

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования

**«Пермский национальный исследовательский
политехнический университет»
(ПНИПУ)**

УТВЕРЖДАЮ

Проректор по образовательной
деятельности



[Signature] /А.Б.Петроченков
25 » *мая* 2022 г.

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ

«Машинное обучение и искусственные нейронные сети»

КВАЛИФИКАЦИЯ

**«Специалист по искусственному интеллекту
и машинному обучению»**

Пермь 2022

1. Общая характеристика программы

1.1. Цель реализации программы

Целью реализации ДПП профессиональной переподготовки является формирование профессиональных компетенций, необходимых для построения систем, основанных на применении искусственного интеллекта (ИИ). Элементы искусственного интеллекта используются в управлении, прогнозировании, распознавании, оказании услуг, поиске и т.д. Поэтому, современному инженеру необходимо владеть инструментами построения систем на основе ИИ.

Успешное освоение курса необходимо для решения задач по созданию и сопровождению промышленного ПО, связанного с применением ИИ, в разнообразных прикладных областях.

1.2. Требования к уровню подготовки поступающего на обучение, необходимому для освоения программы (категория слушателей)

Уровень подготовки поступающего на обучение:

- должен иметь диплом о высшем образовании в области точных и естественных наук;
- должен успешно освоить университетские курсы по дисциплинам: «Высшая математика», «Теория вероятности и математическая статистика», «Информатика», «Языки программирования»;
- должен иметь опыт программирования;

К необязательным требованиям относится владение языком программирования Python.

1.3. Перечень нормативных документов, определяющих требования к выпускнику программы

Образовательные стандарты:

- Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.03 Прикладная информатика, утвержденный Приказом Министерством науки и высшего образования РФ от 26.11.2020 № 1456.

Профессиональные стандарты:

- «Специалист по большим данным» утвержденный Приказом Министерства труда РФ от 06.07.2020 № 405н
- «Руководитель в области информационных технологий», утвержденный Приказом Министерства труда РФ от 18.11.2014 г. №893н.
- «Специалист по интеграции прикладных решений» утвержденный Приказом Министерства труда РФ от 05.09.2017 № 658н.

1.4. Характеристика новой квалификации

Квалификация «Специалист по искусственному интеллекту и машинному обучению» дает ее обладателю право выполнять следующие трудовые функции:

- Подготовка данных для анализа в соответствии с решаемой прикладной задачей, ТФ А/02.5).

- Интеллектуальная обработка данных («Старший Data-аналитик», ТФ В).
- Классификация данных, ТФ В/01.6.
- Прогнозирование на основе особенностей данных, ТФ В/03.6.
- Анализ отклонений в анализируемых данных, ТФ В/04.6.
- Разработка моделей машинного обучения для решений с использованием компьютерного зрения, ТФ С.
- Разработка моделей машинного обучения для решений с использованием обработки естественного языка, ТФ Д.
- Разработка моделей машинного обучения для решений с использованием распознавания аудиосигналов и синтеза речи, ТФ Е.

1.5. Характеристика нового вида профессиональной деятельности

Область профессиональной деятельности:

- 40.011 Специалист по научно-исследовательским и опытно-конструкторским разработкам

Объекты профессиональной деятельности:

- информационно-аналитические системы,
- оказание интеллектуальных услуг, чат-боты,
- робототехника,
- роботизированная автоматизация бизнес-процессов (RPA)

Виды и задачи профессиональной деятельности:

- научно-исследовательская деятельность,
- разработка программного обеспечения

Уровень квалификации: 5-6 уровень в соответствии с Приказом Минтруда России от 12.04.2013 №148н «Об утверждении уровней квалификации в целях разработки проектов профессиональных стандартов»

1.6. Планируемые результаты обучения

Слушатель, успешно завершивший обучение по данной программе, должен обладать следующими профессиональными компетенциями:

- **Применяет принципы и основы алгоритмизации, самостоятельно разрабатывает алгоритмы любой сложности, использует доступный опыт других разработчиков (интернет, литература) (ОК 1)**
- **Анализирует большие данные в проектах под контролем опытных специалистов (ПК 1)**
- **Участвует в проектах применения искусственного интеллекта и машинного обучения под контролем опытных специалистов (ПК 2)**
- **Самостоятельно применяет математическую логику и методы оптимизации для решения задач по оценке и разработки модели (ПК 3)**
- **Самостоятельно подбирает программные и технические средства для визуализации больших данных и использует их в работе, эпизодически прибегая к экспертной консультации (ПК 4)**
- **Идентифицирует принадлежность проблемной области к классу решаемых задач с помощью ИИ (ПК 5)**
- **Использует методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных (ПК 6)**

1.7. Трудоемкость обучения

Трудоемкость программы, включая все виды учебной работы слушателя, составляет 302 часа

1.8. Форма обучения

Обучение по программе осуществляется в очной форме с частичным отрывом от работы, с использованием дистанционных образовательных технологий.

1.9. Документ, выдаваемый по результатам освоения программы

Слушателям, завершившим обучение по соответствующей программе профессиональной переподготовки и успешно прошедшим итоговую аттестацию, выдается диплом о профессиональной переподготовке установленного в ПНИПУ образца:

- диплом о профессиональной переподготовке с правом ведения профессиональной деятельности в сфере создания интеллектуальных информационных систем и цифровых сервисов и присвоением квалификации «Специалист по искусственному интеллекту и машинному обучению».

2. Содержание программы

2.1. Учебный план программы профессиональной переподготовки

Учебный план ДПП профессиональной переподготовки отображает логическую последовательность освоения циклов и разделов программы (дисциплин, модулей, практик), обеспечивающих формирование компетенций.

Наименование дисциплин	Общая трудоемкость, час.	По учебному плану с использованием дистанционных образовательных технологий, час.						СРС, час	Текущий контроль	Промежуточная аттестация
		Аудиторные занятия			Из них дистанционные занятия					
		всего	Л	ПЗ	всего	Л	ПЗ		ИЗ	Зачет
1	2	3	4	5	6	7	8	9	10	
Модуль 1 (обязательный) Основы машинного обучения и искусственных нейронных сетей	82	43	24	18	10	2	8	39	2	1(Д)
М1.1 Основы теории информации и математической статистики	32	16	7	7	4	2	2	16	1	1
М1.2 Основы теории оптимизации	28	14	8	6	2	-	2	14	-	-
М1.3 Основы ИНС	22	13	8	4	4	-	4	9	1	-
Модуль 2 (обязательный) Методы решения задач с ассоциативной памятью	84	45	22	20	12	-	12	39	2	1(Д)
М2.1 Отдельные главы математической статистики	42	22	10	10	4	-	4	20	1	1
М2.2 Рекуррентные нейронные сети, автоэнкодеры и сети глубокого	42	23	12	10	10	-	10	19	1	-

6	Классические методы решения задач компьютерного зрения	Самостоятельно подбирает программные и технические средства для визуализации больших данных и использует их в работе, эпизодически прибегая к экспертной консультации (ПК 4)
7	Сверточные нейронные сети	Идентифицирует принадлежность проблемной области к классу решаемых задач с помощью ИИ (ПК 5)
8	Основы математической лингвистики	Самостоятельно применяет математическую логику и методы оптимизации для решения задач по оценке и разработки модели (ПК 3)
9	ИНС для решения задач NLU	Применяет принципы и основы алгоритмизации, самостоятельно разрабатывает алгоритмы любой сложности, использует доступный опыт других разработчиков (интернет, литература) (ОК 1)
10	Анализ и модели временных рядов	Применяет принципы и основы алгоритмизации, самостоятельно разрабатывает алгоритмы любой сложности, использует доступный опыт других разработчиков (интернет, литература) (ОК 1)
11	Предиктивный анализ временных рядов с использованием ИНС	Применяет принципы и основы алгоритмизации, самостоятельно разрабатывает алгоритмы любой сложности, использует доступный опыт других разработчиков (интернет, литература) (ОК 1)

2.4. Рабочие программы разделов, дисциплин (модулей)

Модуль 1 (обязательный). Основы машинного обучения и искусственных нейронных сетей.

Модуль «**Основы машинного обучения и искусственных нейронных сетей**» включен в обязательную часть образовательной программы и предусматривает знакомство студентов с основами машинного обучения, используемого для анализа данных. Модуль включает три дисциплины:

- основы теории информации и математической статистики;
- основы теории оптимизации;
- основы ИНС.

Цель изучения дисциплины «**Основы теории информации и математической статистики**» - освоить методологические основы прикладной математической статистики и научиться применять ее для решения задач анализа данных и машинного обучения.

Основные задачи обучения:

- **знать** комбинаторный и дискретный анализ;
- **знать** основы решения задачи классификации;
- **знать** основы статистического оценивания.

В результате изучения дисциплины «**Основы теории информации и математической статистики**» студент должен **уметь**:

- формулировать задачу машинного обучения в терминах математической статистики
- строить регрессионные модели.

В результате изучения дисциплины «**Основы теории информации и математической статистики**» студент должен **владеть**:

- методами нахождения несмещенной оценки распределения с минимальной дисперсией (НОРМД);
- линейными методами решения задачи классификации.

Цель изучения дисциплины «**Основы теории оптимизации**» – научиться выполнять работы по созданию содержательной, концептуальной и математической постановок задач для решения трудноформализуемых прикладных проблем.

Основные задачи обучения:

- **знать** основные методы теории оптимизации;
- **знать** логический алгоритм классификации, решающий задачи классификации и регрессии;
- **знать** композиционный метаалгоритм машинного обучения, применяющийся для уменьшения смещения, а также дисперсии в обучении с учителем.

В результате изучения дисциплины «Основы теории оптимизации» студент должен **уметь**:

- использовать метод главных компонент;
- использовать алгоритм AdaBoost для повышения производительности любого алгоритма машинного обучения.

