

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Галунин Сергей Александрович
Должность: проректор по учебной работе
Дата подписания: 28.06.2023 14:55:53
Уникальный программный ключ:
08ef34338325bdb0ac5a47baa5472ce36cc3fc3b

Приложение к ОПОП
«Математические методы в ин-
формационных технологиях»



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

«МАТЕМАТИЧЕСКИЕ ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ»

для подготовки бакалавров

по направлению

01.03.02 «Прикладная математика и информатика»

по профилю

«Математические методы в информационных технологиях»

Санкт-Петербург

2023

ЛИСТ СОГЛАСОВАНИЯ

Разработчики:

доцент, к.ф.-м.н. Рыбин С.В.

Рабочая программа рассмотрена и одобрена на заседании кафедры АМ
12.01.2023, протокол № 6

Рабочая программа рассмотрена и одобрена учебно-методической комиссией
ФКТИ, 16.02.2023, протокол № 2

Согласовано в ИС ИОТ

Начальник ОМОЛА Загороднюк О.В.

1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	АМ
Общая трудоемкость (ЗЕТ)	3
Курс	3
Семестр	6
Виды занятий	
Лекции (академ. часов)	34
Практические занятия (академ. часов)	34
Иная контактная работа (академ. часов)	1
Все контактные часы (академ. часов)	69
Самостоятельная работа, включая часы на контроль (академ. часов)	39
Всего (академ. часов)	108
Вид промежуточной аттестации	
Экзамен (курс)	3

2 АННОТАЦИЯ ДИСЦИПЛИНЫ

«МАТЕМАТИЧЕСКИЕ ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ»

Курс призван познакомить студентов с основами машинного обучения, включая дискриминантный, кластерный и регрессионный анализ, овладение навыками практического решения задач интеллектуального анализа данных.

Курс должен подготовить специалистов в области анализа данных и машинного обучения, владеющих теоретическими основами методов машинного обучения, обладающих навыками построения обучающихся моделей с использованием современных программных средств и способных применять методы машинного обучения для решения прикладных задач.

Студенты научатся пользоваться современными аналитическими инструментами и адаптировать их под особенности конкретных задач.

В курсе рассматриваются основные понятия машинного обучения: постановки задач обучения, методы подготовки данных для обучения, принципы обучения, методы статистического анализа обучающих данных и результатов обучения, методы оценки качества обученных моделей

Слушатели подробно знакомятся с алгоритмами для решения задач классификации, регрессии и кластеризации. В число изучаемых алгоритмов классификации входят алгоритмы ближайшего соседа, SVM, байесовские методы, деревья решений.

SUBJECT SUMMARY

«MATHEMATICAL FOUNDATIONS OF MACHINE LEARNIN»

The course aims to introduce students to the basics of machine learning, including discriminant, cluster and regression analysis, mastering the skills of practical solutions to data mining problems.

The course should prepare specialists in the field of data analysis and machine learning, who possess the theoretical foundations of machine learning methods, have the

skills to build student models using modern software tools and are able to apply machine learning methods for solving applied problems.

Students learn how to use modern analytical tools and adapt them to the characteristics of specific tasks.

The course discusses the basic concepts of machine learning: the formulation of training tasks, methods of preparing data for training, learning principles, methods of statistical analysis of training data and learning outcomes, methods of assessing the quality of trained models

Students learn in detail about algorithms for solving classification, regression and clustering problems. The classification algorithms studied include nearest neighbor algorithms, SVM, Bayesian methods, decision trees.

3 ОБЩИЕ ПОЛОЖЕНИЯ

3.1 Цели и задачи дисциплины

1. Целями изучения дисциплины является приобретение теоретических знаний в области классических методов машинного обучения и формирование практических умений и навыков применения алгоритмов классификации, кластеризации, дисперсионного и регрессионного анализа.

