1) در ابتدا اطلاعات لازم برای محاسبات را بدست می‌آوریم.

$$\mu_1 = \begin{pmatrix} -5.33 \\ -1.83 \end{pmatrix} , \quad \mu_2 = \begin{pmatrix} 3.29 \\ 5 \end{pmatrix}$$

$$\Sigma_1 = \begin{pmatrix} 4.66 & 3.07 \\ 3.07 & 4.97 \end{pmatrix} , \quad \Sigma_2 = \begin{pmatrix} 2.24 & 1.67 \\ 1.67 & 4 \end{pmatrix}$$

باتوجه به محاسبات پیچیده در ادامه از روابط لازم برای محاسبه μ و Σ خودداری کردیم. محاسبه میانگین که ساده و برحسب $\frac{\text{مجموع}}{\text{تعداد}}$ برای هر بعد به دست می‌آید، Σ هم براین اساس که:

$$\Sigma = \begin{array}{c|cc} & x & y \\ \hline x & \sigma_x^2 & \sigma_{xy} \\ y & \sigma_{yx} & \sigma_y^2 \end{array}$$

که رابطه محاسبه σ_{xy} هم به راحتی از مجموع: اختلاف مقادیر با میانگین وضرب درهم، به دست می‌آید.

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1} (x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i|$$

$$+ \ln(P(\omega_i)) + \ln(\lambda_{ji} - \lambda_{ii})$$

احتمالات اولیه (prior) هم براساس تعداد نقاط برابر با $\frac{6}{13}$ و $\frac{7}{13}$ می‌باشد.

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

$$(2-1) P(x|\omega_1) P(\omega_1) = (5-1) P(x|\omega_2) P(\omega_2)$$

$$\frac{6}{13} P(x|\omega_1) = \frac{4\times 7}{13} P(x|\omega_2) \rightarrow 6P(x|\omega_1) = 28P(x|\omega_2)$$

$\frac{3}{=}$ $g_1(x) = g_2(x) \rightarrow -\frac{1}{2} x^t \Sigma_1^{-1} x + \mu_1^t \Sigma_1^{-1} x - \frac{1}{2} \mu_1^t \Sigma_1^{-1} \mu_1 - \cancel{\frac{d}{2} \ln 2\pi}$

$-\frac{1}{2} \ln (|\Sigma_1|) + \ln P(\omega_1) + \ln(\lambda_{21} - \lambda_{11})$

$= -\frac{1}{2} x^T \Sigma_2^{-1} x + \mu_2^t \Sigma_2^{-1} x + \frac{1}{2} \mu_2^t \Sigma_2^{-1} \mu_2 - \cancel{\frac{d}{2} \ln 2\pi} - \frac{1}{2} \ln(|\Sigma_2|)$

