



امتحان پایان‌ترم

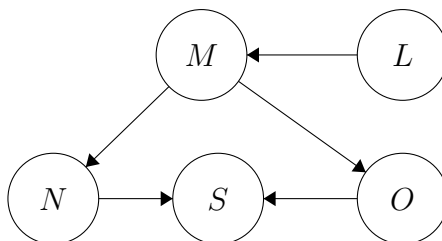
- زمان در نظر گرفته شده برای امتحان ۱۲۰ دقیقه است.
- لطفاً پاسخ‌های خود را به صورت خوانا و خوش خط بنویسید.
- پاسخ هر سوال باید در یک برگه جداگانه نوشته شود. در بالای هر برگه پاسخ‌نامه، نام و شماره دانشجویی خود را به صورت واضح بنویسید. دقت کنید که برگه‌های هر سوال برای تصحیح از سایر برگه‌ها جدا خواهند شد، در نتیجه مسئولیت عدم دریافت نمره در اثر نوشتن پاسخ یک سوال در کنار سوال دیگر و یا نبود مشخصات بر روی برگه بر عهده خودتان خواهد بود.
- در صورت به همراه داشتن Cheatsheet مشخصات خود را روی آن نوشته و همراه پاسخ‌برگ و صورت سوال‌ها تحویل دهید.

سوالات (۵۰ نمره)

۱. (۱۰ نمره)

سوالات کوتاه پاسخ

(آ) شبکه بیزین زیر در شکل ۱ را در نظر بگیرید.



شکل ۱: شکل سوال ۱ قسمت آ

فرض کنید که می‌خواهیم با استفاده از تعداد محدودی نمونه، دو احتمال $p(M=1, O=1 | L=0)$ و $P(M=1, O=1 | S=1)$ را به دست آوریم. محاسبه یکی از این دو باید به کمک Likelihood Weighting و دیگری به کمک Rejection Sampling انجام شود. برای آنکه دقت بیشتری را در محاسبه هر دو کوثری داشته باشیم، کدام یک از احتمالات باید به کمک Likelihood Weighting انجام شود؟ دلیل آن را بیان کنید. توجه کنید که L حالت Deterministic ندارد.

(ب) اگر در مسئله‌ی دسته‌بندی متون ضریب Smoothing را بسیار بالا ببریم، احتمال منتسب شدن هر متن به هر دسته چه تغییری می‌کند؟

(ج) تعداد پارامترهای شبکه عصبی ۴ لایه (۲، ۵، ۷، ۴) که ورودی آن اندازه ۱۰ دارد را بدست آورید. در هر لایه این شبکه Bias داریم.

(د) آیا می‌توان از Q-Learning بدون تغییر برای مسائلی با تعداد حالات نامحدود استفاده کرد؟

حل.

(آ) کوثری دوم. زیرا $s+$ در برگ‌های شبکه بیزین است و در نتیجه ممکن است تمام سمپلینگ انجام شود و در انتها رد شود. با استفاده از Likelihood Weighting می‌توان آن مشاهده را ثابت کرد و بقیه را نمونه گرفت.

(ب) با افزایش ضریب Smoothing احتمال انتساب هر متن به هر کلاس به احتمال Prior نزدیک‌تر می‌شود. اگر ضریب خیلی بالا برود، احتمال انتساب به همان احتمالات Prior میل می‌کند.

(ج) تعداد پارامترها در لایه ۱: $44 = 4 + 4 \times 10$

تعداد پارامترها در لایه ۲: $35 = 7 + 4 \times 7$

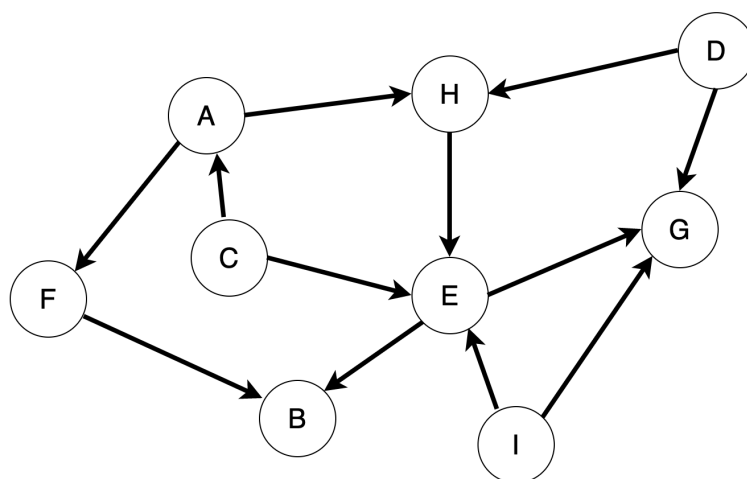
تعداد پارامترها در لایه ۳: $40 = 5 + 7 \times 5$

