

Подписано электронной подписью:  
Вержицкий Данил Григорьевич  
Должность: Директор КГПИ КемГУ  
Дата и время: 2025-09-24 00:00:00  
471086fad29a3b30e244c728abc3661ab35c9d50210dcf0e75e03a5b6fdf6436

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение высшего образования  
«КЕМЕРОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
Кузбасский гуманитарно-педагогический институт  
Факультет физической культуры, естествознания и природопользования

«УТВЕРЖДАЮ»  
Декан  
В. А. Рябов  
«23» января 2025г

### **Рабочая программа дисциплины**

#### **К.М.08.07 Машинное обучение в здравоохранении**

Специальность  
30.05.03 Медицинская кибернетика

Направленность (профиль)  
«Медицинские информационные системы»

Программа специалитета

Квалификация выпускника  
Врач-кибернетик

Форма обучения  
Очная

Год набора 2026

Новокузнецк 2025

## Оглавление

1 Цель дисциплины. ....	4
1.1 Формируемые компетенции, индикаторы достижения компетенций, знания, умения, навыки.....	4
1.2 Место дисциплины .....	4
2 Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации. ....	4
3. Учебно-тематический план и содержание дисциплины. ....	5
3.1 Учебно-тематический план .....	5
4 Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации.....	6
5 Материально-техническое, программное и учебно-методическое обеспечение дисциплины. ....	6
5.1 Учебная литература.....	6
5.2 Материально-техническое и программное обеспечение дисциплины.....	7
5.3 Современные профессиональные базы данных и информационные справочные системы. ....	8
6 Иные сведения и (или) материалы. ....	9
6.1. Темы письменных учебных работ .....	9
6.2. Примерные вопросы и задания / задачи для промежуточной аттестации .....	9

## **Лист внесения изменений в РПД**

### **Сведения об утверждении:**

РПД утверждена Учёным советом факультета физической культуры, естествознания и природопользования  
протокол Учёного совета факультета № 7 от 23.01.2025 г.

Одобрена на заседании методической комиссии факультета физической культуры, естествознания и природопользования  
протокол методической комиссии факультета № 4 от 23.01.2025г.

Рассмотрена на заседании обеспечивающей кафедры математики, физики и математического моделирования  
протокол №5 от 18.12.2024 г. Зав. кафедрой Решетникова Е.В.

## 1 Цель дисциплины.

В результате освоения данной дисциплины у обучающегося должны быть сформированы компетенции основной профессиональной образовательной программы специалитета (далее - ОПОП): ПК-3, ПК-4.

### 1.1 Формируемые компетенции, индикаторы достижения компетенций, знания, умения, навыки

Таблица 1 – Индикаторы достижения компетенций, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции по ОПОП	Знания, умения, навыки (ЗУВ), формируемые дисциплиной
ПК-3 Способен решать системно-аналитические задачи в области здравоохранения	ПК 3.1. Планирует, проводит и обрабатывает результаты медико-биологических исследований ПК 3.2. Способен использовать статистические методы и методы прикладной математики для обработки клинических и экспериментальных данных при решении задач вычислительной диагностики и прогнозирования в здравоохранении	<b>Знать:</b> - необходимые методы машинного обучения для решения задач вычислительной диагностики и прогнозирования в здравоохранении <b>Уметь:</b> - использовать методы машинного обучения для обработки клинических и экспериментальных данных при решении задач вычислительной диагностики и прогнозирования состояний, - обрабатывать медико-технологические данные изучения динамики популяций, биохимических и биофизических процессов в организме, прогнозировать развитие эпидемиологических процессов <b>Владеть:</b> - навыками использования современных методов машинного обучения для обработки клинических и экспериментальных данных при решении задач вычислительной диагностики и прогнозирования в здравоохранении
ПК-4 Способен обеспечивать информационно-технологическую поддержку в области здравоохранения	ПК 4.2 Способен осуществлять поддержку деятельности медицинских специалистов, принятия клинических и управленческих решений на основе использования информационных технологий	<b>Знать:</b> – современные информационные технологии для реализации методов поддержки принятия клинических и управленческих решений на основе машинного обучения <b>Владеть:</b> - современными информационными и коммуникационными технологиями для автоматизации обработки медико-технологических данных, изучения динамики популяций, биохимических и биофизических процессов в организме

### 1.2 Место дисциплины

Дисциплина включена в модуль «Математическое моделирование в задачах профессиональной деятельности» ОПОП ВО, часть, формируемая участниками образовательных отношений. Дисциплина осваивается на 4 курсе в 8 семестре.

## 2 Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации.

Таблица 2 – Объем и трудоемкость дисциплины по видам учебных занятий

Общая трудоемкость и виды учебной работы по дисциплине, проводимые в разных формах	Объём часов по формам обучения
	ОФО
1 Общая трудоемкость дисциплины	108
2 Контактная работа обучающихся с преподавателем (по видам учебных занятий) (всего)	64
Аудиторная работа (всего):	64

в том числе:	
лекции	28
практические занятия, семинары	
практикумы	
лабораторные работы	36
в интерактивной форме	
в электронной форме	
Внеаудиторная работа (всего):	44
в том числе, индивидуальная работа обучающихся с преподавателем	
подготовка курсовой работы /контактная работа	
групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие групповую или индивидуальную работу обучающихся с преподавателем)	
творческая работа (эссе)	
3 Самостоятельная работа обучающихся (всего)	44
4 Промежуточная аттестация обучающегося и объем часов, выделенный на промежуточную аттестацию: - зачет с оценкой	-

