



Государственное бюджетное образовательное учреждение высшего образования
Московской области

ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ
имени дважды Героя Советского Союза, летчика-космонавта А.А. Леонова



«УТВЕРЖДАЮ»
Проректор по учебной работе
Е.К. Самаров
2021 г.

**ИНСТИТУТ
ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ**

**КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И
УПРАВЛЯЮЩИХ СИСТЕМ**

**РАБОЧАЯ ПРОГРАММА
ДИСЦИПЛИНЫ
«ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ»**

Направление подготовки: 09.03.02 Информационные системы и технологии

Профиль: информационные технологии в технических системах

Уровень высшего образования: бакалавриат

Форма обучения: очная, заочная

Королев 2021

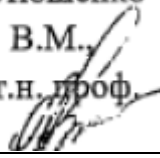
Автор: к.ф.-м.н., Мосалов О.П. Рабочая программа дисциплины: Основы машинного обучения: МГОТУ, 2021 – 12 с.

Рецензент:

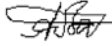
Рабочая программа составлена в соответствии с требованиями федерального государственного образовательного стандарта высшего профессионального образования (ФГОС ВО) по направлению подготовки бакалавров 09.03.02 «Информационные системы и технологии» и Учебного плана, утвержденного Ученым советом МГОТУ.

Протокол № 13 от 22.06.2021 г. Протокол № 13 от 22.06.2021 г.

Рабочая программа рассмотрена и одобрена на заседании кафедры:

Заведующий кафедрой (ФИО, ученая степень, звание, подпись)	Артошенко В.М. д.т.н. проф. 			
Год утверждения (переутверждения)	2021	2022	2023	2024
Номер и дата протокола заседания кафедры	№15 от 02.06.2021			

Рабочая программа согласована:

Руководитель ОПОП  к.т.н., доц. Т.С. Аббасова

Рабочая программа рекомендована на заседании УМС:

Год утверждения (переутверждения)	2021	2022	2023	2024
Номер и дата протокола заседания УМС	№7 от 15.06.2021 г			

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОПОП

Целью изучения дисциплины является изучение основных алгоритмов и методов машинного обучения.

В процессе обучения обучающийся приобретает и совершенствует следующие компетенции:

профессиональные компетенции (ПК)

- Способность выполнять интеграцию программных модулей и компонент (ПК-2);
- Способность оценивать качество программного обеспечения, в том числе проведение тестирования и исследование результатов (ПК-3);
- Способность проводить анализ требований к программному обеспечению, выполнять работы по проектированию программного обеспечения (ПК-11).

Основными задачами дисциплины являются:

1. Изучение различных типов задач, входящих в предметную область «машинное обучение»;
2. Изучение различных алгоритмов для решения задач машинного обучения;
3. Освоение методов сбора и обработки данных для их использования в задачах машинного обучения;
4. Освоение методов для сравнения эффективности различных алгоритмов и для подбора их параметров при решении задач машинного обучения;
5. Описание современного состояния исследований и последних достижений в области машинного обучения.

Показатель освоения компетенции отражают следующие индикаторы:

Трудовые действия:

- На основании специфических особенностей задачи выбирать наиболее эффективные алгоритмы; сравнивать эффективность работы отдельных алгоритмов.
- Интерпретировать результат работы моделей машинного обучения.
- Использовать стандартные возможности и библиотеки языка программирования Python, применяемые для реализации алгоритмов машинного обучения.

Необходимые умения:

- Сбор и обработка данных для их использования в задачах машинного обучения.
- Оценка качества данных; исправление найденные в данных дефектов.

Необходимые знания:

- Алгоритмы для построения моделей машинного обучения, включая их сильные и слабые стороны, а также специфические особенности.
- Метрики производительности для основных типов задач машинного обучения.

2. Место дисциплины в структуре ОПОП ВО

Дисциплина **«Основы машинного обучения»** относится к дисциплинам по выбору части Б1.В.ДВ, формируемой участниками образовательных отношений, основной профессиональной образовательной программы подготовки бакалавров по направлению подготовки 09.03.02 «Информационные системы и технологии».

Изучение данной дисциплины базируется на ранее изученных дисциплинах «Основы информатики», «Основы алгоритмизации и программирования» модуля «Инструментальные средства информационных систем», «Линейная алгебра и аналитическая геометрия», «Математический анализ» и «Теория вероятностей и математическая статистика» модуля «Математика» и компетенциях ОПК-1, ОПК-2, ОПК-6.

Знания и компетенции, полученные при освоении дисциплины, являются базовыми при изучении дисциплин: «Интеллектуальные системы и технологии», «Искусственный интеллект» и выполнении выпускной квалификационной работы бакалавра.

3. Объем дисциплины и виды учебной работы

Общая трудоемкость дисциплины для студентов очной формы составляет 8 зачетных единиц, 288 часа.

Таблица 1

Виды занятий	Всего часов	Семестр 7	Семестр 8	Семестр	Семестр
Общая трудоемкость	288	144	144		
ОЧНАЯ ФОРМА ОБУЧЕНИЯ					
Аудиторные занятия	100	50	50		
Лекции (Л)	44	22	22		
Практические занятия (ПЗ)	56	28	28		
Лабораторные работы (ЛР)	–	–	–		
Практическая подготовка	20	10	10		
Самостоятельная работа	188	94	94		
Контроль самостоятельной работы	+	+	+		
Курсовые работы (проекты)	+	–	+		
Расчетно-графические работы	–	–	–		
Контрольная работа	+	+	–		
Текущий контроль знаний (7-8, 15-16 недели)	Тест	+	+		
Вид итогового контроля	Зачет/ Экзамен	Зачет	Экзамен		
ЗАОЧНАЯ ФОРМА ОБУЧЕНИЯ					
Аудиторные занятия	40	40			
Лекции (Л)	16	16			
Практические занятия (ПЗ)	24	24			
Лабораторные работы (ЛР)	–	–			
Контроль самостоятельной работы	+	+			
Практическая подготовка					
Самостоятельная работа	284	284			
Курсовые работы (проекты)	–	–			
Расчетно-графические работы	–	–			
Контрольная работа	+	+			
Текущий контроль знаний (7 - 8, 15 - 16 недели)	–	–			
Вид итогового контроля	Зачет/ Экзамен	Экзамен			

4. Содержание дисциплины

4.1. Темы дисциплины и виды занятий

Таблица 2

Наименование разделов и тем	Лекции, час. очн/заочн, час	Практические занятия, очн/заочн, час	Занятия в интерактивной форме очн/заочн час	Практическая подготовка, очн/заочн, час.	Код компетенций
Первый семестр					
Тема 1. Машинное обучение. Общая постановка задачи.	2/2	3/3	1/2,5	2/0	ПК-2 ПК-3 ПК-11
Тема 2. Способы машинного обучения. Типы решаемых задач. Метрики производительности.	2/2	3/3	1/2,5	2/0	
Тема 3. Подготовка данных для машинного обучения.	3/2	3/3	2/2,5	1/0	
Тема 4. Линейная регрессия.	3/2	3/3	2/2,5	1/0	
Тема 5. Логистическая регрессия.	3/2	4/3	2/2,5	1/0	
Тема 6. Наивный байесовский классификатор.	3/2	4/3	2/2,5	1/0	
Тема 7. Метод k -ближайших соседей.	3/2	4/3	2/2,5	1/0	
Тема 8. Метод k -средних.	3/2	4/3	2/2,5	1/0	
Итого по первому семестру:	22/16	28/24	14/20	10/0	
Второй семестр					
Тема 9. Метод опорных векторов.	3/0	4/0	2/0	1/0	ПК-2 ПК-3 ПК-11
Тема 10. Деревья принятия решений.	3/0	4/0	2/0	1/0	
Тема 11. Ансамбли.	3/0	4/0	2/0	1/0	
Тема 12. Искусственные нейронные сети и глубокое обучение.	3/0	4/0	2/0	1/0	
Тема 13. Обработка естественного языка.	3/0	3/0	2/0	1/0	
Тема 14. Рекуррентные нейронные сети. Механизм внимания. Трансформеры.	3/0	3/0	2/0	1/0	
Тема 15. Компьютерное	2/0	3/0	1/0	2/0	

зрение.					
Тема 16. Сильный искусственный интеллект. Обучение с подкреплением.	2/0	3/0	1/0	2/0	
Итого по второму семестру:	22/0	28/0	14/0	10/0	
Итого:	44/16	56/24	28/20	20/0	

4.2. Содержание тем дисциплины

Тема 1. Машинное обучение. Общая постановка задачи.

Что такое машинное обучение. Как работает машинное обучение: данные, модель, алгоритм. Параметры и гиперпараметры модели. Переобучение модели. Недообучение модели. Кросс-валидация модели. Интерпретация моделей машинного обучения.

Тема 2. Способы машинного обучения. Типы решаемых задач. Метрики производительности.

Способы машинного обучения: обучение с учителем, обучение без учителя, обучение с подкреплением. Типы задач машинного обучения: регрессия, классификация, кластеризация. Метрики производительности: среднеквадратичная ошибка, средняя абсолютная ошибка, матрица запутанности, коэффициент силуэта.

Тема 3. Подготовка данных для машинного обучения.

Сбор данных. Структурирование и оцифровка данных. Очистка и исправление данных. Очистка данных. Уменьшение размерности. Нормализация. Обогащение данных. Формирование обучающей, валидационной и тестовой выборок.

Тема 4. Линейная регрессия.

Определение. Метод наименьших квадратов. Аналитическое решение. Градиентный спуск. Допущения линейной регрессии. Множественная линейная регрессия. Ограничения линейной регрессии. Интерпретация линейной регрессии. Преимущества и недостатки.

Тема 5. Логистическая регрессия.

Метод максимального правдоподобия. Определение. Логистическая функция потерь. Многоклассовая логистическая регрессия. Смещение и разброс. Регуляризация. L1-регуляризация (lasso). L2-регуляризация (ridge). Elastic Net регуляризация. Преимущества и недостатки.

Тема 6. Наивный байесовский классификатор.

Условная вероятность. Теорема Байеса. Определение. Функции правдоподобия. Сглаживание Лапласа. Преимущества и недостатки.

Тема 7. Метод k -ближайших соседей.

Определение. Метод k -ближайших соседей в задаче классификации. Гипотеза компактности. Метод k -ближайших соседей в задаче регрессии. Периферийные элементы и выбросы. Несбалансированные данные. Перебалансировка. Преимущества и недостатки.

Тема 8. Метод k -средних.

Определение. Расстояние для числовых и категориальных признаков. Алгоритм построения кластеров. Преимущества и недостатки.

Тема 9. Метод опорных векторов.

