به نام خدا

تمرین درس شناسایی آماری الگو -سری دوم

فردین آیار

شماره دانشجویی: ۹۹۱۳۱۰۴۰

استاد: دکتر رحمتی

دانشکده کامپیوتر- زمستان۹۹

فایل های مربوط به تمریناتی که نیاز به پیاده سازی دارند، در پوشه هایی با شماره تمرین قراردارند.

۱)

---

a) درصد های مربوط به FP rate و FN rate در صورت سوال جابه‌جا داده شده اند که در اینجا اصلاح شده‌است.

$$\begin{cases} p(positive|covid) = 0.83 \\ p(negative|covid) = 0.17 \\ p(positive|notCovid) = \ 0.11 \\ p(negative|notCovid) = 0.89 \\ p(covid) = 0.004 \end{cases} \Rightarrow p(covid|positive)$$

$$= \frac{p(positive|covid) \times p(covid)}{p(positive)}$$

$$= \frac{p(positive|covid) \times p(covid)}{p(positive|covid) \times p(covid) + p(positive|notCovid) \times p(notCovid)}$$

$$= \frac{0.83 \times 0.004}{0.83 \times 0.004 + 0.11 \times 0.996} = 0.029$$

b)

$$\begin{cases} p(positive|covid) = 0.73 \\ p(negative|covid) = 0.27 \\ p(positive|notCovid) = \ 0.19 \\ p(negative|notCovid) = 0.81 \\ p(covid) = 0.001 \end{cases} \Rightarrow p(covid|positive)$$

$$= \frac{p(positive|covid) \times p(covid)}{p(positive)}$$

$$= \frac{p(positive|covid) \times p(covid)}{p(positive|covid) \times p(covid) + p(positive|notCovid) \times p(notCovid)}$$

$$= \frac{0.73 \times 0.001}{0.73 \times 0.001 + 0.19 \times 0.999} = 0.004$$

c)

$$\frac{0.73 \times 0.05}{0.73 \times 0.05 + 0.19 \times 0.95} = 0.168$$

d) در صورت سوال ۰.۰۲۵ آمار مربوط به اشخاصی است که تست کرونای آنها مثبت شده‌است. به نظر می‌رسد منظور اشخاصی است که واقعا به کرونا مبتلا هستند، بنابراین این مورد را اصلاح می‌کنیم.

$$\begin{cases} p(positive|covid) = 0.76 \\ p(negative|covid) = 0.24 \\ p(positive|notCovid) = \ 0.32 \Rightarrow p(covid|negative) \\ p(negative|notCovid) = 0.68 \\ p(covid) = 0.025 \end{cases}$$

$$= \frac{p(negative|covid) \times p(covid)}{p(negative)}$$

$$= \frac{p(negative|covid) \times p(covid)}{p(negative|covid) \times p(covid) + p(negative|notCovid) \times p(notCovid)}$$

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

e)

$$p(covid|positive) =$$

$$= \frac{p(positive|covid) \times p(covid)}{p(positive|covid) \times p(covid) + p(positive|notCovid) \times p(notCovid)}$$

$$= \frac{0.76 \times 0.025}{0.76 \times 0.025 + 0.32 \times 0.975} = 0.057$$

f)

$$p(notCovid|positive) = 1 - 0.057 > p(covid|positive)$$

g)

$$p(notCovid|negative) = 1 - 0.009 > p(covid|negative)$$

h) با توجه به فرمولهای ارائه شده در بخش‌های قبلی:

$$p(covid|positive) = \frac{p \times 0.01}{p \times 0.01 + (1 - P) \times 0.99} = \frac{0.01p}{0.99 - 0.98p}$$

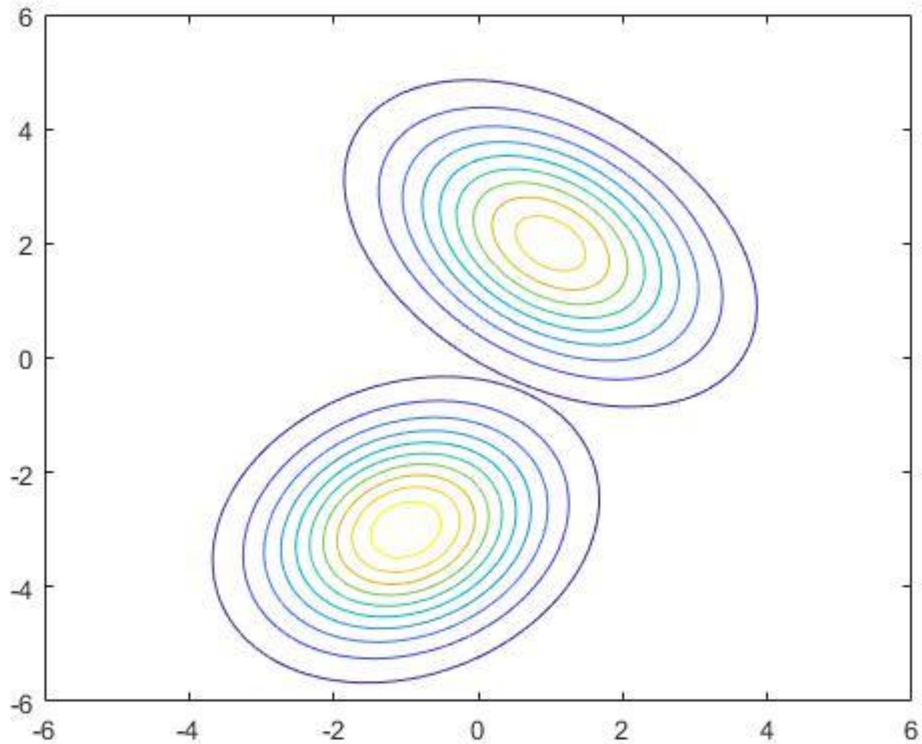
i) با توجه به اهمیت نتیجه تست بیماری کرونا، این مقدار را ۰.۵ در نظر میگیریم.

j)

$$\frac{0.01p}{0.99 - 0.98p} = 0.5 \Rightarrow p = 0.99$$

۲)

