CS535 HW1 933271081, Chih-Hsiang, Wang

1. for 1-dimension : $X_1, \dots, X_N \sim \text{Unif}([0,1])$ $\xrightarrow{cdf}$ $P(X_i \le x) = F_{X_i} = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}$
(d(1,N))

$P(\min_i \{X_i\} \le r) = F_{\min_i\{X_i\}}(r) = 1 - P(\min_i\{X_i\} > r) = 1 - P(X_1 > r, X_2 > r, \dots, X_N > r)$

$(\|r\|^2 \le 1)$ $= 1 - \prod_{i=1}^{N} P(X_i > r) = 1 - \prod_{i=1}^{N} (1-r)$ (by cdf above)

$= 1 - (1-r)^N$

median distance $= r_{0.5}$ (d(1,N)), by median is the 0.5 quantile of its cdf

$\Rightarrow 1-(1-r_{0.5})^N = 0.5 \Rightarrow (1-r_{0.5})^N = 0.5 \Rightarrow -r_{0.5} = \frac{1}{2}^{\frac{1}{N}} - 1 \Rightarrow r_{0.5} = (1-\frac{1}{2}^{\frac{1}{N}})^{\frac{1}{p}}$, $p=1 \Rightarrow$ proved

for p-dimension : $P(X \in S_p(r)) = (\frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2}+1)} r^p) / (\frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2}+1)} 1^p) = r^p \Rightarrow P(\|X\| \le r) = r^p$
(d(p,N))

$P(\min_i\{\|X_i\|\} \le r) = F_{\min_i\{\|X_i\|\}}(r) = 1 - P(\min_i\{\|X_i\|\} > r) = 1 - \prod_{i=1}^{N} P(\|X_i\| > r)$

$= 1 - \prod_{i=1}^{N} (1 - P(\|X\| \le r)) = 1 - (1-r^p)^N$

median distance $= r^p_{0.5}$ (d(p,N)). by median is the 0.5 quantile of its cdf

$\Rightarrow 1-(1-r^p_{0.5})^N = 0.5 \Rightarrow 1 - r^p_{0.5} = 0.5^{\frac{1}{N}} \Rightarrow r^p_{0.5} = 1 - \frac{1}{2}^{\frac{1}{N}} \Rightarrow r_{0.5} = (1-\frac{1}{2}^{\frac{1}{N}})^{\frac{1}{p}} \Rightarrow$ proved

$d(p,N) = (1-\frac{1}{2}^{\frac{1}{N}})^{\frac{1}{p}}$, for $N=10000$, $p=1000 \Rightarrow d(1000, 10000) = (1-\frac{1}{2}^{\frac{1}{10000}})^{\frac{1}{1000}} \doteq 0.9904688$

2. $f(x) = (x_1+x_2)(x_1x_2+x_1x_2^2)$

$$\nabla f(x) = \begin{bmatrix} (x_1x_2+x_1x_2^2)+(x_1+x_2)(x_2+x_2^2) \\ (x_1x_2+x_1x_2^2)+(x_1+x_2)(x_1+2x_1x_2) \end{bmatrix} = \begin{bmatrix} x_2\{(x_1+x_1x_2)+(x_1+x_2)(1+x_2)\} \\ x_1\{(x_2+x_2^2)+(x_1+x_2)(1+2x_2)\} \end{bmatrix}$$

$$= \begin{bmatrix} x_2(1+x_2)(2x_1+x_2) \\ x_1(x_2+x_2^2+x_1+x_2+2x_1x_2+2x_2^2) \end{bmatrix} = \begin{bmatrix} x_2(1+x_2)(2x_1+x_2) \\ x_1(3x_2^2+2x_1x_2+x_1+2x_2) \end{bmatrix} *$$

$$\nabla^2 f(x) = \begin{bmatrix} 2x_2^2+2x_2 & (1+2x_2)(2x_1+x_2)+(x_2+x_2^2) \\ (3x_2^2+2x_1x_2+x_1+2x_2)+x_1(2x_2+1) & x_1(6x_2+2x_1+2) \end{bmatrix}$$

$$= \begin{bmatrix} 2x_2^2+2x_2 & 3x_2^2+4x_1x_2+2x_1+2x_2 \\ 3x_2^2+4x_1x_2+2x_1+2x_2 & 2x_1^2+6x_1x_2+2x_1 \end{bmatrix} *$$

stationary points $\Rightarrow \nabla f(x) = 0$

$x_1 = 0 \Rightarrow x_2(1+x_2)(x_2) = 0 \Rightarrow x_2 = 0$ or $-1 \Rightarrow (0,0), (0,-1)$

$x_2 = -1 \Rightarrow x_1(3-2x_1+x_1-2) = 0 \Rightarrow x_1 = 1 \Rightarrow (1,-1)$

$\Rightarrow (0,0), (0,-1), (1,-1)$ ...... *

$(0,0)$: $\nabla^2 f(x) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $|\nabla^2 f(x)| = 0$

$(0,-1)$: $\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $|\nabla^2 f(x)| = -1$

$(1,-1)$: $\nabla^2 f(x) = \begin{bmatrix} 0 & -1 \\ -1 & -2 \end{bmatrix}$, $|\nabla^2 f(x)| = -1$