تعداد پارامترها در لایه ۴: $12 = 2 + 5 \times 2$

در نتیجه تعداد کل پارامترهای شبکه برابر با ۱۳۱ است. توجه کنید که سوال نگفته شبکه عصبی ۴ لایه نهان دارد بلکه کلاً ۴ لایه دارد. در نتیجه نباید لایه اضافی‌ای برای خروجی در نظر گرفته بشود. تنها بخش اندکی از نمره در صورت در نظر گرفتن لایه اضافی کم شده است.

(د) خیر، Q-Learning نیاز به تعداد حالات محدود دارد تا بتوان مقادیر Q را به طور تجربی محاسبه کرد.

۲. (۱۰ نمره) با توجه به شبکه بیزین زیر در شکل ۲، درستی یا نادرستی عبارات زیر را با ذکر دلیل مشخص کنید.

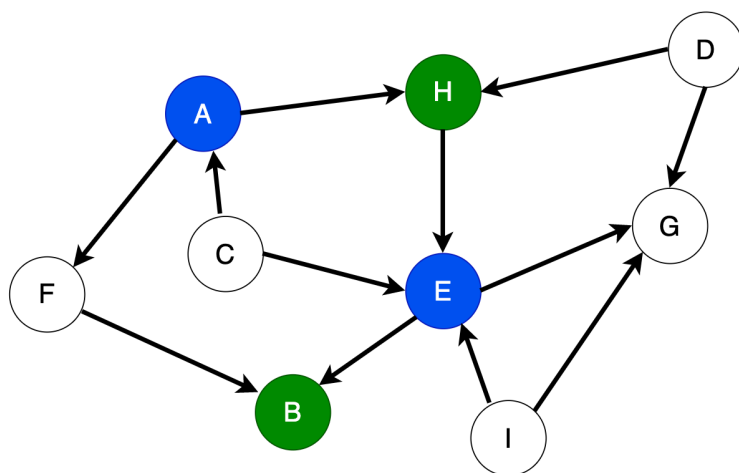


شکل ۲: شکل سوال ۲

- (آ) $B \perp\!\!\!\perp H | A, E$
 (ب) $A \perp\!\!\!\perp I | C, G$
 (ج) $F \perp\!\!\!\perp D | A, G$

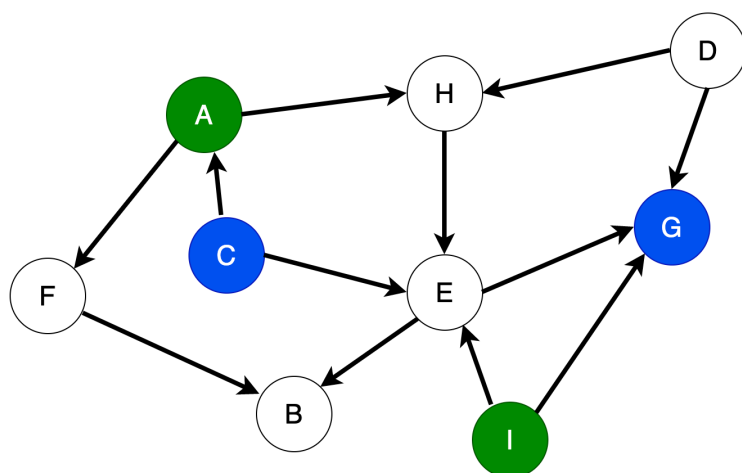
حل.