2. Задачами изучения дисциплины является получение знаний о существующих подходах в области обработки данных методами машинного обучения, умений обоснования применимости конкретных методов к рассматриваемым эмпирическим данным и владения навыками написания, применения и модификации исходного кода существующих библиотек машинного обучения (scikits-learn, scipy, numpy).

3. Знания:

- методов обработки и визуализации данных на языке Python;
- методов оценки качества работы алгоритмов анализа данных;
- методик построения цикла обработки данных.

4. Умения:

- формировать обучающую выборку;
- планировать эксперименты по машинному обучению.

5. Навыки работы в интерактивной среде Jupyter-notebook.

3.2 Место дисциплины в структуре ОПОП

Дисциплина изучается на основе ранее освоенных дисциплин учебного плана:

1. «Дискретная математика и теоретическая информатика»
2. «Объектно-ориентированное программирование»

3. «Комбинаторика и теория графов»

4. «Теория вероятностей и математическая статистика»

и обеспечивает изучение последующих дисциплин:

1. «Большие данные»

2. «Машинное обучение с подкреплением»

3. «Обработка аудиосигналов»

4. «Современные архитектуры глубоких искусственных нейронных сетей»

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:

Код компетенции/ индикатора компетенции	Наименование компетенции/индикатора компетенции
СПК-11	Способен использовать различные математические формализации машинного обучения для создания и развития технологий искусственного интеллекта
<i>СПК-11.1</i>	<i>Разрабатывает модели решения прикладных задач на основе статистической формализации машинного обучения</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	Пр, ач	ИКР, ач	СР, ач
1	Введение в машинное обучение	8	4		9
2	Алгоритмы кластеризации	8	8		10
3	Методы классификации	10	14	1	11
4	Методы регрессии	8	8		9
	Итого, ач	34	34	1	39
	Из них ач на контроль	0	0	0	35
	Общая трудоемкость освоения, ач/зе	108/3			

4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	Введение в машинное обучение	Этапы развития классического ML. Основные компоненты ML. Основные виды ML. Исходные данные. Предобработка и извлечение признаков. Виды и типы признаков. Кодирование категориальных признаков. Открытые наборы данных. Ресурсы для поиска в ML. Сообщества по ML. Библиотеки для ML в Python. Метод главных компонент Эффективная размерность PCA: правило сломанной трости, правило Кайзера, оценка по числу обусловленности. Достоинства и недостатки PCA.

№ п/п	Наименование темы дисциплины	Содержание
2	Алгоритмы кластеризации	<p>Формальная постановка задачи. Алгоритм k-средних. Выбор числа кластеров. ”Метод локтя” Силуэтный метод. Алгоритм DBSCAN. Достоинства и недостатки. Метрики расстояния. Агломеративная (восходящая) кластеризация. Метрики сходства. Дендрограмма. Дивизимная (нисходящая) кластеризация. Кластеризация с помощью графов. Оценка качества кластеризации. Внешние: индексы Рэнда, Жаккара, Фоулкса-Мэллоуза, F-мера. Внутренние: компактность кластеров, индексы Данна, Дэвиса-Болдуина.</p>
3	Методы классификации	<p>Формальная постановка задачи классификации. Переобучение и недообучение. Типы классификации. Метод опорных векторов. Выбор гиперплоскости Линейная неразделимость. Спрямяющее отображение. Трюк с ядром. Часто используемые ядра. Мультиклассовый метод опорных векторов: каждый против всех, каждый против каждого. Классификатор k-ближайших соседей. Взвешенный KNN. Достоинства и недостатки. Деревья решений. Построение дерева решений. Критерий Джини. Алгоритм CART. Ансамбли. Random Forest. Байесовский классификатор. Теорема Байеса. Классификация спама. Преимущества и недостатки. Оценка качества в бинарной классификации. Матрица ошибок, Accuracy. Balanced Accuracy. Precision, Recall, F-score. ROC-кривая, AUC. Оценка качества многоклассовой классификации. Кросс-валидация. Балансировка классов Сокращение мажоритарного класса Увеличение миноритарного класса</p>

№ п/п	Наименование темы дисциплины	Содержание
4	Методы регрессии	Линейная регрессия. Постановка задачи. Парная линейная регрессия. Многомерная линейная регрессия. Точное решение. Сингулярное разложение. Градиентный спуск (пакетный и стохастический). Регуляризация. Гребневая регрессия. Регуляризация LASSO. Эластичные сети. Метрики качества линейной регрессии. Логистическая регрессия. Регуляризация для логистической регрессии. Метод максимального правдоподобия.

