



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение высшего
образования
«Дальневосточный федеральный университет»
(ДФУ)

ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ (ШКОЛА)

«СОГЛАСОВАНО»

Руководитель ОП

Артемьева И.Л.

«УТВЕРЖДАЮ»

И.о. директора департамента

Смагин С.В.



«15» июля 2021 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Методы машинного обучения

Направление подготовки **09.03.04 Программная инженерия**
(Программная инженерия)

Форма подготовки очная

курс 4 семестр 7
лекции 18 час.
практические занятия 00 час.
лабораторные работы 36 час.
в том числе с использованием МАО лек. 0 /пр. 0 /лаб. 18 час.
всего часов аудиторной нагрузки 54 час.
в том числе с использованием МАО 18 час.
самостоятельная работа 54 час.
в том числе на подготовку к экзамену 27 час.
контрольные работы (количество) не предусмотрены
курсовая работа / курсовой проект не предусмотрены
зачет _ семестр
экзамен 7 семестр

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.04 Программная инженерия, утвержденный приказом Министерства образования и науки Российской Федерации от 19 сентября 2017 г. № 920 (с изменениями и дополнениями).

Рабочая программа обсуждена на заседании кафедры прикладной математики, механики, управления и программного обеспечения, протокол № 7.1 от « 04 » июля 2019 г.

Заведующий кафедрой прикладной математики, механики, управления и программного обеспечения д.т.н., профессор Артемьева И.Л.

Составитель: доцент кафедры ПММУиПО Смагин С.В., к.т.н.

Владивосток
2021

Оборотная сторона титульного листа РПД

I. Рабочая программа пересмотрена на заседании кафедры прикладной математики, механики, управления и программного обеспечения:

Протокол от «09» июля 2021 г. № 7.1

Заведующий кафедрой _____ Артемяева И.Л.
(подпись) (И.О. Фамилия)

II. Рабочая программа пересмотрена на заседании департамента программной инженерии и искусственного интеллекта:

Протокол от «17» сентября 2021 г. № 9.1

И.о. директора департамента _____ Смагин С.В.
(подпись) (И.О. Фамилия)

III. Рабочая программа пересмотрена на заседании кафедры:

Протокол от « ____ » _____ 20__ г. № ____

Заведующий кафедрой _____
(подпись) (И.О. Фамилия)

IV. Рабочая программа пересмотрена на заседании кафедры:

Протокол от « ____ » _____ 20__ г. № ____

Заведующий кафедрой _____
(подпись) (И.О. Фамилия)

ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель дисциплины – познакомить студентов с прикладными задачами обучения по прецедентам, изучить основные методы решения этих задач и базовые алгоритмы, реализующие эти методы.

Задачи дисциплины:

1. Ознакомление с наукой о данных (Data Science), большими данными (Big Data), а также с существующими направлениями исследований и областями применения машинного обучения (Machine Learning).
2. Изучение базовых понятий и терминов машинного обучения.
3. Рассмотрение основных типов и примеров прикладных задач.
4. Проведение обзора современных приложений машинного обучения.
5. Изучение основных групп методов и базовых алгоритмов решения задач классификации, регрессии, прогнозирования, кластеризации.
6. Реализация и проведение исследования одного из рассмотренных в курсе алгоритмов на основе обучающих выборок реальных данных.

Планируемые результаты обучения по данной дисциплине (знания, умения, владения), соотнесенные с планируемыми результатами освоения образовательной программы.

Общепрофессиональные компетенции выпускников и индикаторы их достижения:

Наименование категории (группы) общепрофессиональных компетенций	Код и наименование общепрофессиональной компетенции	Код и наименование индикатора достижения общепрофессиональной компетенции
	ОПК-1. Способен применять естественнонаучные и общепрофессиональные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1.1. Знает основные группы методов машинного обучения, базовые алгоритмы, реализующие эти методы, а также их достоинства и недостатки. ОПК-1.2. Умеет программировать, настраивать и применять алгоритмы машинного обучения, необходимые для обработки данных. ОПК-1.3. Владеет методами создания программных средств для решения задач машинного обучения и подходами к их верификации.

	<p>ОПК-2. Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности</p>	<p>ОПК-2.1. Знает модели и методы машинного обучения, применяемые при решении практических задач.</p> <p>ОПК-2.2. Умеет использовать и сравнивать различные модели и методы машинного обучения.</p> <p>ОПК-2.3. Владеет навыками применения программных средств при решении практических задач, связанных с машинным обучением.</p>
--	---	---

Профессиональные компетенции выпускников и индикаторы их достижения:

Задача профессиональной деятельности	Объект или область знания	Код и наименование профессиональной компетенции	Код и наименование индикатора достижения профессиональной компетенции	Основание (ПС, анализ иных требований, предъявляемых к выпускникам)
Тип задач профессиональной деятельности: проектный				

