Vehicule reinforcement learning

Nicolas Casademont, Teo Stocco

Unsupervised and reinforcement learning in neural networks 2016 class, EPFL.

Abstract

Using reward-based learning, this project shows how a car agent can learn to climb a steep hill by accelerating forwards and backwards at appropriate times. It analyses methods for hyperparameter tuning and visualize progresses across various plots.

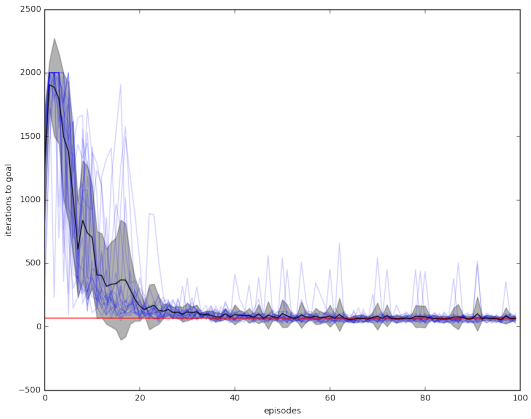
Solution

Red line mean last 60y run

Escape latency

One episode starts with the knowledge of past q-values, another initial states and resetted eligibility traces. It contains many iterations or trials that modify the state until it eventually converges.

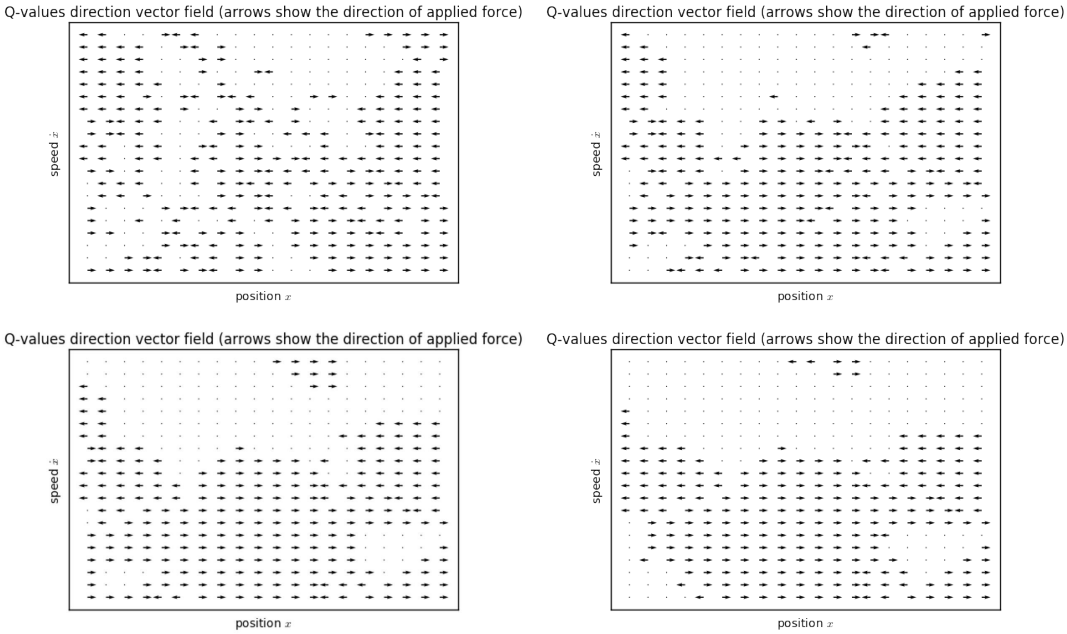
> Simulate at least 10 agents learning the task, and plot the escape latency (time to solve the task), averaged across agents, as a function of trial number (i.e., the learning curve). How long does it take the agent to learn the task?



Q-values visualization

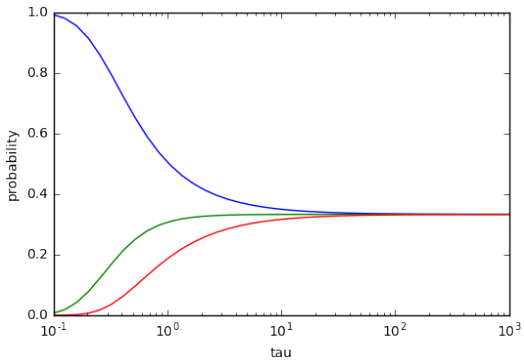
> Visualize the behavior of the agent (the policy) by plotting a vector field (0-length vector for the neutral

> action) given by the direction with the highest Q-value as a function of each possible state (x,$\dot{x}$). Plot examples after different number of trials and comment what you see.



