9913009 محمد محمد باقری تمرین دوم مبانی سامانه‌های الگو

سوال اول

a

مشخص کردن متغیرهای تصادفی:

R : فرد سالم است، به Covid-19 مبتلا نشده است.

S : جواب تست مثبت است.

$$P(R) = \frac{996}{1000} \quad , \quad P(\tilde{R}) = \frac{4}{1000}$$

حال باید احتمال نمونه‌های زیر را مقایسه کنیم تا بتوانیم تصمیم بگیریم نیاز به نگران شدن هست یا خیر؟

$$P(R|S) \underset{\tilde{R}}{\overset{R}{\gtrless}} P(\tilde{R}|S) \Rightarrow$$

$$\frac{P(S|R)\,P(R)}{\cancel{P(S)}} \underset{\tilde{R}}{\overset{R}{\gtrless}} \frac{P(S|\tilde{R})\,P(\tilde{R})}{\cancel{P(S)}}$$

$P(S|R)$: یعنی به شرط اینکه فرد سالم باشد، جواب تست مثبت باشد

یعنی تشخیص اشتباه $\Leftarrow FPR$

$P(S|\tilde{R})$: یعنی فرد مبتلا باشد و جواب تست مثبت باشد، یعنی تشخیص درست $\Leftarrow TPR$

$$TPR = 1 - \overset{0.17}{FNR} = 0.83$$

$\Rightarrow$ نوشتن تابع

likelihood: $\overset{0.11}{FPR} \times \frac{996}{1000} \boxed{>} \overset{0.83}{TPR} \times \frac{4}{1000} \Rightarrow$

در کلاس R قرار میگیریم $\Leftarrow$ فرد سالم است $\Leftarrow$ نیازی به نگران بودن نیست

b

$$P(R) = \frac{999}{1000} \quad , \quad P(\tilde{R}) = \frac{1}{1000}$$

$$P(\tilde{R}|S) = ? \qquad P(\tilde{R}|S) = \frac{P(S|\tilde{R})\,P(\tilde{R})}{P(S)}$$

ادامه‌ی b

$$P(S) = P(S|R)P(R) + P(S|\tilde{R})P(\tilde{R})$$

$$= FPR \times P(R) + TPR \times P(\tilde{R}) =$$

$$0.27 \times 0.999 + 0.81 \times 0.001 = 0.27$$

$$\Rightarrow P(\tilde{R}|S) = \frac{0.81 \times 0.001}{0.27} = \boxed{0.003}$$

c

در این سوال $\frac{TP+TN}{P+N}$ برابر با 0.77 می‌شود درحالی که در سوال ذکر شده مقدار accuracy، 0.73 است **

چون از فرد در شهر قبلی تست کرونا داده است پس همان احتمالی که برای قسمت قبلی داریم

همان احتمال را نیاز است. ← دو حالت کلی برای واقعاً کرونا داشتن وجود دارد:

$\underbrace{\text{1 در شهر قبلی مبتلا شده است}}_{P(\tilde{R}|S)} \times \underbrace{\text{2 در شهر قبلی مبتلا نشده است}}_{P(\tilde{B})}$ یا و در این شهر مبتلا شده $P(B) = 0.05$

$$P(\text{مورد سوال}) = 0.003 \times 0.95 + 0.05 = \boxed{0.05285}$$

اطلاعات سوال.

d

$P(S|\tilde{R}) = 0.76$, $P(\tilde{S}|R) = 0.68$

$P(\tilde{R}) = 0.025$, $P(\tilde{R}|\tilde{S}) = ?$,

$P(\tilde{S}|\tilde{R}) = 0.24$, $P(S|R) = 0.32$, $P(R) = 0.975$

$$P(\tilde{R}|\tilde{S}) = \frac{P(\tilde{S}|\tilde{R})P(\tilde{R})}{P(\tilde{S})} = \frac{P(\tilde{S}|\tilde{R})P(\tilde{R})}{P(\tilde{S}|R)P(R) + P(\tilde{S}|\tilde{R})P(\tilde{R})}$$

$$= \frac{0.24 \times 0.025}{0.68 \times 0.975 + 0.24 \times 0.025} = \boxed{0.009}$$

---

e $P(\tilde{R}|S) = ?$

$$P(\tilde{R}|S) = \frac{P(S|\tilde{R})P(\tilde{R})}{P(S)} = \frac{0.76 \times 0.025}{P(S|R)P(R) + P(S|\tilde{R})P(\tilde{R})}$$

$$= 0.057$$

---

f

همان طور که از قسمت قبل داریم:

$P(\tilde{R}|S) = 0.057$, $P(R|S) = 0.943$

احتمال کرونا نداشتن بیشتر است.

