

330E

کد کنترل

330

E

دفترچه شماره (۱)
صبح جمعه
۹۸/۱۲/۹



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

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

آزمون ورودی دوره دکتری (نیمه‌تمکن) - سال ۱۳۹۹

رشته مهندسی کامپیوتر - هوش مصنوعی - کد (۲۳۵۶)

مدت پاسخ‌گویی: ۱۵۰ دقیقه

تعداد سؤال: ۴۵

عنوان مواد امتحانی، تعداد و شماره سوالات

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

این آزمون نمره منفی دارد.

استفاده از ماشین حساب مجاز نیست.

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

۱۳۹۹

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

اینجانب با شماره داوطلبی با آگاهی کامل، یکسان بودن شماره صندلی خود را با شماره داوطلبی مندرج در بالای کارت ورود به جلسه، بالای پاسخ‌نامه و دفترچه سوالات، نوع و کد کنترل درج شده بر روی دفترچه سوالات و پائین پاسخ‌نامه‌ام را تأیید می‌نمایم.

امضا:

۱- فرض کنید متنی به طول n در اختیار داریم. درخصوص گزاره‌های زیر کدام گزینه صحیح است؟

الف) کد هافمن یک کاراکتر یک بیتی است، اگر و فقط اگر تعداد تکرار آن کاراکتر کمتر از جمع تعداد تکرار بقیه کاراکترها نباشد.

ب) اگر کاراکتری بیشترین تکرار را داشته باشد و تعداد تکرارهای آن بیش از $\frac{n}{3}$ باشد، آنگاه کد هافمن آن کاراکتر تک بیتی است.

(۱) (الف) درست و (ب) درست

(۲) (الف) نادرست و (ب) نادرست

۲- یک گراف کامل 10 رأسی را درنظر بگیرید. که رأس‌های آن از 1 تا 10 شماره‌گذاری شده‌اند. فرض کنید وزن یال بین 0 و 2 برابر $+1$ است. آخرین یال درخت پوشای کمینه که توسط الگوریتم پریم با شروع از رأس 10 اضافه می‌شود، چه وزنی دارد؟

(۱) ۹

(۲) ۱۰

(۳) ۱۱

(۴) ۱۷

۳- کدام الگوریتم مرتبسازی در بهترین حالت، زمان اجرای کمتری دارد؟

Quick Sort (۴) Merge Sort (۳) Selection Sort (۲) Insertion Sort (۱)

برای پیاده‌سازی یک لیست پیوندی حلقوی، کدام ساختمان داده قابل استفاده است؟

(۱) پشته (۲) صف (۳) صف و پشته (۴) هیچ یک از صف و پشته

۴- در پیاده‌سازی متعارف جستجوی عمق اول و جستجوی سطح اول، بهترین از کدام داده ساختارها استفاده می‌شود؟

(۱) پشته و صف (۲) صف و پشته (۳) پشته و لیست (۴) لیست و پشته

۵- مسئله جمع زیرمجموعه بدين شکل تعریف می‌شود: یک مجموعه از اعداد مثبت $S = \{a_1, \dots, a_n\}$ به همراه عدد

W داده شده است. آیا زیرمجموعه‌ای از S بپیدا می‌شود که جمع اعضای آن W شود؟

برای حل این مسئله بروش برنامه‌ریزی پویا یک آرایه دو بعدی $X[1..n, 0..W]$ تعریف می‌کنیم که $X[i, j] = \text{True}$ است. اگر زیرمجموعه‌ای از $S = \{a_1, \dots, a_n\}$ وجود داشته باشد که جمع اعضای آن j شود، در

این خصوصی کدام رابطه درست است؟

$$X[i, j] = X[i-1, j] \wedge X[i, j-a_i] \quad (۲)$$

$$X[i, j] = X[i-1, j] \wedge X[i-1, j-a_i] \quad (۴)$$

$$X[i, j] = X[i-1, j] \vee X[i, j-a_i] \quad (۱)$$

$$X[i, j] = X[i-1, j] \vee X[i-1, j-a_i] \quad (۳)$$

-۷ برای گراف بدون جهت G با n رأس دو مسئله زیر را در نظر بگیرید:

