CS535 Ting-Liang Huang HW1

1. ① According to CDF of n-dim unit ball with r where $r \in [0, 1]$ is

$$P(S_p(r)) = \frac{V_p(r)}{V_p(1)} = r^p$$

Define R is the closet data point to origin in data points, $R = \min_i (\|X_i\|)$, $i = 1 \dots N$ i.i.d.

$$P(R \le r) = 1 - P(\min_i(\|X_i\|) > r)$$

$$= 1 - \prod_{i=1}^{N} (1 - P(\|X_i\| < r))$$

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

We're looking for the median distance $P(R \le r) = 0.5$

$$1 - (1 - r^p)^N = 0.5$$

$$r = \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{N}}\right)^{\frac{1}{p}} = d(p, N) \quad \#$$

② Compute $N = 10000$, $p = 1000$.

$$r = \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{10000}}\right)^{\frac{1}{1000}} \doteqdot 0.99 \quad \#$$

2. ① $f(x) = (x_1 + x_2)(x_1 x_2 + x_1 x_2^2) = x_1^2 x_2 + x_1^2 x_2^2 + x_1 x_2^2 + x_1 x_2^3$

$$\nabla f(x) = \left[\frac{df(x)}{d(x_1)}, \frac{df(x)}{d(x_2)}\right]^T = \left[2x_1x_2 + 2x_1x_2^2 + x_2^2 + x_2^3,\ x_1^2 + 2x_1^2x_2 + 2x_1x_2 + 3x_1x_2^2\right]^T$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{d^2 f(x)}{d^2(x_1)} & \frac{d^2 f(x)}{d(x_1)d(x_2)} \\ \frac{d^2 f(x)}{d(x_1)d(x_2)} & \frac{d^2(f)}{d^2(x_2)} \end{bmatrix} = \begin{bmatrix} 2x_2 + 2x_2^2 & 2x_1 + 4x_1x_2 + 2x_2 + 3x_2^2 \\ 2x_1 + 4x_1x_2 + 2x_2 + 3x_2^2 & 2x_1^2 + 2x_1 + 6x_1x_2 \end{bmatrix}$$

When $\nabla f(x) = 0$, we can find stationary point (x_1, x_2)

$x_1 = 0, x_2 = 0 \Rightarrow \nabla f(x) = [0, 0]^T$ ✓

$x_1 = 0, x_2 = -1 \Rightarrow \nabla f(x) = [0, 0]^T$ ✓

$x_1 = \frac{3}{8}, x_2 = -\frac{3}{4} \Rightarrow \nabla f(x) = [0, 0]^T$ ✓ #

stationary points

② Compute Hessian at each stationary point I found.

$$x_1=0,\ x_2=0 \Rightarrow \nabla^2 f(x) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \#$$

$$x_1=0,\ x_2=-1 \Rightarrow \nabla^2 f(x) = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \#$$

$$x_1=\frac{3}{8},\ x_2=-\frac{3}{4} \Rightarrow \nabla^2 f(x) = \begin{bmatrix} 2\cdot(-\frac{3}{4})+2\cdot(-\frac{3}{4})^2 & 2\cdot\frac{3}{8}+4\cdot\frac{3}{8}\cdot(-\frac{3}{4})+2\cdot(-\frac{3}{4})+3\cdot(-\frac{3}{4})^2 \\ 2\cdot\frac{3}{8}+4\times\frac{3}{8}\times(-\frac{3}{4})+2\times(-\frac{3}{4})+3\times(-\frac{3}{4})^2 & 2\cdot(\frac{3}{8})^2+2\cdot\frac{3}{8}+6\cdot\frac{3}{8}\cdot(-\frac{3}{4}) \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{3}{8} & -\frac{3}{16} \\ -\frac{3}{16} & -\frac{21}{32} \end{bmatrix} \#$$

③ Show that $x=[\frac{3}{8}, -\frac{6}{8}]^T$ is the only local maximum of this function.

