112753207 資碩計一 張詠軒

2-9 (a.) $P(error) = \int_{-\infty}^{(a_1+a_2)/2} P(w_2|x)\,dx + \int_{(a_1+a_2)/2}^{\infty} P(w_1|x)\,dx$

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

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

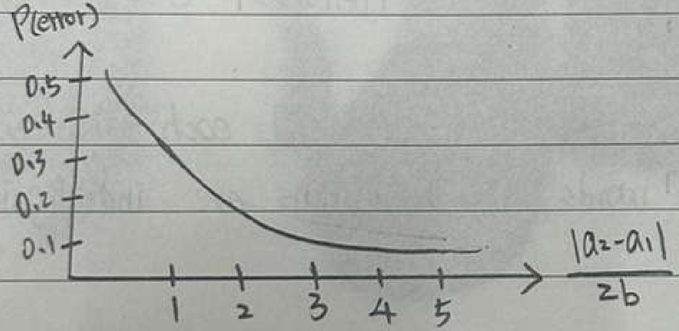
$$= \frac{1}{\pi}\int_{-\infty}^{(a_1+a_2)/2} \frac{1}{1+y^2}\,dy$$

$y = \frac{x-a_2}{b}$

$$\Rightarrow P(error) = \frac{1}{\pi}\left[\tan^{-1}\left|\frac{a_1-a_2}{2b}\right| - \tan^{-1}[-\infty]\right]$$

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

(b.)



(C). The maximum value of the probability of error: $P_{max}\left(\frac{a_2-a_1}{2b}\right) = \frac{1}{2}$ which occures for $\left|\frac{a_2-a_1}{b}\right| = 0$, this occures when either the two distributions are the same, which can happen because $a_1 = a_2$, or even if $a_1 \neq a_2$ because $b = \infty$ and both distributions are flat.

2-12. $\sum_{i=1}^{c} P(\omega_i/x)=1$

(a.) if $P(\omega_i/x)=P(\omega_j/x)$ for all i and j, then $P(\omega_i/x)=\frac{1}{c}$ and hence $P(\omega_{max}/x)=\frac{1}{c}$, if one of the $P(\omega_i/x)<\frac{1}{c}$, then by our normalization condition we must have that $P(\omega_{max}/x)>\frac{1}{c}$

(b.) $P(error)=1-\int P(\omega_{max}/x)\,p(x)\,dx$

(c.) $P(error)=1-\int \underbrace{P(\omega_{max}/x)}_{=g\geq\frac{1}{c}}\,p(x)\,dx$

$=1-g\int p(x)\,dx=1-g \quad \therefore P(error)\leq 1-\frac{1}{c}=(c-1)/c$

(d.) all categories have the same prior probability and each distribution has the same form, in other words, the distributions are indistinguishable.

2-13. If we choose ω_{max}

$\Rightarrow \lambda_s \sum_{j\neq max} P(\omega_j/x)=\lambda_s[1-P(\omega_{max}/x)]$

if we reject, our risk is λ_r, if we choose a non-maximal category ω_k

$\Rightarrow \lambda_s \sum_{j\neq k} P(\omega_j|x)=\lambda_s[1-P(\omega_k|x)]\geq\lambda_s[1-P(\omega_{max}|x)]$

This last inequality shows that we should never decide on a category other than the one that has the maximum posterior probability, as we know from our Bayes analysis, Consequently, we should either choose ω_{max} or we should reject, depending upon which is smaller: $\lambda_s[1-P(\omega_{max}|x)]$ or λ_r

we reject if $\lambda_r\leq\lambda_s[1-P(\omega_{max}|x)]$, that is, if $P(\omega_{max}|x)\geq 1-\frac{\lambda_r}{\lambda_s}$

