810198358 امیرمهدی انصاری پور

Subject Date

۱) طبق قاعده بیز :

الف :

$$P(\omega_i|x) = \frac{P(x|\omega_i)P(\omega_i)}{P(x)}$$

$$\left.\begin{aligned} P(\omega_1|x) &= \frac{P(x|\omega_1)P(\omega_1)}{P(x)} \\ P(\omega_2|x) &= \frac{P(x|\omega_2)P(\omega_2)}{P(x)} \end{aligned}\right\} \xRightarrow{\text{I},\ \text{II}} \boxed{\begin{gathered} P(\omega_1|x) = P(\omega_2|x) \\ x = \frac{a_1+a_2}{2} \end{gathered}}$$

$$P(\omega_1) = P(\omega_2) \quad \text{I}$$

$$\left.\begin{aligned} P(x|\omega_1) &= \frac{1}{\pi b}\left(\frac{1}{1+\left(\frac{x-a_1}{b}\right)^2}\right) \\ P(x|\omega_2) &= \frac{1}{\pi b}\left(\frac{1}{1+\left(\frac{x-a_2}{b}\right)^2}\right) \\ x \rightarrow \frac{a_1+a_2}{2} &\Rightarrow \left(\frac{x-a_1}{b}\right)^2 = \left(\frac{x-a_2}{b}\right)^2 = \left(\frac{a_1-a_2}{2b}\right)^2 \end{aligned}\right\} \Rightarrow$$

$$P(x|\omega_1) = P(x|\omega_2) \quad \text{II}$$

۱) ب : مرز تقسیم در جایی قرار می گیرد که دو نمودار بهم برخورد

می کنند. در بخش قبل دیدیم که این مرز در $x = \frac{a_1+a_2}{2}$ قرار می گیرد.

Subject　　　　　　　　　　　　　　　　　　　　　　　　　Date

$$P(\text{error}) = \int_{-\infty}^{\frac{a_1+a_2}{2}} P(x|\omega_2) \times P(\omega_2)\, dx + \int_{\frac{a_1+a_2}{2}}^{+\infty} P(x|\omega_1) \times P(\omega_1)\, dx =$$

$$\frac{0.5}{\pi \cancel{b}} \left[\int_{-\infty}^{\frac{a_1+a_2}{2}} b^{\cancel{2}} \frac{1}{b^2+(x-a_2)^2}\, dx + \int_{\frac{a_1+a_2}{2}}^{+\infty} b^{\cancel{2}} \frac{1}{b^2+(x-a_1)^2}\, dx \right]$$

$$= \frac{0.5}{\pi} \left[\left[\tan^{-1}\left(\frac{x-a_2}{|b|}\right) \right]_{-\infty}^{\frac{a_1+a_2}{2}} + \left[\tan^{-1}\left(\frac{x-a_1}{|b|}\right) \right]_{\frac{a_1+a_2}{2}}^{+\infty} \right]$$

$$= \frac{0.5}{\pi} \left[\tan^{-1}\left(\frac{a_1-a_2}{|2b|}\right) - \underbrace{\cancel{\tan^{-1}(-\infty)}}_{-\pi/2} + \underbrace{\cancel{\tan^{-1}(+\infty)}}_{+\pi/2} - \tan^{-1}\left(\frac{a_2-a_1}{|2b|}\right) \right] \xRightarrow{\tan^{-1}(-x) = -\tan^{-1}(x)}$$

$$\frac{1}{2} - \frac{1}{\pi} \tan^{-1}\left(\left| \frac{a_2-a_1}{2b} \right| \right)$$

با فرض $a_2 > a_1$: $\quad |a_2 - a_1| = a_2 - a_1$

پ) طبق فرمول بدست آمده در بخش قبل، $P(\text{error})$ زمانی

بیشینه می‌شود که : $\quad \tan^{-1}\left(\left|\frac{a_2-a_1}{2b}\right|\right) = 0 \Rightarrow$

$$\left|\frac{a_2-a_1}{2b}\right| = 0 \Rightarrow \begin{cases} a_1 = a_2 \\ b \to \infty \text{ or } b \to -\infty \end{cases}$$

مقدار بیشینه خطا بنابراین برابر $1/2$ میشود.

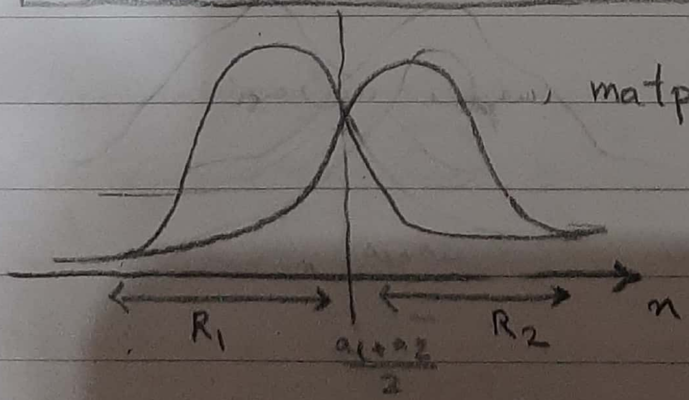
ت) در طبقه بندی بیزی داریم:

$$\frac{P(x|\omega_1)}{P(x|\omega_2)} \underset{2}{\overset{1}{\gtrless}} \cancel{\frac{P(\omega_2)}{P(\omega_1)}}^{\,1} \Rightarrow$$

$$\left(\frac{x-a_2}{b}\right)^2 \underset{2}{\overset{1}{\gtrless}} \left(\frac{x-a_1}{b}\right)^2 \rightarrow$$ طبق فرض مسئله $a_2 > a_1$ $\Rightarrow$

$$2x(\cancel{a_1-a_2}) \underset{2}{\overset{1}{\gtrless}} (\cancel{a_1-a_2})(a_1+a_2) \Longrightarrow$$

$$\boxed{\begin{array}{l} \text{if } x > \frac{a_1+a_2}{2}, \text{ decide } \omega_2 \\ \text{if } x < \frac{a_1+a_2}{2}, \text{ decide } \omega_1 \end{array}}$$



