

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Ректор

Дата подписания: 28.03.2023 15:02:53

Уникальный программный ключ:


c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ

Директор Института магистратуры

 Иванова Е.А.  
« 30 » 03 2024 г.

**Рабочая программа дисциплины  
Обучение с подкреплением и его приложения**

Направление 01.04.02 Прикладная математика и информатика  
магистерская программа 01.04.02.03 "Искусственный интеллект в цифровой экономике"

Для набора 2021 года

Квалификация  
Магистр


**КАФЕДРА      Фундаментальная и прикладная математика**

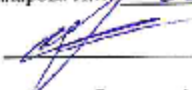
**Распределение часов дисциплины по семестрам**

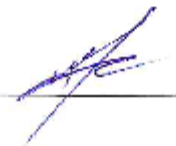
Семестр («Курс»-«Семестр на курсе»)	4 (2.2)		Итого:	
	13 4/6			
Вид занятий	зп	пз	зп	пз
Лекции	10	10	10	10
Практические	12	12	12	12
Итог з.ч.	22	22	22	22
Контактная работа	22	22	22	22
Сам. работа	77	77	77	77
Часы на контроль	9	9	9	9
Итого	108	108	108	108

**ОСНОВАНИЕ**

Учебный план утвержден ученым советом вуза от 30.08.2021 протокол № 1.

Программу составил(и): д.ф.-м.н., проф., Сахарова Л.В. 

Зав. кафедрой: д.ф.-м.н., доц. Стрюков М.Б. 

Методическим советом направления: д.ф.-м.н., доц. Стрюков М.Б. 

### 1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	выработка у студентов компетенций, связанных с применением алгоритмов обучения с подкреплением при решении практических задач.
-----	--

### 2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

**ПК-2: Способен разрабатывать и анализировать концептуальные и теоретические модели решаемых научных проблем и задач**

**ПК-3: Способен разрабатывать и применять математические методы, системное и прикладное программное обеспечение для решения задач научной и проектно-технологической деятельности**

**ПК-4: Способен разрабатывать и анализировать концептуальные и теоретические модели решаемых задач проектной и производственно-технологической деятельности**

**ПК-6: способен разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов**

#### В результате освоения дисциплины обучающийся должен:

**Знать:**

- методы интеллектуального анализа данных, необходимых для построения моделей решаемых научных задач (соотнесено с индикатором ПК-2.1)
- функциональность современных инструментальных средств и систем программирования в области создания моделей искусственных нейронных сетей.(соотнесено с индикатором ПК-3.1)
- методы оптимизации построения концептуальных и теоретических моделей решаемых задач проектной и производственно-технологической деятельности.(соотнесено с индикатором ПК-4.1)
- теоретические основы оценки эффективности и учета рисков инвестиционных проектов, основы современных информационных технологий, применяемых в экономическом анализе (соотнесено с индикатором ПК-6.1)

**Уметь:**

- применять методы интеллектуального анализа данных, необходимых для построения моделей решаемых научных задач (соотнесено с индикатором ПК-2.2)
- проводить оценку и выбор моделей искусственных нейронных сетей и инструментальных средств для решения задач машинного обучения.(соотнесено с индикатором ПК-3.2)
- применять методы оптимизации построения концептуальных и теоретических моделей решаемых задач проектной и производственно-технологической деятельности.(соотнесено с индикатором ПК-4.2)
- моделировать и анализировать бизнес-процессы предприятия, находить организационно управленческие решения в моделях управления.(соотнесено с индикатором ПК-6.2)

**Владеть:**

- навыками применения методов интеллектуального анализа данных, необходимых для построения моделей решаемых научных задач (соотнесено с индикатором ПК-2.3)
- навыками применения современных инструментальных средств и систем программирования для разработки и обучения моделей искусственных нейронных сетей.(соотнесено с индикатором ПК-3.3)
- навыками применения методов оптимизации построения концептуальных и теоретических моделей решаемых задач проектной и производственно-технологической деятельности.(соотнесено с индикатором ПК-4.3)
- навыками планирования научно-исследовательской деятельности, анализа инвестиционных рисков, управления командой проекта; навыками работы в коллективе (соотнесено с индикатором ПК-6.3)

