

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Смирнов Сергей Николаевич
Должность: врио ректора
Дата подписания: 02.04.2025 12:27:28
Уникальный программный ключ:
69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Тверской государственный университет»



Утверждаю:

Проректор по УВР

Сердитова Н.Е.

«01» сентября 2020 г.

Рабочая программа факультативной дисциплины (с аннотацией)
**МАТЕМАТИЧЕСКИЙ АППАРАТ И ИНСТРУМЕНТАРИЙ
ДЛЯ АНАЛИЗА ДАННЫХ**

Направление подготовки

02.03.01 Математика и компьютерные науки

02.04.01 Математика и компьютерные науки

Профиль подготовки

Математическое и компьютерное моделирование

Для студентов: 2 - 4 курса, очной формы обучения бакалавриата
1 курса магистратуры

Составители:

к.ф.-м.н. Д.В. Подгайный

к.ф.-м.н. О.И. Стрельцова

М.И. Зуев

Тверь, 2020

І.Аннотация

1. Цель и задачи дисциплины

Целью освоения дисциплины является:

Освоение основных вычислительных и статистических методов и инструментов для решения широкого круга задач в сфере анализа данных (Data Science), практических навыков по разработке программного обеспечения на языке программирования Python, освоение работы с библиотеками и фреймворками для анализа данных, изучением методов машинного обучения/глубокого обучения и получение практического опыта в решении прикладных задач.

Преподавание учебной дисциплины «**Математический аппарат и инструментарий для анализа данных**» строится на сочетании лекций, практических занятий и различных форм самостоятельной работы студентов. На практических занятиях формируются умения применять важнейшие математические модели организационных систем, используемые при решении задач управления сбытом; условия применения методов линейного и нелинейного программирования для формализации динамических процессов. Предусмотрены аудиторные самостоятельные работы по основным темам курса, вебинары, а также домашние задания.

Задачи курса:

- Изучить основные модели и методы машинного обучения и разработки данных
- Дать умение адекватно применять указанные модели и методы, а также программные средства, в которых они реализованы
- Приобретение опыта анализа реальных данных с помощью изученных методов

2. Место дисциплины в структуре ООП

Дисциплина «**Математическое моделирование и проектная деятельность**» является факультативной. Дисциплина может изучаться в течении 2 семестров в 3-8 семестрах бакалавриата и 1 – 2 семестрах магистратуры, заканчивается зачетом, при успешной сдаче которого выдается соответствующий сертификат.

3. Объем дисциплины 6 зачетных единиц, 216 академических часов,

в том числе:

контактная аудиторная работа: 32 часа лекции, 32 часа практические работы, 32 вебинары/ on-line консультации;

самостоятельная работа: 120 часов.

4. Планируемые результаты обучения по дисциплине

Для студентов бакалавриата

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.2 Определяет, интерпретирует и ранжирует информацию, требуемую для решения поставленной задачи УК-1.3 Осуществляет поиск информации для решения поставленной задачи по различным типам запросов УК-1.4 При обработке информации отличает факты от мнений, интерпретаций, оценок, формирует собственные мнения и суждения, аргументирует свои выводы и точку зрения УК-1.5 Рассматривает и предлагает возможные варианты решения поставленной задачи, оценивая их достоинства и недостатки
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, ис-	УК-2.1 Определяет круг задач в рамках поставленной цели, определяет связи между ними УК-2.2 Предлагает способы решения поставлен-

ходя из действующих правовых норм, имеющихся ресурсов и ограничений	ных задач и ожидаемые результаты; оценивает предложенные способы с точки зрения соответствия цели проекта УК-2.3 Планирует реализацию задач в зоне своей ответственности с учетом имеющихся ресурсов и ограничений, действующих правовых норм
ОПК-4 Способен находить, анализировать, реализовывать программно и использовать на практике математические алгоритмы, в том числе с применением современных вычислительных систем	ОПК-4.1 Использует основы современного математического аппарата, связанного с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в области математического и компьютерного моделирования естественных и социально-экономических процессов ОПК-4.2 Применяет современный математический аппарат, связанный с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в области математического и компьютерного моделирования естественных и социально-экономических процессов в профессиональной деятельности
ОПК-5 Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий, в том числе отечественного производителя, и с учетом основных требований информационной безопасности	ОПК-5.1 Использует информационно-коммуникационные технологии для применения программных продуктов и комплексов программ в области математического и компьютерного моделирования естественных и социально-экономических процессов с учетом основных требований информационной безопасности ОПК-5.2 Осуществляет поиск необходимой информации для решения задач профессиональной деятельности на основе информационной и библиографической культуры