مرز تقسیم همانطور که در matplotlib رسم شده بدین صورت است،

$$P(error|x) = 1 - \min\{P(\omega_1|x), P(\omega_2|x)\}$$ میزان احتمال خطا

که اگر نسبت به x انتگرال بگیریم:

$$\boxed{\int P(error|x)\,dx = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left(\left|\frac{a_2-a_1}{2b}\right|\right)}$$

$$\text{Risk}(\omega_1|x) \underset{1}{\overset{2}{\gtrless}} \text{Risk}(\omega_2|x)$$ ث)

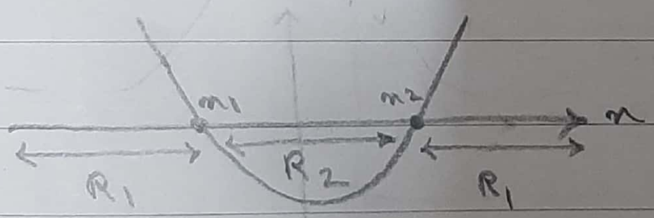
$$\Rightarrow \lambda_{12} P(\omega_2|x) + \lambda_{11} \underbrace{\cancel{P(\omega_1|x)}}_{0} \underset{1}{\overset{2}{\gtrless}} \lambda_{21} P(\omega_1|x) + \lambda_{22} \underbrace{\cancel{P(\omega_2|x)}}_{0}$$

$$\Rightarrow \frac{P(\omega_2|x)}{P(\omega_1|x)} \underset{1}{\overset{2}{\gtrless}} \frac{\lambda_{21}}{\lambda_{12}} \Rightarrow \frac{P(x|\omega_2)}{P(x|\omega_1)} \underset{1}{\overset{2}{\gtrless}} \frac{\overset{2}{\cancel{\lambda_{21}}}\, \cancel{P(\omega_1)}}{\underset{1}{\cancel{\lambda_{12}}}\, \cancel{P(\omega_2)}}$$

$$\Rightarrow \frac{b^2 + (x-a_1)^2}{b^2 + (x-a_2)^2} \underset{1}{\overset{2}{\gtrless}} 2$$

$$\Rightarrow 0 \underset{1}{\overset{2}{\gtrless}} x^2 + (2a_1 - 4a_2)x + b^2 - a_1^2 + 2a_2^2$$

$$\Rightarrow x_1, x_2 = 2a_2 - a_1 \pm \sqrt{8(a_1 - a_2)^2 - 4b^2}$$



همانطور که دیده میشود، بدلیل هزینه زیاد λ_{21} (یعنی پیش‌بینی ω_2 در حالیکه nature، ω_1 است)، R_2 یا همان فضایی که اگر x در آن قرار بگیرد، کلاس 2 پیش‌بینی میشود محدود شده. به عبارت دیگر، الگوریتم کمتر مایل است ω_2 پیش‌بینی کند.

Subject Date

$$P(\text{error}) = \int_{R_1} P(x, \omega_2)\, dx + \int_{R_2} P(x, \omega_1)\, dx =$$

$$\int_{R_1} P(x|\omega_2)\, P(\omega_2)\, dx + \int_{R_2} P(x|\omega_1)\, P(\omega_1)\, dx$$

$$= \frac{1}{2}\left[\int_{-\infty}^{x_1} P(x|\omega_2)\, dx + \int_{x_2}^{\infty} P(x|\omega_2)\, dx + \int_{x_1}^{x_2} P(x|\omega_1)\, dx \right]$$

$$= \frac{1}{2\pi}\left[\tan^{-1}\left(\frac{x_1 - a_2}{b}\right) + \pi - \tan^{-1}\left(\frac{x_2 - a_2}{b}\right) + \tan^{-1}\left(\frac{x_1 - a_1}{b}\right) - \tan^{-1}\left(\frac{x_2 - a_1}{b}\right) \right]$$

$$x_1 = 2a_2 - a_1 + \sqrt{8(a_1 - a_2)^2 - 4b^2}$$

$$x_2 = 2a_2 - a_1 - \sqrt{8(a_1 - a_2)^2 - 4b^2}$$

In [72]:
```
import math
pi = math.pi
import numpy as np
import matplotlib.pyplot as plt
```