- مسئله A: آیا G یک زیرمجموعه مستقل ۴ رأسی دارد؟

- مسئله B: آیا G یک زیرمجموعه مستقل ۴ – n رأسی دارد؟

در خصوص این دو مسئله کدام مرحله درست است؟

(۱) مسئله A عضو کلاس P و مسئله B عضو کلاس NP-Complete است.

(۲) مسئله A عضو کلاس NP-Complete و مسئله B عضو کلاس P است.

(۳) هر دو مسئله عضو کلاس NP-Complete هستند.

(۴) هر دو مسئله عضو کلاس P هستند.

-۸ فرض کنید $G = (V, E)$ یک گراف بدون جهت و گراف $G' = (V', E')$ یک زیرگراف G است. یال‌های G را بهین

شکل وزن دار می‌کنیم: اگر $e \in E'$ باشد، وزن آن را صفر و در غیر اینصورت ۱ می‌گذاریم. از رأس دلخواه $v \in V'$

الگوریتم دایکسترا را برای محاسبه کوتاهترین مسیر به بقیه رئوس اجرا می‌کنیم. کدام مسئله را می‌توان با استفاده

از طول کوتاهترین مسیرهای محاسبه شده، حل کرد؟

(۱) آیا G' درخت است؟

(۲) آیا G' همبند است؟

(۳) آیا G' تشکیل خوشه می‌دهد؟

(۴) تعداد یال‌ها در کوتاهترین مسیر از V به بقیه رئوس چند است؟

-۹ فرض کنید در داخل یک درخت دودویی جستجو، اعداد ۱ تا ۱۰۰۰ ذخیره شده‌اند و ما می‌خواهیم دنبال عدد

۳۶۵ بگردیم. کدام دنباله (از چپ به راست) نمی‌تواند مسیر جستجو باشد؟

(۱) ۴, ۴۰۱, ۳۸۹, ۲۲۱, ۲۶۸, ۳۸۴, ۳۸۳, ۲۸۰, ۳۶۵

(۲) ۹۲۶, ۲۲۲, ۹۱۲, ۲۴۶, ۹۰۰, ۲۶۰, ۳۶۵

(۳) ۴, ۲۵۴, ۴۰۳, ۴۰۵, ۳۳۲, ۳۴۶, ۳۹۹, ۳۶۵

(۴) ۹۲۷, ۲۰۴, ۹۱۲, ۲۴۲, ۹۱۴, ۲۴۷, ۳۶۵

-۱۰ فرض کنید یک آرایه مرتب از n عدد در اختیار داریم. به ازای یک k داده شده، می‌خواهیم دو عدد a و b از آرایه

را پیدا کنیم که $|a - b| = k$ شود. سریع‌ترین الگوریتم برای حل این مسئله دارای چه مرتبه زمانی است؟

$O(n)$ (۱)

$O(n^2)$ (۲)

$O(\log n)$ (۳)

$O(n \log n)$ (۴)

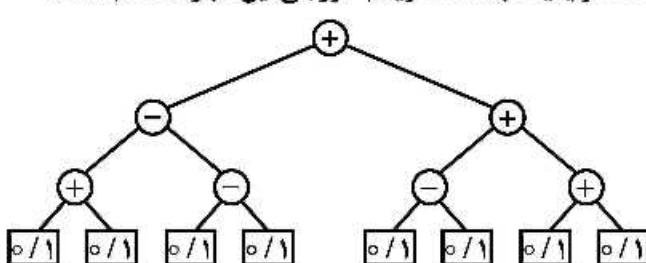
-۱۱ در درخت میانوندی داده شده، مقدار هر برگ می‌تواند صفر یا یک باشد. ماکزیمم خروجی این عبارت کدام است؟

