



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Дальневосточный федеральный университет»
(ДВФУ)

ШКОЛА ЕСТЕСТВЕННЫХ НАУК

«СОГЛАСОВАНО»
Руководитель ОП


(подпись)
« »

Пустовалов Е.В.
(Ф.И.О. рук. ОП)
2018 г.



«УТВЕРЖДАЮ»

Заведующий (ая) кафедрой
компьютерных систем
(название кафедры)


(подпись)
« »

Пустовалов Е.В.
(Ф.И.О. зав. каф.)
2018 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Методы машинного обучения

Направление подготовки 09.04.02 Информационные системы и технологии

магистерская программа «Информационные процессы в науке, промышленности и образовании»

Форма подготовки очная

курс I семестр 2

лекции 0 час.

практические занятия 36 час.

лабораторные работы 0 час.

в том числе с использованием МАО лек. /пр. 18 /лаб. час.

всего часов аудиторной нагрузки 36 час.

в том числе с использованием МАО 18 час.

самостоятельная работа 72 час.

в том числе на подготовку к экзамену час.

контрольные работы (количество) шт.

курсовая работа / курсовой проект семестр

зачет 2 семестр

экзамен не предусмотрен

Рабочая программа составлена в соответствии с требованиями образовательного стандарта, самостоятельно устанавливаемого ДВФУ, утвержденного приказом ректора от 07.07.2015 г. № 12-13-1282.

Рабочая программа обсуждена на заседании кафедры компьютерных систем, протокол № 18 от «16» июля 2018 г.

Заведующий кафедрой Пустовалов Е.В.

Составитель: ассистент кафедры компьютерных систем Солдатов К.С.

Оборотная сторона титульного листа РПУД

I. Рабочая программа пересмотрена на заседании кафедры:

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

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

II. Рабочая программа пересмотрена на заседании кафедры:

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

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

АННОТАЦИЯ

Рабочая учебная программа дисциплины «Методы машинного обучения» разработана для магистров первого года обучения и предназначена для изучения во втором семестре по направлению 09.04.02 «Информационные системы и технологии», профиль «Информационные процессы в науке, промышленности и образовании». В соответствии с требованиями соответствующих образовательных стандартов. Трудоемкость дисциплины составляет 3 зач.ед., 108 ак.часов.

Цель.

Дисциплина является теоретической и практической частью подготовки магистров по использованию информационных и коммуникационных технологий (ИКТ) в профессиональной деятельности.

Задачи:

- Изучение основных положений нейронных сетей для дискретных событий
- Изучение основных положений нейронных сетей для непрерывных событий
- Изучения методов кодирования и характеристик кодов

В результате изучения данной дисциплины у обучающихся формируются следующие общекультурные/ общепрофессиональные/профессиональные элементы компетенций.

Код и формулировка компетенции	Этапы формирования компетенции	
ОПК-3 способностью анализировать и оценивать уровни своих компетенций в сочетании со способностью и готовностью к саморегулированию дальнейшего образования и профессиональной мобильности	Знает	Техники сбора информации, методы дедуктивной и индуктивной обработки информации
	Умеет	воспринимать математические, естественнонаучные, социально-экономические и профессиональные знания
	Владеет	умением самостоятельно приобретать, развивать и применять знания для решения нестандартных задач

ОПК-6 способностью анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	Знает	Что такое интернет-браузер, всемирная паутина, интернет-сайт, поисковая система, wifi-сеть, электронная почта
	Умеет	Печатать тексты на клавиатуре, преобразовывать мысленный поток в последовательный и содержательный поисковый запрос, переходить по интернет-ссылкам, пользоваться текстовым редактором, почтовым клиентом
	Владеет	иностранном языком на уровне профессионального общения
ПК-14 умением осуществлять моделирование процессов и объектов на базе стандартных пакетов автоматизированного проектирования и исследований	Знает	Методы проектирования информационных систем
	Умеет	Применять методы проектирования информационных систем
	Владеет	Информацией о передовых технологиях разработки ИС

Для формирования вышеуказанных компетенций в рамках дисциплины «Методы машинного обучения» применяются следующие методы активного обучения: Работа в малых группах, Творческое задание по разработке и презентации одного из подразделов.

I. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

Лекции не предусмотрены

II. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА

Практические занятия (36 час.)

Занятие 1. Задачи классификации (3 час.)

1. Имеются две урны, содержащие по 20 шаров — 10 белых, 5 черных и 5 красных в первой и 8 белых, 8 черных и 4 красных во второй. Из каждой урны вытаскивают по одному шару. Исход какого из этих двух опытов следует считать более неопределенным?
2. Пусть из многолетних наблюдений за погодой известно, что для определенного пункта вероятность того, что 15 июня будет идти дождь, равна 0,4, а вероятность того, что в указанный день дождя не будет, равна 0,6. Пусть далее для этого же пункта вероятность того, что 15 ноября будет идти дождь равна 0,65, вероятность того, что 15 ноября будет идти снег, равна 0,15 и вероятность того, что 15 ноября вообще не будет осадков, равна 0,2. Если из всех характеристик погоды интересоваться лишь вопросом о наличии и о характере осадков, то в какой из двух перечисленных дней погоду в рассматриваемом пункте следует считать более неопределенной?