<p>Формирование требований к информатизации и автоматизации и прикладных процессов, формализация предметной области проекта; технико-экономическое обоснование проектных решений и составление технического задания на разработку программного продукта; проектирование программно-аппаратных средств в соответствии с техническим заданием; применение современных инструментальных средств при разработке программного обеспечения; документирование компонентов информационной системы на стадии жизненного цикла</p>	<p>Прикладные и информационные процессы. Информационные технологии. Программное обеспечение</p>	<p>ПК-6. Владение навыками моделирования, анализа и использования формальных методов конструирования программного обеспечения</p>	<p>ПК-6.1. Знает особенности выбора признаков моделей и предварительной обработки данных. ПК-6.2. Умеет формировать набор признаков модели и проводить предварительную обработку данных. ПК-6.3. Владеет технологиями оценивания и выявления информативных признаков модели.</p>	<p>06.028 Системный программист 06.004 Специалист по тестированию в области информационных технологий 06.001 Программист</p>
<p>Тип задач профессиональной деятельности: производственно-технологический</p>				

<p>Проведение работ по инсталляции программного обеспечения автоматизированных систем и загрузки баз данных; настройка параметров ИС и тестирование результатов настройки; ведение технической документации; техническое сопровождение ИС в процессе эксплуатации; применение Web технологий при реализации удаленного доступа в системах клиент – сервер и распределенных вычислений</p>	<p>Программное обеспечение</p>	<p>ПК-10. Владение навыками использования различных технологий разработки программного обеспечения</p>	<p>ПК-10.1. Знает сильные и слабые стороны методов и алгоритмов машинного обучения при решении конкретных практических задач. ПК-10.2. Умеет оценивать степень эффективности применения современных методов и алгоритмов машинного обучения при решении конкретных практических задач. ПК-10.3. Владеет навыками отбора подходящего метода и алгоритма машинного обучения, в зависимости от решаемой задачи.</p>	<p>06.028 Системный программист 06.022 Системный аналитик 06.004 Специалист по тестированию в области информационных технологий 06.001 Программист</p>
---	--------------------------------	--	--	---

Для формирования вышеуказанных компетенций в рамках дисциплины «Методы машинного обучения» применяются следующие методы активного/интерактивного обучения: метод круглого стола, метод проектов, дискуссия, анализ конкретных ситуаций, ситуационный анализ.

I. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

Лекционный материал (18 час.)

Тема 1. Введение, цели и задачи курса, основные термины (2 час.)

Рассматриваются понятия: модель предметной области, задача обучения по прецедентам, алгоритм обучения, решающее правило, критерий качества обучения, обучающая и контрольная выборки, модельные и реальные данные, объекты, признаки, свойства, классы, кластеры, типы шкал (бинарные, номинальные, порядковые, количественные).

Тема 2. Современная концепция анализа данных, разбор основных типов прикладных задач (2 час.)

Рассматриваются типы задач обучения с учителем: классификация, регрессия, ранжирование, прогнозирование.

Рассматриваются типы задач обучения без учителя: кластеризация, поиск ассоциативных правил, фильтрация выбросов, сокращение размерности, заполнение пропущенных значений.

Каждый тип задач сопровождается одним или несколькими детально расписанными примерами.

Тема 3. Формальные определения терминов, введение в проблему переобучение (2 час.)

Рассматриваются формальные (математические) определения терминов: модель алгоритмов, алгоритм обучения, этапы обучения, ошибка, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль, проблема переобучения.

Тема 4. Метрические методы классификации (2 час.)

Рассматриваются понятия: гипотеза компактности, метрика, метод ближайших соседей и его обобщения, окно Парзена и потенциальные функции, отбор эталонов (понятие отступа, алгоритм отбора эталонных объектов STOLP), отбор признаков и оптимизация метрики (задача выбора метрики, жадный алгоритм отбора признаков, полный скользящий контроль CCV).

Тема 5. Логические методы классификации (2 час.)

Рассматриваются понятия: закономерность, информативность, интерпретируемость, решающие деревья (алгоритм ID3, небрежные решающие деревья ODT, бинаризация данных).

Тема 6. Линейные методы классификации. Метод стохастического градиента (2 час.)

Рассматриваются понятия: минимизация эмпирического риска для градиентных методов обучения, линейный классификатор, метод стохастического градиента SG.

Тема 7. Линейные методы классификации. Метод опорных векторов (2 час.)

Рассматриваются понятия: принцип оптимальной разделяющей гиперплоскости, двойственная задача, понятие опорного вектора, обобщения линейного SVM (ядра и спрямляющие пространства, нейронные сети и SVM, обзор регуляризаторов для SVM).

Тема 8. Байесовские методы классификации (2 час.)

Рассматриваются понятия: оптимальный байесовский классификатор, вероятностная постановка задачи классификации, задача восстановления плотности распределения, наивный байесовский классификатор.

Тема 9. Методы кластеризации (2 час.)

Рассматриваются понятия: цели кластеризации, некорректность задачи кластеризации, типы кластерных структур, проблема чувствительности к выбору метрики, графовые методы кластеризации (алгоритм выделения связанных компонент, алгоритм ФОРЭЛ, функционалы качества кластеризации), иерархическая кластеризация (таксономия) (агломеративная иерархическая кластеризация, дендрограмма и свойство монотонности, свойства сжатия, растяжения и редуктивности).

II. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА

Практические занятия (0 час.)

Не предусмотрено.

Лабораторные работы (36 час.)

Лабораторная работа №1. Метрические методы классификации (4 час.)

Реализация метода ближайшего соседа и метода k ближайших соседей.

Лабораторная работа №2. Метрические методы классификации (4 час.)

Реализация метода окна Парзена (фиксированной ширины и переменной ширины).

