

Федеральное государственное бюджетное образовательное
учреждение высшего образования
Московский государственный университет имени М.В. Ломоносова
ФИЛИАЛ МГУ В Г. ДУБНЕ

УТВЕРЖДАЮ

И.о. директора
филиала МГУ в г.Дубне
/ Э.Э. Боос /
«01» сентября 2024 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины:

Обработка экспериментальных данных

Уровень высшего образования:

Магистратура

Направление подготовки:

03.04.02 Физика

Направленность (профиль)/специализация образовательной программы:

Физика элементарных частиц

Форма обучения:

Очная

Дубна 2024

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению 03.04.02 «Физика», утвержденным приказом МГУ от 30.12.2020 г. № 1366.

Год (годы) приема на обучение _____

Авторы–составители:

1. д.т.н. Кореньков В.В.
2. к.т.н. Ужинский А.В.

Руководитель магистерской программы:

Доктор физ.-мат. наук, профессор академик РАН В.А. Матвеев, заведующий кафедрой физического факультета МГУ

Аннотация к рабочей программе дисциплины

Содержанием дисциплины являются основной математический формализм и современные методы анализа экспериментальных данных физики высоких энергий, в которых в недавнем времени произошла революция в методике, которая привела к повсеместному использованию машинного обучения и нейронных сетей. Курс является введением в популярные направления обработки данных с помощью этих средств.

Дисциплина реализуется на 2 курсе в 3 семестре и входит в состав вариативной части.

Объем дисциплины составляет 2 з.е., в том числе 36 академических часа, отведенных на контактную работу обучающихся с преподавателем, 36 академических часа, отведенных на самостоятельную работу обучающихся.

Промежуточная аттестация по дисциплине (модулю) – зачет в 3 семестре.

1. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Обработка экспериментальных данных» входит в состав вариативной части и реализуется в 3 семестре 2 года обучения.

2. Входные требования для освоения дисциплины (модуля), предварительные условия

Базовые знания в области общей физики, высшей математики и линейной алгебры в объеме классических университетских курсов

3. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников

В результате освоения дисциплины у обучающихся должны быть сформированы:

Компетенции	Результаты обучения
ОПК-4	<p><u>Знать</u> типовые процедуры применения проблемно-ориентированных прикладных программных средств в области современной ядерной физики</p> <p><u>Уметь</u> использовать современные информационные и компьютерные технологии, средства коммуникаций, способствующие повышению эффективности научной деятельности в области современной ядерной физики</p> <p><u>Владеть</u> методами научного моделирования при решении поставленных исследовательских задач с использованием современных информационных технологий</p>
ПК-2	<p><u>Знать</u> современное программное и аппаратное обеспечение информационных и систем, используемых в профильной области научного исследования</p> <p><u>Уметь</u> модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем для решения поставленной научной задачи, проводить разведочный анализ данных (exploratory data analysis, EDA)</p> <p><u>Владеть</u> навыками адаптирования программного и аппаратного обеспечения информационных и автоматизированных систем для решения поставленной научной задачи, разработки и анализа моделей машинного обучения с библиотекой Scikit-learn</p>
МПК-3	<p><u>Знать</u> основные используемые функции распределения, их свойства, а также математические методы, используемые при обработке экспериментальных данных, методы и технологии построения нейронных сетей</p> <p><u>Уметь</u> создавать модели нейронных сетей прямого и обратного распространения для обработки данных</p> <p><u>Владеть</u> методами оценки точности экспериментальных данных, методами их обработки, разработки гетерогенных нейронных сетей для решения научных задач крупных проектов, включая проекты класса мегасайенс</p>

4. Объем дисциплины (модуля) составляет 2 з.е., в том числе: 36 академических часов, отведенных на контактную работу обучающихся с преподавателем, 36 академических часов, отведенных на самостоятельную работу обучающихся.

Контактная работа включает в себя: занятия лекционного типа и занятия семинарского типа.

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

Наименование и краткое содержание разделов и тем дисциплины (модуля), Форма промежуточной аттестации по дисциплине (модулю)	Всего (ак.ч.)	В том числе					Самостоятельная работа обучающегося, академические часы	Форма текущего контроля успеваемости, наименование
		Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, академические часы ¹						
		Занятия лекционного типа (лекции)	Занятия семинарского типа			Всего		
Семинары	Лабораторные занятия*		Практические занятия*					
Раздел 1. Наука о данных. Инструментарий для работы с данными	8	2	2			4	4	Опрос
Раздел 2. Разведочный анализ данных (exploratory data analysis, EDA)	8	2	2			4	4	Опрос
Раздел 3. Машинное обучение с библиотекой Scikit-learn: обучение с учителем,	8	2	2			4	4	Опрос

¹Текущий контроль успеваемости может быть реализован в рамках занятий лекционного и(или) семинарского типа.