### 3. Учебно-тематический план и содержание дисциплины.

#### 3.1 Учебно-тематический план

Таблица 3 - Учебно-тематический план очной формы обучения

№ недел и п/п	Разделы и темы дисциплины по занятиям	Общая трудоём кость (всего час.)	Трудоемкость занятий (час.)			Формы текущего контроля и промежуточной аттестации успеваемости
			Аудиторн. занятия		СРС	
			лекц.	лаб.		
1-2	1. Введение в машинное обучение	26	6	6	8	Отчеты по лаб. работам (№1-2)
1	1.1. Искусственный интеллект и машинное обучение.	13	2	-	2	
2	1.2. Признаки. Типы признаков. Подготовка данных для обучения.		2	4	4	
3	1.3. Критерии оценки качества работы алгоритмов машинного обучения.	13	2	2	2	
4-6	2. Линейные методы машинного обучения	26	6	10	12	Отчеты по лабораторны м работам (№3-№5)
4	2.1. Постановка задачи линейной регрессии. Регуляризация задач регрессии.		2	4	4	
5	2.2. Постановка задачи линейной классификации. Логистическая регрессия.		2	2	4	
6	2.3. Метод опорных векторов.		2	4	4	
7-8	3. Метрические методы машинного обучения	28	4	6	6	Отчет по лаб. раб. №6
9-10	4. Деревья решений. Ансамблирование моделей		4	6	8	Отчеты по лаб. работам (№7-9)
9	4.1. Построение решающих деревьев.		2	2	4	
10	4.2. Ансамблирование моделей. Случайный лес.		2	4	4	
11-12	5. Искусственные нейронные сети		4	4	4	Отчет по лаб. работе №10
11	5.1. Введение в искусственные нейронные сети.		2	2	2	
12	5.2. Классификация искусственных нейронных сетей		2	2	2	
13-14	6. Кластеризация и вероятностное моделирование данных	28	4	4	6	Отчет по лабораторно й работе №11
13	6.1 Постановка задачи кластеризации. Метод k-средних.	14	2	2	2	
14	6.2. Иерархическая кластеризация. Агломеративная кластеризация	14	2	2	4	
	Промежуточная аттестация (зачет с оценкой)					
	Всего:	108	28	36	44	Зачет с оценкой

#### 4 Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации.

Для получения положительной оценки по результатам освоения дисциплины обучающемуся необходимо выполнить все установленные виды учебной работы. Оценка результатов работы обучающегося в баллах (по видам) приведена в таблице 4.

Таблица 4 - Шкала и показатели оценивания результатов учебной работы обучающихся по видам в балльно-рейтинговой системе (БРС)

Учебная работа (виды)	Сумма баллов	Виды и результаты учебной работы	Оценка в аттестации (шкала и показатели оценивания)	Баллы
Текущая учебная работа в семестре (Посещение занятий по расписанию и выполнение заданий)	100	Лекционные занятия (конспект). (14 занятий)	1 балл посещение и конспект 1 лекционного занятия	7-14
		Лабораторные работы (18 занятий)	1 балл – посещение 1 занятия и выполнение работы на 51–65% 1,5 балла – посещение 1 занятия и выполнение работы на 66–100%	18 – 27
		Отчет по лабораторной работе (11 отчетов)	3 – 4 б. (выполнено 51 - 65% заданий) 4 – 5 б. (выполнено 66 - 100% заданий)	33-59
		Итого по текущей работе в семестре		
Промежуточная аттестация (зачет)	20	Ответ на теоретический вопрос	5 баллов (пороговое значение) 10 баллов (максимальное значение)	5-10
		Решение задачи	5 баллов (пороговое значение) 10 баллов (максимальное значение)	5-10
Итого по промежуточной аттестации (зачет)				10-20
Суммарная оценка по дисциплине: Сумма баллов текущей и промежуточной аттестации 51 – 100 б.				

Если к моменту проведения зачета/ экзамена студент набирает 51 балл и более баллов, оценка может быть выставлена ему в ведомость и в зачетную книжку без процедуры принятия зачета/ экзамена. Выставление оценок производится на последней неделе теоретического обучения по данной дисциплине.

В промежуточной аттестации оценка выставляется в ведомость в 100-балльной шкале и в буквенном эквиваленте (таблица 5).

Таблица 5 – Соотнесение 100-балльной шкалы и буквенного эквивалента оценки

Сумма набранных баллов	Уровни освоения дисциплины и компетенций	Экзамен		Зачет
		Оценка	Буквенный эквивалент	Буквенный эквивалент
86 - 100	Продвинутый	5	отлично	Зачтено
66 - 85	Повышенный	4	хорошо	
51 - 65	Пороговый	3	удовлетворительно	
0 - 50	Первый	2	неудовлетворительно	Не зачтено

#### 5 Материально-техническое, программное и учебно-методическое обеспечение дисциплины.

##### 5.1 Учебная литература

##### Основная учебная литература

1. Митяков, Е. С. Искусственный интеллект и машинное обучение : учебное пособие для вузов / Е. С. Митяков, А. Г. Шмелева, А. И. Ладынин. — Санкт-Петербург : Лань, 2025. — 252 с. — ISBN 978-5-507-51465-6. — Текст : электронный // Лань :

электронно-библиотечная система. — URL: <https://reader.lanbook.com/book/507451#1> — Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный.

2. Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/131721> . — Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный.

#### **Дополнительная учебная литература**

1. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. — Москва : ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060-273-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/69955> . — Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный

2. Гладилин, П. Е. Технологии машинного обучения : учебно-методическое пособие / П. Е. Гладилин, К. О. Боченина. — Санкт-Петербург : НИУ ИТМО, 2020. — 75 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/190885> — Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный.

3. Шалев-Шварц, Ш. Идеи машинного обучения : учебное пособие / Ш. Шалев-Шварц, Бен-Давид Ш. ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 436 с. — ISBN 978-5-97060-673-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/131686> — Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный.

4. Баймуратов, И. Р. Методы автоматизации машинного обучения : учебное пособие / И. Р. Баймуратов. — Санкт-Петербург : НИУ ИТМО, 2020. — 40 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/190871> . — Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный.