Занятие 2. Задачи кластеризации (3 час.)

1. Известно, что некоторой болезнью в среднем болеют 2 человека из 100. Для выявления болезни используется определенная реакция, которая всегда оказывается положительной в том случае, когда человек болен; если же человек здоров, то она столь же часто бывает положительной, как и отрицательной. Пусть опыт b стоит в определении того, болен или здоров человек, а опыт a — в определении результата указанной реакции. Спрашивается, какова будет энтропия $H(b)$ опыта b и условная энтропия $H(b|a)$ опыта b при условии осуществления a .
2. Пусть опыты a и b состоят в последовательном извлечении двух шаров из урны, содержащей m черных и $n - m$ белых шаров (a — извлечение (определение его цвета) первого шара, b — извлечение (определение его цвета) второго шара). Чему равны энтропии $H(a)$ и

$H(b)$ опытов a и b и условные энтропии $H(b|a)$ и $H(a|b)$ тех же опытов? Решите ту же задачу при условии, что опыт a состоит в извлечении k шаров из урны, а опыт b — в последующем извлечении еще одного шара.

Занятие 3. Задачи регрессии и прогнозирования (3 час.)

1. Определить дифференциальную энтропию равномерного на интервале $\{-W_1; +W_2\}$ распределения
2. Как изменится дифференциальная энтропия случайного распределения с нормальным распределением если: а) среднее значение увеличится в 2 раза; б) дисперсия уменьшится в 2 раза.

Занятие 4. Модель нейрона (3 час.)

1. Пусть для некоторого пункта вероятность того, что 15 июня будет идти дождь, равна 0,4, а вероятность того, что дождя не будет, равна 0,6. Пусть далее для этого же пункта вероятность дождя 15 октября равна 0,8, а вероятность отсутствия дождя в этот день — всего 0,2. Предположим, что определенный метод прогноза погоды 15 июня оказывается правильным в $3/5$, всех тех случаев, в которых предсказывается дождь, и в $4/5$ тех случаев, в которых предсказывается отсутствие осадков; в применении же к погоде 15 октября этот метод оказывается правильным в $9/10$ тех случаев, в которых предсказывается дождь, и в половине случаев, в которых предсказывается отсутствие дождя (сравнительно большой процент ошибок в последнем случае естественно объясняется тем, что предсказывается маловероятное событие, предугадать которое довольно трудно). Спрашивается, в какой из двух указанных дней прогноз дает нам больше информации о реальной погоде?

2. Пусть опыт β состоит в извлечении одного шара из урны, содержащей 5 черных и 10 белых шаров, опыт α_k — в предварительном извлечении из той же урны (без возвращения обратно) k шаров. Чему равна энтропия опыта β и информация об этом опыте, содержащаяся в опытах $\alpha_1, \alpha_2, \alpha_{13}, \alpha_{14}$

Занятие 5. Решение задач по нахождению количества информации в непрерывном случае (3 час.)

1. Чему равно приращение энтропии системы ΔH при переходе системы из состояния, характеризуемого среднеквадратическим отклонением σ_1 , в состояние, характеризуемое величиной σ_2 : а) в случае нормально распределения координаты, б) в случае равномерного распределения координаты?
2. Информация передается при помощи частотно-модулированных сигналов, рабочая частота F которых изменяется с равной вероятностью в пределах от $F_1=10\text{МГц}$ до $F_2=50\text{МГц}$. Определить энтропию частоты, если точность измерения частоты ΔF равна 2кГц .

Занятие 6. Закрепление материала по предыдущим занятиям (3 час.)

1. Максимальная энтропия опыта с количеством исходов M
2. Энтропия случайного события
3. Энтропия сложного опыта, независимые события
4. Условная энтропия
5. Средняя условная энтропия
6. Энтропия сложного опыта, зависимые события
7. Дифференциальная энтропия случайного события
8. Взаимная информация в дискретном случае
9. Информация случайного события в непрерывном случае

Занятие 7. Решение задач по определению пропускной способности линии связи без помех (3 час.)

1. Радиостанция может работать на волне λ_1 (событие A_1) или на волне λ_2 (событие A_2); в импульсном (событие B_1) или непрерывном (событие B_2) режимах. Вероятности совместных событий имеют следующие значения: $P(A_1B_1)=0.7$; $P(A_1B_2)=0.15$; $P(A_2B_1)=0.05$; $P(A_2B_2)=0.1$. Вычислить количество информации, получаемой относительно режима работы станции, если станет известна длина волны станции.
2. Определить среднюю взаимную информацию между двумя буквами алфавита, если известно, что средняя энтропия алфавита равна 5 бит, а энтропия на пару букв равняется 8.3 бита.

