CS-535 Deep Learning

Assignment 1 : By Aashish Adhikari

Q1. Solution

Let R be the radius of the ball.
Since we are considering a unit ball, $R=1$.
The points in N are uniformly distributed and are iids.
The probability that a point x_i lies in the sphere is directly proportional to the volume of this sphere.

Say that the volume of the sphere is $V_p(R)$.

The probability that a point x_i lies in this sphere is thus given as

$$P(d_{x_i} < R) = \frac{k \times R^p}{k}$$

where k is a dimension-dependent constant

Hence, $P(d_{x_i} \geq R) = 1 - R^p$

Since all the points in N are iid's, the probability of these points being further away from the median is therefore the product of the individual probabilities.

$$\prod_{i=1}^{N} P(d_{x_p} \geq R)$$

which gives us the cumulative distribution function.
At the median, we expect this to be 1/2.
Thus, equating them

$$\prod_{i=1}^{N} P(d_{x_i} \geq R) = \frac{1}{2}$$

Scanned by CamScanner

Solving for R gives us

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

Hence, proved.

The median distance for $N=10000$ & $p=1000$ is

$$d(p=1000, N=10000) = \left(1 - \frac{1}{2}^{1/10000}\right)^{1/1000}$$

$$= 0.9904688244 \text{ unit. } \#$$

Q.2. Solution

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

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

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

To find the stationary points, we need to set the gradient to 0 and solve for the values of x_1 & x_2.

$$\Rightarrow 2x_1x_2 + x_2^2 + 2x_1x_2^2 + x_2^3 = 0$$

Or, $x_2(x_2+1)(2x_1+x_2) = 0$

$\Rightarrow x_2 = 0$, $x_2 = -1$, $x_2 = -2x_1$

$$\frac{\partial f(x)}{\partial x_2} = 0$$

or, $x_1(x_1 + 2x_1x_2 + 2x_2 + 3x_2^2) = 0$.

either $x_1 = 0$ or, $x_1 + 2x_1x_2 + 2x_2 + 3x_2^2 = 0$.

put $x_2 = -1$

$0 = x_1 - 2x_1 - 2 + 3$

or, $0 = -x_1 + 1$

or, $x_1 = 1$

Thus, $(1,-1)$

put $x_2 = -2x_1$

or, $0 = x_1(-3x_1 + 3x_1^2)$

or, $0 = x_1^2(8x_1 - 3)$

$\Rightarrow x_1 = 0$

$\Rightarrow x_1 = 3/8$.

Thus, $(0,0)$

$(3/8, -6/8)$

$= (3/8, -3/4)$

So, 3 stationary points are $(0,0)$, $(1,-1)$ & $(3/8, -6/8)$.

Hessian Matrix: $H(f_{xy}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$

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

Scanned by CamScanner

$\underline{At (0,0)}$ $\quad H\left(f x_1, x_2\right) = 0$

At $(1,-1)$ $\quad H\left(f x_1, x_2\right) = \begin{bmatrix} 0 & 2-4+2+3 \\ 3 & 2+2-6 \end{bmatrix}$

$$= \begin{bmatrix} 0 & -1 \\ -1 & -2 \end{bmatrix}$$

Hessian $= -1$ ~~$0 - 9 = -9$~~

At $\left(3/8, -6/8\right)$

$$\begin{bmatrix} 2\times\left(-6/8\right) + 2\times\left(\frac{-6}{8}\times\frac{-6}{8}\right) & 2\times\frac{3}{8} + 4\times\frac{3}{8}\times\left(\frac{-6}{8}\right) + 2\times\left(\frac{-6}{8}\right) + 3\times\left(\frac{-6}{8}\right)\times\left(\frac{-6}{8}\right) \\ 2\times\left(3/8\right) + 4\times 3/8\times\left(-6/8\right) + 2\left(-6/8\right) + 3\times\left(-6/8\right)\left(-6/8\right) & 2\times\frac{3}{8} + 6\times\left(3/8\right)\left(-6/8\right) + 2\times\left(3/8\right) \end{bmatrix}$$

$$\begin{bmatrix} \frac{-6}{4} + \frac{9}{8} & \frac{3}{4} + \frac{3}{2}\times\left(\frac{-6}{8}\right) + \left(\frac{-6}{4}\right) + \frac{108}{8} \\ 3/8 & 3/4 - 3\times\frac{36}{64} + 6/8 \end{bmatrix}$$

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

$\therefore$ Hessian $= \frac{63}{256} - \frac{9}{64}$

$= \frac{27}{128}$ #

At $(0,0)$, Hessian $=0$
At $(1,-1)$, Hessian $<0 \implies$ can't be a local maxima
At $\left(3/8, -6/8\right)$, $\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2} < 0$ & $\Delta > 0$,

thus, it is the only local maxima.