В результате изучения дисциплины «Основы теории оптимизации» студент должен **владеть**:

- методом машинного обучения - градиентным бустингом для создания решающей модели прогнозирования в виде ансамбля слабых моделей прогнозирования.

Цель изучения дисциплины «Основы ИНС» - освоить методы решения трудно формулируемых задач с помощью искусственных нейронных сетей первого поколения.

Основные задачи обучения:

- **знать** классификацию видов и архитектур искусственных нейронных сетей (ИНС);
- **знать** устройство искусственного нейрона и архитектуру многослойного персептрона
- **знать** алгоритмы обучения ИНС;
- **знать** основные прикладные проблемы, решаемые с помощью ИНС.

В результате изучения дисциплины «Основы ИНС» студент должен **уметь**:

- выполнять постановку трудно формулируемой проблемы с целью решения ее с помощью методов ИНС;
- подбирать архитектуру ИНС;
- выполнять обучение сети.

В результате изучения дисциплины «Основы ИНС» студент должен **владеть**:

- методами обучения ИНС “с учителем”;
- инструментальной средой Jupyter Notebook;
- методами предварительного анализа и предобработки обучающих данных.

М1.1. Основы теории информации и математической статистики

Лекции - 7 часов, практические занятия - 7 часов, СРС - 16 часов.

Тема 1. Введение в теорию информации. Основные объекты математической статистики.

Введение. Количество информации. Информационная энтропия. Энтропия непрерывного источника информации. Мера количества информации в дискретном сигнале. Статистическая мера количества информации. Кросс-энтропия. Количество информации, как мера снятой неопределенности. Основные понятия математической статистики: статистические оценки (точечные и интервальные), их свойства, проверка гипотез.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 2. Машинное обучение как математическое моделирование.

Вероятностная постановка задачи классификации. Основные понятия: априорная вероятность, апостериорная вероятность, функция правдоподобия класса. Функционал

среднего риска. Теорема Колмогорова — Арнольда. Ошибки I и II рода. Статистические модели. Теоретико-вероятностная постановка задачи обучения с учителем. Минимизация ожидаемой ошибки. No free lunch theorem. Пример: задача регрессии, минимизация квадрата отклонения. Регрессионная функция: условное математическое ожидание. Линейная регрессия. Переобучение и недообучение. Проклятие размерности. Методы оценивания обобщающей способности, кросс-валидация.
Лекции – 3 часа. Практические занятия – 3 часа. Лабораторные работы – 0 часов.

Тема 3. Линейные модели и задача классификации.

Градиентный спуск, методы оценивания градиента. Функции потерь. Метрики качества регрессии. Линейная регрессия, метод наименьших квадратов. Теорема Гаусса—Маркова. Явный вид решения в методе наименьших квадратов. Ковариационная матрица для коэффициентов. Регуляризация. Общая постановка. 0-1 ошибка.
Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
1	Количество информации. Энтропия непрерывного источника информации. Статистическая мера количества информации. Количество информации как мера снятой неопределенности. Решение задач.
2	Оценивание ошибок первого и второго рода. Разложение ошибки на шум, смещение и разброс. Методы оценивания обобщающей способности, кросс-валидация.
2	Условное математическое ожидание. Построение модели линейной регрессии.
3	Оценки на основе МНК.

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
1	Найти интервальные и точечные оценки параметров распределения, по выборке случайных величин, смоделированных с помощью пакета прикладных программ MATLAB. Оценить погрешность полученных результатов.	6
2	Провести классификацию методом k ближайших соседей для смоделированной ранее выборки.	6
3	Написать реферат с иллюстрирующими примерами на тему: Точное решение vs градиентный спуск.	4

М1.2. Основы теории оптимизации

Лекции - 8 часов, практические занятия - 6 часов, СРС - 14 часов.

Тема 1. Выбор и оценка моделей, работа с признаками.

Введение. Постановка задачи оптимизации. Критерии. Задачи минимакса, транспортная задача, задача коммивояжера. Кросс-валидация: тонкости (отбор переменных, переобучение на валидационное множество). Оценки ожидаемой ошибки для линейной регрессии: AIC и другие. L1 и L2 регуляризация. Методы отбора признаков. Метод главных компонент и singular spectrum analysis. Ядровые методы. Ядра и спрямляющие пространства, методы их построения. Операции в спрямляющих пространствах. Принцип Беллмана.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 2. Деревья и ансамбли.

Ограничения линейных методов (пример: XOR). Решающие деревья. CART. Ансамбли. Бутстреп. Бэггинг. Случайный лес.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 3. Бустинг.

Градиентный бустинг. XGBoost. Алгоритм AdaBoost. Теорема о сходимости. Обоснование малой переобучаемости алгоритма.

Лекции – 2 часа. Практические занятия – 0 часов. Лабораторные работы – 0 часов.

Тема 4. Обучение нейросетей.

Batch Gradient Descent. Стохастический градиентный спуск. Скорость сходимости методов градиентного спуска. Алгоритмы Momentum и модификация Nesterov Momentum. Подход к нормализации в методах AdaGrad и RMSProp. Метод ADAM и его преимущества. Ансамблевые методы.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
1	Разбор решения задачи минимакса. Решение транспортной задачи. Модель линейной регрессии. Динамическое программирование.
2	Построение деревьев. Линейная модель.
4	Методы градиентного спуска. Сравнение скорости сходимости, обсуждение областей применимости.

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
3	Когда можно повысить точность классификатора. Описание подходов к увеличению точности классификатора.	6

4	Регуляризация Dropout для методов оптимизации.	8
---	--	---

М1.3. Основы ИНС

Лекций - 8 часов, практические занятия - 4 часа, СРС - 9 часов.

Тема 1. Определение и предназначение ИНС

Введение. Определение и предназначение ИНС. Биологический нейрон. Модели биологического нейрона. Искусственный нейрон МакКаллока-Питса. Персептрон и однослойные сети. Задача линейной сепарабельности. История ИНС.

Тема 2. Базовые алгоритмы обучения

Обучение «с учителем» и «без учителя», правило Хебба, обратное распространение ошибки

Тема 3. Сети прямого распространения

Многослойный персептрон(MLP), обучение многослойных сетей, сеть радиальных базисных функций (RBF), гибридный алгоритм обучение RBF, переобучение, проблема затухающего градиента

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
1	Знакомство со средой JupyterLab. Разбор примера ИНС состоящей из одного нейрона МакКаллока-Питса. Решение задачи регрессии (определения температуры в градусах Фаренгейта).
2	Разбор примера многослойного персептрона. Подготовка и анализ обучающего множества. Решение задачи классификации (определение стоимости домов)

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
1	Работа над Индивидуальным заданием модуля. Постановка задачи, подбор и предобработка обучающего dataset, определение архитектуры ИНС, написание кода, обучение, анализ результатов. Подготовка итоговой презентации.	9

Модуль 2 (обязательный). Методы решения задач с ассоциативной памятью

Модуль «Методы решения задач с ассоциативной памятью» включен в обязательную часть образовательной программы и предусматривает знакомство студентов с основами машинного обучения, используемого для поиска, сжатия и восстановления данных. Модуль включает три дисциплины:

- отдельные главы математической статистики;

- рекуррентные нейронные сети и сети доверия;
- автоэнкодеры.

Цель изучения дисциплины «**Отдельные главы математической статистики**» - освоить методологические основы прикладной математической статистики и научиться применять ее для решения задач анализа данных и машинного обучения.

Основные задачи обучения:

- **знать** методы нахождения оптимальных оценок для неизвестных параметров распределения;
- **знать** основы теории случайных процессов;
- **знать** основы статистического моделирования.

В результате изучения дисциплины «**Отдельные главы математической статистики**» студент должен **уметь:**

- формулировать задачу машинного обучения в терминах математической статистики;
- использовать аппарат цепей Маркова для решения прикладных задач.

В результате изучения дисциплины «**Отдельные главы математической статистики**» студент должен **владеть:**

- методами сэмплирования в статистике;
- методом Монте-Карло.

Цель изучения дисциплины “**Рекуррентные нейронные сети, автоэнкодеры и сети глубокого доверия**” - освоить методы решения задач, которые используют ассоциативную память. Это следующие задачи: поиск скрытых закономерностей в данных, сжатие данных, построение семантических индексов, фильтрация и восстановление поврежденных данных. К таким задачам, традиционно, относят задачи классификации, устойчивые к шумам в данных.

Основные задачи обучения:

- **знать** основы глубокого обучения (Deep Learning);
- **знать** принципы построения рекуррентных сетей и сетей доверия;
- **знать** основы построения автоэнкодеров;
- **знать** принципы обучения вероятностных сетей;
- **знать** принципы предобучения других многослойных ИНС с помощью RBM и автоэнкодеров;

В результате изучения дисциплины “**Рекуррентные нейронные сети, автоэнкодеры и сети глубокого доверия**” студент должен **уметь:**

- распознавать задачи, в которых необходимо применять рекуррентные сети, вероятностные модели или автоэнкодеры;
- выбирать алгоритм обучения для вероятностных моделей;
- выбирать архитектуру автоэнкодера под исследовательскую задачу;
- проводить предобучения MLP с помощью автоэнкодеров или RBM.

В результате изучения дисциплины “**Рекуррентные нейронные сети, автоэнкодеры и сети глубокого доверия**” студент должен **владеть:**

- методами преобучения глубоких сетей;
- методами построения семантических хешей;
- методами восстановления из зашумленных образов.

М 2.1. Отдельные главы математической статистики

Лекции - 12 часов, практические занятия - 12 часов, СРС - 24 часов.

Тема 1. Понятие оценок параметров выборки.

