سؤالات و پاسخنامه آزمون گروه فنی و مهندسی دکتری ۱۴۰۰

زبان عمومی

PART A: Grammar

Directions: Choose the word or phrase (1), (2), (3), or (4) that best completes the blank. Then, mark the correct choice on your answer sheet.

1- The rate that bright comets enter the solar system implies there should be around 3000 of them buzzing around, ... only 25 are known.

1) nonetheless 2) regardless of the fact 3) and yet 4) as there are

2- Contemporary theories of interpretation require that, in our analyses of texts, we consider not only what the text says .....“made.”

1) also its meaning gets and 2) but also gets the meaning of it
3) but its meaning also gets 4) but how its meaning gets

3- ..... individual behavior is influenced by social networks is beyond dispute.

1) That 2) An 3) The 4) It is that

4- Plant scientists have been trying for years to genetically modify flowers for aesthetic purposes. The first to go on sale were blue carnations ..... in Australia, in 1996.

1) were produced 2) produced 3) had been produced 4) to produce

5- Weapons have been carried and delivered by a wide variety of vehicles, ... weapon platforms.

1) they are often called 2) often called 3) called they are often 4) that are called often

6- Articulating what the difference between humans and other creatures consists of ..... behind it have formed a large and difficult project tackled by biologists, anthropologists, psychologists, and philosophers.

1) uncovering the biology 2) the biology of uncovering
3) the biology uncovering 4) and uncovering the biology

7- Most healthcare professionals view depression as “just part of getting old and argue that this illness, ..., can have serious, even fatal consequences.

1) untreated then 2) untreated whether it is 3) if untreated 4) that is untreated

8- Ted had a terrible habit of boasting so much about his smallest accomplishments ..... his vainglory became renowned throughout the small college campus.

1) that 2) as 3) in that 4) as though

PART B: Vocabulary

Directions: Choose the word or phrase (1), (2), (3), or (4) that best completes the blank. Then, mark the correct choice on your answer sheet.

9- Dogs growl and show their teeth in an attempt to frighten the animal or person they perceive as a .....

1) habitat 2) prey 3) suspicion 4) threat

10- Based on his recent poor decisions, it was obvious that Seth lacked even a modicum of good .....

1) sentiment 2) sense 3) sensation 4) sensitivity

11- The judge ..... the extraneous evidence because it was not pertinent to the trial.

1) disclosed 2) distended 3) dismissed 4) distorted

12- The more frequently employees take time to exercise during working hours each week, the fewer sick days they .....

1) expend 2) save 3) take 4) recall

13- Classic psychology experiments have shown that when rats are first ..... with an electrical shock to fear a tone when it sounds, they later fear the tone even without the associated shock.

1) conditioned 2) sparkled 3) displayed 4) intended

14- In 1998 Gordon Sinclair, the owner of a well-known restaurant, was struggling with a problem that ..... all restaurateurs. Patrons frequently reserve a table but, without notice, fail to appear.

1) delegates 2) afflicts 3) intensifies 4) evades

15- Despite what the scientist said, the volcano eruption is not ..... , so do not be concerned!

1) impassive 2) negotiable 3) vulnerable 4) imminent

16- At the landfill, the ..... process is in full swing, turning much of the garbage into gasses.

1) conversion 2) restoration 3) decomposition 4) pressurization

17- Because I am an extreme planner who needs to control everything, I never engage in ..... .

1) justification 2) pretention 3) coincidence 4) spontaneity

18- The roads in our town already have too much traffic; building a new shopping mall will ..... the problem.

1) frustrate 2) exacerbate 3) preserve 4) exploit

19- The movie *Close Encounters of the Third Kind* tells the story of the first contact between beings from outer space and ..... creatures, that is, those living on earth.

1) terrestrial 2) dominant 3) ingenious 4) affable

20- There is ..... agreement that an airport is needed; no one disputes that, but there is fundamental disagreement about where to build it.

1) uniform 2) utilitarian 3) unique 4) unanimous

پاسخنامه زبان عمومی

بخش اول: دستور زبان

در سؤالات زیر، از بین گزینه‌های (۱)، (۲)، (۳) و (۴) پاسخی را انتخاب کنید که به بهترین نحو جای خالی را پر کند. آنگاه پاسخ‌تـان را روی پاسخنامه علامت بزنید.

۱ـ گزینه «۳» با توجه به سرعت و تعداد ورود ستاره‌های دنباله‌دار به منظومه شمسی می‌توان حدس زد که باید تقریباً ۳۰۰۰ مورد از آن‌ها وجود داشته؛ با این حال تنها ۲۵ عدد از آنها شناسایی شده‌اند.

توضیح: همان‌طور که می‌دانید nonetheless قید ربط است؛ یعنی قبل از آن باید نقطه یا نقطه‌ویرگول و بعد از آن باید حتماً کاما بیاید. با این حساب گزینه (۱) نادرست است. گزینه (۲) در صورتی ارزش بررسی کردن دارد که طراح بعد از fact از حرف ربط that استفاده می‌کرد. گزینه (۳) صحیح است، هم با توجه به مفهوم جمله و هم با توجه به اینکه قبل از and کاما می‌آید. و گزینه (۴) نادرست است چون بعد از as دو تا فعل داریم؛ یکی are و یکی know.

۲ـ گزینه «۴» نظریه‌پردازان معاصر در زمینه ترجمه شفاهی باور دارند ما در آنالیز متن، علاوه بر چیزی که متن می‌گوید، باید به نحوه شکل‌گیری معنی آن نیز توجه داشته باشیم.

توضیح: همان‌طور که می‌بینیم این تست با مبحث not only … but also سروکار دارد. اول از همه اینکه در این ساختار but also می‌تواند به صورت but یا also هم به کار برود. پس امیدوارم فوری گزینه (۲) را نزده باشید.

ضمناً می‌دانیم ساختار not only …. but (also) مستلزم رعایت ساختار موازی است؛ بنابراین چون بعد از not only کلمه پرسشی what را داریم باید بعد از but(also) هم از کلمه‌ی پرسشی how استفاده کنیم:

…. **not only what** the text says **but how** its meaning gets made.

۳ـ گزینه «۱» اینکه شبکه‌های اجتماعی بر روی رفتار افراد تاثیرگذار هستند، قابل تردید نمی‌باشد.

توضیح: تست بسیار ساده‌ای است. توی مبحث جمله‌واره‌ی اسمی گفتیم یکی از کاربردهای that clause این است که قبل از فعل be به عنوان فاعل استفاده شوند. گفتیم در این موارد that به صورت «اینکه» ترجمه می‌شود:

That individual behavior is influenced by social networks is beyond dispute.

مثال بیشتر:

That coffee grows in Brazil is well known.

۴ـ گزینه «۲» گیاه‌شناسان سال‌هاست که با استفاده از اصلاح ژنتیک به دنبال زیباتر ساختن گل‌ها هستند. گل میخک آبی اولین موردی بود که برای فروش عرضه شد. این گل در سال ۱۹۹۶ در استرالیا تولید شد.

توضیح: این تست از دو جمله تشکیل شده که برای پاسخگویی به آن فقط به جمله دوم نیاز داریم. جمله دوم دارای فعل اصلی were می‌باشد، با این حساب به هیچ فعل اصلی دوم دیگری نیاز نداریم چون هر جمله فقط‌وفقط باید یک فعل اصلی داشته باشد. این یعنی حذف همزمان گزینه‌های (۱) و (۳). گزینه (۴) نادرست است چون قصد بیان هدف نداریم.

ضمناً شکل اولیه گزینه (۲) این‌طوری بوده:

The first to go on sale were blue carnations **that were produced** in Australia, in 1996.

اگر that were را حذف کنیم، به گزینه (۲) می‌رسیم.

۵ـ گزینه «۲» سلاح‌ها از طریق وسایل نقلیه مختلفی حمل و تحویل داده می‌شوند. این وسایل نقلیه اغلب با نام پلتفرم سلاح شناخته می‌شوند.

توضیح: تقریباً هر سال از این مبحث سؤال می‌آید و ما هم هر سال می‌گوییم بعد از کاما کاربرد that ممنوع است. (این یعنی حذف گزینه (۴)). گزینه (۱) در صورتی صحیح است که کاما به نقطه تبدیل بشود و they هم به They. مهم‌ترین دلیل رد گزینه (۳) کاربرد they بعد از called است.

ضمناً شکل اولیه‌ی گزینه‌ی ۲ این طوری بوده:

Weapons have been carried and delivered by a wide variety of vehicles, **which are often called** weapon platforms.

اگر which are را حذف کنیم، به گزینه (۲) می‌رسیم.

۶ـ گزینه «۴» درک تفاوت بین انسان و سایر موجودات و مسائل بیولوژیکی نهفته در آن باعث بوجود آمدن مباحث و تحقیقات دشوار و گسترده‌ای شده است که دانشمندانی از رشته‌های مختلف مانند زیست‌شناسی، انسان‌شناسی، روانشناسی و فلسفه به آن می‌پردازند.

توضیح: توی تست‌هایی که این‌قدر طولانی هستند، اولین کار این است که به دنبال فعل اصلی باشیم. فعل اصلی سوال ما have formed است. پس به خاطر حضور have باید فاعلمون جمع باشد. اما articulating به تنهایی به فعل مفرد نیاز دارد، این یعنی باید articulating را با and به یک ساختار ing دار موازی دیگر متصل کنیم تا آن موقع کاربرد فعل have هم معنی پیدا کند. و چون فقط گزینه (۴) است که دارای and می‌باشد، می‌توانیم باقی گزینه‌ها را رد کنیم.

۷ـ گزینه «۳» اکثر متخصصین حوزه بهداشت و درمان، افسردگی را بخشی از پروسه افزایش سن می‌دادند و اعتقاد دارند که در صورت عدم درمان می‌تواند عواقب بسیار وخیمی داشته و یا حتی باعث مرگ بیمار شود.

توضیح: اول از همه اینکه طراح سؤال ظاهراً یادش رفته آن (") را که باز کرده ببندد. باید این علامت را قبل از and بیاورد. حالا می‌رسیم به رد گزینه‌ها. کاربرد that بعد از کاما ممنوع است (یعنی رد گزینه (۴)). گزینه ۱ نادرست است چون معلوم نیست طراح سوال آن then را بابت چی استفاده کرده. گزینه (۲) هم کنار می‌رود چون بعد از is هیچ عبارت کامل‌کننده‌ای نداریم.

اما برای اینکه ببینیم چرا گزینه (۳) صحیح است باید اصل جمله را پیدا کنیم.