h

$P(R) = 0.99 \quad , P(\tilde{R}) = 0.01$

$P(S \mid \tilde{R}) = p \quad , P(S \mid R) = 1-p$

$P(\tilde{R} \mid S) = ?$

$$P(\tilde{R} \mid S) = \frac{P(S \mid \tilde{R})\, P(\tilde{R})}{P(S)} = \frac{p \times 0.01}{P(S \mid R) P(R) + P(S \mid \tilde{R}) P(\tilde{R})} =$$

$$\frac{0.01 p}{(1-p) \times 0.99 + p \times 0.01} = \boxed{\frac{p}{99 - 98p}}$$

i

$\underbrace{P(\tilde{R} \mid S)}_{TPR} = a \qquad \underbrace{P(\tilde{R} \mid \tilde{S})}_{FNR} = 1-a$

برای باور کردن جواب مثبت حداقل باید a برابر با 0.9 باشد.

j

$$\frac{p}{99-98p} \geqslant \frac{9}{10} \Rightarrow 10p \geqslant 891 - 882p \Rightarrow$$

$$892p \geqslant 891 \Rightarrow p \geqslant \frac{891}{892} \Rightarrow p \geqslant 0.998$$

---

g

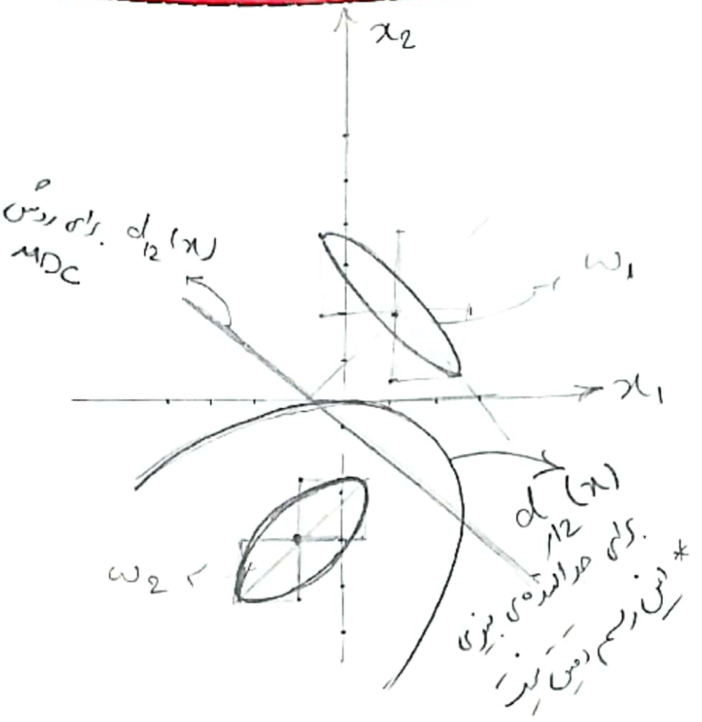
همانطور که از قسمت d داشتیم:

$P(\tilde{R} \mid \tilde{S}) = 0.009 \; , \; P(R \mid \tilde{S}) = 0.991$

بنابراین در صورت منفی بودن جواب تست، احتمال کرونا نداشتن بیشتر از احتمال داشتن آن است.

سوال دوم

a



b

decision boundary for bayesian classifier

$$d_{12}(\underline{x}) = -\frac{1}{2}(\underline{x}-\underline{\mu}_1)^T \Sigma_1^{-1} (\underline{x}-\underline{\mu}_1) + \frac{1}{2}(\underline{x}-\underline{\mu}_2)^T \Sigma_2^{-1} (\underline{x}-\underline{\mu}_2)$$

$$\Sigma_1^{-1} = \begin{bmatrix} 0.65 & 0.25 \\ 0.25 & 0.65 \end{bmatrix}$$

$$\Sigma_2^{-1} = \begin{bmatrix} 0.69 & -0.13 \\ -0.13 & 0.69 \end{bmatrix}$$

$$d_{12}(x) = -\frac{1}{2}\begin{bmatrix} x_1-1 & x_2-2 \end{bmatrix}\begin{bmatrix} 0.65 & 0.25 \\ 0.25 & 0.65 \end{bmatrix}\begin{bmatrix} x_1-1 \\ x_2-2 \end{bmatrix} +$$

$$\frac{1}{2}\begin{bmatrix} x_1+1 & x_2+3 \end{bmatrix}\begin{bmatrix} 0.69 & -0.13 \\ -0.13 & 0.69 \end{bmatrix}\begin{bmatrix} x_1+1 \\ x_2+3 \end{bmatrix} =$$

$$x_1^2 - x_2^2 - 19x_1x_2 + 72.5x_1 + 174.5x_2 + 46.75 = 0$$

که یک تابع درجه دوم است , در شکل قسمت قبل این خط نمایش داده شده است.

decision boundary for MDC :

$$d_{12}(\underline{x}) = \mu_1^T \underline{x} - \frac{1}{2}\mu_1^T\mu_1 - \mu_2^T \underline{x} + \frac{1}{2}\mu_2^T\mu_2$$

$$= 2x_1 + 5x_2 + 2.5 = 0$$

این خط در شکل صفحه‌ی قبل رسم شده است.