задачи классификации								
Раздел 4. Машинное обучение с библиотекой Scikit-learn: обучение с учителем, задачи восстановления регрессии	10	3	3			6	4	Опрос
Раздел 5. Машинное обучение с библиотекой Scikit-learn: обучение без учителя	10	3	3			6	4	Опрос
Раздел 6. Введение в компьютерное зрение.	8	2	2			4	4	Опрос
Раздел 7. Введение в глубокое обучение.	8	2	2			4	4	Опрос
Раздел 8. Сверточные нейронные сети.	8	2	2			4	4	Опрос
Промежуточная аттестация							4²	Зачет
Итого	72	36					36	

*Лабораторные занятия, практические занятия относятся к практической подготовке обучающихся.

²Часы на проведение промежуточной аттестации выделяются из часов самостоятельной работы обучающегося

Раздел 1. Инструментарий для работы с данными.

Работа в Jupyter Notebook – 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. Основы Python: базовые типы данных, типы коллекций, управляющие структуры и функции. Библиотека NumPy: работа с массивами (создание, индексация, срезы, слияние и разбиение), выполнение вычислений над массивами, агрегирование, операции над массивами, сортировка массивов, структурированные массивы. Визуализация с помощью библиотек matplotlib и seaborn: графики, диаграммы, гистограммы и т.д. Работа с данными с библиотекой Pandas: объекты библиотеки, операции над данными, иерархическая индексация, объединение наборов данных, агрегирование и группировка, сводные таблицы, работа с временными рядами.

Раздел 2. Разведочный анализ данных (exploratory data analysis, EDA).

Предобработка данных, анализ на выявление закономерностей, выбросов. Библиотеки для проведения EDA.

Раздел 3. Машинное обучение с библиотекой Scikit-learn: обучение с учителем, задачи классификации

Алгоритмы машинного обучения: задачи классификации (наивная байесовская классификация, гауссов байесовский классификатор, полиномиальный байесовский классификатор). Метод опорных векторов (Support Vector Machines). Задачи машинного обучения: деревья принятия решений и случайные леса. Инструментарий для интерпретации предсказания моделей: техника SHAP (SHapley Additive exPlanations) и др.

Раздел 4. Машинное обучение с библиотекой Scikit-learn: обучение с учителем, задачи восстановления регрессии

Алгоритмы машинного обучения: линейная регрессия, регрессия по комбинации базисных функций, вопросы регуляризации.

Раздел 5. Машинное обучение с библиотекой Scikit-learn: обучение без учителя

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

Раздел 6. Введение в компьютерное зрение

Компьютерное зрение с Python: алгоритмы компьютерного зрения, обработка изображений и видеоданных с OpenCV, scikitimage и др.

Раздел 7. Введение в глубокое обучение

Математические основы нейросетевого подхода, метод обратного распространения ошибки. Фреймворки и библиотеки для глубокого обучения в Python: TensorFlow, Keras, PyTorch.

Раздел 8. Сверточные нейронные сети.

Глубокое обучение в технологиях компьютерного зрения. Нейросетевые архитектуры для детекции объектов: YOLO, SSD, Mask-R-CNN и др. Особенности, структуры и обучение моделей.

6. Фонд оценочных средств для оценивания результатов обучения по дисциплине (модулю)

6.1. Типовые задания и иные материалы, необходимые для оценки результатов обучения:

Типовые вопросы для проведения текущей проверки успеваемости:

- Создание одномерных и многомерных массивов
- Примеры использования (`np.arange()`, `np.linspace()`, `np.random.normal()` и т.д.):
- Атрибуты массивов (`ndim`, `shape`, `size`)
- Срезы: напечатать строку, столбец и т.д.
- Простые линейные графики: цвета и стили линий, настройка графиков, подписи к осям, метки на графиках;
- Простые диаграммы рассеяния;
- Визуализация погрешностей;
- Объекты библиотеки Pandas: объект `DataFrame`, объект `Series`;
- Объект `Series` как специализированный словарь;
- Как происходит индексация и выборка данных;
- Как происходит чтение данных из CSV файла;
- Создания модели с использованием функционального API;
- Подготовка данных для обучения, создание генератора данных;
- Как происходит обучение нейросети и оценка качества получившейся модели.
- Аугментация данных для обучения;
- Регуляризация сети с помощью слоев Dropout;
- Fine-tuning нейросетевых моделей на основе предварительно обученной сети
- Примеры на использование сверточных нейронных сетей для решения прикладных задач.

