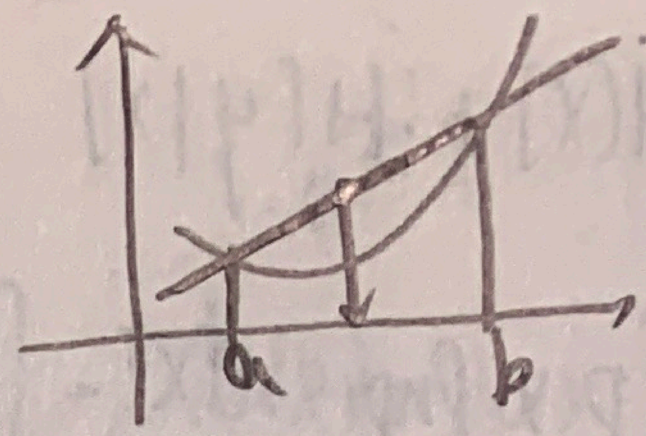
1.

① convex function: $f(\lambda a+(1-\lambda)b) \le \lambda f(a)+(1-\lambda)f(b)$
for $\lambda > 0$



concave function: $f(\lambda a+(1-\lambda)b) \ge \lambda f(a)+(1-\lambda)f(b)$

+3

②

Jensen's inequality: $f\left(\sum_{i=1}^{M}\lambda_i x_i\right) \le \sum_{i=1}^{M}\lambda_i f(x_i)$

where $\lambda_i \ge 0$, $\sum_{i=1}^{M}\lambda_i = 1$.

+3

③

$$KL(p\|q) = -\int p(x)\ln q(x)dx - \left(-\int p(x)\ln p(x)\,dx\right)$$

$$= -\int p(x)\ln\left\{\frac{q(x)}{p(x)}\right\}dx$$

$$f(\lambda a+(1-\lambda)b) \le \lambda f(a)+(1-\lambda)f(b)$$

$$\rightarrow f(E[x]) \le E[f(x)]$$

$$\xrightarrow{\text{continue}} f\left(\int x p(x)dx\right) \le \int f(x)p(x)dx$$

$$KL(p\|q) = -\int p(x)\ln\left\{\frac{q(x)}{p(x)}\right\}dx \ge -\ln\int q(x)dx = 0 \quad ※$$

+4

④

mutual information:

$$I[x,y] = H[x] - H[x|y] = H[y] - H[y|x]$$

$$I[x,y] \equiv KL(p(x,y)\|p(x)p(y))$$

$$= -\iint p(x,y)\ln\left(\frac{p(x)p(y)}{p(x,y)}\right)dx\,dy \ge 0$$

x 和 y indep. 時 "=" 成立

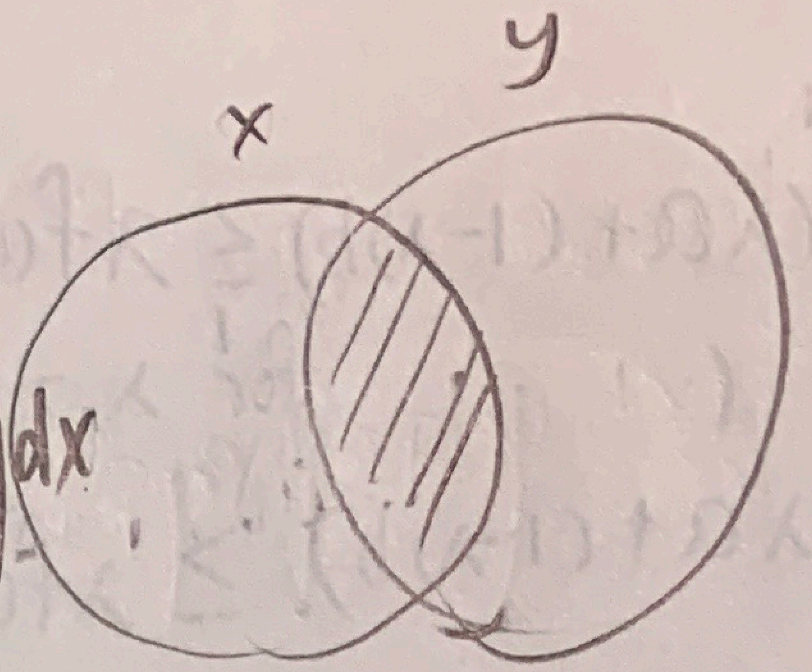
※

+5

2.

