


МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФГБОУ ВО "БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ"  
ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

*Утверждено:*

на заседании кафедры ИТиКМ  
протокол № 9 от 22 апреля 2020 г.

Зав. кафедрой

 А.М. Болотнов

*Согласовано:*

Председатель УМК  
ФМ и ИТ

 А.М. Ефимов

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

*Методы машинного обучения и интеллектуального анализа данных*

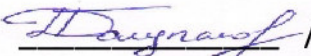
Обязательная часть

**программа магистратуры**

Направление подготовки (специальность)  
**09.04.03. «Прикладная информатика»**

Направленность (профиль) подготовки  
**Интеллектуальное управление и обработка информации**  
Квалификация  
**магистр**

Разработчик (составитель)  
доцент, к.т.н., доцент



Полупанов Д.В.

Для приема: 2020

Уфа 2020

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

Рабочая программа дисциплины *утверждена* на заседании кафедры протокол № 11 от 19 июня 2019 г.

Дополнения и изменения, внесенные в рабочую программу дисциплины, утверждены на заседании кафедры информационных технологий и компьютерной математики, протокол № \_\_\_\_\_ от « \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

Заведующий кафедрой \_\_\_\_\_ / \_\_\_\_\_.

Дополнения и изменения, внесенные в рабочую программу дисциплины, утверждены на заседании кафедры информационных технологий и компьютерной математики, протокол № \_\_\_\_\_ от « \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

Заведующий кафедрой \_\_\_\_\_ / \_\_\_\_\_.

Дополнения и изменения, внесенные в рабочую программу дисциплины, утверждены на заседании кафедры информационных технологий и компьютерной математики, протокол № \_\_\_\_\_ от « \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

Заведующий кафедрой \_\_\_\_\_ / \_\_\_\_\_.

Дополнения и изменения, внесенные в рабочую программу дисциплины, утверждены на заседании кафедры информационных технологий и компьютерной математики, протокол № \_\_\_\_\_ от « \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

Заведующий кафедрой \_\_\_\_\_ / \_\_\_\_\_.

## 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:

Формируемая компетенция (с указанием кода)	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
<i>ОПК-2. Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач</i>	<i>ОПК-2.1. Знать современные интеллектуальные технологии для решения профессиональных задач.</i>	<i>Демонстрирует знания основных технологий машинного обучения и интеллектуального анализа данных</i>
	<i>ОПК-2.2. Уметь обосновывать выбор современных интеллектуальных технологий и программной среды при разработке оригинальных программных средств для решения профессиональных задач</i>	<i>Выбирает подходящий инструментарий для реализации алгоритмов машинного обучения и интеллектуального анализа данных,</i>
	<i>ОПК-2.3. Владеет навыками разработки оригинальных алгоритмов с использованием современных интеллектуальных технологий для решения профессиональных задач</i>	<i>Имеет практический опыт разработки и реализации систем машинного обучения и интеллектуального анализа данных на базе языков программирования</i>

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

Дисциплина «Методы машинного обучения и интеллектуального анализа данных» относится к обязательной части.

Дисциплина изучается на 2 курсе в 3 и 4 семестре.

Целью изучения дисциплины «Методы машинного обучения и интеллектуального анализа данных» является ознакомление магистрантов с основами построения, разработки и применения к различным научным и прикладным профессиональным задачам систем машинного обучения и интеллектуального анализа данных.

