پاسخ تمرین ۲ یادگیری ماشین

امین اسدی

پاییز ۱۳۹۹

سوال اول) (۱۰ نمره)

1

$$D_{Red} = [(1,1),(1,2),(2,1),(2,2),(2,3),(4,3)]$$

$$D_{Blue} = [(-1,-1),(-1,-2),(-2,-1),(3,1),(-3,-2)]$$

$$\mu_{Red} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \Sigma_{Red} = \begin{bmatrix} 1.2 & 0.6 \\ 0.6 & 0.8 \end{bmatrix} \qquad \underline{ddof = 1}$$

$$\mu_{Blue} = \begin{bmatrix} -2 \\ -1 \end{bmatrix} \quad \Sigma_{Blue} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1.5 \end{bmatrix}$$

$$P_{(Red)} = \frac{n_{Red}}{n_{Red}+n_{Blue}} = \frac{6}{6+5} = \frac{6}{11} \qquad P_{Blue} = 1 - P_{Red} = \frac{5}{11}$$

$$g_i(x) = \frac{-1}{2}(x-\mu)^T \Sigma_i^{-1} (x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{\ln|\Sigma_i|}{2} + \ln P_{(w_i)}$$

Gaussian $\qquad i \in \{Red, Blue\}$

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

$$W_i = \frac{-1}{2}\Sigma^{-1} \qquad \Sigma_{Red}^{-1} = \begin{bmatrix} 1.33 & -1 \\ -1 & 2 \end{bmatrix}$$

$$w_i = \Sigma_i^{-1}\mu_i$$

$$w_{i0} = \frac{-1}{2}\mu_i^T \Sigma_i^{-1}\mu_i - \frac{\ln|\Sigma_i|}{2} + \ln P_{(\omega_i)} \qquad \Sigma_{Blue}^{-1} = \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 0.8 \end{bmatrix}$$

$$W_{Red} = \begin{bmatrix} -0.67 & 0.5 \\ 0.5 & -1 \end{bmatrix} \qquad W_{Blue} = \begin{bmatrix} -0.6 & -0.2 \\ -0.2 & -0.4 \end{bmatrix}$$

$$w_{Red} = \begin{bmatrix} 0.67 \\ 2 \end{bmatrix} \qquad w_{Blue} = \begin{bmatrix} -2.8 \\ -1.6 \end{bmatrix}$$

$$w_{Red,0} = -3.01 \qquad w_{Blue,0} \approx -4.5$$

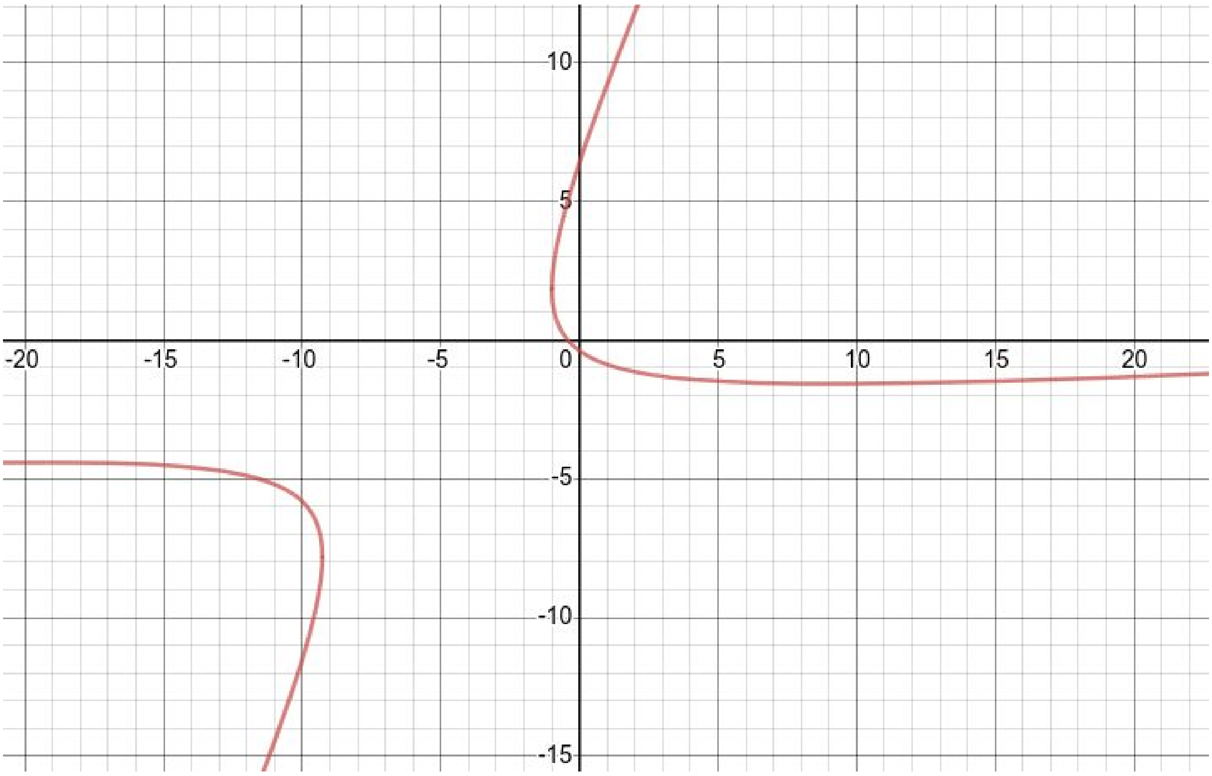
desision boundary $\rightarrow g_{Red}(x) = g_{Blue}(x)$

