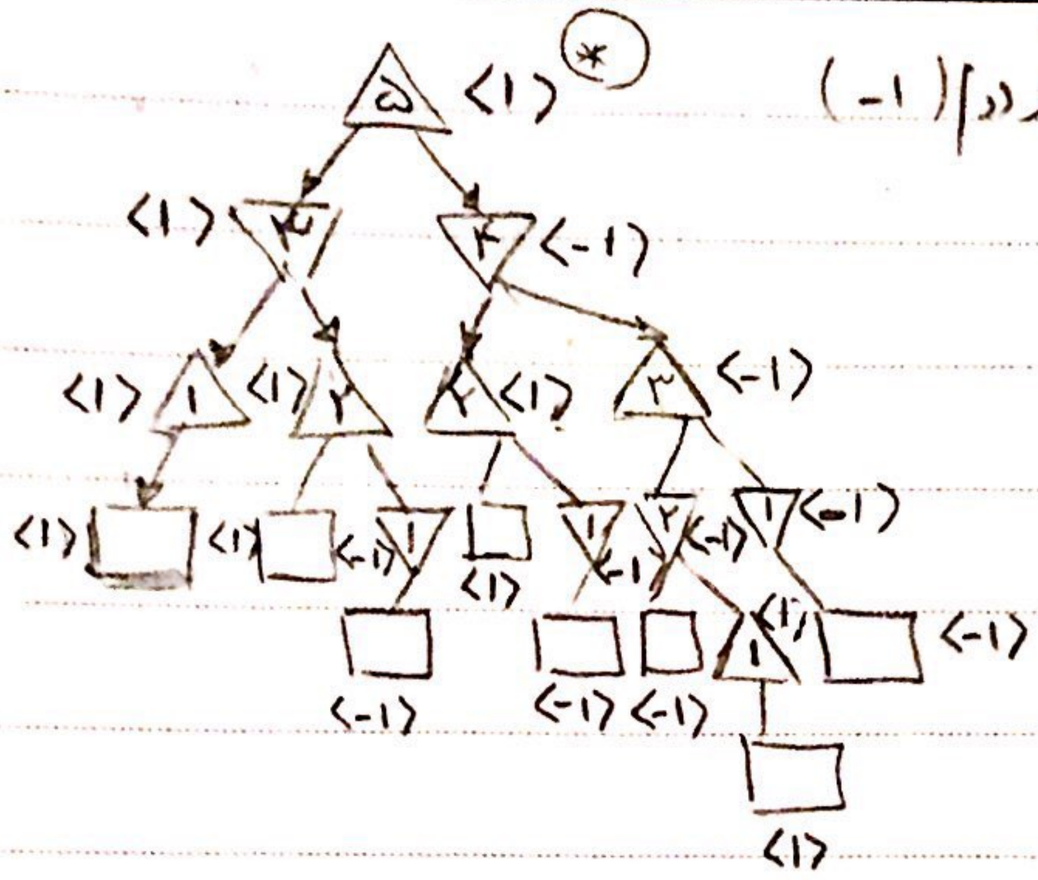
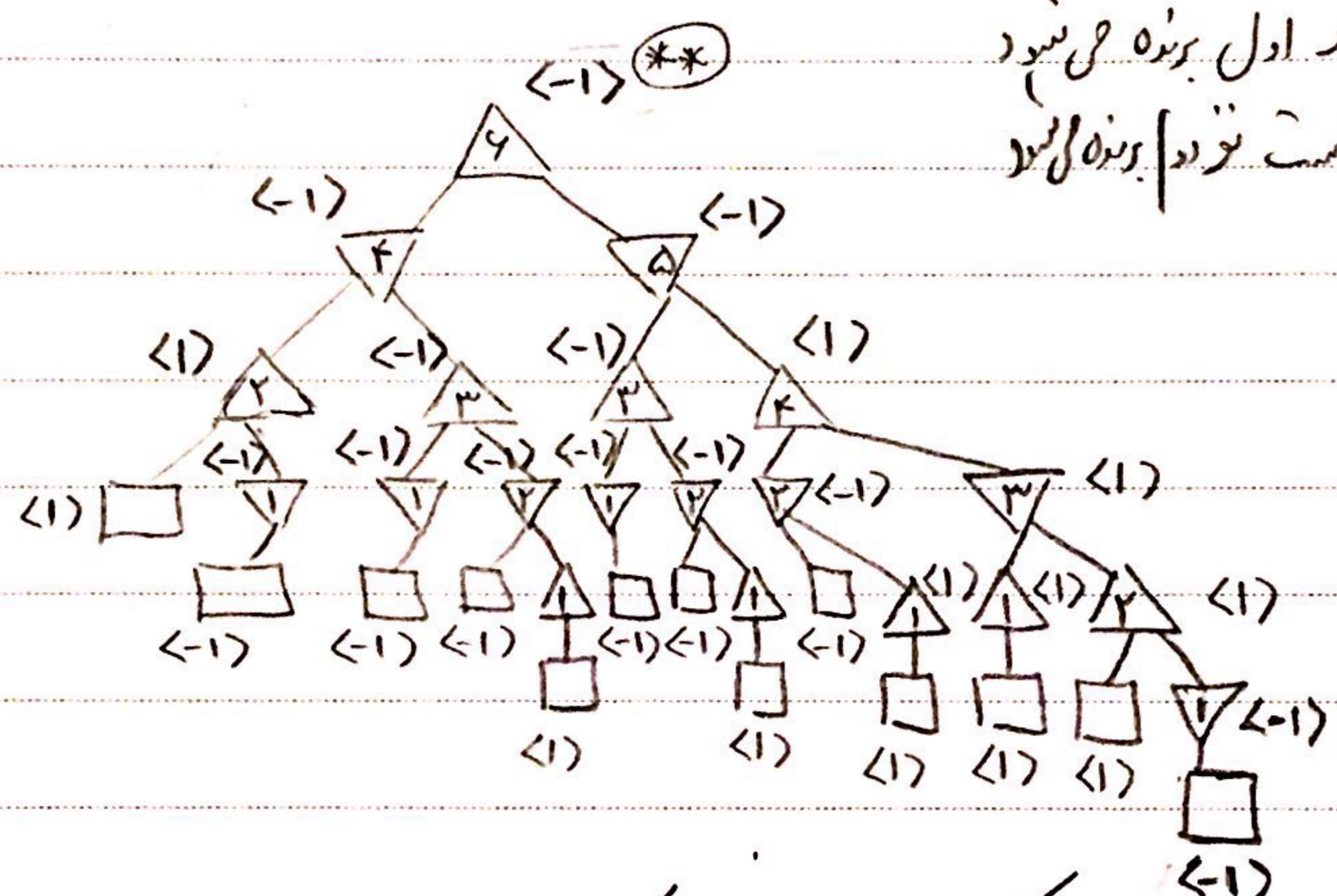
سوال ۱) آ) امتیازهای هرکدام برحسب برد نفر اول (۱) یا نفر دوم (۱-) به ترتیب به صورت ⟨۱⟩ و ⟨۱-⟩ مشخص شده اند و امتیاز داخل هر state بیانگر عدد مثبت طبیعی مسئله است



