Q1(a) $W_1 = (0.01, 0.02, 0.03)$

$W_2 = (0.03, 0.01, 0.02)$

$W_3 = (0.02, 0.03, 0.01)$

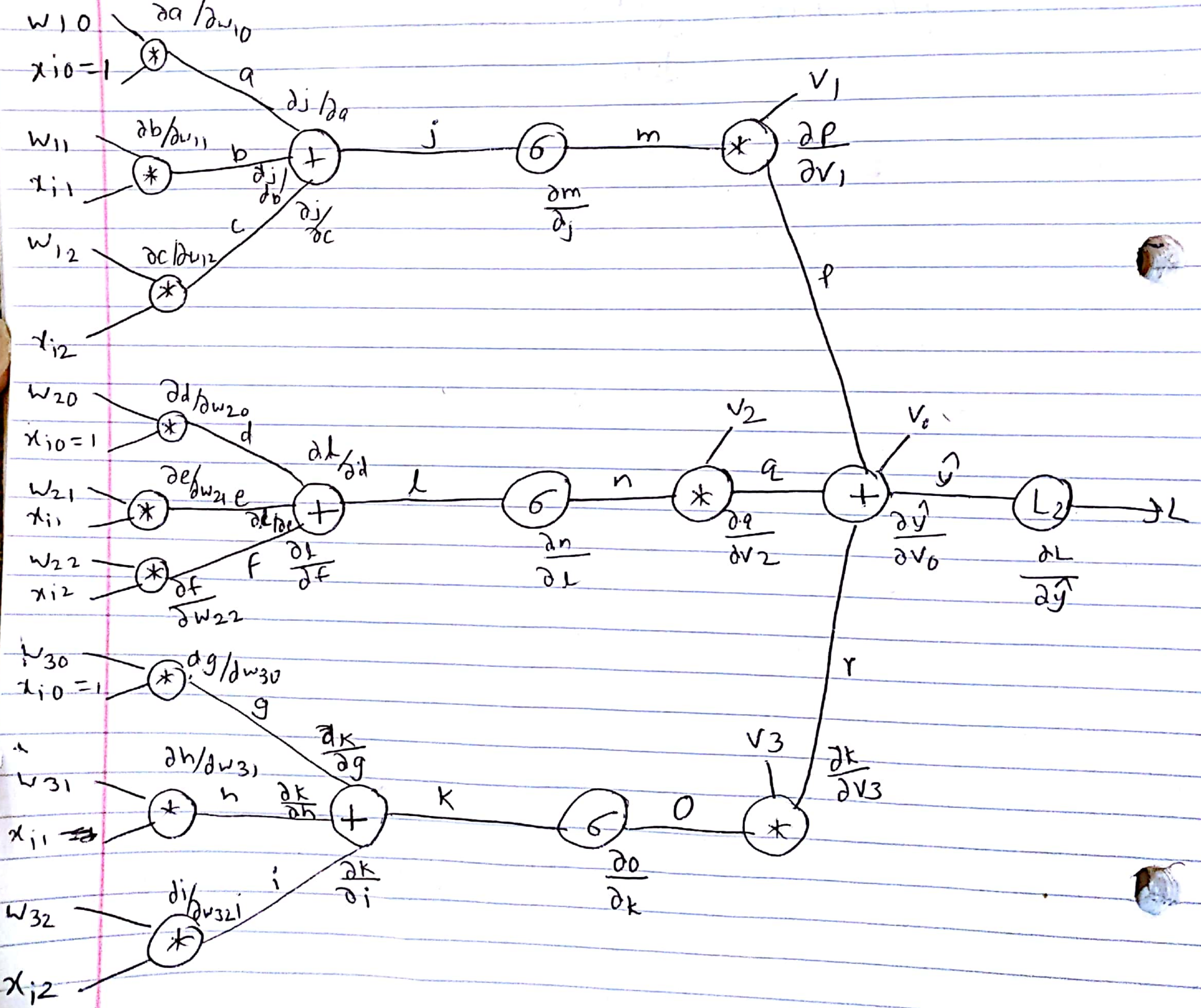
$V = (0.01, 0.02, 0.03, 0.04)$

$(x_1, y_1) = [(1,2), 8)]$

$(x_2, y_2) = [(1,3), 11]$

$(x_3$

Computation graph



b) Values for the forward pass network

1) for first network input layer : $\boxed{i=1}$

$a = w_{10} * x_{i0} = 0.01 \times 1 = 0.01$

$b = w_{11} * x_{i1} = 0.02 \times 1 = 0.02$

$c = w_{12} * x_{i2} = 0.03 \times 2 = 0.06$

$j = a+b+c = 0.01 + 0.02 + 0.06 = 0.09$

$m = \sigma(wx) = \text{Sigmoid}(j) = \frac{1}{1+e^{-j}} = \frac{1}{1+e^{-0.09}} = 0.52248.$

