23/03/23

# Lecture 03
## Actor critic methods in Deep RL

Basic policy gradient: $\nabla_\theta \log \pi(a_t|s_t;\theta)\, G_t$ — called $R_t$ in the A3C paper

Difficult to distinguish between $G_t$ when results are similar, more stable gradients:

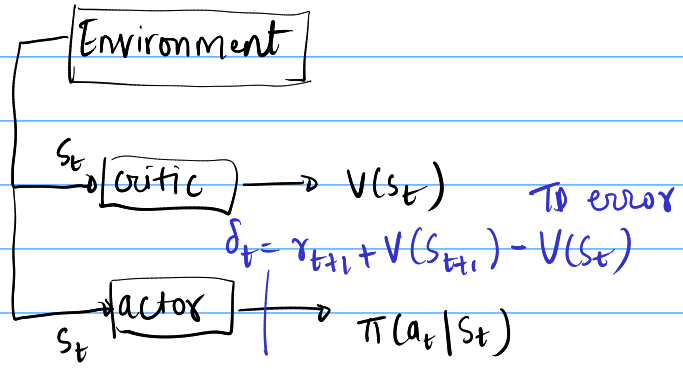
$$\nabla_\theta \log \pi(a_t|s_t;\theta)\big(G_t - b_t(s_t)\big)$$

Advantage

For us, $\theta$ → parameters (weights, biases) of a neural network

$\nabla_\theta$ → gradient ascent

Two implementation options

① 


* In both cases, one can use a common backbone for actor & critic


②