5. Баланов, А. Н. Цифровизация в здравоохранении. Разработка, интеграция и внедрение современных систем : учебное пособие для вузов / А. Н. Баланов. — Санкт-Петербург : Лань, 2024. — 428 с. — ISBN 978-5-507-49376-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/417788> . — Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный.

#### **5.2 Материально-техническое и программное обеспечение дисциплины.**

Учебные занятия по дисциплине проводятся в учебных аудиториях КГПИ КемГУ:

Наименование аудитории, оборудование	адрес
<b>410 аудитория. Специализированная многофункциональная учебная аудитория для проведения учебных занятий лекционного типа, семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, в том числе для организации практической подготовки обучающихся с перечнем основного оборудования:</b> <i>Специализированная (учебная) мебель:</i> доска меловая, кафедра, моноблоки аудиторные. <i>Оборудование для презентации учебного материала:</i> компьютер с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду вуза, экран, проектор, акустическая система.	654079, Кемеровская область, г. Новокузнецк, пр-кт Металлургов, д. 19
<b>502 Лаборатория компьютерного моделирования. Специализированная многофункциональная учебная аудитория для проведения учебных занятий лекционного типа, семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, в том числе для организации практической подготовки обучающихся с перечнем основного оборудования:</b> <i>Специализированная (учебная) мебель:</i> доска меловая, кафедра, столы, стулья. <i>Оборудование для презентации учебного материала:</i> компьютер с возможностью	654079, Кемеровская область, г. Новокузнецк, пр-кт Металлургов, д. 19

<p>подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду вуза, проектор, экран.</p> <p><i>Лабораторное оборудование:</i> компьютеры для обучающихся с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду вуза</p>	
<p><b>508 аудитория. Помещение для самостоятельной работы обучающихся с перечнем основного оборудования:</b></p> <p><i>Специализированная (учебная) мебель:</i> доска меловая, кафедра, столы, стулья.</p> <p><i>Оборудование для презентации учебного материала:</i> компьютер преподавателя с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду вуза, проектор, экран.</p> <p><i>Лабораторное оборудование:</i> компьютеры для обучающихся с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду вуза.</p>	<p>654079, Кемеровская область, г. Новокузнецк, пр-кт Металлургов, д. 19</p>

### 5.3 Современные профессиональные базы данных и информационные справочные системы.

#### Электронные библиотечные ресурсы:

1. Электронная полнотекстовая база данных периодических изданий по общественным и гуманитарным наукам ООО «ИВИС», <https://eivis.ru/basic/details> Договор № 427 – П от 13.01.2025 г период подписки с 01.01.2025 г. по 31.12.2025 г., – Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный.

2. Научная электронная библиотека – <http://elibrary.ru>. Доступ к отдельным периодическим изданиям. Доступ к отдельным периодическим изданиям. Договор № SU-365/2025 от 20.12.2024 г. период подписки с 01.01.2025 г. по 31.12.2025 г. – Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный.

3. Межвузовская электронная библиотека (МЭБ) - <https://icdlib.nspu.ru> КГПИ КемГУ является участником и пользователем МЭБ. Договор № 34 от 30.09.2020 г. (договор бессрочный). – Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный.

4. Электронная библиотека КГПИ КемГУ – <https://elib.nbikemsu.ru/MegaPro/Web> – Доступ из локальной сети КГПИ КемГУ свободный, с домашних ПК – авторизованный.

#### Информационные справочные системы

1. Math-Net.Ru – Информационная система «Общероссийский математический портал». Режим доступа свободный : <http://www.mathnet.ru/>

2. Информационная система «Экспонента» - центр инженерных технологий и моделирования. Режим доступа свободный :: <https://exponenta.ru/>

3. Базы наборов данных UCI Machine Learning Repository — репозиторий наборов данных для машинного обучения. Режим доступа свободный : <http://archive.ics.uci.edu/ml/>

4. Образовательный комитет IAPR Education Committee & Resources — коллекция ссылок на образовательные ресурсы по распознаванию образов, машинному обучению, обработке сигналов, обработке изображений и компьютерному зрению, поддерживаемая Международной ассоциацией распознавания образов . Режим доступа свободный : <http://homepages.inf.ed.ac.uk/rbf/IAPR/index.php>

5. Data Science, Machine Learning. AI & Analytics – Kdnuggets – Портал по интеллектуальному анализу данных, поддерживаемый Григорием Пятецким-Шапиро . Режим доступа свободный : <http://www.kdnuggets.com/>

6. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. . Режим доступа свободный : <http://machinelearning.ru>



## 6 Иные сведения и (или) материалы.

### 6.1. Темы письменных учебных работ

Таблица 5 - Темы письменных учебных работ

Раздел	Темы	Контрольные точки
Введение в машинное обучение	Подготовка табличных медицинских данных для обучения.	Отчет по лабораторной работе №1
	Критерии оценки качества алгоритмов машинного обучения для обработки медицинских данных	Отчет по лабораторной работе №2
Линейные методы машинного обучения	Применение линейной регрессии для построения прогнозной модели при решении задач здравоохранения	Отчет по лабораторной работе №3
	Применение логистической регрессии для построения прогнозной модели при решении задач здравоохранения	Отчет по лабораторной работе №4
	Применение метода опорных векторов для построения прогнозной модели при решении задач здравоохранения	Отчет по лабораторной работе №5
Метрические методы машинного обучения	Применение метода ближайших соседей для построения прогнозной модели при решении задач здравоохранения	Отчет по лабораторной работе №6
Деревья решений. Ансамблирование моделей	Построение дерева решений для прогнозной модели при решении задач здравоохранения	Отчет по лабораторной работе №7
	Построение случайного леса для прогнозной модели при решении задач здравоохранения	Отчет по лабораторной работе №8
	Применение ансамблей моделей для решения прогнозных задач здравоохранения	Отчет по лабораторной работе №9
Искусственные нейронные сети	Применение искусственной нейронной сети для построения прогнозной модели при решении задач здравоохранения	Отчет по лабораторной работе №10
Кластеризация и вероятностное моделирование данных	Применение метода k-средних для решения задачи кластеризации в области здравоохранения	Отчет по лабораторной работе №11

### 6.2. Примерные вопросы и задания / задачи для промежуточной аттестации

Форма промежуточной аттестации зачет с оценкой.

