

ФГБОУ ВО «БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Утверждено:  
на заседании кафедры ИТиКМ  
протокол № 7 от 28.02.2022 г.

Зав. кафедрой  Болотнов А.М.

Согласовано:  
Председатель УМК  
ФМиИТ

 Ефимов А.М.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

дисциплина Машинное обучение и анализ данных

Часть, формируемая участниками образовательных отношений

**программа бакалавриата**

Направление подготовки

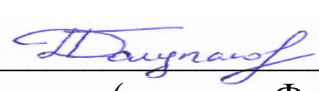
09.03.03 Прикладная информатика

Направленность (профиль) подготовки

"Информационные и вычислительные технологии"

Квалификация

бакалавр

Разработчик (составитель) доцент, к.т.н., доцент	 / <u>Полупанов Д.В.</u> (подпись, Фамилия И.О.)
---	---

Для приема: 2022

Уфа 2022 г.

Составитель / составители: к.т.н., доцент Полупанов Д.В.

Рабочая программа дисциплины *утверждена* на заседании кафедры информационных технологий и компьютерной математики, протокол № 6 от 26 января 2021 г.

Рабочая программа дисциплины *актуализирована* на основании Приказа Минобрнауки России от 26.11.2020 № 1456 "О внесении изменений в федеральные государственные образовательные стандарты высшего образования", Приказа БашГУ от 09.06.2021 № 770 "О внесении изменений в образовательные программы высшего образования – программы бакалавриата, программы специалитета и программы магистратуры". Ученый совет факультета математики и информационных технологий – протокол № 8 от 15.06.2021 г.

Заведующий кафедрой



Болотнов А.М.

## **Список документов и материалов**

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций
2. Цель и место дисциплины в структуре образовательной программы
3. Содержание рабочей программы (объем дисциплины, типы и виды учебных занятий, учебно-методическое обеспечение самостоятельной работы обучающихся)
4. Фонд оценочных средств по дисциплине
  - 4.1. Перечень компетенций и индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.
  - 4.2. Типовые контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине.
5. Учебно-методическое и информационное обеспечение дисциплины
  - 5.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины
  - 5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины, включая профессиональные базы данных и информационные справочные системы
6. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

## 1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций

По итогам освоения дисциплины обучающийся должен достичь следующих результатов обучения:

Категория (группа) компетенций (при наличии ОПК)	Формируемая компетенция (с указанием кода)	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
	<i>ПК-2: Способность использовать современные методы разработки и реализации алгоритмов конк-ретных математических моделей на базе языков программирования и пакетов прикладных программ</i>	<i>ПК-2.1. Знает современные методы разработки алгоритмов, математические модели, языки программирования и пакеты прикладных программ</i>	<i>Демонстрирует знания основных алгоритмов машинного обучения и анализа данных</i>
		<i>ПК-2.2. Умеет разрабатывать алгоритмы математических моделей на базе языков программирования и пакетов прикладных программ.</i>	<i>Реализует с помощью языков программирования алгоритмы обучения машинного обучения и анализа данных</i>
		<i>ПК-2.3. Владеет навыками разработки и алгоритмов на базе языков программирования и пакетов прикладных программ.</i>	<i>Имеет практический опыт разработки и реализации систем машинного обучения и анализа данных на базе языков программирования</i>

## 2. Цель и место дисциплины в структуре образовательной программы

Дисциплина «Машинное обучение и анализ данных» относится к части, формируемой участниками образовательных отношений.

Дисциплина изучается на 4 курсе(ах) в 7 и 8 семестрах.

Целью изучения дисциплины «Машинное обучение и анализ данных» является ознакомление с основами построения, разработки и обучения нейронных сетей и их применения к различным научным и прикладным проблемам.

## 3. Содержание рабочей программы (объем дисциплины, типы и виды учебных занятий, учебно-методическое обеспечение самостоятельной работы обучающихся)

Содержание рабочей программы представлено в Приложении № 1.

## 4. Фонд оценочных средств по дисциплине

### 4.1. Перечень компетенций и индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.

Код и формулировка компетенции ПК-2: *Способность использовать современные методы разработки и реализации алгоритмов конкретных математических моделей на базе языков программирования и пакетов прикладных программ*

Для зачета

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Критерии оценивания результатов обучения	
		«Зачтено»	«Не зачтено»
<i>ПК-2.1. Знает современные методы разработки алгоритмов, математические модели, языки программирования и пакеты прикладных программ</i>	<i>Демонстрирует знания основных алгоритмов машинного обучения и анализа данных</i>	Сформированные, возможно содержащие незначительные пробелы, знания основных алгоритмов машинного обучения и анализа данных	Фрагментарные, неполные, несистематические знания основных алгоритмов машинного обучения и анализа данных
<i>ПК-2.2. Умеет разрабатывать алгоритмы математических моделей на базе языков программирования и пакетов прикладных программ.</i>	<i>Реализует с помощью языков программирования алгоритмы обучения машинного обучения и анализа данных</i>	Сформированное, возможно содержащее незначительные пробелы, умение реализовывать алгоритмы машинного обучения и анализа данных на базе языков программирования	Фрагментарные, неполные, несистематические умения реализовывать алгоритмы машинного обучения и анализа данных на базе языков программирования
<i>ПК-2.3. Владеет навыками разработки и алгоритмов на базе языков программирования и пакетов прикладных программ.</i>	<i>Имеет практический опыт разработки и реализации систем машинного обучения и анализа данных на базе языков программирования</i>	Успешное и систематическое, возможно содержащее незначительные пробелы, применение навыков реализации систем машинного обучения на базе языков программирования	Фрагментарное, неполное, несистематическое владение навыками реализации систем машинного обучения на базе языков программирования

