

МИНОБРНАУКИ РОССИИ

Ярославский государственный университет им. П.Г. Демидова

Кафедра вычислительных и программных систем

УТВЕРЖДАЮ

Декан факультета ИВТ

 Д.Ю. Чалый

«23» мая 2023 г.

Рабочая программа дисциплины

«Машинное обучение»

Направление подготовки

02.04.02 Фундаментальная информатика и информационные технологии

Направленность (профиль)

«Искусственный интеллект и компьютерные науки»

Квалификация выпускника

Магистр

Форма обучения

очная

Программа рассмотрена
на заседании кафедры
от «21» апреля 2023 г.,
протокол № 8

Программа одобрена НМК
факультета ИВТ
протокол № 6 от
«28» апреля 2023 г.

Ярославль

1. Цели освоения дисциплины

Дисциплина «Машинное обучение» относится к обязательной части ОП магистратуры.

Курс содержит подробный разбор наиболее распространенных современных методов машинного обучения. Излагаются основные алгоритмы классификации и подходы к построению систем классификации. В рамках курса изучаются: обучение без учителя, линейная регрессия, градиентный спуск, классификация на основе деревьев, линейная классификация и классификация на основе ядер. Рассматриваются подходы к классификации на много классов и ее особенности, обсуждаются различные метрики качества классификации и проблемы переобучения.

2. Место дисциплины в структуре ОП магистратуры

Дисциплина «Машинное обучение» относится к базовой части ОП магистратуры

Для освоения данной дисциплины студентам необходимы знания, полученные при изучении дисциплин «Основы программирования», «Высокоуровневое программирование», «Языки программирования».

Знания и навыки, полученные при изучении дисциплины «Машинное обучение», используются учащимися при изучении последующих дисциплин, таких как «Компьютерное моделирование», «Интеллектуальные системы», «Программная инженерия».

3. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения ОП магистратуры.

Процесс изучения дисциплины направлен на формирование следующих элементов компетенций в соответствии с ФГОС ВО, ОП ВО и приобретения следующих знаний, умений, навыков и (или) опыта деятельности:

Код компетенции	Формулировка компетенции	Перечень планируемых результатов обучения
Общепрофессиональные компетенции		
ОПК-2. Способен применять компьютерные/суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности	ОПК-2.1. Применяет современные информационно-коммуникационные интеллектуальные компьютерные технологии, инструментальные среды, программно-технические платформы для решения задач в области создания и применения искусственного интеллекта ОПК-2.2. Обосновывает выбор современных информационно-коммуникационных и интеллектуальных компьютерных технологий ОПК-2.3 Разрабатывает оригинальные программные средства, в том числе с использованием	

	современных информационно-коммуникационных и интеллектуальных компьютерных технологий, для решения задач в области создания и применения искусственного интеллекта	
ОПК-6. Способен разрабатывать алгоритмы и программные средства для решения задач в области создания и применения искусственного интеллекта	ОПК-6.1. Применяет инструментальные среды, программно-технические платформы для решения задач в области создания и применения искусственного интеллекта ОПК-6.2. Разрабатывает оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта	
ОПК-7. Способен адаптировать и применять на практике классические и новые научные принципы и методы исследования для решения задач в области создания и применения технологий и систем искусственного интеллекта и методы исследований	ОПК-7.1. Адаптирует известные научные принципы и методы исследований с целью их практического применения ОПК-7.2. Решает профессиональные задачи на основе применения новых научных принципов и методов исследования	
УК-7. Способен понимать фундаментальные принципы работы современных систем искусственного интеллекта, разрабатывать правила и стандарты взаимодействия человека и искусственного интеллекта и использовать их в социальной и профессиональной деятельности	УК-7.1. Использует нормативно-правовую базу, правовые, этические правила, стандарты при решении задач искусственного интеллекта УК-7.2. Разрабатывает стандарты, правила в сфере искусственного интеллекта и смежных областях и использует их в социальной и профессиональной деятельности УК-7.3. Применяет современные методы и инструменты для представления результатов научно-исследовательской деятельности УК-7.4. Владеет нормами международного и	

	<p>российского законодательства в сфере интеллектуальной собственности УК-7.5. Проводит поиск зарегистрированных результатов интеллектуальной деятельности и средств индивидуализации при создании инновационных продуктов в профессиональной деятельности УК-7.6. Осуществляет защиту прав результатов интеллектуальной деятельности и средств индивидуализации при создании инновационных продуктов в профессиональной деятельности</p>	
--	---	--

4. Объем структура и содержание дисциплины «Машинное обучение»

Общая трудоемкость дисциплины составляет 6 зачетные единицы, 216 часов.