گزارش قسمت عملی سؤال دوم

c

خطای بدست آمده با روش MDC : 0.003

〃 〃 〃 〃 بیز : 0.01

d

F-score با در نظر گرفتن هزینه برای اشتباه دسته‌بندی شدن : 0.984

〃 بدون 〃 〃 〃 〃 〃 〃 : 0.991

e

میانگین خطا در روش بیز : 0.0129

〃 〃 〃 MDC : 0.0296

فایل‌ها در پوشه‌ی تمرین قرار داده شده‌اند و با اجرا کردن نتایج نمایش داده می‌شوند.

f

نتایج در فایل تمرین قرار داده شده‌اند. خطوط سبز، کانتور لاین توزیع‌ها است. چون در سؤال گفته شده $\Sigma = \sigma^2 I$ <= توزیع‌ها یا دایره‌ای هستند یا بیضی عمودی یا افقی. در تمام موارد بیز موفق به دسته‌بندی شده است. چون در هیچکدام توزیع‌ها کاملاً بر هم منطبق نیستند. در رسم خطوط تقسیم سعی شده احتمال کلاس بزرگتر در تقسیم‌بندی نواحی مشترک تأثیرش مشخص باشد.

خطوط مشکی، مرز تقسیم هستند. در تصویر دوم میانگین‌ها بر هم منطبق‌اند ولی چون واریانس‌ها متفاوت است و در فرمول توزیع نرمال فاصله‌ی ماهالانوبیس وجود دارد <= بیز در تصویر 2 هم شکست نمی‌خورد.

سوال سوم

a

ω_1: Petal , ω_2: stamen , ω_3: stigma

$$P(\omega_i) = \frac{1}{3}$$

$$\underline{\mu}_1^T = [-48.74 \quad 11.6 \quad 0 \quad 0.74 \quad 0.62 \quad 0.69] .$$

$$\underline{\mu}_2^T = [56.86 \quad 26 \quad 6.4 \quad 0.09 \quad 0.81 \quad 0.8]$$

$$\underline{\mu}_3^T = [46.08 \quad 22 \quad 1 \quad 0.2 \quad 0.82 \quad 0.89]$$

مجموعه ویژگی اول:

$$P(x|\omega_i) = N(\mu, \sigma^2)$$

$$P(x|\omega_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+48.74)^2}$$

$$P(x|\omega_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-56.86)^2}$$

$$P(x|\omega_3) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-46.08)^2}$$

مجموعه ویژگی دوم و سوم:

$$g_i = \ln(P(x|\omega_i)\, P(\omega_i))$$

$$P(x|\omega_i) = \frac{1}{\sqrt{2\pi}\, |\Sigma_i|^{\frac{1}{2}}} \exp\{(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)\}$$

$$d_{ij}(x) = g_i(x) - g_j(x$$

تنها کافی است در فرمول بالا مقادیر را جایگذاری کنیم و معادله‌ی مرزهای تقسیم را بیابیم.

به این نکته دقت شود که برای حالت دو بعدی یعنی مجموعه ویژگی دوم باید $\mu_i = \begin{bmatrix} \mu_{i2} \\ \mu_{i3} \end{bmatrix}$

و برای حالت سه بعدی یعنی مجموعه ویژگی سوم، $\mu_i = \begin{bmatrix} \mu_{i4} \\ \mu_{i5} \\ \mu_{i6} \end{bmatrix}$ انتخاب شود که اعداد 2 تا 6 بیانگر مؤلفه‌ی چندم بردارهای میانگین است که در بالای صفحه نوشته شده است.

b

اعداد محاسبه‌شده برای هر کلاسیفایر :

اول :

1	0	0.1	5.6
2	0	5.5	3.7
3	5.7	0	0
4	0	0.01	4.8
5	3.2	0	0
6	0	7.3	3
7	1	0	0
8	0	0.9	3.6

دوم :

1	-74.7	-12.1	-246.9
2	-26.6	-11.64	-135.4
3	-71	-4	-207
4	-74	-12	-246
5	-82	-7	-252
6	-214	-62	-372
7	-94	-6	-268
8	-88	-6	-257

سوم :

1	-4.1	-3.85	-3.87
2	-4	-3.86	-3.89
3	-3	-4.14	-4.07
4	-4	-3.85	-3.87
5	-3.86	-4.07	-4.00
6	-4.1	-3.84	-3.85
7	-3.84	-4.05	-3.99
8	-4.1	-3.86	-3.88

b)

برچسب در دسته‌بندی یک بعدی	برچسب در دسته‌بندی دو بعدی	برچسب در دسته‌بندی سه بعدی
Stamen ✓	Stamen ✓	stamen ✓
Stigma ✓	stamen ✗	Stamen ✗
Petal ✓	stamen ✗	Petal ✓
Stamen ✗	Stamen ✗	stamen ✗
Petal ✓	stamen ✗	petal ✓
stamen ✓	stamen ✓	stamen ✓
Petal ✓	stamen ✗	petal ✓
Stamen ✓	stamen ✓	stamen ✓

c)

