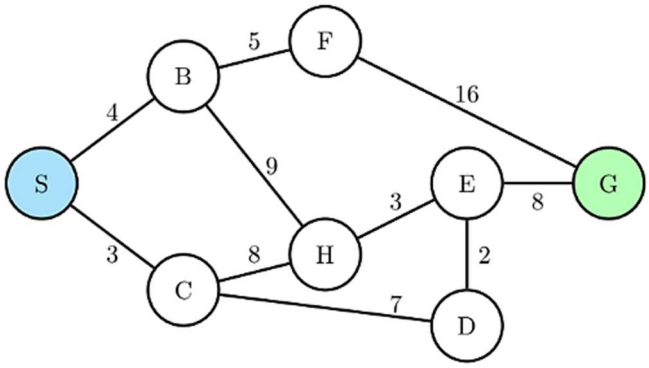
امتحان پایان ترم هوش مصنوعی	دانشکده ریاضی و علوم کامپیوتر
نام و نام خانوادگی.....	دانشگاه صنعتی خواجه نصیرطوسی
شماره دانشجویی.....	مدت زمان: ۱۲۰ دقیقه

۱. گراف فضای حالت زیر را در نظر بگیرید، که در آن **S**حالت شروع و **G**حالت هدف است. هزینه‌ی هر یال روی گراف داده شده است و یال‌ها دوطرفه هستند. با استفاده از جدول هیوریستیک (heuristic) ارائه‌شده، الگوریتم‌های زیر را روی گراف داده‌شده اجرا کنید، درخت جستجو را ترسیم کرده و صفِ frontierو explored را بسازید:

Node	h
S	14
B	12
C	11
D	6
E	?
F	11
G	0
H	6



الف) جستجوی اول-سطح Breadth-First Searchیا BFS

ب) جستجوی با هزینه یکنواخت Uniform Cost Searchیا UCS

ج) الگوریتم A* در حالتی که h(E)= 6 باشد

د) به ازای چه مقادیری برای h(E) هیوریستیک داده شده هم admissible و هم consistent است؟

الف) Breadth-First Search (BFS)

Expanded order: S → B → C → F → H → D → G

Returned path: S – B – F – G

Total path cost: 25

Step-by-step frontier/explored:

Step	Expand	Frontier	Explored
1	S	B,C	S
2	B	C,F,H	S,B
3	C	F,H,D	S,B,C
4	F	H,D,G	S,B,C,F
5	H	D,G,E	S,B,C,F,H
6	D	G,E	S,B,C,F,H,D
7	G	E	S,B,C,F,H,D,G

ب) Uniform-Cost Search (UCS)

Expanded order: S → C → B → F → D → H → E → G

Returned path: S – C – D – E – G

Total path cost: 20

Step-by-step frontier/explored:

Step	Expand	Frontier	Explored
1	S	3:C; 4:B	S
2	C	4:B; 10:D; 11:H	S,C
3	B	9:F; 10:D; 11:H	S,C,B
4	F	10:D; 11:H; 25:G	S,C,B,F
5	D	11:H; 12:E; 25:G	S,C,B,F,D
6	H	12:E; 25:G	S,C,B,F,D,H
7	E	20:G; 25:G	S,C,B,F,D,H,E
8	G	25:G	S,C,B,F,D,H,E,G

ج) A* Search (with h(E) = 6)

Expanded order: S → C → D → B → H → E → G

Returned path: S – C – D – E – G

Total path cost: 20

Step-by-step frontier/explored:

Step	Expand	Frontier	Explored
1	S	14:C; 16:B	S
2	C	16:D; 16:B; 17:H	S,C
3	D	16:B; 17:H; 18:E	S,C,D
4	B	17:H; 18:E; 20:F	S,C,D,B
5	H	18:E; 20:F	S,C,D,B,H
6	E	20:G; 20:F	S,C,D,B,H,E
7	G	20:F	S,C,D,B,H,E,G

د)