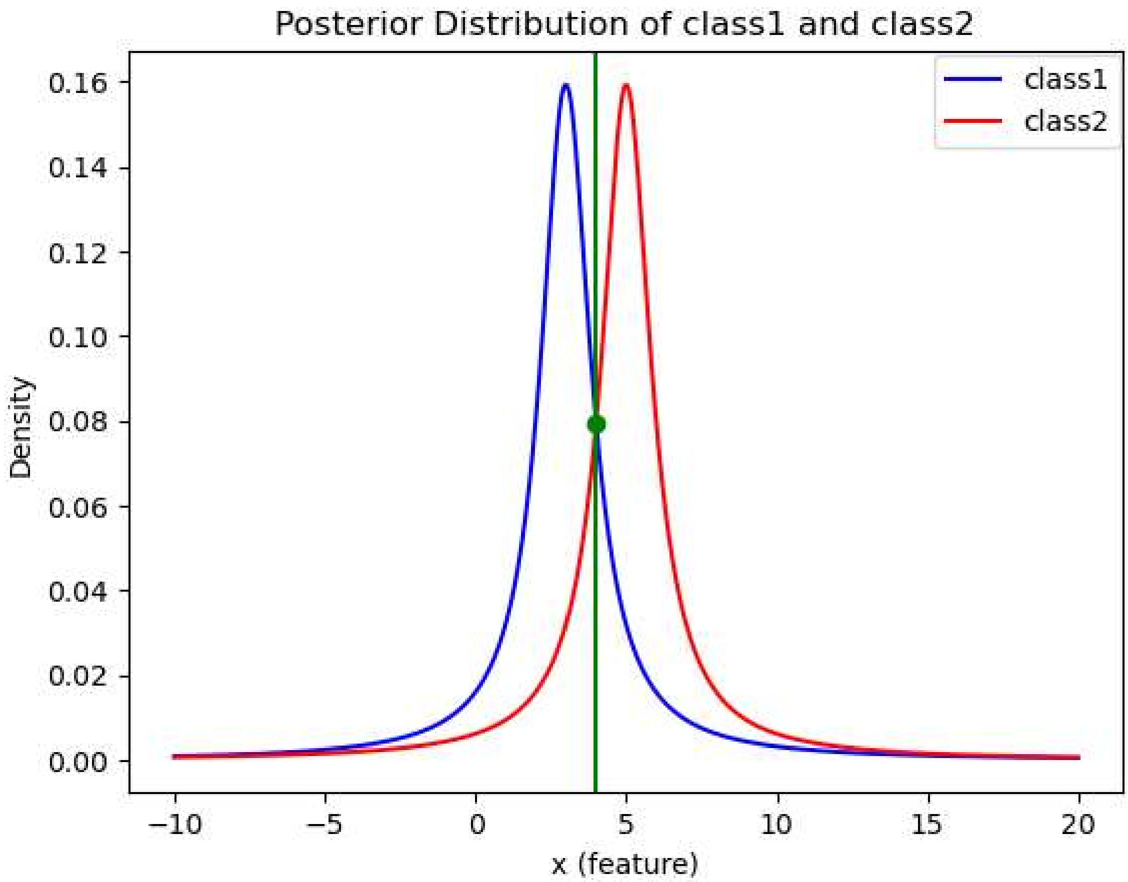
In [73]:
```
def drawDistributions(posterior1, posterior2):
    x = np.arange(-10, 20, 0.01)
    fig = plt.figure(0)
    idx = np.argwhere(np.diff(np.sign(posterior1 - posterior2))).flatten()
    print('Point of intersection: ', x[idx])

    plt.title('Posterior Distribution of class1 and class2')
    plt.xlabel('x (feature)')
    plt.ylabel('Density')
    plt.plot(x, posterior1, color = 'b')
    plt.plot(x, posterior2, color = 'r')
    plt.legend(['class1', 'class2'], loc = (0.81, 0.87))
    plt.axvline(x = x[idx], color = 'g', label = 'decision boundry')
    plt.plot(x[idx], posterior1[idx], 'go')
    plt.show()
```

In [74]:
```
prior1 = prior2 = 0.5
likelihood1 = (pi * (1 + ((x - 3)**2)))**(-1)
likelihood2 = (pi * (1 + ((x - 5)**2)))**(-1)
posterior1 = likelihood1 * prior1
posterior2 = likelihood2 * prior2

drawDistributions(posterior1, posterior2)
```

```
Point of intersection:  [4.]
```


Posterior Distribution of class1 and class2
class1
class2
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Density
−10
−5
0
5
10
15
20
x (feature)

In [ ]:

2) برای پاسخ مرز تقسیم داریم:

$$\frac{P(x|w_A)}{P(x|w_B)} \underset{B}{\overset{A}{\gtrless}} \frac{P(w_B)}{P(w_A)} = 1$$

$$\overset{x>0}{\Longrightarrow} \quad \frac{x}{\sigma_A^2} e^{-x^2/2\sigma_A^2} \underset{B}{\overset{A}{\gtrless}} \frac{x}{\sigma_B^2} e^{-x^2/2\sigma_B^2} \Longrightarrow$$

واضح است که $x>0$ زیرا $P(x|w_i)$ در $x<0$ ها صفر است $\Leftarrow$

$$\frac{\sigma_B^2}{\sigma_A^2} \underset{B}{\overset{A}{\gtrless}} e^{\frac{x^2}{2}\left(\frac{\sigma_B^2-\sigma_A^2}{\sigma_A^2\sigma_B^2}\right)} \Longrightarrow$$ از طرفین ln می‌گیریم $\Rightarrow$