(۱) ۲

(۲) ۳

(۳) ۴

(۴) ۶



-۱۲ می‌دانیم ترتیب شروع و پایان فعالیت‌های H و G و F و C و D و E و A از چپ به راست به صورت $x_s a_s b_s c_s a_e d_s c_e e_s f_s b_e d_e g_s c_e f_e h_s g_e h_e$ می‌باشد. می‌خواهیم این فعالیت‌ها را در تعدادی اتفاق که در اختیار داریم انجام دهیم. یک فعالیت در یک اتفاق قابل انجام است، اگر در تمام مدت زمان آن فعالیت اتفاق بطور کامل در اختیارش باشد. حداقل تعداد اتفاق‌های مورد نیاز برای انجام همه فعالیت‌ها کدام است؟

- (۱) ۳
- (۲) ۴
- (۳) ۵
- (۴) ۶

-۱۳ اعداد صحیح بین ۱ تا ۱۳۹۸ به عنوان ورودی داده شده است. کدام تابع در هم‌ساز، اعداد داده شده را بطور یکنواخت بین ۱۰ خانه جدول در هم‌سازی توزیع می‌کند؟ (یک توزیع یکنواخت است، اگر تفاضل تعداد اعداد نگاشت شده به هر دو خانه از جدول حداقل ۱ باشد.)

$$h(i) = i^7 \bmod 10 \quad (1)$$

$$h(i) = i^4 \bmod 10 \quad (2)$$

$$h(i) = 12i \bmod 10 \quad (3)$$

$$h(i) = 4i^7 + 6 \bmod 10 \quad (4)$$

-۱۴ آرایه A داده شده است. می‌توانیم هر بار دو خانه دلخواه از این آرایه را با هم جایه‌جا کنیم. با حداقل چند جایه‌جایی می‌توان این آرایه را به یک هرم بیشینه تبدیل کرد؟

- (۱) ۰
- (۲) ۱
- (۳) ۲
- (۴) ۳

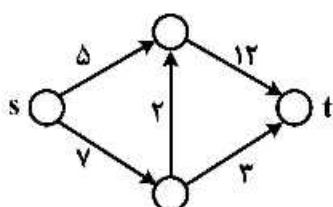
-۱۵ فرض کنید در گراف وزن دار و جهت دار G با n رأس، تنها وزن یال‌های خارج شده از رأس s ممکن است منفی باشد. (البته می‌دانیم گراف دور منفی ندارد). بزرگ‌ترین n که به ازای آن الگوریتم دایکسترا روی هر گراف n رأسی با فرض‌های گفته شده کوتاه‌ترین مسیر از s به بقیه رئوس را درست محاسبه می‌کند، کدام است؟

- (۱) ۲
- (۲) ۳
- (۳) ۴

(۴) به ازای هر n همیشه درست کار می‌کند.

-۱۶ در شبکه داده شده فقط مجاز هستیم ظرفیت یک یال را به هر میزان که بخواهیم افزایش دهیم. با این کار شار بیشینه از s به t را حداقل چه میزان می‌توان افزایش داد؟

- (۱) ∞
- (۲) ۵
- (۳) ۲
- (۴) ۱



- ۱۷- اگر مسئله X عضو کلاس NP-Complete به مسئله Y عضو کلاس P در زمان چندجمله‌ای تبدیل شود، کدام گزینه نادرست است؟

NP = P (۱)

NP – Complete = P (۲)

NP – Hard = NP (۳)

(۴) مسئله ۳ – SAT در زمان چندجمله‌ای حل می‌شود.