Most healthcare professionals argue that this illness, **if it is untreated**, can have serious, even fatal consequences.

چون it به this illness برمی‌گردد، می‌توانیم با فرض اینکه فاعل‌ها یکسان هستند، فاعل جمله‌واره‌ی وابسته یعنی it و فعل is را حذف کنیم و یک وجه وصفی بسازیم:

Most healthcare professionals argue that this illness, **if untreated**, can have serious, even fatal consequences.

۸ـ گزینه «۱» تد اخلاق بسیار زشتی داشت و به خاطر کوچک‌ترین موفقیت‌هایش به‌قدری فخرفروشی می‌کرد که عادت خودستایی او در سرتاسر محوطه‌ی کوچک دانشگاه زبانزد عام و خاص بود.

توضیح: از ساختار so ….that استفاده شده.

….**so** much about ….. **that ….**

بخش دوم: واژگان

دستورالعمل: در سؤالات زیر، از بین گزینه‌های (۱)، (۲)، (۳) و (۴) پاسخی را انتخاب کنید که به بهترین نحو جای خالی را پر کند. آنگاه پاسخ‌تان را روی پاسخنامه علامت بزنید.

۹ـ گزینه «۴» سگ‌ها در هنگام مواجهه با خطر / تهدید، پارس می‌کنند و دندان‌های خود را نشان می‌دهند تا حیوان یا شخص مورد نظر را بترسانند.

۱) زیستگاه، زیست‌بوم ۲) طعمه ۳) سوءظن، تردید ۴) خطر، تهدید

۱۰ـ گزینه «۲» ضعفِ تصمیمات اخیر **سِت** نشان می‌دهد که کوچکترین درکی نسبت به مسائل مختلف ندارد.

۱) تمایل، گرایش، احساس ۲) شعور، معنی، ادراک ۳) احساس، هیجان ۴) حساسیت

۱۱ـ گزینه «۳» قاضی شواهد غیرضروری را مردود اعلام کرد زیرا ارتباط چندانی با روال دادرسی نداشت.

۱) افشاء کردن، فاش کردن ۲) بزرگ کردن، منبسط کردن ۳) مردود شمردن، رد کردن ۴) کج کردن، تحریف کردن

۱۲ـ گزینه «۳» کارمندان هرچقدر در طول هفته بیشتر ورزش کنند، کمتر به مرخصی استعلاجی نیاز پیدا می‌کنند.

توضیح: جواب این سؤال عیناً توی خود سؤال آمده. یعنی اول بوده take time حالا شده take days.

اصطلاح take sick days یعنی «استعلاجی گرفتن.»

۱۳ـ گزینه «۱» طبق آزمایشات روانشناسی کلاسیک، وقتی موش‌ها برای ترسیدن از یک صدای بخصوص به وسیله شوک الکتریکی شرطی شوند، بعدها بدون وجود شوک الکتریکی هم از آن صدا می‌ترسند.

۱) شرطی کردن ۲) درخشیدن، برق زدن ۳) نمایش دادن ۴) قصد داشتن

۱۴ـ گزینه «۲» گوردن سینکلر، صاحب یکی از رستوران‌های مشهور، در سال ۱۹۹۸ با یکی از مشکلاتی که تمامی صاحب رستوران‌ها را آزار می‌داد، دست و پنجه نرم می‌کرد. مشتریان، میز رزرو می‌کردند اما بدون اطلاع در رستوران حاضر نمی‌شدند.

۱) تفویض کردن، سپردن ۲) رنجاندن، آزردن ۳) تشدید کردن ۴) طفره رفتن، شانه خالی کردن

۱۵ـ گزینه «۴» علیرغم دیدگاه آن دانشمند، فوران آتشفشان در آینده‌ای نزدیک اتفاق نخواهد افتاد. بنابراین جایی برای نگرانی نیست!

۱) خونسرد، بی‌عاطفه ۲) قابل مذاکره ۳) آسیب‌پذیر ۴) قریب‌الوقوع

۱۶ـ گزینه «۳» در مکان دفن زباله، پروسه‌ی تجزیه با قدرت در حال اجرا است. در این پروسه بخش زیادی از زباله به گاز تبدیل می‌شود.

۱) تبدیل، تغییر ۲) ترمیم، بازسازی ۳) تجزیه ۴) تحت فشار قرار دادن

۱۷ـ گزینه «۴» من، به دلیل اینکه برنامه‌ریز بسیار دقیقی هستم، باید همه چیز را تحت نظر داشته باشم. بنابراین در زندگی من هرگز چیزی بدون برنامه‌ریزی (فی‌البداهه) اتفاق نمی‌افتد.

۱) توجیه، دلیل آوردن ۲) تظاهر، وانمود ۳) اتفاق، تصادف، تقارن ۴) بدون برنامه، خودبخودی

سؤالات مهندسی کامپیوتر ــ هوش مصنوعی

مجموعه دروس تخصصی (ساختمان داده‌ها و طراحی الگوریتم‌ها، شناسایی الگو، یادگیری ماشین)

۱ــ الگوریتم فلوید ــ وارشال از یک الگوریتم ..... برای حل مسئله کوتاه‌ترین مسیرهای تمام جفت رئوس در یک گراف جهت‌دار $G=(V,E)$ در زمان ..... استفاده می‌کند.

۱) حریصانه، $\Theta(v^3)$ ۲) حریصانه، $\Theta(V^2\log E)$ ۳) برنامه‌نویسی پویا، $\Theta(v^3)$ ۴) برنامه‌نویسی پویا، $\Theta(V^2\log E)$

۲ــ با فرض اینکه $P\neq NP$ باشد، کدام مورد درست است؟

۱) $NP-hard=NP$ ۲) $NP-complete=P$ ۳) $NP-complete=NP$ ۴) $NP-complete\cap P=\varnothing$

۳ــ کمترین تعداد مقایسه موردنیاز برای تعیین اینکه یک عدد صحیح بیش از $\frac{n}{2}$ مرتبه در یک آرایه مرتب از اعداد صحیح به طول n ظاهر می‌شود، از کدام مرتبه است؟

۱) $\Theta(1)$ ۲) $\Theta(\log n)$ ۳) $\Theta(n)$ ۴) $\Theta(n\log n)$

۴ــ n آرایه نامرتب $A_1,\ldots,A_n$ را در نظر بگیرید (n عددی فرد است.) هرکدام از این آرایه‌ها دارای n عنصر متمایز است. هیچ عنصر مشترکی میان هیچ دو آرایه‌ای وجود ندارد. کمترین پیچیدگی زمانی الگوریتمی برای محاسبه میانه این آرایه‌ها از چه مرتبه‌ای است؟

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

۵ــ فرض کنید $W(n)$ و $A(n)$، به ترتیب، نشان‌دهنده بدترین حالت و میانگین زمان اجرای الگوریتم اجراشده بر روی ورودی با اندازه n باشند. کدام مورد همواره درست است؟

۱) $A(n)=O(W(n))$ ۲) $A(n)=\Theta(W(n))$ ۳) $A(n)=\Omega(W(n))$ ۴) $A(n)=o(W(n))$

۶ــ یک آرایه مرتب‌شده از اعداد داریم. می‌خواهیم دو عدد در این آرایه پیدا کنیم که جمع آن دو عدد مساوی یک عدد داده‌شده x باشد. کمترین پیچیدگی زمانی حل این مسئله کدام است؟

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

۷ــ فرض کنید آرایه‌ای از اعداد صحیح $A=[a_1:a_2;\ldots a_n]$ داده شود. فرض کنید یک اندیس (ناشناخته) k وجود دارد به‌طوری‌که زیر آرایه $A=[a_1:a_2;\ldots a_k]$ به ترتیب اکیداً افزایشی مرتب شده است و زیر آرایه $A=[a_k;a_{k+1};\ldots a_n]$ به ترتیب اکیداً نزولی مرتب شده است (یعنی اگر $1\le i<j\le k$، آنگاه $a_i<a_j$، و اگر $k\le i<j\le n$، آنگاه $a_i>a_j$) هدف شما تعیین k است. یک الگوریتم بهینه برای حل این مسئله چه زمان اجرایی دارد؟

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

۸ــ کدام‌یک از موارد زیر درست است؟

۱) آرایه $A=[10;3;5;1;4;2]$ یک max heap است.

۲) هر مسئله محاسباتی با اندازه ورودی n را می‌توان با یک الگوریتمی با زمان چندجمله‌ای برحسب n حل کرد.

۳) برای تمام توابع مثبت $f(n)$، $g(n)$ و $h(n)$، اگر $f(n)=O(g(n))$ و $f(n)=\Omega(h(n))$ باشد، آنگاه $g(n)+h(n)=\Omega(f(n))$ است.

۴) اگر هر رقم جداگانه در RADIX SORT را با استفاده از INSERTION SORT به‌جای COUNTING SORT مرتب کنیم، آنگاه RADIX SORT به‌درستی کار نمی‌کند (یعنی خروجی صحیح را تولید نمی‌کند).

۹ــ کدام مورد زیر مطمئناً عبارت $f(n)=\Omega(g(n))$ را پشتیبانی می‌کند؟

۱) $f(n)\le 4\times g(n)$ برای تمام $n\ge 1$

۲) $f(n)\ge 4\times g(n)$ برای تمام $n\ge 136$

۳) $f(n)\le 4\times g(n)$ برای تمام $n\ge 100$

۴) $\lim\limits_{n\to\infty}\frac{f(n)}{g(n)}=0$

۱۰ـ کدام‌یک از گزاره‌های زیر درست است؟

الف) اگر مسئله P_1 بتواند به مسئله P_2 در زمان خطی کاهش (reduce) یابد، آنگاه اگر P_2 یک مسئله NP-hard باشد، می‌توان نتیجه گرفت P_1 نیز NP-hard است.

ب) یک Clique در یک گراف بدون جهت لزوماً یک vertex cover در گراف مکمل نیست.

۱) فقط گزاره «الف» درست است.
۲) فقط گزاره «ب» درست است.
۳) هر دو گزاره «الف» و «ب» درست است.
۴) هر دو گزاره «الف» و «ب» نادرست است.

۱۱ـ کدام‌یک از گزاره‌های زیر درست است؟

الف) اینکه زمان حل یک مسئله P حد پایین $\Omega(n^2)$ دارد به این معنی است که برای هر الگوریتم A که P را حل می‌کند فقط برخی از نمونه‌های P وقتی به‌عنوان ورودی به A داده شوند، باعث می‌شود A زمان $\Omega(n^2)$ صرف کند.

