Subject: Date:

810100355 تمرین ML _ سری ۱ سارا رستمی

①

الف)

$$a \leqslant b \xrightarrow[(a>0)]{\times a} a^2 \leqslant ab \xrightarrow{\sqrt{}} \sqrt{a^2} \leqslant \sqrt{ab} \rightarrow a \leqslant \sqrt{ab}$$

ب)

$$p(\text{error}) = p(\text{mistake}) = \int_{R_1} \underbrace{P(x, C_2)\,dx}_{P(C_2|x)P(x)dx} + \int_{R_2} \underbrace{P(x, C_1)\,dx}_{P(C_1|x)P(x)dx} \quad (*)$$

می دانیم در ناحیه ی مربوط به کلاس 1 (در R1):

$$P(C_2|x) < P(C_1|x) \xrightarrow{\text{طبق بخش الف}} P(C_2|x) \leqslant \sqrt{P(C_2|x)P(C_1|x)}$$

همچنین در ناحیه ی R2 داریم:

$$P(C_1|x) < P(C_2|x) \xrightarrow{\text{طبق بخش الف}} P(C_1|x) \leqslant \sqrt{P(C_1|x)P(C_2|x)}$$

آنگاه داریم:

$$\int_{R_1} \overbrace{P(x, C_2)}^{P(C_2|x)P(x)}dx \leqslant \int_{R_1} \sqrt{P(C_2|x)P(C_1|x)}\,.\,p(x)dx$$

$$\int_{R_2} \underbrace{P(x, C_1)}_{P(C_1|x)P(x)}dx \leqslant \int_{R_2} \sqrt{P(C_1|x)P(C_2|x)}\,.\,p(x)dx$$

$$\xrightarrow{(*)} p(\text{error}) = p(\text{mistake}) \leqslant \int_{R_1} \sqrt{P(C_1|x)P(C_2|x)}\,.\,p(x)dx + \int_{R_2} \sqrt{P(C_1|x)P(C_2|x)}\,.\,p(x)dx = \int \sqrt{P(C_1|x)P(C_2|x)}\,.\,p(x)dx$$

Subject: Date:

زیر رادیکال می آوریم

$$\int \sqrt{P(C_1|x)P(C_2|x)}\, P(x)\, dx = \int \sqrt{\underbrace{P(C_1|x).P(x)}_{P(x,C_1)} . \underbrace{P(C_2|x)P(x)}_{P(x,C_2)}}\, .dx$$

$$= \int \{P(x,C_1).P(x,C_2)\}^{1/2}\, dx \Rightarrow P(\text{mistake}) \leqslant \int \{P(x,C_1)P(x,C_2)\}^{1/2}\, dx.$$

---

②

الف)

Recall: کسری از داده های متعلق به کلاس هدف (positive) که توسط مدل ما به درستی پیش بینی شده اند.

$$\text{Recal} = \frac{TP}{TP+FN}$$

precision: کسری از داده هایی که عضو کلاس هدف پیش بینی شده اند (positive پیش بینی شده اند) که واقعاً عضو کلاس هدف بوده اند. (از بین داده هایی که + پیش بینی شده اند، چه کسری از آن ها واقعاً درست پیش بینی شده و عضو کلاس بوده اند)

$$\text{precision} = \frac{TP}{TP+FP}$$

accuracy: کسر داده هایی که به درستی پیش بینی شده اند. (به درستی تشخیص داده شده که عضو کلاس هستند یا عضو کلاس نیستند)

$$\text{Accuracy} = \frac{TN+TP}{TN+FN+TP+FP}$$

T: یعنی تشخیص درست

F: یعنی تشخیص (پیش بینی) اشتباه

P: عضو کلاس + تشخیص داده شده

N: عضو کلاس − تشخیص داده شده (در واقع تشخیص داده شده که عضو کلاس + نیست)

ب) ریسک تشخیص ندادن بیماری یک فرد بیمار بیشتر از ریسک تشخیص ندادن سلامت یک فرد سالم است. اگر بیماری یک فرد را نتوانیم تشخیص دهیم مسلماً خطرناک‌تر از این است که به یک فرد سالم بگوییم بیمار است.

