به نام خداوند بخشنده مهربان

هوش مصنوعی

دانشکده مهندسی کامپیوتر

محمدحسین رهبان

بهار ۱۴۰۳

۳ تیر ۱۴۰۳، ساعت ۰۰ : ۹

آزمون پایان‌ترم

زمان آزمون: ۱۸۰ دقیقه

۱. لطفا پاسخ خود را با خط خوانا بنویسید.

۲. پاسخ هر سوال را در یک صفحه جدا و شماره پرسش را به صورت واضح در بالای هر صفحه بنویسید.

۳. نوشته‌های شما در قسمت چرک‌نویس یا برگه سوال به هیچ عنوان تصحیح نخواهد شد.

۴. استفاده از منابع و لوازم الکترونیکی حین پاسخگویی به سوالات آزمون غیرمجاز است.

۵. آزمون از ۱۱۰ نمره می‌باشد و دریافت ۱۰۰ نمره از ۱۱۰ نمره به منزله‌ی کسب نمره‌ی کامل خواهد بود. دقت کنید که نمره‌ی بالای ۱۰۰ سرریز نخواهد کرد.

پرسش‌های آزمون (۱۰۰ + ۱۰ نمره)

پرسش ۱ (۱۲ نمره) به سوالات زیر پاسخ دهید.

(آ) (۴ نمره) درست یا نادرست بودن جملات زیر را با ذکر دلیل مشخص کنید.

- الگوریتم یادگیری‌ـQ [1] می‌تواند تابع بهینه‌ی Q یا همان Q^* را بدون اجرای سیاست بهینه یا همان π^*، یاد بگیرد.
- الگوریتم Value iteration تضمین به همگرایی می‌دهد در صورتی که ضریب تخفیف[2] (γ) در نابرابری $۰ < \gamma < ۱$ قرار گیرد.
- در یک MDP [3] که دارای مدل انتقال T است که به هر سه تایی $T(s,a,s')$ احتمالی غیرصفر اختصاص می‌دهد، الگوریتم یادگیری‌ـQ شکست خواهد خورد.
- در الگوریتم approximate Q-learning اگر نرخ یادگیری (α) و ضریب تخفیف (γ) هر دو کاهش یابند مقادیر Q به پاداش اخیر حساس‌تر می‌شوند.

(ب) (۲ نمره) در هنگام استفاده از یک مدل Naive Bayes به همراه Laplace Smoothing، با افزایش مقدار K کدام یک از موارد زیر <u>می‌تواند</u> رخ دهد؟ دلیل خود را برای هرکدام ذکر کنید.

- افزایش مقدار خطای آموزشی
- کاهش مقدار خطای آموزشی
- افزایش مقدار خطای تست
- کاهش مقدار خطای تست

(ج) (۳ نمره) با ذکر دلیل مشخص کنید که کدام یک از موارد زیر روشی مناسب برای جلوگیری کردن از overfit شدن مدل می‌باشد؟ (می‌توانید بیشتر از یک مورد را انتخاب کنید.)

- کاهش تعداد epoch ها به هنگام آموزش مدل با داده‌های آموزشی و استفاده از SGD
- محدود کردن نرم بردار وزن ($\|W\| \leq ۱$)
- کاهش داده‌های آموزش

(د) (۳ نمره) با ذکر دلیل مشخص کنید که کدام یک از موارد زیر در رابطه با تابع فعال‌سازی Sigmoid و ReLU درست هستند؟ (می‌توانید بیشتر از یک مورد را انتخاب کنید.)

- هر دو تابع فعال‌سازی به طور پیوسته غیرکاهشی هستند.
- در مقایسه با تابع Sigmoid، تابع ReLU از لحاظ محاسباتی پرهزینه‌تر است.
- هر دو تابع دارای مشتق اول یکنوا هستند.

پاسخ

(آ)
- **<u>درست است.</u>** قابل توجه است که الگوریتم Q-lerning در واقع یک یادگیری Off-policy است. در رویکرد Off-policy سیاست بهینه بدون در نظر گرفتن اقدامات عامل یا انگیزه آن برای اقدام بعدی تعیین می‌شود. به عبارتی می‌تواند مقدار اقدام بهینه را بدون نیاز به اجرای سیاست بهینه در تمام طول الگوریتم تعیین کند.

