

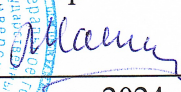
Министерство науки и высшего образования
Российской Федерации

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Донецкий государственный университет»

Физико-технический факультет
Кафедра компьютерных технологий



УТВЕРЖДАЮ
проректор

 П.А. Машаров
«29» марта 2024 г.

МП

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

МАШИННОЕ ОБУЧЕНИЕ

Укрупненная группа направлений
подготовки

Программа высшего образования

Направление подготовки

Профиль подготовки

Квалификация

Форма обучения

09.00.00 Информатика и вычислительная
техника

Программа магистратуры

09.04.01 Информатика и вычислительная
техника

Информатика и вычислительная техника

Технологии искусственного интеллекта

Магистр

Очная, заочная

Рабочая программа адаптирована для лиц
с ограниченными возможностями здоровья и инвалидов

Донецк 2024

Рабочая программа дисциплины «Машинное обучение» для обучающихся по направлению подготовки 09.04.01 Информатика и вычислительная техника, магистерских программ (Профиль подготовки: Информатика и вычислительная техника, Технологии искусственного интеллекта), составлена на основании Федерального государственного образовательного стандарта высшего образования – магистратура по направлению подготовки 09.04.01 Информатика и вычислительная техника, утвержденного приказом Министерства образования и науки Российской Федерации от 19 сентября 2017 г. № 918 (с изм. и доп.), Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 06 апреля 2021 г. № 245 (с изм. и доп.), в соответствии с учебным планом, утвержденным Ученым советом ФГБОУ ВО «ДонГУ» для набора 2024 года.

Разработчики:

доцент кафедры компьютерных технологий,
канд. техн. наук, доцент



Т.В. Ермоленко

Рабочая программа одобрена на заседании кафедры компьютерных технологий.
Протокол от 26.03.2024 г. № 12

Заведующий кафедрой



Г.В. Аверин

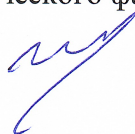
СОГЛАСОВАНО:

Декан физико-технического факультета
28.03.2024 г.



С.А. Фоменко

Учебно-методическая комиссия физико-технического факультета.
Протокол от 27.03.2024 г. № 2
Председатель



В. Н. Котенко

Руководитель основной профессиональной образовательной программы,
д-р технических наук, проф.
26.03.2024 г.



Г.В. Аверин

1. ОБЛАСТЬ ПРИМЕНЕНИЯ И МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ

Учебная дисциплина «Машинное обучение» относится к вариативной части блока «Дисциплины (модули)» и состоит из двух содержательных модулей: модуль 1 – «Методы классификации и регрессии», модуль 2 – «Ансамбли решающих правил. Критерии выбора моделей и методы отбора признаков».

Дисциплина реализуется на физико-техническом факультете кафедрой компьютерных технологий.

Этот курс опирается на математическую подготовку студентов, полученную при изучении дисциплин бакалавриата: «Математика», «Теория вероятности, математическая статистика», «Математическая логика», «Современные информационные системы и технологии», на знания технологий разработки современного программного обеспечения, полученные при изучении дисциплин бакалавриата: «Технологии разработки программного обеспечения» «Объектно-ориентированное программирование», а также на знания технологий искусственного интеллекта, полученные в результате освоения дисциплины магистратуры «Технологии извлечения знаний», закладывает фундамент научно-прикладной подготовки будущих исследователей в области инженерии знаний.

Полученные знания используются студентами при изучении дисциплины «Распознавание речи», а также во время выполнения научно-исследовательской работы и при написании магистерской диссертации.

2. СТРУКТУРА ДИСЦИПЛИНЫ

<i>Характеристика учебной дисциплины</i>		
Направление подготовки	09.04.01 Информатика и вычислительная техника	
Магистерская программа	Информатика и вычислительная техника	
Программа подготовки	академическая магистратура	
Квалификация	магистр	
Количество содержательных модулей	2	
Дисциплина базовой / вариативной части образовательной программы	дисциплина вариативной части	
Формы контроля	1 модульный контроль, 1 экзамен	
Показатели	очная форма обучения	заочная форма обучения
Количество зачетных единиц (кредитов)	4,5	4,5
Год подготовки	1	1
Семестр	1	2
Количество часов	162	162
- лекционных	18	4
- практических, семинарских		
- лабораторных	36	6
- самостоятельной работы	108	152
в т.ч. индивидуальное задание		
Недельное количество часов,	10	10
в т.ч. аудиторных	3	1,2

3. ОПИСАНИЕ ДИСЦИПЛИНЫ

Цели и задачи

Цель – формирование у магистрантов знаний основных математических методов и алгоритмов машинного обучения, развить интуицию для лучшего понимания основных идей, лежащих за этими методами, привить навыки работы с программным обеспечением, реализующим алгоритмы машинного обучения, и умение применять полученные знания при проектировании и реализации интеллектуальных информационных систем.

