Machine Learning HW-1

NO.

DATE

1. Jensen's inequality

已知：$f(\lambda a + (1-\lambda) b) \le \lambda f(a) + (1-\lambda) f(b)$ ——— equation(1)

when $M=1$

$$f\left(\sum_{i=1}^{M}\lambda_i x_i\right) = f\left(\sum_{i=1}^{1}\lambda_i x_i\right) = f(\lambda_1 x_1) \overset{\because \sum \lambda_i = 1}{=\!=\!=} f(x_1) \le \lambda_1 \cdot f(x_1)$$

$\because \lambda_i \ge 0$

$\therefore f(\lambda_1 x_1) \le \lambda_1 f(x_1)$ 成立!

when $M=2$

$$f\left(\sum_{i=1}^{M}\lambda_i x_i\right) = f\left(\sum_{i=1}^{2}\lambda_i x_i\right) = f(\lambda_1 x_1 + \lambda_2 x_2) \underset{\therefore \lambda_1+\lambda_2=1}{\overset{\because \sum \lambda_i = 1}{=\!=\!=}} f(\lambda x_1 + (1-\lambda) x_2)$$

$\Rightarrow$ 令 $\begin{cases}\lambda_1 = \lambda \\ \lambda_2 = 1-\lambda\end{cases}$

由 equation(1)，可知 $f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$

$\therefore M=2$ 時亦成立 ②

when $M=3$

$$f\left(\sum_{i=1}^{M}\lambda_i x_i\right) = f\left(\sum_{i=1}^{3}\lambda_i x_i\right) = f(\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3)$$

$$= f\left((\lambda_1+\lambda_2) \times \left(\frac{\lambda_1}{\lambda_1+\lambda_2}x_1 + \frac{\lambda_2}{\lambda_1+\lambda_2}x_2\right) + \lambda_3 x_3\right)$$

$\because \lambda_1+\lambda_2+\lambda_3=1$

$\therefore$ 令 $\lambda = \lambda_1+\lambda_2$, $1-\lambda = \lambda_3$

$$= f\left(\lambda \times \left(\frac{\lambda_1}{\lambda_1+\lambda_2}x_1 + \frac{\lambda_2}{\lambda_1+\lambda_2}x_2\right) + (1-\lambda) x_3\right)$$

$$\le \lambda f\left(\frac{\lambda_1}{\lambda_1+\lambda_2}x_1 + \frac{\lambda_2}{\lambda_1+\lambda_2}x_2\right) + (1-\lambda) f(x_3)$$

由②可得 $\le \lambda\left(\frac{\lambda_1}{\lambda_1+\lambda_2}f(x_1) + \frac{\lambda_2}{\lambda_1+\lambda_2}f(x_2)\right) + (1-\lambda) f(x_3)$

$\because \frac{\lambda_1}{\lambda_1+\lambda_2} + \frac{\lambda_2}{\lambda_1+\lambda_2} = 1$

NO.

DATE

又 ∵ $\lambda = \lambda_1 + \lambda_2$

∴ $\lambda\left(\frac{\lambda_1}{\cancel{\lambda_1+\lambda_2}} f(x_1) + \frac{\lambda_2}{\cancel{\lambda_1+\lambda_2}} f(x_2)\right) + (1-\lambda) f(x_3)$

$= \lambda_1 f(x_1) + \lambda_2 f(x_2) + (1-\lambda_1-\lambda_2) f(x_3)$

$= \lambda_1 f(x_1) + \lambda_2 f(x_2) + \lambda_3 f(x_3)$

⇒ 得證 $f\left(\sum_{i=1}^{3} \lambda_i x_i\right) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \lambda_3 f(x_3) = \sum_{i=1}^{3} \lambda_i f(x_i)$

∴ $M=3$ 亦成立

⋮

when $M = k-1$,

$$f\left(\sum_{i=1}^{k-1} \lambda_i x_i\right) \leq \sum_{i=1}^{k-1} \lambda_i f(x_i) \text{ 成立}$$

when $M = k$.

$$f\left(\sum_{i=1}^{k} \lambda_i x_i\right) = f\left(\sum_{i=1}^{k-1} \lambda_i x_i + \lambda_k x_k\right)$$

$$= f\left((1-\lambda_k) \times \frac{\sum_{i=1}^{k-1} \lambda_i x_i}{(1-\lambda_k)} + \lambda_k x_k\right)$$

$$\leq (1-\lambda_k) \times f\left(\frac{\sum_{i=1}^{k-1} \lambda_i x_i}{1-\lambda_k}\right) + \lambda_k f(x_k)$$

$$= (1-\lambda_k) \times f\left(\sum_{i=1}^{k-1} \frac{\lambda_i}{1-\lambda_k} x_i\right) + \lambda_k f(x_k)$$

根據 $M=k-1$ 的結論

$$\leq (1-\lambda_k) \times \sum_{i=1}^{k-1} \frac{\lambda_i}{1-\lambda_k} f(x_i) + \lambda_k f(x_k)$$

NO.

DATE

$$= \cancel{(1-\lambda_k)} \cdot \frac{1}{\cancel{(1-\lambda_k)}} \sum_{i=1}^{k-1} \lambda_i f(x_i) + \lambda_k f(x_k)$$

$$= \sum_{i=1}^{k-1} \lambda_i f(x_i) + \lambda_k f(x_k)$$

$$= \sum_{i=1}^{k} \lambda_i f(x_i)$$

\# 得證!

$\forall M, \quad f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \leq \sum_{i=1}^{M} \lambda_i f(x_i)$, which $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$ #

2. Derive the Entropy of the Universe Gaussian.

Base on definition:

$$H[x] = -\int p(x) \ln(p(x))\, dx$$

$$= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \times \left(\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \ln\left(e^{\frac{-(x-\mu)^2}{2\sigma^2}}\right)\right) dx$$

$$= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \times \left(-\frac{1}{2} \ln(2\pi\sigma^2) + \frac{-(x-\mu)^2}{2\sigma^2}\right) dx$$

$$= \frac{1}{2} \ln(2\pi\sigma^2) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \times \frac{-(x-\mu)^2}{2\sigma^2} dx$$

$$= \frac{1}{2} \ln(2\pi\sigma^2) \int_{-\infty}^{\infty} p(x)\, dx - \int_{-\infty}^{\infty} p(x) \times \frac{-(x-\mu)^2}{2\sigma^2}\, dx$$