2-31. choose ω_1 if $P(\omega_1|x) > P(\omega_2|x)$; otherwise choose ω_2

(a.) $P(\omega_1) = P(\omega_2) = \frac{1}{2}$, $P(\omega_1|x) \propto p(x|\omega_1)$

therefore the decision boundary is:

choose ω_1 if $|x-\mu_1| < |x-\mu_2|$, otherwise choose ω_2

without loss of generality, we assume $\mu_1 < \mu_2$,

$\Rightarrow P(\text{error}) = P(|x-\mu_1| > |x-\mu_2| \mid \omega_1)P(\omega_1) + P(|x-\mu_2| > |x-\mu_1| \mid \omega_2)P(\omega_2)$

$$= \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-u^2/2} du$$

where $a = \dfrac{|\mu_2-\mu_1|}{2\sigma}$

(b.) $\lim\limits_{a\to\infty} \dfrac{1}{\sqrt{2\pi}a} e^{\frac{-a^2}{2}} = 0$

$$P_e = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{\frac{-u^2}{2}} du \leq \frac{1}{2\pi} \int_a^\infty \frac{u}{a} e^{\frac{-u^2}{2}} du$$

$$= \frac{1}{\sqrt{2\pi}a} e^{\frac{-a^2}{2}}$$

$\Rightarrow P_e$ goes to 0 as $a = \dfrac{|\mu_2-\mu_1|}{\sigma}$ goes infinity.

2.44.

choose w_k if $g_k(x) \geq g_j(x)$ for all $j \neq k$

use function : $g_j(x) = \ln p(x/w_j) + \ln P(w_j)$

$P(x|w_j) = P((x_1, \ldots, x_d)|w_j) = \prod_{i=1}^{d} P(x_i|x_j)$,

where $P_{ij} = P_r[x_i = 1 | w_j]$,

$q_{ij} = P_r[x_i = 0 | w_j]$,

$r_{ij} = P_r[x_i = -1 | w_j]$

$$\Rightarrow P(x_i|w_j) = p_{ij}^{\frac{1}{2}x_i + \frac{1}{2}x_i^2} q_{ij}^{1-x_i^2} r_{ij}^{-\frac{1}{2}x_i + \frac{1}{2}x_i^2}, \quad i = 1, \ldots, d$$
$$j = 1, \ldots, c$$

thus for the full vector x the conditional probability is

$$P(x|w_j) = \prod_{i=1}^{d} p_{ij}^{\frac{1}{2}x_i + \frac{1}{2}x_i^2} q_{ij}^{1-x_i^2} r_{ij}^{-\frac{1}{2}x_i + \frac{1}{2}x_i^2}$$

$$\Rightarrow g_j(x) = \ln P(x|w_j) + \ln P(w_j)$$
$$= \sum_{i=1}^{d} \left[\left(\frac{1}{2}x_i + \frac{1}{2}x_i^2\right) \ln p_{ij} + (1 - x_i^2) \ln q_{ij} + \left(-\frac{1}{2}x_i + \frac{1}{2}x_i^2 \ln r_{ij}\right) \right] + \ln P(w_j)$$

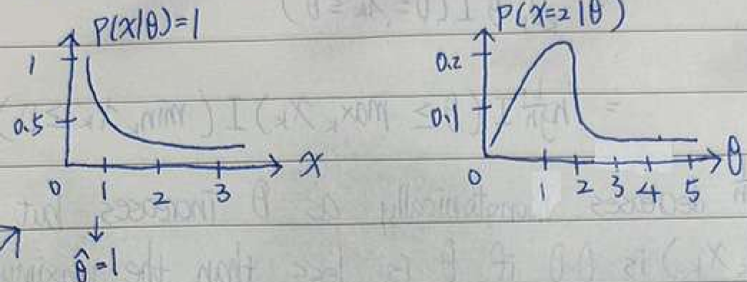
$$= \sum_{i=1}^{d} x_i^2 \ln \frac{\sqrt{p_{ij} r_{ij}}}{q_{ij}} + \frac{1}{2} \sum_{i=1}^{d} x_i \ln \frac{p_{ij}}{r_{ij}} + \sum_{i=1}^{d} \ln q_{ij} + \ln P(w_j)$$

which are quadratic functions of the components x_i.

3-1.

$$P(x|\theta) = \begin{cases} \theta e^{-\theta x}, & x \ge 0 \\ 0, & \text{otherwise} \end{cases}$$

(a.) $P(x=2|\theta)$ is not maximized when $\theta = 2$ but instead for a valus less than 1.0.



