۸۱۰۱۹۹۴۲۵ ایمان رسولی پرتو

به نام خدا

ML
HW1

۱-

آ) $P(x|\omega_i) = \frac{1}{\pi b} \frac{1}{1+\left(\frac{x-a_i}{b}\right)^2}$; $a_2 > a_1$

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

$$\longrightarrow P(\omega_1|x) = P(\omega_2|x)$$

$$\frac{P(x|\omega_1)\cancel{P(\omega_1)}}{\cancel{P(x)}} = \frac{P(x|\omega_2)\cancel{P(\omega_2)}}{\cancel{P(x)}}$$

$$\longrightarrow P(x|\omega_1) = P(x|\omega_2)$$

$$\cancel{\frac{1}{\pi b}} \frac{1}{1+\left(\frac{x-a_1}{b}\right)^2} = \cancel{\frac{1}{\pi b}} \frac{1}{1+\left(\frac{x-a_2}{b}\right)}$$

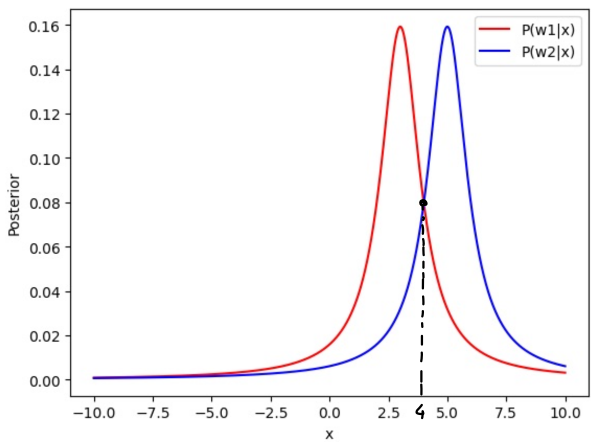
$$\longrightarrow \cancel{1}+\left(\frac{x-a_1}{\cancel{b}}\right)^2 = \cancel{1}+\left(\frac{x-a_2}{\cancel{b}}\right)^2 \xrightarrow{\sqrt{\ }} x-a_1 = \pm(x-a_2)$$

$$\longrightarrow \cancel{x}-a_1 = \cancel{x}-a_2 \longrightarrow a_1 = a_2 \; \times$$

طبق فرض $a_2 > a_1$

$$\longrightarrow x-a_1 = -x+a_2$$

$$\longrightarrow \boxed{x = \frac{a_1+a_2}{2}}$$



ب) $P(error) = P(\omega \neq \hat{\omega}) = P(\{\omega=1, \hat{\omega}=2\} \cup \{\omega=2, \hat{\omega}=1\})$ ‏ که احتمال خطا در Bayse classifier رخ میدهد:

برچسب واقعی

برچسب زده شده توسط classifier

$$\xrightarrow{\text{disjoint}} = P(\omega=1, \hat{\omega}=2) + P(\omega=2, \hat{\omega}=1)$$

$$+ P(\omega=1, \hat{\omega}=1) - P(\omega=1, \hat{\omega}=1)$$

$$= P(\omega=1) + P(\omega=2, r\in R_1) - P(\omega=1, r\in R_1)$$

$$= P(\omega=1) + P(\omega=2 \mid r\in R_1)P(r\in R_1) - P(\omega=1 \mid r\in R_1)P(r\in R_1)$$

$$= P(\omega=1) - \int_{R_1} P(r)\left[P(\omega_1|r) - P(\omega_2|r)\right] dr$$

قاعده بیز

$$\int_{-\infty}^{\frac{a_1+a_2}{2}} P(r|\omega_1)P(\omega_1) - P(r|\omega_2)P(\omega_2)$$

$$= \frac{1}{2}\int_{-\infty}^{\frac{a_1+a_2}{2}} \frac{1}{\pi b}\left(\frac{1}{1+\left(\frac{r-a_1}{b}\right)^2} - \frac{1}{1+\left(\frac{r-a_2}{b}\right)^2}\right) dr$$

$$= \frac{1}{2\pi}\left(\tan^{-1}\frac{r-a_1}{b} - \tan^{-1}\frac{r-a_2}{b}\right)\Bigg|_{-\infty}^{\frac{a_1+a_2}{2}}$$

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

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

قسمت پ $\longrightarrow P(error) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2-a_1}{2b}\right|$

پ) با توجه به رابطه بدست آمده در بخش ب، بیشترین مقدار خطا زمانی اتفاق می‌افتد که ترم $\tan^{-1}$ صفر شود چون تابع $\tan^{-1}|\theta|$ همواره مثبت است. این حالت زمانی رخ می‌دهد که $a_1=a_2$ شود یعنی دو توزیع دقیقاً روی یکدیگر می‌افتند و احتمال خطا هم $\frac{1}{2}$ می‌شود (با توجه به رابطه). چون دو توزیع کاملاً روی هم قرار گرفته‌اند به احتمال $\frac{1}{2}$ برچسب Predict شده با برچسب واقعی یکسان خواهد بود.