## 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 *ОПК-2. Способен разрабатывать оригинальные*

алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач

Для зачета

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Критерии оценивания результатов обучения	
		«Зачтено»	«Не зачтено»
<i>ОПК-2.1. Знать современные интеллектуальные технологии для решения профессиональных задач.</i>	<i>Демонстрирует знания основных технологий машинного обучения и интеллектуального анализа данных</i>	Сформированные, возможно содержащее незначительные пробелы, знания основных алгоритмов машинного обучения и анализа данных	Фрагментарные, неполные, несистематические знания основных алгоритмов машинного обучения и анализа данных
<i>ОПК-2.2. Уметь обосновывать выбор современных интеллектуальных технологий и программной среды при разработке оригинальных программных средств для решения профессиональных задач</i>	<i>Выбирает подходящих инструментарий для реализации алгоритмов машинного обучения и интеллектуального 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>ОПК-2.3. Владеет навыками разработки оригинальных алгоритмов с использованием современных интеллектуальных технологий для решения профессиональных задач</i>	<i>Имеет практический опыт разработки и реализации систем машинного обучения и интеллектуального 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>ОПК-2.1. Знать современные интеллектуальные технологии для решения профессиональных задач.</i>	<i>Демонстрирует знания основных технологий машинного обучения и интеллектуального 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>ОПК-2.2. Уметь обосновывать выбор современных интеллектуальных технологий и программной среды при</i>	<i>Выбирает подходящий инструментарий для реализации алгоритмов ма-</i>	Отсутствие умений или фрагментарные умения реализовывать алгоритмы машинного обучения и анализа данных на базе	В целом успешное, но не систематическое умение реализовывать алгоритмы машинного обучения и анализа данных на базе	В целом успешное, но содержащее отдельные пробелы, умение реализовывать алгоритмы машинного обучения и анализа	Сформированное реализовывать алгоритмы машинного обучения и анализа данных на базе языков

<i>разработке оригинальных программных средств для решения профессиональных задач</i>	<i>шинного обучения и интеллектуального анализа данных,</i>	языков программирования	языков программирования	данных на базе языков программирования	программирования
<i>ОПК-2.3. Владеет навыками разработки оригинальных алгоритмов с использованием современных интеллектуальных технологий для решения профессиональных задач</i>	<i>Имеет практический опыт разработки и реализации систем машинного обучения и интеллектуального анализа данных на базе языков программирования</i>	Отсутствие или фрагментарное применение навыков реализации систем машинного обучения на базе языков программирования	В целом успешное, но не систематическое применение навыков реализации систем машинного обучения на базе языков программирования	В целом успешное, но содержащее отдельные пробелы применение навыков реализации систем машинного обучения на базе языков программирования	Успешное и систематическое применение навыков реализации систем машинного обучения на базе языков программирования

**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>ОПК-1.1. Знать математические, естественнонаучные и социально-экономические методы для использования в профессиональной деятельности.</i>	<i>Демонстрирует знания основных математических и инструментальных методов поддержки принятия решений, критериев выбора инструментов системы поддержки принятия решений</i>	Индивидуальный и групповой опрос, РГР, зачет, экзамен
<i>ОПК-1.2. Уметь решать нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных социально-экономических и профессиональных знаний.</i>	<i>Формализует процесс обоснования и принятия решений; выбирает инструментарий для каждого этапа принятия решения; использует инструментарий мониторинга исполнения решений</i>	Индивидуальный и групповой опрос, лабораторные работы, РГР, зачет, экзамен
<i>ОПК-1.3. Имеет практический опыт применения математических, естественнонаучных социально-экономических и профессиональных знаний.</i>	<i>Владеет навыками формулирования требований к системам поддержки принятия решений, разработки их отдельных элементов, оценки вариантов закупок, внедрения и эксплуатации систем в области принятия решений</i>	Лабораторные работы, РГР, зачет, экзамен

**Экзаменационные билеты**

*Примерный перечень вопросов к зачету*

1. Различные задачи машинного обучения – классификации, кластеризации, идентификации, прогнозирования, извлечения знаний.
2. Основные проблемы машинного обучения
3. Вероятностная постановка задачи распознавания образов. Байесов классификатор
4. Обобщенные линейные модели. Логистическая регрессия
5. Метод опорных векторов и беспризнаковое распознавание образов
6. Задачи выбора модели. Кроссвалидация.
7. Теория Вапника-Червоненкиса. Информационные критерии
8. Байесовский подход к теории вероятности. Примеры байесовских рассуждений
9. Решение задачи выбора модели по Байесу. Обоснованность модели.
10. Метод релевантных векторов
11. Недиагональная регуляризация обобщенных линейных моделей.
12. Общее решение для недиагональной регуляризации
13. Методы оценки обоснованности.
14. Графические модели.
15. Гауссовские процессы в машинном обучении

#### **Шкала оценивания:**

- «**Зачтено**» выставляется магистранту, если он без затруднений ответил на все дополнительные вопросы, практическая часть работы выполнена полностью без неточностей и ошибок или если магистрант раскрыл в основном теоретические вопросы, однако допущены неточности в определении основных понятий. При ответе на дополнительные вопросы допущены небольшие неточности, но при выполнении практической части работы допущены несущественные ошибки;

- «**Незачтено**» выставляется магистранту, если 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. Магистрант не смог ответить ни на один дополнительный вопрос.