(آ) به دلیل وجود ۳ تایی‌های غیرفعال $C \rightarrow A \rightarrow H$ و $F \leftarrow A \rightarrow H$ ، $C \rightarrow A \rightarrow F$ تمام مسیرهای H به B که از A می‌گذرند غیرفعال هستند. همین‌طور به دلیل وجود ۳ تایی‌های غیرفعال $H \rightarrow E \rightarrow B$ ، $I \rightarrow E \rightarrow B$ و $B \leftarrow E \rightarrow G$ تمام مسیرهای B به H که از E می‌گذرند نیز غیرفعال هستند. بنابراین هیچ مسیر فعالی از B به H وجود ندارد و عبارت گفته شده درست است.



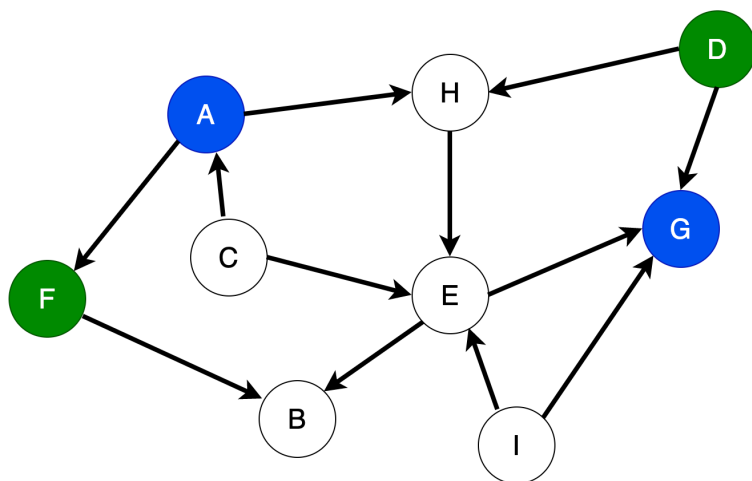
شکل ۳: شبکه بیزین قسمت آ

(ب) در مسیر $A \rightarrow H \rightarrow E \rightarrow G \leftarrow I$ تمامی ۳ تایی‌ها فعال هستند، در نتیجه یک مسیر فعال از A به I وجود دارد و عبارت گفته شده نادرست است.



شکل ۴: شبکه بیزین قسمت ب

(ج) به دلیل وجود ۳ تایی‌های غیرفعال $F \leftarrow A \rightarrow H$ و $C \rightarrow A \rightarrow F$ تمام مسیرهای F به D که از A می‌گذرند غیرفعال است. همچنین به دلیل وجود ۳ تایی غیرفعال $F \rightarrow B \leftarrow E$ تمام مسیرهای F به D که از B می‌گذرند غیرفعال است. در نتیجه تمامی مسیرهای F به D غیرفعال است و عبارت گفته شده درست است.



شکل ۵: شبکه بیزین قسمت ج

(آ) فرض کنید برای داده‌های جدول ۱ یک درخت تصمیم آموزش می‌دهیم تا X را به وسیله A, B, C پیش‌بینی کنیم. درصد خطای مدل پس از آموزش بر روی داده‌های آموزش چقدر خواهد بود؟

C	B	A	X
۰	۰	۰	۰
۱	۰	۰	۰
۱	۰	۰	۰
۰	۱	۰	۰
۱	۱	۰	۰
۱	۱	۰	۱
۱	۱	۰	۱
۰	۰	۱	۰
۱	۰	۱	۱
۰	۱	۱	۱
۰	۱	۱	۱
۱	۱	۱	۰
۱	۱	۱	۱

جدول ۱: داده‌های مدل درخت تصمیم سوال ۳

(ب) فرض کنید روی مجموعه‌ی داده‌ی دلخواهی، درخت تصمیمی برای دسته‌بندی بین k داده، آموزش می‌دهیم. حداکثر خطایی که ممکن است این مدل روی داده‌های آموزش داشته باشد چقدر خواهد بود؟ (پاسخ را به صورت کسری بنویسید)

حل.

(آ) در درخت تصمیم، در دادگان آموزش خطا تنها زمانی اتفاق می‌افتد که چند داده با فیچرهای یکسان وجود داشته باشد اما برچسب آن‌ها متفاوت باشد.

