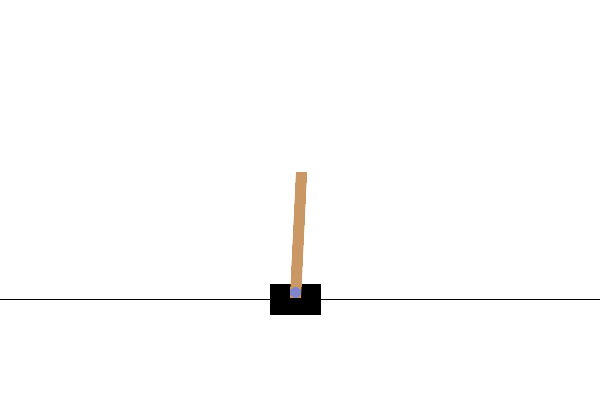
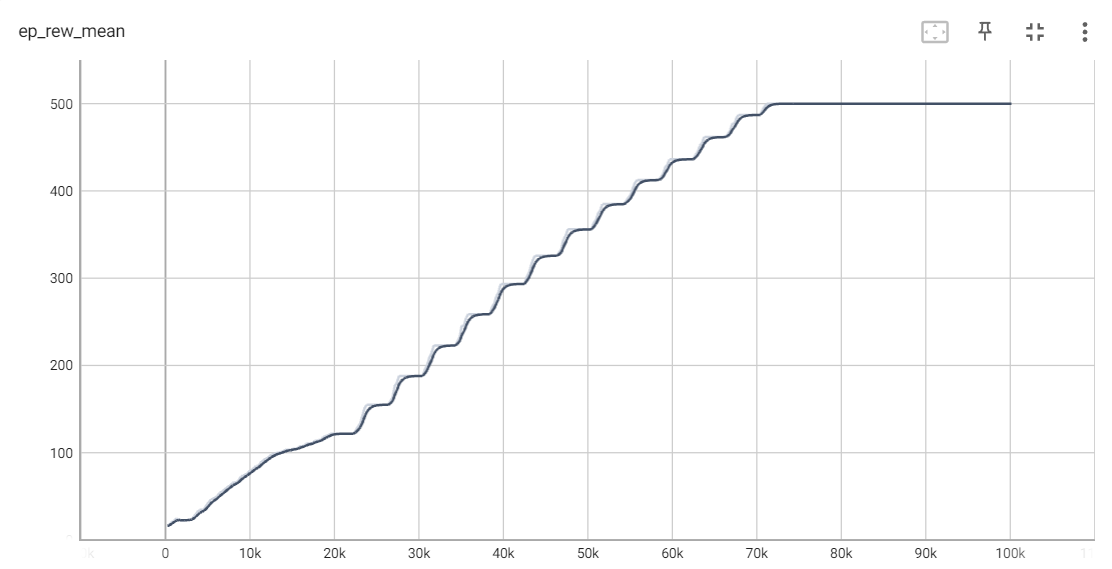
**CartPole**

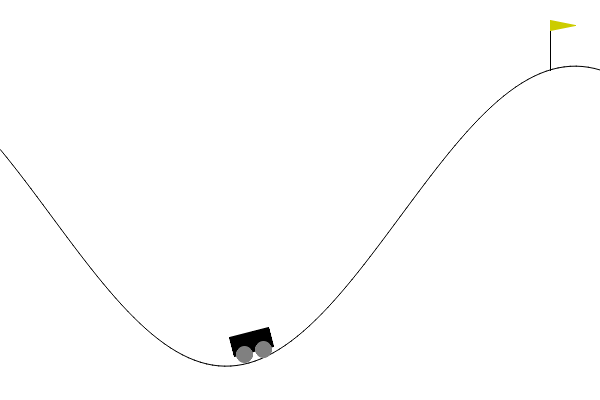


Ένας στύλος συνδέεται με έναν μη ενεργοποιημένο σύνδεσμο σε ένα καρότσι, το οποίο κινείται κατά μήκος μιας τροχιάς χωρίς τριβές. Το σύστημα ελέγχεται με την εφαρμογή δύναμης +1 ή -1 στο καρότσι. Το εκκρεμές ξεκινά όρθιο και ο στόχος είναι να μην πέσει. Μια ανταμοιβή +1 παρέχεται για κάθε χρονικό βήμα που ο στύλος παραμένει όρθιος. Το επεισόδιο τελειώνει όταν ο πόλος απέχει περισσότερο από 15 μοίρες από την κατακόρυφο ή το καρότσι μετακινηθεί πάνω από 2,4 μονάδες από το κέντρο.