Для экзамена

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Критерии оценивания результатов обучения			
		2 («Не удовлетворительно»)	3 («Удовлетворительно»)	4 («Хорошо»)	5 («Отлично»)
<i>ПК-2.1. Знает современные методы разработки алгоритмов, математические модели, языки программирования и пакеты прикладных программ</i>	<i>Демонстрирует знания основных алгоритмов машинного обучения и анализа данных</i>	Отсутствие знаний или фрагментарные представления об основных алгоритмах машинного обучения и анализа данных	Неполные представления об основных алгоритмах машинного обучения и анализа данных	Сформированные, но содержащие отдельные пробелы представления об основных алгоритмах машинного обучения и анализа данных	Сформированные систематические представления об основных алгоритмах машинного обучения и анализа данных
<i>ПК-2.2. Умеет разрабатывать ал-</i>	<i>Реализует с помощью языков про-</i>	Отсутствие умений или фрагментарные умения реализовывать	В целом успешное, но не систематическое умение реализовывать алгоритмы	В целом успешное, но содержащее отдельные пробелы, умение	Сформированное умение реализовывать алгоритмы машинного обучения и анализа

<i>горитмы математических моделей на базе языков программирования и пакетов прикладных программ.</i>	<i>граммирования алгоритмы обучения машинного обучения и анализа данных</i>	<i>алгоритмы машинного обучения и анализа данных на базе языков программирования</i>	<i>машинного обучения и анализа данных на базе языков программирования</i>	<i>применять в реализовывать алгоритмы машинного обучения и анализа данных на базе языков программирования</i>	<i>данных на базе языков программирования</i>
<i>ПК-2.3. Владеет навыками разработки и алгоритмов на базе языков программирования и пакетов прикладных программ.</i>	<i>Имеет практический опыт разработки и реализации систем машинного обучения и анализа данных на базе языков программирования</i>	<i>Отсутствие или фрагментарное применение навыков реализации систем машинного обучения на базе языков программирования</i>	<i>В целом успешное, но не систематическое применение навыков реализации систем машинного обучения на базе языков программирования</i>	<i>В целом успешное, но содержащее отдельные пробелы применение навыков реализации систем машинного обучения на базе языков программирования</i>	<i>Успешное и систематическое применение навыков реализации систем машинного обучения на базе языков программирования</i>

**4.2. Типовые контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине.**

<b>Код и наименование индикатора достижения компетенции</b>	<b>Результаты обучения по дисциплине</b>	<b>Оценочные средства</b>
<i>ПК-2.1. Знает современные методы разработки алгоритмов, математические модели, языки программирования и пакеты прикладных программ</i>	<i>Демонстрирует знания основных алгоритмов машинного обучения и анализа данных</i>	<i>Групповой и индивидуальный опрос, РГР, тестирование, экзамен</i>
<i>ПК-2.2. Умеет разрабатывать алгоритмы математических моделей на базе языков программирования и пакетов прикладных программ.</i>	<i>Реализует с помощью языков программирования алгоритмы обучения машинного обучения и анализа данных</i>	<i>Лабораторные работы, тестирование, РГР, экзамен</i>
<i>ПК-2.3. Владеет навыками разработки и алгоритмов на базе языков программирования и пакетов прикладных программ.</i>	<i>Имеет практический опыт разработки и реализации систем машинного обучения и анализа данных на базе языков программирования</i>	<i>Лабораторные работы, РГР, тестирование, экзамен</i>

Критериями оценивания при *модульно-рейтинговой системе* являются баллы, которые выставляются преподавателем за виды деятельности (оценочные средства) по итогам изучения модулей (разделов дисциплины), перечисленных в рейтинг-плане дисциплины (*для экзамена*: текущий контроль – максимум 40 баллов; рубежный контроль – максимум 30 баллов,

поощрительные баллы – максимум 10; для зачета: текущий контроль – максимум 50 баллов; рубежный контроль – максимум 50 баллов, поощрительные баллы – максимум 10).

Шкалы оценивания:

(для экзамена:

от 45 до 59 баллов – «удовлетворительно»;

от 60 до 79 баллов – «хорошо»;

от 80 баллов – «отлично».

для зачета:

зачтено – от 60 до 110 рейтинговых баллов (включая 10 поощрительных баллов),

не зачтено – от 0 до 59 рейтинговых баллов).

### Рейтинг – план дисциплины

#### «Машинное обучение и анализ данных»

направление/специальность 09.03.03. «Прикладная информатика»

курс 4, семестр 7

Виды учебной деятельности студентов	Балл за конкретное задание	Число заданий за семестр	Баллы	
			Минимальный	Максимальный
<b>Модуль 1. Базовые методы машинного обучения и анализа данных</b>				
<b>Текущий контроль</b>				
Групповой и индивидуальный опрос	1	20	0	20
Лабораторная работа	5	1	0	5
<b>Рубежный контроль</b>				
Тестовый контроль	1	25	0	25
<b>Модуль 2. Нейронные сети прямого распространения, обучающиеся «с учителем»</b>				
<b>Текущий контроль</b>				
Групповой и индивидуальный опрос	1	15	0	15
Лабораторная работа	5	2	0	10
<b>Рубежный контроль</b>				
Тестовый контроль	1	25	0	25
<b>Поощрительные баллы</b>				
1. Студенческая олимпиада				<b>10</b>
2. Публикация статей				<b>10</b>
3. Задания повышенной сложности				<b>10</b>
<b>Посещаемость (баллы вычитаются из общей суммы набранных баллов)</b>				
1. Посещение лекционных занятий			<b>0</b>	<b>-6</b>
2. Посещение практических (семинарских, лабораторных занятий)			<b>0</b>	<b>-10</b>

### Рейтинг – план дисциплины

#### «Машинное обучение и анализ данных»

направление/специальность 09.03.03. «Прикладная информатика»

