@RayanAzmoon

کد کنترل

731

A

731A

«اگر دانشگاه اصلاح شود مملکت اصلاح می‌شود.»
امام خمینی (ره)

صبح جمعه
۹۷/۱۲/۳
دفترچهٔ شمارهٔ (۱)

جمهوری اسلامی ایران
وزارت علوم، تحقیقات و فنّاوری
سازمان سنجش آموزش کشور

آزمون ورودی دورهٔ دکتری (نیمه‌متمرکز) ــ سال ۱۳۹۸

رشتهٔ مهندسی کامپیوتر ــ هوش مصنوعی ــ کد (۲۳۵۶)

تعداد سؤال: ۴۵

مدت پاسخ‌گویی: ۱۵۰ دقیقه

عنوان مواد امتحانی، تعداد و شمارهٔ سؤالات

ردیف	مواد امتحانی	تعداد سؤال	از شمارهٔ	تا شمارهٔ
۱	مجموعه دروس تخصصی: ساختمان داده‌ها و طراحی الگوریتم‌ها ــ شناسایی الگو ــ یادگیری ماشین	۴۵	۱	۴۵

این آزمون نمرهٔ منفی دارد.

استفاده از ماشین‌حساب مجاز نیست.

حق چاپ، تکثیر و انتشار سؤالات به هر روش (الکترونیکی و...) پس از برگزاری آزمون، برای تمامی اشخاص حقیقی و حقوقی تنها با مجوز این سازمان مجاز می‌باشد و با متخلفین برابر مقررات رفتار می‌شود.

۱۳۹۸

* داوطلب گرامی، عدم درج مشخصات و امضا در مندرجات جدول ذیل، به‌منزلهٔ عدم حضور شما در جلسهٔ آزمون است.

اینجانب ..... با شمارهٔ داوطلبی ..... در جلسهٔ این آزمون شرکت می‌نمایم.

امضا:

۱- یک ماتریس دو بعدی $n \times n$ از اعداد داده شده، که اعداد هر سطر و هر ستون آن مرتب شده است. به ازای عدد داده شدهٔ x، جست و جوی x در این ماتریس در چه زمانی امکان‌پذیر است؟

۱) $O(n)$ ۲) $O(\log n)$ ۳) $O(\log^2 n)$ ۴) $O(n \log n)$

۲- می‌خواهیم بزرگ‌ترین زیر دنبالهٔ مشترک دو دنبالهٔ $a_1,...,a_n$ و $b_1,...,b_m$ را محاسبه کنیم. فرض کنید $L(i,j)$ برابر طول بزرگترین زیر دنبالهٔ مشترک $a_1,...,a_i$ و $b_1,...,b_j$ باشد. کدام‌یک از تعاریف بازگشتی زیر درست است؟

الف) $L(n,m) = \max(L(n-1,m), L(n,m-1), L(n-1,m-1)+1 \text{ if } a_n = b_m)$

ب) $L(n,m) = \max(L(n-1,m), L(n-1,k-1)+1)$ که k برابر بزرگ‌ترین عددی است که $a_n = b_k$ (درصورت عدم وجود $k = 0$ خواهد بود)

فرض کنید $L(i,0) = L(0,i) = 0$ و $L(i,-1) = -1$ برای هر $i \geq 0$.

۱) فقط الف ۲) فقط ب ۳) الف و ب ۴) هیچ‌یک از الف و ب

۳- فرض کنید یک آرایهٔ دوبعدی $m \times n$ در اختیار داریم که هر ردیف آن مرتب شده است. فرض کنید همهٔ اعداد متمایز هستند. می‌خواهیم k-اُمین عدد در آرایه را پیدا کنیم. در چه زمانی این کار امکان‌پذیر است؟

۱) $O(m.n)$ ۲) $O(\log n \log m)$

۳) $O(\log n + \log m)$ ۴) $O(m(\log n + \log m))$

۴- اگر ظرفیت همهٔ یال‌ها در یک شبکه برابر C باشد، زمان اجرای الگوریتم فورد ـ فالکرسون برای محاسبه شار بیشینه از مبدا s به مقصد t در بدترین حالت کدام مورد خواهد بود؟

(فرض کنید تعداد رئوس و یال‌های گراف به ترتیب n و m هستند و درجه خروجی s برابر k باشد. همچنین فرض کنید در هر مرحله الگوریتم بیش‌ترین شار ممکن را از مسیر انتخاب شده، عبور می‌دهد.)