Определение. Бинарная классификация. Опорный вектор. Отступ. Линейная неразделимость. Ядро преобразования. Многоклассовая классификация. Парный подход. Один против всех. Применение к задаче регрессии. Преимущества и недостатки.

Тема 10. Деревья принятия решений.

Определение дерева. Правила. Описание метода. Энтропия. Коэффициент Джини. Отсечение ветвей. Основные алгоритмы построения деревьев принятия решений. Преимущества и недостатки.

Тема 11. Ансамбли.

Определение. Типы ансамблей. Бэггинг. Бутстрэп. Агрегация результатов. Случайный лес. Бустинг. Адабуст. Градиентный бустинг. Преимущества и недостатки.

Тема 12. Искусственные нейронные сети и глубокое обучение.

Формальный нейрон. Простой перцептрон Розенблатта. Метод коррекции ошибки. Глубокое обучение. Сеть прямого распространения. Метод обратного распространения ошибки. Функции активации. Функция ошибки. Эпохи и батчи. Алгоритм оптимизации. Гиперпараметры искусственной нейронной сети. Дропаут. Начальные значения весов. Преимущества и недостатки.

Тема 13. Обработка естественного языка.

Распознавание речи. Синтез речи. Анализ текста. Генерация текста. Предварительная обработка текста. Векторные представления слов. Bag of words. Hashing trick. N -граммы. Шумовые слова. TF-IDF. Word2vec. Continuous bag of words. Skip-gram. Преимущества и недостатки.

Тема 14. Рекуррентные нейронные сети. Механизм внимания. Трансформеры.

Подход seq2seq. Рекуррентные нейронные сети. Сети с долгой краткосрочной памяти. Механизм внимания. Трансформеры. Преимущества и недостатки.

Тема 15. Компьютерное зрение.

Компьютерное зрение. Машинное зрение. Распознавание образов. Поиск объектов на изображении. Классификация изображений. Детектирование объектов. Сегментация изображений. Обработка найденных объектов. Отслеживание движущихся объектов (трекинг). Распознавание лиц (идентификация). Распознавание символов или текста. Восстановление

(реконструкция) изображений. Удаление шума с изображения. Восстановление или добавление цвета. Поворот объекта. Построение 3D-формы. Генерация изображений. Свёрточная нейронная сеть. Генеративно-сопоставительная сеть.

Тема 16. Сильный искусственный интеллект. Обучение с подкреплением.

Искусственный интеллект – определение и типизация. Сильный искусственный интеллект. Тест Тьюринга. Обучение с подкреплением. Динамическое программирование. Метод Монте-Карло. Метод временной разности. SARSA. Q-обучение. DQN.

5. Перечень учебно-методического обеспечения для самостоятельной работы по дисциплине

Методические указания для обучающихся по освоению дисциплины.

6. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Структура фонда оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведена в Приложении 1.

7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

1. Вьюгин, В. В. Математические основы машинного обучения и прогнозирования: Учебное пособие / Вьюгин В.В. - Москва :МЦНМО, 2014. - 304 с.: ISBN 978-5-4439-2014-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/958689> (дата обращения: 16.04.2021). – Режим доступа: по подписке.
2. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / С. Рашка ; пер. с англ. А.В. Логунова. - Москва : ДМК Пресс, 2017. - 418 с. - ISBN 978-5-97060-409-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027758> (дата обращения: 16.04.2021). – Режим доступа: по подписке.

Дополнительная литература:

1. Коэльо, Луис Педро Построение систем машинного обучения на языке Python / Луис Педро Коэльо, Вилли Ричарт ; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва : ДМК Пресс, 2016. - 302 с. - ISBN 978-5-97060-330-7. -

Текст : электронный. - URL: <https://znanium.com/catalog/product/1027824> (дата обращения: 16.04.2021). – Режим доступа: по подписке.

2. Саттон, Р. С. Обучение с подкреплением: введение : практическое руководство / Р. С. Саттон, Э. Барто ; пер. с англ. А. А. Слинкина. - Москва : ДМК Пресс, 2020. - 552 с. - ISBN 978-5-97060-097-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1210617> (дата обращения: 16.04.2021). – Режим доступа: по подписке.

Рекомендуемая литература:

1. Танцов, П. Н. Интеллектуальные информационные системы : лабораторный практикум / П. Н. Танцов. - Москва : Изд. Дом МИСиС, 2015. - 86 с. - ISBN 978-5-87623-898-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1232708> (дата обращения: 16.04.2021). – Режим доступа: по подписке.

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. www.machinelearning.ru – профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.
2. <https://neurohive.io/ru/osnovy-data-science/> – базовый курс «Основы Data Science».
3. <https://dyakonov.org/> – блог д.ф.-м.н. Александра Дьяконова.

9. Методические указания для обучающихся по освоению дисциплины

Методические указания для обучающихся по освоению дисциплины (модуля) приведены в Приложении 2.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Перечень программного обеспечения: MS Office, Jupyter Notebook, Python, GitHub.

Информационные справочные системы: не предусмотрено курсом данной дисциплины

Ресурсы информационно-образовательной среды МГОТУ:

Рабочая программа и методическое обеспечение по дисциплине «Основы машинного обучения».

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

Лекционные занятия:

- аудитория, оснащенная презентационной техникой (проектор, экран);
- комплект электронных презентаций / слайдов.

Практические занятия:

- учебный класс, оснащенный вычислительной техникой (ПК);
- рабочее место преподавателя, оснащенное компьютером с доступом в Интернет;
- рабочие места студентов, оснащенные компьютерами с доступом в Интернет.

Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

**ИНСТИТУТ
ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ**

**КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
И УПРАВЛЯЮЩИХ СИСТЕМ**

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ
ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО
ДИСЦИПЛИНЕ**

«ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ»

(Приложение 1 к рабочей программе)

Направление подготовки: 09.03.02 Информационные системы и технологии

Профиль: информационные технологии в технических системах

Уровень высшего образования: бакалавриат

Форма обучения: очная, заочная

Королев 2021

1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

№ п/п	Индекс компетенции	Содержание компетенции (или ее части)*	Раздел дисциплины, обеспечивающий формирование компетенции (или ее части)	В результате изучения раздела дисциплины, обеспечивающего формирование компетенции (или ее части), обучающийся должен:		
				трудовые действия	необходимые умения	необходимые знания
1	ПК-2	Способность выполнять интеграцию программных модулей и компонент	Тема 4. Тема 5. Тема 6. Тема 7. Тема 8. Тема 9. Тема 10. Тема 11. Тема 12.	На основании специфических особенностей задачи выбирать наиболее эффективные алгоритмы. Сравнить эффективность работы отдельных алгоритмов. Интерпретировать результат работы моделей машинного обучения. Использовать стандартные возможности и библиотеки языка программирования Python, применяемые для реализации алгоритмов машинного обучения.	Оценка качеств данных; исправление найденные в данных дефектов.	Алгоритмы для построения моделей машинного обучения, включая их сильные и слабые стороны, а также специфические особенности. Метрики производительности для основных типов задач машинного обучения.
2	ПК-3	Способность оценивать качество программного обеспечения, в том числе проведение тестирования и исследование результатов	Тема 1. Тема 2. Тема 3.	Использовать стандартные возможности и библиотеки языка программирования Python, применяемые для реализации алгоритмов машинного обучения.	Сбор и обработка данных для их использования в задачах машинного обучения. Оценка качеств данных; исправление найденные в данных дефектов.	Метрики производительности для основных типов задач машинного обучения.

3	ПК-11	Способность проводить анализ требований к программному обеспечению, выполнять работы по проектированию программного обеспечения	Тема 13. Тема 14. Тема 15. Тема 16.	Использовать стандартные возможности и библиотеки языка программирования Python, применяемые для реализации алгоритмов машинного обучения. На основании специфических особенностей задачи выбирать наиболее эффективные алгоритмы; сравнивать эффективность работы отдельных алгоритмов. Интерпретировать результат работы моделей машинного обучения.	Сбор и обработка данных для их использования в задачах машинного обучения. Оценка качества данных; исправление найденных в данных дефектов.	Алгоритмы для построения моделей машинного обучения, включая их сильные и слабые стороны, а также специфические особенности.
---	-------	---	--	--	---	--

2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Код компетенции	Инструменты, оценивающие сформированность компетенции	Показатель оценивания компетенции	Критерии оценки
ПК-2	Контрольная работа	<p>А) полностью сформирована (компетенция освоена на высоком уровне) – 5 баллов</p> <p>Б) частично сформирована:</p> <ul style="list-style-type: none"> •компетенция освоена на продвинутом уровне – 4 балла; •компетенция освоена на базовом уровне – 3 балла; <p>В) не сформирована (компетенция не освоена) – 2 и менее баллов</p>	<p>Проводится в форме письменной работы.</p> <p>Время, отведенное на процедуру – 90 мин.</p> <p>Неявка – 0.</p> <p>Критерии оценки:</p> <ol style="list-style-type: none"> 1.Соответствие содержания контрольной работы заявленной тематике (1 балл). 2.Качество источников и их количество при подготовке работы (1 балл). 3.Владение информацией и способность отвечать на вопросы аудитории (1 балл). 4.Качество самой представленной работы (1 балл). 5. Использование специализированного программного обеспечения (1 балл). <p>Максимальная сумма баллов – 5 баллов.</p> <p>Результаты оценочной процедуры представляются обучающимся в срок не позднее 1 недели после проведения процедуры – для текущего контроля. Оценка проставляется в электронный журнал.</p>
ПК-3	Доклад в форме презентации	<p>А) полностью сформирована (компетенция освоена на высоком уровне) – 5 баллов</p> <p>Б) частично сформирована:</p> <ul style="list-style-type: none"> •компетенция освоена на продвинутом уровне – 4 балла; •компетенция освоена на базовом уровне – 3 балла; <p>В) не сформирована (компетенция не освоена) – 2 и менее баллов</p>	<p>Проводится в устной форме.</p> <p>Время, отведенное на процедуру – 15-20 минут.</p> <p>Неявка на защиту доклада – 0.</p> <p>Критерии оценки:</p> <ol style="list-style-type: none"> 1.Соответствие представленной презентации заявленной тематике (1 балл). 2.Качество источников и их количество при подготовке доклада и разработке презентации (1 балл). 3.Владение информацией и способность отвечать на вопросы аудитории (1 балл). 4.Качество самой представленной презентации (1 балл). 5.Оригинальность подхода и всестороннее раскрытие выбранной тематике (1 балл). <p>Максимальная сумма баллов – 5</p>

			баллов. Результаты оценочной процедуры представляются обучающимся в срок не позднее 1 недели после проведения процедуры – для текущего контроля. Оценка проставляется в электронный журнал.
ПК-11	Курсовая работа	<p>А) полностью сформирована (компетенция освоена на высоком уровне) – 5 баллов</p> <p>Б) частично сформирована: •компетенция освоена на продвинутом уровне – 4 балла; •компетенция освоена на базовом уровне – 3 балла;</p> <p>В) не сформирована (компетенция не освоена) – 2 и менее баллов</p>	<p>Проводится в форме письменной работы</p> <p>Время, отведенное на процедуру – 1 семестр.</p> <p>Неявка на защиту работы – 0.</p> <p>Критерии оценки:</p> <p>1.Соответствие содержания курсовой работы заявленной тематике (1 балл).</p> <p>2.Качество источников и их количество при подготовке работы (1 балл).</p> <p>3.Владение информацией и способность отвечать на вопросы аудитории (1 балл).</p> <p>4.Качество самой представленной работы (1 балл).</p> <p>5. Использование специализированного программного обеспечения (1 балл).</p> <p>Максимальная сумма баллов – 5 баллов.</p> <p>Результаты оценочной процедуры представляются обучающимся в срок не позднее 1 недели после проведения процедуры – для текущего контроля. Оценка проставляется в электронный журнал.</p>

3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Тематика докладов в презентационной форме

1. Задачи ИИ, не входящие в машинное обучение.
2. Кроссвалидация – пример использования.
3. Задача ранжирования.
4. Задача уменьшения размерности.
5. Задача выявления аномалий.
6. Обучение с частичным привлечением учителя (Semi-supervised learning).
7. Перенос знаний (Transfer learning).
8. Метрика классификации AUC (area under curve).