4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Освоение интерактивной среды Jupyter-notebook.	2
2. Обработка и визуализация данных на языке Python.	2
3. Кластеризация алгоритмом k – средних на базе музыкальных фрагментов. Выборка из открытого набора данных Free Music Archive.	4
4. Агломеративная кластеризация на базе акустических шумов – на основе открытого набора данных Mivvia Audio Events Dataset.	4
5. Соревнования по анализу данных на платформе KAGGLE: House Prices.	6
6. Соревнования по анализу данных на платформе KAGGLE: Digit Recognizer.	6
7. Соревнования по анализу данных на платформе KAGGLE: Titanic.	6
8. Классификации речи по эмоциональным состояниям на базе The Ryerson Audio-Visual Database of Emotional Speech and Song.	4
Итого	34

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 учебными, научными источниками и информационными ресурсами сети Интернет.

Планирование времени для изучения дисциплины осуществляется на весь период обучения. 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.

Одной из форм обучения и контроля самостоятельной работы является консультирование по новым научным направлениям, осуществляемое руководителями научных направлений работы кафедры, реализующей данную дисциплину.

плину. Обычное консультирование предполагает особым образом организованное взаимодействие между преподавателем и студентами, при этом предполагается, что консультант знает готовое решение, которое он может предписать консультируемому, и он владеет способами деятельности, которые указывают путь решения проблемы. Научное же консультирование в рамках личного общения или в рамках заседания научного семинара не предполагает знание готовых решений, так как связано с актуальной научной проблематикой.

Текущая СРС	Примерная трудоемкость, ач
Работа с лекционным материалом, с учебной литературой	0
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	4
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
ИТОГО СРС	39

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 библи.
Основная литература		
1	Николенко С. Глубокое обучение [Электронный ресурс] / С. Николенко, А. Кадурын, Е. Архангельская, 2019. -480 с.	неогр.
2	Рашка С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения [Электронный ресурс], 2017. -418 с.	неогр.
3	Андрей Бурков Машинное обучение без лишних слов [Электронный ресурс] / Бурков Андрей, 2020. -192 с.	неогр.
4	Элбон, Крис. Машинное обучение с использованием Python. Сборник рецептов [Текст] : [пер. с англ.] / К. Элбон, 2020. -369 с.	20
Дополнительная литература		
1	Литвинова, Виктория Викторовна. Лекции по теории вероятностей. Случайные события [Текст] : учеб. пособие / В. В. Литвинова, Ж. Т. Беленкова, Е. А. Толкачева, 2020. -30, [1] с.	300
2	Письменный, Дмитрий Трофимович. Конспект лекций по теории вероятностей, математической статистике и случайным процессам [Текст] / Д.Т. Письменный, 2010. -287 с.	18
3	Гладков Л. Л. Теория вероятностей и математическая статистика [Электронный ресурс] : учебное пособие / Л. Л. Гладков, Г. А. Гладкова, 2020. -196 с.	неогр.
4	Шакла Нишант Машинное обучение и TensorFlow [Электронный ресурс] / Нишант Шакла, 2019. -336 с.	неогр.

