

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

Утверждено
Директор Института математики и информатики
Арамян Р.Г.



«21» мая 2025, протокол №9.1

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ДИСЦИПЛИНЫ

Наименование дисциплины: Введение в машинное обучение

Автор (ы) **Аветисян К.И.**
Ф.И.О, ученое звание (при наличии), ученая степень (при наличии)

Направление подготовки: 01.04.02 Прикладная математика и информатика
Наименование образовательной программы: «Искусственный интеллект и робототехника»

1. АННОТАЦИЯ

1.1. Краткое описание содержания данной дисциплины;

Курс «Вступление в Машинное Обучение» предназначен для магистрантов. В рамках этого курса студенты изучат основные концепции и методы глубокого обучения, включая базовые архитектуры, такие как конволюционные сети, рекуррентные, трансформерные и т.д.. Этот курс является базовым для последующего перехода к курсам связанным с углубленным пониманием больших языковых и зрительных моделей.

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

Виды учебной работы	В акад. часах
Общее количество часов	32
Лекция	16
Практика	16
Итоговый контроль	Экзамен

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

Данная дисциплина связана с дисциплиной «Вступление в большие языковые модели»

1.4. Результаты освоения программы дисциплины:

Код компетенции	Наименование компетенции	Код индикатора достижения компетенций	Наименование индикатора достижений компетенций
УК-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	1	Разрабатывает командную стратегию для достижения поставленной цели
		2	Умеет организовывать и руководить работой команды
		3	Демонстрирует понимание результатов работы команды

			и личных действий в ней
ОПК-1	Способен решать актуальные задачи фундаментальной и прикладной математики	1	Имеет представление об основных подходах к решению актуальных задач фундаментальной и прикладной математики
		2	Демонстрирует умение применять математический аппарат для решения задач.
		3	Имеет навыки выбора подходящих методов решения задач фундаментальной и прикладной математики
ОПК-2	Способен совершенствовать и реализовывать новые математические методы решения прикладных задач	1	Обладает знаниями о существующих математических методах, применяемых для решения прикладных задач.
		2	Демонстрирует умение использования математического языка и математической символики, построения цепочки рассуждений, формулировки математических утверждений для решения прикладных задач.
		3	Имеет практический опыт совершенствования и реализации различных математических методов

			решения прикладных задач
ОПК-3	Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности	1	Формулирует основные теоретические положения в области математического моделирования.
		2	Демонстрирует умения давать содержательную интерпретацию полученных результатов при проведении анализа математических моделей.
		3	Имеет практический опыт разработки и проведения анализа математических моделей при решении задач
ОПК-4	Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности	1	Обладает знаниями о существующих информационно-коммуникационных технологиях и основных требованиях информационной безопасности.
		2	Демонстрирует умения комбинировать и адаптировать существующие информационно-коммуникационные технологии, а также умение учитывать основные требования информационной безопасности при решении

			прикладных задач
		3	Имеет практический опыт комбинирования и адаптирования существующих информационно коммуникационных технологий и учета основных требований информационной безопасности при решении прикладных задач
УК-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	1	Критически анализирует проблемную ситуацию с целью выработки стратегии действий, аргументировано формулирует собственные суждения и оценки
		2	Использует критический анализ, систематизацию и обобщение информации для решения проблемной ситуации
		3	Владеет навыками критического анализа проблемных ситуаций на основе системного подхода, вырабатывания стратегии действий
УК-2	Способен управлять проектом на всех этапах его жизненного цикла	1	Определяет этапы жизненного цикла проекта и выстраивает последовательность их

			реализации
		2	Формулирует проблему, на решение которой направлен проект, грамотно определяет цель проекта.
		3	Проектирует решение конкретных задач проекта, выбирая оптимальный способ их решения.
УК-4	Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	1	Обладает знаниями особенностей и правил личной и профессиональной устной и письменной коммуникации, в том числе на иностранном(ых) языке(ах)
		2	Демонстрирует умение применять современные коммуникативные технологии для академического и профессионального взаимодействия в ситуации устной и письменной коммуникации, в том числе на иностранном(ых) языке(ах)
		3	Имеет навыки академического и профессионального взаимодействия, в том числе на иностранном(ых)