9. Различные метрики кластеризации.
10. Обогащение данных в машинном обучении.
11. Коэффициент детерминации (R-квадрат) как метрика линейной регрессии.
12. Полиномиальная регрессия.
13. Различные алгоритмы кластеризации.
14. Структура и возможности библиотеки sklearn.
15. Структура и возможности библиотеки keras.
16. Структура и возможности библиотеки xgboost.
17. Алгоритмы отсечения ветвей в деревьях принятия решений.
18. Ансамбли – стекинг.
19. Алгоритмы оптимизации нейронных сетей.
20. Варианты реализации дропаута.

Указания по проведению контрольных работ

Учебным планом данного курса предусмотрено написание одной контрольной работы, что является одним из условий успешного освоения основных положений данной дисциплины и служит допуском к сдаче зачёта по курсу во время зачетной сессии.

Цель выполняемой работы: Продемонстрировать знания и умения в изучении дисциплины.

Порядок выполнения контрольной работы:

Контрольная работа излагается логически последовательно, грамотно, разборчиво. Работа должна иметь титульный лист. Он содержит полное название высшего учебного заведения, кафедра, реализующая данную дисциплину, название (тема) контрольной работы, фамилию, инициалы автора, также необходимо указать номер группы, фамилию и инициалы, а также должность, ученое звание и степень научного руководителя (преподавателя), проверяющего контрольную работу.

На следующем листе излагается план контрольной работы, который включает в себя: название всех разделов, введение и заключение, а также список литературы. Излагая вопрос (раздел) каждый смысловой абзац необходимо начать с красной строки.

В конце работы ставится подпись студента и дата сдачи. Страницы контрольной работы должны быть пронумерованы. Номер страницы ставится в нижнем правом углу.

Размер шрифта №14 (Times New Roman), полуторный интервал, стандартный лист формата А4. Поля: верхнее – 20 мм, нижнее – 20 мм, левое – 30 мм, правое – 15 мм.

Дополнительно контрольная работа может иметь приложения (формулы, схемы, графики, диаграммы).

По всем возникающим вопросам обучающемуся следует обращаться за консультацией на кафедру. Срок выполнения контрольной работы определяется кафедрой. Срок проверки контрольной работы – 3 дня с момента необходимой фиксированной даты сдачи.

Задания контрольной работы

1. Некая модель регрессии вернула для четырёх элементов тестовой выборки следующие значения: $z_{11}, z_{12}, z_{13}, z_{14}$.
При этом целевые значения (ожидаемый результат) для тех же четырёх элементов следующие: 0, 1, 2, 3.
Вычислите среднеквадратичную ошибку этой модели.

2. Некая модель бинарной классификации вернула для девяти элементов тестовой выборки следующие значения:
 $z_{21}, z_{22}, z_{23}, z_{24}, z_{25}, z_{26}, z_{27}, z_{28}, z_{29}$.
При этом целевые значения (ожидаемый результат) для тех же девяти элементов тестовой выборки следующие:
0, 0, 1, 1, 0, 0, 1, 1, 1.

Вычислите следующие метрики классификации для этой модели:

- accuracy,
- precision,
- recall,
- F1.

3. Некая модель кластеризации разделила четыре элемента набора данных на два кластера следующим образом:

(z_{31}, z_{32})	0
(z_{33}, z_{34})	0
(z_{35}, z_{36})	1
(z_{37}, z_{38})	1

Вычислите коэффициент силуэта для каждого из элементов и их среднее значение.

4. Решите аналитически задачу линейной регрессии (то есть вычислите значения a и b) для следующего набора данных (x_i, y_i) :
 $(0, z_{41}), (1, z_{42}), (2, z_{43})$.

Контрольная работа выполняется по вариантам, каждый из которых содержит свои данные для решения вышеперечисленных заданий.

Примеры вариантов данных для контрольной работы:

Вариант №1

$$z_1 = [3, 10, 2, 5]$$

$$z_2 = [0, 1, 1, 1, 1, 0, 0, 1, 0]$$

$$z_3 = [(3, 3), (4, 0), (3, 2), (1, 4)]$$

$$z_4 = [0, 2, 0]$$

Вариант №2

$z1 = [1, 1, 9, 1]$
 $z2 = [1, 0, 1, 0, 0, 1, 1, 0, 1]$
 $z3 = [(1, 1), (3, 2), (0, 3), (4, 0)]$
 $z4 = [1, 2, 0]$

Вариант №3

$z1 = [6, 9, 7, 9]$
 $z2 = [0, 1, 1, 1, 1, 0, 1, 0, 1]$
 $z3 = [(3, 1), (2, 4), (2, 0), (3, 4)]$
 $z4 = [0, 1, 4]$

Вариант №4

$z1 = [7, 6, 8, 1]$
 $z2 = [1, 0, 1, 1, 0, 0, 0, 0, 0]$
 $z3 = [(4, 4), (1, 3), (4, 2), (3, 2)]$
 $z4 = [4, 4, 0]$

Вариант №5

$z1 = [7, 9, 3, 9]$
 $z2 = [0, 1, 0, 1, 1, 0, 1, 1, 1]$
 $z3 = [(3, 2), (0, 4), (4, 4), (4, 2)]$
 $z4 = [3, 4, 0]$

Вариант №6

$z1 = [4, 3, 9, 10]$
 $z2 = [0, 0, 1, 0, 0, 0, 0, 1, 0]$
 $z3 = [(2, 1), (2, 0), (4, 1), (2, 2)]$
 $z4 = [0, 1, 1]$

Вариант №7

$z1 = [5, 9, 3, 5]$
 $z2 = [1, 1, 1, 1, 1, 0, 0, 1, 1]$
 $z3 = [(2, 3), (1, 2), (0, 2), (4, 1)]$
 $z4 = [4, 3, 0]$

Вариант №8

$$z1 = [4, 1, 7, 3]$$

$$z2 = [0, 0, 1, 1, 0, 1, 0, 0, 1]$$

$$z3 = [(4, 2), (3, 0), (2, 1), (1, 0)]$$

$$z4 = [2, 0, 0]$$

Вариант №9

$$z1 = [5, 5, 3, 7]$$

$$z2 = [1, 0, 0, 0, 0, 1, 0, 0, 1]$$

$$z3 = [(1, 2), (0, 1), (4, 3), (4, 1)]$$

$$z4 = [3, 0, 3]$$

Вариант №10

$$z1 = [5, 9, 8, 1]$$

$$z2 = [1, 1, 1, 0, 0, 0, 1, 0, 1]$$

$$z3 = [(3, 1), (2, 0), (3, 4), (2, 4)]$$

$$z4 = [3, 4, 1]$$

Указания по проведению курсовых работ

Задание на курсовую работу для сравнения двух изученных алгоритмов машинного обучения выдается на первом практическом занятии 6-го семестра. Защита курсовой работы проводится при очной встрече в конце семестра в соответствии с расписанием.

В обязательном порядке курсовая работа должна включать следующие разделы:

1. Введение;
2. Аналитическая часть:
 - описание сравниваемых алгоритмов;
 - анализ сильных и слабых сторон алгоритмов;
3. Практическая часть:
 - описание набора данных;
 - описание вычислительных экспериментов;
 - описание полученных результатов;
4. Заключение;
5. Список использованной литературы;
6. Приложения.

Во введении необходимо сформулировать цель и задачи работы, охарактеризовать объект и методы исследования. Объем введения составляет 1—2 страницы.

Аналитическая часть пишется на основании глубокого изучения литературы, посвященной рассматриваемым алгоритмам машинного обучения. Глава включает краткое изложение принципов их работы, сильных и слабых сторон.

При написании этой главы необходимо использовать справочники, учебники и учебные пособия, другие литературные источники, посвященные вопросам машинного обучения. Объем аналитической части 15-20 стр.

В практической части курсовой работы студент осуществляет анализ и подготовку выбранного набора данных с использованием изученных методов. Набор данных может быть выбран самостоятельно или предоставлен по запросу студента.

В данной главе отражаются результаты вычислительных экспериментов, описываются полученные результаты и делаются выводы. Объем главы 15-20 стр.

В заключении должны быть сформулированы краткие выводы по всей работе. Объем заключения — 1 страница.

Оформление курсовой работы

Курсовая работа должна быть представлена студентом в виде отчета в установленный срок. Отчет оформляется на бумаге формата А4 в печатном виде с установкой следующих полей: правое – 10 мм, левое – 30 мм, верхнее – 20 мм, нижнее – 25 мм. Номера страниц проставляются арабскими цифрами в правом верхнем углу.

Нумерация разделов производится арабскими цифрами с точкой после номера. Введение и заключение не нумеруются. Точку в конце заголовка раздела не ставят. Каждый раздел начинается с новой страницы. Расстояние между заголовком и текстом – 15 мм.

Рисунки располагаются после ссылки на них в тексте. Под рисунком указывается его номер и название. **Таблицы** должны иметь номер и название. Все рисунки и таблицы должны иметь сквозную нумерацию.