Задачи – освоение логических, метрических и вероятностных методов классификации и регрессии, алгоритмов построения разделяющих гиперплоскостей и нейросетей, критериев выбора моделей и методов отбора признаков, методов построения композиций классификаторов.

Требования к результатам освоения дисциплины:

Процесс изучения дисциплины «Машинное обучение» направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО РФ, ГОС ВО ДНР (проект) по направлению подготовки 09.04.01 Информатика и вычислительная техника и основной профессиональной образовательной программы высшего образования направления подготовки 09.04.01 Информатика и вычислительная техника:

Общепрофессиональные компетенции (ОПК):	
ОПК-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач.

Индикаторы достижения компетенций и результаты обучения. Достижение компетенций оценивается на основе следующих индикаторов и соответствующих им результатов обучения:

Общепрофессиональные компетенции	Индикаторы	Результаты обучения
ОПК-2. Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач.	ОПК-2.1. Знать: современные информационно-коммуникационные и интеллектуальные технологии, инструментальные среды, программно-технические платформы для решения профессиональных задач	Знает возможности алгоритмов машинного обучения и способы их применения для решения практических задач управления и распознавания образов.
		Знает логические, метрические и вероятностные методы классификации и регрессии.
		Знает основные понятия, связанные с нейросетевыми технологиями к построению интеллектуальных систем.
		Знает композиции классификаторов, методы бустинга.
		Знает критерии выбора моделей и методы отбора признаков.
	ОПК-2.2. Уметь: обосновывать выбор современных информационно-коммуникационных и интеллектуальных технологий, разрабатывать оригинальные	Умеет выполнять грамотную постановку задач, возникающих в практической деятельности, для их решения с помощью методов машинного обучения.
		Умеет обосновать применение того или иного алгоритма машинного обучения для решения конкретной задачи.
		Умеет проводить анализ работы методов машинного обучения с выявлением их сильных и слабых сторон.

	программные средства для решения профессиональных задач	Умеет анализировать результаты обучения алгоритма машинного обучения, предлагать пути повышения его точности.
	ОПК-2.3. Владеть: навыками разработки оригинальных программных средств, в том числе с использованием современных информационно-коммуникационных и интеллектуальных технологий, для решения профессиональных задач	Владеет навыками работы с инструментальными средствами и технологиями работы машинного обучения.
		Владеет навыками построения интеллектуальных систем на основе нейросетей.

В результате изучения учебной дисциплины студент должен

Знать:

- постановку основных задач машинного обучения;
- возможности алгоритмов машинного обучения и способы их применения для решения практических задач управления и распознавания образов;
- логические, метрические и вероятностные методы классификации и регрессии;
- основные понятия и принципы работы искусственных нейронных сетей;
- композиции классификаторов, методы бустинга;
- критерии выбора моделей и методы отбора признаков;

Уметь:

- выполнять грамотную постановку задач, возникающих в практической деятельности, для их решения с помощью методов машинного обучения;
- обосновать применение того или иного алгоритма машинного обучения для решения конкретной задачи;
- проводить анализ работы методов машинного обучения с выявлением их сильных и слабых сторон;
- анализировать результаты обучения алгоритма машинного обучения, предлагать пути повышения его точности;
- программно реализовывать алгоритмы машинного обучения;

Владеть:

- базовым инструментарием машинного обучения;
- навыками построения различного рода классификаторов.

4. Содержание дисциплины и формы организации учебного процесса

Порядковый номер и тема	Краткое содержание темы
1	2
<p align="center">Содержательный модуль 1. Методы классификации и регрессии</p>	

<p>Тема 1. Основные понятия и примеры прикладных задач.</p>	<p>Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал. Типы задач. Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль.</p>
<p>1</p>	<p>2</p>
<p>Тема 2. Методы классификации и регрессии.</p>	<p>Метрические и логические методы классификации и регрессии: метод эталонных образов, метод ближайших соседей kNN и его обобщения; подбор числа k по критерию скользящего контроля; логическая закономерность; параметрические семейства закономерностей; переборные алгоритмы синтеза конъюнкций; решающее дерево; вывод критериев ветвления; энтропийный критерий, критерий Джини. Линейный классификатор: разделяющая поверхность; метод решающих функций и опорных векторов; линейные модели регрессии и классификации; метод наименьших квадратов; полиномиальная регрессия. Байесовская теория классификации: наивный байесовский классификатор; непараметрические и параметрические методы оценивания плотности; смеси. Нейронные сети: модель биологического нейрона; функции активации; алгоритм обратного распространения ошибок; свёрточные сети, рекуррентные сети.</p>