### 3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература
	<b>Раздел 1. Мета эвристики. Классическая теория</b>				
1.1	Тема 1.1. Модель взаимодействия агента со средой Связь с оптимальным уравнением. Марковская цепь. Среда. Действия. Траектории. Марковский процесс принятия решений (MDP). Эпизодичность. Дисконтирование. /Лек/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.2	Тема 1.2. Алгоритмы обучения с подкреплением Условия задачи RL. On-policy. Концепция model-free алгоритмов. Классификация RL-алгоритмов. Критерии оценки RL-алгоритмов. Сложности задачи RL. Бенчмарки. /Лек/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3

1.3	Тема 1.2. Алгоритмы обучения с подкреплением Условия задачи RL. On-policy. Концепция model-free алгоритмов. Классификация RL-алгоритмов. Критерии оценки RL-алгоритмов. Сложности задачи RL. Бенчмарки. /Пр/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.4	Тема 1.2. Алгоритмы обучения с подкреплением Условия задачи RL. On-policy. Концепция model-free алгоритмов. Классификация RL-алгоритмов. Критерии оценки RL-алгоритмов. Сложности задачи RL. Бенчмарки. /Ср/	4	6		
1.5	Тема 1.3. Бэйзлайны Задача безградиентной оптимизации. Случайный поиск. Hill Climbing. Имитация отжига. Эволюционные алгоритмы. Weight Agnostic Neural Networks (WANN). Видовая спецификация. Генетические алгоритмы. /Пр/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.6	Тема 1.3. Бэйзлайны Задача безградиентной оптимизации. Случайный поиск. Hill Climbing. Имитация отжига. Эволюционные алгоритмы. Weight Agnostic Neural Networks (WANN). Видовая спецификация. Генетические алгоритмы. /Ср/	4	6	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.7	Тема 1.4. Эволюционные стратегии Идея эволюционных стратегий. Оценка вероятности редкого события. Метод Кросс-Энтропии для стохастической оптимизации. Метод Кросс-Энтропии для обучения с подкреплением (CEM). Натуральные эволюционные стратегии (NES). OpenAI-ES. Адаптация матрицы ковариации (CMA-ES). /Пр/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.8	Тема 1.4. Эволюционные стратегии Идея эволюционных стратегий. Оценка вероятности редкого события. Метод Кросс-Энтропии для стохастической оптимизации. Метод Кросс-Энтропии для обучения с подкреплением (CEM). Натуральные эволюционные стратегии (NES). OpenAI-ES. Адаптация матрицы ковариации (CMA-ES). /Ср/	4	6	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.9	Тема 2.1. Оценочные функции Свойства траекторий. V-функция. Уравнение Беллмана. Оптимальная стратегия. Q-функция. Принцип оптимальности Беллмана. Отказ от однородности. Вид оптимальной стратегии (доказательство через отказ от однородности). Уравнение оптимальности Беллмана. Критерий оптимальности Беллмана. /Лек/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.10	Тема 2.2. Relative Performance Identity Advantage-функция. Relative Performance Identity (RPI). Policy Improvement. /Ср/	4	6	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.11	Тема 2.3. Динамическое программирование Метод простой итерации. Policy Evaluation. Value Iteration. Policy Iteration. Generalized Policy Iteration. /Ср/	4	6	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.12	Тема 2.4. Табличные алгоритмы Монте-Карло алгоритм. Экспоненциальное сглаживание. Стохастическая аппроксимация. Temporal Difference. Exploration-exploitation дилемма. Q-learning. SARSA. /Ср/	4	4	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.13	Тема 2.5. Bias-Variance Trade-Off Дилемма смещения-разброса. N-step Temporal Difference. Backward View. Eligibility Trace. TD (lambda). /Пр/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3