- ۱۸- زمان اجرای الگوریتمی به صورت $T(n) = T(4n/11) + T(6n/11) + n$ است. مرتبه زمانی اجرای این الگوریتم کدام است؟

O(n) (۱)

O(n log n) (۲)

O(log n) (۳)

- ۱۹- فرض کنید G یک گراف وزن‌دار و جهت‌دار با n رأس و m یال است. با فرض اینکه فاصله کوتاهترین مسیر برای هر دو رأس را در یک ماتریس $n \times n$ در اختیار داریم، مطلع شده‌ایم که وزن تنها یک یال (u, v) تغییر پیدا کرده است. می‌خواهیم به‌ازای دو رأس مشخص s و t طول کوتاهترین مسیر بین این دو رأس را به‌روز کنیم. این کار را در چه زمانی می‌توان انجام داد؟

O(n) (۱)

O(n+m) (۲)

O(1) (۳)

O(n log n + m) (۴)

- ۲۰- دو هرم کمینه در اختیار داریم که هر یک شامل n عدد است. می‌خواهیم یک هرم کمینه برای همه این $2n$ عدد بسازیم. با چه مرتبه زمانی می‌توان این کار را انجام داد؟ (فرض کنید هرم‌های کمینه با آرایه پیاده‌سازی شده‌اند.)

O(n) (۱)

O(n log n) (۲)

$O(n \log^* n)$ (۳)

$O(n \log \log n)$ (۴)

- ۲۱- کدام گزینه در مورد پیچیدگی زمانی الگوریتم خوشبندی K-Means صحیح است؟ (فرض کنید t تعداد تکرار، k تعداد خوشبندی و n تعداد نمونه‌ها است).

$O(t * n * k^2)$ (۱)

$O(t * n * k)$ (۲)

$O(t * n^2 * k)$ (۴)

- ۲۲- کدام گزینه در خصوص روش SVM نرم (hard margin) و سخت (soft margin) با هسته گاوسی صحیح است؟

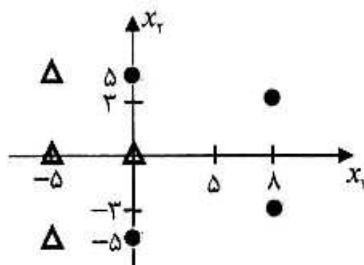
(۱) با کاهش پارامتر σ ، امکان رخداد بیش برآذش بیشتر می‌شود.

(۲) با افزایش پارامتر σ در هسته گاوسی، نرخ خطأ در داده‌های آزمون افزایش می‌یابد.

(۳) با کاهش ضریب جریمه در الگوریتم SVM soft margin، دسته‌بندی دچار بیش برآذش می‌شود.

(۴) با افزایش ضریب جریمه در soft margin SVM میزان داده‌های آموزش نادرست دسته بندی شده کاهش، اما داده‌های آزمون درست دسته بندی شده افزایش می‌یابد.

-۲۳- در مسئله دسته بندی شکل زیر به روش NN-1، با معیارهای فاصله $d_1(x, y)$ و $d_2(x, y)$ در مسئله دسته بندی شکل زیر به روش NN-1، با معیارهای فاصله $d_1(x, y) = \max_i |x_i - y_i|$ و $d_2(x, y) = \sum_{i=1}^D |x_i - y_i|$ به ترتیب از چپ به راست کدام است؟



- (۱) دایره، دایره
- (۲) دایره، مثلث
- (۳) مثلث، دایره
- (۴) مثلث، مثلث

-۲۴- در تخمین چگالی به صورت غیرپارامتریک در فضای \mathcal{X} از پنجره پارزن گاووسی با مقیاس ۱ استفاده شده است. به

$$p(x) = \frac{1}{N} \sum_{n=1}^N \frac{1}{(2\pi)^{D/2}} \exp\left\{-\frac{1}{2}\|x - x_n\|^2\right\}$$