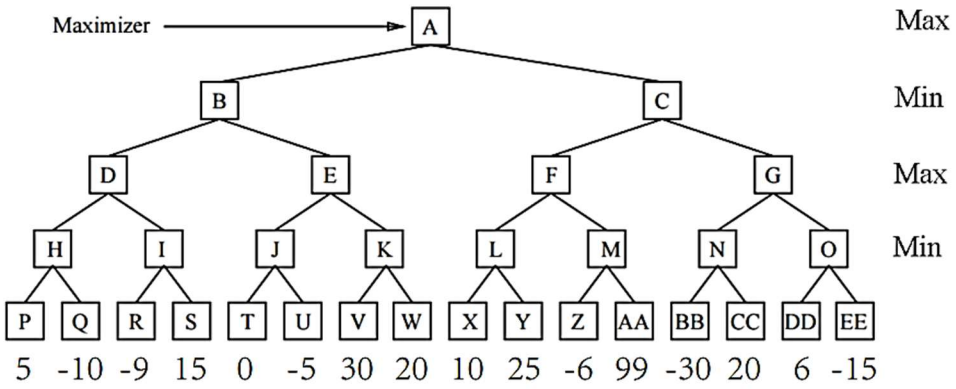
Let h*(E) be the true optimal cost from E to G. The cheapest path from E to G is the direct edge E–G with cost 8, so admissibility requires h(E) ≤ 8.

For consistency, for every edge (u, v) we require h(u) ≤ c(u, v) + h(v). Considering the edges touching E gives the tightest constraints:

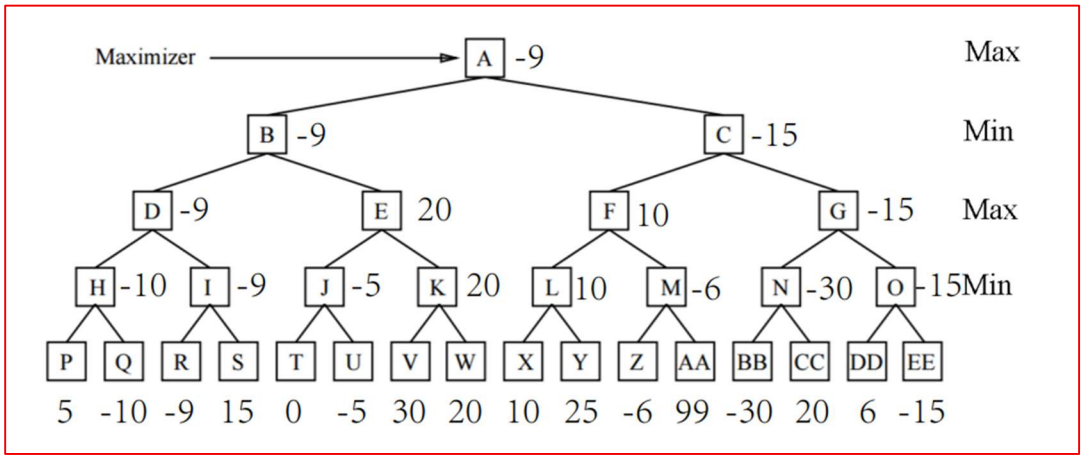
From E to D: h(E) ≤ 2 + h(D) = 8 → h(E) ≤ 8.
From D to E: h(D) = 6 ≤ 2 + h(E) → h(E) ≥ 4.
From E to H: h(E) ≤ 3 + h(H) = 9 (looser than 8).
From H to E: h(H) = 6 ≤ 3 + h(E) → h(E) ≥ 3 (looser than 4).
From E to G: h(E) ≤ 8 + h(G) = 8.
Therefore the range that is both admissible and consistent is: 4 ≤ h(E) ≤ 8.

---

۲. درخت مین‌مکس زیر را در نظر بگیرید. گره ها با A تا EE نامگذاری شده اند و مقادیر برگ ها در آخرین ردیف نشان داده شده است.



الف) در گره ریشه باید چه حرکتی انجام شود (چپ یا راست)؟ مقدار آن چقدر است؟ مراحل اعمال الگوریتم مین‌مکس را در ادامه نشان دهید و برای هر گره مقدار مشخص را بنویسید. (به‌عنوان مثال، P = 3، ... Q = −10)



ب) اگر از هرس آلفا-بتا از چپ به راست استفاده کنیم کدام نودها هرس میشوند؟ برای هر گره مقادیر آلفا و بتا را بنویسید.

Solution: Node: K, V, W, AA, CC

The value of α and β are $-\infty$ and -9 for the pruning of E-K. The value of α and β are 10 and $+\infty$ for the pruning of M-AA. The value of α and β are -9 and 10 for the pruning of N-CC.

