

**ГОУ ВПО Российско-Армянский (Славянский)
университет**



УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ДИСЦИПЛИНЫ

Наименование дисциплины: Б1.0.07 «Методы машинного обучения в материаловедении»

Автор (ы) д.ф.-м.н. Мамасахлисов Евгений Шамильевич
Ф.И.О, ученое звание (при наличии), ученая степень (при наличии)

Направление подготовки: 11.04.04 Электроника и наноэлектроника
Магистерская программа: Квантовая и оптическая электроника

Согласовано:

Заведующий Кафедрой общей физики и квантовых наноструктур

Айрапетян Д.Б.



(подпись)

1. АННОТАЦИЯ

1.1. Краткое описание содержания данной дисциплины;

В рамках дисциплины «Методы машинного обучения в материаловедении» изучаются основы современных методов машинного обучения (МО), используемых при исследовании материалов, а также даются представления о различных таких материалах.

1.2. Трудоемкость в академических кредитах и часах, формы итогового контроля (экзамен/зачет);

4 академических кредита / 144 часа. Форма итогового контроля — экзамен.

1.3. Взаимосвязь дисциплины с другими дисциплинами учебного плана специальности (направления)

Квантоворазмерные системы нанoeлектроники, Статистический анализ данных.

1.4. Результаты освоения программы дисциплины:

Код компетенции (в соответствии рабочим с учебным планом)	Наименование компетенции (в соответствии рабочим с учебным планом)	Код индикатора достижения компетенций (в соответствии рабочим с учебным планом)	Наименование индикатора достижений компетенций (в соответствии рабочим с учебным планом)
ПК-1	Готов формулировать цели и задачи научных исследований в соответствии с тенденциями и перспективами развития электроники и нанoeлектроники, а также смежных областей науки и техники, способностью обоснованно выбирать теоретические и экспериментальные методы и средства решения сформулированных задач	ПК-1.1	Знает принципы построения и функционирования изделий микро- и нанoeлектроники
		ПК-1.2	Умеет рассчитывать предельно-допустимые и предельные режимы работы изделий микро- и нанoeлектроники
		ПК-1.3	Владеет навыками выбора теоретических и экспериментальных методов исследования изделий микро- и нанoeлектроники
ПК-2	Способен разрабатывать эффективные	ПК-2.1	Знает методы разработки

	алгоритмы решения сформулированных задач с использованием современных языков программирования и обеспечивать их программную реализацию		эффективных алгоритмов решения научно-исследовательских задач
		ПК-2.2	Умеет использовать алгоритмы решения исследовательских задач с использованием современных языков программирования
		ПК-2.3	Владеет навыками разработки стратегии и методологии исследования изделий микро- и нанoeлектроники

2. УЧЕБНАЯ ПРОГРАММА

2.1. Цели и задачи дисциплины

Цель преподавания дисциплин: Целью изучения дисциплины является подготовка высококвалифицированных специалистов, владеющих теоретическими знаниями и практическими навыками в области исследования механических, электронных и прочих свойств материалов, включая органические и неорганические кристаллы, полимерные материалы, углеродные нанотрубки и т. д. методами машинного обучения, и способных на основе полученных знаний к активной творческой работе в области технической физики и нанотехнологий как в научно-исследовательских учреждениях, так и в условиях промышленного производства.

Учебная задача: Задачи курса состоят в изложении основных понятий, необходимых для описания и последующего моделирования физических процессов, протекающих в различных материалах, методами машинного обучения.

2.2. Трудоемкость дисциплины и виды учебной работы (в академических часах и зачетных единицах) *(удалить строки, которые не будут применены в рамках дисциплины)*

Виды учебной работы	Всего, в акад. часах	Распределение по семестрам
		1 сем
1	2	3
1.Общая трудоемкость изучения	144	144

дисциплины по семестрам, в т. ч.:		
1.1. Аудиторные занятия, в т. ч.:	32	32
1.1.1. Лекции	16	16
1.1.2. Практические занятия, в т. ч.	16	16
1.1.2.1. Контрольные работы	54	54
1.2. Самостоятельная работа, в т. ч.:	58	58
1.3. Консультации	54	54
Итоговый контроль (Экзамен, Зачет, диф. зачет - указать)	Экзамен 54	Экзамен 54