Свойства оценок, методы нахождения оценок. Понятие об оценивании неизвестных параметров выборки. Несмещенность и состоятельность оценок. Метод моментов и метод

максимального правдоподобия. Свойства соответствующих оценок. Примеры вычисления. Оценки с наименьшей дисперсией. Неравенство Рао-Крамера. Эффективные оценки. Асимптотические свойства оценок максимального правдоподобия.
Лекции – 1 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 2. Понятие о достаточных статистиках.

Контрпример для эмпирического среднего (как оценки среднего равномерного распределения). Примеры и контрпример для модели Коши. Достаточные статистики и теорема факторизации. Существование достаточных статистик. Построение достаточных статистик для неизвестных параметров выборки из нормальной совокупности.
Лекции – 1 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 3. Интервальные оценки.

Оценка параметров выборки из нормального распределения. Интервальные оценки неизвестных параметров выборки. Доверительные интервалы для различных параметров. Построение доверительного интервала для среднего известной дисперсии и для среднего при неизвестной дисперсии. Построение доверительного интервала для дисперсии при известном среднем и для дисперсии при неизвестном среднем). Связь понятия состоятельной оценки и понятия доверительного интервала. Доверительные оценки в общей ситуации. Метод главных компонент.
Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 4. Критерии значимости.

Проверка статистических гипотез. Общие понятия из теории Пирсона-Неймана проверки гипотез. Критерии значимости. Проверка статистических гипотез. Критерий согласия хи-квадрат. Теорема Фишера. Критерии Стьюдента (одновыборочный и двухвыборочный) и Фишера. Применение для проверки сопряженности признаков. Критерий однородности. Критерий знаков. Ранговые критерии. Простые и сложные гипотезы. Ошибки первого и второго рода в задаче различения гипотез. Мощность критерия. Теорема Неймана-Пирсона. Равномерно наиболее мощные критерии.
Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 5. Понятие о регрессионном анализе.

Понятие о регрессионном анализе. Задача регрессии. Линейная регрессия и метод наименьших квадратов. Значимость параметров линейной регрессии. Различные критерии для проверки гипотезы о значимости линейной регрессии. Критерий Стьюдента о независимости выборок. Критерий Фишера для проверки гипотезы о значимости линейной регрессии. Корреляционный анализ. Выборки с отклонением. Расстояние Кульбака — Лейблера. Построение функции правдоподобия. Нахождение оценок по методу максимального правдоподобия. Определение класса распределений, позволяющих использовать ММП. Байесовская статистика.
Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 6. Случайные процессы и моделирование.

Стационарные случайные процессы. Понятие Марковости. Цепи Маркова. Моделирование для решения прикладных задач. Сэмплирование. Сэмплирование по Гиббсу. Алгоритм Метрополиса-Гастингса. Метод Монте-Карло.
Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
1	Нахождение несмещенных оценок неизвестных параметров распределения. Проверка их эффективности. Состоятельность статистических оценок.
2	Достаточные статистики. Проверка статистики на достаточность.
3	Построение доверительных интервалов для неизвестного параметра наблюдаемой случайной величины.
4	Критерий Хи-квадрат и критерий Колмогорова-Смирнова для проверки статистической гипотезы.
5	Байесовская статистика. ОМП. Критерий Фишера.
6	Применение метода Монте-Карло для прикладных задач. Расчет параметров для цепей Маркова.

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
2	Улучшение неравенства Крамера – Рао. Неравенство Баттачария.	6
3	Теорема Рао – Блекуэлла - Колмогорова.	6
5	Оценка параметров сдвига и масштаба. Оценка Питмена.	6
6	Понятие стохастической сходимости.	6

М 2.2. Рекуррентные нейронные сети и сети глубокого доверия

Лекции - 12 часов, практические занятия - 10 часов, СРС - 19 часов.

Тема 1. Рекуррентные сети первого поколения

Сеть Хопфилда. Обучения сети Хопфилда. Диафония. Метод проекций и дельта-проекции. Оценка емкости ассоциативной памяти. Сеть Хемминга. Гетероассоциативные запоминающие устройства. Расстояние Хемминга. Нейроны WTA. Сеть Эльмана.

Тема 2. Вероятностные ИНС и сети глубокого доверия

Машина Больцмана. Алгоритм имитации отжига. Ограниченная машина Больцмана (RBM). Алгоритм контрастной дивергенции. Эффективное предобучение многослойной нейронной сети. Сети глубокого доверия (DBN). Понятие глубокого обучения (Deep Learning)

Тема 3. Автоэнкодеры

Структура автоэнкодера. Использование автоэнкодера для обучения DBN. Регуляризованные варианты автоэнкодеров: разреженный и шумоподавляющий. Задача восстановления зашумленных образов. Многослойные автоэнкодеры. Глубокие автоэнкодеры. Задача семантического хеширования.

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
1	Разбор примера сети Хопфилда. Решение задачи распознавания цифр
2	Разбор примера предобучения многослойного персептрона с помощью RBM. Решение задачи классификации
3	Разбор примера регуляризованного автоэнкодера. Решение задачи восстановления образа
4	Разбор примера сжатия данных с помощью DBN. Решение задачи создания семантического индекса.

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
1	Работа над Индивидуальным заданием модуля. Постановка задачи, подбор и предобработка обучающего dataset, определение архитектуры ИНС, написание кода, обучение, анализ результатов. Подготовка итоговой презентации.	19

Модуль 3 (вариативный). Решение задач компьютерного зрения

Этот модуль следует выбирать студентам, которые специализируются в области использования нейронных сетей и задач компьютерного зрения.

Модуль «**Решение задач компьютерного зрения**» включен в вариативную (по выбору) часть образовательной программы и предусматривает знакомство студентов с основами компьютерного зрения, используемого для анализа изображений. Модуль включает две дисциплины:

- классические методы решения задач компьютерного зрения;
- сверточные нейронные сети.

Цель изучения дисциплины «**Классические методы решения задач компьютерного зрения**» – изучить основы цифровой обработки изображений, освоить базовые знания в области компьютерного зрения, а также применяемый в ней математический аппарат, и получить практические навыки обработки изображений и видео.

Основные задачи обучения:

- **знать** основные подходы компьютерного зрения для предварительной обработки и улучшения изображений;
- **знать** основные вычислительные методы, применяемые в компьютерном зрении;
- **знать** основные метрики, применяемые для оценки качества моделей;
- **знать** методы поиска и распознавания объектов на изображении и видеопоследовательности;
- **знать** принципы построения моделей сегментации.

В результате изучения дисциплины «**Классические методы решения задач компьютерного зрения**» студент должен **уметь:**

- выбирать правильную метрику оценки качества модели;
- готовить датасеты для различных задач компьютерного зрения;
- подбирать и реализовывать алгоритмы машинного обучения под прикладную задачу.

В результате изучения дисциплины «**Классические методы решения задач компьютерного зрения**» студент должен **владеть:**

- методами предварительной обработки исходных изображений;
- методами построения и обучения классифицирующих, детектирующих и сегментирующих моделей.

Цель изучения дисциплины «**Сверточные нейронные сети**» – изучить современные подходы и методы в области компьютерного зрения и научиться применять их в различных задачах.

Основные задачи обучения:

- **знать** принципы построения сверточных нейронных сетей;
- **знать** принципы подготовки датасетов для различных задач компьютерного зрения;
- **знать** принципы обучения сверточных нейронных сетей;
- **знать** популярные архитектуры сверточных нейронных сетей для задач классификации, детектирования и сегментации.

В результате изучения дисциплины «**Сверточные нейронные сети**» студент должен **уметь:**

- готовить датасеты для современных задач компьютерного зрения;
- выбирать архитектуру сверточной нейронной сети под исследовательскую задачу;
- проводить дообучение сверточных нейронных сетей под конкретную задачу.

В результате изучения дисциплины «**Сверточные нейронные сети**» студент должен **владеть:**

- методами подготовки датасетов к различным задачам компьютерного зрения;
- методами построения и обучения современных моделей глубоких сверточных нейронных сетей для решения различных задач компьютерного зрения.

М 3.1. Классические методы решения задач компьютерного зрения

Лекции – 12 часов, практические занятия – 8 часов, СРС - 17 часов.

Тема 1. Введение в компьютерное зрение. Основы цифрового представления изображений.

Введение в понятие предмета компьютерного зрения, общие подходы и классификация решаемых задач. Показатели эффективности в задачах компьютерного зрения: классификации, детектирования, сегментации (accuracy, precision, recall, F-score, ROC, IoU, AP, mAP). Компьютерное и человеческое зрение. Представления цифрового изображения.

Лекции – 2 часа. Практические занятия – 0 часов. Лабораторные работы – 0 часов.

Тема 2. Обработка изображений.

Тональная и цветовая коррекции. Подавление шума. Свертка изображений, линейные и нелинейные фильтры. Медианная фильтрация. Адаптивная фильтрация изображений. Сглаживание и повышение резкости. Выделение границ. Алгоритм Кэнни. Оператор Собеля. Бинаризация изображений: понятие и методы.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 3. Особые точки и дескрипторы.

Понятие особой точки изображения. Особые точки Харисса. Выбор значений порога для выделения точки. Способ определения схожих участков в разных изображениях. Понятие дескриптора, представление и назначение. Принципы построения дескрипторов. Метод SIFT. Производные дескрипторы – RIFT/G-RIF/SURF.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 4. Распознавание объектов: классификация и детектирование изображений.

Понятие классификации, проблемы и задачи. Построение датасета: репрезентативность выборки, разделение на поднаборы (обучение, валидация, тестирование). Понятие детектирования, проблемы и задачи. Гистограммы ориентированных градиентов (Histograms of Oriented Gradients), наивный Байесовский ближайший сосед (Naive Bayesian Nearest Neighbor), мешок слов (Bag of Words). Признаки Хаара. Поиск лиц – метод Виолы-Джонса. Бустинг. Каскады классификаторов.

Лекции – 4 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 5. Сегментация изображений.