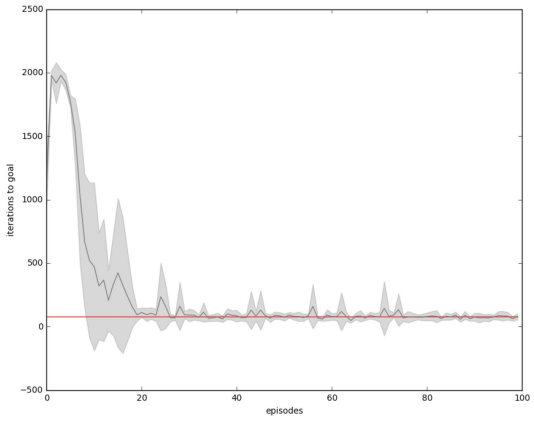
Temperature

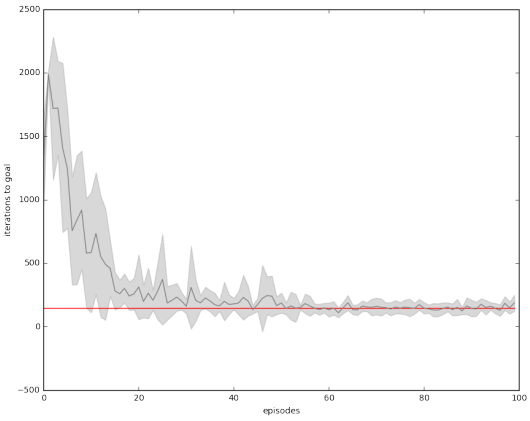
> Investigatetheexplorationtemperatureparameterτ,comparingthelearningcurves. Try fixed values such as, τ = 1, τ = ∞, τ = 0, and time decaying functions. Explain its relation to exploration and exploitation.

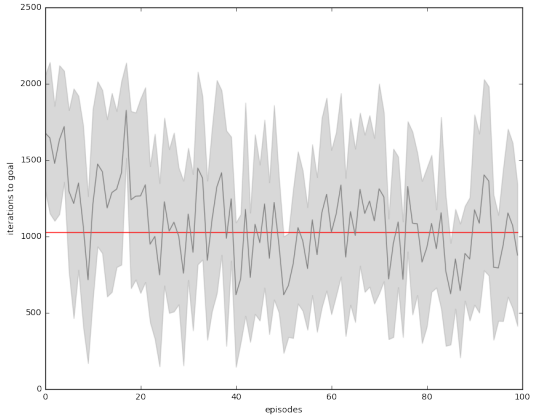


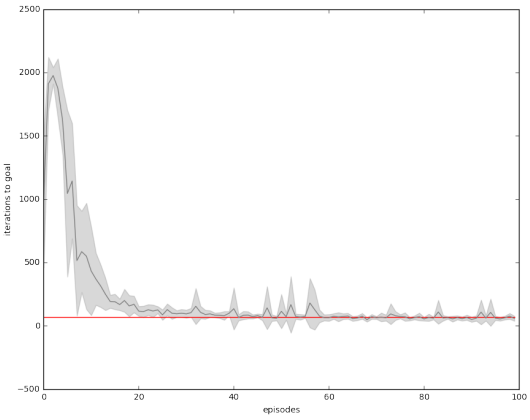
qs = np.array([0.3, 0.2, 0.1])

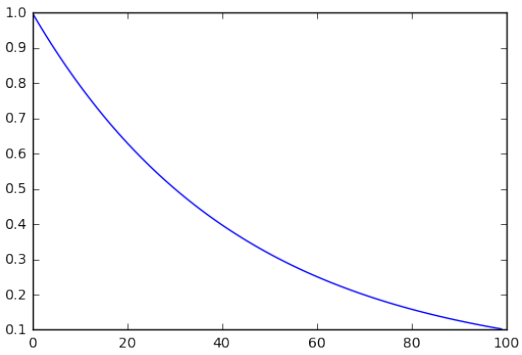
‘tau’: ‘10.6157894737’, ‘lmbda’: 0.05, ‘fill’: 0 tau = 1