۱) $O(kC + m + n)$ ۲) $O(kC(m+n))$

۳) $O(C(m+n))$ ۴) $O(k(m+n))$

۵- فرض کنید ۱۳۹۷ نقطه متمایز روی محور اعداد حقیقی داده شده است. می‌خواهیم این ۱۳۹۷ نقطه را طوری رنگ‌آمیزی کنیم که به ازای هر بازهٔ $[a,b]$ روی محور اعداد حقیقی، از بین نقاطی که در این بازه قرار گرفته‌اند حداقل یک نقطه وجود داشته باشد که رنگ آن با بقیه نقاط داخل بازه متفاوت باشد. حداقل چند رنگ برای این کار نیاز است؟

۱) ۶ ۲) ۱۱ ۳) ۳۸ ۴) ۱۳۹۷

۶- فرض کنید یک B-tree داریم با n برگ که درجهٔ هر گره حداقل $\log n$ و حداکثر $2\log n - 1$ است. هزینهٔ جست‌وجوی یک عدد در این درخت کدام است؟ (فرض کنید کلیدها داخل هر گره میانی در یک لیست پیوندی یک سویه ذخیره شده‌اند.)

۱) $O(\log n)$ ۲) $O(\log n \log\log n)$

۳) $O(\log n \log^2 \log n)$ ۴) $O(\log^2 n / \log\log n)$

۷- فرض کنید یک گراف وزن‌دار همبند داده شده است که وزن یال‌ها متمایز است. یک یال را امن گوییم اگر در هیچ دوری حضور نداشته باشد و یک یال را خطرناک گوییم اگر سنگین‌ترین یال در یک دور باشد. کدام‌یک از دو گزاره زیر درست است؟

الف) هر یال امن عضو درخت پوشای کمینه است.

ب) هر یال خطرناک عضو درخت پوشای کمینه نیست.

۱) الف ۲) ب ۳) الف و ب ۴) هیچ‌یک از الف و ب

۸- گراف جهت‌دار G با n رأس و m یال داده شده است. هر رأس i از گراف ارزشی به اندازهٔ V_i دارد. به‌ازای هر رأس i از گراف، با ارزش‌ترین رأسی که از رأس i قابل دسترسی است را W_i می‌نامیم. می‌خواهیم تمام W_iها را به‌ازای i از ۱ تا n محاسبه کنیم. این کار در چه زمانی قابل انجام است؟ (بهترین گزینه را انتخاب کنید.)

۱) $O(m+n)$ ۲) $O(m+n^2)$

۳) $O(n(m+n))$ ۴) $O(m+n\log n)$

۹- یک درخت جست‌وجوی دودویی با n گره داریم که به علت نویز، اعداد ذخیره شده در برخی از گره‌های آن تغییر کرده است. تنها عملی که می‌توان برای اصلاح این درخت انجام داد جابه‌جا کردن مقادیر ذخیره شده در یک گره و یکی از فرزندان آن است. کمینهٔ تعداد اعمال مورد نیاز برای تبدیل درخت به یک درخت دودویی جست‌وجو در بدترین حالت کدام است؟ (دقت کنید که درخت اولیه لزوماً متوازن نیست.)

۱) $O(n)$ ۲) $O(n^2)$

۳) $O(n\log n)$ ۴) $O(n\log\log n)$

۱۰- زوج‌های مرتب زیر را در نظر بگیرید:

$$(10,A),(2,B),(5,C),(7,D),(8,E),(1,F),(4,G)$$

فرض کنید درختی داریم که براساس مؤلفه‌های اول این زوج‌ها یک هرم کمینه، و براساس مؤلفه‌های دوم یک درخت جست‌وجوی دودویی است. ارتفاع این درخت کدام است؟

۱) ۲ ۲) ۳ ۳) ۴ ۴) ۵

۱۱- در گراف همبند و بدون جهت G با n رأس، از یک رأس مشخص BFS و DFS را اجرا می‌کنیم، ترتیب ملاقات رئوس در هر دو اجرا یکسان شده است. در این خصوص کدام مورد درست است؟

۱) گراف G فقط ستاره‌ای است.

۲) گراف G فقط یک مسیر است.

۳) تعداد یال‌های G از $O(n)$ است.

۴) تعداد یال‌های G می‌تواند $\Omega(n\log n)$ باشد.

@RayanAzmoon

۱۲- فرض کنید گراف G همبند، بدون جهت و وزن‌دار است به‌طوری‌که می‌تواند دور منفی هم داشته باشد. در مورد مسئله پیدا کردن کوتاه‌ترین مسیر از یک رأس به رأس دیگر طوری که از هر رأسی حداکثر یک‌بار عبور کند، چه می‌توان گفت؟

۱) یک مسئله ان‌پی ـ تمام است.

۲) یک مسئله ان‌پی ـ سخت است.

۳) به علت وجود دور منفی در گراف لزوماً چنین مسیری وجود ندارد.

۴) در زمان چندجمله‌ای برحسب اندازهٔ ورودی می‌توان مسئله را حل کرد.