Таблица 6 – Типовые (примерные) контрольные вопросы и задания

Разделы и темы	Примерные теоретические вопросы	Примерные практические задания / задачи
1. Введение в машинное обучение  1.1. Искусственный интеллект и машинное обучение.	1. Какие основные направления искусственного интеллекта выделяют. 2. Каковы ключевые исторические этапы развития искусственного интеллекта? 3. Какие основные задачи решаются с помощью машинного обучения? 4. Каково содержание задач классификации, регрессии, кластеризации, снижения размерности, выявления аномалий? 5. Перечислите основные виды машинного обучения. 6. В чем отличие машинного обучения с учителем и без учителя? 7. Для решения каких практических задач	1. Дана выборка, содержащая генетические данные пациентов, к какой задаче машинного обучения относится задача прогнозирования вероятности сахарного диабета? 2. В предыдущей задаче специалист после прогноза может подтвердить или же опровергнуть опасения модели, отдавая ей результат — права модель или нет. Какой вид машинного обучения реализуется в этом примере? 3. Многие модели ориентируются на данные экспрессии генов. Проблема, с которой сталкиваются исследователи — слишком большое количество генов

	используется обучение с подкреплением? 8. Какие прогнозные задачи здравоохранения решаются методами машинного обучения?	(например, у человека их около 28000). Вследствие этого очень тяжело обучать модели. Какая задача машинного обучения в этом случае требует решения?																																				
1.2. Признаки. Типы признаков. Подготовка данных для обучения.	1. Каковы основные требования к данным? Какие типы данных выделяют? 2. Какие задачи анализа данных необходимо решить перед построением модели машинного обучения? 3. Какие основные ошибки в данных могут повлиять на результат применения моделей машинного обучения? 4. Какими могут быть причины возникновения ошибок и пропусков в данных? 5. Какие существуют методы заполнения пропусков в данных? 6. Какие применяются методы обработки категориальных данных? 7. Какие применяются методы обработки числовых данных?	1. Выполните кодирование двоичными векторами признака «характер получаемого лечения»: <table><tr><td>id</td><td>f</td></tr><tr><td>1</td><td>хирургическое лечение</td></tr><tr><td>2</td><td>лучевая терапия</td></tr><tr><td>3</td><td>лучевая терапия</td></tr><tr><td>4</td><td>гормональная терапия</td></tr><tr><td>5</td><td>хирургическое лечение</td></tr></table> 2. Проведите нормализацию признаков, характеризующих клиники: x – средняя длительность пребывания пациента (в днях), y – летальность в %: <table><tr><td>id</td><td>x</td><td>y</td></tr><tr><td>1</td><td>11</td><td>1,7</td></tr><tr><td>2</td><td>12</td><td>1,8</td></tr><tr><td>3</td><td>10</td><td>1,6</td></tr><tr><td>4</td><td>11</td><td>1,7</td></tr><tr><td>5</td><td>9</td><td>1,8</td></tr></table>	id	f	1	хирургическое лечение	2	лучевая терапия	3	лучевая терапия	4	гормональная терапия	5	хирургическое лечение	id	x	y	1	11	1,7	2	12	1,8	3	10	1,6	4	11	1,7	5	9	1,8						
id	f																																					
1	хирургическое лечение																																					
2	лучевая терапия																																					
3	лучевая терапия																																					
4	гормональная терапия																																					
5	хирургическое лечение																																					
id	x	y																																				
1	11	1,7																																				
2	12	1,8																																				
3	10	1,6																																				
4	11	1,7																																				
5	9	1,8																																				
1.3. Критерии оценки качества работы алгоритмов машинного обучения.	1. Что такое функция ошибки (функция качества)? Что такое функционал ошибки (функционал качества)? 2. Какие методы используются для минимизации функционала ошибки? 3. Что называют переобучением? Как проявляется переобучение при оценке качества модели? 4. Что называют недообучением? Как проявляется недообучение при оценке качества? 5. Какие есть методы валидации моделей? 6. В чем заключается кросс-валидация? 7. Какие метрики используются для оценки качества при решении задач классификации? 8. В чем разница между точностью и полнотой в контексте метрик задачи классификации? 9. Какие метрики используются для оценки качества при решении задачи регрессии? 10. Какой диапазон значений коэффициента детерминации и как эти значения трактуются с точки зрения качества модели?	1. В таблице приведены реальные и спрогнозированные значения количества инфекционных заболеваний в текущем году по регионам. Оцените результаты прогнозирования при помощи коэффициента детерминации. <table><tr><td>id</td><td>y (точное)</td><td>y (прогноз)</td></tr><tr><td>1</td><td>57316</td><td>50879</td></tr><tr><td>2</td><td>168892</td><td>125765</td></tr><tr><td>3</td><td>160690</td><td>146876</td></tr><tr><td>4</td><td>14277</td><td>10987</td></tr><tr><td>5</td><td>170676</td><td>165987</td></tr></table> 2. В таблице приведены результаты прогноза наличия сахарного диабета по результатам генетических данных пациентов, и реальные сведения о наличии этого заболевания. Рассчитайте F-меру результатов прогнозирования. <table><tr><td>id</td><td>y (точное)</td><td>y (прогноз)</td></tr><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>2</td><td>1</td><td>1</td></tr><tr><td>3</td><td>0</td><td>1</td></tr><tr><td>4</td><td>0</td><td>0</td></tr><tr><td>5</td><td>1</td><td>0</td></tr></table>	id	y (точное)	y (прогноз)	1	57316	50879	2	168892	125765	3	160690	146876	4	14277	10987	5	170676	165987	id	y (точное)	y (прогноз)	1	0	0	2	1	1	3	0	1	4	0	0	5	1	0
id	y (точное)	y (прогноз)																																				
1	57316	50879																																				
2	168892	125765																																				
3	160690	146876																																				
4	14277	10987																																				
5	170676	165987																																				
id	y (точное)	y (прогноз)																																				
1	0	0																																				
2	1	1																																				
3	0	1																																				
4	0	0																																				
5	1	0																																				
2. Линейные методы машинного обучения  2.1. Постановка задачи линейной регрессии. Регуляризация задач регрессии.	1. Каковы преимущества и недостатки линейных моделей? 2. Модель линейной регрессии. Постановка задачи. Матричная формулировка. Точное решение. 3. Метод градиентного спуска. Стохастический градиентный спуск. 4. Что такое регуляризация в линейной регрессии? 5. Какие есть виды регуляризаторов? 6. Как использовать линейную регрессию для построения нелинейных зависимостей? 7. Что называется остатками регрессии?	1. Найдите коэффициенты линейной регрессии для прогнозирования значения признака y от x (x – средняя длительность пребывания пациента (в днях), y – летальность в %): <table><tr><td>id</td><td>x</td><td>y</td></tr><tr><td>1</td><td>11</td><td>1,7</td></tr><tr><td>2</td><td>12</td><td>1,8</td></tr><tr><td>3</td><td>10</td><td>1,6</td></tr><tr><td>4</td><td>11</td><td>1,7</td></tr><tr><td>5</td><td>9</td><td>1,8</td></tr></table> 2. Для задачи 1 рассчитайте функционал потерь.	id	x	y	1	11	1,7	2	12	1,8	3	10	1,6	4	11	1,7	5	9	1,8																		
id	x	y																																				
1	11	1,7																																				
2	12	1,8																																				
3	10	1,6																																				
4	11	1,7																																				
5	9	1,8																																				