(b.) log-likelihood function is

$$l(\theta) \sum_{k=1}^{n} \ln p(x_k|\theta) = \sum_{k=1}^{n} [\ln\theta - \theta x_k] = n\ln\theta - \theta \sum_{k=1}^{n} x_k$$

$$\nabla_\theta l(\theta) = 0, \quad \nabla_\theta l(\theta) = \frac{\partial}{\partial\theta}[n\ln\theta - \theta \sum_{k=1}^{n} x_k]$$

$$= \frac{n}{\theta} - \sum_{k=1}^{n} x_k = 0$$

Maximum-likelihood : $\hat{\theta} = \dfrac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k}$

(c.) approximate the mean : $\frac{1}{n}\sum_{k=1}^{n} x_n$

by the integral: $\int_0^\infty x p(x) dx$

$\Rightarrow \int_0^\infty x e^{-x} dx = 1$

3-2. $P(x|\theta) = \begin{cases} \frac{1}{\theta} &, 0 \le x \le \theta \\ 0 &, \text{otherwise} \end{cases}$

(a) use the notation of an indicator function $I(\cdot)$

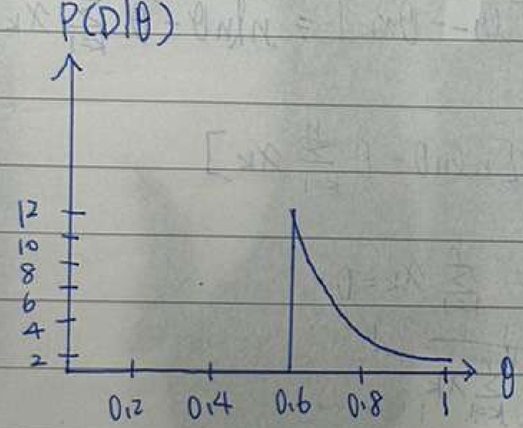
$$P(D|\theta) = \prod_{k=1}^{n} p(x_k|\theta)$$

$$= \prod_{k=1}^{n} \frac{1}{\theta} I(0 \le x_k \le \theta)$$

$$= \frac{1}{\theta^n} I(\theta \ge \max_k x_k) I(\min_k x_k \ge 0)$$

we note that $\frac{1}{\theta^n}$ decreases monotonically as θ increases but also that $I(\theta \ge \max_k x_k)$ is 0.0 if θ is less than the maximum value of x_k, therefore, our likelihood function is maximized at $\hat{\theta} = \max_k x_k$

(b)



3-3

$$z_{ik} = \begin{cases} 1 & \text{, if the state of nature for the } k^{th} \text{ sample is } \omega_i \\ 0 & \text{, otherwise} \end{cases}$$

(a.) $\Pr[z_{ik}=1 \mid P(\omega_i)] = P(\omega_i)$

and $\Pr[z_{ik}=0 \mid P(\omega_i)] = 1-P(\omega_i)$

$\Rightarrow P(z_{ik} \mid P(\omega_i)) = [P(\omega_i)]^{z_{ik}}[1-P(\omega_i)]^{1-z_{ik}}$

$$P(z_{i1},\dots,z_{ik} \mid P(\omega_i)) = \prod_{k=1}^{n} P(z_{ik} \mid P(\omega_i))$$

$$= \prod_{k=1}^{n} [P(\omega_i)]^{z_{ik}}[1-P(\omega_i)]^{1-z_{ik}}$$

(b.). log-likelihood as function of $P(\omega_i)$

$$l(P(\omega_i)) = \ln P(z_{i1},\dots,z_{in} \mid P(\omega_i))$$

$$= \ln\left[\prod_{k=1}^{n}[P(\omega_i)]^{z_{ik}}[1-P(\omega_i)]^{1-z_{ik}}\right]$$