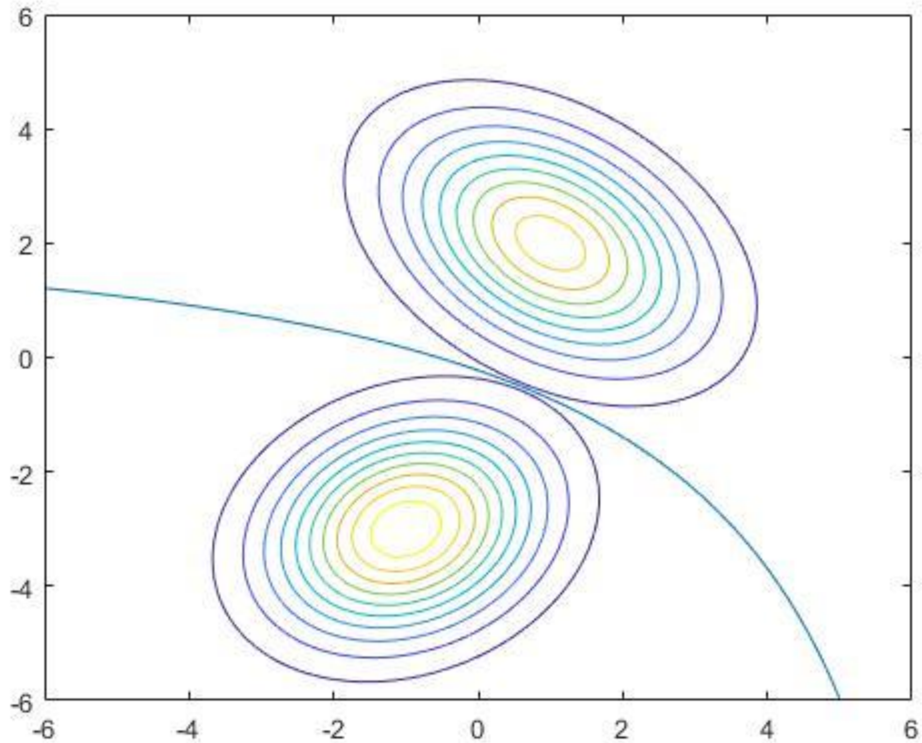
a)کد مربوط به این قسمت در اسکریپت a.m قراردارد. کانتور لاین‌های مربوط به دو کلاس در شکل زیر ارائه شده است.



b) کد مربوط به این قسمت در اسکریپت b.m قرار دارد. شکل کلی توابع discriminant برای حالت عمومی(ماتریس‌های کوواریانس متفاوت و غیر قطری) به صورت زیر است.

$$\begin{cases} g_i(x) = x^t W_i x + w_i^t x + w_{i0} \\ W_i = -\frac{1}{2}\Sigma_i^{-1} \\ w_i = \Sigma_i^{-1}\mu_i \\ w_{i0} = -\frac{1}{2}\mu_i^t \Sigma_i^{-1}\mu_i - \frac{1}{2}\ln|\Sigma_i| + \ln p(\omega_i) \end{cases}$$

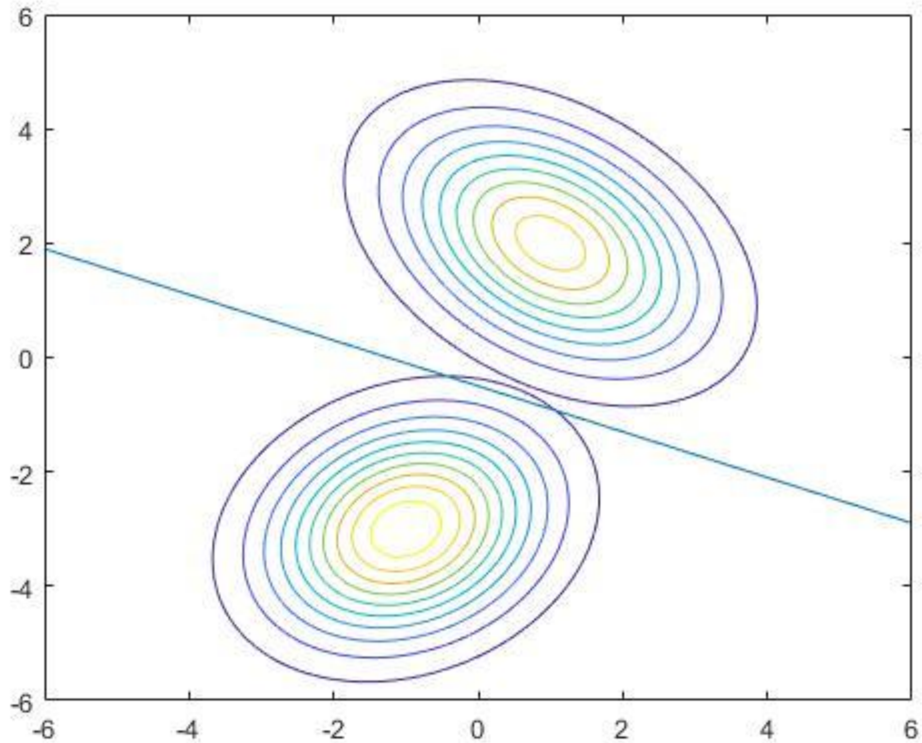
با مساوی قراردادن توابع فوق برای کلاس‌ها، شکل تابع جداکننده به صورت زیر به دست می‌آید.



تابع جداکننده برای MDC به صورت زیر می‌باشد:

$$d(x) = (\mu_1 - \mu_2)^t x - \frac{1}{2}(\mu_1^t \mu_1 - \mu_1^t \mu_1)$$

شکل تابع فوق به این صورت می‌باشد:



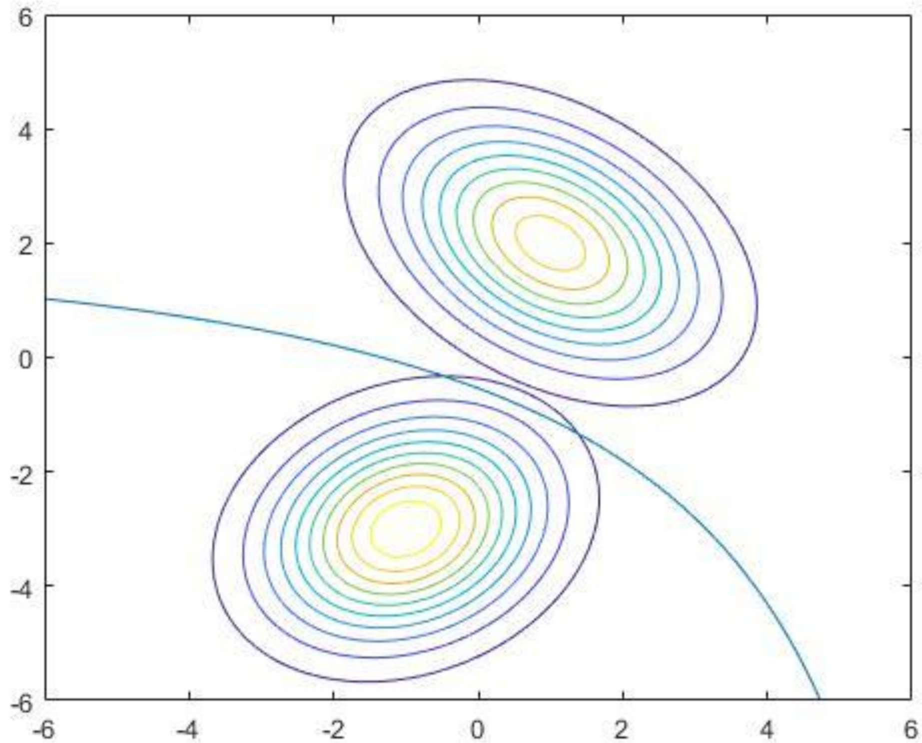
c) کد مربوط به این قسمت در فایل c.m قرار دارد. برچسب داده های کلاس ۱، برابر با ۱ و برچسب داده‌های کلاس ۲ برابر با صفر فرض شده است. برچسب نمونه ها با استفاده از توابع جداکننده پیش‌بینی شده و سپس مقدار خطا محاسبه شده است. خطای به دست آمده برای دسته‌بندی‌کننده بیز ۰.۰۱۲۵ و برای MDC ۰.۰۱۳ می باشد. با توجه به تصادفی بودن نمونه ها، امکان تغییر در این خطا وجود دارد.