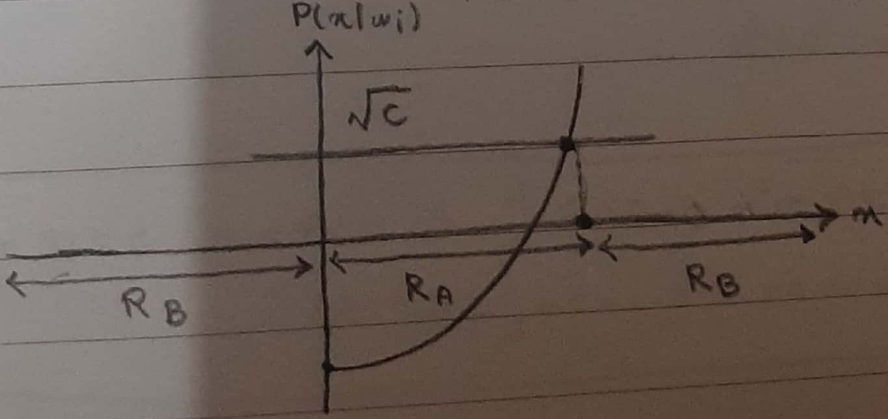
$$4\ln(\sigma_B) - 4\ln(\sigma_A) \underset{B}{\overset{A}{\gtrless}} x^2\left(\frac{\sigma_B^2-\sigma_A^2}{\sigma_B^2\sigma_A^2}\right) \Longrightarrow$$

با فرض $\sigma_A < \sigma_B$ برای عوض نشدن جهت $\underset{B}{\overset{A}{\gtrless}}$: $(\sigma_B^2-\sigma_A^2>0)$

$$\frac{\sigma_B^2\cdot\sigma_A^2}{\sigma_B^2-\sigma_A^2}\left[4\ln(\sigma_B) - 4\ln(\sigma_A)\right] \underset{B}{\overset{A}{\gtrless}} x^2 \overset{\text{جذر}}{\Longrightarrow}$$

$$x < \sqrt{\underbrace{\frac{\sigma_B^2\sigma_A^2}{\sigma_B^2-\sigma_A^2}\left[4\ln(\sigma_B) - 4\ln(\sigma_A)\right]}_{C\ (\text{constant})}}$$

$$x > -\sqrt{C}$$ (درست نبود $x>0$ ، پس به این نماد نیاز نیست)



3

$$\mu = \frac{1}{n}\sum_i x_i \qquad \Sigma_1 = \frac{1}{n}\sum_i (x_i-\mu)(x_i-\mu)^T \qquad (3$$

$$\underbrace{x_A}_{\text{1 colw}} = \begin{bmatrix}0.5\\0.5\end{bmatrix}, \begin{bmatrix}1.5\\0.5\end{bmatrix}, \begin{bmatrix}0.5\\-0.5\end{bmatrix}, \begin{bmatrix}1.5\\-0.5\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix},$$

$$\begin{bmatrix}-0.5\\0.5\end{bmatrix}, \begin{bmatrix}-1.5\\1\end{bmatrix}, \begin{bmatrix}-1\\-1\end{bmatrix}, \begin{bmatrix}-2\\-1\end{bmatrix}, \begin{bmatrix}-1.5\\0\end{bmatrix}$$

$$\Rightarrow \mu_A = \frac{1}{10}\begin{bmatrix}-1.5\\-1.5\end{bmatrix} = \begin{bmatrix}-0.15\\-0.15\end{bmatrix},$$

$$\Sigma_A = \begin{bmatrix}15.525 & 0.43\\0.43 & 5.025\end{bmatrix}$$

Subject

Date

$$x_B = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 1.5 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 2 \end{bmatrix}, \begin{bmatrix} 1.5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2.5 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\mu_B = \frac{1}{9} \begin{bmatrix} 12 \\ 14.5 \end{bmatrix} = \begin{bmatrix} 1.3 \\ 1.61 \end{bmatrix}$$

$$\Sigma_B = \begin{bmatrix} 5.01 & 1.667 \\ 1.667 & 8.888 \end{bmatrix}$$

$$g_i(x) = x^T W_i x + w_i^T x + w_{i0}$$

$$W_i = -\frac{1}{2} \Sigma_i^{-1} \qquad w_i = \Sigma_i^{-1} \mu_i$$

$$g_A(x): \quad W_A = -\frac{1}{2} \left(\frac{1}{77.82} \times \begin{bmatrix} 5.025 & -0.43 \\ -0.43 & 15.525 \end{bmatrix} \right)$$

$$w_A = \frac{1}{77.82} \underset{2\times2}{\begin{bmatrix} 5.025 & -0.43 \\ -0.43 & 15.525 \end{bmatrix}} \underset{2\times1}{\begin{bmatrix} -0.15 \\ -0.15 \end{bmatrix}}$$

$$\Rightarrow \boxed{W_A = \begin{bmatrix} -0.032 & 0.002 \\ 0.002 & -0.099 \end{bmatrix}} \quad \times -2$$

$$w_A = \frac{1}{77.82} \begin{bmatrix} -0.68 \\ -2.26 \end{bmatrix} = \boxed{\begin{bmatrix} -0.008 \\ -0.029 \end{bmatrix}}$$

$$w_{i0} = -\frac{1}{2} \mu_i^T \Sigma_i^{-1} \mu_i - \frac{1}{2} \ln |\Sigma_i| + \ln P(w_i)$$