2.3. Содержание дисциплины

2.3.1. Тематический план и трудоемкость аудиторных занятий (модули, разделы дисциплины и виды занятий) по рабочему учебному плану

Разделы и темы дисциплины	Всего (ак. часов)	Лекции(ак. часов)	Практ. Занятия (ак. часов)
1	2=3+4+5+6+7	3	4
МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В МАТЕРИАЛОВЕДЕНИИ			
Введение			
<u>Раздел 1. Вероятность и случайные процессы</u>	6	3	3
<i>Тема 1.1. Вероятность и теорема Байеса</i>			
<i>Тема 1.2. Типы случайных процессов</i>			
<i>Тема 1.3. Основные свойства марковских процессов</i>			
<u>Раздел 2. Основные подходы к машинному обучению</u>	8	4	4
<i>Тема 2.1. МО «с учителем»</i>			
<i>Тема 2.2. МО «без учителя»</i>			
<i>Тема 2.3. Обучение с подкреплением</i>			
<u>Раздел 3. Методы понижения размерности</u>	6	3	3

<i>Тема 3.1. Метод главных компонент</i>			
<i>Тема 3.2. Метод стохастического вложения соседей с t-распределением</i>			
<u>Раздел 4. Глубокое обучение</u>	6	3	3
<i>Тема 4.1. Искусственные нейронные сети (НС) прямого распространения</i>			
<i>Тема 4.2. Обратное распространение ошибки</i>			
<i>Тема 4.3. Конволюционные нейронные сети</i>			
<i>Тема 4.4. Основные алгоритмы оптимизации</i>			
<u>Раздел 5. Основы и приложения теории информации</u>	6	3	3
<i>Тема 5.1. Информационная энтропия и количество информации</i>			
<i>Тема 5.2. Взаимная энтропия и расхождение Кульбака - Лейблера</i>			
<u>Раздел 6. Применение методов МО к актуальным задачам материаловедения</u>			
<i>Тема 6.1. Выбор свойств модели, ответственных за свойства материала</i>			
<i>Тема 6.2. Потенциалы взаимодействия, основанные на МО</i>			
<i>Тема 6.3. Нейросетевые модели и предсказание свойств материалов</i>			
ИТОГО	32	16	16

2.3.2. Краткое содержание разделов дисциплины в виде тематического плана

МОДУЛЬ 1.

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В МАТЕРИАЛОВЕДЕНИИ

Введение

Общие сведения о методах МО в материаловедении. Их значение для фундаментальной и прикладной науки. Области применений.

Раздел 1. Вероятность и случайные процессы

Тема 1.1. Вероятность и теорема Байеса

Основные понятия теории вероятности и математической статистики. Теорема Байеса и ее применение.

Тема 1.2. Типы случайных процессов

Введение понятия случайного процесса для решения физических задач. Основные типы случайных процессов. Свойства процессов Бернулли. Их возможное применение.

Тема 1.3. Основные свойства марковских процессов

Условные вероятности и определение марковского процесса (МП). Уравнение Колмогорова — Чепмена. Основное кинетическое уравнение и принцип детального баланса. МП с непрерывным и дискретным временем.

Раздел 2. Основные подходы к машинному обучению

Тема 2.1. МО «с учителем»

Задачи классификации и регрессии. Вероятностная интерпретация МО «с учителем». Функция потерь.

Тема 2.2. МО «без учителя»

Задачи кластеризации. Алгоритм К — средних. Вероятностная интерпретация МО «без учителя». Шумоподавление. Автокодировщики.

Тема 2.3. Обучение с подкреплением

Марковские процессы принятия решений и обучение с подкреплением.

Раздел 3. Методы понижения размерности

Тема 3.1. Метод главных компонент (РСА)

Сингулярные значения матрицы. РСА как линейное преобразование. Вариация и выделение главных компонент.

Тема 3.2. Метод стохастического вложения соседей с t -распределением

Гауссово распределение. Расхождение Кульбака — Лейблера. Физическая интерпретация.