Формулы помещаются на отдельной строке и нумеруются цифрами в скобках с правой стороны от формулы (нумерация сквозная). Каждая формула должна содержать расшифровку условных обозначений с указанием размерности. При ссылке в тексте на формулу указывается ее номер.

Список литературы оформляется в следующем порядке: первыми указываются законы и государственные документы, затем в алфавитном порядке перечисляются учебники, справочники, словари и т. п. Все литературные источники нумеруются, в тексте дается ссылка на номер. Последовательность записи такова: фамилия и инициалы авторов; полное название книги или статьи; место выпуска и

наименование издательства; год выпуска; номер журнала; количество страниц.

Приложения помещаются после списка литературы на последующих

страницах. Каждое приложение начинается с нового листа с указанием в правом верхнем углу слова «Приложение» с порядковым номером. Нумеруются приложения буквами русского алфавита. Приложения имеет заголовок.

Защита курсовой работы

Курсовая работа считается выполненной после исправления обнаруженных при проверке преподавателем недостатков и ошибок.

Основными критериями при выставлении оценки являются:

самостоятельность выполнения проекта;
теоретическая и практическая подготовка студента в области основных алгоритмов и методов машинного обучения;
навыки исследовательского характера;
грамотность изложения и свободное владение материалом;
качество оформления курсовой работы;
правильность ответов на вопросы при защите курсовой работы.

Порядок защиты курсовой работы:

1. Доклад студента о результатах своих исследований (до 5 мин.);
2. Ответы на вопросы.

Методические рекомендации по выполнению теоретической главы курсовой работы

Первая глава курсовой работы предполагает реферативное изложение материала, изученного по учебникам, учебным пособиям, справочникам и другим литературным источникам по теме курсовой работы в соответствии с вопросами, представленными в плане. Перечень вопросов может быть изменен (расширен, сужен) по желанию студента при согласовании с руководителем.

Примерная тематика курсовых работ

1. Сравнение метода k ближайших соседей и метода опорных векторов при решении задачи классификации.
2. Сравнение линейной регрессии и дерева принятия решений при решении задачи регрессии.
3. Сравнение логистической регрессии и наивного байесовского классификатора при решении задачи классификации.
4. Сравнение метода опорных векторов и дерева принятия решений при решении задачи классификации.
5. Сравнение наивного байесовского классификатора и градиентного бустинга при решении задачи классификации.
6. Сравнение дерева принятия решений и градиентного бустинга при решении задачи регрессии.
7. Сравнение искусственной нейронной сети и метода опорных векторов при решении задачи классификации.
8. Сравнение метода k ближайших соседей и линейной регрессии при решении задачи регрессии.

9. Сравнение дерева принятия решений и метода k ближайших соседей при решении задачи классификации.
10. Сравнение дерева принятия решений и метода k ближайших соседей при решении задачи регрессии.
11. Сравнение метода опорных векторов и наивного байесовского классификатора при решении задачи классификации.
12. Сравнение линейной регрессии и случайного леса при решении задачи регрессии.
13. Сравнение метода k ближайших соседей и случайного леса при решении задачи классификации.
14. Сравнение искусственной нейронной сети и дерева принятия решений при решении задачи регрессии.
15. Сравнение градиентного бустинга и метода опорных векторов при решении задачи классификации.
16. Сравнение метода k ближайших соседей и искусственной нейронной сети при решении задачи регрессии.
17. Сравнение логистической регрессии и случайного леса при решении задачи классификации.
18. Сравнение дерева принятия решений и логистической регрессии при решении задачи классификации.
19. Сравнение искусственной нейронной сети и логистической регрессии при решении задачи классификации.
20. Сравнение искусственной нейронной сети и случайного леса при решении задачи регрессии.
21. Сравнение метода k ближайших соседей и градиентного бустинга при решении задачи классификации.
22. Сравнение метода k ближайших соседей и градиентного бустинга при решении задачи регрессии.
23. Сравнение наивного байесовского классификатора и искусственной нейронной сети при решении задачи классификации.
24. Сравнение логистической регрессии и метода k ближайших соседей при решении задачи классификации.
25. Сравнение метода опорных векторов и случайного леса при решении задачи классификации.
26. Сравнение градиентного бустинга и искусственной нейронной сети при решении задачи регрессии.
27. Сравнение метода k ближайших соседей и наивного байесовского классификатора при решении задачи классификации.
28. Сравнение искусственной нейронной сети и линейной регрессии при решении задачи регрессии.
29. Сравнение случайного леса и метода k ближайших соседей при решении задачи регрессии.
30. Сравнение логистической регрессии и метода опорных векторов при решении задачи классификации.

31. Сравнение искусственной нейронной сети и метода k ближайших соседей при решении задачи классификации.
32. Сравнение наивного байесовского классификатора и дерева принятия решений при решении задачи классификации.
33. Сравнение линейной регрессии и градиентного бустинга при решении задачи регрессии.
34. Сравнение градиентного бустинга и логистической регрессии при решении задачи классификации.

4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Формой контроля знаний по дисциплине «Основы машинного обучения» являются две текущие аттестации в виде тестов каждый семестр, промежуточные аттестации в виде зачета с оценкой в 5-ом семестре и экзамена в 6-ом семестре для очной формы обучения; промежуточные аттестации в виде зачета и экзамена в 4-ом семестре для заочной формы обучения.

Неделя текущего контроля	Вид оценочного средства	Код компетенций, оценивающий знания, умения, навыки	Содержание оценочного средства	Требования к выполнению	Срок сдачи (неделя семестра)	Критерии оценки по содержанию и качеству с указанием баллов
7-8	тестирование	ПК-2 ПК-3 ПК-11	27-28 вопросов	Компьютерное тестирование; время отведенное на процедуру – 30 минут.	Результаты тестирования предоставляются в день проведения процедуры.	Критерии оценки определяются процентным соотношением. Не явка – 0. Удовлетворительно – от 51% правильных ответов. Хорошо – от 70%. Отлично – от 90%. Максимальная оценка – 5 баллов.
15-16	тестирование	ПК-2 ПК-3 ПК-11	27-28 вопросов	Компьютерное тестирование; время отведенное на процедуру – 30 минут.	Результаты тестирования предоставляются в день проведения процедуры.	Критерии оценки определяются процентным соотношением. Не явка – 0. Удовлетворительно – от 51% правильных ответов. Хорошо – от 70%. Отлично – от 90%. Максимальная оценка – 5 баллов.

18	зачет с оценкой	ПК-2 ПК-3 ПК-11	2 вопроса	Зачет проводится в устной форме, путем ответа на вопросы. Время отведенное на процедуру – 0,25 часа на студента.	Результаты предоставляются в день проведения зачета.	<p>Критерии оценки:</p> <p>«Отлично»:</p> <ul style="list-style-type: none"> • знание основных понятий предмета; • умение использовать и применять полученные знания на практике; • работа на практических занятиях; • знание основных научных теорий, изучаемых предметов; • ответ на вопросы билета. <p>«Хорошо»:</p> <ul style="list-style-type: none"> • знание основных понятий предмета; • умение использовать и применять полученные знания на практике; • работа на практических занятиях; • знание основных научных теорий, изучаемых предметов; • ответы на вопросы билета • неправильно решено практическое задание <p>«Удовлетворительно»:</p> <ul style="list-style-type: none"> • демонстрирует частичные знания по темам дисциплин; • незнание неумение использовать и применять полученные знания на практике; • не работал на практических занятиях; <p>«Неудовлетворительно»:</p> <ul style="list-style-type: none"> • демонстрирует частичные знания по темам дисциплин; • незнание основных понятий предмета; • неумение использовать и применять полученные знания на практике; • не работал на практических занятиях;
----	-----------------	-----------------------	-----------	--	--	---

						<ul style="list-style-type: none"> • не отвечает на вопросы.
18	экзамен	ПК-2 ПК-3 ПК-11	2 вопроса	Экзамен проводится в устной форме, путем ответа на вопросы. Время, отведенное на процедуру – 30 минут.	Результаты предоставляются в день проведения экзамена.	<p>Критерии оценки:</p> <p>«Отлично»:</p> <ul style="list-style-type: none"> • знание основных понятий предмета; • умение использовать и применять полученные знания на практике; • работа на практических занятиях; • знание основных научных теорий, изучаемых предметов; • ответ на вопросы билета. <p>«Хорошо»:</p> <ul style="list-style-type: none"> • знание основных понятий предмета; • умение использовать и применять полученные знания на практике; • работа на практических занятиях; • знание основных научных теорий, изучаемых предметов; • ответы на вопросы билета • неправильно решено практическое задание <p>«Удовлетворительно»:</p> <ul style="list-style-type: none"> • демонстрирует частичные знания по темам дисциплин; • незнание неумение использовать и применять полученные знания на практике; • не работал на практических занятиях; <p>«Неудовлетворительно»:</p> <ul style="list-style-type: none"> • демонстрирует частичные знания по темам дисциплин; • незнание основных понятий предмета; • неумение использовать и

						применять полученные знания на практике; • не работал на практических занятиях; • не отвечает на вопросы.
--	--	--	--	--	--	---

4.1. Типовые вопросы, выносимые на тестирование

Тесты используются в режиме промежуточного контроля. По форме заданий выбраны закрытые тесты (с выборочным ответом). Каждому вопросу соответствует один вариант ответа.

Первое тестирование (5-й семестр)

1. Как соотносятся машинное обучение и искусственный интеллект?

(?) Это одно и то же.

(!) Машинное обучение – это часть искусственного интеллекта.

(?) Искусственный интеллект – это часть машинного обучения.

(?) Это не связанные между собой области.

2. В чём отличие машинного обучения от обычного программирования?

(?) Машинное обучение преобразует входные данные в логику и выходные данные.

(?) Машинное обучение преобразует входные данные и логику в выходные данные.

(!) Машинное обучение преобразует входные и выходные данные в логику.

(?) Машинное обучение преобразует логику в данные.

3. Что такое задача машинного обучения?

(!) Используя данные научиться решать задачу так, чтобы метрика производительности улучшалась.

(?) Используя данные и задачу, построить метрику производительности, которая будет улучшаться.

(?) Подобрать данные, при которых для данной задачи метрика производительности улучшается.

(?) Используя данные и задачу, выбрать метрику производительность, которая улучшается сильнее других.

4. Как связаны модель и алгоритм в машинном обучении?

(?) Модель используется для построения алгоритма.

(!) Алгоритм используется для построения модели.

(?) Модель и алгоритма независимы и выводятся из данных.

(?) Модель и алгоритм – это одно и то же.

5. Как ведут себя параметры и гиперпараметры модели при её обучении?

(?) Параметры и гиперпараметры могут изменяться.

(?) Параметры и гиперпараметры не могут изменяться.