d) کد مربوط به این قسمت در فایل d.m قرار دارد. با درنظر گرفتن هزینه متفاوت برای کلاس های مختلف و فرض مسئله دو کلاسه، ضریب w_{i0} از توابعdiscriminant که در بخش b ارائه شد، به شکل زیر خواهد بود.

$$w_{10} = -\frac{1}{2}\mu_i^t\Sigma_i^{-1}\mu_i - \frac{1}{2}\ln|\Sigma_i| + \ln p(\omega_i) + \ln \lambda_{12}$$

$$w_{20} = -\frac{1}{2}\mu_i^t\Sigma_i^{-1}\mu_i - \frac{1}{2}\ln|\Sigma_i| + \ln p(\omega_i) + \ln \lambda_{21}$$

سایر پارامترها بدون تغییر باقی می‌مانند. با اعمال این تغییر، تابع جداکننده جدید را رسم می‌کنیم.



با توجه به اینکه کلاس ها نسبتا از هم جدا هستند، اضافه کردن ماتریس هزینه باعث تغییر زیادی در شکل نهایی تابع جداکننده نشده است. برای محاسبه f-score خروجی هر دو دسته‌بندی‌کننده در فایل csv ذخیره شده و با استفاده از فایل f -score.py مقدار f-score آنها حساب شده است. خروجی به این صورت می باشد:

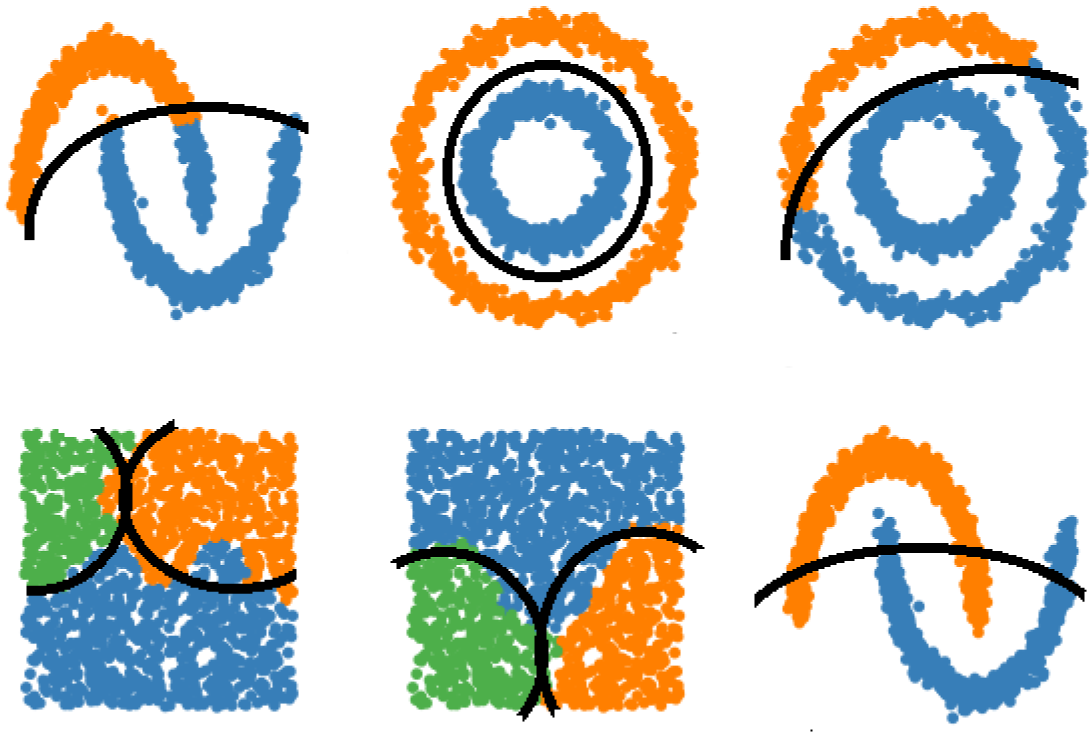
f-score با هزینه یکسان: ۰.۹۸۱۲

f-score با هزینه متفاوت: ۰.۹۸۵۵

همانطور که مشاهده می‌شود تفاوت زیادی در خروجی وجود ندارد. دلیل این امر همانطور که قبلا گفته شد، این است که کلاس ها به خوبی از هم جدا شده‌اند.

e) کد مربوط به این بخش در فایل e.m قرار دارد. پس از ۲۰ بار اجرا، میانگین خطای بیز ۰.۰۰۹۵ و MDC ۰.۰۱ بدست آمد.

f) با توجه به فرضیات سوال، شکل مرزهای تصمیم غیر خطی خواهد بود. با توجه به معلوم نبودن ماتریس‌های کوواریانس، به صورت تقریبی مرزها را رسم می‌کنیم:



۳)

---

a)

Feature Set 1:

$$\begin{cases} \mu_{stamen} = 56.86 \\ \mu_{stigma} = 46.08 \\ \mu_{petal} = -48.75 \end{cases}$$

با توجه به یکسان بودن احتمالات اولیه و همانی بودن کوواریانس(واریانس) داریم:

$$p(\omega_i|x) \sim p(x|\omega_i) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-\mu_i)^2}{2}}$$

$$g_i(x) = \ln\big(p(x|\omega_i)\big) = -\frac{1}{2} ln2\pi - \frac{x^2 - 2x\mu_i + \mu_i^2}{2} \Rightarrow g_i(x)$$

$$= \mu_i x - \frac{\mu_i^2}{2} \quad (linear)$$

Feature Set 2:

$$\begin{cases} \mu_{stamen} = \begin{bmatrix} 2.6 \\ 6.4 \end{bmatrix} \\ \mu_{stigma} = \begin{bmatrix} 22 \\ 1 \end{bmatrix} \\ \mu_{petal} = \begin{bmatrix} 11.4 \\ 0 \end{bmatrix} \end{cases}$$

$$p(\omega_i|x) \sim p(x|\omega_i) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-\mu_i)^t(x-\mu_i)}{2}}$$

$$g_i(x) = \ln\big(p(x|\omega_i)\big) = -\frac{1}{2} ln2\pi - \frac{x^t x - 2\mu_i^t x + \mu_i^t \mu_i}{2} \Rightarrow g_i(x)$$
$$= \mu_i^t x - \frac{\mu_i^t \mu_i}{2} \quad (linear)$$