	<p>Как по анализу остатков оценить качество модели?</p> <p>8. Какие существуют методы оценки влияния признаков на модель?</p>	<p>3. Напишите код на Python для обучения модели линейной регрессии по данным из задания 1. Используйте методы библиотеки Scikit-learn (sklearn).</p>																																
<p>2.2. Постановка задачи линейной классификации. Логистическая регрессия.</p>	<p>1. Как определяется понятие классификация в контексте машинного обучения?</p> <p>2. Чем отличаются бинарная и мультиклассовая классификации?</p> <p>3. Как происходит классификация с перекрывающимися и неперекрывающимися классами?</p> <p>4. В чем отличие четкой и нечеткой классификации?</p> <p>5. Линейно разделимые выборки. В чем суть линейного классификатора? Жесткая классификация на основе линейного классификатора.</p> <p>6. Мягкая классификация на основе логистической регрессии.</p> <p>7. Кросс-энтропия и функционал потерь логистической регрессии.</p> <p>8. Регуляризация логистической регрессии.</p> <p>9. В каких случаях возникает необходимость варьировать вероятностный порог принятия решения?</p>	<p>1. Определите является ли выборка линейно разделимой (<math>x</math> – вес человека, <math>y</math> – рост, 1 – индекс массы тела соответствует норме, 0 – дефицит массы тела):</p> <table><tr><td>id</td><td>x</td><td>y</td><td>ИМТ</td></tr><tr><td>1</td><td>45</td><td>150</td><td>1</td></tr><tr><td>2</td><td>48</td><td>155</td><td>1</td></tr><tr><td>3</td><td>45</td><td>165</td><td>0</td></tr><tr><td>4</td><td>45</td><td>160</td><td>0</td></tr><tr><td>5</td><td>51</td><td>170</td><td>0</td></tr><tr><td>6</td><td>51</td><td>160</td><td>1</td></tr></table> <p>2. Для выборки из задачи 1 постройте линейный классификатор.</p> <p>3. Напишите код на Python для обучения модели логистической регрессии по данным из задания 1. Используйте методы библиотеки Scikit-learn (sklearn).</p>	id	x	y	ИМТ	1	45	150	1	2	48	155	1	3	45	165	0	4	45	160	0	5	51	170	0	6	51	160	1				
id	x	y	ИМТ																															
1	45	150	1																															
2	48	155	1																															
3	45	165	0																															
4	45	160	0																															
5	51	170	0																															
6	51	160	1																															
<p>2.3. Метод опорных векторов.</p>	<p>1. В чем состоит идея метода опорных векторов для линейно разделимой выборки? Разделяющая полоса и опорные векторы.</p> <p>2. Понятие отступа для линейного классификатора. Эталонные, неинформативные объекты и выбросы.</p> <p>3. Задача условной оптимизации метода опорных векторов в случае линейно-разделимой выборки. Сведение условной оптимизации к безусловной.</p> <p>4. Опорные векторы для линейно-неразделимой выборки.</p> <p>5. Задача условной оптимизации метода опорных векторов в случае линейно-неразделимой выборки. Сведение условной оптимизации к безусловной.</p> <p>6. Нелинейные обобщения метода опорных векторов. Возможные виды ядер.</p>	<p>1. Постройте разделяющую полосу для выборки (<math>x</math> – вес человека, <math>y</math> – рост, 1 – индекс массы тела соответствует норме, 0 – дефицит массы тела):</p> <table><tr><td>id</td><td>x</td><td>y</td><td>ИМТ</td></tr><tr><td>1</td><td>45</td><td>150</td><td>1</td></tr><tr><td>2</td><td>48</td><td>155</td><td>1</td></tr><tr><td>3</td><td>45</td><td>165</td><td>0</td></tr><tr><td>4</td><td>45</td><td>160</td><td>0</td></tr><tr><td>5</td><td>51</td><td>170</td><td>0</td></tr><tr><td>6</td><td>51</td><td>160</td><td>1</td></tr></table> <p>2. Определите опорные векторы выборки из задания 1 (используя построенную разделяющую полосу)</p> <p>3. Напишите код на Python для обучения модели методом опорных векторов по данным из задания 1. Используйте методы библиотеки Scikit-learn (sklearn).</p>	id	x	y	ИМТ	1	45	150	1	2	48	155	1	3	45	165	0	4	45	160	0	5	51	170	0	6	51	160	1				
id	x	y	ИМТ																															
1	45	150	1																															
2	48	155	1																															
3	45	165	0																															
4	45	160	0																															
5	51	170	0																															
6	51	160	1																															
<p>3. Метрические методы машинного обучения</p>	<p>1. Какие методы измерения расстояний между объектами используют в задачах классификации? В чем заключается их суть?</p> <p>2. В чем суть метода ближайшего соседа для задачи классификации.</p> <p>3. Метод k ближайших соседей. Особенности подбора гиперпараметра k.</p> <p>4. Взвешенный метод k ближайших соседей.</p> <p>5. Отбор эталонных объектов.</p> <p>6. Регрессия по соседним объектам.</p>	<p>1. Определите класс в пустой ячейке, используя разное количество соседей, таблицы выборки (<math>x</math> – вес человека, <math>y</math> – рост, 1 – индекс массы тела соответствует норме, 0 – дефицит массы тела):</p> <table><tr><td>id</td><td>x</td><td>y</td><td>ИМТ</td></tr><tr><td>1</td><td>45</td><td>150</td><td>1</td></tr><tr><td>2</td><td>48</td><td>155</td><td>1</td></tr><tr><td>3</td><td>45</td><td>165</td><td>0</td></tr><tr><td>4</td><td>45</td><td>160</td><td>0</td></tr><tr><td>5</td><td>51</td><td>170</td><td>0</td></tr><tr><td>6</td><td>51</td><td>160</td><td>1</td></tr><tr><td>7</td><td>48</td><td>160</td><td>?</td></tr></table> <p>2. Постройте регрессию методом k-ближайших соседей (используя разное количество соседей) для прогнозирования значения признака <math>y</math> от <math>x</math> (<math>x</math> – средняя</p>	id	x	y	ИМТ	1	45	150	1	2	48	155	1	3	45	165	0	4	45	160	0	5	51	170	0	6	51	160	1	7	48	160	?
id	x	y	ИМТ																															
1	45	150	1																															
2	48	155	1																															
3	45	165	0																															
4	45	160	0																															
5	51	170	0																															
6	51	160	1																															
7	48	160	?																															

