$\boxed{3}$ $f(x) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2 \rightarrow (1)$

$$\frac{\partial f}{\partial x_1} = 2x_1 + 8$$

$\partial\, df = 2x_1 - 4x_2 + 20 \rightarrow (2)$

$$\frac{\partial f}{\partial x_2} = -4x_2 + 12$$

$$\frac{\partial f}{\partial x_1} = 0 \Rightarrow x_1 = -4$$

$$\frac{\partial f}{\partial x_2} = 0 \Rightarrow x_2 = 3$$

$\therefore$ Stationary point := $(-4, 3)$

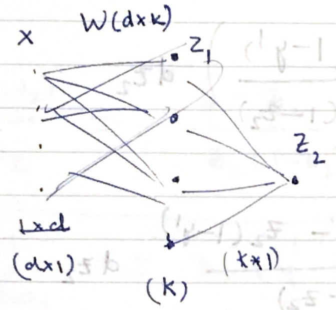
Since (1) is quadratic; its differential (2) is linear.
And two lines can intersect at only one point.
$\therefore$ The stationary point is unique.

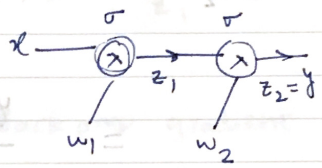
$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix}$$

$\Delta = |H| = -8 - 0 = -8 < 0$

$\therefore$ It is a saddle point

$\boxed{5}$





$$Z_1 = \cancel{Wx+b}$$

$$a_1 = w_1 x + b_1$$

$$z_1 = \sigma(a_1) = \frac{1}{1+\exp(-\underbrace{(w_1^T x + b_1)}_{a_1})}$$

$$z_2 = \cancel{\sigma(z_1)}\ \sigma(\underbrace{w_2 z_1 + b_2}_{a_2})$$

$$= \frac{1}{1+\exp(-w_2 z_1 - b_2)}$$

Forward computing

$$\boxed{z_2(x) = \frac{1}{1+\exp\left(-\dfrac{w_2}{1+\exp(-w_1^T x - b_1)} - b_2\right)}}$$

$$L = y' \log z_2 + (1-y')\log(1-z_2)$$

$$\frac{dL}{dw_2} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_2}$$

$$= \frac{y'}{z_2} dz_2 + \frac{(1-y')}{(1-z_2)}(-dz_2)$$

$$\frac{dL}{dw_2} = \left(\frac{y'}{z_2} - \frac{(1-y')}{(1-z_2)} \right) dz_2$$

$$= \frac{y'(1-z_2) - z_2(1-y')}{z_2(1-z_2)} \, dz_2$$

$$= \left(\frac{y' - \cancel{z_2 y'} - z_2 + \cancel{z_2 y'}}{z_2(1-z_2)} \right) dz_2$$

$$= \left(\frac{y' - z_2}{z_2(1-z_2)} \right) dz_2$$

$$\cancel{z_2 = \frac{1}{1+e}} \qquad z_2 = \frac{1}{1 + \exp(-(\underbrace{w_2 z_1 + b_2}_{h_2}))}$$

$$z_2(h_2) = \frac{1}{1 + \exp(-h_2)}$$

$$\frac{\partial z_2}{\partial h_2} = z_2(1-z_2)$$

$$\frac{\partial h_2}{\partial w_2} = z_1$$

$$\therefore \frac{\partial L(w_2)}{\partial w_2} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_2} \cdot \frac{\partial h_2}{\partial w_2}$$

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

$$\boxed{\frac{\partial L}{\partial w_2} = z_1 (y' - z_2)}$$ Back prop gradient at output

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1}$$

$$L = y' \log(z_2) + (1-y') \log(1-z_2)$$

lets put $z_2 = \sigma(z_1)$

$$\Rightarrow L = y' \log(\sigma(z_1)) + (1-y') \log(1-\sigma(z_1))$$

$$\frac{\partial L}{\partial z_1} = \frac{y'}{\sigma(z_1)} \sigma'(z_1)$$ $$\boxed{\sigma'(z) = \sigma(z)(1-\sigma(z))}$$

$$= \frac{y'}{\cancel{\sigma(z_1)}} \cdot \cancel{\sigma(z)} \cdot (1-\sigma(z_1))$$

$$= (1-\sigma(z_1))\, y'$$

$$\frac{\partial z_1}{\partial h_1} = z_1(1-z_1) \; ; \quad \frac{\partial h_1}{\partial w_1} = x$$

$$\boxed{\therefore \frac{\partial L}{\partial w_1} = y'\left(1-\sigma(z_1)\right) \cdot z_1(1-z_1) \cdot x}$$

Back prop gradient at hidden layer.

$\boxed{2}$

$f = (x_1+x_2)(x_1x_2 + x_1x_2^2) = x_1x_2^3 + x_1^2x_2^2 + x_1^2x_2 + x_1x_2^2$

$\nabla f = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}$

$= 2x_1x_2^2 + x_2^3 + 2x_1x_2 + x_2^2$
$+ 2x_1^2x_2 + x_1^2 + 3x_1x_2^2 + 2x_1x_2$

$= 5x_1x_2^2 + 4x_1x_2 + x_1^2 + x_2^2 + 2x_1^2x_2 + x_2^3$

$\nabla^2 f = \frac{\partial \nabla f}{\partial x_1} + \frac{\partial \nabla f}{\partial x_2}$

$= 2x_1 + 4x_2 + 5x_2^2 + 4x_1x_2$
$+ 2x_1^2 + 10x_1x_2 + 4x_1 + 3x_2^2 + 2x_2$

$= 6x_1 + 6x_2 + 8x_2^2 + 14x_1x_2 + 2x_2^2$

Stationary points $\Rightarrow$ $\cancel{\nabla f = 0}$ $\frac{\partial f}{\partial x_1} = 0$ and $\frac{\partial f}{\partial x_2} = 0$.

