مشاوره تحصیلی هیوا

تخصصی‌ترین سایت مشاوره کشور

باسمه تعالی
جمهوری اسلامی ایران
وزارت آموزش و پرورش
باشگاه دانش‌پژوهان جوان

کد دفترچه : ۱

دفترچه سوال رسمی آزمون
واحد سنجش و ارزیابی باشگاه دانش‌پژوهان جوان

علم برای یک ملّت مهم‌ترین ابزار آبرو، پیشرفت و اقتدار است. «امام خامنه‌ای (ره)»

دفترچهٔ سؤالات مرحلهٔ دوم سال تحصیلی ۱۴۰۵-۱۴۰۴

دومین دورهٔ المپیاد هوش مصنوعی

مدت پاسخگویی: ۱۲۰ دقیقه	نوع آزمون: چندگزینه‌ای
	تعداد سؤالات: ۱۵

توضیحات مهم

استفاده از هرنوع ماشین‌حساب ممنوع است.

۱- مشخصات خود را با اطلاعات بالای هر صفحه تطبیق دهید در صورتی که حتی یکی از صفحات پاسخ نامه با مشخصات شما هم‌خوانی ندارد بلافاصله مراقبین را مطلع نمایید.

۲- پاسخ هر سوال را در محل تعیین شده خود بنویسید. چنانچه همه یا قسمتی از جواب سوال را در محل پاسخ سوال دیگری بنویسید به شما نمره ای تعلق نمی گیرد.

۳- با توجه به آنکه برگه های پاسخ نامه به نام شما صادر شده است امکان ارائه هیچ‌گونه برگه اضافه وجود نخواهد داشت. لذا توصیه می‌شود ابتدا سوالات را در برگه چرک نویس، حل کرده و آنگاه در پاسخ‌برگ پاکنویس نمایید.

۴- عملیات تصحیح توسط مصححین پس از برش سربرگ به صورت ناشناس انجام خواهد شد. لذا از درج هرگونه نوشته یا علامت مشخصه که نشان دهنده صاحب برگه باشد، خودداری نمایید. در غیر این صورت تقلب محسوب شده و در هر مرحله ای که باشید از ادامه حضور در المپیاد محروم خواهید شد.

۵- از مخدوش کردن **بارکدها و مربع‌ها در چهارگوشهٔ صفحه در دفترچهٔ پاسخ‌برگ** جداً خودداری کنید. در غیر این صورت برگهٔ شما تصحیح نخواهد شد.

۶- همراه داشتن هر گونه کتاب جزوه یادداشت و لوازم الکترونیکی نظیر تلفن همراه، ساعت هوشمند، دستبند هوشمند و لپ تاپ ممنوع است همراه داشتن این قبیل وسایل حتی اگر از آن استفاده نکنید یا خاموش باشد تقلب محسوب خواهد شد.

۷- این دفترچه شامل ۱۵ سوال و با احتساب جلد ۵ برگ است.

این صفحه جهت استفاده به عنوان چرک‌نویس در نظر گرفته شده است.

۱ کیمیاگر بازار اعظم در تلاش است تا با ساخت یک سامانهٔ هوشمند، میزان رضایت قهرمانان از معجون‌های مختلف را پیش از مصرف تخمین بزند. هر قهرمان دارای یک بردار ویژگی‌های ذاتی u در فضای dـ بعدی است. هر معجون نیز با یک بردار مواد اولیه x در فضای mـ بعدی توصیف می‌شود. سامانهٔ کیمیاگر دارای یک هستهٔ پردازشی (ماتریس W به ابعاد $(d \times m)$ است که مواد اولیه را به یک بردار ویژگی‌های نهایی معجون $v = Wx$ تبدیل می‌کند. میزان رضایت پیش‌بینی‌شدهٔ قهرمان از معجون، برابر با ضرب داخلی بردار ویژگی‌های ذاتی قهرمان و بردار ویژگی‌های نهایی معجون است:

$$\hat{r} = u^T v = u^T W x$$

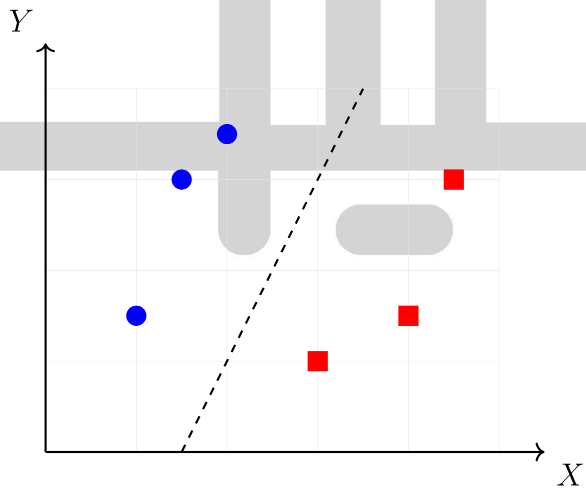
کیمیاگر پس از مصرف معجون توسط قهرمان، میزان رضایت واقعی او یعنی r را از وی می‌پرسد. او برای اینکه سامانهٔ پیش‌بینی خود را دقیق‌تر کند تا در آینده تخمین‌های بهتری داشته باشد، تابع زیان (Loss function) مجذور خطا را برای یک رویارویی به صورت زیر تعریف کرده است:

$$L = \frac{۱}{۲}(r - \hat{r})^{۲}$$

کیمیاگر با الهام از روش نزول گرادیان (Gradient descent)، ماتریس W را به‌روزرسانی می‌کند تا خطای پیش‌بینی را کاهش دهد. هنگامی که کیمیاگر ماتریس W را بر اساس خطای تنها یک قهرمان و یک معجون به‌روزرسانی می‌کند، ماتریس تغییرات اعمال‌شده (یعنی ماتریس گرادیان تابع زیان نسبت به ماتریس W) از نظر هندسی و جبر خطی چه ویژگی بارزی دارد؟

۱) تغییرات به صورت یک ماتریس قطری است که تنها مقادیر ویژهٔ ماتریس W را تغییر می‌دهد.
۲) تغییرات اعمال‌شده همواره یک ماتریس از مرتبهٔ یک (رتبه یک یا Rank-1) است که فضا را تنها در ترکیب یک راستای ورودی و یک راستای خروجی تغییر می‌دهد.
۳) تغییرات از مرتبهٔ کامل (Full-rank) است و تمامی ابعاد ماتریس W را به طور مستقل درگیر می‌کند.
۴) تغییرات تنها بردار ویژگی‌های ذاتی قهرمان را دوران می‌دهد و تأثیری روی فضای مواد اولیه ندارد.

۲ شکل ۱ مجموعه‌ای از داده‌های دوبعدی را به همراه مرز تصمیم یک الگوریتم 1-NN با استفاده از فاصله اقلیدسی نشان می‌دهد.



شکل ۱: مرزهای تصمیم اولیه 1-NN

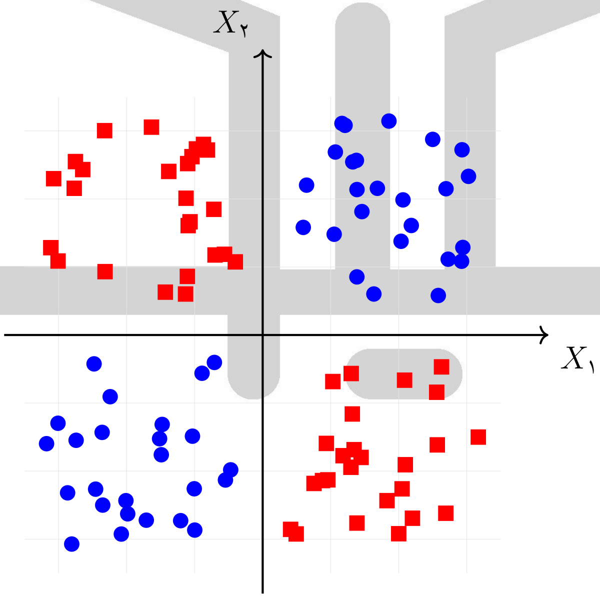
فرض کنید پیش از آموزش مجدد مدل، مقیاس محور X را در تمام داده‌های آموزش در عدد ۱۰ ضرب کنیم (یعنی $X'_{new} = ۱۰X$)، اما مقیاس محور Y ثابت بماند. از نظر هندسی، مرزهای تصمیم جدید نسبت به فضای اولیه چه تغییری خواهند کرد؟

۱) تغییری نمی‌کنند، زیرا الگوریتم KNN نسبت به مقیاس‌بندی خطی مقاوم (Scale-invariant) است.
۲) مرزها بیشتر به موازات محور Y کشیده می‌شوند (تقریباً عمودی می‌شوند).
۳) مرزها بیشتر به موازات محور X کشیده می‌شوند (تقریباً افقی می‌شوند).
۴) نواحی مربوط به یک کلاس، کاملاً به شکل دایره‌های کامل در می‌آیند.