ت) مرز تصمیم $\longrightarrow$ محل برخورد $P(\omega_i|r)$ $\longrightarrow P(\omega_1|r) = P(\omega_2|r)$

$$\xrightarrow[P(\omega_1)=P(\omega_2)]{\text{طبق بیز}} P(r|\omega_1) = P(r|\omega_2)$$

$$\longrightarrow \cancel{\frac{1}{\pi b}}\frac{1}{1+\left(\frac{r-a_1}{b}\right)^2} = \cancel{\frac{1}{\pi b}}\frac{1}{1+\left(\frac{r-a_2}{b}\right)^2}$$

$$\longrightarrow r-a_1 = \pm(r-a_2)$$

$r-a_1 = r-a_2 \longrightarrow a_1 = a_2$ ✗

$r-a_1 = -r+a_2$

$\longrightarrow r = \frac{a_1+a_2}{2}$

مرز تصمیم

$$P(error) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2-a_1}{2b}\right| \longrightarrow \text{Baysian Classifier}$$

ت) $$\frac{P(r|\omega_1)}{P(r|\omega_2)} > \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}\;\frac{\cancel{P(\omega_2)}^{\,1}}{\cancel{P(\omega_1)}} \longrightarrow \frac{P(r|\omega_1)}{P(r|\omega_2)} > \frac{1}{2} \longrightarrow \frac{\cancel{\frac{1}{\pi b}}\;\frac{1}{1+\left(\frac{r-a_1}{b}\right)^2}}{\cancel{\frac{1}{\pi b}}\;\frac{1}{1+\left(\frac{r-a_2}{b}\right)^2}} > \frac{1}{2}$$

$$\longrightarrow \frac{1+\left(\frac{r-a_2}{b}\right)^2}{1+\left(\frac{r-a_1}{b}\right)^2} > \frac{1}{2} \xrightarrow{\text{مرز تصمیم}} \frac{1+\left(\frac{r-a_2}{b}\right)^2}{1+\left(\frac{r-a_1}{b}\right)^2} = \frac{1}{2} \longrightarrow \cancel{2}^{\,1} + 2\frac{(r-a_2)^2}{b^2} = \cancel{1} + \frac{(r-a_1)^2}{b^2}$$

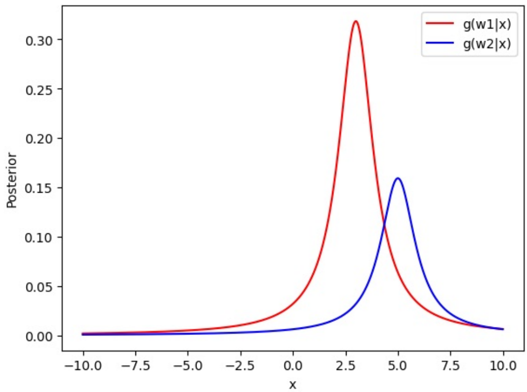
$$\xrightarrow{\times b^2} b^2 + 2(r^2 - 2a_2 r + a_2^2) = r^2 - 2a_1 r + a_1^2$$

$$\longrightarrow r^2 + 2(a_1 - 2a_2)r + b^2 + 2a_2^2 - a_1^2 = 0$$

$$\text{مرز تصمیم} \longleftarrow r = -(a_1-2a_2) \pm \sqrt{(a_1-2a_2)^2 - 2a_2^2 - b^2 + a_1^2} = -(a_1-2a_2) \pm \sqrt{2(a_1-a_2)^2 - b^2}$$

با مثال عددی داریم:

$$a_1 = 3,\quad a_2 = 5,\quad b = 1 \longrightarrow r = 7 \pm \sqrt{7}$$



$$P(error) = P(r \in R_1, \omega_2) + P(r \in R_2, \omega_1) = \int_{R_1} P(r|\omega_2)P(\omega_2)\,dr + \int_{R_2} P(r|\omega_1)P(\omega_1)\,dr$$

$$= \int_{-\infty}^{7-\sqrt{7}} P(r|\omega_2)P(\omega_2)\,dr + \int_{7+\sqrt{7}}^{+\infty} P(r|\omega_2)P(\omega_2)\,dr + \int_{7-\sqrt{7}}^{7+\sqrt{7}} P(r|\omega_1)P(\omega_1)\,dr$$

$$= \frac{1}{2\pi}\left(\tan^{-1}(2-\sqrt{7}) + \frac{\pi}{2} + \frac{\pi}{2} - \tan^{-1}(2-\sqrt{7}) + \tan^{-1}(4+\sqrt{7}) - \tan^{-1}(4-\sqrt{7})\right)$$

$$\longrightarrow P(error) = 0.27$$

2 –

الف)

$$R(1|r) = \lambda_{11} P(w_1|r) + \lambda_{12} P(w_2|r)$$

$$R(2|r) = \lambda_{21} P(w_1|r) + \lambda_{22} P(w_2|r)$$

$$\longrightarrow R(2|r) \overset{1}{\underset{2}{\gtrless}} R(1|r)$$

$$\xrightarrow{\text{جاگذاری}} \lambda_{21} P(w_1|r) + \lambda_{22} P(w_2|r) \overset{1}{\underset{2}{\gtrless}} \lambda_{11} P(w_1|r) + \lambda_{12} P(w_2|r)$$

$$\longrightarrow (\lambda_{21} - \lambda_{11}) P(w_1|r) \overset{1}{\underset{2}{\gtrless}} (\lambda_{12} - \lambda_{22}) P(w_2|r)$$

