

Федеральное государственное бюджетное образовательное учреждение
инклюзивного высшего образования
«Московский государственный гуманитарно-экономический университет»

Факультет Прикладной математики и информатики
Кафедра Информационных технологий и прикладной математики

УТВЕРЖДАЮ

И.о. проректора по учебно-
методической работе
Хакимов Р.М.



«30» августа 2021г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ**

образовательная программа направления подготовки
09.03.03 "Прикладная информатика"
Блок Б1.В.ДВ.05.02 «Дисциплины (модули)», часть, формируемая
участниками образовательных отношений, дисциплины (модули) по выбору

Профиль подготовки
Прикладная информатика в биоинформационных технологиях

Квалификация (степень) выпускника
Бакалавр

Форма обучения: очная

Курс 3, 4 семестр 6, 7

Москва
2021

Рабочая программа составлена на основании федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.03 «Прикладная информатика (уровень бакалавриата)», утвержденного приказом Министерства образования и науки Российской Федерации № 922 от 19 сентября 2017 г. Зарегистрировано в Минюсте России 12 октября 2017 г. №48531.

Составители рабочей программы: МГГЭУ, доцент кафедры ИТиПМ

место работы, занимаемая должность


подпись

Никольский А.Е.
Ф.И.О.

«30» августа 2021 г.
Дата

Рецензент: МГГЭУ, профессор кафедры информационных технологий и прикладной математики

место работы, занимаемая должность


подпись

Истомина Т.В.
Ф.И.О.

«30» августа 2021 г.
Дата

Рабочая программа утверждена на заседании кафедры Информационных технологий и прикладной математики (протокол № 2 от «30» августа 2021 г.)

Зав. кафедрой ИТиПМ


подпись

Митрофанов Е.П. «30» августа 2021 г.

Ф.И.О.

Дата

СОГЛАСОВАНО

Начальник
учебного отдела

«30» августа 2021 г.

Дата


подпись

И.Г.Дмитриева

Ф.И.О.

СОГЛАСОВАНО

Декан факультета ПМий

«30» августа 2021 г.

Дата


подпись

Е.В. Петрунина

Ф.И.О.

СОГЛАСОВАНО

Заведующая библиотекой

«30» августа 2021 г.

Дата


подпись

В.А. Ахтырская

Ф.И.О.

1. ОРГАНИЗАЦИОННО-МЕТОДИЧЕСКИЙ РАЗДЕЛ

1.1. Цель и задачи изучения учебной дисциплины (модуля)

- сформировать теоретические знания по основам машинного обучения для построения формальных математических моделей и интерпретации результатов моделирования;

- выработать умения по практическому применению методов машинного обучения для построения формальных математических моделей и интерпретации результатов моделирования при решении прикладных задач в различных прикладных областях;

- выработать умения и навыки использования различных программных инструментов анализа баз данных и систем машинного обучения.

1.2. Требования к результатам освоения дисциплины

Изучение данной дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Планируемые результаты обучения по дисциплине (модулю), характеризующие этапы формирования компетенций
ПК-10. Способен применять системный подход и математические методы в формализации решения прикладных задач	ПК-10.1. Знает базовые положения фундаментальных разделов системного анализа и математики в объеме, необходимом для обработки информации и анализа данных в прикладной области; принципы и методы проведения исследований в области информационных систем и технологий; техники планирования и проведения вычислительного эксперимента.
	ПК-10.2. Умеет формулировать и доказывать наиболее важные результаты в прикладных областях; применять численные методы для решения прикладных задач; программно реализовать вычислительный эксперимент посредством языков программирования или с использованием специализированных пакетов прикладных программ; разрабатывать алгоритмы решения конкретных задач.
	ПК-10.3. Владеет навыками постановки задачи; навыками работы с библиографическими источниками информации; навыками решения поставленных задач в предметной области в рамках выбранного профиля.

1.3. Место дисциплины (модуля) в структуре образовательной программы направления подготовки 09.03.03 Прикладная информатика (бакалавриат).

Учебная дисциплина «Методы машинного обучения» относится к части, формируемой участниками образовательных отношений блока Б1. «Дисциплин (модулей)» и является дисциплиной по выбору.