۱۳- تعدادی فایل با اندازه‌های مشخص را می‌خواهیم روی نوار ذخیره کنیم. فرض کنید $f_1,\ldots,f_n$ به‌ترتیب (از راست به چپ) روی نوار ذخیره شده باشند، هزینه خواندن فایل f_i برابر $\sum_{k=1}^{i} |f_k|$ خواهد بود که $|f_k|$ برابر طول فایل f_k می‌باشد. فرض کنید قرار است هر فایل تنها یک‌بار خوانده شود. می‌خواهیم مجموع هزینه را کمینه کنیم. بدین منظور از الگوریتم حریصانه زیر استفاده می‌کنیم. فایل‌ها را به‌ترتیب اندازه از کوچک به بزرگ روی نوار ذخیره می‌کنیم. اگر n تعداد فایل‌ها باشد، کم‌ترین n که به ازای آن الگوریتم فوق لزوماً درست کار نمی‌کند، کدام است؟

۱) ۲

۲) ۳

۳) ۴

۴) به‌ازای هر n، الگوریتم فوق بهینه عمل می‌کند.

۱۴- آرایهٔ A شامل n عنصر داده شده است. می‌دانیم که تمام عناصر به‌جز $\sqrt[4]{n}$ عنصر، در محل مرتب‌شدهٔ خود هستند ولی مکان عناصر نامرتب را نمی‌دانیم. این آرایه را در چه زمانی می‌توان مرتب کرد؟

۱) $O(n)$ ۲) $O(n\sqrt[4]{n})$

۳) $O(n\log n)$ ۴) $O(\sqrt{n}\log n)$

۱۵- پیمایش‌های پیش ترتیب و پس‌ترتیب یک درخت دودویی به‌صورت زیر است:

preorder: abcdefg , postorder : cbfgeda

با فرض ذخیره‌سازی درخت در آرایه (ریشه در خانه‌ی ۱ و فرزندان گره اندیس در اندیس‌های $2i$ و $2i+1$)، حداکثر تعداد خانه‌های بلا استفاده قبل از محل آخرین گره در آرایه کدام است؟

۱) ۵

۲) ۶

۳) ۷

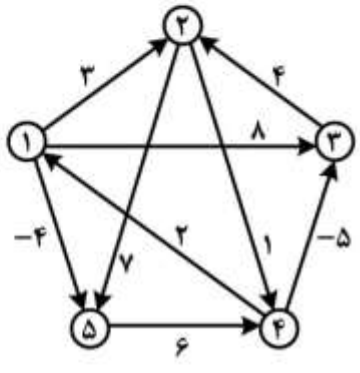
۴) ۸

@RayanAzmoon

۱۶- یک ساختمان داده را درنظر بگیرید که از دو پشته S_1 و S_2 تشکیل شده است. این ساختمان داده دو عمل درج و استخراج را پشتیبانی می‌کند. به هنگام درج عنصر x در این ساختمان داده، $push(S_1, x)$ را اجرا می‌کنیم، به هنگام اسخراج اگر S_2 خالی نبود، $Pop(S_2)$ را اجرا می‌کنیم، در غیر این‌صورت همهٔ عناصر داخل S_1 را پاپ و داخل S_2 پوش می‌کنیم و بعد دستور $Pop(S_2)$ را اجرا و به عنوان خروجی دستور استخراج در نظر می‌گیریم. اگر دو پشته در ابتدا خالی باشد و n عمل درج و استخراج به‌ترتیب دلخواه انجام شود، هزینه سرشکن این عمل‌ها کدام است و ساختمان داده فوق چه ساختمان داده‌ای را پیاده‌سازی می‌کند؟

۱) $O(1)$ و صف
۲) $O(n)$ و صف
۳) $O(1)$ و پشته
۴) $O(n)$ و پشته

۱۷- اگر الگوریتم جانسون برای یافتن کوتاه‌ترین مسیر بین تمام رأس‌های گراف را روی گراف وزن‌دار زیر اجرا کنیم، پس از اجرای مرحلهٔ تغییر وزن یال‌ها در الگوریتم، وزن جدید یال بین رأس‌های ۱ و ۵ که وزن اولیهٔ آن ۴- است، کدام مقدار خواهد شد؟



۱) ۴-
۲) ۴
۳) ۳
۴) ۰

۱۸- اگر رشتهٔ **ababbcbaabdadad** را به‌وسیلهٔ الگوریتم هافمن کدگذاری کنیم، طول رشتهٔ حاصل چند بیت خواهد بود؟

۱) ۲۰
۲) ۲۴
۳) ۲۸
۴) ۳۰

۱۹- فرض کنید n عدد صحیح k بیتی داریم. فرض کنید هزینه جمع، تفریق و مقایسه دو عدد k بیتی $O(k)$ است. اگر $k = O(\log n)$ باشد، کدام گزینه در مورد الگوریتم‌های مرتب‌سازی درست است؟

۱) زمان اجرای الگوریتم مرتب‌سازی شمارشی $O(n)$ است.
۲) زمان اجرای الگوریتم مرتب‌سازی سریع $O(n \log n)$ است.
۳) زمان اجرای الگوریتم مرتب‌سازی ادغامی $O(n \log^2 n)$ است.
۴) زمان اجرای الگوریتم مرتب‌سازی درجی $\Omega(n^2 \log^2 n)$ است.

