

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Галунин Сергей Александрович  
Должность: проректор по учебной работе  
Дата подписания: 04.05.2023 12:02:31  
Уникальный программный ключ:  
08ef34338325bdb0ac5a47baa5472ce36cc3fc3b

Приложение к ОПОП  
«Распределенные интеллектуаль-  
ные системы и технологии»



**СПбГЭТУ «ЛЭТИ»**  
ПЕРВЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ

МИНОБРНАУКИ РОССИИ

федеральное государственное автономное образовательное учреждение высшего образования  
**«Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И.Ульянова (Ленина)»**  
**(СПбГЭТУ «ЛЭТИ»)**

---

## **РАБОЧАЯ ПРОГРАММА**

**ДИСЦИПЛИНЫ**

**«ИНСТРУМЕНТЫ ВИЗУАЛИЗАЦИИ И АНАЛИЗА ДАННЫХ»**

**для подготовки магистров**

**по направлению**

**09.04.01 «Информатика и вычислительная техника»**

**по программе**

**«Распределенные интеллектуальные системы и технологии»**

Санкт-Петербург

2022

## ЛИСТ СОГЛАСОВАНИЯ

Разработчики:

заведующий кафедрой, к.т.н., доцент Кринкин К.В.

старший преподаватель Берленко Т.А.

Рабочая программа рассмотрена и одобрена на заседании кафедры МОЭВМ

15.02.2022, протокол № 2

Рабочая программа рассмотрена и одобрена учебно-методической комиссией

ФКТИ, 24.02.2022, протокол № 2

Согласовано в ИС ИОТ

Начальник ОМОЛА Загороднюк О.В.

## 1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	МОЭВМ
Общая трудоемкость (ЗЕТ)	4
Курс	1
Семестр	1
<b>Виды занятий</b>	
Лекции (академ. часов)	8
Электронные лекции (акад. часов)	8
Практические занятия (академ. часов)	18
Иная контактная работа (академ. часов)	1
Все контактные часы (академ. часов)	27
Самостоятельная работа, включая часы на контроль (академ. часов)	109
Всего (академ. часов)	144
<b>Вид промежуточной аттестации</b>	
Экзамен (курс)	1

## **2 АННОТАЦИЯ ДИСЦИПЛИНЫ**

### **«ИНСТРУМЕНТЫ ВИЗУАЛИЗАЦИИ И АНАЛИЗА ДАННЫХ»**

Дисциплина включает в себя две области, посвященные работе с данными. Рассматриваются инструменты и методы визуализации многомерных данных, включая интерактивные методы, направленные на исследование и оценку структуры и зависимостей в данных. Изучаются инструменты для анализа данных (R, Pandas), которые используются в статистическом обучении. Статистическое обучение относится к набору инструментов для моделирования и понимания сложных наборов данных. Это недавно разработанная область в области статистики, она сочетается с параллельными разработками в области компьютерных наук и, в частности, машинного обучения. Эта область охватывает многие методы, такие как регрессия лассо и разреженная регрессия, деревья классификации и регрессии, а также метод опорных векторов. С ростом проблем «больших данных» статистическое обучение стало очень востребованной областью во многих научных областях. Дисциплина входит в состав модуля ”Инструментарий реализации систем искусственного интеллекта”.

## **SUBJECT SUMMARY**

### **«DATA VISUALIZATION AND ANALYSIS TOOLS»**

The discipline includes two sections devoted to working with data. Tools and methods for visualizing multidimensional data are considered, including interactive methods aimed at researching and evaluating the structure and dependencies in data. We study data analysis tools (R, Pandas) that are used in statistical learning. Statistical learning refers to a set of tools for modeling and understanding complex datasets. This is a newly developed field in statistics, and is coupled with parallel developments in computer science and, in particular, machine learning. This area covers many techniques such as lasso and sparse regression, classification and regression

trees, and support vector machines. With the rise of big data challenges, statistical learning has become a highly sought-after field in many scientific fields. The discipline is part of the "Tools for the implementation of artificial intelligence systems" module.

## 3 ОБЩИЕ ПОЛОЖЕНИЯ

### 3.1 Цели и задачи дисциплины

1. Дисциплина нацелена на получение теоретических знаний об инструментах визуализации данных и статистического обучения, приобретение практических навыков применения полученных теоретических знаний для решения задач профессиональной деятельности.