[1] Q-Learning

[2] discount factor

[3] Markov Decision Process

- **درست است.** دلیل آن این است که وقتی γ در این محدوده باشد، پاداش‌های آینده‌تر به درستی تخفیف داده می‌شوند(تنزیل می‌شوند) و این اطمینان را می‌دهد که الگوریتم می‌تواند به درستی بین پاداش‌های نزدیک و آینده تعادل برقرار کند. این اتفاق اجازه می دهد تا الگوریتم به یک راه حل منحصر به فرد همگرا شود، همانطور که توسط معادلات بلمن در یک MDP محدود، ثابت می‌شود.
- **نادرست است.** درواقع اینکه بدانیم که هر سه تایی $T(s, a, s')$ احتمالی غیر از صفر دارد باعث می‌شود که الگوریتم Q-learning بهتر اجرا شود چرا که این یعنی تمام استیت ها در MDP با یک دنباله‌ای از حرکت‌ها از استیت دلخواه دیگر قابل دسترس است که این یکی از شرط‌هایی است که سبب آموزش بهتر خواهد شد؛ به این شکل که الگوریتم می‌تواند در تمام فضا اکتشاف انجام دهد.
- **نادرست است** چرا که با کاهش γ به پاداش‌های اخیر ارزش کمتری نسبت به قبل می‌دهیم و همچنین با کاهش α فرایند یادگیری را کندتر کرده ایم.

(ب)
- افزایش مقدار خطای آموزشی می‌تواند رخ دهد. درواقع با افزایش مقدار K مدل کمتر روی داده های آموزشی فیت می‌شود که سبب می‌شود خطای آموزشی افزایش پیدا کند.
- افزایش مقدار خطای تست می‌تواند رخ دهد. در صورتی که مقدار K بیش از اندازه افزایش یابد، مدل بیش از اندازه هموار خواهد شد که باعث عملکرد ضعیف آن و underfit خواهد شد.
- کاهش مقدار خطای آموزشی نمی‌تواند رخ دهد. قابل توجه هست که افزایش مقدار K باعث می‌شود مدل روی داده های آموزشی کمتر فیت شود و به همین دلیل نمی‌تواند روی داده های آموزشی نتیجه بهتری بگیرد از حالتی که مقدار K کمتر بوده است.
- کاهش مقدار خطای تست می‌تواند رخ دهد. در واقع در صورتی که پیش از افزایش مقدار K مدل ما روی داده‌های آموزشی overfit شده باشد، پس از افزایش مقدار K سبب بهبود generalizatoin مدل خواهد شد و خطای تست را کاهش خواهد داد.

(ج) دو مورد اول روش‌هایی مناسب برای جلوگیری از overfit شدن مدل هستند. روش آخر مناسب نیست چرا که ما با دیتای کمتر بیشتر به داده آموزش وابسته می‌شویم.

(د) مورد اول صحیح است. درواقع مورد دوم نادرست است چرا که تابع Relu به ازای مقادیر منفی ۰ و به ازای مقادیر مثبت ۱ برمی‌گرداند و هزینه محاسباتی ندارد ولیکن تابع sigmoid نیاز به محاسبه مقادیر نمایی دارد. همچنین مورد سوم نیز نادرست است چون sigmoid مشتق اول یکنوا ندارد.

پرسش ۲ (۱۰ نمره) با توجه به داده‌های آموزشی زیر می‌خواهیم یک درخت تصمیم جهت دسته‌بندی داده‌ها طراحی کنیم.

X_1	X_2	X_3	y
F	F	T	−
F	T	F	+
F	T	T	−
F	T	F	−
T	F	F	+
T	T	T	+

(آ) (۸ نمره) درخت تصمیم را برای این مجموعه داده رسم کنید. توجه کنید در گره‌هایی که تعداد داده‌ها با لیبل مثبت و منفی برابر است، لیبل آن گره را مثبت در نظر می‌گیریم. همچنین مقدار بهره‌وری اطلاعات[۴] را برای هر سه ویژگی در ریشه حساب کنید. (الزامی به محاسبه‌ی بهره‌وری اطلاعات در سایر گره‌ها نیست.)
برای محاسبه‌ی لگاریتم‌ها می‌توانید از مقادیر تقریبی زیر استفاده کنید.

$$log_2(\frac{1}{4}) = -2, log_2(\frac{3}{4}) = -0.4, log_2(\frac{2}{3}) = -0.6, log_2(\frac{1}{3}) = -1.6$$

(ب) (۲ نمره) دقت آموزش درخت تصمیم را بدست آورید. دلیل وجود خطا چیست؟

پاسخ

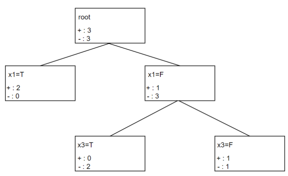
(آ)

$$H(y) = -(\frac{1}{2}log(\frac{1}{2}) + \frac{1}{2}log(\frac{1}{2})) = 1$$
$$H(y|x_1) = -(\frac{1}{3}(0) + \frac{2}{3}(\frac{1}{4}log(\frac{1}{4}) + \frac{3}{4}log(\frac{3}{4})) \approx 0.53 \Rightarrow IG(x_1) = 0.47$$
$$H(y|x_2) = -(\frac{2}{3}(-1) + \frac{1}{3}(-1)) = 1 \Rightarrow IG(x_2) = 0$$
$$H(y|x_3) = -(\frac{1}{2}(\frac{1}{3}log(\frac{1}{3}) + \frac{2}{3}log(\frac{2}{3})) + \frac{1}{2}(\frac{1}{3}log(\frac{1}{3}) + \frac{2}{3}log(\frac{2}{3}))) \approx 0.93 \Rightarrow IG(x_3) = 0.07$$

