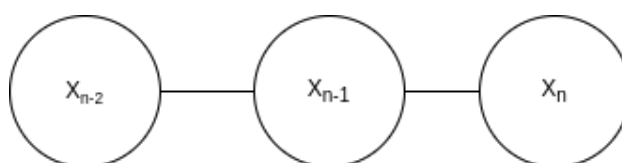




مساله ۱.

الف) نادرست. به دو دلیل:

۱. تعداد زیادی از حالت‌ها در درخت جستجو تکرار میشوند. ۲. عمق درخت میتواند بینهایت شود.
- ب) نادرست. در واقع X_n میتواند به X_{n-2} وابسته باشد چون یک مسیر زنجیره‌ای بین آنها وجود دارد. طبیعتاً X_{n-2} نیز وابسته به X_n است. (طبق مطالب یادگرفته در bayesian network ها)



شکل ۱: زنجیره مارکوف

ج) درست. چون در reinforcement learning امکان بهینه عمل کردن از ابتدا وجود ندارد (چون در RL تابع ریوارد R) و احتمال گذر T بین حالات را نمی‌دانیم) و یادگیری به صورت آنلاین صورت می‌گیرد با استفاده از تعریف معیار پیشیمانی که تفاوت میانگین ریوارد سیاست بهینه و میانگین ریوارد کسب شده است میزان ریوارد از دست داده را اندازه‌گیری می‌کنیم.

مساله ۲.

با توجه به اینکه تنها منبع دریافت اطلاعات ما در این مساله دوربین نصب شده در جلوی هواپیما است پس استیت، آرایه‌ی اعداد پیکسل‌های دوربین خواهد بود. مثلاً اگر خروجی دوربین یک عکس 200×200 باشد استیت، یک آرایه $3 \times 200 \times 200$ از اعداد بین ۰ تا ۲۵۵ خواهد بود زیرا هر پیکسل شامل سه عدد rgb است. ۵ نوع اکشن داریم:

- حرکت رو به جلو
- چرخش ۹۰ درجه به راست
- چرخش ۹۰ درجه به چپ
- چرخش ۴۵ درجه به راست
- چرخش ۴۵ درجه به چپ

تابع reward نیز به این گونه در نظر می‌گیریم:

- برخورد به دیوار ۱ - دریافت می‌کند.
- حرکت مستقیم بدون برخورد به دیوار ۱ + دریافت می‌کند.
- حرکت چرخشی بدون برخورد به دیوار ۰/۲ + دریافت می‌کند.

- در صورت انجام حرکت چرخشی در صورتی که حرکت قبلی نیز حرکتی چرخشی بوده باشد $0/2$ - دریافت می کند. (برای جلوگیری از چرخش مداوم به دور خود)

* پاسخ این سوال محدود به موارد گفته شده نیست و هر جواب درست دیگر نیز قابل قبول است.