(!) Параметры могут изменяться, гиперпараметры не могут изменяться.

(?) Параметры не могут изменяться, гиперпараметры могут изменяться.

6. Что из этого не является методом борьбы с переобучением модели?

(?) Увеличение количества данных.

(?) Упрощение модели.

(!) Увеличение тестовой выборки.

(?) Уменьшение времени обучения.

7. Что из этого не является методом борьбы с недообучением модели?

(?) Усложнение модели.

(?) Увеличение размерности данных.

(?) Уменьшение шума в данных.

(!) Уменьшение времени обучения.

8. В чём основная цель кросс-валидации?

(?) Найти самый эффективный размер обучающей выборки.

(!) Оценить эффективности модели, используя все имеющиеся данные.

(?) Настроить гиперпараметры модели.

(?) Убедиться, что ошибка работы модели минимальна.

9. Что из этого не является требованием к интерпретации модели машинного обучения?

(?) Соответствие экспертным ожиданиям.

(!) Высокая скорость.

(?) Возможность сравнения результатов работы модели для разных данных.

(?) Краткость выводов.

10. Что из перечисленного не является способом машинного обучения?

(?) Обучение без учителя.

(?) Обучение с учителем.

(!) Обучение без подкрепления.

(?) Обучение с подкреплением.

11. Какая ключевая особенность данных, используемых в обучении с учителем?

- (?) Описание объекта содержит значения характеристик объекта.
- (!) Для каждого описания объекта известен ожидаемый ответ модели.
- (?) Различные описания объектов не могут иметь одинаковый ожидаемый ответ.
- (?) Количество различных ожидаемых ответов должно быть значительно меньше количества описаний объектов.

12. В чём состоит задача, решаемая обучением без учителя?

- (!) Обнаружить неизвестные зависимости между данными.
- (?) Оценить качество данных.
- (?) Предсказать значение некоторой величины для каждого элемента данных.
- (?) Уменьшить размерность данных.

13. С помощью чего обучается агент при использовании обучения с подкреплением?

- (?) Заранее известные ожидаемые ответы.
- (!) Информация о реакции внешней среды на действия агента.
- (?) Информация о предыдущих действиях агента.
- (?) Экспертная оценка действий агента.

14. Что из перечисленного не является задачей машинного обучения?

- (?) Регрессия.
- (?) Кластеризация.
- (!) Прогрессия.
- (?) Классификация.

15. Что представляет собой задача регрессии?

- (?) Разбивка множества объектов на группы исходя из их схожести.
- (?) Определение принадлежности объекта к одной из заранее известных групп объектов.
- (?) Вычисление уникального идентификатора на основании описания объекта.
- (!) Вычисление числа или числового вектора на основании описания объекта.

16. Что из перечисленного неверно при решении задачи классификации?

- (?) Все классы должны быть известны заранее.
- (!) Количество классов может быть бесконечно.

- (?) Классификация – это задача обучения с учителем.
- (?) Каждый объект должен относиться хотя бы к одному классу.

17. Что из перечисленного не применяется в качестве метрики производительности задачи классификации?

- (?) Precision.
- (!) Reward.
- (?) Accuracy.
- (?) Recall.

18. Что из перечисленного верно при решении задачи кластеризации?

- (?) Элементы данных должны иметь заранее известные метки.
- (?) Все метрики основаны на экспертной информации.
- (?) Количество кластеров определяется в процессе решения задачи.
- (!) Необходимо задать функцию похожести элементов.

19. Из чего состоит набор данных?

- (?) Значение параметра.
- (!) Элемент данных.
- (?) Параметр.
- (?) Тип данных.

20. Что не является способом борьбы с несбалансированными данными?

- (?) Использование различных алгоритмов.
- (?) Сбор дополнительных данных.
- (!) Уменьшение размерности данных.
- (?) Использование специальной метрики производительности.

21. Если набор данных на 90% относится к классу А и на 10% – к классу В, то что представляет из себя undersampling?

- (!) Удаление элементов класс А из набора данных.
- (?) Удаление элементов класс В из набора данных.
- (?) Добавление элементов класс А в набор данных.
- (?) Добавление элементов класс В в набор данных.

22. Преобразование данных в формат, приспособленный для хранения и обработки с помощью компьютера – это:

- (?) Структурирование данных.
- (!) Оцифровка данных.
- (?) Очистка данных.
- (?) Обогащение данных.

23. С какой проблемой не борется очистка данных?

- (?) Шум в данных.
- (!) Несбалансированность данных.
- (?) Несоответствие значений.
- (?) Неполнота данных.

24. Какой процесс помогает улучшить визуализируемость данных?

- (?) Очистка данных.
- (?) Нормализация данных.
- (?) Синтетическая генерация новых данных.
- (!) Уменьшение размерности данных.

25. Какие из перечисленных данных являются категориальными?

- (?) Имена.
- (?) Даты рождения.
- (?) Города рождения.
- (!) Всё вышеперечисленное.

26. Унитарное кодирование (one-hot encoding) позволяет преобразовать

- (?) Числовой параметр в категориальный.
- (?) Категориальный параметр в числовой.
- (?) Числовой параметр в несколько категориальных.
- (!) Категориальный параметр в несколько числовых.

27. Обогащение данных приводит к

- (?) Увеличению количества элементов в наборе данных.
- (?) Уменьшению количества элементов в наборе данных.
- (!) Увеличению количества параметров элементов в наборе данных.
- (?) Уменьшению количества параметров элементов в наборе данных.

Второе тестирование (5-й семестр)

1. К какому способу машинного обучения относится линейная регрессия?

- (?) Обучение без учителя
- (!) Обучение с учителем
- (?) Обучение с подкреплением
- (?) Обучение без подкрепления

2. Что обычно используется в качестве метрики производительности для линейной регрессии?

- (?) Средняя абсолютная ошибка
- (?) Логистическая функция потерь

(!) Среднеквадратичная ошибка

(?) F1-мера

3. В случае одного регрессора сколько параметров необходимо определить для решения задачи линейной регрессии?

(?) 1

(!) 2

(?) 3

(?) 4

4. При градиентном спуске как на каждом шаге изменяются искомые параметры?

(?) В сторону увеличения первой производной

(!) В сторону уменьшения первой производной

(?) В сторону увеличения второй производной

(?) В сторону уменьшения второй производной

5. В чем особенность стохастического градиентного спуска?

(?) Значения частных производных вычисляются по всему набору данных

(?) Значения частных производных заменяются на случайные числа

(?) Значения частных производных вычисляются по небольшой группе случайно выбранных элементов данных

(!) Значения частных производных вычисляются по одному, случайно выбранному элементу данных

6. В множественной линейной регрессии

(?) Один регрессор и одна зависимая переменная

(!) Несколько регрессоров и одна зависимая переменная

(?) Один регрессор и несколько зависимых переменных

(?) Несколько регрессоров и несколько зависимых переменных

7. Какой тип задач машинного обучения решает логистическая регрессия?

(?) Регрессия

(!) Классификация

(?) Кластеризация

(?) Всё вышеперечисленное

8. К чему приводит регуляризация?

(?) Смещение и разброс уменьшаются

(?) Смещение уменьшается, разброс увеличивается

(!) Смещение увеличивается, разброс уменьшается

(?) Смещение и разброс увеличиваются

9. Что из нижеперечисленного не является типом регуляризации?

(?) Lasso

(!) Logloss

(?) Elastic Net

(?) Ridge

10. Условная вероятность $P(A|B)$ – это

(!) Вероятность наступления A при условии, что B уже произошло.

(?) Вероятность наступления B при условии, что A уже произошло.

(?) Вероятность наступления A при условии, что B не произошло.

(?) Вероятность наступления B при условии, что A не произошло.

11. Если события A и B независимы, то чему равна условная вероятность $P(A|B)$?

(?) $P(A) P(B)$

(!) $P(A)$

(?) $P(B)$

(?) $P(A) / P(B)$

12. Для определения $P(A|B)$ с помощью теоремы Байеса нужно знать

(?) $P(A)$

(?) $P(B)$

(?) $P(B|A)$

(!) Всё вышеперечисленное.

13. Теорема Байеса позволяет

(?) Уточнить вероятность гипотезы до проведения эксперимента.

(?) Вычислить вероятность успешного проведения эксперимента.

(!) Уточнить вероятность гипотезы после проведения эксперимента.

(?) Вычислить вероятность неуспешного проведения эксперимента.

14. В чём заключается «наивность» Байесовского классификатора?

(?) Значения параметров в элементе не могут повторяться.

(!) Параметры элементов в наборе данных считаются независимыми.

(?) Количество классов равно двум.

(?) Порядок следования параметров в элемент определяет их важность.

15. С каким минусом Байесовского классификатора борется сглаживание Лапласа?

- (?) Низкая скорость обучения.
- (?) Параметры должны быть независимыми.
- (?) Значения параметров должны быть категориальными.
- (!) Невозможно работать со значениями параметров, которых не было в обучающей выборке.

16. Что из перечисленного не является плюсом Байесовского классификатора?

- (?) Хорошо работает в случае малого количества элементов.
- (?) Хорошо работает в случае большого количества параметров.
- (!) Возможность работы с числовыми значениями параметров.
- (?) Быстрые обучения и работа.

17. Какие типы задач решают наивный Байесовский классификатор и логистическая регрессия?

- (?) Наивный Байесовский классификатор – задачу классификации, логистическая регрессия – задачу регрессии.
- (?) Наивный Байесовский классификатор – задачу регрессии, логистическая регрессия – задачу классификации.
- (?) Оба алгоритма решают задачу регрессии.
- (!) Оба алгоритма решают задачу классификации.

18. Какие из алгоритмов относятся к генеративному и дискриминативному подходу?

- (?) Наивный Байесовский классификатор и логистическая регрессия оба относятся к генеративному подходу.
- (?) Наивный Байесовский классификатор и логистическая регрессия оба относятся к дискриминативному подходу.
- (!) Наивный Байесовский классификатор – генеративный подход, логистическая регрессия – дискриминативный подход.
- (?) Наивный Байесовский классификатор – дискриминативный подход, логистическая регрессия – генеративный подход.

19. Какие типы задач машинного обучения решаются методом k -ближайших соседей?

- (?) Кластеризация и классификация.
- (!) Классификация и регрессия.
- (?) Регрессия и кластеризация.
- (?) Только кластеризация.

20. Что означает k в методе k -ближайших соседей?

- (?) Количество используемых параметров из набора данных.

- (?) Размер обучающей выборки.
- (!) Количество элементов, необходимых для принятия решения.
- (?) Размерность пространства, в котором производятся вычисления.

21. Что утверждает гипотеза компактности?

