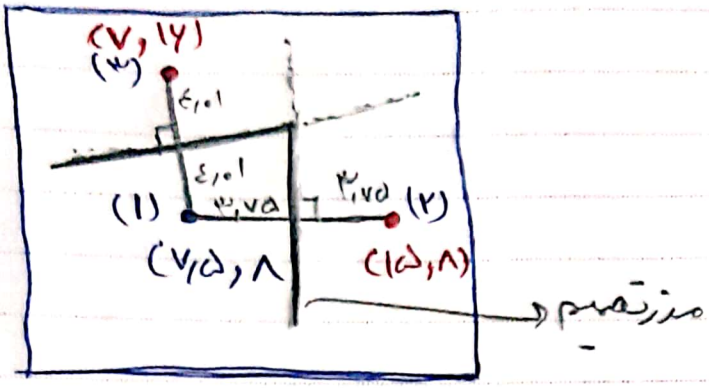
سوال ۱ - الف):

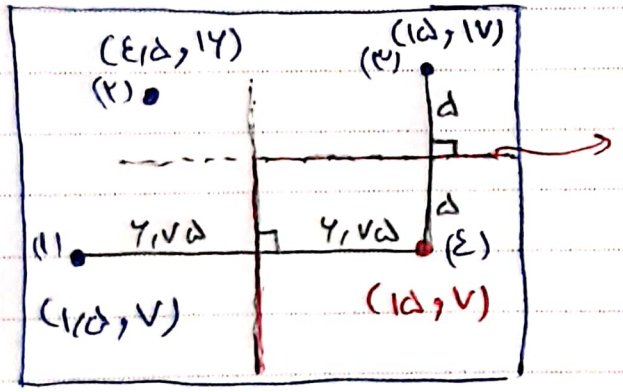
قسمت (۲):



$$\Delta 1,3 = \sqrt{(7.5)^2 + (8)^2} \simeq 8.02 \xrightarrow{\div 2} 4.01$$

$$\Delta 1,2 = \sqrt{(7.5)^2 + 0} = \sqrt{56.25} = 7.5 \xrightarrow{\div 2} 3.75$$

قسمت (۴)



$$\Delta 4,3 = \sqrt{5^2 + 10^2} = \sqrt{100} = 10 \xrightarrow{\div 2} d$$

$$\Delta 4,1 = \sqrt{(13.5)^2 + 5^2} = 13.5 \xrightarrow{\div 2} 6.75d$$

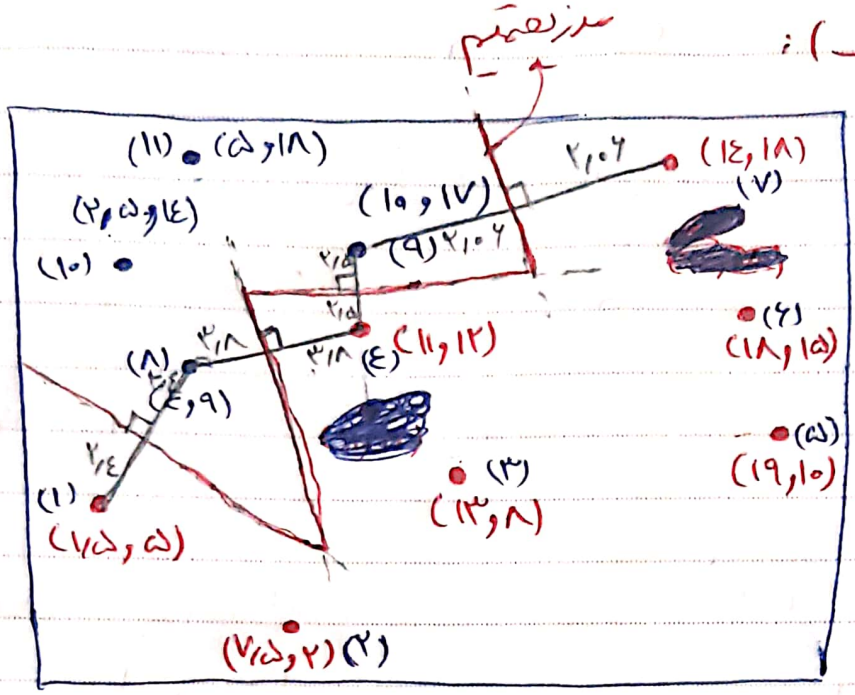
$$\Delta 4,2 = \sqrt{(10.5)^2 + 9^2} = \sqrt{191.25} = 13.82$$