1	2
<p>Тема 3. Прогнозирование временных рядов.</p>	<p>Задача прогнозирования временных рядов. Примеры приложений. Экспоненциальное скользящее среднее. Адаптивная авторегрессионная модель. Адаптивная селективная модель. Адаптивная композиция моделей.</p>
<p align="center">Содержательный модуль 2. Ансамбли решающих правил. Критерии выбора моделей и методы отбора признаков</p>	
<p>Тема 4. Линейные композиции, бустинг.</p>	<p>Основные понятия: базовый алгоритм (алгоритмический оператор), корректирующая операция. Взвешенное голосование. Алгоритм AdaBoost. Экспоненциальная аппроксимация пороговой функции потерь. Процесс последовательного обучения базовых алгоритмов. Теорема о сходимости бустинга. Базовые алгоритмы в бустинге. Алгоритм AnyBoost. Алгоритм ComBoost.</p>
<p>Тема 5. Критерии выбора моделей и методы отбора признаков.</p>	<p>Внутренние и внешние критерии. Эмпирические, статистические и аналитические критерии. Сложность задачи отбора признаков. Полный перебор. Метод добавления и удаления, шаговая регрессия. Поиск в глубину, метод ветвей и границ. Анализ главных компонент и факторный анализ.</p>

Тематический план

Содержательный модуль 1												
Названия содержательных модулей и тем	Количество часов											
	Очная форма обучения						Заочная форма обучения					
	всего	в т.ч.					всего	в т.ч.				
		лекции	практические	лабораторные	самостоятельная работа	индивидуальная работа		лекции	практические	лабораторные	самостоятельная работа	индивидуальная работа
Тема 1. Основные понятия и примеры прикладных задач.	4	2			2		4				4	
Тема 2. Методы классификации и регрессии.	40	4		12	24		40	1		2	37	
Тема 3. Прогнозирование временных рядов.	20	2		8	10		20			2	18	
Итого по содержательному модулю 1	64	8		20	36		64	1		4	59	
Содержательный модуль 2												
Тема 4. Линейные композиции, бустинг.	32	4			28		32	1			31	
Тема 5. Критерии выбора моделей и методы отбора признаков.	30	2		8	20		30			2	28	
Итого по содержательному модулю 2	62	6		8	48		62	1		2	59	
Всего часов	126	14		28	84		126	2		6	118	

Курс дисциплины «Машинное обучение» предусматривает следующие **формы организации учебного процесса**:

- лекции;
- практические занятия;
- лабораторные занятия;
- самостоятельная работа студента.

По источнику передачи и восприятия учебной информации используются словесные (лекция, беседа), наглядные (слайды, иллюстрации, коды программ), практические (исследования, упражнения, лабораторные работы) методы.

По характеру познавательной деятельности студентов используются объяснительно-иллюстративные и репродуктивные методы, проблемное преподавание, частично-поисковый и исследовательский методы.

В зависимости от основной дидактической цели и задач используются методы устного изложения знаний, закрепление учебного материала, самостоятельной работы студентов по осмыслению и усвоению нового материала, работы по применению знаний на практике и выработке умений и навыков, проверки и оценки знаний, умений и навыков.

Используются следующие методы контроля:

- устный контроль (экспресс-опрос на лекциях);
- защита лабораторных работ;
- модульная контрольная работа.
- итоговый контроль (экзаменационные билеты).

5. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПРОВЕДЕНИЯ ЛЕКЦИОННЫХ, ПРАКТИЧЕСКИХ И ЛАБОРАТОРНЫХ ЗАНЯТИЙ

Темы лекционных занятий

<i>№ п/п</i>	<i>Название темы</i>	<i>Количество часов</i>
1	Основные понятия и примеры прикладных задач.	2
2	Метрические и логические методы классификации и регрессии.	1
3	Линейные и вероятностные методы классификации и регрессии.	1
4	Искусственные нейронные сети.	2
5	Прогнозирование временных рядов.	2
6	Алгоритмы Adaboost, Anyboost.	2
7	Голосование по большинству и старшинству.	2
8	Отбор информативных признаков.	1
9	Синтез информативных признаков.	1
	ВСЕГО	14

Темы лабораторных занятий

<i>№ п/п</i>	<i>Название темы</i>	<i>Количество часов</i>
1	Построение модели парной и множественной линейной регрессии средствами языка R.	8
2	Построение наивного байесовского классификатора.	4
3	Анализ временных рядов средствами языка R.	8
4	Анализ данных с помощью метода главных компонент средствами языка R.	8
	ВСЕГО	28

6. Методические рекомендации по организации самостоятельной работы студентов

Самостоятельная работа студентов по курсу «Машинное обучение» предусматривает:

- систематическое посещение лекционных занятий, ведение конспекта лекций;
- повседневное изучение лекционного материала и содержания технической литературы, рекомендуемые этой программой и рабочим учебным планом;
- добросовестную подготовку к лабораторным и практическим занятиям;
- своевременное и качественное оформление отчётов по лабораторным работам.
- самостоятельную разработку алгоритмов и текстов программ лабораторных работ.