- (!) Более близкие объекты чаще относятся к одному и тому же классу, чем к разным.
- (?) Более близкие объекты чаще относятся к разным классам, чем к одному и тому же.
- (?) Количество элементов в каждом классе примерно одинаково.
- (?) Количество классов не превышает заранее выбранное значение.

22. Что из перечисленного не является проблемой метода k -ближайших соседей?

- (?) Алгоритм медленно работает с большими наборами данных.
- (?) Алгоритм плохо работает с большим числом параметров.
- (!) Алгоритм плохо интерпретируем.
- (?) Необходимо заранее выбрать функцию расстояния.

23. Для решения какой проблемы метода k -ближайших соседей применяется кросс-валидация?

- (!) Необходимо заранее выбрать значение k .
- (?) Алгоритм медленно работает из-за периферийных элементов.
- (?) Алгоритм плохо работает с несбалансированными данными.
- (?) Алгоритм совершает ошибки из-за выбросов.

24. Что из перечисленного не относится к преимуществам метода k -ближайших соседей?

- (?) Нет требований к значениям параметров.
- (?) Обучающая выборка может легко дополняться.
- (!) Устойчивость к несбалансированным данным.
- (?) Нет явного процесса обучения.

25. Что происходит с числом кластеров в процессе работы метода k -средних?

- (?) Число кластеров увеличивается от 1 до k .
- (?) Число кластеров уменьшается от k до 1.
- (!) Число кластеров всё время равно k .
- (?) Число кластеров никак не зависит от k .

26. Что из перечисленного не является недостатком метода k -средних?

- (!) Алгоритм плохо распараллеливается.
- (?) Алгоритм останавливается в первом достигнутом локальном минимуме.
- (?) Нужно заранее знать число кластеров k .
- (?) Результат зависит от изначального выбора центроидов.

27. Что применяется для определения наилучшего значения k в методе k -средних?

- (!) Коэффициент силуэта.
- (?) Перебалансировка данных.
- (?) Кросс-валидация.
- (?) Гипотеза компактности.

Первое тестирование (6-й семестр)

1. Какую из этих задач машинного обучения не решает метод опорных векторов?
 - (?) Бинарная классификация
 - (?) Многоклассовая классификация
 - (!) Кластеризация
 - (?) Регрессия
2. В методе опорных векторов если размерность элемента данных равна N , то чему равна размерность разделяющей поверхности?
 - (?) N
 - (!) $N-1$
 - (?) $N/2$
 - (?) 1
3. К чему сводится задача бинарной классификации методом опорных векторов в случае линейной разделимости?
 - (?) Минимизация зазора.
 - (!) Максимизация зазора.
 - (?) Минимизация количества опорных векторов.
 - (?) Максимизация количества опорных векторов.
4. Как решается проблема линейной неразделимости в методе опорных векторов?
 - (!) Увеличение размерности пространства элементов данных.
 - (?) Уменьшение размерности пространства элементов данных.
 - (?) Уменьшение количества элементов данных.
 - (?) Данная проблема неразрешима.

5. Что такое ядро при решении линейно неразделимой задачи бинарной классификации методом опорных векторов?
- (?) Сумма отображений элементов.
 - (?) Сумма элементов.
 - (!) Скалярное произведение отображений элементов.
 - (?) Скалярное произведение элементов.
6. В чем заключается парный подход при решении задачи многоклассовой классификации?
- (?) Решение задачи бинарной классификации для двух самых больших классов.
 - (?) Решение задачи бинарной классификации для каждого класса и случайно выбранной пары классов.
 - (!) Решение задачи бинарной классификации для всех возможных пар классов.
 - (?) Решение задачи бинарной классификации для двух случайно выбранных групп классов.
7. При решении задачи регрессии методом опорных векторов чему соответствует зазор?
- (!) Допустимому отклонению значения от найденного решения.
 - (?) Значению среднеквадратичной ошибки.
 - (?) Значению средней абсолютной ошибки.
 - (?) Количеству элементов, для которых отклонение равно нулю.
8. Какой граф, называется деревом?
- (?) Циклический и связный.
 - (!) Ациклический и связный.
 - (?) Циклический и несвязный.
 - (?) Ациклический и несвязный.
9. Какие типы задач машинного обучения можно решать с помощью деревьев принятия решений?
- (?) Кластеризацию и классификацию.
 - (!) Классификацию и регрессию.
 - (?) Регрессию и кластеризацию.
 - (?) Кластеризацию, классификацию и регрессию.
10. Для чего используются энтропия и коэффициент Джини?
- (?) Для вычисления количества ветвей.
 - (?) Для отсечения ветвей.
 - (!) Для построения новых узлов.
 - (?) Для борьбы с переобучением.

11. Что из перечисленного не используется как критерий остановки алгоритма построения дерева?
- (?) Ограничение глубины.
 - (!) Максимальное количество ветвей, выходящих из узла.
 - (?) Ранняя остановка.
 - (?) Минимальное количество элементов в листе.
12. В чём заключается основная идея отсечения ветвей?
- (?) Удаление части ветвей случайным образом.
 - (?) Удаление листьев, содержащих один элемент данных.
 - (?) Удаление листьев, расположенных дальше всего от корня дерева.
 - (!) Удаление листьев, которые минимально влияют на качество работы дерева.
13. Какой из алгоритмов построения деревьев позволяет решать задачи регрессии?
- (?) ID3.
 - (?) C4.5.
 - (!) CART.
 - (?) Ни один из этих алгоритмов.
14. Что не является преимуществом деревьев принятия решений?
- (?) Низкие требования к предобработке данных.
 - (!) Устойчивость к переобучению.
 - (?) Хорошая интерпретируемость.
 - (?) Быстрые обучение и работа.
15. Что из нижеперечисленного не является ансамблем?
- (?) Случайный лес.
 - (!) Наивный Байесовский классификатор.
 - (?) Адабуст.
 - (?) Градиентный бустинг.
16. Что из недостатков «слабых» моделей: высокий разброс и высокое смещение – может компенсировать ансамбль?
- (?) Высокий разброс.
 - (?) Высокое смещение.
 - (!) И то, и другое.
 - (?) Ни то, ни другое.
17. Для чего в бэггинге формируются бутстреп-выборки?
- (?) Для определения порядка объединения моделей в ансамбль.
 - (?) Для предотвращения переобучения.

(?) Для вычисления коэффициентов, с которыми модели объединяются в ансамбль.

(!) Для получения нескольких приблизительно независимых наборов данных.

18. Какие «слабые» модели используются при построении случайного леса?

(!) Деревья принятия решений.

(?) Искусственные нейронные сети.

(?) Модели линейной регрессии.

(?) Наивные Байесовские классификаторы.

19. Что из нижеперечисленного не является достоинством случайного леса?

(?) Возможность параллельного обучения.

(!) Небольшой размер моделей.

(?) Устойчивость к переобучению.

(?) Не требуется настройка гиперпараметров.

20. Что из нижеперечисленного не является недостатком Адабуст?

(!) Сложность реализации.

(?) Склонность к переобучению при наличии шума в данных.

(?) Требование большой длины обучающей выборки.

(?) Сложность интерпретации.

21. Что из нижеперечисленного не является программной реализацией бустинга?

(?) CatBoost.

(?) XGBoost.

(!) AdaBoost.

(?) LightGBM.

22. Какая часть формального нейрона делает его нелинейным?

(?) Порог.

(!) Функция активации.

(?) Входной сигнал.

(?) Синаптические веса.

23. Что означает слово «прямой» в «сети прямого распространения»?

(!) Сигнал передается от слоя к слою слева направо.

(?) Сигнал передается от нейрона к нейрону сверху вниз.

(?) Функция активации имеет вид прямой.

(?) Выходные значения передаются на выходные нейроны напрямую, без изменений.

24. Какое главное свойство активационной функции в многослойной нейронной сети?
- (?) Возрастание.
 - (?) Ограниченная область значений.
 - (!) Дифференцируемость.
 - (?) Симметричность относительно начала координат.
25. Какую из перечисленных функций активации имеет смысл применять только в выходном слое?
- (!) Линейная.
 - (?) ReLU.
 - (?) Сигмоида.
 - (?) Softmax.
26. Что из нижеперечисленного не является недостатком градиентного спуска?
- (?) Чувствительность к скорости обучения.
 - (?) Застревание в локальных минимумах.
 - (!) Необходимость знать ландшафт пространства поиска.
 - (?) Нерегулярность обновления важных, но редких, параметров.
27. Что из нижеперечисленного не является гиперпараметров алгоритма построения искусственной нейронной сети?
- (?) Количество эпох обучения.
 - (?) Начальные значения весов.
 - (?) Количество скрытых слоёв сети.
 - (!) Конечные значения весов.
28. Что из нижеперечисленного не является недостатком искусственных нейронных сетей?
- (!) Низкая скорость работы обученной сети.
 - (?) Необходимость большого количества данных для обучения.
 - (?) Плохая интерпретируемость.
 - (?) Длительный процесс обучения.

Второе тестирование (6-й семестр)

1. К какой задаче в рамках обработки естественного языка относится голосовая биометрия?
- (!) Распознавание речи.
 - (?) Синтез речи.
 - (?) Анализ текста.
 - (?) Генерация текста.

2. К какой задаче в рамках обработки естественного языка относится машинный перевод?
 - (?) Распознавание речи.
 - (?) Синтез речи.
 - (?) Анализ текста.
 - (!) Генерация текста.

3. Как называется разбивка текста на предложения и отдельные слова?
 - (?) Лемматизация.
 - (?) Нормирование.
 - (!) Токенизация.
 - (?) Оцифровка.

4. Чему равна длина векторных представлений при использовании Bag of words, если размер словаря равен N ?
 - (?) $N/2$.
 - (!) N .
 - (?) N^2 .
 - (?) $2N$.

5. Что используется для того, чтобы Bag of words мог работать с неизвестными ранее словами?
 - (!) Hashing trick.
 - (?) n -граммы.
 - (?) Шумовые слова.
 - (?) TF-IDF.

6. Если одно и то же слово встречается в каждом тексте из набора, то что верно для этого слова?
 - (?) $tf = 0, idf = 0$.
 - (!) $tf \neq 0, idf = 0$.
 - (?) $tf = 0, idf \neq 0$.
 - (?) $tf \neq 0, idf \neq 0$.

7. Что используется в качестве меры близости в word2vec?
 - (?) Евклидово расстояние между векторными представлениями.
 - (?) Модуль разности векторных представлений.
 - (?) Количество совпадающих букв.
 - (!) Косинусное расстояние между векторными представлениями.

8. В каком случае на вход seq2seq модели подается не текст?
 - (?) Разговорная модель.
 - (!) Генерация названий.
 - (?) Резюмирование текста.