Понятие сегментации, виды, проблемы и задачи. Пороговая обработка в задачах сегментации. Обработка с глобальным порогом. Метод Оцу оптимального глобального порогового преобразования. Сегментация методом кластеризации, метод К-средних, метод водораздела. Текстурная сегментация: описание, методы. Характеристики Харалика. Фильтры Габора.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
2	Решение задачи выделения границ с помощью детекторов Кэнни и Собеля.
3	Применение SIFT дескрипторов для выделения похожих изображений.
4	Решение задачи детектирования лиц в реальном времени методом Виолы-Джонса.
5	Разбор примера сегментации изображений несколькими методами.

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
1	Работа над Индивидуальным заданием модуля. Постановка задачи, подбор и предобработка обучающего dataset, определение метода решения, написание кода, обучение, анализ результатов. Подготовка итоговой презентации.	17

М 3.2. Сверточные нейронные сети

Лекции – 12 часов, практические занятия – 10 часов, СРС - 22 часа.

Тема 1. Основы сверточных нейронных сетей. Подготовка данных для обучения моделей.

Операция свертки. Операция пулинга. Ядро свертки. Карта признаков. Сверточный слой. Слой пулинга. Полносвязные слои (классифицирующие). Дропаут. Батч-нормализация. Построение сверточных сетей. Виды обучения искусственных нейронных сетей. Обучение с учителем. Подготовка датасета для решения задач классификации, детектирования и сегментации: разметка, загрузка и предварительная обработка данных. Репрезентативность датасета. Несбалансированность данных. Расширение данных (аугментация).

Лекции – 4 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 2. Задача классификации.

Типы классификации. Постановка задачи классификации. Разбор популярных архитектур сверточных нейронных сетей: LeNet-5, AlexNet, GoogLeNet, VGGNet, ResNet, Xception и т.д. Использование предварительно обученной сверточной нейронной сети. Дообучение.

Лекции – 4 часа. Практические занятия – 4 часа. Лабораторные работы – 0 часов.

Тема 3. Детектирование и сегментация.

Постановка задачи детектирования. Подготовка датасета для задачи детектирования (обучение с учителем). Разбор популярных методов детектирования объектов: R-CNN, Fast R-CNN, Faster R-CNN, YOLO, SSD. Типы сегментации. Постановка задачи сегментации. Подготовка датасета для задачи сегментации (обучение с учителем). Деконволюция. Анпулинг. Разбор популярных методов сегментации объектов: FCN, U-Net, Mask R-CNN.

Лекции – 4 часа. Практические занятия – 4 часа. Лабораторные работы – 0 часов.

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
1	Реализация подготовки датасета для различных задач компьютерного зрения: разметка, загрузка и предварительная обработка данных. Аугментация данных. Построение сверточных нейронных сетей.

2	Реализация дообучения популярной архитектуры сверточной нейронной сети на новом датасете. Сравнение результатов с сетью, обученной с нуля.
3	Решение задачи детектирования объектов в режиме реального времени. Решение задачи сегментации объектов. Для задач использовать любой доступный датасет.

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
1	Работа над Индивидуальным заданием модуля. Постановка задачи, подбор и предобработка обучающего dataset, определение архитектуры ИНС, написание кода, обучение, анализ результатов. Подготовка итоговой презентации.	22

Модуль 4 (вариативный). Решение задач понимания естественного языка

Этот модуль следует выбирать студентам специализирующимся в области решения задач распознавания, для использования компьютерной лингвистики для прикладных задач.

Модуль «**Решение задач понимания естественного языка**» включен в вариативную (по выбору) часть образовательной программы и предусматривает знакомство студентов с основами машинного обучения, используемого для анализа данных. Модуль включает две дисциплины:

- основы математической лингвистики;
- ИНС для решения задач NLU.

Цель изучения дисциплины «Основы математической лингвистики» - познакомить студента с теоретическими знаниями о прагматически-ориентированных лингвистических формальных и концептуальных моделях, этапах и уровнях формализации в лингвистике, о современном состоянии и о наиболее перспективных разработках в области математической лингвистики.

Основные задачи обучения:

- **знать** основные понятия;
- **знать** структурное строение языка;
- **знать** методы решения задач МЛ.

В результате изучения дисциплины «Основы математической лингвистики» студент должен **уметь**:

- ориентироваться в области смежных наук - лингвистики, компьютерных систем и технологий и математики.

В результате изучения дисциплины «Основы математической лингвистики» студент должен **владеть**:

- теоретическими знаниями о прагматически-ориентированных лингвистических формальных и концептуальных моделях, этапах и уровнях формализации в лингвистике, о современном состоянии и о наиболее перспективных разработках в области математической лингвистики.

Цель изучения дисциплины «**ИНС для решения задач NLU**» - освоить методы решения задач NLP и NLU с применением современных нейросетей. К таким задачам относятся следующие прикладные задачи: классификация текста, извлечение полезных данных из текста, ведение диалога, генерация резюме.

Основные задачи обучения:

- **знать** основные архитектуры сетей, решающих задачи NLP и NLU с высокой точностью;
- **знать** методы обучения на текстовых данных;
- **знать** методы предобработки текстовых данных.

В результате изучения дисциплины «**ИНС для решения задач NLU**» студент должен **уметь**:

- выбирать архитектуру сети для решения конкретной задачи;
- выполнять предобработку текста;
- использовать предварительно обученные нейросети.

В результате изучения дисциплины «**ИНС для решения задач NLU**» студент должен **владеть**:

- методами разметки и предобработки текста;
- методами построения классификатора текста;
- методами построения чат-ботов.

М 4.1. Основы математической лингвистики

Лекции - 12 часов, практические занятия - 10 часов, СРС - 22 часа.

Тема 1. Определение базовых понятий.

Квантитативная лингвистика. Когнитивная лингвистика. Психолингвистика. Математическая лингвистика. Компьютерная лингвистика (МЛ). Квантитативная лингвистика. Когнитивная лингвистика. Психолингвистика. Математическая лингвистика. Компьютерная лингвистика (МЛ). Natural language processing (NLP) Теоретическая часть. Основные понятия и определения. Различное толкование базовых понятий.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 2. Основные задачи МЛ.

Исследование ЕЯ и разработка лингвистических моделей исходя из принципа минимальной достаточности. Разработка технологий и средств обработки ЕЯ. Создание машиночитаемых лингвистических ресурсов. Математические модели и инструментарий из математической логики, теории множеств, теории алгоритмов

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 3. Понятие NLP.

Описание функционирования языка. Разные определения языка (включая из теории формальных языков). NLP как научно-прикладная дисциплина по исследованию языков и построению систем обработки ЕЯ (СМП, СПЕЯ).

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 4. Формальные грамматики.

Порождающие и распознающие грамматики. Главные задачи в рамках ФГ

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 5. Базовая классификация грамматик по Хомскому.

Четыре класса грамматик и языков. Грамматики типа "0". Неукорачивающие грамматики (НГ) и свойство легкораспознаваемости языков, порожденных НГ. КЗ-грамматики и НС-грамматики. КС-грамматики, А-грамматики.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 6. Оценка сложности вывода.

Временная и емкостная сложности и соответствующие сигнализирующие функции для всех классов грамматик.

Лекции – 2 часа. Практические занятия – 0 часов. Лабораторные работы – 0 часов.

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
1	Математическая лингвистика. Компьютерная лингвистика (МЛ). Natural language processing (NLP). Основные понятия и определения. Различное толкование базовых понятий.
2	Математические модели и инструментарий из математической логики, теории множеств, теории алгоритмов.
3	NLP как научно-прикладная дисциплина по исследованию языков и построению систем обработки естественных языков ЕЯ (СМП, СПЕЯ).
4	Построение цепочек вывода языков из формальных грамматик. Построение цепочек распознавания принадлежности языку из формальных грамматик.
5	Построение цепочек вывода языков из формальных грамматик. КЗ-грамматики и НС-грамматики. КС-грамматики, А-грамматики.

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
2	Создание машиночитаемых лингвистических ресурсов.	6
4	Формально-концептуальные лингвистические модели.	4
5	Лингвистическая концепция И. А. Мельчука	6

6	Для чего нужен автоматический морфологический разбор слов? Парсинг (синтаксический разбор) стандартный метод, используемый в АОЕЯ.	6
---	--	---

М 4.2. ИНС для решения задач NLU

Лекции - 8 часов, практические занятия - 4 часов, СРС - 17 часов.

Тема 1. Основные задачи NLU и NLP

Классификация намерений (Intent Classification) и тональности (Sentiment analysis). Распознавание именованных сущностей (Named Entity Recognition). Предсказание следующего предложения (Next Sentence Prediction). Вопросно-ответная система (Question-answering system). SQuAD. Диалоговая система (Dialog Management). Машинный перевод (Machine Translation). Классификация sequence-to-sequence задач.

Тема 2. Предобработка текстов

Конвертирование регистров букв. Удаление знаков пунктуации, стоп-слов и лишних whitespaces. Токенизация. Стемминг или лемматизация. Векторизация. Прямое кодирование токенов и мешок слов (BoW - Bag of words). TF-IDF (Term Frequency. – Inverse Document Frequency). Word Embedding. Обучение word2vec: CBoW и Skip-Gram. Отрицательная выборка. Алгоритм FastText и subword-модели.

Тема 3. Подходы, основанные на применении рекуррентных сетей

Рекуррентный нейрон (RNN). Long-short term memory (LSTM). Gated recurrent unit (GRU). Архитектура ELMo (Embedding's from Language Models).

Тема 4. Механизм внимания и трансформеры

Трансформеры и механизм внимания (Attention). Self-attention. Архитектура BERT (Bidirectional Encoder Representations from Transformers). Предобработка Byte Pair Encoding (BPE). Обучение BERT.