۲۰- آرایه‌ای شامل n عدد داریم. اگر در اجرای الگوریتم مرتب‌سازی ادغامی روی این آرایه هرگاه تعداد اعداد کمتر از $\sqrt{n}$ شد، روال بازگشتی را متوقف و از الگوریتم مرتب‌سازی درجی استفاده کنیم، زمان اجرای الگوریتم کدام مورد خواهد بود؟ (فرض کنید زمان اجرای الگوریتم مرتب‌سازی درجی از مرتبهٔ $O(m^2)$ است که m تعداد اعداد می‌باشد.)

۱) $O(n^2)$
۲) $O(n\sqrt{n})$
۳) $O(n \log n)$
۴) $O(n \log n \sqrt{n})$

۲۱- ماتریس $X_{N\times d}$ حاوی N نمونه d ـ بعدی است. چنان‌چه ω اولین مؤلفه اساسی این ماتریس باشد، اولین بردار ویژه (به‌ترتیب نزولی مقادیر ویژه) ماتریس xx^{+} چه خواهد بود؟

۱) ω ۲) $X\omega$ ۳) $\lambda\omega$ ۴) $\frac{\omega}{\|\omega\|}$

۲۲- فرض کنید $x_1, x_2, \ldots, x_n$ نمونه‌هایی باشند که به‌صورت i.i.d از یک توزیع گوسی انتخاب شده‌اند و هر یک دارای میانگین ناشناختهٔ θ و واریانس معلوم σ_0^2 هستند. اگر پارامتر θ خود دارای یک توزیع گوسی با میانگین μ و واریانس σ^2 باشد، تخمین MAP برای پارامتر θ کدام است؟

۱) $\hat{\theta} = \frac{\sigma^2(\frac{1}{n}\sum_{i=1}^{n} x_i)}{\sigma^2 n + \sigma_0^2}$

۲) $\hat{\theta} = \frac{\sigma^2(\sum_{i=1}^{n} x_i) + \sigma_0^2\mu}{\sigma^2 n + \sigma_0^2}$

۳) $\hat{\theta} = \frac{\sigma^2(\sum_{i=1}^{n} x_i) + \sigma_0^2\mu}{n(\sigma^2 + \sigma_0^2)}$

۴) $\hat{\theta} = \frac{\sigma^2(\frac{1}{n}\sum_{i=1}^{n} x_i) + \sigma_0^2\mu}{\sigma_0^2 n + \sigma^2}$

۲۳- مجموعه داده $S = S_1 \cup S_2$ به‌صورت زیر داده شده است. اولین مؤلفه اساسی این مجموعه داده کدام است؟

$S_1 = \{(x, 0/1x) \mid -10 \le x \le 10\}$ و $S_2 = \{(x, -0/1x) \mid -10 \le x \le 10\}$

۱) $[1, 0]$

۲) $[1, 0/1]$

۳) $(0, 1)$

۴) $(1, -0/1)$

۲۴- کدام جمله در مورد قانون یادگیری پرسپترون در شبکه‌های عصبی نادرست است؟

۱) قانون پرسپترون به ازای هر داده آموزشی که نادرست دسته‌بندی شده باشد، وزن‌های شبکه را یک دور به هنگام می‌کند.

۲) قانون پرسپترون برای آموزش نورون با خروجی دو حالته (نورون با خروجی sign) به‌کار می‌رود.

۳) اگر داده‌ها به‌صورت خطی تفکیک‌پذیر باشند و نرخ یادگیری به اندازه کافی کوچک باشد، استفاده از قانون پرسپترون به خطای صفر همگرا می‌شود.

۴) اگر داده‌ها به‌صورت خطی تفکیک‌پذیر نباشند و نرخ یادگیری به اندازه کافی کوچک باشد، استفاده از قانون پرسپترون به حداقل خطای ممکن همگرا می‌شود.

@RayanAzmoon

۲۵- یک مسئله دو کلاسه با احتمال پیشین $P(c_1)=P(c_2)=\frac{1}{2}$ را در نظر بگیرید که $p(x|c_1)$ به‌صورت یکنواخت درون یک ابرکره به شعاع یک و $p(x|c_2)$ به‌صورت یکنواخت درون ابرکره دیگر به شعاع یک توزیع شده‌اند. این دو ابرکره ۱۰ واحد از هم فاصله دارند. فرض کنید $D=\{x_1,...,x_n\}$ مجموعه‌ای از n نمونه مستقل بر چسب‌دار از این دو کلاس باشد و $D_k(x)=\{x'_1,...,x'_k\}$ مجموعه k نزدیک‌ترین همسایه x باشد. متوسط احتمال خطا $(P_n(e))$ وقتی k عددی فرد باشد، کدام است؟