ب) اینکه زمان حل یک مسئله P حد پایین $\Omega(n^2)$ دارد به این معنی است که برای هر الگوریتم A که P را حل می‌کند هر نمونه از P که به‌عنوان ورودی به A داده شود، باعث می‌شود A زمان $\Omega(n^2)$ صرف کند.

۱) فقط گزاره «الف» درست است.
۲) فقط گزاره «ب» درست است.
۳) هر دو گزاره «الف» و «ب» درست هستند.
۴) هر دو گزاره «الف» و «ب» نادرست هستند.

۱۲ـ کدام‌یک از گزاره‌های زیر درست است؟

الف) اگر یک الگوریتم زمان چندجمله‌ای برای یک مسئله که NP-hard است ارائه شود، آنگاه می‌توان نتیجه گرفت P = NP است.

ب) اگر یک مسئله NP-complete است، آنگاه می‌توان نتیجه گرفت که آن مسئله هیچ راه‌حلی ندارد.

۱) فقط گزاره «الف» درست است.
۲) فقط گزاره «ب» درست است.
۳) هر دو گزاره «الف» و «ب» درست هستند.
۴) هر دو گزاره «الف» و «ب» نادرست هستند.

۱۳ـ کدام‌یک از گزاره‌های زیر درست است؟

الف) اگر یک مسئله که به کلاس NP تعلق دارد یک راه‌حل زمان چندجمله‌ای داشته باشد، آنگاه P = NP است.

ب) اگر کسی یک حد پایین زمان نمایی برای یک مسئله که NP-complete است بدهد، آنگاه P ≠ NP است.

۱) فقط گزاره «الف» درست است.
۲) فقط گزاره «ب» درست است.
۳) هر دو گزاره «الف» و «ب» درست هستند.
۴) هر دو گزاره «الف» و «ب» نادرست هستند.

۱۴ـ مسئله کوله‌پشتی ۱ـ۰ را در نظر بگیرید که n شیء با وزن صحیح داریم و گنجایش کوله‌پشتی عدد صحیح M است. این مسئله دارای یک الگوریتم مبتنی بر روش برنامه‌ریزی پویا با زمان O(M.n) است. این مرتبه زمانی برحسب اندازه مسئله چگونه است؟

۱) خطی
۲) درجه دو
۳) نمایی
۴) شبه چندجمله‌ای

۱۵ـ برای پیدا کردن k امین عدد در میان n عدد که به‌عنوان کلید در گره‌های یک درخت جستجویی دودویی متوازن ذخیره شده‌اند، کمترین پیچیدگی زمانی ممکن کدام است؟ (هر گره درخت فقط شامل کلید و اشاره‌گر به پدر و فرزند چپ و راست است.)

۱) $\Theta(n)$
۲) $\Theta(\log n)$
۳) $\Theta(n \log n)$
۴) $\Theta(\log \log n)$

۱۶ـ n عدد به‌عنوان کلید در گره‌های یک درخت جستجویی دودویی متوازن ذخیره شده‌اند. هر گره علاوه بر کلید و اشاره‌گر به پدر و فرزند چپ و راست، تعداد گره‌های زیر درخت خود را هم نگهداری می‌کند. برای پیدا کردن rank کلید یک گره (یعنی اینکه کلید گره چندمین عدد در بین n عدد است) کمترین پیچیدگی زمانی ممکن کدام است؟

۱) $\Theta(\log \log n)$
۲) $\Theta(\log n)$
۳) $\Theta(n)$
۴) $\Theta(n \log n)$

۱۷ـ در یک درخت قرمز ـ سیاه، طول طولانی‌ترین مسیر ساده از یک گره x به یک برگ در زیر درخت خودش حداکثر چند برابر طول کوتاه‌ترین مسیر از گره x به یک برگ در زیر درخت خودش است؟

۱) ۴
۲) ۳
۳) ۲
۴) ۱

۱۸ـ کدام‌یک از گزاره‌های زیر درست است؟

الف) هر درخت جستجوی دودویی دلخواه با n گره می‌تواند به یک درخت جستجوی دودویی دلخواه دیگر با n گره با انجام O(n) عمل rotation تبدیل شود.

ب) برای هر دو تابع f(n) و g(n) یکی از سه حالت (۱) $f(n) \in o(g(n))$، (۲) $f(n) \in w(g(n))$ و (۳) $f(n) \in \theta(g(n))$ برقرار است.

۱) فقط گزاره «الف» درست است.
۲) فقط گزاره «ب» درست است.
۳) هر دو گزاره «الف» و «ب» درست هستند.
۴) هر دو گزاره «الف» و «ب» نادرست هستند.

۱۹ـ کمترین پیچیدگی زمانی ممکن برای مرتب‌سازی n عدد طبیعی که مقادیر کمتر از $n^۲$ دارند، کدام است؟

۱) $\Theta(n\log^۲ n)$ ۲) $\Theta(n^۲)$ ۳) $\Theta(n\log n)$ ۴) $\Theta(n)$

۲۰ـ اگر عدد ۳۶۳ را در یک درخت جستجوی دودویی، جستجو کنیم، کدام دنباله زیر نمی‌تواند دنباله‌ای از کلید گره‌هایی باشد که بررسی می‌شوند؟ (ترتیب از راست به چپ است.)

۱) ۳۶۳،۳۹۷،۳۴۴،۳۳۰،۳۹۸،۴۰۱،۲۵۲،۲
۲) ۳۶۳،۳۶۲،۲۵۸،۸۹۸،۲۴۴،۹۱۱،۲۲۰،۹۲۴
۳) ۳۶۳،۲۷۸،۳۸۱،۳۸۲،۲۶۶،۲۱۹،۳۸۷،۳۹۹،۲
۴) ۳۶۳،۲۴۵،۹۱۲،۲۴۰،۹۱۱،۲۰۲،۹۲۵

۲۱ـ نزدیک‌ترین فاصله نقطه z از ابرصفحه $W^T x = ۰$ کدام است؟

۱) $W^T z$ ۲) $\frac{W^T z}{\|W\|}$ ۳) $\|W\|.\|z\|$ ۴) $\frac{W^T z}{\|W\|^۲}$

۲۲ـ در دسته‌بندی یک مسئله M دسته‌ای به کمک روش بیزین، کدام رابطه در مورد احتمال خطای دسته‌بندی دسته‌بند بهینه یا همان P_e صادق است؟

۱) $\frac{۱}{M} \le P_e \le ۱-\frac{۱}{M}$ ۲) $P_e \le ۱-\frac{M}{۱+M}$ ۳) $P_e \le ۱-\frac{۱}{M}$ ۴) $P_e \le \frac{۱}{M}$

۲۳ـ در ساخت یک جنگل تصادفی هر درخت را براساس نمونه‌گیری Bootstrap از داده‌های آموزشی می‌سازیم. اگر n تعداد کل نمونه‌های آموزشی و d تعداد ویژگی باشد، به‌طور متوسط هر درخت براساس چه تعداد نمونه آموزشی ساخته می‌شود؟

۱) $(۱-\frac{۱}{n})^n$ ۲) $(۱-\frac{۱}{n})^d$ ۳) $۱-(۱-\frac{۱}{n})^d$ ۴) $۱-(۱-\frac{۱}{n})^n$

۲۴ـ فرض کنید در یک مسئله دسته‌بندی، داده‌های دارای دو ویژگی $x_۱$ و $x_۲$ هستند که هرکدام مقادیر ۱−، ۰ یا ۱+ را می‌گیرند و برچسب طبقه هم مقدار صفر یا یک را می‌گیرد. سه فرضیه $h_۱, h_۲, h_۳$ به‌صورت زیر تعریف شده‌اند:

$$h_۱(x)=\begin{cases}۱ & \text{if } x_۱=۱\\ ۰ & \text{else}\end{cases} \quad h_۲(x)=\begin{cases}۱ & \text{if } x_۱\neq x_۲\\ ۰ & \text{else}\end{cases} \quad h_۳(x)=\begin{cases}۱ & \text{if } x_۱.x_۲=۰\\ ۰ & \text{else}\end{cases}$$

کدام‌یک از فرضیه‌های فوق را می‌توان از طریق یک پرسپترون (با گرفتن وزن‌های مناسب) پیاده‌سازی کرد؟

۱) $h_۱$ ۲) $h_۲$ ۳) $h_۳$ ۴) $h_۲$ و $h_۳$

۲۵ـ در دسته‌بندی یک مسئله دو دسته‌ای با داده‌های آموزش به شرح جدول زیر، دو دسته‌بند Bayes و Naïve Bayes داده آزمون (Weight = ۸۰, Height = ۱۸۰) را در کدام دسته، دسته‌بندی می‌کنند؟

#	Height	Weight	Class
۱	۸۰	۱۸۰	+
۲	۷۰	۱۸۰	+
۳	۸۰	۱۸۰	−
۴	۸۰	۱۸۰	+
۵	۷۰	۱۷۰	+
۶	۵۰	۱۷۰	−
۷	۴۰	۱۶۰	−
۸	۳۰	۱۴۰	−
۹	۱۰۰	۱۷۰	+
۱۰	۱۱۰	۱۶۰	+
۱۱	۵۰	۱۶۰	−
۱۲	۱۱۰	۱۶۰	+

۱) Naïve Bayes $\rightarrow -$, Bayes $\rightarrow +$
۲) Naïve Bayes $\rightarrow +$, Bayes $\rightarrow -$
۳) هر دو مثبت
۴) هر دو منفی

پاسخنامه مهندسی کامپیوتر ــ هوش مصنوعی

مجموعه دروس تخصصی (ساختمان داده‌ها و طراحی الگوریتم‌ها، شناسایی الگو، یادگیری ماشین)

۱ــ گزینه «۳» الگوریتم فلوید ـ وارشال یک روش حل برنامه‌نویسی پویاست (از حافظه استفاده می‌کند) که هزینه محاسباتی آن $\Theta(v^3)$ است.

۲ــ گزینه «۴» اگر P برابر با NP نباشد، P و NP-Complete اشتراکی ندارند.

۳ــ گزینه «۱» این کار را می‌توان با یک جدول درهم‌سازی انجام داد. توجه داشته باشید که تعداد مقایسه مدنظر است نه هزینه محاسباتی. تعداد مقایسه‌های مورد نیاز $\Theta(1)$ است.

۴ــ گزینه «۳» یافتن میانه میانه‌ها برای n^2 عنصر هزینه $\Theta(n^2)$ دارد.