- (?) Машинный перевод.
9. Что не относится к достоинствам рекуррентных нейронных сетей?
(?) Размер модели не зависит от длины входной последовательности данных.
(?) Данные могут сохраняться в процессе работы сети.
(?) Нет ограничений на длину входной последовательности данных.
(!) Высокая скорость работы сети.
10. Какая функция активации используется в фильтрах сети с долгой краткосрочной памятью?
(!) Сигмоида.
(?) Линейная.
(?) Гиперболический тангенс.
(?) ReLU.
11. Где в seq2seq моделях применяется механизм внимания?
(?) Внутри энкодера.
(?) Внутри декодера.
(!) Между энкодером и декодером.
(?) После декодера.
12. Где применяется механизм внимания в архитектуре трансформеров?
(?) В энкодере и между энкодером и декодером.
(?) В декодере и между энкодером и декодером.
(?) В энкодере и декодере.
(!) В энкодере, декодере и между ними.
13. Какое утверждение относительно архитектуры трансформеров верно?
(?) И энкодер, и декодер состоят из одного блока каждый.
(?) Энкодер состоит из одного блока, декодер – из нескольких.
(?) Энкодер состоит из нескольких блоков, декодер – из одного.
(!) И энкодер, и декодер состоят из нескольких блоков.
14. Какие существуют способы использования архитектуры трансформеров?
(!) Полная архитектура и отдельные энкодер или декодер.
(?) Полная архитектура и отдельный энкодер.
(?) Полная архитектура и отдельный декодер.
(?) Только полная архитектура.
15. Что не относится к поиску объектов на изображении?
(?) Детектирование объектов.
(!) Генерация изображений.

- (?) Классификация изображений.
- (?) Сегментация изображений.

16. Как называется построение на изображении прямоугольника, полностью содержащего объект?

- (?) Классификация изображения.
- (?) Семантическая сегментация.
- (?) Сегментация экземпляров.
- (!) Детектирование объекта.

17. Что происходит при семантической сегментации изображения с двумя объектами одного класса?

- (?) Выделяется область, содержащая один из объектов.
- (!) Выделяется область, содержащая оба объекта.
- (?) Выделяется область, содержащая пересечение объектов.
- (?) Семантическая сегментация не применима в таком случае.

18. К какому типу задач относится распознавание CAPTCHA?

- (?) Отслеживание движущихся объектов.
- (?) Распознавание лиц.
- (!) Распознавание символов или текста.
- (?) Сегментация экземпляров.

19. Что не относится к задаче генерации изображений?

- (!) Удаление шума с изображения.
- (?) Преобразование набросков в фотореалистичные изображения.
- (?) Создание повторяющегося контента.
- (?) Автоматическое редактирование.

20. В каком формате подаются данные на вход свёрточной сети?

- (?) В виде двумерной матрицы.
- (!) В виде нескольких двумерных матриц одного размера.
- (?) В виде нескольких двумерных матриц разного размера.
- (?) В виде одномерного вектора.

21. В генеративно-состязательной сети что подаётся на вход генератора?

- (?) Обучающий набор образцов.
- (!) Случайные числа.
- (?) Результат работы дискриминатора.
- (?) Результат работы самого генератора.

22. Какой из типов искусственного интеллекта должен решать тест Тьюринга?

- (!) Сильный искусственный интеллект.

- (?) Слабый искусственный интеллект.
- (?) Искусственный интеллект общего назначения (AGI)
- (?) Ни один из приведённых вариантов.

23. Что не являются одним из путей к построению AGI?

- (?) Обработка естественного языка.
- (!) Рекуррентные нейронные сети.
- (?) Обучение с подкреплением.
- (?) Компьютерное зрение.

24. Что не является частью входных данных в задаче обучения с подкреплением?

- (?) Множество действий.
- (?) Функция подкрепления.
- (!) Стратегия выбора действий.
- (?) Множество состояний среды.

25. Какой из перечисленных методов обучения с подкреплением является основанным на модели?

- (?) Q-обучение.
- (?) Метод Монте-Карло.
- (?) SARSA.
- (!) Динамическое программирование.

26. В чём заключается основное ограничение метода Монте-Карло при решении задачи обучения с подкреплением?

- (!) Необходимы завершённые эпизоды опыта агента.
- (?) Количество действий должно быть меньше количества состояний среды.
- (?) Необходимо иметь полную модель среды.
- (?) Стратегия выбора действия должна описываться линейной функцией.

27. Какой из перечисленных методов решения задачи обучения с подкреплением не использует понятие временной разности?

- (?) SARSA.
- (?) Deep Q-Network.
- (!) Метод Монте-Карло.
- (?) Q-обучение.

28. В чём различие между SARSA и Q-обучением?

- (!) В способ обновления значений качества $Q(s,a)$.
- (?) В способе выбора действия.
- (?) В учёте влияния подкрепления.

(?) В использовании коэффициента дисконтирования.

4.2. Типовые вопросы, выносимые на зачет

1. Параметры и гиперпараметры модели машинного обучения.
2. Переобучение и недообучение модели машинного обучения.
3. Интерпретация модели машинного обучения.
4. Способы машинного обучения.
5. Типы задач машинного обучения.
6. Метрики производительности.
7. Типы данных и требования к ним.
8. Несбалансированные данные.
9. Формирование обучающей, валидационной и тестовой выборок.
10. Смещение и разброс.
11. Регуляризация.
12. Линейная регрессия.
13. Логистическая регрессия.
14. Наивный Байесовский классификатор.
15. Метод k ближайших соседей.
16. Метод k средних.

4.3. Типовые вопросы, выносимые на экзамен

1. Как работает машинное обучение? Алгоритмы и модели машинного обучения.
2. Кроссвалидация. Применение и типы.
3. Задача регрессии.
4. Задача классификации.
5. Задача кластеризации.
6. Очистка данных.
7. Уменьшение размерности.
8. Линейная регрессия.
9. Логистическая регрессия.
10. Наивный Байесовский классификатор.
11. Метод k ближайших соседей.
12. Метод k средних.
13. Метод опорных векторов.
14. Деревья принятия решений.
15. Ансамбли – бэггинг.
16. Ансамбли – случайный лес.
17. Ансамбли – бустинг.
18. Искусственные нейронные сети прямого распространения.
19. Bag of words.
20. TF-IDF.
21. Wor2vec.

- 22.Рекуррентные нейронные сети.
- 23.Сети с долгой краткосрочной памятью.
- 24.Механизм внимания. Трансформеры.
- 25.Свёрточные нейронные сети.
- 26.Генративно-состязательные сети.
- 27.Обучение с подкреплением. Метод Монте-Карло.
- 28.Обучение с подкреплением. Метод временной разности.

Методические указания для обучающихся по освоению дисциплины

**ИНСТИТУТ
ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ**

**КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
И УПРАВЛЯЮЩИХ СИСТЕМ**

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО
ОСВОЕНИЮ ДИСЦИПЛИНЫ**

«ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ»

(Приложение 2 к рабочей программе)

Направление подготовки: 09.03.02 Информационные системы и технологии

Профиль: информационные технологии в технических системах

Уровень высшего образования: бакалавриат

Форма обучения: очная, заочная

Королев 2021

1. Общие положения

Целью изучения дисциплины является изучение основных алгоритмов и методов машинного обучения.

Задачами дисциплины являются:

1. Изучение различных типов задач, входящих в предметную область «машинное обучение»;
2. Изучение различных алгоритмов для решения задач машинного обучения;
3. Освоение методов сбора и обработки данных для их использования в задачах машинного обучения;
4. Освоение методов для сравнения эффективности различных алгоритмов и для подбора их параметров при решении задач машинного обучения;
5. Описание современного состояния исследований и последних достижений в области машинного обучения.

2. Указания по проведению практических занятий

1 СЕМЕСТР

Практическое занятие 1

Вид практического занятия: смешанная форма практического занятия.

Тема и содержание практического занятия:

1. Изучить устройство и базовые команды Git.
2. Научиться создавать новый репозиторий и копировать уже существующий.
3. Научиться сохранять изменения в репозитории.
4. Изучить основные типы данных и операции Python.
5. Научиться работать с Jupyter notebook.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 4 ч / 3 ч.

Практическое занятие 2

Вид практического занятия: смешанная форма практического занятия.

Тема и содержание практического занятия:

1. Изучить способ обновления своего форка из оригинального репозитория.
2. Реализовать и сравнить два способа формирования обучающей выборки.
3. Определить параметры и гиперпараметры приведённой модели машинного обучения.
4. По данным графикам ошибок на обучающей и тестовой выборках определить, наблюдается ли эффект переобучения или недообучения.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 4 ч / 3 ч.

Практическое занятие 3

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Используя результаты работы моделей машинного обучения, решающих задачи регрессии, классификации и кластеризации, научиться вычислять значения основных метрик производительности.

2. Реализовать программный код для вычисления компонентов матрицы путаницы.

3. Для задачи бинарной классификации провести и проанализировать эксперименты по применимости различных метрик.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 4 ч / 3 ч.

Практическое занятие 4

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Научиться находить проблемы в данных.

2. Научиться исправлять найденные проблемы в данных.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 4 ч / 3 ч.

Практическое занятие 5

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Научиться аналитически решать задачу линейной регрессии.

2. Научиться решать задачу линейной регрессии с помощью библиотеки sklearn.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 4 ч / 3 ч.

Практическое занятие 6

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Научиться решать задачу классификации с помощью библиотеки sklearn, используя наивный Байесовский классификатор.

2. Научиться решать задачу классификации с помощью библиотеки sklearn, используя логистическую регрессию.

3. Научиться сравнивать результаты работы моделей классификации, используя F1-меру.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 4 ч / 3 ч.

Практическое занятие 7

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Научиться решать задачу регрессии с помощью библиотеки sklearn, используя метод k -ближайших соседей.

2. Научиться подбирать оптимальное значение параметра k с помощью среднеквадратичной ошибки.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 4 ч / 3 ч.

Практическое занятие 8

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Научиться использовать метод k -средних.

2. Научиться анализировать качество кластеризации методов k -средних.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 4 ч / 3 ч.

ВТОРОЙ СЕМЕСТР

Практическое занятие 9

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Научиться решать задачу регрессии с помощью библиотеки sklearn, используя метод опорных векторов.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 6 ч / 0 ч.

Практическое занятие 10

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Получить практические навыки решения задач регрессии и классификации с помощью деревьев принятия решений.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 6 ч / 0 ч.

Практическое занятие 11

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Сравнить несколько моделей для решения задачи регрессии с помощью ансамблей.

2. Сравнить несколько моделей для решения задачи классификации с помощью ансамблей.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 6 ч / 0 ч.