۳ یک سیستم هوشمند کشاورزی برای دسته‌بندی کیفیت سیب‌ها از یک بردار ویژگی ۳ـ بعدی استفاده می‌کند. فرض کنید ویژگی‌های سیب‌های کلاس صادراتی منحصراً در فضای داخلی یک چندوجهی محدب (Convex polytope) قرار دارند که با ۵ وجه مسطح محصور شده است. تیم توسعه می‌خواهد یک شبکه عصبی پیش‌خور (Feedforward neural network) با یک لایه پنهان و یک نورون در لایه خروجی طراحی کند.
اگر تمام نورون‌های شبکه (لایه پنهان و خروجی) از تابع فعال‌ساز آستانه (Threshold) استفاده کنند و فرض کنیم تمامی نورون‌ها دارای پارامتر بایاس (Bias) هستند، برای طراحی این شبکه به گونه‌ای که تضمین‌کننده جداپذیری دقیق و هندسی این چندوجهی باشد، در مجموع حداقل به چند نورون نیاز است؟

۱) ۶
۲) ۵
۳) ۱۵
۴) چنین تفکیکی با این ساختار شبکه غیرممکن است.

۴ مهندسان یک شرکت خودروسازی خودران در حال طراحی ماژول هم‌جوشی حسگرها (Sensor fusion) هستند. توزیع داده‌های مربوط به دو حسگر در یک فضای ویژگی دوبعدی $(X_۱, X_۲)$ در شکل ۲ نشان داده شده است که در آن کلاس وضعیت عادی (دایره‌ها) در ربع اول و سوم، و کلاس وضعیت خطر (مربع‌ها) در ربع دوم و چهارم سیستم مختصات قرار دارند (پدیده XOR).



شکل ۲: توزیع داده‌ها در مسئله غیرخطی XOR

برای تفکیک دقیق و ۱۰۰ درصدی این داده‌ها بر روی مجموعه آموزش، حداقل تعداد لایه‌های پنهان (Hidden layers) برای یک شبکه عصبی پیش‌خور (Feedforward neural network) با توابع فعال‌ساز ReLU و حداقل عمق (Depth) یک درخت تصمیم (Decision tree) استاندارد (با شرط‌های تک‌متغیره موازی با محورها و فرض اینکه گره ریشه در عمق صفر قرار دارد) به ترتیب از راست به چپ چقدر باید باشد؟

۱) ۱ لایه پنهان - عمق ۳
۲) ۱ لایه پنهان - عمق ۲
۳) ۲ لایه پنهان - عمق ۲
۴) ۲ لایه پنهان - عمق ۳

۵ یک مهندس یادگیری ماشین در حال پیاده‌سازی یک خط لوله فشرده‌سازی (Compression Pipeline) برای بردارهای تعبیه (Embeddings) استخراج‌شده از یک مدل زبانی بزرگ (LLM) است. او ماتریس کوواریانس داده‌ها (C) را در فضای d بعدی محاسبه می‌کند. فرض کنید مقادیر ویژه این ماتریس به صورت متمایز

$$\lambda_۱ > \lambda_۲ > \cdots > \lambda_d > ۰$$

مرتب شده‌اند.

به دلیل یک خطای برنامه‌نویسی (استفاده از argmin به جای argmax در توابع کتابخانه‌ای)، الگوریتم او به جای انتخاب مهم‌ترین مؤلفه‌ها برای کاهش ابعاد (PCA)، کل داده‌های d بعدی را تنها روی یک محور یک‌بعدی، که بردار ویژه متناظر با **کوچکترین** مقدار ویژه (λ_d) است، تصویر (Project) می‌کند.

از نظر ریاضی، مقدار دقیق واریانس داده‌های جدیدِ تصویرشده در این راستای اشتباه، چقدر خواهد بود؟

۱) $\lambda_d^۲$ ۲) $\frac{\lambda_d}{\sum_{i=۱}^{d} \lambda_i}$ ۳) $\mathrm{Tr}(C) - \lambda_d$ ۴) λ_d

۶ در رصدخانهٔ اعظم، طیف نوری سیارات ناشناخته به صورت یک بردار ویژگی x در فضای d بعدی ثبت می‌شود. هدف، تشخیص وجود یا عدم وجود K ناهنجاری کیهانی مختلف در این سیارات است. یک سیاره می‌تواند همزمان چندین ناهنجاری داشته باشد یا کاملاً عادی باشد (دسته‌بندی چندبرچسبی یا Multi-label classification با بردار برچسب $y \in \{۰, ۱\}^K$).

مدل یادگیری ماشین برای هر ناهنجاری (kـام)، یک امتیاز خام (Logit) به نام z_k تولید می‌کند. احتمال وقوع هر ناهنجاری با تابع هموارساز سیگموید (Sigmoid) به صورت $p_k = \frac{۱}{۱+e^{-z_k}}$ محاسبه می‌شود. برای مقابله با شرایط خاص کیهان، تابع زیان (Loss function) برای هر سیاره با استفاده از یک ضریب جبرانی β به شکل زیر تعریف شده است:

$$L = -\sum_{k=۱}^{K} \left[\beta y_k \ln(p_k) + (۱ - y_k)\ln(۱ - p_k)\right]$$

سیستم با محاسبهٔ مشتقات (گرادیان) این تابع زیان، پارامترهای مدل را در خلاف جهت شیب به‌روزرسانی می‌کند.

