به نام خداوند بخشنده مهربان

هوش مصنوعی

دانشکده مهندسی کامپیوتر

بهار۱۴۰۴

۱شهریور ۱۴۰۴ | **امتحان پایانترم** | زمان آزمون: ۱۲۰ 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. آزمون از ۱۱۰ نمره است. دریافت ۱۰۰ نمره از ۱۱۰ نمره به منزله دریافت نمره کامل خواهد بود. نمره بالای ۱۰۰ سرریز نخواهد کرد.

۴. نوشته‌های شما در قسمت چرک‌نویس به هیچ عنوان تصحیح نخواهد شد.

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

پرسش‌های آزمون (۱۱۰ نمره)

پرسش ۱ (۲۳ نمره) درستی یا نادرستی موارد زیر را همراه با ذکر دلیل مشخص کنید

(آ) (۳نمره) الگوریتم $Q-learning$ می تواند تابع بهینه ی Q یا همان Q^* را بدون اجرای سیاست بهینه یا همان π^*، یاد بگیرد.

(ب) (۳نمره) الگوریتم Value iteration تضمین به همگرایی می دهد در صورتی که ضریب تخفیف (γ) در نابرابری $۰ < \gamma < ۱$ قرار گیرد

(ج) (۳نمره) در الگوریتم approximate Q-learning اگر نرخ یادگیری (α) و ضریب تخفیف (γ) هر دو کاهش یابند، مقادیر Q به پاداش اخیر حساس‌تر می‌شوند.

(د) (۳نمره) در صورتی که داده های آموزش بدون نویز باشند یک درخت تصمیم در صورتی که اجازه داده شود بدون شرط توقف و محدودیتی روی عمق رشد کند، توانایی دارد که هر داده آموزشی را با دقت ۱۰۰ درصد دسته‌بندی کند ولی ممکن است نتواند Generalize کند.

(ه) (۳نمره) در هنگام استفاده از یک مدل Naive Bayes به همراه Laplace Smoothing با افزایش مقدار K مقدار خطای آموزش می تواند افزایش و مقدار خطای تست می تواند افزایش یابد .

(و) (۴نمره) در یک MDP با حالات محدود و پاداش با باند مشخص و $۰ < \gamma < ۱$ در صورتی همه پاداش‌ها با عدد ثابت c جمع شود، سیاست بهینه تغییر نمی‌کند.

(ز) (۴نمره) اگر بدانیم که $|A| << |S|$ آنگاه پیچیدگی زمانی هر iteration در policy iteration و value iteration باهم برابر می شود.

پاسخ

(آ) **درست است:** قابل توجه است که الگوریتم Q-learning 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 می‌توان مقدار اقدام بهینه را بدون نیاز به اجرای سیاست بهینه در تمام طول الگوریتم تعیین کند.

(ب) **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 می‌دهد تا الگوریتم به یک راه حل منحصربه‌فرد همگرا شود، مشروط بر اینکه MDP محدود و ثابت باشد.

(ج) **نادرست است:** چرا که با کاهش γ پاداش‌های اخیر تاثیر کمتری نسبت به قبل می‌دهیم و همچنین با کاهش α فرآیند یادگیری را کندتر کرده‌ایم.

(د) **درست است:** درخت تصمیم با توجه به فرایند ساخته شدنش تا جایی پیش می‌رود که داده‌های رسیده به برگ‌ها فقط از یک دسته باشد و بنابراین روی داده آموزشی به دقت ۱۰۰ درصد دست می‌یابد. بدین‌علت چنین درختی ممکن است به علت واریانس بالا، روی داده‌های جدید عملکرد خوبی نداشته باشد.

(ه) **درست است:** هر دو می توانند رخ دهند. با افزایش مقدار K مدل کمتر روی داده های آموزشی فیت می شود که سبب می شود خطای آموزش افزایش یابد . و در صورتی که مقدار K بیش از اندازه افزایش یابد مدل بیش از اندازه هموار خواهد شد که باعث عملکرد ضعیف آن underfit خواهد شد .

(و) **درست است:** افزودن یک مقدار ثابت به تمام پاداش‌ها تأثیری بر روی ترتیب نسبی ارزش‌های حالت‌ها ندارد. (اگر فرض می‌کردیم که MDP می‌تواند یک استیت Terminal داشته باشد، گزاره غلط می‌شد. پس به جواب‌هایی که برای این حالت مثال نقض آورده‌اند نیز نمره داده می شود.)