			языке(ах)
УК-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия	1	Обладает необходимыми знаниями о разнообразии культур и об основных принципах межкультурного взаимодействия
		2	Демонстрирует умение самостоятельно добывать профессиональные знания с использованием иностранного языка для развития способности межкультурного взаимодействия
		3	Имеет навыки межкультурного взаимодействия при выполнении профессиональных задач
УК-6	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	1	Применяет рефлексивные методы в процессе оценки разнообразных ресурсов, используемых для решения задач самоорганизации и саморазвития
		2	Определяет цели и приоритеты собственной деятельности и способы их достижения
		3	Планирует результаты собственной деятельности с учетом необходимых

			ресурсов
--	--	--	----------

2. УЧЕБНАЯ ПРОГРАММА

2.1. Цели и задачи дисциплины

Цель дисциплины дать базовые знания глубокого обучения для последующей возможности перехода к рассмотрению тем связанных с большими языковыми и зрительными моделями.

Задачи

- Ознакомить студентов с базовыми архитектурами глубокого обучения
- Ознакомить с современными решениями на базе данных архитектур
- Ознакомить с современными практиками в машинном и глубоком обучении

2.2. Трудоемкость дисциплины и виды учебной работы (в академических часах и зачетных единицах)

Виды учебной работы	Всего, в акад. часах	Распределение по семестрам	
		1 сем	
1	2	3	
1. Общая трудоемкость изучения дисциплины по семестрам, в т. ч.:	32	32	
1.1. Аудиторные занятия, в т. ч.:			
1.1.1. Лекции	16	16	
1.1.2. Практические занятия, в т. ч.	16	16	
1.2. Самостоятельная работа, в т. ч.:	62	62	
Итоговый контроль	Экзамен/	Экзамен	

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)	Семинары (ак. часов)	Лабор. (ак. часов)
1	2=3+4+5+6 +7	3	4	5	6
Тема 1. Вариации конволюционных	1	1			
Тема 2. Состязание ImageNet и подходы	1	1			
Тема 3. Реализация конволюционных	1	1			

Тема 4. Векторные представление.	1	1			
Тема 5. Предобработка данных.	1	1			
Тема 6. Реализация процессов обработки	1	1			
Тема 7. Рекуррентные нейронные сети.	1	1			
Тема 8. LSTM, xLSTM.	1	1			
Тема 9. Реализация рекуррентных	1	1			
Тема 10. Механизм внимания.	1	1			
Тема 11. Энкодер-декодер архитектуры.	1	1			
Тема 12. Реализация механизма	1	1			
Тема 13. Архитектура трансформера.	1	1			
Тема 14. Архитектура BERT и его	1	1			
Тема 15. Тонкая настройка языковых	1	1			
Тема 16. Мультиязычные языковые	1	1			
Тема 17. Авторегрессивные модели.	1		1		
Тема 18. Тонкая настройка	1		1		
Тема 19. Модели семейства GPT.	1		1		
Тема 20. Модели семейства GPT.	1		1		
Тема 21. Реализация модели GPT.	1		1		
Тема 22. Модель Llama.	1		1		
Тема 23. Модели Mistral.	1		1		
Тема 24. Реализация модели GPT.	1		1		
Тема 25. Архитектура ResNet.	1		1		
Тема 26. Архитектура DenseNet.	1		1		
Тема 27. Реализация ResNet.	1		1		
Тема 28. Архитектура Vision	1		1		
Тема 29. Архитектура Vision	1		1		
Тема 30. Реализация DenseNet.	1		1		
Тема 31. Вариационные Автоэнкодеры.	1		1		
Тема 32. Вариационные Автоэнкодеры.	1		1		
ИТОГО	32	16	16		

2.3.2. Краткое содержание разделов дисциплины в виде тематического плана

Тема 1. Вариации конволюционных нейронных сетей.

Рассматриваются различные архитектуры и модификации конволюционных нейронных сетей (CNN), используемых для обработки изображений и других данных. Объясняются популярные архитектуры, включая VGG, Inception, и другие. Приводятся примеры применения этих вариаций в задачах распознавания образов, классификации, детектирования объектов и сегментации изображений.