Лабораторная работа №3. Метрические методы классификации (4 час.)

Реализация метода потенциальных функций и алгоритма STOLP.

Лабораторная работа №4. Логические методы классификации (4 час.)

Реализация жадного алгоритма построения решающего дерева ID3.

Лабораторная работа №5. Логические методы классификации (4 час.)

Реализация алгоритма построения небрежных решающих деревьев ODT.

Лабораторная работа №6. Линейные методы классификации (4 час.)

Реализация метода стохастического градиента SG.

Лабораторная работа №7. Байесовские методы классификации (4 час.)

Реализация наивного байесовского классификатора.

Лабораторная работа №8. Методы кластеризации (4 час.)

Реализация алгоритма ФОРЭЛ (ФОРмальные ЭЛементы).

Лабораторная работа №9. Методы кластеризации (4 час.)

Разбор функционалов качества кластеризации и реализация алгоритма Ланса-Уильямса.

Ш. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине «Методы машинного обучения» представлено в разделе VIII и включает в себя: план-график выполнения самостоятельной работы по дисциплине, в том числе примерные нормы времени на выполнение по каждому заданию; характеристика заданий для самостоятельной работы обучающихся и методические рекомендации по их

выполнению; требования к представлению и оформлению результатов самостоятельной работы; критерии оценки выполнения самостоятельной работы.

III. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

№ п/п	Контролируемые разделы/темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства – наименование	
				текущий контроль	промежуточная аттестация
1	Темы 1-3	ОПК2	знает	Собеседование УО-1, Круглый стол УО-4,	Экзамен Вопросы 1-11 (теоретическая часть) и 1-3 (практическая)
			умеет	Практические задания 1-3 ПР-6	
2	Темы 4-9	ПК6	знает	Собеседование УО-1, Круглый стол УО-4,	Экзамен Вопросы 4-6 (практическая часть)
			умеет	Практические задания 4-5 ПР-6	
		ОПК1	знает	Собеседование УО-1, Круглый стол УО-4,	Экзамен Вопросы 7-10 (практическая часть)
			умеет	Практические задания 6-7 ПР-6	
		ПК10	знает	Собеседование УО-1, Круглый стол УО-4,	Экзамен Вопросы 11-13 (практическая часть)
			умеет	Практические задания 8-9 ПР-6	

Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков и (или) опыта деятельности, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в разделе IX.

V. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

1. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных – М.: ДМК Пресс, 2015. – 400 с. <http://www.studentlibrary.ru/book/ISBN9785970602737.html>
<https://e.lanbook.com/book/69955>
2. Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. 2-е издание / пер. с англ. Слинкин А.А. – М.: ДМК Пресс, 2016. – 302 с.: ил. <http://www.studentlibrary.ru/book/ISBN9785970603307.html>
3. Вьюгин В.В. Математические основы машинного обучения и прогнозирования [Электронный ресурс]: учеб. пособие – М.: МЦНМО, 2013. – 304 с. <https://e.lanbook.com/book/56397>
4. Рашка С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения [Электронный ресурс]: рук. / С. Рашка; пер. с англ. Логунова А.В. – М.: ДМК Пресс, 2017. – 418 с. <https://e.lanbook.com/book/100905>
5. Кулаичев А.П. Методы и средства комплексного анализа данных: учебное пособие для вузов. – М.: Форум: ИНФРА-М, 2014. 511 с. <https://lib.dvfu.ru:8443/lib/item?id=chamo:795113&theme=FEFU>

Дополнительная литература

1. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. – Новосибирск: ИМ СО РАН, 1999. – 270 с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:10172&theme=FEFU>
2. Кухаренко Б.Г. Интеллектуальные системы и технологии – М.: Московская государственная академия водного транспорта, 2015. – 116 с.
<http://www.iprbookshop.ru/47933.html>
3. Дедукция и обобщение в системах принятия решений / В.Н. Вагин. М: Наука, 1988. – 383 с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:55526&theme=FEFU>
4. Давнис В.В., Тинякова В.И., Мокшина С.И., Алексеева А.И. Компьютерные решения задач многомерной статистики. Часть 1. Кластерный и дискриминантный анализ: Учебное пособие. – Воронеж: Изд-во ВГУ, 2005. – 37 с. <http://window.edu.ru/resource/417/40417>
5. Симчера В.М. Методы многомерного анализа статистических данных: учеб. пособие. – М.: Финансы и статистика, 2008. – 400 с.: ил.
<http://www.studentlibrary.ru/book/ISBN9785279031849.html>

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. <http://window.edu.ru/resource/417/40417> Давнис В.В., Тинякова В.И., Мокшина С.И., Алексеева А.И. Компьютерные решения задач многомерной статистики. Часть 1. Кластерный и дискриминантный анализ: Учебное пособие. – Воронеж: Изд-во ВГУ, 2005. – 37 с.
1. <http://machinelearning.ru/> MachineLearning.ru Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.
2. http://shad.yandex.ru/lectures/machine_learning.xml Видеолекции курса «Машинное обучение» Школы анализа данных.
3. <http://www.studentlibrary.ru/book/ISBN9785940575061.html> Николенко С.И., Тулупьев А.Л. Самообучающиеся системы. – М.: МЦНМО, 2009. – 288 с.: 24 ил.
4. <http://znanium.com/catalog.php?bookinfo=451186> Информационные аналитические системы: 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.