Занятие 8. Решение задач по определению пропускной способности линии связи с помехами (3 час.)

1. Сигнал S подается на вход канала с вероятностью 0.6 и отсутствует на входе с вероятностью 0.4. Поступивший сигнал воспроизводится на выходе канала с вероятностью 0.8 и теряется с вероятностью 0.2. при отсутствии сигнала на выходе возможен ложный сигнал S' на выходе с вероятностью 0.3. Определить среднее количество информации о входном сигнале по фиксируемому выходному.
2. Определить количество информации, передаваемое по двоичному симметричному каналу, если $p_1=0,2$; $p_0=0.8$.

Занятие 9. Решение задач по кодированию информации (3 час.)

1. Для двоичного источника с вероятностью появления «1» 0.1 исследовать зависимость эффективности равномерного кода от длины кодируемых блоков (до 4)

Занятие 10. Построение кодов Шеннона-Фано и Хаффмана (3 час.)

1. Построить код Шеннона-Фано для ансамбля сообщений с вероятностями $1/6, 1/6, \dots, 1/6$. Определить характеристики кода.
2. В сообщениях используются символы алфавита A_1, A_2, A_3, A_4 с вероятностями соответственно $0.45; 0.1; 0.15; 0.3$. Для передачи сообщения по каналу связи могут быть применены два кода. В первом символам алфавита соответствуют кодовые последовательности a, b, c, d , во втором a, d, b, c . Длительности кодовых последовательностей в условных единицах равны: $t_a=8, t_b=6, t_c=5, t_d=3$. Определить количество информации, передаваемое каждым кодом в единицу времени. Построить, используя данные кодовые последовательности оптимальный код.

Занятие 11. Построение (N, M) кодов (3 час.)

1. Построить (N, M) код с длиной информационной части 3, обнаруживающий одиночные ошибки.
2. Определить эффективность кода обнаруживающего и исправляющего одиночные ошибки, если его общая длина равна 3.

Занятие 12. Закрепление материала по предыдущим занятиям (3 час.)

1. Линия связи без помех
2. Линия связи с помехами
3. Мгновенные коды. Равномерные и неравномерные коды. Экономность кода.
4. Коды Шеннона-Фано, код Хаффмана.
5. (N, M) коды, матричное задание кодов.

III. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине «Методы машинного обучения» представлено в Приложении 1 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.

IV. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Раздел I. Обучение простых алгоритмов классификации	ОПК-3 ОПК-6 ПК-14	знает	ПР-1	Зачет, вопросы 1-26
			умеет	ПР-1	Зачет, задание 1-20
			владеет	УО-1	Решение задач
	Раздел II. Обзор классификаторов на основе машинного обучения	ОПК-3 ОПК-6 ПК-14	знает	ПР-2	Зачет, вопросы 1-26
			умеет	ПР-2	Зачет, задание 1-20
			владеет	УО-2	Решение задач
	Раздел III. Построение обучающих наборов	ОПК-3 ОПК-6 ПК-14	знает	ПР-3	Зачет, вопросы 1-26
			умеет	ПР-3	Зачет, задание 1-20
			владеет	УО-3	Решение задач
	Раздел IV. Дополнительн	ОПК-3 ОПК-6 ПК-14	знает	ПР-4	Зачет, вопросы 1-26

	ые API интерфейсы		умеет	ПР-4	Зачет, задание 1-20
			владеет	УО-4	Решение задач

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 Приложении 2.

V. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

(электронные и печатные издания)

№	Название	Ссылка в ЭК НБ ДВФУ	Внешняя ссылка
1	Шарден Б., Массарон Л., Боскетти А., «Крупномасштабное машинное обучение вместе с Python», ДМК Пресс, 2018, 358 стр.	https://lib.dvfu.ru:8443/lib/item?id=Lan:Lan-105836&theme=FEFU	
2	Флах П. «Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных», ДМК Пресс, 2015, 400 стр.	https://lib.dvfu.ru:8443/lib/item?id=Lan:Lan-69955&theme=FEFU	
3	Рашка С. «Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения», ДМК Пресс, 2017, 418 стр.	https://lib.dvfu.ru:8443/lib/item?id=Lan:Lan-100905&theme=FEFU	

Дополнительная литература

(печатные и электронные издания)

№	Название	Ссылка в ЭК НБ ДВФУ	Внешняя ссылка
1	Федотов И.Е. Модели параллельного программирования/ Федотов И.Е.— М.: СОЛОН-ПРЕСС, 2012.— 384 с.		http://www.iprbookshop.ru/20877
2	Алексеев А.А. Основы параллельного программирования с использованием Visual Studio 2010/ Алексеев А.А.— М.: Интернет-Университет Информационных Технологий (ИНТУИТ), 2013.— 138 с.		http://www.iprbookshop.ru/16714
3	Среда разработки параллельных программ ЭВМ: учеб. пособие / [А.О. Блинов и др.] под ред. А.О. Блинова. – М. : ЮНИТИ-ДАНА, 2014. – 341 с.		http://www.iprbookshop.ru/16437
4	Богачев К.Ю. Основы параллельного программирования: учебное пособие/ Богачев К.Ю.— Электрон. текстовые данные.— М.: БИНОМ. Лаборатория знаний, 2013.— 342 с.		http://www.iprbookshop.ru/20702