Организация самостоятельной работы студентов

№ п/п	Название темы	Количество часов
1	2	3
1	Изучение темы: Обучение по прецедентам. Обучение без учителя. Примеры практических задач	2
2	Изучение темы: Метрические методы классификации и регрессии. В том числе вопросов, не освещаемых на лекциях: – метод окна Парзена с постоянной и переменной шириной окна; – непараметрическая регрессия. Локально взвешенный метод наименьших квадратов. Ядерное сглаживание; – робастная непараметрическая регрессия. Алгоритм LOWESS.	4
3	Изучение темы: Логические методы классификации и регрессии.	4
4	Подготовка к лабораторной работе №1: Изучение возможностей языка R для построения моделей парной и множественной линейной регрессии и проведение их анализа.	4
5	Изучение темы: Вероятностные методы классификации и регрессии. В том числе вопросов, не освещаемых на лекциях: – непараметрическое оценивание плотности: ядерная оценка плотности Парзена-Розенблатта. – параметрическое оценивание плотности: нормальный дискриминантный анализ; – выбор числа компонентов смеси. Пошаговая стратегия. Априорное распределение Дирихле; – смесь многомерных нормальных распределений. Сеть радиальных базисных функций (RBF) и применение EM-алгоритма для её настройки; – сравнение RBF-сети и SVM с гауссовским ядром.	6
6	Подготовка к лабораторной работе №2: Изучение байесовского подхода к классификации.	4
7	Изучение темы: Нейронные сети. В том числе вопросов, не освещаемых на лекциях: – быстрые методы стохастического градиента; – методы постепенного усложнения сети, оптимальное прореживание нейронных сетей; – функции активации ReLU и PReLU. Проблема «паралича» сети. – обучение рекуррентных сетей: Backpropagation Through Time (BPTT).	6
8	Изучение темы: Прогнозирование временных рядов. В том числе вопросов, не освещаемых на лекциях: – прогнозирование временных рядов с помощью рекуррентной сети LSTM	4
9	Подготовка к лабораторной работе №3: Изучение возможностей языка R для прогнозирования временных рядов.	2
10	Изучение темы: Ансамбли моделей классификаторов. В том числе вопросов, не освещаемых на лекциях: – бэггинг и метод случайных подпространств; – случайный лес; – градиентный бустинг; – алгоритм LogitBoost; – смесь алгоритмов.	14

1	2	3
11	Изучение темы: Методы голосования. В том числе вопросов, не освещаемых на лекциях: – простое голосование: алгоритм ComBoost.	14
12	Изучение темы: Критерии выбора модели. В том числе вопросов, не освещаемых на лекциях: – критерий непротиворечивости моделей; – критерий регуляризации.	8
13	Изучение темы: Отбор информативных признаков. В том числе вопросов, не освещаемых на лекциях: – отбор признаков методом стохастического поиска.	8
114	Подготовка к лабораторной работе №4: Изучение метода анализа главных компонент. Изучение возможностей языка R для проведения анализа данных с помощью анализа главных компонент.	4
	ВСЕГО	84

7. ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ

Индивидуальные задания не предусмотрены.

8. КОНТРОЛЬНЫЕ ВОПРОСЫ К ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

1. Приведите функциональную и вероятностную постановку задачи обучения по прецедентам. Какие типы признаков вы знаете?
2. Что такое модель алгоритмов и метод обучения? Что такое функция потерь и функционал качества? Назовите наиболее употребительные функции потерь.
3. Чем характеризуется обобщающая способность метода? Какие эмпирические оценки обобщающей способности вы знаете?
4. Изложите суть метода ближайшего соседа. Какими недостатками он обладает?
5. Приведите алгоритм k ближайших соседей и его модификацию – k взвешенных ближайших соседей. Какие недостатки простейших метрических алгоритмов типа kNN?
6. Что такое отступ объекта относительно алгоритма? На какие типы делятся объекты в порядке убывания отступа?
7. В чем заключается алгоритм STOLP для отбора эталонных объектов?
8. Каковы критерии информативности, позволяющие называть предикаты закономерностями? Приведите эвристическое, статистическое, энтропийное определения информативности.
9. В чем заключается бинаризация количественных признаков? В чем заключается разбиение диапазона значений признака на информативные зоны?
10. Какие вы знаете алгоритмы синтеза конъюнкций?
11. Что такое решающие списки? В чем заключается жадный алгоритм их построения?
12. Какие разновидности решающих списков? В чем заключаются их достоинства?
13. Что такое решающие деревья? В чем преимущества метода деревьев решений?
14. Из каких этапов состоит синтез решающих деревьев?
15. Какие критерии расщепления вы знаете?
16. Назовите основные этапы алгоритма построения решающего дерева ID3? В чем его преимущества и недостатки?
17. В чем суть алгоритма CART?
18. В чем суть редукции решающих деревьев?