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

№ п/п	Электронный адрес
1	Курс С.И. Николенко https://logic.pdmi.ras.ru/~sergey/teaching/mlhse2017.html
2	Вики-ресурс по машинному обучению www.MachineLearning.ru
3	Набор данных Free Music Archive https://github.com/mdeff/fma
4	Набор данных Mivia Audio Events Dataset https://mivia.unisa.it/datasets/audio-analysis/mivia-audio-events/

№ п/п	Электронный адрес
5	Данные для соревнования по анализу данных на платформе KAGGLE: House Prices https://www.kaggle.com/c/house-prices-advanced-regression-techniques
6	Данные для соревнования по анализу данных на платформе KAGGLE: Digit Recognizer https://www.kaggle.com/c/digit-recognizer
7	Данные для соревнования по анализу данных на платформе KAGGLE: Titanic https://www.kaggle.com/c/titanic
8	Набор данных The Ryerson Audio-Visual Database of Emotional Speech and Song https://zenodo.org/record/1188976#.YRZcnEBn1aQ

5.3 Адрес сайта курса

Адрес сайта курса: <https://vec.etu.ru/moodle/course/view.php?id=13120>

6 Критерии оценивания и оценочные материалы

6.1 Критерии оценивания

Для дисциплины «Математические основы машинного обучения» предусмотрены следующие формы промежуточной аттестации: экзамен.

Экзамен

Оценка	Описание
Неудовлетворительно	Курс не освоен. Студент испытывает серьезные трудности при ответе на ключевые вопросы дисциплины
Удовлетворительно	Студент в целом овладел курсом, но некоторые разделы освоены на уровне определений и формулировок
Хорошо	Студент демонстрирует полное овладение курсом, способен применять полученные знания при решении конкретных задач
Отлично	Студент демонстрирует полное овладение курсом, способен применять полученные знания при решении конкретных задач.

Особенности допуска

Форма проведения, регламент экзамена и правила выставления оценки определяются лектором и сдаются на кафедру вместе с экзаменационными вопросами до начала экзаменационной сессии.

Текущая аттестация студентов может учитываться при проведении экзамена в одном из двух видов:

-«допуск» до экзамена. Допущенными до экзамена считаются студенты, получившие по итогам текущей аттестации оценку выше порогового уровня (имеющие в сводных электронных ведомостях кафедры итоговую оценку 3, 4, 5).

-«часть экзамена». Доля участия суммарной многобалльной оценки текущей аттестации (оценки за практические занятия) в экзаменационной оценке определяется лектором и доводится до сведения студентов и ассистентов в начале семестра.

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

Вопросы к экзамену

№ п/п	Описание
1	Этапы развития классического машинного обучения.
2	Основные компоненты и виды ML.
3	Предобработка и извлечение признаков. Виды и типы признаков. Кодирование категориальных признаков.
4	Метод главных компонент.
5	Эффективная размерность PCA.
6	Правило сломанной трости. Правило Кайзера. Оценка по числу обусловленности.
7	Достоинства и недостатки PCA.
8	Описательная статистика. Диаграммы рассеивания. Гистограммы и графики плотности. Проверка на нормальное распределение.
9	Алгоритм k-средних. Модификации.
10	Выбор числа кластеров. "Метод локтя". Силуэтный метод.
11	Алгоритм DBSCAN. Достоинства и недостатки. Метрики расстояния.
12	Агломеративная (восходящая) кластеризация. Метрики сходства. Дендрограмма.
13	Дивизимная (нисходящая) кластеризация.
14	Кластеризация с помощью графов.