Для студентов магистратуры

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.1 Формулирует на основе поставленной проблемы проектную задачу и способ ее решения через реализацию проектного управления; УК-2.2 Разрабатывает концепцию проекта в рамках обозначенной проблемы: формулирует цель, задачи, обосновывает актуальность, значимость, ожидаемые результаты и возможные сферы их применения; УК-2.3 Разрабатывает план реализации проекта с учетом возможных рисков реализации и возможностей их устранения, планирует необходимые ресурсы, в том числе с учетом их заменяемости; УК-2.4 Осуществляет мониторинг хода реализации про-

	<p>екта, корректирует отклонения, вносит дополнительные изменения в план реализации проекта, уточняет зоны ответственности участников проекта;</p> <p>УК-2.5 Предлагает процедуры и механизмы оценки качества проекта, инфраструктурные условия для внедрения результатов проекта.</p>
<p>ОПК-1</p> <p>Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики</p>	<p>ОПК-1.1</p> <p>Осуществляет поиск актуальных и значимых проблем прикладной и компьютерной математики в области математического и компьютерного моделирования социально-экономических и природных систем, а также известные, актуальные методы и алгоритмы для их решения</p> <p>ОПК-1.2</p> <p>Формулирует и формализует конкретные актуальные, значимые проблемы прикладной и компьютерной математики в области математического и компьютерного моделирования социально-экономических и природных систем</p> <p>ОПК-1.3</p> <p>Решает конкретные актуальные, значимые задачи прикладной и компьютерной математики в области математического и компьютерного моделирования социально-экономических и природных систем</p>
<p>ОПК-3</p> <p>Способен самостоятельно создавать прикладные программные средства на основе современных информационных технологий и сетевых ресурсов, в том числе отечественного производства</p>	<p>ОПК-3.1</p> <p>Разрабатывает алгоритмы по вычислению параметров математических моделей динамических систем в естественных науках и исследованию их характера поведения для создания прикладных программ на основе современных информационных технологий и сетевых ресурсов</p> <p>ОПК-3.2</p> <p>Создает прикладные программы для вычисления параметров математических моделей динамических систем в естественных науках и исследованию их характера поведения на основе современных информационных технологий и сетевых ресурсов</p> <p>ОПК-3.3</p> <p>Проводит тестирование и верификацию используемых программных средств</p>

5. Форма контроля

Зачет

6. Язык преподавания русский.

II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

Учебная программа – наименование разделов и тем	Всего (час.)	Контактная работа (час.)			Самостоятельная работа, в том числе Контроль (час.)
		Лекции	Практические работы	вебинары/ on-line консультации	
Язык программирования Python, библиотеки для анализа и визуализации данных NumPy, matplotlib, pandas	19	3	3	3	10
Библиотека NumPy: работа с массивами (создание, индексация, срезы, слияние и разбиение), выполнение вычислений над массивами, агрегирование, операции над массивами, сортировка массивов, структурированные массивы	19	3	3	3	10
Работа с данными с библиотекой pandas: объекты библиотеки, операции над данными, иерархическая индексация, объединение набор данных, агрегирование и группировка, сводные таблицы, работа с временными рядами.	21	3	3	3	12
Визуализация с помощью библиотек matplotlib и seaborn: графики, гистограммы и т.д.	21	3	3	3	12
Машинное обучение с библиотекой Scikit-learn.	26	4	4	4	14

Задачи машинного обучения: кластеризация, методы понижения размерности. Метод главных компонент.	26	4	4	4	14
Математические основы нейросетевого подхода.	28	4	4	4	16
Фреймворки и библиотеки TensorFlow, Keras: библиотека глубокого обучения в Python.	28	4	4	4	16
Глубокое обучение для задач анализа текстов	28	4	4	4	16
ИТОГО	216	32	32	32	120

Тема 1. Язык программирования Python, библиотеки для анализа и визуализации данных NumPy, matplotlib, pandas

Синтаксис языка Python для основных алгоритмических конструкций, литералов, выражений. Описание встроенных типов данных, особенности общепринятого в Python стиля программирования.