Subject Date

$$w_{A0} = \underset{1\times2}{[-0.15 \;\; -0.15]} \begin{bmatrix}-0.032 & 0.002\\ 0.002 & -0.099\end{bmatrix}\begin{bmatrix}-0.15\\ -0.15\end{bmatrix}$$

$$-\frac{1}{2}\ln|77.82| + \ln\left(\frac{1}{2}\right)$$

احتمال های پیشین برابر

$$= -0.002 - 2.177 - 0.69 = -2.87$$

$$g_A(x) = \underset{1\times2}{[x_1 \;\; x_2]} \underset{2\times2}{\begin{bmatrix}-0.032 & 0.002\\ 0.002 & -0.099\end{bmatrix}}\begin{bmatrix}x_1\\ x_2\end{bmatrix} +$$

$$[-0.008 \;\; -0.029]\begin{bmatrix}x_1\\ x_2\end{bmatrix} - 2.87 =$$

$$\boxed{-0.032x_1^2 - 0.099x_2^2 + 0.004x_1x_2 - 0.008x_1 - 0.029x_2 - 2.87 \quad (\mathrm{I})}$$

$$g_B(x):\; W_B = -\frac{1}{2}\left(\frac{1}{41.74}\begin{bmatrix}8.88 & -1.667\\ -1.667 & 5.01\end{bmatrix}\right)$$

$$= \boxed{\begin{bmatrix}-0.106 & 0.019\\ 0.019 & -0.06\end{bmatrix}}$$

$$w_B = \frac{1}{41.74}\begin{bmatrix}8.88 & -1.667\\ -1.667 & 5.01\end{bmatrix}\begin{bmatrix}1.3\\ 1.61\end{bmatrix} = \boxed{\begin{bmatrix}0.21\\ 0.14\end{bmatrix}}$$

$$w_{B0} = [1.3 \;\; 1.61]\begin{bmatrix}-0.106 & 0.019\\ 0.019 & -0.06\end{bmatrix}\begin{bmatrix}1.3\\ 1.61\end{bmatrix} - \frac{1}{2}\ln(41.74)$$

$$+\ln\left(\frac{1}{2}\right) = -0.253 - 1.86 - 0.69 = \boxed{2.803}$$

Subject Date

$$g_B(x) = [x_1 \quad x_2] \begin{bmatrix} -0.106 & 0.019 \\ 0.019 & -0.06 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} +$$

$$[0.21 \quad 0.14] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 2.806 =$$

$$\boxed{-0.106 x_1^2 + 0.038 x_1 x_2 - 0.06 x_2^2 + 0.21 x_1 + 0.14 x_2 - 2.803 \quad \mathrm{II}}$$

I, II

$$\boxed{g_A(x) = g_B(x) \Longrightarrow g(x) = 0.074 x_1^2 - 0.039 x_2^2 - 0.034 x_1 x_2 - 0.218 x_1 - 0.169 x_2 - 0.067 = 0}$$

مرز تقسیم

ت) برای یافتن مرز تصمیم داریم:

$$Risk(\omega_1|x) = Risk(\omega_2|x)$$

$$\Rightarrow \lambda_{11} \cancel{P(\omega_1|x)} + \lambda_{12} P(\omega_2|x) = \lambda_{21} P(\omega_1|x) + \lambda_{22} \cancel{P(\omega_2|x)}$$

Subject Date

همانند قسمت قبل w_A را با کلاس 1 (نقاط سیاه) و w_B را با کلاس 2 (نقاط قرمز)

در نظر می‌گیریم:

$$\frac{P(w_A|x)}{P(w_B|x)} = \frac{\lambda_{AB}}{\lambda_{BA}} 2 \Rightarrow$$

$$\ln(P(x|w_A)P(w_A)) - \ln(P(x|w_B)P(w_B)) = \overset{0.69}{\ln(2)}$$

نتیجه مانند حالت ب است با این فرق که:

$$0.074x_1^2 - 0.039x_2^2 - 0.034x_1x_2$$

$$-0.218x_1 - 0.169x_2 - 0.757 = 0$$

ث) اینجا prior ها در term های w_{0A} و w_{0B} موثر اند پس داریم:

$$w_{0A}(x) = -\frac{1}{2}\mu_A^T \Sigma_A^{-1}\mu_A - \frac{1}{2}\ln|\Sigma_A| + \ln(\overset{1/3}{P_A})$$