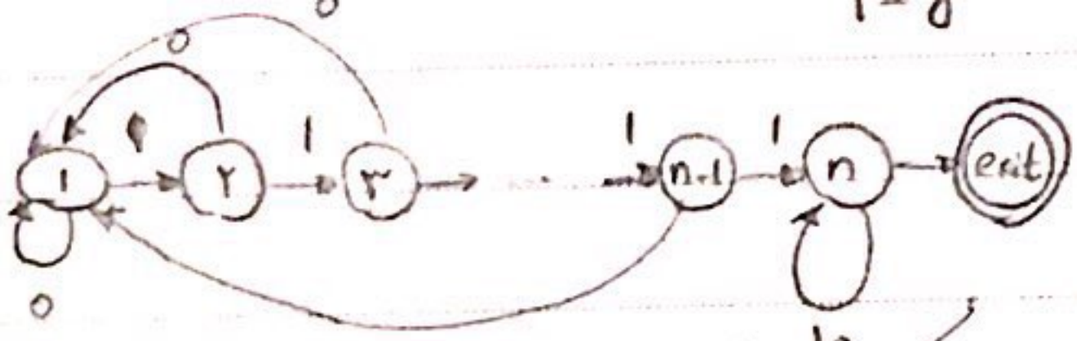
* همانطور که مشخص است نفر اول برنده می شود