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
1	Разбор примера сети LSTM. Решение задачи классификации текста
2	Разбор примера сети, основанной на предобученной модели RuBERT. Решение задачи классификации текста.

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
1	Работа над Индивидуальным заданием модуля. Постановка задачи, подбор и предобработка обучающего dataset, определение архитектуры ИНС, написание кода, обучение, анализ результатов. Подготовка итоговой презентации.	15

Модуль 5 (вариативный). Решение задач предиктивного анализа на временных рядах

Этот модуль следует выбирать обучающимся, специализирующимся на построении прогнозов на основе статистических данных, для анализа и интерпретации результатов, полученных с помощью машинного обучения.

Модуль «Решение задач предиктивного анализа на временных рядах» включен в вариативную (по выбору) часть образовательной программы и предусматривает знакомство студентов с основами машинного обучения, используемого для анализа данных. Модуль включает две дисциплины:

- анализ и модели временных рядов;
- предиктивный анализ временных рядов с использованием ИНС.

Цель изучения дисциплины «**Анализ и модели временных рядов**» - обработка массивов экономических данных в динамике в соответствии с задачей, анализ, оценка, интерпретация полученных результатов и обоснование выводов; построение эконометрических моделей временных рядов, относящихся к области профессиональной деятельности, анализ и интерпретация полученных результатов.

Основные задачи обучения:

- **знать** основную терминологию и модели, применяемые для анализа временных рядов.

В результате изучения дисциплины «Анализ и модели временных рядов» студент должен:

- **уметь** проводить предварительный анализ временного ряда, подбирать модель под особенности временного ряда и экономической проблемы.

В результате изучения дисциплины «Анализ и модели временных рядов» студент должен:

- **владеть** навыками анализа и прогнозирования экономических процессов на основе данных со структурой временного ряда с применением соответствующих эконометрических моделей.

Цель изучения дисциплины «**Предиктивный анализ временных рядов с использованием ИНС**» - освоить методы решения задач прогнозирования поведения системы по историческим данным ее параметров с применением современных инструментов машинного обучения. К таким задачам относятся следующие прикладные задачи: прогнозирование отказов технических систем, прогнозирование котировок финансового инструмента на рынке, прогнозирование спроса и т.д.

М 5.1. Анализ и модели временных рядов

Лекции - 12 часов, практические занятия - 12 часов, СРС - 24 часа.

Тема 1. Стационарные случайные процессы. Временные ряды. Линейные стационарные модели.

Случайные процессы. Понятие о стационарности процесса. Стационарность в широком и узком смысле. Процесс "белого шума". Линейные стационарные процессы. Оценки характеристик временного ряда - среднего, дисперсии, автоковариационной и автокорреляционной функций. Основные модели линейных стационарных процессов.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 2. Модели авторегрессии. Условия стационарности и обратимости уравнения Юла-Уокера.

Регрессия. Экспоненциальное сглаживание. Авторегрессия. Оценка параметров авторегрессионной модели. Процессы авторегрессии первого и второго порядка. Условия стационарности модели. Вывод оценок характеристик временного ряда -

среднего, дисперсии, автоковариационной и автокорреляционной функций. Исследование поведения автокорреляционной функций. Получение уравнений Юла-Уокера, использование их для оценок параметров процесса. Методология Бокса-Дженкинса подбора ARIMA модели для ряда наблюдений.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 3. Модели скользящего среднего. Условия стационарности и обратимости, оценка параметров.

Процессы скользящего среднего первого и второго порядка. Условия стационарности модели. Вывод оценок характеристик временного ряда - среднего, дисперсии, автоковариационной и автокорреляционной функций. Исследование поведения автокорреляционной функций. Вывод формул для получения оценок параметров.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 4. Смешанные модели. Условия стационарности, оценка параметров.

Смешанные процессы. Исследование поведения смешанных процессов. Условия стационарности модели. Вывод оценок характеристик временного ряда - среднего, дисперсии, автоковариационной и автокорреляционной функций. Исследование поведения автокорреляционной функций. Выбор числа неизвестных параметров модели.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 5. Линейные нестационарные модели. Понятие разности процесса. Интегрированные модели.

Тренды и сезонность. Нестационарные модели, причины нарушения условий стационарности. Понятие разности процесса. Фильтр Кальмана. Понятие об интегрированных линейных нестационарных процессах. Сведение исследования нестационарных процессов к исследованию линейных стационарных процессов. Выявление трендов и сезонных компонент нестационарной модели.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Тема 6. Прогнозирование временных рядов. Оценка ошибки прогноза, коррекция прогноза.

Понятие прогноза. Условные математические ожидания как наилучший в среднем квадратическом прогноз линейной модели. Прогноз для моделей авторегрессии, скользящего среднего, смешанной, нестационарной интегрированной. Вычисление ошибок прогноза. Коррекция прогноза на основе наблюдений и новый прогноз. Метод сингулярных функций. Фурье анализ. Вейвлет анализ.

Лекции – 2 часа. Практические занятия – 2 часа. Лабораторные работы – 0 часов.

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
1	Линейные стационарные процессы.
2	Экспоненциальное сглаживание. Авторегрессия. Оценка параметров авторегрессионной модели.
3	Процессы скользящего среднего первого и второго порядка.
4	Вывод оценок характеристик временного ряда - среднего, дисперсии, автоковариационной и автокорреляционной функций.

5	Выявление трендов и сезонных компонент нестационарной модели.
6	Фурье анализ. Вейвлет анализ.

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
2	Сходимость для стохастической аппроксимации авторегрессионного процесса.	6
3	Скользящая средняя в техническом анализе.	6
4	Выбор оптимальной размерности стохастической модели.	6
6	Вейвлет-фильтрация сигналов и изображений. Примеры в MATLAB.	6

М 5.2. Предиктивный анализ временных рядов с использованием ИНС

Лекции - 8 часов, практические занятия - 7 часов, СРС - 18 часов.

Тема 1. Введение в предиктивный анализ временных рядов с помощью ИНС.

Задачи анализа временных рядов. Прогнозирование будущих значений временных рядов. Подготовка обучающих данных (dataset). Предварительная обработка данных для формирования входного пространства ИНС. Нормализация данных. Критерии оценки качества обученных моделей.

Тема 2. Применение рекуррентных сетей.

Рекуррентная нейронная сеть (RNN): определение и принципы. Глубокая рекуррентная сеть. Долгая краткосрочная память (LSTM). Управляемый рекуррентный блок (GRU). Обучение LSTM и GRU. Алгоритм backpropagation through time (обратное распространение во времени)

Тема 3. Применение сетей прямого распространения.

Многослойный персептрон (MLP) в задачах предиктивного анализа временных рядов. Временная свертка. Глубокие сети 1D-CNN. Модели Wavenet.

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий
1	Разбор примера сети RNN. Решение задачи прогнозирования температуры воздуха на основе метеорологических данных.
2	Разбор примера глубокой сверточной сети 1D-CNN. Решение задачи прогнозирования отказа двигателя на основе анализа виброускорений подшипников

Виды самостоятельной работы слушателей (СРС)

№ темы	Вид СРС	Трудоемкость, ч.
	Работа над Индивидуальным заданием модуля. Постановка задачи, подбор и предобработка обучающего dataset, определение архитектуры ИНС, написание кода, обучение, анализ результатов. Подготовка итоговой презентации	18

Стажировка

Стажировка обучающихся осуществляется в целях непрерывного совершенствования их профессионального мастерства. Ее главной задачей является обновление и углубление знаний в научно-профессиональной области на основе ознакомления с современными достижениями науки, прогрессивной техники и технологии.

В течение восьми недель стажер машинного обучения / искусственного интеллекта будет изучать машинное обучение и работать с ним. Он получит практический опыт работы с большими наборами данных, разработки решений на базе искусственного интеллекта и отточит свои навыки моделирования.

Стажер будет исследовать, разрабатывать, внедрять, оптимизировать и использовать модели глубокого обучения для улучшения восприятия и контроля автономного управления.

Типичные повседневные задачи и действия стажировки по машинному обучению включают чтение статей по глубокому обучению, реализацию описанных моделей и алгоритмов, их адаптацию к настройкам и повышение внутренних показателей.

Он также будет разрабатывать алгоритмы в следующих областях: глубокое обучение (сверточные нейронные сети), обнаружение объектов, отслеживание, слияние нескольких датчиков и т. д.

По завершению стажировки специалист получает письменный отзыв от руководителя стажировки (ответственного) в Принимающей организации. Содержание отчета о стажировке определяется индивидуальным заданием стажировки специалиста и должно включать систематизированный материал, полученный и собранный специалистом в период прохождения стажировки, а также может включать краткое описание Принимающей организации, направления ее деятельности, выводы и предложения по совершенствованию производства и др.

Результаты работы стажера защищаются на предприятии и докладываются куратору стажировки. По результатам прохождения стажировки выставляется отметка «зачтено/не зачтено», согласно уровню выполненной работы и оценке работы представителем предприятия.

Методические рекомендации по изучению дисциплин

Вид работ	Методические рекомендации
Лекции	Необходимо вести конспект лекций, который ведется в отдельной общей тетради, рекомендуется оставлять место для заметок, например, в виде полей. Требуется знание основного материала предыдущих лекций, включая знание основных определений и ключевых теорем. Рекомендуется выделять в тексте ключевые слова, определения, леммы и теоремы.