عبارت دیگر

است. فرض کنید داده‌ها را با تبدیل $y = Ax$ نگاشت کنیم. کدام گزینه تخمین چگالی معادل در فضای جدید بر حسب $B = A^{-1}$ است؟

$$p(y) = \frac{1}{N} \sum_{n=1}^N \frac{|B|}{(2\pi)^{D/2}} \exp\left\{-\frac{1}{2}(y - y_n)^T B (y - y_n)\right\} \quad (1)$$

$$p(y) = \frac{1}{N} \sum_{n=1}^N \frac{1}{(2\pi)^{D/2}|B|} \exp\left\{-\frac{1}{2}(y - y_n)^T B (y - y_n)\right\} \quad (2)$$

$$p(y) = \frac{1}{N} \sum_{n=1}^N \frac{|B|}{(2\pi)^{D/2}} \exp\left\{-\frac{1}{2}(y - y_n)^T B^T B (y - y_n)\right\} \quad (3)$$

$$p(y) = \frac{1}{N} \sum_{n=1}^N \frac{1}{(2\pi)^{D/2}|B|} \exp\left\{-\frac{1}{2}(y - y_n)^T B^T B (y - y_n)\right\} \quad (4)$$

-۲۵- فرض کنید در یک مسئله دسته بندی، از تابع $y = w^T x$ برای تخمین برچسب داده‌ها استفاده کنیم. برای یافتن w بهینه از روی داده‌های آموزش $\{(x_1, y_1), \dots, (x_n, y_n)\}$ ، تابع ضرر $L(w)$ را بهینه می‌کنیم. در صورتی که داده‌های آموزش خطی جداپذیر باشند، کدام $L(w)$ تعداد جواب‌های بهینه کمتری دارد؟

$$L_1(w) = \frac{1}{n} \sum_{i=1}^n \max(0, -y_i w^T x_i) \quad (1)$$

$$L_2(w) = \frac{1}{n} \sum_{i=1}^n (w^T x_i)^2 u(-y_i w^T x_i) \quad (2)$$

$$u(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad \text{که در آن } L_2(w) = \frac{1}{n} \sum_{i=1}^n u(-y_i w^T x_i) \quad (3)$$

$$L_3(w) = \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \max(0, -y_i w^T x_i) \quad (4)$$

که در آن C به سمت بینهایت میل می‌کند.

- ۲۶- در یک مسئله دسته بندی دو دسته‌ای، داده‌های دو دسته دارای توزیع گاوی هستند. با توجه به پارامترهای

$$x = \begin{bmatrix} 1/5 \\ 1/5 \end{bmatrix} \text{ از میانگین دسته‌ها } (d_1^T, d_2^T) \text{ کدام است؟}$$

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \sum_1 = \begin{bmatrix} 3 & -3 \\ -3 & 3/5 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \sum_2 = \begin{bmatrix} 3 & 3 \\ 3 & 3/5 \end{bmatrix}$$

(۱۸/۷۵ ۰/۷۵) (۲)

(۴/۵ ۱۸/۷۵) (۴)

(۴/۵ ۴/۵) (۱)

(۰/۷۵ ۱۸/۷۵) (۳)

- ۲۷- یک ماشین بردار پشتیبان خطی سخت (SVM – Hard linear SVM) خطی بدون در نظر گرفتن جریمه در یک مسئله دسته بندی دو دسته‌ای در فضای دوبعدی با n داده، آموزش داده شده است. نتیجه حاصل $k = 2$ بردار پشتیبان بوده است. در صورتی که یک داده برچسبدار دلخواه به مجموعه داده‌های قبلی اضافه کرده و مجدد دسته‌بند را آموزش دهیم، حداکثر چند بردار پشتیبان ممکن است به دست آید؟

k (۱)

n (۳)

$k+1$ (۳)

$n+1$ (۴)

- ۲۸- مسئله دسته بندی دو دسته‌ای با احتمال‌های پیشین برابر ($p(\omega_1) = p(\omega_2)$) و توابع توزیع گاوی به صورت زیر را در نظر بگیرید.

$$p(x|\omega_1) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\right) \quad p(x|\omega_2) = N\left(\begin{bmatrix} 4 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