Тема 2. Состязание ImageNet и подходы решения задач.

Рассматривается состязание ImageNet (ImageNet Large Scale Visual Recognition Challenge, ILSVRC), которое является ежегодным конкурсом по классификации и обнаружению объектов в изображениях. Основные подходы к решению задач включают глубокие сверточные нейронные сети (CNN), такие как AlexNet, VGG, ResNet и Inception, которые значительно улучшили точность распознавания объектов.

Тема 3. Реализация конволюционных нейронных сетей.

Рассматриваются конволюционные нейронные сети (CNN), которые реализуются с использованием слоев сверток, которые применяют фильтры к входным изображениям для извлечения признаков. Слой активации, обычно ReLU, вводит нелинейность, помогая сети обучать сложные зависимости. Слой пулинга уменьшает размер карт признаков, сохраняя важную информацию и снижая вычислительные затраты. Полносвязные слои завершают сеть, преобразуя двумерные карты признаков в одномерный вектор для классификации. В конце сети добавляется выходной слой с функцией softmax для классификации изображений по заданным категориям.

Тема 4. Векторные представление.

Рассматриваются векторные представления, которые преобразуют объекты, такие как слова или изображения, в числовые векторы, что позволяет моделям машинного обучения обрабатывать и анализировать данные. В обработке естественного языка (NLP) популярны методы, такие как Word2Vec, GloVe и BERT, которые преобразуют слова в плотные векторы, отражающие их семантическое значение и контекст. В компьютерном зрении, CNN используются для извлечения признаков изображений и представления их в виде векторов. Векторные представления улучшают качество моделей, обеспечивая компактное и информативное кодирование данных.

Тема 5. Предобработка данных.

Рассказывается предобработка данных, которая включает в себя этапы подготовки данных для улучшения качества и эффективности моделей машинного обучения. Основные шаги включают очистку данных от пропущенных или ошибочных значений, нормализацию или стандартизацию признаков для обеспечения единообразного масштаба, а также кодирование категориальных переменных с использованием методов, таких как one-hot encoding. Также может применяться уменьшение размерности, например, с помощью методов PCA или SVD, для уменьшения сложности данных и устранения избыточности. Предобработка данных является критическим этапом, влияющим на точность и производительность моделей.

Тема 6. Реализация процессов обработки данных.

Рассматривается реализация процессов обработки данных, которая включает сбор, очистку, преобразование и загрузку данных для их дальнейшего анализа и моделирования. Сначала данные собираются из различных источников, затем проводится очистка данных, устраняются пропущенные, дублированные или некорректные значения. Преобразование данных включает нормализацию, стандартизацию, кодирование категориальных признаков и уменьшение размерности. Эффективная обработка данных повышает точность и производительность аналитических решений.

Тема 7. Рекуррентные нейронные сети.

Рассматриваются рекуррентные нейронные сети (RNN), которые используются для обработки последовательных данных, таких как временные ряды или текст. В отличие от обычных нейронных сетей, RNN имеют петли, позволяющие сохранять информацию о предыдущих состояниях и учитывать контекст при обработке текущего входа. Основные типы RNN включают простые RNN, LSTM (Long Short-Term Memory) и GRU (Gated Recurrent Unit), которые решают проблему долгосрочной зависимости и затухания градиента.

Эти сети широко применяются в задачах предсказания временных рядов, машинного перевода, распознавания речи и генерации текста.

Тема 8. LSTM, xLSTM.

Рассматриваются рекуррентные нейронные сети с долговременной кратковременной памятью (LSTM) и их усовершенствованная версия xLSTM. Объясняется структура и механизм работы LSTM, включая входные, забывающие и выходные гейты, что позволяет эффективно обрабатывать последовательные данные и сохранять долгосрочные зависимости. Описываются преимущества xLSTM, которые включают улучшение стабильности и эффективности обучения по сравнению с традиционными LSTM.

Тема 9. Реализация рекуррентных нейронных сетей.

Рассматривается реализация рекуррентных нейронных сетей (RNN) и их модификаций (LSTM, xLSTM) для обработки последовательных данных.