دسته‌بندی کننده‌ی اول

True \ Predict	Petal	stamen	stigma
Petal	3	0	0
stamen	0	3	0
stigma	0	1	1

دسته‌بندی کننده‌ی دوم

True \ Predict	ω_1	ω_2	ω_3
ω_1	0	3	0
ω_2	0	3	0
ω_3	0	2	0

دسته‌بندی کننده‌ی سوم

True \ Predict	ω_1	ω_2	ω_3
ω_1	3	0	0
ω_2	0	3	0
ω_3	0	2	0

d)

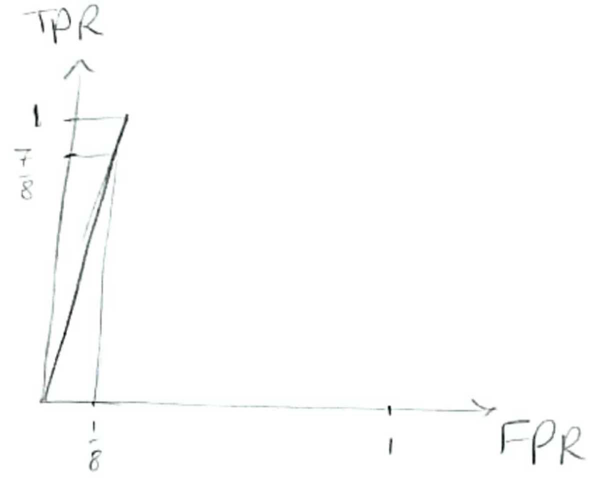
ارور برای اولی: $\frac{1}{8}$

ارور برای دومی: $\frac{5}{8}$

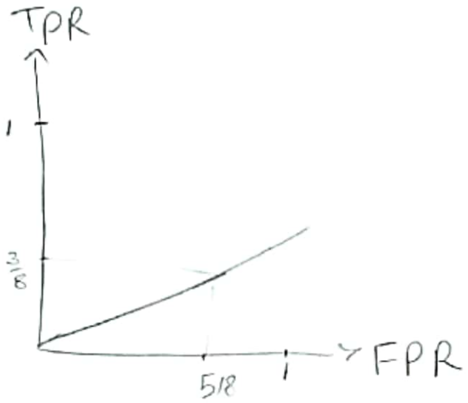
ارور برای سومی: $\frac{2}{8}$

e

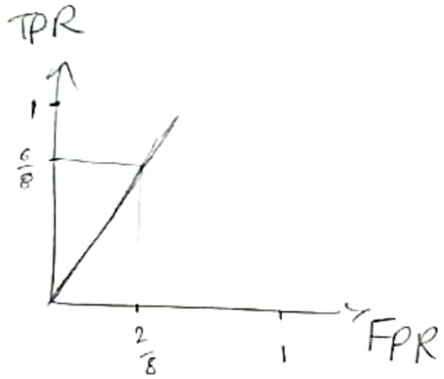
Roc برای اولی :



Roc برای دومی



Roc برای سومی :



f

در این سوال چون $\Sigma = I$ است ⇐ نتایج دو روش یکسان هستند. چون

در صورتی که $I = \Sigma$ باشد فاصله‌ی ماهالانوبیس به فاصله‌ی اقلیدسی تبدیل می‌شود.

عبارت زیر تابع تصمیم‌گیری را نشان می‌دهد :

$$g_i(\underline{x}) = \underline{x}^T \underline{\mu}_i - \frac{1}{2}\underline{\mu}_i^T \underline{\mu}_i$$

$$d_{ij} = g_i(x) - g_j(x)$$

سوال چهارم

a

$$g(\omega_1): \quad P(\omega_1)P(\underline{x}|\omega_1) > P(\omega_2)\,P(\underline{x}|\omega_2)$$

$$g(\omega_2). \quad P(\omega_2)P(x|\omega_2) > P(\omega_1)\,P(\underline{x}|\omega_1)$$

$$P(x|\omega_1) = \begin{cases} \frac{x}{8} & 0 \leqslant x \leqslant 1 \\ \frac{1}{8} & 1 \leqslant x \leqslant 8 \\ \frac{9}{8} - \frac{x}{8} & 8 \leqslant x \leqslant 9 \end{cases}$$

$$P(x|\omega_2) = \begin{cases} \frac{x}{2} - \frac{3}{2} & 3 \leqslant x \leqslant 4 \\ \frac{1}{2} & 4 \leqslant x \leqslant 5 \\ 3 - \frac{x}{2} & 5 \leqslant x \leqslant 6 \end{cases}$$

$$P(\omega_2) = \frac{10}{13} \quad , \quad P(\omega_1) = \frac{3}{13}$$

در روش دسته بندی بیز، ماکزیمم احتمال ها را بدست می آوریم. بنابراین برای یافتن ناحیه ی هر کلاس:

مجهول

$$\frac{10}{13}\underbrace{\left(\frac{x_1}{2} - \frac{3}{2}\right)} = \underbrace{\frac{1}{8}} \times \frac{3}{13} \Rightarrow x_1 = 3.075$$