19. Что такое решающая функция и дискриминантная поверхность? Каким свойством обладает гиперплоскость, построенная методом SVM? Зачем в SVM используются ядра?

20. Что такое регрессия? Приведите общий вид регрессионной модели. Приведите модель линейной регрессии. Какие критерии оценки качества регрессионной модели вы знаете?

21. Могут ли зависимости в регрессионной модели быть нелинейными? В чем заключается общий подход при нахождении оптимальных оценок параметров регрессионной модели?

22. Изложите общую идею метода потенциальных функций.

23. Каковы основные этапы алгоритма метода потенциальных функций?

24. В чем суть Байесовского подхода к распознаванию образов?

25. Изложите идею параметрических методов оценивания плотности вероятностей.

26. Изложите идею непараметрических методов оценивания плотности вероятностей.

Что такое смеси?

27. Опишите математическую модель нейрона. По каким формулам определяется функционирование нейрона? Каковы ограничения модели нейрона? Опишите модель многослойного персептрона. Что такое глубокая нейросеть?

28. Что такое нейронная сеть Хопфилда? Принципы ее работы и ограничения.

29. Приведите алгоритм обратного распространения ошибки.

30. Опишите общую архитектуру сверточных нейросетей. Что такое двумерная свёртка? Зачем используют слои пулинга в сверточных нейронных сетях?

31. Какие техники свёртки применяются в сверточных нейронных сетях? Объясните разницу в терминах «фильтр» и «ядро».

32. Что такое рекуррентные сети? Опишите их общую архитектуру. В каких задачах они используются? Приведите схему однослойной рекуррентной сети. Что отличало рекуррентную сеть Элмана от предыдущих концепций рекуррентных сетей?

33. Каковы особенности архитектуры LSTM-сети? В чем преимущество LSTM-сети? Что понимается под обучением нейросети? Назовите основные правила обучения нейросетей.

34. Приведите алгоритм обучения методами Хебба.

35. В чем заключается принцип работы сети Кохонена.

36. Что такое временной ряд? Каковы его составляющие? Модель временного ряда? Назовите типы моделей временного ряда.

37. В каком случае применяют трансформацию данных ряда? Какие виды трансформации данных вам известны?

38. Что такое стационарный временной ряд и коэффициент автокорреляции? Что является его статистической оценкой?

39. Что такое коррелограмма? Какие выводы можно сделать после ее анализа?

40. Что такое частичная автокорреляция? Какие тесты вы знаете для определения стационарности временного ряда?

41. Что такое единичный корень и как проверить его отсутствие?

42. Какие наиболее часто используемые на практике критерии проверки «наличия-отсутствия» тренда?

43. С помощью каких критериев оценивают информативность модели ряда? Из каких этапов состоит алгоритм анализа временного ряда?

44. Приведите модель AR(p). Как определяется ее порядок?

45. Приведите модель MA(p). Как определяется ее порядок.

46. Приведите модель ARIMA. Как определяются ее порядки?

9. ОБРАЗЕЦ МОДУЛЬНОГО КОНТРОЛЯ

ГОУ ВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ»

Физико-технический факультет

Направление подготовки: **09.04.01 Информатика и вычислительная техника**
 Магистерская программа: **Информатика и вычислительная техника**
 Программа подготовки: **академическая магистратура**
 Семестр: **3**
 Учебная дисциплина: **Машинное обучение**

МОДУЛЬНАЯ КОНТРОЛЬНАЯ РАБОТА

ВАРИАНТ №1

1. Приведите функциональную и вероятностную постановку задачи обучения по прецедентам. Какие типы признаков вы знаете?
2. Что такое нейронная сеть Хопфилда? Принципы ее работы и ограничения.
3. В каком случае применяют трансформацию данных ряда? Какие виды трансформации данных вам известны?

Утверждено на заседании кафедрой компьютерных технологий,
 протокол № ____ от “__” _____ 20__ г.