Практическое занятие	В ходе подготовки к практическим занятиям студент должен изучить основную литературу, лекции. Внимательно слушать и конспектировать базовые примеры, разбираемые преподавателем. Задавать уточняющие вопросы в ходе решения базовых задач преподавателем. При решении домашних заданий периодически возвращаться к разобранным на практических занятиях примерах. Своевременно и полностью решать задачи на самостоятельную работу. Своевременное и качественное выполнение самостоятельной работы базируется на соблюдении настоящих рекомендаций и изучении рекомендованной литературы. Желательно задавать вопросы про те сложные места в решениях задач, которые вызвали затруднение при самостоятельной работе. Студент может дополнить список использованной литературы современными источниками, не представленными в списке рекомендованной литературы.
Самостоятельная работа	Эта работа ведется в той же тетради, что и практические занятия. Самостоятельная работа является отдельным блоком, который выделяется заголовком, например, "Домашнее задание". Рекомендуется прорабатывать материал непосредственно после практических занятий. При решении задач и примеров рекомендуется их выполнение по образцу из практического занятия.
Индивидуальное задание	Подготовка к промежуточной аттестации, предусмотренной в виде индивидуального задания, ведется на основе курса лекций и рекомендованной литературы. Необходимо знание и понимание всех понятий, определений, утверждений, лемм и теорем. Необходимо умение формулировать теоремы в форме непротиворечивых логических конструкций. Желательно уметь строить и приводить примеры к соответствующим определениям и утверждениям. Необходимо знание доказательства теорем и остальных утверждений. Студенту предлагается решить реальную прикладную задачу с использованием изученных методов.

3. Организационно - педагогические условия реализации программы

В процессе изучения тем по данной образовательной программе используются различные образовательные технологии (технологии тестирования, технологии проектного обучения, информационно-коммуникационные технологии, а также дистанционные образовательные технологии) как в проведении лекционных, практических занятий, так и самостоятельной работы, итоговой аттестации слушателей. Применение технологий и их сочетание определяется преподавателями, ведущими обучение по темам программы, самостоятельно.

Итоговая аттестация слушателей может проводиться в традиционном и\или дистанционном режиме.

ИКТ и дистанционные образовательные технологии применяются посредством работы слушателей и преподавателей с использованием различных программ (BigBlueButton, эл. почты, Telegram)

Комплексное использование в учебном процессе всех вышеназванных технологий стимулирует личностную, интеллектуальную активность, развивает познавательные процессы, способствует формированию профессиональных компетенций.

Тип контроля	Форма контроля	Параметры
Текущий	Домашнее задание	Решение учебных задач
Итоговый	Защита индивидуального проекта	Презентация проекта. Скрипт Jupyter notebook с описанием на Markdown

3.1. Учебно-методическое обеспечение программы

Основная литература

1. Боровков А. А. Математическая статистика. [электронный ресурс] Издательство "Лань" ISBN 978-5-8114-7677-0, 5-е изд., стер., 704 с., 2021 <https://e.lanbook.com/book/164711>
2. Зубков А. М., Севастьянов Б. А., Чистяков В. Сборник задач по теории вероятностей [электронный ресурс] Издательство "Лань" ISBN 978-5-8114-9085-1, 4-е изд., стер., 320 с., 2022 <https://e.lanbook.com/book/184062?category=913>
3. Ивченко Г.И., Медведев Ю.И. Введение в математическую статистику: [электронный ресурс] Учебник. М.: Издательство ЛКИ, 2010. – 600 с. - Режим доступа: <https://libren.org/d/math-stud/math-st1024.htm>
4. Розанов Ю.А. Теория вероятностей, случайные процессы и математическая статистика: [электронный ресурс] Учебник для вузов. - М.: Наука, 1985, 320 с. - Режим доступа: <https://bookree.org/reader?file=792869&pg=4>
5. Лагутин М.Б. Наглядная математическая статистика: [электронный ресурс] учебное пособие. Издательство "Лаборатория знаний" ISBN 978-5-00101-642-7, 7-е изд., 475 с., 2019. - Режим доступа: <https://e.lanbook.com/book/116104>
6. Свешников А. А. Прикладные методы теории марковских процессов [электронный ресурс] Издательство "Лань" ISBN 978-5-8114-9341-8, 2-е изд, стер. 192 с., 2022. - Режим доступа: <https://e.lanbook.com/book/189439?category=913>
7. Дерр В. Я. Теория вероятностей и математическая статистика: [электронный ресурс] Учебное пособие для вузов. Издательство "Лань" ISBN 978-5-8114-6515-6, 596 с., 2021. - Режим доступа: <https://e.lanbook.com/book/159475?category=913>
8. Методы оптимизации : учебное пособие / И.В. Гребенникова.— Екатеринбург : УрФУ, 2017.— 148 с. ISBN 978-5-7996-2090-5 Режим доступа: https://elar.urfu.ru/bitstream/10995/48965/1/978-5-7996-2090-5_2017.pdf
9. Данилин А.И. Основы теории оптимизации (постановки задач) [Электронный ресурс]: электрон. учеб. пособие / Самар. гос.аэрокосм. ун-т им. С.П.Королева – Самара, 2011 - Режим доступа: <http://repo.ssau.ru/bitstream/Uchebnye-posobiya/Osnovy-teorii-optimizacii-postanovki-zadach-Elektronnyi-resurs-elektron-ucheb-posobie-54681/1/Данилин%20А.И.%20Основы%20теории%20оптимизации.pdf>
10. Аббасов М. Э. Методы оптимизации: Учеб. пособие / Аббасов М. Э. — СПб.: Издательство “ВВМ”, 2014. — 64 с. ISBN 978-5-9651-0875-6 Режим доступа: http://www.apmath.spbu.ru/ru/staff/abbasov_m_e/publ/publ4.pdf
11. Микони, С.В. Дискретная математика для бакалавра: множества, отношения, функции, графы [Электронный ресурс] : учебное пособие / С.В. Микони. - Электрон. дан. - Санкт-Петербург : Лань, 2012. - 192 с. - Режим доступа: <https://e.lanbook.com/book/4316>

12. Попов, С.В. Прикладная логика [Электронный ресурс] : учебное пособие / С.В. Попов, Н.Л. Брошкова. - Электрон. дан. - Москва : Физматлит, 2011. - 216 с. - Режим доступа: <https://e.lanbook.com/book/5286>
13. Математика и информатика: Учебник / В.Я. Турецкий; Уральский государственный университет им. А.М. Горького. - 3-е изд., испр. и доп. - М.: ИНФРА-М, 2010. - 558 с. - (Высшее образование). ISBN 978-5-16-000171-5 - Режим доступа: <http://znanium.com/catalog/product/206346>
14. С. Осовский Нейронные сети для обработки информации М.: Финансы и статистика, 2002. - 344 с.
15. С. Хайкин Нейронные сети: полный курс М.: Вильямс, 2006. – 1104 с.
16. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение М.: ДМК Пресс, 2017. – 654 с.
17. Хобсон, Ханнес Коул Обработка естественного языка в действии. Санкт-Петербург: Питер. 2020. 576 с.
18. Christopher Olah Understanding LSTM Networks // Colah's blogs, August 2015 [<http://colah.github.io/posts/2015-08-Understanding-LSTMs>]
19. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin Attention Is All You Need // arXiv:1706.03762.2017 [<https://arxiv.org/pdf/1706.03762.pdf>]
20. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // arXiv:1810.04805. 2019. [<https://arxiv.org/pdf/1810.04805.pdf>]
21. Гонсалес Р., Вудс Р. Цифровая обработка изображений. - М.: Техносфера, 2005. - 1072 с.

Дополнительная литература

1. Плотников А. Н. Элементарная теория анализа и статистическое моделирование временных рядов [электронный ресурс] Издательство "Лань" ISBN 978-5-8114-7748-7, 212 с., 2021. - Режим доступа: <https://e.lanbook.com/book/179030>
2. Ларионова И.А. Статистика. Анализ временных рядов: [электронный ресурс] Учебное пособие. Издательство "МИСИС", 72 с., 2001. - Режим доступа: <https://e.lanbook.com/book/117457>
3. Васильев Ф.П. Методы оптимизации. М.: Факториал Пресс, Гл. ред. физ.-мат. лит., 2002.– 824 с.
4. Пшеничный Б.Н., Данилин Ю.М. Численные методы в экстремальных задачах. М.: Наука, 1975. – 320 с.
5. Мину М. Математическое программирование. Теория и алгоритмы. М.: Наука. Гл. ред. физ.-мат. лит., 1990. 488 с.
6. Моисеев Н.Н., Иванилов Ю.П., Столярова Е.М. Методы оптимизации. М.: Наука. Гл. ред. физ.-мат. лит., 1978. 352 с. 5. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. Пер. с англ. — М.: Мир, 1985. 509 с.
7. Информационные системы: Учебное пособие / О.Л. Голицына, Н.В. Максимов, И.И. Попов. - 2-е изд. - М.: Форум: НИЦ ИНФРА-М, 2014. - 448 с. - (Высшее образование). ISBN 978-5-91134-833-5 - Режим доступа: <http://znanium.com/catalog/product/435900>
8. Теоретические основы информационных процессов и систем / Душин В.К., - 5-е изд. - М.: Дашков и К, 2018. - 348 с.: ISBN 978-5-394-01748-3 - Режим доступа: <http://znanium.com/catalog/product/450784>
9. Информационные технологии и системы: Учебное пособие / Е.Л. Федотова. - М.: ИД ФОРУМ: НИЦ ИНФРА-М, 2014. - 352 с. - (Высшее образование). ISBN 978-5-8199-0376-6 - Режим доступа: <http://znanium.com/catalog/product/429113>

10. С. И. Николенко, А. Кадулин, Е. В. Архангельская Глубокое обучение. Погружение в мир нейронных сетей. СПб.: Питер, 2018. — 480 с
11. Головкин В.А. Нейронные сети: обучение, организация и применение. Кн.4: Учеб. пособие для вузов /Общая ред. А.И.Галушкина.-М.:ИПРЖР,2001.-256 с.:ил.
12. Chiu, Jason P. C. and Eric Nichols. Named Entity Recognition with Bidirectional LSTM-CNNs. TACL 4 (2016): 357-370.
13. Автоматическая обработка текстов на естественном языке и анализ данных: учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с
14. Шолле, Ф. Глубокое обучение на Python // СПб.: Питер. – 2018. – 400 с.
15. Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. – O'Reilly Media, 2019.