Раздел 4. Глубокое обучение

Тема 4.1. Искусственные нейронные сети (НС) прямого распространения

Модель искусственного нейрона. Перцептрон. Градиентные методы обучения.
Функция активации и функция потерь.

Тема 4.2. Обратное распространение ошибки

Графы вычислений. Дифференцирование сложной функции и алгоритм обратного распространения.

Тема 4.3. Конволюционные нейронные сети

Операция свертки. Мотивация. Пулинг. Эффективные алгоритмы свертки.

Тема 4.4. Основные алгоритмы оптимизации

Метод градиентного спуска. Метод стохастического градиентного спуска. ADAM. RMSProp.

Раздел 5. Основы и приложения теории информации

Тема 5.1. Информационная энтропия и количество информации

Количество информации и вероятность события. Информационная энтропия.
Максимум энтропии и равновероятные события.

Тема 5.2. Взаимная энтропия и расхождение Кульбака - Лейблера

Расхождение Кульбака — Лейблера как мера близости функций распределения.
Вероятностная интерпретация задач МО.

Раздел 6. Применение методов МО к актуальным задачам материаловедения

Тема 6.1. Выбор свойств модели, ответственных за свойства материала

Кулоновская матрица. Структура. Химический состав. Использование методов глубокого обучения.

Тема 6.2. Потенциалы взаимодействия, основанные на МО

MLIP. Потенциалы LAMMPS.

Тема 6.3. Нейросетевые модели и предсказание свойств материалов

Примеры удачного применения моделей нейронных сетей, ранее использовавшихся при исследовании материалов различного происхождения.

2.3.3. Краткое содержание семинарских/практических занятий/лабораторного практикума

Практические занятия направлены на освоение базовых и прикладных методов машинного обучения в контексте задач материаловедения. В ходе семинаров студенты изучают методы работы с реальными материалами данных, получают навыки программной реализации моделей, обсуждают физико-химические особенности параметров материалов, участвующих в предсказании их свойств.

2.3.4. Материально-техническое обеспечение дисциплины

Компьютерный класс, проектор.

2.4. Модульная структура дисциплины с распределением весов по формам контролей

Формы контролей	Вес формы (форм) текущего контроля в результирующей оценке текущего контроля (по модулям)		Вес формы промежуточного контроля в итоговой оценке промежуточного контроля		Вес итоговой оценки промежуточного контроля в результирующей оценке промежуточных контролей		Вес итоговой оценки промежуточного контроля в результирующей оценке промежуточных контролей (семестровой оценке)	Веса результирующей оценки промежуточных контролей и оценки итогового контроля в результирующей оценке итогового контроля
	M1 ¹	M2	M1	M2	M1	M2		
Вид учебной работы/контроля								
Контрольная работа (при наличии)			0.5	0.5				
Устный опрос (при наличии)								
Лабораторные работы (при наличии)	0.5	0.5						
Письменные домашние задания (при наличии)								
Решение задач	0.5	0.5						
Веса результирующих оценок текущих контролей в итоговых оценках промежуточных контролей					0.5	0.5		
Веса оценок промежуточных контролей в итоговых оценках промежуточных контролей								
Вес итоговой оценки 1-го промежуточного контроля в результирующей оценке промежуточных контролей							0.5	
Вес итоговой оценки 2-го промежуточного контроля в результирующей оценке промежуточных контролей							0.5	
Вес результирующей оценки промежуточных контролей в результирующей оценке итогового контроля								0.5
Вес итогового контроля (Экзамен/зачет) в результирующей оценке итогового контроля								0.5
	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$

¹ Учебный Модуль

3. Теоретический блок (указываются материалы, необходимые для освоения учебной программы дисциплины)

3.1. Материалы по теоретической части курса

3.1.1. Учебник(и);

1. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. ДМК Пресс, 2018
2. C. M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006
3. J. VanderPlas. Python Data Science Handbook. O'Reily, 2023

4. Фонды оценочных средств (указываются материалы, необходимые для проверки уровня знаний в соответствии с содержанием учебной программы дисциплины).

