به نام خدا

تمرین سری سوم، فصل دوم کتاب مرجع – الگوشناسی آماری

سوال ۷)

a)

$$\int_{-\infty}^{+\infty} P(x \mid \omega_i)dx = \frac{1}{\pi b}\int_{-\infty}^{+\infty} \frac{dx}{1+(\frac{x-a_i}{b})^2}$$

$$u = \frac{x-a_i}{b},\ du = \frac{dx}{b}$$

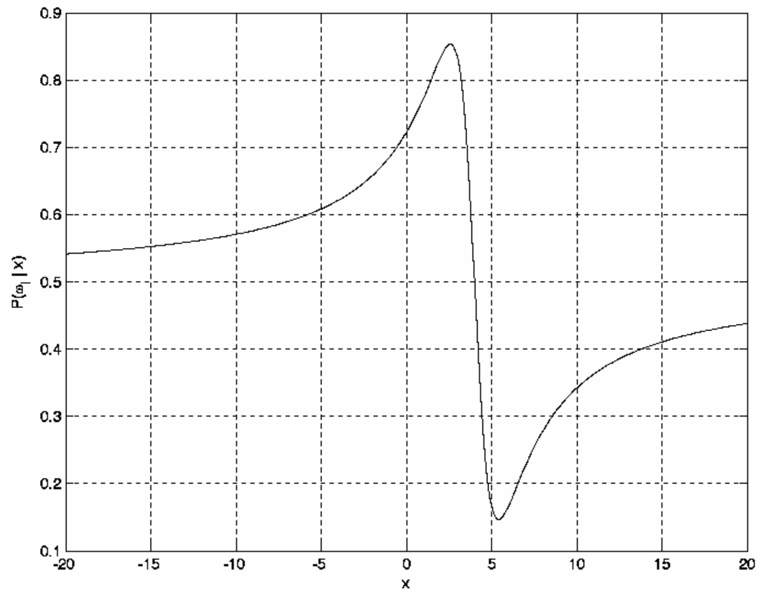
$$\int_{-\infty}^{+\infty} P(x \mid \omega_i)dx = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{dx}{1+u^2} = \frac{1}{\pi}(\tan^{-1} u)_{-\infty}^{+\infty} = 1$$

b)

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

c)

$$P(\omega_1 \mid x) = \frac{P(x \mid \omega_1)P(\omega_1)}{P(x \mid \omega_1)P(\omega_1) + P(x \mid \omega_2)P(\omega_2)}$$



d) هر دو به سمت 0.5 نزدیک می شوند. دلیل این مسئله این است که likelihood مربوط به x در بینهایت به صفر نزدیک می شود و بنابراین، احتمال هر کلاس برابر احتمال پیشین آن کلاس می شود.

سوال ۸)

a)

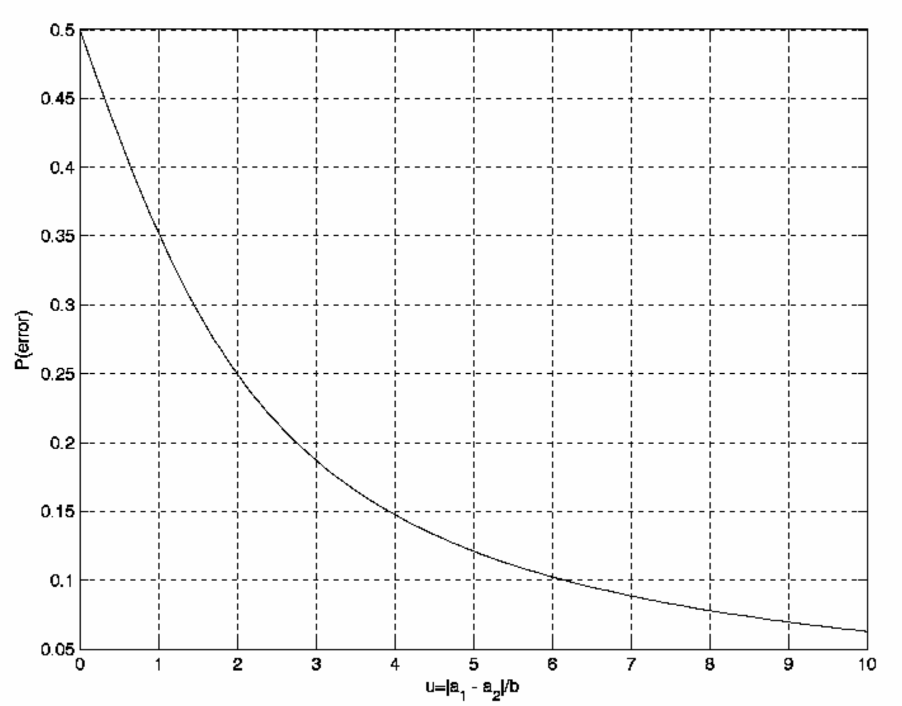
$$P(error) = \int_{R_1} P(\omega_2 \mid x)P(x)dx + \int_{R_2} P(\omega_1 \mid x)P(x)dx$$

