



FIȘA UNITĂȚII DE CURS/MODULULUI

1. Date despre program

|  |   |
|--|---|
| 1.1. Instituția de învățământ superior | Universitatea de Stat din Moldova                     |
| 1.2. Facultatea                        | Matematică și Informatică                             |
| 1.3. Departamentul                     | Informatică   |
| 1.4. Domeniul de formare profesională  | 0613 Dezvoltarea produselor program și a aplicațiilor |
| 1.5. Ciclul de studii                  | Studii superioare de master, ciclul II                |
| 1.6. Programul de studii               | Inteligența Artificială și Știința Datelor            |

2. Date despre unitatea de curs/modul

|   |                          |                |   |                        |   |                          |    |
|---|--------------------------|----------------|---|------------------------|---|--------------------------|----|
| 2.1. Denumirea disciplinei                | Învățare prin recompensă |                |   |                        |   |                          |    |
| 2.2. Titularul activităților de curs      | ?                        |                |   |                        |   |                          |    |
| 2.3. Titularul activităților de laborator | ?                        |                |   |                        |   |                          |    |
| 2.4. Anul de studii                       | 2                        | 2.5. Semestrul | 1 | 2.6. Tipul de evaluare | E | 2.7. Categoria formativă | SA |

3. Timpul total estimat (ore pe semestru al activităților didactice)

|  |     |                    |    |                |            |
|--|-----|--------------------|----|----------------|------------|
| 3.1. Număr de ore pe săptămână   | 4   | din care: 3.2 curs | 2  | 3.3. laborator | 2          |
| 3.4. Total ore în planul de învățământ   | 40  | din care: 3.5 curs | 20 | 3.6. laborator | 20         |
| <b>Distribuția fondului de timp pentru studiul individual:</b>                     |     |                    |    |                | <b>ore</b> |
| Studiu după manual, suport de curs, bibliografie și notițe                         |     |                    |    |                | 40         |
| Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate |     |                    |    |                | 35         |
| Pregătire laboratoare, teme, referate, portofolii                                  |     |                    |    |                | 35         |
| 3.7. Total ore studiu individual   | 110 |                    |    |                |            |
| 3.8. Total ore pe semestru   | 150 |                    |    |                |            |
| 3.9. Număr de credite  | 5   |                    |    |                |            |

4. Precondiții de acces la unitatea de curs/modul

|                    |   |
|--------------------|---|
| 4.1. de curriculum | Baze de date, programare(Python), statistică pentru știința datelor, analiza și vizualizarea datelor, învățarea automată, depozite de date și data mining, învățare profundă. |
| 4.2. de competențe | Cunoștințe de programare, statistică, învățare automată și învățare profundă.   |

5. Condiții de desfășurare

|                      |  |
|----------------------|--|
| 5.1. a cursului      | Sală de curs cu tablă mare și proiector, laptop bibliotecile Python instalate, |
| 5.2. a laboratorului | Sală de laborator, calculatoare cu bibliotecile Python instalate.              |

6. Obiectivele disciplinei – rezultate așteptate ale învățării la formarea cărora contribuie parcurgerea și promovarea disciplinei

|            |   |
|------------|---|
| Cunoștințe | (C1) Dobândirea unor cunoștințe fundamentale despre conceptele cheie ale RL, cum ar fi agentul, mediul, starea, acțiunea și recompensa.<br>(C2) Dobândirea unor cunoștințe aprofundate a MDPs, inclusiv structura acestora și relația cu RL.<br>(C3) Dobândirea unor cunoștințe aprofundate ale algoritmilor principali folosiți în RL, cum ar fi SARSA, Q-learning, Metodele Monte Carlo și Învățarea prin Diferență |
|------------|---|