Зав. кафедрой _____

Преподаватель _____

Критерии оценивания модульного контроля

<i>Номер задания</i>	<i>Количество баллов</i>
Задание 1	2
Задание 2	4
Задание 3	4
Всего	10

10. ОБРАЗЕЦ ЭКЗАМЕНАЦИОННОГО БИЛЕТА

Теоретические вопросы к экзамену

1. Приведите функциональную и вероятностную постановку задачи обучения по прецедентам. Какие типы признаков вы знаете?
2. Что такое модель алгоритмов и метод обучения? Что такое функция потерь и функционал качества? Назовите наиболее употребительные функции потерь.
3. Чем характеризуется обобщающая способность метода? Какие эмпирические оценки обобщающей способности вы знаете?
4. Изложите суть метода ближайшего соседа. Какими недостатками он обладает?
5. Приведите алгоритм k ближайших соседей и его модификацию – k взвешенных ближайших соседей. Какие недостатки простейших метрических алгоритмов типа kNN?
6. Что такое отступ объекта относительно алгоритма? На какие типы делятся объекты в порядке убывания отступа?
7. В чем заключается алгоритм STOLP для отбора эталонных объектов?

8. Каковы критерии информативности, позволяющие называть предикаты закономерностями? Приведите эвристическое, статистическое, энтропийное определения информативности.

9. В чем заключается бинаризация количественных признаков? В чем заключается разбиение диапазона значений признака на информативные зоны?

10. Какие вы знаете алгоритмы синтеза конъюнкций?

11. Что такое решающие списки? В чем заключается жадный алгоритм их построения?

12. Какие разновидности решающих списков? В чем заключаются их достоинства?

13. Что такое решающие деревья? В чем преимущества метода деревьев решений?

14. Из каких этапов состоит синтез решающих деревьев?

15. Какие критерии расщепления вы знаете?

16. Назовите основные этапы алгоритма построения решающего дерева ID3? В чем его преимущества и недостатки?

17. В чем суть алгоритма CART?

18. В чем суть редукции решающих деревьев?

19. Что такое решающая функция и дискриминантная поверхность? Каким свойством обладает гиперплоскость, построенная методом SVM? Зачем в SVM используются ядра?

20. Что такое регрессия? Приведите общий вид регрессионной модели. Приведите модель линейной регрессии. Какие критерии оценки качества регрессионной модели вы знаете?

21. Могут ли зависимости в регрессионной модели быть нелинейными? В чем заключается общий подход при нахождении оптимальных оценок параметров регрессионной модели?

22. Изложите общую идею метода потенциальных функций.

23. Каковы основные этапы алгоритма метода потенциальных функций?

24. В чем суть Байесовского подхода к распознаванию образов?

25. Изложите идею параметрических методов оценивания плотности вероятностей.

26. Изложите идею непараметрических методов оценивания плотности вероятностей. Что такое смеси?

27. Опишите математическую модель нейрона. По каким формулам определяется функционирование нейрона? Каковы ограничения модели нейрона? Опишите модель многослойного персептрона. Что такое глубокая нейросеть?

28. Что такое нейронная сеть Хопфилда? Принципы ее работы и ограничения.

29. Приведите алгоритм обратного распространения ошибки.

30. Опишите общую архитектуру сверточных нейросетей. Что такое двумерная свёртка? Зачем используют слои пулинга в сверточных нейронных сетях?

31. Какие техники свёртки применяются в сверточных нейронных сетях? Объясните разницу в терминах «фильтр» и «ядро».

32. Что такое рекуррентные сети? Опишите их общую архитектуру. В каких задачах они используются? Приведите схему однослойной рекуррентной сети. Что отличало рекуррентную сеть Элмана от предыдущих концепций рекуррентных сетей?

33. Каковы особенности архитектуры LSTM-сети? В чем преимущество LSTM-сети? Что понимается под обучением нейросети? Назовите основные правила обучения нейросетей.

34. Приведите алгоритм обучения методами Хебба.

35. В чем заключается принцип работы сети Кохонена.

36. Что такое временной ряд? Каковы его составляющие? Модель временного ряда? Назовите типы моделей временного ряда.

