Actor Critic Methods: From Paper to Code

REINFORCE: Monte Carlo Policy Gradients

Gradient Ascent in J

Turns into expectation value

$$\nabla J(\theta) \propto \sum_s \mu(s) \sum_a q_\pi(s,a) \nabla_\theta \pi(a|s,\theta)$$

Calculated by PyTorch

$$\nabla J(\theta) = E_\pi[\sum_a q_\pi(S_t,a) \nabla_\theta \pi(a|S_t,\theta)]$$

Gradient Ascent in J

$$\nabla J(\theta) = E_\pi[\sum_a q_\pi(S_t, a) \nabla_\theta \pi(a|S_t, \theta)]$$

Multiply and divide by policy → multiply by 1

$$\nabla J(\theta) = E_\pi[\sum_a \pi(a|S_t, \theta) q_\pi(S_t, a) \frac{\nabla_\theta \pi(a|S_t, \theta)}{\pi(a|S_t, \theta)}]$$

$$\nabla J(\theta) = E_\pi[q_\pi(S_t, A_t) \frac{\nabla_\theta \pi(A_t|S_t, \theta)}{\pi(A_t|S_t, \theta)}]$$

$$\nabla J(\theta) = E_\pi[G_t \frac{\nabla_\theta \pi(A_t|S_t, \theta)}{\pi(A_t|S_t, \theta)}]$$

All known or easily calculable quantities!

Gradient Ascent in J

Direction of change in parameter space

$$\nabla J(\theta) = E_{\pi}\left[G_t \frac{\nabla_{\theta}\pi(A_t|S_t,\theta)}{\pi(A_t|S_t,\theta)}\right]$$

Weight gradients by return

Prevent sampling bias

Implementation Notes

$$\nabla J(\theta) = E_\pi[G_t \frac{\nabla_\theta \pi(A_t|S_t,\theta)}{\pi(A_t|S_t,\theta)}]$$

$$\frac{\nabla x}{x} = \nabla \ln x \Rightarrow \frac{\nabla_\theta \pi(A_t|S_t,\theta)}{\pi(A_t|S_t,\theta)} = \nabla_\theta \ln \pi(A_t|S_t,\theta)$$

$$\nabla_\theta J(\theta) = E_\pi[G_t \nabla_\theta \ln \pi(A_t|S_t,\theta)]$$

$$\theta_{t+1} = \theta_t + \alpha G_t \nabla_\theta \ln \pi(A_t|S_t,\theta)$$

Generated by playing game

Output of our N.N.

Algorithm Overview

Initialize deep N.N. to model agent's policy

Repeat for large number of episodes:

Generate episode using policy, keep track of probabilities

For each step in the agent's memory:

Calculate the return G for the episode

$$\theta_{t+1} = \theta_t + \alpha G_t \nabla_\theta \ln \pi(A_t|S_t, \theta)$$

One step at a time; will come back to this for review

Conclusion

- Don’t need distribution of states
- Update rule in terms of known quantities


Created by:
Cristian Ibarra Santillán