2. Задачами дисциплины является приобретение знаний, умений и навыков для решения практических задач, связанных с применением интерактивных методов, направленных на выявление и оценку структуры и зависимостей в данных; применением инструментов для моделирования сложных наборов данных.

3. Получение знаний о методах визуализации многомерных данных и программных инструментах анализа данных, основанных на языке программирования R и программной библиотеке Pandas.

4. Выработка умений:

-применять интерактивные методы, направленные на исследование и оценку структуры и зависимостей в данных;

-использовать статистическое обучение в таких областях как маркетинг, финансы и других областях знаний.

5. Освоение навыков:

-применения методов поиска групп в данных, включая традиционные и современные методы кластерного анализа. методов уменьшения размерностей, включая многомерное масштабирование, нелинейные и другие методы;

-применения методов регрессии лассо и разреженной регрессии, деревьев классификации и регрессии, а также метода опорных векторов.

### **3.2 Место дисциплины в структуре ОПОП**

Дисциплина изучается на основе знаний, полученных при освоении программы бакалавриата или специалитета.

и обеспечивает изучение последующих дисциплин:

1. «Методы обработки данных (классические байесовские фильтры)»

### 3.3 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения образовательной программы обучающийся должен достичь следующие результаты обучения по дисциплине:

<b>Код компетенции/ индикатора компетенции</b>	<b>Наименование компетенции/индикатора компетенции</b>
ПК-9	Способен осуществлять руководство по созданию и развитию систем и комплексов обработки данных, в том числе больших данных, для корпоративных и государственных заказчиков
<i>ПК-9.3</i>	<i>Участствует в создании (модернизации) общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных; хранения наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области искусственного интеллекта</i>

## 4 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

### 4.1 Содержание разделов дисциплины

#### 4.1.1 Наименование тем и часы на все виды нагрузки

№ п/п	Наименование темы дисциплины	Лек, ач	ЭЛек, ач	Пр, ач	ИКР, ач	СР, ач
1	Введение	1	1			
2	Тема 1. Алгоритмы без учителя, алгоритмы с учителем	1	1	3		14
3	Тема 2. Ядерные методы и сенсорные сети	1	1	3		14
4	Тема 3. Статистическое обучение и язык R	1	1	4		14
5	Тема 4. Линейная регрессия и классификация	2	2	4		18
6	Тема 5. Выход за пределы линейности	1	1	4		14
7	Заключение	1	1	0	1	35
	Итого, ач	8	8	18	1	109
	Из них ач на контроль	0	0	0	0	0
	Общая трудоемкость освоения, ач/зе	144/4				

#### 4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	Введение	Место дисциплины в ООП. Виды контроля и аттестации.
2	Тема 1. Алгоритмы без учителя, алгоритмы с учителем	Линейные методы: метод главных компонент обычный, двойной; метрическое многомерное масштабирование; ориентир в метрическом многомерном масштабировании (аппроксимация Нистрома); неотрицательная матричная факторизация. Нелинейные методы: кусочно-линейное представление; ISOMAP; локальные методы многомерного масштабирования; собственные карты Лапласа; связь со спектральной кластеризацией; стохастическое вложение соседей (с $t$ -распределением); глубокие автокодировщики. Единая структура алгоритмов уменьшения размерности. Метод соседних компонент. Метод соответствующих компонент. Максимизация прибыли в выборе ближайшего соседа.

№ п/п	Наименование темы дисциплины	Содержание
3	Тема 2. Ядерные методы и сенсорные сети	Ядерное выравнивание и изучение ядра с помощью полуопределённого программирования. Развёртывание максимальной дисперсии. Раскрашенное развёртывание максимальной дисперсии. Реализация графов и сенсорные сети. Разметка графов (Кластеризация, классификация). Кластеризация микрочипов. Поиск, ранжирование. Обработка изображений. Совместная фильтрация. Многозадачное обучение Выбор переменных и изучение словаря
4	Тема 3. Статистическое обучение и язык R	Что такое статистическое обучение. Оценка точности модели. Введение в R
5	Тема 4. Линейная регрессия и классификация	Простая линейная регрессия. Множественная линейная регрессия. Другие соображения в модели регрессии. Сравнение линейной регрессии с алгоритмом K-ближайших соседей. Обзор классификации. Почему не линейная регрессия. Логистическая регрессия. Линейный дискриминантный анализ. Сравнение методов классификации. Методы передискретизации. Кросс-валидация. Bootstrap. Линейный выбор модели и регуляризация. Выбор подмножества. Методы усадки. Методы уменьшения размерности. Соображения в больших размерностях.
6	Тема 5. Выход за пределы линейности	Полиномиальная регрессия. Степ-функция. Базисные функции. Регрессионные сплайны. Сглаживание сплайнами. Локальная регрессия. Обогащенные аддитивные модели. Древовидные методы. Основы деревьев выбора. Бэггинг, случайные леса. Метод опорных векторов. Классификатор опорных векторов. Соотношение с логистической регрессией.
7	Заключение	Основные выводы по курсу