Практическое занятие 12

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Сравнить несколько моделей для решения задачи регрессии с помощью искусственных нейронных сетей.

2. Сравнить несколько моделей для решения задачи классификации с помощью искусственных нейронных сетей.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 6 ч / 0 ч.

Практическое занятие 13

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Сравнить несколько алгоритмов для решения задачи регрессии.

2. Сравнить несколько алгоритмов для решения задачи классификации.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 6 ч / 0 ч.

Практическое занятие 14

Вид практического занятия: смешенная форма практического занятия.

Тема и содержание практического занятия:

1. Изучить предложенный базовый вариант (baseline) решения задачи.

2. Применить полученные ранее навыки для улучшения baseline.

Образовательные технологии: технологии компьютерного обучения.

Продолжительность занятия – 18 ч / 0 ч.

3. Указания по проведению лабораторного практикума

Не предусмотрен учебным планом.

4. Указания по проведению самостоятельной работы студентов

№ п/п	Наименование блока (раздела) дисциплины	Виды СРС
1.	Тема 1. Машинное обучение. Общая постановка задачи.	<ol style="list-style-type: none"> 1. Что такое машинное обучение. 2. Как работает машинное обучение: данные, модель, алгоритм. 3. Параметры и гиперпараметры модели. 4. Переобучение модели. 5. Недообучение модели. 6. Кросс-валидация модели. 7. Интерпретация моделей машинного обучения.
2.	Тема 2. Способы машинного обучения. Типы решаемых задач. Метрики производительности.	<ol style="list-style-type: none"> 1. Способы машинного обучения: обучение с учителем, обучение без учителя, обучение с подкреплением. 2. Типы задач машинного обучения: регрессия, классификация, кластеризация. 3. Метрики производительности: среднеквадратичная ошибка, средняя абсолютная ошибка, матрица запутанности, коэффициент силуэта.
3	Тема 3. Подготовка данных для машинного обучения.	<ol style="list-style-type: none"> 1. Сбор данных. 2. Структурирование и оцифровка данных. 3. Очистка и исправление данных. 4. Очистка данных. 5. Уменьшение размерности. 6. Нормализация. 7. Обогащение данных. 8. Формирование обучающей, валидационной и тестовой выборок.
4	Тема 4. Линейная регрессия.	<ol style="list-style-type: none"> 1. Метод наименьших квадратов. 2. Аналитическое решение. 3. Градиентный спуск. 4. Допущения линейной регрессии. 5. Множественная линейная регрессия. 6. Ограничения линейной регрессии. 7. Интерпретация линейной регрессии. 8. Преимущества и недостатки.
5.	Тема 5. Логистическая регрессия.	<ol style="list-style-type: none"> 1. Метод максимального правдоподобия. 2. Логистическая функция потерь. 3. Многоклассовая логистическая регрессия. 4. Смещение и разброс. 5. Регуляризация. 6. L1-регуляризация (lasso). 7. L2-регуляризация (ridge). 8. Elastic Net регуляризация. 9. Преимущества и недостатки.
6.	Тема 6. Наивный байесовский классификатор.	<ol style="list-style-type: none"> 1. Условная вероятность. 2. Теорема Байеса. 3. Функции правдоподобия. 4. Сглаживание Лапласа.

		5. Преимущества и недостатки.
7.	Тема 7. Метод ближайших соседей. k -	1. Метод k -ближайших соседей в задаче классификации. 2. Гипотеза компактности. 3. Метод k -ближайших соседей в задаче регрессии. 4. Периферийные элементы и выбросы. 5. Несбалансированные данные. 6. Перебалансировка. 7. Преимущества и недостатки.
8.	Тема 8. Метод k -средних.	1. Расстояние для числовых и категориальных признаков. 2. Алгоритм построения кластеров. 3. Преимущества и недостатки.
9.	Тема 9. Метод опорных векторов.	1. Бинарная классификация. 2. Опорный вектор. 3. Отступ. 4. Линейная неразделимость. 5. Ядро преобразования. 6. Многоклассовая классификация. 7. Парный подход. 8. Один против всех. 9. Применение к задаче регрессии. 10. Преимущества и недостатки.
10.	Тема 10. Деревья принятия решений.	1. Определение дерева. 2. Правила. 3. Энтропия. 4. Коэффициент Джини. 5. Отсечение ветвей. 6. Основные алгоритмы построения деревьев принятия решений. 7. Преимущества и недостатки.
11.	Тема 11. Ансамбли.	1. Типы ансамблей. 2. Бэггинг. 3. Бутстрэп. 4. Агрегация результатов. 5. Случайный лес. 6. Бустинг. 7. Адабуст. 8. Градиентный бустинг. 9. Преимущества и недостатки.
12.	Тема 12. Искусственные нейронные сети и глубокое обучение.	1. Формальный нейрон. 2. Простой перцептрон Розенблатта. 3. Метод коррекции ошибки. 4. Глубокое обучение. 5. Сеть прямого распространения. 6. Метод обратного распространения ошибки. 7. Функции активации. 8. Функция ошибки. 9. Эпохи и батчи. 10. Алгоритм оптимизации. 11. Гиперпараметры искусственной нейронной сети. 12. Дропаут. 13. Начальные значения весов. 14. Преимущества и недостатки.

13.	Тема 13. Обработка естественного языка.	<ol style="list-style-type: none"> 1. Распознавание речи. 2. Синтез речи. 3. Анализ текста. 4. Генерация текста. 5. Предварительная обработка текста. 6. Векторные представления слов. 7. Bag of words. 8. Hashing trick. 9. N-граммы. 10. Шумовые слова. 11. TF-IDF. 12. Word2vec. 13. Continuous bag of words. 14. Skip-gram. 15. Преимущества и недостатки.
14.	Тема 14. Рекуррентные нейронные сети. Механизм внимания. Трансформеры.	<ol style="list-style-type: none"> 1. Подход seq2seq. 2. Рекуррентные нейронные сети. 3. Сети с долгой краткосрочной памяти. 4. Механизм внимания. 5. Трансформеры. 6. Преимущества и недостатки.
15.	Тема 15. Компьютерное зрение.	<ol style="list-style-type: none"> 1. Компьютерное зрение. Машинное зрение. Распознавание образов. 2. Поиск объектов на изображении. Классификация изображений. Детектирование объектов. Сегментация изображений. 3. Обработка найденных объектов. Отслеживание движущихся объектов (трекинг). Распознавание лиц (идентификация). Распознавание символов или текста. 4. Восстановление (реконструкция) изображений. Удаление шума с изображения. Восстановление или добавление цвета. Поворот объекта. Построение 3D-формы. 5. Генерация изображений. 6. Свёрточная нейронная сеть. 7. Генеративно-сопоставительная сеть.
16.	Тема 16. Сильный искусственный интеллект. Обучение с подкреплением.	<ol style="list-style-type: none"> 1. Искусственный интеллект – определение и типизация. 2. Сильный искусственный интеллект. Тест Тьюринга. 3. Обучение с подкреплением. 4. Динамическое программирование. 5. Метод Монте-Карло. 6. Метод временной разности. SARSA. Q-обучение. DQN.

5. Указания по проведению контрольных работ

Цель выполняемой работы: Продемонстрировать знания и умения в изучения дисциплины.

Порядок выполнения контрольной работы:

Контрольная работа излагается логически последовательно, грамотно, разборчиво. Работа должна иметь титульный лист. Он содержит полное

название высшего учебного заведения, кафедры, реализующая данную дисциплину, название (тема) контрольной работы, фамилию, инициалы автора, также необходимо указать номер группы, фамилию и инициалы, а также должность, ученое звание и степень научного руководителя (преподавателя), проверяющего контрольную работу.

На следующем листе излагается план контрольной работы, который включает в себя: название всех разделов, введение и заключение, а также список литературы. Излагая вопрос (раздел) каждый смысловой абзац необходимо начать с красной строки.

В конце работы ставится подпись студента и дата сдачи. Страницы контрольной работы должны быть пронумерованы. Номер страницы ставится в нижнем правом углу.

Размер шрифта №14 (Times New Roman), полуторный интервал, стандартный лист формата А4. Поля: верхнее – 20 мм, нижнее – 20 мм, левое – 30 мм, правое – 15 мм.

Дополнительно контрольная работа может иметь приложения (формулы, схемы, графики, диаграммы).

По всем возникающим вопросам обучающемуся следует обращаться за консультацией на кафедру. Срок выполнения контрольной работы определяется кафедрой. Срок проверки контрольной работы – 3 дня с момента необходимой фиксированной даты сдачи.

6. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

1. Вьюгин, В. В. Математические основы машинного обучения и прогнозирования: Учебное пособие / Вьюгин В.В. - Москва :МЦНМО, 2014. - 304 с.: ISBN 978-5-4439-2014-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/958689> (дата обращения: 16.04.2021). – Режим доступа: по подписке.
2. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / С. Рашка ; пер. с англ. А.В. Логунова. - Москва : ДМК Пресс, 2017. - 418 с. - ISBN 978-5-97060-409-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027758> (дата обращения: 16.04.2021). – Режим доступа: по подписке.

Дополнительная литература:

1. Коэльо, Луис Педро Построение систем машинного обучения на языке Python / Луис Педро Коэльо, Вилли Ричарт ; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва : ДМК Пресс, 2016. - 302 с. - ISBN 978-5-97060-330-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027824> (дата обращения: 16.04.2021). – Режим доступа: по подписке.

2. Саттон, Р. С. Обучение с подкреплением: введение : практическое руководство / Р. С. Саттон, Э. Барто ; пер. с англ. А. А. Слинкина. - Москва : ДМК Пресс, 2020. - 552 с. - ISBN 978-5-97060-097-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1210617> (дата обращения: 16.04.2021). – Режим доступа: по подписке.

Рекомендуемая литература:

1. Танцов, П. Н. Интеллектуальные информационные системы : лабораторный практикум / П. Н. Танцов. - Москва : Изд. Дом МИСиС, 2015. - 86 с. - ISBN 978-5-87623-898-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1232708> (дата обращения: 16.04.2021). – Режим доступа: по подписке.

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. www.machinelearning.ru – профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.
2. <https://neurohive.io/ru/osnovy-data-science/> – базовый курс «Основы Data Science».
3. <https://dyakonov.org/> – блог д.ф.-м.н. Александра Дьяконова.

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Перечень программного обеспечения: MS Office, Jupyter Notebook, Python, GitHub.

Информационные справочные системы: не предусмотрено курсом данной дисциплины

Ресурсы информационно-образовательной среды МГОТУ:

Рабочая программа и методическое обеспечение по дисциплине «Основы машинного обучения».