فرض کنید ناهنجاری‌های کیهانی پدیده‌هایی به‌شدت نادر هستند؛ یعنی از بین هزاران ناهنجاری ممکن (K بسیار بزرگ)، هر سیاره به ندرت حتی یک ناهنجاری دارد. در روزهای اول یادگیری که پارامترهای مدل با اعداد تصادفی بسیار نزدیک به صفر مقداردهی شده‌اند، اگر کیهان‌شناس به اشتباه ضریب جبرانی را خنثی کند (یعنی $\beta = ۱$)، از نظر ریاضی و هندسی چه اتفاقی برای فرآیند پیش‌بینی سیستم می‌افتد؟

۱) سیستم به دلیل تعداد زیاد برچسب‌های صفر، در یک نقطهٔ بهینهٔ محلی (Local minimum) گرفتار شده و مشتقات در همان گام اول کاملاً صفر می‌شوند.
۲) به دلیل استقلال عبارات در تابع زیان، برچسب‌های صفر تنها بر ناهنجاری‌های غایب تأثیر گذاشته و اختلالی در روند یادگیری ناهنجاری‌های نادر ایجاد نمی‌کنند.
۳) خروجی‌های مدل به سرعت به سمت احتمال ۰/۵ همگرا شده و سیستم به دلیل تعادل گرادیان‌ها در این نقطه، قابلیت تفکیک ناهنجاری‌ها را از دست می‌دهد.
۴) یک نیروی عظیم ناشی از خیل عظیم برچسب‌های صفر، تمام امتیازات را به شدت در جهت منفی هل داده و مدل

دچار کوری (رد کردن همهٔ ناهنجاری‌ها) می‌شود.

۷ در یک پژوهشگاه هوش مصنوعی، دو گروه از محققان برای پیش‌بینی یک پدیده پیچیده بر اساس d ویژگی ورودی ($\mathbf{x} \in \mathbb{R}^d$)، دو معماری متفاوت برای لایه اول شبکه عصبی خود پیشنهاد داده‌اند.

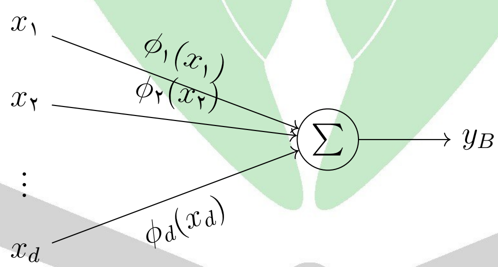
گروه اول پیروان شبکه‌های کلاسیک هستند. در این معماری، ویژگی‌های ورودی ابتدا با ضرایب وزنیِ قابل یادگیری ترکیب خطی شده و سپس از یک تابع فعال‌ساز ثابت (σ) عبور می‌کنند:

$$y_A = \sigma(w_۱x_۱ + w_۲x_۲ + \cdots + w_dx_d)$$

گروه دوم ساختاری نوین طراحی کرده‌اند. آن‌ها به جای ترکیب خطی، روی مسیر هر ویژگی یک شبکه کوچک و مستقل (ϕ_i) قرار داده‌اند که غیرخطی بوده و در طول آموزش تکامل می‌یابد. خروجی نهایی آن‌ها، صرفاً مجموع خروجی این شبکه‌های مستقل است:

$$y_B = \phi_۱(x_۱) + \phi_۲(x_۲) + \cdots + \phi_d(x_d)$$

فرض کنید توابع ϕ_i و σ اسکالر و دو بار مشتق‌پذیر هستند. شمای هندسی معماری نوین در شکل زیر ترسیم شده است:



برای ارزیابی اینکه آیا این مدل‌ها ظرفیت درک «تعاملات متقاطع» (Cross-interactions) بین ویژگی‌ها را دارند، محققان از یک ماتریس ارزیابی (ماتریس هسین) متشکل از مشتقات دوم استفاده می‌کنند. در این ماتریس مربعی، درایه سطر i و ستون j به صورت مشتق دوم $\frac{\partial}{\partial x_i}\left(\frac{\partial y}{\partial x_j}\right)$ محاسبه می‌شود.

درایه‌های خارج از قطر اصلی نشان می‌دهند که آیا تغییر در ویژگی iام، حساسیتِ مدل نسبت به ویژگی jام را تغییر می‌دهد یا خیر. اگر این مقادیر صفر باشند، شبکه متغیرها را کاملاً مستقل از هم می‌بیند و توانایی درک الگوهای ترکیبی را در این لایه ندارد.

