CS294 HW2

ktian08

November 2019

1

$\sum_{t=1}^{T} E_{\tau \sim p_\theta(\tau)}[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)] =$

$\sum_{t=1}^{T} E_{\tau \sim p_\theta(\tau/s_t,a_t|s_t,a_t)} E_{(s_t,a_t)\sim p_\theta(s_t,a_t)}[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)] =$

$\sum_{t=1}^{T} E_{\tau \sim p_\theta(\tau/s_t,a_t|s_t,a_t)} \int\int p_\theta(s_t,a_t)[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)]\, da_t\, ds_t =$

$\sum_{t=1}^{T} E_{\tau \sim p_\theta(\tau/s_t,a_t|s_t,a_t)} \int\int p_\theta(s_t)\pi_\theta(a_t|s_t)[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)]\, da_t\, ds_t =$

$\sum_{t=1}^{T} E_{\tau \sim p_\theta(\tau/s_t,a_t|s_t,a_t)} \int p_\theta(s_t)(\int \pi_\theta(a_t|s_t)[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)]\, da_t)\, ds_t =$

$\sum_{t=1}^{T} E_{\tau \sim p_\theta(\tau/s_t,a_t|s_t,a_t)} \int p_\theta(s_t) b(s_t)(\int \pi_\theta(a_t|s_t)[\nabla_\theta \log \pi_\theta(a_t|s_t)]\, da_t)\, ds_t =$

$\sum_{t=1}^{T} E_{\tau \sim p_\theta(\tau/s_t,a_t|s_t,a_t)} \int p_\theta(s_t) b(s_t)(\int [\nabla_\theta \pi_\theta(a_t|s_t)]\, da_t)\, ds_t =$

$\sum_{t=1}^{T} E_{\tau \sim p_\theta(\tau/s_t,a_t|s_t,a_t)} \int p_\theta(s_t) b(s_t) \nabla_\theta(\int [\pi_\theta(a_t|s_t)]\, da_t)\, ds_t =$

$\sum_{t=1}^{T} E_{\tau \sim p_\theta(\tau/s_t,a_t|s_t,a_t)} \int p_\theta(s_t) b(s_t) \nabla_\theta 1\, ds_t =$

0. QED.

2

2.1

Markov assumption

2.2

$\sum_{t=1}^{T} E_{\tau \sim p_\theta(\tau)}[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)] =$

$\sum_{t=1}^{T} E_{s_t \sim p_\theta(s_t)} E_{a_t \sim \pi_\theta(a_t|s_t)}[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)] =$

$\sum_{t=1}^{T} \int p_\theta(s_t)(\int \pi_\theta(a_t|s_t)[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)]\, da_t)\, ds_t =$

$\sum_{t=1}^{T} \int p_\theta(s_t) b(s_t)(\int \pi_\theta(a_t|s_t)[\nabla_\theta \log \pi_\theta(a_t|s_t)]\, da_t)\, ds_t =$

$\sum_{t=1}^{T} \int p_\theta(s_t) b(s_t) (\int [\nabla_\theta \pi_\theta(a_t|s_t)]\, da_t)\, ds_t =$

$\sum_{t=1}^{T} \int p_\theta(s_t) b(s_t) \nabla_\theta (\int [\pi_\theta(a_t|s_t)]\, da_t)\, ds_t =$

$\sum_{t=1}^{T} \int p_\theta(s_t) b(s_t) \nabla_\theta 1\, ds_t =$

0. QED.