Перечень информационных технологий и программного обеспечения

Для проведения лабораторных занятий на рабочей станции (или сервере) должны быть установлены:

- ОС семейства Windows/Linux/UNIX
- текстовый редактор (mcedit, nano или аналоги)
- компилятор C/C++ (gcc/g++, icc или аналоги)
- Python 3+
- офисное ПО для оформления отчетов (LibreOffice или аналоги)

веб-браузер для доступа к документации через сеть Интернет

VI. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

При изучении дисциплины 36 часов отводится на аудиторные занятия и 72 часа на самостоятельную работу студентов. Из 36 часов аудиторных занятий 36 практических с использованием методов активного обучения. С целью более эффективного освоения материала рекомендуется практические занятия проводить в конце семестра, когда студенты будут знакомы с теоретическим содержанием курса.

При подготовке к зачету рекомендуется просмотреть материалы лекций и собственные конспекты, разбить вопросы по разделам и темам, затем определить содержание ответов на вопросы.

Выполнение заданий в малых группах и творческого задания предусматривает выбор подраздела курса и согласование его с преподавателем.

VII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Для проведения исследований, связанных с выполнением задания по практике, а также для организации самостоятельной работы студентам доступно следующее лабораторное оборудование и специализированные кабинеты, соответствующие действующим санитарным и противопожарным

нормам, а также требованиям техники безопасности при проведении учебных и научно-производственных работ:

Наименование оборудованных помещений и помещений для самостоятельной работы	Перечень основного оборудования
Владивосток, о. Русский, п. Аякс д.10, корпус L, ауд. L 502 учебная аудитория для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации	Мультимедийное оборудование: ЖК-панель 47", Full HD, LG M4716 CCBA - 1 шт. Парты и стулья
г. Владивосток, о. Русский, п. Аякс д.10, корпус D, ауд. D734 учебная аудитория для проведения занятий семинарского типа, практических занятий: компьютерный класс	Моноблок HPP-B0G08ES#ACB/8200E AIO i52400S 500G 4.0G 28 PC - 15 шт Мультимедийное оборудование: Экран проекционный ScreenLine Trim White Ice 50 см черная кайма сверху, размер рабочей области 236x147 см Документ-камера Avervision CP355AF ЖК-панель 47", Full HD, LG M4716 CCBA Мультимедийный проектор Mitsubishi EW330U, 3000 ANSI Lumen, 1280x800 Сетевая видеочка Multipix MP-HD718
Читальные залы Научной библиотеки ДВФУ с открытым доступом к фонду (корпус А - уровень 10) Аудитории для самостоятельной работы	Читальные залы Научной библиотеки ДВФУ с открытым доступом к фонду: Моноблок Lenovo C360G-i34164G500UDK – 15 шт. Интегрированный сенсорный дисплей Polymedia FlipBox - 1 шт. Копир-принтер-цветной сканер в e-mail с 4 лотками Xerox WorkCentre 5330 (WC5330C – 1 шт. Скорость доступа в Интернет 500 Мбит/сек. Рабочие места для людей с ограниченными возможностями здоровья оснащены дисплеями и принтерами Брайля; оборудованы: портативными устройствами для чтения плоскочечатных текстов, сканирующими и читающими машинами видеоувеличителем с возможностью регуляции цветовых спектров; увеличивающими электронными лупами и ультразвуковыми маркировщиками

В целях обеспечения специальных условий обучения инвалидов и лиц с ограниченными возможностями здоровья в ДВФУ все здания оборудованы

пандусами, лифтами, подъемниками, специализированными местами, оснащенными туалетными комнатами, табличками информационно-навигационной поддержки.



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

ШКОЛА ЕСТЕСТВЕННЫХ НАУК

**УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ
РАБОТЫ ОБУЧАЮЩИХСЯ**

по дисциплине «Методы машинного обучения»

**Направление подготовки 09.04.02 Информационные системы и
технологии**

**магистерская программа «Информационные процессы в науке,
промышленности и образовании»**

Форма подготовки очная

Владивосток

2018

План-график выполнения самостоятельной работы по дисциплине

№ п/п	Дата/сроки выполнения	Вид самостоятельной работы	Примерные нормы времени на выполнение	Форма контроля
1	1-18 неделя обучения, еженедельно (2 часа в неделю)	Домашнее задание	36	Проверка ДЗ
2	12-16 неделя обучения	Подготовка творческого задания	18	Выступление по результатам
3	18 неделя, зачетная неделя	Подготовка к экзамену	18	зачет

Самостоятельная работа по выполнению домашнего задания должна включать в себя повторение лекционного материала, повторение формул по разделу, повторение решенных задач по разделу, решение задач домашнего задания по разделу. Выполненное задание должно быть оформлено в соответствии с требованиями по оформлению решения задач, текст, формулы легко читаемы.