خط پیشنهادی روش Fisher discriminant به عنوانی که داده‌های دو دسته پس از تصویر بر روی آن خط، به صورت بینهایه جدا شوند، کدام است؟

$$W = \frac{1}{3} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad (۲)$$

$$W = -\frac{2}{3} \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad (۴)$$

$$W = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (۱)$$

$$W = \frac{1}{2} \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad (۳)$$

- ۲۹- روش SVM (hard margin) به صورت زیر به دست آمده است، که ورودی با استفاده از نگاشت Φ به فضایی با ابعاد بالاتر منتقل می‌شود. آیا $y = [1, -1]^T$, $x = [+1, 1]^T$ می‌توانند بردار پشتیبان باشند؟

$$f(\Phi(x)) = 4\Phi_1(x) + 9\Phi_2(x) + 4\Phi_3(x) \quad \Phi(x) = \begin{bmatrix} \Phi_1(x) = x_1^2 \\ \Phi_2(x) = x_2^2 \\ \Phi_3(x) = x_1 x_2 \\ \Phi_4(x) = -x_1 \end{bmatrix}$$

X بله - Y خیر (۲)

X خیر - Y بله (۴)

X بله - Y بله (۱)

X خیر - Y بله (۳)

- ۳۰- تابع هسته $K(x_i, x_j)$ در واقع بیانگر نگاشت ϕ از داده $x \in \mathbb{R}^d$ به فضای مقصد ϕ است، به صورتی که در فضای مقصد داریم $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$. اگر در حل یک مسئله از تابع هسته RBF به صورت

$$K(x_i, x_j) = \exp\left(-\frac{1}{2} \|x_i - x_j\|^2\right)$$

صحیح است؟

(۱) $|\phi(x_i) - \phi(x_j)|^2 < \frac{1}{2}$

(۲) $|\phi(x_i) - \phi(x_j)|^2 < 1$

(۳) $|\phi(x_i) - \phi(x_j)|^2 < 2$

(۴) هیچ رابطه مشخصی بین هر دو داده x_i و x_j در فضای هسته وجود ندارد.

- ۳۱- در یک مسئله یادگیری، مجموعه داده‌ای با n نمونه داریم، که هر نمونه دارای p ویژگی است. می‌خواهیم از این مجموعه داده برای ساخت یک دسته بند با روش Bagging استفاده کنیم. به همین منظور می‌بایست از این مجموعه آموزشی یک مجموعه آموزشی جدید ایجاد کنیم. کدام روش برای ایجاد مجموعه آموزشی جدید درست است؟

(۱) نمونه‌برداری P نمونه با جایگذاری n نمونه برداری

(۲) نمونه‌برداری \sqrt{P} نمونه بدون جایگذاری

- ۳۲- یک دسته‌بند برای دسته بندی دو دسته‌ای داریم، که خروجی آن احتمال تعلق یک نمونه به یک دسته را مشخص می‌کند. کدام گزینه در مورد ROC و AUC درست نیست؟

(۱) منحنی ROC یک محنت اکیداً صعودی است.

(۲) منحنی ROC با تغییر مقدار آستانه تصمیم‌گیری می‌سازیم.

(۳) مقدار $AUC = 0.5$ نشان می‌دهد که این دسته‌بند بدتر از دسته‌بند تصادفی است.

(۴) منحنی ROC اجازه می‌دهد که مصالحه بین True Positive و False Positive را مصورسازی نماییم.