		<p>длительность пребывания пациента (в днях), <math>y</math> – летальность в %):</p> <table><tr><td>id</td><td>x</td><td>y</td></tr><tr><td>1</td><td>11</td><td>1,7</td></tr><tr><td>2</td><td>12</td><td>1,8</td></tr><tr><td>3</td><td>10</td><td>1,6</td></tr><tr><td>4</td><td>11</td><td>1,7</td></tr><tr><td>5</td><td>9</td><td>1,8</td></tr></table> <p>3. В задаче 2 вычислите средние квадратические ошибки при разных значениях гиперпараметра <math>k</math>.</p> <p>4. Напишите код на Python для обучения модели методом ближайших соседей по данным из заданий 1 и 2. Используйте методы библиотеки Scikit-learn (sklearn).</p>	id	x	y	1	11	1,7	2	12	1,8	3	10	1,6	4	11	1,7	5	9	1,8																																						
id	x	y																																																								
1	11	1,7																																																								
2	12	1,8																																																								
3	10	1,6																																																								
4	11	1,7																																																								
5	9	1,8																																																								
<p>4. Деревья решений. Ансамблирование моделей</p> <p>4.1. Построение решающих деревьев.</p>	<p>1. В чем суть построения дерева решений? Определите понятия: решающее правило (предикат), корень, узел, лист.</p> <p>2. Построение дерева классификации и дерева регрессии.</p> <p>3. Что называют глубиной дерева? Как глубина дерева влияет на результат обучения?</p> <p>4. Методика усечения деревьев при построении по предварительном ограничениям.</p> <p>5. Пост усечение дерева - регуляризация.</p> <p>6. Как влияет выбор решающего правила на результат обучения? Выбор решающего правила на основании уменьшения энтропии и на основании прироста информации.</p>	<p>1.Вычислив дисперсии признаков <math>x</math> и <math>y</math> (предварительно нормализуйте) для таблицы выборки (<math>x</math> – вес человека, <math>y</math> – рост, <math>1</math> – индекс массы тела соответствует норме, <math>0</math> – дефицит массы тела), определите по какому признаку следует сделать первое разбиение:</p> <table><tr><td>id</td><td>x</td><td>y</td><td>ИМТ</td></tr><tr><td>1</td><td>45</td><td>150</td><td>1</td></tr><tr><td>2</td><td>48</td><td>155</td><td>1</td></tr><tr><td>3</td><td>45</td><td>165</td><td>0</td></tr><tr><td>4</td><td>45</td><td>160</td><td>0</td></tr><tr><td>5</td><td>51</td><td>170</td><td>0</td></tr><tr><td>6</td><td>51</td><td>160</td><td>1</td></tr><tr><td>7</td><td>48</td><td>160</td><td>1</td></tr></table> <p>2. Вычислив энтропию каждого признака, определите по какому признаку следует сделать первое разбиение:</p> <table><tr><td>id</td><td>характер получаемого лечения</td><td>возраст</td><td>ремиссия</td></tr><tr><td>1</td><td>хирургическое лечение</td><td>25</td><td>1</td></tr><tr><td>2</td><td>лучевая терапия</td><td>35</td><td>1</td></tr><tr><td>3</td><td>лучевая терапия</td><td>25</td><td>0</td></tr><tr><td>4</td><td>гормональная терапия</td><td>30</td><td>0</td></tr><tr><td>5</td><td>хирургическое лечение</td><td>35</td><td>1</td></tr></table> <p>3. В задаче 2 определите по какому признаку следует сделать первое разбиение, вычислив меру Джини.</p> <p>4. Напишите код на Python для построения деревьев решений по данным из заданий 1 и 2. Используйте методы библиотеки Scikit-learn (sklearn).</p>	id	x	y	ИМТ	1	45	150	1	2	48	155	1	3	45	165	0	4	45	160	0	5	51	170	0	6	51	160	1	7	48	160	1	id	характер получаемого лечения	возраст	ремиссия	1	хирургическое лечение	25	1	2	лучевая терапия	35	1	3	лучевая терапия	25	0	4	гормональная терапия	30	0	5	хирургическое лечение	35	1
id	x	y	ИМТ																																																							
1	45	150	1																																																							
2	48	155	1																																																							
3	45	165	0																																																							
4	45	160	0																																																							
5	51	170	0																																																							
6	51	160	1																																																							
7	48	160	1																																																							
id	характер получаемого лечения	возраст	ремиссия																																																							
1	хирургическое лечение	25	1																																																							
2	лучевая терапия	35	1																																																							
3	лучевая терапия	25	0																																																							
4	гормональная терапия	30	0																																																							
5	хирургическое лечение	35	1																																																							
<p>4.2. Ансамблирование моделей. Случайный лес.</p>	<p>1. Кросс-валидация и ансамблирование моделей, построенных одним методом. Голосование. Усреднение.</p> <p>2. Построение бутстрэп выборки. Ансамблирование на основе голосования (бэггинг).</p> <p>3. Применение бэггинга для решающих деревьев.</p> <p>4. Метод случайных подпространств.</p>	<p>1. Применив кросс-валидацию постройте не менее трех моделей регрессии для прогнозирования значения признака <math>y</math> от <math>x</math> (<math>x</math> – средняя длительность пребывания пациента (в днях), <math>y</math> – летальность в %). Используя ансамблирование, получите прогноз для <math>x = 13</math>.</p> <table><tr><td>id</td><td>x</td><td>y</td></tr><tr><td>1</td><td>11</td><td>1,7</td></tr></table>	id	x	y	1	11	1,7																																																		
id	x	y																																																								
1	11	1,7																																																								

