

Регрессионные деревья. Деревья классификации.

Энтропия как мера информации. Коэффициент Джини. Алгоритм CART. Алгоритм C4.5. Пример построения дерева классификации. Решающие правила. Обрезка деревьев. Проблема мультиколлинеарности. Подбор факторов для включения в модель. Метод LASSO. Множественная логистическая регрессия.

Тема 6. Нейронные сети. Применение пакета R для реализации алгоритмов нейронных сетей.

Линейные классификаторы. Алгоритм обучения перцептрона. Теорема Новикова. 2-х-слойный перцептрон. Многослойные нейронные сети. Обучение нейронной сети. Градиентный спуск и модели искусственных нейронных сетей. Метод обратного распространения ошибки.

Реализация примеров использования нейронной сети в среде R. Обзор перспективных задач машинного обучения и подходов к их решению.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства образования и науки Российской Федерации от 5 апреля 2017 года №301)

Письмо Министерства образования Российской Федерации №14-55-996ин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

Вьюгин, В.В. Математические основы машинного обучения и прогнозирования [Электронный ресурс] : учеб. пособие Москва : МЦНМО, 2013. 304 с. - <https://e.lanbook.com/book/56397>

Лагутин, М.Б. Наглядная математическая статистика [Электронный ресурс] : учеб. пособие Москва : Издательство 'Лаборатория знаний', 2015. 475 с. - <https://e.lanbook.com/book/70706>

Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Электронный ресурс] Москва : ДМК Пресс, 2015. 400 с. - <https://e.lanbook.com/book/69955>

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);

- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;
- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Байесовские_методы_машинного_обучения_(курс_лекций_Д.П. Ветрова) - <http://www.machinelearning.ru/wiki/index.php?title=>

Байесовские_методы_машинного_обучения_(курс_лекций)_/_2017_Д.П. Ветров

Машинное обучение (курс лекций, Н.Ю.Золотых) - <http://www.uic.unn.ru/~zny/ml/>

Машинное_обучение_(курс_лекций_К.В.Воронцова) -

[http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_\(курс_лекций%2C_К.В.Воронцов\)](http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_(курс_лекций%2C_К.В.Воронцов))

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лекции	После прослушивания каждой лекции студент должен дома самостоятельно ее проработать, используя конспект и ту основную и дополнительную литературу, которая была рекомендована преподавателем. Такая самостоятельная работа даст возможность студенту на очередной лекции лучше понимать и усваивать новый материал.
практические занятия	Каждое практическое занятие предполагает знание изложенного на лекции теоретического материала, относящегося к данному практическому занятию. Поэтому для лучшего усвоения студенту желательно перед каждым практическим занятием изучить предстоящую тему, используя конспект и литературу, которая была рекомендована преподавателем.
самостоятельная работа	Самостоятельная работа студентов опирается на приведенную основную и дополнительную литературу по изучаемому курсу, а также на предложенные интернет-ресурсы. Для помощи студенту при освоении им тем, вынесенных на самостоятельное обучение, предполагается проведение консультаций, как общих, так и индивидуальных.
зачет	Для подготовки к зачету по дисциплине студенту будет предложена подробная программа курса и вопросы, выносимые на зачет. При подготовке к зачету предполагается, что студент, изучающий курс, сначала ознакомится с лекциями, просмотрит материалы тем, выносимых на зачет, в предлагаемой учебно-методической литературе.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

Мультимедийная аудитория.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 01.04.02 "Прикладная математика и информатика" и магистерской программе "Методы прикладной математической статистики".

Приложение 2
к рабочей программе дисциплины (модуля)
Б1.Б.9 Введение в машинное обучение

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Методы прикладной математической статистики

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2017

Основная литература:

1. Вьюгин, В.В. Математические основы машинного обучения и прогнозирования [Электронный ресурс]: учебное пособие / В.В. Вьюгин. - Электрон. дан. - Москва : МЦНМО, 2014. - 304 с. - Режим доступа: <https://e.lanbook.com/book/56397>
2. Флах П., Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Электронный ресурс] / Флах П. - М. : ДМК Пресс, 2015. - 400 с. - ISBN 978-5-97060-273-7 - Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785970602737.html>
3. Кашина О.А., Миссаров М.Д. Электронный курс 'Анализ данных в среде R', 2014. - Режим доступа: <https://edu.kpfu.ru/course/view.php?id=833>

Дополнительная литература:

1. Крянев, А.В. Метрический анализ и обработка данных [Электронный ресурс] / А.В. Крянев, Г.В. Лукин, Д.К. Удумян. - Электрон. дан. - Москва : Физматлит, 2012. - 308 с. - Режим доступа: <https://e.lanbook.com/book/59523>
2. Компьютерные технологии анализа данных в эконометрике / Д.М. Дайитбегов. - 2-е изд., испр. и доп. - М.: Вузовский учебник: ИНФРА-М, 2010. - 578 с. - (Научная книга). ISBN 978-5-9558-0191-9 -Режим доступа: <http://www.znaniium.com/bookread.php?book=251791>
3. Чураков, Е.П. Введение в многомерные статистические методы [Электронный ресурс] : учеб. пособие - Санкт-Петербург: Лань, 2016. -148 с. - Режим доступа: <https://e.lanbook.com/book/87598>

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.Б.9 Введение в машинное обучение

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Методы прикладной математической статистики

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2017

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. 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, учебно-методические комплексы, 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 образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен обучающимся. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.