$$\rightarrow [x\ y]\begin{bmatrix} -0.67 & 0.5 \\ 0.5 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + [0.67\ \ 2]\begin{bmatrix} x \\ y \end{bmatrix} - 3.01 = [x\ y]\begin{bmatrix} -0.6 & -0.2 \\ -0.2 & -0.4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + [-2.8\ -1.6]\begin{bmatrix} x \\ y \end{bmatrix} - 4.5$$

$$\rightarrow [x\ y] \begin{bmatrix} -0.07 & 0.7 \\ 0.7 & -0.6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + [3.47\ 3.6] \begin{bmatrix} x \\ y \end{bmatrix} + 1.5 = 0$$

$$\rightarrow -0.07x^2 + 1.4xy - 0.6y^2 + 3.47x + 3.6y + 1.5 = 0$$

نمودار مرز تصمیم گیری:



سوال دوم) (۱۰ نمره)

2 الف)

$$P(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

$$\rightarrow P\left(\begin{bmatrix} 2 \\ 0.5 \\ 3 \end{bmatrix}\right) = \underline{0.002}$$

$$x-\mu = \begin{bmatrix} 2 \\ 0.5 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1.5 \\ 2 \end{bmatrix}$$

$$\Sigma^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.13 & -0.05 \\ 0 & -0.05 & 0.22 \end{bmatrix}$$

ب)

سوال سوم) (۱۵ نمره)

3 الف)

$$\omega_i \to \text{minimum risk} \to R_i < R_j \quad \text{for all } j$$

$$R_i = \sum_{k=1}^{c} \lambda_s P(\omega_k|x) = \sum_{\substack{k=1 \\ k\neq j}}^{c} \lambda_s P(\omega_k|x) + \lambda_s P(\omega_j|x)$$

$$\underset{j\neq i}{R_j} = \sum_{\substack{k=1 \\ k\neq i}}^{c} \lambda_s P(\omega_k|x) + \lambda_s P(\omega_i|x)$$

$$R_i - R_j < 0 \to \lambda_s P(\omega_j|x) - \lambda_s P(\omega_i|x) < 0$$

$$\to \underbrace{P(\omega_i|x) > P(\omega_j|x)}$$

$$\underset{w\geq c+1}{R_w} = \sum_{k=1}^{c} \lambda_r P(\omega_k|x) = \lambda_r \sum_{k=1}^{c} P(\omega_k|x) \overset{\sum_{k=1}^{c} P(\omega_k|x)=1}{=\!=} \lambda_r$$

$$R_i < R_w \to \sum_{\substack{k=1 \\ k\neq i}}^{c} \lambda_s P(\omega_k|x) < \lambda_r \to \lambda_s \sum_{\substack{k=1 \\ k\neq i}}^{c} P(\omega_k|x) < \lambda_r$$

$$\to \lambda_s (1 - P(\omega_i|x)) < \lambda_r \to \underbrace{P(\omega_i|x) \geq 1 - \frac{\lambda_r}{\lambda_s}}$$

ب)

$$g_i(x) = \begin{cases} P(\omega_i|x)\, P(x) \\ \left(1 - \frac{\lambda_r}{\lambda_s}\right) \sum_{j=1}^{c} P(\omega_j|x)\, P(x) \end{cases} = \begin{cases} P(\omega_i|x)\, P(x) & 1\leq i\leq c \\ P(x)\left(1 - \frac{\lambda_r}{\lambda_s}\right) \underbrace{\cancel{\sum_{j=1}^{c} P(\omega_j|x)}}_{1} & i\geq c+1 \end{cases}$$