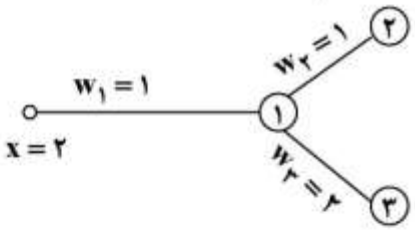
۱) $\frac{1}{2^n}\sum_{j=0}^{k}\binom{n}{j}$

۲) $\frac{1}{2^k}\sum_{j=0}^{k}\binom{n}{j}$

۳) $\frac{1}{2^n}\sum_{j=0}^{(k-1)/2}\binom{n}{j}$

۴) $\frac{1}{2^k}\sum_{j=0}^{(k-1)/2}\binom{n}{j}$

۲۶- در شبکه عصبی زیر نورون ۱ دارای تابع فعالیت خطی و نورون‌های شماره ۲ و ۳ دارای تابع فعالیت سیگموئید $(\sigma(x)=\frac{1}{1+e^{-x}})$ هستند. در لحظهٔ t مقادیر ورودی و وزن‌های شبکه در شکل نشان داده شده است. اگر در زمان آموزش مقادیر $\Delta w_2=2$ و $\Delta w_3=1$ به‌دست آمده باشد، مقدار Δw_1 کدام است؟



۱) ۱

۲) ۲

۳) ۳

۴) ۴

۲۷- داده‌های زیر را در فضای برداری دو بعدی برای کلاس‌های A و B درنظر بگیرید.

A : {(۲,۱) (۳ ,۲) (۴,۳)}

B : {(۳,۳) (۱ ,۳) (۱,۱) (۰,۱)}

با استفاده از الگوریتم دسته‌بندی بیز ساده به‌ترتیب مشخص کنید، داده (۳,۱) در کدام کلاس و با چه احتمالی قرار می‌گیرد؟ (احتمال اولیه دو کلاس A و B را مساوی لحاظ کنید)

۱) A , $\frac{1}{14}$

۲) A , $\frac{3}{14}$

۳) B , $\frac{1}{21}$

۴) B , $\frac{2}{21}$

۲۸- فرض کنید مسأله‌ای داریم که از الگوریتم **EM** برای پیدا کردن تخمین بیشینه درست نمایی **(Maximum likelihood estimates)** با متغیرهای پنهان استفاده می‌کند. حال می‌خواهیم الگوریتم را تصحیح کنیم که تخمین **MAP** را پیدا نماید. کدام گام الگوریتم **EM** نیاز به تصحیح دارد؟

۱) گام Expectation

۲) گام Maximization

۳) گام Expectation, Maximization

۴) الگوریتم نیاز به تصحیح ندارد.

۲۹- اگر نقاط (x_1, x_2) در فضای دو بعدی دارای توزیع گوسی با میانگین $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ و ماتریس کوواریانس $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma \\ \sigma & \sigma_2^2 \end{bmatrix}$ باشند، میانگین توزیع شرطی $p(x_1 \mid x_2 = a)$ کدام است؟

۱) $p(x_1 \mid x_2 = a) = N(\bar{\mu}, \bar{\sigma}^2), \bar{\mu} = \mu_1 + \frac{\sigma}{\sigma_2^2}(a - \mu_2)$

۲) $p(x_1 \mid x_2 = a) = N(\bar{\mu}, \bar{\sigma}^2), \bar{\mu} = \mu_1 + \frac{\sigma}{\sigma_2^2}(a + \mu_2)$

۳) $p(x_1 \mid x_2 = a) = N(\bar{\mu}, \bar{\sigma}^2), \bar{\mu} = \mu_1 + \frac{\sigma}{\sigma_1^2}(a - \mu_2)$

۴) $p(x_1 \mid x_2 = a) = N(\bar{\mu}, \bar{\sigma}^2), \bar{\mu} = \mu_1 + \frac{\sigma}{\sigma_1^2}(a + \mu_2)$

۳۰- یک هسته گاوسی با رابطهٔ $k(x_1, x_2) = e^{\frac{-\|x_1 - x_2\|^2}{2}}$ را در نظر بگیرید. فاصله بین دو بردار $x_1 = [1,1]^T$ و $x_2 = [2,2]^T$ در فضای جدید (فضایی که حاصل‌ضرب داخلی نگاشت برابر $k(x_1, x_2)$ است) کدام است؟

۱) ۰

۲) $\sqrt{2}$

۳) e^{-1}

۴) $(2 - \frac{2}{e^2})^{-2}$

۳۱- کدام گزینه درباره مستقل از مقیاس بودن (به این معنی که با ضرب همه مقادیر در یک عدد ثابت نتیجه تغییر نکند) منظم سازها درست است؟

۱) L_1, L_0 مستقل از مقیاس هستند اما L_2 مستقل از مقیاس نیست.

۲) L_2, L_1 مستقل از مقیاس هستند اما L_0 مستقل از مقیاس نیست.

