

ΙΔΡΥΜΑ	ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ																			
ΣΧΟΛΗ	ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ																			
ΤΜΗΜΑ	ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ																			
ΕΠΙΠΕΔΟ ΣΠΟΥΔΩΝ	ΠΡΟΠΤΥΧΙΑΚΟ και ΜΕΤΑΠΤΥΧΙΑΚΟ (ΣΥΝΔΙΔΑΣΚΑΛΙΑ)																			
ΤΙΤΛΟΣ ΜΑΘΗΜΑΤΟΣ	ΕΝΙΣΧΥΤΙΚΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΚΑΙ ΣΤΟΧΑΣΤΙΚΑ ΠΑΙΓΝΙΑ																			
ΚΩΔΙΚΟΣ ΜΑΘΗΜΑΤΟΣ	ΕΠ22α	Εξάμηνο	7	ECTS	6															
ΩΡΕΣ ΔΙΔΑΣΚΑΛΙΑΣ	ΘΕΩΡ.	3	ΦΡΟΝΤ.	1	ΕΡΓΑΣΤ.															
ΤΥΠΟΣ ΜΑΘΗΜΑΤΟΣ	Επιλέξτε ένα από τα ακόλουθα και διαγράψτε τα υπόλοιπα Προαιρετικό Μάθημα (ΠΜ) <table border="1"> <thead> <tr> <th>Κ</th><th>Ε1</th><th>Ε2</th><th>Ε3</th><th>Ε4</th><th>Ε5</th><th>Ε6</th></tr> </thead> <tbody> <tr> <td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> </tbody> </table> Συμπληρώστε τον πίνακα όπως στο πρόγραμμα σπουδών: Κατεύθυνση (Α, Β) / Υποχρεωτικό Ειδίκευσης (Υ) / Βασικό Ειδίκευσης (Β)/ Επιλογής Ειδίκευσης (Ε)						Κ	Ε1	Ε2	Ε3	Ε4	Ε5	Ε6							
Κ	Ε1	Ε2	Ε3	Ε4	Ε5	Ε6														
ΗΛΕΚΤΡΟΝΙΚΗ ΣΕΛΙΔΑ ΜΑΘΗΜΑΤΟΣ (URL)	eclass <at> uoa.gr																			
ΠΡΟΑΠΑΙΤΟΥΜΕΝΑ ΜΑΘΗΜΑΤΑ	ΔΕΝ ΥΠΑΡΧΟΥΝ																			
ΓΛΩΣΣΑ ΔΙΔΑΣΚΑΛΙΑΣ και ΕΞΕΤΑΣΕΩΝ:	ΕΛΛΗΝΙΚΑ																			
ΤΟ ΜΑΘΗΜΑ ΠΡΟΣΦΕΡΕΤΑΙ ΣΕ ΦΟΙΤΗΤΕΣ ERASMUS	Θα μπορούσε																			

ΠΕΡΙΕΧΟΜΕΝΟ ΜΑΘΗΜΑΤΟΣ
Συμπληρώστε (λεκτική περιγραφή) το περιεχόμενο του μαθήματος αναφέροντας τη δομή και τα θέματα που καλύπτονται. Το μάθημα εξετάζει την ακόλουθη ερώτηση: - Πως μπορεί μια μονάδα/πράκτορας (agent) να μάθει να ενεργεί για την επίτευξη κάποιου σκοπού, σε ένα δυναμικό αβέβαιο και μερικώς παρατηρήσιμο περιβάλλον, παρουσία και άλλων δρώντων μηχανισμών που επιδιώκουν τους δικούς τους σκοπούς? Το ερώτημα και οι ειδικότερες εκδοχές του (ένας πράκτορας, κοινός σκοπός και συντονισμός ενεργειών των μονάδων) έχει κεντρικό ενδιαφέρον σε μεγάλο φάσμα εφαρμογών: - Υπολογιστική όραση, ρομποτική και αυτόνομη οδήγηση - Επεξεργασία φυσικής γλώσσας - Παιχνίδια (επιτραπέζια, παιχνίδια με τράπουλες, βιντεο) - Κυβερνοφυσικά συστήματα - Ανίχνευση σπάνιων γεγονότων (ανίχνευση και διάγνωση στην ιατρική φροντίδα) - Επικοινωνίες και ασφάλεια (αντιμετώπιση απειλών και επιθέσεων σε αισθητήρες, κανάλια και υπολογιστικούς κόμβους) - Το μάθημα ασχολείται με εκείνες τις απαντήσεις στο παραπάνω ερώτημα οι οποίες στηρίζονται στη θεωρία τις μεθόδους και τους αλγορίθμους της ενισχυτικής μηχανικής μάθησης (reinforcement learning). Ετσι, η αλληλεπίδραση μεταξύ των μονάδων και του περιβάλλοντος στο οποίο ενεργούν είναι συνεχής. Οι μονάδες παρατηρούν την κατάσταση του περιβάλλοντος μέσω σχετικών μετρήσεων και έμμεσα αντλούν πληροφορίες για τις ενέργειες