*Структура экзаменационного билета:*

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

*Образец экзаменационного билета:*

<p>Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования <b>БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ</b> <i>Факультет математики и информационных технологий</i> <i>Кафедра информационных технологий и компьютерной математики</i></p>	
<p>2 курс, 4 семестр, 20__/20__ учебный год</p>	
<p>Дисциплина <i>Методы машинного обучения и интеллектуального анализа данных</i> Направление <i>09.04.03 - Прикладная информатика</i></p>	
<p><b>ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 1</b></p>	
<p>1. Основные алгоритмы и стратегии обучения нейронных сетей. Обучение «с учителем» 2. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере урожая п сельхоз культур на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно</p>	
<p><i>Зав. кафедрой ИТ и КМ</i></p>	<p><i>А.М. Болотнов</i></p>

*Примерный перечень вопросов к экзамену*

1. Основные понятия искусственного интеллекта. Основные задачи, методы и алгоритмы.
2. Логические основы построения систем искусственного интеллекта
3. Построение систем искусственного интеллекта на основе нечеткого вывода
4. Ансамблевые методы машинного обучения.
5. Б эггинг, бустинг, стекинг
6. Развитие методов деревьев решений в машинном обучении.

7. Метод случайного леса.
8. Нейронные сети как инструментальный машинного обучения.
9. Классификация, основные архитектуры нейронных сетей
10. Основные алгоритмы и стратегии обучения нейронных сетей. Обучение «с учителем»
11. Основные алгоритмы и стратегии обучения нейронных сетей. Обучение «без учителя»
12. Основные алгоритмы и стратегии обучения нейронных сетей. Комбинирование стратегий
13. Сверточные нейронные сети
14. Нейронные сети глубокого обучения
12. Импульсные (спайковые) нейронные сети
13. Нечеткие нейронные сети
14. Генетические алгоритмы
15. Нечеткие нейронные сети с генетическим проектированием
16. Нечеткие системы с генетической настройкой

### **Шкала оценивания:**

- **«Отлично»** выставляется магистранту, если им дан полный, развернутый ответ на теоретический вопрос билета, продемонстрировал знание функциональных возможностей, терминологии, основных элементов, умение применять теоретические знания при выполнении практического задания. Магистрант без затруднений ответил на все дополнительные вопросы. Практическая часть выполнена полностью без неточностей и ошибок;

- **«Хорошо»** выставляется магистранту, если им раскрыт в основном теоретический вопрос, однако допущены неточности в определении основных понятий. При ответе на дополнительные вопросы допущены небольшие неточности. При выполнении практической части допущены несущественные ошибки;

- **«Удовлетворительно»** выставляется магистранту, если при ответе на теоретический вопрос допущено несколько существенных ошибок в толковании основных понятий. Логика и полнота ответа страдают заметными изъянами. Заметны пробелы в знании основных методов. Теоретический вопрос в целом изложен достаточно, но с пропусками материала. Имеются принципиальные ошибки в логике построения ответа на вопрос. Магистрант не решил задачу или при решении допущены грубые ошибки;

- **«Неудовлетворительно»** выставляется магистранту, если ответ на теоретический вопрос свидетельствует о непонимании и крайне неполном знании основных понятий и методов. Обнаруживается отсутствие навыков применения теоретических знаний при выполнении практического задания. Магистрант не смог ответить ни на один дополнительный вопрос.

### **Образцы заданий для индивидуального и группового опроса**

*Тест «Общие сведения о машинном обучении»*

1. Выберите верные утверждения
  - a) Одна из задач машинного обучения - научиться делать прогнозы для признаков
  - b) Объекты описываются с помощью признаков
  - c) Одна из задач машинного обучения - научиться делать прогнозы для объектов
  - d) Признаки описываются с помощью объектов
2. Какие из этих задач являются задачами классификации?
  - a) Прогноз температуры на следующий день
  - b) Разделение книг, хранящихся в электронной библиотеке, на научные и художественные
  - c) Поиск групп похожих пользователей интернет-магазина
  - d) Прогноз оценки студента по пятибалльной шкале на экзамене по машинному обучению в следующей сессии
3. Какие свойства данных препятствуют однозначному построению разделяющей поверхности?
  - a) Ортогональность
  - b) Мультиколлинеарность
  - c) Противоречивость
  - d) Категориальность



4. Какая способность людей и систем позволяет получать им новые знания по наблюдению отдельных прецедентов (примеров)?