курс 4, семестр 8

Виды учебной деятельности студентов	Балл за конкретное задание	Число заданий за семестр	Баллы	
			Минимальный	Максимальный
<b>Модуль 3. Обучение нейронных сетей «без учителя». Сети и карты Кохонена</b>				
<b>Текущий контроль</b>				
Групповой и индивидуальный опрос	1	10	0	10
Лабораторная работа	5	2	0	10
<b>Рубежный контроль</b>				
Тестовый контроль	1	15	0	15
<b>Модуль 4. Радиально-базисные нейронные сети, сети Хопфилда и другие архитектуры</b>				
<b>Текущий контроль</b>				
Групповой и индивидуальный опрос	1	5	0	5
Лабораторная работа	5	3	0	15

<b>Рубежный контроль</b>				
Тестовый контроль	1	15	0	15
<b>Поощрительные баллы</b>				
1. Студенческая олимпиада				<b>10</b>
2. Публикация статей				<b>10</b>
3. Задания повышенной сложности				<b>10</b>
<b>Посещаемость (баллы вычитаются из общей суммы набранных баллов)</b>				
1. Посещение лекционных занятий			<b>0</b>	<b>-6</b>
2. Посещение практических (семинарских, лабораторных занятий)			<b>0</b>	<b>-10</b>
<b>Итоговый контроль</b>				
Экзамен			<b>0</b>	<b>30</b>

### Экзаменационные билеты

*Структура экзаменационного билета:*

Экзаменационный билет содержит в себе два вопроса. Первый вопрос – теоретический, второй – практический.

*Примерные вопросы для экзамена:*

1. Понятие искусственного нейрона. Основные элементы искусственного нейрона. Свойства нейронных сетей.
2. Формальная математическая модель нейронной сети.
3. Теорема Колмогорова.
4. Теорема Стоуна-Вейерштрасса.
5. Обобщение теоремы Стоуна. Теорема А.Н. Горбаня о полноте.
6. Нейронные сети как универсальный аппроксиматор функций. Теорема о вычислении функций нейронными сетями.
7. Персептрон Розенблата. Задача обучения однослойного персептрона.
8. Проблема «Исключающего ИЛИ» и её нейросетевое решение
9. Многослойный персептрон. Основные виды активационных функций.
10. Обучение многослойного персептрона. Алгоритм обратного распространения ошибки
11. Теоремы об ошибке обучения в алгоритме обратного распространения ошибки
12. Другие алгоритмы обучения многослойного персептрона. Градиентные методы. Конструктивные алгоритмы
13. Основные виды задач анализа данных. Классификация, кластеризация, регрессия.
14. Деревья решений.
15. Модель Липпмана-Хемминга. Принцип «Победитель забирает все». Обучение нейронной сети Липпмана-Хемминга.
16. Обучение на основе самоорганизации. Карта самоорганизации Кохонена.
17. Нейронная сеть встречного распространения Хехт-Нильсена. Обучение нейронных сетей встречного распространения.
18. Гибридная нейронная сеть самоорганизации на основе конкуренции. Архитектура гибридной сети, обучение гибридной сети.
19. Нейронная сеть Хопфилда. Модель функционирования нейронной сети Хопфилда
20. Правило обучения Хебба. Модификация правила Хебба.
21. Двухнаправленная ассоциативная нейронная сеть Барта Коско.
22. Применение модели Хопфилда. Задача распознавания образов. Задача комбинаторной оптимизации.
23. Радиально-базисные нейронные сети как особый класс нейронных сетей. Радиальные активационные функции
24. Теорема Ковера о распознавании образов
25. Задача аппроксимации функции радиально-базисной сетью.
26. Архитектура радиально-базисной сети. Гиперрадиальная нейронная сеть.



27. Методы обучения радиально-базисных нейронных сетей. Гибридный алгоритм обучения.
28. Методы обучения радиально-базисных нейронных сетей. Градиентный алгоритм обучения

*Образец экзаменационного билета:*

Федеральное государственное бюджетное образовательное учреждение  
высшего профессионального образования

**БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

*Факультет математики и информационных технологий  
Кафедра информационных технологий и компьютерной математики*

4 курс, 8 семестр, 20\_/20\_ учебный год

Дисциплина *Машинное обучение и анализ данных*  
Направление *09.03.03 - Прикладная информатика*

**ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 1**

1. Алгоритмы построения деревьев решений
2. Нейрон  $j$  получил на вход сигнал от четырех других нейронов уровня возбуждения, значения которых равны 10, -20, 5, 4 и соответствующие веса связей равны 0.8, 0.5, 0.7 и -0.5 соответственно. Вычислите сигнал на выходе данного нейрона в случае если функция активации нейронов есть гиперболический тангенс (параметр  $\alpha=0,5$ ).

*Зав. кафедрой ИТ и КМ*

*А.М. Болотнов*

***Критерии оценки (в баллах):***

- **25-30 баллов** *выставляется студенту, если студент дал полные, развернутые ответы на все теоретические вопросы билета, продемонстрировал знание функциональных возможностей, терминологии, основных элементов, умение применять теоретические знания при выполнении практических заданий. Студент без затруднений ответил на все дополнительные вопросы. Практическая часть работы выполнена полностью без неточностей и ошибок;*

- **17-24 баллов** *выставляется студенту, если студент раскрыл в основном теоретические вопросы, однако допущены неточности в определении основных понятий. При ответе на дополнительные вопросы допущены небольшие неточности. При выполнении практической части работы допущены несущественные ошибки;*