ج) اگر در هنگام هرس از راست به چپ حرکت کنیم کدام نودها هرس میشوند؟

Solution: Node: L, P, T, X, Y

The value of α and β are -9 and 20 for the pruning of H-P. The value of α and β are 20 and $+\infty$ for the pruning of J-T. The value of α and β are $-\infty$ and -15 for the pruning of F-L.

---

۳. "سید امیر" یک ورزشکار حرفه‌ای است که در یک ورزش انفرادی بازی می‌کند. وضعیت امیر می‌تواند یا کاملاً آماده، نیمه‌آماده، یا مصدوم باشد. صرف‌نظر از وضعیتش، امیر می‌تواند انتخاب کند که در یک تورنمنت شرکت کند، زمانی را صرف تمرین کند یا تصمیم بگیرد که استراحت کامل داشته باشد. تیم مربی‌گری امیر یک **فرآیند تصمیم‌گیری مارکوف (MDP)** طراحی می‌کند تا وضعیت‌ها، کنش‌ها، پاداش‌ها و گذارها را دنبال کند. فرض می‌شود که ضریب تنزیل (gamma discount factor) برابر ۱ است (مگر اینکه طور دیگری بیان شود) و درست قبل از تورنمنت بزرگ بعدی، امیر در وضعیت کاملاً آماده قرار دارد. تیم به مدل زیر برای امیر می‌رسد:

(الف) پاداش‌ها برای زوج‌های (وضعیت، کنش):

وقتی کاملاً آماده است:

- اگر امیر تصمیم بگیرد بازی کند ⟶ پاداش +۱۰۰
- اگر تصمیم بگیرد تمرین کند ⟶ پاداش -۱۰
- اگر تصمیم بگیرد استراحت کند ⟶ پاداش ۰

وقتی نیمه‌آماده است:

- اگر تصمیم بگیرد بازی کند ⟶ پاداش +۲۰
- اگر تصمیم بگیرد تمرین کند ⟶ پاداش -۱۰
- اگر تصمیم بگیرد استراحت کند ⟶ پاداش -۲۰

وقتی مصدوم است:

- اگر تصمیم بگیرد بازی کند ⟶ پاداش -۶۰
- اگر تصمیم بگیرد تمرین کند ⟶ پاداش -۳۰
- اگر تصمیم بگیرد استراحت کند ⟶ پاداش ۰

(ب) احتمال‌های گذار:

وقتی کاملاً آماده است:

- اگر تصمیم بگیرد بازی کند ← ٪۸۰ احتمال ماندن در حالت کاملاً آماده، ٪۲۰ احتمال مصدوم شدن.
- اگر تصمیم بگیرد تمرین کند ← ٪۹۰ احتمال ماندن در حالت کاملاً آماده، ٪۱۰ احتمال مصدوم شدن.
- اگر تصمیم بگیرد استراحت کند ← ٪۵۰ احتمال ماندن کاملاً آماده، ٪۵۰ احتمال نیمه‌آماده شدن.

وقتی نیمه‌آماده است:

- اگر تصمیم بگیرد بازی کند ← ٪۵۰ احتمال ماندن نیمه‌آماده، ٪۵۰ احتمال مصدوم شدن.
- اگر تصمیم بگیرد تمرین کند ← ٪۴۰ احتمال ماندن نیمه‌آماده، ٪۶۰ احتمال کاملاً آماده شدن.
- اگر تصمیم بگیرد استراحت کند ← همیشه نیمه‌آماده باقی می‌ماند.

وقتی مصدوم است:

- اگر تصمیم بگیرد بازی کند ← همیشه مصدوم باقی می‌ماند.
- اگر تصمیم بگیرد تمرین کند ← همیشه مصدوم باقی می‌ماند.
- اگر تصمیم بگیرد استراحت کند ← ٪۵۰ احتمال ماندن در حالت مصدوم، ٪۵۰ احتمال نیمه‌آماده شدن.