Типовые вопросы, задания для проведения промежуточной аттестации (зачета):

1. Привести классификацию задач машинного обучения, примеры использования алгоритмах в различных информационных системах.
2. Дать краткое описание особенностей задач, решаемых методами машинного обучения.
3. Дайте краткое описание возможностей библиотеки Scikit-Learn.
4. Дайте краткую характеристику библиотек для предобработки данных, визуализации данных и построения моделей, их обучения и анализа полученных результатов.
5. Привести примеры линейных моделей.
6. Привести алгоритмы для решения задачи классификации.
7. Дать определение проблемы переобучения и методов ее обнаружения и разрешения (регуляризация).

8. Привести примеры алгоритмов машинного обучения - решающие деревья и их композиции.
9. Привести описание метода K-ближайших соседей.
10. Дать описание метода опорных векторов (SVM) и привести основные этапы его вывода, указать виды ядер в методе опорных векторов.
11. Привести байесовские модели в задачах машинного обучения
12. Привести критерии оценки эффективности моделей машинного обучения в задачах классификации и регрессии.
13. Привести математическую модель биологического нейрона, привести примеры функций активации.
14. Дать определения основных объектов нейронных сети (слои, функции активации, функция потерь, оптимизаторы), дать доказательство алгоритма обратного распространения ошибки.
15. Дать определения основных элементов сверточной нейронной сети: свертка (2D и 3D), pooling.
16. Привести примеры различных архитектур нейронных сетей и классов задач, решаемых на их основе.
17. Дайте описание библиотек и фреймворков для построения нейросетевых моделей, их обучения: TensorFlow, Keras, PyTorch.
18. Привести описание возможностей библиотеки компьютерного зрения OpenCV.
19. Описать основные этапы при работе с данными и применяемый инструментарий (с примерами).
20. Привести описание источников данных и методов их получения.
21. Описать технологии по получению данных и описать подходы к разработке признаков объектов в задачах анализа данных.

6.2. Шкала и критерии оценивания

Результат освоения дисциплины	Критерии оценивания знаний, умений и навыков			
	2/ не зачтено	3/ зачтено	4/ зачтено	5/ зачтено
Знания	Отсутствие знаний	В целом успешные, но не систематические знания	В целом успешное, но содержащие отдельные пробелы знания	Успешные и систематические знания
Умения	Отсутствие умения применять знания фундаментальных и актуальных проблем.	В целом успешное, но не систематическое умение применять знания	В целом успешное, но содержащие отдельные пробелы	Успешное и систематическое умение применять знания

			умение применять знания	
Навыки	Отсутствие/фрагментарные навыки в решении задач	В целом успешные, но не систематические навыки в решении задач	В целом успешные, но содержащие отдельные пробелы навыки в решении задач	Успешные и систематические навыки в решении задач

7. Ресурсное обеспечение

Перечень основной и дополнительной учебной литературы

Основная литература

1. Плас Дж. В. Python для сложных задач: наука о данных и машинное обучение / Плас Дж.Вандер. - СПб.: Питер, 2018. - 576 с.: ил. - (Бестселлеры O'Reilly). - ISBN 978-5-496-03068-7.
2. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / Флах Петер; перевод с английского А. А. Слинкина. - М.: ДМК Пресс, 2015. - 400 с.: ил. - Библиогр.:с.376.-Предм.указ.:с.387. - ISBN 978-1-107-09639-4.
3. Николенко С.И. Глубокое обучение. Погружение в мир нейронных сетей / Николенко С. И., Кадури А. А., Архангельская Е. О. - СПб.: Питер, 2020. - 480 с.: ил. - (Библиотека программиста). - ISBN 978-5-4461-1537-2.
4. Шолле Ф. Глубокое обучение на Python / Шолле Франсуа. - СПб.: Питер, 2018. - 400 с.: ил. - (Библиотека программиста). - ISBN 978-5-4461-0770-4.
5. The official home of the Python Programming Language: <https://www.python.org/>
6. Саммерфилд М. Программирование на Python 3. Подробное руководство. -СПБ.: Символ-Плюс. 2015.-608 с.
7. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных: <http://www.machinelearning.ru>
8. Айвазян С.В., Мхитарян В.С. Прикладная статистика в задачах и упражнениях. –М.: ЮНИТИ-ДАНА . 2001. -270 с.
9. Библиотека научных вычислений: <http://www.numpy.org/>
10. Python Data Analysis Library: <https://pandas.pydata.org/>
11. Библиотека scikit-learn: <https://scikit-learn.org/stable/>
12. Keras: The Python Deep Learning library: <https://keras.io/>