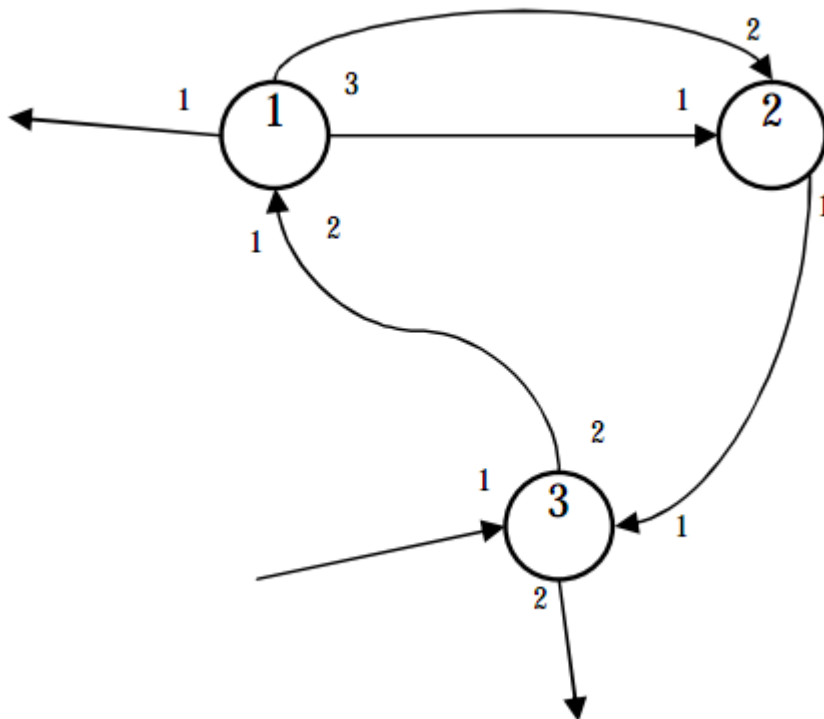
- **10-16 баллов** *выставляется студенту, если при ответе на теоретические вопросы студентом допущено несколько существенных ошибок в толковании основных понятий. Логика и полнота ответа страдают заметными изъянами. Заметны пробелы в знании основных методов. Теоретические вопросы в целом изложены достаточно, но с пропусками материала. Имеются принципиальные ошибки в логике построения ответа на вопрос. Студент не решил задачу или при решении допущены грубые ошибки;*

- **0-10 баллов** *выставляется студенту, если он отказался от ответа или не смог ответить на вопросы билета, ответ на теоретические вопросы свидетельствует о непонимании и крайне неполном знании основных понятий и методов. Обнаруживается отсутствие навыков применения теоретических знаний при выполнении практических заданий. Студент не смог ответить ни на один дополнительный вопрос.*

**Образцы заданий для группового опроса (контрольной работы)**

1. Укажите возможные значения весов и порога однослойного персептрона с двумя входами, реализующего логическую функцию OR.
2. Выведите производную активационной функции гиперболический тангенс через саму функцию

3. Нейрон  $j$  получил на вход сигнал от четырех других нейронов уровня возбуждения, значения которых равны  $-10, 20, 15, 4$  и соответствующие веса связей равны  $0.5, 0.8, -0.7$  и  $-0.5$  соответственно. Вычислите сигнал на выходе данного нейрона в случае если функция активации нейронов есть логистическая сигмоида (параметр  $\alpha=0,5$ ).
4. Можно ли применять гауссову функцию активации при методе обучения обратного распространения ошибки?
5. Выпишите матрицу внешних связей, если нейроны связаны, как показано на рисунке



6. Дайте формулировку теоремы об универсальной аппроксимации нейросетями. Какой вывод из этой теоремы нельзя сделать?
7. В чем заключается "обучение без учителя"?

*Описание методики оценивания:*

- «Отлично» выставляется студенту, если более 90% заданий выполнены верно.
- «Хорошо» выставляется студенту, если более 75% заданий выполнены верно.
- «Удовлетворительно» выставляется студенту, если более половины заданий выполнены верно.
- «Неудовлетворительно» выставляется студенту, если менее половины заданий выполнены верно.

### Задания для лабораторных работ

Лабораторные работы посвящены программной реализации различных нейросетевых архитектур и решению с их помощью прикладных задач.

*Лабораторная работа 1. Построение дерева решений*

Для своего варианта описать структуру исходных данных:

- общие характеристики массива данных: предметная область, количество записей
- входные параметры: названия и типы
- выходной класс: название и значения

Провести серию экспериментов с построением и тестированием деревьев решений

В качестве выборки взять известную базу данных «пассажиры Титаника». суть задачи состоит в том, чтобы с помощью методов машинного обучения построить модель, которая прогнозировала бы спасется человек или нет. К задаче прилагаются 2 файла:

*train.csv* - набор данных на основании которого будет строиться модель (обучающая выборка)

*test.csv* - набор данных для проверки модели

PassengerId - идентификатор пассажира

Survival - поле в котором указано спасся человек(1) или нет (0)

Pclass - содержит социально-экономический статус:

высокий

средний

низкий

Name - имя пассажира

Sex - пол пассажира

Age - возраст

SibSp - содержит информацию о количестве родственников 2-го порядка (муж, жена, братья, сестры)

Parch - содержит информацию о количестве родственников на борту 1-го порядка (мать, отец, дети)

Ticket - номер билета

Fare - цена билета

Cabin - каюта

Embarked - порт посадки

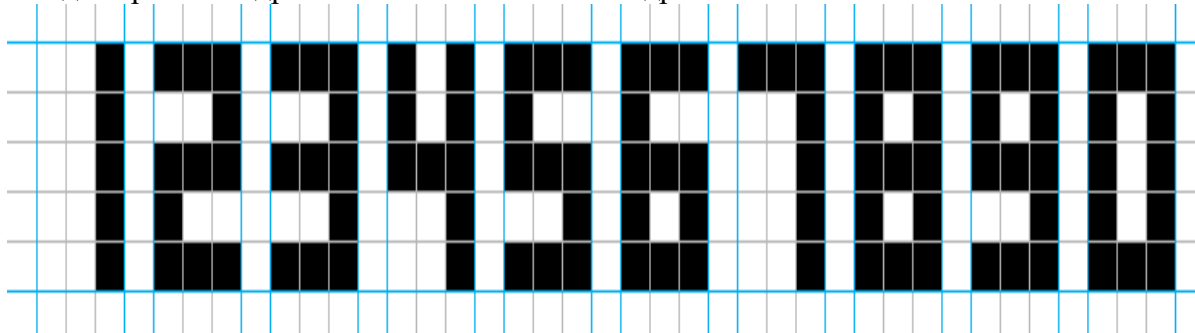
C - Cherbourg

Q - Queenstown

S - Southampton

*Лабораторная 2 Распознавание цифр с помощью перцептрона Розенблатта.*

Описание задач. Требуется распознать черно-белые цифры от 0 до 9, представленные в виде черных квадратиков в табличке 3×5 квадратов.