پس ما باید به دنبال معیاری باشیم که FN را کمینه کند و TP را بیشینه کند. TP همان تشخیص درست بیماری بوده و FN عدم تشخیص بیماری یک فرد بیمار می‌باشد.

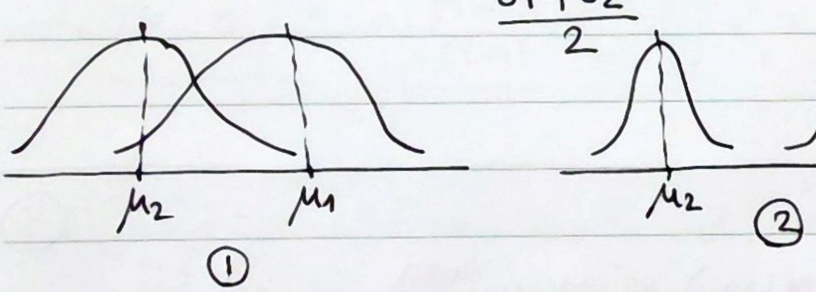
با افزایش مقدار Recall، این اهداف محقق می‌شوند.

$$Recall = \frac{TP}{TP+FN}$$

ج) discriminability به معنی قابلیت تمایز دو توزیع (در اینجا به طور خاص توزیع) می‌باشد.

هرچه فاصله‌ی میانگین دو توزیع از هم بیشتر بوده و واریانس آن‌ها کمتر باشد، توزیع‌ها از هم discriminable تر هستند و بیشتر قابل تمیز می‌باشند.

$$discriminability = \frac{|\mu_2 - \mu_1|}{\frac{\sigma_1+\sigma_2}{2}} \longrightarrow$$ هرچه بیشتر باشد، دو کلاس راحت‌تر قابل تمیز هستند



discriminability بین دو توزیع در شکل ② بیشتر از شکل ① است

د) هدف از classification این است که برای هر داده‌ی ورودی unseen مثل x بتوانیم یک کلاس (همان label) assign کنیم. به تعبیری تشخیص دهیم این داده‌ی جدید به چه کلاسی تعلق دارد.

روش‌های discriminative، مدل از احتمال شرطی ($P(w_i|x)$) برای پیش‌بینی داده‌های unseen استفاده می‌کند. این روش‌ها برای بدست آوردن احتمال، به طور مستقیم یک توزیع فرضی برای $P(w_i|x)$ در نظر می‌گیرند و سپس پارامترهای این توزیع را با کمک داده‌های train تخمین می‌زنند.

برخی مدل‌های discriminative عبارتند از، logistic Regression، SVMs، nearest neighbor، decision Trees & Random Forest و ...

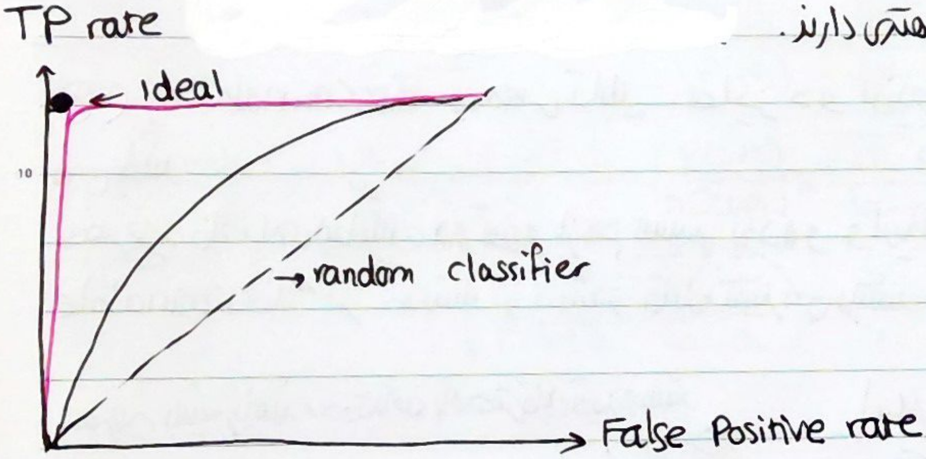
روش‌های Generative، برای بدست آوردن احتمال شرطی $P(Y|X)$، احتمال Prior $P(Y)$ و likelihood، $P(X|Y)$ را به کمک داده‌ی train تخمین می‌زند و سپس