Перечень информационных технологий и программного обеспечения

При осуществлении образовательного процесса студентами и профессорско-преподавательским составом используется следующее программное обеспечение:

1. Microsoft Office (Access, Excel, PowerPoint, Word и т. д).
2. Open Office.
3. Skype.

При осуществлении образовательного процесса студентами и профессорско-преподавательским составом используются следующие информационно-справочные системы:

1. Научная электронная библиотека eLIBRARY.
2. Электронно-библиотечная система издательства «Лань».
3. Электронная библиотека "Консультант студента".
4. Электронно-библиотечная система IPRbooks.
5. Информационная система "ЕДИНОЕ ОКНО доступа к образовательным ресурсам".
6. Доступ к электронному заказу книг в библиотеке ДВФУ, доступ к нормативным документам ДВФУ, расписанию, рассылке писем.

Лабораторные занятия проводятся в специализированном компьютерном классе. Для проведения практических занятий требуется следующее программное обеспечение: CPython, RGui, RStudio.

VI. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Дисциплина «Методы машинного обучения» изучается в следующих организационных формах: лекционное занятие; практическое занятие; самостоятельное изучение теоретического материала; самостоятельное выполнение индивидуального проекта; индивидуальные и групповые консультации.

Основной формой самостоятельной работы студента является изучение конспекта лекций, их дополнение рекомендованной литературой, выполнение проекта, а также активная работа на лабораторных занятиях.

К прослушиванию лекции следует готовиться, для этого необходимо знать программу курса и рекомендованную литературу. Тогда в процессе лекции легче отделить главное от второстепенного, легче сориентироваться: что записать, что самостоятельно проработать, что является трудным для понимания, а что легко усвоить.

Контроль за выполнением самостоятельной работы студента производится в виде контроля каждого этапа работы, отраженного в документации и защиты проекта.

Студент должен планировать график самостоятельной работы по дисциплине и придерживаться его.

Подготовка к экзамену

К аттестации допускаются студенты, которые систематически в течение всего семестра посещали и работали на занятиях и показали уверенные знания в ходе выполнении лабораторных работ.

Непосредственная подготовка к аттестации осуществляется по вопросам, представленным в рабочей учебной программе. Тщательно изучите формулировку каждого вопроса, вникните в его суть, составьте план ответа. Обычно план включает в себя:

- определение сущности рассматриваемого вопроса, основных положений, утверждений, определение необходимости их доказательства;
- графический материал (таблицы, рисунки, графики), необходимые для раскрытия сущности вопроса;
- роль и значение рассматриваемого материала для практической деятельности, примеры использования в практической деятельности.

VII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Лекции проводятся с использованием проектора и внутренней системы портала ДВФУ. Лабораторные занятия проходят в аудиториях, оборудованных компьютерами типа Lenovo C360G-i34164G500UDK с лицензионными программами Microsoft Office 2013 и аудиовизуальными средствами проектор Panasonic DLPProjectorPT-D2110XE, плазма LG FLATRON M4716CCBAM4716CJ. Для выполнения самостоятельной работы студенты в жилых корпусах ДВФУ обеспечены Wi-Fi.

VIII. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

План-график выполнения самостоятельной работы по дисциплине

№	Дата/сроки	Вид самостоятельной	Примерные	Форма контроля
----------	-------------------	----------------------------	------------------	-----------------------

п/п	выполнения	работы	нормы времени на выполнение	
1.	7 неделя обучения	Формулировка задачи, выбор предметной области	1 час	Собеседование УО-1
2.	8 неделя обучения	Разработка модели предметной области	2 час	Проект ПР-9
3.	9 неделя обучения	Исследование свойств модели предметной области	3 час	Собеседование УО-1
4.	10 неделя обучения	Выделение набора информативных признаков	3 час	Собеседование УО-1
5.	11 неделя обучения	Постановка задачи машинного обучения	3 час	Собеседование УО-1
6.	12 неделя обучения	Выбор метода машинного обучения и анализ его свойств	3 час	Собеседование УО-1
7.	13 неделя обучения	Формирование спецификации программного средства, в котором будет реализован выбранный метод машинного обучения	3 час	Проверка отчетов, собеседование
8.	14 неделя обучения	Выбор/генерация обучающей выборки	3 час	Проверка отчетов, собеседование
9.	15 неделя обучения	Реализация выбранного метода машинного обучения	3 час	Проверка отчетов, собеседование
10.	17 неделя обучения	Оценка свойств метода	3 час	Проверка отчетов, собеседование
	18 неделя обучения	Подготовка к экзамену	27 час	
	Всего		54 час	Экзамен

Рекомендации по самостоятельной работе студентов

Индивидуальные задания на самостоятельную работу студенты получают в ходе процесса администрирования и сопровождения программного продукта, в соответствии с теми ролями, которые они играют, с этапом сопровождения программного продукта, в зависимости от потребности бригады и в соответствии с планом работы бригады. Распределение заданий производит бригадир под контролем преподавателя.