$\Rightarrow$

$13.5 < 13.82 \Rightarrow$ بین فاصله نقطه ۴ با نقطه ۱ کمتر از فاصله نقطه ۴ با نقطه ۲ است پس نقطه ۴ به نقطه ۱ نزدیک‌تر است

ادامه سوال ۱ـ الف) :

قسمت (ج) :



$\Delta 7,9 = \sqrt{4^2 + 1^2} = \sqrt{17} \simeq 4,12$ ⟩ $4,12 < 9 \Rightarrow$ نقطه ۷ به ۹ متصل می‌شود

$\Delta 7,11 = \sqrt{9^2 + 0^2} = \sqrt{81} = 9$ ⟩

$\Delta 4,9 = \sqrt{1^2 + 5^2} = \sqrt{26} = 5,02$ ⟩

$\Delta 4,11 = \sqrt{6^2 + 6^2} = \sqrt{72}$

$\Delta 4,10 = \sqrt{(1,5)^2 + 2^2} = \sqrt{27,25}$

$\Delta 4,8 = \sqrt{7^2 + 3^2} = \sqrt{58}$ ⟩

$\sqrt{26}$ از بقیه کوتاه‌تر است و بین نقطه ۴ به نقطه ۹ باید متصل شود

$\Delta 1,8 = \sqrt{(2,5)^2 + 4^2} = \sqrt{22,25}$ ⟩ $\sqrt{22,25} < \sqrt{82} \Rightarrow$ بین نقطه ۱ به نقطه ۸ متصل می‌شود

$\Delta 1,10 = \sqrt{1^2 + 9^2} = \sqrt{82}$ ⟩

حال با این به تابع مورد نظر رسیده ایم اکنون یک نقطه از بین همین جواب ها
و مدار تقاطع گرفته و بعد این نیاز به حداقل یک عدد منفی و مدل نیاز است
تا این نقطه از کدام دو مدار گذشته. این نقطه ۸ و ۱ با نقاط ۴ و ۳ و ۲
بررسی و کنیم تا کوتاه ترین مشخص شود

$\sqrt{58}$ از سایر کوتاه تر است.

این نقطه ۸ با نقطه ۴ در تقاطع گرفته شده و این دو به یکدیگر متصل می شوند.

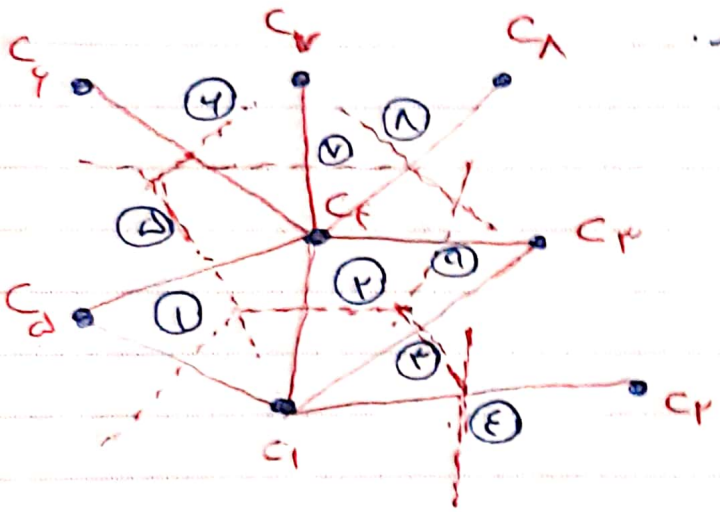
$\sqrt{58} = \sqrt{7^2 + 3^2} = \sqrt{58}$ ۵ ۸,۴