Тема 2. Библиотека NumPy: работа с массивами (создание, индексация, срезы, слияние и разбиение), выполнение вычислений над массивами, агрегирование, операции над массивами, сортировка массивов, структурированные массивы

Разнообразие типов-коллекций в языке Python. Общие подходы и особенности при их использовании при написании программ: индексирование, срезы, функции вычисления длины, максимального и минимального значений, сортировки, смены порядка следования элементов и др. Особенности каждого из типов данных.

Тема 3. Работа с данными с библиотекой pandas: объекты библиотеки, операции над данными, иерархическая индексация, объединение набор данных, агрегирование и группировка, сводные таблицы, работа с временными рядами.

Функции открытия, закрытия, чтения, записи в текстовый файл. Обработка файлов в формате UNICODE. Использование итераторов при работе с файлами. Хранение и обработка информации при помощи двоичных файлов.

Тема 4. Визуализация с помощью библиотек matplotlib и seaborn: графики, гистограммы и т.д.

Формулировка двойственной задачи линейного программирования, её экономическая интерпретация. Теоремы двойственности и их экономическое значение. Понятие двойственной оценки ограничения и объективно обусловленной оценки ресурса.

Стоимостная интерпретация двойственных оценок.

Проверка адекватности линейной математической модели с помощью двойственных оценок.

Тема 5. Машинное обучение с библиотекой Scikit-learn.

Пакет NumPy для осуществления численных расчетов и выполнения матричных вычислений, обзор других пакетов для научных вычислений.

Тема 6. Задачи машинного обучения: кластеризация, методы понижения размерности. Метод главных компонент.

Формулировка общей задачи математического программирования. Классификация задач нелинейного программирования.

Понятие о функции Лагранжа. Теорема Куна-Таккера для общей и выпуклой задач математического программирования. Общая интерпретация множителей Лагранжа в оптимуме задачи математического программирования.

Функциональная матрица задачи математического программирования в точке оптимума и её свойства.

Тема 7. Математические основы нейросетевого подхода.

Градиентные методы численного решения задач выпуклого программирования. Программное обеспечение выпуклого программирования.

Линеаризация задач выпуклого программирования. Сепарабельное программирование и его применение для приближённого решения невыпуклых задач математического программирования.

Практические приложения числовых моделей нелинейного программирования. Значение нелинейного программирования в моделировании с. Анализ компенсационных эффектов при исследовании потребительского спроса. Уравнение Слуцкого.

Тема 8. Фреймворки и библиотеки TensorFlow, Keras: библиотека глубокого обучения в Python.

Приложения с графическим интерфейсом пользователя. Для построения интерфейса не применяются визуальные средства ("построители интерфейса"), а используются возможности графической библиотеки виджетов (Tk).

III. Образовательные технологии

Учебная программа – наименование разделов и тем (в строгом соответствии с разделом II РПД)	Вид занятия	Образовательные технологии
Язык программирования Python, библиотеки для анализа и визуализации данных NumPy, matplotlib, pandas	Практическая работа	1. <i>Активное слушание</i> 2. <i>Метод case-study</i>
	Лекции	Лекция (традиционная, лекция-визуализация, лекция с запланированными ошибками)
Библиотека NumPy: работа с массивами (создание, индексация, срезы, слияние и разбиение), выполнение вычислений над массивами, агрегирование, операции над массивами, сортировка массивов, структурированные массивы	Практическая работа	1. <i>Информационные (цифровые)</i>
	Лекции	Лекция (традиционная, проблемная, лекция-визуализация, лекция-консультация); Активное слушание
Работа с данными с библиотекой pandas: объекты библиотеки, операции над данными, иерархическая индексация, объединение набор данных, агрегирование и группировка, сводные таблицы, работа с временными рядами.	Практическая работа	1. <i>Информационные (цифровые)</i>
	Лекции	Лекция (традиционная, проблемная, лекция-визуализация, лекция с запланированными ошибками); Активное слушание
Визуализация с помощью библиотек matplotlib и seaborn: графики, гистограммы и т.д.	Практическая работа	1. <i>Активное слушание</i> 2. <i>Метод case-study</i>
	Лекции	Лекция (традиционная, лекция-визуализация)
Машинное обучение с библиотекой Scikit-learn.	Практическая работа	1. <i>Информационные (цифровые)</i> 2. <i>Методы группового решения творческих задач (метод Дельфи, метод б–б, метод развивающей кооперации, мозговой штурм (метод генерации идей), нетворкинг и т.д.)</i>