№ п/п	Темы (разделы) дисциплины, их содержание	Семестр	Виды учебных занятий, включая самостоятельную работу студентов, и их трудоемкость (в академических часах)						Формы текущего контроля успеваемости Форма промежуточной аттестации (по семестрам)
			Контактная работа						
			лекции	практические	лабораторные	консультации	аттестационные испытания	самостоятельная работа	
1	Введение в машинное обучение.	1	2					3	
2	Язык Python введение.	1	1	2				3	
3	Знакомство с библиотекой Numpy.	1		2				3	
4	Знакомство с библиотекой Pandas.	1		2				3	

5	Первичный анализ данных	1	1	2			3		
6	Метрические методы.	1	2	4		1	3		
7	Линейные методы.	1	2	4			3		
8	Решающие деревья.	1	1	2			3		
9	Методы понижения размерности.	1	1	2		1	3		
10	Обучение без учителя.	1	1	2			3		
11	Рекомендательные системы.	1	1	2			3,7		
	Итого, 1й семестр:		12	24		2	0,3	33,7	
12	Введение в обработку изображений.	2	2	4			10		
13	Нахождение границ.	2	2	4			10		
14	Простые методы детекции объектов.	2	1	2			10		
15	Вычисление признаков из изображения.	2	1	2			10		
16	Метод опорных векторов (SVM).	2	2	4		1	10		
17	Композиция алгоритмов.	2	2	4			10		
18	Знакомство в нейронными сетями.	2	2	4		1	10		
						2		Экзамен	
	Всего за 2 семестр		12	24		4	34	70	Экзамен
	Всего		24	48		4	34,3	103,7	

Содержание разделов дисциплины:

Введение в машинное обучение.

Классификация задач машинного обучения: обучение с учителем, обучение без учителя, рекомендательные системы. Основные термины: объекты, признаки, метки (ответы), решающее правило, обобщающая способность, переобучение.

Постановки задач классификации и регрессии. Виды признаков (бинарные, числовые, порядковые, номинальные, количественные). Примеры задач. Современное состояние области, успешные проекты. Связь с другими областями, в частности компьютерным зрением.

Язык Python введение. Основные характеристики языка, синтаксис. Базовые типы данных и операции с ними. Преобразования типов. Условные операторы (ветвления). Циклы в Python (for и while). Функция range. Структуры данных: списки, кортежи, словари и множества. Функции, модули и библиотеки, импорт.

<p>Знакомство с библиотекой NumPy. Сравнение numpy.array и list. Срезы. Логическая индексация. Особенности операций +, *, **. Сравнение range, arrange, xrange. Генерация случайных чисел из разных распределений. Знакомство с библиотекой Matplotlib. Рисование графиков plot, subplot, подписи к осям, название, легенда, задание цвета и вида графика. Диаграммы: bar, hist, bxpplot, pie. Рисование тепловой карты imshow.</p>
<p>Знакомство с библиотекой Pandas. Тип данных Series: объявление, индексация, обращение к элементам, отбор по условию. Пропуски и их заполнению (fillna). Тип данных DataFrame. Индексация, названия столбцов. Обращение к элементам (at, loc, iloc). Добавление/удаление элементов. Проверка на пропуски, удаление, заполнение. Нахождение и удаление дубликатов. Переиндексация. Методы apply, sort, groupby.</p>
<p>Первичный анализ данных (библиотека sklearn). Этапы анализа данных. Чтение csv файла, информация о наборе данных (info, describe, head/tail). Подходы к заполнению пропусков. Преобразование номинальных признаков (Label Encoding, One-Hot Encoding, Hashing). Нормировка числовых признаков (StandardScaling, MinMax Scaling). Работа с текстовыми признаками (map, мешок слов, N-граммы, TF-IDF). Как правильно оценивать качество модели (отложенная выборка, кросс-валидация, LOO). Метрики качества (accuracy, точность, полнота, F1-мера).</p>
<p>Метрические методы. Гипотеза компактности и непрерывности. Меры близости, обобщенная метрика Минковского. Пример задачи классификации цветков ириса. Обобщенный метрический классификатор. Метод kNN. Окно Парзена. Задача регрессии, метод наименьших квадратов. Непараметрическая регрессия, ядерное сглаживание Надарая-Ватсона. Виды функции ядра. Алгоритм LOWESS. Применение методов библиотеки sklearn: train_test_split, cross_val_score, accuracy_score. Применение метрических методов с использованием библиотеки sklearn. Методы KNeighborsClassifier, GridSearchCV, Pipeline (нормировка признаков и kNN).</p>
<p>Линейные методы. Линейная регрессия. Метрики качества регрессии. Градиентный спуск и способы оценивания градиента. Переобучение и регуляризация. Свойства L1, L2 регуляризации. Методы библиотеки sklearn: LinearRegression, Lasso, Ridge, ElasticNet. Линейная классификация. Логистическая регрессия и оценки вероятности классов. Многоклассовая классификация, сведение к бинарным задачам. Многоклассовая логистическая регрессия. Подходы к вычислению метрик качества macro micro усреднение. Понятие PR, ROC-кривая, метрики AUC-PR, AUC-ROC.</p>
<p>Решающие деревья. Жадный алгоритм построения. Выбор лучшего разбиения с помощью критерия информативности. Критерии информативности для регрессии и классификации. Учёт пропусков в деревьях. Решающие деревья и категориальные признаки. Усечение дерева. Деревья для задачи регрессии. Небрежные решающие деревья.</p>
<p>Методы понижения размерности. Линейные методы, метод главных компонент (PCA). Нелинейные методы MDS (многомерное шкалирование), SNE, Tsne.</p>
<p>Обучение без учителя. Кластеризация. Метрики близости. Алгоритмы: k средних, fuzzy k-means, иерархические методы. Нелинейные методы кластеризации DB-SCAN. Обсуждение метрик качества кластеризации (бизнес-метрики).</p>