- a) Корректировать ошибку
- b) Обобщать
- c) Запоминать
- d) Распознавать образы

5. Какая задача лучше всего подходит под следующее описание. Нахождение такой функции  $F$ , которая бы наилучшим образом

254

отображала неизвестные ранее объекты  $X$  в конечное множество целочисленных номеров (имен, меток), на основании обучающих пар  $(X, Y)$ ?

- a) Прогнозирование денежных затрат
- b) Кластеризация клиентов
- c) Классификация образов
- d) Выявление особенностей в данных

6. Почему для обучения моделей используются такие методы как Градиентный спуск?

- a) Потому что метод позволяет корректировать параметры модели постепенно
- b) Потому что аналитические решения не всегда дают корректное решение
- c) Потому что такой подход позволяет получать более точные решения (Глобальный экстремум в отличие от локального)
- d) Потому что при большой размерности входных данных подобные методы работают быстрее

7. Выберите верные утверждения

- a) Метод Байеса - это во многом классический подход к классификации, основанный на оценке частоты встреч объектов со схожими признаками
- b) Благодаря универсальности статистического подхода метод Байеса позволяет решать любые задачи без априорной информации
- c) Данный метод позволяет очень хорошо обобщать высокоуровневые признаки
- d) Закон, задающий распределение вероятностей, который используется в предсказательной модели сильно влияет на способ обобщения

*Тест «Нейронные сети»*

1. Выберите верные утверждения

- a) НС проще подобрать под любую нелинейную задачу. Все что нужно сделать это увеличивать число слоев пропорционально числу признаков
- b) НС позволяют обрабатывать более высокоуровневые признаки за счет нелинейной функции активации и последовательным слоям
- c) По сравнению с Регрессией НС практически не подвержены Переобучению при любом количестве нейронов
- d) С точки зрения математического аппарата НС это комбинация полиномиальной регрессии высокого порядка и формулы Байеса
- e) НС может аппроксимировать любую нелинейную непрерывную функцию, но это еще не гарантирует 100% сходимости на произвольных данных
- f) НС в отличие от регрессии может хорошо обрабатывать высокую степень мультиколлинеарности и противоречивости в данных

2. Сеть какого типа лучше использовать для прогнозирования временных рядов?

- a) Сверточную
- b) ART MAP
- c) Импульсную
- d) MLP
- e) Рекуррентную
- f) Когнитрон

3. Сеть какого типа лучше использовать для обработки трехмерных сцен?

- a) MLP
- b) Рекуррентную

- c) ART MAP
- d) Сверточную
- e) Когнитрон
- f) Импульсную

4. Сеть какого типа лучше использовать для решения задачи классификации клиентов по одиночному вектору их характеристик (с учетом того, что этот вектор содержит большое количество категориальных признаков)?

- a) Автокодировщик
- b) MLP
- c) Когнитрон
- d) ART MAP
- e) Сверточную
- f) Рекуррентную
- g) Импульсную

5. Сеть какого типа можно использовать в условиях постоянного изменения данных, когда точной выборки еще не существует и сеть приходится постоянно дообучивать на новых классах?

- a) MLP
- b) Сверточную
- c) Когнитрон
- d) Рекуррентную
- e) ART MAP
- f) Автокодировщик
- g) Импульсную

Описание методики оценивания:

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

Вопросы по теме «Генетические алгоритмы»

1. Поясните происхождение названия «генетические алгоритмы».
2. Опишите сферу применения генетических алгоритмов.
3. Дайте определение гена в контексте генетических алгоритмов.
4. Дайте определение хромосоме в контексте генетических алгоритмов.
5. Дайте определение популяции в контексте генетических алгоритмов.
6. Дайте определение степени приспособленности в контексте генетических алгоритмов.
7. Дайте определение кроссовера в контексте генетических алгоритмов.

## РГР

РГР представляет собой разработку относительно универсального программного продукта, реализующего ту или иную систему машинного обучения, например, 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.

Описание методики оценивания:

- «**Зачтено**» выставляется магистранту, если работа выполнена полностью без существенных погрешностей и ошибок. Программный продукт обладает достаточной универсальностью и пригоден для практического использования. При ответе на дополнительные вопросы при защите возможно допущены небольшие неточности;

- «**Не зачтено**» выставляется магистранту, если обнаруживается отсутствие навыков применения теоретических знаний при выполнении задания. Программа пригодна только для решения конкретной задачи и не обладает универсальностью. Магистрант не смог ответить ни на один дополнительный вопрос.