طبق رابطه‌ی Bayes، احتمال posterior، $P(Y|X)$ را محاسبه می‌کند.

برخی مدل‌های Generative عبارتند از: Naive Bayes، Bayesion networks، HMMS و ...

ه) برای یک طبقه‌بند مطلوب ← ایده‌آل، سطح زیر نمودار ROC بیشتر خواهد بود، به طوری که در حالت ایده‌آل (طبقه‌بندی که همیشه همه‌ی داده‌ها را درست دسته‌بندی کند) سطح زیر نمودار $+\infty$ خواهد شد.

classifier هایی که منحنی آن‌ها به گوشه‌ی بالا سمت چپ نزدیک‌تر است، عملکرد بهتری دارند.



3

$$P(x|w_2) \sim U(a, b) \quad \text{و} \quad P(x|w_1) \sim N(\mu, \delta^2)$$

$$g(x) = g_1(x) - g_2(x) = P(w_1|x) - P(w_2|x) = P(x|w_1)P(w_1) - P(x|w_2)P(w_2)$$

$$\Rightarrow \text{if } g(x) \begin{cases} > 0 & \text{choose } w_1 \\ < 0 & \text{choose } w_2 \end{cases} \Rightarrow P(x|w_1)P(w_1) > P(x|w_2)P(w_2)$$

$$\Rightarrow \frac{P(x|w_1)}{P(x|w_2)} > \frac{P(w_2)}{P(w_1)} \Longrightarrow$$ در این صورت w_1

ممکن است $P(x|w_2) = 0$ در این صورت نباید در مخرج باشد. پس می‌نویسیم

$$\text{decision} \begin{cases} w_1 \; ; & P(x|w_1) > \frac{P(w_2)}{P(w_1)} \times P(x|w_2) \\ w_2 \; ; & \text{otherwise} \end{cases}$$

ادامه ③

- if $x \notin [a,b] \Longrightarrow p(x|\omega_2)=0 \xrightarrow{\text{Decision}} \omega_1$

- if $x \in [a,b]$:

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} > \frac{P(\omega_2)}{P(\omega_1)} \times \frac{1}{b-a} \longrightarrow e^{-\frac{(x-\mu)^2}{2\sigma^2}} > \frac{P(\omega_2)}{P(\omega_1)} \times \frac{\sqrt{2\pi}\,\sigma}{b-a}$$

$$\xrightarrow{\ln} -\frac{(x-\mu)^2}{2\sigma^2} > \ln\left(\frac{P(\omega_2)}{P(\omega_1)} \times \frac{\sqrt{2\pi}\,\sigma}{b-a}\right)$$

$$\Longrightarrow (x-\mu)^2 < -2\sigma^2 \ln\left(\frac{P(\omega_2)}{P(\omega_1)} \times \frac{\sqrt{2\pi}\,\sigma}{b-a}\right)$$

$$\Longrightarrow -\sqrt{-2\sigma^2 \ln\left(\frac{P(\omega_2)}{P(\omega_1)} \times \frac{\sqrt{2\pi}\,\sigma}{b-a}\right)} < x-\mu < \sqrt{-2\sigma^2 \ln\left(\frac{P(\omega_2)}{P(\omega_1)} \times \frac{\sqrt{2\pi}\,\sigma}{b-a}\right)}$$