Тема 10. Механизм внимания.

Рассматривается механизм внимания (attention mechanism) в нейронных сетях, который позволяет моделям фокусироваться на наиболее значимых частях входных данных при обработке последовательностей. Объясняется, как механизм внимания взвешивает различные части входных данных, чтобы улучшить точность и интерпретируемость модели. Описываются различные виды механизмов внимания, такие как self-attention и multi-head attention.

Тема 11. Энкодер-декодер архитектуры.

Рассматриваются энкодер-декодер архитектуры, которые широко применяются в задачах преобразования последовательностей, таких как машинный перевод и обработка естественного языка. Объясняется принцип работы энкодера, который считывает и кодирует входную последовательность в фиксированное представление, и декодера, который преобразует это представление обратно в выходную последовательность.

Тема 12. Реализация механизма внимания.

Рассматривается реализация механизма внимания в нейронных сетях для улучшения обработки последовательных данных. Объясняется процесс вычисления весов внимания, который включает в себя оценку значимости каждой части входных данных и их взвешивание для создания контекстного представления. Описываются основные шаги реализации, такие как вычисление сходства, нормализация с использованием softmax функции и вычисление взвешенной суммы. Приводятся примеры реализации различных типов внимания, включая self-attention и multi-head attention.

Тема 13. Архитектура трансформера.

Рассматривается архитектура трансформера, которая используется для обработки последовательных данных и является основой многих современных моделей, таких как BERT и GPT. Объясняется принцип работы трансформера, основанный на механизмах внимания, который позволяет параллельно обрабатывать элементы последовательности, что значительно ускоряет обучение и повышает производительность. Описывается архитектура, у которой ключевыми компонентами являются механизмы multi-head attention и позиционные кодировки, которые помогают моделировать последовательные зависимости.

Тема 14. Архитектура BERT и его вариации.

Рассматривается архитектура BERT (Bidirectional Encoder Representations from Transformers) и её вариации, применяемые для задач обработки естественного языка. Объясняется принцип работы BERT, который включает двунаправленное обучение трансформеров для лучшего понимания контекста слов в предложении. Описываются ключевые компоненты BERT, такие как masked language modeling (MLM) и next sentence prediction (NSP), а также преимущества использования предобученных моделей для различных NLP задач. Приводятся примеры и

особенности вариаций BERT, включая RoBERTa, DistilBERT и ALBERT, которые предлагают улучшения в производительности и эффективности модели.

Тема 15. Тонкая настройка языковых моделей.

Рассматривается тонкая настройка (fine-tuning) языковых моделей, используемая для адаптации предобученных моделей к специфическим задачам и данным. Объясняется процесс тонкой настройки, включающий дообучение модели на целевом датасете с использованием небольшого количества эпох, что позволяет модели более точно решать заданные задачи, такие как классификация текста, извлечение информации или машинный перевод. Описываются преимущества тонкой настройки, включая экономию вычислительных ресурсов и повышение точности на специфических задачах, а также примеры успешного применения этого метода в различных областях.

Тема 16. Мультиязычные языковые модели.

Рассматриваются мультиязычные языковые модели, предназначенные для обработки текста на различных языках с использованием единой модели. Объясняется принцип работы таких моделей, которые обучаются на многоязычных корпусах данных и могут понимать и генерировать текст на нескольких языках. Описываются ключевые особенности мультиязычных моделей, такие как их способность к перекрестному обучению, что позволяет моделям переносить знания с одного языка на другой, улучшая производительность. Приводятся примеры популярных мультиязычных моделей, таких как mBERT и XLM-R, и их применение в задачах машинного перевода, мультиязычного поиска и других многоязычных приложениях.

Тема 17. Авторегрессивные модели.

Рассматриваются авторегрессивные модели, используемые для генерации последовательных данных, таких как текст, аудио или временные ряды. Объясняется принцип работы авторегрессивных моделей, которые предсказывают следующий элемент последовательности на основе предыдущих элементов, что позволяет моделировать временные зависимости в данных. Описываются ключевые преимущества этих моделей, включая их способность к генерации последовательностей высокой точности и гибкость в применении к различным типам данных. Приводятся примеры популярных авторегрессивных моделей, таких как GPT (Generative Pre-trained Transformer) и их успешное использование в задачах генерации текста и других приложениях.