<p>Рекомендательные системы. Подходы content-base, collaborative filtering, user-based, item-base, на основе SVD разложения. Использование регуляризации. Оценка качества рекомендаций</p>
<p>Введение в обработку изображений. Цветовые пространства: RGB, CMY, CIE Lab, HSB/HSL/HSI. Понятие гистограммы яркости. Фильтрация изображений. Понятие свертка, подходы к вычислению. Фильтр Гаусса, медианный фильтр (нелинейный). Повышение резкости. Знакомство с библиотекой OpenCV. Общее описание что содержит библиотека, какие языки и платформы поддерживает. Модули. Пример простого приложения с загрузкой и отображением картинки. Практическое использование изученных понятий.</p>
<p>Нахождение границ. Понятие градиента изображения. Методы вычисления, операторы Робертса, Превитта, Собеля, сила и направление градиента. Детектор Кэнни выделение границ. Гистограмма градиентов (HOG).</p>
<p>Простые методы детекции объектов. Бинаризация изображений. Морфологические операции сужение, расширение. Морфологическая фильтрация, нахождение границы, скелета. Алгоритм вычисления компонент связности (объектов).</p>
<p>Вычисление признаков из изображения. Цветовые признаки, гистограмма яркости HOG как признаки. Геометрические признаки объектов (площадь, периметр, вытянутость, компактность, моменты). Сопоставление контуров, Distance Transform для быстрого вычисления расстояния между контурами. Текстуриный анализ, идея банка фильтров (фильтры разных направлений).</p>
<p>Метод опорных векторов (SVM). Линейно разделимая выборка. Постановка задачи регрессии, аппроксимация функции для задачи классификации. Случай линейной отделимости, постановка задачи оптимизации и ее решение. Понятия опорного вектора, опорные нарушители. Нелинейное обобщение, ядерный трюк. Виды ядер. SVM-регрессия. Практическое применение метода SVM для классификации и регрессии (библиотека sklearn). Подбор параметра C. Как влияет выбор ядра на форму разделяющей поверхности, тестирование разных ядер. Детекция пешеходов. Идея скользящего окна. Вычисление признаков для фрагмента изображения (HOG). Применение метода SVM для классификации пешеход/НЕ пешеход. Улучшения: использование блоков, улучшение выборки при обучении за счет использования сложных примеров с предыдущего шага. Проблема дисбаланса классов, выбор правильной метрики качества.</p>
<p>Композиция алгоритмов. Разные подходы: взвешенное голосование, Bagging, Boosting, Staking. Случайный лес. Бустинг над деревьями. XGBoost. Детектор лиц Виоло-Джонса. Идея слабых классификаторов. Понятие интегральной матрицы. Алгоритм AdaBoost. Каскад классификаторов, требования к их качеству.</p>
<p>Знакомство в нейронными сетями. Математическая модель нейрона. Идея суммирования входов и применение функции активации. Виды функций активации. Многослойный перцептрон. Проблема исключяющего или. Возможности двухслойного перцептрона. Понятие функции потерь, сведение к задаче оптимизации, обратное распространение ошибки, градиентный спуск. Сверточные НС. Сверточный слой, слой прореживания, функция softmax. Современные архитектуры СНС. Примеры успешно решенных задач: распознавание рукописных цифр, набор ImageNet, распознавание лиц.</p>