بنابراین از فیچر x_1 در ریشه استفاده خواهیم کرد. برای گره‌ی بعدی بهروری اطلاعات به شکل زیر است.

$$H(y) = -(\frac{1}{4}log(\frac{1}{4}) + \frac{3}{4}log(\frac{3}{4})) = 0.8$$
$$H(y|x_1) = 0.8 \Rightarrow IG(x_1) = 0$$
$$H(y|x_2) = -(\frac{3}{4}(\frac{1}{3}log(\frac{1}{3}) + \frac{2}{3}log(\frac{2}{3})) + \frac{1}{4}(0)) \approx 0.69 \Rightarrow IG(x_2) = 0.11$$
$$H(y|x_3) = -(\frac{1}{2}(0) + \frac{1}{2}(\frac{1}{2}log(\frac{1}{2}) + \frac{1}{2}log(\frac{1}{2}))) = 0.5 \Rightarrow IG(x_3) = 0.3$$

در گره دوم از فیچر x_3 برای تقسیم استفاده می‌کنیم. با اینکه در یکی از فرزندان x_3 به برگ نرسیده‌ایم، چون این دو داده فیچرهای یکسان و لیبل‌های متفاوتی دارند به حالت پایه دوم می‌رسیم و دیگر این گره را باز نمی‌کنیم.

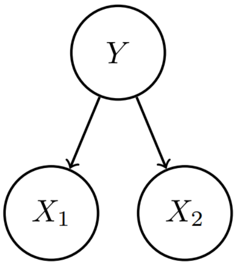


شکل ۱: decision tree

(ب) دقت درخت برابر $0.83 = \frac{5}{6}$ است. این خطا به دلیل وجود نویز در دیتا است چون به ازای دو ورودی با فیچرهای کاملا یکسان، برچسب‌های متفاوتی داریم.

[۴] Information gain

پرسش ۳ (۱۰ نمره) مدلی naive bayes مطابق شبکه زیر داریم که دارای دو ویژگی x_1 و x_2 است. همچنین جداول احتمالاتی زیر که بر حسب پارامترهای p_1, p_2, p_3 هستند را در اختیار داریم.



شکل ۲: naive bayes

y	$p(y)$
۰	$1-p_3$
۱	p_3

X_2	y	$p(x_2\|y)$
۰	۰	p_2
۱	۰	$1-p_2$
۰	۱	$1-p_2$
۱	۱	p_2

X_1	y	$p(x_1\|y)$
۰	۰	p_1
۱	۰	$1-p_1$
۰	۱	$1-p_1$
۱	۱	p_1

(آ) (۹ نمره) با استفاده از داده‌های آموزشی زیر، پارامترهای p_1, p_2, p_3 را با استفاده از maximum likelihood تخمین بزنید.

X_1	X_2	y
۱	۰	۱
۰	۱	۱
۱	۰	۱
۰	۰	۰
۰	۰	۰

(ب) (۱ نمره) داده‌ای با $x_1 = 1$ و $x_2 = 1$ به کدام کلاس تعلق دارد؟

پاسخ

(آ) می‌دانیم در naive bayes، common cause برقرار است. بنابراین : $p(y, x_1, x_2) = p(x_1|y)p(x_2|y)p(y)$.
همچنین احتمالات موجود در CPTها را می‌توان به این شکل نوشت:(هرگاه مقدار x_i و y یکسان باشد، احتمال آنها نیز یکسان است.)

$$p(x_1 = i|y = j) = p_1^{1[i=j]}(1 - p_1)^{1-1[i=j]}$$
$$p(x_2 = i|y = j) = p_2^{1[i=j]}(1 - p_2)^{1-1[i=j]}$$
$$p(y = i) = p_3^{i}(1 - p_3)^{1-i}$$

$$\begin{aligned}
L(D; p_1, p_2, p_3) &= p(x_1, x_2, y|p_1, p_2, p_3) \\
&= \prod_{i=1}^{5} p(x_1^i, x_2^i, y^i) \\
&= \prod_{i=1}^{5} p(x_1^i|y^i)p(x_2^i|y^i)p(y^i) \\
&= \prod_{i=1}^{5} p_1^{1[x_1^i=y^i]}(1 - p_1)^{1-1[x_1^i=y^i]}p_2^{1[x_2^i=y^i]}(1 - p_2)^{1-1[x_2^i=y^i]}p_3^{y^i}(1 - p_3)^{1-y^i} \\
log(L(D; p_1, p_2, p_3)) &= \sum_{i=1}^{5} 1[x_1^i = y^i]log(p_1) + (1 - 1[x_1^i = y^i])log(1 - p_1)+ \\
&1[x_2^i = y^i]log(p_2) + (1 - 1[x_2^i = y^i])log(1 - p_2) + y^i log(p_3) + (1 - y^i)log(1 - p_3)
\end{aligned}$$