Каждая цифра представляет собой всего пятнадцать квадратиков, причем только двух возможных цветов. За белый квадратик отвечает 0, а черный квадратик – 1. Можно представить цифры в виде последовательности нулей и единиц

0 → 111101101101111

1 → 001001001001001

2 → 111001111100111

3 → 111001111001111

4 → 101101111001001

5 → 111100111001111

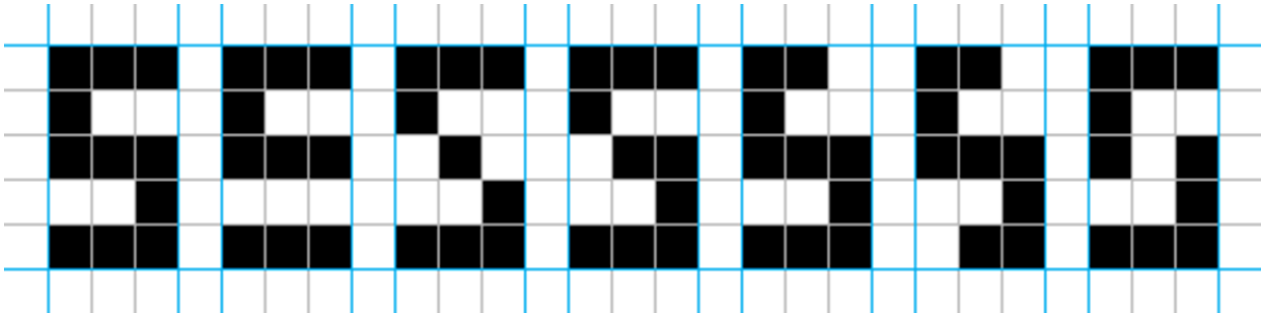
6 → 111100111101111

7 → 111001001001001

8 → 111101111101111

9 → 111101111001111

Необходимо разработать программу, эмулирующую перцептрон Розенблатта с 15 входами и одним выходом, позволяющего распознать по одной цифре. В качестве примера приведем слегка искаженные варианты цифры 5



```

1111001110001111
1111000100011111
1111000110011111
1101001110011111
1101001110010111
1111001010011111

```

### *Лабораторная работа 3. Оценка итогов выборов президентов США с помощью нейронных сетей*

Описание задачи. Данная задача давно является тестовой при создании новых программных реализаций нейронных сетей, обучающихся с учителем. Оказывается, чтобы ответить, кто станет президентом США – кандидат от правящей партии или кандидат от оппозиционной партии, достаточно знать ответы на нижеследующие 12 вопросов.

Разумеется, предполагается, что предвыборные компании кандидатов отработаны добросовестно и все участники сделали все возможное.

Вот эти вопросы:

- 1) Правящая партия у власти более 1 срока?
- 2) Правящая партия получила больше 50 % на прошлых выборах?
- 3) В год выборов была активна третья партия?
- 4) Была серьезная конкуренция при выдвижении кандидата от правящей партии?
- 5) Кандидат от правящей партии был президентом в год выборов?
- 6) Был ли год выборов временем спада или депрессии?
- 7) Был ли рост среднего национального валового продукта на душу населения более 2,1%?
- 8) Произвел ли правящий президент существенные изменения в политике?
- 9) Во время правления были существенные социальные волнения'?
- 10) Администрация правящей партии виновна в серьезной ошибке или скандале?
- 11) Кандидат правящей партии - национальный герой?
- 12) Кандидат оппозиционной партии - национальный герой?

Данные представлены в файле lab3.txt.

Требуется построить программы, реализующую многослойный персептрон, для предсказания итогов выборов.

### *Лабораторная работа 4. Кластеризация «ирисов Фишера».*

В 1936 году Р. Фишер предложил метод дискриминантного анализа, проиллюстрированный им на базе данных, позже названной «ирисы Фишера». Это данные о 150 экземплярах ириса, по 50 экземпляров из трёх видов:

- Ирис щетинистый (*Iris setosa*);
- Ирис разноцветный (*Iris versicolor*);
- Ирис виргинский (*Iris virginica*).

Для каждого экземпляра измеряются четыре характеристики:

- x1 длина чашелистика ;
- x2 ширина чашелистика;
- x3 длина лепестка;
- x4 ширина лепестка.

Традиционным стало испытывать новые методы классификации и кластеризации на данной выборке. Данные представлены в файле lab4.txt.

Требуется построить программу, реализующую самоорганизующуюся карту Кохонена. Осуществить кластеризацию объектов и интерпретацию результатов.

*Лабораторная работа 5. Сегментация клиентов телекоммуникационной компании с помощью самоорганизующихся карт Кохонена*

Описание задачи. Телекоммуникационная компания, предоставляющая услуги мобильной связи, ставит перед собой задачу сегментации абонентской базы для улучшения положения на рынке.