3.2. Материально-технические условия реализации программы

Требования к рабочему месту слушателя при использовании дистанционных образовательных технологий:

- компьютер или мобильное устройство, подключенное к сети Интернет. Для участия в вебинарах обязательно наличие веб-камеры и микрофона;
- программное обеспечение: Интернет-браузер (Mozilla Firefox, Google Chrome, Safari и т.д.), Adobe Reader, программа для проигрывания видеофайлов (например, Windows Media player), Python 3, Jupyterlab.

Материально-техническое обеспечение образовательного процесса по дисциплине для обучающихся из числа лиц с ограниченными возможностями здоровья осуществляется согласно «Порядку организации и осуществления образовательной деятельности по образовательным программам для инвалидов и лиц с ограниченными возможностями здоровья в Пермском национальном исследовательском политехническом университете».

3.3. Кадровый состав

Кадровое обеспечение необходимое для проведения занятий по программе:

- 2(два) кандидата/доктора физ.-мат. наук/тех. наук, специализирующихся в соответствующих областях и имеющих опыт преподавания не менее 5 лет,
- 2(два) ассистента/ведущих специалиста и практика компаний, без степени, специализирующихся в соответствующих областях

4. Оценка качества освоения программы

4.1. Формы аттестации

Оценка качества освоения ДПП профессиональной переподготовки проводится в отношении соответствия результатов освоения программы заявленным целям и планируемым результатам обучения.

Оценки по всем формам текущего контроля выставляются по 5-ти балльной шкале. В качестве текущего контроля знаний преподаватель оценивает работу студентов на семинарских занятиях: активность на занятиях, правильность выполнения заданий, использование дополнительного материала, участие в дискуссиях и групповой работе.

Тип контроля	Форма контроля	Параметры
Текущий	Домашнее задание	Решение учебных задач
Промежуточный	Устный зачет по дисциплине	Собеседование по комплексному применению навыков, полученных в результате обучения, для решения прикладных задач
Итоговый	Защита индивидуального проекта	Презентация проекта. Скрипт Jupyter notebook с описанием на Markdown

4.2. Оценочные материалы

Вынесены в Приложение 1.

Порядок проведения промежуточной аттестации

Задания для промежуточной аттестации выкладываются на странице курса и в группе курса. Задания нацелены на практическое применение изученных на занятиях методов и алгоритмов. Выполненные задания сдаются в электронном виде. В каждом задании есть теоретическая и практическая часть. Итоговое задание формулируется преподавателем и представляет из себя задачу из предметной области курса (анализ данных и машинное обучение). Задание выполняется с использованием языка программирования Python и специализированных библиотек, использованных данных, указание источника, загрузку данных, разведочный анализ, визуализацию описательных статистик, исследование с помощью методов снижения размерности, описание постановки регрессионной модели (в зависимости от типа задачи), обоснование выбора метода и анализ качества полученной модели. Работа должна быть выполнена методологически корректно, без грубых ошибок. Программный код должен исполняться без фатальных ошибок. Студент должен понимать суть выполненной работы и быть готов дать пояснения по программному коду. Задание сдается либо в виде файла *.ipynb – iPython Notebook, либо в виде презентации с приложением программного кода. Для решения задания допускается использовать язык программирования R. Необходимым условием для прохождения промежуточной аттестации является оценка «зачтено» за портфолио. Оценка «зачтено» за портфолио выставляется при условии выполнения и защиты всех заданий, входящих в портфолио.

Промежуточная аттестация по дисциплине проводится в виде зачета. Оценка «зачтено» соответствуют успешному прохождению промежуточной аттестации.

Порядок проведения итогового экзамена

Итоговый экзамен проводится в форме защиты индивидуального проекта по решению конкретной прикладной задачи, требующей навыков приобретенных во время изучения курса. При этом, возможна сдача итоговой аттестации с использованием дистанционной образовательной платформы.

Слушатель регистрируется на платформе Discord. За 10 минут до начала итоговой аттестации обучающийся присоединяется к видеоконференции, пройдя по высланной за день до защиты на электронную почту ссылке. Включает режим видео и показывает развернутый паспорт для идентификации личности комиссии. Загружает свою презентацию и в форме доклада приводит решение своего индивидуального задания. Для

доклада отводится 15 минут. Потом в режиме дискуссии идет обсуждение работы, задаются уточняющие вопросы по выполнению индивидуального задания. На обсуждение каждой работы отводится 20 минут. После защиты всех студентов комиссия обсуждает преимущества и недостатки выслушанных проектов и оценивает уровень выполнения индивидуального задания по проекту и на закрытом заседании принимает решение об освоении слушателями программы. После выставления всех баллов слушатели вновь подключаются к видеоконференции, где каждому озвучивают окончательную оценку за индивидуальный проект и за весь курс.

Система рейтинговой оценки и контроля знаний студентов

№ п/п	Структура	Баллы по каждому модулю
1.	Оценка за активное участие в учебном процессе и посещение занятий: всех занятий не менее 75% не менее 50 % не менее 25% итого:	5 4 3 2 до 5
2.	Устный опрос в форме собеседования Письменная работа – отчет о выполнении индивидуального задания по модулю, содержащая решение прикладной проблемы, выполненная с помощью методов и технологий освоенных студентами во время изучения модуля Итого	20 20 40
3.	Защита Индивидуального задания	55
	Всего	100

Пересчет на 5 балльную систему

2 (неудовлетворительно)	3 (удовлетворительно)	4 (хорошо)	5 (отлично)
<50	50-64	65-84	85-100

По окончании программы обучения слушателям, успешно прошедшим итоговую аттестацию, получившим положительную оценку, выдается диплом о профессиональной переподготовке.

Критерии оценки итогового экзамена:

Критерии оценки	«Отлично»	«Хорошо»	«Удовлетворительно»	«Неудовлетворительно»
Степень освоения результатов обучения	Освоены полностью	Освоены в большей степени	Освоены частично	Не освоены
Содержание ИАР	Теоретический материал программы изучен полностью, слушатель показал углубленные знания	Теоретический материал программы изучен в достаточной степени, слушатель	Теоретический материал программы изучен частично, слушатель имеет	Теоретический материал программы не изучен. Слушатель

	по освоению программы	показал хорошие знания по освоению программы	знания только по нескольким дисциплинам по освоению программы	имеет общее понятийное представление о программе
	Слушатель владеет навыками оценки, анализа и применения в практической деятельности, полученных в процессе проектирования, данных	Слушатель владеет навыками систематизации полученных в процессе проектирования данных	Слушатель имеет представление о полученных в процессе проектирования данных	Отсутствует представление о полученных в процессе проектирования данных
Доклад	Доклад четкий, регламентированный, дающий полное представление о выполненной работе	Доклад четкий, регламентированный показывает достаточное представление о выполненной работе	Доклад не четкий, с отступлениями, показывает частичное представление о выполненной работе	Доклад с отступлениями, не показывает представление о выполненной работе
Ответы на вопросы	Ответы полные, уверенные	Ответы недостаточно полные, уверенные	Ответы не полные, не уверенные	Не может ответить на вопросы

Итоговая аттестационная комиссия рассматривает работы слушателей, оценивает и на закрытом заседании принимает решение об освоении слушателями программы.

По окончании программы обучения слушателям, успешно прошедшим итоговую аттестацию, получившим положительную оценку, выдается диплом о профессиональной переподготовке.

5. Составители программы

Данелян Вадим Самвелович, магистрант 2 курса кафедры «Вычислительная математика, механика и биомеханика» (модуль 3)

Клюев Андрей Владимирович, канд. физ.-мат. наук, доцент кафедры «Вычислительная математика, механика и биомеханика» (разделы М1.3, разделы М2.2, разделы М4.2, раздел М5.2)

Кротова Елена Львовна, канд. физ.-мат. наук, доцент кафедры «Высшая математика» (модуль 1 раздел 1.1-1.2, модуль 2 раздел 2.1, модуль 4 раздел 4.1, модуль 5 раздел 5.1)

Программа обсуждена на заседании кафедры высшей математики. Протокол № 10 от 26 мая 2022 г.

Секретарь

 Ю.В. Уразбаева

Руководитель программы

 Е.Л. Кротова

Директор ЦФО

 Е.Л. Кротова

СОГЛАСОВАНО

Заместитель начальника УМУ

 И.Л. Герасимчук

Оценочные материалы

Раздел М 1.1 Основы теории информации и математической статистики

Примеры практических заданий

1. Равномерно распределённая случайная величина X принимает значения из алфавита $\{0000, 0001, \dots, 1001\}$ (числа от 0 до 9, записанные с помощью четырех битов). Вычислить энтропию каждого бита.
2. Сформировать набор векторов ошибок и опознавателей группового кода, предназначенного для передачи 15 символов, позволяющего исправлять одиночные и обнаруживать двойные ошибки.
3. Найти частоту квантования по времени сигнала $x(t)$ (равномерное распределение, Гамма распределение), если относительная величина площади отсекаемой части энергетического спектра не превышает определенную величину.

Пример Индивидуального задания

По выборке из 2×100 наблюдений провести полный статистический анализ. Проверить данные на случайность. Построить гипотезу и типе распределения случайных величин (X и Y) и выбрать альтернативную гипотезу. Оценить неизвестные параметры выбранных распределений разными методами. Выбрать оптимальные оценки. Оценить ошибки первого и второго рода для построенных статистических решающих правил. Проверить существует ли зависимость между столбцами данных. Определить ее вид и построить сглаживающую кривую. Оценить ошибку построенной аппроксимации.

