Oppsummering

Johan Martin Skavhaug

Del 1

Torsdag 17. okt. Undersøkte muligheten for å trene RL-miljøet Ant fra Pybullet. Fikk lastet ned pybullet og kjørt eksempelkode med ferdigtrent Ant. Sendte mail til Jonathan om prosjektoppgaven.

Tirsdag 22. okt Studerte forskjellige algoritmer, og hvilke av disse som kunne brukes på Ant. SAC og PPO er to interessante. Undersøkte hvilke biblioteker det finnes for å trene i RL.Jobbet gjennom `https://github.com/tensorflow/agents/blob/master/tf_agents/colabs/7_SAC_minitaur_tutorial.ipynb`

Tirsdag 29. okt Veiledningsmøte, ble oppfordret til å saumfare nettet etter andre implementasjoner for å se hvordan andre har gjort det. . Fant og jobbet med TF Agents: `https://github.com/tensorflow/agents/blob/master/tf_agents/agents/ppo/examples/v2/train_eval.py` Skrevet om eksempelet til å trene Ant med ppo, men ikke oppnådd noen gode resultater. Usikker på om rendering faktisk fungerer siden Ant står i ro, eller om modellen er nødt til å trene lengre.

Tirsdag 5. nov Veiledningsmøte. Sjekket hvordan hyperparametre som tf_agents bruker i PPO. Satte opp koden til å endre på noen hyperparametre og begynte å se på hvordan læringen endret seg med forskjellige hyperparametre. Søkte mye på nett for å finne forklaringer på forskjellige hyperparametre og forstå hvordan de er implementert i TF agents.

Onsdag 6 nov. Satte opp flere forskjellige hyperparametre til trening i VM. Code refactor for å få alle hyperparametre justerbar fra kommandolinje slik at vi kan trene forskjellige kombinasjoner vha. scripts, og bedre holde styr på hvilke kombinasjoner vi har prøvd, og hvilke kombinasjoner som gjør hva.

Del 2

Fredag 8. nov Satte opp en hel haug med hyperparametre for trening over helga.

Lørdag 9. nov Studert en del av resultatene av treningene, ble skuffet over de foreløpige resultatene.

Mandag 11. nov Leste meg opp på hva de forskjellige hyperparametrene i ppo er i detalj og prøvd å finne ut hvordan de er implementert i TF-Agents. Studert mye kildekode for å finne ut hva som må til for at maura skal gå. Fant ikke ut hvorfor modellen ikke gir noe godt resultat. Sette opp adaptive learning_rate og endre nettverksstørrelsene og forskjellige verdier for kl tolerance og target. Fant mye info og sammenlignet kildekode med dette repoet: `https://github.com/pat-coady/trpo`
lest litt: `https://arxiv.org/abs/1506.02438`

Tirsdag 12. nov Veiledningsmøte. Rotet i litt kildekode for å legge til en ekstra penalty for at mauren ikke rører på seg. Satte opp en ny batch med hyperparametre med nettverk som faktisk reflekterer dimensjonene til ant. networks280x80 er med stillpenalty-2

Onsdag 13. nov Sett mer på hvorfor maura ikke går, satt opp en ny batch med trening med adaptive learning rate og entropy regularization = 0.01

Mandag 18. nov Fyllt ut deler av rapporten med korte tanker. Begynt å skrive teori og samle kilder vi har brukt.

Tirsdag 19. nov Veiledningsmøte. Flyttet fokus fra ant og ppo over på grid-search av hyperparametre. Begynt å studere resultatene av alle kombinasjonene av hyperparametre nærmere, skrevet ny introduksjon og fyllt ut mer teori.

Onsdag 20. nov Gått gjennom alle resultatene og lagt inn plotter av forskjellige kombinasjoner av hyperparametre. Isolerte de for å bedre se forskjellene på hver enkelt parameter.

Torsdag 21. nov Skrevet diskusjon om resultatene, konklusjon, ryddet kode osv.