國立交通大學考試試卷
Exam Paper of National Chiao Tung University

題號 NO	分數 Score
1	14.5
2	10
3	15
4	19
5	18
6	20.
7	
8	
9	
10	
11	
12	
13	
14	
15	
總分 Total	96.5

課程名稱 Course Name: 機器學習

姓名 Name: 簡詠倫

學號 Student ID: 309513061

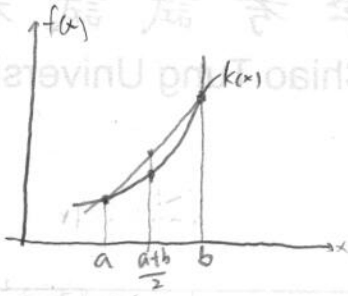
班別 Department: 電信甲

日期 Date: 11/13

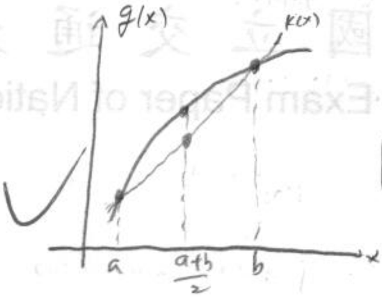
※考試作弊者將受記大過以上處分
※If you cheat in exam, you will be punished.

人

Convex funtion



concave function



+3

若f(x)為convex-function
則有一通過a,b兩點的 (b>a)
直線方程式k(x)使 $k(\frac{a+b}{2}) > f(\frac{a+b}{2})$,
f(x)二次微分後仍有大於零的值

若g(x)為concave function
則有一通過a,b兩點的直 (b>a)
線方程式k(x)使 $k(\frac{a+b}{2}) < g(\frac{a+b}{2})$
g(x)二次微分後仍有小於零的值

Jensen's inequality: $f(\sum_{i=1}^{M} \lambda_i x_i) \le \sum_{i=1}^{M} \lambda_i f(x_i)$, $f(E(x)) \le E[f(x)]$ # +3

KL divergence: $KL(p\|q) = -\int p(x) \ln(\frac{q(x)}{p(x)}) dx$,

use Jensen's inequality $f(E(x)) \le E[f(x)]$ +4

$\Rightarrow -\int p(x) \ln\frac{q(x)}{p(x)} dx \ge -\ln\int p(x) \frac{q(x)}{p(x)} dx = 0$

$\therefore -\int p(x) \ln\frac{q(x)}{p(x)} dx \ge 0$ #

Mutual information of two random variables x and y:

$I(x,y) \equiv KL(p(x,y)\|p(x)p(y)) \ge 0$ +4.5

if two random variables are independent, $I(x,y) = -\iint p(x,y) \ln\frac{p(x)p(y)}{p(x,y)} dx$

$= -\int p(x,y) \ln\frac{p(x)p(y)}{p(x)p(y)} dx$

$= 0$ #

2、 $p(x,y)$, $H[x,y] \le H[x] + H[y]$

$H[x] = -\sum_i p(x_i)\ln p(x_i)$ or $-\int p(x_i)\ln p(x_i)\,dx$, $H[y|x] = -\iint p(x,y)\ln p(y|x)\,dy\,dx$ +2

$H[x,y] = H[y|x] + H[x]$

if x and y are statistically independent,

$$H[y|x] = -\iint p(x,y)\ln p(y|x)\,dy\,dx = -\iint p(x)p(y)\ln\frac{p(x)p(y)}{p(x)}\,dy\,dx$$ +6

$$= -\underbrace{\int p(x)\,dx}_{1}\int p(y)\ln p(y)\,dy$$

$$= -\int p(y)\ln p(y)\,dy = H[y]$$

if x and y are statistically independent $\Rightarrow H[x,y] = H[y] + H[x]$ #

3、 $$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 u^{a-1}(1-u)^{b-1}\,du$$

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

$\xrightarrow{\text{分部積分}}$ $u = e^{-x}$ $v = x^a$

$\Rightarrow a\int_0^\infty e^{-x}x^{a-1}\,dx = \underbrace{e^{-x}x^a\Big|_0^\infty}_{0} + \int_0^\infty x^a e^{-x}\,dx$

$\Rightarrow \Gamma(a+1) = a\Gamma(a)$

$\xrightarrow{\text{得}}$ $\Gamma(a) = (a-1)\Gamma(a-1)$

$\Gamma(b) = (b-1)\Gamma(b-1)$

$\Gamma(a+b) = (a+b-1)\Gamma(a+b-1)$

$$\Rightarrow \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \frac{(a-1)(b-1)\Gamma(a-1)\Gamma(b-1)}{(a+b-1)\Gamma(a+b-1)}$$

$$= \frac{\cancel{(a-1)!}(b-1)!}{(a+b-1)(a+b-2)\dots(a+1)a\cancel{(a-1)!}}$$

$$= \frac{(b-1)!}{(a+b-1)(a+b-2)\cdots a}$$

$\int_0^1 u^{a-1}(1-u)^{b-1}\,du$

$\xrightarrow{\text{分部積分}}$ $u = (1-u)^{b-1}$ $v = u^a$

$\Rightarrow a\int_0^1 u^{a-1}(1-u)^{b-1}\,du$

$= \underbrace{(1-u)^{b-1}u^a\Big|_0^1}_{0} + (b-1)\int_0^1 u^a(1-u)^{b-2}\,du$

$\Rightarrow \int_0^1 u^{a-1}(1-u)^{b-1}\,du = \frac{b-1}{a}\int_0^1 u^a(1-u)^{b-2}\,du$

$$= \frac{(b-1)(b-2)\dots}{(a+b-1)(a+b-2)\dots a}$$

相等

#

4.