اکنون از log likelihood نسبت به تک تک پارامترها مشتق می‌گیریم و آنها را برابر صفر قرار می‌دهیم تا مقادیری که به ازای آنها likelihood بیشینه می‌شود را بیابیم.

$$\frac{\partial log(L(D;p_۱,p_۲,p_۳))}{\partial p_۱} = \sum_{i=۱}^{۵} \frac{۱[x_۱^i = y^i]}{p_۱} - \frac{(۱ - ۱[x_۱^i = y^i])}{۱ - p_۱} = ۰ \Rightarrow p_۱ = \frac{\sum_{i=۱}^{۵} ۱[x_۱^i = y^i]}{۵}$$

$$\frac{\partial log(L(D;p_۱,p_۲,p_۳))}{\partial p_۲} = \sum_{i=۱}^{۵} \frac{۱[x_۲^i = y^i]}{p_۲} - \frac{(۱ - ۱[x_۲^i = y^i])}{۱ - p_۲} = ۰ \Rightarrow p_۲ = \frac{\sum_{i=۱}^{۵} ۱[x_۲^i = y^i]}{۵}$$

$$\frac{\partial log(L(D;p_۱,p_۲,p_۳))}{\partial p_۳} = \sum_{i=۱}^{۵} \frac{y^i}{p_۳} - \frac{۱ - y^i}{۱ - p_۳} = ۰ \Rightarrow p_۳ = \frac{\sum_{i=۱}^{۵} y^i}{۵}$$

بنابراین تخمین maximum likelihood از این پارامترها برابر است با :

$$p_۱ = \frac{۴}{۵}, p_۲ = \frac{۳}{۵}, p_۳ = \frac{۳}{۵}$$

(ب)

$$p(y = ۱|x_۱ = ۱, x_۲ = ۱) = \frac{p(y = ۱, x_۱ = ۱, x_۲ = ۱)}{p(y = ۱, x_۱ = ۱, x_۲ = ۱) + p(y = ۰, x_۱ = ۱, x_۲ = ۱)} = \frac{\frac{۳}{۵} \times \frac{۴}{۵} \times \frac{۳}{۵}}{\frac{۲}{۵} \times \frac{۱}{۵} \times \frac{۲}{۵} + \frac{۳}{۵} \times \frac{۴}{۵} \times \frac{۳}{۵}} = \frac{۹}{۱۰}$$

$$p(y = ۰|x_۱ = ۱, x_۲ = ۱) = \frac{۱}{۱۰}$$

بنابراین کلاس ورودی $x_۱ = ۱, x_۲ = ۱$ ۱ پیش بینی می‌شود.

پرسش ۴ (۲۲ نمره) قصد داریم با استفاده از لاجستیک رگرشن داده‌های آموزشی $\{(x_i, y_i), i = ۱, \dots, n\}$ را به طوری که $x_i \in \mathbb{R}^d$ یک بردار ویژگی [۵] و $y_i \in \{۰, ۱\}$ یک برچسب [۶] دودویی است، طبقه‌بندی کنیم. به همین منظور به سوالات زیر در رابطه با این الگوریتم پاسخ دهید.

(آ) (۲ نمره) لاجستیک رگرشن سعی در پیش‌بینی چه چیزی دارد؟ به صورتی احتمالی بررسی کنید.

(ب) (۸ نمره) نشان دهید بیشینه کردن log likelihood بر روی داده‌های آموزشی معادل کمینه کردن تابع هزینه‌ی زیر است. ($w \in \mathbb{R}^d$ بردار وزن‌های مدل است که سعی در تخمین آن داریم)

$$J(w) = -\sum_{i=۱}^{n} y_i log(p(y_i|x_i; w)) + (۱ - y_i) log(۱ - p(y_i|x_i; w))$$

(ج) (۸ نمره) با استفاده از بخش قبل، مشتق تابع log likelihood را نسبت به w بدست آورید. همچنین توجه کنید که این مساله بهینه‌سازی دارای جواب به فرم بسته نمی‌باشد. با توجه به این موضوع روشی را برای بدست آوردن جواب برای این الگوریتم پیشنهاد دهید.

(د) (۴ نمره) نشان دهید مرز تصمیم‌گیری (منحنی که در فضای ویژگی، داده‌های دو کلاس را از هم تفکیک می‌کند) [۷] در این الگوریتم یک تابع خطی است.