Примерные виды индивидуальных заданий: подготовка изменений спецификаций подсистемы, переделка существующих и разработка новых тестовых ситуаций, составление набора тестов для одной тестовой ситуации, программирование/ модификация одной или нескольких функций/дефектов

подсистемы, составление и редактирование документов разработки, инспекция документов разработки, разработка/модификация схемы базы данных программного продукта, выпуск нового релиза, разработка плана внедрения новой версии программного продукта, и т.п.

Контроль самостоятельной работы осуществляется не только преподавателем, но и всей бригадой разработчиков.

Рекомендации по планированию и организации времени, отведенного на изучение дисциплины

Освоение дисциплины следует начинать с изучения рабочей учебной программы, которая содержит основные требования к знаниям, умениям и навыкам. Обязательно следует учитывать рекомендации преподавателя, данные в ходе установочных занятий. Затем – приступить к изучению отдельных разделов и тем в порядке, предусмотренном программой.

Получив представление об основном содержании раздела, темы, необходимо изучить материал с помощью рекомендуемой основной литературы. Целесообразно составить краткий конспект или схему, отображающую смысл и связи основных понятий данного раздела и включенных в него тем. Обязательно следует записывать возникшие вопросы, на которые не удалось ответить самостоятельно.

Подготовку к началу обучения включает несколько необходимых пунктов:

1) Необходимо создать для себя рациональный и эмоционально достаточный уровень мотивации к последовательному и планомерному изучению дисциплины.

2) Необходимо изучить список рекомендованной основной и дополнительной литературы и убедиться в её наличии у себя дома или в библиотеке в бумажном или электронном виде.

3) Желательно в самом начале периода обучения возможно тщательнее спланировать время, отводимое на работу с источниками и литературой по дисциплине, представить этот план в наглядной форме (график работы с датами) и в дальнейшем его придерживаться, не допуская срывов графика индивидуальной работы и «аврала» в предсессионный период. Пренебрежение этим пунктом приводит к переутомлению и резкому снижению качества усвоения учебного материала.

Рекомендации по работе с литературой

1) Всю учебную литературу желательно изучать «под конспект». Цель написания конспекта по дисциплине – сформировать навыки по поиску, отбору, анализу и формулированию учебного материала.

2) Написание конспекта должно быть творческим – нужно не переписывать текст из источников, но пытаться кратко излагать своими словами содержание ответа, при этом максимально структурируя конспект, используя символы и условные обозначения.

3) При написании конспекта каждый новый вопрос начинается с нового листа, для каждого экзаменационного вопроса отводится 1-2 страницы конспекта.

5) При работе над конспектом обязательно выявляются и отмечаются трудные для самостоятельного изучения вопросы, с которыми уместно обратиться к преподавателю при посещении установочных лекций и консультаций, либо в индивидуальном порядке.

6) При чтении учебной и научной литературы всегда следить за точным и полным пониманием значения терминов и содержания понятий, используемых в тексте.

Требования к представлению и оформлению результатов самостоятельной работы

Самостоятельная работа включает в себя повторение теоретического и практического материала дисциплины, заслушиваемого и конспектируемого в ходе аудиторных занятий; изучение основной и дополнительной литературы, указанной в рабочей учебной программе дисциплины, самоконтроль ответов на основные проблемные вопросы по темам занятий, самостоятельный повтор действий, осуществляемых в ходе выполнения лабораторных работ, в том числе при работе со специальным программным обеспечением.

Результаты самостоятельной работы представляются и оформляются в виде документации, по теме лабораторной работы.

Критерии оценки выполнения самостоятельной работы

Критерии оценки выполнения самостоятельной работы – правильность ответов на вопросы по темам теоретической части дисциплины, достижение правильного результата при осуществлении собственных действий по лабораторным работам.

Подготовка к промежуточной аттестации по дисциплине: экзамену (зачету)

К аттестации допускаются студенты, которые систематически в течение всего семестра посещали и работали на занятиях и показали уверенные знания в ходе выполнения лабораторных работ.

Непосредственная подготовка к аттестации осуществляется по вопросам, представленным в рабочей учебной программе. Тщательно изучите формулировку каждого вопроса, вникните в его суть, составьте план ответа. Обычно план включает в себя:

- определение сущности рассматриваемого вопроса, основных положений, утверждений, определение необходимости их доказательства;
- запись обозначений, формул, необходимых для полного раскрытия вопроса;
- графический материал (таблицы, рисунки, графики), необходимые для раскрытия сущности вопроса;
- роль и значение рассматриваемого материала для практической деятельности, примеры использования в практической деятельности.

Вопросы для собеседования

1. Определение основных терминов (модель предметной области, объект, класс, алгоритм обучения, решающее правило, признаковое описание, обучающая и контрольная выборка, модельные и реальные данные).
2. Классификатор, ошибки 1-ого и 2-ого рода, проблема переобучения.
3. Основные типы задач машинного обучения: обучение с учителем.
4. Основные типы задач машинного обучения: обучение без учителя.
5. Примеры задач: классификации, восстановления регрессии, ранжирования, кластеризации.
6. Слабая и сильная постановки задачи чемпионата мира среди алгоритмов классификации.
7. Постановка задачи обучения по прецедентам (объекты и признаки, ответы и типы задач, модель алгоритмов, метод обучения, этап обучения и этап применения, функционалы качества).