15	Оценка качества кластеризации.
16	Индексы Рэнда, Жаккара, Фоулкса-Мэллоуза, F-мера.
17	Компактность кластеров, индексы Данна, Дэвиса-Болдуина.
18	Переобучение и недообучение. Типы классификации.
19	Метод опорных векторов. Выбор гиперплоскости. Линейная неразделимость. Спрямяющее отображение.
20	Трюк с ядром. Часто используемые ядра.
21	Мультиклассовый метод опорных векторов. Каждый против всех. Каждый против каждого.
22	Классификатор k-ближайших соседей (KNN). Взвешенный KNN. Достоинства и недостатки.
23	Деревья решений. Построение дерева решений. Критерий Джини.
24	Алгоритм CART. Ансамбли. Random Forest.
25	Байесовский классификатор. Теорема Байеса. Номинальные признаки. Классификация спама.
26	Преимущества и недостатки. Оценка качества в бинарной классификации.
27	Матрица ошибок, Ассигасу. Balanced Accuracy. Precision. Recall. F-score. ROC-кривая, AUC.
28	Оценка качества многоклассовой классификации. Кросс-валидация.
29	Балансировка классов. Методы сокращения мажоритарного класса. Методы увеличения миноритарного класса.
30	Линейная регрессия. Постановка задачи.
31	Парная линейная регрессия.
32	Многомерная линейная регрессия. Точное решение. Сингулярное разложение. Градиентный спуск (пакетный и стохастический).
33	Регуляризация. Гребневая регрессия. Регуляризация LASSO. Эластичные сети.
34	Метрики качества линейной регрессии.
35	Логистическая регрессия. Постановка задачи. Типы логистической регрессии.
36	Регуляризация для логистической регрессии.
37	Метод максимального правдоподобия.

Форма билета

Министерство науки и высшего образования Российской Федерации
 ФГАОУ ВО «Санкт-Петербургский государственный электротехнический
 университет «ЛЭТИ» имени В.И. Ульянова (Ленина)»

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № _____

Дисциплина **Математические основы машинного обучения** ФКТИ

1. Метод опорных векторов. Выбор гиперплоскости. Линейная неразделимость. Спрямяющее отображение.

2. Балансировка классов. Методы сокращения мажоритарного класса. Методы увеличения миноритарного класса.

УТВЕРЖДАЮ

Заведующий кафедрой

С.Н. Поздняков

Весь комплект контрольно-измерительных материалов для проверки сформированности 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	Введение в машинное обучение	
2		Практическая работа
3	Алгоритмы кластеризации	
4		
5		
6		Практическая работа
7	Методы классификации	
8		
9		
10		
11		
12		Практическая работа
13	Методы регрессии	
14		
15		
16		Практическая работа

6.4 Методика текущего контроля

На лекционных занятиях.

Текущий контроль включает в себя контроль посещаемости (не менее **66** % занятий).

На практических занятиях

В процессе обучения по дисциплине «**Математические основы машинного обучения**» студент обязан выполнить **все** практические задания по темам курса. Предусматривается проверка заданий преподавателем, а далее защита выполненной работы и работа над ошибками, если это необходимо. Оформление решений студентами осуществляется индивидуально в формате **Jupyter-notebook**.

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**.

Методика оценивания практических заданий:

”неудовлетворительно” (или 2), если получена оценка менее **0.6**;

”удовлетворительно” (или 3), если получена оценка не менее **0.6**;

”хорошо” (или 4), если получена оценка не менее **0.7**;

”отлично” (или 5), если получена оценка не менее **0.8**.

самостоятельной работы студентов

Контроль самостоятельной работы студентов осуществляется на лекционных и практических занятиях студентов по методикам, описанным выше.

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, проектор, компьютер или ноутбук, экран, маркерная или меловая доска.	Свободно распространяемое ПО или ПО, разработанное в РФ, соответствующее по характеристикам Windows XP, Microsoft Office 2007 и выше, Jupyter Notebook.
Практические занятия	Аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя, проектор, компьютер или ноутбук, экран, маркерная или меловая доска.	Свободно распространяемое ПО или ПО, разработанное в РФ, соответствующее по характеристикам Windows XP, Microsoft Office 2007 и выше, Jupyter Notebook.
Самостоятельная работа	Помещение для самостоятельной работы	Оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.	Свободно распространяемое ПО или ПО, разработанное в РФ, соответствующее по характеристикам Windows XP, Microsoft Office 2007 и выше, Jupyter Notebook.

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

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