Изучение учебной дисциплины «Методы машинного обучения» базируется на знаниях, умениях и навыках, полученных обучающимися при изучении предшествующих курсов: «Информационные системы и технологий», «Высокоуровневые методы информатики и программирования». Изучение учебной дисциплины «Методы машинного обучения» необходимо для освоения практически всех последующих дисциплин учебного плана и защиты ВКР.

2. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

2.1. Объем дисциплины и виды учебной работы в соответствии с формами обучения

Объем дисциплины «Методы машинного обучения» составляет 6 з.е. / 216 часов.

Вид учебной работы	Всего, часов	Курс, часов	
		3 курс 6 сем	4 курс 7 сем.
Аудиторная работа обучающихся с преподавателем (по видам учебных занятий), всего в том числе:	104	58	46
Лекции	38	20	18
Практические занятия	66	38	28
Лабораторные занятия			
Самостоятельная работа обучающихся	76	50	26
Промежуточная аттестация (подготовка и сдача), всего:			
Контрольная работа			
Курсовая работа			
Зачет			
Экзамен	36		36
Итого:	216/6	108/3	108/3

2.2. Содержание дисциплины по темам (разделам)

№ п/п	Наименование раздела (темы)	Содержание раздела (тематика занятий)	Формируемые компетенции (индекс)
1.	Раздел 1. Понятийный аппарат и методологическая база методов машинного обучения.	Тема 1. Основные понятия и определения в методах машинного обучения. Тема 2. История развития методов машинного обучения. Тема 3. Методологическая база методов машинного обучения.	ПК-10
2.	Раздел 2. Основные технологии, используемые в методах машинного обучения.	Тема 1. Типы и способы представления методов машинного обучения Тема 2. Базовые алгоритмы решения задач машинного обучения Тема 3. Основные программно-информационные ресурсы методов машинного обучения	ПК-10
3.	Раздел 3. Анализ многомерных данных. Корреляционные и причинно-следственные связи	Тема 1. Анализ многомерных данных. Корреляционные и причинно-следственные связи. Корреляция признаков и структура данных. Тема 2. Регрессия. Метод наименьших квадратов. Теорема Гаусса-Маркова. Обобщенный метод наименьших квадратов. Тема 3. Многомерная регрессия. Особенности построения регрессии по	ПК-10

		<p>многомерным данным. Множественная линейная регрессия, ее преимущества и недостатки.</p> <p>Тема 4. Кластеризация. Кластеризация как классификация без учителя. Меры сходства и меры различия образов. Метод К средних. Метод ISODATA. Метод FOREL.</p>	
4.	Раздел 4. Графовые методы. Нейронные сети.	<p>Тема 1. Графовые методы. Иерархическая кластеризация. Агломеративные и разделяющие алгоритмы кластеризации. Дендрограммы.</p> <p>Тема 2. Нейронные сети. Предпосылки возникновения нейросетей. Перцептрон Розенблатта. Многослойный перцептрон. Карты Кохонена. Сети Хопфилда. Методы обучения нейросетей. Метод опорных векторов.</p> <p>Тема 3. Машинное обучение и теория Вапника-Червоненкиса. Принцип структурной минимизации риска. Метод опорных векторов. Политика назначения штрафов.</p>	ПК-10

2.3. Разделы дисциплин и виды занятий

№ п/п	Наименование темы дисциплины	Лекционные занятия	Практические занятия	Самостоятельная работа	Всего часов	Формы текущего контроля успеваемости
1.	Понятийный аппарат и методологическая база методов машинного обучения.	10	16	20	46	Устный опрос, отчет о практической работе
2.	Основные технологии, используемые в методах машинного обучения.	10	22	30	62	Устный опрос, реферат
Итого: (6 семестр)		20	38	50	108	
3.	Анализ многомерных данных. Корреляционные и причинно-следственные связи	8	14	14	36	Устный опрос, отчет о практической работе
4.	Графовые методы. Нейронные сети.	10	14	12	36	Устный опрос,
Экзамен					36	
Итого: (7 семестр)		18	28	26	108	