۵ــ گزینه «۱» اگر بخواهیم بگوییم بهترین حالت زمان اجرای یک الگوریتم، در بدترین حالت، بیشتر از بدترین حالت زمان اجرای آن نخواهد بود می‌توانیم بگوییم:

$$A(n) = O(W(n))$$

۶ــ گزینه «۱» این دو آرایه را A و B درنظر بگیرید که هردو صعودی هستند. اعضای B را از b به x-b تغییر می‌دهیم و آن را C می‌نامیم. برای حفظ صعودی بودنش، از آخر به اول می‌چینیم. حال روالی مانند merge از مرتب‌سازی ادغامی را روی آن می‌یابیم تا یک عنصر تکراری در A و C بیابیم. در صورت یافتن این دو عنصر با فرض اینکه مقدار این عنصر برابر a است، این دو عنصر برابر a از مجموعه A و x-a از مجموعه B هستند.

۷ــ گزینه «۴» به روشی مشابه با جست‌وجوی دودویی می‌توان عنصری که از دو عنصر مجاورش بزرگ‌تر باشد را بیابیم. هزینه این کار $\Theta(\log n)$ است.

۸ــ گزینه «۳» عبارت گزینه (۱) نادرست است، زیرا عنصر ۴ نمی‌تواند فرزند عنصر ۳ در هرم بیشینه باشد. عبارت گزینه (۲) نادرست است و هر مسئله‌ای راه‌حل چندجمله‌ای ندارد. عبارت گزینه (۳) صحیح است، زیرا اگر $f(n) = O(g(n))$، آنگاه $g(n) = \Omega(f(n))$ خواهد بود. یعنی رشد g از f کمتر نیست. اضافه شدن یک مقدار مثبت به g نیز این شرط را رد نمی‌کند. عبارت گزینه (۴) نادرست است. استفاده از مرتب‌سازی درجی منجر به افزایش هزینه مرتب‌سازی می‌شود و خروجی نادرست نمی‌دهد.

۹ــ گزینه «۲» اینکه رشد f ناکمتر از g است تنها در عبارت گزینه (۲) قابل مشاهده است.

۱۰ــ گزینه «۲» عبارت اول نادرست است و مسئله A می‌تواند NP-Hard نباشد. عبارت دوم صحیح است و یک vertex cover الزاماً یک clique نیست. clique یک زیرگراف کامل است ولی vertex cover مجموعه رئوسی است که هر یال از گراف حداقل به یکی از آن‌ها متصل است.

۱۱ــ گزینه «۲» تعریف «ب» تعریف درستی برای نماد Ω است.

۱۲ــ گزینه «۱» عبارت «الف» درست است. عبارت «ب» نادرست است. منظور از NP-Complete غیرقابل حل بودن مسئله نیست.

۱۳ــ گزینه «۲» عبارت «الف» نادرست است. مسئله NP راه حل چندجمله‌ای ندارد.

۱۴ــ گزینه «۴» مرتبه $O(M.n)$ شبه چندجمله‌ای است.

۱۵ــ گزینه «۱» نیاز به پیمایش درخت داریم که هزینه آن $\Theta(n)$ است.

۱۶ـ گزینه «۲» با جست‌وجوی دودویی در درخت متوازن با هزینه $\Theta(\log n)$ به عنصر مطلوب می‌رسیم. rank این گره با توجه به تعداد گره‌های زیردرخت اجداد و فرزند چپ گره موردنظر تعیین می‌شود. اجداد گره موردنظر وگره‌های فرزند چپ آن که گره موردنظر در زیردرخت راست آن قرار دارد، قبل از گره موردنظر هستند. برای مثال اگر گره هدف عدد ۵۰ باشد و یکی از اجداد آن که گره ۵۰ در زیردرخت راست آن قرار دارد ۳۰ باشد، تمام فرزندان زیردرخت چپ ۳۰ و خود گره ۳۰ قبل از ۵۰ خواهند بود. همچنین فرزندان زیردرخت چپ ۵۰ نیز قبل از ۵۰ خواهند بود. تعداد گره‌های این مسیر $\Theta(\log n)$ است.

۱۷ـ گزینه «۳» درخت قرمز ـ سیاه یک درخت متوازن است. نسبت بزرگ‌ترین و کوچک‌ترین مسیر از یک گره تا برگ‌ها حداکثر برابر ۲ است.

۱۸ـ گزینه «۱» عبارت «الف» درست است. ولی عبارت «ب» با توجه به اینکه توابع می‌توانند نوسانی باشند نادرست است.

۱۹ـ گزینه «۴» می‌توان از مرتب‌سازی radix sort استفاده کرد. هزینه این کار معادل با دو مرتبه مرتب‌سازی شمارشی برای اعداد ۰ تا n است که یکبار باقی‌مانده و یکبار خارج‌قسمت اعداد به n را برای مرتب‌سازی در نظر می‌گیریم. هزینه محاسبات $\Theta(n)$ خواهد بود.

۲۰ـ گزینه «۴» عدد ۹۱۲ نمی‌تواند بعد از ۹۱۱ و ۲۴۰ بیاید. به همین دلیل دنباله عبارت گزینه (۴) مورد قبول نیست.

۲۱ـ گزینه «۲» در معادله ابر صفحه جداکننده، W برابر بردار وزن و X نیز برابر بردار ویژگی می‌باشد. ضرب نقطه‌ای $W^t.z$ بیانگر میزان شباهت نقطه z با ابر صفحه می‌باشد. نتیجه حاصل از این ضرب اگر مقداری مثبت باشد، این نقطه در یک طرف ابرصفحه بوده و اگر مقدار منفی، نقطه در طرف دیگر ابرصفحه می‌باشد. حال اگر مقدار به‌دست آمده تقسیم بر بزرگی بردار وزن $(\| W \|)$ گردد، نتیجه به‌دست آمده بیانگر نزدیک‌ترین فاصله نقطه z نسبت به ابرصفحه جداکننده می‌باشد.

در گزینه (۱) فاصله به‌دست آمده بدون در نظر گرفتن بزرگی بردار می‌باشد.

در گزینه (۳) تنها بزرگی دو بردار در یکدیگر ضرب شده است که ارتباطی با فاصله دو بردار ندارد.

در گزینه (۴) نیز فرمول به صورت نادرست نوشته شده است.

۲۲ـ گزینه «۳» یک مسئله کلاس‌بندی با دو کلاس را در نظر بگیرید. در صورتی که مدل به صورت شانسی و تصادفی هر داده ورودی را به یک کلاس تعلق دهد بیانگر عدم یادگیری مدل و بالاترین میزان خطا می‌باشد. در این حالت احتمال انتخاب هر کلاس $\frac{1}{2}$ می‌باشد. برعکس این قضیه، در صورتی که مدل کاملاً آموزش دیده باشد و بتواند تمامی نمونه‌های ورودی را به کلاس صحیح اختصاص دهد، احتمال دقت مدل برابر ۱ می‌گردد و احتمال خطا برابر ۰. حال این قضیه قابل تعمیم به مسائلی با تعداد کلاس‌های بیشتر نیز می‌باشد.

با توجه مثال فوق، در بدترین حالت احتمال تشخیص درست مدل برابر $\frac{1}{M}$ می‌باشد و بنابراین در این حالت احتمال خطا برابر $1-\frac{1}{M}$ می‌باشد. در نتیجه گزینه (۳) صحیح است.

قابل ذکر است که گزینه (۱) نیز نادرست می‌باشد. زیرا در صورتی که مدل تمام نمونه‌ها را درست دسته‌بندی نمایید، مقدار خطا برابر صفر می‌گردد اما در این گزینه این اتفاق رخ نمی‌دهد.

۲۳ـ گزینه «۴» مفهوم سؤال کمی مبهم مطرح شده است. زیرا روش Bootstrap روشی با جایگذاری می‌باشد و در این صورت برای ساخت هر درخت به طور میانگین از کل داده‌ها استفاده می‌گردد که برابر است با n. با توجه به گزینه‌های مطرح شده، می‌توان به تعبیری توضیحات زیر را ارائه نمود:

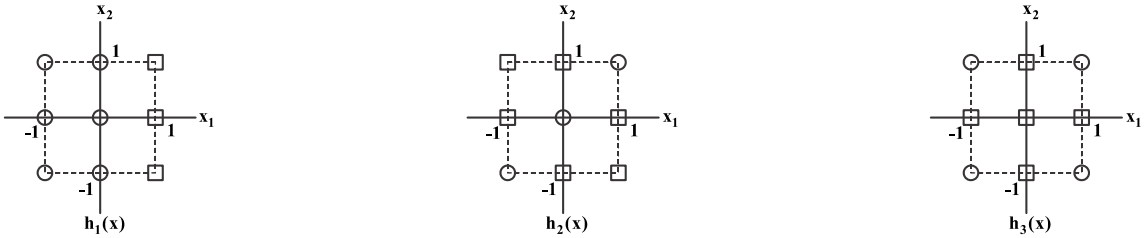
در روش نمونه‌گیری Bootstrap، احتمال انتخاب هر نمونه $\frac{1}{n}$ می‌باشد و در این صورت احتمال این‌که آن نمونه انتخاب نگردد نیز برابر $1-\frac{1}{n}$ است. از آنجایی‌که ما n نمونه را با جایگذاری نمونه‌برداری می‌کنیم، احتمال این‌که یک مثال خاص در هیچ‌یک از n نمونه انتخاب نشده باشد، برابر می‌گردد با $(1-\frac{1}{n})^n$؛ بنابراین در نقطه مقابل، احتمال این‌که یک مثال خاص حداقل یک بار انتخاب گردد برابر می‌گردد با $1-(1-\frac{1}{n})^n$. پس به طور میانگین هر مثال به اندازه‌ی $1-(1-\frac{1}{n})^n$ بار در نمونه‌ی Bootstrap انتخاب می‌گردد.

از آنجاییکه نمونه Bootstrap در کل شامل n مثال است، تعداد مورد انتظار نمونه‌های متمایز آن $n\times[1-(1-\frac{1}{n})^n]$ می‌باشد. زمانی که n مقداری بزرگ است، مقدار این معادله نیز برابر با n می‌گردد.

در نهایت به طور خلاصه، میانگین اندازه‌ی نمونه Bootstrap برای ساخت هر درخت n می‌باشد که به طور تقریبی، مقدار $1-(1-\frac{1}{n})^n$ نیز قابل نمایش است.

نکته: تعداد ویژگی‌ها (d) تأثیری در اندازه‌ی نمونه‌های Bootstrap ندارد.