4.1. Планы практических занятий

№	Тема занятия	Краткое содержание
1	Условная вероятность и теорема Байеса	Решение задач на применение теоремы Байеса и расчет условных вероятностей.
2	Работа с нормальным распределением	Расчет параметров нормального распределения, стандартное отклонение, дисперсия.
3	Принципы классификации и регрессии	Реализация линейной регрессии. Метод наименьших квадратов.
4	Алгоритмы обучения без учителя	Кластеризация, реализация К-средних и визуализация результатов.
5	Персептрон и градиентный спуск	Простая нейросеть: обучение, активация, ошибка.
6	Метод главных компонент (РСА)	Снижение размерности, работа с данными в многомерном пространстве.
7	Оценка вероятностей и правдоподобия	Оценка максимального правдоподобия и KL-дивергенция.
8	Информационная энтропия	Расчет количества информации, применение в задачах МО.

4.2. Материалы по практической части курса

4.2.1. Задачники (практикумы);

Ward L., Wolverton C., *Atomistic Materials Modeling: From Basics to Machine Learning*. — Springer, 2021. — 450 p.

Butler K. T., Davies D. W., Cartwright H., Isayev O., Walsh A., *Machine Learning for Molecular and Materials Science*. — Nature, 2018.

4.3. Вопросы и задания для самостоятельной работы студентов

- [1] Выведите основное уравнение марковского процесса и сформулируйте принцип детального баланса.
- [2] Реализуйте алгоритм К-средних для кластеризации набора точек в двумерном пространстве.
- [3] Постройте персептрон и обучите его на выборке AND/OR.
- [4] Объясните разницу между плотностью вероятности и вероятностью события.
- [5] Приведите пример применения центральной предельной теоремы в машинном обучении.
- [6] Рассчитайте количество информации, содержащейся в событии с вероятностью 0.1.
- [7] Выполните понижение размерности с помощью PCA и интерпретируйте результат.
- [8] Найдите взаимную информацию между двумя случайными величинами.

4.4. Образцы вариантов контрольных работ, тестов и/или других форм текущих и промежуточных контролей

Контрольная работа (вариант)

Задание 1. Определите понятие марковского процесса и приведите пример.

Задание 2. Выведите уравнение Колмогорова–Чепмена.

Задание 3. Реализуйте линейную регрессию на простом наборе данных и интерпретируйте коэффициенты.

Задание 4. Рассчитайте информационную энтропию распределения: $\{0.5, 0.25, 0.25\}$.

4.5. Перечень экзаменационных вопросов

- 1. Определение марковского процесса
- 2. Основное уравнение и детальный принцип баланса
- 3. Описание алгоритма К-средних
- 4. Персептрон как простейшая нейронная сеть
- 5. Линейная регрессия и метод наименьших квадратов
- 6. Оценка максимального правдоподобия
- 7. Максимальное правдоподобие и расхождение Кульбака — Лейблера
- 8. Параметры нормального распределения. Одномерный случай
- 9. Дисперсия случайной величины. Среднеквадратичное отклонение
- 10. Параметры нормального распределения. Многомерный случай

11. Вероятностная интерпретация контролируемого обучения
12. Вероятностная интерпретация обучения без учителя
13. Что такое ковариация?
14. Центральная предельная теорема.
15. Количество информации и информационная энтропия
16. Условная вероятность и теорема Байеса
17. Плотность вероятности
18. Проблемы классификации и кластеризации
19. Метод Монте-Карло цепи Маркова
20. Анализ основных компонентов

4.6. Образцы экзаменационных билетов

ИНЖЕНЕРНО-ФИЗИЧЕСКИЙ ИНСТИТУТ
Кафедра общей физики и квантовых наноструктур

Направление: Электроника и наноэлектроника
Дисциплина: Методы машинного обучения в материаловедении
(магистратура II-ой курс, I-ый семестр)

Экзаменационный билет № **

1. Описание алгоритма K-средних.
2. Параметры нормального распределения. Многомерный случай.
3. Проблемы классификации и кластеризации.

Зав. кафедрой ОФКН _____ Д.Б. Айрапетян
20__ г.
