**Программа учебной дисциплины \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Утверждена

Академическим советом ООП

Протокол № от «\_\_»\_\_\_\_\_20\_\_ г.

|  |  |
| --- | --- |
| Автор | Ратников Фёдор Дмитриевич, к. ф.-м. н., [fratnikov@hse.ru](mailto:fratnikov@hse.ru) |
| Число кредитов | 4 |
| Контактная работа (час.) | 112 |
| Самостоятельная работа (час.) | 96 |
| Курс | Обучение с подкреплением |
| Формат изучения дисциплины | без использования онлайн курса |

1. **ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ**

Основная цель освоения дисциплины «Обучение с подкреплением» – научиться использовать методы одноимённой области машинного обучения в практических и исследовательских задачах. Что не менее важно — научиться понимать границы применимости методов обучения с подкреплением и ситуации, когда лучше держаться от них подальше.Более детальную информацию интересующемуся студенту предлагается узнать со <https://bit.ly/2HjLPnn>.

Изучение данной дисциплины базируется на следующих пререквизитах:

* Машинное обучение 1
* Теория вероятностей

Также есть научные области, знание которых не является необходимым, однако могут упростить освоение курса:

* Глубинное обучение
* Непрерывная оптимизация
* Основы теории игр
* Байесовские методы в машинном обучении
* Параллельные и распределённые вычисления

Основные положения дисциплины могут быть использованы в дальнейшем при изучении дисциплин:

* Машинное обучение 2
* Методы искусственного интеллекта в робототехнических системах
* Байесовские методы в глубинном обучении
* Автоматическая обработка текстов

В результате освоения данной дисциплины студент будет

**Знать:**

* Постановки задач обучения с подкреплением, их отличие от задач в других областях ML.
* Формализмы MAB, MDP, POMDP
* Основные value-based и policy-based алгоритмы обучения с подкреплением
* Отличия между on-policy и off-policy обучением

**Уметь:**

* Сводить прикладные задачи к формализму обучения с подкреплением
* Реализовывать простейшие методы обучения с подкреплением для MDP
* Масштабировать алгоритмы обучения с подкреплением с использованием моделей машинного обучения, в т.ч. линейных моделей и нейросетей

**Владеть:**

* Формализмом уравнений Беллмана, фреймворком value-based методов
* Навыками дифференцирования матожидания недифференцируемых функций по параметрам распределния аргумента.

# Содержание УЧЕБНОЙ ДИСЦИПЛИНЫ

*Тема 1. Reinforcement learning problems; MDP; Crossentropy method*

*Тема 2. Value-based RL; Bellman equations; Value Iteraion; Policy Iteration*

*Тема 3. Model-free RL; Qlearning and SARSA; On- and Off-policy learning*

*Тема 4. Approximate RL; Deep Q-learning; training stability & tricks;*

*Тема 5. Exploration; Exploration in contextual bandits; Exporation & uncertainty*

*Тема 6. Policy-based methods; REINFORCE; Actor-critic; Value- vs Policy-based RL*

*Тема 7. Applications I: Reinforcement learning for Natural Language Processing*

*Тема 8. Applications II: Deep Architectures, Nearest Neighbor Search, etc.*

*Тема 9. RL in Partially Observable Markov Decision Processes;*

*Тема 10. Advanced policy-based methods: TRPO, PPO, DPG; Constrained RL*

# ОЦЕНИВАНИЕ

Оценка за курс зависит только от баллов, накопленных за домашние задания. Экзамен по данной дисциплине не предусмотрен.

Результирующая оценка по дисциплине рассчитывается по формуле

O\_итог=O\_накопл

Накопленная и итоговая оценки округляются арифметически.

Накопленная оценка рассчитывается по формуле

O\_накопл=O\_дз

# ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Практические задания выполняются студентами в виде jupyter notebook. В курсе используется стандартный набор библиотек машинного обучения (numpy, scipy, sklearn, pandas, matplotlib), вычислительные графы в pytorch / tensorflow на выбор студента. Стандартные среды для отладки алгоритмов взяты из openai gym. Проверка домашних заданий осуществляется в системе anytask.

Текущий контроль требует от студентов решать задания, в основном оформленные в виде jupyter-notebooks (примеры: http://bit.ly/2B9dpkT , http://bit.ly/2BtrTJg , http://bit.ly/2kGMlyk ).

1. **РЕСУРСЫ**
   1. **Основная литература**

* Bishop C.M. Pattern Recognition and Machine Learning, 2006. (<https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>)
  1. **Программное обеспечение**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование** | **Условия доступа** |
| 1. | Anaconda | *ПО с открытым исходным кодом* |
| 2. | OpenAI Gym | *ПО с открытым исходным кодом* |
| 3. | Docker | *ПО с открытым исходным кодом* |

* 1. **Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование** | **Условия доступа** |
|  | ***Интернет-ресурсы (электронные образовательные ресурсы)*** | |
| 1. | Открытое образование | URL: https://openedu.ru/ |
| 2. | Coursera | URL: https://www.coursera.org |

* 1. **Материально-техническое обеспечение дисциплины**

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);

мультимедийный проектор с дистанционным управлением.

Учебные аудитории для лабораторных и самостоятельных занятий по дисциплине оснащены ­­­­­­­­­­­­­­­­­­­­­­­­ ПЭВМ (операционная система, офисные программы), с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.