$+ \ln P(\omega_2) + \ln(\lambda_{21} - \lambda_{22})$

$\circledast$

$\rightarrow -\frac{1}{2} (0.36 x^2 - 0.44 xy + 0.34 y^2) + (-1.52 x + 0.57 y)$

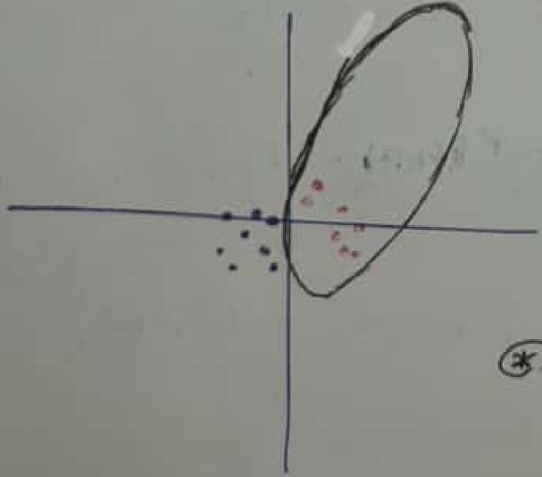
$- \frac{1}{2} (7.04) + 1.33 - 0.77$

$= -\frac{1}{2} (0.65 x^2 - 0.54 xy + 0.36 y^2) + (0.78 x + 0.93 y)$

$-\frac{1}{2} (7.19) + 0.92 - 0.62 + 1.36$

$\rightarrow 0.145 x^2 - 0.05 xy + 0.01 y^2 - 2.3 x - 0.36 y - 1.05 = 0$

بیضی



$\circledast$ $|\Sigma_1| = 0.07$, $|\Sigma_2| = 0.16$

$$\Sigma_1^{-1} = \begin{pmatrix} 0.36 & -0.22 \\ -0.22 & 0.34 \end{pmatrix}$$

$$\Sigma_2^{-1} = \begin{pmatrix} 0.65 & -0.27 \\ -0.27 & 0.36 \end{pmatrix}$$

Scanned with CamScanner

3

a (٢

M classes : $C_1, \dots, C_M$

x : input vector

regions : $R_1, \dots, R_M$

if $M=2$: $P_e = P(x \in R_2, C_1) + P(x \in R_1, C_2)$

$$= P(x \in R_2 | C_1) P(C_2) + P(x \in R_1 | C_2) P(C_1)$$

$$= P(C_2) \int_{R_2} P(x \in R_2 | C_1) + P(C_1) \int_{R_1} P(x \in R_1 | C_2) \quad (*)$$

Since $R_1 \cup R_2$ covers all feature space:

$$P(C_1) = \int_{R_1} P(C_1|x) P(x)\, dx + \int_{R_2} P(C_1|x) P(x)\, dx \quad (**)$$

$(*)\ (**) \rightarrow P_e = P(C_1) - \int_{R_1} [P(C_1|x) - P(C_2|x)]\, P(x)\, dx$

$$P_e = P(C_1) - \int_{R_1} [P(C_1|x) - P(C_2|x)]\, P(x)\, dx$$

$P_e \rightarrow$ will be minimized if $\rightarrow \begin{cases} [P(C_1|x) - P(C_2|x)] > 0 : R_1 \\ [P(C_1|x) - P(C_2|x)] < 0 : R_2 \end{cases}$

This the Bayesian decision theory.

Now for $M > 2 \rightarrow$ consider one class agains others (2 class classification), so based on what we said above we will have: $P(C_k | x) > P(C_j | x) \quad \forall j \neq k$

$\rightarrow$ Bayesian decision theory

b) مسئله را بر اساس برهان خلف حل می‌کنیم.

مقدار خطای داده شده زمانی به دست می‌آید که:

$$\underbrace{\frac{1}{m} \quad \cdots \quad \frac{1}{m}}_{\text{احتمال برای هر کلاس}} \rightarrow P_e = 1 - \frac{1}{m} = \frac{m-1}{m}$$

حال فرض می‌کنیم که می‌توان خطای بیشتری از این را هم به دست آورد.

برای این منظور احتمال یکی از کلاس را $P > \frac{1}{m}$ قرار می‌دهیم.

در نتیجه برای کلاس هدف $P' < \frac{1}{m}$ خواهد شد.

الزاماً $P_e < \frac{m-1}{m}$ اما مشکلی که وجود دارد این است که

بر اساس Bayesian theory دیگر کلاس انتخاب شده کلاس هدف نیست و کلاس هدف آن کلاسی است که

$P > \frac{1}{m}$ چون مجموع احتمال کلاس های دیگر از بقیه حالات کمتر است پس در واقع $P_e < \frac{m-1}{m}$ ← خلف

پس max برابر $\frac{m-1}{m}$ است.

---

c) برای این حالت به سادگی از روش one vs all استفاده می‌کنیم.