(ز) **نادرست است:** در صورتی که تعداد حالت ها بسیار بیشتر از تعداد اقدامات باشد،پیچیدگی زمانی این دو الگوریتم می تواند متفاوت باشد . اردر زمانی value iteration برابر است با $|S|^۲|A|$ و برای policy iteration برابر است با $|S|^۲|A| + |S|^۳$

پرسش ۲ (۱۵ نمره)

در این مسئله، ما به تخمین گر maximum likelihood با استفاده از naive bayes می‌پردازیم. در اینجا، ویژگی‌های باینری $x_j,\ j = 1, \ldots, n$ در مدل ما متغیرهای گسسته و دودویی هستند، یعنی $x_j \in \{0, 1\}$. و $x = [x_1\ x_2\ \cdots\ x_n]^T$ را به عنوان بردار ورودی در نظر می‌گیریم.

برای هر نمونهٔ آموزشی، خروجی‌های ما یک مقدار دودویی منفرد $y \in \{0, 1\}$ هستند.

مدل ما با پارامترهای $\phi_{j|y=0} = p(x_j = 1 \mid y = 0)$، $\phi_{j|y=1} = p(x_j = 1 \mid y = 1)$، و $\phi_y = p(y = 1)$ مشخص می‌شود.

ما توزیع مشترک (x, y) را بر اساس موارد زیر مدل‌سازی می‌کنیم:

$$p(y) = (\phi_y)^y (1 - \phi_y)^{1-y}$$

$$\begin{aligned} p(x \mid y = 0) &= \prod_{j=1}^{n} p(x_j \mid y = 0) \\ &= \prod_{j=1}^{n} (\phi_{j|y=0})^{x_j} (1 - \phi_{j|y=0})^{1-x_j} \end{aligned}$$

$$\begin{aligned} p(x \mid y = 1) &= \prod_{j=1}^{n} p(x_j \mid y = 1) \\ &= \prod_{j=1}^{n} (\phi_{j|y=1})^{x_j} (1 - \phi_{j|y=1})^{1-x_j} \end{aligned}$$

(آ) (۴نمره) تابع likelihood مشترک

$$\ell(\varphi) = \log \prod_{i=1}^{m} p(x^{(i)}, y^{(i)}; \varphi)$$

را بر حسب پارامترهای مدل داده‌شده در بالا بیابید . φ کل مجموعه پارامترها را نشان می‌دهد:

$$\{\phi_y,\ \phi_{j|y=0},\ \phi_{j|y=1},\ j = 1, \ldots, n\}.$$

(ب) (۴نمزه) نشان دهید پارامتر هایی که تابع likelihood را بیشینه می کنند به صورت زیر هستند:

$$\phi_{j|y=0} = \frac{\sum_{i=1}^{m} 1\{x_j^{(i)} = 1 \wedge y^{(i)} = 0\}}{\sum_{i=1}^{m} 1\{y^{(i)} = 0\}}$$

$$\phi_{j|y=1} = \frac{\sum_{i=1}^{m} 1\{x_j^{(i)} = 1 \wedge y^{(i)} = 1\}}{\sum_{i=1}^{m} 1\{y^{(i)} = 1\}}$$

$$\phi_y = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 1\}}{m}.$$

(ج) (۷نمره) در نظر بگیرید که می‌خواهیم پیش‌بینی‌ای روی یک داده‌ی جدید x با استفاده از محتمل‌ترین برچسب کلاسی که توسط الگوریتم naive bayes تولید می‌شود، انجام دهیم. نشان دهید اگر $p(y = 0 \mid x)$ و $p(y = 1 \mid x)$ احتمالات کلاس بازگردانده‌شده توسط naive bayes باشند، برداری $\theta \in \mathbb{R}^{n+1}$ وجود دارد به‌طوری‌که

$$p(y = 1 \mid x) \geq p(y = 0 \mid x) \quad \text{اگر و تنها اگر} \quad \theta^T \begin{bmatrix} 1 \\ x \end{bmatrix} \geq 0$$