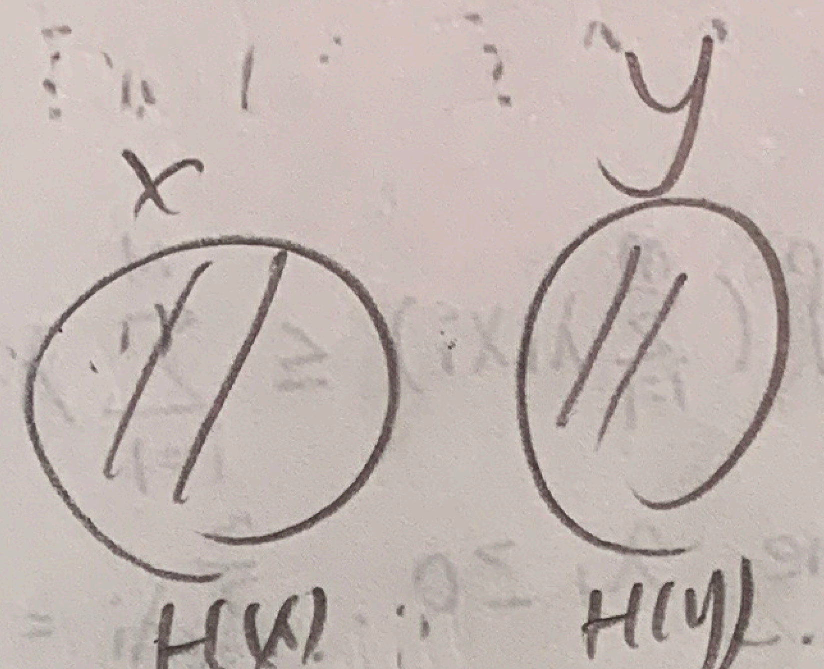
$$H[x,y] = H(x) + H[y|x]$$ +2

$$= -\int p(x)\ln p(x)\,dx - \iint p(y,x)\ln p(y|x)\,dy\,dx$$



$$\because H[y|x] \le H[y]$$

"=" iff x, y are indep.



+2 +12

$$\therefore H[x,y] \le H(x) + H[y]$$

"=" iff x, y are statistically independent. #

3. $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\,dx$, $\Gamma(a+1) = a!$

$$\Gamma(b) = \int_0^\infty y^{b-1} e^{-y}\,dy$$

$$\mathrm{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,\mu^{a-1}(1-\mu)^{b-1}$$

$$\underset{\text{posterior}}{P(\mu|m,l,a,b)} \propto \underset{\text{likelihood}}{\mathrm{Bin}(m|N,\mu)}\;\underset{\text{prior}}{\mathrm{Beta}(\mu|a,b)}$$

$$\propto \frac{\Gamma(m+a+l+b)}{\Gamma(m+a)\Gamma(l+b)}\,\boxed{\mu^{(m+a+1)}(1-\mu)^{(l+b-1)}}$$

$$\lim_{m,l\to\infty} P(x=1|D) = \int_0^1 \mu\, p(\mu|D)\,d\mu = E(\mu|D) = \frac{m+a}{m+a+l+b} \to \frac{m}{N}$$

$$\frac{\prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}}{\binom{N}{m}\mu^N(1-\mu)^{N-m}} = \frac{\mu^a(1-\mu)^{1-a}\,\mu^b(1-\mu)^{1-b}}{\mu^{b+1}(1-\mu)^{1-a}(1-\mu)^{1-b}\mu^a} = \frac{\mu^{a-1}(1-\mu)^{b-1}}{\int_0^1 \mu^{a-1}(1-\mu)^{b-1}\,d\mu}$$

$$\Rightarrow \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 \mu^{a-1}(1-\mu)^{b-1}\,d\mu$$ #

???

4. $P(x, \eta) = h(x) g(\eta) \exp\{\eta^T u(x)\}$.