پاسخ

(آ) لاجستیک رگرشن سعی دارد براساس داده‌های آموزشی احتمال اینکه ورودی عضوی از کلاس ۱ باشد را تخمین بزند.

$$p(y = ۱|x) = \frac{۱}{۱ + e^{-w.x}}$$

(ب)

$$\begin{aligned} L(w) &= \prod_{i:y_i=۱} p(y_i|x_i, w) \prod_{i:y_i=۰} (۱ - p(y_i|x_i, w)) \\ &= \prod_{i=۱}^{n} p(y_i|x_i, w)^{y_i} (۱ - p(y_i|x_i, w))^{۱-y_i} \\ log(L(w)) &= \sum_{i=۱}^{n} y_i p(y_i|x_i, w) + (۱ - y_i) log(۱ - p(y_i|x_i, w)) \end{aligned}$$

بنابراین بیشینه کردن log likelihood معادل کمینه کردن تابع هزینه گفته شده است.

[۵] Feature vector
[۶] Label
[۷] boundary Decision

(ج)

$$
\begin{aligned}
log(L(w)) &= \sum_{i=1}^{n} y_i p(y_i|x_i,w) + (1-y_i) log(1-p(y_i|x_i,w)) \\
&= \sum_{i=1}^{n} log(1-p(y_i|x_i,w)) + \sum_{i=1}^{n} y_i log(\frac{p(y_i|x_i,w)}{1-p(y_i|x_i,w)})
\end{aligned}
$$

همچنین با توجه به اینکه $p(y=1|x,w)=\frac{1}{1+e^{-w.x}}$ است، داریم:

$$
\begin{aligned}
1-p(y=0|x,w) &= 1-\frac{1}{1+e^{-w.x}} \\
&= \frac{e^{-w.x}}{1+e^{-w.x}} \\
&\overset{\times\frac{e^{w.x}}{e^{w.x}}}{=} \frac{1}{1+e^{w.x}}
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
log(\frac{p(y=1|x,w)}{1-p(y=0|x,w)}) &= log(\frac{1}{e^{-w.x}}) \\
&= w.x
\end{aligned} \tag{2}
$$

با استفاده از ۱ و ۲ داریم:

$$
\begin{aligned}
log(L(w)) &= \sum_{i=1}^{n} -log(1+e^{w.x_i}) + \sum_{i=1}^{n} y_i w.x_i \Rightarrow \\
\frac{\partial log(L(w))}{\partial w} &= \sum_{i=1}^{n} -\frac{e^{w.x_i}}{1+e^{w.x_i}} x_i + \sum_{i=1}^{n} y_i x_i \\
&= \sum_{i=1}^{n} -\frac{1}{1+e^{-w.x_i}} x_i + \sum_{i=1}^{n} y_i x_i \\
&= \sum_{i=1}^{n} -p(y_i|x_i,w) x_i + \sum_{i=1}^{n} y_i x_i \\
&= \sum_{i=1}^{n} (y_i - p(y_i|x_i,w)) x_i
\end{aligned} \tag{3}
$$

این مساله قابل حل به صورت مستقیم نیست به همین دلیل می‌توان از روش‌های iterative مانند gradient descent استفاده کرد. (می‌توان نشان داد که تابع log likelihood یک تابع مقعر است و تنها یک نقطه بیشینه دارد. برای اثبات می‌توان حالتی که x یک اسکالر است را در نظر گرفت.)

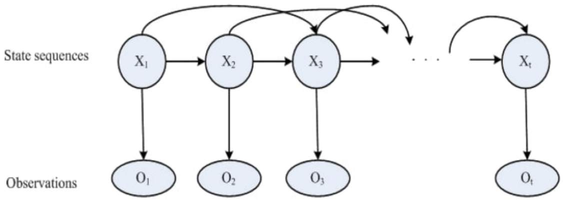
(د) مرز تصمیم گیری ناحیه‌ای است که در آن احتمال اینکه ورودی از کلاس ۰ یا ۱ باشد برابر است. نشان می‌دهیم برای اینکه یک ورودی از کلاس ۱ پیش بینی شود کافی است که در بالای این صفحه‌ی جداکننده قرار بگیرد و این صفحه از ترکیب خطی ورودی‌ها ایجاد می‌شود.

$$
\begin{aligned}
p(y=1|x,w) &> p(y=0|x,w) \\
\frac{1}{1+e^{-w.x}} &> \frac{e^{-w.x}}{1+e^{-w.x}} \\
log(1) - log(1+e^{-w.x}) &> log(e^{-w.x}) - log(1+e^{-w.x}) \\
0 &< w.x
\end{aligned}
$$

بنابراین مرز تصمیم گیری یک تابع خطی از ورودی است.

پرسش ۵ (۱۰ نمره) مدل مارکوف پنهان [8] زیر را در نظر بگیرید که دامنه‌ی متغیرهای آن دودویی [9] است. در این مدل مطابق شکل ۳ هر متغیر حالت به دو متغیر حالت پیشین خود وابسته است.