احتمال کلاس دوم — تابع توزیع کلاس دوم — تابع توزیع کلاس اول

$$\frac{10}{13}\left(3 - \frac{x_2}{2}\right) = \frac{1}{8} \times \frac{3}{13} \Rightarrow x_2 = 5.925$$

$\Leftarrow$

if $\quad x \in [3.075, 5.925] \Rightarrow x \in \omega_2$

else if $\quad x \in [0, 3.075] \cup [5.925, 9] \Rightarrow x \in \omega_1$

b

$$P(\varepsilon) = P(\omega_1)\int_{3.075}^{5.925} \frac{1}{8}\,dx + P(\omega_2)\int_{3}^{3.075} \left(\frac{x}{2} - \frac{3}{2}\right)dx +$$

$$P(\omega_2)\int_{5.925}^{6} \left(3 - \frac{x}{2}\right)dx = \frac{3}{13}(0.35) + \frac{10}{13}(0.0014) +$$

$$\frac{10}{13}(0.0014) = \boxed{0.082}$$

c

$$\varepsilon_{UB} = \sqrt{P(\omega_1)P(\omega_2)} \int_{\infty}^{+\infty} \sqrt{p(x|\omega_1)\, p(x|\omega_2)}\, dx$$

$$= \sqrt{\frac{3}{13} \times \frac{10}{13}} \left[\int_3^4 \sqrt{\frac{1}{8}\left(\frac{x}{2} - \frac{3}{2}\right)}\, dx + \int_4^5 \sqrt{\frac{1}{8} \times \frac{1}{2}}\, dx + \int_5^6 \sqrt{\frac{1}{8}\left(3 - \frac{x}{2}\right)}\, dx \right] = \boxed{0.245}$$

d

$$\varepsilon_{CB} = P(\omega_1)^{s}\, P(\omega_2)^{1-s} \int_{-\infty}^{+\infty} p(x|\omega_1)^{s}\, p(x|\omega_2)^{1-s}\, dx$$

$$= \left(\frac{3}{13}\right)^{s} \left(\frac{10}{13}\right)^{1-s} \left[\int_3^4 \left(\frac{1}{8}\right)^{s} \left(\frac{x}{2} - \frac{3}{2}\right)^{1-s} dx + \int_4^5 \left(\frac{1}{8}\right)^{s} \left(\frac{1}{2}\right)^{1-s} dx + \int_5^6 \left(\frac{1}{8}\right)^{s} \left(3 - \frac{x}{2}\right)^{1-s} dx \right] =$$

$$\frac{3^{s} \times 10^{1-s}}{13} \left[\frac{1}{2^{1+2s}} \left[\frac{(x-3)^{2-s}}{2-s} \right]_3^4 + \frac{1}{2^{1+2s}} + \frac{1}{2^{1+2s}} \left[\frac{-(6-x)^{2-s}}{2-s} \right]_5^6 \right]$$

$$= \frac{3^{s} \times 10^{1-s}}{13 \times 2^{1+2s}} \left[\frac{1}{2-s} + 1 + \frac{1}{2-s} \right] = \frac{3^{s} \times 10^{1-s}}{13 \times 2^{2s+1}} \left[\frac{2}{2-s} + 1 \right]$$

$$= \left(\frac{10}{13}\right) \left(\frac{3}{10}\right)^{s} \left(\frac{1}{2}\right)^{1+2s} \left[1 + \frac{2}{2-s} \right]$$

$$P_e \leq \varepsilon_{CB}$$

e

استاد درس ندادن ـ گفتن امتیازی هستش !

گزارش سوال پنجم

در قسمت اول و در فایل p5a&b.py ، ابتدا تصویر آموزشی با استفاده از کتابخانه‌ی openCV خوانده شده و سپس بلوک های 8x8 جدا شده و عملیات DCT روی هر بلوک انجام و الگوی زیگزاگ اعمال شده و بردار مد نظر بدست آمده. سپس دومین بزرگترین عدد این بردار به عنوان نماینده‌ی بلوک در نظر گرفته شده است. همزمان تصویر mask نیز خوانده شده و در بلوک متناظر بررسی شده که تعداد پیکسل های سیاه بیشتر است یا سفید. متناسب با این موضوع به نماینده‌ی بلوک ، برچسب مسی یا زمین زده شده است. در آخر نیز هیستوگرام هرکدام رسم شده و داده های بدست آمده یعنی مسی|x و زمین|x در فایل ذخیره شده.

در قسمت سوم و در فایل p5c.py ، تصویر تست خوانده شده و مانند گام قبل نماینده‌ی هر بلوک استخراج شده و در فایل Test_x.txt ذخیره شده است.

در قسمت چهارم و در فایل p5d.py ، فایل های ذخیره شده در گام های قبل بارگذاری شده و سعی شده است که به نماینده های تصویر تست برچسب زده شود. طبق مشاهدات مقدار x برای مسی در تصویر آموزشی بین ۰ و 620 بود و برای زمین بین 1 تا 126.

توزیع xها به شرط هر کلاس توزیع نرمال نبود.