Feature Set 3:

$$\begin{cases} \mu_{stamen} = \begin{bmatrix} 0.09 \\ 0.81 \\ 0.808 \end{bmatrix} \\ \mu_{stigma} = \begin{bmatrix} 0.20 \\ 0.82 \\ 0.89 \end{bmatrix} \\ \mu_{petal} = \begin{bmatrix} 0.74 \\ 0.62 \\ 0.69 \end{bmatrix} \end{cases}$$

$$p(\omega_i|x) \sim p(x|\omega_i) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-\mu_i)^t(x-\mu_i)}{2}}$$

$$g_i(x) = \ln\big(p(x|\omega_i)\big) = -\frac{1}{2} ln2\pi - \frac{x^t x - 2\mu_i^t x + \mu_i^t \mu_i}{2} \Rightarrow g_i(x)$$
$$= \mu_i^t x - \frac{\mu_i^t \mu_i}{2} \quad (linear)$$

b) با استفاده از توابع بدست آمده در قسمت قبلی داریم: (محاسبات در فایل اکسل Book3 موجود می‌باشد.)

#	Pred. FS1	Pred. FS2	Pred. FS3	True Label
1	Stamen	Stamen	Stamen	Stamen

2	Stigma	Stamen	Stamen	Stigma
3	Petal	Stamen	Petal	Petal
4	Stamen	Stamen	Stamen	Stigma
5	Petal	Stamen	Petal	Petal
6	Stamen	Stamen	Stamen	Stamen
7	Petal	Stamen	Petal	Petal
8	Stamen	Stamen	Stamen	Stamen

c) با توجه به نتایج قسمت قبل داریم:

Confusion Matrix for feature set 1				
		Actual		
		Stamen	Stigma	Petal
Predicted	Stamen	3	1	0
	Stigma	0	1	0
	Petal	0	0	3
Confusion Matrix for feature set 2				
		Actual		
		Stamen	Stigma	Petal
Predicted	Stamen	3	2	3
	Stigma	0	0	0
	Petal	0	0	0

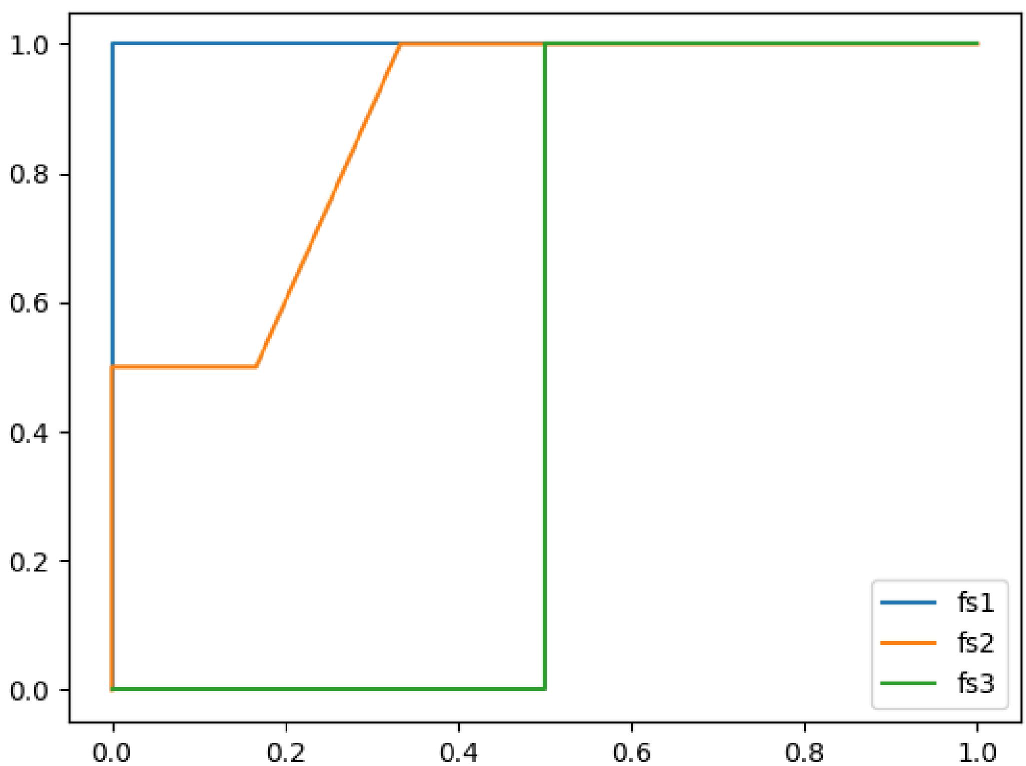
Confusion Matrix for feature set 3				
		Actual		
		Stamen	Stigma	Petal
Predicted	Stamen	3	2	0
	Stigma	0	0	0
	Petal	0	0	3

d) خطا برای هر دسته‌بندی‌کننده به شرح زیر است.

$$\begin{cases} E_{FS1} = \dfrac{number\ of\ misclassified\ samples}{total\ numebr\ of\ sampels} = \dfrac{1}{8} \\ E_{FS2} = \dfrac{5}{8} \\ E_{FS3} = \dfrac{2}{8} \end{cases}$$

e) برای رسم نمودار roc در مسائل چند کلاس رویکردهای متفاوتی وجود دارد. در اینجا با توجه به اینکه قسمت باارزش زعفران stigma است، آن را کلاس مثبت و دو دسته دیگر را کلاس منفی می‌نامیم. در واقع دو دسته stamen و petal را با هم ترکیب کرده و خروجی g متناظر با دسته جدید، همان ماکسیمم g آن ها به ازای هر نمونه است. این داده ها را به صورت csv ذخیره کرده و با استفاده از فایل roc.py و با توجه به برچسب صحیح،roc را رسم می‌کنیم.

همانطور که مشاهده می‌شود به علت کم بودن داده ها(به خصوص داده های stigma) شکل نمودارها به صورت خطهای شکسته می‌باشد. مطابق این شکل، دسته‌بندی‌کننده ۱ بهترین عملکرد را دارد.



f) با توجه به اینکه کوواریانس داده ها همانی است، روش بیز و MDC یکسان هستند. بنابراین نتایج مراحل قبل برای MDC برقرار است. شکل کلی توابع discriminant در دسته بندی‌کننده MDC به صورت زیر است:

$$g_i(x) = -\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu) = -\frac{x^t x - 2\mu_i^t x + \mu_i^t \mu_i}{2} \Rightarrow g_i(x) = \mu_i^t x - \frac{\mu_i^t \mu_i}{2}$$

که دقیقا برابر با توابع جداکننده قبلی است. بنابراین تمامی نتایج فوق بدون تغییر تکرار می‌شود.