$$\Longrightarrow \underbrace{\mu - \sigma\sqrt{-2\ln\left(\frac{P(\omega_2)}{P(\omega_1)} \times \frac{\sqrt{2\pi}\,\sigma}{b-a}\right)}}_{A} < x < \underbrace{\mu + \sigma\sqrt{-2\ln\left(\frac{P(\omega_2)}{P(\omega_1)} \times \frac{\sqrt{2\pi}\,\sigma}{b-a}\right)}}_{B}$$

حالت ① اگر A و B حقیقی نباشند ، بازه‌ی R_1 و R_2 را می‌توان به صورت زیر تعریف کرد:

$R_1 = (-\infty, a) \cup (b, +\infty)$ و $R_2 = (a,b)$

$$P(\text{error}) = P(\omega_1) \times \int_{R_2} p(x|\omega_1)\,dx + P(\omega_2) \times \int_{R_1} \cancel{p(x|\omega_2)}^{\,0}\,dx = \ldots$$

$$\ldots = P(\omega_1) \times \int_{R_2} p(x|\omega_1)\,dx = P(\omega_1) \int_a^b p(x|\omega_1)\,dx \leqslant \int_a^b p(x|\omega_1)\,dx$$

حالت ② اگر A و B حقیقی باشند $\Longrightarrow -2\ln\left(\frac{P(\omega_2)}{P(\omega_1)} \times \frac{\sqrt{2\pi}\,\sigma}{b-a}\right) > 0$

$$\ln\left(\frac{P(\omega_2)}{P(\omega_1)} \times \frac{\sqrt{2\pi}\,\sigma}{b-a}\right) < 0 \Longrightarrow \frac{P(\omega_2)}{P(\omega_1)} \times \frac{\sqrt{2\pi}\,\sigma}{b-a} < e^0 = 1$$

$$\frac{P(\omega_2)}{P(\omega_1)} \times \frac{1}{b-a} < \frac{1}{\sqrt{2\pi}\,\sigma} \Longrightarrow \frac{P(\omega_2)}{P(\omega_1)} \times \frac{1}{b-a} < \frac{1}{\sqrt{2\pi}\,\sigma} \underbrace{\left(e^{-\frac{(\mu-\mu)^2}{2\sigma^2}}\right)}_{1}$$

$$x=\mu$$

چون این نامعادله به ازای $x \in (A,B)$ است ← تمام نقاط این بازه به کلاس ω_1 متعلق هستند. در این صورت بازه ی R_1 و R_2 را می‌توان تعریف کرد:

$$R_1 = (-\infty, a) \cup (b, +\infty) \cup (A,B)$$

$$R_2 = (a,b) - (A,B) = (a,b) \setminus (A,B) \longrightarrow \text{semi minus}$$

$$P(\text{error}) = P(\omega_1)\int_{R_2} p(x|\omega_1)dx + P(\omega_2)\int_{R_1} p(x|\omega_2)dx = P(\omega_1)\int_{(a,b)\setminus(A,B)} p(x|\omega_1)dx$$

$$+ P(\omega_2)\int_{(a,b)\cap(A,B)} p(x|\omega_2)dx \xrightarrow[\text{چون}]{} p(x|\omega_2)=0 \text{ if } x \notin (a,b)$$