Методические указания к выполнению проекта

Целью индивидуального проекта является исследование свойств метода машинного обучения на различных наборах данных. Для достижения этой цели необходимо выбрать один из методов машинного обучения, реализовать его на любом языке программирования (либо воспользоваться существующей реализацией), после чего выбрать не менее 3 наборов данных

с сайта UC Machine Learning Repository (<http://archive.ics.uci.edu/ml/index.php>) и провести исследование. Для этого нужно разбить каждый набор данных на обучающую и контрольную выборки, после чего провести обучение и контроль метода. В результате необходимо сформировать оценку свойств исследуемого метода в зависимости от типа и состава использованных наборов данных – построить сравнительную таблицу с указанием процента ошибок на контрольной выборке. Также нужно определить те значения параметров метода, при которых достигаются наилучшие результаты (наименьший процент ошибок на контрольной выборке).

Темы индивидуальных заданий (проектов)

1. Метрические методы классификации: метод ближайшего соседа, метод k ближайший соседей, метод окна Парзена, метод потенциальных функций.
2. Логические методы классификации: жадный алгоритм построения дерева ID3, достоинства и недостатки решающих деревьев ID3.
3. Байесовские методы классификации: наивный байесовский классификатор, оптимальный байесовский классификатор.
4. Методы кластеризации: графовые методы кластеризации (алгоритм выделения связанных компонент, алгоритм КНП (кратчайший незамкнутый путь), алгоритм ФОРЭЛ (ФОРмальные Элементы)).
5. Методы кластеризации: иерархическая кластеризация (агломеративная иерархическая кластеризация, формула Ланса-Уильямса, визуализация кластерной структуры).

Критерии оценки проектов

- 100-86 баллов выставляется, если студент/группа точно определили содержание и составляющие части задания, умеют аргументировано отвечать на вопросы, связанные с заданием. Продемонстрировано знание и владение навыками самостоятельной исследовательской работы по теме. Фактических ошибок, связанных с пониманием проблемы, нет.

- 85-76 - баллов - работа студента/группы характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет.

- 75-61 балл – проведен достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимание базовых основ и теоретического обоснования выбранной темы. Привлечены основные

источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы

- 60-50 баллов - если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая темы. Допущено три или более трех ошибок смыслового содержания раскрываемой проблемы

IX. ФОНДЫ ОЦЕНОЧНЫХ СРЕДСТВ

№ п/п	Контролируемые разделы/темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства – наименование	
				текущий контроль	промежуточная аттестация
1	Темы 1-3	ОПК2	знает	Собеседование УО-1, Круглый стол УО-4,	Экзамен Вопросы 1-11 (теоретическая часть) и 1-3 (практическая)
			умеет	Практические задания 1-3 ПР-6	
2	Темы 4-9	ПК6	знает	Собеседование УО-1, Круглый стол УО-4,	Экзамен Вопросы 4-6 (практическая часть)
			умеет	Практические задания 4-5 ПР-6	
		ОПК1	знает	Собеседование УО-1, Круглый стол УО-4,	Экзамен Вопросы 7-10 (практическая часть)
			умеет	Практические задания 6-7 ПР-6	
		ПК10	знает	Собеседование УО-1, Круглый стол УО-4,	Экзамен Вопросы 11-13 (практическая часть)
			умеет	Практические задания 8-9 ПР-6	

Шкала оценивания уровня сформированности компетенций

Код и формулировка компетенции	Этапы формирования компетенции		критерии	показатели
ОПК-1. Способен применять	знает (пороговый)	основные группы методов машинного	Знание областей применения, преимуществ и	Способность выбрать наиболее подходящий метод машинного

естественнонаучные и инженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	уровень)	обучения, базовые алгоритмы, реализующие эти методы, а также их достоинства и недостатки.	недостатков основных методов машинного обучения.	обучения для произвольной предметной области.
	умеет (продвинутый)	программировать, настраивать и применять алгоритмы машинного обучения, необходимые для обработки данных.	Умение реализовать заданный метод машинного обучения и применить его к заданной обучающей выборке.	Способность подготовить набор данных и обучить на его основе заданный метод машинного обучения.
	владеет (высокий)	методами создания программных средств для решения задач машинного обучения и подходами к их верификации	Владение навыками формирования решающих правил и способами оценки их свойств (верификации).	Способность оценить решающее правило с точки зрения оценки времени и качества его работы на контрольной выборке.
ОПК-2. Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	знает (пороговый уровень)	модели и методы машинного обучения, применяемые при решении практических задач.	Знание способов получения оценок моделей и методов машинного обучения, а также критериев их сравнения.	Способность выбрать один из набора методов машинного обучения для произвольной предметной области на основе оценки и анализа их свойств.
	умеет (продвинутый)	использовать и сравнивать различные модели и методы машинного обучения.	Умение поставить задачу машинного обучения в заданной предметной области и предложить метод ее решения.	Способность проанализировать произвольную предметную область на предмет необходимости применения в ней методов машинного обучения, способность поставить и обосновать задачу.
	владеет (высокий)	навыками применения программных средств при решении практических задач,	Владение навыками работы с программными средствами и средами, реализующими методы	Способность реализовать заданный метод машинного обучения при помощи программного средства, а также получить результат на основе