[8] Hidden Markov Model
[9] Binary



شکل ۳

همچنین جدول توزیع احتمالات شرطی این مدل نیز به شکل زیر است:

$x_{t-۲}$	$x_{t-۱}$	$p(x_t = ۱ \mid x_{t-۱}, x_{t-۲})$
۰	۰	۰/۸
۱	۰	۰/۳
۰	۱	۰/۶
۱	۱	۰/۱

x_t	$p(o_t = ۱ \mid x_t)$
۰	۰/۲
۱	۰/۴

یک مدل مارکوف پنهان مرتبه اول [۱۰] (عادی) معادل با مدل مارکوف مطرح شده طراحی نمایید.

پاسخ برای هر i متغیر y_i را معادل جفت $x_{i-۱}, x_i$ قرار می‌دهیم. در این صورت برای پیدا کردن $p(y_i|y_{i-۱})$ کافی است دقت کنید هر y_i ۴ حالت متفاوت $(۰,۰), (۰,۱), (۱,۰), (۱,۱)$ دارد. حال پس باید رقم سمت راست y_i برابر رقم سمت چپ $y_{i-۱}$ باشد پس اگر اینطور نباشد $p(y_i|y_{i-۱})$ صفر می‌شود. اگر $y_i = (x_i, x_{i-۱})$ و $y_{i-۱} = (x_{i-۱}, x_{i-۲})$ آنگاه داریم:

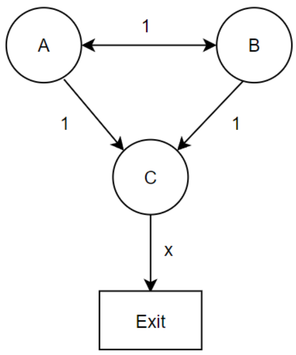
$$p(y_i|y_{i-۱}) = p(x_i, x_{i-۱}|x_{i-۱}, x_{i-۲}) = p(x_i|x_{i-۱}, x_{i-۱}) \quad (۴)$$

حال ترم Observation برابر O_i قرار می دهیم.

$$p(O_i|y_i) = p(O_i|x_i, x_{i-۱}) = p(O_i|x_i) \quad (۵)$$

پرسش ۶ (۳۰ نمره)

(آ) (۱۰ نمره) به MDP کشیده شده در شکل ۴ توجه کنید. در این MDP کنش های قابل انجام در هر حالت با توجه به جهت یال‌ها مشخص می‌شود. بنابراین به عنوان مثال از وضعیت A می‌توان به وضعیت B یا C رفت. همچنین میزان پاداش هر کنش بر روی یال آن نوشته شده است. تنها کنشی که در وضعیت C می‌توان انجام داد، خروج است که معادل ورود به وضعیت ترمینال است و پاداش x را به همراه دارد.



شکل ۴

اکنون فرض کنید هر کنشی غیر از کنش خروج با احتمال ۰/۵ موفقیت آمیز باشد و در صورت شکست، عامل در سر جای خود بماند و پاداش ۰ را دریافت کند. توجه شود کنش خروج همچنان به صورت قطعی انجام می‌شود. همچنین $\gamma = ۰/۵$ است.

در صورتی که حرکت از وضعیت ۱ به وضعیت ۲ را با ۲ $\rightarrow$ ۱ نمایش دهیم، به ازای چه مقداری از x ، $Q^*(A, A \rightarrow B) = Q^*(A, A \rightarrow C)$ خواهد بود؟

راهنمایی: به رابطه بلمن برای استیت A و کنش‌هایش توجه کنید.

(ب) (۱۲ نمره) یک MDP محدود با پاداش‌های دارای کران مشخص را در نظر بگیرید و فرض کنید این MDP یک سیاست بهینه قطعی دارد. حال از روی این MDP یک MDP جدید می‌سازیم به این صورت که اگر کنش a در یک وضعیت s بهینه نباشد، $r(s,a)$ را از مقدار ثابت و مثبت c کم می‌کنیم و در صورتی که a کنش بهینه باشد مقدار پاداش آن تغییری نمی‌کند. آیا سیاست بهینه در MDP جدید با سیاست بهینه در MDP اولیه برابر است؟ اگر پاسخ شما مثبت است آن را اثبات کرده و در غیر این صورت مثال نقض بیاورید.

---
[۱۰] first order

(ج) (۸ نمره) معادله‌ی بلمن را معکوس کرده‌ایم به گونه‌ای که مقدار یک حالت را بر اساس حالت‌های قبلی به شکل زیر محاسبه می‌کنیم. با ارائه‌ی یک مثال نقض نشان دهید که این رابطه در حالت کلی صحیح نمی‌باشد.

$$V^\pi(s') = \sum_s \sum_a P(s'|s,a) \left(\frac{V^\pi(s) - R(s,a)}{\gamma} \right)$$

پاسخ

(آ)

$$Q^*(A, A \to B) = \frac{۱}{۲}(۰ + \lambda V^*(A)) + \frac{۱}{۲}(۱ + \lambda V^*(B))$$
$$Q^*(A, A \to C) = \frac{۱}{۲}(۰ + \lambda V^*(A)) + \frac{۱}{۲}(۱ + \lambda x)$$