|                               |  |
|-------------------------------|--|
|                               | <p>Temporală (TD).</p> <p>(C4) Dobândirea unor cunoștințe aprofundate ale ecuațiilor Bellman și a utilizării programării dinamice în evaluarea și îmbunătățirea politicilor.</p> <p>(C5) Dobândirea unor cunoștințe despre rețele Q profunde (DQN), metode de gradient al politicii și învățarea prin întărire ierarhică.</p> <p>(C6) Dobândirea unor cunoștințe ale aplicabilității RL în diverse domenii, cum ar fi robotică, inteligență artificială și sisteme autonome.</p>   |
| Abilități                     | <p>(A1) Abilitatea de a formula probleme reale ca probleme de învățare prin recompensă și de a identifica corect elementele esențiale (stări, acțiuni, recompense).</p> <p>(A2) Abilitatea de dezvoltare și implementare a algoritmilor RL, inclusiv SARSA, Q-learning și DQN, folosind limbaje de programare precum Python.</p> <p>(A3) Abilitatea de a evalua și optimiza politici în cadrul MDPs folosind metodele și algoritmi învățați.</p> <p>(A4) Abilitatea în dezvoltarea de strategii eficiente pentru explorarea mediului și exploatarea cunoștințelor dobândite de agent.</p> <p>(A5) Abilitatea în dezvoltarea și aplicarea conceptelor RL în sisteme multi-agent, în medii competitive și cooperative.</p> <p>(A6) Abilitatea de a integra rețele neuronale în procesele RL, inclusiv în construirea de rețele Q profunde.</p> <p>(A7) Abilitatea de a analiza datele rezultate din procesele RL și de a interpreta eficiența algoritmilor folosiți.</p>   |
| Responsabilitate și autonomie | <p>(R1) Dezvoltarea capacității de a analiza și evalua critic algoritmi și strategiile RL utilizate în diferite contexte.</p> <p>(R2) Capacitatea de a învăța și de a se adapta continuu la noile metode și tehnologii emergente în RL.</p> <p>(R3) (R2) Capacitatea de a învăța și de a se adapta continuu la noile metode și tehnologii emergente în RL.</p> <p>de a iniția și dezvolta proiecte complexe de RL în mod independent, de la definirea problemei până la implementarea soluției.</p> <p>(R4) Capacitatea de a colabora eficient în echipe multidisciplinare, contribuind la rezolvarea problemelor complexe prin utilizarea RL.</p> <p>(R5) Dezvoltarea capacității de a aplica RL în scenarii reale, în care deciziile trebuie luate rapid și eficient.</p> <p>(R6) Înțelegerea impactului deciziilor bazate pe RL asupra mediului înconjurător și asupra societății, și asumarea responsabilității pentru soluțiile dezvoltate.</p> <p>(R7) Stimularea creativității și inovării prin utilizarea RL pentru a dezvolta soluții noi și eficiente la problemele emergente.</p> |



| 7.1. Curs   | Numărul de ore |
|---|----------------|
| <i>C1. Introducere în Învățarea prin Recompensă.</i> Prezentarea generală a învățării prin întărire (RL). Diferențele dintre RL și alte paradigme de învățare automată. Concepte cheie: agent, mediu, stare, acțiune, recompensă.   | 1              |
| <i>C2. Procese Decizionale Markov (MDPs).</i> Fundamentele MDPs. Probabilități de tranziție și funcții de recompensă. Politici și funcții de valoare.   | 2              |
| <i>C3. Programarea Dinamică în RL.</i> Evaluarea și îmbunătățirea politicii. Ecuațiile Bellman. Algoritmi iterativi pentru rezolvarea MDPs.   | 1,5            |
| <i>C4. Metode Monte Carlo pentru RL.</i> Introducere în metodele Monte Carlo în RL. Estimarea funcțiilor de valoare și a politicilor. Strategii de explorare vs. Exploatare.  | 1,5            |
| <i>C5. Învățare prin Diferență Temporală (TD).</i> Bazele învățării TD. Algoritmi TD: SARSA și Q-learning. Avantajele învățării TD față de metodele Monte Carlo.  | 1,5            |
| <i>C6. Aproximarea Funcției în RL.</i> Trecerea de la metode tabulare la aproximarea funcției. Aproximarea funcției liniare. Introducere în învățarea profundă pentru aproximarea funcției.   | 1,5            |
| <i>C7. Rețele Q Profunde (DQN).</i> Combinarea Q-learning cu rețele neuronale profunde. Replay de experiență și ținte Q fixe. Implementarea unui DQN de bază.   | 2,5            |
| <i>C8. Metode de Gradient al Politicii.</i> Introducere în metodele de gradient al politicii. Algoritmul REINFORCE. Metode Actor-Critic.  | 1,5            |
| <i>C9. Variante Avansate ale DQN.</i> Îmbunătățiri și variații ale DQN: Double DQN, Dueling DQN. Replay de experiență prioritizat. Învățare multi-pas.  | 1,5            |
| <i>C10. Optimizarea Politicii cu Regiune de Încredere (TRPO) și Optimizarea Politicii Proximale (PPO).</i> Prezentarea TRPO și PPO. Optimizarea politicii cu regiuni de încredere. Implementarea PPO.   | 1,5            |
| <i>C11. Învățarea prin Întărire Multi-Agent.</i> Concepte în RL multi-agent. Medii cooperative și competitive. Învățarea în sisteme multi-agent.  | 1              |
| <i>C12. Învățarea prin Întărire Ierarhică.</i> Introducere în RL ierarhic. Cadru de opțiuni. Aplicații și beneficii ale RL ierarhic.  | 1              |
| <i>C13. Învățarea prin Întărire Inversă.</i> Bazele RL invers. Învățarea din demonstrații. Aplicații în robotică și sisteme autonome.   | 1              |
| <i>C14. Tehnici de Explorare în RL.</i> Provocarea explorării în RL. Strategii de explorare: $\epsilon$ -greedy, UCB, eșantionare Thompson. Motivație intrinsecă.   | 1              |
| <i>Total prelegeri:</i>   | 20             |
| <b>Bibliografie:</b> <ol style="list-style-type: none"><li>1. Richard S. Sutton, Andrew G. Barto, Reinforcement Learning: An Introduction Second edition, in progress, The MIT Press Cambridge, Massachusetts London, England.</li><li>2. Sutton, R. S., &amp; Barto, A. G. (2018). <i>Reinforcement Learning: An Introduction</i> (2nd ed.). MIT Press.</li><li>3. Szepesvári, C. (2010). Algorithms for Reinforcement Learning. Morgan &amp; Claypool Publishers.</li><li>4. Silver, D. (2015). UCL Course on Reinforcement Learning [Lecture Notes]. University College London. Retrieved from <a href="https://www.davidsilver.uk/teaching/">https://www.davidsilver.uk/teaching/</a></li><li>5. Haarnoja, T., Zhou, A., Abbeel, P., &amp; Levine, S. (2018). Soft Actor-Critic: Off-Policy Maximum</li></ol> |                |