$\sqrt{82} = \sqrt{9^2 + 1^2} = \sqrt{82}$ ۵ ۸,۳

$\sqrt{61/25} = \sqrt{(7,5)^2 + 7^2} = \sqrt{61/25}$ ۵ ۸,۲

$\sqrt{58} = 7/62 \div 2 \rightarrow 3,8$

سوال 1 - ب) خیر. همواره نمیتوان گفت حتما اگر تعداد نمونه های آموزشی بیشتر باشد دقت بیشتر خواهد شد. ممکن است تعداد نمونه های آموزشی زیاد باشد ولی تعداد ویژگی ها کم باشد و در این صورت ممکن است حاشیه ها نتوانند به خوبی کلاس ها را از هم جدا کنند. مثلا اگر تعداد کلاس های مختلف زیاد باشد و نقاط هر کلاس به صورت پراکنده توزیع شده باشند ممکن است حاشیه های زیادی برای تفکیک کلاس ها نیاز باشد.



همانطور که در این مثال مشاهده می کنید تعداد نمونه های آموزشی ما ۹ تا است ولی برای جدا کردن ۸ کلاس مختلف از هم نیاز به ۸ حاشیه داریم و اگر تعداد حاشیه ها ۱۱ تا باشد نیاز به ۹ نمونه داریم تا حاشیه ها بتوانند کلاس ها را از هم جدا کنند. پس در این مثال تعداد نمونه های مختلف از کلاس های مختلف خیلی پراکنده بوده و نمونه های از کلاس های یکسان نزدیک هم نبوده تا حاشیه های زیادی تولید نشود.

SUBJECT:

Year: Month: Data: ( )

سوال ۱ - ج) : خیر. KNN این مشکل را ندارد. چون درخت تصمیم به اصطلاح یک یادگیرنده مشتاق است که زمانی که مجموعه ای از تابل های آموزشی وجود داشته باشد این داده های آموزشی را یادگرفته و با استفاده از آن مدل یا یک مدل پیش بینی کننده می سازد که اگر داده های جدید آموزشی وارد می شوند می توانند به روی ساختار کلی درخت تأثیر گذار باشند و آن را تغییر دهند. در مقابل KNN یک شیوه یادگیری تنبل است یعنی زمانی که تابل های آموزشی معلوم باشند صرفاً آن ها را ذخیره کرده و منتظر می ماند تا یک تابل تست وارد شود آنگاه شروع به کلاس بندی تابل تست بر پایه شباهت آن با تابل های ذخیره شده می کند و برخلاف یادگیرنده های مشتاق (مثل درخت تصمیم) یادگیرنده های تنبل که ربیشتر اوقات هنگام پیش بینی یا کلاس بندی انجام می دهند و نیازی به ساخت مدل یا یادگیری از روی تابل های آموزشی ندارند که این مشکل باعث می شود با ورود تابل های آموزشی جدید مشکلی به وجود نیاید و صرفاً تابل تست وارد شده، ممکن است در موقع کلاس بندی با تابل های آموزشی بیشتری مقایسه شده و محاسبات افزایش یابد.

SUBJECT:

Year: Month: Data: ( )

سوال ۲: مقرر درخت تصمیم ساخته می‌شود بسیاری از ویژگی‌ها می‌توانند نشان دهنده آنومالی‌ها در داده‌های آموزشی به علت نویز یا داده‌های پرت باشند. هرس درخت می‌تواند ویژگی‌هایی با کمترین اعتبار را حذف کند. از طرف دیگر هرس درخت می‌تواند مشکل overfitting داده‌ها را برطرف کند. اگر درخت خیلی ریشه کند به صحت پیش‌بینی داده‌های آموزشی اضافه کرده و در واقع اثرهای متقابل موجود در داده‌ها را یاد نگرفته است. به‌طوری که بتواند بر آن‌ها داده‌ی تست را تعمیم دهد. این مشکل باعث می‌شود خطای تست افزایش یابد درحالی که خطای آموزش کاهش است. این حالت overfitting خوانده می‌شود که هرس درخت می‌تواند این مشکل را برطرف کند. از طرف دیگر درخت هرس شده نسبت به کوچک‌تر شدن و ساده‌شدن به صحت کمتری دارند که باعث می‌شود روند کلی بین داده‌های تست را بدرستی دریافت. بیشتر انجام دهند روش‌های ساده‌تر قابل تعمیم‌تر هستند.