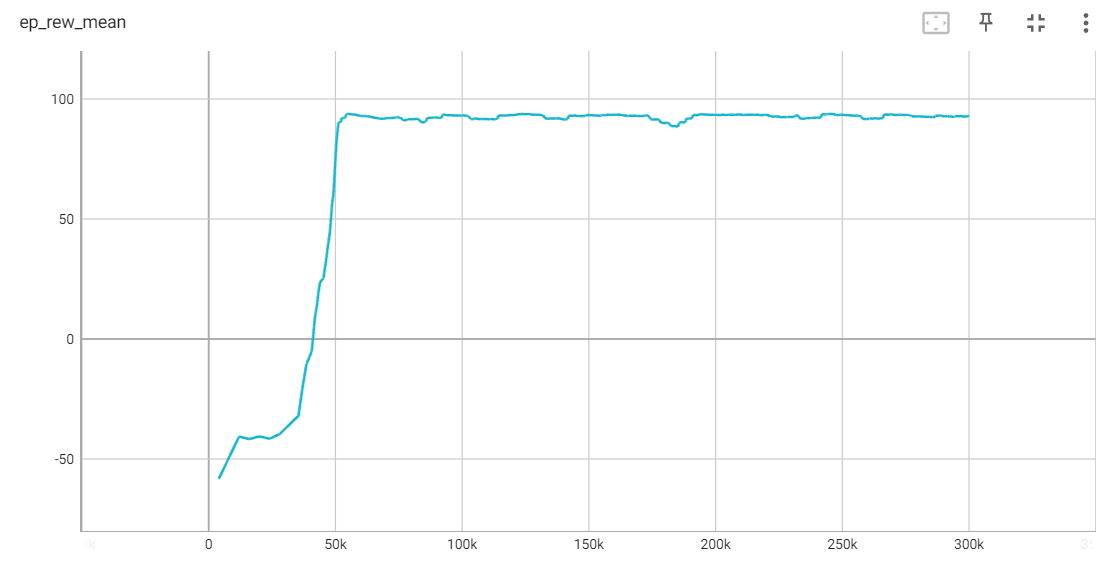
Το CartPole-v0 ορίζει την "λύση" ως τη λήψη μέσης ανταμοιβής 195,0 σε 100 διαδοχικές δοκιμές.Εμείς φτάνουμε στα 500 άρα το έχουμε λύσει.