Exponential family

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

Bernoulli distribution:

$$\mu^x (1-\mu)^{1-x}$$

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

$$= (1-\mu) \exp\{\ln(\frac{\mu}{1-\mu}) \cdot x\}$$

$$= \sigma(-\eta) \exp(\eta x)$$

其中 $h(x) = 1$, $\eta = \ln(\frac{\mu}{1-\mu})$, $u(x) = x$.

$g(\eta) = \sigma(-\eta)$, $\mu = \frac{1}{1+\exp(-\eta)} = \sigma(\eta)$

$\Rightarrow$ Bernoulli distribution is exponential family

+6

Gaussian distribution

$$\frac{1}{(2\pi\sigma^2)^{1/2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

$$= \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}\mu^2\}$$

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

其中. $h(x) = \frac{1}{(2\pi)^{1/2}}$, $\eta^T = \begin{array}{c} \begin{pmatrix} \frac{\mu}{\sigma^2} \\ \frac{-1}{2\sigma^2} \end{pmatrix} \end{array} \begin{array}{c} \eta_1 \\ \eta_2 \end{array}$

$u(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$, $g(\eta) = (-2\eta_2)^{1/2} \exp(\frac{\eta_1^2}{4\eta_2})$

$\Rightarrow$ Gaussian distribution is exponential family #

+6

ML principle, $p(x|\eta) = \left(\prod_{n=1}^{N} h(x_n)\right) g(\eta)^N \exp(\eta^T \sum_{n=1}^{N} u(x_n))$

$$\xrightarrow{\ln} \ln p(x|\eta) = \sum_{n=1}^{N} h(x_n) + N \ln g(\eta) + \eta^T \sum_{n=1}^{N} u(x_n)$$

$$\Longrightarrow \left.\frac{\partial \ln p(x|\eta)}{\partial \eta}\right|_{\eta_{ML}} = 0 \Rightarrow \nabla N \ln g(\eta) + \sum_{n=1}^{N} u(x_n) = 0$$

$$\Rightarrow -\nabla \ln g(\eta_{ML}) = \frac{1}{N} \sum_{n=1}^{N} u(x_n) \#$$

+7

可由充分统计量 $\sum_n u(x_n)$ 求得 η

5、 $y(x,w) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(x) = w^T\phi(x)$, likelihood $p(t|x,w,\beta) = N(t|y(x,w),\beta^{-1})$

prior $p(w|\alpha) = N(w|0,\alpha^{-1})$

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

(a) (b)

Evolution:

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

$$= \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp\{-E(w)\}\, dw \text{ , where } E(w) = \beta E_D(w) + \alpha E_W(w)$$

$$= \frac{\beta}{2}\|t - \Phi w\|^2 + \frac{\alpha}{2}w^T w$$

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

$m_N = \beta A^{-1}\Phi^T t$ (∵ prior 的 mean 為 0)

$A = S_N^{-1} = \alpha I + \beta \Phi^T\Phi$

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

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

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

+10

* Maximize the evidence function,

eigenvector equation : $(\beta\Phi^T\Phi)u_i = \lambda_i u_i \Rightarrow A$ has eigenvalues $(\alpha + \lambda_i)$

$$\frac{d}{d\alpha}\ln|A| = \frac{d}{d\alpha}\ln \mathrm{Tr}(\lambda_i + \alpha) = \frac{d}{d\alpha}\sum_i \ln(\lambda_i + \alpha) = \sum_i \frac{1}{\lambda_i + \alpha}$$

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

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

求 β ; $\frac{d}{d\beta}\ln|A| = \frac{d}{d\beta}\sum_i \ln(\lambda_i + \alpha) = \frac{1}{\beta}\sum_i \frac{\lambda_i}{\lambda_i + \alpha}$ $\left(\frac{d\lambda_i}{d\beta} = \frac{\lambda_i}{\beta}\right)$

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

$$\Rightarrow \beta^{-1} = \frac{1}{N-\gamma}\sum_{n=1}^{N}\{t_n - m_N^T\phi(x_n)\}^2 \ \#$$

6.

likelihood function $p(t|\pi, \mu_1, \mu_2, \Sigma)$

(+20)

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

$$\frac{\partial p(t|\pi, \mu_1, \mu_2, \Sigma)}{\partial \pi} = 0 \Rightarrow \pi = \frac{1}{N}\sum_{n=1}^{N} t_n = \frac{N_1}{N} = \frac{N_1}{N_1+N_2} \#$$

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

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

類推 $\nabla_{\mu_2} = 0 \Rightarrow \mu_2 = \frac{1}{N_2}\sum_{n=1}^{N} (1-t_n) x_n$ C_2 #

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

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

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

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

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

$$\nabla_{\Sigma^{-1}} = 0 \Rightarrow \Sigma = S \#$$

ML is not robust to outliers #