به‌طور کلی هرس کردن درخت باعث می‌شود generalization error کاهش یابد. بنابراین داده‌های تست با احتمال بیشتری به درستی دسته‌بندی شوند.

تقویه بین پیش هرس و پس هرس: در پیش هرس قبل از اینکه درخت بزرگ شود

ریشه کند با validation set ای که از داده های train جدا شده بود درخت را ارزیابی

می کنیم هر یا روحفای train و خطا روی validation set بررسی کرده تا در صورت

نیاز جهت ایجاد شاخه ها و نودهای بی از پیش در درخت گرفته شود.

ولی در پس هرس اجازه رشد درخت تا انتها پیش رفته و ریشه کنی شاخه ها

اضافی را از پس انجام می دهیم.

پیش هرس به صورت forwarding یا از بالا به پایین پیش می رود و پس هرس به صورت

backwarding یا پایین به بالا پیش می رود.

هم پیش هرس و هم پس هرس هردو به منظور برطرف کردن مشکل overfitting و

جلوگیری از پیچیدگی زیاد مدل مشترکا هستند اما در مرحله هرس ضریب خطی

یا Generalized error که دقت پیش بینی مدل است محاسبه کرده و براساس آن

شاخه ها را اضافه یا حذف می کند.

SUBJECT:
Year Month Data ( )

سوال ۳: $X = ($ تب = دارد ، سردرد = ندارد ، سرفه = دارد $)$

$$P(C_1 = \text{آری} \mid X) = \frac{P(X \mid C_1 = \text{آری})\, P(C_1 = \text{آری})}{P(X)}$$

$$P(X \mid C_1 = \text{آری}) = P(\text{تب} = \text{دارد} \mid C_1 = \text{آری}) \times$$

$$P(\text{سردرد} = \text{دارد} \mid C_1 = \text{آری}) \times P(\text{سرفه} = \text{ندارد} \mid C_1 = \text{آری}) =$$

$$\frac{4}{5} \times \frac{1}{5} \times \frac{2}{5} = \frac{6}{125}$$

$$P(C_1 = \text{آری}) = \frac{5}{10} = \frac{1}{2}$$

$$P(X \mid C_1 = \text{آری})\, P(C_1 = \text{آری}) = \frac{6}{125} \times \frac{1}{2} = \frac{6}{250}$$

$$P(C_2 = \text{خیر} \mid X) = \frac{P(X \mid C_2 = \text{خیر})\, P(C_2 = \text{خیر})}{P(X)}$$

$$P(X \mid C_2 = \text{خیر}) = P(\text{تب} = \text{دارد} \mid C_2 = \text{خیر}) \times$$

$$P(\text{سردرد} = \text{دارد} \mid C_2 = \text{خیر}) \times P(\text{سرفه} = \text{ندارد} \mid C_2 = \text{خیر})$$

$$= \frac{2}{5} \times \frac{3}{5} \times \frac{3}{5} = \frac{18}{125}$$

$$P(C_2 = \text{خیر}) = \frac{5}{10}$$

$$P(X \mid C_2 = \text{خیر})\, P(C_2 = \text{خیر}) = \frac{8}{125} \times \frac{1}{2} = \frac{8}{250}$$

$\frac{8}{250} > \frac{6}{250}$ $\Rightarrow$ پس این نمونه به کلاس خیر = C_2 اختصاص می‌یابد.