1.14	Тема 2.5. Bias-Variance Trade-Off Дилемма смещения-разброса. N-step Temporal Difference. Backward View. Eligibility Trace. TD ( $\lambda$ ). /Ср/	4	6	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
	<b>Раздел 2. Value-based подход. Policy Gradient подход</b>				
2.1	Тема 3.1. Deep Q-learning Q-сетка. Переход к параметрической Q-функции. Декорреляция сэмплов. Таргет-сеть. DQN. /Лек/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.2	Тема 3.1. Deep Q-learning Q-сетка. Переход к параметрической Q-функции. Декорреляция сэмплов. Таргет-сеть. DQN. /Ср/	4	4	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.3	Тема 3.2. Модификации DQN Twin DQN. Clipped Twin DQN. Double DQN. Dueling DQN. Шумные сети (Noisy Nets). Приоритизированный реплей (Prioritized DQN). Multi- step DQN. /Лек/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.4	Тема 3.2. Модификации DQN Twin DQN. Clipped Twin DQN. Double DQN. Dueling DQN. Шумные сети (Noisy Nets). Приоритизированный реплей (Prioritized DQN). Multi- step DQN. /Пр/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.5	Тема 3.2. Модификации DQN Twin DQN. Clipped Twin DQN. Double DQN. Dueling DQN. Шумные сети (Noisy Nets). Приоритизированный реплей (Prioritized DQN). Multi- step DQN. /Ср/	4	6	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.6	Тема 3.3. Distributional RL Идея Distributional подхода. Z-функция. Distributional-форма. Distributional Policy Evaluation. Distributional Value Iteration. Категориальная аппроксимация Z- функций. Categorical DQN. Квантильная аппроксимация Z-функций. Quantile Regression DQN. Implicit Quantile Networks. Rainbow DQN. /Пр/	4	2	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.7	Тема 3.3. Distributional RL Идея Distributional подхода. Z-функция. Distributional-форма. Distributional Policy Evaluation. Distributional Value Iteration. Категориальная аппроксимация Z- функций. Categorical DQN. Квантильная аппроксимация Z-функций. Quantile Regression DQN. Implicit Quantile Networks. Rainbow DQN. /Ср/	4	6		
2.8	Тема 4.1. Policy Gradient Theorem Вывод первым способом. Вывод вторым способом. Физический смысл. REINFORCE. State visitation frequency. Расцепление внешней и внутренней стохастики. Связь с policy improvement. /Ср/	4	6	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.9	Тема 4.2. Схемы «Актёр-критик» Бейзлайн. Введение критика. Bias-variance trade-off. Generalized Advantage Estimation (GAE). Обучение критика. Advantage Actor-Critic (A2C). /Ср/	4	7	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.10	Тема 4.3. Policy Gradient с переиспользованием сэмплов Проблема переиспользования сэмплов. Суррогатная функция. Нижняя оценка. Оптимизация нижней оценки. Trust Region Policy Optimization (TRPO). Proximal Policy Loss. Clipped Value Loss. Proximal Policy Optimization (PRO). /Ср/	4	8	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3

2.11	/Экзамен/	4	9	ПК-2 ПК-3 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
------	-----------	---	---	---------------------	-------------------------

#### 4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущей и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

#### 5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

##### 5.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л1.1	Пол Дейтел, Харви Дейтел	Python: Искусственный интеллект, большие данные и облачные вычисления	Санкт-Петербург: Питер, 2021	<a href="https://ibooks.ru/reading.php?short=1&amp;productid=371701">https://ibooks.ru/reading.php?short=1&amp;productid=371701</a> неограниченный доступ для зарегистрированных пользователей
Л1.2	Павлова, А. И.	Искусственные нейронные сети: учебное пособие	Москва: Ай Пи Ар Медиа, 2021	<a href="https://www.iprbookshop.ru/108228.html">https://www.iprbookshop.ru/108228.html</a> неограниченный доступ для зарегистрированных пользователей

##### 5.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л2.1	Гладков Л. А., Курейчик В. В., Курейчик В. М., Курейчик В. М.	Генетические алгоритмы: учебник	Москва: Физматлит, 2010	<a href="https://biblioclub.ru/index.php?page=book&amp;id=68417">https://biblioclub.ru/index.php?page=book&amp;id=68417</a> неограниченный доступ для зарегистрированных пользователей
Л2.2	Барский, А. Б.	Логические нейронные сети: учебное пособие	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020	<a href="https://www.iprbookshop.ru/97547.html">https://www.iprbookshop.ru/97547.html</a> неограниченный доступ для зарегистрированных пользователей
Л2.3		Студент. Аспирант. Исследователь: всероссийский научный журнал: журнал	Владивосток: Эксперт-Наука, 2021	<a href="https://biblioclub.ru/index.php?page=book&amp;id=685687">https://biblioclub.ru/index.php?page=book&amp;id=685687</a> неограниченный доступ для зарегистрированных пользователей