برای برچسب زدن به داده های تست، مقدار x در بازه های مختلف بررسی شد:

if $1 \leq x \leq 126 \Rightarrow$ p(x|مسی) p(مسی) $\boxed{?}$ p(x|زمین) p(زمین)

else $\Rightarrow$ $x \in$ "مسی"

همچنین این احتمال های اولیه از روی تصویر leo1_mask در گام اول یعنی در فایل p5a&b.py محاسبه و در یک فایل متنی ذخیره شد.

پس از آنکه برچسب زدن به داده ها تست انجام شد متناظر با برچسب تعاشیده، کل بلوک مقدار صفر یا 255 را گرفت و در نهایت تصویر با اسم result.png ذخیره شد.

در قسمت آخر در فایل p5e.py، تصویر 1002_mask بارگذاری شده با تصویر ذخیره شده در گام قبل به صورت پیکسل به پیکسل مقایسه ی عددی شد و خطا محاسبه شد. خطای بدست آمده در فاز تست برابر با 0.24 است.

سوال ششم

a

$$L(\theta) = \prod_{K=1}^{n} p(x_K|\theta) = \underline{\left(\frac{3\theta}{5}\right)^4 \left(\frac{2(1-\theta)}{5}\right)^2 \left(\frac{2\theta}{5}\right)^3 \left(\frac{3(1-\theta)}{5}\right)}$$

b

$$\ln L(\theta) = \sum_{K=1}^{n} \ln P(x_K|\theta) = 4\ln\left(\frac{3\theta}{5}\right) + 2\ln\left(\frac{2-2\theta}{5}\right) +$$

$$3\ln\left(\frac{2\theta}{5}\right) + \ln\left(\frac{3-3\theta}{5}\right)$$

c

$$\left[\ln L(\theta)\right]' = \frac{4}{\theta} + \frac{2}{1-\theta} + \frac{3}{\theta} - \frac{1}{1-\theta} =$$

$$\frac{7}{\theta} - \frac{3}{1-\theta} = \frac{7-7\theta-3\theta}{\theta(1-\theta)} = \frac{7-10\theta}{\theta-\theta^2} = 0 \Rightarrow$$

$$\boxed{\hat{\theta} = \frac{7}{10}}$$

d

$$\ln L(\theta) = \sum_{K=1}^{n} \left[\ln \frac{1}{(m-1)!} - m\ln\theta + (m-1)\ln x_K - \frac{x_K}{\theta}\right]$$

$$\Rightarrow \frac{\partial\left[\ln L(\theta)\right]}{\partial\theta} = \sum_{K=1}^{n}\left[-\frac{m}{\theta} + \frac{x_K}{\theta^2}\right] = 0$$

$$\Rightarrow \frac{-nm}{\theta} + \frac{\sum_{K=1}^{n} x_K}{\theta^2} = \frac{\sum_{K=1}^{n} x_K - nm\theta}{\theta^2} = 0$$

$$\Rightarrow \boxed{\hat{\theta} = \frac{1}{nm}\sum_{K=1}^{n} x_K}$$

e

$$\ln L(\theta) = \sum_{K=1}^{n} \ln P(y_K|\theta) = \sum_{K=1}^{n}\left[\ln\left(e^{-\theta}\theta^{y_K}\right) - \ln y_K!\right]$$

$$= \sum\left[-\theta + y_K\ln\theta - \ln y_K!\right]$$

$$\Rightarrow \frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{k=1}^{n}\left(-1+\frac{y_k}{\theta}\right) = -n + \frac{\sum y_k}{\theta} = 0$$

$$\Rightarrow \boxed{\hat{\theta} = \frac{1}{n}\sum_{k=1}^{n} y_k} \rightarrow = \mu$$

f $E(\hat{\theta}) - \theta$?

$$E(\hat{\theta}) = E\left\{\frac{1}{n}\sum_{k=1}^{n} y_k\right\} = \frac{1}{n} E\left\{\sum_{k=1}^{n} y_k\right\} =$$

$$\frac{1}{n}\sum_{k=1}^{n} \underbrace{E[y_k]}_{=\mu} = \frac{1}{n}\sum \mu = \frac{1}{n} \times n\mu = \boxed{\mu}$$

$$E(\hat{\theta}) = \mu$$

میدانیم که تابع $f(y_i, \theta)$ تابع پواسون است، و پارامتر مجهول، میانگین توزیع داده هاست. یعنی $\mu = \theta$، از روابط بالا داریم که $E(\hat{\theta}) = \mu$ ⇐ با تخمینی که زده ایم میزان خطا کمترین حالت است (منظور از خطا فاصله‌ی بین داده های واقعی و توزیع تخمین زده شده است) ⇐ $E(\hat{\theta}) - \theta = 0$ ⇐ این تخمین unbiased است.

