Assignment 1

Thursday, 14 October 2021 3:38 PM

1) $q^*(s,\pi(s)) - \tilde{q}(s,\pi(s))$ (using $q_\pi(s,a) = r(s,a) + \gamma \sum_{s\in S} P(s'|s,a)\, v_\pi(s')$)

$= \gamma \sum_{s'\in S} P(s'|s,a)\,(v^*(s') - v_\pi(s'))$ ($\because v^*(s') - v_\pi(s') \le \max(v^*(s') - v_\pi(s')) = \|v^* - v_\pi\|_\infty$)

$\le \gamma \sum P(s'|s,a)\, \|v^* - v_\pi\|_\infty$

$\le \gamma \|v^* - v_\pi\|_\infty$ ($\because \sum P(s'|s,a) = 1$) — ①

$\tilde{q}(s,\pi(s)) - q^*(s,\pi(s)) \le \|q^* - \tilde{q}\|_\infty$ — ②

$\tilde{q}(s,\pi(s)) \ge \tilde{q}(s,\pi^*(s))$ ($\because \pi$ is greedy policy w.r.t $\tilde{q}$)

$\Rightarrow q^*(s,\pi^*(s)) - \tilde{q}(s,\pi(s)) \le q^*(s,\pi^*(s)) - \tilde{q}(s,\pi^*(s))$
$\le \|q^* - \tilde{q}\|_\infty$ — ③

Now, from ② & ③

$v^*(s) - q^*(s,\pi(s)) = q^*(s,\pi^*(s)) - \tilde{q}(s,\pi(s)) + \tilde{q}(s,\pi(s)) - q^*(s,\pi(s))$

$\le \|q^* - \tilde{q}\|_\infty + \|q^* - \tilde{q}\|_\infty$

$\le 2\|q^* - \tilde{q}\|_\infty$ — ④

Now, from ① & ④

$v^*(s) - v_\pi(s) = v^*(s) - q^*(s,\pi(s)) + q^*(s,\pi(s)) - \tilde{q}(s,\pi(s))$

$\le 2\|q^* - \tilde{q}\|_\infty + \gamma\|v^* - v_\pi\|_\infty$

H/P

2) $J^{\pi_{old}}(\pi_{old}) = D_{KL}\left(\pi_{old}(x|s_t) \,\Big\|\, \frac{\exp(Q^{\pi_{old}}(s_t,x))}{Z^{\pi_{old}}(s_t)}\right)$

$= \int \pi_{old}(x|s_t)\left[\log \pi_{old}(x|s_t) - Q^{\pi_{old}}(s_t,x) + \log Z^{\pi_{old}}(s_t)\right] dx$

$= -V^{\pi_{old}}(s_t) + \underbrace{\int \log Z^{\pi_{old}}(s_t)\, dx}_{\text{gets cancelled}}$

$J^{\pi_{old}}(\pi_{new}) = D_{KL}\left(\pi_{new}(x|s_t) \,\Big\|\, \frac{\exp(Q^{\pi_{old}}(s_t,x))}{Z^{\pi_{old}}(s_t)}\right)$

$= \int \pi_{new}(x|s)\left[\log \pi_{new}(x|s_t) - Q^{\pi_{old}}(s_t,x) + \log Z^{\pi_{old}}(s_t)\right] dx$

$= \int \pi_{new}(x|s)\left[\log \pi_{new}(x|s_t) - Q^{\pi_{old}}(s_t,x)\right] dx + \underbrace{\int \log Z^{\pi_{old}}(s_t)\, dx}_{\text{gets cancelled}}$

$J^{\pi_{old}}(\pi_{new}) \le J^{\pi_{old}}(\pi_{old})$

$\Rightarrow \int \pi_{new}(x|s)\left[\log \pi_{new}(x|s_t) - Q^{\pi_{old}}(s_t,x)\right] dx \le -V^{\pi_{old}}(s_t)$

$\Rightarrow V^{\pi_{old}}(s_t) \le \int \pi_{new}(x|s_t)\left[Q^{\pi_{old}}(s_t,x) - \log \pi_{new}(x|s_t)\right] dx$ — ①

Now, using Bellman eq^{ns} & (using ①)

$Q^{\pi_{old}}(s,a) = r(s,a) + \gamma \sum_{s'\in S} P(s'|s,a)\, V^{\pi_{old}}(s')$

$\le r(s,a) + \gamma \sum_{s'\in S} P(s'|s,a)\left(\int \pi_{new}(x|s')\left[Q^{\pi_{old}}(s',x) - \log \pi_{new}(x|s')\right] dx\right)$

$= Q^{\pi_{new}}(s,a) + \gamma \sum_{s'\in S} P(s'|s,a)\left(\int \pi_{new}(x|s')\left[Q^{\pi_{old}}(s',x) - Q^{\pi_{new}}(s',x)\right] dx\right)$

$\Rightarrow Q^{\pi_{old}}(s,a) - Q^{\pi_{new}}(s,a) \le \gamma E\left(Q^{\pi_{old}}(s',x) - Q^{\pi_{new}}(s',x)\right)$ — ②

$\forall\, s, a$

E → expection wrt $s' \in S$ (→ state) & $x \in A$ (→ action)

Now after application of E again on both sides

$\Rightarrow E\left(Q^{\pi_{old}}(s,a) - Q^{\pi_{new}}(s,a)\right) \le \gamma E\left(Q^{\pi_{old}}(s,a) - Q^{\pi_{new}}(s,a)\right)$

$\Rightarrow (1-\gamma)\, E\left(Q^{\pi_{old}}(s,a) - Q^{\pi_{new}}(s,a)\right) \le 0$

$\Rightarrow E\left(Q^{\pi_{old}}(s,a) - Q^{\pi_{new}}(s,a)\right) \le 0$ ($\because \gamma < 1$)

$\Rightarrow \gamma E\left(Q^{\pi_{old}}(s,a) - Q^{\pi_{new}}(s,a)\right) \le 0$ ($\because \gamma > 0$)

$\Rightarrow Q^{\pi_{old}}(s,a) - Q^{\pi_{new}}(s,a) \le 0$ (using ②)

$\Rightarrow Q^{\pi_{old}}(s,a) \le Q^{\pi_{new}}(s,a)\ \forall\, s, a$

H/P