BOOSTAN

SUBJECT:

Year: Month: Data: ( )

سوال ۴ ← ب) :

$$P(C_1 = \text{bad loss} \mid x) = \frac{P(x \mid C_1)\, P(C_1)}{P(x)}$$

$$P(\text{income} = ۲۵۰۰۰ \mid C_1) = \frac{1}{\sqrt{۲\pi}\,\sigma}\, e^{\frac{(۲۵۰۰۰ - \mu)^۲}{۲\sigma^۲}}$$

if class = badloss && attri: income

$$\mu = \frac{۶۴۱۵۷۰/۹۸ + ۲۴۱۸۸/۱ + ۲۸۷۸۷/۳۴ + ۲۳۸۸۷/۷۲ + ۶۷۲۸۱/۴۴}{۵} = \boxed{۳۴۰۷۰/۱۱۶ = \mu}$$

$$\sigma^۲ = \frac{۱۴۷۳۳۴۱۱۸/۱۶}{(۱۲۰۹۴/۸۴۴)^۲} + \frac{۹۷۶۲۷۹۹/۹}{(۹۸۷۲۱۰۱۶)^۲} + \frac{۲۷۸۰۲۱۴۹۷۴۶/۲}{(۵۲۷۲/۷۷۶)^۲}$$

$$+ \frac{۱۰۲۶۹۷۹۸۶/۲}{(۱۰۱۷۳/۴)^۲} + \frac{۱۷۴۰۳۶۰۸/۲}{(۱۳۲۱۱/۳۵)^۲} = \frac{}{۵}$$

$$\boxed{۵۶۶۴۸۵۱۷۹۱/۸۴ = \sigma^۲} \qquad \boxed{\sigma = ۷۵۲۷۶/۲}$$

$$P(۲۵۰۰۰ \mid C_1) = \frac{1}{\sqrt{۲\pi} \times ۷۵۲۷۶/۲}\, e^{\frac{(۲۵۰۰۰ - ۳۴۰۷۰/۱۱۶)^۲}{۲ \times ۵۶۶۴۸۵۱۷۹۱/۸}}$$

BOOSTAN

=

SUBJECT

Year: Month: Date: ( )

ادامه سوال ۴ بخش ب):

$(۲۵۰۰۰ - ۴۰۶۰,۱۱۴)^۲ = ۸۲۰۸۵۷۰۱,۹$

$۲ \times ۵۲۲۴۸۵۱۷۹۱,۸ = ۱۱۳۲۹۷۰۳۵۸۳,۶$

$\frac{۸۲۰۸۵۷۰۱,۹}{۱۱۳۲۹۷۰۳۵۸۳,۶} = \%۰۰۷$ $e^{\%۰۰۷} = ۱,۰۰۷$

$\frac{۱/۰۰۷}{\sqrt{۲\pi} \times ۷۵۲۲۵۵,۲} = \frac{۱/۰۰۷}{۱۸۸,۱۶۳} = \%۰۰۵ = P(۲۵...|C_۱)$

$۲/۵$

---

if class = badloss && attri = age

$\mu = \frac{۲۲ + ۳۳ + ۲۸ + ۵۱ + ۲۵}{۵} = ۳۱,۸$

$\sigma^۲ = \frac{(۹۸)^۲ + (۱۱,۴)^۲ + (۳۸)^۲ + (۱۹۴)^۲ + (۷۸)^۲}{۵}$ (۹۸, ۱,۴, ۱۴,۴, ۲۴۸,۲, ۶۸)

$= ۱۰۵,۳۲ = \sigma^۲$ $\sigma = ۱۰,۲۶$

$\frac{(۲۴ - ۳۱,۸)^۲}{۲ \times ۱۰۵,۳۲}$ ۷۰,۱۸۴

$P(age = ۲۴ | C_۱) = \frac{۱}{\sqrt{۲\pi} \times ۱۰,۲۶} e$ =

$\frac{۷۰,۱۸۴}{۲۵۰,۱۴۴} = \%۲۸۹$ $e^{\%۲۸۹} = ۱,۳۳۵$

BOOSTAN

SUBJECT:

Year: Month: Data: ( )