Самостоятельная работа по подготовке творческого задания должна включать в себя поиск информации в сети Интернет и рекомендуемых источниках, обсуждение основных характеристик, подготовка черновиков презентаций и текста выступления. Презентация должна быть не менее 15 слайдов, выступление продолжительностью 15-18 минут. Каждая малая группа должна подготовить вопросы для остальных групп по их темам.

Самостоятельная работа по подготовке к экзамену должна включать повторение теоретического материала, подготовку ответов на вопросы с использованием лекций и рекомендуемых источников.

Оценка результатов самостоятельной работы по подготовке творческого задания выполняется по следующим критериям:

5 баллов выставляется, если студент выразил своё мнение по сформулированной проблеме, аргументировал его, точно определив ее содержание и составляющие. Приведены данные отечественной и зарубежной литературы, статистические сведения, информация нормативно-правового характера. Продемонстрированы знания и владения навыками самостоятельной исследовательской работы по теме исследования; методами и приемами анализа теоретических и/или практических аспектов изучаемой области. Фактических ошибок, связанных с пониманием проблемы, нет; графически работа оформлена правильно

4 балла - работа характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Для аргументации приводятся данные отечественных и зарубежных авторов. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет. Допущены одна-две ошибки в оформлении работы

3 балла - Студент провел достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимает базовые основы и теоретическое обоснование выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы, оформлении работы

2 балла - 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 в смысловом содержании раскрываемой проблемы, в оформлении работы.

Оценка результатов самостоятельной работы в малых группах выполняется по следующим критериям:

5 баллов выставляется, если студент/группа выразили своё мнение по сформулированной проблеме, аргументировали его, точно определив ее содержание и составляющие. Приведены данные отечественной и зарубежной литературы, статистические сведения, информация нормативно-правового характера. Продемонстрировано знание и владение навыком самостоятельной исследовательской работы по теме исследования; методами и приемами анализа международно-политической практики. Фактических ошибок, связанных с пониманием проблемы, нет

4 балла - работа студента/группы характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Для аргументации приводятся данные отечественных и зарубежных авторов. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет.

3 балла - 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

2 балла - 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



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
Высшего образования
«Дальневосточный федеральный университет»
(ДФУ)

ШКОЛА ЕСТЕСТВЕННЫХ НАУК

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине «Методы машинного обучения»

**Направление подготовки 09.04.02 Информационные системы и
технологии**

**магистерская программа «Информационные процессы в науке,
промышленности и образовании»**

Форма подготовки очная

Владивосток

2018

Паспорт ФОС

Код и формулировка компетенции	Этапы формирования компетенции	
ОПК-3 способностью анализировать и оценивать уровни своих компетенций в сочетании со способностью и готовностью к саморегулированию дальнейшего образования и профессиональной мобильности	Знает	Техники сбора информации, методы дедуктивной и индуктивной обработки информации
	Умеет	воспринимать математические, естественнонаучные, социально-экономические и профессиональные знания
	Владеет	умением самостоятельно приобретать, развивать и применять знания для решения нестандартных задач
ОПК-6 способностью анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	Знает	Что такое интернет-браузер, всемирная паутина, интернет-сайт, поисковая система, wifi-сеть, электронная почта
	Умеет	Печатать тексты на клавиатуре, преобразовывать мысленный поток в последовательный и содержательный поисковый запрос, переходить по интернет-ссылкам, пользоваться текстовым редактором, почтовым клиентом
	Владеет	иностранным языком на уровне профессионального общения
ПК-14 умением осуществлять моделирование процессов и объектов на базе стандартных пакетов автоматизированного проектирования и исследований	Знает	Методы проектирования информационных систем
	Умеет	Применять методы проектирования информационных систем
	Владеет	Информацией о передовых технологиях разработки ИС

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций	Оценочные средства		
			текущий контроль	промежуточная аттестация	
1	Раздел I. Обучение простых алгоритмов классификации	ОПК-3 ОПК-6 ПК-14	знает	ПР-1	Зачет, вопросы 1-26
			умеет	ПР-1	Зачет, задание 1-20
			владеет	УО-1	Решение задач
	Раздел II. Обзор классификаторов на основе машинного обучения	ОПК-3 ОПК-6 ПК-14	знает	ПР-2	Зачет, вопросы 1-26
			умеет	ПР-2	Зачет, задание 1-20
			владеет	УО-2	Решение задач
	Раздел III. Построение обучающих наборов	ОПК-3 ОПК-6 ПК-14	знает	ПР-3	Зачет, вопросы 1-26
			умеет	ПР-3	Зачет, задание 1-20
			владеет	УО-3	Решение задач
	Раздел IV. Дополнительные API интерфейсы	ОПК-3 ОПК-6 ПК-14	знает	ПР-4	Зачет, вопросы 1-26
			умеет	ПР-4	Зачет, задание 1-20
			владеет	УО-4	Решение задач

