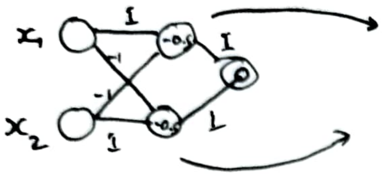
Assigenment- 3

AI20BTECH11006

1)

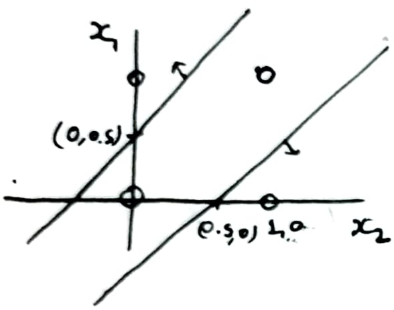
a) XOR

$XOR(x_1, X_2) = x_1 \cdot \bar{x}_2 + \bar{x}_1 \cdot \bar{x}_2$



$x_1 - x_2 \geq 0.5$

$x_2 - x_1 \geq 0.5$



$x_1 = 0 \quad x_2 = 0$

then $x_1 - x_2 \geq 0.5 \rightarrow 0$

$x_2 - x_1 \geq 0.5 \rightarrow 0$

taking or of this $\rightarrow 0$

$x_1 = 1 \quad x_2 = 1$

then $x_1 - x_2 \geq 0.5 \rightarrow 0$

$x_2 - x_1 \geq 0.5 \rightarrow 0$

taking or of this $\rightarrow 0$

$x_1 = 0 \quad x_2 = 1$

then

$x_1 - x_2 \geq 0.5 \rightarrow 0$

$x_2 - x_1 \geq 0.5 \rightarrow 1$

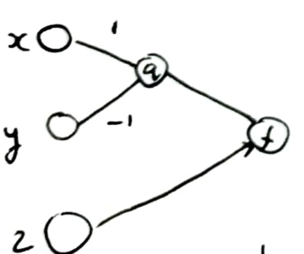
output: 1

b) $q = x - y$

$f = q * z$

$q = -2 - 5 = -7$

$f = -7 \times 2 = 28$



$f = q * z$

(the product of inputs)

$f = z(x - y) = xz - yz = 28$

$\frac{\partial f}{\partial x} = z = -4$

$\frac{\partial f}{\partial y} = -z = 4$

$\frac{\partial f}{\partial z} = x - y = -7 = q$

2)

Error function

$$E(\underline{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \ln(y_k(\underline{x}_n, \underline{w})) \quad — (1)$$

The softmax function is given as

$$y_k(\underline{x}, \underline{w}) = \frac{\exp(a_k(\underline{x}, \underline{w}))}{\sum_j \exp(a_j(\underline{x}, \underline{w}))} \quad — (2)$$

from (2)

$$\frac{\partial y_k}{\partial a_k} = \frac{\partial \frac{\exp(a_k(\underline{x}, \underline{w}))}{\sum_j \exp(a_j(\underline{x}, \underline{w}))}}{\partial a_k} = \frac{\exp(a_k(\underline{x}, \underline{w}))\left(\sum_j \exp(a_j(\underline{x}, \underline{w})) - \exp(a_k(\underline{x}, \underline{w}))\right)}{\left(\sum_j \exp(a_j(\underline{x}, \underline{w}))\right)^2}$$

$$= y_k(1 - y_k) \quad — (3)$$

$$\frac{\partial y_k}{\partial a_t} \; (t \neq k) = \frac{\partial \frac{\exp(a_k(\underline{x}, \underline{w}))}{\sum_j \exp(a_j(\underline{x}, \underline{w}))}}{\partial a_t} = \frac{-\exp(a_k(\underline{x}, \underline{w}))\exp(a_t(\underline{x}, \underline{w}))}{\left(\sum_j \exp(a_j(\underline{x}, \underline{w}))\right)^2}$$