- ۳۳- یک دسته‌بند برای دسته بندی دو دسته‌ای داریم، که خروجی آن احتمال تعلق یک نمونه به دسته با برچسب C_1 را مشخص می‌کند. این احتمال به صورت زیر بیان می‌شود.

$$P(y = C_1) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } 1 \leq x \end{cases}$$

اگر ماتریس هزینه دسته بندی اشتباه به صورت زیر باشد، برای کدام مقدار x ریسک کمینه می‌شود؟

دسته پیش‌بینی شده

| | | C2 | C1 |
|----|----|----|----|
| C2 | C2 | 0 | 10 |
| | C1 | 30 | 0 |

$\frac{3}{4}$ (۴)

$\frac{1}{2}$ (۳)

$\frac{1}{4}$ (۲)

0 (۱)

۳۴ - برای ساخت یک درخت تصمیم، ماتریس داده $X = \begin{bmatrix} 6 & 3 \\ 2 & 7 \\ 9 & 6 \\ 4 & 2 \end{bmatrix}$ و برحسب آن بهصورت $Y = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ داده شده است.

هر نمونه دارای ۲ ویژگی x_1 و x_2 هست که بهترتیب در ستون‌های اول و دوم ماتریس X آورده شده است. کدام گزینه را اگر بهعنوان ریشه درخت تصمیم انتخاب کنیم بیشترین بهره اطلاعاتی را خواهیم داشت؟

۱) $x_1 > 2$

۲) $x_2 > 3$

۳) $x_1 > 4$

۴) $x_2 > 6$

۳۵ - فرض کنید که ماتریس کوواریانس زیر را داریم.

$$\Sigma = \begin{bmatrix} 5 & \alpha \\ \alpha & 4 \end{bmatrix}$$

برای چه مقادیری از α این ماتریس یک ماتریس کوواریانس معتبر است؟

۱) $-\sqrt{20} < \alpha < \sqrt{20}$ ۲) $-\sqrt{20} \leq \alpha \leq \sqrt{20}$

۳) برای همه مقادیر حقیقی α ۴) برای همه مقادیر حقیقی α مثبت

۳۶ - فرض کنید که ماتریس کوواریانس مجموعه داده‌ای بهصورت زیر باشد. می‌خواهیم با کمک الگوریتم PCA این داده‌ها را به یک بعد کاهش بعد دهیم. کدام یک از بردارهای زیر بردار تبدیل PCA است؟

$$\Sigma = \begin{bmatrix} 82 & -80 \\ -80 & 82 \end{bmatrix}$$

$$W = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$W = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

$$W = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$W = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

۳۷ - کدام گزینه بعد VC یک درخت تصمیم با سه گره در فضای R^1 است؟

۱) ۵ ۲) ۳ ۳) ۴ ۴) ۵

۳۸ - برای یک مسئله دسته بندی ۲ دسته‌ای، مجموعه داده‌ای با یک میلیون نمونه داریم، که هر نمونه دارای ۴ ویژگی و هر ویژگی دارای ۳ مقدار است. می‌خواهیم برای دسته بندی این داده از الگوریتم بیز ساده (Naïve Bayes) استفاده کنیم. برای یادگیری این دسته‌بند چند پارامتر باید یادگرفته شود؟

۱) ۹ ۲) ۱۳ ۳) ۱۵ ۴) ۱۷

- ۳۹- اگر K_1 و K_2 دوتابع هسته معتبر، و p یک چندجمله‌ای براساس پارامترش که دارای ضرایب مثبت است، و f یک تابع با مقدار حقیقی و a یک مقدار مثبت باشد، کدام هسته معتبر نیست؟

$$K(x, z) = p(K_1(x, z)) \quad (2)$$

$$K(x, z) = f(x)f(z) \quad (1)$$

$$K(x, z) = K_2(x, z) - K_1(x, z) \quad (4)$$

$$K(x, z) = K_2(x, z)K_1(x, z) \quad (3)$$

- ۴۰- برای یک مسئله دسته بندی در فضای ۲ بعدی، بعد VC دسته‌بند توزیع مختلف و توزیع گاووسی با ماتریس کوواریانس برابر کدام است؟

۵ (۴)

۴ (۳)

۳ (۲)

۲ (۱)