### **Лабораторные работы**

Лабораторные работы посвящены программной реализации различных систем машинного обучения и решению с их помощью прикладных задач.

#### *Примеры лабораторных работ*

Лабораторная работа. Работа с файлом данных «Титаника»

Приведенные ниже задания основаны на данных 'titanic.csv', где содержатся сведения о пассажирах «Титаника».

Задачи:

1. Какое количество мужчин и женщин ехало на корабле? В качестве ответа приведите два числа через пробел.
2. Какой части пассажиров удалось выжить? Посчитайте долю выживших пассажиров. Ответ приведите в процентах (число в интервале от 0 до 100, знак процента не нужен), округлив до двух знаков.
3. Какую долю пассажиры первого класса составляли среди всех пассажиров? Ответ приведите в процентах (число в интервале от 0 до 100, знак процента не нужен), округлив до двух знаков.
4. Какого возраста были пассажиры? Посчитайте среднее и медиану возраста пассажиров. В качестве ответа приведите два числа через пробел.
5. Коррелируют ли число братьев/сестер/супругов с числом родителей/детей? Посчитайте корреляцию Пирсона между признаками SibSp и Parch.
6. Какое самое популярное женское имя на корабле? Извлеките из полного имени пассажира (колонка Name) его личное имя (First Name). Это задание – типичный пример того, с чем сталкивается специалист по анализу данных. Данные очень разнородные и шумные, но из них требуется извлечь необходимую информацию. Попробуйте вручную разобрать несколько значений столбца Name и выработать правило для извлечения имен, а также разделения их на женские и мужские.

Лабораторная работа. Генетические алгоритмы

1. На языке Python разработайте скрипт, который с помощью генетического алгоритма решает следующую задачу. Дано  $N$  наименований продуктов, для каждого из которых известно  $m$  характеристик. Необходимо получить самый дешевый рацион из  $k$  наименований, удовлетворяющий заданным медицинским нормам для каждой из  $m$  характеристик.
2. На языке Python разработайте скрипт, который с помощью генетического алгоритма решает следующую задачу. Дано  $n$  пунктов производства продуктов и  $k$  городов, которые в них нуждаются. Каждый город может потребить  $x$  продуктов, а каждый пункт произвести  $y$  продуктов. Необходимо получить оптимальный маршрут, так, чтобы все города получили нужный им объем продуктов без сильного его превышения, а транспортные расходы были минимальными.
3. На языке Python разработайте скрипт, который с помощью генетического алгоритма решает следующую задачу. Дано  $N$  наименований продуктов, для каждого из которых известно  $m$  характеристик. Необходимо получить самый лучший по характеристикам рацион из  $k$  наименований, удовлетворяющий заданным ценовым рамкам. Лучшим считается рацион с минимальным отклонением от нормы.
4. На языке Python разработайте скрипт, который с помощью генетического алгоритма решает следующую задачу. Дано  $n$  пунктов производства продуктов и  $k$  городов, которые в них нуждаются. Каждый город может потребить  $x$  продуктов, а каждый пункт произвести  $y$  про-

дуктов. Необходимо получить оптимальный маршрут, так, чтобы все города получили нужный им объем продуктов с минимальным его превышением, а транспортные расходы укладывались в определенные рамки.

5. На языке Python разработайте скрипт, который с помощью генетического алгоритма решает следующую задачу. Дано  $N$  полей для  $i$  и  $k$  культур для посева. Для каждого поля известна характеристика урожайности каждой из  $k$  культур, а для каждой культуры – его закупочная стоимость. Необходимо получить самый лучший урожай за наименьшую стоимость.

*Описание методики оценивания:*

- «**Отлично**» выставляется, если все задачи решены верно, возможно допустить одну незначительную ошибку.

- «**Хорошо**» выставляется, если задачи решены в целом верно, но имеются две незначительные ошибки

- «**Удовлетворительно**» выставляется, если более половины задач решено верно.