	5. Случайный лес. 6. Гиперпараметры случайного леса. Методы Python для проведения экспериментов по подбору гиперпараметров. 7. Применение ансамблирования для моделей, построенных разными методами. Бустинг. Градиентный бустинг.	<table><tr><td>2</td><td>12</td><td>1,8</td></tr><tr><td>3</td><td>10</td><td>1,6</td></tr><tr><td>4</td><td>11</td><td>1,7</td></tr><tr><td>5</td><td>9</td><td>1,8</td></tr></table> 2. Напишите код на Python для построения случайного леса по данным из задания 1. Используйте методы библиотеки Scikit-learn (sklearn).	2	12	1,8	3	10	1,6	4	11	1,7	5	9	1,8																																						
2	12	1,8																																																		
3	10	1,6																																																		
4	11	1,7																																																		
5	9	1,8																																																		
5. Искусственные нейронные сети  5.1. Введение в искусственные нейронные сети.	1. Что представляет собой искусственная нейронная сеть (ИНС) и в каких областях применима? 2. Какие основные компоненты составляют структуру ИНС и какие роли выполняют? 3. Какую роль выполняют функции активации в ИНС и как выбор функции влияет на процесс обучения? 4. Что такое дельта-правило в контексте обучения ИНС и как оно связано с коррекцией весов? 5. Как работает алгоритм обратного распространения ошибки и каким образом он применяется для обучения многослойных ИНС? 6. Какие преимущества и ограничения существуют у метода обратного распространения ошибки?	1. Для сети с одним нейроном, имеющим 2 входа и один выход и линейной функцией активации с нулевым сдвигом, скорректируйте начальные веса методом обратного распространения ошибки (в качестве функции ошибки возьмите среднее квадратичное отклонение), для элемента выборки (x – вес человека, y – рост, 1 – индекс массы тела соответствует норме, 0 – дефицит массы тела): <table><tr><td>id</td><td>x</td><td>y</td><td>ИМТ</td></tr><tr><td>1</td><td>45</td><td>150</td><td>1</td></tr></table> Проверьте, что ошибка действительно уменьшилась. Начальные веса возьмите произвольно. 2. Напишите программу на Python (без использования библиотечных ИНС), реализующую обучение персептрона для данных из задания 1.	id	x	y	ИМТ	1	45	150	1																																										
id	x	y	ИМТ																																																	
1	45	150	1																																																	
5.2. Классификация искусственных нейронных сетей	1. Какие виды функций активации существуют? Как зависит выбор функции от решаемой задачи? 2. В чем заключается задача классификации к контексте ИНС и какие типы ИНС обычно применяют для ее решения? 3. Какие гиперпараметры влияют на эффективность ИНС? 4. Какие основные категории ИНС с точки зрения их архитектуры существуют и для каких задач они наиболее подходят?	1. Напишите код на Python, реализующий нейросеть для классификации по таблице выборки (x – вес человека, y – рост, 1 – индекс массы тела соответствует норме, 0 – дефицит массы тела): <table><tr><td>id</td><td>x</td><td>y</td><td>ИМТ</td></tr><tr><td>1</td><td>45</td><td>150</td><td>1</td></tr><tr><td>2</td><td>48</td><td>155</td><td>1</td></tr><tr><td>3</td><td>45</td><td>165</td><td>0</td></tr><tr><td>4</td><td>45</td><td>160</td><td>0</td></tr><tr><td>5</td><td>51</td><td>170</td><td>0</td></tr><tr><td>6</td><td>51</td><td>160</td><td>1</td></tr><tr><td>7</td><td>48</td><td>160</td><td>1</td></tr></table> Используйте методы библиотеки Scikit-learn (sklearn). 2. Напишите код на Python, реализующий нейросеть для прогнозирования значения признака y от x (x – средняя длительность пребывания пациента (в днях), y – летальность в %). <table><tr><td>id</td><td>x</td><td>y</td></tr><tr><td>1</td><td>11</td><td>1,7</td></tr><tr><td>2</td><td>12</td><td>1,8</td></tr><tr><td>3</td><td>10</td><td>1,6</td></tr><tr><td>4</td><td>11</td><td>1,7</td></tr><tr><td>5</td><td>9</td><td>1,8</td></tr></table> Используйте методы библиотеки Scikit-learn (sklearn).	id	x	y	ИМТ	1	45	150	1	2	48	155	1	3	45	165	0	4	45	160	0	5	51	170	0	6	51	160	1	7	48	160	1	id	x	y	1	11	1,7	2	12	1,8	3	10	1,6	4	11	1,7	5	9	1,8
id	x	y	ИМТ																																																	
1	45	150	1																																																	
2	48	155	1																																																	
3	45	165	0																																																	
4	45	160	0																																																	
5	51	170	0																																																	
6	51	160	1																																																	
7	48	160	1																																																	
id	x	y																																																		
1	11	1,7																																																		
2	12	1,8																																																		
3	10	1,6																																																		
4	11	1,7																																																		
5	9	1,8																																																		
6. Кластеризация и вероятностное моделирование данных  6.1. Постановка	1. Какова основная постановка задачи кластеризации и какие цели преследуются при применении методов кластерного анализа? 2. Функционал качества кластеризации. 3. ЕМ-алгоритм в кластерном анализе	1. Дана двухклассовая выборка (x – вес человека, y – рост, 1 – индекс массы тела соответствует норме, 0 – дефицит массы тела): <table><tr><td>id</td><td>x</td><td>y</td><td>ИМТ</td></tr><tr><td>1</td><td>45</td><td>150</td><td>1</td></tr></table>	id	x	y	ИМТ	1	45	150	1																																										
id	x	y	ИМТ																																																	
1	45	150	1																																																	