۲۴ــ گزینه «۱» یک پرسپترون یک دسته‌بند خطی است، یعنی تنها قادر به دسته‌بندی خطی داده‌ها می‌باشد. بنابراین داده‌ها می‌بایست خطی جداپذیر باشند. برای بررسی این موضوع کافی است هر سه فرضیه داده شده را مصورسازی نماییم سپس مشخص می‌گردد کدام فرضیه‌ها خطی جداپذیر نیستند. بر طبق اطلاعات سؤال، تنها فرضیه ۱ خطی جداپذیر بوده و دو فرضیه دیگر به صورت خطی قابل تفکیک نیستند. (در نمودارها یک کلاس با استفاده از دایره و دیگری با استفاده از مربع رسم شده‌اند)



۲۵ــ گزینه «۳» در الگوریتم بیز، ویژگی‌ها به صورت وابسته در نظر گرفته شده و احتمال توأمان برای آن‌ها محاسبه می‌گردد، در صورتی که در روش بیز ساده، ویژگی‌ها مستقل در نظر گرفته شود، پس از محاسبه احتمال هر ویژگی به صورت جداگانه، آن‌ها را در یکدیگر ضرب می‌نماییم. در نهایت در هر دو روش، داده جدید به کلاسی متعلق می‌گردد که احتمال Posterior بیشتری داشته باشد.

ابتدا احتمال Prior هر کلاس را محاسبه می‌نماییم: $p(+)=\frac{۷}{۱۲}$ $p(-)=\frac{۵}{۱۲}$

نکته: در جدول داده شده برای این سؤال، به اشتباه عنوان ویژگی Weight و Height جابه‌جا نوشته شده است.

روش بیز: $$P(+\mid \text{Weight}=۸۰,\ \text{Height}=۱۸۰)=\frac{P(\text{Weight}=۸۰,\ \text{Height}=۱۸۰\mid +)P(+)}{P(\text{Weight}=۸۰,\ \text{Height}=۱۸۰)} \quad (۱)$$

چون مخرج معادله در پاسخ تأثیری ندارد آن را نادیده می‌گیریم. حال به صورت جداگانه مقدار بیشینه‌ی درستنمایی (ML) توأمان را برای هر دو کلاس محاسبه می‌نماییم:

$$P(\text{Weight}=۸۰,\ \text{Height}=۱۸۰\mid +)=\frac{۲}{۷}$$

$$P(\text{Weight}=۸۰,\ \text{Height}=۱۸۰\mid -)=\frac{۱}{۵}$$

حال با جایگذاری مقادیر فوق در رابطه‌ی (۱) مقدار احتمال Posterior را برای نقطه داده شده محاسبه می‌نماییم (در صورت تبدیل + در رابطه (۱) به -، احتمال کلاس منفی محاسبه می‌گردد):

$$P(+\mid \text{Weight}=۸۰,\ \text{Height}=۱۸۰)=\frac{۲}{۷}\times\frac{۷}{۱۲}\simeq ۰/۱۶$$

$$P(-\mid \text{Weight}=۸۰,\ \text{Height}=۱۸۰)=\frac{۱}{۵}\times\frac{۵}{۱۲}\simeq ۰/۰۸$$

با توجه به مقدار احتمال Posterior بیشتر برای کلاس +، نتیجه می‌گیریم با روش بیز، داده به کلاس مثبت تعلق دارد.

روش بیز ساده: $$P(+\mid \text{Weight}=۸۰,\ \text{Height}=۱۸۰)=\frac{P(\text{Weight}=۸۰\mid +)\times P(\text{Height}=۱۸۰\mid +)\times P(+)}{P(\text{Weight}=۸۰,\ \text{Height}=۱۸۰)} \quad (۲)$$

شناسایی الگو ــ یادگیری ماشین

۲۱ــ برای یک طبقه‌بند Hard-SVM دو کلاسه، ۴ عدد داده آموزشی متعادل (balanced) از هر دو کلاس، که به‌طور خطی جدایی‌پذیر هستند موجود است. کدام‌یک از مقادیر زیر نمی‌تواند تعداد بردارهای پشتیبان آن پس از اتمام فاز آموزش باشد؟

۱) ۲ ۲) ۱ ۳) ۴ ۴) ۳

۲۲ــ در آموزش پرسپترون که داده‌ها یکی‌یکی در هر مرحله به آن داده می‌شوند و ضریب یادگیری η با افزایش مراحل i با فرمول $\eta_i = \frac{1}{1+i}$ کاهش می‌یابد، کدام مورد درخصوص تأثیر تغییرات ضریب یادگیری η با i درست است؟

۱) اصلاً ارتباطی بر تأثیر داده‌ها ندارد.
۲) تأثیر داده‌های ابتدایی، کمتر از داده‌های انتهایی می‌شود.
۳) تأثیر داده‌های انتهایی، کمتر از داده‌های ابتدایی می‌شود.
۴) تأثیر داده‌های انتهایی و داده‌های ابتدایی، تقریباً برابر می‌شود.

۲۳ــ کدام مورد زیر، می‌تواند ماتریس کواریانس یک مجموعه داده باشد؟

۱) $\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}$ ۲) $\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}$ ۳) $\begin{bmatrix} 1 & 4 \\ 1 & 2 \end{bmatrix}$ ۴) $\begin{bmatrix} -3 & 3 \\ 3 & 3 \end{bmatrix}$

۲۴ــ فرض کنید می‌خواهیم سیستمی مبتنی بر یادگیری برای تبدیل صوت به متن ایجاد کنیم که برای هر صوت، متن مرتبط با آن را نیز داریم. داده‌های ما صوت‌هایی به زبان فارسی هستند، ولی لهجه‌های مختلفی در آنها وجود دارد که در مجموعه آموزشی، برچسبی برای لهجه‌ها وجود ندارد. اگر بخواهیم از تأثیر لهجه‌ها در یادگیری استفاده شود، فرایند یادگیری در این مسئله باید چگونه باشد؟

۱) ابتدا مبتنی بر دسته‌بندی و سپس مبتنی بر خوشه‌بندی
۲) ابتدا مبتنی بر خوشه‌بندی و سپس مبتنی بر دسته‌بندی
۳) تنها مبتنی بر دسته‌بندی
۴) تنها مبتنی بر خوشه‌بندی

۲۵ــ در کدام‌یک از طبقه‌بندهای زیر، برای فاز تست لازم است همه داده‌های آموزشی یا بخشی از آنها استفاده شود؟

۱) درخت تصمیم ۲) شبکه عصبی MLP ۳) Hard-SVM ۴) K-nn

۲۶ــ در الگوریتم C-SVM، مسئله بهینه‌سازی مقید زیر را داریم:

$$J(\underline{w}, w_0, \underline{\alpha}) = \frac{1}{2}\|\underline{w}\|^2 + c\sum_{i=1}^{N}\alpha_i$$

$$\text{subject to}: \begin{cases} y_i(\underline{w}^T\underline{x}_i + w_0) - 1 + \alpha_i \geq 0 & i = 1,2,...,N \\ \alpha_i \geq 0 & i = 1,2,...,N \end{cases}$$

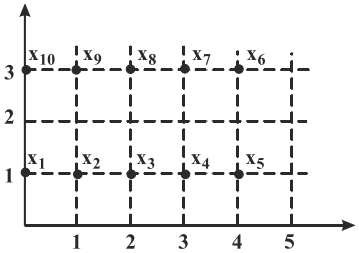
تابع لاگرانژ هم به‌صورت زیر نوشته شده است:

$$L(\underline{w}, w_0, \underline{\alpha}, \underline{\lambda}, \underline{\mu}) = \frac{1}{2}\underline{w}^T\underline{w} + c\sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\lambda_i(y_i(\underline{w}^T x_i + w_0) - 1 + \alpha_i) - \sum_{i=1}^{N}\mu_i\alpha_i$$

اگر برای $C = 5$ مسئله حل شده باشد و برای داده $\underline{x}_5$ داشته باشیم $\lambda_5 = 2$، آنگاه کدام گزاره نادرست است؟

۱) $\alpha_5 = 0$ و این داده درست طبقه‌بندی شده است.
۲) $\mu_5 > 0$ و این داده درست طبقه‌بندی شده است.
۳) $\alpha_5 > 0$ و این داده روی مارجین (حاشیه) قرار دارد.
۴) $\mu_5 = 3$ و این داده روی مارجین (حاشیه) قرار دارد.

۲۷ــ در فضای دوبعدی، ۱۰ داده داریم که می‌خواهیم با روش K-means با معیار فاصله اقلیدسی یک خوشه‌بندی دوکلاسی انجام دهیم. برای شروع کار، دو نقطه دلخواه را به‌عنوان مرکز دو خوشه قرار می‌دهیم و الگوریتم را اجرا می‌کنیم و داده‌ها به‌جز دو نقطه اولیه، به‌ترتیب شماره اندیس به الگوریتم وارد می‌شود. نتیجه نهایی خوشه‌بندی با دو نقطه اولیه x_1 و x_{10} با نتیجه خوشه‌بندی با کدام دو نقطه اولیه در گزینه‌های زیر، یکسان نیست؟

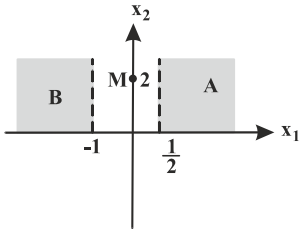


۱) x_1 و x_6
۲) x_5 و x_6
۳) x_4 و x_7
۴) x_2 و x_9

۲۸ــ برای ارزیابی یک تخمین‌زننده، بایاس و واریانس آن را محاسبه می‌کنند. کدام تخمین‌زننده بهتر است؟

۱) بایاس آن کم و واریانس آن زیاد باشد.
۲) بایاس و واریانس آن کم باشد.
۳) بایاس آن زیاد و واریانس آن کم باشد.
۴) بایاس و واریانس آن زیاد باشد.

۲۹ـ در یک طبقه‌بندی دوکلاسی با دو ویژگی، از کلاس ۱ به تعداد ۱۰۰ داده آموزشی داریم که در ناحیه A در شکل زیر و از کلاس ۲ به تعداد ۱۰۰ داده آموزشی داریم که در ناحیه B در شکل زیر هر دو به‌صورت تصادفی قرار دارند. می‌خواهیم به داده‌ای که در شکل با نقطه M مشخص شده است، با طبقه KNN با K = ۱ و فاصله اقلیدسی یک برچسب بدهیم. کدام مورد درست است؟



۱) بدون دیدن داده‌های آموزش، نمی‌توان برچسب را تعیین کرد.
۲) همواره احتمال تعلق داده به هر دو کلاس، برابر است.
۳) برچسب این داده، کلاس ۱ است.
۴) برچسب این داده، کلاس ۲ است.

