

Εργασία Εξαμήνου - Φάση Β

Το gridworld πρόβλημα

Προθεσμία Υποβολής: 08/12/2025

(Η προθεσμία είναι ενδεικτική. Η καταληκτική ημερομηνία για την παράδοση όλων των εργασιών θα είναι προς το τέλος του εξαμήνου όπου θα βγει σχετική ανακοίνωση).

Οι μη τερματικές καταστάσεις είναι $S = \{1, 2, \dots, 14\}$. Υπάρχουν τέσσερις ενέργειες σε κάθε κατάσταση (εκτός από τις τερματικές καταστάσεις που δεν κάνουμε τίποτα), $A = \{up, down, right, left\}$. Οι ενέργειες προκαλούν με ντετερμινιστικό τρόπο μεταβάσεις καταστάσεων, εκτός από αυτές τις ενέργειες που θα βγάλουν τον πράκτορα εκτός του πλέγματος. Ισχύει ότι $p(6, -1|5, right) = 1$, $p(7, -1|7, right) = 1$, $p(10, r|5, right) = 0$ for all $r \in R$. Η ανταμοιβή είναι -1 για όλες τις μεταβάσεις.

- 1) Ακολουθήστε μία ισοπίθανη τυχαία πολιτική και εφαρμόστε την επαναληπτική αξιολόγηση πολιτικής για να βρείτε την λύση.
 - a. Χρησιμοποιήστε 2 πίνακες για να αποθηκεύετε τις συναρτήσεις τιμής και να κάνετε τις ενημερώσεις
 - b. Χρησιμοποιήστε μόνο 1 πίνακα για να αποθηκεύετε τις συναρτήσεις τιμής και να κάνετε τις ενημερώσεις
- 2) Τυπώστε την συνάρτηση τιμής για κάθε κατάσταση.
- 3) Βρείτε τη (ις) βέλτιστη (ες) ενέργεια (ες) για κάθε κατάσταση