$d = w_{20} * x_{10} = 0.03 \times 1 = 0.03$

$e = w_{21} * x_{i1} = 0.01 \times 1 = 0.01$

$f = w_{22} * x_{i2} = 0.02 \times 2 = 0.04$

$k = d+e+f = 0.08$

$n = \sigma(wx) = \text{sigmoid}(l) = 0.5199$

~~w30 h =~~

$g = w_{30} * x_{i0} = 0.02 \times 1 = 0.02$

$h = w_{31} * x_{i1} = 0.03 \times 1 = 0.03$

$i = w_{32} * x_{i2} = 0.01 \times 2 = 0.02$

$g+h+i = 0.07$

$\sigma(wx) = \text{sigmoid}(o) = 0.517$

(ii)

$p = m \times v_1 = 0.52 \times 0.02 = 0.0104$

$q = n \times v_2 \quad 0.52 \times 0.03 = 0.0156$

$r = o \times v_3 = 0.517 \times 0.04 = 0.0206.$

$\hat{y}_1 = v_0 + p + q + r = 0.01 + 0.0104 + 0.0156 + 0.0206 = \boxed{0.0568}$

Scanned by CamScanner

ii) x_2 : for second input layer : $\boxed{i=2}$

$a = w_{10} \times x_{20} = 0.01 \times 1 = 0.01$
$b = w_{11} \times x_{21} = 0.02 \times 1 = 0.02$
$c = w_{12} \times x_{22} = 0.03 \times 3 = 0.09$

$j = a+b+c = 0.01 + 0.02 + 0.09 = 0.12$
$m = 6(wx) = \text{sigmoid}(j) = \frac{1}{1+e^{-j}} = 0.52.$

$d = w_{20} \times x_{20} = 0.03 \times 1 = 0.03$
$e = w_{21} \times x_{21} = 0.01 \times 1 = 0.01$
$f = w_{22} \times x_{22} = 0.02 \times 3 = 0.06$

$l = d+e+f = 0.03 + 0.01 + 0.06 = 0.1$
$n = 6(wx) = \text{sigmoid}(l) = \frac{1}{1+e^{-l}} = \text{sigmoid}(0.1) = 0.52$

$g = w_{30} \times x_{20} = 0.02 \times 1 = 0.02$
$h = w_{31} \times x_{21} = 0.03 \times 1 = 0.03$
$i = w_{32} \times x_{22} = 0.01 \times 3 = 0.03$

$o = 6(wx) = \text{sigmoid}(k) = \text{sigmoid}(0.08) = 0.519$

$k = 0.02 + 0.03 + 0.02 = 0.08.$

$p = m \times v_1 = 0.52 \times 0.02 = 0.0104$
$q = n \times v_2 = 0.52 \times 0.03 = 0.0156$
$r = o \times v_3 = 0.519 \times 0.04 = 0.0207$

$\hat{y}_2 = v_0 + p+q+r = 0.01 + 0.0104 + 0.0156 + 0.0207$

$\boxed{\hat{y}_2 = 0.0207}$

iii) x_3 : for third input layer $\boxed{i=3}$

$a = w_{10} \times x_{30} = 0.01 \times 1 = 0.01$
$b = w_{11} \times x_{31} = 0.02 \times 2 = 0.04$
$c = w_{12} \times x_{32} = 0.03 \times 2 = 0.06$

$j = a+b+c = 0.01 + 0.04 + 0.06 = 0.11$
$m = \sigma(wx) = \text{sigmoid}(j) = 0.52$

$d = w_{20} \times x_{30} = 0.03 \times 1 = 0.03$

$e = w_{21} \times x_{31} = 0.01 \times 2 = 0.02$
$f = w_{22} \times x_{32} = 0.02 \times 2 = 0.04$

$k = d+e+f = 0.03 + 0.02 + 0.04 = 0.09$
$n = \sigma(wx) = \text{sigmoid}(k) = 0.52$

$g = w_{30} \times x_{30} = 0.02 \times 1 = 0.02$
$h = w_{31} \times x_{31} = 0.03 \times 2 = 0.06$
$i = w_{32} \times x_{32} = 0.01 \times 2 = 0.02$

$K = g+h+i = 0.1$
$o = \sigma(wx) = \text{sigmoid}(K) = \text{sigmoid}(0.1) = 0.52.$

$p = m \times v_1 = 0.52 \times 0.02 = 0.0104$
$q = n \times v_2 = 0.52 \times 0.03 = 0.0156$
$r = o \times v_3 = 0.52 \times 0.04 = 0.0207$