- «**Неудовлетворительно**» выставляется, если менее половины задач решено верно

### **Самостоятельная работа**

Самостоятельная работа магистрантов 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. Нечеткие нейронные сети на основе нечетких нейронов
2. Нечеткие нейронные сети на основе нейронов, реализующих нечеткие операции
3. Нечеткие продукционные сети в нейронных сетях
4. Нечеткие ситуационные сети
5. Алгоритмы М.А. Айзермана и Э.М. Бравермана обучения нейронных сетей
6. Метод двойственности обучения нейронных сетей
7. Представление нейронных сетей с помощью ориентированных графов
8. Устойчивость нейронных сетей
9. Рекуррентные нейронные сети, имитирование конечных автоматов
10. Конструктивные алгоритмы обучения нейронных сетей
11. Самоорганизующиеся нейронные сети, конгитрон и неоконгитрон. Алгоритм обучения
12. Стохастические методы обучения нейронных сетей
13. Сигма-пи нейронные сети

## **5. Учебно-методическое и информационное обеспечение дисциплины**

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

#### **Основная литература:**

1. Барский, А.Б. Введение в нейронные сети [Электронный ресурс]: учеб. пособие — Электрон. дан. — М.: , 2016. — 358 с. — Режим доступа: <https://e.lanbook.com/book/100684>. — Загл. с экрана.
2. Болотова, Л.С. Системы искусственного интеллекта: модели и технологии, основанные на знаниях [Электронный ресурс]: учеб. — Электрон. дан. — М.: Финансы и статистика, 2012. — 664 с. — Режим доступа: <https://e.lanbook.com/book/65936>. — Загл. с экрана.
3. Борисов, В.В. Нечеткие модели и сети [Электронный ресурс] : учеб. пособие / В.В. Борисов, В.В. Круглов, А.С. Федулов. — Электрон. дан. — М.: Горячая линия-Телеком, 2012. — 284 с. — Режим доступа: <https://e.lanbook.com/book/5126>. — Загл. с экрана.
4. Галушкин, А.И. Нейронные сети: основы теории [Электронный ресурс]: учеб. пособие — Электрон. дан. — М.: Горячая линия-Телеком, 2012. — 496 с. — Режим доступа: <https://e.lanbook.com/book/5144>. — Загл. с экрана.

5. Ярушкина, Н.Г. Основы теории нечетких и гибридных систем. Учебное пособие [Электронный ресурс]: учеб. пособие — Электрон. дан. — М. : Финансы и статистика, 2004. — 320 с. — Режим доступа: <https://e.lanbook.com/book/28372>. — Загл. с экрана.

#### Дополнительная литература:

6. 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)

7. Коэльо, Л.П. Построение систем машинного обучения на языке Python [Электронный ресурс] / Л.П. Коэльо, В. Ричарт. — Электрон. дан. — М.: ДМК Пресс, 2016. — 302 с. — Режим доступа: <https://e.lanbook.com/book/82818>. — Загл. с экрана.

8. 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. — М.: Горячая линия-Телеком, 2013. — 384 с. — Режим доступа: <https://e.lanbook.com/book/11843>. — Загл. с экрана.

9. Харахан, О.Г. Системы искусственного интеллекта. Практикум для проведения лабораторных работ [Электронный ресурс]: учеб. пособие — Электрон. дан. — М.: Горная книга, 2006. — 80 с. — Режим доступа: <https://e.lanbook.com/book/3508>. — 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

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

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 № 531 (физмат корпус- учебное), аудитория № 515(физмат корпус- учебное).</p> <p><b>2. Учебная аудитория для проведения занятий семинарского типа:</b> аудитория № 526 (физмат корпус- учебное), аудитория № 527(физмат корпус- учебное), аудитория № 522 (физмат корпус- учебное).</p> <p><b>4. Учебная аудитория для текущего контроля и промежуточной аттестации:</b> аудитория № 531 (физмат корпус- учебное), аудитория № 526 (физмат корпус- учебное), аудитория № 527(физмат корпус- учебное).</p> <p><b>4. Помещения для самостоятельной работы:</b> читальный зал</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>Аудитория №526</b> Учебная мебель, доска настенная меловая.</p> <p><b>Аудитория №527</b> Учебная мебель, доска настенная меловая.</p> <p><b>Читальный зал №2</b> 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> Учебная мебель, доска, персональный компьютер Lenovo Think Centre A70z Intel Pentium E 5800, 320 Gb, 19" – 13 шт., кондиционер Lessar LS/LU-H24KB2.</p> <p><b>Читальный зал №2</b></p>	<p>1. Windows 8 Russian. Windows Professional 8 Russian Upgrade. Договор № 104 от 17.06.2013 г. Лицензии бессрочные.</p> <p>2. Microsoft Office Standard 2013 Russian. Договор № 114 от 12.11.2014 г. Лицензии бессрочные.</p> <p>3. Python (лицензия Python SoftwareFoundation License, свободное программное обеспечение).</p>
---	--	--