- ۴۱- درخت تصمیم حاصل از روش ID3، بر مجموعه داده زیر دارای چند گره برگ است؟

Label Features

| Y | | | | |
|----|----------------|----------------|----------------|----------------|
| | X _۱ | X _۲ | X _۳ | X _۴ |
| +1 | 0 | 1 | 0 | 1 |
| +1 | 1 | 0 | 1 | 0 |
| +1 | 1 | 1 | 1 | 0 |
| +1 | 0 | 0 | 0 | 1 |
| +1 | 1 | 1 | 1 | 0 |
| -1 | 0 | 0 | 1 | 1 |
| -1 | 0 | 0 | 0 | 0 |
| -1 | 0 | 0 | 1 | 0 |
| -1 | 1 | 0 | 0 | 0 |
| -1 | 0 | 0 | 1 | 1 |

- ۴۲- می‌خواهیم بر روی مجموعه داده زیر رگرسیون خطی به صورت $Y = 2\beta^T X - X$ انجام دهیم. مقدار به دست آمده برای β کدام است؟

$$\begin{array}{|c|ccc|} \hline x & -1 & 0 & 2 \\ \hline y & 1 & -1 & 1 \\ \hline \end{array}$$

$$\sqrt{\frac{3}{8}} \quad (2)$$

$$\sqrt{\frac{3}{10}} \quad (1)$$

$$\sqrt{\frac{1}{4}} \quad (4)$$

$$\sqrt{\frac{3}{5}} \quad (3)$$

- ۴۳- کدام مجموعه گیت‌های منطقی است که نمی‌توان تک‌تک اعضایش را با یک پرسپترون پیاده‌سازی کرد؟

AND , XOR , OR (۱)

AND , OR , NOR (۲)

AND , OR , NAND (۳)

NAND , NOR , AND (۴)

- ۴۴- مثال‌های آموزشی زیر با سه ویژگی x , y و z و خروجی U را در نظر بگیرید. اگر برای یادگیری U از دسته‌بندی کننده ساده بیزی (Naive Bayes Classifier) استفاده شود، پس از خاتمه یادگیری، کدام مورد مقادیر احتمالاتی زیر را به درستی نشان می‌دهد؟ (برای سادگی احتمال $P(D)$ را ۱ در نظر بگیرید).

$$T_1 = P(U = 0 | x = 0, y = 1, z = 0) = ?$$

$$T_2 = P(U = 1 | x = 0) = ?$$

| x | y | z | U |
|-----|-----|-----|-----|
| ۱ | ۰ | ۰ | ۰ |
| ۰ | ۱ | ۱ | ۰ |
| ۰ | ۰ | ۱ | ۰ |
| ۱ | ۰ | ۰ | ۱ |
| ۰ | ۰ | ۱ | ۱ |
| ۰ | ۱ | ۰ | ۱ |
| ۱ | ۱ | ۰ | ۱ |

$$T_2 = \frac{۱۶}{۴۹}, T_1 = \frac{۱۴۴}{۲۴۰۱} \quad (۱)$$

$$T_2 = \frac{۱۶}{۴۹}, T_1 = \frac{۱}{۲۱} \quad (۲)$$

$$T_2 = \frac{۳}{۲۱}, T_1 = \frac{۱}{۲۱} \quad (۳)$$

$$T_2 = \frac{۲}{۷}, T_1 = \frac{۱}{۲۱} \quad (۴)$$

- ۴۵- در یک مسئله دسته‌بندی، اگر مجموعه فرضیه‌ها را به صورت $H = \text{sgn}(ax^2 + bx + c) = 1$ در نظر بگیریم، که پارامترهای a , b و c اعداد حقیقی هستند و $\text{sgn}(z) = 1$ اگر $z > 0$ و $\text{sgn}(z) = -1$ اگر $z \leq 0$. مقدار بعد VC فضای H کدام است؟

۱) ۱

۲) ۲

۳) ۳

۴) ۴

tahsilatetakmili.com