درصورتی که داشته باشیم $1\leq i,j\leq c$ و بخواهیم g_i و g_j را مقایسه کنیم $P(\omega_i|x)P(x)$ و $P(\omega_j|x)P(x)$ مقایسه خواهند شد. پس برای انتخاب کلاس ω_i:
باید $\underbrace{P(\omega_i|x) > P(\omega_j|x)}$ باشد (به ازای هر j). همچنین اگر $1\leq i\leq c$ و $j\geq c+1$ و آنگاه
برای انتخاب کلاس ω_i باید داشته باشیم: $P(x)\left(1-\frac{\lambda_r}{\lambda_s}\right) < P(\omega_i|x)P(x) \to \underbrace{P(\omega_i|x) > 1-\frac{\lambda_r}{\lambda_s}}$
در نتیجه طبق بخش الف دو شرط برقرار هستند و انتخاب کلاس ω_i بهینه است.

ج) درصورتی که $\frac{\lambda_r}{\lambda_s} = 1$ باشد مقدار $g_i(x)$ برای $i \geq c+1$ برابر صفر خواهد شد در نتیجه

این طبقه‌بند هیچ‌گاه نقطه‌ای را رد نمی‌کند چون $\underset{1 \leq i \leq c}{g_i(x)} \geq \underset{j \geq c+1}{g_j(x)} = 0$

درصورتی که $\frac{\lambda_r}{\lambda_s} = 0$ باشد مقدار $g_i(x)$ برای $i \geq c+1$ برابر $P(x)$ خواهد شد در نتیجه:

$$\underset{1 \leq i \leq c}{g_i(x)} = P(x|\omega_i)P(x) \leq P(x) = \underset{j \geq c+1}{g_j(x)}$$

پس همیشه این طبقه‌بند نقطه‌ها را رد خواهد کرد.

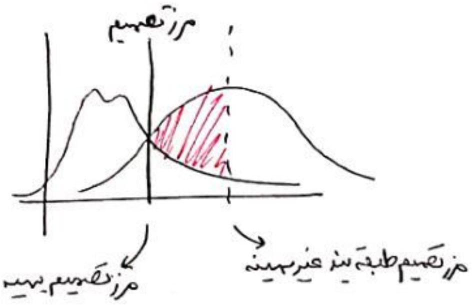
هرچه از نسبت ۰ به ۱ نزدیک شویم احتمال رد شدن نقطه کاهش می‌یابد.

سوال پنجم) (۵ نمره)

سوال ۵)

الف) هرچه نسبت $\frac{\lambda_{12}}{\lambda_{21}}$ بیشتر باشد یعنی اینکه هزینه اینکه کلاس واقعی 2 باشد ولی ما 1 انتخاب کنیم زیاد است بنابراین مرز تصمیم گیری به سمت کوچکتر کردن ناحیه انتخاب کلاس 1 جابجا می شود.

ب)



ناحیه‌ی قرمز رنگ reducible error را نشان می دهد که ناحیه‌ای است که با جابجایی مرز تصمیم می توان آن را کوچکتر کرد و به طبقه بند optimal نزدیکتر شد.

ج) می خواهیم بسنجیم که دو توزیعی که برای حضور و عدم حضور یک محرک خارجی در نظر گرفته ایم چقدر از هم فاصله دارند که هرچه متمایزتر باشد طبقه بندی بهینه تر است. این تمایز را می توان با معیاری مثل زیر اندازه گرفت:

$$d' = \frac{|\mu_2 - \mu_1|}{\sigma}$$

تمایزپذیری

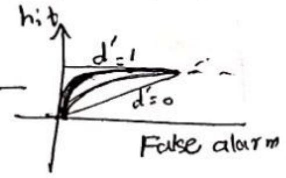
فاصله میانگین دو توزیع

واریانس دو توزیع

اگر واریانس ثابت باشد هر چه میانگین ها از هم دورتر باشند، کلاس ها تمایز بیشتری از هم دارند. ← مشهود

د) در حالتی که میانگین توزیع داده ها را نمی‌دانیم ولی state of nature و تصمیم سیستم را می دانیم و می خواهیم d' را بدست آوریم می توانیم تعداد زیادی آزمون و خطا انجام دهیم و hit rate و false rate را بدست آوریم. حال اگر density را ثابت بگیریم و threshold را تغییر دهیم ← hit و false alarm نیز تغییر می کنند و به ازای هر d' یک خم بدست می آید و خم های مربوط به d' های بیشتر، بالاتر خواهند بود.

صعودی اکید ←



۵) generative approach : در این روش تابع احتمال توأم بردار ویژگی و کلاس را بدست آوریم و از روی آن می توانیم احتمال شرطی و احتمال های مارجینال را هم بدست آوریم که این روش نیاز به تعداد نمونه بسیار زیاد است

اما در حالت Discriminate : در این رویکرد به صورت مستقیم خود توابع احتمال شرطی و مارجینال را تخمین بزنیم که تعداد نمونه کمتری می خواهد و عملی تر است.