##### 5.3 Профессиональные базы данных и информационные справочные системы

Консультант +

Официальная документация библиотеки pybrain. - <http://pybrain.org/docs/index.html>

Образовательная платформа по Python - <http://pybrain.org/docs/index.html>

сайт федеральной государственной службы статистики - <https://rosstat.gov.ru/databases>

##### 5.4. Перечень программного обеспечения

Python

##### 5.5. Учебно-методические материалы для студентов с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

#### 6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:

- столы, стулья;

- персональный компьютер / ноутбук (переносной);

- проектор;

- экран / интерактивная доска.

#### **7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)**

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

## ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

### 1 Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

#### 1.1 Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
<b>ПК-2: Способен разрабатывать и анализировать концептуальные и теоретические модели решаемых научных проблем и задач</b>			
З. методы интеллектуального анализа данных, необходимых для построения моделей решаемых научных задач	изучает основную и дополнительную литературу, содержащую материал об основных понятиях инструментальных средствах и математических методах, используемых при решении профессиональных задач, для подготовки к зачету, и устному опросу	полнота и содержательность ответа на зачете, устном опросе, соответствие ответов материалу, содержащемуся в изученной литературе	УО (Раздел 1 в. 1-16 Раздел 2 в. 1-16) Э (1-32)
У. применять методы интеллектуального анализа данных, необходимых для построения моделей решаемых научных задач	решение практико-ориентированных и практических заданий: составление программ на Python с использованием базовых конструкций, условий, циклов, массивов	правильность решения заданий на применение составление программ на Python с использованием базовых конструкций, условий, циклов, массивов	Раздел 1. ПЗ 1-7 Раздел 2. ПЗ 1-5
В. навыками применения методов интеллектуального анализа данных, необходимых для	решение практико-ориентированных и практических заданий: составление программ на	обоснованность применения методов для: решения заданий	Раздел 1. ПЗ 1-7 Раздел 2.



построения моделей решаемых научных задач	Phyton с использованием базовых конструкций, условий, циклов, массивов)	на составление программ на Phyton с использованием базовых конструкций, условий, циклов, массивов)	ПЗ 1-5
<b>ПК-3:Способен разрабатывать и применять математические методы, системное и прикладное программное обеспечение для решения задач научной и проектно-технологической деятельности</b>			
З. функциональность современных инструментальных средств и систем программирования в области создания моделей искусственных нейронных сетей	изучает основную и дополнительную литературу, содержащую материал об основных понятиях инструментальных средствах и математических методах, используемых при решении профессиональных задач, для подготовки к зачету, и устному опросу	полнота и содержательность ответа на зачете, устном опросе, соответствие ответов материалу, содержащемся в изученной литературе	УО (Раздел 1 в. 1-16 Раздел 2 в. 1-16) Э (1-32)
У. проводить оценку и выбор моделей искусственных нейронных сетей и инструментальных средств для решения задач машинного обучения	решение практико-ориентированных и практических заданий: составление программ на Phyton с использованием базовых конструкций, условий, циклов, массивов	правильность решения заданий на составление программ на Phyton с использованием базовых конструкций, условий, циклов, массивов	Раздел 1. ПЗ 1-7 Раздел 2. ПЗ 1-5
В. навыками применения современных инструментальных средств и систем программирования для разработки и обучения моделей искусственных нейронных сетей	решение практико-ориентированных и практических заданий: составление программ на Phyton с использованием базовых конструкций, условий, циклов, массивов)	обоснованность применения методов для: решения заданий на составление программ на Phyton с использованием базовых конструкций, условий, циклов,	Раздел 1. ПЗ 1-7 Раздел 2. ПЗ 1-5