Раздел М 1.2 Основы теории оптимизации

Пример практических заданий

1. Решить задачу линейного программирования Симплекс-методом и графическим методом.
2. Автомобиль выезжает из пункта A и едет с постоянной скоростью v км/ч до пункта B , отстоящего от пункта A на 24,5 км. В пункте B автомобиль переходит на равнозамедленное движение, причем за каждый час его скорость уменьшается на 54 км/ч, и движется так до полной остановки. Затем автомобиль сразу же поворачивает обратно и возвращается в A с постоянной скоростью v . Какова должна быть скорость v , чтобы автомобиль за наименьшее время проезжал путь от A до полной остановки и обратно до пункта A указанным способом?
3. Показать, что любая из точек локального минимума функции $f(x)$, унимодальной на отрезке $[a, b]$, является и точкой ее глобального минимума.
4. Может ли применение методов сокращения отрезков поиска привести к неверному определению точки минимума x^* , если функция $f(x)$ не является унимодальной? Ответ пояснить иллюстрацией.

Пример индивидуального задания.

Для квадратичной трехмерной функции найти точку минимума, с точностью 0,000001. Задание выполнить с использованием метода наискорейшего градиентного спуска и наискорейшего покоординатного спуска.

Пример модульного индивидуального задания

Раздел М 1 Основы машинного обучения и искусственных нейронных сетей

Решить задачу прогнозирования поведения финансового временного ряда с помощью многослойной сети (MLP). Исследовать эффективность различных методов обучения

(алгоритм переменной метрики, алгоритм Левенберга-Марквардта, алгоритм сопряженных градиентов) по сравнению с классическим алгоритмом.

Раздел М 2.1 Отдельные главы математической статистики

Примеры практических заданий.

1. Декан полагает, что среди студентов девушки чаще, чем юноши, прогуливают занятия. Выборочное исследование показало, что девушки пропустили в среднем 3,9 дня в семестр, а юноши 3,6 дня. В исследовании участвовало 16 девушек и 22 юношей. Стандартные отклонения составили 0,6 и 0,8 дня соответственно. Проверьте предположение исследователя на уровне значимости $\alpha = 0,01$. Считать, что генеральные дисперсии равны.
2. По выборке из двумерного нормального распределения проверить гипотезу независимости компонентов наблюдаемого случайного вектора (X, Y) . Построить линии регрессии одного из признаков по другому признаку. Найти наилучший прогноз признака Y при фиксированном значении признака X .

Пример индивидуального задания.

Для выборки наблюдений случайного вектора (X, Y)

1. Определить выборочный коэффициент корреляции величин X и Y .
2. Составить уравнение линейной регрессии Y на X . Построить график уравнения линейной регрессии на одном чертеже с опытными данными.
3. Оценить качество линейной модели регрессии по коэффициенту детерминации
4. Провести оценку значимости параметров уравнения регрессии при $\alpha = 0,1$.
5. Провести оценку значимости параметров уравнения регрессии при $\alpha = 0,1$.
6. На уровне значимости $\alpha = 0,1$ найти доверительный интервал, в который попадает прогнозное значение фактора y .

Раздел М 4.1 Основы математической лингвистики

Примеры практических задач

1. Задана КС-грамматика $G = \langle V, W, S, R \rangle$, где V – терминальный алфавит, W – нетерминальный алфавит, S – аксиома, R – множество правил. Обозначим через L – язык, порождаемый грамматикой G . Какие из следующих утверждений являются верными?

1. Если в правой части (то есть после стрелки) любого правила из R находятся только символы из W^+ , то грамматика G порождает непустой язык.
2. Если в грамматике G возможен хотя бы один полный вывод, то грамматика G порождает пустой язык.
3. Если в грамматике G возможен хотя бы один полный вывод, то грамматика G порождает непустой язык.
4. Если $R = \{S \rightarrow S\}$, то грамматика G порождает пустой язык.
5. L всегда совпадает с V^* .
6. Если терминальные символы грамматики G не встречаются в правых частях правил из R , то грамматика G порождает пустой язык.
7. Множества V и W могут содержать некоторые общие элементы.
8. Если некоторая цепочка содержится в L , то её полный вывод не начинается с аксиомы S .
9. Некоторые терминальные символы могут содержаться в W .
10. Если некоторая цепочка содержится в L , то некоторые входящие в её состав символы могут быть нетерминальными.

11. Существует цепочка, содержащаяся в V^* , но не содержащаяся в V^+ .
12. Если x – элемент множества V^* , то x может содержать нетерминальные символы.

2. Рассмотрим высказывание Рано встаёт охрана. Смысл не изменится, если наречие рано переместить в последнюю позицию: Встаёт охрана рано. В тех же двух позициях можно вместо рано употребить другие наречия: весело, бодро, лениво, медленно, быстро. Рассмотрим множество предложений, образующихся путём добавления к словосочетанию встаёт охрана в первой или последней позиции одного (и только одного) из названных шести наречий. Постройте КС-грамматику, которая порождает бы в точности это множество предложений.

3. Ниже написано, как комментирует происхождение слова «орангутанг» А. Э. Брэм в своей книге Жизнь животных (текст немного изменён в соответствии с целями задания).

«Название «орангутанг» в переводе означает «лесной человек», так как по мнению туземцев эти обезьяны – настоящие люди и они хорошо говорили бы, если бы захотели, но это не происходит, так как они боятся, как бы не заставили их работать»

Постройте логическую форму этого предложения, воспользовавшись средствами пропозициональной логики.

4. Посредством логических языков можно описывать не только те смыслы, которые явно выражены в тексте, но и те смыслы, которые лишь подразумеваются или считаются само собой разумеющимися, но в тексте явным образом не отражены. Следующий заголовок был опубликован в одной из газет города N:

В городе N штрафуют владельцев собак, гуляющих без намордников и поводков

Скорее всего, автор статьи не заметил неоднозначности заголовка своей заметки, поскольку подразумевал (как само собой разумеющееся) следующую пропозицию 'намордники и поводки носят собаки, а не их хозяева'. Такие пропозиции, не высказываемые в предложении явным образом и считающиеся заведомо истинными, называется пресуппозициями данного предложения. Представьте средствами логики предикатов оба варианта смысла заданного предложения. Приведем два предложения

(1) Лектор рассуждал о естественном и искусственном интеллекте

(2) Лектор рассуждал о редком и малоизвестном языке

По своему устройству эти предложения похожи, однако им были поставлены в соответствие разные логические формы. На основе каких пресуппозиций это было сделано? Опишите смысл использованных пресуппозиций в виде формул предикатной логики.

Пример Индивидуального задания

Резюмируйте сложности применения логических языков, которые вы заметили в ходе изучения раздела.

Пример модульного индивидуального задания

Раздел М 2 Методы решения задач с ассоциативной памятью

Решить задачу классификации рукописных цифр (по базе MNIST). Использовать сети MLP, Хопфилда, глубокий автоэнкодер. Провести сравнительный анализ эффективности работы данных сетей.

Раздел М 3.1 Классические методы решения задач компьютерного зрения

Примеры практических заданий.

1. Взять несколько разнообразных изображений и решить задачу выделения границ с помощью детекторов Кэнни и Собеля.
2. Реализовать метод k-средних для сегментации изображений.

Пример индивидуального задания.

Решить задачу детектирования автомобильных номеров в режиме реального времени с применением каскадов Хаара.

Раздел М 3.2 Сверточные нейронные сети

Примеры практических заданий.

1. Реализовать аугментацию небольшого набора данных (например, взять кошек и собак). Проанализировать полученные изображения.
2. Решить задачу классификации кошек и собак. Использовать любую из изученных архитектур сверточных нейронных сетей.

Пример индивидуального задания.

Решить задачу детектирования автомобильных номеров в режиме реального времени с применением нейронной сети YOLO.

Пример модульного индивидуального задания

Раздел М 3 Решение задач компьютерного зрения

Решить задачу классификации изображений банкнот различных валют. Использовать несколько сетей CNN разных архитектур. Провести сравнительный анализ эффективности работы. Проверочное множество должно состоять из снимков частей банкнот, снимков в разных масштабах, снимков под разными углами и освещением.

Пример модульного индивидуального задания

Раздел М 4 Решение задач понимания естественного языка

С помощью ИНС построить и обучить родовидовой классификатор текстов художественной литературы. Использовать RNN/LSTM/BERT.

Раздел М 5.1 Анализ и модели временных рядов

Примеры практических задач

1. Программа выдала следующие характеристики ряда остатков: длина ряда $n = 20$; коэффициент асимметрии $A = 0,6$; коэффициент эксцесса $\mathcal{E} = 0,7$. На основании этих характеристик проверить гипотезу о нормальном законе распределения остаточной последовательности.
2. Для прогнозирования численности промышленно-производственного персонала предприятия была выбрана линейная модель. Оценка параметров трендовой модели осуществлялась по квартальным данным за период с I квартала 2018 г. по IV квартал 2021 г. Значение статистики Дарбина-Уотсона для ряда остатков $d =$

1,39. Проверить гипотезу об отсутствии в остатках автокорреляции первого порядка (уровень значимости $\alpha = 0,05$).

Пример индивидуального задания

Имеются обезличенные (смоделированные) данные, состоящие из 1000 наблюдений за значением случайной величины, описывающей процесс N за последние 16 периодов (кварталов).

Требуется:

1. Построить автокорреляционную функцию и сделать вывод о наличии сезонных колебаний.
2. Построить аддитивную модель временного ряда и мультипликативную модель временного ряда.
3. Сделать прогноз на 2 квартала вперед.
4. Составить модель временного ряда с сезонной составляющей.
5. Сравнить с прогнозом, построенным с помощью скользящей средней.
6. Оценить ошибки построенных прогнозов.

Раздел М 5 Решение задач предиктивного анализа на временных рядах

Используя LSTM и MLP решить задачу прогнозирования цен актива на фондовом рынке. Провести сравнительный анализ эффективности работы.