بردار θ را بیابید همچنین θ_0 را جمله ثابت در نظر بگیرید

پاسخ

(آ)

$$
\begin{aligned}
\ell(\varphi) &= \log \prod_{i=۱}^{m} p(x^{(i)}, y^{(i)}; \varphi) \\
&= \log \prod_{i=۱}^{m} p(x^{(i)}|y^{(i)}; \varphi) p(y^{(i)}; \varphi) \\
&= \log \prod_{i=۱}^{m} \left(\prod_{j=۱}^{n} p(x_j^{(i)}|y^{(i)}; \varphi) \right) p(y^{(i)}; \varphi) \\
&= \sum_{i=۱}^{m} \left(\log p(y^{(i)}; \varphi) + \sum_{j=۱}^{n} \log p(x_j^{(i)}|y^{(i)}; \varphi) \right) \\
&= \sum_{i=۱}^{m} \Bigg[y^{(i)} \log \phi_y + (۱ - y^{(i)}) \log(۱ - \phi_y) \\
&\qquad + \sum_{j=۱}^{n} \left(x_j^{(i)} \log \phi_{j|y^{(i)}} + (۱ - x_j^{(i)}) \log(۱ - \phi_{j|y^{(i)}}) \right) \Bigg]
\end{aligned}
$$

(ب) تنها جمله‌هایی در $\ell(\varphi)$ که گرادیان غیر صفر نسبت به $\phi_{j|y=۰}$ دارند، آن‌هایی هستند که شامل $\phi_{j|y^{(i)}}$ می‌شوند. بنابراین:

$$
\nabla_{\phi_{j|y=۰}} \ell(\varphi) = \nabla_{\phi_{j|y=۰}} \sum_{i=۱}^{m} \left(x_j^{(i)} \log \phi_{j|y^{(i)}} + (۱ - x_j^{(i)}) \log(۱ - \phi_{j|y^{(i)}}) \right)
$$

$$
= \nabla_{\phi_{j|y=۰}} \sum_{i=۱}^{m} \left(x_j^{(i)} \log(\phi_{j|y=۰}) ۱\{y^{(i)} = ۰\} + (۱ - x_j^{(i)}) \log(۱ - \phi_{j|y=۰}) ۱\{y^{(i)} = ۰\} \right).
$$

$$
= \sum_{i=۱}^{m} \left(\frac{x_j^{(i)}}{\phi_{j|y=۰}} ۱\{y^{(i)} = ۰\} - \frac{۱ - x_j^{(i)}}{۱ - \phi_{j|y=۰}} ۱\{y^{(i)} = ۰\} \right).
$$

با قرار دادن $\nabla_{\phi_{j|y=۰}} \ell(\varphi) = ۰$ داریم:

$$
۰ = \sum_{i=۱}^{m} \left(\frac{x_j^{(i)}}{\phi_{j|y=۰}} ۱\{y^{(i)} = ۰\} - \frac{۱ - x_j^{(i)}}{۱ - \phi_{j|y=۰}} ۱\{y^{(i)} = ۰\} \right).
$$

$$
= \sum_{i=۱}^{m} \left(x_j^{(i)} (۱ - \phi_{j|y=۰}) ۱\{y^{(i)} = ۰\} - (۱ - x_j^{(i)}) \phi_{j|y=۰} ۱\{y^{(i)} = ۰\} \right).
$$

$$
= \sum_{i=۱}^{m} \left((x_j^{(i)} - \phi_{j|y=۰}) ۱\{y^{(i)} = ۰\} \right).
$$

$$
= \sum_{i=۱}^{m} x_j^{(i)} ۱\{y^{(i)} = ۰\} - \phi_{j|y=۰} \sum_{i=۱}^{m} ۱\{y^{(i)} = ۰\}.
$$

$$
= \sum_{i=۱}^{m} ۱\{x_j^{(i)} = ۱ \wedge y^{(i)} = ۰\} - \phi_{j|y=۰} \sum_{i=۱}^{m} ۱\{y^{(i)} = ۰\}.
$$

در نتیجه به جواب مطلوب خود می‌رسیم:

$$
\phi_{j|y=۰} = \frac{\sum_{i=۱}^{m} ۱\{x_j^{(i)} = ۱ \wedge y^{(i)} = ۰\}}{\sum_{i=۱}^{m} ۱\{y^{(i)} = ۰\}}.
$$

