به نام خدا

یادگیری تقویتی

نیم‌سال بهار ۴۰۲ـ۱۴۰۱

اساتید: دکتر رهبان، آقای حسنی

دانشکده‌ی مهندسی کامپیوتر

مدت: ۲۰ دقیقه | کوییز ۶ | تاریخ: ۷ خرداد

نام و نام خانوادگی : شماره دانشجویی:

سوال ۱: Offline Policy

(آ) دو تا از چالش‌های Offline Policy را نام برده و توضیح دهید.

(ب) می‌توان با استفاده از KL-divergence یک محدودیت روی پالیسی گذاشت اما این کار در عین اینکه پیاده سازی راحتی دارد یک مشکلی دارد، آن مشکل را توضیح دهید.

(ج) در روش Implicit Q-learning (IQL) توضیح دهید چگونه value function ها به روزرسانی می‌شوند بدون اینکه درگیر مشکلات Offline Policy شویم؟

(د) ایده روش Conservative Q-learning (CQL) را توضیح دهید.

سوال ۲: Unsupervised Rl

در مقاله DIAYN هدف بیشینه کردن Mutual Information بین استیت‌ها و اسکیل‌ها هست که به صورت زیر تعریف می‌شود:

$I(S, Z) = H(Z) - H(S|Z)$

اما در مقاله DADS هدف بیشینه کردن رابطه زیر است :

$I(S', Z \mid S) = H(S'|S) - H(S'|S,Z)$

رابطه بالا رو توضیح دهید و بیان کنید این روش چه مزیتی نسبت به روش DIAYN دارد. همچنین بیان کنید چه تابع پاداشی در روش DADS وجود دارد.