① $P(x|\mu) = \mu^x (1-\mu)^{1-x}$ ← Bernoulli distribution

$= \exp\{x \ln\mu + (1-x)\ln(1-\mu)\}$ $h(x)$ $h(x)=1$

$= (1-\mu) \exp\{\underbrace{\ln\left(\frac{\mu}{1-\mu}\right)}_{\eta} \cdot \underbrace{x}_{u(x)}\} = \underbrace{\sigma(-\eta)}_{g(\eta)} \exp(\eta x)$

+5

Gaussian distribution

$P(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left\{\frac{-1}{2\sigma^2}(x-\mu)^2\right\}$.

+6

$= h(x) g(\eta) \exp\{\eta^T u(x)\}$

$\eta = \begin{pmatrix} \frac{\mu}{\sigma^2} \\ \frac{-1}{2\sigma^2} \end{pmatrix} \begin{matrix} \eta_1 \\ \eta_2 \end{matrix}$ ✓ $g(\eta) = (-2\eta_2)^{\frac{1}{2}} \left(\frac{\eta_1^2}{4\eta_2}\right)$, $u(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$, $h(x) = (2\pi)^{-\frac{1}{2}}$

② $g(\eta) \int h(x) \exp\{\eta^T u(x)\} dx = 1$

$= \nabla g(\eta) \int h(x) \exp\{\eta^T u(x)\} dx + g(\eta) \int h(x) \exp\{\eta^T u(x)\} u(x) dx = 0$

$\to \frac{-1}{g(\eta)} \nabla g(\eta) = g(\eta) \int h(x) \exp\{\eta^T u(x)\} u(x) dx$

$= E[u(x)] = -\nabla \ln g(\eta)$.

+8

exponential family

$\Rightarrow P(X|\eta) = \left(\prod_{n=1}^{N} h(x_n)\right) g(\eta)^N \exp\left\{\eta^T \sum_{n=1}^{N} u(x_n)\right\}$, $X_n = \{x_1 \cdots x_N\}$.

$\frac{\nabla \ln P(X|\eta)}{\nabla \eta} = 0 \to -\nabla \ln g(\eta_{ML}) = \frac{1}{N} \underbrace{\sum u(x_n)}_{\text{sufficiant statistics.}}$

5.

① $p(t|\alpha,\beta) = \int p(t|w,\beta)\,p(w|\alpha)\,dw$

$= \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \left(\frac{\alpha}{2\pi}\right)^{\frac{M}{2}} \int \exp\{-E(w)\}\,dw$ $\circ\, m_W = \beta A^{-1} \Phi t$

$m_W = S_N(S_0^{-1} m_0 + \beta \Phi^T t)$

$E(w) = \beta E_D(w) + \alpha E_W(w) = \frac{\beta}{2}\|t-\Phi w\|^2 + \frac{\alpha}{2} w^T w$

$\circ\, A = S_N^{-1} = \underset{I}{S_0^{-1}} + \underset{\nabla\nabla E(w)}{\beta \Phi^T \Phi}$

$= \underbrace{E(m_W)}_{\downarrow} + \frac{1}{2}(w-m_N)^T A (w-m_N)$

$\frac{\beta}{2}\|t-\Phi m_W\|^2 + \frac{\alpha}{2} m_W^T m_W = \beta E_D(w) + \alpha E_W(w)$, $A = \alpha I + \beta \Phi^T \Phi$

$\int \exp\{-E(w)\}\,dw = \exp\{-E(m_N)\} \int \exp\{\frac{1}{2}(w-m_N)^T A (w-m_N)\}\,dw$

$= \exp\{-E(m_N)\}\,(2\pi)^{M/2}|A|^{-1/2}$

$\ln p(t|\alpha,\beta) = \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi) + \frac{M}{2}\ln\alpha - E(m_N) - \frac{1}{2}\ln|A|$

(M is the order of the poly)

②

$\hat{\alpha} = \operatorname{argmax} p(t|\alpha,\beta)$

$(\beta\Phi^T\Phi)u_i = \lambda_i u_i$ A has eigenvalue $\{\alpha . \lambda_i\}$.