٤)

a) با توجه به اینکه $\int_{-\infty}^{+\infty} P(x|w_1)dx=1$ ، مساحت زیر هر دو نمودار را برابر یک قرار می‌دهیم تا ارتفاع نمودارها مشخص شود.

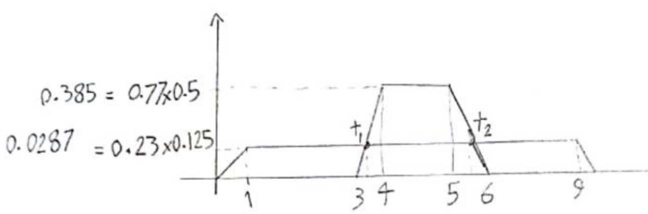
$$w_1 : (3+1)\times\frac{c_1}{2}=1 \longrightarrow c_1=0.5$$

$$w_0 : (9+7)\times\frac{c_2}{2}=1 \longrightarrow c_2=0.125$$

حال نمودار $P(w_i)P(x|w_i)$ را رسم می‌کنیم.

$$P(w_0)=0.3P(w_1)$$

$$\rightarrow \begin{cases} P(w_0)=0.23 \\ P(w_1)=0.77 \end{cases}$$



$$z\times 0.385=0.0287\rightarrow z\simeq 0.074 \begin{cases} t_1=3+0.074=3.074 \\ t_2=6-0.074=5.926 \end{cases}$$

$$x\in\begin{cases} w_1 & 3.074<x<5.926 \\ w_0 & \text{o.w} \end{cases}$$

b)

$$\mathcal{E}_1=\int_{R_2}P(x|w_1)\overset{\text{تقارن}}{=}2\int_{3}^{3.074}P(x|w_1) = 2\times 0.074\times 0.125/2=0.0287$$

$$\mathcal{E}_0=\int_{R_1}P(x|w_0)=\int_{3.074}^{5.926}P(x|w_0)=(5.926-3.074)\times 0.125=0.3565$$

$$P(error) = P(w_1)\mathcal{E}_1+P(w_0)\mathcal{E}_0=0.104$$

(c

$$\varepsilon_u = \sqrt{P(\omega_1)\ P(\omega_0)} \int \sqrt{P(x|\omega_1)P(x|\omega_0)}\ dx$$

$$= \sqrt{0.23 \times 0.77} \left[2 \int_3^4 \sqrt{0.5(x-3) \times 0.125}\, dx + \int_4^5 \sqrt{0.5 \times 0.125}\ dx \right]$$

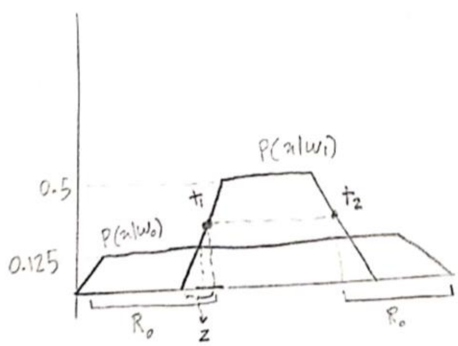
$$= \sqrt{0.177} \left[0.5 \int_3^4 \sqrt{x-3}\ dx + 0.25 \right]$$

$$= \sqrt{0.177} \left[0.5 \times \frac{2}{3} \left((4-3)^{3/2} - 0^{3/2} \right) + 0.25 \right] = 0.245$$

(d

$$\varepsilon_u = 0.23^s \times 0.77^{1-s} \left[2 \int_3^4 (0.5(x-3))^{1-s} \times 0.125^s\ dx + \int_4^5 0.5^{1-s} \times 0.125^s\ dx \right]$$

(e ε_1 را برابر 0.05 قرار می‌دهیم تا از این طریق خطای ε_0 مینیمم شود.



$$\varepsilon_1 = 2 \int_3^{3+Z} P(x|\omega_1)\, dx \longrightarrow \text{مساحت مثلث}$$

$$= 2 \times Z \times 0.5 \times Z/2 = 0.05 \rightarrow Z = 0.31$$

$$x \in \begin{cases} \omega_1 & 3.31 < x < 5.69 \\ \omega_0 & \text{O.W} \end{cases}$$

$$\varepsilon_0 = (5.69 - 3.31) \times 0.125 = 0.285$$

$$P(error) = 0.23 \times 0.285 + 0.05 \times 0.77 = 0.104$$

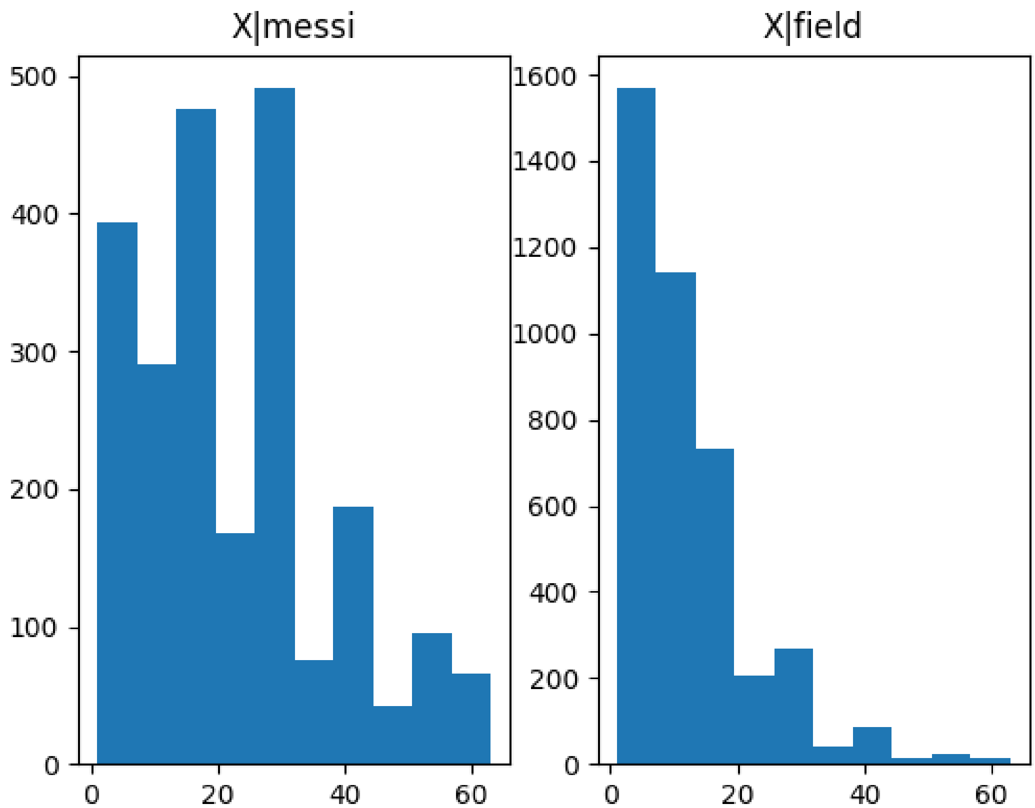
۵) کد مربوط به این سوال در فایل 2.py قراردارد.