2.4. Планы теоретических (лекционных) занятий

№	Наименование тем лекций	Кол-во часов в 6,7 семестрах
6 семестр		
Раздел 1. Понятийный аппарат и методологическая база методов машинного обучения		
1.	Основные понятия и определения технологии, используемые в методах машинного обучения. История развития основных технологии, используемых в методах машинного обучения.	6
2.	Методологическая база используемой в методах машинного обучения.	4
Раздел 2. Основные технологии, используемые в методах машинного обучения.		
1.	Типы и способы представления данных в методах машинного обучения.	2
2.	Базовые алгоритмы решения задач, используемых в методах машинного обучения.	4
3.	Основные программно-информационные ресурсы, используемые в методах машинного обучения.	4
7 семестр		
Раздел 3. Анализ многомерных данных. Корреляционные и причинно-следственные связи		
1.	Анализ многомерных данных. Корреляционные и причинно-следственные связи. Корреляция признаков и структура данных.	2
2.	Регрессия. Метод наименьших квадратов. Теорема Гаусса-Маркова. Обобщенный метод наименьших квадратов.	2
3.	Многомерная регрессия. Особенности построения регрессии по многомерным данным. Множественная линейная регрессия, ее преимущества и недостатки.	2
4.	Кластеризация. Кластеризация как классификация без учителя. Меры сходства и меры различия образов. Метод K средних. Метод ISODATA. Метод FOREL.	2
Раздел 4. Графовые методы. Нейронные сети.		
1.	Графовые методы. Иерархическая кластеризация. Агломеративные и разделяющие алгоритмы кластеризации. Дендрограммы.	2
2.	Нейронные сети. Предпосылки возникновения нейросетей. Перцептрон Розенблатта. Многослойный перцептрон. Карты Кохонена. Сети Хопфилда. Методы обучения нейросетей. Метод опорных векторов.	4
3.	Машинное обучение и теория Вапника-Червоненкиса. Принцип структурной минимизации риска. Метод опорных векторов. Политика назначения штрафов.	4

2.5. Планы практических (семинарских) занятий

№	Наименование тем практических (семинарских) занятий	Кол-во часов в 6,7 семестрах
6 семестр		
Раздел 1. Понятийный аппарат и методологическая база используемая в методах машинного обучения		
1.	История развития информационного менеджмента в медицине.	6

2.	Методологическая база информационного менеджмента в медицине.	10
Раздел 2. Основные информационные технологии, используемые в методах машинного обучения		
1.	Базовые алгоритмы решения задач в методах машинного обучения	4
2.	Основные программно-информационные ресурсы в методах машинного обучения	4
3.	Решение задач методами машинного обучения	6
4.	Решение задач методами машинного обучения в программной среде SiLab.	8
7 семестр		
Раздел 3 Анализ многомерных данных. Корреляционные и причинно-следственные связи		
1.	Анализ многомерных данных. Корреляционные и причинно-следственные связи. Корреляция признаков и структура данных.	4
2.	Регрессия. Метод наименьших квадратов. Теорема Гаусса-Маркова. Обобщенный метод наименьших квадратов.	4
3.	Многомерная регрессия. Особенности построения регрессии по многомерным данным. Множественная линейная регрессия, ее преимущества и недостатки.	6
Раздел 4. Графовые методы. Нейронные сети.		
1.	Графовые методы. Иерархическая кластеризация. Дендрограммы.	4
2.	Нейронные сети. Перцептрон Розенблатта. Многослойный перцептрон. Карты Кохонена. Сети Хопфилда. Методы обучения нейросетей.	4
3.	Машинное обучение и теория Вапника-Червоненкиса. Принцип структурной минимизации риска. Метод опорных векторов	6

2.6. Планы лабораторных работ - не предусмотрены учебным планом

2.7. Планы самостоятельной работы обучающегося по дисциплине (модулю)

№	Название разделов и тем	Виды самостоятельной работы	Трудоемкость	Формируемые компетенции	Формы контроля
1.	Понятийный аппарат и методологическая база в методах машинного обучения.	Работа с источниками	20	ПК-10	Устный опрос, отчет о практической работе
2.	Основные технологии, используемые в методах машинного обучения	Оформление отчетов	30	ПК-10	Устный опрос реферат
3.	Анализ многомерных данных. Корреляционные и причинно-следственные связи	Работа с источниками	14	ПК-10	Устный опрос, отчет о практической работе
4.	Графовые методы. Нейронные сети.	Оформление отчетов	12	ПК-10	Устный опрос, отчет о практической работе

3. ОСОБЕННОСТИ ОБУЧЕНИЯ ЛИЦ С ИНВАЛИДНОСТЬЮ И ОВЗ

Обучение лиц с ограниченными возможностями здоровья осуществляется с учетом индивидуальных психофизических особенностей, а для инвалидов также в соответствии с индивидуальной программой реабилитации инвалида.