الف) درحالت "horizon 1" فقط پاداش فوری اهمیت دارد و آینده در نظر گرفته نمی‌شود. بنابراین در هر وضعیت باید کنشی انتخاب شود که بیشترین پاداش همان لحظه را بدهد. سیاست بهینه horizon 1 برای امیر چیست؟

$$\pi_1^*(\text{fully fit}) = \text{play}$$
$$\pi_1^*(\text{partially fit}) = \text{play}$$
$$\pi_1^*(\text{injured}) = \text{break}$$

ب) برای horizon 2، وقتی امیر نیمه‌آماده است، بهترین کنش چیست و پاداش مورد انتظار horizon 2 برای انجام آن بهترین کنش در حالت نیمه‌آماده چقدر است؟ فرض کنید ضریب تنزیل ۱ است. نشان دهید چگونه به پاسخ رسیده‌اید.

$$
\begin{aligned}
Q_2(\text{partially fit}, \text{play}) &= R(\text{partially fit}, \text{play}) + \sum_{s'} T(\text{partially fit}, \text{play}, s') \max_{a'} Q_1(s', a') \\
&= 20 + (0.5 * 20 + 0.5 * 0) \\
&= 30 \\
Q_2(\text{partially fit}, \text{train}) &= R(\text{partially fit}, \text{train}) + \sum_{s'} T(\text{partially fit}, \text{train}, s') \max_{a'} Q_1(s', a' \\
&= -10 + (0.4 * 20 + 0.6 * 100) \\
&= 58 \\
Q_2(\text{partially fit}, \text{rest}) &= R(\text{partially fit}, \text{rest}) + \sum_{s'} T(\text{partially fit}, \text{rest}, s') \max_{a'} Q_1(s', a') \\
&= -20 + (1 * 20) \\
&= 0 \\
\pi_2^*(\text{partially fit}) &= \arg\max_a Q_2(\text{partially fit}, \text{a}) \\
&= \text{train}
\end{aligned}
$$

ج) اگر ضریب تنزیل ۰.۵ باشد، آیا پاسخ بخش (ب) تغییر می‌کند ؟ نشان دهید چرا یا چرا نه.

Solution: Yes it changes. The new best Q(partially fit, a) for all actions a is now **play** because the discount of 0.5 ensures that the large expected reward on step is not enough to overcome the negative immediate negative reward of training in **partially fit** state.

$$
\begin{aligned}
Q_2(\text{partially fit}, \text{play}) &= R(\text{partially fit}, \text{play}) + \delta \sum_{s'} T(\text{partially fit}, \text{play}, s') \max_{a'} Q_1(s', a') \\
&= 20 + 0.5(0.5 * 20 + 0.5 * 0) \\
&= 25 \\
Q_2(\text{partially fit}, \text{train}) &= R(\text{partially fit}, \text{train}) + \delta \sum_{s'} T(\text{partially fit}, \text{train}, s') \max_{a'} Q_1(s', a') \\
&= -10 + 0.5(0.4 * 20 + 0.6 * 100) \\
&= 24 \\
Q_2(\text{partially fit}, \text{rest}) &= R(\text{partially fit}, \text{rest}) + \delta \sum_{s'} T(\text{partially fit}, \text{rest}, s') \max_{a'} Q_1(s', a') \\
&= -20 + 0.5(1 * 20) \\
&= -10 \\
\pi_2^*(\text{partially fit}) &= \arg\max_a Q_2(\text{partially fit}, \text{a}) \\
&= \text{play}
\end{aligned}
$$

د) سیاست بهینه infinite horizon برای امیر چیست؟ فرض کنید ضریب تنزیل ۱ است.

$$
\begin{aligned}
\pi_\infty^*(\text{fully fit}) &= \text{play} \\
\pi_\infty^*(\text{partially fit}) &= \text{train} \\
\pi_\infty^*(\text{injured}) &= \text{break}
\end{aligned}
$$

ه) آیا سیاستی وجود دارد که در افق بی‌نهایت پاداش مورد انتظار را بیشینه کند و طبق آن امیر در حالت مصدوم باید بازی کند؟ توضیح دهید.

Solution: No there isn't. Both other actions have a negative reward and they both keep **Amir** in the injured state.

و) سهیل ورزشکار دیگری است که همان ورزش سید امیر را انجام می‌دهد. مدل MDP سهیل دقیقاً مانند مدل سید امیر است، به جز اینکه تیم سهیل پاداش مربوط به *بازی کردن* در حالت کاملاً آماده را فراموش کرده‌اند. همچنین تیم سهیل به یاد دارند که بهترین کنش در حالت horizon 2 در حالت نیمه‌آماده دقیقاً همان است که برای سید امیر به‌دست آمده (در بخش ب).

با توجه به این اطلاعات، بازه مقادیر ممکن برای R(ready to play, play) برای سهیل چیست؟ فرض کنید ضریب تنزیل برابر ۱ است.