سوال ششم (۲۰ نمره)

Data = [(-5,5), (-5,0), (-5,-5), (0,0), (0,5), (0,-5), (8,3), (8,-3)]

Point = (5,0)

فاصله ها:

- d_1 :
  - (-5,5) : 10
  - (-5,0) : 10
  - (-5, -5) : 10
  - (0,0) : 5
  - (0,5) : 5
  - (0,-5) : 5
  - (8,3) : 3
  - (8,-3) : 3

کمترین فاصله مربوط به ۲ نقطه آخر است. هر دو از کلاس قرمز هستند.

پس کلاس پیشبینی شده قرمز است.

- d_2 :
  - (-5,5) : 15
  - (-5,0) : 10
  - (-5, -5) : 15
  - (0,0) : 5
  - (0,5) : 10
  - (0,-5) : 10
  - (8,3) : 6
  - (8,-3) : 6

کمترین فاصله مربوط به نقطه (0,0) است که از کلاس آبی است.

پس کلاس پیشبینی شده آبی است.

- d_3 :
    - (-5,5)   : 125
    - (-5,0)   : 100
    - (-5, -5) : 125
    - (0,0)   : 25
    - (0,5)   : 50
    - (0,-5)   : 50
    - (8,3)   : 18
    - (8,-3)   : 18

کمترین فاصله مربوط به ۲ نقطه آخر است. هر دو از کلاس قرمز هستند.

پس کلاس پیشبینی شده قرمز است.

ب)

$x = [1.5,\ 1.5]$

$x - \mu_1 = [1.5 - 0,\ 1.5 - 0] = [1.5,\ 1.5]$

$\Sigma^{-1} = [[2.33\ 2],\ [2,\ 2]]$

$\text{mahalonobis } 1 = (x - \mu_1)^T . \Sigma^{-1} . (x - \mu_1) = 4.33$

$x = [1.5,\ 1.5]$

$x - \mu_2 = [1.5 - 0,\ 1.5 - 0] = [1.5,\ 1.5]$

$\Sigma^{-1} = [[2.33\ -2],\ [-2,\ 2]]$

$\text{mahalonobis } 2 = (x - \mu_2)^T . \Sigma^{-1} . (x - \mu_2) = 0.68$

ج)

$$x = \begin{bmatrix} w \\ z \end{bmatrix} \qquad \forall x , \quad x^T \begin{bmatrix} 5 & \alpha \\ \alpha & 4 \end{bmatrix} x > 0$$

$$\rightarrow [w \ z] \begin{bmatrix} 5 & \alpha \\ \alpha & 4 \end{bmatrix} \begin{bmatrix} w \\ z \end{bmatrix} > 0 \rightarrow [5w + \alpha z \quad \alpha w + 4z] \begin{bmatrix} w \\ z \end{bmatrix} > 0$$

$$\rightarrow 5w^2 + \underbrace{\alpha w z + w\alpha z}_{2\alpha z w} + 4z^2 > 0$$

$$\rightarrow \Delta \leq 0 \rightarrow \cancel{4}\alpha^2 z^2 - \cancel{4} \times 5 \times 4z^2 \leq 0$$

$$\rightarrow \alpha^2 \leq 20 \rightarrow |\alpha| \leq \sqrt{20}$$

عمل انتخاب w_1

$$R(\alpha_1|x) = \lambda_{11} P(w_1|x) + \lambda_{12} P(w_2|x)$$ د)

$$R(\alpha_2|x) = \lambda_{21} P(w_2|x) + \lambda_{22} P(w_2|x)$$

decide w_1

$$\longrightarrow R(\alpha_1|x) < R(\alpha_2|x) \qquad P(w_1) = 1 - P(w_2) = \frac{3}{4}$$

$$\rightarrow \frac{P(x|w_1)}{P(x|w_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(w_2)}{P w_1}$$

$$\rightarrow \frac{x + \frac{1}{2}}{\cancel{\frac{3}{4}}(x^2+1)} > \frac{\overset{1}{\cancel{2-1}}}{\underset{2}{\cancel{3-1}}} \frac{\frac{1}{4}}{\cancel{\frac{3}{4}}} \rightarrow 8(x + \tfrac{1}{2}) > x^2 + 1$$

$$\rightarrow 8x + 4 > x^2 + 1 \rightarrow x^2 - 8x - 3 < 0$$

$$\rightarrow \frac{8 - \sqrt{76}}{2} < x < \frac{8 + \sqrt{76}}{2}$$

کد و گزارش سوالات ۷ تا ۱۱ در فایل codes.ipynb قرار دارد.