Шкала оценивания уровня сформированности компетенций

Код и формулировка компетенции	Этапы формирования компетенции		критерии	показатели
	знает (пороговый уровень)	новые научные принципы и методы исследований при использовании интеллектуальный анализа и хранилищ данных;		
ОПК-3, способностью анализировать и оценивать уровни своих компетенций в сочетании со способностью, и	знает (пороговый уровень)	новые научные принципы и методы исследований при использовании интеллектуальный анализа и хранилищ данных;	воспроизводить и объяснять учебный материал с требуем	способность показать базовые знания и основные умения в использовании: - принципов системного анализа к исследованию систем; - принципов технологии интеллектуального анализа и хранилищ данных;

готовностью к саморегулированию дальнейшего образования и профессиональной мобильности		инструментальные средства технологии интеллектуального анализа и хранилищ данных	ой степенью научной точности и полноты	- характеристик программных средств по реализации технологии интеллектуального анализа и хранилищ данных
	умеет (продвинутый)	использовать современное программное обеспечение для решения научных и производственных задач методами интеллектуального анализа и хранилищ данных	выполнять типичные задачи на основе воспроизведения стандартных алгоритмов решения	способность применить знания и практические умения в задачах, связанных с выбором и обоснованием проектных решений по автоматизации прикладных и информационных процессов предприятий, используя теоретические методы исследования и методы компьютерного моделирования, современные программные инструментальные средства и информационные ресурсы Интернет, используя технологии интеллектуального анализа и хранилищ данных
	владеет (высокий)	навыками применения современных информационных технологий в научно-исследовательской и производственной деятельности на основе интеллектуального анализа и хранилищ данных	решать усложненные задачи в нетипичных ситуациях на основе приобретенных знаний, умений и навыков	способность применить фактическое и теоретическое знание, практические умения по разработке проектных материалов в профессиональной области прикладной информатики, связанных с выбором и обоснованием проектных решений по автоматизации прикладных и информационных процессов предприятий, используя теоретические методы исследования и методы компьютерного моделирования, современные программные инструментальные средства и информационные ресурсы Интернет, используя технологии интеллектуального анализа и хранилищ данных
ОПК-6, способность анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	знает (пороговый уровень)	принципы проведения и оценки научных экспериментов; методы поддержки принятия решений;	воспроизводить и объяснять учебный материал с требуемой степенью научной точности и полноты	способность показать базовые знания и основные умения в использовании: - характеристик методов проведения научных экспериментов; - принципов оценивания результатов научных исследований; - характеристик методов поддержки принятия решений при организации научных экспериментов
	умеет (продвинутый)	планировать проведение научных экспериментов;	выполнять типичные	способность применить знания и практические умения в задачах, связанных с выбором и

		оценивать результаты исследований: использовать методы и средства поддержки принятия решений	е задачи на основе воспроизведения стандартных алгоритмов решения	обоснованием проектных решений по автоматизации прикладных и информационных процессов предприятий, используя научные принципы проектных исследований, методы и средства поддержки принятия решений, методы оценивания решений по критериям социальной, технической и экономической эффективности
	владеет (высокий)	компьютерными технологиями поддержки принятия решений; инструментарием проведения и оценки научных экспериментов	решать сложные задачи в нетипичных ситуациях на основе приобретенных знаний, умений и навыков	способность применить фактическое и теоретическое знание, практические умения по разработке проектных материалов в профессиональной области прикладной информатики, связанных с выбором и обоснованием проектных решений по автоматизации прикладных и информационных процессов предприятий, используя научные принципы проектных исследований, методы и средства поддержки принятия решений, методы оценивания решений по критериям социальной, технической и экономической эффективности
ПК-14, умением осуществлять моделирование процессов и объектов на базе стандартных пакетов автоматизированного проектирования и исследований	знает (пороговый уровень)	методы управления проектами по решению прикладных задач ИС; программные средства управления проектами по информатизации прикладных задач и созданию ИС предприятий	воспроизводить и объяснять учебный материал с требуемой степенью научной точности и полноты	способность показать базовые знания и основные умения в использовании: - принципов создания ИС; - общесистемных принципов по управлению ИТ проектами; - характеристик программных систем по управлению проектами; - принципов организации и управления разработкой ИС
	умеет (продвинутый)	выбирать подходы и инструментарий для проектирования информационных систем; применять программные средства для управления проектами по информатизации прикладных задач и созданию ИС	выполнять типичные задачи на основе воспроизведения стандартных алгоритмов решения	способность применить знания и практические умения в задачах, связанных с выбором и обоснованием стратегии и проектных решений по автоматизации прикладных и информационных процессов предприятий, используя требования по управлению ИТ проектами предприятий и организаций, принципы организации и управления разработкой ИС