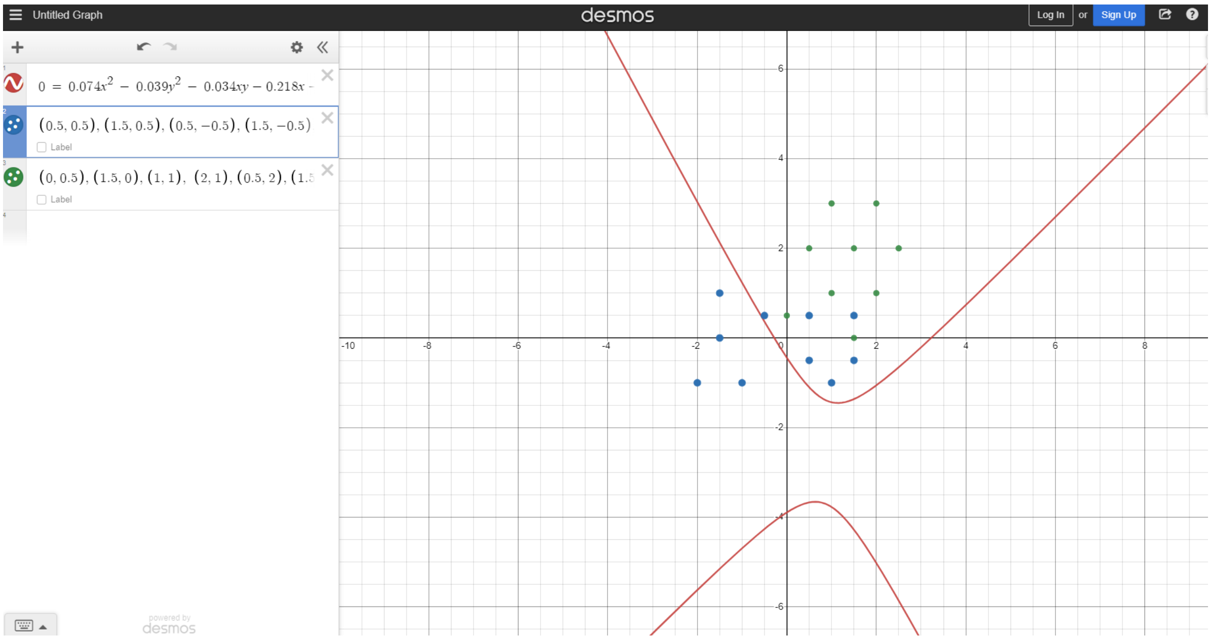
$$= -0.002 - 2.177 - 1.09 = -3.269$$

$$w_{0B}(x) = -2.518$$

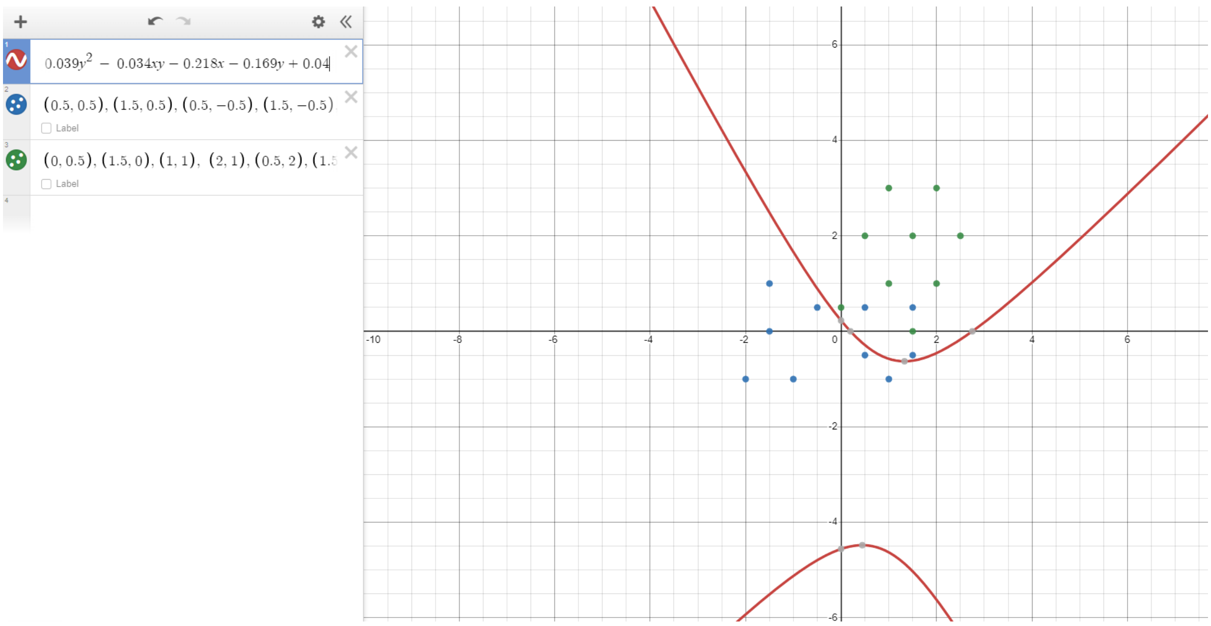
$$\Rightarrow \boxed{g(x) = 0.074x_1^2 - 0.039x_2^2 - 0.034x_1x_2 - 0.218x_1 - 0.218x_1 - 0.169x_2 - 0.751}$$

رسم نمودارهای سوال 3:

3-پ) همانطور که دیده می شود، تعدادی از sample های کلاس 1 misclassified شده اند. البته این موضوع در تصویر بعدی بهتر می شود، زیرا در آنجا احتمال های prior دقیق بصورت $P(w1) \ = \ 10 \ / \ 19$ و $P(w2) \ = \ 9/19$ در رابطه $g(w1)$ و $g(w2)$ جایگذاری شده اند.