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

Solution

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

$$\frac{\partial f(x)}{\partial x_2} = 12 - 4x_2$$

Setting to zero;

$x_1 = -4$ $\qquad$ $x_2 = 3$

$\Rightarrow (-4,3)$ is the only stationary point

Hessian Matrix $= \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix}$

Hessian $= -8$

Scanned by CamScanner

Since $\Delta < 0$, it is a saddle point

Q.4 Solution:

Say $C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$

Since A & B are positive ~~semi~~-definite, let

$z_a^T A z_a > 0$ for all compatible nonzero column vectors z_a.

Lets assume A & B are 2x2 matrices.

$A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$ $B = \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}$

Also $z_b^T B z_b > 0$ for all compatible nonzero column vectors z_b.

So, $C = \begin{bmatrix} a_1 & a_2 & 0 & 0 \\ a_3 & a_4 & 0 & 0 \\ 0 & 0 & b_1 & b_2 \\ 0 & 0 & b_3 & b_4 \end{bmatrix}$

Let $z_c^T C z_c$ for all compatible non-zero column vectors z_c.

$z_c^T = [z_{c1} \ z_{c2} \ z_{c3} \ z_{c4}] = [z_{a1} \ z_{a2} \ z_{b1} \ z_{b2}]$

We need to show $z_c^T C z_c > 0$.

A

$\therefore z_c^T C z_c$

$$= \begin{bmatrix} z_{c1} & z_{c2} & z_{c3} & z_{c4} \end{bmatrix}_{1\times4} \begin{bmatrix} a_1 & a_2 & 0 & 0 \\ a_3 & a_4 & 0 & 0 \\ 0 & 0 & b_1 & b_2 \\ 0 & 0 & b_3 & b_4 \end{bmatrix}_{4\times4} \begin{bmatrix} z_{c1} \\ z_{c2} \\ z_{c3} \\ z_{c4} \end{bmatrix}$$

$$= \begin{bmatrix} z_{c1}a_1 + z_{c2}a_3 & z_{c1}a_2 + z_{c2}a_4 & z_{c3}b_1 + z_{c4}b_3 & z_{c3}b_2 + z_{c4}b_4 \end{bmatrix} \begin{bmatrix} z_{c1} \\ z_{c2} \\ z_{c3} \\ z_{c4} \end{bmatrix}$$

$$= \begin{bmatrix} \text{---} \; P \; \text{---} \end{bmatrix} \begin{bmatrix} | \\ Q \\ | \end{bmatrix}$$

However, since A is true for any z_{a1} & z_{a2} z_{b1} & z_{b2},

$z_{c1} = z_{a1}$ $\quad$ $z_{c3} = z_{b1}$

$z_{c2} = z_{a2}$ $\quad$ $z_{c4} = z_{b2}$

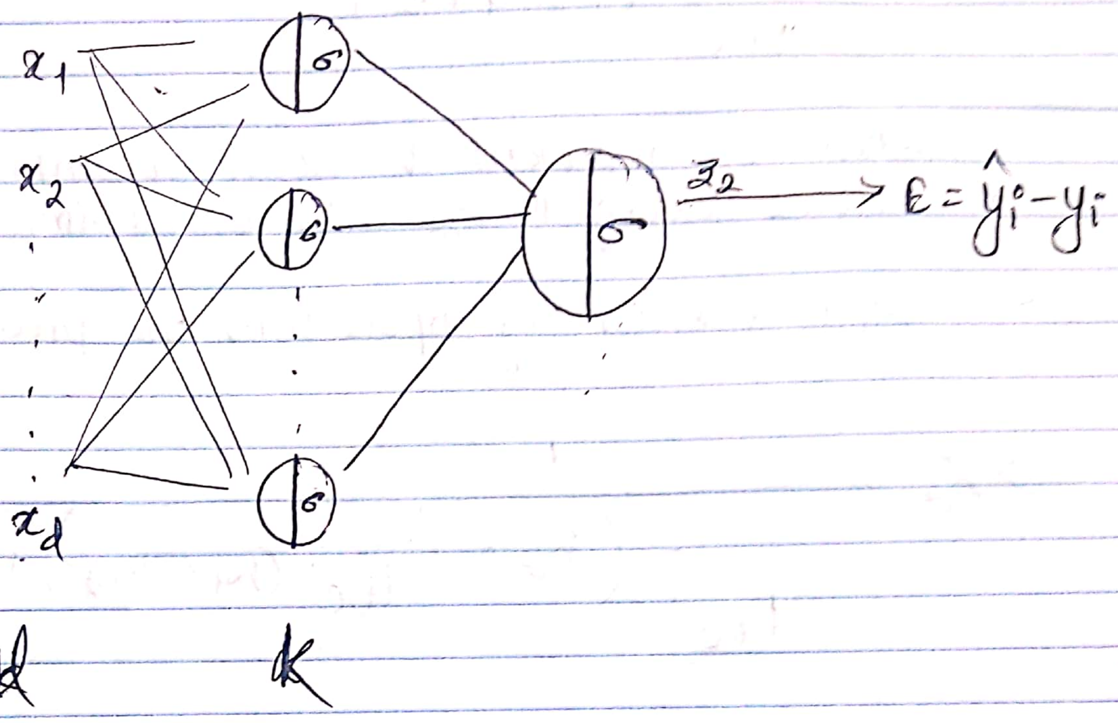
Hence all elements in P above are positive and all elements in Q are also positive by the definition.