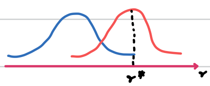
$$\xrightarrow{\text{قانون بیز}} (\lambda_{21} - \lambda_{11}) P(r|w_1) P(w_1) \overset{1}{\underset{2}{\gtrless}} (\lambda_{12} - \lambda_{22}) P(r|w_2) P(w_2)$$

$$\longrightarrow \frac{P(r|w_1)}{P(r|w_2)} \overset{1}{\underset{2}{\gtrless}} \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(w_2)}{P(w_1)}$$

$$\xrightarrow{\text{ناحیه تصمیم کلاس 1}} 0 < \frac{P(r|w_2)}{P(r|w_1)} < \frac{\lambda_{21} - \lambda_{11}}{\lambda_{12} - \lambda_{22}} \frac{P(w_1)}{P(w_2)}$$

ب)

$$\frac{f(r|w_2)}{f(r|w_1)} < \frac{\lambda_{21} - \lambda_{11}}{\lambda_{12} - \lambda_{22}} \frac{P(w_1)}{P(w_2)} < \frac{\lambda_{21}}{\lambda_{12}} \frac{P(w_1)}{P(w_2)}$$



حداکثر $\frac{f(r|w_2)}{f(r|w_1)}$ زمانی رخ می‌دهد

که r، $f(r|w_2)$ را Max کند.

3 - $P(\omega_i|r) = \dfrac{P(r|\omega_i)P(\omega_i)}{P(r)}$, $P(r|\omega_i) = \dfrac{r}{\sigma_i^2} e^{\frac{-r^2}{2\sigma_i^2}}$ $\quad r \geq 0$

$$\xrightarrow[P(\omega_1)=P(\omega_2)]{\text{مرز تصمیم}} P(r|\omega_1) = P(r|\omega_2)$$

$$\longrightarrow \frac{\cancel{r}}{\sigma_1^2} e^{-\frac{r^2}{2\sigma_1^2}} = \frac{\cancel{r}}{\sigma_2^2} e^{-\frac{r^2}{2\sigma_2^2}}$$

$$\longrightarrow \sigma_2^2 e^{-\frac{r^2}{2\sigma_1^2}} = \sigma_1^2 e^{\frac{-r^2}{2\sigma_2^2}}$$

$$\xrightarrow{\log} 2\log\sigma_2 - \frac{r^2}{2\sigma_1^2} = 2\log\sigma_1 - \frac{r^2}{2\sigma_2^2}$$

$$\longrightarrow r^2\left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right) = 2\log\frac{\sigma_1}{\sigma_2}$$

$$\longrightarrow r^2 = \frac{4\log\frac{\sigma_1}{\sigma_1}}{\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}} = 4\log\frac{\sigma_1}{\sigma_2}\,\frac{(\sigma_1\sigma_2)^2}{\sigma_1^2-\sigma_2^2}$$

$$\longrightarrow r = \pm\sqrt{4\log\frac{\sigma_1}{\sigma_2}\,\frac{(\sigma_1\sigma_2)^2}{\sigma_1^2-\sigma_2^2}} = \pm 2\sigma_1\sigma_2\sqrt{\frac{\log\frac{\sigma_1}{\sigma_2}}{\sigma_1^2-\sigma_2^2}} \rightarrow \text{مرز تصمیم}$$

4–

$$X_1 \sim N(r \mid \mu_1, \sigma^2)$$

$$X_2 \sim N(r \mid \mu_2, \sigma^2)$$

$$P(\omega_1 \mid r) = P(\omega_2 \mid r)$$

$$\xrightarrow{\text{قاعده بیز}} P(r \mid \omega_1) P(\omega_1) = P(r \mid \omega_2) P(\omega_2)$$

$$\longrightarrow \frac{P(r \mid \omega_1)}{P(r \mid \omega_2)} = \frac{P(\omega_2)}{P(\omega_1)} \longrightarrow \frac{\cancel{\frac{1}{\sqrt{2\pi}\sigma}}\, e^{-\frac{(r-\mu_1)^2}{2\sigma^2}}}{\cancel{\frac{1}{\sqrt{2\pi}\sigma}}\, e^{-\frac{(r-\mu_2)^2}{2\sigma^2}}} = \frac{P(\omega_2)}{P(\omega_1)}$$

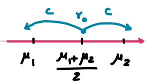
$$\xrightarrow{\log} -\frac{(r-\mu_1)^2}{2\sigma^2} + \frac{(r-\mu_2)^2}{2\sigma^2} = \log \frac{P(\omega_2)}{P(\omega_1)}$$

$$\xrightarrow{\times 2\sigma^2} (r-\mu_2)^2 - (r-\mu_1)^2 = 2\sigma^2 \log \frac{P(\omega_2)}{P(\omega_1)}$$

$$\cancel{r^2} - 2\mu_2 r + \mu_2^2 - \cancel{r^2} + 2\mu_1 r - \mu_1^2 = 2\sigma^2 \log \frac{P(\omega_2)}{P(\omega_1)}$$

$$2(\mu_1 - \mu_2) r = 2\sigma^2 \log \frac{P(\omega_2)}{P(\omega_1)} + \mu_1^2 - \mu_2^2$$

$$\xrightarrow{\mu_1 + \mu_2} r = \underbrace{\frac{\sigma^2 \log \frac{P(\omega_2)}{P(\omega_1)}}{\mu_1 - \mu_2}}_{c} + \underbrace{\frac{\mu_1 + \mu_2}{2}}_{r_0}$$



(فرض: $\mu_2 > \mu_1$)