Для получения обучающимися, имеющими ограниченные физические возможности, качественного образования должны выполняться следующие важные условия: обучающийся должен иметь возможность беспрепятственно посещать образовательное учреждение и использовать в своём обучении дистанционные образовательные технологии.

Для обучения и контроля обучающихся с нарушениями координации движений предусмотрено проведение тестирования с использованием компьютера.

Во время аудиторных занятий обязательно использование средств обеспечения наглядности учебного материала с помощью мультимедийного проектора. Скорость изложения материала должна учитывать ограниченные физические возможности студентов.

4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Учебно-методическое обеспечение дисциплины для организации самостоятельной работы студентов (содержит перечень основной литературы, дополнительной литературы, программного обеспечения и Интернет-ресурсы).

В распоряжении преподавателей и обучающихся имеется основное необходимое материально-техническое оборудование, Интернет-ресурсы, доступ к полнотекстовым электронным базам, книжный фонд библиотеки Московского государственного гуманитарно-экономического университета.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1. Основная литература

1. Исаев, С.В. Интеллектуальные системы : учеб. пособие / С.В. Исаев, О.С. Исаева. - Красноярск : Сиб. федер. ун-т, 2017. - 120 с. - ISBN 978-5-7638-3781-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1032129>

2. Пятаева, А. В. Интеллектуальные системы и технологии : учеб. пособие / А. В. Пятаева, К. В. Раевич. - Красноярск : Сиб. федер. ун-т, 2018. - 144 с. - ISBN 978-5-7638-3873-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1032131>

5.2. Дополнительная литература

1. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2021. — 243 с. — (Высшее образование). — ISBN 978-5-534-01042-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/469867>

2. Новиков, Ф. А. Символический искусственный интеллект: математические основы представления знаний : учебное пособие для вузов / Ф. А. Новиков. — Москва : Издательство Юрайт, 2021. — 278 с. — (Высшее образование). — ISBN 978-5-534-00734-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/470241>

3. Станкевич, Л. А. Интеллектуальные системы и технологии : учебник и практикум для вузов / Л. А. Станкевич. — Москва : Издательство Юрайт, 2021. — 397 с. — (Высшее образование). — ISBN 978-5-534-02126-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/469517>

5.3. Программное обеспечение

Scilab

R STUDIO

Python с расширениями PIL, Py OpenGL

Microsoft Office

Microsoft Windows

7-Zip

AcrobatReader

5.4. Электронные ресурсы

1. Национальный открытый университет ИНТУИТ [Электронный ресурс]. URL: <http://www.intuit.ru>

2. Хабрахабр [Электронный ресурс]. URL: <http://habrahabr.ru/>.

3. <http://www.lessons-tva.info/> - На сайте представлены различные учебные материалы, в том числе онлайн учебники (авторские курсы) по дисциплинам: экономическая информатика, компьютерные сети и телекоммуникации, основы электронного бизнеса, информатика и компьютерная техника.

4. Электронная библиотека: <http://elibrary.ru/defaultx.asp>

5. Java портал Sun Microsystems – <http://java.sun.com>.

6. Programmer's Forum: <http://www.programmist.net>

7. Портал разработчиков андроид: <http://developer.android.com>

8. Библиотека ТехНэт: <http://technet.microsoft.com/ru-ru/library/aa991542>

9. Электронная библиотека «Знаниум»: <https://znanium.com/>

10. Электронная библиотека «Юрайт»: <https://urait.ru/>

11. Научная электронная библиотека «Elibrary.ru»: <https://www.elibrary.ru/defaultx.asp>

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ (МОДУЛЯ)

№ п/п	Наименование оборудованных учебных кабинетов, лабораторий	Перечень оборудования и технических средств обучения
1	Лекционная аудитория	Мультимедийный проектор, интерактивная доска
2	Компьютерный класс	Компьютеры МХР Pentium, мониторы LG), принтеры, мультимедиа проектор –1. Терминалы подключения к сети Internet.