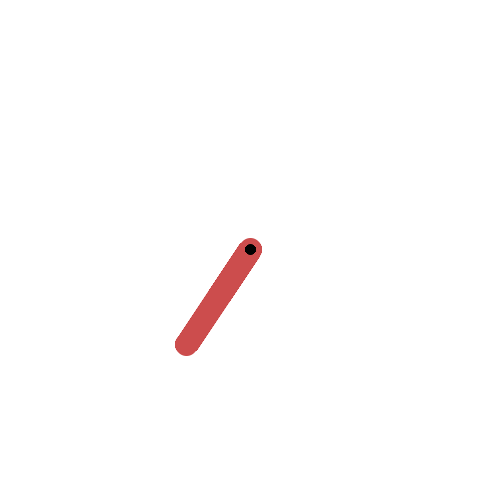


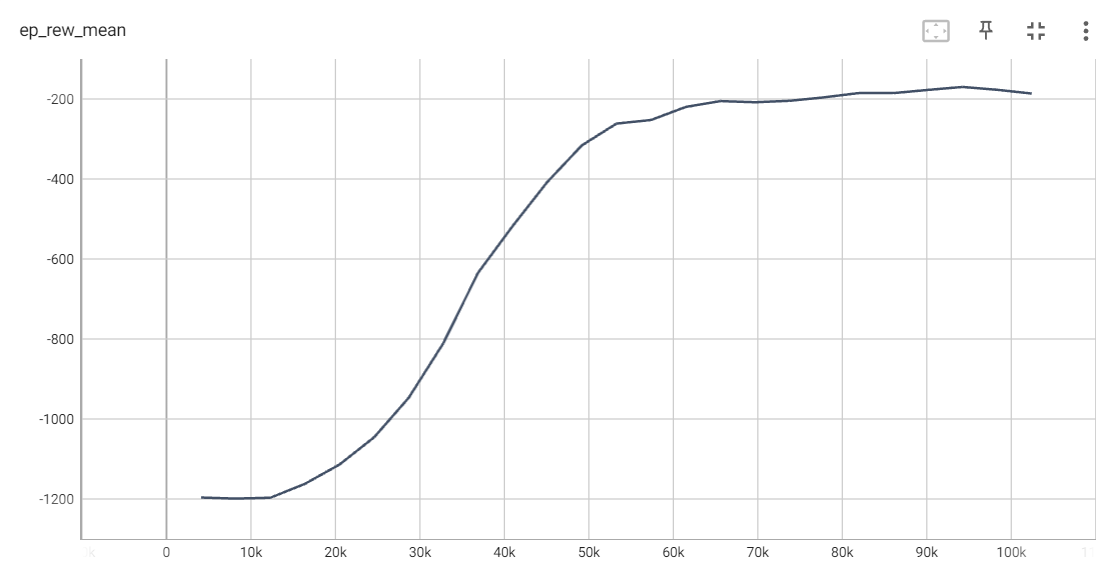
**Mountain Car (Continuous** **)**

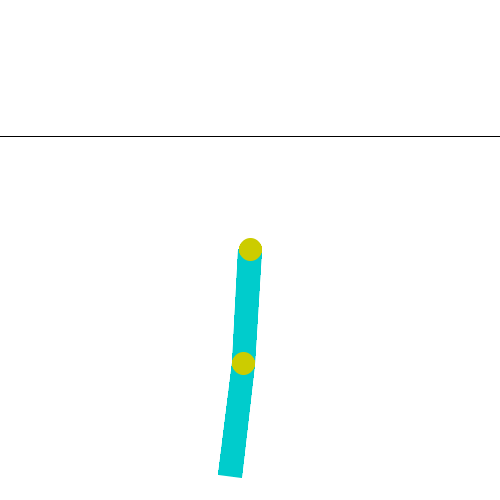
 Ένα αυτοκίνητο βρίσκεται σε μονοδιάστατη πίστα, τοποθετημένο ανάμεσα σε δύο «βουνά». Ο στόχος είναι να ανεβείτε το βουνό στα δεξιά. Ωστόσο, ο κινητήρας του αυτοκινήτου δεν είναι αρκετά δυνατός για να ανέβει στο βουνό με ένα μόνο πέρασμα. Επομένως, ο μόνος τρόπος για να πετύχετε είναι να οδηγείτε εμπρός και πίσω για να αποκτήσετε δυναμική. Εδώ, η ανταμοιβή είναι μεγαλύτερη αν ξοδέψετε λιγότερη ενέργεια για να φτάσετε τον στόχοΕδώ, αυτή είναι η συνεχής έκδοση.

Το MountainCarContinuous-v0 ορίζει την "λύση" ως τη λήψη μέσης ανταμοιβής 90,0 σε 100 διαδοχικές δοκιμές.Εμείς έχουμε 90 άρα το έχουμε λύσει.

