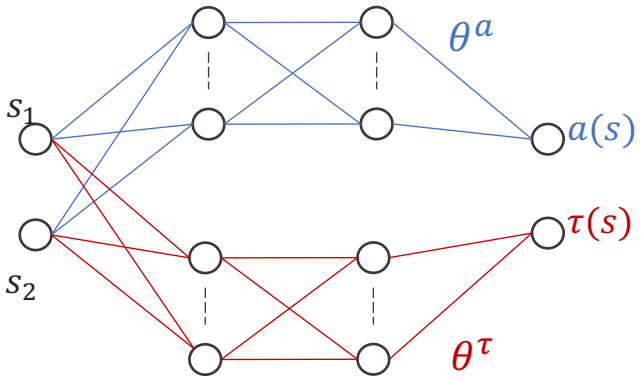
Weekly Report

M2 Ibuki Takeuchi

- Last week
  - Preparation for seminar
  - Calculate policy gradient



- Analytical calculation
  - Use dummy parameter θ to discuss what θ^a, θ^τ have in common
  - θ^π is a tuple of θ^a, θ^τ
  - $$\nabla_\theta V^{\theta^\pi}(s) = \nabla_\theta [r\left(s, \pi(s|\theta^\pi)\right) + \gamma V^{\theta^\pi}\left(s'(\theta^\pi)\right)]$$
    $$= \nabla_\theta r\left(s, \pi(s|\theta^\pi)\right) + \gamma \nabla_\theta \left\{V^{\theta^\pi}\left(s'(\theta^\pi)\right)\right\}$$
    $$= \nabla_\theta r\left(s, \pi(s|\theta^\pi)\right) + \gamma \nabla_\theta s'(\theta^\pi) \nabla_{s'} V^{\theta^\pi}(s')|_{s'=s'(\theta^\pi)}$$

    1st step's term

    $$+\gamma \nabla_\theta V^{\theta^\pi}(s')|_{s'=s'(\theta^\pi)}$$

    recursive term

Weekly Report

M2 Ibuki Takeuchi

- Detail calculation of $\nabla_{\theta^\tau} V^{\theta^\pi}(s)$
  - $s'(\theta^\pi) = s'(\theta^\pi|s)$: next state ($\tau$ second after)
  - $\nabla_{\theta^\tau} r\big(s, \pi(s|\theta^\pi)\big) = \nabla_{\theta^\tau}\tau(s|\theta^\tau)\{\underbrace{-s'(\theta^\pi)^T Q s'(\theta^\pi) - a(s|\theta^\pi)^T R a(s|\theta^\pi) + \lambda}_{\nabla_\tau r\left(s, \pi(s|\theta^\pi)\right)}\}$
  - $\nabla_{\theta^\tau} s'(\theta^\pi) = \nabla_{\theta^\tau}\tau(s|\theta^\tau)\{\underbrace{f\big(s'(\theta^\pi)\big) + g(s'(\theta^\pi))a(s|\theta^a))}_{\nabla_\tau s'(\theta^\pi)}\}$
  - $\nabla_{\theta^\tau} V^{\theta^\pi}(s) = \nabla_{\theta^\tau}\tau(s|\theta^\tau)\{$
    $- s'(\theta^\pi)^T Q s'(\theta^\pi) - a(s|\theta^\pi)^T R a(s|\theta^\pi) + \lambda$
    $+ \gamma(f\big(s'(\theta^\pi)\big) + g(s'(\theta^\pi))a(s|\theta^a)))\nabla_{s'} V^{\theta^\pi}(s')|_{s'=s'(\theta^\pi)}\}$
    $+ \gamma\nabla_\theta V^{\theta^\pi}(s')|_{s'=s'(\theta^\pi)}$
  - Considering the equation
    - Is there region in parameter space where policy gradient steep?

Weekly Report

M2 Ibuki Takeuchi

- Strategy
  - If there is the reason, why the policy gradient suddenly changed, in $\nabla_{\theta^\tau} V^{\theta^\pi}(s), \nabla_{\theta^a} V^{\theta^\pi}(s)$
    - Ahead for master thesis(?)
  - If not, it means the reason above is in DDPG's approximation
    - For master thesis, compare $\nabla_{\theta^\tau} V^{\theta^\pi}(s), \nabla_{\theta^a} V^{\theta^\pi}(s)$ for theoretical hyper parameter settings