5. Образовательные технологии, используемые при осуществлении образовательного процесса

В процессе обучения используются следующие образовательные технологии:

Вводная лекция – дает первое целостное представление о дисциплине и ориентирует студента в системе изучения данной дисциплины. Студенты знакомятся с назначением и задачами курса, его ролью и местом в системе учебных дисциплин и в системе подготовки в целом. Дается краткий обзор курса, история развития науки и практики, достижения в этой сфере, имена известных ученых, излагаются перспективные направления исследований. На этой лекции высказываются методические и организационные особенности работы в рамках данной дисциплины, а также дается анализ рекомендуемой учебно-методической литературы.

Академическая лекция (или лекция общего курса) – последовательное изложение материала, осуществляемое преимущественно в виде монолога преподавателя. Требования к академической лекции: современный научный уровень и насыщенная информативность, убедительная аргументация, доступная и понятная речь, четкая структура и логика, наличие ярких примеров, научных доказательств, обоснований, фактов.

Лабораторная работа – организация учебной работы с реальными материальными и информационными объектами, экспериментальная работа с аналоговыми моделями реальных объектов.

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

В процессе осуществления образовательного процесса используются: для разработки документов, презентаций, для работы с электронными таблицами OfficeStd 2013 RUS OLP NL Acdmc 021-10232

LibreOffice (свободное)

– для поиска учебной литературы библиотеки ЯрГУ – Автоматизированная библиотечная информационная система "БУКИ-НEXТ" (АБИС "Буки-Next");

– для проведения лабораторных занятий используется CASE-среда Enterprise Architect (разработчик Sparx Systems).

7. Перечень основной и дополнительной учебной литературы, ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

Основная литература:

Маккинли, У. Python и анализ данных / Уэс Маккинли - Москва : ДМК Пресс, 2015. - 482 с. - ISBN 978-5-97060-315-4. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970603154.html>

Гонсалес, Р. Цифровая обработка изображений / Гонсалес Р. , Вудс Р. - Издание 3-е, исправленное и дополненное. - Москва : Техносфера, 2012. - 1104 с. - ISBN 978-5-94836-331-8. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785948363318.html>

Шапиро, Л. , Стокман Дж. Компьютерное зрение / Л. Шапиро, Дж. Стокман; пер. с англ. - 4-е изд. - Москва : Лаборатория знаний, 2020. - 763 с. Систем. требования: Adobe Reader XI ; экран 10". (Лучший зарубежный учебник) - ISBN 978-5-00101-696-0. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785001016960.html>

Дополнительная литература:

Кэлер, А. , Брэдски Г. Изучаем OpenCV 3 / Кэлер А. , Брэдски Г. , пер. с англ. А. А. Слинкина. - Москва : ДМК Пресс, 2017. - 826 с. - ISBN 978-5-97060-471-7. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970604717.html>
в) ресурсы сети «Интернет»

1. Электронная библиотека учебных материалов ЯрГУ (http://www.lib.uniya.ac.ru/opac/bk_cat_find.php).

2. Информационная система "Единое окно доступа к образовательным ресурсам" (<http://www.edu.ru> (раздел Учебно-методическая библиотека) или по прямой ссылке <http://window.edu.ru/library>).

8. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине включает в свой состав специальные помещения:

- учебные аудитории для проведения занятий лекционного типа и практических занятий (семинаров);

- учебные аудитории для проведения лабораторных занятий;

- учебные аудитории для проведения групповых и индивидуальных консультаций,

- учебные аудитории для проведения текущего контроля и промежуточной аттестации;

- помещения для самостоятельной работы;

- помещения для хранения и профилактического обслуживания технических средств обучения.

Специальные помещения укомплектованы средствами обучения, служащими для представления учебной информации большой аудитории.

Для проведения занятий лекционного типа предлагаются наборы демонстрационного оборудования и учебно-наглядных пособий, хранящиеся на электронных носителях и

обеспечивающие тематические иллюстрации, соответствующие рабочим программам дисциплин.

Помещения для лабораторных занятий и самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Число посадочных мест в лекционной аудитории больше либо равно списочному составу потока, а в аудитории для практических занятий (семинаров) – списочному составу группы обучающихся.