	Лекции	Лекция (традиционная, лекция-визуализация, лекция с запланированными ошибками); Активное слушание
Задачи машинного обучения: кластеризация, методы понижения размерности. Метод главных компонент.	Практическая работа	1. <i>Активное слушание</i> 2. <i>Метод case-study</i>
	Лекции	Лекция (традиционная, лекция-визуализация)
Математические основы нейросетевого подхода.	Практическая работа	1. <i>Активное слушание</i> 2. <i>Метод case-study</i>
	Лекции	Лекция (традиционная, проблемная, лекция-визуализация, лекция-консультация); Активное слушание
Фреймворки и библиотеки TensorFlow, Keras: библиотека глубокого обучения в Python.	Практическая работа	1. <i>Активное слушание</i> 2. <i>Метод case-study</i>
	Лекции	Лекция (традиционная, лекция-визуализация, лекция с запланированными ошибками); Активное слушание
Глубокое обучение для задач анализа текстов	Практическая работа	1. <i>Дискуссионные технологии (форум, симпозиум, дебаты, аквариумная дискуссия, панельная дискуссия, круглый стол, фасилитированная и т.д.)</i>
	Лекции	Лекция (традиционная, проблемная, лекция-визуализация, лекция с запланированными ошибками); Активное слушание

Перечень педагогических и информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

В качестве традиционных форм обучения дисциплине выступают лекции и практические работы. На занятиях практикуется самостоятельная работа студентов, выполнение заданий в малых группах, письменные работы, моделирование дискуссионных ситуаций, работа с раздаточным материалом, привлекаются ресурсы сети INTERNET, ролевые и деловые игры, кейс-анализ, презентация, видеофильмы, видеокурсы, мультимедийные курсы, тестирование как метод контроля. Курс предусматривает выполнение

контрольных и самостоятельных работ, письменных домашних заданий. В качестве форм контроля используются различные варианты взаимопроверки и взаимоконтроля.

В процессе освоения дисциплины используются следующие образовательные технологии, способы и методы формирования компетенций:

- 1) информационно-рецептивные:
 - чтение и конспектирование литературы;
- 2) репродуктивные технологии:
 - анализ и написание текстов,
 - выполнение проблемных и творческих заданий;
- 3) рейтинговая система контроля успеваемости;
- 4) интерактивные технологии:
 - тренинг в малых группах,
 - дискуссии (пресс-конференция и круглый стол).

IV. Оценочные материалы для проведения текущей и промежуточной аттестации

Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине:

- методические рекомендации по организации самостоятельной работы студентов;
- методические рекомендации по организации учебной работы студентов;
- примеры практических заданий;
- вопросы для самоконтроля.

Практические (семинарские) занятия

Практические занятия по курсу «Математический аппарат и инструментарий для анализа данных» имеют цель познакомить студентов с основными вычислительными и статистическими методами и инструментами для решения широкого круга задач в сфере анализа данных (Data Science), приобрести практические навыки по разработке программного обеспечения на языке программирования Python, освоить работу с библиотеками и фреймворками для анализа данных. В ходе изучения курса студенты получают практический опыт решения прикладных задач с использованием методов машинного обучения и глубокого обучения (ML/DL).

Практические занятия будут проводиться на гетерогенной вычислительной платформе HybriLIT (<http://hlit.jinr.ru/>) в рамках экосистемы для ML/DL.

Прохождение всего цикла семинарских занятий является необходимым условием получения студентом допуска к экзамену.

Список семинаров.

№	Наименование практических занятий (семинаров)
1.	Вводное занятие по работе на гетерогенной вычислительной платформе ЛИТ ОИЯИ <i>HybriLIT</i> в рамках экосистемы для ML/DL.
2.	Основы <i>Python</i> : базовые типы данных, типы коллекций, управляющие структуры и функции.
3.	Язык программирования <i>Python</i> , библиотеки для анализа и визуализации данных: <i>NumPy</i> , <i>matplotlib</i> , <i>pandas</i> (основы).
4.	Библиотека <i>NumPy</i> : работа с числовыми массивами.
5.	Работа с данными с библиотекой <i>pandas</i> .
6.	Визуализация с помощью библиотек <i>matplotlib</i> и <i>seaborn</i> : графики, гистограммы и т.д.
7.	Машинное обучение с библиотекой <i>Scikit-learn</i> .
8.	Алгоритмы машинного обучения: задачи классификации.
9.	Алгоритмы машинного обучения: Линейная регрессия, регрессия по комбинации базисных функций, вопросы регуляризации
10.	Алгоритмы машинного обучения: метод опорных векторов (<i>Support Vector Machines</i>).
11.	Задачи машинного обучения: деревья принятия решений и случайные леса.
12.	Задачи машинного обучения: кластеризация, методы понижения размерности. Метод главных компонент.
13.	Фреймворки и библиотеки TensorFlow, Keras: библиотеки глубокого обучения Python.
14.	Машинное и глубокое обучение для задач анализа текстов.
15.	Компьютерное зрение с <i>Python</i> : работа с <i>opencv</i> .
16.	Компьютерное зрение с <i>Python</i> : глубокое обучение в технологиях компьютерного зрения.