راه‌حل برای $\phi_{j|y=۱}$ نیز به همان روش به دست می‌آید.
برای حلِ ϕ_y،

$$
\begin{aligned}
\nabla_{\phi_y} \ell(\varphi) &= \nabla_{\phi_y} \sum_{i=۱}^{m} \left(y^{(i)} \log \phi_y + (۱ - y^{(i)}) \log(۱ - \phi_y) \right) \\
&= \sum_{i=۱}^{m} \left(y^{(i)} \frac{۱}{\phi_y} - (۱ - y^{(i)}) \frac{۱}{۱ - \phi_y} \right).
\end{aligned}
$$

سپس با قرار دادن $\nabla_{\phi_y}\ell(\varphi) = ۰$ خواهیم داشت

$$
\begin{aligned}
۰ &= \sum_{i=۱}^{m}\left(y^{(i)}\frac{۱}{\phi_y} - (۱-y^{(i)})\frac{۱}{۱-\phi_y}\right)\\
&= \sum_{i=۱}^{m}\left(y^{(i)}(۱-\phi_y) - (۱-y^{(i)})\phi_y\right)\\
&= \sum_{i=۱}^{m} y^{(i)} - \sum_{i=۱}^{m}\phi_y\\
&= \sum_{i=۱}^{m} y^{(i)} - m\,\phi_y.
\end{aligned}
$$

در نتیجه،

$$
\phi_y \;=\; \frac{\sum_{i=۱}^{m} ۱\{\,y^{(i)} = ۱\,\}}{m}\,.
$$

(ج)

$$
\begin{aligned}
& p(y=۱|x) \geq p(y=۰|x)\\
\iff & \frac{p(y=۱|x)}{p(y=۰|x)} \geq ۱\\
\iff & \frac{\left(\prod_{j=۱}^{n} p(x_j|y=۱)\right)p(y=۱)}{\left(\prod_{j=۱}^{n} p(x_j|y=۰)\right)p(y=۰)} \geq ۱\\
\iff & \frac{\left(\prod_{j=۱}^{n} (\phi_{j|y=۰})^{x_j}(۱-\phi_{j|y=۰})^{۱-x_j}\right)\phi_y}{\left(\prod_{j=۱}^{n} (\phi_{j|y=۱})^{x_j}(۱-\phi_{j|y=۱})^{۱-x_j}\right)(۱-\phi_y)} \geq ۱\\
\iff & \sum_{j=۱}^{n}\left(x_j\log\left(\frac{\phi_{j|y=۱}}{\phi_{j|y=۰}}\right) + (۱-x_j)\log\left(\frac{۱-\phi_{j|y=۰}}{۱-\phi_{j|y=۱}}\right)\right) + \log\left(\frac{\phi_y}{۱-\phi_y}\right) \geq ۰\\
\iff & \sum_{j=۱}^{n} x_j\log\left(\frac{\phi_{j|y=۱}(۱-\phi_{j|y=۰})}{\phi_{j|y=۰}(۱-\phi_{j|y=۱})}\right) + \sum_{j=۱}^{n}\log\left(\frac{۱-\phi_{j|y=۱}}{۱-\phi_{j|y=۰}}\right) + \log\left(\frac{\phi_y}{۱-\phi_y}\right) \geq ۰\\
\iff & \theta^T\begin{bmatrix}۱\\ x\end{bmatrix} \geq ۰
\end{aligned}
$$

که

$$
\begin{aligned}
\theta_۰ \;&=\; \sum_{j=۱}^{n}\log\left(\frac{۱-\phi_{j|y=۱}}{۱-\phi_{j|y=۰}}\right) \;+\; \log\left(\frac{\phi_y}{۱-\phi_y}\right)\\
\theta_j \;&=\; \log\left(\frac{(\phi_{j|y=۱})(۱-\phi_{j|y=۰})}{(\phi_{j|y=۰})(۱-\phi_{j|y=۱})}\right),\quad j=۱,\ldots,n.
\end{aligned}
$$

پرسش ۳ (۱۵ نمره) دو مدل زیر از رگرسیون لجستیک را برای یک دسته‌بندی دودویی با تابع سیگموید در نظر بگیرید

$$
g(z) = \frac{۱}{۱+e^{-z}}\ :
$$

- مدل ۱:

$$
P(Y=۱ \mid X, w_۱, w_۲) = g(w_۱X_۱ + w_۲X_۲)
$$

- مدل ۲:

$$
P(Y=۱ \mid X, w_۱, w_۲) = g(w_۰ + w_۱X_۱ + w_۲X_۲)
$$

ما سه نمونه‌ی آموزشی داریم:

$$
\begin{gathered}
x^{(۱)} = [۱,۱]^T,\quad x^{(۲)} = [۱,۰]^T,\quad x^{(۳)} = [۰,۰]^T\\
y^{(۱)} = ۱,\quad y^{(۲)} = -۱,\quad y^{(۳)} = ۱
\end{gathered}
$$