**Приложение №1 к рабочей программе дисциплины
«Машинное обучение»**

**Фонд оценочных средств
для проведения текущей и промежуточной аттестации студентов
по дисциплине**

**1. Типовые контрольные задания или иные материалы,
необходимые для оценки знаний, умений, навыков и (или) опыта
деятельности,
характеризующих этапы формирования компетенций**

**1.1 Контрольные задания и иные материалы,
используемые в процессе текущей аттестации**

Пример теста

1. Чему будет равен корень из среднеквадратичной ошибки для набора из 3 наблюдений, где отклонение предсказания линейной регрессии от реальных значений равны: -1, 2, -2?
 - а) 2
 - б) 3
 - в) 0
 - г) 1.
2. Рассмотрим признак “Образовательная программа” при анализе данных по студентам университета. Этот признак может принимать три значения: “Экономика”, “Математика”, “Философия”. Воспользуемся one-hot кодированием и заменим этот признак на три бинарных, которые будут соответствовать категориям в том порядке, в котором они перечислены выше. Как будет закодирован признак со значением “Философия”?
 - а) (1, 0, 0)
 - б) (1, 1, 0)
 - в) (1, 1, 1)
 - г) (0, 1, 0)
 - д) (1, 0, 1)
 - е) (0, 0, 1)
 - ж) (0, 1, 1).
3. Модель линейной регрессии выглядит так: $a(x)=w_0+w_1 x_1+\dots+w_d x_d$. Сколько у неё параметров?
 - а) d ;
 - б) $d+1$;
 - в) 1;
 - г) $d-1$.
4. Предположим, что мы строим модель предсказания роста по возрасту и весу человека. Модель с какими коэффициентами вероятнее всего переобучилась?
 - а) $0.001 * (\text{возраст}) + 0.5 * (\text{вес})$;
 - б) $0.1 * (\text{возраст}) + 0.33 (\text{вес})$;
 - в) $1402325.3 * (\text{возраст}) + -1404370.5 (\text{вес})$.

5. Предположим, что мы строим модель предсказания стоимости дома по количеству комнат и средней цене дома в районе. Перед количеством комнат коэффициент равен 1400230, а перед средней ценой дома в районе 0.8. Можно ли утверждать, что количество комнат — более важный признак для качества предсказания, чем средняя цена в районе и почему?
- Нет, так как коэффициенты несравнимы, поскольку признаки имеют разный масштаб;
 - Нет, так как средняя цена дома в районе — это признак с большим разбросом, а именно разброс характеризует ценность признака;
 - Да, так как количество комнат — это признак, который может принимать небольшое количество значений, а значит, каждое значение содержит в себе больше информации;
 - Да, так как коэффициент перед количеством комнат больше.
6. Какая из моделей приводит к отбору признаков?
- Линейная регрессия без регуляризации
 - Ridge-регрессия;
 - Lasso-регрессия;
 - ElasticNet.

Правильные ответы

Вопрос №	Правильный ответ	Вопрос №	Правильный ответ
1	б	4	в
2	е	5	а
3	б	6	в

Критерии оценки

- «Отлично» – 6 правильных ответов;
- «Хорошо» – 5 правильных ответов;
- «Удовлетворительно» – 4 правильных ответов;
- «Неудовлетворительно» – 3 и менее правильных ответов.

Примеры лабораторных работ

Написать программу с использованием библиотек, которая решает следующую задачу:

1. Бинаризируйте изображение `gobo3.jpg`, чтобы выделить красные кружки, выбор порога бинаризации должен быть автоматизирован, выбор цветового пространства остается за автором. Проверьте работу Вашего алгоритма на изображениях `gobo2.jpg`, `gobo1.jpg`.
2. Выделите внутреннюю и внешнюю границы на изображении `binar1.jpg` `binar2.jpg` с помощью морфологической обработки. Поясните результаты.
3. Выделите объекты на изображении `binar1.jpg`, вычислите площадь каждого объекта.
4. Продемонстрируйте промежуточные результаты работы алгоритма.
5. Выделите объекты на изображении `Clusters.jpg`. На выходе должно получиться изображение, на котором разные объекты помечены разными цветами, а фон черным.
6. Для изображения `circles.jpg` реализуйте морфологический алгоритм для построения трех изображений, которые бы содержали соответственно: 1) только частицы, касающиеся краев изображения; 2) только группы перекрывающихся частиц; 3) только одиночные круглые частицы.