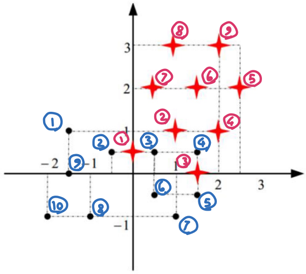
$$\xrightarrow{\text{مرز بین دو میانگین قرار نگیرد}} c > \mu_2 - r_0 \;,\; c < r_0 - \mu_1 \longrightarrow |c| > \frac{\mu_2 - \mu_1}{2}$$

$$\longrightarrow \left| \frac{\sigma^2 \log \frac{P(\omega_2)}{P(\omega_1)}}{\mu_1 - \mu_2} \right| > \frac{\mu_2 - \mu_1}{2}$$

$$\xrightarrow{\text{با فرض } \mu_2 > \mu_1} \boxed{\log \frac{P(\omega_2)}{P(\omega_1)} > \frac{(\mu_2 - \mu_1)^2}{2\sigma^2}}$$

شرط لازم برای اینکه مرز تصمیم بین دو میانگین قرار نگیرد

5_



$\omega_1: \; r_1 = \begin{bmatrix} -1,5 \\ 1 \end{bmatrix}, \; r_2 = \begin{bmatrix} -0,5 \\ 0,5 \end{bmatrix}, \; r_3 = \begin{bmatrix} 0,5 \\ 0,5 \end{bmatrix}, \; r_4 = \begin{bmatrix} 1,5 \\ 0,5 \end{bmatrix},$

$r_5 = \begin{bmatrix} 1,5 \\ -0,5 \end{bmatrix}, \; r_6 = \begin{bmatrix} 0,5 \\ -0,5 \end{bmatrix}, \; r_7 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \; r_8 = \begin{bmatrix} -1 \\ -1 \end{bmatrix},$

$r_9 = \begin{bmatrix} -1,5 \\ 0 \end{bmatrix}, \; r_{10} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$

$\omega_2: \; r'_1 = \begin{bmatrix} 0 \\ 0,5 \end{bmatrix}, \; r'_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; r'_3 = \begin{bmatrix} 1,5 \\ 0 \end{bmatrix}, \; r'_4 = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$

$r'_5 = \begin{bmatrix} 2,5 \\ 2 \end{bmatrix}, \; r'_6 = \begin{bmatrix} 1,5 \\ 2 \end{bmatrix}, \; r'_7 = \begin{bmatrix} 0,5 \\ 2 \end{bmatrix}, \; r'_8 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \; r'_9 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

الف) $\mu = \frac{1}{n} \sum_{i=1}^{n} r_i \longrightarrow \mu_1 = \begin{bmatrix} -0,15 \\ -0,15 \end{bmatrix}, \; \mu_2 = \begin{bmatrix} 1,33 \\ 1,61 \end{bmatrix}$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} ; \quad \sigma_{ii} = Var(X) = \frac{1}{n} \sum_{i=1}^{n} (r_i - \mu)^2 , \quad \sigma_{ij} = Cov(X,Y) = \frac{1}{n} \sum_{i=n} (r_i - \mu_r)(y_i - \mu_y)$$

$$\longrightarrow \Sigma_1 = \begin{bmatrix} 1,553 & 0,0025 \\ 0,0025 & 0,503 \end{bmatrix} , \quad \Sigma_2 = \begin{bmatrix} 0,556 & 0,185 \\ 0,185 & 0,988 \end{bmatrix}$$

ب)

$$P(r|\omega_1) = \frac{1}{2\pi |\Sigma_1|^{\frac{1}{2}}} e^{\underbrace{-(\frac{1}{2}(r-\mu_1)^T \Sigma_1^{-1} (r-\mu_1))}_{g_1(r)}} , \; |\Sigma_1| = 0,78 \longrightarrow \Sigma_1^{-1} = \frac{1}{0,78} \begin{bmatrix} 0,503 & -0,0025 \\ -0,0025 & 1,553 \end{bmatrix} = \begin{bmatrix} 0,644 & -0,0032 \\ -0,0032 & 1,988 \end{bmatrix}$$

$$r = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \rightarrow (r - \mu_1^T) \Sigma_1^{-1} (r - \mu_1) = \begin{bmatrix} r_1 + 0,15 & r_2 + 0,15 \end{bmatrix} \begin{bmatrix} 0,644 & -0,0032 \\ -0,0032 & 1,988 \end{bmatrix} \begin{bmatrix} r_1 + 0,15 \\ r_2 + 0,15 \end{bmatrix}$$

$$= \begin{bmatrix} 0,644 r_1 - 0,0032 r_2 + 0,096 & -0,0032 r_1 + 1,988 r_2 + 0,3 \end{bmatrix} \begin{bmatrix} r_1 + 0,15 \\ r_2 + 0,15 \end{bmatrix}$$

$$= 0,644 r_1^2 + 0,097 r_1 - 0,0032 r_1 r_2 - 0,0005 r_2 + 0,096 r_1 + 0,144$$

$$- 0,0032 r_1 r_2 - 0,0005 r_1 + 1,988 r_2^2 + 0,3 r_2 + 0,3 r_2 + 0,045$$

$$= 0,644 r_1^2 + 1,988 r_2^2 - 0,0064 r_1 r_2 + 0,1925 r_1 + 0,6 r_2 + 0,189$$

$$\longrightarrow g_1(r) = 0,322 r_1^2 + 0,994 r_2^2 - 0,0032 r_1 r_2 + 0,096 r_1 + 0,3 r_2 + 0,095$$