(آ) (۷نمره) آیا مهم است که نمونه‌ی سوم در مدل ۱ چه برچسبی داشته باشد؟ یعنی اگر برچسب نمونه‌ی سوم را به ۱− تغییر دهیم، آیا مقدار آموخته‌شده‌ی $\mathbf{w} = (w_۱, w_۲)$ متفاوت خواهد شد؟ آیا در مدل ۲ مهم است؟ پاسخ خود را به طور خلاصه توضیح دهید. (راهنما: به مرز تصمیم روی صفحه‌ی دوبعدی فکر کنید.)

(ب) (۸نمره) اکنون فرض کنید مدل رگرسیون لجستیک (مدل ۲) را بر اساس n نمونه‌ی آموزشی $x^{(۱)}, \dots, x^{(n)}$ و برچسب‌ها $y^{(۱)}, \dots, y^{(n)}$ با بیشینه‌سازی log-likelihood جریمه‌دار زیر آموزش دهیم:

$$\sum_i \log P(y^{(i)} \mid x^{(i)}, \mathbf{w}) - \frac{\lambda}{۲}\|\mathbf{w}\|^۲ = \sum_i \log g(y^{(i)}\mathbf{w}^T x^{(i)}) - \frac{\lambda}{۲}\|\mathbf{w}\|^۲$$

برای λ بزرگ ، جمله‌های log-likelihood به صورت توابع خطی از $\mathbf{w}$ رفتار می‌کنند:

$$\log g(y^{(i)}\mathbf{w}^T x^{(i)}) \approx \tfrac{۱}{۲} y^{(i)}\mathbf{w}^T x^{(i)}.$$

log-likelihood جریمه‌دار را با استفاده از این تقریب (در مدل ۱) بیان کنید، و عبارت برآوردگر MLE $\hat{\mathbf{w}}$ را بر حسب λ و داده‌های آموزشی $\{x^{(i)}, y^{(i)}\}_{i=۱}^n$ به‌دست آورید. بر اساس این، توضیح دهید $\mathbf{w}$ با افزایش λ چگونه رفتار می‌کند. (فرض کنید هر $x^{(i)} = (x_۱^{(i)}, x_۲^{(i)})^T$ و $y^{(i)}$ یا ۱ است یا ۱−.)

پاسخ

(آ) در مدل ۱ اهمیتی ندارد، زیرا $x^{(۳)} = (۰, ۰)$ باعث می‌شود که $w_۱x_۱ + w_۲x_۲$ صفر باشد و بنابراین likelihood مدل به مقدار $\mathbf{w}$ وابسته نباشد. اما در مدل ۲ اهمیت دارد.

(ب)

$$\log l(\mathbf{w}) \approx \sum_i \tfrac{۱}{۲} y^{(i)}\mathbf{w}^T\mathbf{x}^{(i)} - \tfrac{\lambda}{۲}\|\mathbf{w}\|^۲$$

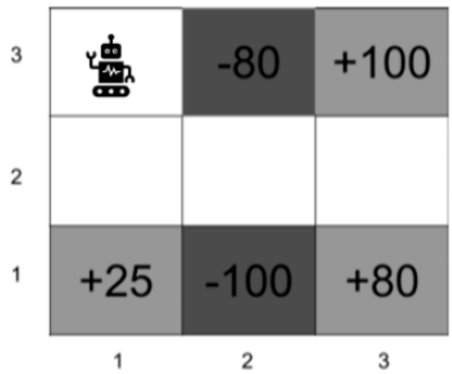
$$\frac{\partial}{\partial w_۱}\log l(\mathbf{w}) \approx \tfrac{۱}{۲}\sum_i y^{(i)}x_۱^{(i)} - \lambda w_۱ = ۰$$

$$\frac{\partial}{\partial w_۲}\log l(\mathbf{w}) \approx \tfrac{۱}{۲}\sum_i y^{(i)}x_۲^{(i)} - \lambda w_۲ = ۰$$

$$\mathbf{w} = \frac{۱}{۲\lambda}\sum_i y^{(i)}\mathbf{x}^{(i)}$$

پرسش ۴ (۲۴ نمره)

