

سوال یک- معادلات بلمن چیست و چگونه در حل MDP مورد استفاده قرار می گیرد؟

سوال دو- با جستجو در اینترنت و خواندن درباره الگوریتم SARSA، تفاوت آن را با Q-learning شرح دهید.

سوال سه- تریدآف میان استخراج و اکتشاف (exploration-exploitation) چیست و در RL چگونه به دنبال حل کردن آن هستیم؟