	владеет (высокий)	навыками проектирования информационных систем с использованием инструментальных средств; навыками использования пакетов прикладных программ для управления проектами по информатизации прикладных задач и созданию ИС предприятий	решать усложненные задачи в нетипичных ситуациях на основе приобретенных знаний, умений и навыков	способность применить фактическое и теоретическое знание, практические умения по разработке проектных материалов в профессиональной области прикладной информатики, связанных с выбором и обоснованием стратегии и проектных решений по автоматизации прикладных и информационных процессов предприятий, используя требования по управлению ИТ проектами предприятий и организаций, принципы организации и управления разработкой ИС
--	-------------------	---	---	--

Методические материалы, определяющие процедуры оценивания знаний, умений и навыков

Методическим материалом является Положение о ФОС ДВФУ

Вопросы к промежуточной аттестации (зачет)

1. Энтропия как мера неопределенности.
2. Введение понятия энтропии.
3. Мера неопределенности по Хартли. Мера неопределенности по Шеннону.
4. Статистический смысл энтропии.
5. Средняя условная энтропия двух опытов. Свойства.
6. Энтропия сложных событий. Правило сложения энтропии зависимые события.
7. Энтропия сложных событий. Правило сложения энтропии независимые события.
8. Дифференциальная энтропия, дифференциальная энтропия нормального распределения.
9. Энтропия источников непрерывных сообщений.
10. Понятие об информации. Соотношение понятий энтропии и информации.

11. Информация в случае непрерывного опыта.
12. Основные понятия передачи информации по линиям связи.
Определение кода. Код Морзе. Код Бодо.
13. Линия связи без помех.
14. Линия связи с помехами.
15. Пропускная способность линии связи с помехами.
16. Двоичная симметричная линия. Двоичная симметричная линия со стиранием.
17. Основная теорема о кодировании.
18. Теорема о кодировании при наличии помех.
19. Формула Шеннона для пропускной способности непрерывного канала при наличии аддитивного шума.
20. Мгновенные коды. Равномерные и неравномерные коды. Экономность кода.
21. Коды Шеннона — Фано и Хаффмана, доказательство оптимальности кодов Хаффмана.
22. Коды исправляющие все одиночные ошибки, (N, M) –коды.
23. Коды обнаруживающие и исправляющие ошибки.
24. Коды обнаруживающие ошибки, коды исправляющие ошибки, (N, M) –коды.
25. Неравенство Хемминга, неравенство Варшамова-Гилберта.
26. Верхняя граница Хемминга числа кодовых обозначений.

Критерии оценки (зачет)

приводить примеры современных проблем изучаемой области.

4 балла - ответ, обнаруживающий прочные знания основных процессов изучаемой предметной области, отличается глубиной и полнотой раскрытия темы; владение терминологическим аппаратом; умение объяснять сущность, явлений, процессов, событий, делать выводы и обобщения, давать аргументированные ответы, приводить примеры; свободное владение

монологической речью, логичность и последовательность ответа. Однако допускается одна - две неточности в ответе.

3 балла - оценивается ответ, свидетельствующий в основном о знании процессов изучаемой предметной области, отличающийся недостаточной глубиной и полнотой раскрытия темы; знанием основных вопросов теории; слабо сформированными навыками анализа явлений, процессов, недостаточным умением давать аргументированные ответы и приводить примеры; недостаточно свободным владением монологической речью, логичностью и последовательностью ответа. Допускается несколько ошибок в содержании ответа; неумение привести пример развития ситуации, провести связь с другими аспектами изучаемой области.

2 балла - ответ, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы; незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов; неумением давать аргументированные ответы, слабым владением монологической речью, отсутствием логичности и последовательности. Допускаются серьезные ошибки в содержании ответа; незнание современной проблематики изучаемой области.

Типовые задания для текущего контроля

1. Имеется два игральных кубика, определите вероятность выпадения числа семь и энтропию такого опыта.

2. Ключ шифрования содержит 256 бит, сколько времени понадобится для его вскрытия методом перебора, если известно, что за одну секунду компьютер перебирает 10000 вариантов.

3. 1 сентября на первом курсе одного из факультетов запланировано по расписанию три лекции по разным предметам. Всего на первом курсе изучается 7 предметов. Студент, не успевший ознакомиться с расписанием, пытается его угадать. Какова вероятность успеха в данном эксперименте, если

считать, что любое расписание из четырех предметов равновероятно? Какова неопределенность такого опыта?

4. В подъезде дома установлен замок с кодом. Дверь автоматически отпирается, если в определенной последовательности набрать три цифры из имеющихся десяти. На набор одной комбинации уходит 20 секунд. Какова вероятность открыть дверь за 40 минут? Какова неопределенность такого опыта?

5. Из набора домино (28 штук) выбирают 7 штук. Какова неопределенность данного события?

6. Из колоды карт (52 карты) извлекают три карты. Какова неопределенность такого события?

7. Бросают две игральные кости. Какова неопределенность данного события?