Тематика практических работ.

№	Тема работы
ПР1	Вводное занятие по работе на гетерогенной вычислительной платформе ЛИТ ОИЯИ <i>HybriLIT</i> в рамках экосистемы для ML/DL. Работа в <i>Jupyter Notebook</i> – интерактивной web-среде для разработки приложений на Python, анализа и визуализации данных, решения задач математического моделирования, в том числе на базе методов машинного и глубокого обучения.
ПР2	Основы Python : базовые типы данных, типы коллекций, управляющие структуры и функции.
ПР3	Язык программирования <i>Python</i> , библиотеки для анализа и визуализации данных: <i>NumPy</i> , <i>matplotlib</i> , <i>pandas</i> (основы).
ПР4	Библиотека NumPy : работа с массивами (создание, индексация, срезы, слияние и разбиение), выполнение вычислений над массивами, агрегирование, операции над массивами, сортировка массивов, структурированные массивы. Примеры:

	<ul style="list-style-type: none"> • Создание одномерных и многомерных массивов (<code>np.arange()</code>, <code>np.linspace()</code>, <code>np.random.normal()</code> и т.д.): • Атрибуты массивов (<code>ndim</code>, <code>shape</code>, <code>size</code>) • Срезы: напечатать строку, столбец и т.д. • Маскирование и «прихотливая» индексация.
ПР5	<p>Работа с данными с библиотекой Pandas: объекты библиотеки, операции над данными, иерархическая индексация, объединение набор данных, агрегирование и группировка, сводные таблицы, работа с временными рядами.</p> <p>Примеры:</p> <ul style="list-style-type: none"> • Объекты библиотеки <i>Pandas</i>: объект <code>DataFrame</code>, объект <code>Series</code>; • объект <code>Series</code> как специализированный словарь; • индексация и выборка данных; • чтение данных из CSV файла; • сводные таблицы на примере данных о рождаемости.
ПР6	<p>Визуализация с помощью библиотек matplotlib и seaborn: графики, гистограммы и т.д.</p> <p>Примеры:</p> <ul style="list-style-type: none"> • простые линейные графики: цвета и стили линий, настройка графиков, подписи к осям, метки на графиках; • простые диаграммы рассеяния; • визуализация погрешностей; • графики плотности и контурные графики.
ПР7	<p>Машинное обучение с библиотекой Scikit-learn. Лабораторная работа № 1: задача о конверсии сайта.</p>
ПР8	<p>Алгоритмы машинного обучения: задачи классификации (наивная байесовская классификация, гауссов байесовский классификатор, полиномиальный баесовский классификатор).</p>
ПР9	<p>Алгоритмы машинного обучения: линейная регрессия, регрессия по комбинации базисных функций, вопросы регуляризации.</p>
ПР10	<p>Алгоритмы машинного обучения: метод опорных векторов (<i>Support Vector Machines</i>).</p>
ПР11	<p>Задачи машинного обучения: деревья принятия решений и случайные леса.</p>
ПР12	<p>Задачи машинного обучения: кластеризация, методы понижения размерности. Метод главных компонент.</p>
ПР13	<p>Фреймворки и библиотеки TensorFlow, Keras: библиотека глубокого обучения <i>Python</i>.</p> <p>На примерах рассматривается процесс создания нейронной сети в фреймворке Keras:</p> <ul style="list-style-type: none"> • создания модели с использованием функционального API; • подготовку данных для обучения, создание генератора данных; • обучение нейросети и оценка качества получившейся модели. <p>Отдельно рассматриваются необходимые приёмы повышения точности результатов:</p> <ul style="list-style-type: none"> • аугментация данных для обучения; • регуляризация сети с помощью слоев Dropout; • <i>Fine-tuning</i> нейросетевых моделей на основе