7. ОЦЕНКА КОМПЕТЕНЦИЙ ПО ИЗУЧАЕМОЙ ДИСЦИПЛИНЕ

№	Критерии оценки			
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
ЗНАТЬ				
1	<p>Не знает основные понятия и определения, используемые в методах машинного обучения, истории развития методов машинного обучения, базовых алгоритмов решения задач, машинного обучения, основных программно-информационных ресурсов, используемых в методах машинного обучения.</p>	<p>Не знает точных определений в области анализа многомерных данных и особенностей построения регрессии по многомерным данным, множественной линейной регрессии.</p> <p>Знает базовые алгоритмы решения задач машинного обучения и базовые положения фундаментальных разделов системного анализа в объеме, необходимом для обработки информации и анализа данных в прикладной области. Не знает программно-информационные ресурсы в методах машинного обучения.</p>	<p>Студент знает основные понятия и определения технологии, используемые в методах машинного обучения.</p> <p>Студент усвоил основное содержание материала дисциплины, но имеет пробелы в изложении материала. Усвоил принципы и методы проведения исследований в области информационных систем и проведения вычислительного эксперимента. Знает методы решения задач машинного обучения в программной среде SiLab</p>	<p>Студент знает, понимает, выделяет главные положения в изученном материале и способен дать краткую характеристику основным идеям проработанного материала дисциплины.</p> <p>Знает основные понятия и определения технологии, используемые в методах машинного обучения.</p> <p>Показывает глубокое знание и понимание сущности работы и настройки нейронных сетей.</p>
УМЕТЬ				
2	<p>Студент не умеет проводить анализ многомерных данных, определять корреляционные и причинно-следственные связи.</p>	<p>Студент испытывает затруднения при систематизированном изложении теории обучения нейронных сетей</p> <p>Студент непоследовательно излагает корреляционные и</p>	<p>Студент умеет самостоятельно анализировать многомерные данные, корреляционные и причинно-следственные связи.</p> <p>Студент умеет использовать методы машинного обучения в</p>	<p>Студент умеет формулировать и доказывать наиболее важные результаты в прикладных областях; применять численные методы для решения задач машинного обучения; программно реализовать</p>

		причинно-следственные связи	практических задачах.	вычислительный эксперимент посредством языков программирования или с использованием специализированных пакетов прикладных программ машинного обучения; разрабатывать алгоритмы решения конкретных задач Студент умеет анализировать элементы, устанавливать связи между ними в различных задачах машинного обучения.
ВЛАДЕТЬ				
3	Студент не владеет навыками постановки задачи; навыками работы с библиографическими источниками информации; навыками решения поставленных задач в предметной области в рамках выбранного профиля. Студент не владеет навыками построения нейронных сетей, методами обучения нейросетей.	Студент владеет навыками постановки задачи; навыками работы с библиографическими источниками информации; навыками решения поставленных задач в предметной области в рамках выбранного профиля. Студент не владеет навыками построения нейронных сетей, графовых моделей в задачах машинного обучения.	Студент владеет знаниями всего изученного материала, владеет навыками построения нейронных сетей, графовых моделей, но допускает незначительные ошибки. Студент владеет навыками постановки задачи; навыками работы с библиографическими источниками информации; навыками решения поставленных задач в предметной области в рамках выбранного профиля в задачах машинного обучения.	Студент владеет понятийным аппаратом и методологической базой в методах машинного обучения, владеет навыками построения нейронных сетей, графовых моделей. Студент владеет концептуально-понятийным аппаратом, научным языком и терминологией построения нейронных сетей, анализом многомерных данных, анализом корреляционных и причинно-следственных связей в задачах машинного обучения.
	Компетенция или ее часть не сформирована	Компетенция или ее часть сформирована на базовом уровне	Компетенция или ее часть сформирована на среднем уровне	Компетенция или ее часть сформирована на высоком уровне

8. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Интерактивные образовательные технологии, используемые в аудиторных занятиях и самостоятельной работе обучающихся не предусмотрены.

9. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

9.1. Организация входного, текущего и промежуточного контроля обучения

Входное тестирование – не предусмотрено.

Текущий контроль – опрос, реферат, отчет о практической работе.

Промежуточная аттестация – экзамен.

9.2. Тематика рефератов, проектов, творческих заданий, эссе и т.п.