$$a_1 > a_2$$

$$P(error) = \int_{-\infty}^{\frac{a_1+a_2}{2}} \frac{0.5}{\pi b} \frac{dx}{1+(\frac{x-a_1}{b})^2} + \int_{\frac{a_1+a_2}{2}}^{+\infty} \frac{0.5}{\pi b} \frac{dx}{1+(\frac{x-a_1}{b})^2}$$

$$P(error) = \frac{0.5}{\pi}(\tan^{-1} u|_{-\infty}^{\frac{a_1+a_2}{2}} + \tan^{-1} u\,|_{\frac{a_1+a_2}{2}}^{+\infty}) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_1-a_2}{2b}\right|$$

b)



c) وقتی $a_1 = a_2$ باشد، خطا برابر 0.5 خواهد بود. در این حالت تمام x ها را به یک کلاس نسبت می دهیم. چون احتمال پیشین کلاس ها 0.5 است، میزان خطا برابر 0.5 خواهد شد.

سوال ۹)

a)

$$P(error)=\int_{R_1}P(\omega_2\mid x)P(x)dx+\int_{R_2}P(\omega_1\mid x)P(x)dx$$
$$P(error)=\int_{-\infty}^{\theta}P(x\mid\omega_1)P(\omega_1)dx+\int_{\theta}^{+\infty}P(x\mid\omega_2)P(\omega_2)dx$$

b)

$$P(error)=\int_{-\infty}^{\theta}P(x\mid\omega_1)P(\omega_1)dx-\int_{+\infty}^{\theta}P(x\mid\omega_2)P(\omega_2)dx$$
$$\frac{\partial P(error)}{\partial\theta}=P(\theta\mid\omega_1)P(\omega_1)-P(\theta\mid\omega_2)P(\omega_2)=0$$
$$\Rightarrow P(\theta\mid\omega_1)P(\omega_1)=P(\theta\mid\omega_2)P(\omega_2)$$

c) خیر، چون ممکن است این معادله چند جواب برای θ داشته باشد. مثلا دو تابع Gaussian که میانگین یکسان و واریانس متفاوتی دارند، دو جواب به ازاء θ بدست می آید.

d) حالتی را در نظر بگیرید که $P(\omega_1|x)$ یک تابع Gaussian با میانگین 1- و واریانس 1 و $P(\omega_2|x)$ هم یک تابع Gaussian با میانگین 1 و واریانس 1 باشد، در این صورت انتخاب $\theta = 0$ ، خطا را بیشینه می کند.

سوال ۱۱)

a) برای میزان ریسک یک جداکننده داریم :

$$R=\sum_{i=1}^{a}\int_{R_i}R(\alpha_i\mid x)P(x)dx$$

چون نواحی R_i به صورت تصادفی انتخاب می شوند، مقدار R یک مقدار تصادفی خواهد بود. لذا مقدار امید ریاضی آن را پیدا می کنیم :

$$R=\int[\sum_{i=1}^{a}R(\alpha_i\mid x)P(\alpha_i\mid x)]P(x)dx$$

b) داریم :

$$R(x)=\sum_{i=1}^{a}P(\alpha_i\mid x)R(\alpha_i\mid x)$$

فرض کنید به ازاء i=j کمترین ریسک باشد، در آن صورت اگر احتمال انتخاب عمل j ام غیر ۱ باشد، عمل دیگری چون k وجود دارد، که ریسک آن بزرگتر از ریسک عمل j ام است و احتمال انتخاب آن عمل غیر صفر است. اگر این احتمال انتخاب این عمل را صفر کنیم و به احتمال انتخاب عمل j ام اضافه کنیم، میزان ریسک کاهش می یابد. لذا باید ریسک عمل j ام یک باشد.

c) بله، ممکن است. فرض کنید قانونی که انتخاب شده، کمترین ریسک را انتخاب نکند، در آن صورت با پایین آوردن احتمال اعمال دیگر و بالا بردن احتمال عمل متناظر با کمترین ریسک، ریسک کلی کمتر می شود.