۳۰ـ اگر در یک مسئله طبقه‌بندی دوکلاسی در فضای ویژگی دوبعدی، ماتریس کوواریانس و میانگین کلاس‌ها مقادیر زیر باشند:

$$\sum=\begin{bmatrix}۴ & ۰\\ ۰ & ۹\end{bmatrix} \quad \mu_۱=\begin{bmatrix}-۱۰\\ -۱۰\end{bmatrix} \quad \mu_۲=\begin{bmatrix}۱۰\\ ۱۰\end{bmatrix}$$

آنگاه نقاطی که فاصله ماهالانوبیس ثابت ۱۰ را از میانگین کلاس خود دارند، حداکثر مقدار فاصله اقلیدسی آنها از همان میانگین، چقدر خواهد بود؟

۱) ۳۰ ۲) ۴۰ ۳) ۵۰ ۴) ۶۰

۳۱ـ برای طبقه‌بندی داده‌های XOR در فضای دوبعدی، کدام‌یک از طبقه‌بندهای زیر مناسب هستند؟

a) طبقه‌بند درخت تصمیم
b) طبقه‌بند Hard SVM
c) یک نرون عصبی مصنوعی پرسپترون
d) طبقه‌بند شبکه عصبی MLP

۱) a و b ۲) b و c ۳) a و d ۴) c و d

۳۲ـ اگر برای متغیر تصادفی x تابع چگالی احتمال داده نشده باشد، ولی میانگین μ و واریانس $\sigma^۲$ آن معلوم باشد، با روش حداکثر آنتروپی، کدام‌یک از توابع چگالی احتمال زیر به‌دست می‌آید؟

۱) تابع $f(x)=\begin{cases}\frac{۱}{x_۲-x_۱} & x_۱\le x\le x_۲\\ ۰ & \text{در غیر این صورت}\end{cases}$ $\quad x_۲=\mu+۳\sigma \quad x_۱=\mu-۳\sigma$

۲) تابع اگر $f(x)=\begin{cases}\frac{x-\mu+\sigma}{۲\sigma^۲} & \mu-\sigma\le x\le\mu+\sigma\\ ۰ & \text{در غیر این صورت}\end{cases}$

۳) تابع $f(x)=\frac{\mu}{۲\sigma}e^{x}$

۴) $f(x)=\frac{۱}{\sqrt{۲\pi\sigma^۲}}e^{-\frac{(x-\mu)^۲}{۲\sigma^۲}}$

۳۳ـ کدام مورد، درخصوص ماشین‌های بردار پشتیبان SVM درست است؟

۱) در یک مسئله جدایی‌ناپذیر خطی با افزایش C در روش C-SVM حاشیه (margin) بین دو کلاس و تعداد نقاط خطا افزایش می‌یابد.
۲) روش C-SVM و روش v-SVM، با انتخاب مناسب پارامترها ممکن است به جواب یکسانی برسند.
۳) ضرایب لاگرانژ پاسخ سیستم برای یک پایگاه داده خاص D، همواره منحصر به فرد است.
۴) با افزودن داده آموزشی جدید به سیستم، همواره پاسخ سیستم تغییر می‌کند.

۳۴ـ یک ماتریس کوواریانس کدام ویژگی یا ویژگی‌ها از ۵ ویژگی زیر را همیشه دارا است؟

a) وارون‌پذیری
b) قطری بودن
c) همه درایه‌ها مثبت هستند.
d) متقارن بودن
e) داشتن حداقل یک مقدار ویژه منفی

۱) a و d ۲) d ۳) b ،d و e ۴) a ،b ،c و e

۱۵ـ گزینه «۳» الگوریتم BFS از صف استفاده می‌کند. درخت‌های عمق اول و سطح اول یک گراف الزاماً مشابه نیستند. عبارت‌های دوم و سوم صحیح هستند.

۱۶ـ گزینه «۲» در هرم بیشینه، بزرگ‌ترین مقدار در ریشه قرار دارد و دومین بزرگ‌ترین مقدار، یکی از فرزندان ریشه است. هزینه یافتن این عنصر $O(1)$ خواهد بود.

۱۷ـ گزینه «۱» هر گره در مسئله MAZE یک مسیر از ابتدا تا یک خانه‌ای را نشان می‌دهد که طول آن برابر ارتفاع آن گره است. با توجه به اینکه از هر گره حداکثر ۴ خروجی وجود دارد، $O(۴^n)$ حد بالای پیچیدگی زمانی خواهد بود.

۱۸ـ گزینه «۴» در این مسئله، هزینه مسیر معنی ندارد و تنها رسیدن به اولین جواب مهم است. در نتیجه روش عقب‌گرد انتخاب مناسبی برای این مسئله است و شاخه و حد راه‌حل خوبی نیست. با درنظر گرفتن صفحه $n\times n$ و نیاز به عبور دقیقاً یک مرتبه از تمام خانه‌ها، $(n^۲)!$ تعداد حالت‌های مسئله را نشان می‌دهد که البته تعداد قابل توجهی از آن‌ها هرس می‌شوند.

۱۹ـ گزینه «۴» حالت‌های ممکن تقسیم ماتریس‌ها، شاخه‌های درخت هستند. مجموع ضرب‌ها نیز حدپایین هستند تا بتوان مسیرهایی که به جواب بهینه نخواهند رسید را هرس کرد.

۲۰ـ گزینه «۱» هر سه عبارت صحیح هستند.

شناسایی الگو ـ یادگیری ماشین

۲۱ـ گزینه «۲» در Hard-SVM برای داده‌های خطی جدایی‌پذیر، مرز تصمیم‌گیری به‌گونه‌ای رسم می‌شود که حاشیه بین دو کلاس بیشینه شود. بردارهای پشتیبان نقاطی هستند که دقیقاً روی حاشیه قرار می‌گیرند یا درون آن هستند. بنابراین حداقل تعداد بردارهای پشتیبان ۲ (یک نقطه از هر کلاس) می‌باشد. همچنین حداکثر تعداد در تئوری می‌تواند به اندازه کل داده‌ها باشد، اما در عمل معمولاً کمتر است.

با توجه به داده‌های متعادل (۴ نقطه از هر کلاس) و خطی جدایی‌پذیر بودن، امکان وجود تنها یک بردار پشتیبان وجود ندارد، زیرا مرز تصمیم‌گیری نیاز به نقاطی از هر دو کلاس دارد تا حاشیه را تعریف کند.

۲۲ـ گزینه «۳» در الگوریتم پرسپترون با ضریب یادگیری کاهشی، مقدار ضریب یادگیری با افزایش مراحل (افزایش i در مخرج کسر) کاهش می‌یابد. در مراحل ابتدایی (مقدار i کوچک‌تر) ضریب یادگیری بزرگ‌تر است، بنابراین تغییرات وزن‌ها تحت‌تأثیر قوی‌تر داده‌های اولیه قرار می‌گیرند.

در مراحل انتهایی (مقدار i بزرگ‌تر) ضریب یادگیری کوچک‌تر می‌شود و تغییرات ناشی از داده‌های جدید کاهش می‌یابد. در نتیجه، داده‌های ابتدایی تأثیر بیشتری بر مدل نهایی دارند، درحالی‌که داده‌های انتهایی تأثیر کمتری خواهند داشت. این استدلال بر گزینه (۳) دلالت دارد.

سازمان سنجش گزینه (۴) را به‌عنوان پاسخ صحیح اعلام کرده است، اما با توجه به توضیحات فوق، گزینه (۳) صحیح می‌باشد.

۲۳ـ گزینه «۱» ماتریس کواریانس باید دو شرط اصلی داشته باشد:

۱ـ متقارن بودن: عناصر غیرقطری سطر و ستون یکسان باشند.

۲ـ نیمه معین مثبت (Positive Semi-Definite): همه مقادیر ویژه ماتریس باید غیرمنفی باشند.

حال به بررسی گزینه‌ها می‌پردازیم:

گزینه (۱): ـ متقارن است $(۲ = ۲)$.

ـ مقادیر ویژه: $\lambda = \frac{۷ \pm \sqrt{۱۷}}{۲} \approx ۱/۴۴, ۵/۵۶$ (هردو مثبت)

گزینه (۲): ـ متقارن است $(۳ = ۳)$.

ـ مقادیر ویژه: $\lambda = \frac{۳ \pm \sqrt{۳۷}}{۲} \approx ۵/۵۴, -۲/۵۴$ (یکی منفی)

گزینه (۳): ـ متقارن نیست $(۴ \neq ۱)$.

گزینه (۴): ـ متقارن است $(۳ = ۳)$.

ـ مقادیر ویژه: $\lambda = \pm\sqrt{-۱۸}$ (موهومی)

۲۴ـ گزینه «۲» از آنجا که لهجه‌ها در داده‌های آموزشی برچسب‌گذاری نشده‌اند، می‌توان با خوشه‌بندی، صوت‌ها را بر اساس شباهت‌های آوایی (مانند ویژگی‌های فرکانسی، الگوی تلفظ و غیره) به گروه‌هایی تقسیم کرد. این گروه‌ها احتمالاً نماینده لهجه‌های مختلف هستند. پس از شناسایی خوشه‌ها (لهجه‌ها)، می‌توان از این اطلاعات به‌عنوان ویژگی‌های اضافی در مدل دسته‌بندی (تبدیل صوت به متن) استفاده کرد. برای مثال، مدل می‌آموزد که چگونه هر لهجه بر تلفظ کلمات تأثیر می‌گذارد و دقت تبدیل صوت به متن را بهبود می‌بخشد.

۲۵ـ گزینه «۴» K-nn (K ـ نزدیک‌ترین همسایه): این الگوریتم مبتنی بر نمونه (Instance-based) است و در فاز تست نیاز دارد تمام یا بخشی از داده‌های آموزشی را در حافظه نگه دارد تا نزدیک‌ترین همسایه‌ها را برای نمونه‌های جدید محاسبه کند. بدون دسترسی به داده‌های آموزشی، K-nn نمی‌تواند عمل کند.

در رابطه با سایر گزینه‌ها (درخت تصمیم، Hard-SVM، MLP): این مدل‌ها پس از آموزش، پارامترهای نهایی (مانند ساختار درخت، ابرصفحه، یا وزن‌های شبکه عصبی) را ذخیره می‌کنند. در فاز تست، تنها از این پارامترها استفاده می‌شود و نیازی به داده‌های آموزشی اصلی نیست.