37. В каком случае применяют трансформацию данных ряда? Какие виды трансформации данных вам известны?
38. Что такое стационарный временной ряд и коэффициент автокорреляции? Что является его статистической оценкой?
39. Что такое коррелограмма? Какие выводы можно сделать после ее анализа?
40. Что такое частичная автокорреляция? Какие тесты вы знаете для определения стационарности временного ряда?
41. Что такое единичный корень и как проверить его отсутствие?
42. Какие наиболее часто используемые на практике критерии проверки «наличия-отсутствия» тренда?
43. С помощью каких критериев оценивают информативность модели ряда? Из каких этапов состоит алгоритм анализа временного ряда?
44. Приведите модель $AR(p)$. Как определяется ее порядок?
45. Приведите модель $MA(p)$. Как определяется ее порядок.
46. Приведите модель $ARIMA$. Как определяются ее порядки?
47. Что называется композицией алгоритмов? В чем суть взвешенного голосования?
48. Для чего используется процедура *bootstrap*? Приведите алгоритм этой процедуры.
49. Для чего используется процедура *bagging*? Приведите алгоритм этой процедуры.
50. Приведите классический алгоритм Шапире.
51. Каковы шаги алгоритма *AdaBoost*? Назовите достоинства и недостатки *AdaBoost*.
52. Как можно адаптировать градиентный спуск к задачам бустинга?
53. Приведите алгоритм построения композиции для голосования по большинству. Изложите идею преобразования простого голосования во взвешенное.
54. Приведите алгоритм построения композиции для голосования по старшинству. Какие стратегии выбора параметра алгоритма голосования по старшинству?
55. Какие стратегии выбора класса в алгоритме голосования по старшинству?
56. Какие типы критериев задачах отбора признаков и моделей? Приведите критерий средней ошибки на контрольных данных. Приведите критерий скользящего контроля.
57. Что такое *bootstrap*? Приведите алгоритм полного перебора всевозможных наборов признаков.
58. Приведите алгоритм *Add*. Каковы его достоинства и недостатки?
59. Приведите алгоритм *Add-Del*. Каковы его достоинства и недостатки?
60. Каковы стратегии обхода дерева возможных наборов признаков?
61. Назовите критерии выбора модели, основанные на оценках обобщающей способности.
62. Что строит алгоритм *PCA*, выполняя обработку всей матрицы наблюдений? Зачем используют *PCA*? Что такое ковариационная матрица? Что из себя представляют главные компоненты? Являются ли они коррелированными?
63. Что такое матрица счетов? Что такое матрица нагрузок?
64. Из каких соображений выбирают количество главных компонент, которым стоит ограничиться? Оптимум каких критериев позволяет достичь метод *PCA*?
65. В чем суть критерия *broken stick model*?

Образец экзаменационного билета

ГОУ ВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ»

Физико-технический факультет

Направление подготовки: **09.04.01 Информатика и вычислительная техника**
 Магистерская программа: **Информатика и вычислительная техника**
 Программа подготовки: **академическая магистратура**
 Семестр: **3**
 Учебная дисциплина: **Машинное обучение**

БИЛЕТ №1

1. Что называется композицией алгоритмов? В чем суть взвешенного голосования?
2. Приведите функциональную и вероятностную постановку задачи обучения по прецедентам. Какие типы признаков вы знаете?
3. Рассмотрим задачу классификации с двумя классами 0 и 1. Пусть пространство признаков двумерное. Объекты каждого класса имеют нормальное распределение с математическим ожиданием (0, 0.5) и (0.4, 1) соответственно и матрицей ковариации

$$\Sigma = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$$

Априорные вероятности классов равны $P\{Y=0\}=1/3$ и $P\{Y=1\}=2/3$

Найти: уравнение разделяющей поверхности байесова классификатора; собственное разложение матрицы Σ .

Перейти к новым координатам, оси которых совпадают с собственными векторами матрицы Σ . Выписать уравнение разделяющей поверхности байесова классификатора в новых координатах.

Утверждено на заседании кафедрой компьютерных технологий,
 протокол № ____ от “__” _____ 20__ г.

Зав. кафедрой _____

Преподаватель _____

Критерии оценивания экзамена

Номер задания	Количество баллов
Задание 1	10
Задание 2	10
Задание 3	30
Всего	50 баллов

11. КРИТЕРИИ ОЦЕНИВАНИЯ

Распределение баллов, которые могут получить студенты в процессе изучения дисциплины

Форма контроля	Максимальное количество баллов
Лабораторная работа №1	20
Лабораторная работа №2	20
Лабораторная работа №3	20
Лабораторная работа №4	20
Модульный контроль	10
Организационно-учебная работа студента	10

Согласно модульному принципу организации учебного процесса, содержание дисциплины «Машинное обучение» включает в себя два содержательных модуля. Каждый содержательный модуль состоит из теоретического материала и лабораторных работ, выполнение которых требует овладения теорией в указанном в модуле объеме.

К модульному контролю студент должен защитить 3 лабораторные работы, каждая из которых оценивается в 20 баллов.

На модульном контроле студент имеет возможность получить 10 баллов, ответив на 2 теоретических вопроса, каждый из которых оценивается в 5 баллов.

К концу семестра студент должен защитить еще 1 лабораторную работу, которая оценивается в 20 баллов.

Дополнительно 10 баллов студент может получить в течение семестра, посещая лекции и участвуя в устных опросах.

Экзамен сдают студенты с целью повышения рейтинга. На экзамене студент может получить 50 баллов, ответив на два теоретических вопроса, каждый из которых оценивается в 10 баллов, и решив задачу, оцениваемую в 30 баллов.