Дополнительная литература

1. Траск Э. Грожаем глубокое обучение / Траск Эндрю. - СПб.: Питер, 2019. - 352 с.: ил. - (Библиотека программиста). - ISBN 978-5-4461-1334-7.
2. Силен Д. Основы Data Science и Big Data. Python и наука о данных / Силен Дэви, Мейсман Арно, Али Мохамед. - СПб.: Питер, 2018. - 338 с.: ил. - (Библиотека программиста). - ISBN 9785496025171.
3. Шарден Б. Крупномасштабное машинное обучение вместе с Python: учитесь быстро развивать мощные модели машинного обучения и развертывать крупномасштабные приложения прогнозирования / Шарден Бастиан, Массарон Лука, Боскетти Альберто; перевод с английского А. В. Логунова. - М.: ДМК Пресс, 2018. - 358 с.: ил. - Предм.указ.:с.350. - ISBN 978-5-97060-506-6.
4. Брюс П. Практическая статистика для специалистов Data Science. 50 важнейших понятий / Брюс Питер, Брюс Эндрю. - СПб.: БХВ-Петербург, 2018. - 304 с.: ил. - Библиогр.:с.295.-Предм.указ.:с.297. - ISBN 978-5-9775-3974-6.
5. Вьюгин В. В. Математические основы машинного обучения и прогнозирования / Вьюгин Владимир Вячеславович; МФТИ. Лаборатория структурных методов анализа данных в предсказательном моделировании (ПреМоЛаб); Институт проблем передачи информации им. А. А. Харкевича РАН. - М.: МЦНМО, 2013. - 304 с. - Лит.:с.301. - ISBN 978-5-4439-0111-4.
6. Джонс М.Т. Программирование искусственного интеллекта в приложениях. Пер.с англ. А.И.Осипова. - М.: Пресс, 2006. - 312с. ISBN 5-94074-275-0. - ISBN 1-58450-278-9.

Перечень профессиональных баз данных и информационных справочных систем

1. <http://www.edu.ru> – портал Министерства образования и науки РФ
2. <http://www.ict.edu.ru> – система федеральных образовательных порталов «ИКТ в образовании»
3. <http://www.openet.ru> - Российский портал открытого образования
4. <http://www.mon.gov.ru> - Министерство образования и науки Российской Федерации
5. <http://www.fasi.gov.ru> - Федеральное агентство по науке и инновациям

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Math-Net.Ru [Электронный ресурс] : общероссийский математический портал / Математический институт им. В. А. Стеклова РАН ; Российская академия наук, Отделение математических наук. - М. : [б. и.], 2010. - Загл. с титул. экрана. - Б. ц. URL: <http://www.mathnet.ru>
2. Университетская библиотека Online [Электронный ресурс] : электронная библиотечная система / ООО "Директ-Медиа" . - М. : [б. и.], 2001. - Загл. с титул. экрана. - Б. ц. URL: www.biblioclub.ru
3. Универсальные базы данных East View [Электронный ресурс] : информационный ресурс / East View Information Services. - М. : [б. и.], 2012. - Загл. с титул. экрана. - Б. ц. URL: www.ebiblioteka.ru
4. Научная электронная библиотека eLIBRARY.RU [Электронный ресурс] : информационный

портал / ООО "РУНЭБ" ; Санкт-Петербургский государственный университет. - М. : [б. и.], 2005. -

Загл. с титул. экрана. - Б. ц.

URL: www.eLibrary.ru

Описание материально-технического обеспечения:

Курс может быть прочитан в поточной аудитории при наличии: работающих электрических розеток, компьютера, проектора, экрана, учебной доски

Для семинарских занятий необходима современная вычислительная техника, объединенную в локальную вычислительную сеть, имеющую выход в Интернет.

Используются специализированные компьютерные классы, оснащенные современным оборудованием.

8. Язык преподавания: русский