$$\longrightarrow P(r|\omega_1) = 0,18 \, e^{-g_1(r)}$$

$$P(r|\omega_2) = \frac{1}{2\pi |\Sigma_2|^{\frac{1}{2}}} e^{\underbrace{-(\frac{1}{2}(r-\mu_2)^T \Sigma_2^{-1} (r-\mu_2))}_{g_2(r)}} , \; |\Sigma_2| = 0,52 \longrightarrow \Sigma_2^{-1} = \frac{1}{0,52} \begin{bmatrix} 0,988 & -0,185 \\ -0,185 & 0,556 \end{bmatrix} = \begin{bmatrix} 1,92 & -0,36 \\ -0,36 & 1,08 \end{bmatrix}$$

$$\begin{bmatrix} r_1 - 1,33 & r_2 - 1,67 \end{bmatrix} \begin{bmatrix} 1,92 & -0,36 \\ -0,36 & 1,08 \end{bmatrix} \begin{bmatrix} r_1 - 1,33 \\ r_2 - 1,67 \end{bmatrix} = \begin{bmatrix} 1,92 r_1 - 0,36 r_2 - 1,95 & -0,36 r_1 + 1,08 r_2 - 1,325 \end{bmatrix} \begin{bmatrix} r_1 - 1,33 \\ r_2 - 1,67 \end{bmatrix}$$

$$= 1,92 r_1^2 - 2,55 r_1 - 0,36 r_1 r_2 + 0,48 r_2 - 1,95 r_1 + 2,6$$

$$- 0,36 r_1 r_2 + 0,6 r_1 + 1,08 r_2^2 - 1,8 r_2 - 1,325 r_2 + 2,21$$

$$= 1,92 r_1^2 + 1,08 r_2^2 - 0,72 r_1 r_2 - 3,9 r_1 - 2,645 r_2 + 4,8$$

$$\longrightarrow g_2(r) = 0,96 r_1^2 + 0,54 r_2^2 - 0,36 r_1 r_2 - 1,95 r_1 - 1,32 r_2 + 2,4$$

$$\longrightarrow P(r|\omega_2) = 0,22 \, e^{-g_2(r)}$$

$$P(\omega_1|x) \overset{1}{\underset{2}{\gtrless}} P(\omega_2|x)$$

$$\xrightarrow[P(\omega_1)=P(\omega_2)=0.5]{\text{با فرض}} P(x|\omega_1) \overset{1}{\underset{2}{\gtrless}} P(x|\omega_2) \longrightarrow \frac{P(x|\omega_1)}{P(x|\omega_2)} \overset{1}{\underset{2}{\gtrless}} 1 \longrightarrow \frac{0.18}{0.22} e^{-(g_1(x)-g_2(x))} \overset{1}{\underset{2}{\gtrless}} 1$$

$$\xrightarrow{\log} \underbrace{\cancel{\log 0.82}}_{-0.2} - (g_1(x)-g_2(x)) \overset{1}{\underset{2}{\gtrless}} 0$$

$$\longrightarrow g_1(x) - g_2(x) + 0.2 \overset{2}{\underset{1}{\gtrless}} 0$$

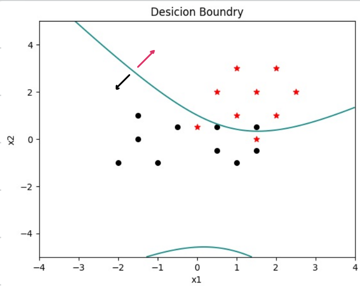
$$\longrightarrow -0.638x_1^2 + 0.454x_2^2 + 0.3568x_1x_2 + 1.854x_1 + 1.62x_2 - 2.11 \overset{2}{\underset{1}{\gtrless}} 0$$

Decision Boundry

از کلاس 1، یک نمونه به اشتباه طبقه‌بندی شده

از کلاس 2، دو نمونه به اشتباه طبقه‌بندی شده

$\longrightarrow$ درصد خطای تجربی $= \frac{1+2}{10+9} \times 100 = 15.8\%$



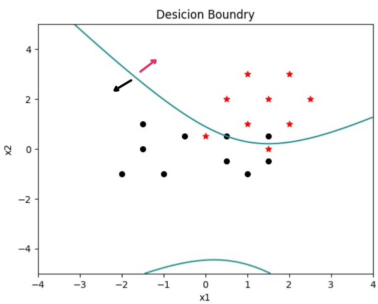
ب) Minimum Risk Classification : $\frac{P(x|\omega_1)}{P(x|\omega_2)} \overset{1}{\underset{2}{\gtrless}} \frac{\overset{2a}{\lambda_{12}} - \overset{0}{\cancel{\lambda_{22}}}}{\underset{a}{\lambda_{21}} - \underset{0}{\cancel{\lambda_{11}}}} \cdot \overset{1}{\cancel{\frac{P(\omega_2)}{P(\omega_1)}}}$

$$\longrightarrow \frac{P(x|\omega_1)}{P(x|\omega_2)} \overset{1}{\underset{2}{\gtrless}} 2 \longrightarrow 0.82\, e^{-(g_1(x)-g_2(x))} \overset{1}{\underset{2}{\gtrless}} 2$$

$$\xrightarrow{\log} 0.2 - (g_1(x)-g_2(x)) \overset{1}{\underset{2}{\gtrless}} 0.7$$

$$\longrightarrow g_1(x) - g_2(x) + 0.5 \overset{2}{\underset{1}{\gtrless}} 0$$

$$\longrightarrow -0.638x_1^2 + 0.454x_2^2 + 0.3568x_1x_2 + 1.854x_1 + 1.62x_2 - 1.81 \overset{2}{\underset{1}{\gtrless}} 0$$