If the horizon 2 optimal policy in state partially fit is the same for both amir and soheil, then π*2(partially fit) = train. This implies that Q2(partially fit, train) > Q2(partially fit, play) and Q2(partially fit, train) > Q2(partially fit, rest).

Therefore, we have:

$$Q_2(\text{partially fit, train}) > Q_2(\text{partially fit, play})$$

$$R(\text{partially fit, train}) + \sum_{s'} T(\text{partially fit, train}, s') \max_{a'} Q_1(s', a') >$$

$$R(\text{partially fit, play}) + \sum_{s'} T(\text{partially fit, play}, s') \max_{a'} Q_1(s', a')$$

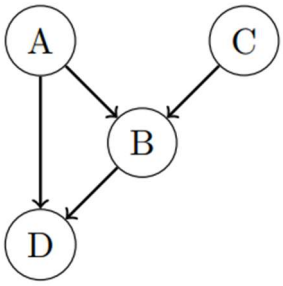
$$-10 + (0.4 * 20 + 0.6 * \text{R(fully fit, play)}) > 20 + (0.5 * 20 + 0.5 * 0)$$

$$0.6 * \text{R(fully fit, play)} > 32$$

$$R(\text{fully fit, play}) > \frac{32 * 10}{6} = 53.33.$$

---

۴. شبکه بیزی نشان‌داده‌شده در نمودار زیر را در نظر بگیرید.



الف) تمام استقلال‌های شرطی که توسط این نمودار شبکه بیزی برقرار شده‌اند را انتخاب کنید.

- ☐ A ⊥ C | B
- ☐ D ⊥ C | A, B
- ☐ D ⊥ C
- ☐ A ⊥ C
- ☐ A ⊥ C | D
- ☐ A ⊥ C | B, D
- ☐ D ⊥ C | B

ب) به‌دلیل این استقلال‌های شرطی، برخی توزیع‌ها نمی‌توانند توسط این شبکه بیزی نمایش داده شوند. کمینه مجموعهٔ یال‌هایی که باید اضافه شوند تا شبکه بیزی حاصل بتواند هر توزیعی را نمایش دهد چیست؟

Either (C → A AND C → D) OR (A → C AND C → D)

ج)در ادامه، برخی جدول‌های احتمال شرطی نیمه‌پر شده برای متغیرهایA ، B، C و D آورده شده است. توجه داشته باشید که این جدول‌ها لزوماً عوامل شبکه بیزی نیستند. شش خانه خالی را پر کنید به طوری که این توزیع بتواند توسط شبکه بیزی نمایش داده شود.

A	C	$P(C \mid A)$
$+a$	$+c$	0.8
$+a$	$-c$	0.2
$-a$	$+c$	0.8
$-a$	$-c$	0.2

A	B	D	$P(D \mid A, B)$
$+a$	$+b$	$+d$	0.60
$+a$	$+b$	$-d$	0.40
$+a$	$-b$	$+d$	0.10
$+a$	$-b$	$-d$	0.90
$-a$	$+b$	$+d$	0.20
$-a$	$+b$	$-d$	0.80
$-a$	$-b$	$+d$	0.50
$-a$	$-b$	$-d$	0.50

A	B	C	$P(C \mid A, B)$
$+a$	$+b$	$+c$	0.50
$+a$	$+b$	$-c$	0.50
$+a$	$-b$	$+c$	0.20
$+a$	$-b$	$-c$	0.80
$-a$	$+b$	$+c$	0.90
$-a$	$+b$	$-c$	0.10
$-a$	$-b$	$+c$	0.40
$-a$	$-b$	$-c$	0.60

C	$P(C)$
$+c$	**(i)**
$-c$	**(ii)**

A	B	C	D	$P(D, C \mid A, B)$
$+a$	$+b$	$+c$	$+d$	**(iii)**
$+a$	$+b$	$-c$	$-d$	**(iv)**
$+a$	$-b$	$+c$	$+d$	**(v)**
$+a$	$-b$	$-c$	$-d$	**(vi)**
⋮	⋮	⋮	⋮	⋮

(i): 0.8 (ii): 0.2 (iii): 0.6 * 0.5 = 0.3 (iv): 0.4 * 0.5 = 0.2 (v): 0.1 * 0.2 = 0.02 (vi): 0.9 * 0.8 = 0.72

موفق باشید :))