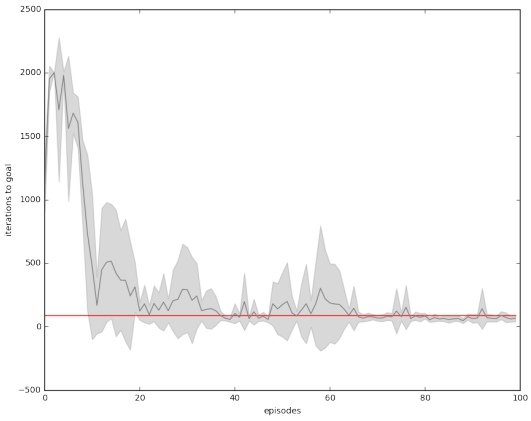


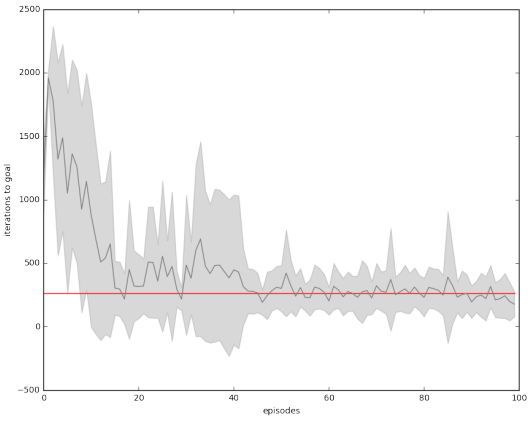


tau 1 -> 0.1 taum

Learning curve and eligibility traces

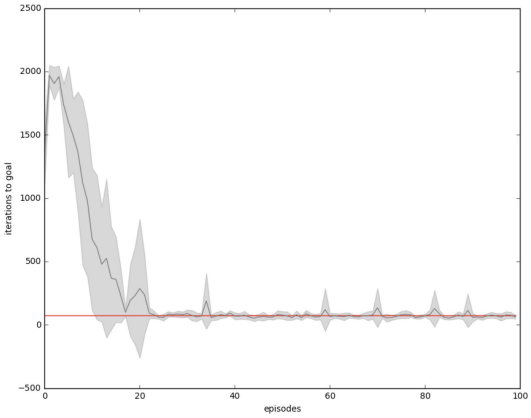
> Compare the learning curves for different values of the eligibity trace decay rate, e.g., λ = 0.95 and λ = 0. What is the role of the eligibility trace?

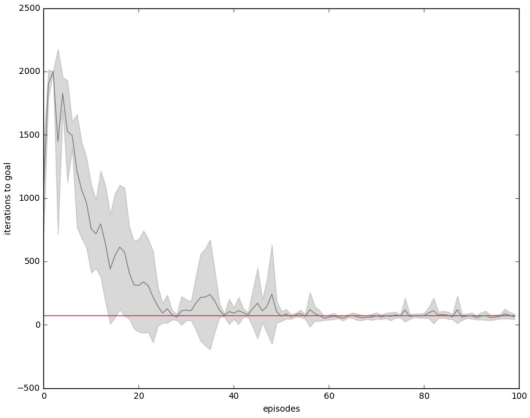


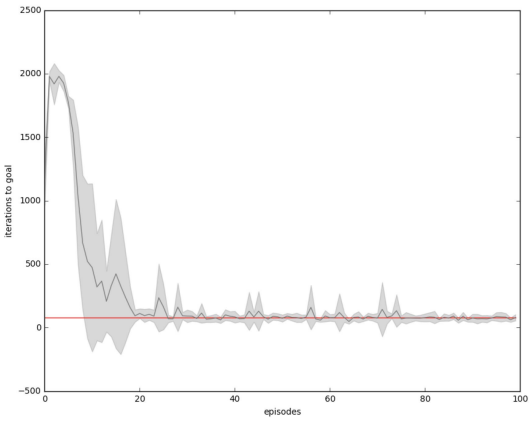


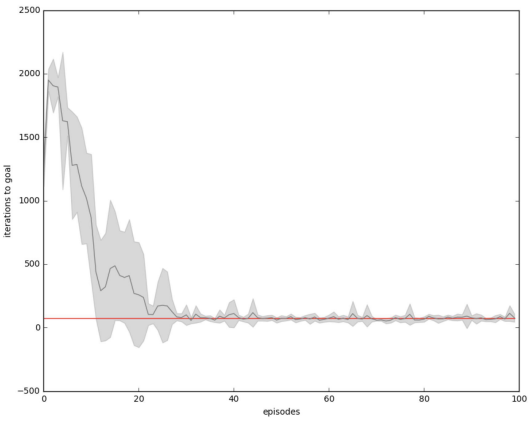
Initialization

> Try different initialization of the weights waj = 0 and waj = 1 What is the effect on the learning curves? Explain why.









Conclusion