$H = f_{x_1x_1}\cdot f_{x_2x_2} - (f_{x_1x_2})^2$

$H(0,0) = 0\cdot 0 - 0^2 = 0$

$H(0,-1) = 0\cdot 0 - (-1)^2 = -1.$

$H(\frac{3}{8},-\frac{3}{4}) = (-\frac{3}{8})\times(-\frac{21}{32}) - (-\frac{3}{16})^2 = \frac{63-9}{16\times 16} = \frac{\cancel{54}^{27}}{16\times\cancel{16}_{8}} = \frac{27}{128}$

If $f_{x_1x_2}<0$ and $H>0$, then it's local maximum, $\therefore H(\frac{3}{8},-\frac{3}{4})$ is local maximum #

3. ① Show that the function $f(x) = 8x_1+12x_2+x_1^2-2x_2^2$ has only one stationary point; and it's neither a minimum nor a maximum, but a saddle point.

$$\nabla f(x) = \left[\frac{df(x)}{d(x_1)},\ \frac{df(x)}{d(x_2)}\right]^T = [8+2x_1,\ 12-4x_2]^T = 0$$

$$\begin{cases} 8+2x_1=0 \\ 12-4x_2=0 \end{cases},\quad x_1=-4,\ x_2=3 \text{ (only have one stationary point)} \#$$

② $H = f_{x_1x_1}\cdot f_{x_2x_2} - (f_{x_1x_2})^2$ $\qquad \nabla^2 f(x) = \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix}$

$= 2\times(-4) - 0^2 = -8$

Since $H=-8<0$ and $f_{x_1x_2}=2>0$

$\therefore$ It's a saddle point #

4. if A and B are definite matrics, prove that the matrix $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ is also a positive definite:

(1) A matrix is positive definite if $x^T A x > 0$ for all $x \neq 0$

(2) Assume two vectors x_A, x_B where $x_A^T A x_A > 0$ and $x_B^T B x_B > 0$

(3) combine x_A, x_B to bigger vectors $x_c = [x_a, x_b]$ and $x_c \neq 0$

(4) $x_c^T \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} x_c > 0 \quad \therefore \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ is positive definite. #

5.

① Derive the forward computing $z_2 = \dfrac{1}{1+\exp(-(w_2^T z_1 + b_2))} = \dfrac{1}{1+\exp\left(-\left(w_2^T \dfrac{1}{1+\exp(-w_1^T x - b_1)} + b_2\right)\right)}$ #

② Backprop:

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

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_2} = \frac{y^* - z_2}{z_2(1-z_2)} \cdot \frac{\partial(\sigma(w_2^T z_1 + b_2))}{\partial(w_2^T z_1 + b_2)} \cdot \frac{\partial(w_2^T z_1 + b_2)}{\partial w_2} = \frac{y^* - z_2}{\cancel{z_2(1-z_2)}} \cdot \cancel{z_2(1-z_2)}\, z_1 = z_1(y^* - z_2) \#$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial b_2} = \frac{y^* - z_2}{z_2(1-z_2)} \cdot \frac{\partial(\sigma(w_2^T z_1 + b_2))}{\partial(w_2^T z_1 + b_2)} \cdot \frac{\partial(w_2^T z_1 + b_2)}{\partial b_2} = \frac{y^* - z_2}{\cancel{(1-z_2)z_2}} \cdot \cancel{z_2(1-z_2)} \cdot 1 = y^* - z_2 \#$$

$$\frac{\partial L}{\partial z_1} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial z_1} = \frac{y^* - z_2}{z_2(1-z_2)} \cdot \frac{\partial(\sigma(w_2^T z_1 + b_2))}{\partial(w_2^T z_1 + b_2)} \cdot \frac{\partial(w_2^T z_1 + b_2)}{\partial z_1} = \frac{y^* - z_2}{\cancel{z_2(1-z_2)}} \cdot \cancel{z_2(1-z_2)} \cdot w_2^T = w_2^T(y^* - z_2)$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1} = w_2^T(y^* - z_2) \cdot \frac{\partial(\sigma(w_1^T x + b_1))}{\partial(w_1^T x + b_1)} \cdot \frac{\partial(w_1^T x + b_1)}{\partial w_1} = w_2^T(y^* - z_2) \cdot z_1(1-z_1) \cdot x \#$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial b_1} = w_2^T(y^* - z_2) \cdot \frac{\partial(\sigma(w_1^T x + b_1))}{\partial(w_1^T x + b_1)} \cdot \frac{\partial(w_1^T x + b_1)}{\partial b_1} = w_2^T(y^* - z_2) \cdot z_1(1-z_1) \cdot 1 \#$$