Тема 18. Тонкая настройка мультиязычных языковых моделей.

Рассматривается тонкая настройка мультиязычных языковых моделей для адаптации их к специфическим задачам на нескольких языках. Объясняется процесс тонкой настройки, который включает дообучение предобученных моделей на целевых многоязычных датасетах, что позволяет улучшить производительность моделей в различных языковых контекстах. Описываются преимущества такого подхода, включая возможность эффективного использования одной модели для множества языков и улучшение качества обработки языка в специфических задачах.

Тема 19. Модели семейства GPT.

Рассматриваются модели семейства GPT-1 (Generative Pre-trained Transformer 1), которые являются первыми в серии генеративных предобученных трансформеров. Объясняется принцип работы GPT-1, основанный на архитектуре трансформера, и его способность генерировать текст на основе предшествующего контекста. Описываются ключевые особенности GPT-1, включая его предобучение на большом корпусе данных и последующую тонкую настройку для выполнения специфических задач, таких как машинный перевод и текстовая генерация.

Тема 20. Модели семейства GPT.

Рассматриваются модели семейства GPT-2 (Generative Pre-trained Transformer 2), которые представляют собой улучшенную версию оригинальной модели GPT. Объясняется принцип работы GPT-2, основанный на масштабной архитектуре трансформера с большим количеством параметров, что позволяет генерировать более связный и осмысленный текст. Описываются ключевые особенности GPT-2, такие как его способность продолжать текст, задаваемый пользователем, и высокая точность в различных задачах, включая ответ на вопросы, перевод текста и создание творческого контента. Приводятся примеры применения GPT-2 в различных областях, демонстрирующие её мощные возможности в генерации естественного языка.

Тема 21. Реализация модели GPT.

Рассматривается реализация модели GPT-1 (Generative Pre-trained Transformer 1). Объясняется архитектура многослойного трансформера с механизмом внимания для генерации текста. Описываются этапы предобучения на большом корпусе текстов и тонкой настройки для специфических задач.

Тема 22. Модель Llama.

Рассматривается модель Llama, предназначенная для обработки и генерации текста. Объясняется архитектура модели, основанная на многослойном трансформере с механизмом внимания, что позволяет эффективно работать с последовательными данными. Описываются основные особенности и преимущества Llama, такие как улучшенная производительность и точность в сравнении с предыдущими моделями. Приводятся примеры применения модели Llama в различных задачах обработки естественного языка, включая генерацию текста и перевод.

Тема 23. Модели Mistral.

Рассматриваются модели Mistral, разработанные для обработки и генерации текста с высокой точностью. Объясняется их архитектура, основанная на улучшенных трансформерах, что обеспечивает эффективную работу с большими объемами данных. Описываются ключевые особенности моделей Mistral, такие как повышенная производительность и оптимизация для различных задач обработки естественного языка. Приводятся примеры применения моделей Mistral в задачах, таких как генерация текста, машинный перевод и анализ текста.

Тема 24. Реализация модели GPT.

Рассматривается реализация модели GPT-2 (Generative Pre-trained Transformer 2). Объясняется архитектура многослойного трансформера с механизмом внимания, что позволяет генерировать связный и осмысленный текст. Описываются этапы предобучения модели на большом корпусе данных и тонкой настройки для конкретных задач. Приводятся примеры кода для реализации GPT-2.

Тема 25. Архитектура ResNet.

Рассматривается архитектура ResNet (Residual Network), широко используемая в задачах компьютерного зрения. Объясняется принцип работы резидуальных блоков, которые помогают решать проблему исчезающих градиентов при обучении глубоких нейронных сетей. Описываются ключевые компоненты ResNet, такие как skip-connections, которые позволяют пропускать информацию через слои, улучшая обучение и производительность сети.

Тема 26. Архитектура DenseNet.

Рассматривается архитектура DenseNet (Densely Connected Convolutional Networks), применяемая в задачах компьютерного зрения. Описываются ключевые компоненты DenseNet, такие как dense connections, которые позволяют передавать информацию и градиенты напрямую между слоями, улучшая обучение и производительность сети.