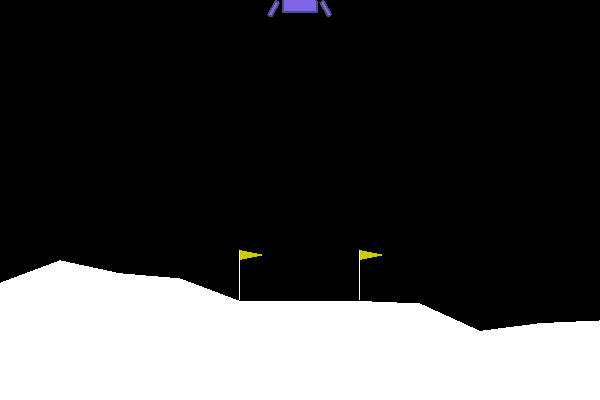


**Pendulum** 

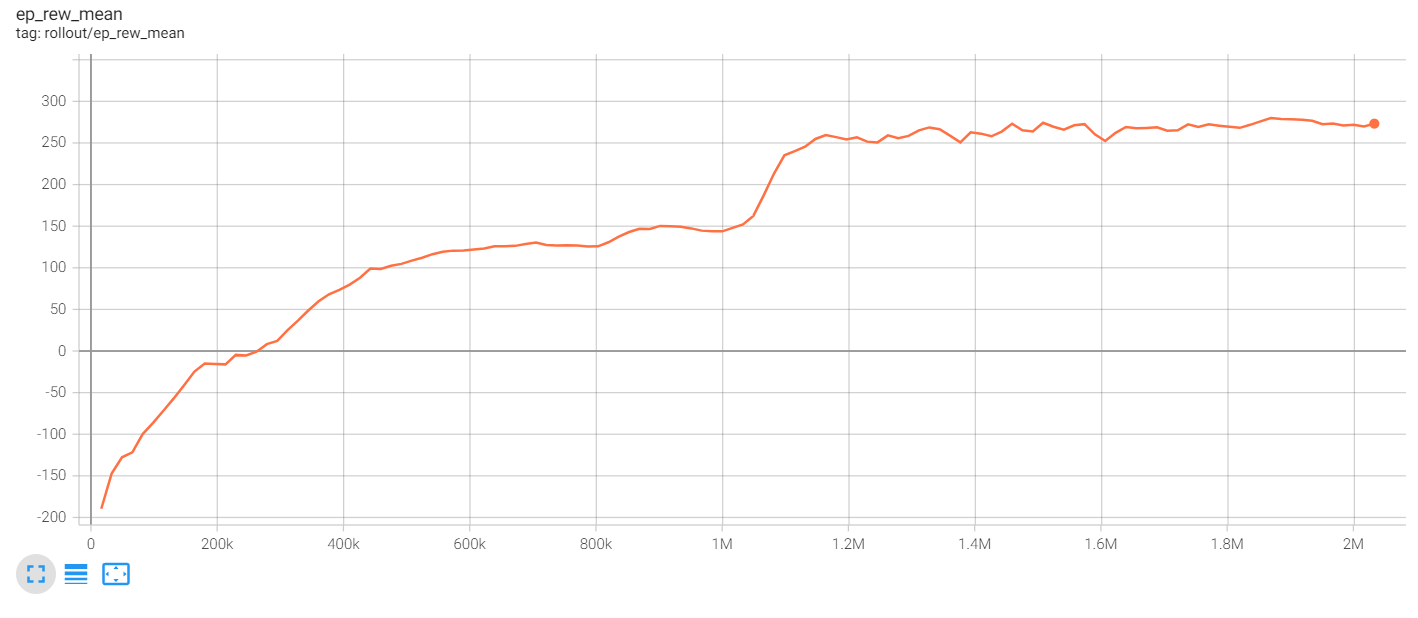
Το πρόβλημα της αιώρησης του ανεστραμμένου εκκρεμούς είναι ένα κλασικό πρόβλημα στη βιβλιογραφία ελέγχου. Σε αυτήν την έκδοση του προβλήματος, το εκκρεμές ξεκινά σε μια τυχαία θέση και ο στόχος είναι να το αιωρήσετε προς τα πάνω ώστε να παραμείνει όρθιο.Το Pendulum-v0 είναι ένα άλυτο περιβάλλον, που σημαίνει ότι δεν έχει καθορισμένο όριο ανταμοιβής στο οποίο θεωρείται λυμένο. Εμείς έχουμε -200.

**Acrobot**

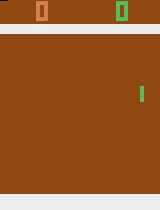
Το σύστημα ακρομποτ περιλαμβάνει δύο συνδέσμους και δύο συνδέσμους, όπου ενεργοποιείται η άρθρωση μεταξύ των δύο συνδέσμων. Αρχικά, οι σύνδεσμοι κρέμονται προς τα κάτω και ο στόχος είναι να ταλαντευτεί το άκρο του κάτω συνδέσμου μέχρι ένα δεδομένο ύψος. Το Acrobot-v1 είναι ένα άλυτο περιβάλλον, που σημαίνει ότι δεν έχει καθορισμένο όριο ανταμοιβής στο οποίο θεωρείται λυμένο. Εμείς έχουμε -85.