با محاسبه تحلیلی این ماتریس برای هر دو خروجی y_A و y_B، کدام گزاره در توصیف تفاوت ظرفیت یادگیری این دو معماری اکیداً درست است؟

۱) ماتریس مشتقات مدل نوین یک ماتریس چگال (Dense) از رتبه کامل (Full-rank) است که نشان می‌دهد این مدل در همان لایَه اول تمام تعاملات متقاطع متغیرها را درک می‌کند؛ اما این ماتریس در مدل کلاسیک قطری (Diagonal) است و تنها اثرات مستقل را می‌بیند.

۲) ماتریس مشتقاتِ مدل نوین یک ماتریس قطری (Diagonal) است، به این معنا که مدل صرفاً جمع‌پذیر است و تعاملات متقاطعِ در آن صفر است؛ در مقابل، این ماتریس در مدل کلاسیک یک ماتریس چگال (Dense) اما از رتبه یک (Rank-1) است که تعاملات را تنها در راستای بردار وزن‌ها مدل می‌کند.

۳) ماتریس هر دو مدل قطری (Diagonal) است، اما درایه‌های روی قطر در مدل نوین توابعی به شدت غیرخطی هستند در حالی که در مدل کلاسیک این درایه‌ها مقادیری ثابت‌اند که باعث محدودیت در یادگیری می‌شود.

۴) در هر دو معماری، ماتریس مشتقات همواره یک ماتریس متعامد (Orthogonal) است تا مقادیر گرادیان در

طول فرآیند یادگیری دچار پدیده محوشدگی (Vanishing Gradient) نشوند.

۸ در یک سیاره دوردست، «کتابدار اعظم» یک مدل زبانی بزرگ (LLM) مبتنی بر معماری ترانسفورمر با میلیاردها پارامتر است که تمام دانش بشری را در وزن‌های خود ذخیره کرده است. سندیکای مورخان متوجه می‌شود که این مدل به دلیل وجود داده‌های مسموم در زمان آموزش، یک واقعیت تاریخی مهم را کاملاً اشتباه یاد گرفته است. به دلیل هزینه نجومی محاسبات، امکان آموزش مجدد (Retraining) یا تنظیم ظریف (Fine-tuning) کل مدل وجود ندارد.

مهندسان هوش مصنوعی تصمیم می‌گیرند از تکنیک نوین «ویرایش جراحی دانش» استفاده کنند. آن‌ها کشف می‌کنند که دانش‌های واقعی در لایه‌های پیش‌خور (Feed-Forward) شبکه به عنوان یک حافظه انجمنی خطی (Linear Associative Memory) ذخیره می‌شوند؛ جایی که ماتریس وزن W، بردار نمایشگر یک مفهوم مانند x (مثلاً 'پایتخت مریخ') را به بردار خروجی y (مثلاً 'شهر المپوس') نگاشت می‌کند ($y = Wx$).

برای تصحیح این تک‌واقعیت، مهندسان می‌خواهند خروجی مدل برای ورودی خاص x_* به مقدار جدید و صحیح y_* تغییر کند ($W_{new}x_* = y_*$)، اما همزمان خروجی برای تمام مفاهیم قبلی که بردارهایشان بر x_* عمود هستند، کاملاً دست‌نخورده باقی بماند تا مدل دچار فراموشی فاجعه‌بار نشود.

اگر بردار خطای خروجی را به صورت $e = y_* - Wx_*$ تعریف کنیم، از منظر جبر خطی، ماتریس تغییرات وزن‌ها (یعنی $\Delta W = W_{new} - W$) باید چه رتبه‌ای (Rank) داشته باشد و فرمول ریاضی آن کدام است؟

۱) یک ماتریس رتبه کامل (Full-Rank) با فرمول $\Delta W = ex_*^T + x_*e^T$

۲) یک ماتریس رتبه کامل (Full-Rank) با فرمول $\Delta W = I - \frac{ex_*^T}{\|x_*\|^۲}$

۳) یک ماتریس رتبه یک (Rank-1) با فرمول $\Delta W = \frac{ex_*^T}{\|x_*\|^۲}$

۴) یک ماتریس رتبه یک (Rank-1) با فرمول $\Delta W = \frac{x_*e^T}{\|e\|^۲}$

۹ در یک بازارچه تاریخی، رقابتی پنهان میان دو عامل برقرار است: ابتدا «سندیکای عطاران» قرار دارد؛ آن‌ها عصاره‌های معطری می‌سازند و تلاش می‌کنند توزیع «رده‌های بویایی» محصولاتشان (P_S) را دقیقاً منطبق بر توزیع رده‌های بویایی عطرهای اصیل دربار (P_R) جلوه دهند. در سوی دیگر، «بازرس اعظم دربار» قرار دارد؛ او یک نمونه عطر با رده بویایی مشخص x را استشمام می‌کند و یک عدد احتمالاتی $D(x) \in (۰,۱)$ خروجی می‌دهد. این عدد، میزان اطمینان بازرس به «اصیل بودن» عطر است.