		связанных с машинным обучением	машинного обучения в своих библиотеках.	выборки реальных данных.
ПК-6. Владение навыками моделирования, анализа и использования формальных методов конструирования программного обеспечения	знает (пороговый уровень)	особенности выбора признаков моделей и предварительной обработки данных	Знание основных принципов моделирования предметных областей.	Способность сформировать набор признаков упрощенной математической модели для произвольной предметной области.
	умеет (продвинутый)	формировать набор признаков модели и проводить предварительную обработку данных	Умение декомпозировать систему на подсистемы.	Способность выделить в системе подсистемы, способность для каждой подсистемы выделить наиболее характерные ее свойства и характеристики ее объектов.
	владеет (высокий)	технологиями оценивания и выявления информативных признаков модели	Владение навыками выделения и анализа совокупностей признаков, имеющих зависимости в изменении их значений, а также в их влиянии на конечный результат.	Способность выделить из группы признаков подгруппу, в наибольшей степени влияющей на результат решения конкретной задачи машинного обучения.
ПК-10. Владение навыками использования различных технологий разработки программного обеспечения	знает (пороговый уровень)	сильные и слабые стороны методов машинного обучения при решении конкретных практических задач	Знание назначений и ограничений методов машинного обучения.	Способность сформулировать преимущества и недостатки применения выбранного метода машинного обучения данных для решения конкретной прикладной задачи.
	умеет (продвинутый)	оценивать степень эффективности применения современных методов машинного обучения при решении конкретных практических	Умение оценить степень применимости, а также временную и вычислительную сложности выбранного метода машинного обучения для решения конкретной	Способность оценить временные и ресурсные затраты на решения конкретной прикладной задачи выбранным методом машинного обучения.

		задач	прикладной задачи.	
	владеет (высокой)	навыками отбора подходящего метода машинного обучения в зависимости от решаемой задачи	Владение способностью провести анализ степени применимости имеющихся методов машинного обучения для решения конкретной прикладной задачи.	Способность выбрать из имеющегося набора методов машинного обучения одного, наиболее подходящего с точки зрения экономии ресурсов, а также качества получаемого результата.

Методические рекомендации, определяющие процедуры оценивания результатов освоения дисциплины

Текущая аттестация студентов. Текущая аттестация студентов проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

Текущая аттестация проводится в форме защиты проекта и осуществляется ведущим преподавателем.

Объектами оценивания выступают:

- степень усвоения теоретических знаний - оценивается в форме собеседования;
- уровень овладения практическими умениями и навыками – оценивается в форме защиты проекта.

Промежуточная аттестация студентов. Промежуточная аттестация студентов проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

По дисциплине «Методы машинного обучения» предусмотрен экзамен (устный опрос в форме ответов на вопросы экзаменационных билетов).

Текущий контроль

Текущий контроль предполагает систематическую проверку усвоения учебного материала, сформированности компетенций или их элементов, регулярно осуществляемую на протяжении изучения дисциплины, в соответствии с ее рабочей программой.

Состоит в проверке правильности выполнения заданий по самостоятельной работе. Задание зачтено, если нет ошибок. По текущим ошибкам даются пояснения.

Тесты предназначены для проверки знаний по компетенциям. Проверка достижения умений и навыков по компетенциям проверяется выполнением практических работ и курсовой работы.

Промежуточный контроль

Промежуточный контроль осуществляется в конце семестра и завершает изучение дисциплины. Помогает оценить более крупные совокупности знаний и умений, сформированность определенных профессиональных компетенций по дисциплине. Промежуточный контроль проводится в форме экзамена, допуск к экзамену возможен для обучающихся, получивших оценку «зачтено» в результате выполнения самостоятельной работы и успешно выполнившие все лабораторные работы.

Оценочные средства для текущей аттестации

Вопросы для собеседования

1. Определение основных терминов (модель предметной области, объект, класс, алгоритм обучения, решающее правило, признаковое описание, обучающая и контрольная выборка, модельные и реальные данные).
2. Классификатор, ошибки 1-ого и 2-ого рода, проблема переобучения.
3. Основные типы задач машинного обучения: обучение с учителем.
4. Основные типы задач машинного обучения: обучение без учителя.
5. Примеры задач: классификации, восстановления регрессии, ранжирования, кластеризации.
6. Слабая и сильная постановки задачи чемпионата мира среди алгоритмов классификации.
7. Постановка задачи обучения по прецедентам (объекты и признаки, ответы и типы задач, модель алгоритмов, метод обучения, этап обучения и этап применения, функционалы качества).

Темы проектов

1. Метрические методы классификации: метод ближайшего соседа, метод k ближайший соседей, метод окна Парзена, метод потенциальных функций.
2. Логические методы классификации: жадный алгоритм построения дерева ID3, достоинства и недостатки решающих деревьев ID3.
3. Байесовские методы классификации: наивный байесовский классификатор, оптимальный байесовский классификатор.
4. Методы кластеризации: графовые методы кластеризации (алгоритм выделения связных компонент, алгоритм КНП (кратчайший незамкнутый путь), алгоритм ФОРЭЛ (ФОРмальные ЭЛементы)).
5. Методы кластеризации: иерархическая кластеризация (агломеративная иерархическая кластеризация, формула Ланса-Уильямса, визуализация кластерной структуры).