سوال ۱۳)

$$R(\alpha_i \mid x) = \sum_{j=1, j\neq i}^{c} \lambda_s P(\omega_j \mid x) = \lambda_s (1 - P(\omega_i \mid x)), \quad i \neq c+1$$
$$R(\alpha_{c+1} \mid x) = \lambda_r$$
$$\alpha(x) = \alpha_i \Leftrightarrow R(\alpha_i \mid x) = \min_j \{R(\alpha_j \mid x)\}$$
$$\Leftrightarrow \begin{cases} \lambda_s (1 - P(\omega_i \mid x)) \leq \lambda_c (1 - P(\omega_j \mid x)) \Leftrightarrow P(\omega_i \mid x) \geq P(\omega_j \mid x) \\ \lambda_s (1 - P(\omega_i \mid x)) \leq \lambda_r \Leftrightarrow P(\omega_i \mid x) \geq 1 - \dfrac{\lambda_r}{\lambda_s} \end{cases}$$

اگر $\lambda_r=0$ حالت بهینه حالتی است که همواره reject می کند، چون در آن صورت ریسک برابر صفر است. ولی اگر $\lambda_r/\lambda_s>1$ در آنصورت هیچگاه reject نخواهیم کرد چون هر کدام از اعمال را که انتخاب کنیم ریسک کمتری نسبت به حالت reject خواهیم داشت.

سوال ۲۷)

$$-\frac{1}{2}(x-\mu_1)^t \Sigma^{-1} (x-\mu_1) + \ln P(\omega_1) = -\frac{1}{2}(x-\mu_2)^t \Sigma^{-1} (x-\mu_2) + \ln P(\omega_2)$$
$$\frac{1}{2} x^t \Sigma^{-1} \mu_1 + \frac{1}{2} \mu_1^t \Sigma^{-1} x + \ln P(\omega_1) = \frac{1}{2} x^t \Sigma^{-1} \mu_2 + \frac{1}{2} \mu_2^t \Sigma^{-1} x + \ln P(\omega_2)$$
$$\mu_1^t \Sigma^{-1} x + \ln P(\omega_1) = \mu_2^t \Sigma^{-1} x + \ln P(\omega_2)$$
$$(\mu_1 - \mu_2)^t \Sigma^{-1} x + \ln \frac{P(\omega_1)}{P(\omega_2)} = 0$$

معادله اخیر، معادله یک ابرصفحه در فضای x است. برای اینکه اتفاق مورد نظر بیافتد باید هر دو میانگین ها در یک طرف این ابرصفحه قرار بگیرند. لذا داریم :

$$\begin{cases}(\mu_1-\mu_2)^t\Sigma^{-1}\mu_1+\ln\dfrac{P(\omega_1)}{P(\omega_2)}\geq 0,(\mu_1-\mu_2)^t\Sigma^{-1}\mu_2+\ln\dfrac{P(\omega_1)}{P(\omega_2)}\geq 0\\(\mu_1-\mu_2)^t\Sigma^{-1}\mu_1+\ln\dfrac{P(\omega_1)}{P(\omega_2)}\leq 0,(\mu_1-\mu_2)^t\Sigma^{-1}\mu_2+\ln\dfrac{P(\omega_1)}{P(\omega_2)}\leq 0\end{cases}$$

$$\begin{cases}\dfrac{P(\omega_1)}{P(\omega_2)}\geq \max\{\exp\{(\mu_2-\mu_1)^t\Sigma^{-1}\mu_1\},\exp\{(\mu_2-\mu_1)^t\Sigma^{-1}\mu_2\}\}\\\dfrac{P(\omega_1)}{P(\omega_2)}\leq \min\{\exp\{(\mu_2-\mu_1)^t\Sigma^{-1}\mu_1\},\exp\{(\mu_2-\mu_1)^t\Sigma^{-1}\mu_2\}\}\end{cases}$$

هر کدام از دو شرط فوق که برآورده شود، این اتفاق می افتد.