		массивов)	
<b>ПК-4: Способен разрабатывать и анализировать концептуальные и теоретические модели решаемых задач проектной и производственно-технологической деятельности</b>			
3. методы оптимизации построения концептуальных и теоретических моделей решаемых задач проектной и производственно-технологической деятельности	изучает основную и дополнительную литературу, содержащую материал об основных понятиях инструментальных средствах и математических методах, используемых при решении профессиональных задач, для подготовки к зачету, и устному опросу	полнота и содержательность ответа на зачете, устном опросе, соответствие ответов материалу, содержащемуся в изученной литературе	УО (Раздел 1 в. 1-16 Раздел 2 в. 1-16) Э (1-32)
У. применять методы оптимизации построения концептуальных и теоретических моделей решаемых задач проектной и производственно-технологической деятельности	решение практико-ориентированных и практических заданий:  составление программ на Phyton с использованием базовых конструкций, условий, циклов, массивов	правильность решения заданий на  составление программ на Phyton с использованием базовых конструкций, условий, циклов, массивов	Раздел 1. ПЗ 1-7  Раздел 2. ПЗ 1-5
В. навыками применения методов оптимизации построения концептуальных и теоретических моделей решаемых задач проектной и производственно-технологической деятельности	решение практико-ориентированных и практических заданий:  применение различных методов с использованием Loginom, составление программ на Phyton с использованием базовых конструкций, условий, циклов, массивов)	обоснованность применения методов для:  решения заданий с использованием Loginom,  решения заданий на составление программ на Phyton с использованием базовых конструкций, условий, циклов, массивов)	Раздел 1. ПЗ 1-7  Раздел 2. ПЗ 1-5
<b>ПК-6: способен разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов</b>			

3. теоретические основы оценки эффективности и учета рисков инвестиционных проектов, основы современных информационных технологий, применяемых в экономическом анализе	изучает основную и дополнительную литературу, содержащую материал в библиотеках Phyton и математических методах, используемых при решении профессиональных задач, для подготовки к зачету, и устному опросу	полнота и содержательность ответа на зачете, устном опросе, соответствие ответов материалу, содержащемуся в изученной литературе	УО (Раздел 1 в. 1-16 Раздел 2 в. 1-16) Э (1-32)
У. моделировать и анализировать бизнес-процессы предприятия, находить организационно управленческие решения в моделях управления	решение практико-ориентированных и практических заданий: применение составление программ на Phyton с использованием библиотек	правильность решения заданий на составление программ на Phyton с использованием библиотек	Раздел 1. ПЗ 1-7 Раздел 2. ПЗ 1-5
В. навыками планирования научно-исследовательской деятельности, анализа инвестиционных рисков, управления командой проекта; навыками работы в коллективе	решение практико-ориентированных и практических заданий: составление программ на Phyton с использованием библиотек (NumPy, Pandas, matplotlib, PyBrian)	обоснованность применения методов для: решения заданий на составление программ на Phyton с использованием библиотек (NumPy, Pandas, matplotlib, PyBrian)	Раздел 1. ПЗ 1-7 Раздел 2. ПЗ 1-5

*Э – вопросы к экзамену, ПЗ-практическое задание, УО- устный опрос*

### 1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале:

84-100 баллов (оценка «отлично»)

67-83 баллов (оценка «хорошо»)

50-66 баллов (оценка «удовлетворительно»)

0-49 баллов (оценка «неудовлетворительно»)

## 2 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

В разделе приводятся типовые варианты оценочных средств: вопросы к экзамену, практические задания, вопросы для устного опроса