Данные за последние месяцы представляют собой таблицу со следующими полями:

- Возраст абонента;
- Среднемесячный расход;
- Средняя продолжительность разговора;
- Количество звонков днем за месяц;
- Количество звонков вечером за месяц;
- Количество звонков ночью за месяц;
- Звонки в другие города;
- Звонки в другие страны;
- Доля звонков на стационарные телефоны;
- Количество SMS в месяц;

Отобраны только активные абоненты, которые регулярно пользовались услугами сотовой связи за последний период.

Исходные данные представлены в файле lab4.txt

Требуется построить программу, реализующую самоорганизующуюся карту Кохонена. Осуществить кластеризацию объектов и интерпретацию результатов.

*Лабораторная работа 6. Оценка итогов выборов президентов США с помощью радиально-базисных нейронных сетей*

Описание задачи. Работа совпадает с лабораторной работой № 3, в отличие от которой требуется программная реализация другого типа нейронной сети – с радиально-базисной функцией активации. Данная задача давно является тестовой при создании новых программных реализаций нейронных сетей, обучающихся с учителем. Оказывается, чтобы ответить, кто станет президентом США – кандидат от правящей партии или кандидат от оппозиционной партии, достаточно знать ответы на нижеследующие 12 вопросов.

Разумеется, предполагается, что предвыборные кампании кандидатов отработаны добросовестно и все участники сделали все возможное.

Вот эти вопросы:

- 1) Правящая партия у власти более 1 срока?
- 2) Правящая партия получила больше 50 % на прошлых выборах?
- 3) В год выборов была активна третья партия?
- 4) Была серьезная конкуренция при выдвижении кандидата от правящей партии?
- 5) Кандидат от правящей партии был президентом в год выборов?
- 6) Был ли год выборов временем спада или депрессии?
- 7) Был ли рост среднего национального валового продукта на душу населения более 2,1%?
- 8) Произвел ли правящий президент существенные изменения в политике?
- 9) Во время правления были существенные социальные волнения'?
- 10) Администрация правящей партии виновна в серьезной ошибке или скандале?
- 11) Кандидат правящей партии - национальный герой?
- 12) Кандидат оппозиционной партии - национальный герой?

Данные представлены в файле lab3.txt.

Требуется построить программы, реализующую радиально-базисную нейронную сеть, для предсказания итогов выборов.

*Лабораторная работа 7. Оценка эффективности массовой рассылки клиентам с помощью радиально-базисной нейронной сети*

Описание задачи.. Торговая компания, осуществляющая продажу товаров, располагает информацией о своих клиентах и их покупках. Компания провела рекламную рассылку 135-4

клиентам и получила отклик в 14,5% случаев. Необходимо построить модель отклика и проанализировать результаты, с целью предложить способы минимальных издержек на новые рассылки.

Имеются данные о клиентах, содержащие такие сведения как пол, возраст, сколько лет данная персона клиент компании, суммарная стоимость заказов клиента, общее число покупок, факты обращения в службу поддержки. Также известны расходы на одну рассылку, издержки на обслуживание клиента, ожидаемая выручка с одного заказа. Данные представлены в файле lab7.txt.

Требуется построить программу, реализующую радиально-базисную нейронную сеть, для предсказания отклика рассылки.

#### *Лабораторная работа 8*

Требуется выполнить задания лабораторных работ 2 – 7 (на выбор), используя какую-нибудь из нейросетевых библиотек Python

#### *РГР*

РГР представляет собой разработку относительно универсального программного продукта, реализующего ту или иную нейросетевую архитектуру. В отличие от лабораторных работ, ориентированных на решение конкретной задачи, здесь данные могут быть любыми – произвольного объема и размерности. Программный продукт должен обладать полноценным графическим интерфейсом, предусматривать возможность чтения данных из файла и запись результатов в файл, выбор конкретных параметров нейронных сетей – например числа слоев, нейронов в них, активационных функций, шага обучения и т.п.

#### *Тестирование*

*Примерные вопросы для тестового контроля*

#### **1. Из теоремы Стоуна можно сделать следующий вывод:**

1. с помощью операции суперпозиции и линейных комбинаций из любых нелинейных элементов можно получить любую функцию многих переменных с любой заданной точностью
2. можно получить произвольную непрерывную функцию многих переменных с помощью операций сложения, умножения и суперпозиции из непрерывных функций двух переменных
3. с помощью нейронной сети можно задать любую непрерывную функцию, с любой, заданной точностью
4. можно получить любую непрерывную функцию многих переменных с помощью операций сложения, умножения и суперпозиции из непрерывных функций одного переменного

#### **1. К классу нейросетей, обучающихся "с учителем" относятся**

**а. многослойные персептроны**

**3. карты Кохонена**

**в. радиально-базисные сети**

**г. сети Хопфилда**

**д. персептроны Розенблата**

**Верными ответами являются:**

1. а, в, д

2. а, б, г

3. а, в, д

4. все ответы являются верными

#### **4. Многослойные персептроны относятся**

1. к нейронным обучающимся "без учителя"

2. к рекуррентным нейронным сетям

3. к нейронным сетям обучающимся "с учителем"

4. к нейронным сетям смешанной стратегии обучения

#### **5. С помощью классического персептрона Розенблата можно решить задачу**

**а) линейной регрессии**

**б) классификации для логической функций AND**

**в) классификации для логической функции OR**

#### г) линейного разделения двух классов

Верными ответами являются

1. б, в
- 2 а, г
3. б, в, г
4. все ответы являются верными

Всего предлагается два теста по 25 вопросов в первом семестре и два теста по 15 вопросов во втором семестре.