Entropy Deep Reinforcement Learning with a Stochastic Actor. In Proceedings of the 35th International Conference on Machine Learning (ICML). Retrieved from

<https://arxiv.org/abs/1801.01290>

6. Aske Plaat (2022). Deep Reinforcement Learning. LIACS, Leiden University, Leiden, The Netherlands
7. Warren B. Powell (2022). Reinforcement Learning and Stochastic Optimization. John Wiley & Sons
8. Shengbo Eben Li (2023). Reinforcement Learning for Sequential Decision and Optimal Control. Springer
9. Miguel Morales (2020). Grokking Deep Reinforcement Learning. Manning

### 7.2. Laborator

### Numărul de ore

|   |    |
|---|----|
| L1. Proiectarea unui sistem de recompensă pentru un joc simplu sau o simulație.             | 2  |
| L2. Configurarea unui mediu de învățare prin întărire folosind OpenAI Gym.                  | 3  |
| L3. Construirea unui Agent Q-Learning de Bază.  | 3  |
| L4. Implementarea unei Rețele Q Profunde (DQN).   | 4  |
| L5. Experimentarea cu diferite strategii de explorare: $\epsilon$ -greedy, Softmax, UCB.    | 2  |
| L6. Simularea unui mediu multi-agent și implementarea agenților cooperativi sau competitivi | 3  |
| L7. Proiectarea unei sarcini de învățare prin întărire ierarhică                            | 3  |
| <i>Total laborator:</i>   | 20 |

### Bibliografie:

1. Richard S. Sutton, Andrew G. Barto, Reinforcement Learning: An Introduction Second edition, in progress, The MIT Press Cambridge, Massachusetts London, England.
2. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.
3. Szepesvári, C. (2010). Algorithms for Reinforcement Learning. Morgan & Claypool Publishers.
4. Silver, D. (2015). UCL Course on Reinforcement Learning [Lecture Notes]. University College London. Retrieved from <https://www.davidsilver.uk/teaching/>
5. Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In Proceedings of the 35th International Conference on Machine Learning (ICML). Retrieved from <https://arxiv.org/abs/1801.01290>
6. Sudharsan Ravichandiran (2020). Deep Reinforcement Learning with Python. Packt Publishing
7. Graesser Laura, Keng Wah Loon (2020). Foundations of deep reinforcement learning : theory and practice in Python. Addison-Wesley
8. Andrea Lonza (2019). Reinforcement Learning Algorithms with Python: Learn, understand, and develop smart algorithms for addressing AI challenges. Packt Publishing
9. Sean Saito, Yang Wenzhuo, Rajalingappaa Shanmugamani (2018). Python Reinforcement Learning Projects. Packt Publishing
10. Maxim Lapan (2020). Deep Reinforcement Learning Hands-On. Packt Publishing
11. Mohit Sewak (2019). Deep Reinforcement Learning. Frontiers of Artificial Intelligence. Springer

### 8. Evaluare

| Nota semestrială |             |                  |                   |
|------------------|-------------|------------------|-------------------|
| Atestarea 1      | Atestarea 2 | Evaluare curentă | Lucrul individual |
| 25%              | 25%         | 25%              | 25%               |

Nota finală



# Universitatea de Stat din Moldova



|                  |                |
|------------------|----------------|
| Nota semestrială | Nota la examen |
| 60%              | 40%            |

## Standard minim de performanță

Standard minim (cunoștințe și aptitudini necesare pentru nota 5)

- prezența la cursuri și seminarii conform cerințelor generale ale facultății
- cunoașterea conceptelor de bază privind învățarea prin recompensă
- abilitatea de a proiecta un sistem de recompensă
- capacitatea de a implementa o rețea Q Profundă

Evaluarea în cadrul disciplinei „Învățare prin recompensă” se realizează formativ - activitatea la orele de laborator, verificarea orală sau scrisă curentă, două lucrări de atestare (la mijloc și la sfârșit de semestru), și sumativ (final) – 4 lucrări independente, realizate pe parcursul semestrului și examenul final la disciplină. Nota generală la disciplină însumează nota de la examen și cea semestrială, în proporție de 40 și, respectiv, 60 la sută, și apreciază gradul de corespundere cu finalitățile scontate: cunoștințele și competențele acumulate, abilitatea de a aplica cunoștințele, gradul de integrare a cunoștințelor de către studenți etc.

Data completării

25.08.2024

Titular de disciplină

Dr., conf. univ. ?

Data avizării în departament

Șef departament