در این جا سه داده‌ی $(A = 0, B = 1, C = 1)$ داریم که برچسب دو تا از آن‌ها ۱ و دیگری صفر داده شده است. پس درخت مقدار ۱ را پیش‌بینی می‌کند و یک خطا به ازای این داده اتفاق می‌افتد. هم‌چنین دو داده‌ی $(A = 1, B = 1, C = 1)$ داریم که برچسب یکی صفر و دیگری یک است و درخت به صورت رندم یکی از این برچسب‌ها را برای پیش‌بینی انتخاب می‌کند که باعث می‌شود یک خطا هم در اینجا داشته باشیم.

پس درخت در پیش‌بینی دو تا از داده‌ها خطا دارد که درصد خطای آن برابر خواهد بود با: $\frac{2}{13} \times 100 \approx 15.4\%$

(ب) هدف: می‌خواهیم اثبات کنیم که زمانی که هیچ فیچر جداکننده‌ای نداریم، حداکثر خطا را داریم. فرض کنید که c تعداد کلاس‌ها و k تعداد داده‌ها باشد.

زمانی که مقدار فیچرهای تمام داده‌ها یکسان است: حداقل تعداد درست $\lceil \frac{k}{c} \rceil$ بوده و در نتیجه خطا $\frac{k - \lceil \frac{k}{c} \rceil}{k}$

اثبات را با استقرا انجام می‌دهیم:

پایه: $k = 2$

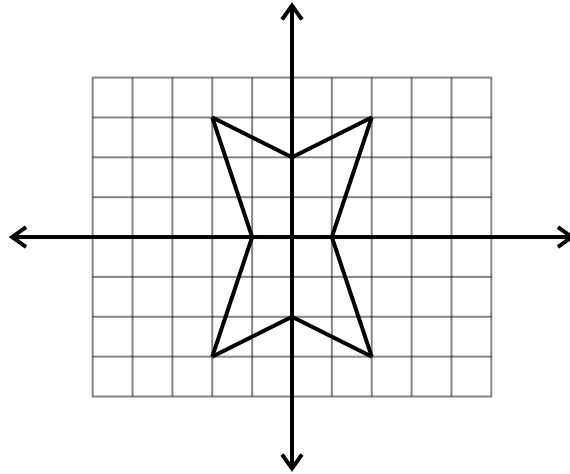
فرض: برای $k < n$ برقرار است.

حکم: $k = n$. برهان خلف: فرض می‌کنیم زمانی که فیچر جداکننده‌ای داریم و درخت براساس آن جدا می‌شود حداکثر خطا را داریم. در این صورت مثلاً اگر تعداد درست در دسته i ام $\beta_i < n$ باشد آن‌گاه تعداد درست $\lceil \frac{\beta_i}{c} \rceil \leq$ در نتیجه $\sum_i \lceil \frac{\beta_i}{c} \rceil \geq$ تعداد درست کل خواهد بود.

همچنین $\sum_i \lceil \frac{\beta_i}{c} \rceil \geq \lceil \frac{\sum_i \beta_i}{c} \rceil$
 زمانی که فیچر جدا کننده نداریم، حداقل تعداد درست $\lceil \frac{\sum_i \beta_i}{c} \rceil = \lceil \frac{n}{c} \rceil$
 در نتیجه این موضوع تناقض بوده و حکم ثابت می شود.

۴. (۱۰ نمره)

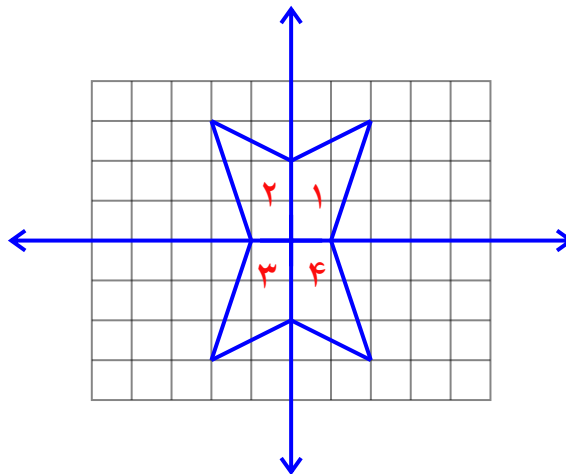
آیا می‌توان شبکه‌ای با دو ورودی x و y طراحی کرد به طوری که برای ورودی‌ای که در ناحیه داخلی شکل ۶ قرار بگیرد، خروجی ۱ بدهد و در غیر این صورت، صفر؟ ساختار شبکه و مقادیر وزن‌ها و بایاس‌ها را به طور دقیق مشخص کنید.