محیط زیر را در نظر بگیرید. در هر یک از استیت‌ها عامل می‌تواند یکی از کنش‌های بالا، پایین، چپ و راست را به طور قطعی انجام دهد و به استیت هدف برسد. همچنین اگر یک کنش موجب شود عامل از محیط خارج شود، آن کنش انجام نمی‌شود و عامل سر جای خود باقی می‌ماند. استیت‌هایی که در محیط زیر با عدد مشخص شده‌اند، استیت‌های ترمینال هستند و با ورود عامل به آنها، پاداشی برابر با عدد داخل آنها دریافت می‌کند. ورود به سایر حالات پاداشی برابر صفر خواهد داشت. همچنین فرض کنید که ضریب تخفیف γ برابر 0.5 است، نرخ یادگیری α برابر 0.5است. همچنین فرض کنید عامل از خانه (۱،۳) شروع به کار می‌کند.



(آ) (۶نمره)
مقدار $V^*((۳, ۲))$، $V^*((۲, ۲))$ و $V^*((۱, ۲))$ را بدست آورید. همچنین بر اساس این مقادیر، سیاست بهینه را برای این سه استیت تعیین کنید.
اکنون می‌خواهیم با استفاده از روش Approximate Q-Learning این مساله را بررسی کنیم. مجموعه فیچرها و وزن‌های اولیه آنها در پایین داده شده است.

Feature		Initial Weight (w_i)
$f_0(s,a) = 1$ (Bias feature)		۱۰/۰
$f_1(s,a) = \begin{cases} 1 & \text{a=Left} \\ 2 & \text{a=Right} \\ 0 & \text{otherwise.} \end{cases}$		۲۰/۰
$f_2(s,a) = \begin{cases} 1 & \text{a=Up} \\ 1 & \text{a=Down} \\ 0 & \text{otherwise.} \end{cases}$		۵۰/۰

(ب) (۶نمره) با توجه به نمونه $[s, a, s', r] = [(3,2), Up, (3,3), 100]$ وزن‌ها را آپدیت کنید و مقدار $Q((3,2), Up)$ و $Q((2,2), Up)$ را محاسبه کنید.

(ج) (۶نمره) حالا فرض کنید که وزن‌ها به تعداد n بار آپدیت می‌شوند و بردار $\vec{w}$ حاصل می‌شود. بردار وزن‌ها و کنش a به ازای $Q_{\vec{w}}((2,2), a) = Q_{\vec{w}}((3,2), a)$ برقرار است؟ توضیح دهید.

(د) (۶نمره) با توجه به بخش‌های قبلی، آیا الگوریتم Approximate Q-Learning در این مساله می‌تواند به سیاست بهینه همگرا شود؟ توضیح دهید.

پاسخ

(آ)

$$V^*((3,2)) = 100$$
$$V^*((2,2)) = 0 + \gamma 100 = 50$$
$$V^*((1,2)) = 0 + \gamma^2 100 = 25$$

بنابراین سیاست بهینه به شکل زیر خواهد بود. توجه شود برای حالت (۱, ۲) دو کنش وجود دارند که action-value function آنها برابر با مقدار V^* می‌شود.

$$\pi((3,2)) = \text{Up}$$
$$\pi((2,2)) = \text{Right}$$
$$\pi((1,2)) = \text{Down or Right}$$

(ب) توجه شود استیت (۳, ۳) یک استیت ترمینال است و بنابراین: $Q((3,3), a) = 0$ است.

$$\text{Difference} = r + \gamma \max_{a'} Q(s', a') - Q(s,a)$$
$$= 100 + 0.5 \times 0 - 60 = 40$$

$$w_0 = 10 + 0.5(\text{difference}) f_0((3,2), Up) = 30$$
$$w_1 = 20 + 0.5(\text{difference}) f_1((3,2), Up) = 20$$
$$w_2 = 50 + 0.5(\text{difference}) f_2((3,2), Up) = 70$$

$$Q((3,2), Up) = 30 + 0 + 70 = 100$$
$$Q((2,2), Up) = 30 + 0 + 70 = 100$$

(ج) این تساوی همواره برقرار خواهد بود، چون فیچر های استفاده‌شده به استیت عامل وابسته نیستند و به‌ازای یک کنش یکسان، مقدار Q ها برابر خواهد بود توجه شود که $Q(s,a) = \sum_i w_i f_i(s,a)$