#### Самостоятельная работа

Самостоятельная работа студентов заключается в подготовке к занятиям и выполнении зачетных заданий с использованием рекомендованной учебно-методической литературы. В качестве дополнительных заданий предлагаются следующие темы докладов или рефератов:

1. Представление нейронных сетей с помощью ориентированных графов
2. Алгоритмы М.А. Айзермана и Э.М. Бравермана обучения нейронных сетей
3. Метод двойственности обучения нейронных сетей
4. Теорема о сходимости персептрона
5. Теорема об универсальной аппроксимации. Результаты G. Cybenko, K. Funahashi, K. Hornik и др.
6. Устойчивость нейронных сетей
7. Теорема Ковера о разделимости образов. Выводы из теоремы
8. Радиально-базисные функции. Теорема Мичелли
9. Универсальная теорема об аппроксимации для радиально-базисных нейронных сетей
10. Нейродинамическое программирование. Теоремы Ляпунова
11. Теорема Козна-Госсберга. Модель Хопфилда как частный случай теоремы Козна-Гроссберга
12. Рекуррентные нейронные сети, имитирование конечных автоматов
13. Конструктивные алгоритмы обучения нейронных сетей
14. Нейросетевые нечеткие системы
15. Аппаратная реализация нейрокомпьютеров
16. Самоорганизующиеся нейронные сети, конгитрон и неоконгитрон. Алгоритм обучения
17. Стохастические методы обучения нейронных сетей
18. Надежность нейронных сетей
19. Библиотеки машинного обучения для языка Python (на выбор)
20. Алгоритмы построения нечетких деревьев решений
21. Алгоритм «случайный лес»

## 5. Учебно-методическое и информационное обеспечение дисциплины

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

#### Основная литература:

1. Коэльо, Л.П. Построение систем машинного обучения на языке Python [Электронный ресурс] / Л.П. Коэльо, В. Ричарт. — Электрон. дан. — Москва : ДМК Пресс, 2016. — 302 с. — Режим доступа: <https://e.lanbook.com/book/82818>. — Загл. с экрана.
2. . Галушкин, А.И. Нейронные сети: основы теории [Электронный ресурс] : учеб. пособие — Электрон. дан. — Москва : Горячая линия-Телеком, 2012. — 496 с. — Режим доступа: <https://e.lanbook.com/book/5144>. — Загл. с экрана.

#### Дополнительная литература:

3. Ежов, А.А. Нейрокомпьютинг и его применения в экономике и бизнесе / А.А. Ежов, С.А. Шумский ; Национальный Открытый Университет "ИНТУИТ". - М. : Интернет-Университет Информационных Технологий, 2006. - 268 с. : ил. ; То же [Электронный ресурс]. - URL: [/biblioclub.ru/index.php?page=book&id=233761](http://biblioclub.ru/index.php?page=book&id=233761)
4. Барский, А.Б. Введение в нейронные сети [Электронный ресурс] : учеб. пособие — Электрон. дан. — Москва : , 2016. — 358 с. — Режим доступа: <https://e.lanbook.com/book/100684>. — Загл. с экрана.
5. Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер.с польск. И.Д.Рудинского [Электронный ресурс] : учеб. пособие / Д. Рутковская, М. Пилиньский, Л. Рутковский. — Электрон. дан. — Москва : Горячая линия-Телеком, 2013. — 384 с. — Режим доступа: <https://e.lanbook.com/book/11843>. — Загл. с экрана..

### 5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины, включая профессиональные базы данных и информационные справочные системы

<http://neuronus.com>

<https://basegroup.ru>

Microsoft Office

Python 3 – свободно распространяемый интерпретатор языка программирования Python

PyCharm - Свободно-распространяемая среда для разработки программного обеспечения

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

<p><b>1. Учебная аудитория для проведения занятий лекционного типа:</b> аудитория № 511(физмат корпус- учебное), аудитория № 531(физмат корпус- учебное).</p> <p><b>2. Учебная аудитория для проведения занятий семинарского типа:</b> аудитория № 511 (физмат корпус- учебное), № 531 (физмат корпус- учебное), аудитория № 522 (физмат корпус- учебное).</p>	<p><b>Аудитория №511</b> Учебная мебель, доска настенная меловая, мультимедиа проектор Mitsubishi EX 320U 3D 2.4кг., экран на штативе Draper Diplomat (1:1) 84/84* 213*213 MW, компьютер в составе: системный блок DEPO 460MD/3-540/T500G/DVD-RW, монитор 20".</p> <p><b>Аудитория №531</b> Учебная мебель, доска настенная меловая, мультимедиа-проектор Sony VPL-EX120, XGA, 2600 ANSI, 3,2 кг, потолочное крепление для проектора (2101068302), доска аудитор. ДА32.</p> <p><b>Читальный зал №2</b></p>	<ol style="list-style-type: none"><li>1. Windows 8 Russian. Windows Professional 8 Russian Upgrade. Договор № 104 от 17.06.2013 г. Лицензии бессрочные.</li><li>2. Microsoft Office Standard 2013 Russian. Договор № 114 от 12.11.2014 г. Лицензии бессрочные.</li><li>3. Python 3 (лицензия Python Software Foundation License, ссвободное программное обеспечение).</li></ol>
--	--	---



<p><b>3. Учебная аудитория для проведения групповых и индивидуальных консультаций:</b> аудитория № 511 (физмат корпус-учебное), аудитория № 531 (физмат корпус-учебное).</p> <p><b>4. Учебная аудитория для текущего контроля и промежуточной аттестации:</b> аудитория № 511 (физмат корпус-учебное), аудитория № 531 (физмат корпус-учебное)</p> <p><b>5. Помещения для самостоятельной работы:</b> читальный зал №2 (физмат корпус-учебное).</p>	<p>Учебная мебель, учебно-наглядные пособия, стенд по пожарной безопасности, моноблоки стационарные – 8 шт, принтер – 1 шт., сканер – 1 шт.</p> <p><b>Аудитория №522 (лаборатория компьютерного моделирования)</b></p> <p>Учебная мебель, доска, персональный компьютер Lenovo Think Centre A70z Intel Pentium E 5800, 320 Gb, 19" – 13 шт., кондиционер Lessar LS/LU-H24KB2.</p>	
---	---	--