$$\mathrm{Var}\{\hat{\theta}\} = E\{\hat{\theta}^2\} - \overbrace{(E\{\hat{\theta}\})^2}^{=\mu^2} \Rightarrow$$

$$E\{\hat{\theta}^2\} = E\left\{\left(\frac{1}{n}\sum_{k=1}^{n} y_k\right)\left(\frac{1}{n}\sum_{j=1}^{n} y_j\right)\right\} =$$

$$E\left\{\frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n} y_k y_j\right\} = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n} E\{y_k y_j\} =$$

$$\frac{1}{n^2}\left[(n^2-n)\mu^2 + n\mu^2 + n\sigma^2\right] = \mu^2 + \frac{\sigma^2}{n}$$

$$\Rightarrow \mathrm{Var}\{\hat{\theta}\} = \cancel{\mu^2} + \frac{\sigma^2}{n} - \cancel{\mu^2} = \boxed{\frac{\sigma^2}{n}}$$

$$\Rightarrow \sigma^2 = \boxed{n\,\mathrm{Var}\{\hat{\theta}\}}$$

$$P(\hat{\theta} \mid Y) \,?$$

$$P(\theta) = \frac{1}{\Gamma(\alpha)\,\beta^{\alpha}}\, \theta^{\alpha-1} e^{-\frac{\theta}{\beta}}$$

$$P(\theta \mid Y) = \frac{P(Y\mid\theta)\,P(\theta)}{P(Y)} \Rightarrow$$

$$\hat{\theta}_{Map} \Rightarrow \frac{\partial}{\partial\theta} \ln\left(\prod_{k=1}^{n} P(y_k \mid \theta)\, P(\theta)\right) = 0$$

$$\Rightarrow \sum_{k=1}^{n}\left(-1 + \frac{y_k}{\theta}\right) + \frac{\alpha-1}{\theta} - \frac{1}{\beta} = 0$$

$$\Rightarrow \frac{\sum y_k}{\theta} + \frac{\alpha-1}{\theta} = n - \frac{1}{\beta} \Rightarrow$$

$$\frac{1}{\theta}\left(\sum y_k + \alpha - 1\right) = n - \frac{1}{\beta} \Rightarrow \boxed{\hat{\theta}_{MAP} = \frac{\sum_{k=1}^{n} y_k + \alpha - 1}{n - \frac{1}{\beta}}}$$

$$\hat{\theta}_{MAP} = \hat{\theta}_{ML} \Leftarrow \beta, \alpha = 1 \text{ اگر}$$

$$E[\hat{\theta}] = \frac{1}{n - \frac{1}{\beta}} + n\alpha$$

h

استاد این مطلب رو درس نداده بودن ولی چیزی که از ویکی‌پدیا خوندم :

$$\text{if}\left(L_\theta(y) = h(y)\, g_\theta(T(y))\right) \Rightarrow \text{sufficient.}$$

$$L_\theta(y) = \prod_{k=1}^{n} \frac{e^{-\theta}\, \theta^{y_k}}{y_k!} \quad (\text{طبق سمت } e)$$

حال اگه $h(y)$ را بصورت زیر تعریف کنم :

$$h(y) = \prod_{k=1}^{n} \frac{1}{y_k!} \Rightarrow$$

$$L_\theta(y) = h(y)\, \underbrace{e^{-n\theta}\, \theta^{\overbrace{\sum_{i=1}^{n} y_i}^{T(y)}}}_{g_\theta(T(y))} = h(y)\, g_\theta(T(y))$$

پس $T(y)$ ، برای θ sufficient statistic هست.

سوال هفتم

a

n تا تانک داریم، K تا از این n تا مشاهده کرده‌ایم یعنی: $\binom{n}{K}$ حالت
برای مشاهده‌ی K تا تانک وجود دارد.
حال اگر بزرگترین سریال تانک مشاهده شده m باشد یعنی K-1 تا تانک
دیگر اعدادی بین 1 تا m-1 دارند $\Leftarrow \binom{m-1}{K-1}$

پس با توجه به این مطالب داریم:

$$P(m,K\mid n)=\frac{\binom{m-1}{K-1}}{\binom{n}{K}}$$

b

$$P(n\mid m,K)=\frac{P(m,K\mid n)\,P(n)}{P(m,K)}$$

$$P(m,K)=\sum_{m<n<\Omega}P(m,K\mid n)\,P(n)=\sum_{j=m}^{\Omega}\frac{\binom{m-1}{K-1}}{\binom{j}{K}}\,P(j)$$

$$=\frac{1}{\Omega}\binom{m-1}{K-1}\sum_{j=m}^{\Omega-1}\frac{1}{\binom{j}{K}}=\frac{1}{\Omega}\binom{m-1}{K-1}\left(\frac{1}{\binom{m-1}{K-1}}-\frac{1}{\binom{\Omega-1}{K-1}}\right)\left(\frac{K}{K-1}\right)$$

$$\Rightarrow P(n\mid m,K)=\frac{\dfrac{\cancel{\binom{m-1}{K-1}}}{\binom{n}{K}}\times\cancel{\dfrac{1}{\Omega}}}{\cancel{\dfrac{1}{\Omega}}\binom{m-1}{K-1}\left(\dfrac{K}{K-1}\right)\left(\dfrac{1}{\binom{m-1}{K-1}}-\dfrac{1}{\binom{\Omega-1}{K-1}}\right)}=$$

$$\frac{1}{\frac{K}{K-1}\times\binom{n}{K}\left(\frac{1}{\binom{m-1}{K-1}}-\frac{1}{\binom{\Omega-1}{K-1}}\right)}$$