**Lunar Lander**

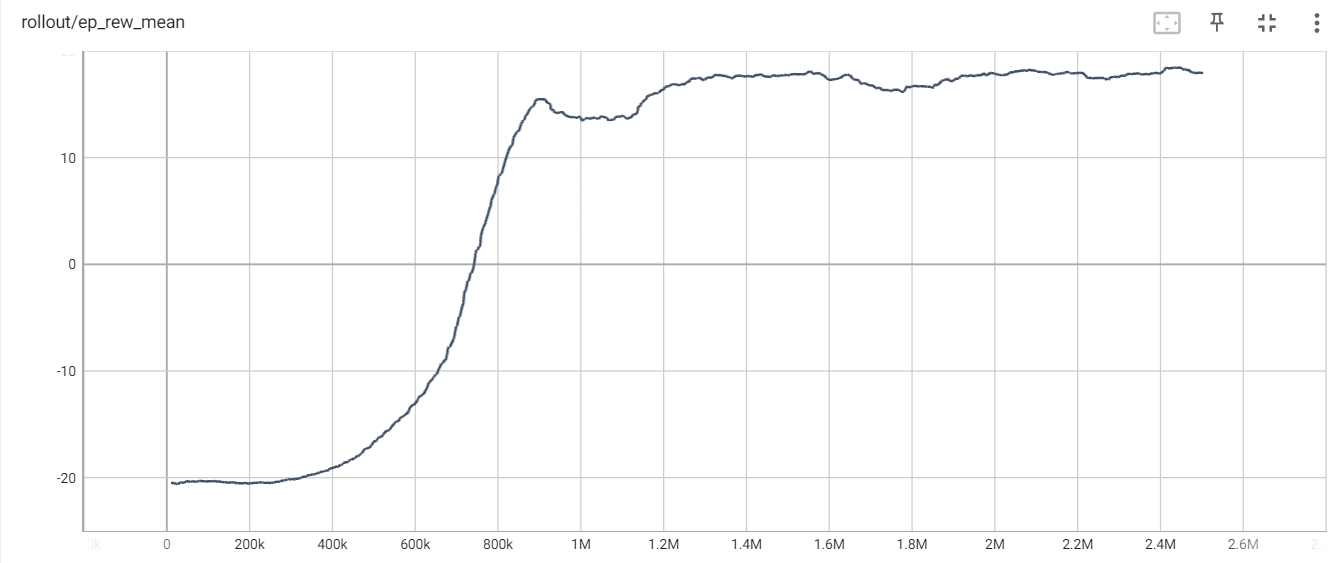
Το σημείο προσγείωσης είναι πάντα στις συντεταγμένες (0,0). Οι συντεταγμένες είναι οι δύο πρώτοι αριθμοί στο διάνυσμα κατάστασης. Η ανταμοιβή για τη μετάβαση από την κορυφή της οθόνης στο σημείο προσγείωσης και η μηδενική ταχύτητα είναι περίπου 100..140 βαθμοί. Εάν ο προσγειωτής απομακρυνθεί από το σημείο προσγείωσης, χάνει την ανταμοιβή πίσω. Το επεισόδιο τελειώνει εάν το σκάφος προσγειωθεί ή ακινητοποιηθεί, λαμβάνοντας επιπλέον -100 ή +100 πόντους. Κάθε επαφή γείωσης ποδιού είναι +10. Ο κύριος κινητήρας πυροδότησης είναι -0,3 βαθμοί κάθε καρέ. Η λύση είναι 200 βαθμοί. Είναι δυνατή η προσγείωση έξω από το σημείο προσγείωσης. Το καύσιμο είναι άπειρο, επομένως ένας πράκτορας μπορεί να μάθει να πετάει και στη συνέχεια να προσγειωθεί στην πρώτη του προσπάθεια. Τέσσερις διακριτές ενέργειες είναι διαθέσιμες: μην κάνετε τίποτα, πυροβολήστε κινητήρα αριστερού προσανατολισμού, πυροδότηση κύριας μηχανής, πυροδότηση μηχανής δεξιού προσανατολισμού. Το LunarLander-v2 ορίζει την "λύση" ως τη λήψη μέσης ανταμοιβής 200 σε 100 διαδοχικές δοκιμές. Εμείς έχουμε 270 άρα το έχουμε λύσει.