$\frac{d}{d\alpha}\ln|A| = \frac{d}{d\alpha}\ln\left(\prod_i(\lambda_i+\alpha)\right) = \frac{d}{d\alpha}\sum \ln(\lambda_i+\alpha) = \sum_i \frac{1}{\lambda_i+\alpha}$

$\frac{d}{d\alpha}\ln p(t|\alpha,\beta) = \frac{M}{2\alpha} - \frac{1}{2}m_N^T m_N - \frac{1}{2}\sum_i \frac{1}{\lambda_i+\alpha} = 0.$

$\alpha m_N^T m_N = M - \alpha\sum_i \frac{1}{\lambda_i+\alpha} = \gamma = \sum_i \frac{\lambda_i}{\alpha+\lambda_i}$

$\rightarrow \alpha = \frac{\gamma}{m_N^T m_N}$. implict solution ✗

$(\beta\Phi^T\Phi)u_i = \lambda_i u_i$

$\hat{\beta} = \operatorname{argmax} p(t|\alpha,\beta)$

$\frac{d}{d\beta}\ln|A| = \frac{d}{d\beta}\ln(\lambda_i+\alpha) = \frac{1}{\beta}\sum_i\left(\frac{\lambda_i}{\lambda_i+\alpha}\right)\left(\frac{d\lambda_i}{d\beta}\right) = \frac{\gamma}{\beta}.$, λ_i is propotioned to β.

$\frac{d}{d\beta}\ln p(t|\alpha,\beta) = \frac{N}{2\beta} - \frac{1}{2}\sum_{n=1}^{N}\{t_n - m_N^T\phi(x_n)\}^2 - \frac{\gamma}{2\beta} = 0.$

$\rightarrow \beta^{-1} = \frac{1}{N-\gamma}\sum_{n=1}^{N}\{t_n - m_N^T\phi(x_n)\}^2$, implict solution

✗

6. $\{x_n, t_n\}_{n=1}^{N}$ $t_n=1$ C_1
$t_n=0$ C_2

+20

Gaussian class - conditional density

$P(C_1)=\pi$,
$P(C_2)=1-\pi$,

$P(x_n, C_1) = P(C_1)p(x_n|C_1) = \pi N(x_n|\mu_1, \Sigma)$, $t_n=1$

$P(x_n, C_2) = P(C_2)P(x_n|C_2) = (1-\pi)N(x_n|\mu_2, \Sigma)$, $t_n=0$.

① likelihood function:

$$P(t_1,\dots,t_N|\pi,\mu_1,\mu_2,\Sigma)$$

$$= \prod_{n=1}^{N}[\pi N(x_n|\mu_1,\Sigma)]^{t_n}[(1-\pi)N(x_n|\mu_2,\Sigma)]^{1-t_n}$$ ※

②

$$\frac{\partial P(t|\pi,\mu_1,\mu_2,\Sigma)}{\partial \pi}=0.$$ (function depends on π)

$$\Rightarrow \pi = \frac{1}{N}\sum_{n=1}^{N} t_n = \frac{N_1}{N_1+N_2}.$$ ※

$$\mu_1: \sum_{n=1}^{N} t_n \ln N(x_n|\mu_1,\Sigma) = \frac{-1}{2}\sum_{n=1}^{N} t_n(x_n-\mu_1)^T\Sigma^{-1}(x_k-\mu_1) + 常数.$$

$$\nabla\mu_1 = 0 \Rightarrow \mu_1 = \frac{1}{N_1}\sum_{n=1}^{N} t_n x_n$$, sample mean of C_1

$$\nabla\mu_2 = 0 \Rightarrow \mu_2 = \frac{1}{N_2}\sum_{n=1}^{N}(1-t_n)x_n$$, sample mean of C_2 ※

$$\frac{-1}{2}\sum_{n=1}^{N} t_n \ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N} t_n(x_n-\mu_1)^T\Sigma^{-1}(x_n-\mu_1)$$

$$-\frac{1}{2}\sum_{n=1}^{N}(1-t_n)\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(1-t_n)(x_n-\mu_2)^T\Sigma^{-1}(x_n-\mu_2)$$

$$= \frac{-N}{2}\ln|\Sigma| - \frac{N}{2}\mathrm{Tr}\{\Sigma^{-1}S\}$$

$$S = \frac{N_1}{N}S_1 + \frac{N_2}{N}S_2, \quad S_1 = \frac{1}{N_1}\sum_{n\in C_1}(x_n-\mu_1)(x_n-\mu_1)^T$$

$$S_2 = \frac{1}{N_2}\sum_{n\in C_2}(x_n-\mu_2)(x_n-\mu_2)^T$$

$$\nabla\Sigma^{-1}=0.$$

$$\Rightarrow \Sigma = S$$

※※