## 4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

## 4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Линейные методы.	3
2. Нелинейные методы.	3
3. Ядерные методы.	3
4. Реализация приложений	9
Итого	18

#### **4.4 Курсовое проектирование**

Курсовая работа (проект) не предусмотрены.

#### **4.5 Реферат**

Реферат не предусмотрен.

#### **4.6 Индивидуальное домашнее задание**

Индивидуальное домашнее задание не предусмотрено.

#### **4.7 Доклад**

Доклад не предусмотрен.

#### **4.8 Кейс**

Кейс не предусмотрен.

#### **4.9 Организация и учебно-методическое обеспечение самостоятельной работы**

Изучение дисциплины сопровождается самостоятельной работой студентов с рекомендованными преподавателем литературными источниками и информационными ресурсами сети Интернет.

Планирование времени для изучения дисциплины осуществляется на весь период обучения, предусматривая при этом регулярное повторение пройденного материала. Обучающимся, в рамках внеаудиторной самостоятельной работы, необходимо регулярно дополнять сведениями из литературных источников материал, законспектированный на лекциях. При этом на основе изучения рекомендованной литературы целесообразно составить конспект основных положений, терминов и определений, необходимых для освоения разделов учебной

дисциплины.

Особое место уделяется консультированию, как одной из форм обучения и контроля самостоятельной работы. Консультирование предполагает особым образом организованное взаимодействие между преподавателем и студентами, при этом предполагается, что консультант либо знает готовое решение, которое он может предписать консультируемому, либо он владеет способами деятельности, которые указывают путь решения проблемы.

<b>Текущая СРС</b>	<b>Примерная трудоемкость, ач</b>
Работа с лекционным материалом, с учебной литературой	35
Опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	0
Самостоятельное изучение разделов дисциплины	0
Выполнение домашних заданий, домашних контрольных работ	0
Подготовка к лабораторным работам, к практическим и семинарским занятиям	39
Подготовка к контрольным работам, коллоквиумам	0
Выполнение расчетно-графических работ	0
Выполнение курсового проекта или курсовой работы	0
Поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	0
Работа над междисциплинарным проектом	0
Анализ данных по заданной теме, выполнение расчетов, составление схем и моделей, на основе собранных данных	0
Подготовка к зачету, дифференцированному зачету, экзамену	35
<b>ИТОГО СРС</b>	<b>109</b>

## 5 Учебно-методическое обеспечение дисциплины

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

№ п/п	Название, библиографическое описание	К-во экз. в библ.
Основная литература		
1	Роберт И. R в действии. Анализ и визуализация данных в программе R [Электронный ресурс] / И. Роберт, Кабаков, 2014. -588 с.	неогр.
2	Мастицкий С. Э. Статистический анализ и визуализация данных с помощью R [Электронный ресурс], 2015. -496 с.	неогр.
Дополнительная литература		
1	Эрик Мэтиз Изучаем Python: программирование игр, визуализация данных, веб-приложения. 3-е изд. [Электронный ресурс] / Мэтиз Эрик, 2021. -512 с.	неогр.

### 5.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», используемых при освоении дисциплины

№ п/п	Электронный адрес
1	Машинное обучение <a href="https://machinelearningmastery.ru/">https://machinelearningmastery.ru/</a>
2	Основы программирования на R <a href="https://stepik.org/course/497/">https://stepik.org/course/497/</a>

### 5.3 Адрес сайта курса

Адрес сайта курса: <https://vec.etu.ru/moodle/course/view.php?id=12944>

## 6 Критерии оценивания и оценочные материалы

### 6.1 Критерии оценивания

Для дисциплины «Инструменты визуализации и анализа данных» формой промежуточной аттестации является экзамен. Оценивание качества освоения дисциплины производится с использованием рейтинговой системы.