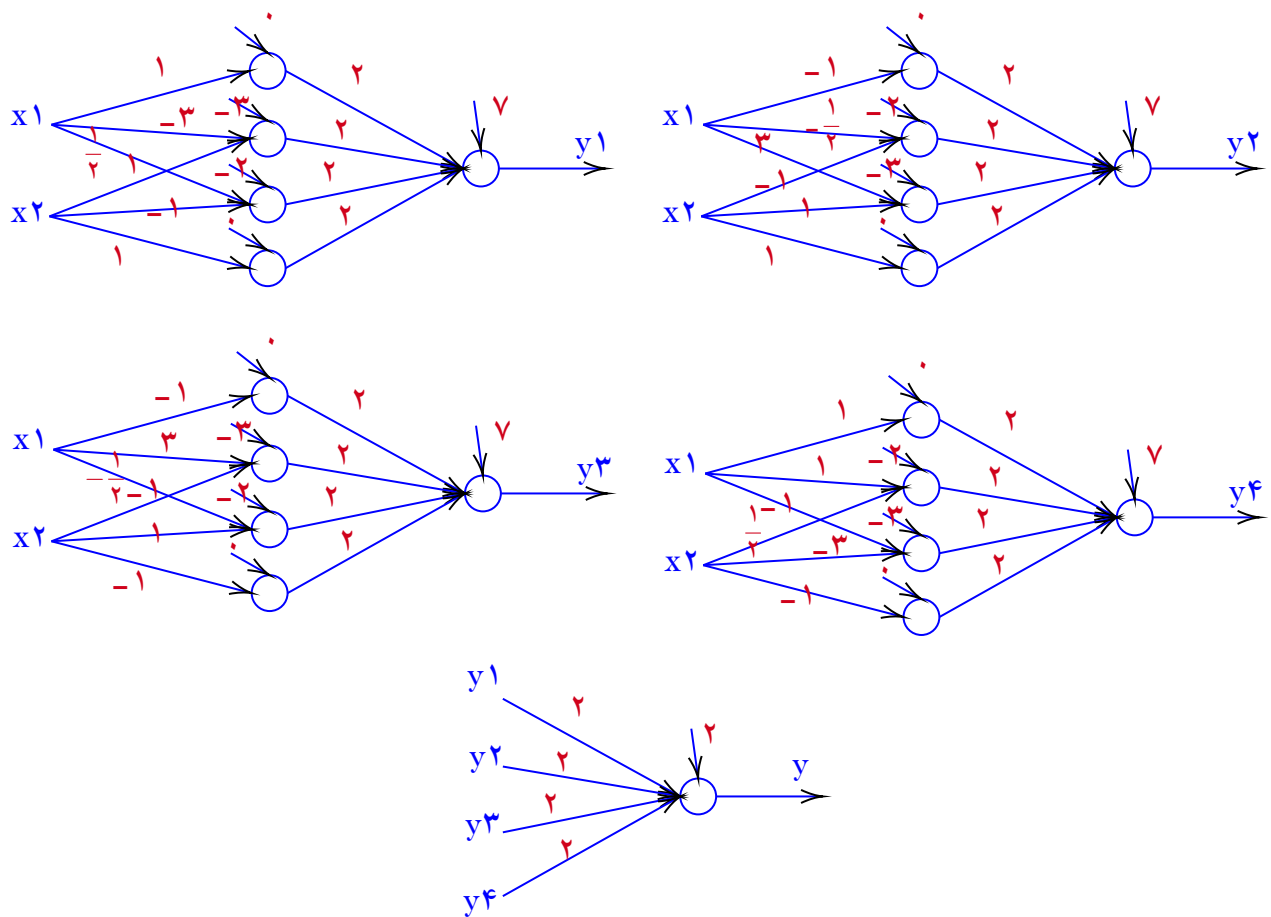


شکل ۶: شکل سوال ۴

حل.