Hessian

$(x_1, x_2) : \Rightarrow (0,0) \rightarrow \cancel{(0,0)}$ 0

$\Rightarrow \left(\frac{3}{8}, \frac{-6}{8}\right) \rightarrow \cancel{1.125}$ -0.56

$\Rightarrow (-0.2417, 1) \rightarrow \cancel{11.58}$ 11.16

[P/O]

$$\frac{\partial f}{\partial x_1} = 2x_1x_2^2 + x_2^3 + 2x_1x_2 + x_2^2$$

$$\frac{\partial f}{\partial x_2} = 2x_1^2x_2 + x_1^2 + 2x_1x_2 + 3x_1x_2^2$$

Solving $\frac{\partial f}{\partial x_1} = 0$ and $\frac{\partial f}{\partial x_2} = 0$ simultaneously

$(x_1, x_2) = (0, -1) = \begin{bmatrix} 0 & 5 \\ -8 & 0 \end{bmatrix}$

$(x_1, x_2) = (0, 0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

$(x_1, x_2) = (1, -1) = \begin{bmatrix} 0 & 3 \\ -10 & -2 \end{bmatrix}$

Stationary points and their Hessians

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial x_1^2} = 2x_2^2 + 0 + 2x_2 + 0$$

$$\frac{\partial^2 f}{\partial x_2^2} = 2x_1^2 + 0 + 2x_1 + 6x_1x_2$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = 4x_1x_2 + 3x_2^2 + 2x_1 + 2x_2^2$$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = 4x_1x_2 + 2x_1 + 2x_2 + 6x_2$$

$$H\left(\frac{3}{8}, -\frac{6}{8}\right) = \begin{bmatrix} -0.375 & 2.43 \\ -6.37 & -0.65 \end{bmatrix}$$

we can cleary see that $|H| > 0$ and $\frac{\partial^2 f}{\partial x_1^2} < 0$

$\Rightarrow \left(\frac{3}{8}, \frac{-6}{8}\right)$ is a maximum.

There are only 4 stationary points for this function,

All others have either $|H| \leq 0$ ~~or~~ and/or

$\frac{\partial^2 f}{\partial x_1^2} \geq 0$; making those points anything

but maximum.

$\therefore$ only local maximum of 'f' is $\left(\frac{3}{8}, \frac{-6}{8}\right)$

1

let random variable x represents a random point in $\mathbb{B}$

~~Accordi~~ if $x_1, x_2, \dots x_N$ are N points in $\mathbb{B}$

such that $x_1 \leq x_2 \leq \dots x_N$;

then marginal dist of x_1 (minimum point from origin)

is given by : $f_{x_1}(x) = N[1 - F(x)]^{N-1} f(x) \rightarrow (1)$

CDF of x PDF of x

we know area under PDF (1) is 1

$\therefore$ median value would be at $\frac{1}{2}$ way through

CDF of PDF (1). $\rightarrow$ (2)

$\therefore$ Area under CDF $(f_{x_1}(x)) = \frac{1}{2} \longrightarrow$ (2)

Probability that x_k lies anywhere at a distance d in $\mathbb{B}$ is

$$F(d) = \frac{V_p(d)}{V_p(1)}$$

$$F(d) = d^p \hookrightarrow (4)$$

derivative of CDF is PDF

$$f(d) = p d^{p-1} \hookrightarrow (3)$$

Independently for 2, deriving CDF of $f_{x_1}(x)$ ~~for all N~~

$$\int N[1-F_x(x)]^{N-1} f(x)\, dx$$

putting 3, 4;

$$\int N[\underbrace{1-d^p}_{v}]^{N-1} \cdot p \cdot d^{p-1} \cdot dp \rightarrow 5$$

$$v = 1-d^p$$
$$dv = -pd^{p-1} \cdot dp.$$

From above substitution 5 $\Rightarrow$

$$= -\int N v^{N-1} \cdot dv$$

$$\boxed{d(1000, 10000) = 0.99}$$

$$= -\frac{N}{N} v^N$$

$$CDF(f_{x_1}(x)) = -(1-d^p)^N$$

From 2;

$$-(1-d^p)^N = \frac{1}{2}$$

$$1-d^p = \frac{1}{2^{1/N}}$$

$$\left(1-\frac{1}{2^{1/N}}\right) = d^p \Rightarrow \boxed{d = \left(1-\frac{1}{2^{1/N}}\right)^{1/p}}$$

$\boxed{4}$ why do we care about positive definiteness?

$$M = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} = A \cdot B$$

$$MI = \lambda I.$$

The eigenvalues of $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ are determined by A, B.

By definition; if all eigen values of a matrix M are positive, then the M is positive definite.

It is given A, B are positive definite, the determinant of positive definite matrix is always positive

This makes eigen values of M positive

$\therefore$ M is pos. definite.