Приводятся примеры применения DenseNet в классификации изображений и других задачах компьютерного зрения.

Тема 27. Реализация ResNet.

Рассматривается реализация модели ResNet (Residual Network) для задач компьютерного зрения. Объясняется структура модели, включающая резидуальные блоки, которые помогают преодолевать проблему исчезающих градиентов. Описываются этапы построения модели, такие как определение skip-connections для прямого пропуска информации через слои. Приводятся примеры кода для реализации ResNet.

Тема 28. Архитектура Vision Transformers.

Рассматривается архитектура Vision Transformers (ViT), предназначенная для обработки изображений. Объясняется принцип работы ViT, который отличается от традиционных свёрточных нейронных сетей (CNN) и использует механизм внимания. Изображение разрезается на небольшие патчи (обычно 16x16 пикселей), и каждый патч линейно преобразуется в векторное представление. Эти векторы, вместе с позиционными кодировками, подаются на вход трансформера, который обрабатывает их как последовательность, аналогично обработке слов в тексте.

Тема 29. Архитектура Vision Transformers.

Описываются ключевые компоненты Vision Transformers, такие как энкодер трансформера, который состоит из многослойных блоков внимания и слоёв нормализации. Механизм внимания (self-attention) позволяет модели учитывать взаимосвязи между различными частями изображения, что улучшает точность и качество обработки. Приводятся примеры применения ViT в задачах классификации изображений и других областях компьютерного зрения, демонстрирующие его высокую производительность и эффективность в сравнении с традиционными CNN.

Тема 30. Реализация DenseNet.

Рассматривается реализация модели DenseNet (Densely Connected Convolutional Networks) для задач компьютерного зрения. Описываются этапы построения модели, такие как определение dense connections для прямой передачи информации между слоями. Приводятся примеры кода для реализации DenseNet.

Тема 31. Вариационные Автоэнкодеры.

Рассматриваются вариационные автоэнкодеры (VAE), которые сочетают в себе свойства автоэнкодеров и вероятностного моделирования для обучения компактных представлений данных и их генерации. Объясняется структура VAE, включающая энкодер, который преобразует входные данные в параметры распределения (среднее и дисперсия) в латентном пространстве, и декодер, который восстанавливает данные из случайного вектора, сгенерированного на основе этих параметров. VAE обучаются с использованием комбинированной функции потерь, включающей ошибки реконструкции и регуляризацию на основе дивергенции Кульбака-Лейблера.

Тема 32. Вариационные Автоэнкодеры.

Описываются ключевые компоненты VAE, такие как латентное пространство, где энкодер создаёт распределение для каждого входного образца, и метод семплирования, который позволяет получать случайные векторы для генерации новых данных. Приводятся примеры применения VAE в задачах генерации изображений, обработки естественного языка и других областях, где требуется обучение на небольших или неполных данных. VAE используются для создания новых образцов данных, улучшения качества генерации и моделирования сложных распределений в высокоразмерных пространствах.

2.3.3. Краткое содержание семинарских/практических занятий/лабораторного практикума

(Кратко изложить форму/формы проведения семинарских занятий).

2.3.4. Материально-техническое обеспечение дисциплины

Компьютеры с интернет-браузером.

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 <i>(при наличии)</i>										
Устный опрос <i>(при наличии)</i>										
Тест <i>(при наличии)</i>										
Лабораторные работы <i>(при наличии)</i>										
Письменные домашние задания <i>(при наличии)</i>										
Реферат <i>(при наличии)</i>										
Эссе <i>(при наличии)</i>										
Проект <i>(при наличии)</i>										
Другие формы <i>(при наличии)</i>										

¹ Учебный Модуль

Веса результирующих оценок текущих контролей в итоговых оценках промежуточных контролей								
Веса оценок промежуточных контролей в итоговых оценках промежуточных контролей								
Вес итоговой оценки 1-го промежуточного контроля в результирующей оценке промежуточных контролей								
Вес итоговой оценки 2-го промежуточного контроля в результирующей оценке промежуточных контролей								
Вес результирующей оценки промежуточных контролей в результирующей оценке итогового контроля								
Вес итогового контроля (Экзамен/зачет) в результирующей оценке итогового контроля								
	$\sum = 1$	$\sum = 1$	$\sum = 1$	$\sum = 1$	$\sum = 1$	$\sum = 1$	$\sum = 1$	$\sum = 1$