$(\frac{3}{8}, -\frac{6}{8})$: $\nabla^2 f(x) = \begin{bmatrix} 2\cdot\frac{9}{16}+2\cdot(-\frac{3}{4}) & 3\cdot\frac{9}{16}-\frac{9}{8}-\frac{3}{4} \\ 3\cdot\frac{9}{16}+4\cdot\frac{3}{8}\cdot(-\frac{3}{4})+\frac{3}{4}-\frac{6}{4} & 2\cdot\frac{9}{64}+6\cdot\frac{3}{8}\cdot(-\frac{6}{8})+\frac{3}{4} \end{bmatrix} = \begin{bmatrix} -\frac{3}{8} & -\frac{3}{16} \\ -\frac{3}{16} & -\frac{21}{32} \end{bmatrix}$, $|\nabla^2 f(x)| = \frac{27}{128}$

$\because |\nabla^2 f(x)| > 0$ only when $x = (\frac{3}{8}, -\frac{6}{8})$ for points $(0,0), (0,-1), (1,-1), (\frac{3}{8}, -\frac{6}{8})$, and $-\frac{3}{8} < 0$ (f_{xx})

$\Rightarrow$ only $x = [\frac{3}{8}, -\frac{6}{8}]^T$ is the local maximum of $f(x)$ *

$f(x) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$

$\nabla f(x) = \begin{bmatrix} 8+2x_1 \\ 12-4x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow x_1 = -4,\ x_2 = 3 \Rightarrow$ stationary point is only $(-4, 3)$

$\Downarrow$

$|\nabla^2 f(x)| = \left| \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix} \right| = -8 < 0 \Rightarrow$ only one stationary point $(-4, 3)$ for $f(x)$,

and it is neither a minimum nor a maximum, but is a saddle point $_*$

4.

let $x = (x_1, x_2)^T$, $x \neq 0$

$x^T \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} x = (x_1^T\ x_2^T) \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^T A x_1 + x_2^T B x_2$

$\because A, B$ are positive definite matrices $\Rightarrow x_1^T A x_1 > 0,\ x_2^T B x_2 > 0 \Rightarrow x_1^T A x_1 + x_2^T B x_2 > 0$

$\Rightarrow x^T \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} x > 0$. (M is positive definite $\Leftrightarrow z^T M z > 0$)

$\Rightarrow \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ is also positive definite $_*$

5. forward function :

$$z_2 = \frac{1}{1+\exp(-w_2^T \cdot \frac{1}{1+\exp(-w^T x - b_1)} - b_2)} = \frac{1}{1+\exp -(\frac{w_2^T}{1+\exp-(w^T x+b_1)} + b_2)}$$

backpropagation functions .

$$\frac{\partial L}{\partial z_2} = y^* \frac{1}{z_2} + (1-y^*) \frac{-1}{1-z_2} = \frac{y^* - y^* z_2 - z_2 + y^* z_2}{z_2(1-z_2)} = \frac{y^* - z_2}{z_2(1-z_2)}$$

$$\left(\sigma(x) = \frac{1}{1+e^{-x}} \Rightarrow \frac{\partial \sigma}{\partial x} = \sigma(x)(1-\sigma(x)) \right)$$

$$\frac{\partial z_2}{\partial w_2} = z_2(1-z_2) z_1 , \quad \frac{\partial z_2}{\partial b_2} = z_2(1-z_2) , \quad \frac{\partial z_2}{\partial z_1} = z_2(1-z_2) w_2^T$$

$$\frac{\partial z_1}{\partial w} = z_1(1-z_1) x , \quad \frac{\partial z_1}{\partial b_1} = z_1(1-z_1)$$

$$\left(L(\theta) = \sum_{n=1}^{N} C^n(\theta) \quad \frac{\partial L(\theta)}{\partial w} = \sum_{n=1}^{N} \frac{\partial C^n(\theta)}{\partial w} \right)$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial z_2} \frac{\partial z_2}{\partial w_2} = \frac{y^* - z_2}{z_2(1-z_2)} \cdot z_2(1-z_2) z_1 = z_1 (y^* - z_2)$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial z_2} \frac{\partial z_2}{\partial b_2} = \frac{y^* - z_2}{z_2(1-z_2)} \cdot z_2(1-z_2) = y^* - z_2$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial z_2} \frac{\partial z_2}{\partial z_1} \frac{\partial z_1}{\partial w} = \frac{y^* - z_2}{z_2(1-z_2)} \cdot w_2^T \cdot z_2(1-z_2) \cdot z_1(1-z_1) x = w_2^T z_1 x (y^* - z_2)(1-z_1)$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial z_2} \frac{\partial z_2}{\partial z_1} \frac{\partial z_1}{\partial b_1} = \frac{y^* - z_2}{z_2(1-z_2)} \cdot w_2^T \cdot z_2(1-z_2) \cdot z_1(1-z_1) = w_2^T z_1 (y^* - z_2)(1-z_1)$$

$$\left(\begin{aligned} w_2^t &= w_2^{t-1} - \alpha \frac{\partial L}{\partial w_2} \\ w^t &= w^{t-1} - \alpha \frac{\partial L}{\partial w} \\ b_2^t &= b_2^{t-1} - \alpha \frac{\partial L}{\partial b_2} \\ b_1^t &= b_1^{t-1} - \alpha \frac{\partial L}{\partial b_1} \end{aligned} \right)$$