رده‌های بویایی عطرها تنها در سه سطح دسته‌بندی می‌شوند: $x \in \{۱,۲,۳\}$. توزیع واقعی عطرهای اصیل دربار (P_R) در این رده‌ها به این شرح است: رده اول (رایحه پایه) ۱۰٪، رده دوم (رایحه میانی) ۶۰٪، رده سوم (رایحه بالایی) ۳۰٪.

اما توزیع محصولات فعلی سندیکای عطاران (P_S) هنوز بی‌نقص نیست و با الگوی اصیل تفاوت دارد: رده اول ۴۰٪، رده دوم ۴۰٪، رده سوم ۲۰٪.

بازرس اعظم، تابع سود خود را بر اساس ارزش امید ریاضیِ تشخیص‌های صحیح تعریف کرده و همواره سعی می‌کند مقدار رابطه زیر را بیشینه کند:

$$V(D) = \sum_{x\in\{۱,۲,۳\}} [P_R(x)\ln(D(x)) + P_S(x)\ln(۱ - D(x))]$$

الف) اگر بازرس اعظم، استراتژی قضاوت خود را به صورت کاملاً بهینه تنظیم کند تا تابع سود $V(D)$ در شرایط فعلی بازار بیشینه شود، او به عطری با رایحه میانی ($x = ۲$) چه احتمال خروجی‌ای ($D^*(۲)$) اختصاص خواهد داد؟

ب) فرض کنید در آینده، سندیکای عطاران به قدری در ترکیب عناصر پیشرفت کنند که توزیع محصولات آن‌ها با توزیع عطرهای اصیل دقیقاً برابر شود ($P_S = P_R$). در این نقطه تعادل، اگر بازرس همچنان استراتژی خود را

بهینه انتخاب کند، مقدار نهایی تابع سود او (V) دقیقاً به چه عددی همگرا خواهد شد؟

۱) **الف)** $D^*(۲) = ۰/۶۰$ ، **ب)** تابع سود نهایی برابر با $-\ln(۲)$ است.
۲) **الف)** $D^*(۲) = ۰/۴۰$ ، **ب)** تابع سود نهایی برابر با $-\ln(۲)$ است.
۳) **الف)** $D^*(۲) = ۰/۶۰$ ، **ب)** تابع سود نهایی برابر با $-\ln(۴)$ است.
۴) **الف)** $D^*(۲) = ۰/۴۰$ ، **ب)** تابع سود نهایی برابر با $-\ln(۴)$ است.

۱۰ یک سیستم هوشمند پزشکی، داروی بیماران را از بین دو داروی A و B انتخاب می‌کند. میزان موفقیت داروی A قطعاً برابر با ۰/۷ است. داروی B تازه وارد بازار شده است و با توجه به شرایط تولید، ممکن است از نوع باکیفیت (با احتمال موفقیت ۰/۹) یا از نوع کم‌کیفیت (با احتمال موفقیت ۰/۳) باشد. پیش از شروع کار، سیستم احتمال پیشین (Prior probability) باکیفیت یا کم‌کیفیت بودن داروی B را برابر (هر کدام ۰/۵) فرض می‌کند.
سیستم در دو آزمایش اولیه، داروی B را تجویز کرده است که نتیجه آن یک موفقیت و یک شکست بوده است. سیستم با استفاده از این مشاهدات، باور خود نسبت به باکیفیت بودن داروی B را به روش بیزی به‌روزرسانی می‌کند.
اکنون برای تجویز به بیمار سوم، سیستم از استراتژی $\epsilon-\text{greedy}$ با $\epsilon = ۰/۲$ استفاده می‌کند. در این استراتژی، سیستم ابتدا برای هر دارو برآورد فعلی خود از احتمال موفقیت آن (امید ریاضی موفقیت با توجه به اطلاعات به‌روزرسانی‌شده) را محاسبه می‌کند؛ سپس با احتمال $۱-\epsilon$ دارویی را که بالاترین احتمال موفقیت موردانتظار را دارد به عنوان داروی برتر انتخاب کرده و با احتمال ϵ یکی از دو دارو را به صورت کاملاً تصادفی (هر کدام با شانس مساوی) برمی‌گزیند.
احتمال اینکه درمان برای بیمار سوم موفقیت‌آمیز باشد، چقدر است؟