ت) Minimum Risk Classification:

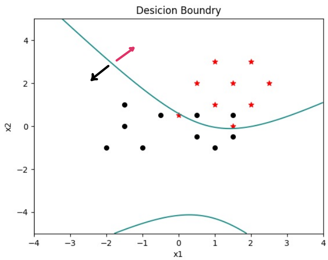
$$\frac{P(r|\omega_1)}{P(r|\omega_2)} \underset{2}{\overset{1}{\gtrless}} \frac{\overset{2a}{\cancel{\lambda_{12}}}-\overset{0}{\cancel{\lambda_{22}}}}{\underset{a}{\cancel{\lambda_{21}}}-\overset{0}{\cancel{\lambda_{11}}}} \, \frac{\overset{2}{\cancel{P(\omega_2)}}}{\cancel{P(\omega_1)}} \longrightarrow \frac{P(r|\omega_1)}{P(r|\omega_2)} \underset{2}{\overset{1}{\gtrless}} 4$$

$$\longrightarrow 0/82\, e^{-(g_1(r)-g_2(r))} \underset{2}{\overset{1}{\gtrless}} 4$$

$$\xrightarrow{\log} 0/2 - (g_1(r)-g_2(r)) \underset{2}{\overset{1}{\gtrless}} 1/4$$

$$\longrightarrow g_1(r) - g_2(r) + 1/2 \underset{1}{\overset{2}{\gtrless}} 0$$

$$\longrightarrow -0/638 r_1^2 + 0/454 r_2^2 + 0/3568 r_1 r_2 + 1/854 r_1 + 1/62 r_2 - 1/11 \underset{1}{\overset{2}{\gtrless}} 0$$



مشاهده میکنیم با اضافه شدن ریسک مرز تصمیم جابجا شده و همچنین خطا classifier افزایش مییابد.

۴) $P(r|\lambda) = \dfrac{\lambda^{r} e^{-\lambda}}{r!} \quad , \quad D = \{r_1, \dots, r_n\}$

آ) $L(\lambda) = \sum_{i=1}^{n} \log P(r_i|\lambda) = \sum_{i=1}^{n} r_i \log^{\lambda} - \lambda - \log^{r_i!}$ با فرض i.i.d $\{r_i\}$.

$$= \log^{\lambda} \sum_{i=1}^{n} r_i - n\lambda - \sum_{i=1}^{n} \log r_i!$$

$$\hat{\lambda}_{ML} = \arg\max_{\lambda} L(\lambda) \longrightarrow \frac{dL}{d\lambda} = 0 \longrightarrow \frac{\sum_{i=1}^{n} r_i}{\lambda} - n = 0$$

$$\longrightarrow \hat{\lambda}_{ML} = \frac{1}{n} \sum_{i=1}^{n} r_i$$

ب) $P(\lambda|D) \propto P(D|\lambda) P(\lambda)$

$$\downarrow$$

$$P(r_1, \dots, r_n|\lambda) P(\lambda)$$

$$\downarrow$$

$$\prod_{i=1}^{n} P(r_i|\lambda) P(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{r_i} e^{-\lambda}}{r_i!} c \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$= \frac{\lambda^{\sum_{i=1}^{n} r_i} e^{-n\lambda}}{\prod_{i=1}^{n} r_i!} c \lambda^{\alpha-1} e^{-\beta\lambda}$$

نسبت به λ، ثابت c'

$$= c' \lambda^{\alpha + \sum_{i=1}^{n} r_i - 1} e^{-(\beta+n)\lambda}$$

$$= \text{Gamma}(\lambda \mid \alpha_{new}, \beta_{new}) ; \quad \alpha_{new} = \alpha + \sum_{i=1}^{n} r_i , \quad \beta_{new} = \beta + n$$

پ) توزیع احتمال پسین ($P(\lambda|D)$) و احتمال پیشین ($P(\lambda)$) هردو از جنس توزیع گاما هستند. درنتیجه این توزیع برای پارامتر λ Conjugate Prior است.

ت) $\hat{\lambda}_{MAP} = \arg\max_{\lambda} P(\lambda|D) = \arg\max_{\lambda} \text{Gamma}(\lambda \mid \alpha_{new}, \beta_{new})$

$$\longrightarrow \hat{\lambda}_{MAP} = \frac{\alpha_{new} - 1}{\beta_{new}} = \frac{\alpha + \sum_{i=1}^{n} r_i - 1}{n + \beta}$$

ث) $\hat{\lambda}_{MAP} = \dfrac{\alpha + \sum_{i=1}^{n} x_i - 1}{n+\beta} \xrightarrow{n \to \infty} \hat{\lambda}_{MAP} = \dfrac{1}{n} \sum_{i=1}^{n} x_i = \hat{\lambda}_{ML}$

با میل کردن $n \to \infty$ ترم $\alpha - 1$ در صورت و β در مخرج قابل صرف‌نظر کردن هستند و تخمین‌گر MAP به MLE میل می‌کند.