**Atari Pong**



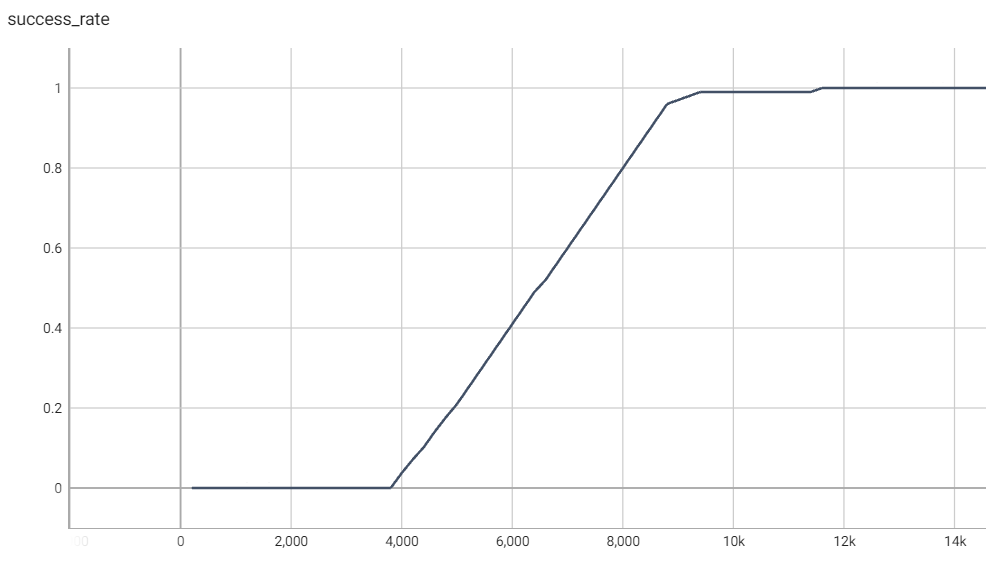
Το παιχνίδι Pong της Atari.Θεωρείται λυμένο για σκορ πάνω από 20. Εμείς έχουμε 18 άρα δεν το έχουμε λύσει.



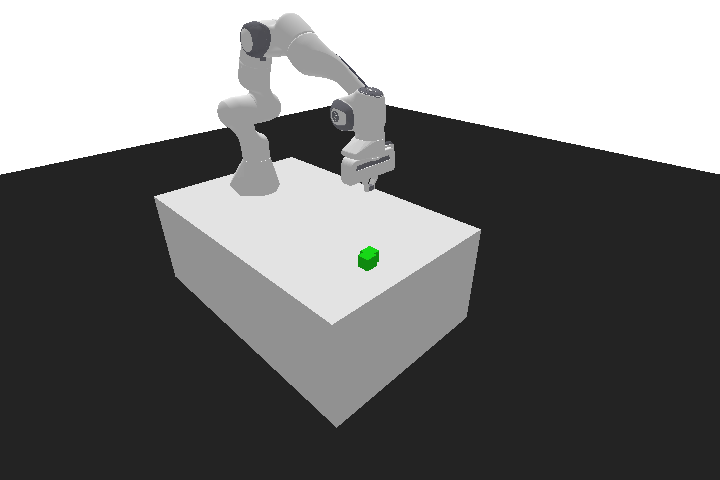
**Panda Gym Reach**



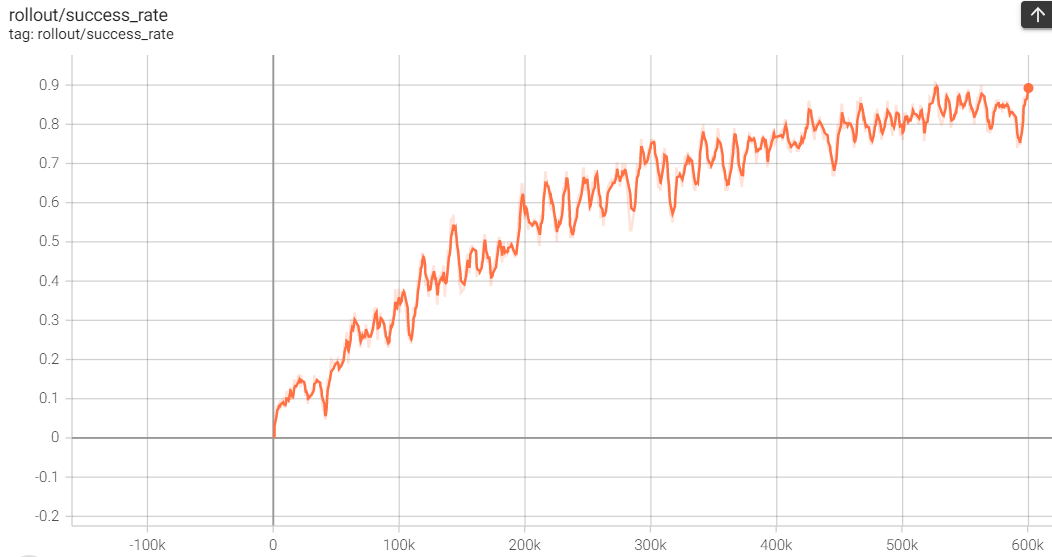
Στόχος είναι ο ρομποτικός βραχίονας να φτάσει σε μια συγκεκριμένη θέση στον χώρο.Έχουμε ποσοστό επιτυχίας 100% άρα το έχουμε λύσει.