### Вопросы к экзамену

1. Связь с оптимальным уравнением. Марковская цепь. Среда. Действия. Траектории. Марковский процесс принятия решений (MDP). Эпизодичность. Дисконтирование.
2. Условия задачи RL. On-policy. Концепция model-free алгоритмов. Классификация RL-алгоритмов.
3. Критерии оценки RL-алгоритмов. Сложности задачи RL. Бенчмарки.
4. Задача безградиентной оптимизации. Случайный поиск. Hill Climbing. Имитация отжига.
5. Эволюционные алгоритмы. Weight Agnostic Neural Networks (WANN). Видовая спецификация. Генетические алгоритмы.
6. Идея эволюционных стратегий. Оценка вероятности редкого события. Метод Кросс-Энтропии для стохастической оптимизации.
7. Метод Кросс-Энтропии для обучения с подкреплением (CEM). Натуральные эволюционные стратегии (NES). OpenAI-ES. Адаптация матрицы ковариации (CMA-ES).
8. Свойства траекторий. V-функция. Уравнение Беллмана.
9. Оптимальная стратегия. Q-функция. Принцип оптимальности Беллмана.
10. Отказ от однородности. Вид оптимальной стратегии (доказательство через отказ от однородности).
11. Уравнение оптимальности Беллмана. Критерий оптимальности Беллмана.
12. Advantage-функция. Relative Performance Identity (RPI). Policy Improvement.
13. Метод простой итерации. Policy Evaluation. Value Iteration. Policy Iteration. Generalized Policy Iteration.
14. Монте-Карло алгоритм. Экспоненциальное сглаживание. Стохастическая аппроксимация. Temporal Difference. Exploration-exploitation дилемма. Q-learning. SARSA.
15. Дилемма смещения-разброса. N-step Temporal Difference.
16. Backward View. Eligibility Trace. TD( $\lambda$ ).
17. Q-сетка. Переход к параметрической Q-функции. Декорреляция сэмплов. Таргет-сеть. DQN.
18. Twin DQN. Clipped Twin DQN. Double DQN. Dueling DQN.
19. Шумные сети (Noisy Nets). Приоритизированный реплей (Prioritized DQN). Multi-step DQN.
20. Идея Distributional подхода. Z-функция. Distributional-форма. Distributional Policy Evaluation. Distributional Value Iteration.
21. Категориальная аппроксимация Z-функций. Categorical DQN.
22. Квантильная аппроксимация Z-функций. Quantile Regression DQN.
23. Implicit Quantile Networks. Rainbow DQN.
24. Вывод первым способом. Вывод вторым способом. Физический смысл.
25. REINFORCE. State visitation frequency.
26. Расцепление внешней и внутренней стохастики. Связь с policy improvement.
27. Бейзлайн. Введение критика. Bias-variance trade-off.
28. Generalized Advantage Estimation (GAE).
29. Обучение критика. Advantage Actor-Critic (A2C).
30. Проблема переиспользования сэмплов. Суррогатная функция.

31. Нижняя оценка. Оптимизация нижней оценки. Trust Region Policy Optimization (TRPO). Proximal Policy Loss.
32. Clipped Value Loss. Proximal Policy Optimization (PRO).

#### **Критерии оценивания:**

Основой для определения баллов, набранных при промежуточной аттестации, служит объём и уровень усвоения материала, предусмотренного рабочей программой дисциплины. При этом необходимо руководствоваться следующим:

- 84-100 баллов (оценка **«отлично»**) - изложенный материал фактически верен, наличие глубоких исчерпывающих знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; правильные, уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе, усвоение основной и знакомство с дополнительной литературой;
- 67-83 баллов (оценка **«хорошо»**) - наличие твердых и достаточно полных знаний в объеме пройденной программы дисциплины в соответствии с целями обучения, правильные действия по применению знаний на практике, четкое изложение материала, допускаются отдельные логические и стилистические погрешности, обучающийся усвоил основную литературу, рекомендованную в рабочей программе дисциплины;
- 50-66 баллов (оценка **«удовлетворительно»**) - наличие твердых знаний в объеме пройденного курса в соответствии с целями обучения, изложение ответов с отдельными ошибками, уверенно исправленными после дополнительных вопросов; правильные в целом действия по применению знаний на практике;
- 0-49 баллов (оценка **«неудовлетворительно»**) - ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и «наводящие» вопросы.

### **Практические задания**

#### **Раздел 1. Мета эвристики. Классическая теория**

##### Тема 1.2. Алгоритмы обучения с подкреплением

Условия задачи RL. On-policy. Концепция model-free алгоритмов. Классификация RL-алгоритмов. Критерии оценки RL-алгоритмов. Сложности задачи RL. Бенчмарки

##### Тема 1.3. Бэйзлайны

Задача безградиентной оптимизации. Случайный поиск. Hill Climbing. Имитация отжига. Эволюционные алгоритмы. Weight Agnostic Neural Networks (WANN). Видовая спецификация. Генетические алгоритмы.

