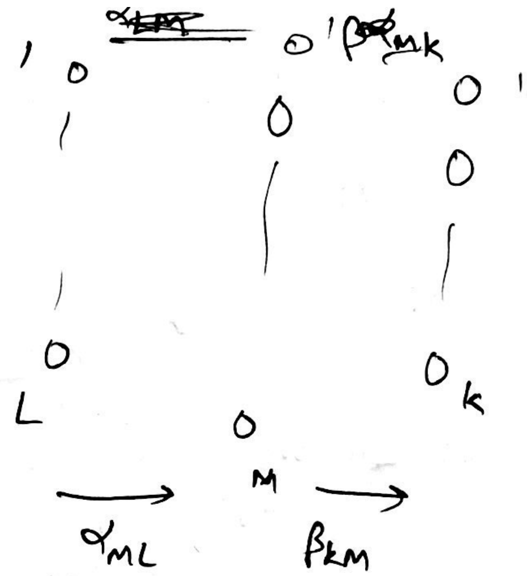
~~We can use the~~

1A) Given configuration of the multi layer Perceptron.



$$\frac{\partial R_i(\theta)}{\partial \beta_{km}} = \frac{\partial}{\partial \beta_{kM}}\left[\sum_{i=1}^{k} (y_{ki} - f_k(x_i,\theta))^2\right]$$

$$= \frac{\partial}{\partial \beta_{km}}\left[\left(\sum_{k=1}^{k} y_{ki} - g_k\left(\begin{bmatrix} \beta_1^T z_i + \beta_{10} \\ | \\ \beta_k z_i + \beta_{k0} \end{bmatrix}\right)\right)^2\right]$$

$$= -2\left[(y_{ki} - g_k(I))\right] \cdot g_k'(I) \cdot \_ z_{mi}$$

$$\beta_k^T z_i + \beta_{k0} = \beta_{k1} z_{1i} + \beta_{k2} z_{2i} + \_ + \beta_{km} z_{mi} + \_ + \beta_{k0}$$

$$\boxed{\therefore \frac{\partial R_i(\theta)}{\partial \beta_{km}} = -2\left[(y_{ki} - g_k(I))\right] \cdot g_k'(T)\, z_{mi}}$$ → gradients for final layer

$$\frac{\partial R_i(\theta)}{\partial \alpha_{ml}} = \frac{\partial}{\partial \alpha_{ml}}\left[\sum_{k=1}^{k} (y_{ki} - g_k(T_k))^2\right]$$

$$T_k = \beta_k^T z + \beta_{k0}$$

$$= [\beta_{k1} \; \_ \; \beta_{kM}]\begin{bmatrix} z_1 \\ | \\ z_m \end{bmatrix} + \beta_{k0}$$

$$z_m = \sigma(\alpha_m^T x + \alpha_{m0}) = \sigma\left(\sum_{j=1}^{l} \alpha_{mj} x_j + \alpha_{m0}\right)$$

$$= \sigma(\alpha_{m1} x_1 + \alpha_{m2} x_2 + \_ + \alpha_{ml} x_l + \alpha_{m0})$$

$$\frac{\partial R_i(\theta)}{\partial \alpha_{ml}} = \frac{\partial}{\partial \alpha_{ml}}\left[\sum_{k=1}^{k} (y_{ki} - g_k(T))^2\right]$$

$$\boxed{\frac{\partial R_i(\theta)}{\partial \alpha_{ml}} = -2\sum_{k=1}^{k}\left[(y_{ki} - g_k(T))\right] \cdot g_k'(T) \cdot \beta_{km}\, \sigma'(\alpha_m^T x_i + \alpha_{m0}) \cdot x_{li}}$$

gradients for hidden layer

we can use those gradients to approach at our final parameters.

Scanned with CamScanner

(2) $E[\omega] \approx E[\tilde{\omega}] + (\omega - \tilde{\omega})^T \nabla E|_{\omega = \tilde{\omega}} + \frac{1}{2}(\omega - \tilde{\omega})^T H (\omega - \tilde{\omega})$

$H \rightarrow$ hessian

$\omega = [\omega_1, \text{——}, \omega_n)$

$\tilde{\omega} = [\tilde{\omega}_1', \omega_2', - \omega_n']$

$E(\omega) \rightarrow$ Quadratic error function

$= \sqrt{\Sigma(\omega_i - \tilde{\omega}_i)^2}$

$E[\tilde{\omega}] \rightarrow$ Constant

$\nabla E|_{\omega = \tilde{\omega}} \rightarrow$ gradient of error function w.r.t ~~each~~ each of the term in the weight vector

so, there will be 'ω' independent elements

$$\frac{1}{2}(\omega - \tilde{\omega})(H(\omega - \tilde{\omega})) \longrightarrow \begin{bmatrix} \frac{\partial^2 E}{\partial \omega_1^2} & & \text{——} & \frac{\partial^2 E}{\partial \omega_1 \partial \omega_\omega} \\ | & \frac{\partial^2 E}{\partial \omega_2^2} & & | \\ | & & \ddots & | \\ \frac{\partial^2 E}{\partial \omega_\omega \partial \omega_1} & & \text{——} & \frac{\partial^2 E}{\partial \omega_\omega^2} \end{bmatrix}$$

This hessian contains

$\omega + \frac{\omega^2 - \omega}{2}$ independent terms

so, the total independent terms in the error function are

$\omega + \omega + \frac{\omega^2 - \omega}{2} = \frac{\omega^2 + 3\omega}{2} = \omega(\omega + 3)/2$

③ A Convex hull is defined as all points x such that

$$x = \sum_i \alpha_i x_i$$

$\alpha_i \geq 0$, $\sum_i \alpha_i = 1$

To show that $\exists\ \bar{\omega}$ such what $\bar{\omega} x + b_1 > 0$, $\bar{\omega} z + b_2 < 0$ ~~Such~~ then the two convex hulls do not intersect.

Assumption: Convex hulls intersect, pts linearly seperable

ⓐ Proof by Contradiction

if the two hulls intersect,

then there will be some common points between the two hulls

$\cancel{w^T x_i + w_0 > 0}$,

$$\sum \alpha_i x_i = \sum \beta_i z_i$$

$$f(x) = \sum_i (w^T \alpha_i x_i) + w_0 = \sum_i \alpha\, w^T x_i + w_0$$

$$= \sum_i \beta\, w^T z_i + w_0$$

$$= f(z)$$

but since both the hulls are linearly seperable it should not be the case that $f(x) = f(z)$

it should be either $f(x) > 0$, $f(z) < 0$ or vice versa

which contradicts with our assumptions that the two convex hulls intersect.

ⓑ Assume the Convex hulls are linearly seperable, and the convex hulls intersect

Since the hulls are linearly seperable.

for any of α $f(x) = \sum_i (w^T \alpha_i x_i + w_0) > 0$ and for every 'β'

$$f(z) = \sum_i (w^T \beta_i z_i) + w_0 < 0$$

if their hulls intercect,

there should exist atleast one set of α, β

such that $f(x) = f(z)$

but since $f(x) > 0$, $f(z) < 0$

is a direct contradiction to the assumption $(f(x) = f(z))$

which means the hulls does not overlap.

from the above two methods we proved that if two convex hulls are linearly seperable, then they do not overlap and so is their vice versa