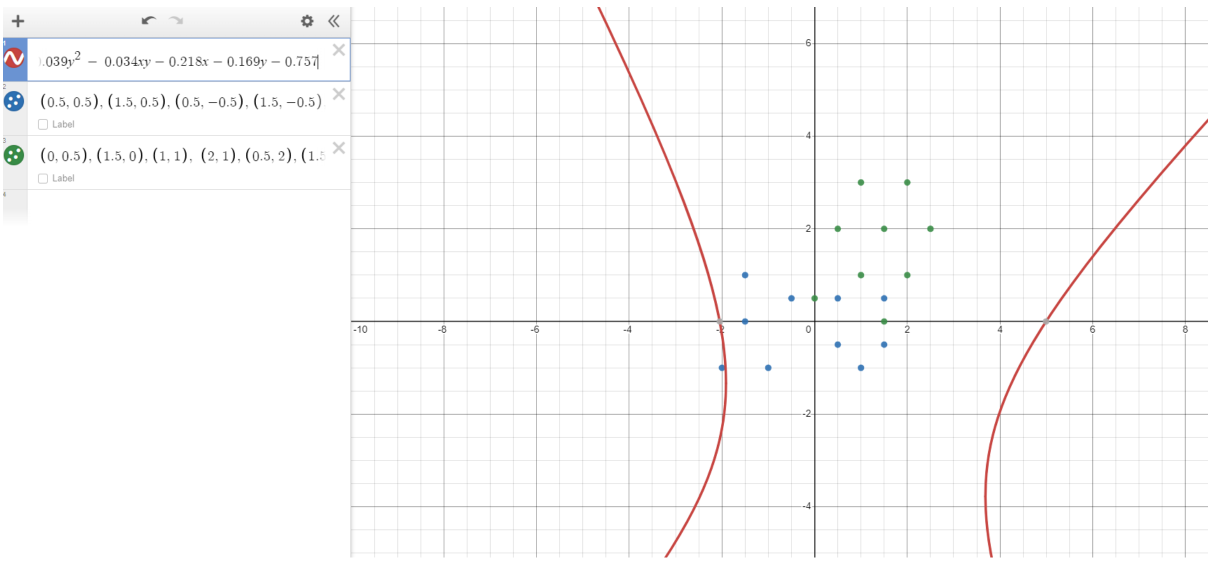


مرز تصمیم گیرنده در حالت $P(w1) \ = \ P(w2)$ (خطای تجربی = 6 / 19)



مرز تصمیم گیرنده در حالت $P(w1) = 10/19 \; and \; P(w2) = 9/19$
(خطای تجربی = 3 / 19)

3-ت) همانطور که دیده می شود، به دلیل هزینه زیاد پیش بینی اشتباه $w1$ در حالی که $w2 = nature$ می باشد ($\lambda12$) نسبت به حالت پیش بینی کلاس 2 برای نمونه کلاس 1 ($\lambda21$) مدل تمایل کمتری به پیش بینی کلاس 1 دارد. (تنها یک نمونه را به $w1$ map کرده که بسیار دور است از sample های $w2$).

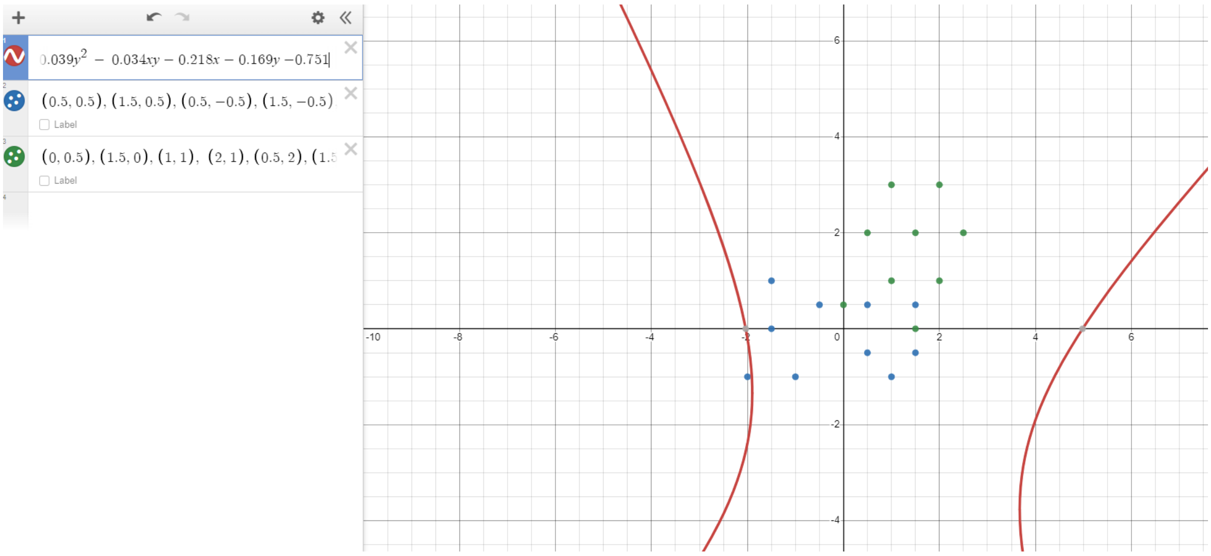


مرز تصمیم در حالت $\lambda12 = 2\lambda21$ (خطای تجربی = 9 / 19)

3-ث) همانطور که در تصویر دیده می شود، به دلیل بالا بودن مقدار threshold در رابطه Bayes classifier:

$$P(x|w2) \, / \, P(x|w1) \;\; >< \;\; P(w1) \, / \, P(w2)$$

مدل بر اساس احتمال های پیشین بیشتر کلاس 2 را پیش بینی میکند.