(د) با توجه به بخش قبل می‌توان نتیجه گرفت که سیاستِ به‌دست‌آمده با این الگوریتم برای همهٔ استیت ها یکسان خواهد بود؛ زیرا $\pi(s) = \arg\max_a Q(s,a)$ و طبق بخش قبل برای تمام استیت ها، به‌ازای یک کنشِ خاص، مقدار Q برابر خواهند داشت؛ پس اگر در یک استیت یک کنش مقدارِ Q بیشینه شود، این شرایط در سایر استیت ها نیز برقرار خواهد بود و آن کنش به‌عنوان سیاستِ تمامِ استیت ها انتخاب می‌شود. با این حال،با توجه به بخش الف می‌دانیم که سیاستِ بهینه در مثالِ استیت های (۳, ۲) و (۲, ۲) یکسان نمی باشد؛ ازاین‌رو این بدین معناست که الگوریتمِ Approximate Q-learning با استفاده از این فیچر ها نمی‌تواند همگرا شود و به سیاستِ بهینه برسد.

پرسش ۵ (۱۵ نمره)

شما يك شبكه عصبي دو لايه داريد كه براي طبقه بندي حيوانات آموزش داده شده است. بردار ويژگي ورودي $\vec{x} = [x_1, x_2, x_3, 1]$ است؛ در آن x_1, x_2, x_3 ويژگی‌ها هستند و عدد ۱ نقش باياس را دارد. دو گره پنهان و سه گره خروجي داريم و خروجي شبكه $\vec{y}^* = [y_1^*, y_2^*, y_3^*]$ متناظر با سه كلاس خروجي است؛ به طوري كه $y_1^* = \Pr(\text{dog} \mid \vec{x})$، $y_2^* = \Pr(\text{cat} \mid \vec{x})$ و $y_3^* = \Pr(\text{skunk} \mid \vec{x})$. لايه پنهان از سيگمويد (sigmoid) و لايه خروجي از سافت مكس (softmax) استفاده مي كند.

(آ) (۵نمره) يك توله سگ مالتيز بردار ويژگي $\vec{x} = [۲, ۲۰, -۱, ۱]^\top$ دارد. فرض كنيد همه وزن ها و باياس ها در ابتدا صفر هستند. $\vec{y}^*$ را به دست آوريد.

(ب) (۱۰نمره) فرض كنيد w_{ij} وزني باشد كه گره خروجي i ام را به گره پنهان j ام وصل مي كند. مشتق $\dfrac{\partial y_۲^*}{\partial w_{۲۱}}$ را به دست آوريد. پاسخ را بر حسب y_i^*، w_{ij} و توابع فعال‌سازی h_j براي مقادير مناسب i و j بيان كنيد.

پاسخ

(آ) اگر تمام وزن‌ها و بایاس‌ها صفر باشند، آنگاه خروجی هر گرهِ پنهان برابر است با

$$۰ \times ۲ + ۰ \times ۲۰ + ۰ \times (-۱) + ۰ \times ۱ = ۰$$

با ورودیِ صفر، مقدار سیگموید برابر $۰/۵ = \frac{۱}{۱+\exp(۰)} = \frac{۱}{۱+\exp(-f)}$ می‌شود؛ اما چون وزن‌های لایهٔ آخر نیز همگی صفرند، خروجی‌های لایهٔ آخر نیز همگی صفر می‌شوند. با عبور از تابع فعال‌سازی softmax، هر گرهٔ خروجی مقدار

$$\frac{\exp(۰)}{\sum_{i=۱}^{۳} \exp(۰)} = \frac{۱}{۳}$$

را محاسبه می‌کند. پس $\mathbf{y}^* = \begin{bmatrix} \frac{۱}{۳} \\ \frac{۱}{۳} \\ \frac{۱}{۳} \end{bmatrix}$ می باشد.