#### Экзамен

Оценка	Количество баллов	Описание
Неудовлетворительно	0 – 49	теоретическое содержание курса не освоено, необходимые практически навыки и умения не сформированы, выполненные учебные задания содержат грубые ошибки, дополнительная самостоятельная работа над курсом не приведет к существенному повышению качества выполнения учебных заданий
Удовлетворительно	50 – 69	теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки и умения работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий содержат ошибки
Хорошо	70 – 84	теоретическое содержание курса освоено полностью, без пробелов, некоторые практические навыки и умения сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками
Отлично	85 – 100	теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки и умения сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено количеством баллов, близким к максимальному

## Особенности допуска

Для допуска к экзамену необходимо посетить не менее 80% лекций и практических занятий, выполнить и защитить 4 практические работы с начислением баллов за каждую. Максимальное количество баллов -100. Для допуска к экзамену необходимо набрать не менее 50 баллов. Экзамен проводится по билетам с учетом результатов текущего контроля.

## 6.2 Оценочные материалы для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине

### Вопросы к экзамену

№ п/п	Описание
1	Метод главных компонент обычный, двойной
2	Метрическое многомерное масштабирование
3	Локальные методы многомерного масштабирования
4	Метод соседних компонент.
5	Реализация графов и сенсорные сети.
6	Кластеризация микрочипов.
7	Обработка изображений. Совместная фильтрация.
8	Многозадачное обучение. Выбор переменных и изучение словаря
9	Что такое статистическое обучение.
10	Оценка точности модели.
11	Введение в R
12	Сравнение линейной регрессии с алгоритмом K-ближайших соседей.
13	Линейный дискриминантный анализ.
14	Методы передискретизации.
15	Линейный выбор модели и регуляризация.
16	Методы уменьшения размерности.
17	Полиномиальная регрессия.
18	Локальная регрессия.
19	Древовидные методы.
20	Классификатор опорных векторов.

### Форма билета

Министерство науки и высшего образования Российской Федерации  
ФГАОУ ВО «Санкт-Петербургский государственный электротехнический

## ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 1

Дисциплина **Инструменты визуализации и анализа данных** ФКТИ

1. Линейные методы: метод главных компонент обычный, двойной; метрическое многомерное масштабирование; аппроксимация Нистрома; неотрицательная матричная факторизация.

2. Множественная линейная регрессия.

УТВЕРЖДАЮ

Заведующий кафедрой

К.В. Кринкин

**Образцы задач (заданий) для контрольных (проверочных) работ**

**Примерные вопросы при защите практических работ**

1. Как разбить данные на обучающую и контрольную выборку?
2. Какая ошибка линейной модели по методу наименьших квадратов получается для контрольной выборки?
3. Вычислите ошибку лассо-модели по контрольной выборке.
4. Примените метод отбора оптимального подмножества переменных к обучающей выборке и постройте график зависимости MSE на обучающей выборке от размера оптимальной модели
5. Каковы отличия модели с наименьшим значением MSE на контрольной выборке от истинной модели, использованной для создания данных?

Весь комплект контрольно-измерительных материалов для проверки сфор-

мированности компетенции (индикатора компетенции) размещен в закрытой части по адресу, указанному в п. 5.3

### 6.3 График текущего контроля успеваемости

Неделя	Темы занятий	Вид контроля
1	Тема 1. Алгоритмы без учителя, алгоритмы с учителем	
2		
3		
4		Практическая работа
5	Тема 2. Ядерные методы и сенсорные сети	
6		
7		
8		Практическая работа
9	Тема 3. Статистическое обучение и язык R	
10		
11		
12		Практическая работа
13	Тема 4. Линейная регрессия и классификация Тема 5. Выход за пределы линейности	
14		
15		
16		Практическая работа

### 6.4 Методика текущего контроля

#### на лекционных занятиях

Текущий контроль включает в себя контроль посещаемости (не менее 80% занятий).