حال، برای همه ی نقاط x به طوری که $x \in (A,B)$:

$$P(\omega_1)P(x|\omega_1) > P(\omega_2)P(x|\omega_2)$$

پس

$$\longrightarrow P(\omega_1)\int_{(a,b)\cap(A,B)} p(x|\omega_1)dx > P(\omega_2)\int_{(a,b)\cap(A,B)} p(x|\omega_2)dx$$

$$\Longrightarrow p(\text{error}) \leqslant P(\omega_1)\int_{(a,b)\setminus(A,B)} p(x|\omega_1)dx + P(\omega_1)\int_{(a,b)\cap(A,B)} p(x|\omega_1)dx =$$

$$P(\omega_1)\int_a^b p(x|\omega_1)dx \leqslant \int_a^b p(x|\omega_1)dx$$

برای هر دو حالت ① و ② ، upper bound خطا برابر شد با:

$$\boxed{p(\text{error}) \leqslant \int_a^b p(x|\omega_1)dx}$$

ادامه‌ی سؤال ③ طبق صورت سؤال $p(x|w_1)$ گوسی است پس:

$$\text{upper bound} = \int_a^b p(x|w_1)\,dx \underset{\text{normalize}}{=} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}z^2}\,dz =$$

upper bound

$$= \int_{-\infty}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}z^2}\,dz - \int_{-\infty}^{\frac{a-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}z^2}\,dz = G\left(\frac{b-\mu}{\sigma}\right) - G\left(\frac{a-\mu}{\sigma}\right)$$

---

④ فرمول هر یک از distance ها به صورت زیر است (برای دو کلاس B, R):

Euclidean distance $\rightarrow \sqrt{\sum_{i=1}^{d} (R_i - B_i)^2}$

Manhattan distance $\rightarrow \sum_{i=1}^{d} |R_i - B_i|$

Chebyshev distance $\rightarrow \max\limits_{i \in (1,2,\dots,d)} \{ |R_i - B_i| \}$

مقدار centroid برای کلاس های Blue و Red به صورت زیر است:

$$\vec{\mu}_{Blue} = \frac{1}{4}\left(\begin{bmatrix} 2 \\ 5 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} 4 \\ 1 \end{bmatrix} + \begin{bmatrix} 8 \\ 4 \end{bmatrix} \right) = \begin{bmatrix} 4 \\ 2.25 \end{bmatrix}$$

$$\vec{\mu}_{Red} = \frac{1}{4}\left(\begin{bmatrix} -3 \\ 9 \end{bmatrix} + \begin{bmatrix} -3 \\ 4 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right) = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$$

Nearest centroid _ Chebyshev:

$(2,2)$, $(4, 2.25)$: $\max\{2, 0.25\} = 2$ class B

$(2,2)$, $(-2, 4)$: $\max\{4, 2\} = 4$ class Red $\Longrightarrow$ class B

Nearest centroid _ Manhattan distance:

(2,2) , (4, 2.25) : 2 + 0.25 = 2.25 class B
(2,2) , (−2,4) : 4 + 2 = 6 class R

$\Longrightarrow$ $\boxed{\text{class B}}$

Nearest centroid _ euclidean distance:

(2,2) , (4, 2.25) : $\sqrt{4 + 0.0625} = 2.0155$ class B
(2,2) , (−2,4) : $\sqrt{16 + 4} = 4.472$ class R

$\Longrightarrow$ $\boxed{\text{class B}}$

Nearest neighbor _ chebyshev distance:

(2,2) , (1,4) : max(1,2) = 2 class R
(2,2) , (2,5) : max(0,3) = 3 class B
(2,2) , (4,1) : max(2,1) = 2 class B

فرقی نداره. تصادفی
یک کلاس انتخاب میشه
(فاصله‌ی کمینه از هردو کلاس برابره)

Nearest neighbor _ Manhattan distance:

(2,2) , (1,4) : 1 + 2 = 3 class R
(2,2) , (2,5) : 0 + 3 = 3 class B
(2,2) , (4,1) : 2 + 1 = 3 class B

اینجاهم فاصله‌ها برابر شده
(برای هردو کلاس). پس
به طور تصادفی یک کلاس انتخاب میشه

Nearest neighbor _ Euclidean distance:

(2,2) , (1,4) : $\sqrt{1+4} = \sqrt{5}$ class R
(2,2) , (2,5) : $\sqrt{0+9} = 3$ class B
(2,2) , (4,1) : $\sqrt{4+1} = \sqrt{5}$ class B

//

	Nearest neighbor	Nearest centroid
chebyshev	Red or Blue	Blue
Manhattan	Red or Blue	Blue
Euclidean	Red or Blue	Blue

5

$$P(\mu|X) = \frac{P(X|\mu)P(\mu)}{P(X)}$$

ثابت K

$P(X)$ ثابته پس می‌تونیم آن را درنظر نگیریم. همچنین چون داده‌ها به صورت i.i.d هستند، می‌نویسیم:

ثابت

$$P(\mu|X) = K\prod_{i=1}^{N} P(x_i|\mu)P(\mu) = K\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\Sigma}} e^{-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)}$$

ثابت

$$\times \frac{1}{\sqrt{2\pi\Sigma_0}} e^{-\frac{1}{2}(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)} = K' e^{-\frac{1}{2}\left(\sum_{i=1}^{N}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) + (\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)\right)}$$

ثابت

$$= K'\exp\left\{\frac{-1}{2}\left(\sum_{i=1}^{N}\left(x_i^T\Sigma^{-1}x_i - \mu^T\Sigma^{-1}x_i - x_i^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu\right) + \right.\right.$$

cte ثابت

$$\left.\left. + \left(\mu^T\Sigma_0^{-1}\mu - \mu_0^T\Sigma_0^{-1}\mu - \mu^T\Sigma_0^{-1}\mu_0 + \mu_0^T\Sigma^{-1}\mu_0\right)\right)\right\} = \quad (*)$$

$$= \boxed{K''\exp\left\{\frac{-1}{2}\mu^T(N\Sigma^{-1}+\Sigma_0^{-1})\mu + \mu^T\left(\Sigma^{-1}\sum_{i=1}^{N}x_i + \Sigma_0^{-1}\mu_0\right) + cte\right\}}$$

می‌دانیم:

$$P(\mu|x) = K''\exp\left\{\frac{-1}{2}(\mu-\mu_n)^T\Sigma_n^{-1}(\mu-\mu_n)\right\} =$$

$$= K''\exp\left\{\frac{-1}{2}\left(\mu^T\Sigma_n^{-1}\mu - \mu^T\Sigma_n^{-1}\mu_n - \mu_n^T\Sigma_n^{-1}\mu + \overbrace{\mu_n^T\Sigma_n^{-1}\mu_n}^{cte}\right)\right\}$$

$$= \boxed{K''\exp\left\{\frac{-1}{2}\mu^T\Sigma_n^{-1}\mu + \mu^T\Sigma_n^{-1}\mu_n + cte\right\}} \quad (\bigstar)$$

$$\underline{(*) = (\bigstar)} \Longrightarrow \Sigma_n^{-1} = N\Sigma^{-1} + \Sigma_0^{-1} \xrightarrow{(a+b)^{-1} = b^{-1}(a^{-1}+b^{-1})a^{-1}} \boxed{\Sigma_n = \Sigma_0\left(\frac{1}{N}\Sigma + \Sigma_0\right)^{-1}\frac{1}{N}\Sigma}$$

$$\mu^T\left(\Sigma^{-1}\sum_{i=1}^{N}x_i + \Sigma_0^{-1}\mu_0\right) = \mu^T\Sigma_n^{-1}\mu_n \Longrightarrow \mu_n = \Sigma_n\left(\Sigma^{-1}\sum_{i=1}^{N}x_i + \Sigma_0^{-1}\mu_0\right)$$

$$\Longrightarrow \boxed{\mu_n = \Sigma_0\left(\frac{1}{N}\Sigma + \Sigma_0\right)^{-1}\frac{1}{N}\Sigma\left(\Sigma^{-1}\sum_{i=1}^{N}x_i + \Sigma_0^{-1}\mu_0\right)}$$