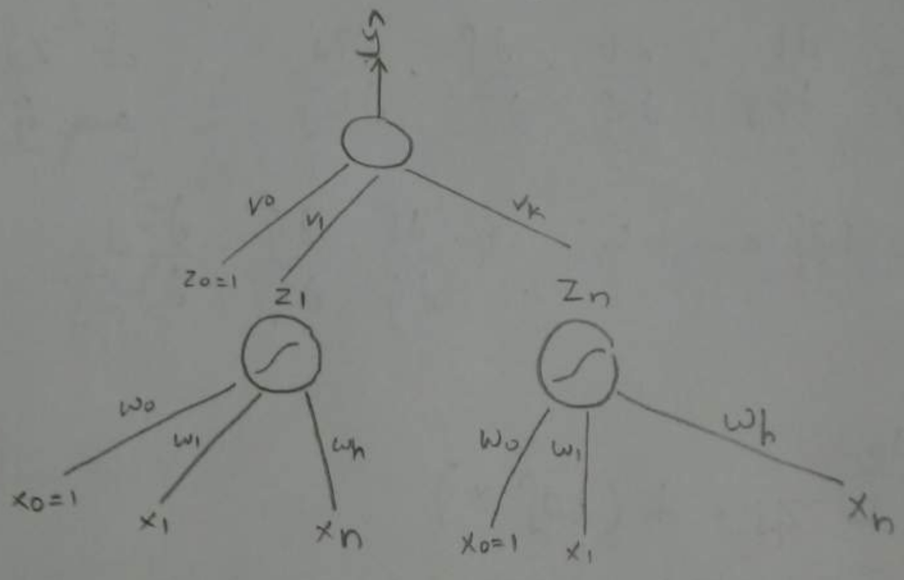
2.1] MLP

Solution :

Given : Error function : $E(\{w_i\}, v)$

$$= \frac{1}{2} \sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)})^2$$

→ Using Back propagation algorithm, after updating the parameter of the i^{th} layer, we update the parameter of the $(i-1)^{th}$ Layer.



Here;

X represents the original input

Z represents the desired input which is the output of the input layer

W is the parameter of input layer

V is the parameter of hidden Layer

$\hat{y}$ is the predicted final output.

We update the parameters to minimize the Error function.

→ Updating V : $V \leftarrow V - \eta \cdot \frac{\partial E}{\partial V}$

Now $\frac{\partial E}{\partial V} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V}$ ... ~~here~~ as $\hat{y}$ depends on V and E depends on $\hat{y}$

$\therefore V \leftarrow V - \eta \cdot \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V}$

→ Updating W : $w_j \leftarrow w_j - \eta \cdot \frac{\partial E}{\partial w_j}$

Now $\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_j}$ ... as z_j depends on w_j and $\hat{y}$ depends on z_j

$\therefore w_j \leftarrow w_j - \eta \cdot \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_j}$

→ Now

↳ $z_j = h(w_j^T x)$

$\therefore \frac{\partial z_j}{\partial w_j} = z_j(1 - z_j)\, x$

↳ $\hat{y} = v^T z$

$\therefore \frac{\partial \hat{y}}{\partial v} = z$

- Expanding the derivatives

i) $$\frac{\partial E}{\partial V} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V}$$

$$= 2 \cdot \frac{1}{2} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) \cdot z^{(i)}$$

$$= \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) \cdot Z^{(i)}$$

ii) $$\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_j}$$

$$= \cancel{2} \cdot \frac{1}{\cancel{2}} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) \cdot V_j \cdot Z_j^{(i)} \left(1 - Z_j^{(i)}\right) x^{(i)}$$

$$= \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) \cdot V_j \cdot Z_j^{(i)} \left(1 - Z_j^{(i)}\right) x^{(i)}$$

- Update rule for V:

$$V \leftarrow V - \eta \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) \cdot z^{(i)}$$

- Update rule for w_j:

$$w_j \leftarrow w_j - \eta \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) V_j \cdot Z_j^{(i)} \left(1 - Z_j^{(i)}\right) x^{(i)}$$

→ ∴ Update eqns are the same as those for maximum Likelihood estimation.

→ ∴ Update equations for regression & classification are <u>same</u>.