Reinforcement Learning

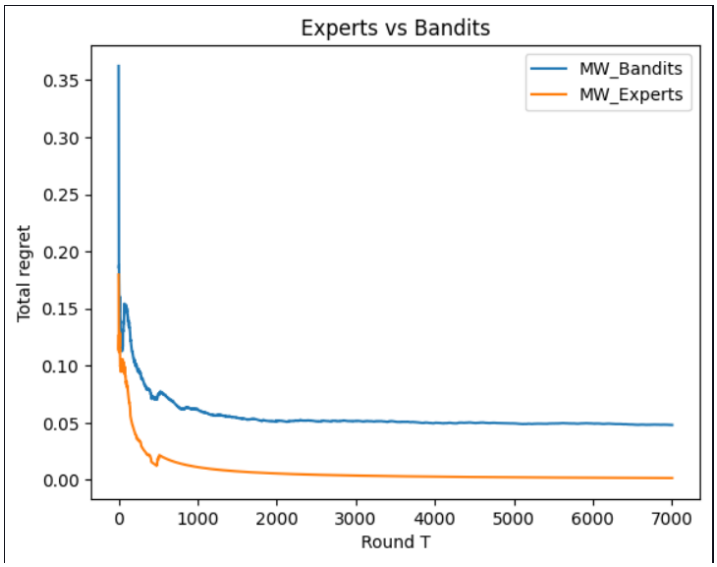
2nd Programming Assignment

ΚΩΝΣΤΑΝΤΙΝΟΣ ΝΙΚΟΛΟΣ
ΑΜ:2019030096

Περιγραφή της άσκησης:

Ζητείται η υλοποίηση του αλγορίθμου Multiplicate Weights σε περιβάλλον Experts και Bandits. Βασική διαφορά των 2 αποτελεί η έλλειψη γνώσης για όλους τους experts. Στο εχθρικό περιβάλλον Bandits γνωρίζουμε το loss μόνο του expert που επιλέχθηκε σε κάθε γύρο επομένως αλλάζουμε μόνο εκείνη την τιμή καθώς και μόνο το δικό του βάρος. Αποτέλεσμα αυτού η αργότερη 'μάθηση' σε περιβάλλον bandits σε αντίθεση με τος Experts.

Γνωρίζουμε ότι το Multiplicate Weights Algorithm σε Experts περιβάλλον αναμένεται να έχει $\text{RegretT} = \leq 2\sqrt{\text{Tlnk}}$.
Ενώ σε Bandits αναμένεται να έχει $\text{RegretT} = \sqrt{\text{kTlnk}} = 5.47\sqrt{\text{Tlnk}} = 2.73 * 2\sqrt{\text{Tlnk}}$
γεγονός που μπορεί να επιβεβαιωθεί από το παρακάτω διάγραμμα



Ακόμη ένα μέτρο σύγκρισης που μας ενδιαφέρει είναι πως στο Experts παρατηρούνται

4820 σωστές προβλέψεις σε αντίθεση με τις 3711