ФГБОУ ВО «БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

**СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ**

Дисциплины «Машинное обучение и анализ данных» на 7 семестр

очная

форма обучения

<b>Вид работы</b>	<b>Объем дисциплины</b>
Общая трудоемкость дисциплины (з.е. / часов)	3/108
Учебных часов на контактную работу с преподавателем:	
лекций	18
практических/ семинарских	
лабораторных	18
других (групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие работу обучающихся с преподавателем) (ФКР)	0,2
из них, предусмотренные на выполнение курсовой работы / курсового проекта	
Учебных часов на самостоятельную работу обучающихся (СР)	17,8
из них, предусмотренные на выполнение курсовой работы / курсового проекта	
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	

Форма(ы) контроля:

Зачет 7 семестр

№ п/п	Тема и содержание	Форма изучения материалов: лекции, практические занятия, семинарские занятия, лабораторные работы, самостоятельная работа и трудоемкость (в часах)				Задания по самостоятельной работе студентов	Форма текущего контроля успеваемости (коллоквиумы, контрольные работы, компьютерные тесты и т.п.)
		ЛК	ПР/СЕМ	ЛР	СР		
1	2	3	4	5	6	8	9
1.	Введение в машинное обучение. Классы систем машинного обучения. Анализ данных, основные понятия и определения. Основные задачи, методы и алгоритмы.	2			7,8	Проработка лекционного материала, литературных источников.	Групповой и индивидуальный опрос, тестирование
2.	Байесовский классификатор. Деревья решений.	2		6	8	Проработка лекционного материала, литературных источников, подготовка к лабораторной работе	Групповой и индивидуальный опрос, тестирование, лабораторная работа
3.	Кластеризация. Алгоритм k-средних и его модификации.	2			8	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, тестирование
4.	Непараметрическая, линейная и нелинейная регрессия. Метод опорных векторов.	2			8	Проработка лекционного материала, литературных источников.	Групповой и индивидуальный опрос, тестирование
5.	Машинное обучение с помощью нейронных сетей. Искусственный нейрон, возможности и ограничения. Идеи обучения нейронных сетей. Математические модели нейронных сетей.	2			8	Проработка лекционного материала, литературных источников.	Групповой и индивидуальный опрос, тестирование
6.	От искусственного нейрона к персептрон Розенблатта. Теорема о сходимости персептрона. Проблема	2			8	Проработка лекционного материала,	Групповой и индивидуальный опрос, тестирование

	«Исключающего ИЛИ» (XOR).					литературных источников.	
7.	Основные теоремы нейроинформатики. Теорема Колмогорова. Теорема Стоуна-Вейерштрасса и её обобщение для нейронных сетей.	2			8	Проработка лекционного материала, литературных источников.	Групповой и индивидуальный опрос, тестирование
8.	Многослойный персептрон. Общий подход к обучению нейронных сетей «с учителем». Алгоритм обратного распространения ошибки.	2			8	Проработка лекционного материала, литературных источников.	Групповой и индивидуальный опрос, тестирование
9.	Объектно-ориентированная модель и программная реализация многослойного персептрона	2		12	8	Проработка лекционного материала, литературных источников. Подготовка к лабораторной работе	Групповой и индивидуальный опрос, тестирование, лабораторная работа
	<b>Всего часов:</b>	18		18	71,8		

## СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ

Дисциплины «Машинное обучение и анализ данных» на 8 семестр

очная

форма обучения

Вид работы	Объем дисциплины
Общая трудоемкость дисциплины (з.е. / часов)	4/144
Учебных часов на контактную работу с преподавателем:	
лекций	18
практических/ семинарских	
лабораторных	36
других (групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие работу обучающихся с преподавателем) (ФКР)	1,7
из них, предусмотренные на выполнение курсовой работы / курсового проекта <sup>1</sup>	
Учебных часов на самостоятельную работу обучающихся (СР)	53,5
из них, предусмотренные на выполнение курсовой работы / курсового проекта <sup>2</sup>	
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	34,8

Форма(ы) контроля:

Экзамен 8 семестр

<sup>1</sup> Контактных часов – 2

<sup>2</sup> Количество часов на самостоятельную работу указывается на усмотрение разработчика, но не более 20 часов

№ п/п	Тема и содержание	Форма изучения материалов: лекции, практические занятия, семинарские занятия, лабораторные работы, самостоятельная работа и трудоемкость (в часах)				Задания по самостоятельной работе студентов	Форма текущего контроля успеваемости (коллоквиумы, контрольные работы, компьютерные тесты и т.п.)
		ЛК	ПР/СЕМ	ЛР	СР		
1	2	3	4	5	6	8	9
1.	Обучение нейронных сетей на основе принципов самоорганизации. Задача кластеризации	2			4	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, тестирование
2.	Звезды Гроссберга. Сети и карты Кохонена.	2			4	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, тестирование
3.	Самоорганизующиеся карты Кохонена. Алгоритмы обучения	2			4	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, тестирование
4.	Объектно-ориентированная модель и программная реализация самоорганизующихся карт Кохонена	2		6	4	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, тестирование
5.	Нейронные сети с радиально-базисной функцией активации	2			4	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, тестирование
6.	Алгоритмы обучения нейронных сетей с радиально-базисной функцией активации	2		6	4	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, тестирование
7.	Нейронные сети Хопфилда. Правило обучение Хебба	2			4	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, тестирование

8.	Генетические алгоритмы и их применение к обучению нейронных сетей	2			4	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, тестирование
9.	Нейросетевые библиотеки для языка программирования Python	2		12	4	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, тестирование
10.	РГР				17,5	Выполнение расчетно-графической работы	
	<b>Всего часов:</b>	18		36	53,5		