۲۶ـ گزینه «۳» شرایط بهینه‌سازی (KKT) شامل موارد زیر است:

ـ شرط ایستایی نسبت به α_i:

$$\frac{\partial L}{\partial \alpha_i} = c - \lambda_i - \mu_i = 0 \Rightarrow \lambda_i + \mu_i = c$$

ـ شرط تکمیلی:

$$\lambda_i \left[y_i(w^T x_i + w_0) - 1 + \alpha_i \right] = 0 \ , \ \mu_i \alpha_i = 0$$

حال برای داده X_5 با شرایط داده شده تحلیل می‌کنیم. مقدار $C = 5$ است. داده X_5 برای آن $\lambda_5 = 2$ داده‌شده است. طبق شرط ایستایی:

$$\lambda_5 + \mu_5 = 5 \Rightarrow \mu_5 = 5 - 2 = 3$$

همراه با شرط تکمیلی برای α_5:

$$\mu_5 \alpha_5 = 0$$

چون $\mu_5 = 3 > 0$ داریم، بنابراین به ناچار $\alpha_5 = 0$ همچنین، از تکمیلی بودن قید اصلی نتیجه می‌گیریم:

$$y_5(w^T x_5 + w_0) - 1 + \alpha_5 = 0$$

با قرار دادن $\alpha_5 = 0$، خواهیم داشت:

$$y_5(w^T x_5 + w_0) = 1$$

بنابراین، داده X_5 دقیقاً روی مارجین قرار دارد و طبقه‌بندی آن صحیح است چون شرط $y_i(w^T x_i + w_0) \geq 1$ برقرار است.

حال به بررسی گزاره‌ها می‌پردازیم:

گزاره ۱) درست است، چون $\alpha_5 = 0$ و داده طبق شرط روی مارجین (که به‌عنوان طبقه‌بندی صحیح در نظر گرفته می‌شود) قرار دارد.

گزاره ۲) درست است، چون $\mu_5 = 3 > 0$

گزاره ۳) نادرست است، چرا که محاسبات نشان می‌دهد $\alpha_5 = 0$ و نه بزرگ‌تر از صفر.

گزاره ۴) درست است، همانطور که محاسبه کردیم $\mu_5 = 3$ و داده روی مارجین قرار دارد.

۲۷ـ گزینه «۱» با فرض اینکه الگوریتم k-means به روش کلاسیک (ابتدا تخصیص داده به نزدیک‌ترین مرکز و سپس به‌روزرسانی مراکز) اجرا شود و داده‌های باقی‌مانده (x_2 تا x_9) به‌ترتیب شماره وارد شوند، روند کار به شرح زیر خواهد بود:

گام ۱: تعیین مراکز اولیه

- خوشه ۱: مرکز اولیه $= x_1 = (0,1)$
- خوشه ۲: مرکز اولیه $= x_{10} = (0,3)$

گام ۲: تخصیص نقاط به خوشه‌ها (با توجه به فاصله اقلیدسی)

برای هر نقطه باقی‌مانده، فاصله از مراکز فعلی را محاسبه می‌کنیم:

۱ـ $x_2 = (1,1)$:

ـ فاصله از $(0,1)$: ۱

ـ فاصله از $(0,3)$: $\sqrt{(0-1)^2 + (3-1)^2} \approx 2/236$

$\rightarrow$ تخصیص به خوشه ۱

۲ـ $x_3 = (2,1)$:

ـ فاصله از $(0,1)$: ۲

ـ فاصله از $(0,3)$: $\approx 2/828$

$\rightarrow$ خوشه ۱

۳ـ $x_4 = (3,1)$:

ـ فاصله از $(0,1)$: ۳

ـ فاصله از $(0,3)$: $\approx 3/606$

$\rightarrow$ خوشه ۱

۴ـ $x_5 = (4,1)$:

ـ فاصله از $(0,1)$: ۴

ـ فاصله از $(0,3)$: $\approx 4/472$

$\rightarrow$ خوشه ۱

۵ـ $x_۶ = (۴,۳)$:
ـ فاصله از $(۰,۱)$: $۴/۴۷۲ \approx ۲۰$
ـ فاصله از $(۰,۳)$: ۴
← خوشه ۲

۶ـ $x_۷ = (۳,۳)$:
ـ فاصله از $(۰,۱)$: $۳/۶۰۶ \approx$
ـ فاصله از $(۰,۳)$: ۳
← خوشه ۲

۷ـ $x_۸ = (۲,۳)$:
ـ فاصله از $(۰,۱)$: $۲/۸۲۸ \approx$
ـ فاصله از $(۰,۳)$: ۲
← خوشه ۲

۸ـ $x_۹ = (۱,۳)$:
ـ فاصله از $(۰,۱)$: $۲/۲۳۶ \approx$
ـ فاصله از $(۰,۳)$: ۱
← خوشه ۲

پس از این تخصیص اولیه، خوشه‌بندی اولیه به صورت زیر است:
ـ خوشه ۱: $x_۱, x_۲, x_۳, x_۴, x_۵$
ـ خوشه ۲: $x_{۱۰}$ و $x_۶, x_۷, x_۸, x_۹$

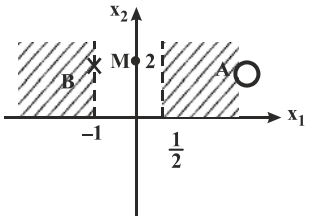
گام ۳: به‌روزرسانی مراکز خوشه‌ها
محاسبه میانگین (centroid) هر خوشه:
ـ برای خوشه ۱ $(x_۱, x_۲, x_۳, x_۴, x_۵)$:
ـ میانگین محور ۲ $= x = \frac{(۰+۱+۲+۳+۴)}{۵} = \frac{۱۰}{۵}$
ـ میانگین محور ۱ $= y = \frac{(۱+۱+۱+۱+۱)}{۵} = \frac{۵}{۵}$
← مرکز جدید خوشه ۱: $(۲,۱)$
ـ برای خوشه ۲ $(x_۶, x_۷, x_۸, x_۹, x_{۱۰})$:
ـ میانگین محور ۲ $= x = \frac{(۰+۴+۳+۲+۱)}{۵} = \frac{۱۰}{۵}$
ـ میانگین محور ۳ $= y = \frac{(۳+۳+۳+۳+۳)}{۵} = \frac{۱۵}{۵}$
← مرکز جدید خوشه ۲: $(۲,۳)$

گام ۴: بررسی تخصیص‌ها با مراکز جدید
با مراکز جدید $(۲,۱)$ و $(۲,۳)$، فاصله هر نقطه به مراکز محاسبه می‌شود. به‌عنوان مثال:
ـ $x_۱ = (۰,۱)$:
ـ فاصله تا $(۲,۱) = ۲$
ـ فاصله تا $(۲,۳) \approx ۲/۸۲۸$
← در خوشه ۱ باقی می‌ماند.

همین منطق برای سایر نقاط هم به همان نتیجه منتهی می‌شود. بنابراین تخصیص نهایی تغییر نمی‌کند و الگوریتم همگرا می‌شود. حال با بررسی هریک از گزینه‌ها با همین روش که توضیح داده شد، می‌توانیم دریابیم که تنها گزینه‌ای که نتیجه خوشه‌بندی با آن، متفاوت از نتایج فعلی می‌شود، گزینه (۱) می‌باشد.

۲۸ـ گزینه «۲» در ارزیابی یک تخمینگر، دو عامل اصلی وجود دارند که باید در نظر گرفته شوند: بایاس و واریانس. بایاس معیاری است که نشان می‌دهد تخمینگر به‌طور سیستماتیک چقدر از مقدار واقعی پارامتر فاصله دارد. به‌عبارتی، اگر میانگین تخمین‌های یک تخمینگر در نمونه‌های متعدد به مقدار واقعی نزدیک باشد، آن تخمینگر دارای بایاس کم است. در نتیجه، تخمینگری با بایاس کمتر، به‌طور میانگین، نزدیک به مقدار واقعی قرار می‌گیرد.
واریانس نوسان تخمین‌ها را نسبت به هم در نمونه‌های مختلف بیان می‌کند. تخمینگری با واریانس پایین نسبت به تغییرات نمونه‌های انتخابی پایدارتر است؛ یعنی نتایج آن در نمونه‌های مختلف تغییر چندانی نمی‌کند. یک تخمینگر با واریانس بالا ممکن است در نمونه‌های مختلف بسیار متفاوت عمل کند، حتی اگر به‌طور میانگین خطای کمتری داشته باشد.
در انتخاب تخمینگر مناسب، هدف این است که هر دو معیار در بهترین حالت باشند؛ به این معنا که تخمینگر هم به‌صورت سیستماتیک نزدیک به مقدار درست باشد (بایاس پایین) و هم از نظر تغییرپذیری در نمونه‌های مختلف پایداری داشته باشد (واریانس پایین). به‌عبارت دیگر، تخمینگری که بتواند هم از خطاهای سیستماتیک جلوگیری کند و هم از نوسانات شدید در نتایج اجتناب کند، از نظر عملکرد آماری بهتر است.

۲۹ــ گزینه «۱» می‌دانیم که در روش 1-NN، با استفاده از نزدیک‌ترین همسایه به داده موردنظر، کلاس داده مورد نظر را تعیین می‌نماییم. با توجه به نمودار، داده M روی محور عمودی واقع شده است $(X_1 = 0)$ و ناحیه A از $X_1 = \frac{1}{2}$ شروع می‌گردد و ناحیه B از $X_1 = -1$. در نگاه اول شاید به نظر برسد که چون مرز شروع ناحیه A به نقطه موردنظر نزدیک‌تر است تا مرز شروع ناحیه B به نقطه موردنظر، پس نقاطی که در ناحیه A وجود دارند به نقطه مورد نظر ما نزدیک‌تر هستند پس این نقاط کلاس نقطه موردنظر را تعیین می‌کنند. ولی با توجه به صورت سؤال «داده‌ها به‌صورت تصادفی در هر ناحیه قرار دارند». یعنی امکان دارد نزدیک‌ترین داده به داده موردنظر ما از کلاس B باشد. برای درک بهتر یک مثال نقض را در شکل زیر نشان می‌دهیم. در شکل زیر فرض شده نزدیک‌ترین داده به نقطه موردنظر، در کلاس A با دایره نشان داده شود و در کلاس B با x:



بنابراین بدون دیدن داده‌های آموزش، نمی‌توان برچسب را تعیین نمود.