#### Тема 1.4. Эволюционные стратегии

Идея эволюционных стратегий. Оценка вероятности редкого события. Метод Кросс-Энтропии для стохастической оптимизации. Метод Кросс-Энтропии для обучения с подкреплением (CEM). Натуральные эволюционные стратегии (NES). OpenAI-ES. Адаптация матрицы ковариации (CMA-ES)

#### Тема 2.2. Relative Performance Identity

Advantage-функция. Relative Performance Identity (RPI). Policy Improvement.

#### Тема 2.3. Динамическое программирование

Метод простой итерации. Policy Evaluation. Value Iteration. Policy Iteration. Generalized Policy Iteration.

#### Тема 2.4. Табличные алгоритмы

Монте-Карло алгоритм. Экспоненциальное сглаживание. Стохастическая аппроксимация. Temporal Difference. Exploration-exploitation дилемма. Q-learning. SARSA.

#### Тема 2.5. Bias-Variance Trade-Off

Дилемма смещения-разброса. N-step Temporal Difference. Backward View. Eligibility Trace. TD ( $\lambda$ ).

Каждое задание оценивается в 7 баллов

Критерии оценивания:

7 б. – задание выполнено верно;

6 б. – при выполнении задания были допущены неточности, не влияющие на результат;

4-5 б. – при выполнении задания были допущены ошибки;

3-1 б. – при выполнении задания были допущены существенные ошибки.

0 б. – задание не выполнено.

Максимальное количество баллов по практическим заданиям раздела 1 – 49 б.

### **Раздел 2. Value-based подход. Policy Gradient подход**

### Тема 3.1. Deep Q-learning

Q-сетка. Переход к параметрической Q-функции. Декорреляция сэмплов. Таргет-сеть. DQN.

### Тема 3.2. Модификации DQN

Twin DQN. Clipped Twin DQN. Double DQN. Dueling DQN. Шумные сети (Noisy Nets). Приоритизированный реплей (Prioritized DQN). Multi- step DQN.

### Тема 3.3. Distributional RL

Идея Distributional подхода. Z-функция. Distributional- форма. Distributional Policy Evaluation. Distributional Value Iteration. Категориальная аппроксимация Z- функций. Categorical DQN. Квантильная аппроксимация Z-функций. Quantile Regression DQN. Implicit Quantile Networks. Rainbow DQN.

### Тема 4.2. Схемы «Актёр-критик»

Бейзлайн. Введение критика. Bias-variance trade-off. Generalized Advantage Estimation (GAE). Обучение критика. Advantage Actor-Critic (A2C).

### Тема 4.3. Policy Gradient с переиспользованием сэмплов

Проблема переиспользования сэмплов. Суррогатная функция. Нижняя оценка. Оптимизация нижней оценки. Trust Region Policy Optimization (TRPO). Proximal Policy Loss. Clipped Value Loss. Proximal Policy Optimization (PRO).

Каждое задание оценивается в 7 баллов

Критерии оценивания:

7 б. – задание выполнено верно;

6 б. – при выполнении задания были допущены неточности, не влияющие на результат;

4-5 б. – при выполнении задания были допущены ошибки;

3-1 б. – при выполнении задания были допущены существенные ошибки.

0 б. – задание не выполнено.

Максимальное количество баллов по практическим заданиям раздела 2 – 35 б.