3. Теоретический блок

3.1. Материалы по теоретической части курса

- 3.1.1. “Deep Learning” by Ian Goodfellow, Yoshua Bengio, Aaron Courville
- 3.1.2. “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow” by Geron Aurelien
- 3.1.3. “Natural Language Processing in Action” by Lane, Howard, and Hapke
- 3.1.4. “Generative Deep Learning” by David Foster

4. Фонды оценочных средств

4.1. Планы практических и семинарских занятий

- 1. Обзор вариаций конволюционных нейросетей (CNN): классификация, сегментация, генерация.
- 2. Состязание ImageNet и эволюция архитектур (AlexNet → EfficientNet).
- 3. Реализация базовой CNN в PyTorch.
- 4. Работа с векторными представлениями слов и изображений: Word2Vec, CLIP.

5. Предобработка данных: нормализация, аугментация, токенизация.
6. Реализация пайплайна загрузки и обработки данных (torchvision, HuggingFace datasets).
7. Рекуррентные нейронные сети (RNN): типы, проблемы, решения.
8. Архитектуры LSTM и xLSTM: особенности, модификации.
9. Реализация RNN и LSTM в PyTorch.
10. Attention-механизм: интуиция, типы, вычисление.
11. Архитектура энкодер-декодер (seq2seq) и Transformer.
12. Реализация self-attention и encoder-decoder моделей.
13. Архитектура трансформера: от Vaswani до современных улучшений.
14. Архитектура BERT и её вариации (RoBERTa, DistilBERT).
15. Тонкая настройка BERT на задачи классификации и NER.
16. Мультиязычные модели (mBERT, XLM-R).
17. Авторегрессия и autoregressive decoding (sampling, beam search).
18. Тонкая настройка мультиязычных моделей на локальных датасетах.
19. GPT: от GPT-1 до GPT-4, архитектура и отличия.
20. Сравнение GPT, ChatGPT, GPT-Neo, GPT-J.
21. Реализация минимального GPT на PyTorch (nanoGPT).
22. Архитектура LLaMA: особенности, отличия от GPT.
23. Обзор моделей Mistral, Mixtral, и их особенности.
24. Реализация inference для модели GPT или LLaMA.
25. Архитектура ResNet: остаточные связи и обучение глубоких сетей.
26. Архитектура DenseNet: плотные соединения и параметры.
27. Реализация ResNet на собственных данных.
28. Архитектура Vision Transformers (ViT): patch embedding и позиционное кодирование.
29. Применение ViT и сравнение с CNN.
30. Реализация DenseNet или ViT на задаче классификации.
31. Вариационные автоэнкодеры (VAE): вероятностный подход.
32. Реализация VAE и визуализация латентного пространства.

4.2. Планы лабораторных работ и практикумов

1. Реализация и обучение базовой CNN на CIFAR-10.
2. Разработка пайплайна предобработки изображений и текста.
3. Реализация LSTM и сравнение с обычной RNN.
4. Реализация механизма внимания вручную.
5. Сборка архитектуры Transformer «с нуля».
6. Использование BERT для классификации текста (Sentiment Analysis).
7. Тонкая настройка mBERT на мультиязычные данные.
8. Генерация текста с GPT и настройка sampling/beam search.
9. Визуализация attention map в трансформере.
10. Сравнение ResNet и DenseNet на задаче классификации.
11. Реализация ViT и сравнение с CNN.
12. Построение и обучение VAE на изображениях (MNIST, Fashion-MNIST).

4.3. Материалы по практической части курса

4.3.1. Учебно-методические пособия

- И. Гудфеллоу, Й. Бенджио, А. Курвилль. *Глубокое обучение*.
- Ч. Ола. *The Illustrated Transformer*.
- А. Карпати. *nanoGPT и минимальные реализации трансформеров*.

