



**سوال ۱: RL as Inference**

تفاوت Soft Q-learning با Q-learning معمولی در چیست؟ روابط آپدیت پالیسی و استیت ها را بنویسید و دو مزیت Soft Q-learning را بیان نمایید.

**سوال ۲: Exploration in RL**

(آ) چگونه می توان از یک Hash Function در روش های Exploration by counting استفاده کرد؟ این کار چه مزیت هایی ایجاد می کند؟ آیا Hash Function باید ویژگی خاصی نسبت به Hash Function های معمولی داشته باشد؟ یک روش برای ساخت این Hash Function را نام ببرید

(ب) با یک مثال یک مزیت Bootstrapped DQN الگوریتم را توضیح دهید.

**سوال ۳: Imitation Learning**

(آ) الگوریتم DAgger را توضیح دهید.

(ب) اگر اجرای یک پالیسی منجر به یک اتفاق برگشت نا پذیر شود چه راه حلی را پیشنهاد می کنید؟ و این راه حل چه مشکلی دارد؟

**سوال ۴: Inverse RL**

(آ) مشکل underspecified در Inverse RL را توضیح دهید.

(ب) چگونه می توان Inverse RL را با کمک شبکه های متخاصم مدل کرد؟