با توجه به تقارن مساله داریم : $V^*(A) = V^*(B)$. همچنین توجه داریم که در استیت A تنها می‌توان دو حرکت را انجام داد و چون q-value آنها برابر است بنابراین : $V^*(A) = Q^*(A, A \to B) = Q^*(A, A \to C)$.
این دو تساوی را در رابطه‌ی اول جایگذاری می‌کنیم:

$$V^*(A) = \frac{۱}{۴}V^*(A) + \frac{۱}{۲} + \frac{۱}{۴}V^*(A) \Rightarrow$$
$$V^*(A) = ۱$$

حالا این مقدار را در رابطه‌ی دوم جایگذاری می‌کنیم و داریم:

$$۱ = \frac{۱}{۴} + \frac{۱}{۲} + \frac{۱}{۴}x \Rightarrow$$
$$x = ۱$$

(ب) ادعای گفته شده صحیح می‌باشد. ابتدا، به صورت کلی می‌توان گفت برای همه $s \in S$ و $a \in A$، $R(s,a) \geq R'(s,a)$، منظور از R' مقدار پاداش در MDP جدید می‌باشد. زیرا بر اساس تعریف مسئله در صورتی که کنش سیاست بهینه باشد $R(s,a) = R'(s,a)$ در غیر این صورت $R(s,a) - R'(s,a) = R(s,a) - c$ حال روابط را به صورت زیر می‌نویسیم:

$$v_M^{\pi^*}(s) = E\left[\sum_{k=۰}^{\infty} \gamma^k R_{t+k} | S_t = s, \pi^*, M\right]$$

$$= E\left[\sum_{k=۰}^{\infty} \gamma^k R_{t+k} | S_t = s, \pi^*, M'\right] = v_{M'}^{\pi^*}(s) \quad s \in S \text{ برای همه}$$

به صورت کلی برای تمام سیاست‌ها نیز می‌توان رابطه زیر را نوشت:

$$E\left[\sum_{k=۰}^{\infty} \gamma^k R_{t+k} | S_t = s, \pi^*, M'\right] = E\left[\sum_{k=۰}^{\infty} \gamma^k R_{t+k} | S_t = s, \pi^*, M\right] \geq E\left[\sum_{k=۰}^{\infty} \gamma^k R_{t+k} | S_t = s, \pi, M\right] \geq E\left[\sum_{k=۰}^{\infty} \gamma^k R_{t+k} | S_t = s, \pi, M'\right]$$

در نتیجه می‌توان گفت برای هر سیاستی:

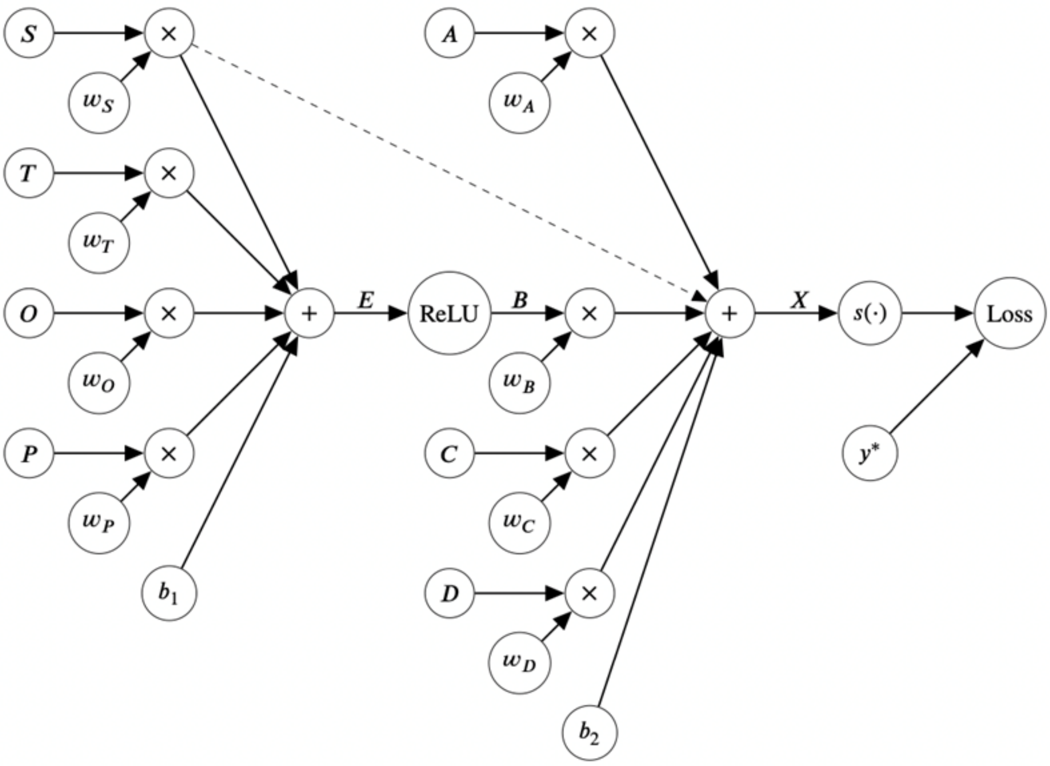
$$v_{M'}^{\pi^*}(s) \geq v_{M'}^{\pi^*}(s)$$

در نتیجه ثابت کردیم که در M' نیز سیاست π^* سیاست بهینه می‌باشد.