4.3.2. Учебные справочники

- [PyTorch documentation](#)
- [HuggingFace Transformers](#)
- [Papers with Code](#)

4.3.3. Задачники (практикумы)

- Kaggle (раздел Deep Learning)
- Google Colab Notebooks по VAE, GPT, ViT
- HuggingFace Datasets + Evaluate: задания по классификации, генерации, переводам

4.3.4. Наглядно-иллюстративные материалы

- Визуализация attention'a (BERTViz, BertVizTool)
- TensorBoard
- Деревья архитектур (cnn.architectureZoo)

4.3.5. Другие виды материалов

- GitHub: karpathy/nanoGPT, lucidrains/transformer, pytorch/vision
- Видеолекции Deep Learning AI, Stanford CS231n и CS224n
- Datasets: MNIST, CIFAR-10, WikiText-103, Common Crawl

4.4. Вопросы и задания для самостоятельной работы студентов

- Сравните ResNet и DenseNet: преимущества и недостатки.
- Реализуйте простой трансформер без использования nn.Transformer.
- Постройте собственный датасет на HuggingFace Datasets.
- Проведите сравнение качества генерации GPT-2 и GPT-3 на задаче суммаризации.
- Реализуйте мини-VAE и визуализируйте латентное пространство.

4.5. Тематика рефератов, эссе и других форм самостоятельных работ

- Эволюция архитектур языковых моделей: от LSTM до GPT-4

- Архитектуры трансформеров в компьютерном зрении
- Плюсы и минусы ViT по сравнению с CNN
- Особенности настройки LLaMA и её обучение
- Мультиязычные модели и проблемы лингвистической адаптации

4.6. Образцы контрольных и промежуточных заданий

Контрольная работа (фрагмент):

1. Назовите отличительные особенности архитектуры BERT.
2. Реализуйте attention-механизм в коде.
3. Сравните подходы autoregressive и encoder-decoder.

4.7. Перечень экзаменационных вопросов

1. Варианты архитектур CNN и их области применения
2. Устройство и обучение трансформера
3. Отличие encoder-only (BERT) и decoder-only (GPT)
4. Архитектура и применение LSTM
5. Attention: типы и реализация
6. Модели ViT и их преимущества
7. Различия GPT, LLaMA и Mistral
8. Авторегрессия и beam search
9. Механизм тонкой настройки
10. Архитектуры VAE и их применение

4.8. Образцы экзаменационных билетов

Билет №2

1. Опишите архитектуру DenseNet.
2. Реализуйте модуль self-attention.
3. Сравните GPT и LLaMA: архитектурные особенности.

4.9. Образцы экзаменационных практических заданий

- Реализовать forward-проход encoder-decoder трансформера.
- Выполнить inference текста через GPT-2 с beam search.
- Реализовать классификатор на основе BERT.
- Сравнить точность ViT и CNN на CIFAR-10.

4.10. Банк тестовых заданий для самоконтроля

1. Что отличает ResNet от DenseNet?
 - ☐ Использование dropout
 - ☒ Остаточные связи / плотные связи
2. Как работает self-attention?
 - ☐ Использует сверточные ядра
 - ☒ Взвешивает входы по схожести
3. Какой слой отвечает за позиционную информацию в ViT?
 - ☐ BatchNorm
 - ☒ Position Embedding
4. Что делает токен [CLS] в BERT?
 - ☐ Обозначает конец текста
 - ☒ Служит для классификации

4.11. Методики решения и ответы к тестам

- **Вопрос:** Что делает nn.MultiheadAttention?
Ответ: Вычисляет attention с несколькими головами, позволяя учитывать разные аспекты входа.
- **Вопрос:** Как интерпретировать attention map?
Ответ: Показывает, на какие слова/участки модель обращает внимание при генерации или классификации.

Если нужно, могу подготовить **оформленный документ в .docx/.pdf** с титульным листом и оглавлением или **сделать выгрузку в таблицу Excel** по каждому компоненту.

5. Методический блок

5.1. Методика преподавания

- 5.1.1. Методические рекомендации для студентов по подготовке к семинарским, практическим или лабораторным занятиям, по организации самостоятельной работы студентов при изучении конкретной дисциплины.