** همانطور که مشخص است نفر دوم برنده می شود

ج) با کمی دقت به بازی می توان متوجه شد که اعدادی وجود دارند که هر مرد در تلاش برای رسیدن به آن (پس از ۵ هشتن) هستند تا بتوانند نفر مقابل را ببرند. برای مثال اگر نفر اول بتواند با مقداری کاهش عدد را به ۳ برساند برد قطعی است لذا در صورتی که عدد نوبت او ۴ یا ۵ باشد او می تواند با کاهش عدد به ۳ برنده شود. حال اگر عدد نوبت او ۶ باشد او می تواند عدد را به ۴ یا ۵ کاهش دهد که در هر صورت نفر دوم می تواند با کاهش عدد به ۳ برنده شود.
لذا استراتژی برد این گونه است که هر مرد تلاش کند تا با کاهش عدد به عددی مضرب ۳ شرایط برد را برای خود فراهم کند
⇐ اگر عدد ابتدای کار مضرب ۳ باشد نفر دوم تحت بازی ایده برنده می شود و در غیر این صورت نفر اول برنده می شود

Scanned with CamScanner



سوال ۲) الف)

همانطور که از شکل بالا مشخص است در هر مرحله جایزه حاصل از برگشت به خانه ① برابر صفر است ولی در ادامه در صورت بازگشت با جایزه های یک بیشتری در مسیر مواجه خواهیم شد که ضریب γ ها می گیرند و این ضریب به صورت توان کاهش می یابد این در حالی است که رفتن به وضعیت n و ماندن در این وضعیت جایزه ۱۰ را به ما می دهد که با توجه به جایزه های یک قبلی منطقی است هر چه سریع تر خود را به وضعیت n برسانیم و در این وضعیت بمانیم. هدف از هر چه سریع تر رسیدن به وضعیت n این است که هر مرحله γ باعث کاهش جایزه می شود و به نوعی بهتر است زودتر به جایزه ۱۰ برسیم.

لذا policy مورد نظر این است که در مراحل ۱ تا $n-1$ به اندازه شماره بعد برویم و در حالت n در این وضعیت بمانیم. بهینه بودن این policy با توجه به توضیحات داده شده مبنی بر تلاش برای رسیدن به جایزه ۱۰ در سریع ترین حالت ثابت می شود

ب)

$$V^*(n) = 10 + \gamma V^*(n) \Rightarrow V^*(n) = \frac{10}{1-\gamma} = 20$$

$$V^*(n-1) = 1 + \gamma V^*(n)$$

$$V^*(n-2) = 1 + \gamma V^*(n-1) = 1 + \gamma + \gamma^2 V^*(n) \Rightarrow V^*(n-i) = \sum_{j=0}^{i-1} \gamma^j + \gamma^i V^*(n)$$

$i \le n-i$

$$\Rightarrow V^*(i) = \sum_{j=0}^{n-i-1} \gamma^j + \gamma^{n-i} V^*(n) = \frac{1-\gamma^{n-i}}{1-\gamma} + \gamma^i \times 20 = 2 - \left(\frac{1}{2}\right)^{n-i-1} + \left(\frac{1}{2}\right)^{n-i} \times 20$$

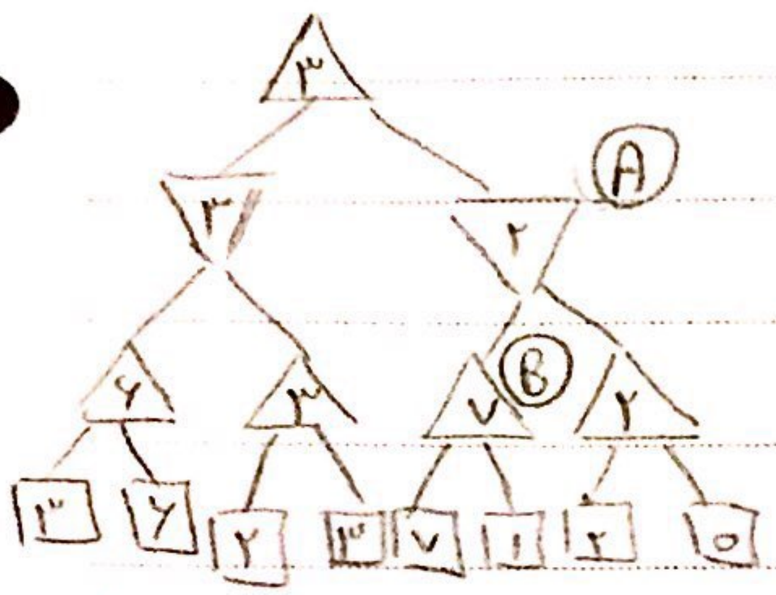
PAPCO

Scanned with CamScanner

Subject :