	предварительно обученной сети.
ПР14	<p>Машинное и глубокое обучение для задач анализа текстов.</p> <p>Рассматривается workflow типичного проекта Data Science, работающего с текстовыми документами: формулировка проблемы, генерация данных, анализ данных на основе соответствующих функций:</p> <p>несколько способов извлечения и описания семантической информации и будет показано как включить / дополнить ее дополнительной несемантической (что может помочь улучшить результаты). Далее мы рассмотрим, сконструируем и применим несколько стандартных моделей машинного обучения (ML) для описания наших данных: мы приведем их к задачам классификации и регрессии. Затем мы анализируем эффективность методов ML, а также роль, влияние и актуальность наших семантических и несемантических признаков. Далее мы покажем, как применять методы глубокого обучения для решения той же проблемы - мы рассмотрим простые модели DNN (глубокая нейронная сеть) и CNN (сверточная нейронная сеть). В конце мы сопоставляем наши результаты по ML и DL, обсуждаем их плюсы и минусы: эффективность, необходимые вычислительные ресурсы, возможные пути их улучшения.</p>
ПР15	Компьютерное зрение с <i>Python</i> : работа с <i>opencv</i> .
ПР16	Компьютерное зрение с <i>Python</i> : глубокое обучение в технологиях компьютерного зрения.

Самостоятельная работа студентов, предусмотренная учебным планом выполняется в ходе семестра в форме выполнения домашних и лабораторных заданий. Отдельные темы теоретического курса прорабатываются студентами самостоятельно в соответствии с планом самостоятельной работы и конкретными заданиями преподавателя с учетом индивидуальных особенностей студентов.

Пример домашней работы

I. Задача. (С.В. Айвазян, В.С. Мхитарян. Прикладная статистика в задачах и упражнениях. –М.: ЮНИТИ-ДАНА . 2001. -270 с.)

Специально исследование показало, что склонность фирм к утаиванию части своих доходов

(и, соответственно, - к уклонению от уплаты части налогов) в существенной мере определяется двумя показателями

- x_1 – соотношением «быстроты активов» и текущих пассивов;
- x_2 – соотношением прибыли и процентных ставок

(оба показателя оцениваются по определенной методике в шкале от 300 до 900 баллов);

Обучающая выборка:

№	Фирмы, уклоняющиеся от налогов		Фирмы, не уклоняющиеся от налогов	
	x_1	x_2	x_1	x_2
1	740	68	750	590

2	670	600	360	600
3	560	550	720	750
4	540	520	540	710
5	590	540	570	700
6	590	700	520	670
7	470	600	590	790
8	560	540	670	700
9	540	630	620	730
10	500	600	690	840
11	-	-	610	680
12	-	-	550	730
13	-	-	590	750

Требуется:

- Предложить (с обоснованием) и реализовать метод, с помощью которого можно определить, к какому классу (1 или 2) следует отнести фирму с показателями:
 $x_1 = 740, x_2 = 590$.
- **предложить (с обоснованием) и вычислить расстояние, измеряющее степень различия двух анализируемых классов**

Алгоритмы классификации :

- SVM (*C-Support Vector Classification* с различными значениями параметра *C*)
- Gaussian Naive Bayes algorithm for classification (*GaussianNB*)
- Метод К-ближайших соседей (*Nearest Neighbors Classification, KNeighborsClassifier*):
провести исследование в зависимости от параметров)

Примеры лабораторных работ

Лабораторная работа № 1

Задачи машинного обучения: Задача о конверсии сайта (корреляция и линейная регрессия).

Цель работы: На примере построения математической модели зависимости количества регистрации (и скачиваний) программного продукта от количества посещений сайта (данные работы [1]) освоить работу с моделями библиотеки *scikit-learn*.

Метод решения:

Для построения линейной модели предлагается использовать метод наименьших квадратов [2].

Библиотеки: для визуализации данных: *matplotlib* и *seaborn* [3], для построения линейной модели *Scikil-Learn* [4].