## Перечень вопросов для устного опроса

### Раздел 1. Мета эвристики. Классическая теория

1. Связь с оптимальным уравнением. Марковская цепь. Среда. Действия. Траектории. Марковский процесс принятия решений (MDP). Эпизодичность. Дисконтирование.
2. Условия задачи RL. On-policy. Концепция model-free алгоритмов. Классификация RL-алгоритмов.
3. Критерии оценки RL-алгоритмов. Сложности задачи RL. Бенчмарки.
4. Задача безградиентной оптимизации. Случайный поиск. Hill Climbing. Имитация отжига.
5. Эволюционные алгоритмы. Weight Agnostic Neural Networks (WANN). Видовая спецификация. Генетические алгоритмы.
6. Идея эволюционных стратегий. Оценка вероятности редкого события. Метод Кросс-Энтропии для стохастической оптимизации.
7. Метод Кросс-Энтропии для обучения с подкреплением (CEM). Натуральные эволюционные стратегии (NES). OpenAI-ES. Адаптация матрицы ковариации (CMA-ES).
8. Свойства траекторий. V-функция. Уравнение Беллмана.
9. Оптимальная стратегия. Q-функция. Принцип оптимальности Беллмана.
10. Отказ от однородности. Вид оптимальной стратегии (доказательство через отказ от однородности).
11. Уравнение оптимальности Беллмана. Критерий оптимальности Беллмана.
12. Advantage-функция. Relative Performance Identity (RPI). Policy Improvement.
13. Метод простой итерации. Policy Evaluation. Value Iteration. Policy Iteration. Generalized Policy Iteration.
14. Монте-Карло алгоритм. Экспоненциальное сглаживание. Стохастическая аппроксимация. Temporal Difference. Exploration-exploitation дилемма. Q-learning. SARSA.
15. Дилемма смещения-разброса. N-step Temporal Difference.
16. Backward View. Eligibility Trace. TD( $\lambda$ ).

Максимальное количество баллов по разделу 1 – 8 б

### Раздел 2. Value-based подход. Policy Gradient подход

1. Q-сетка. Переход к параметрической Q-функции. Декорреляция сэмплов. Таргет-сеть. DQN.
2. Twin DQN. Clipped Twin DQN. Double DQN. Dueling DQN.
3. Шумные сети (Noisy Nets). Приоритизированный реплей (Prioritized DQN). Multi-step DQN.
4. Идея Distributional подхода. Z-функция. Distributional-форма. Distributional Policy Evaluation. Distributional Value Iteration.
5. Категориальная аппроксимация Z-функций. Categorical DQN.
6. Квантильная аппроксимация Z-функций. Quantile Regression DQN.
7. Implicit Quantile Networks. Rainbow DQN.
8. Вывод первым способом. Вывод вторым способом. Физический смысл.
9. REINFORCE. State visitation frequency.
10. Расцепление внешней и внутренней стохастики. Связь с policy improvement.
11. Бейзлайн. Введение критика. Bias-variance trade-off.
12. Generalized Advantage Estimation (GAE).
13. Обучение критика. Advantage Actor-Critic (A2C).
14. Проблема переиспользования сэмплов. Суррогатная функция.



15. Нижняя оценка. Оптимизация нижней оценки. Trust Region Policy Optimization (TRPO). Proximal Policy Loss.
16. Clipped Value Loss. Proximal Policy Optimization (PRO).

Максимальное количество баллов по разделу 2 – 8 б.

**Критерии оценивания:**

Для каждого вопроса:

- 2 балла - дан полный, развёрнутый ответ на поставленный вопрос, изложение материала при ответе – грамотное и логически стройное;
- 1 балл - дан неполный ответ на поставленный вопрос
- 0 баллов - обучающийся не владеет материалом по заданному вопросу.

*Максимальное количество баллов по устному опросу – 30*

**3 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций**

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

**Текущий контроль** успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

**Промежуточная аттестация** проводится в форме экзамена 3 семестр

Экзамен проводится по окончании теоретического обучения до начала экзаменационной сессии в письменном виде. Количество вопросов в задании – 2. Проверка ответов и объявление результатов производится в день экзамена. Результаты аттестации заносятся в зачетную ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать задолженность в установленном порядке.

## МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

- лекции;
- практические занятия.

В ходе практических работ развиваются навыки применения математических методов, выбора инструментальных средств для обработки и анализа экономических данных в профессиональной деятельности

При подготовке к практическим занятиям каждый студент должен:

- изучить рекомендованную учебную литературу;

В процессе подготовки к практическим занятиям студенты могут воспользоваться консультациями преподавателя.

Теоретические вопросы должны быть изучены студентами в ходе самостоятельной работы. Контроль самостоятельной работы студентов над учебной программой курса осуществляется методом устного опроса и выполнения практических заданий. В ходе самостоятельной работы каждый студент обязан прочитать основную и по возможности дополнительную литературу по изучаемой теме. Выделить непонятные термины, найти их значение в литературе.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами. Также обучающиеся могут взять на дом необходимую литературу на абонементе университетской библиотеки или воспользоваться читальными залами.