των άλλων μονάδων. Κάποια πληροφοριακά στοιχεία είναι κοινά σε όλες τις μονάδες, άλλα συνιστούν ιδιωτική πληροφόρηση. Με βάση τις παρατηρήσεις που έχει στη διάθεση της, κάθε μονάδα επιλέγει μια ενέργεια και την εκτελεί. Η κατάσταση του περιβάλλοντος μεταβάλλεται χρονικά και με αβέβαιο (πιθανοτικό) τρόπο, ανάλογα με την μέχρι τώρα εξέλιξη της και βάσει των ενεργειών των μονάδων. Οι αποφάσεις κάθε μονάδας λαμβάνονται ακολουθιακά (στό χρόνο ή σε στάδια), στηρίζονται σε κανόνες, συνιστούν δηλαδή πολιτικές και κατά συνέπεια και η μάθηση είναι ακολουθιακή. Ο κύκλος αυτός επαναλαμβάνεται συνεχώς επιτρέποντας στις μονάδες να μαθαίνουν πως να ενεργούν ώστε να βελτιώνουν, όχι τη στιγμιαία ανταμειβή, αλλά το συνολικό μακροπρόθεσμο όφελος/απόδοση.

Ειδικότερα, θα μελετήσουμε:

1. τις θεμελιώδεις μεθόδους της ενισχυτικής μάθησης:

- Μαρκοβιανές διαδικασίες απόφαση (MDPs), μερικώς παρατηρήσιμες MDPs (POMDPs) και πίστη, Δυναμικός προγραμματισμός και εξίσωση Bellman, Q learning, στοχαστικά παίγνια, κοινή και ιδιωτική πληροφόρηση, πολλαπλοί πράκτορες, ισορροπίες
- Εξομείωση και αλγόριθμοι στοχαστικής προσέγγισης
- Βασικές προσεγγιστικές δομές, βαθιά νευρωνικά δίκτυα με και χωρίς μνήμη.

2. Αντιπρόσωπους αλγορίθμων ενισχυτικής μάθησης (vanilla policy gradient VPG, trust region policy optimization TRPO, proximal policy optimization PPO, Deep deterministic policy gradient DPPG, Twin delayed DDPG, TD3, Soft Actor Critic SAC, DQN, DDQN, και άλλους)

3. Βιβλιοθήκες αλγορίθμων ενισχυτικής μάθησης (Stable baselines3), βιβλιοθήκες με προεγκατεστημένα περιβάλλοντα (OpenAI Gym) και υποκείμενες βιβλιοθήκες νευρωνικών δικτύων (Pytorch).

ΜΑΘΗΣΙΑΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Περιγράψτε τους στόχους ή/και τα μαθησιακά αποτελέσματα του μαθήματος. Μπορείτε να αναφερθείτε στις επιμέρους κατηγορίες των μαθησιακών αποτελεσμάτων σε επίπεδο γνώσεων, δεξιοτήτων (νοητικών, πρακτικών) και ικανοτήτων.

Συμβουλευτείτε το Παράρτημα «Περιληπτικός Οδηγός συγγραφής Μαθησιακών Αποτελεσμάτων» για τη συγγραφή των μαθησιακών αποτελεσμάτων.

Απόκτηση γνώσεων και δεξιοτήτων σε μια σημαντική περιοχή της μηχανικής μάθησης που έχει ισχυρούς δεσμούς με τον βέλτιστο έλεγχο (Optimal control), την επιχειρησιακή έρευνα (Operations Research), τις προσεγγιστικές δομές (στοχαστική προσέγγιση, νευρωνικά δίκτυα), την εξομείωση, τη θεωρία πληροφορίας και τη θεωρία παιγνίων, με εξαιρετικές επιτυχίες τα τελευταία χρόνια

ΔΙΔΑΚΤΙΚΕΣ και ΜΑΘΗΣΙΑΚΕΣ ΜΕΘΟΔΟΙ - ΑΞΙΟΛΟΓΗΣΗ

ΤΡΟΠΟΣ ΠΑΡΑΔΟΣΗΣ

Στην τάξη (Πρόσωπο με πρόσωπο)

ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ

Υποστήριξη μαθησιακής διαδικασίας μέσω της ηλεκτρονικής πλατφόρμας e-class (Παροχή υλικού, Ανακοινώσεις, Ανάθεση εργασιών, Ομάδες φοιτητών)
Επικοινωνία μέσω ηλεκτρονικού ταχυδρομείου
Ζωντανή μετάδοση διαλέξεων
Δυνατότητα παρακολούθησης καταγεγραμμένων διαλέξεων

ΟΡΓΑΝΩΣΗ ΔΙΔΑΣΚΑΛΙΑΣ – ΔΙΔΑΚΤΙΚΕΣ ΤΕΧΝΙΚΕΣ

Περιγράφονται αναλυτικά ο τρόπος και τεχνικές διδασκαλίας, αριθμός ατόμων ανά ομάδα κτλ.

Περιγράφονται αναλυτικά ο τρόπος και τεχνικές διδασκαλίας και αναγράφονται αναλυτικά οι ώρες μελέτης του φοιτητή για κάθε μαθησιακή δραστηριότητα καθώς και οι ώρες μη καθοδηγούμενης μελέτης ώστε ο συνολικός φόρτος εργασίας σε επίπεδο εξαμήνου να αντιστοιχεί στα standards του ECTS (25 ώρες φόρτου εργασίας ανά πιστωτική μονάδα) Εμπλουτισμένες Διαλέξεις, Online Διαλέξεις, Σεμινάρια, Φροντιστήριο, Εργαστήριο, Εργαστηριακή Άσκηση, Μελέτη & ανάλυση βιβλιογραφίας, Πρακτική Εκπόνηση project, Εκπόνηση ατομικών / ομαδικών εργασιών Τηλεσυνεργασία (αναφορά σε εργαλεία) Κλπ	Συμπληρώνεται αναλόγως και ο παρακάτω πίνακας.		
	Δραστηριότητα	Φόρτος (ώρες)	
	Διαλέξεις	39	
	Φροντιστήριο	13	
	Εργαστήριο	13	
	Ομαδική Εργασία σε μελέτη περίπτωσης	30	
	Μικρές ατομικές εργασίες εξάσκησης	10	
	Αυτοτελής Μελέτη	45	
		...	
	Σύνολο ωρών	150	
ΑΞΙΟΛΟΓΗΣΗ ΦΟΙΤΗΤΩΝ Περιγραφή της διαδικασίας αξιολόγησης Μέθοδοι αξιολόγησης (Διαμορφωτική ή/και Τελική), Εργαλεία Αξιολόγησης (Δοκιμασία Πολλαπλής Επιλογής, Ερωτήσεις Σύντομης Απάντησης, Ερωτήσεις Ανάπτυξης Δοκιμίων, Επίλυση Προβλημάτων, Γραπτή Εργασία, Έκθεση/Αναφορά, Προφορική Εξέταση, Δημόσια Παρουσίαση, Εργαστηριακή Εργασία, Άλλη/άλλες,), Παροχή ανατροφοδότησης (περιγραφική, μέσω κλίμακας διαβαθμισμένων κριτηρίων)	Συμμετοχή στο μάθημα (ανά δύο εβδομάδες παρουσιάσεις μαθησιακής προόδου		
	Εργασία		
	Αξιολόγηση	Αριθμός	Ποσοστό
	Γραπτή εξέταση	0	0%
	Μαθησιακή πρόοδος	2	20%
	Ασκήσεις	3	20%
	Εργαστήριο	0	0%
	Τελική εργασία	1	60%

ΣΥΝΙΣΤΩΜΕΝΗ-ΒΙΒΛΙΟΓΡΑΦΙΑ
1.D. Bertsekas: Dynamic Programming and Optimal Control Vols I and II, Athena, Third Edition 2.Sutton and Barto: Reinforcement Learning: an introduction, Draft Textbook Jan 2018. 3.D. Bertsekas and J. Tsitsiklis: Neurodynamic Programming, Athena 1996. 4.D. Silver: Introduction to RL (DeepMind, ucl) Διαλέξεις D. 5. OpenAI SpinningUP documentation 6.Stable-Baselines3 documentation 7. V. Krishnamurthi: Partially observed Markov Decision Processes, Cambridge, 2016 8. J. Filar and Koos Vrieze Competitive Markov Decision Processes, Springer 1996. Σχετικές δημοσιεύσεις θα δοθούν κατά τη διάρκεια του μαθήματος στην η-τάξη.