Hence, $z_c^T C z_c > 0$ for all ~~non-neg~~ z_c

This proof can be extended for any dimension of A & B as long as it is $n \times n$.

Hence, proved. #

Q 5. Solution :



Each node in the hidden layer has its own sigmoid function and produces one element for the vector

$$z_1 = \begin{bmatrix} z_{11} \\ z_{12} \\ \vdots \\ z_{1k} \end{bmatrix}$$

However, z_2 is a single value.

$$z_{1i} = \frac{1}{1+e^{-(W_{1i}^T X + b_{1i})}}$$

where W_{1i} represents the set of weights in the first layer corresponding to node i & b_{1i} follows accordingly.

z_1 = represents all such z_{1i}'s.

$$Z_2 = \frac{1}{1+e^{-(W_2^T Z_1 + b_2)}}$$

where W_2 represents a vector representing all the weights leading to Z_2 & b_2 follows similarly.

Z_1 is a vector of inputs from the previous layer.

$$\Rightarrow \cancel{Z_2 = \frac{1}{1+e^{-\left(W_2 \times \left(\frac{1}{1+e^{-(W_1 X + b_1)}}\right) + b_2\right)}}}$$

$\Rightarrow$ Hence forward propagation gives Z_2 as.

$$Z_2 = \frac{1}{1+e^{-(W_2^T Z_1 + b_2)}}.$$

where $W_2^T = \begin{bmatrix} \text{———} \; W_2^T \; \text{———} \end{bmatrix}_{1 \times k}$

$$= \begin{bmatrix} W_{21} & W_{22} & \cdots \end{bmatrix}_{1 \times k}$$

$$Z_1 = \begin{bmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1k} \end{bmatrix}_{k \times 1}$$

z_2 can also be written as.

$$z_2 = \frac{1}{1+e^{-\left(w_{21}z_{11}+w_{22}z_{12}+w_{23}z_{13}+\dots+b_2\right)}}$$

where each $z_{1i} = \frac{1}{1+e^{-\left(w_{1i}^{T}x+b_{1i}\right)}}$

and ~~every~~
$w_{1i}^{T}x$ is itself a summation as $\sum_{j=1}^{k} w_{1j} \cdot x_j$

Thus, ~~backpropag~~ forward propagation is over at z_2.

Now, back propagation:

$$E = y^* \log z_2 + (1-y^*)\log(1-z_2)$$

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

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

Also, $E = y^* \log\left(\frac{1}{1+e^{-(w_2 z_1+b_2)}}\right) + (1-y^*) \log \left[1 - \frac{1}{1+e^{-(w_2 z_1+b_2)}}\right]$

Now, $\frac{\partial E}{\partial w_2} = \frac{\partial E}{\partial z_2} \times \frac{\partial z_2}{\partial w_2}$

$$= \frac{y^* - z_2}{z_2(1-z_2)} \times \frac{\partial\left(\sigma(w_2^T z_1 + b_2)\right)}{\partial w_2}$$

$$= \frac{y^* - z_2}{z_2(1-z_2)} \times \frac{\partial\left[\sigma(w_2^T z_1 + b_2)\right]}{\partial(w_2^T z_1 + b_2)} \times \frac{\partial(w_2^T z_1 + b_2)}{\partial w_2}$$

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

$$= \frac{y^* - z_2}{z_2(1-z_2)} \times z \times (1-z_2) \times z_1$$

$$= (y^* - z_2) \times z_1$$

Similarly, $\frac{\partial E}{\partial b_2} = (y^* - z_2)$

Now, $\frac{\partial E}{\partial z_1} = \frac{\partial E}{\partial z_2} \times \frac{\partial z_2}{\partial z_1}$

$$= \frac{y^* - z_2}{z_2(1-z_2)} \times z_2 \times (1-z_2) \times w_2^T$$

$$= (y^* - z_2)\, w_2^T$$

$$\frac{\partial E}{\partial W_1} = \frac{\partial E}{\partial Z_1} \times \frac{\partial Z_1}{\partial W_1}$$

$$= (y^* - Z_2) \times W_2^T \times \frac{\partial Z_1}{\partial A_1} \times \frac{\partial A_1}{\partial W_1}$$

$$= (y^* - Z_2) \times W_2^T \times Z_1 \times (1 - Z_1) \times X$$

Similarly, $\frac{\partial E}{\partial b_1} = (y^* - Z_2) \times W_2^T \times Z_1 (1 - Z_1)$

Hence, all required terms were also calculated.