(ب) نماد f_i را به‌عنوان مقدار گرهٔ خروجیِ iـام در نظر بگیرید. در این صورت، softmax را می‌توان به‌شکل زیر نوشت:

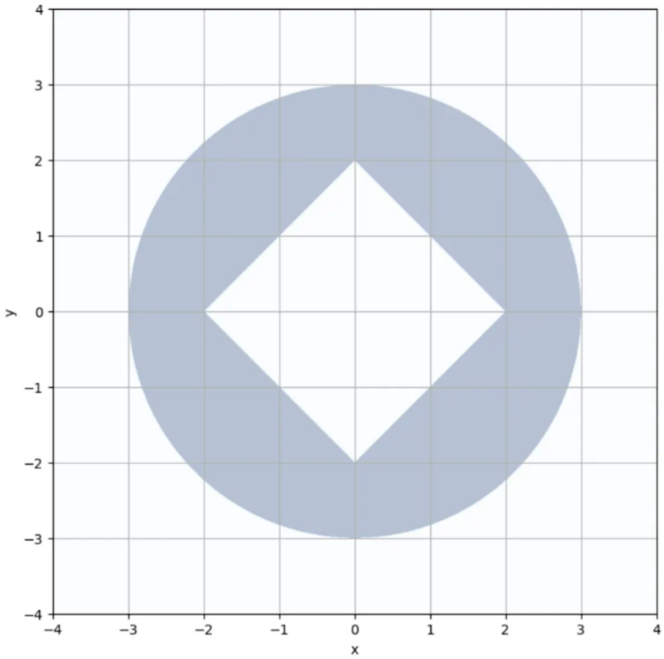
$$y_۲^* = \frac{\exp(f_۲)}{\sum_{j=۱}^{۳} \exp(f_j)}, \qquad f_j = \sum_i w_{ji} h_i.$$

آنگاه:

$$\begin{aligned}
\frac{dy_۲^*}{dw_{۲۱}} &= \frac{۱}{\sum_{i=۱}^{۳} \exp(f_i)} \frac{d\,\exp(f_۲)}{dw_{۲۱}} + \exp(f_۲)\,\frac{d}{dw_{۲۱}}\left(\frac{۱}{\sum_{i=۱}^{۳} \exp(f_i)}\right) \\
&= \frac{۱}{\sum_{i=۱}^{۳} \exp(f_i)}\,\exp(f_۲)\frac{df_۲}{dw_{۲۱}} + \exp(f_۲)\left(-\frac{۱}{\left(\sum_{i=۱}^{۳} \exp(f_i)\right)^۲}\right)\frac{d}{dw_{۲۱}}\left(\sum_{i=۱}^{۳} \exp(f_i)\right) \\
&= \frac{\exp(f_۲)}{\sum_{i=۱}^{۳} \exp(f_i)}\,h_۱ - \frac{\exp(f_۲)}{\left(\sum_{i=۱}^{۳} \exp(f_i)\right)^۲}\frac{d\,\exp(f_۲)}{dw_{۲۱}} \\
&= \frac{\exp(f_۲)}{\sum_{i=۱}^{۳} \exp(f_i)}\,h_۱ - \frac{\exp(f_۲)}{\left(\sum_{i=۱}^{۳} \exp(f_i)\right)^۲}\,\exp(f_۲)\frac{df_۲}{dw_{۲۱}} \\
&= \frac{\exp(f_۲)}{\sum_{i=۱}^{۳} \exp(f_i)}\,h_۱ - \frac{\exp(۲f_۲)}{\left(\sum_{i=۱}^{۳} \exp(f_i)\right)^۲}\,h_۱ \\
&= y_۲^*\,h_۱ - (y_۲^*)^۲\,h_۱.
\end{aligned}$$

در گام سوم از این واقعیت استفاده شده که $\frac{df_۲}{dw_{۲۱}} = h_۱$ (زیرا $f_۲ = \sum_i w_{۲i} h_i$) و هنگام مشتق‌گیری از $\sum_i \exp(f_i)$ تنها جملهٔ متناظر با $f_۲$ به $w_{۲۱}$ وابسته است.

پرسش ۶ (۱۸ نمره) یک شبکه ی عصبی به کمک تابع فعالسازی پله طراحی کنید که به ازای نقاط درون ناحیه مشخص شده $+1$ و در غیراینصورت 0 خروجی دهد. (فرض کنید ورودی های شبکه $x, x^۲, y, y^۲$ هستند.)



پاسخ ناحیه مشخص شده برابر است با تفاضل یک دایره و یک لوزی