شکل را به چهار ناحیه تقسیم می‌کنیم. برای هر ناحیه i ، شبکه عصبی با خروجی y_i را طوری تشکیل می‌دهیم که آن ناحیه را پوشش بدهد و در نهایت خروجی همه را OR می‌کنیم.

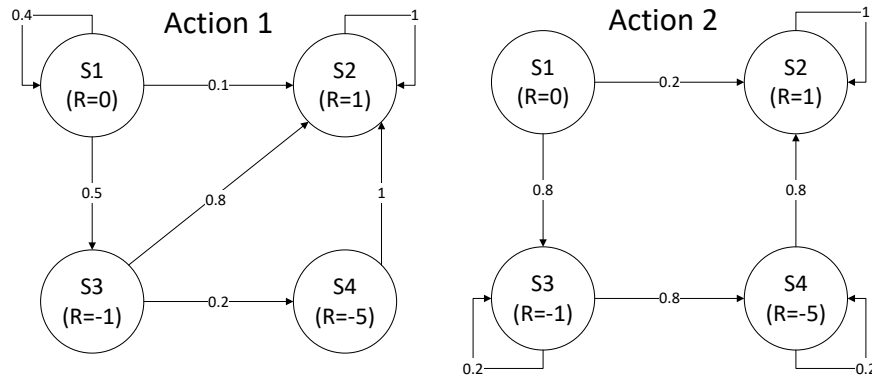




در این شکل‌ها منظور از x_1 و x_2 به ترتیب محور افقی و عمودی مختصات است.
 جواب بالا تنها یکی از جواب‌های ممکن برای این سوال است و به سایر جواب‌های درست هم نمره کامل داده شده است.

۵. (۱۰ نمره)

با توجه به Markov Decision Process در شکل ۷، مقادیر مربوط به هر حالت یا همان Value function را به دست آورید. توجه کنید که حالت‌ها دو بار تکرار شده‌اند (به ازای هر حرکت) تا از پیچیدگی شکل جلوگیری شود. روی هر یال احتمال تغییر از حالت مبدا به مقصد در صورت انتخاب action مربوطه نوشته شده‌است. همچنین پاداش هر حالت که بین دو action مقدار ثابتی است و در هنگام خروج از حالت دریافت می‌شود، روی هر حالت با استفاده از مقدار R نمایش داده شده‌است. (از $\gamma = 0.9$ استفاده کنید.)



شکل ۷: نمودار حالت Markov Decision Process سوال ۵

حل.

$$\begin{aligned}
 V(S_2) &= 1 + 0.9V(S_2) \rightarrow V(S_2) = 10 \\
 V(S_4) &= -5 + 0.9 \max(V(S_2), 0.2V(S_4) + 0.8V(S_2)) = -5 + 0.9 * 10 = 4 \\
 V(S_3) &= -1 + 0.9 \max(0.2V(S_4) + 0.8V(S_2), 0.2V(S_3) + 0.8V(S_4)) \\
 \begin{cases} V(S_3) &= -1 + 0.9(0.8 + 8) = 6.92 \\ V(S_3) &= -1 + 0.9(0.2V(S_3) + 0.8 * 4) \rightarrow V(S_3) < 3 \end{cases} \\
 \rightarrow V(S_3) &= 6.92 \\
 V(S_1) &= 0 + 0.9 \max(0.4V(S_1) + 0.1V(S_2) + 0.5V(S_3), 0.2V(S_2) + 0.8V(S_3)) \\
 \begin{cases} V(S_1) &= 0.9 * (2 + 5.54) = 6.78 \\ V(S_1) &= 0.36V(S_1) + 0.9 * 3.11 \rightarrow V(S_1) = 4.52 \end{cases} \\
 \rightarrow V(S_1) &= 6.78
 \end{aligned}$$

(۱)

توجه: اگر فرض کنید وقتی از s_i به s_j می‌رویم، مقدار پاداش عدد s_j است (یعنی پاداش را در هنگام ورود حساب کنیم) پاسخ‌های نهایی به صورت زیر خواهد بود:

$$V(S_1) = 6.78, V(S_2) = 10, V(S_3) = 6.92, V(S_4) = 4$$

در صورتی که این فرض ذکر شده باشد، به این پاسخ‌ها هم نمره کامل داده شده است.