a) پس از بارگزاری تصویر ماسک و پردازش آن، احتمالات اولیه به شرح زیر است:

p(messi)=0.358

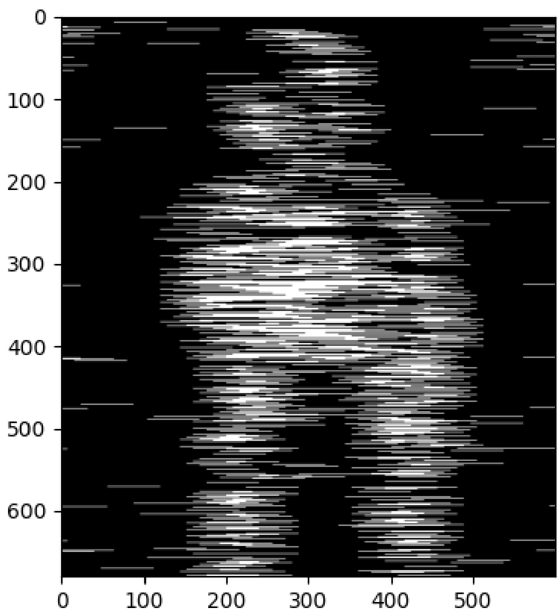
p(field)=0.642

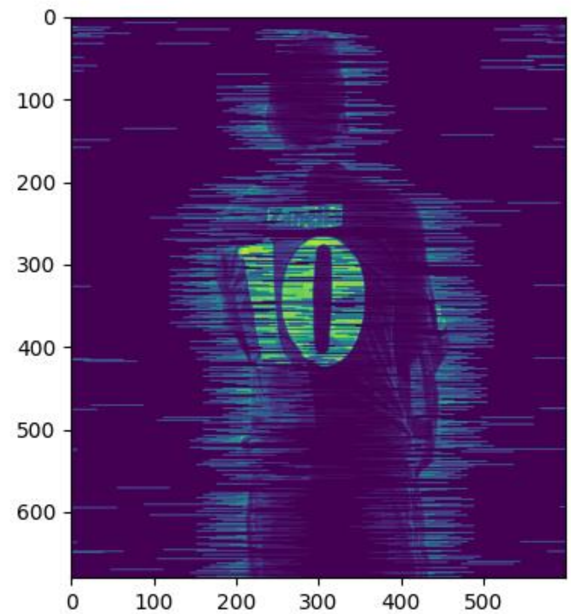
b) به کمک تصویر آموزش و تصویر ماسک، هیستوگرام ها رسم می‌شوند:



c) عملیات فوق روی تصویر آزمون اجرا شده و نتایج در متغیر t2_second_max ذخیره شده‌است.

d)نتایج به صورت زیر است:





همانطور که در تصاویر مشاهده می‌شود، خروجی به صورت بلاک نیست. دلیل این موضوع این است که تابع blockshaped که برای ساخت بلاک‌ها استفاده می شد، بلاک‌های مربعی ایجاد نکرده است و متاسفانه زمان کافی برای اصلاح آن وجود نداشت.

e) با استفاده از تصویر ماسک، میزان خطا ۰.۲۵۷ محاسبه شد که نشان می‌دهد علارغم اینکه شکل بلاک‌ها اشتباه است؛نتایج تا حدی قابل قبول می‌باشد.

٦)

---

a)

$$L(\theta) = \frac{3\theta}{5} \times \frac{2(1-\theta)}{5} \times \frac{3\theta}{5} \times \frac{2\theta}{5} \times \frac{3(1-\theta)}{5} \times \frac{2(1-\theta)}{5} \times \frac{2\theta}{5} \times \frac{3\theta}{5} \times \frac{3\theta}{5} \times \frac{2\theta}{5}$$
$$= \frac{3^5 \times 2^5 (1-\theta)^3 \theta^7}{5^{10}}$$

b)

$$\ln\big(L(\theta)\big) = 5\ln 3 + 5\ln 2 + 3\ln(1-\theta) + 7\ln\theta - 10\ln 5$$

c)

$$\frac{\mathrm{d}ln(L(\theta))}{\mathrm{d\theta}} = \frac{-3}{1-\theta} + \frac{7}{\theta} = 0 \Rightarrow \theta = 0.7$$

d)

$$\ln\big(L(\theta)\big) = \sum -\ln\frac{1}{(m-1)!} - m\ln\theta + (m-1)\ln x - \frac{x}{\theta}$$

$$\frac{\mathrm{d}ln(L(\theta))}{\mathrm{d\theta}} = \sum -\frac{m}{\theta} + \frac{x}{\theta^2} = \frac{-nm}{\theta} + \frac{\sum x}{\theta^2} = 0 \Rightarrow \frac{\sum x}{\theta^2} = \frac{nm}{\theta} \Rightarrow \theta = \frac{\sum x}{nm}$$

e)

$$\ln\big(L(\theta)\big) = \sum -\theta + y\ln\theta - \ln y!$$

$$\frac{\mathrm{d}ln(L(\theta))}{\mathrm{d\theta}} = \sum -1 + \frac{y}{\theta} = 0 \Rightarrow -n + \frac{\sum y}{\theta} = 0 \Rightarrow \theta_{ml} = \frac{\sum y}{n}$$

f) توزیع فوق Poisson distribution نام دارد و مقدار E آن همان θ می باشد. بنابراین:

$$E[\theta_{ml}] = \frac{\sum E[y]}{n} = \theta$$

بنابراین تخمین فوق یک تخمین unbiased برای پارامتر تتا است. برای محاسبه واریانس می دانیم واریانس توزیع Poisson برابر با θ می باشد. از طرفی y ها نمونه های مستقل و با توزیع یکسان می‌باشند؛ بنابراین طبق قضیه حد مرکزی، $var(\theta_{ml}) = \frac{\theta}{n}$

g) با فرض $\alpha = 1$ برای توزیع گاما:

$$f(\theta|x) \sim \frac{likelihood \times prior\ density}{p(x)} = \frac{\frac{e^{-n\theta}\theta^{\sum y}}{\prod y!} \times \lambda e^{-\lambda\theta}}{p(x)} = \frac{\frac{e^{(-n-\lambda)\theta}\theta^{\sum y}\lambda}{\prod y!}}{p(x)}$$

$$p(x) = \int_0^\infty \frac{e^{(-n-\lambda)\theta}\theta^{\sum y}\lambda}{\prod y!} d\theta = \frac{\lambda}{\prod y!}\int_0^\infty e^{-(n+\lambda)\theta}\theta^{\sum y} d\theta$$