Задания для группы из нескольких человек

7. Выделите в изображениях `table1.jpg`, `table2.bmp`, `table3.jpg` границы таблицы с использованием морфологических операций. Результатом обработки должно быть изображение, в котором удален весь

- текст и оставлены только границы линий. Какие были проблемы, какие есть пути их решения, какие выбрали вы.
8. Предложите и реализуйте алгоритм нахождения не до конца заполненных бутылок на изображении FigP1126.tif. Бутылка считается заполненной до конца, если уровень выше середины между началом сужения и низом горлышка.
 9. На изображениях eye_.bmp выделите (постройте бинарное изображение) зрачек/радужку. Алгоритм должен работать приемлемо для всех (4 изображений). Какие были проблемы, какие есть пути их решения, какие выбрали вы. Вычислите характеристики выделенного объекта (площадь, периметр, вытянутость, компактность).
 10. Реализовать алгоритм выделения на изображении объемлющего прямоугольника, в котором располагается штрих-код.

Студент защищает лабораторную работу, при этом готовит презентацию с результатами работы программы на разных этапах (подходах к решению). Студент должен быть готов ответить на вопросы по коду программы и пояснить выбор значений параметров в функциях, которые использует. А также понимать алгоритм, который используется в той или иной, сторонней функции.

Показатели	Критерии
Содержание программы	Анализирует изученный материал, Правильно выбирает путь решения задачи и использует подходящие алгоритмы, Программа работает корректно.
Аргументированно отвечает на вопросы	Знает изложенный материал, Проявляет критическое мышление
Представление лабораторной	Использует иллюстративные, наглядные материалы, Владеет культурой речи.

Критерии оценки

- «Отлично» – программа правильно решает поставленную задачу, лабораторная работа полностью соответствует описанным критериям;
- «Хорошо» – программа правильно решает поставленную задачу, лабораторная работа соответствует описанным критериям за исключением некоторых замечаний не более чем по нескольким пунктам критериев;
- «Удовлетворительно» – программа правильно решает поставленную задачу, лабораторная работа соответствует более чем половине описанных критериев;
- «Неудовлетворительно» – программа не предоставлена, или не верно решает поставленную задачу, лабораторная работа не соответствует большей части описанных критериев.

Вариант билета на зачете

1. Опишите метод «Решающие деревья», перечислите его достоинства и недостатки..
2. Что такое переобучение/недообучение? Как проводить анализ того переобучена или недообучена модель?
3. Опишите метод построения рекомендательной системы на основе SVD разложения.
4. При решении задачи бинарной классификации для 1000 тестовых примеров получены результаты, представленные в таблице ниже. Чему равны точность (Precision), полнота (Recall) и F-мера классификатора? Что показывают эти величины, когда какую лучше стоит использовать?

	Действительный класс	
	1	0
1	85	890
0	15	10

Критерии оценки

- «Отлично» – даны верные ответы на 4 вопроса из билета;
- «Хорошо» – даны верные ответы на 4 вопроса из билета, в некоторых вопросах были допущены существенные недочеты;
- «Удовлетворительно» – даны верные ответы на 3 вопроса из билета;
- «Неудовлетворительно» – даны верные ответы на 2 и менее вопросов из билета.

Вариант билета на экзамене

1. Опишите основные шаги детектора границ Саппу.
2. Какие подходы к композиции алгоритмов Вы знаете?
3. Опишите алгоритм SVM. Перечислите плюсы минусы метода, за что отвечают основные параметры алгоритма?
4. Понятие свертки, свойства, что может вычислять свертка. Понятие сверточного слоя, карта свертки, зачем используют? Сколько параметров содержит сверточный слой со сверткой 5 x 5 для изображения 128 x 128 x 3?

Практическая часть:

Задан набор данных в виде csv-файла. Напишите программу, которая обучает модель (алгоритм) задачу классификации. Оцените качество работы модели.

Набор данных содержит информацию о клиентах сотовой связи (тариф, число минут разговора в месяц, звонков в кол-центр оператора, регион...), необходимо предсказать уйдет ли клиент в течении месяца.

Критерии оценки

- «Отлично» – практическая часть полностью реализована, даны верные ответы на 4 вопроса из билета, возможно допущены небольшие огрехи;
- «Хорошо» – практическая часть реализована почти полностью, даны верные ответы на 4 вопроса из билета, в некоторых вопросах были допущены существенные недочеты;
- «Удовлетворительно» – практическая часть реализована частично, даны верные ответы на 3 вопроса из билета;
- «Неудовлетворительно» – практическая часть не реализована или даны верные ответы лишь на 2 и менее вопросов из билета.