задачи кластеризации. Метод k-средних.	<p>4. Что представляет собой метод k-средних и каким образом он решает задачу кластеризации?</p> <p>5. Как выбирается количество кластеров в методе k-средних и как это влияет на результаты кластеризации?</p>	<table><tr><td>2</td><td>48</td><td>155</td><td>1</td></tr><tr><td>3</td><td>45</td><td>165</td><td>0</td></tr><tr><td>4</td><td>45</td><td>160</td><td>0</td></tr><tr><td>5</td><td>51</td><td>170</td><td>0</td></tr><tr><td>6</td><td>51</td><td>160</td><td>1</td></tr><tr><td>7</td><td>48</td><td>160</td><td>1</td></tr></table> <p>Вычислите среднее внутрикластерное расстояние точки id=1.</p> <p>2. Для выборки из задания 1 вычислите среднее межкластерное расстояние для точки id=1.</p>	2	48	155	1	3	45	165	0	4	45	160	0	5	51	170	0	6	51	160	1	7	48	160	1
2	48	155	1																							
3	45	165	0																							
4	45	160	0																							
5	51	170	0																							
6	51	160	1																							
7	48	160	1																							
6.2. Иерархическая кластеризация. Агломеративная кластеризация	<p>1. В чем заключается быстрая агломеративная кластеризация?</p> <p>2. Какие этапы включает агломеративная кластеризация?</p> <p>3. Как выбирается мера близости между кластерами в агломеративной кластеризации и как она влияет на их формирование?</p> <p>4. Какие существуют методы объединения кластеров в агломеративной кластеризации и как выбрать наиболее подходящий метод для задачи?</p> <p>5. Что такое иерархическая кластеризация и какова ее связь с агломеративной?</p> <p>6. Какие преимущества и ограничения существуют у метода k-средних по сравнению с агломеративной кластеризацией?</p> <p>7. Какие факторы могут повлиять на результаты кластеризации при использовании агломеративной кластеризации?</p> <p>8. Каким образом можно интерпретировать результаты кластеризации, полученные с использованием различных методов и как выбрать наилучший для конкретной задачи?</p>	<p>1. Постройте дендрограмму для выборки (x – вес человека, y – рост):</p> <table><tr><td>id</td><td>x</td><td>y</td></tr><tr><td>1</td><td>45</td><td>150</td></tr><tr><td>2</td><td>48</td><td>155</td></tr><tr><td>3</td><td>45</td><td>165</td></tr><tr><td>4</td><td>45</td><td>160</td></tr><tr><td>5</td><td>51</td><td>170</td></tr><tr><td>6</td><td>51</td><td>160</td></tr><tr><td>7</td><td>48</td><td>160</td></tr></table> <p>2. Напишите код на Python для построения дендрограммы для выборки из задания 1.</p> <p>3. Напишите код на Python для кластеризации выборки из задания 1. Используйте методы библиотеки Scikit-learn (sklearn).</p>	id	x	y	1	45	150	2	48	155	3	45	165	4	45	160	5	51	170	6	51	160	7	48	160
id	x	y																								
1	45	150																								
2	48	155																								
3	45	165																								
4	45	160																								
5	51	170																								
6	51	160																								
7	48	160																								

Составитель (и): Решетникова Е.В., доцент, зав. каф. МФиММ