(ج) می‌توان با یک مثال نقض اثبات کرد که این رابطه درست نیست. یک MDP به نام m را فرض کنید. در این MDP یک وضعیت به نام s' وجود دارد که از هیچ وضعیت دیگری قابل دسترسی نیست ولی $v^\pi(s') \neq ۰$. در این حالت بر اساس رابطه بالا، چون $P(s,a,s')$ برابر صفر می‌شود، آنگاه طبق رابطه گفته شده باید داشته باشیم که $v^\pi(s') = ۰$ در نتیجه عبارت بالا درست نیست.

در صورتی هم که بدون استفاده از رابطه بلمن مثال نقضی زدید نیز قابل قبول است.

پرسش ۷ (۱۶ نمره) شبکه عصبی زیر را در نظر بگیرید:



در این شبکه S ،T ،O ،P ،A ،C و D ورودی‌ها، w_S ،w_T ،w_O ،w_P ،w_A ،w_C ،w_D ،w_B وزن‌ها و $b_۱$ و $b_۲$ مقادیر بایاس هستند. همچنین تابع $s(.)$ به شکل زیر تعریف می‌شود:

$$s(x) = \frac{۱}{۱+e^{-x}}$$

فرض کنید یال مشخص شده با خط چین وجود ندارد. عبارات زیر را برحسب ورودی‌ها و متغیرهای E و $\frac{\partial Loss}{\partial s(X)}$ و $s(X)$ و وزن‌های شبکه بدست آورید:

الف) $\frac{\partial Loss}{\partial w_A}$

ب) $\frac{\partial Loss}{\partial w_S}$

حال فرض کنید یال مشخص شده با خط چین مانند یک یال عادی در شبکه عصبی وجود دارد. عبارات زیر را برحسب ورودی‌ها و متغیرهای E و $\frac{\partial Loss}{\partial s(X)}$ و $s(X)$ و وزن‌های شبکه بدست آورید:

ج) $\frac{\partial Loss}{\partial w_A}$

د) $\frac{\partial Loss}{\partial w_S}$

پاسخ ابتدا دقت کنید که داریم:

$$\frac{\partial s(x)}{\partial x} = s(x)[۱ - s(x)]$$

همچنین تعریف می‌کنیم که:

$$h(x) = \frac{\partial ReLLU(x)}{\partial x} = \begin{cases} ۱ & x \geq ۰ \\ ۰ & o.w \end{cases}$$

الف)

$$\frac{\partial Loss}{\partial w_A} = \frac{\partial Loss}{\partial s(X)} \cdot \frac{\partial s(X)}{\partial X} \cdot \frac{\partial X}{\partial Aw_A} \cdot \frac{\partial Aw_A}{\partial w_A} = \frac{\partial Loss}{\partial s(X)} . s(X)[۱ - s(X)].A$$

ب)

$$\frac{\partial Loss}{\partial w_S} = \frac{\partial Loss}{\partial s(X)} \cdot \frac{\partial s(X)}{\partial X} \cdot \frac{\partial X}{\partial Bw_B} \cdot \frac{\partial Bw_B}{\partial ReLU(E)} \cdot \frac{\partial ReLU(E)}{\partial E} \cdot \frac{\partial E}{\partial Sw_S} \cdot \frac{\partial Sw_S}{\partial w_S} = \frac{\partial Loss}{\partial s(X)} . s(X)[۱ - s(X)].w_B.h(E).S$$

ج) در این حالت مسیرهای w_A تا Loss تغییر نکرده و پاسخ مانند قسمت الف خواهد بود.

د) در این حالت دو مسیر از w_S به Loss وجود دارد که هر دو روی مشتق خواسته شده اثر می‌گذارند:

$$\frac{\partial Loss}{\partial w_S} = \frac{\partial Loss}{\partial s(X)} \cdot \frac{\partial s(X)}{\partial X} \cdot \left(\frac{\partial X}{\partial Bw_B} \cdot \frac{\partial Bw_B}{\partial ReLU(E)} \cdot \frac{\partial ReLU(E)}{\partial E} \cdot \frac{\partial E}{\partial Sw_S} + \frac{\partial X}{\partial Sw_S}\right) \cdot \frac{\partial Sw_S}{\partial w_S} = \frac{\partial Loss}{\partial s(X)} . s(X)[۱ - s(X)].(w_B.h(E) + ۱).S$$