8. Имеется 12 монет одного достоинства, 11 из них имеют одинаковый вес, а одна — фальшивая — отличается по весу от остальных (причем неизвестно, легче ли она или тяжелее настоящих). Каково наименьшее число взвешиваний на чашечных весах без гирь, которое позволяет обнаружить фальшивую монету и выяснить, легче ли она, чем остальные монеты, или тяжелее. Решить тот же вопрос для случая 13 монет.

9. Какова вероятность того, что взятое наугад целое положительное число до 1000, окажется целой степенью другого целого числа с показателем, больше единицы? Вычислите энтропию такого события.

10. Какую неопределенность содержит сообщение о событие – сдача экзамена студентом, если по опыту предыдущих экзаменов известно, что вероятность того, что студент сдал экзамен $7/8$.

11. Пусть X и Y два случайных опыта; $Z=X+Y$. Чему равна условная энтропия $H(x|z)$, если: а) X и Y независимы; б) X и Y зависимы; в) $X \equiv Y$.

12. Определить среднее количество информации, приходящееся на один символ сообщения 01001000101001, при условии, что источник эргодический, а последовательность типичная.

13. Имеются два дискретных источника с независимыми и равновероятными элементами: двоичный и троичный. На выходе первого зафиксированы два символа, на выходе второго три. Чему равны неопределенности полученных последовательностей букв, образованных парами символов первого источника и тройками символов второго?

14. Символы азбуки Морзе могут появиться в сообщении с вероятностями: для точки - 0.51, для тире - 0.31, для промежутка между буквами - 0.12, между словами - 0.06. Определить среднее количество информации в сообщении из 500 символов данного алфавита, считая, что связь между последовательными символами отсутствует.

15. Известно, что из 100 изготовленных деталей в среднем 10 деталей имеют дефекты. Для выявления брака используется метод, дающий всегда отрицательный эффект, если деталь изготовлена с браком. Если брак отсутствует, то деталь признается годной лишь в 80% случаев. Какое количество информации о качестве детали можно получить в среднем по результату такого метода отбраковки?

16. Орудие стреляет по удаленной цели. При каждом выстреле она поражается с вероятностью $p = 0.1$. Разведка может только один раз проверить, поражается ли цель хоть один раз. Через некоторое количество выстрелов k следует провести проверку, чтобы она дала максимальное количество информации?

17. Алфавит состоит из двух букв А, Б, В появляющихся в тексте с вероятностями 0.6, 0.3 и 0.1 соответственно. Закодировать отдельные буквы равномерным кодом. Закодировать пары и тройки букв равномерным кодом. Повторить кодирование одной буквы с не равномерным кодом. Сравнить эффективность кодов. Построить кодовое дерево для не равномерного кода

18. По линии связи передаются сообщения из 5-ти равновероятных букв. Закодировать буквы равномерным кодом. Закодировать тройки букв равномерным кодом. Сравнить эффективности кодов.

19. Имеются два дискретных троичных источника с независимыми элементами. На выходе каждого источника появляются сообщения

одинаковой длины по 25 элементов. Количество различных элементов в сообщении каждого источника постоянно. Сообщения каждого источника отличаются только порядком элементов. Зафиксированы два типичных сообщения: 0212021202120211201120200 – первого источника и 0121012011012012210200120 – второго. Элемент какого источника несет в среднем большее количество информации?

20. Сообщения с вероятностями 0,5; 0,25; 0,0625; 0,0625; 0,0625; 0,0625 кодируются одним из шести различных кодов: 1) 0-10-110-1110-1011-1101; 2) 1-011-010-001-000-110; 3) 0-10-110-1110-11110-111110; 4) 111-110-101-100-011-010; 5) 0-01-011-0111-01111-01111; 6) 1-01-0011-0010-0001-0000. Определить, какие коды являются разделимыми (мгновенными). Вычислить характеристики кодов.

Критерии оценки результатов текущего контроля

5 баллов - если ответ показывает глубокое и систематическое знание всего программного материала и структуры конкретного вопроса, а также основного содержания и новаций лекционного курса по сравнению с учебной литературой. Студент демонстрирует отчетливое и свободное владение концептуально-понятийным аппаратом, научным языком и терминологией соответствующей научной области. Знание основной литературы и знакомство с дополнительно рекомендованной литературой. Логически корректное и убедительное изложение ответа.

4 - балла - знание узловых проблем программы и основного содержания лекционного курса; умение пользоваться концептуально-понятийным аппаратом в процессе анализа основных проблем в рамках данной темы; знание важнейших работ из списка рекомендованной литературы. В целом логически корректное, но не всегда точное и аргументированное изложение ответа.

3 - балла - фрагментарные, поверхностные знания важнейших разделов программы и содержания лекционного курса; затруднения с использованием

научно-понятийного аппарата и терминологии учебной дисциплины; неполное знакомство с рекомендованной литературой; частичные затруднения с выполнением предусмотренных программой заданий; стремление логически определенно и последовательно изложить ответ.

2 балла - незнание, либо отрывочное представление о данной проблеме в рамках учебно-программного материала; неумение использовать понятийный аппарат; отсутствие логической связи в ответе.