۳۰ــ گزینه «۱» برای نوشتن معادله فاصله ماهالانوبیس، فرض می‌کنیم یک نقطه X نسبت به میانگین μ (هر کلاس) دارای فاصله ماهالانوبیس باشد. با توجه به داده‌های مسئله، فاصله ماهالانوبیس ثابت $d_M = 10$ است. بنابراین این معادله به شکل زیر نوشته می‌شود:

$$d_M(x,\mu) = \sqrt{(x-\mu)^T \Sigma^{-1}(x-\mu)}$$
$$(x-\mu)^T \Sigma^{-1}(x-\mu) = 10^2 = 100$$

با توجه به اینکه $\Sigma^{-1} = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{9} \end{bmatrix}$، اگر $y = x - \mu$ قرار دهیم، معادله به صورت مقابل درمی‌آید: $\frac{y_1^2}{4} + \frac{y_2^2}{9} = 100$

برای به‌دست آوردن فرم استاندارد یک بیضی، کل معادله را بر ۱۰۰ تقسیم می‌کنیم: $\frac{y_1^2}{400} + \frac{y_2^2}{900} = 1$

در این فرم، طول‌های هون (نیم‌قطرهای اصلی) بیضی a و b به‌صورت زیر است:

ـ در جهت y_1 (محور افقی): $\sqrt{400} = 20$

ـ در جهت y_2 (محور عمودی): $\sqrt{900} = 30$

یادآوری: این نیم‌قطرها نشان‌دهنده بیشترین فاصله‌های اقلیدسی از میانگین در جهت‌های اصلی بیضی هستند.

پس از میانگین، حداکثر فاصله اقلیدسی (یعنی $\|y\|_2 = \sqrt{y_1^2 + y_2^2}$) برای نقاطی که دارای فاصله ماهالانوبیس ثابت ۱۰ هستند، زمانی به‌دست می‌آید که نقطه روی محور دارای واریانس بزرگ‌تر قرار گیرد (در این مثال، محور y_2 که واریانس آن ۹ است). به‌عبارت دیگر، بعد تبدیل از فاصله ماهالانوبیس به فاصله اقلیدسی در جهت یک واریانس خاص، برابر است با ضرب فاصله ماهالانوبیس در جذر واریانس مربوطه، بنابراین داریم:

$$\max\|y\| = 10 \times \sqrt{\max\{4,9\}} = 10 \times \sqrt{9} = 10 \times 3 = 30$$

سازمان سنجش گزینه (۴) یعنی مقدار ۶۰ را به‌عنوان گزینه صحیح انتخاب نموده است. در رد این گزینه می‌توان بیان نمود که این اشتباه احتمالاً به دلیل تفسیر «حداکثر فاصله» به‌عنوان «قطر» توسط سازمان سنجش می‌باشد. از آنجا که نقطه‌ای در یک جهت از میانگین فاصله ۳۰ دارد، نقطه‌ای در جهت مخالف نیز ۳۰ واحد فاصله دارد؛ بنابراین بیشینه فاصله بین دو نقطه‌ای که هر دو از میانگین به اندازه‌ی ۳۰ فاصله دارند، برابر خواهد بود با $30+30=60$ اما در سؤال خواسته‌شده «فاصله اقلیدسی آن‌ها از همان میانگین» که به معنای شعاع یا فاصله از مرکز (میانگین) است (که برابر با ۳۰ می‌شود)، نه فاصله دو نقطه مقابل.

سازمان سنجش گزینه (۴) را به‌عنوان پاسخ صحیح اعلام کرده است، اما با توجه به توضیحات فوق، گزینه (۱) صحیح می‌باشد.

۳۱ــ گزینه «۳» با توجه به ماهیت مسئله XOR، متوجه می‌شویم که داده‌های XOR به‌طور خطی قابل تفکیک نیستند؛ به‌عبارت دیگر، مسئله XOR غیرخطی است. از این رو، طبقه‌بندهای خطی مانند یک پرسپترون تک‌لایه یا SVM خطی (که معمولاً به‌عنوان hard SVM نیز شناخته می‌شود) قادر به دسته‌بندی صحیح داده‌ها نخواهند بود.

پرسپترون تک‌لایه (مورد c): تنها قادر به پیداکردن مرزهای تصمیم خطی است و برای مسئله غیرخطی XOR ناکافی است.

Hard SVM classifier (مورد b): اگر بدون استفاده از کرنل غیرخطی (مثلاً کرنل گاوسی یا چندجمله‌ای) به‌کار رود، مرز تصمیم خطی ایجاد می‌کند و نمی‌تواند مسئله XOR را تفکیک کند.

MLP (شبکه عصبی چندلایه) (مورد d): به دلیل وجود لایه‌های پنهان و توانایی ایجاد مرزهای تصمیم غیرخطی، می‌تواند مسئله XOR را به‌خوبی دسته‌بندی کند.

درخت تصمیم (مورد a): با تقسیم‌بندی بازگشتی فضای ویژگی‌ها حتی به کمک تقسیم‌بندی‌های محور و ایجاد مرزهای غیرخطی، قادر است مسئله XOR را تفکیک کند.

بنابراین، دو طبقه‌بندی که به‌خوبی می‌توانند داده‌های XOR را دسته‌بندی کنند، طبقه‌بند شبکه عصبی چندلایه (MLP) و طبقه‌بند درخت تصمیم هستند.

۳۲ـ گزینه «۴» برای یک متغیر تصادفی پیوسته که تنها میانگین (μ) و واریانس (σ^2) آن مشخص باشد، اصل حداکثر آنتروپی بیان می‌کند که تابع چگالی احتمالی با بیشترین آنتروپی همان توزیع نرمال (گاوسی) است. از میان چهار تابع ارائه‌شده، تنها تابع گزینه ۴ همان تابع چگالی نرمال است.

۳۳ـ گزینه «۲» بیایید هر گزینه را بررسی کنیم:

گزینه (۱): در یک مسئله جدایی‌ناپذیر خطی با روش C-SVM، پارامتر C کنترل‌کننده مجازات (penalty) نقاطی است که داخل حاشیه یا اشتباه طبقه‌بندی شده‌اند. با افزایش C، مجازات‌ها بیشتر می‌شود؛ این بدان معناست که مدل تمایل دارد اشتباهات را کاهش دهد حتی به هزینه کاهش حاشیه. از این رو، با افزایش C معمولاً حاشیه (margin) کوچک‌تر و تعداد نقاط خطا کمتر می‌شود، نه اینکه افزایش پیدا کند. پس گزینه (۱) نادرست است.

گزینه (۲): روش‌های C-SVM و v-SVM هر دو هدف اصلی یکسانی دارند: یافتن یک hyperplane با حداکثر حاشیه و کنترل خطاهای مجاز. در واقع، این دو مدل با انتخاب مناسب پارامترهای C و v می‌توانند به جواب‌های معادل دست پیدا کنند. بنابراین گزینه (۲) درست است.

گزینه (۳): اگرچه تابع تصمیم نهایی (که شامل ترکیب خطی از داده‌های پشتیبان و آستانه است) یکتا است، اما ضرایب لاگرانژ (Lagrange multipliers) ممکن است برای یک مجموعه داده خاص منحصر به فرد نباشند؛ دلیل آن وجود چندین جواب بهینه در مسئله بهینه‌سازی SVM در شرایط خاص (مانند همبستگی میان داده‌ها) است. پس ادعای «همواره منحصر به فرد» بودن ضرایب لاگرانژ صحیح نیست.

گزینه (۴): افزودن داده‌های آموزش جدید به سیستم همیشه باعث تغییر جواب نمی‌شود. اگر داده‌های افزوده شده در داخل حاشیه قرار نگیرند (یعنی نقاط جدید support vector نشوند)، جواب سیستم تحت‌تأثیر قرار نخواهد گرفت.

۳۴ـ گزینه «۲» برای یک مجموعه داده، ماتریس کوواریانس به‌صورت زیر تعریف می‌شود:

$\Sigma = \text{cov}(X)$ که در آن $\Sigma_{ij} = \text{cov}(X_i, X_j)$

از آنجا که کوواریانس بین X_i و X_j (یعنی $\text{cov}(X_i, X_j)$) برابر با کوواریانس بین X_j و X_i (یعنی $\text{cov}(X_j, X_i)$) است، نتیجه می‌شود که:

$$\Sigma_{ij} = \Sigma_{ji}$$

یعنی هر ماتریس کوواریانس دارای ویژگی متقارن بودن است.

حال به گزینه‌های ارائه‌شده نگاه می‌کنیم:

ـ (a) وارون‌پذیری: یک ماتریس کوواریانس درصورتی وارون‌پذیر است که تمام متغیرها خطی مستقل باشند؛ اما در بسیاری از موارد (مثلاً در حضور همبستگی کامل یا وابستگی خطی بین متغیرها) ممکن است ماتریس کوواریانس تکین (singular) شود. بنابراین وارون‌پذیری ویژگی عمومی و مشترکی نیست.

ـ (b) قطری بودن: یک ماتریس کوواریانس زمانی قطری است که متغیرهای موجود کاملاً بدون همبستگی باشند. اما در حالت کلی، متغیرها همبستگی‌هایی با یکدیگر دارند؛ بنابراین ماتریس کوواریانس لزوماً قطری نیست.

ـ (c) مثبت بودن همه درایه‌ها: این ویژگی مربوط به ماتریس‌های مثلثی (پایین یا بالا) است؛ درحالی‌که در ماتریس کوواریانس به دلیل متقارن بودن، هم درایه‌های بالایی و هم درایه‌های پایینی برابر هستند اما آن را به‌عنوان یک ماتریس مثلثی نمی‌توان در نظر گرفت.

ـ (d) متقارن بودن: همان‌طور که توضیح داده شد، این ویژگی همیشه در ماتریس‌های کوواریانس به قوت عمل می‌کند.

ـ (e) داشتن حداقل یک مقدار ویژه منفی: از آنجایی‌که ماتریس‌های کوواریانس مثبت نیم معین هستند، تمامی مقادیر ویژه آن‌ها صفر یا مثبت است؛ یعنی هیچ مقدار ویژه منفی‌ای وجود ندارد.

۳۵ـ گزینه «۱» با افزایش پیچیدگی مدل به‌طور کلی نتایج زیر به‌دست می‌آید:

ـ بایاس (Bias): با پیچیدگی بالاتر مدل، مدل قادر است الگوهای داده را بهتر به‌دست آورد و تقریب بهتری از تابع واقعی ارائه دهد؛ بنابراین بایاس کاهش می‌یابد.

ـ واریانس (Variance): از سوی دیگر، مدل‌های پیچیده معمولاً حساس‌تر به تغییرات داده‌های آموزشی شده و ممکن است نتایجی ناپایدار و متغیر داشته باشند، یعنی واریانس افزایش می‌یابد.