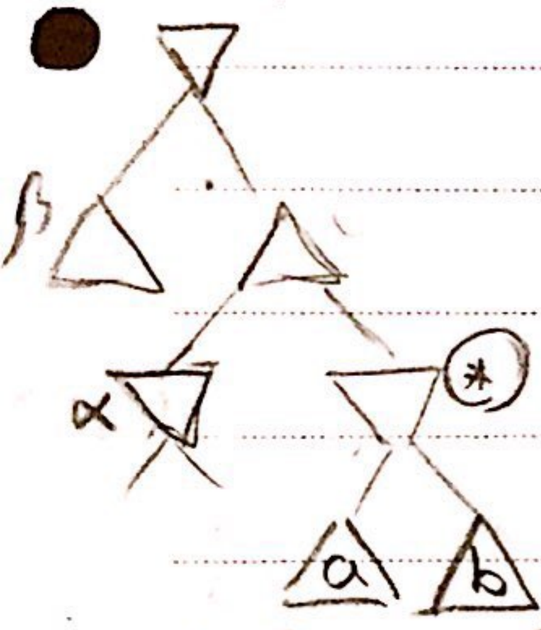
Year . Month . Date . ( )

سوال ۳) آ) ابتدا درخت بازی را تکمیل می‌کنیم



بهترین امتیازی که می‌تواند این بازیکن ببرد ۷ است به این گونه که از نقطه گره (A) را انتخاب می‌کند و سپس با ریشه دادن به گره‌ی موجب این می‌شود که سمت گره (B) را انتخاب کند و لذا از این ۷ امتیاز برنده می‌شود که لذا از آن دلار برنده شده توسط او با احتساب هزینه قمار برابر ۵ - ۲ = ۷ خواهد بود.

ب) اکنون که هزینه ریشه به ۵ دلار افزایش یافته است می‌توان فهمید که دلار شیخ با احتساب رسیدن به بهترین امتیاز با کمک یک ریشه برابر ۲ = ۵ - ۷ خواهد شد که از ذلار این در حالت عادی کمتر است لذا بهتر است با توجه به قیمت جدید ریشه آن به صورت سالم بازی کند و ۳ دلار برنده شود.



ج) در حالت عادی برای حرص آلفا بتا کاری برد بلال قبل حالت بدون

این بیشتر

$\alpha \geq \beta \Rightarrow$ (*) is Pruned

اما در این سوال امکان دارد بتوان با ریشه دادن برای قبل هر حالت (*) به جای $\min(a,b)$ ، $\max(a,b)$ را انتخاب کرد ولی pruning به این نحوه برداشت حالت (*) را prune می‌کند که این کار موجب از دست رفتن امکان رسیدن به بهترین دلار با توجه به امکان ریشه دادن می‌شود

Scanned with CamScanner

Subject:

Year. Month. Date. ( )

سوال ۴) الف)

$$Q(s,a) = \sum_{s'} T(s,a,s')\,(R(s,a,s') + \gamma V(s'))$$

$$\Rightarrow Q(s,a) = R(s,a,s') + \gamma V(s')$$

چون داریم در این سوال $T(s,a,s')=1$ پس نیازی نداریم

از آن جا که برای تعیین خانه‌ی بعدی S، policy مشخص است داریم:

$$Q(S, left) = 100 + \gamma V(s')$$

$$V(s') = R(s, left, s'') + \gamma V(s'')$$

$$\vdots$$

$$\Rightarrow Q(S, left) = 100 + \gamma(-1 + \gamma(-1 + \gamma(\dots))) = 100 - \gamma - \gamma^2 - \dots - \gamma^{98}$$

$$= 100 - \gamma(1 + \dots + \gamma^{97}) \simeq 100 - \frac{\gamma}{1-\gamma}$$

$$Q(S, down) = 80 + \gamma(1 + \gamma(1 + \gamma(\dots))) = 80 + \gamma + \dots + \gamma^{98} \simeq 80 + \frac{\gamma}{1-\gamma}$$

$$Q(S, down) > Q(S, left) \Rightarrow \frac{\gamma}{1-\gamma} > 20 \Rightarrow 24\gamma > 20 \Rightarrow \gamma > \frac{20}{24}$$