Подготовка и анализ данных:

- Создать CSV-файл с данными: `/data/website_stat.csv`
- Напечатать таблицу, данные по посещаемости и регистрации на сайте, сводные статистические показатели
- Построить зависимость зарегистрировавшихся на сайте от количества посещений

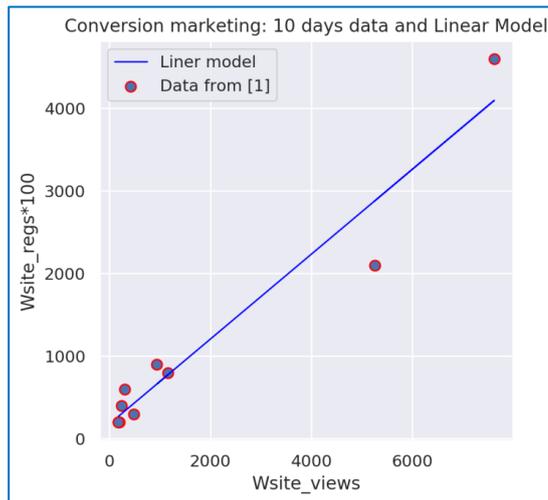
Реализации:

1. Реализация средствами *Python* и *Numpy*.
2. Реализации с использованием *Python*, *Numpy* и *Scikil-Learn*.

Отчет должен содержать:

1. Постановку задачи.
2. Описание метода решения (вывод формул).
3. Описание двух реализаций.
4. Построенные графики зависимостей:

Построить зависимость зарегистрировавшихся на сайте от количества посещений и полученной линейной модели:



5. Ответить на вопросы:
 - При количестве посещений сайта в 8000 человек, сколько планируется получить зарегистрировавшихся (или скачавших) программный продукт;
 - Для обеспечения 500 скачиваний продукта, сколько человек должны зайти на сайт?
6. Анализ полученных результатов и выводы.

Лабораторная работа № 2

Нейроконтроллер для персонажей компьютерных игр

Цель работы: на примере задачи разработки *Нейроконтроллера для персонажей компьютерных игр* (пример взят из [1]) освоить работу с нейросетями, изучить метод обратного распространения ошибки и освоить *workflow* с нейронными сетями в фрейворке *Keras*.

Данные (обучающая и тестовые выборки) представлены в виде таблицы:

- здоровье (числовой параметр, принимающий значения от 0 до 2)
- наличие ножа (числовой параметр от или 1)
- наличие пистолета (числовой параметр от или 1)
- наличие врагов (числовой параметр - количество врагов)
- последний столбец показывает принадлежность классу, т.е. действие
 - уворачиваться
 - атаковать
 - прятаться
 - бежать

Требуется: создания нейронной сети в фреймворка *Keras*:

- создания модели в Keras, моделирующего поведение персонажа игры, в зависимости от значения 4 входных параметров ;
- обучение нейросети и оценка качества получившейся модели.

V. Учебно-методическое и информационное обеспечение дисциплины

1) Рекомендуемая литература

а) Основная литература

1. Программирование на языке высокого уровня. Программирование на языке Object Pascal : учеб. пособие / Т.И. Немцова, С.Ю. Голова, И.В. Абрамова ; под ред. Л.Г. Гагариной. — М. : ИД «ФОРУМ» : ИНФРА-М, 2018. — 496 с. + Доп. Материалы. - [Электронный ресурс]. - Режим доступа: <http://znanium.com/go.php?id=944326>.
2. Технология программирования: учебное пособие. - Тамбов: Издательство ФГБОУ ВПО «ТГТУ», 2013. - 173 с. - [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=277802>.

б) Дополнительная литература

1. Лавлинский В.В. Технология программирования на современных языках программирования. - Воронеж: Воронежская государственная лесотехническая академия, 2012. - 118 с.- [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=142453>.
2. Грацианова Т.Ю. Программирование в примерах и задачах [Электронный ресурс] : учеб. пособие — Электрон. дан. — Москва : Издательство "Лаборатория знаний", 2016. — 373 с. — Режим доступа: <https://e.lanbook.com/book/90242>.