**Panda Gym Push**

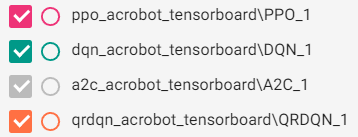
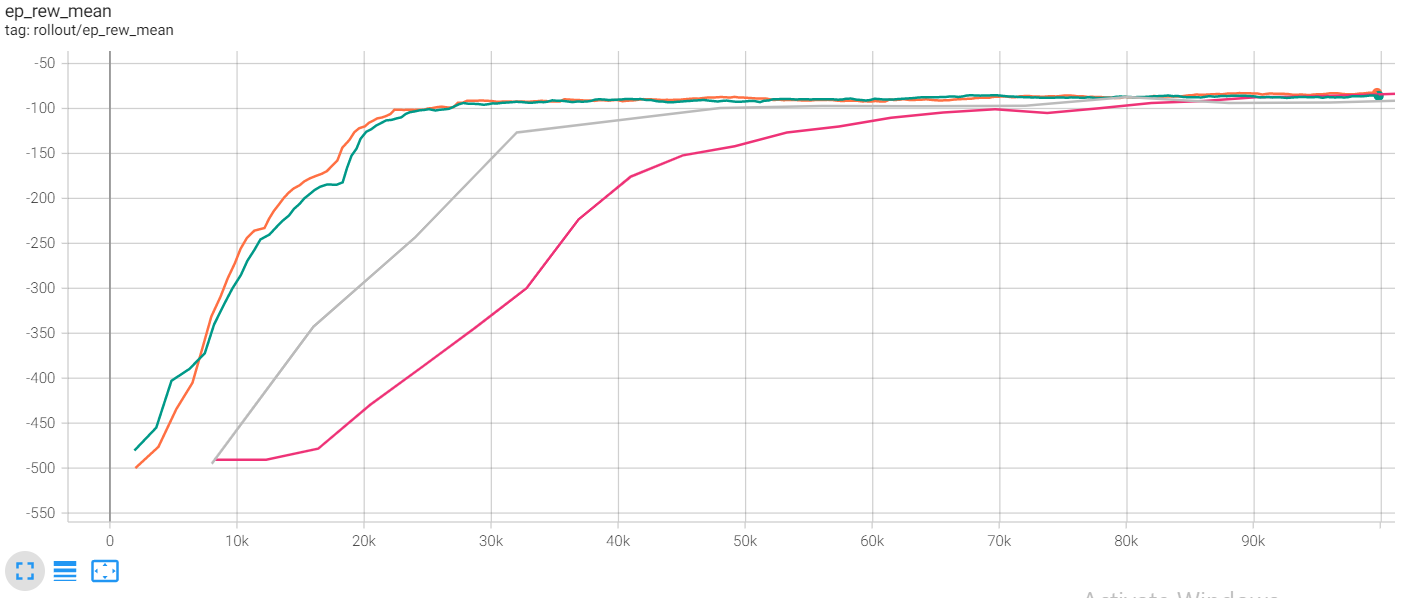


Στόχος είναι ο ρομποτικός βραχίονας να μετακινήσει ένα κουτί σπρώχνοντάς το μέχρι να φτάσει στην επιθυμητή θέση στόχου. Έχουμε ποσοστό επιτυχίας 90% άρα το έχουμε λύσει.



**Σύγκριση αλγορίθμων**

Τρέχουμε 4 διαφορετικούς αλγορίθμους στο περιβάλλον Acrobot.Οι αλγόριθμοι είναι οι : PPO,DQN,A2C,QRDQN.



Βλέπουμε ότι η πιο αργή είναι η PPO και η πιο γρήγορη είναι η QRDQN.