#### на практических занятиях

Текущий контроль включает в себя:

- контроль посещаемости (не менее 80 % занятий);
- выполнение и защита **4 практических работ**. Под выполнением работ подразумевается подготовка к работе, выполнение задания, подготовка отчета и его защита. Отчет оформляется после выполнения задания и представляется преподавателю на проверку в электронном виде. После проверки отчет либо возвращается (при наличии замечаний) на доработку, либо допускается к защите.

Работы защищаются студентами индивидуально во время практических занятий. На защите студент должен показать: понимание постановки задачи, подхода к ее решению, умение объяснять ход решения, выбор тех или иных методик, технологий для решения задачи, качество оформления отчетных материалов. Преподаватель задает вопросы, позволяющие определить глубину понимания теоретического материала практической работы, а также самостоятельность ее выполнения.

**Максимальное количество баллов за 4 практические работы - 100.**

Практические работы оцениваются по следующим критериям:

- *Работа не выполнена* - 0 баллов;
- *Понимание теоретического материала* - 1 балл: наличие ошибок в изложении теоретического материала, 2 балла: хорошо владеет теоретическим материалом;
- *Умение объяснить постановку задачи* - 1 балл: наличие ошибок в последовательности объяснения, 2 балла: полное и логично последовательное объяснение;
- *Умение объяснить ход решения задачи* - 1 балл: наличие ошибок в последовательности объяснения, 2 балла: полное и логично последовательное объяснение;
- *Умение объяснить выбор методики, технологий* - 1 балл: наличие ошибок в последовательности объяснения, 2 балла: полное и логично последовательное объяснение;
- *Качество оформления отчетных материалов* - 1 балл: наличие ошибок и небрежность в оформлении, 2 балла: качественное оформление в соответствии с требованиями, предъявляемыми к оформлению студенческих работ;
- *Доля оригинальности* при проверке в системе обнаружения заимствований Antiplagiat.ru:

- 70 - 80% - 5 баллов;
- 81 - 90% - 10 баллов;
- 91 - 95% - 15 баллов;

Набранное количество баллов соответствует следующим оценкам:

- Неудовлетворительно - менее 60 баллов;
- Удовлетворительно – от 60 до 75 баллов;
- Хорошо – от 76 до 85 баллов;
- Отлично – от 86 до 100 баллов.

В ходе проведения практических занятий целесообразно привлечение студентов к как можно более активному участию в дискуссиях, решении задач, обсуждениях и т. д. При этом активность студентов также может учитываться преподавателем, как один из способов текущего контроля на практических занятиях.

#### **самостоятельной работы студентов.**

Контроль самостоятельной работы студентов осуществляется по методикам, описанным выше.

**Итоговая оценка** формируется как среднее арифметическое оценок текущего контроля (4 практические работы) и ответов на вопросы билета, округленное по правилам математического округления.

## 7 Описание информационных технологий и материально-технической базы

Тип занятий	Тип помещения	Требования к помещению	Требования к программному обеспечению
Лекция	Лекционная аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя, IBM-совместимый компьютер Pentium или выше проектор, экран, меловая или маркерная доска	1) Linux Альт Образование 10 и выше; 2) P7-Офис 7 и выше либо LibreOffice 7 и выше
Практические занятия	Аудитория	Количество посадочных мест, оборудованных персональными IBM-совместимыми компьютерами Pentium или выше в соответствии с контингентом, рабочее место преподавателя, IBM-совместимый компьютер Pentium или выше проектор, экран, меловая или маркерная доска	1) Linux Альт Образование 10 и выше; 2) P7-Офис 7 и выше либо LibreOffice 7 и выше
Самостоятельная работа	Помещение для самостоятельной работы	Оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.	1) Linux Альт Образование 10 и выше; 2) P7-Офис 7 и выше либо LibreOffice 7 и выше

## **8 Адаптация рабочей программы для лиц с ОВЗ**

Адаптированная программа разрабатывается при наличии заявления со стороны обучающегося (родителей, законных представителей) и медицинских показаний (рекомендациями психолого-медико-педагогической комиссии). Для инвалидов адаптированная образовательная программа разрабатывается в соответствии с индивидуальной программой реабилитации.

## ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

<b>№ п/п</b>	<b>Дата</b>	<b>Изменение</b>	<b>Дата и номер протокола заседания УМК</b>	<b>Автор</b>	<b>Начальник ОМОЛА</b>