$$= -y_k y_t \quad — (4)$$

Now, we will calculate the partial derivative of cross entropy wrt a_k

$$E = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \ln(y_k(\underline{x}_n, \underline{w}))$$

for further steps y_{kn} will represent

$y_{kn} = y_k(\underline{x}_n, \underline{w})$

$$E = -\sum_{n=1}^{N} \sum_{k=1}^{K} t_{kn} \ln(y_{kn})$$

$$\frac{\partial E}{\partial a_l} = -\frac{\partial\left(\sum_{n=1}^{N} \sum_{k=1}^{K} t_{kn} \ln(y_{kn})\right)}{\partial a_l} = -\sum_{n=1}^{N}\left(\sum_{k\neq l}^{K} \frac{t_{kn}}{y_{kn}} \frac{\partial y_{kn}}{\partial a_l} + \frac{t_{ln}}{y_{ln}} \frac{\partial y_{ln}}{\partial a_l}\right)$$

Now, using 3 & 4
(both are independent of n)

$$\frac{\partial E}{\partial a_l} = -\sum_{n=1}^{N}\left(\sum_{k\neq l}^{K} \frac{-t_{kn} \cancel{y_{kn}} (y_{ln})}{\cancel{y_{kn}}} + \frac{t_{ln}(1-y_{ln}) \cancel{y_{ln}}}{\cancel{y_{ln}}}\right)$$

$$= -\sum_{n=1}^{N}\left(t_{ln} - \sum_{k=1}^{K} t_{kn} y_{ln}\right)$$

$$\because \sum_{k=1}^{K} t_{kn} = 1$$
since it is hot vector

$$= -\sum_{n=1}^{N}\left(t_{ln} - y_{ln}\right)$$

$$= y_l - t_l$$

Hence Proved

3) Jensen's inequality

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

In probability

$$\varphi(E[X]) \leq E[\varphi(X)] \quad — (1)$$

$$E_{AV} = \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}_x\left[(y_m(x) - f(x))^2\right] \quad \Big| \quad E_{ENS} = E_x\left[\left(\frac{1}{M}\sum_{m=1}^{M} y_m(x) - f(x)\right)^2\right]$$

~~Here let X = ym(x~~

Here $\varphi(x)$ is $E_x\left((y_m(x) - f(x))^2\right)$

$E(\varphi(x))$ basically refers to expected / average mean squared error over M classifiers

Now, $\varphi[E(x)]$

$$= E_x\left(\left(\frac{1}{M}\sum_{m=1}^{M} (y_m(x) - f(x)\right)^2\right]$$

$$= E_x\left(\left(\frac{1}{M}\sum y_m(x) - f(x)\right)^2\right)$$

Using (1)

$E_{ENS} < E_{AV}$ $\because \varphi(x)$ is convex

$\because f(x)$ is convex

Now, what we did is not limited to mean squared error.

expectation conserves convexity

Proof

let $z(x)$ be convex

$\varphi(x) = E_x(z(x))$

$\varphi(\alpha x_1 + (1-\alpha)x_2) = E_{\alpha, x_1, x_2}\left[z(\alpha x_1 + (1-\alpha)x_2)\right]$

$\cancel{\leq E_{\alpha, x_1, x_2}\left[\alpha z(\alpha x_1) + z((1-\alpha)x_2)\right]}$

$\leq E_{\alpha, x_1, x_2}\left[\alpha z(x_1) + (1-\alpha)z(x_2)\right]$

$\because$ Jensen's inequality

$= \alpha E_{\alpha, x_1, x_2}[z(x_1)] + (1-\alpha) E_{\alpha, x_1, x_2}(z(x_2))$

Again from Jensen's inequality, we can say the $\varphi(x)$ is convex.

Now for any $E(y)$ which is convex, its expectation will also be convex

take $\varphi(x) = E_x(E(x))$

then by Jensen's inequality the result holds for all convex functions because $\varphi(x)$ will also be convex.