

MARKOV DÖNTÉSI FOLYAMATOK ÉS MEGERŐSÍTÉSES TANULÁS

záróvizsga 2025/26

1. **Markov döntési problémák (MDP):** Markov láncok definíciója és alapfogalmai; dinamikus rendszer reprezentáció; diszkrét- és Borel MDP-k; irányítási politikák; Markov politikák elégségessége; értékelőfüggvények; irányítási politikák parciális rendezése; optimális értékelőfüggvény.
2. **Véges horizontú problémák:** specifikáció diszkrét- és Borel MDP-k esetére; értékelőfüggvények; politika kiértékelés; optimalitási elv; dinamikus programozás; optimális politikák; demonstratív példák: eszköz eladás, készletgazdálkodás, lineáris-kvadratikus regulátor, Riccati egyenletek.
3. **Sztochasztikus legrövidebb út problémák:** a feladat specifikációja; értékelőfüggvények; Bellman operátorok; megfelelő (proper) irányításai politikák; fő feltevések; optimális politikák tulajdonságai; fő megoldások: érték iteráció és változatai, politikai iteráció és változatai, lineáris programozás.
4. **Diszkontált- és átlagos költségű problémák:** feladat specifikációk; értékelőfüggvények; Bellman operátorok; ekvivalens sztochasztikus legrövidebb út problémák; véges diszkontált és átlagos MDP-k főbb tulajdonságai; fő megoldási módszerek: érték- és politika iteráció, lineáris programozás.
5. **Időbeli különbség („temporal difference”) tanulás:** Monte Carlo politika kiértékelés; TD(0), T(1), és TD(λ); online vs offline; first- vs every visit; TD erős konzisztenciája; TD diszkontált problémák esetén; akció-értékelő függvények; Q-tanulás; SARSA; felfedezés vs kihasználás, puha politikák.
6. **Bandita algoritmusok:** sztochasztikus bandita problémák; megbánás; szub-Gauss eloszlások; koncentrációs egyenlőtlenségek; „Explore-Then-Commit” (ETC) algoritmus; „Upper Confidence Bounds” (UCB) algoritmus és tulajdonságai; kontextuális és lineáris banditák; lineáris UCB.
7. **Sztochasztikus approximáció:** standard problémák; általános feltételek a lépésközökre; Robbins-Monro és Kiefer-Wolfowitz algoritmusok; sztochasztikus gradiens; SPSA; LMS; momentum gyorsítás és Polyak átlagolás; erős konzisztencia garantálásának módszerei, pl., Ljapunov függvények.

Irodalom:

- Bäuerle, N. & Rieder, U.: Markov Decision Processes with Applications to Finance, Springer, 2011
- Bertsekas, D. P.: Dynamic Programming and Optimal Control, 4th ed., Athena Scientific, 2017
- Feinberg, A. E. & Shwartz, A. (eds.): Handbook of Markov Decision Processes, Kluwer, 2002
- Bertsekas, D. P., & Tsitsiklis, J. N.: Neuro-Dynamic Programming, Athena Scientific, 1996
- Lattimore, T., & Szepesvári, Cs.: Bandit Algorithms, Cambridge University Press, 2018