.039y^2 − 0.034xy − 0.218x − 0.169y −0.751
(0.5, 0.5), (1.5, 0.5), (0.5, −0.5), (1.5, −0.5)
Label
(0, 0.5), (1.5, 0), (1, 1), (2, 1), (0.5, 2), (1.5
Label

۴) الف)

$$\hat{\theta}_{ML} = \underset{\theta}{\arg\max} \ \ln P(D|\theta) \ , \ D = \{X_1, \dots, X_n\}$$

$$\Rightarrow \hat{\lambda}_{ML} = \underset{\lambda}{\arg\max} \sum_i \ln(X_i|\lambda) =$$

$$\underset{\lambda}{\arg\max} \left[\sum_i \ln(\lambda^{x_i}) + \ln(e^{-\lambda}) - \ln(x_i!)\right] =$$

$$\underset{\lambda}{\arg\max} \left[\sum_i x_i \ln\lambda - \lambda - \ln(x_i!)\right]$$

حل: برای یافتن $\hat{\lambda}$ بهینه، باید از عبارت بالا نسبت به λ مشتق بگیریم (چون و مساوی صفر قرار دهیم.

پارامتر در اینجا vector نیست، در صورت vector بودن، باید گرادیان می‌گرفتیم)

$$\frac{d}{d\lambda}\left[\sum_i (x_i \ln\lambda - \lambda - \ln(x_i!))\right] = 0 \Rightarrow$$

$$\sum_{i=1}^{n}\left[\frac{x_i}{\lambda} - 1\right] = 0 \overset{\times\lambda}{\Rightarrow} \sum_i [x_i - \lambda] = 0 \Rightarrow$$

$$\boxed{\hat{\lambda}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

ب)

$$P(\lambda|D) = \frac{P(D|\lambda)\,P(\lambda)}{\underbrace{P(D)}_{\text{عدد ثابت } a}} \propto \prod_{i=1}^{N} P(x_i|\lambda)\,P(\lambda)$$

$$= P(\lambda)\prod_{i=1}^{N}\frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = P(\lambda)\frac{\lambda^{\sum_i x_i} e^{-n\lambda}}{\prod_i (x_i)!}$$

Date

$$\Longrightarrow P(\lambda \mid D) \propto \lambda^{\alpha-1+\sum_i x_i} \, e^{-\lambda(n+\beta)} \Big/ \prod_i (x_i)!$$

$$\Longrightarrow P(\lambda \mid D) \propto \lambda^{\alpha-1+\underbrace{n\bar{x}}_{\sum_i x_i}} \, e^{-\lambda(n+\beta)} \Big/ \underbrace{\prod_i (x_i)!}_{\text{عدد ثابت (برحسب } \lambda)}$$

$$\Longrightarrow \boxed{P(\lambda \mid D) \propto e^{-\lambda(n+\beta)} \, \lambda^{\alpha-1+n\bar{x}}}$$

پ) همانطور که دیده می‌شود، توزیع posterior مانند prior از توزیع Gamma

پیروی می‌کند: $P(\lambda \mid D) \sim \text{Gamma}(\lambda \mid \underbrace{\alpha + n\bar{x}, \beta + n}_{\text{پارامترهای جدید}})$

بنابراین احتمال prior، $P(\lambda)$، یک conjugate prior است.

ت) $$\hat{\lambda}_{MAP} = \arg\max_{\lambda} P(\lambda \mid D)$$

برای یافتن $\hat{\lambda}_{MAP}$، از $P(\lambda \mid D)$ نسبت به λ مشتق گرفته و آن را صفر قرار می‌دهیم،

$$\frac{d}{d\lambda}\Big[\underset{\substack{\downarrow \\ \text{constant}}}{C} \, e^{-\lambda(n+\beta)} \, \lambda^{\alpha-1+n\bar{x}} \Big] = 0 \Longrightarrow$$

$$(-n-\beta) e^{-\lambda(n+\beta)} \lambda^{\alpha-1+n\bar{x}} + (\alpha-1+n\bar{x}) e^{-\lambda(n+\beta)} \lambda^{\alpha-2+n\bar{x}} = 0$$

$$\underbrace{e^{-\lambda(n+\beta)}\lambda^{\alpha-2+n\bar{x}}}_{\neq 0}\left[(-n-\beta)\lambda+(\alpha-1+n\bar{x})\right]=0$$

$$\Rightarrow \boxed{\hat{\lambda}_{MAP}=\frac{\alpha-1+n\bar{x}}{n+\beta}}$$

ث) وقتی $n \to \infty$:

$$\hat{\lambda}_{MAP}=\frac{\alpha-1+n\bar{x}}{n+\beta}\longrightarrow\frac{n\bar{x}}{n}=\boxed{\bar{x}}$$

همانطور که دیده می‌شود، $\hat{\lambda}_{MAP}$ به $\bar{x}$ (یا همان $\frac{\sum_i x_i}{n}$) میل می‌کند،

که برابر $\hat{\lambda}_{MLE}$ است.

ج) زمانیکه تعداد sample ها در مجموعه داده $D=\{x_1, \dots, x_n\}$

زیاد نباشد، یا به عبارت دیگر n به بی‌نهایت میل نکند، تفسیر $\hat{\theta}_{MAP}$

که از روش Bayesian بدست می‌آید مناسبتر است.