1. Работа с типами данных в языке Python.
2. Введение в массивы библиотеки NumPy.
3. Выполнение вычислений над массивами библиотеки NumPy.
4. Операции над данными в библиотеке Pandas.
5. Визуализация с помощью библиотеки Matplotlib.
6. Библиотека Scikit-Learn.
7. Смеси Гауссовых распределений.
8. Ядерная оценка плотности распределения.
9. Метод опорных векторов. Оптимальная разделяющая гиперплоскость.
10. Случаи линейной делимости и отсутствия линейной делимости.

Кусочно-линейная функция потерь.

11. Задача квадратичного программирования и двойственная задача.

12. Понятие опорных векторов. Линейные методы классификации.

13. Градиентные методы. Линейный классификатор, связь с методом максимума правдоподобия.

14. Метод стохастического градиента и частные случаи: адаптивный линейный элемент ADALINE, персептрон Розенблатта, правило Хэбба.

15. Метрические методы классификации. Метод ближайших соседей и его обобщения.

16. Постановка задач обучения по прецедентам.

17. Типы задач: классификация, регрессия, прогнозирование, кластеризация.

Примеры прикладных задач.

18. Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества.

19. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных.

20. Полигон алгоритмов классификации.

21. CRISP-DM — межотраслевой стандарт ведения проектов интеллектуального анализа данных.

22. Понятие логической закономерности. Решающие списки и деревья.

Объединение в решающие леса.

23. Сингулярное разложение, метод главных компонент.

24. Наивная байесовская классификация.

25. Машинное обучение с учителем и обучение без учителя.

9.3. Курсовая работа

Не предусмотрено.

9.4. Вопросы к зачету

Не предусмотрено.

9.5. Вопросы к экзамену

1. Основные понятия. Определение предмета машинного обучения. Примеры задач и областей приложения. Образы и признаки.
2. Типы задач предсказания. Регрессия. Таксономия. Классификация. Типы ошибок классификации. Обобщающая способность классификатора.
3. Принцип минимизации эмпирического риска. Недообучение. Переобучение. Статистический, нейросетевой и структурно-лингвистический подходы к распознаванию образов.
4. Структура типичной системы распознавания образов. Цикл построения системы распознавания образов.
5. Классификация. Общие принципы. Этапы классификации. Алгоритмы обучения классификаторов с учителем и без учителя. Дискриминантный анализ. Геометрическая интерпретация задачи классификации.
6. Проективный подход. Метрики в пространстве признаков. Евклидово расстояние. Расстояние Махаланобиса. Ошибки первого и второго рода. Чувствительность и избирательность.
7. Кривая мощности критерия классификации. ROC-кривые. Проверка классификатора. Проверка тестовой выборкой. Перекрестная проверка. Оценка информативности признаков.
8. Основные методы машинного обучения. Байесовская классификация. Условная вероятность. Формула полной вероятности. Формула Байеса. Статистическое распознавание образов. Задача классификации спама. Критерий отношения правдоподобия. Байесовский риск. Критерий Байеса.
9. Критерий максимального правдоподобия. Многоклассовые байесовские классификаторы. Байесовские классификаторы для нормально распределенных классов при различной структуре матрицы ковариации
10. Оценивание функций распределения. Параметрическое оценивание. Метод максимума правдоподобия. Байесовское оценивание. Непараметрическое оценивание. Распознавание рукописных цифр с помощью наивного байесовского
11. Деревья решений. Основные понятия. Классы решаемых задач: описание данных, классификация, регрессия. Общий алгоритм построения дерева решений. Критерии выбора наилучшего атрибута: прирост информации, относительный прирост информации, индекс Гини.
12. Правила останова разбиения дерева. Обрезание дерева. Алгоритм ID3. Переобучение деревьев решений. Обработка непрерывных атрибутов. Обучение на данных с пропусками. Программное обеспечение для построения деревьев решений. Распознавание спамовых писем с помощью деревьев решений
13. Анализ многомерных данных. Корреляционные и причинно-следственные связи. Корреляция признаков и структура данных.
14. Метод главных компонент как декомпозиция матрицы данных. Матрица счетов. Матрица нагрузок. Матрица ошибок. Объясненная и остаточная вариация в данных. Графическая интерпретация метода главных компонент. Критерии выбора количества главных компонент. Понижение размерности признакового пространства методом главных компонент
15. Регрессия. Метод наименьших квадратов. Теорема Гаусса-Маркова. Обобщенный метод наименьших квадратов. Рекурсивный метод наименьших квадратов. Анализ регрессионных остатков.