$$= \sum_{k=1}^{n}[z_{ik}\ln P(\omega_i) + (1-z_{ik})\ln(1-P(\omega_i))]$$

therefore, the maximum-likelihood values for the $P(\omega_i)$ must satisfy

$$\nabla_{P(\omega_i)} l(P(\omega_i)) = \frac{1}{P(\omega_i)}\sum_{k=1}^{n} z_{ik} - \frac{1}{1-P(\omega_i)}\sum_{k=1}^{n}(1-z_{ik}) = 0$$

$$(1-\hat{P}(\omega_i))\sum_{k=1}^{n} z_{ik} = \hat{P}(\omega_i)\sum_{k=1}^{n}(1-z_{ik})$$

$$\Rightarrow \sum_{k=1}^{n} z_{ik} = \hat{P}(\omega_i)\sum_{k=1}^{n} z_{ik} + n\hat{P}(\omega_i) - \hat{P}(\omega_i)\sum_{k=1}^{n} z_{ik}$$

$$\Rightarrow \hat{P}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n} z_{ik}$$

that is, the estimate of the probability of category ω_i is merely the probability of obtaining its indicatory value in the eraining data.

3-9.

$\hat{\theta} = \arg\max_{\theta} P(x|\theta)$

mapping $x \to T(x)$ where $T(\cdot)$ is continuous.

Then we can write $P(T|\theta)dT = P(x|\theta)dx$

$$P(T|\theta) = \frac{P(x|\theta)}{\frac{dT}{dx}}$$

Then we find the value of θ maximizing $P(T(x)|\theta)$ as

$$\arg\max_{\theta} P(T(x)|\theta) = \arg\max_{\theta} \frac{P(x|\theta)}{\frac{dT}{dx}}$$
$$= \arg\max_{\theta} P(x|\theta)$$
$$= \hat{\theta}$$

where we have assumed $\frac{dT}{dx} \neq 0$ at $\theta = \hat{\theta}$

the maximum-likelihood value of $T(\theta)$ is indeed $\hat{\theta}$.

We must check whether the value of $\hat{\theta}$ derived this way gives a maximum or minimum for $P(T|\theta)$

3-11. We assume $P_2(x) \equiv P(x|\omega_2) \sim N(\mu, \Sigma)$ but that $P_1(x) \equiv P(x|\omega_1)$ is arbitrary.

The kullback-Leibler divergence from $P_1(x)$ to $P_2(x)$ is

$$D_{KL}(P_1, P_2) = \int P_1(x) \ln P_1(x) dx + \frac{1}{2} \int P_1(x) [d \ln(2\pi) + \ln|\Sigma| + (x-\mu)^t \Sigma^{-1} (x-\mu)] dx$$

where we used the fact that P_2 is a Gaussian,

$$P_2(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[\frac{(x-\mu)^t \Sigma^{-1} (x-\mu)}{2} \right]$$

now seek μ and Σ to minimize this distence, we set the derivative to zero.

$$\frac{\partial}{\partial \mu} D_{KL}(P_1, P_2) = -\int \Sigma^{-1} (x-\mu) P_1(x) dx = 0,$$

and this implies.

$$\Sigma^{-1} \int P_1(x)(x-\mu) dx = 0$$

we assume Σ is non-singular

$\Rightarrow \int P_1(x)(x-\mu) dx = \mathcal{E}_1[x-\mu] = 0, \quad \mathcal{E}_1[x] = \mu$

The mean of the second distribution should be the same as that of the Gaussian.

Turn to the covariance of the second distribution, we denote $A = \Sigma$, we take a derivative of the kullback-Leibler divergence,

$$\frac{\partial}{\partial A} D_{KL}(P_1, P_2) = 0 = \int P_1(x) [-A^{-1} + (x-\mu)(x-\mu)^t] dx,$$

and thus, $\mathcal{E}_1[\Sigma - (x-\mu)(x-\mu)^t]$, or $\mathcal{E}_1[(x-\mu)(x-\mu)^t] = \Sigma$

$\frac{\partial |A|}{\partial A} = |A| A^{-1}$, we relied on the fact that $A = \Sigma^{-1}$ is symmetric

since Σ is a covariance matrix, More generally, for an arbitrary non-singular matrix $\Rightarrow \frac{\partial |M|}{\partial M} = |M| (M^{-1})^t$