۳) L_0 مستقل از مقیاس است اما L_1, L_2 مستقل از مقیاس نیستند.

۴) L_2 مستقل از مقیاس است اما L_1, L_0 مستقل از مقیاس نیستند.

۳۲- یک مجموعه داده دو بعدی به همراه خروجی مورد انتظار متناظر با هر داده به‌صورت $\{(x_{i,1}, x_{i,2}) : y_i\}$ در اختیار داریم. می‌خواهیم از مدل رگرسیونی $y_i = w_1^2 x_{i,1} + w_2^2 x_{i,2}$ برای پیش‌بینی y استفاده کنیم. مقدار بهینه w_1 وقتی که از معیار کمینه مربعات خطا به عنوان تابع هدف استفاده می‌کنیم، از کدام رابطه حاصل می‌شود؟

۱) $w_1 = \sqrt{\frac{\sum_i x_{i,1} y_i - 2 w_2 x_{i,1} x_{i,2}}{\sum_i x_{i,2}^2}}$

۲) $w_1 = \sqrt{\frac{\sum_i x_{i,1} y_i - w_2^2 x_{i,1} x_{i,2}}{\sum_i x_{i,1}^2}}$

۳) $w_1 = \sqrt{\frac{\sum_i x_{i,1} x_{i,2} - w_2^3 x_{i,1} y_i}{\sum_i x_{i,2}^2}}$

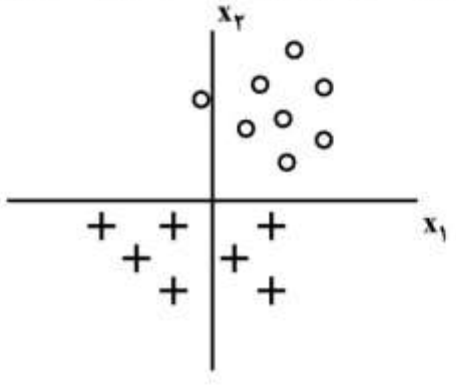
۴) $w_1 = \sqrt{\frac{\sum_i x_{i,1}^2 - w_2^2 x_{i,1} x_{i,2}}{\sum_i x_{i,1} y_i}}$

۳۳- اگر بخواهیم مساله دسته‌بندی دو کلاسه شکل زیر را به کمک مدل Simple linear logistic regression(Simple LR) حل کنیم از آنجا که داده‌ها خطی جدایی‌پذیر هستند، خطای مدل بر روی داده‌های آموزشی برابر با صفر است. حال اگر بخواهیم همین مساله را با بیشینه‌سازی مدل Regularized linear logistic regression(Regularized LR) زیر به ازای مقادیر خیلی بزرگ C حل کنیم، کدام مورد درست است؟

$$\sum_{i=1}^{n} \log(P(y_i \mid x_i, w_0, w_1, w_2)) - C w_j^2$$

$$\text{Where } p(y=1 \mid \vec{x}, \vec{w}) = \frac{1}{1+\exp(-w_0 - w_1 x_1 - w_2 x_2)}$$

$$\text{And } j \in \{0,1,2\}$$



۱) اگر $w_j = w_1$ آنگاه خطای مدل Regularized LR بر روی داده‌های آموزشی نسبت به مدل Simple LR ثابت باقی می‌ماند.

۲) اگر $w_j = w_2$ آنگاه خطای مدل Regularized LR بر روی داده‌های آموزشی نسبت به مدل Simple LR ثابت باقی می‌ماند.

۳) اگر $w_j = w_1$ آنگاه خطای مدل Regularized LR بر روی داده‌های آموزشی نسبت به مدل Simple LR کاهش می‌یابد.

۴) اگر $w_j = w_2$ آنگاه خطای مدل Regularized LR بر روی داده‌های آموزشی نسبت به مدل Simple LR کاهش می‌یابد.

۳۴- هنگامی که تعداد نمونه‌های آموزشی به سمت بی‌نهایت میل می‌کنند، کدام مورد درباره مدل آموزش دیده شده درست است؟

۱) بایاس کاهش می‌یابد اما واریانس ثابت می‌ماند.

۲) واریانس کاهش می‌یابد اما بایاس ثابت می‌ماند.

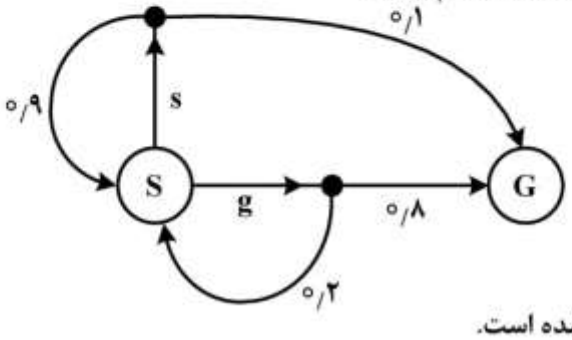
۳) بایاس ثابت می‌ماند اما واریانس افزایش می‌یابد.