۱) ۰/۷۱۸
۲) ۰/۶۵۶
۳) ۰/۶۹۰
۴) ۰/۶۷۸

۱۱ مریخ‌نوردی در حال کاوش سطح مریخ است. حسگرهای این کاوشگر در هر مشاهده از یک سنگ، یک نمونه دادهٔ ۳ـبعدی X استخراج می‌کنند. برای کاهش مصرف انرژی در ارسال اطلاعات به زمین، مریخ‌نورد پیش‌تر یک شبکه خودکدگذار (Autoencoder) خطی با یک لایه پنهان (شامل تنها یک نورون) را روی توزیع آماری داده‌های مریخ آموزش داده است تا این داده‌های ۳ـبعدی را فشرده کرده و سپس در زمین بازسازی کند. فرض کنید میانگین داده‌ها در هر ۳ بعد صفر است.
ماتریس کوواریانس توزیع داده‌های آموزش دارای مقادیر ویژه $\lambda_1 = ۹$، $\lambda_2 = ۴$ و $\lambda_3 = ۱$ با بردارهای ویژهٔ یکه و متعامد متناظر v_1، v_2 و v_3 است. هر نمونه دادهٔ جدید قابل بیان به فرم $X = c_1v_1 + c_2v_2 + c_3v_3$ است. طبق مدل‌سازی آماری، ضرایب c_1, c_2, c_3 در واقع متغیرهای تصادفی مستقلی هستند که هر یک از آن‌ها به صورت مستقل می‌تواند مقادیر ۰، ۱ و ۲ را به ترتیب با احتمالات ۱/۲، ۱/۴ و ۱/۴ اتخاذ کند.
در یکی از مشاهدات جدید، سیستم عیب‌یاب مریخ‌نورد پیش از فشرده‌سازی گزارش می‌دهد که مجذور نرم نمونهٔ مشاهده‌شده دقیقاً برابر با $\|X\|^2 = ۵$ است. با فرض این شرایط، چقدر احتمال دارد که مجذور خطای بازسازی این شبکه ($E = \|X - \hat{X}\|^2$) در زمین، بزرگتر یا مساوی ۴ باشد؟

۱) ۱/۳
۲) ۱/۲
۳) ۳/۴
۴) ۲/۳

۱۲ فرض کنید یک مدل هوش مصنوعی برای تشخیص حملات سایبری در یک شبکه بزرگ طراحی شده است. می‌دانیم احتمال وقوع یک حمله (نرخ پایه یا Prevalence) در کل ترافیک شبکه برابر با π است ($\pi \ll ۱$).
این مدل توانسته است نرخ یادآوری (Recall) معادل R را ثبت کند؛ یعنی کسر R از کل حملات واقعی را به درستی حمله تشخیص می‌دهد. مدیر امنیت شبکه شرط مهمی را تعیین کرده است: «اعتبار هشدارهای سیستم باید بالا باشد»، به این معنی که اگر مدل ترافیکی را به عنوان حمله پیش‌بینی کرد، باید با احتمال حداقل P (معیار Precision)، آن ترافیک واقعاً یک حمله باشد.
با توجه به مقادیر داده شده، بیشترین مقدار مجاز برای نرخ هشدار کاذب (False Positive Rate یا FPR) این مدل برای برآورده کردن شرط مدیر شبکه کدام است؟
یادآوری تعاریف پایه:

- True Positive (TP): تعداد حملاتی که به درستی حمله تشخیص داده شده‌اند.
- False Positive (FP): تعداد ترافیکهای عادی که به غلط حمله تشخیص داده شده‌اند.
- False Negative (FN): تعداد حمله‌هایی که به غلط ترافیک عادی تشخیص داده شده‌اند.
- True Negative (TN): تعداد ترافیکهای عادی که به درستی ترافیک عادی تشخیص داده شده‌اند.
- True Positive Rate یا Recall: کسر حملات واقعی که به درستی تشخیص داده شده‌اند ($\frac{TP}{TP+FN}$).
- Precision: کسر هشدارهای سیستمی که واقعاً حمله بوده‌اند ($\frac{TP}{TP+FP}$).
- False Positive Rate: کسر ترافیک‌های عادی که به اشتباه حمله تشخیص داده شده‌اند ($\frac{FP}{FP+TN}$).