ج) هنگامی که داده آموزشی زیادی در اختیار داشته باشیم، می‌توانیم از MLE استفاده کنیم که محاسبه آن راحت‌تر است. اما اگر n (تعداد داده‌های آموزشی) کم باشد، بهتر است از دیدگاه بیزین و تخمین‌گر MAP مسئله را حل کنیم چون می‌توانیم دانش خود را به عنوان احتمال Prior وارد مسئله کنیم و مستقل داده‌ها بیشتر باشیم.

پاسخ سوال ۷:

الف) طبقه‌بند ساده بیز فرض می‌کند که ویژگی های متغیرهای ورودی مستقل از یکدیگر هستند، به این معنی که وجود یک ویژگی بر احتمال وجود ویژگی دیگر تأثیر نمی‌گذارد. این فرض اغلب در عمل درست نیست، اما این الگوریتم همچنان می‌تواند عملکرد خوبی داشته باشد، به خصوص زمانی که مجموعه داده بزرگ باشد.

در مقابل، طبقه‌بند Naïve Bayes سریع و کارآمد است و در مقایسه با سایر الگوریتم‌های یادگیری ماشین به مقدار کمی از داده‌های آموزشی نیاز دارد. اما زمانی که فرض استقلال نقض می شود، یا داده‌های آموزشی نامتعادل هستند یا دارای نویز هستند، ممکن است عملکرد خوبی نداشته باشد.

از نظر ساختاری، در Naïve Bayes فرض می‌کنیم، feature ها از یکدیگر مستقل هستند در نتیجه برای محاسبه likelihood از رابطه ساده شده زیر استفاده می‌شود:

$$p(\boldsymbol{x}|\omega_i) = p(x_1|\omega_i) \times \dots \times p(x_d|\omega_i)$$

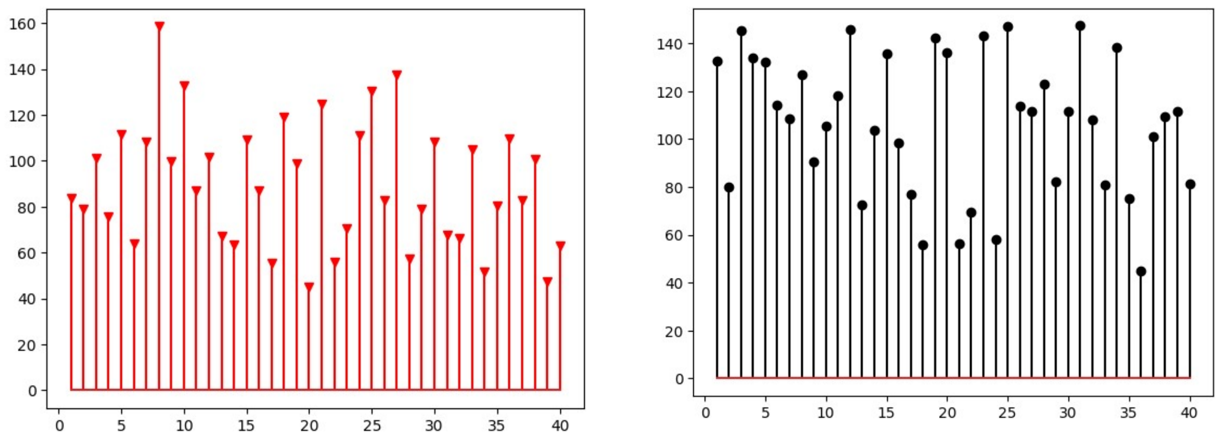
با توجه به تفاوت‌ها در بسیاری از موقعیت‌ها استفاده از طبقه‌بند بیز ساده منطقی‌تر است. اما زمانی که تعداد کلاس‌ها زیاد و داده‌های آموزشی محدود باشد، استفاده از طبقه‌بند Naïve Bayse منطقی‌تر به نظر می‌رسد.

برای پیش‌پردازش داده‌ها، داده‌هایی که برخی از ویژگی‌های آنها با x مشخص شده را از drop dataset میکنیم. چرا که اطلاعاتی از ویژگی‌های آن کلاس در اختیارمان قرار نمی‌دهد.

پاسخ سوال ۸:

معیار اول

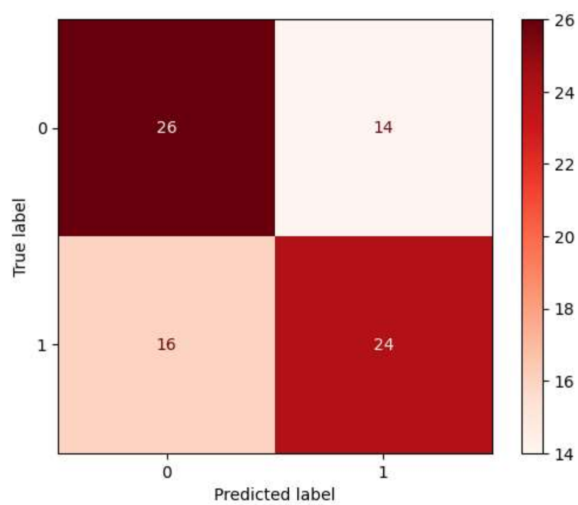
در ابتدا تصاویر را خوانده و میانگین هر یک را محاسبه میکنیم. برای دو کلاس ابری و غروب نتایج زیر بدست می‌آید:



شکل ۱- نمودار میانگین تصاویر ابری (مشکی) و غروب آفتاب (قرمز)

با معیار نخست، می‌توان مرز تصمیم را ۱۰۰ در نظر گرفت. چرا که تعداد خوبی از تصاویر ابری میانگین بالای ۱۰۰ و تصاویر غروب آفتاب میانگین زیر ۱۰۰ دارند.