با اعمال تغییر متغیر $n + \lambda = a$ و $\sum y = b - 1$ ، انتگرال فوق تبدیل به تعریف تابع گاما می‌شود:

$$\int_0^\infty e^{-(a)\theta}\theta^{b-1} d\theta = \frac{\Gamma(b)}{a^b} = \frac{\Gamma(\sum y + 1)}{(n+\lambda)^{\sum y + 1}}$$

با جاگذاری p(x) در تابع posterior داریم:

$$f(\theta|x)=\frac{\frac{e^{(-n-\lambda)\theta}\theta^{\sum y}\lambda}{\prod y!}}{\frac{\lambda}{\prod y!}\frac{\Gamma(\sum y+1)}{(n+\lambda)^{\sum y+1}}}=\frac{(n+\lambda)^{\sum y+1}e^{(-n-\lambda)\theta}\theta^{\sum y}}{\Gamma(\sum y+1)}$$

$$E(\theta|x)=\int_0^\infty \frac{(n+\lambda)^{\sum y+1}e^{(-n-\lambda)\theta}\theta^{\sum y+1}}{\Gamma(\sum y+1)}d\theta$$
$$=\frac{(n+\lambda)^{\sum y+1}}{\Gamma(\sum y+1)}\int_0^\infty e^{(-n-\lambda)\theta}\theta^{\sum y+1}\,d\theta$$

مجددا با اعمال تغییر متغیر $n+\lambda=a$ و $\sum y+1=b-1$ ، انتگرال فوق تبدیل به تعریف تابع گاما می‌شود:

$$\int_0^\infty e^{-a\theta}\theta^{b-1}d\theta=\frac{\Gamma(b)}{a^b}=\frac{\Gamma(\sum y+2)}{(n+\lambda)^{\sum y+2}}$$

درنهایت باجاگذاری در رابطه $E(\theta|x)$ داریم:

$$E(\theta|x)=\frac{(n+\lambda)^{\sum y+1}}{\Gamma(\sum y+1)}\times\frac{\Gamma(\sum y+2)}{(n+\lambda)^{\sum y+2}}=\frac{\sum y+2}{(n+\lambda)}$$

همانطور که مشاهده می‌شود تخمین های ML و MAP تا حد زیادی به هم شبیه هستند. با میل n به سمت بی‌نهایت، نتیجه هر دو یکسان خواهد بود.

h)

$$L(\theta)=\frac{e^{-n\theta}\theta^{\sum y}}{\prod y!}$$

اگر تعریف کنیم $h(y)=\frac{1}{\prod y!}$ ، $T(y)=\sum y$ و $g_\theta\big(T(y)\big)=e^{-n\theta}\theta^{T(y)}$ طبق قضیه Fisher–Neyman T(y) یک sufficient statistic برای θ می باشد.

i)

$$\ln\big(L(\theta)\big)=\sum -\theta+y\ln\theta-\ln y!$$

$$I(\theta)=-nE\left[\frac{\partial^2(-\theta+y\ln\theta-\ln y!)}{\partial\theta^2}\right]=nE\left[\frac{y}{\theta^2}\right]=\frac{n}{\theta^2}E[y]=\frac{n}{\theta}$$

$$Cramér\text{–}Rao\ bound \quad var(\theta_{est}) \geq \frac{1}{I(\theta)} = \frac{\theta}{n}$$

$$efficiency(\theta_{ml}) = \frac{1}{I(\theta) \times var(\theta_{ml})} = \frac{1}{\frac{n}{\theta} \times \frac{\theta}{n}} = 1 \leq 1 \quad \theta_{ml}\ is\ efficient$$

۷)

---

a) فرض می‌کنیم k نمونه از میان تانک های موجود در میدان‌نبرد مشاهده شده و m در مشاهده شماره dام دیده شده‌است. احتمالا چنین اتفاقی برابر است با:

$$\frac{m-1}{n} \times \frac{m-2}{n-1} \times \frac{m-3}{n-2} \times \cdots \times \frac{m-d+1}{n-d+2} \times \frac{\mathbf{1}}{\boldsymbol{n-d+1}} \times \frac{m-d}{n-d} \times \frac{m-d-1}{n-d-1} \times \cdots \times \frac{m-d-(k-d-1)}{n-d-(k-d-1)} = \frac{(n-k)!}{n!} \times \frac{(m-1)!}{(m-k)!}$$

در عبارت فوق، بخشی که پررنگ شده مربوط به مشاهده d است که در آن m مشاهده شده است. بنابراین صورت کسر ۱ می‌باشد.عبارت های قبل و بعد از بخش پررنگ نیز، به ترتیب مربوط به مشاهدات قبل و بعد از آن است. همانطور که مشاهده می‌شود؛ عبارت فوق مستقل از d می‌باشد. بنابراین احتمال اینکه m بزرگترین سریال مشاهده شده باشد برابر است با k برابر احتمال فوق:

$$p(m|n,k) = k\frac{(n-k)!}{n!} \times \frac{(m-1)!}{(m-k)!} = \binom{m-1}{k-1} \times \binom{n}{k}^{-1} \quad (m \leq n, k \leq m)$$

b)

$$p(n|m,k) = \frac{p(m|n,k)p(n|k)}{p(m|k)} = \frac{p(m|n,k)p(n|k)}{\sum_{n=m}^{\infty} p(m|n,k)p(n|k)}$$

ابتدا احتمال p(n|k) را می‌یابیم. این عبارت برابر است با احتمال اینکه ارتش آلمان n تانک داشته باشد، در صورتی که k تانک مشاهده شده باشد. با توجه به p(n) که در صورت سوال داده شده، این احتمال برابر است با:

$$p(n|k) = \frac{1}{\Omega - \mathrm{k}} \quad (k \leq n, n < \Omega)$$

بنابراین:

$$p(n|m,k)=\frac{\binom{m-1}{k-1}\times\binom{n}{k}^{-1}\times(\Omega-\mathrm{k})^{-1}}{\sum_{n=m}^{\Omega-1}\binom{m-1}{k-1}\times\binom{n}{k}^{-1}\times(\Omega-\mathrm{k})^{-1}}=\frac{\binom{n}{k}^{-1}}{\sum_{n=m}^{\Omega-1}\binom{n}{k}^{-1}}$$

با توجه به راهنمایی صورت سوال داریم:

$$p(n|m,k)=\frac{\binom{n}{k}^{-1}}{\sum_{n=m}^{\Omega-1}\binom{n}{k}^{-1}}=\frac{\binom{n}{k}^{-1}}{\frac{k}{k-1}\left(\frac{1}{\binom{m-1}{k-1}}-\frac{1}{\binom{\Omega}{k-1}}\right)}$$

با میل Ω به سمت بی نهایت:

$$p(n|m,k)=\frac{\binom{n}{k}^{-1}}{\frac{k}{k-1}\left(\frac{1}{\binom{m-1}{k-1}}\right)}=\frac{k-1}{k}\times\frac{\binom{m-1}{k-1}}{\binom{n}{k}}\quad(m\le n\,,k\ge 2)$$