۴) بایاس و واریانس هر دو کاهش می‌یابند.

۳۵- در فرایند آموزش دسته‌بند توسط الگوریتم **AdaBoost** یادگیر ضعیف L_0 ساخته شده است. در ابتدا ۲۸۰۰ نمونه آموزشی داریم که احتمال انتخاب هر کدام برابر $\frac{1}{2800}$ است. این یادگیر ۱۰۰ نمونه آموزشی را به درستی دسته‌بندی نمی‌نماید. وزن این دسته‌بند در این الگوریتم برابر کدام است؟

۱) $\ln(27)$ ۲) $\ln(28)$ ۳) $\ln(100)$ ۴) $\ln(\frac{27}{28})$

۳۶- یک عامل یادگیر تقویتی در یک محیط تصادفی فعالیت می‌کند. تعامل عامل یادگیر و محیط توسط **MDP** شکل زیر نشان داده شده است. عامل در حالت **S** دارای دو کنش **g,s** است. هنگامی که عامل در حالت **S** قرار دارد پاداش ۰/۱− و هنگامی که در حالت **G** قرار دارد پاداش ۱/۰ را دریافت می‌کند. اعداد روی یال‌ها احتمال گذر را نشان می‌دهند و ضریب کاهش γ برابر با یک است. ارزش حالت **S** کدام است؟



۱) ۰/۸
۲) ۰/۷۹
۳) ۱/۰
۴) ۰/۸۷۵

۳۷- در مساله رگرسیون خطی تابع هدف به‌صورت زیر تعریف شده است.

$$W^* = \arg\min \| XW - y \|_2^2 + \lambda \| w \|_1$$

افزایش λ چه تأثیری روی بایاس و واریانس دارد؟

۱) افزایش واریانس و افزایش بایاس
۲) افزایش واریانس و کاهش بایاس
۳) کاهش واریانس و افزایش بایاس
۴) کاهش واریانس و کاهش بایاس

۳۸- یک فضای تصمیم مارکوف داریم که فضای حالت آن **S** و فضای کنش آن **A** و تابع پاداش آن $R(s,a,s')$ و نرخ کاهش برابر با γ است. هدف یادگیری تابع سیاست یک ربات در محیط واقعی است. اما ربات در اختیار نداریم و تنها یک نرم افزار شبیه‌ساز ربات داریم که تابع گذر آن $T_{sim}(s,a,s')$ است که با تابع گذر ربات واقعی که برابر $T_{real}(s,a,s')$ است، متفاوت است. می‌خواهیم بدون تغییر نرم‌افزار شبیه‌ساز، تعدادی نمونه از شبیه‌ساز جمع‌آوری و برای یادگیری ربات در محیط واقعی استفاده نماییم. برای یادگیری ربات از الگوریتم یادگیری **Q** استفاده می‌شود که برای نمونه (s,a,s',r) به‌صورت زیر است:

$$Q(s,\alpha) \leftarrow (1-\alpha)Q(s,a) + \alpha\left[r + \gamma \max_{a'}(s',a')\right]$$

برای یادگیری مقادیر واقعی **Q**، کدام مورد می‌تواند به‌کار رود؟(α نرخ یادگیری است.)

۱) $Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha \frac{T_{real}(s,a,s')}{T_{sim}(s,a,s')}\left[r + \gamma \max_{a'} Q(s',a')\right]$

۲) $Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha \frac{T_{sim}(s,a,s')}{T_{real}(s,a,s')}\left[r + \gamma \max_{a'} Q(s',a')\right]$

۳) $Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha T_{sim}(s,a,s')\left[r + \gamma \max_{a'} Q(s',a')\right]$

۴) $Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha T_{real}(s,a,s')\left[r + \gamma \max_{a'} Q(s',a')\right]$

۳۹- در مورد خوشه‌بندی مجموعه دادهٔ $\{۰,۴,۵,۲۰,۲۶,۴۰\}$ به دو خوشه، کدام مورد نادرست است؟

۱) نتیجه خوشه‌بندی با استفاده از روش K-means اگر نقطه شروع مراکز خوشه‌ها به‌ترتیب برابر ۰ و ۵۰ باشد، $\{۰,۴,۵,۲۰\}$ و $\{۲۶,۴۰\}$ خواهد بود.

۲) نتیجه خوشه‌بندی با استفاده از روش K-means اگر نقطه شروع مرکز خوشه‌ها به‌ترتیب برابر ۲۰ و ۴۰ باشد، $\{۰,۴,۵,۲۰,۲۶\}$ و $\{۴۰\}$ خواهد بود.

۳) نتیجه خوشه‌بندی با استفاده از روش Single-link به‌صورت $\{۰,۴,۵\}$ و $\{۲۰,۲۶,۴۰\}$ است.