2) Программное обеспечение

а) Лицензионное программное обеспечение

1. Kaspersky Endpoint Security 10 для Windows Акт на передачу прав №969 18.10.2018
2. Mathcad 15 M010 Акт предоставления прав ИС00000027 от 16.09.2011;
3. MATLAB R2012b Акт предоставления прав № Us000311 от 25.09.2012;

4. Microsoft Visual Studio Ultimate 2013 с обновлением 4 Акт предоставления прав №689 от 05.07.2019 г.

5. Microsoft Windows 10 Enterprise Акт приема-передачи №689 от 05.07.2019 г.

6. MS Office 365 pro plus Акт приема-передачи №689 от 05.07.2019 г.

б) Свободно распространяемое программное обеспечение

1. Adobe Acrobat Reader DC

2. Git version 2.5.2.2

3. Google Chrome бесплатно

4. Lazarus 1.4.0

3) Современные профессиональные базы данных и информационные справочные системы

1. ЭБС "Издательство Лань"

2. ЭБС ZNANIUM.COM

3. ФГБУ "РГБ"

4. ЭБ eLibrary

5. American Institute of Physics

6. American Physical Society - APS Online Journals

7. EBSCO Publishing - INSPEC

8. Web of Science

9. SCOPUS

10. ЭБС "Университетская библиотека онлайн"

ТвГУ имеет подписку на коллекцию из 331 российских журналов в полнотекстовом электронном виде, в том числе:

1. Alma mater (Вестник высшей школы)

2. Вопросы статистики

3. Журнал вычислительной математики и математической физики

4. Известия высших учебных заведений. Математика

5. Известия Российской академии наук. Серия физическая
6. Известия Российской академии наук. Теория и системы управления
7. Инновации в образовании
8. Стандарты и качество
9. Школьные технологии

4) Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. The official home of the Python Programming Language:
<https://www.python.org/>
2. М. Саммерфилд . Программирование на Python. Подробное руководство. СПб.: Символ-Плюс. 2015.-608 с.
3. Петер Флах. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. –М.: ДМК Пресс, 2015.- 400 с.
4. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных: <http://www.machinelearning.ru>
5. С.В. Айвазян, В.С. Мхитарян. Прикладная статистика в задачах и упражнениях. –М.: ЮНИТИ-ДАНА . 2001. -270 с.
6. Библиотека научных вычислений: <http://www.numpy.org/>
7. Python Data Analysis Library: <https://pandas.pydata.org/>
8. Библиотека scikit-learn: <https://scikit-learn.org/stable/>
9. <http://elementy.ru> - «Элементы большой науки»
10. <http://www.astronet.ru/>- Российская астрономическая сеть
11. <https://www.wikipedia.org/> - Википедия - свободная энциклопедия

VI. Методические материалы для обучающихся по освоению дисциплины

Перечень вопросов для самоконтроля

1. Язык программирования Python, библиотеки для анализа и визуализации данных NumPy, matplotlib, pandas
2. Библиотека NumPy: работа с массивами (создание, индексация, срезы, слияние и разбиение).
3. Выполнение вычислений над массивами, агрегирование, операции над массивами, сортировка массивов, структурированные массивы
4. Работа с данными с библиотекой pandas: объекты библиотеки, операции над данными, иерархическая индексация,
5. Работа с данными с библиотекой pandas: объединение набор данных, агрегирование и группировка, сводные таблицы, работа с временными рядами.

6. Визуализация с помощью библиотек matplotlib и seaborn: графики, гистограммы и т.д.
7. Машинное обучение с библиотекой Scikit-learn.
8. Алгоритмы машинного обучения:
9. Задачи классификации (наивная байесовская классификация, гауссов байесовский классификатор, полиномиальный баесовский классификатор)
10. Алгоритмы машинного обучения:
11. Линейная регрессия, регрессия по комбинации базисных функций, вопросы регуляризации
12. Алгоритмы машинного обучения: метод опорных векторов (support vector machines).
13. Задачи машинного обучения: деревья принятия решений и случайные леса.
14. Задачи машинного обучения: кластеризация, методы понижения размерности. Метод главных компонент.
15. Математические основы нейросетевого подхода.
16. Фреймворки и библиотеки TensorFlow, Keras: библиотека глубокого обучения в Python.
17. Глубокое обучение для задач анализа текстов
18. Компьютерное зрение с Python: работа с opencv.
19. Глубокое обучение в технологиях компьютерного зрения.

VII. Материально-техническое обеспечение

Набор учебной мебели, Меловая доска, Переносной ноутбук, Компьютер:(процессор Core i5-2400+монитор LC E2342T (10шт.)

Графопроектор, мультимедийный комплект учебного класса (вариант № 1)
Проектор Casio XJ-M140, кронштейн, кабель, удлинитель, настенный проекц. экран Lumien 180*180.

VIII. Сведения об обновлении рабочей программы дисциплины

№п.п.	Обновленный раздел рабочей программы дисциплины	Описание внесенных изменений	Реквизиты документа, утвердившего изменения
1.			
2.			