با اعمال این معیار روی همین تصاویر، ماتریس آشفتگی به شکل زیر بدست می‌آید:



شکل ۲- ماتریس آشفتگی معیار نخست

$$Precision = \frac{TP}{TP + FP} = \frac{26}{26 + 16} \cong 0.62$$

$$Recall = \frac{TP}{TP+FN} = \frac{26}{26+14} = 0.6$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{26 + 24}{26 + 24 + 16 + 14} = 0.625$$

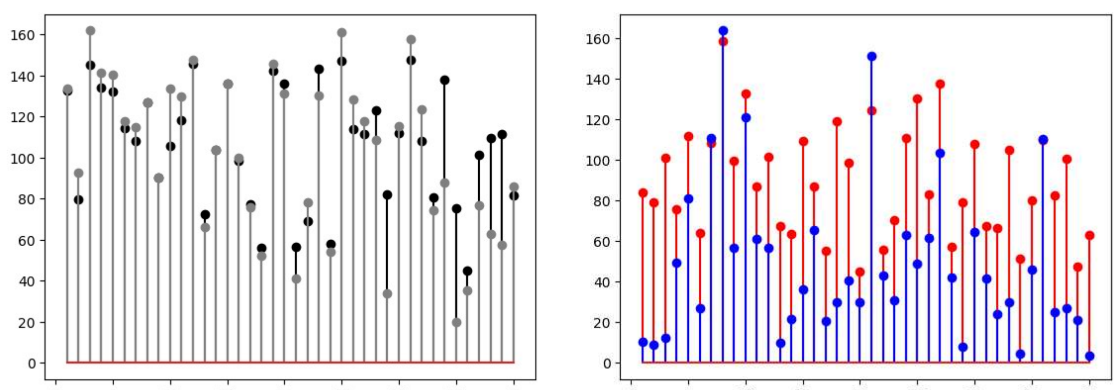
برخی تصاویری که به اشتباه طبقه‌بندی شده‌اند:

شکل ۳- دو نمونه از تصاویری که به اشتباه طبقه‌بندی شده‌اند

با توجه به شکل ۱، با این معیار خطای زیادی را خواهیم داشت. علت خطا تصاویری مثل شکل ۳ هستند که در هوای ابری طیف رنگی متمایل به قرمز بالاست و در غروب طیف رنگی آبی زیادی حضور دارد.

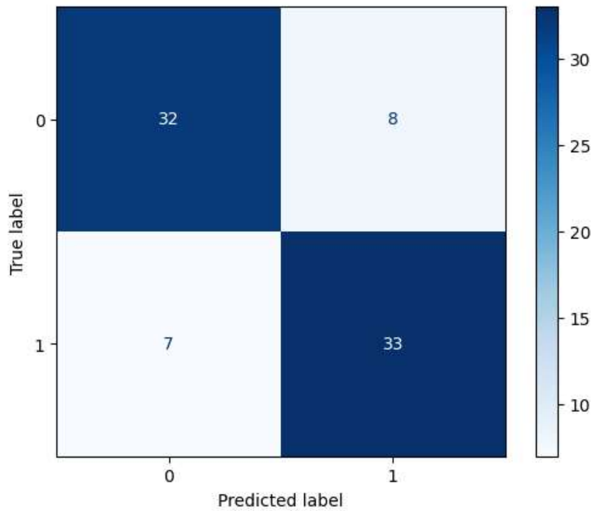
معیار دوم

با توجه به تصاویر یک معیار دقیق‌تر این است که میانگین طیف رنگی آبی را در تصاویر مقایسه کنیم. برای اینکار، میانگین ماتریس slice شده از ماتریس ۳ بعدی RGB را مقایسه میکنیم.



شکل ۴- نمودار میانگین بعد آبی تصاویر نسبت به میانگین اصلی

می‌بینیم که میانگین بعد آبی در تصاویر غروب، اکثرا زیر ۷۰ و برای تصاویر ابری اکثرا بالای ۷۰ است. با آزمون و خطا حول این عدد و اعمال این معیار به مرز تصمیم ۶۶ و ماتریس آشفتگی زیر می‌رسیم:



شکل ۵- ماتریس آشفتگی معیار دوم

$$Precision = \frac{32}{32+7} = 0.82$$

$$Recall = \frac{32}{32+8} = 0.8$$

$$Accuracy = \frac{32+33}{32+33+7+8} = 0.8125$$

مشاهده می‌کنیم با این معیار به دقت بالاتر و متعاقبا خطای کمتری رسیدیم.

برخی از تصاویری که به اشتباه طبقه‌بندی شده‌اند:

شکل ۶- دو نمونه از تصاویری که به اشتباه در معیار دوم طبقه‌بندی شده‌اند

علت خطا، در تصویر غروب، آبی رنگ بودن آسمان است. به همین سبب در بعد آبی میانگین بالایی دارد و از مرز تصمیم بالاتر می‌باشد. در تصویر هوای ابری هم به سبب محیط سبز رنگ میانگین بعد آبی کم است و از مرز تصمیم پایین‌تر قرار می‌گیرد.