سوال ۲۸)

a)

$$\sigma_{ij}=E[(x_i-\mu_i)(x_j-\mu_j)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x_i-\mu_i)(x_j-\mu_j)P(x_i,x_j)dx_idx_j$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x_i-\mu_i)(x_j-\mu_j)P(x_i)P(x_j)dx_idx_j=\int_{-\infty}^{+\infty}(x_i-\mu_i)P(x_i)dx_i\int_{-\infty}^{+\infty}(x_j-\mu_j)P(x_j)dx_j=0$$

b) حالتی را در نظر می گیریم که دو متغیر به صورت Joint دارای توزیع Gaussian باشند :

$$P(x_i,x_j)=\frac{1}{2\pi|\Sigma|^{1/2}}\exp\{-\frac{1}{2}\left(x_i-\mu_i,x_j-\mu_j\right)\Sigma^{-1}\begin{pmatrix}x_i-\mu_i\\x_j-\mu_j\end{pmatrix}\}$$

$$\Sigma=\begin{bmatrix}\sigma_i^2&\sigma_{ij}\\\sigma_{ij}&\sigma_i^2\end{bmatrix}=\begin{bmatrix}\sigma_i^2&0\\0&\sigma_i^2\end{bmatrix},\Sigma^{-1}=\begin{bmatrix}1/\sigma_i^2&0\\0&1/\sigma_i^2\end{bmatrix}$$

$$P(x_i,x_j)=\frac{1}{2\pi\sigma_i\sigma_j}\exp\{-\frac{1}{2}(\frac{x_i-\mu_i}{\sigma_i})^2-\frac{1}{2}(\frac{x_j-\mu_j}{\sigma_j})^2\}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\{-\frac{1}{2}(\frac{x_i-\mu_i}{\sigma_i})^2\}\frac{1}{\sqrt{2\pi}\sigma_j}\exp\{-\frac{1}{2}(\frac{x_j-\mu_j}{\sigma_j})^2\}=P(x_i)P(x_j)$$

c) فرض کنید X یک متغیر تصادفی با توزیع نرمال با میانگین صفر باشد. در آنصورت قرار می دهیم $Y=X^2$. داریم :

$$E[Y] = E[X^2] = \sigma^2$$

$$\sigma_{ij} = E[(X-0)(X^2-\sigma^2)] = E[X^3] - \sigma^2 E[X] = E[X^3]$$

$$E[X^3] = \int_{-\infty}^{+\infty} x^3 N(0,\sigma^2)dx = 0 \Rightarrow \sigma_{ij} = 0$$

دلیل صفر بودن انتگرال فوق، فرد بودن تابع زیر انتگرال است.

سوال ۴۲)

a) احتمال اینکه مولفه i ام در یک نمونه در کلاس j ام برابر یک باشد، p_{ij} است.

b)

$$P(\omega_j \mid x_1, x_2, \cdots, x_d) \propto P(\omega_j)\prod_{i=1}^{d} p_{ij}^{x_i}(1-p_{ij})^{1-x_i}$$

$$\ln P(\omega_j \mid x_1, x_2, \cdots, x_d) = \ln P(\omega_j) + \sum_{i=1}^{d} x_i \ln p_{ij} + (1-x_i)\ln(1-p_{ij})$$

$$= \ln P(\omega_j) + \sum_{i=1}^{d} x_i \ln\frac{p_{ij}}{1-p_{ij}} + \sum_{i=1}^{d}\ln(1-p_{ij})$$

$$x \in \omega_j \Leftrightarrow \forall k \neq j : P(\omega_j \mid x_1, x_2, \cdots, x_d) \geq P(\omega_k \mid x_1, x_2, \cdots, x_d)$$

$$\Leftrightarrow \ln P(\omega_j \mid x_1, x_2, \cdots, x_d) \geq \ln P(\omega_k \mid x_1, x_2, \cdots, x_d)$$

$$g_j(x) = \ln P(\omega_j \mid x_1, x_2, \cdots, x_d) = \ln P(\omega_j) + \sum_{i=1}^{d} x_i \ln\frac{p_{ij}}{1-p_{ij}} + \sum_{i=1}^{d}\ln(1-p_{ij})$$

با تشکر از آقای محمد حسین رهبان