۱) $FPR \leqslant \frac{R\pi(۱-P)}{P(۱-\pi)}$ ۲) $FPR \leqslant \frac{R(۱-\pi)}{P\pi}$ ۳) $FPR \leqslant \frac{P\pi(۱-R)}{R(۱-\pi)}$ ۴) $FPR \leqslant \frac{R\pi}{P(۱-\pi)}$

۱۳ یک بانک برای تشخیص تراکنش‌های متقلبانه، یک مدل رگرسیون لجستیک (Logistic Regression) خطی را روی داده‌هایش آموزش داده و بردار وزن بهینه w بدست آمده است. اگر به دلایل امنیتی، کل بردارهای ویژگی نمونه‌ها در یک ماتریس متعامد (Orthogonal Matrix) Q ضرب شوند ($Q^TQ = I$)، چه اتفاقی برای بردار وزن‌های جدید w' و مرز تصمیم (Decision Boundary) می‌افتد؟

۱) بردار وزن‌ها به Q^Tw تغییر می‌کند و احتمالات خروجی مدل برای هر نقطه تغییر خواهند کرد.
۲) مسئله بهینه‌سازی به دلیل دوران فضا دیگر محدب (Convex) نبوده و همگرایی به جواب بهینه تضمین نمی‌شود.
۳) بردار وزن‌ها به Qw تغییر می‌کند، مرز تصمیم دوران می‌یابد، اما احتمالات پیش‌بینی‌شده برای هر نقطه ثابت می‌ماند.
۴) بردار وزن‌ها ثابت می‌ماند ($w' = w$) زیرا فاصله اقلیدسی (Euclidean Distance) تحت تبدیل متعامد حفظ می‌شود.

۱۴ یک سیستم هوش مصنوعی در یک بیمارستان برای تشخیص یک بیماری نادر بر اساس بردار ویژگی‌های بالینی X، یک مدل رگرسیون لجستیک (Logistic regression) را آموزش داده است. در این مدل، خروجی پیش از اعمال تابع هموارساز سیگموئید معادل با لگاریتم شانس (Log-odds) است و به صورت $W^TX + b$ محاسبه می‌شود.
در زمان آموزش، احتمال شیوع این بیماری در داده‌های بیمارستان ۲۰ درصد بود ($P(Y = ۱) = ۰/۲$). مدل پس از آموزش به‌خوبی کالیبره شده و مقدار بایاس آن برابر با b به دست آمده است.
اکنون این سیستم به یک کلینیک تخصصی منتقل شده که در آن به دلیل نوع مراجعین، احتمال ابتلای اولیه بیماران ۸۰ درصد است ($P(Y = ۱) = ۰/۸$). فرض کنید توزیع ویژگی‌ها به شرط کلاس، یعنی $P(X|Y)$ در هر دو مکان کاملاً یکسان است. برای اینکه مدل بدون تغییر در وزن‌های تخصیص‌یافته به ویژگی‌ها (W) مجدداً برای کلینیک جدید کالیبره شود، مقدار بایاس جدید از نظر ریاضی دقیقاً باید چقدر تنظیم شود؟

۱) $b - \ln(۴)$ ۲) $b - \ln(۱۶)$ ۳) $b + \ln(۴)$ ۴) $b + \ln(۱۶)$

۱۵ یک سیستم هوش مصنوعی در یک طرح کشاورزی دقیق (Precision agriculture) وظیفه دارد یک گلخانه صنعتی را به منظور جایگذاری دو پمپ آبیاری، به دو ناحیه کاملاً مجزا تقسیم کند. این سیستم برای یافتن مکان بهینه پمپ‌ها از الگوریتم K-Means ($K = ۲$) استفاده می‌کند تا تابع هدف، یعنی مجموع مجذور فواصل هر نقطه از زمین تا مرکز پمپاژ ناحیه خود را کمینه کند.
با فرض اینکه N گیاه (نقطه داده) به طور یکنواخت (Uniform) در سراسر گلخانه توزیع شده‌اند، تابع هدف به صورت زیر تعریف می‌شود:

$$J = \sum_{k=۱}^{۲} \sum_{x_i \in C_k} \|x_i - \mu_k\|^۲$$

که در آن C_k نشان‌دهنده ناحیه (خوشه) kـ‌ام، μ_k مختصات مرکز پمپاژ آن ناحیه و x_i مختصات هر گیاه است.
این گلخانه یک مستطیل با طول $X \in [۰, ۴]$ و عرض $Y \in [۰, ۶]$ است و تراکم گیاهان و نیاز آبی در تمام نقاط آن کاملاً یکسان است. پس از اجرای الگوریتم K-Means با مقداردهی‌های اولیه تصادفی متعدد برای مختصات مراکز پمپاژ، اگر سیستم در نهایت به بهینه مطلق سراسری (Global optimum) برسد، مرز هندسی تصمیم (خط جداکننده دو ناحیه آبیاری) دقیقاً چه معادله‌ای خواهد داشت؟

۱) خط افقی $Y = ۳$ (جداسازی در راستای محور Y)
۲) خط عمودی $X = ۲$ (جداسازی در راستای محور X)
۳) خط قطری $Y = \frac{۳}{۲}X$ (جداسازی وابسته به هر دو محور)
۴) خطوط $X = ۲$ و $Y = ۳$ تابع هدف برابری تولید کرده و مرز نهایی به مقداردهی اولیه بستگی دارد.