ادامه سوال ۴ ـ ب) :

$$P(age=24|C_1)=\frac{1,33}{\sqrt{2\pi}\times 10,24}=\frac{1,33}{25,69}=\%052$$

$$P(X|C_1)=P(age=24|C_1)\times P(income=25000|C_1)=$$

$$=\%005\times\%052=\%00024$$

$$P(C_1)=\frac{5}{10}=\frac{1}{2}$$

$$P(X|C_1)P(C_1)=\%00013$$

$$P(C_2=Goodrisk\ |\ X)=\frac{P(X|C_2)P(C_2)}{P(X)}$$

if class = Goodrisk && attri : income

$$\mu=\frac{22994,9+28714,5+4918,75+4772,5+4125,34}{5}$$

$$=\frac{38948,998}{5}=7$$

$$\sigma^2=\frac{(4945,1598)^2+(10233,498)^2+(10411,5744)^2+(10470,5327)^2+(10482,1008)^2}{5}$$

BOOSTAN

$$=50590,2,34=\sigma^2$$

$$\sigma=7778,81$$

SUBJECT:

Year: Month: Data: ( )

ادامه سوال ۴ ـ ب):

$$P(\text{income} = ۲۵۰۰۰ \mid C_۲) = \frac{1}{\sqrt{۲\pi \times ۷۷۷۸,۸۱}} \times e^{-\frac{(۲۵۰۰۰ - ۳۸۹۴۸,۹۹۸)^۲}{۲ \times ۶۰۵۹۹۰۲,۳۴}}$$

$$(۲۵۰۰۰ - ۳۸۹۴۸,۹۹۸)^۲ = ۱۹۴۵۷۴۵۴۵,۲۰۴$$

$$۲ \times ۶۰۵۹۹۰۲,۳۴ = ۱۲۱۰۱۹۸۰۴,۶۸$$

$$\frac{۱۹۴۵۷۴۵۴۵,۲۰۴}{۱۲۱۰۱۹۸۰۴,۶۸} = ۱,۶۰۸$$

$$= P(۲۵۰۰۰ \mid C_۲)$$

$$\frac{۱,۶۰۸}{\sqrt{۲\pi} \times ۷۷۷۸,۸۱} = \frac{۱۱۶۰۸}{۱۹۶۴۷۱۰۲۵} = ۰,۰۰۰۰۸$$

$۲,۵$

if class = Goodrisk && attri: age

$$\mu = \frac{۳۹ + ۳۴ + ۳۵ + ۳۰ + ۴۴}{۵} = ۳۵,۸$$

$$\sigma^۲ = \frac{(۱۳,۸)^۴}{(۱۳,۴)^۲} + (۱۱)^۲ + (۶,۴)^۲ + (۴,۸)^۲ + \frac{۱۹۰,۴}{۱۷۴,۲} + \frac{۱,۴}{۵}$$

$$= ۷۵,۷۲ = \sigma^۲$$

$$\sigma = ۸,۷$$

ادامه سوال ۴ ب):

$$\frac{(۲۴ - ۳۲/۸)^۲}{۲ \times ۷۵/۸۲}$$

$$P(age = ۲۴ | C_۲) = \frac{۱}{\sqrt{۲\pi} \times ۸/۷} \times e$$

$(۲۴ - ۳۲/۸)^۲ = ۸۲۹/۴۴$ $\quad ۲ \times ۷۵/۸۲ = ۱۵۱/۶۴$

$\frac{۸۲۹/۴۴}{۱۵۱/۶۴} = ۵/۴۷$ $\quad e^{۵/۴۷} = ۲۳۷/۴۶$

$$\frac{۲۳۷/۴۶}{\sqrt{۲\pi} \times ۸/۷} = \frac{۲۳۷/۴۶}{۲۱/۷۵} = ۱۰/۹۲ = P(۲۴ | C_۲)$$

$۲/۵$

$P(X | C_۲) = P(age = ۲۴ | C_۲) \times P(Income = ۲۵۰۰۰ | C_۲) =$

$۱۰/۹۲ \times \%۰۰۰۸ = \%۰۰۰۸۷$

$P(C_۲) = \frac{۵}{۱۰}$ $\quad P(X | C_۲) P(C_۲) = \%۰۰۰۴۳۵$

$\%۰۰۰۴۳۵$ $\quad \%۰۰۰۱۳$

$P(X | C_۲) P(C_۲) > P(X | C_۱) P(C_۱)$

$\Rightarrow$ بنابراین نتیجه می‌گیریم که این فرد به $C_۲$ یا Good risk است.

CamScanner