۴) نتیجه خوشه‌بندی با استفاده از روش Complete-link به‌صورت $\{۰,۴,۵\}$ و $\{۲۰,۲۶,۴۰\}$ است.

۴۰- کدام عبارت درست است؟

۱) دو ویژگی مستقل می‌توانند کوواریانس ناصفر داشته باشند.

۲) پس از اعمال PCA ابعاد فضای جدید از یکدیگر مستقل هستند.

۳) دو ویژگی مستقل ممکن است بعد از اعمال تبدیل خطی به هم وابسته شوند.

۴) درحالتی‌که میانگین داده‌ها از داده‌ها کم نشده باشد، پیدا کردن راستایی که واریانس داده‌های افکنش (project) شده روی آن راستا بیشینه است معادل با پیدا کردن راستایی است که افکنش داده‌ها روی آن منجر به کمترین خطای بازسازی می‌شود.

۴۱- تابع هزینه روش خوشه‌بندی K-means به‌صورت $\sum_{k=1}^{k}\sum_{n\in C_k}\| x^{(n)}-\mu_k \|^2$ را در نظر بگیرید. کمینه کردن این تابع معادل با کمینه کردن کدام‌یک از توابع هزینه زیر است؟ (C_k خوشه kام را نشان می‌دهد.)

۱) $\sum_{k=1}^{k}\sum_{n\in C_k}\sum_{n'\in C_k}\| x^{(n)}-x^{(n')} \|^2$

۲) $\sum_{k=1}^{k}\frac{1}{2|C_k|}\sum_{n\in C_k}\sum_{n'\in C_k}\| x^{(n)}-x^{(n')} \|^2$

۳) $\sum_{k=1}^{k}\sum_{n\in C_k}\sum_{n'\in C_k}\| x^{(n)}-x^{(n')} \|^2-\sum_{k=1}^{k}\sum_{k'=1}^{k}\| \mu_k-\mu_{k'} \|^2$

۴) $\sum_{k=1}^{k}\frac{1}{2|C_k|}\sum_{n\in C_k}\sum_{n'\in C_k}\| x^{(n)}-x^{(n')} \|^2-\frac{1}{2k}\sum_{k=1}^{k}\sum_{k'=1}^{k}\| \mu_k-\mu_{k'} \|^2$

۴۲- یک مجموعه داده داریم که دارای N نمونه است و هر کدام از نمونه‌ها دارای n ویژگی دودویی هستند. می‌خواهیم یک درخت تصمیم از روی این مجموعه داده بسازیم. این درخت حداکثر چند برگ خواهد داشت؟

۱) $\max(N,2^n)$ ۲) $\max(\log N,2^n)$

۳) $\min(\log N,2^n)$ ۴) $\min(N,2^n)$

۴۳- در یک مسئله دسته‌بندی دو دسته‌ای، داده‌های هر دسته دارای توزیع گوسی هستند که $\mu_1 = \begin{pmatrix} ۰ \\ ۰ \end{pmatrix}, \mu_2 = \begin{pmatrix} ۳ \\ ۳ \end{pmatrix}$ میانگین دو دسته و ماتریس کوواریانس دسته‌ها برابر و وارون ماتریس کوواریانس برابر $\sum^{-1} = \begin{bmatrix} ۰/۹۵ & -۰/۱۵ \\ -۰/۱۵ & ۰/۵۵ \end{bmatrix}$ است. کدام مورد فاصله نمونه $x = \begin{pmatrix} ۱ \\ ۲/۲ \end{pmatrix}$ از میانگین دسته‌ها که با (d_1^2, d_2^2) نمایش داده می‌شود را نشان می‌دهد؟ (d_i فاصله نمونه x از میانگین دسته iام است.)

۱) $(۵/۸۴, ۴/۶۴)$ ۲) $(۴/۶۴, ۵/۸۴)$ ۳) $(۳/۶۷, ۲/۹۵)$ ۴) $(۲/۹۵, ۳/۶۷)$

۴۴- در یک مسئله یادگیری مجموعه داده‌ای $S = \{x_1, x_2, ..., x_n\}$ را داریم که میانگین آنها در فضای ورودی برابر است با $\mu = \frac{1}{n}\sum_{L=1}^{n} x_i$ و تابع هسته $K(۰,۰)$ داده شده است. کدام مورد میانگین داده‌ها در فضای ویژگی را نشان می‌دهد؟

۱) $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K(x_i, x_j)$ ۲) $\frac{\mu}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K(x_i, x_j)$

۳) $\frac{1}{n}\sum_{i=1}^{n} K(x_i, x_i)$ ۴) $\frac{\mu}{n}\sum_{i=1}^{n} K(x_i, x_i)$

۴۵- بعد VC یک درخت تصمیم به سه گره دودویی در فضای R^1 (یک بعدی) برابر کدام مورد است؟

۱) ۲ ۲) ۳ ۳) ۴ ۴) ۵

@RayanAzmoon