Критерии оценки проектов

- 100-86 баллов выставляется, если студент/группа точно определили содержание и составляющие части задания, умеют аргументировано отвечать на вопросы, связанные с заданием. Продемонстрировано знание и владение навыками самостоятельной исследовательской работы по теме. Фактических ошибок, связанных с пониманием проблемы, нет.
- 85-76 - баллов - работа студента/группы характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет.
- 75-61 балл – проведен достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимание базовых основ и теоретического обоснования выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы
- 60-50 баллов - если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая темы. Допущено три или более трех ошибок смыслового содержания раскрываемой проблемы

Шкала оценивания

Менее 60 баллов	незачтено	неудовлетворительно
-----------------	-----------	---------------------

От 61 до 75 баллов	зачтено	удовлетворительно
От 76 до 85 баллов	зачтено	хорошо
От 86 до 100 баллов	зачтено	отлично

Оценочные средства для промежуточной аттестации

Вопросы к экзамену

Теоретическая часть:

1. Определение машинного обучения как направления исследований, его цели и задачи, особенности, основные области применения методов.
2. Типы обучения (дедуктивное, индуктивное, комбинированное). Типичный сценарий индуктивного обучения.
3. Признаковое описание объектов обучающей и контрольной выборок. Основные типы признаков и критерии их отбора.
4. Задача обучения по прецедентам, обобщающая способность, метод и алгоритм обучения, решающее правило, этапы обучения и применения.
5. Критерий качества обучения (функционал качества), ошибка, эмпирический риск, сведение задачи обучения к задаче оптимизации.
6. Классификатор, ошибки 1-ого и 2-ого рода, проблема переобучения, задача чемпионата мира среди алгоритмов классификации.
7. Основные типы задач обучения с учителем (классификация, регрессия, ранжирование, прогнозирование), их постановки и отличия.
8. Основные типы задач обучения без учителя (кластеризация, поиск ассоциативных правил, фильтрация выбросов, сокращение размерности, заполнение пропущенных значений), их постановки и отличия.
9. Другие типы задач: частичное обучение, обучение с подкреплением, динамическое обучение, активное обучение, метаобучение.
10. Примеры задачи классификации: медицинская диагностика, оценивание заемщиков, предсказание ухода клиента. Пример задачи восстановления регрессии: прогнозирование потребительского спроса.
11. Пример задачи ранжирования: ранжирование текстовых документов. Примеры задачи кластеризации: рубрикация текстов, рекомендации.

Практическая часть:

1. Метрические методы классификации: гипотеза компактности, метрика, метрическое пространство, обобщенный метрический классификатор.

2. Метрические методы классификации: метод ближайшего соседа, метод k ближайших соседей, метод окна Парзена, метод потенциальных функций.

3. Метрические методы классификации: понятие отступа, типы объектов (в зависимости от отступа), алгоритм STOLP.

4. Логические методы классификации: логическая закономерность, интерпретируемость, информативность, часто используемые виды закономерностей, часто используемые критерии информативности, обобщенный алгоритм нахождения информативных закономерностей.

5. Логические методы классификации: определение бинарного решающего дерева, пример решающего дерева, жадный алгоритм построения дерева ID3, достоинства и недостатки решающих деревьев ID3.

6. Логические методы классификации: определение бинарного решающего дерева, редукция дерева, небрежные решающие деревья ODT, задача бинаризации вещественного признака.

7. Линейные методы классификации (метод стохастического градиента): задача построения разделяющей поверхности, линейный классификатор, математическая модель нейрона, градиентный метод численной минимизации, алгоритм SG.

8. Линейные методы классификации (метод опорных векторов): кусочно-линейная аппроксимация, задача SVM, оптимальная разделяющая гиперплоскость, достоинства и недостатки метода.

9. Байесовские методы классификации: байесовский классификатор апостериорная вероятность, вероятностная постановка задачи, функционал среднего риска, оптимальный байесовский классификатор.

10. Байесовские методы классификации: разбиение задачи классификации на две подзадачи и алгоритм решение первой из них, три основных подхода к оцениванию плотностей, наивный байесовский классификатор.

11. Методы кластеризации: постановка задачи кластеризации, некорректность задачи кластеризации, цели кластеризации, типы кластерных структур, проблема чувствительности к выбору метрики.

12. Методы кластеризации: графовые методы кластеризации (алгоритм выделения связанных компонент, алгоритм кратчайшего незамкнутого пути, алгоритм ФОРЭЛ, функционалы качества кластеризации).

13. Методы кластеризации: агломеративная иерархическая кластеризация (алгоритмы Ланса-Уильямса (обычный и редуцированный), формула Ланса-Уильямса и ее частные случаи, визуализация кластерной структуры).

Критерии выставления оценки студенту на экзамене по дисциплине «Методы машинного обучения»

Баллы (рейтинговой оценки)	Оценка зачета/ экзамена (стандартная)	Требования к сформированным компетенциям
86-100	«зачтено»/ «отлично»	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, правильно обосновывает принятое решение, владеет разносторонними навыками и приемами выполнения практических задач.
76-85	«зачтено»/ «хорошо»	Оценка «хорошо» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения.
61-75	«зачтено»/ «удовлетворительно»	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении практических работ.
0-60	«не зачтено»/ «неудовлетворительно»	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет практические работы. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных

		занятий по соответствующей дисциплине.
--	--	--