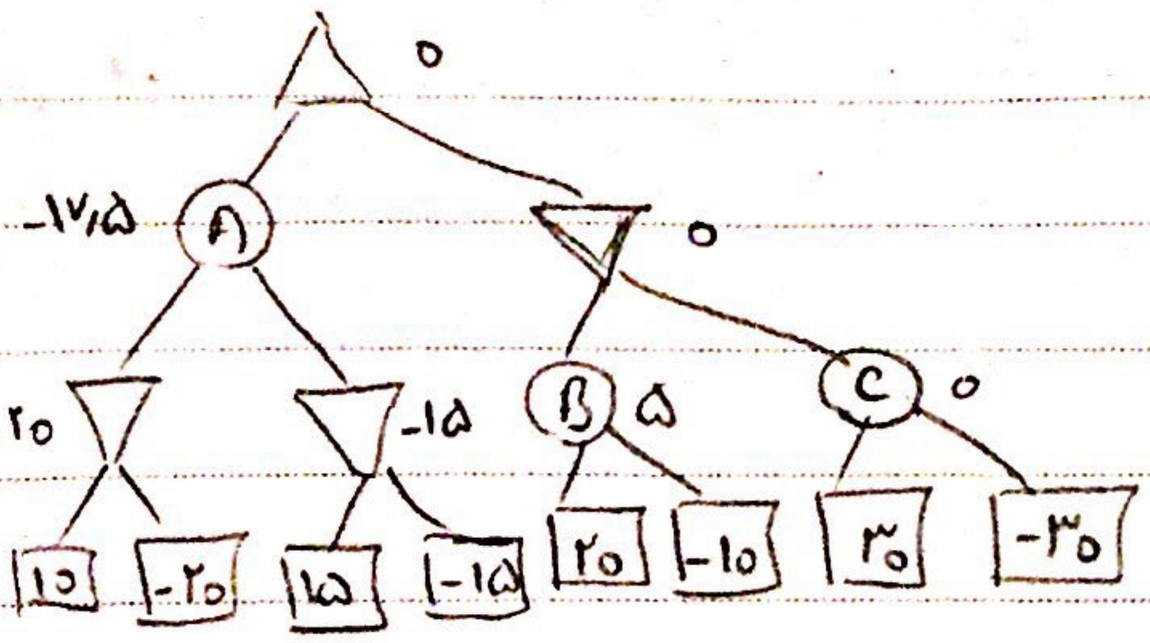
(ب

P4PCO

Scanned with CamScanner

Subject:

Year. Month. Date. ( )



سوال ۵) درخت صحنه‌ی بخاطی را تکمیل میکنیم

$E[B] = ۲۰ \times \frac{۱}{۲} + (-۱۰) \times \frac{۱}{۲} = ۵$

$E[C] = ۳۰ \times \frac{۱}{۲} + (-۳۰) \times \frac{۱}{۲} = ۰$

$E[A] = -۲۰ \times \frac{۱}{۲} + (-۱۵) \times \frac{۱}{۲} = -۱۷/۵$

$\Rightarrow E[root] = ۰$

ب) به ازای هر حالت معلوم بودن هر سلله مهره (مقابل) کدام مهره را می‌بایم

	چپ	راست
A	۰	۰
B	۰	-۱۰
C	۵	-۱۷/۵

با توجه به جدول اهمیت داشتن C بسیار زیاد است چراکه اگر C سمت راست را انتخاب کند نوادل حاصل A را انتخاب می‌کند و اگر C چپ باشد نوادل همان شاخه سمت راست را انتخاب می‌کند و امتیاز او ۵ می‌شود. لذا نتیجه C در انتخاب نوادل مؤثر خواهد بود این درحالی است که اطلاع از نتیجه A برای مثال تأثیری در انتخاب نوادل ندارد چراکه او همواره شاخه راست را انتخاب می‌کند زیرا میانگین آن راست از (A) بیشتر است ۵ > -۱۷/۵

ج) مشخص است بین حالت حالت‌های [راست یا چپ = C یا B یا A] $E[root]$ بهترین حالت برای میانگین رئیس

$E[root \mid C = \text{چپ}] = ۵$

است لذا مشخص است نتیجه سلله C برابر حالت چپ بیشترین منفعت را برای او دارد

PAPCO

Scanned with CamScanner