№2 (физмат корпус- учебное).	Учебная мебель, учебно-наглядные пособия, стенд по пожарной безопасности, моноблоки стационарные – 8 шт, принтер – 1 шт., сканер – 1 шт.	
------------------------------	--	--

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

### СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ

Дисциплины «Методы машинного обучения и интеллектуального анализа данных» на 3 семестр  
очная  
форма обучения

Вид работы	Объем дисциплины
Общая трудоемкость дисциплины (з.е. / часов)	3 / 108
Учебных часов на контактную работу с преподавателем:	

лекций	12
практических/ семинарских	
лабораторных	24
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)	71,8
из них, предусмотренные на выполнение курсовой работы / курсового проекта	
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	

Форма(ы) контроля:  
зачет 3 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.	Различные задачи машинного обучения – классификации, кластеризации, идентификации, прогнозирования, извлечения знаний. Основные проблемы машинного обучения	2		4	5,8	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
2.	Вероятностная постановка задачи распознавания образов. Байесов классификатор	2		4	6	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
3.	Обобщенные линейные модели. Логистическая регрессия	2		4	6	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
4.	Метод опорных векторов и беспризнаковое распознавание образов	2		4	6	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
5.	Задачи выбора модели. Кроссвалидация. Теория Валника-Червоненкиса. Информационные критерии	2		4	6	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
6.	Байесовский подход к теории вероятности. Примеры байесовских рассуждений	2		4	6	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
7.	Решение задачи выбора модели по Байесу. Обоснованность модели.	2		4	6	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
8.	Метод релевантных векторов	2		4	6	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы



									ала, литературных источников	
9.	Недиагональная регуляризация обобщенных линейных моделей.	2		4	6			Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы	
10.	Общее решение для недиагональной регуляризации	2		4	6			Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы	
11.	Методы оценки обоснованности.	2		4	6			Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы	
12.	Графические модели. Гауссовские процессы в машинном обучении	2		4	6			Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы	
	<b>Всего часов:</b>	12		24	71,8					

### СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ

Дисциплины «Методы машинного обучения и интеллектуального анализа данных» на 4  
семестр  
очная  
форма обучения

Вид работы	Объем дисциплины
Общая трудоемкость дисциплины (з.е. / часов)	7 / 252
Учебных часов на контактную работу с преподавателем:	
лекций	12
практических/ семинарских	
лабораторных	24
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	
Учебных часов на самостоятельную работу обучающихся (СР)	180
из них, предусмотренные на выполнение курсовой работы / курсового проекта	
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	34,8

Форма(ы) контроля:  
зачет 3 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.	Основные понятия искусственного интеллекта. Основные задачи, методы и алгоритмы.	2		4	10	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
2.	Логические основы построения систем искусственного интеллекта	2		4	10	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
3.	Построение систем искусственного интеллекта на основе нечеткого вывода	2		4	10	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
4.	Ансамблевые методы машинного обучения. Бэггинг, бустинг, стекинг	2		4	10	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
5.	Развитие методов деревьев решений в машинном обучении. Метод случайного леса.	2		4	10	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
6.	Нейронные сети как инструмент машинного обучения. Классификация, основные архитектуры	2		4	10	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
7.	Основные алгоритмы и стратегии обучения нейронных сетей.	2		4	10	Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы
8.	Сверточные нейронные сети	2		4	10	Проработка лекционного материала,	Групповой и индивидуальный опрос, лабораторные работы

									литературных источников	
9.	Нейронные сети глубокого обучения	2			4	10		Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы	
10.	Импульсные (спайковые) нейронные сети	2			4	10		Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы	
11.	Нечеткие нейронные сети	2			4	10		Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы	
12.	Генетические алгоритмы	2			4	10		Проработка лекционного материала, литературных источников	Групповой и индивидуальный опрос, лабораторные работы	
13.	РГР					60				
	<b>Всего часов:</b>	12			24	180				