2. Перечень компетенций, этапы их формирования, описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкалы оценивания

2.1 Шкала оценивания сформированности компетенций и ее описание

Оценивание уровня сформированности компетенций в процессе освоения дисциплины осуществляется по следующей трехуровневой шкале:

Пороговый уровень - предполагает отражение тех ожидаемых результатов, которые определяют минимальный набор знаний и (или) умений и (или) навыков, полученных студентом в результате освоения дисциплины. Пороговый уровень является обязательным уровнем для студента к моменту завершения им освоения данной дисциплины.

Продвинутый уровень - предполагает способность студента использовать знания, умения, навыки и (или) опыт деятельности, полученные при освоении дисциплины, для решения профессиональных задач. Продвинутый уровень превосходит пороговый уровень по нескольким существенным признакам.

Высокий уровень - предполагает способность студента использовать потенциал интегрированных знаний, умений, навыков и (или) опыта деятельности, полученных при освоении дисциплины, для творческого решения профессиональных задач и самостоятельного поиска новых подходов в их решении путем комбинирования и использования известных способов решения применительно к конкретным условиям. Высокий уровень превосходит пороговый уровень по всем существенным признакам.

3. Методические рекомендации преподавателю по процедуре оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Целью процедуры оценивания является определение степени овладения студентом ожидаемыми результатами обучения (знаниями, умениями, навыками и (или) опытом деятельности).

Процедура оценивания степени овладения студентом ожидаемыми результатами обучения осуществляется с помощью методических материалов, представленных в разделе «Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций»

3.1 Критерии оценивания степени овладения знаниями, умениями, навыками и (или) опытом деятельности, определяющие уровни сформированности компетенций

Пороговый уровень (общие характеристики):

- владение основным объемом знаний по программе дисциплины;
- знание основной терминологии данной области знаний, стилистически грамотное, логически правильное изложение ответа на вопросы без существенных ошибок;
- владение инструментарием дисциплины, умение его использовать в решении стандартных (типовых) задач;
- способность самостоятельно применять типовые решения в рамках рабочей программы дисциплины;
- усвоение основной литературы, рекомендованной рабочей программой дисциплины;
- знание базовых теорий, концепций и направлений по изучаемой дисциплине;
- самостоятельная работа на практических и лабораторных занятиях, периодическое участие в групповых обсуждениях, достаточный уровень культуры исполнения заданий.

Продвинутый уровень (общие характеристики):

- достаточно полные и систематизированные знания в объеме программы дисциплины;
- использование основной терминологии данной области знаний, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать выводы;
- владение инструментарием дисциплины, умение его использовать в решении учебных и профессиональных задач;
- способность самостоятельно решать сложные задачи (проблемы) в рамках рабочей программы дисциплины;
- усвоение основной и дополнительной литературы, рекомендованной рабочей программой дисциплины;
- умение ориентироваться в базовых теориях, концепциях и направлениях по изучаемой дисциплине и давать им сравнительную оценку;
- самостоятельная работа на практических и лабораторных занятиях, участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

Высокий уровень (общие характеристики):

- систематизированные, глубокие и полные знания по всем разделам дисциплины;
- точное использование терминологии данной области знаний, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;
- безупречное владение инструментарием дисциплины, умение его использовать в постановке и решении научных и профессиональных задач;

- способность самостоятельно и творчески решать сложные задачи (проблемы) в рамках рабочей программы дисциплины;
- полное и глубокое усвоение основной и дополнительной литературы, рекомендованной рабочей программой дисциплины;
- умение ориентироваться в основных теориях, концепциях и направлениях по изучаемой дисциплине и давать им критическую оценку;
- активная самостоятельная работа на практических и лабораторных занятиях, творческое участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

3.2 Описание процедуры выставления оценки

В зависимости от уровня сформированности каждой компетенции по окончании освоения дисциплины студенту выставляется оценка «зачтено» или «не зачтено».

Показатели и критерии, используемые при выставлении оценки подробно описаны в разделе «Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций».

Пороговый уровень формирования компетенций соответствует оценке «зачтено» за выполнение лабораторных работ в течение семестра и сдачу зачета в конце семестра.

Оценка «зачет» выставляется студенту, у которого каждая компетенция (полностью или частично формируемая данной дисциплиной) сформирована не ниже, чем на пороговом уровне.

Оценка «незачтено» выставляется студенту, у которого хотя бы одна компетенция (полностью или частично формируемая данной дисциплиной) сформирована ниже, чем на пороговом уровне.

Приложение №2 к рабочей программе дисциплины «Машинное обучение»

Методические указания для студентов по освоению дисциплины

Основной формой изложения учебного материала по дисциплине «Машинное обучение» являются лекции, проводимые в виде электронных презентаций, что позволяет сделать материал лекций более наглядными, улучшает информативность и понимание изучаемого курса.