(c,d

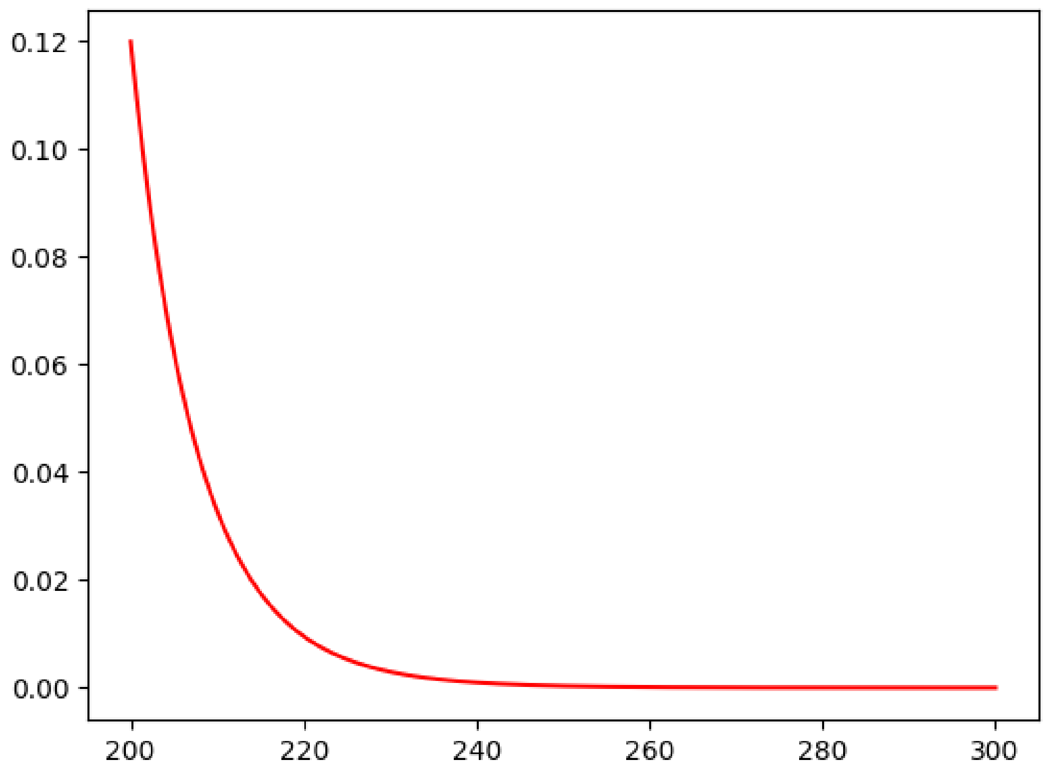
$$\begin{aligned}E(n|m,k)&=\sum_{n=m}^{n=\infty}\frac{k-1}{k}\times\frac{\binom{m-1}{k-1}}{\binom{n}{k}}\times n=\frac{k-1}{k}\binom{m-1}{k-1}\sum_{n=m}^{n=\infty}\frac{n}{\binom{n}{k}}\\&=\frac{k-1}{k}\binom{m-1}{k-1}\sum_{n=m}^{n=\infty}\frac{n}{\frac{n}{k}\binom{n-1}{k-1}}=(k-1)\binom{m-1}{k-1}\sum_{n=m}^{n=\infty}\frac{1}{\binom{n-1}{k-1}}=E\\&=(k-1)\binom{m-1}{k-1}\sum_{n-1=m-1}^{n-1=\infty}\frac{1}{\binom{n-1}{k-1}}\end{aligned}$$

با توجه به راهنمایی صورت سوال برای $k-1\ge 2$ داریم:

$$\begin{aligned}E(n|m,k)&=(k-1)\binom{m-1}{k-1}\frac{k-1}{k-2}\times\frac{1}{\binom{m-2}{k-2}}\\&=(k-1)\frac{m-1}{k-1}\binom{m-2}{k-2}\frac{k-1}{k-2}\times\frac{1}{\binom{m-2}{k-2}}=(k-1)\frac{m-1}{k-2}\quad(k\ge 3)\end{aligned}$$

بنابراین حداقل k برابر با ۳ می باشد.

e)کد مربوط به این بخش در فایل plot.py قرار دارد. مطابق انتظار با افزایش n، احتمال کاهش می‌یابد. طبق فرمولی که برای امیدریاضی در بخش قبل حساب شد؛ میانگین این توزیع باید در حدود نقطه ۲۰۷ باشد. این مورد از روی شکل به صورت نسبی قابل تشخیص است.



۹)

---

a) در شرایطی که ماتریس کوواریانس همه کلاس‌ها مساوی و به شکل σI باشند.

b) الگوریتم MDC در مرحله آموزش میانگین(نماینده) هر کلاس را با توجه به داده های آموزش محاسبه می‌کند. بنابراین دارای فاز آموزش می‌باشد. همچنین دسته‌بندی کننده بیز، در مرحله آموزش باید با توجه به داده های آموزش، احتمالات اولیه و لایکلی هود هر کلاس را محاسبه کند.

c) با توجه به یکسان بودن واریانس ها و با استفاده از روش بیز، نقطه مرز تصمیم ۱.۵ خواهد بود. حال احتمال خطا برابر است با:

$$p(error) = 2 \times \frac{1}{2} \times \int_{1.5}^{\infty} p(x|\omega_1) = p(X > 1.5) = p\left(\frac{X+1}{1} > \frac{1.5+1}{1}\right)$$

که برابر است با p(z>2.5)، با توجه به جدول توزیع نرمال، برابر است با ۰.۰۰۵۲ که همان مقدار خطا در مرحله تست است.

d) رویکردهای مختلفی برای این مسئله وجود دارد. رسم نمودار سه بعدی(برای مسئله سه کلاسه)، رسم نمودار هر کلاس در برابر سایر کلاس ها و رسم نمودار دو بعدی به صورت دو به دو برای کلاس‌ها

e) بیز برای تصمیم گیری به احتمالات اولیه و لایکلی هود به ازای تمام خروجی های ممکن نیاز دارد. در مسئله رگرسیون معمولا دامنه خروجی نامحدود است بنابراین امکان تعیین پارامترهای ذکر شده وجود نخواهد داشت.

f) خروجی بیز در صورتی که احتمالات posterior یکسان بدست آید، ممکن است یکتا نباشد.

g) در صورتی که تخمین اولیه ما درباره احتمال prior پارامتر مجهول اشتباه باشد،روش MAP دچار خطا خواهد شد و روش ML نتایج بهتری خواهد داد.

h) این روش مانند روش عادی است، با این تفاوت که یک عبارت پنالتی از تابع لایکلی هود کم می‌شود. این روش زمانی استفاده می شود که احتمال لایکلی هود نسبتاً flat باشد، بنابراین پیدا کردن تخمین دقیق برای آن دشوار است. عبارت پنالتی شامل اطلاعاتی است که تخمین ما را دقیق تر می‌کند.