در صورت سؤال گفته شده $\Omega \to \infty \Leftarrow \frac{1}{\binom{\Omega-1}{K-1}} \to 0 \Leftarrow$

$$\underset{n\to\infty}{P(n\mid m,K)} = \frac{\binom{m-1}{K-1}}{\frac{K}{K-1}\times\binom{n}{K}} = \frac{m-1}{n}\times\frac{\binom{m-2}{K-2}}{\binom{n-1}{K-1}}$$

$$\frac{K}{K-1}\times\binom{n}{K} = \binom{n}{K-1}$$

---

c

$$E\left[P(n\mid m,k)\right] = \sum_{m\leq n<\infty} n\,P(n\mid m,k) =$$

$$(m-1)\times\cancel{\binom{m-2}{K-2}}\sum_{m\leq n<\infty}\frac{1}{\binom{n-1}{K-1}} = \boxed{\frac{K-1}{K-2}\times\frac{(m-1)}{1}}$$

$$= \frac{K-1}{K-2}\times\frac{1}{\cancel{\binom{m-2}{K-2}}}$$

---

d

باید $K \geq 3$ باشد چون اگر کمتر باشد کسر بودن داده در قسمت قبل درست در نمی‌آید.

سوال پنجم

a

در شرایطی که تابع توزیع $p(x|w_i)$ ها تابعی براساس فاصله‌ی اقلیدسی نقطه x از مرکز داده‌ها باشد و احتمال $p(w_i)$ ها برابر باشد؛

به طور خاص مثلاً اگر $p(x|w_i) = N(\mu, I)$ باشد، فاصله‌ی ماهالانوبیس در تابع توزیع معادل همان تابع اقلیدسی می‌شود $\Leftarrow$ نتیجه‌ی هر دو روش یکی می‌شود.

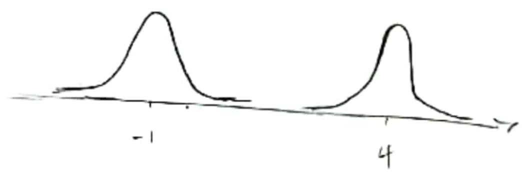
b

MEC: مرحله‌ی آموزش دارد و در آن احتمال $p(x|w_i)$ ها محاسبه می‌شود.

MDC: مرحله‌ی آموزش دارد و در آن میانگین هر کلاس محاسبه می‌شود.

در هر دو مورد در مرحله‌ی تست از تمام داده‌های آموزش استفاده نمی‌شود و فقط از پارامترهای بدست آمده استفاده می‌شود $\Leftarrow$ هیچکدام lazy نیستند.

c



همان‌طور که در شکل بالا مشخص است تداخل و اشتراک این دو کلاس خیلی کم است $\Leftarrow$ در این حالت استفاده از بیز بهترین جواب با کمترین خطا را می‌دهد.

d

بله ممکن است. دو به دو کلاس‌ها را در نظر میگیریم و یکی را هدف و دیگری را غیر هدف می‌نامیم، بعد در هر کدام از این دوتایی‌ها FPR و TPR را محاسبه می‌کنیم و با هم جمع میکنیم و از روی FPR و TPR کلی، نمودار ROC را رسم می‌کنیم.

e

خیر. در مسائل رگرسیون، تخمین یک عدد را انجام می‌دهیم نه کلاس یک داده‌ی تست.
به همین دلیل نمی‌توانیم از بیز به عنوان تخمین زننده‌ی یک عدد استفاده کنیم.

g

اگر مسأله فضای صفر و یک داشته باشد (یعنی $C_{11}=C_{22}=0$, $C_{12}=C_{21}=1$)

در این حالت با داشتن احتمال اولیه‌ی دقیق استفاده از MAP بهتر است ولی اگر فضا صفر و یک نبود، MLE جواب بهتری می‌دهد یعنی خطای کمتری دارد.

h

در حالتی که توزیع ها منطبق بر هم باشند، بیز جواب می‌دهد. ولی در غیر اینصورت، روش بیز همیشه بیشترین احتمال ثانویه را به عنوان جواب انتخاب می‌کند ← همیشه جواب اش یکی است و منحصربفرد. اما اگر احتمال ثانویه‌ها متفاوت نباشد و یکسان باشد یعنی همان توزیع‌ها منطبق باشند بیز جواب منحصربفرد و خوبی نمی‌دهد.

i

در رابطه‌ای که می‌خواهیم با روش MLE، ماکزیمم کنیم یک جمله‌ی $P(\theta)$ که تابع پنالتی نامیده می‌شود، اضافه می‌کنیم و در واقع با این کار از overfit شدن جلوگیری می‌کنیم. رابطه‌ی حاصل به صورت زیر است:

$$\hat{\theta}_{PMLE} = \arg\max \log L(\theta ; x_1, x_2 \dots x_n) + P(\theta)$$