По большинству тем предусмотрены лабораторные работы, на которых происходит закрепление лекционного материала путем применения его к конкретным задачам и получение навыков работы непосредственно с CASE- системой, путем выполнения заданий по построению моделей определенного типа. Выполнение предлагаемых в процессе изучения курса лабораторных работ позволяет не только понять и закрепить теоретический материал, но и приобрести навык анализа предметной области и построения объектно-ориентированных моделей с применением современных технологий.

Для успешного освоения дисциплины очень важно решение достаточно большого количества задач, как в аудитории, так и самостоятельно в качестве домашних заданий. Примеры решения задач разбираются и обсуждаются на лекциях и лабораторных занятиях. Основная цель решения задач – помочь усвоить фундаментальные понятия и основы концепции объектно-ориентированного анализа и проектирования. Для решения всех задач необходимо знать и понимать лекционный материал. Поэтому в процессе изучения дисциплины рекомендуется регулярное повторение пройденного лекционного материала. Материал, законспектированный на лекциях, необходимо дома еще раз прорабатывать и при необходимости дополнять информацией, полученной при выполнении лабораторных работ или из учебной литературы.

Большое внимание должно быть уделено выполнению домашней работы. В качестве заданий для самостоятельной работы дома студентам предлагаются задачи, аналогичные разобранным на лекциях и лабораторных занятиях или немного более сложные, которые являются результатом объединения нескольких базовых задач.

Для проверки и контроля усвоения теоретического материала, приобретенных практических навыков работы с CASE - системой, в течение обучения проводятся мероприятия текущей аттестации в виде двух контрольных работ. Для более успешной подготовки к контрольным работам перед их проведением преподавателем проводятся консультации.

В конце семестра студенты сдают зачет. Зачет по итогам семестра выставляется по результатам написания лабораторных и контрольных работ, а так же краткого собеседования по вопросам теоретического материала.

Освоить вопросы, излагаемые в процессе изучения дисциплины «Машинное обучение» самостоятельно студенту достаточно сложно. Это связано с отсутствием опыта в проведении анализа предметной области и построения объектно-ориентированных моделей. Поэтому посещение всех аудиторных занятий является совершенно необходимым. Без упорных и регулярных занятий в течение семестра сдать зачет по итогам изучения дисциплины студенту будет сложно.

Учебно-методическое обеспечение самостоятельной работы студентов по дисциплине

Для самостоятельной работы особенно рекомендуется использовать учебную литературу, указанную в разделе № 7 данной рабочей программы.

Также для подбора учебной литературы рекомендуется использовать широкий спектр интернет-ресурсов:

1. Электронно-библиотечная система «Университетская библиотека online»

(www.biblioclub.ru) - электронная библиотека, обеспечивающая доступ к наиболее востребованным материалам-первоисточникам, учебной, научной и художественной литературе ведущих издательств (*регистрация в электронной библиотеке – только в сети университета. После регистрации работа с системой возможна с любой точки доступа в Internet.).

2. Для самостоятельного подбора литературы в библиотеке ЯрГУ рекомендуется использовать:

1. Личный кабинет (http://lib.uniyar.ac.ru/opac/bk_login.php) дает возможность получения online доступа к списку выданной в автоматизированном режиме литературы, просмотра и копирования электронных версий изданий сотрудников университета (учеб. и метод. пособия, тексты лекций и т.д.) Для работы в «Личном кабинете» необходимо зайти на сайт Научной библиотеки ЯрГУ с любой точки, имеющей доступ в Internet, в пункт меню «Электронный каталог»; пройти процедуру авторизации, выбрав вкладку «Авторизация», и заполнить представленные поля информации.

2. Электронная библиотека учебных материалов ЯрГУ

(http://www.lib.uniyar.ac.ru/opac/bk_cat_find.php) содержит более 2500 полных текстов учебных и учебно-методических материалов по основным изучаемым дисциплинам, изданных в университете. Доступ в сети университета, либо по логину/паролю.

3. Электронная картотека «[Книгообеспеченность](#)»

(http://www.lib.uniyar.ac.ru/opac/bk_bookreq_find.php) раскрывает учебный фонд научной библиотеки ЯрГУ, предоставляет оперативную информацию о состоянии книгообеспеченности дисциплин основной и дополнительной литературой, а также цикла дисциплин и специальностей. Электронная картотека «[Книгообеспеченность](#)» доступна в сети университета и через Личный кабинет.