Шкала соответствия баллов национальной шкале

Оценка по шкале ECTS	Оценка по 100-балльной шкале	Оценка по государственной шкале (экзамен, дифференцированный зачет)	Оценка по государственной шкале (зачет)
A	90-100	5 (отлично)	зачтено
B	80-89	4 (хорошо)	зачтено
C	75-79	4 (хорошо)	зачтено
D	70-74	3 (удовлетворительно)	зачтено
E	60-69	3 (удовлетворительно)	зачтено
FX	35-59	2 (неудовлетворительно) с возможностью повторной сдачи	не зачтено
F	0-34	2 (неудовлетворительно) с возможностью повторной сдачи при условии обязательного набора дополнительных баллов	не зачтено

Оценка за овладение курса выставляется по следующим критериям:

– Оценку «отлично» заслуживает студент, который обнаружил глубокие знания при ответах на теоретические вопросы по темам курса, а также выполнил лабораторные работы в полном объеме и набрал более 90 баллов.

– Оценку «хорошо» заслуживает студент, сдавший все лабораторные работы и сделавший ошибки в теоретических или практических ответах при их защите, которые могут быть интерпретированы как малосущественные для вопросов, которые рассматривались. Студент должен набрать более 75 баллов.

– Оценку «удовлетворительно» заслуживает студент, который сдал лабораторные работы, отвечающий на вопросы неполно и с ошибками, но при этом набрал более 60 баллов.

– Оценку «неудовлетворительно» заслуживает студент, который не выполнил большинства лабораторных работ и набрал менее 60 баллов.

12. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОГО ПРОЦЕССА

Лекционные занятия проводятся в аудитории, оснащенной мультимедийной техникой и доской. Лабораторные занятия проводятся в компьютерном классе, оборудованном компьютерами с лицензионным программным обеспечением, доступом к сети Интернет, столами, доской.

13. РЕКОМЕНДОВАННАЯ ЛИТЕРАТУРА

№ п/п	Наименование	Кол-во экземпляров в библиотеке ДонНУ	Наличие электронной версии в ЭБС
<i>Основная литература</i>			
1.	Ермоленко Т.В. Методы машинного обучения [Электронный ресурс]: учебное пособие / Т.В. Ермоленко – Донецк: ДонНУ, 2017. – Электронные данные (1 файл).	-	+
2.	Ермоленко Т.В. Введение в машинное обучение [Электронный ресурс]: учебное-методическое пособие / Т.В. Ермоленко – Донецк: ДонНУ, 2017. – Электронные данные (1 файл).	-	+
<i>Дополнительная литература</i>			
3.	Коэлья Л.П. Построение систем машинного обучения на языке Python [Электронная книга] / Л.П. Коэлья, В. Ричарт. – М.: Издательство «ДМК Пресс», 2016. – 302 с. URL: http://mirknig.su/knigi/programming/10054-postroenie-sistem-mashinnogo-obucheniya-na-yazyke-python-2-e-izdanie.html (в свободном доступе)	-	-
4.	Шитиков В.К., Мاستицкий С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R [Электронный ресурс]. / В.К. Шитиков, С.Э. Мاستицкий – 2017. – 351 с. – Электронные данные, URL: https://github.com/ranalytics/data-mining (в свободном доступе)	-	-

14. ИНФОРМАЦИОННЫЕ РЕСУРСЫ

1. К.В. Воронцов. Машинное обучение (курс лекций). URL: <http://www.machinelearning.ru/wiki/index.php?title=ML> (дата обращения 03.01.2017).
2. Портал статей по применению ИТ и машинному обучению - http://habrahabr.ru/hub/machine_learning/ (дата обращения 03.01.2017).
3. Портал искусственного интеллекта URL: <http://www.aiportal.ru/downloads> (дата обращения 03.01.2017).

15. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

1. Язык R – свободная программная среда для статистических вычислений и графики (open source, лицензия GNU GPL 2).
2. RStudio – интегрированная среда разработки (IDE) для R (open source, лицензия GNU GPL 2).

Рабочая программа рассмотрена и переутверждена на заседании кафедры компьютерных технологий с изменениями (без изменений) на 2020 год.

Протокол № 12 от «2» апреля 2020 г.

Заведующий кафедрой

Ермоленко Т.В.

Рабочая программа рассмотрена и переутверждена на заседании кафедры компьютерных технологий с изменениями (без изменений) на 2021 год.

Протокол № ____ от «____» _____ 2021 г.

Заведующий кафедрой

Рабочая программа рассмотрена и переутверждена на заседании кафедры компьютерных технологий с изменениями (без изменений) на 2022 год.

Протокол № ____ от «____» _____ 2022 г.

Заведующий кафедрой

Рабочая программа рассмотрена и переутверждена на заседании кафедры компьютерных технологий с изменениями (без изменений) на 2023 год.

Протокол № ____ от «____» _____ 2023 г.

Заведующий кафедрой