یکی یک کلاس را انتخاب می‌کنیم، بقیه کلاس ها را که یک کلاس فرض می‌کنیم ← شبیه ۲ کلاسه خواهد شد

۵

۳)

a

$$L(x,\lambda) = \|x - x_0\|^2 + 2\lambda(w^T x + b)$$

$$= \|x\|^2 - 2(x_0 - \lambda w)^T x + 2\lambda b + \|x_0\|^2$$

$$\frac{\partial L}{\partial x} = 0 = 2x - 2(x_0 - \lambda w)$$

$$\rightarrow 2x = 2(x_0 - \lambda w) \rightarrow x = x_0 - \lambda w$$

$$\lambda = \frac{w^T x_0 + b}{\|w\|^2}$$

$$\|x_0 - \lambda w - x_0\| = |\lambda| \, \|w\| = \frac{|w^T x_0 + b|}{\|w\|}$$

b

$$x^T A x = 1$$

$$\text{فاصله} \rightarrow \min_{x^T A x = 1} \|x - x_0\|^2$$

$$\rightarrow L = \|x - x_0\|^2 + \lambda(1 - x^T A x)$$

$$\nabla L = 0 \rightarrow 2(x - x_0 + \lambda A x) = 0$$

$$x = (I + \lambda A)^{-1} x_0$$

فرض می‌کنیم A ماتریس قطری باشد (اگر نباشد قطری می‌کنیم و در سیستم مختصات جدید با آن کار می‌کنیم) : $\quad d_i = a_{ii}$

$$x_i = \frac{x_{0i}}{1 + \lambda d_i}$$

در این حالت :

$$x^T A x = 1 \longrightarrow \sum \frac{d_i x_{0i}^2}{(1+\lambda d_i)^2} = 1$$

نمی‌توان λ را به راحتی محاسبه کرد اما با استفاده از روش عددی می‌توان آن را محاسبه کرد.

برای روش عددی دو عدد در نظر می‌گیریم و سعی می‌کنیم این دو عدد را نزدیک می‌کنیم. وقتی اختلاف دو عدد در حد ε شود آنگاه به جواب رسیده‌ایم.

ایده از روش Newton می‌آید که در آن :

$$\theta_1 = \theta - \frac{f(\theta)}{f'(\theta)}$$

چون ما اینجا می‌خواهیم :

$$\underbrace{\sum \frac{d_i x_{0i}^2}{(1+\lambda d_i)^2} - 1}_{f} = 0$$

پس f' را محاسبه می‌کنیم و مقدار f و f' را براساس d_i ها و x_{0i} ها محاسبه می‌کنیم یعنی :

$\varepsilon = 1e{-}10$

$x' = 0$

$x'' = 0.1$

while $abs(x_1 - x_2) > \varepsilon$:

$x' = x''$

$x'' = x' - \frac{f}{f'}$

در نهایت $\longleftarrow \lambda = x''$ و فاصله $= \| x - z \|$

و

$$z = (I + \lambda A)^{-1} x_0$$

7

c

A) $x = (I + \lambda A)^{-1} x_0 = \begin{bmatrix} 1+\lambda & 1 \\ 1 & 1+\lambda \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2+\lambda} \\ \frac{1}{2+\lambda} \end{bmatrix}$

$x^T A x = 1 \rightarrow \begin{bmatrix} \frac{1}{2+\lambda} & \frac{1}{2+\lambda} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2+\lambda} \\ \frac{1}{2+\lambda} \end{bmatrix} = 1$

$\rightarrow 2\left(\frac{1}{2+\lambda}\right)^2 = 1 \rightarrow \frac{1}{2+\lambda} = \frac{\sqrt{2}}{2}$

$\rightarrow x = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$

B) $x = (I + \lambda A)^{-1} x_0 = \begin{pmatrix} 1+2\lambda & 1+3\lambda \\ 1+\lambda & 1+5\lambda \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$= \begin{pmatrix} \frac{+2}{+7\lambda+3} \\ \frac{+1}{+7\lambda+3} \end{pmatrix}$

$x^T A x = 1 \rightarrow \begin{bmatrix} \frac{+2}{7\lambda+3} & \frac{+1}{7\lambda+3} \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \frac{+2}{7\lambda+3} \\ \frac{+1}{7\lambda+3} \end{bmatrix} = 1$

$21\left(\frac{+1}{7\lambda+3}\right)^2 = 1 \rightarrow x = \begin{bmatrix} \frac{2\sqrt{21}}{21} \\ \frac{\sqrt{21}}{21} \end{bmatrix}$