$\hat{y}_3 = v_0 + p + q + r = 0.01 + 0.0104 + 0.0156 + 0.0207$
$\boxed{\hat{y}_3 = 0.0567}$

c)

l_2 loss $= L_2(\sigma(XW) - y)$
$= (\hat{y} - y)^2$

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y)$$

$$\frac{\partial \hat{y}}{\partial v_0} = \frac{\partial \hat{y}}{\partial p} = \frac{\partial \hat{y}}{\partial q} = \frac{\partial \hat{y}}{\partial r} = 1 \quad (\because \hat{y} = v_0 + p + q + r).$$

$$\frac{\partial p}{\partial v_1} = m \; ; \; \frac{\partial m}{\partial j} = m(1-m) \; ; \; \frac{\partial j}{\partial c} = 1 \; ; \; \frac{\partial j}{\partial b} = 1 \; , \; \frac{\partial j}{\partial a} = 1$$

$$\frac{\partial q}{\partial v_2} = n \; ; \; \frac{\partial n}{\partial k} = n(1-n) \; ; \; \frac{\partial k}{\partial f} = 1 \; ; \; \frac{\partial k}{\partial d} = 1 \; ; \; \frac{\partial k}{\partial e} = 1$$

$$\frac{\partial r}{\partial v_3} = o \; ; \; \frac{\partial o}{\partial k} = o(1-o) \; ; \; \frac{\partial k}{\partial g} = 1 \; ; \; \frac{\partial k}{\partial h} = 1 \; ; \; \frac{\partial k}{\partial i} = 1$$

$$\frac{\partial a}{\partial w_{10}} = 1 \; ; \; \frac{\partial b}{\partial w_{11}} = x_{i1} \; ; \; \frac{\partial c}{\partial w_{12}} = x_{i2}$$

$$\frac{\partial d}{\partial w_{20}} = 1 \; ; \; \frac{\partial e}{\partial w_{21}} = x_{i1} \; ; \; \frac{\partial f}{\partial w_{22}} = x_{i2}$$

$$\frac{\partial g}{\partial w_{30}} = 1 \; ; \; \frac{\partial h}{\partial w_{31}} = x_{i1} \; ; \; \frac{\partial i}{\partial w_{32}} = x_{i2}$$

* ~~Gr~~. $\frac{\partial p}{\partial m} = v_1 \; , \; \frac{\partial q}{\partial n} = v_2 \; , \; \frac{\partial r}{\partial o} = v_3$

★ Gradient Descent of the loss (cost) function w.r.t. each weights

$$\frac{\partial \hat{y}}{\partial w_{10}} = \frac{\partial \hat{y}}{\partial p} \times \frac{\partial p}{\partial m} \times \frac{\partial m}{\partial j} \times \frac{\partial j}{\partial a} \times \frac{\partial a}{\partial w_{10}}$$

$$= 1 \times v_1 \times m(1-m) \times 1 \times 1$$

$$\boxed{\frac{\partial \hat{y}}{\partial w_{10}} = v_1 \cdot m(1-m)}$$

$$\frac{\partial \hat{y}}{\partial w_{11}} = \frac{\partial \hat{y}}{\partial p} \times \frac{\partial p}{\partial m} \times \frac{\partial m}{\partial j} \times \frac{\partial j}{\partial b} \times \frac{\partial b}{\partial w_{11}}$$

$$= 1 \times v_1 \times m(1-m) \times 1 \times x_{i1}$$

$$\boxed{\frac{\partial \hat{y}}{\partial w_{11}} = v_1 \times m(1-m)\, x_{i1}}$$

$$\frac{\partial \hat{y}}{\partial w_{12}} = \frac{\partial \hat{y}}{\partial p} \times \frac{\partial p}{\partial m} \times \frac{\partial m}{\partial j} \times \frac{\partial j}{\partial c} \times \frac{\partial c}{\partial w_{12}}$$

$$= 1 \times v_1 \times m(1-m) \times 1 \times x_{i2}$$

$$\boxed{\frac{\partial \hat{y}}{\partial w_{12}} = v_1 \times m \times (1-m) \times x_{i2}}$$

$$\frac{\partial \hat{y}}{\partial w_{20}} = \frac{\partial \hat{y}}{\partial q} \times \frac{\partial q}{\partial n} \times \frac{\partial n}{\partial l} \times \frac{\partial l}{\partial d} \times \frac{\partial d}{\partial w_{20}}$$

$$= 1 \times v_2 \times n(1-n) \times 1 \times 1$$

$$\boxed{\frac{\partial \hat{y}}{\partial w_{20}} = v_2 \times n(1-n)}$$

$$\frac{\partial \hat{y}}{\partial w_{21}} = \frac{\partial \hat{y}}{\partial q} \times \frac{\partial q}{\partial n} \times \frac{\partial n}{\partial l} \times \frac{\partial l}{\partial e} \times \frac{\partial e}{\partial w_{21}}$$

$$= 1 \times v_2 \times n(1-n) \times 1 \times x_{i1}$$

$$\boxed{\frac{\partial \hat{y}}{\partial w_{21}} = v_2 \times n(1-n) \times x_{i1}}$$

$$\frac{\partial \hat{y}}{\partial w_{22}} = \frac{\partial \hat{y}}{\partial q} \times \frac{\partial q}{\partial n} \times \frac{\partial n}{\partial k} \times \frac{\partial k}{\partial f} \times \frac{\partial f}{\partial w_{22}}$$

$$\frac{\partial \hat{y}}{\partial w_{22}} = \cancel{1 \times v_2 \times (1} \;\; \cancel{1 \times n} \;\; 1 \times v_2 \times n \times (1-n) \times 1 \times x_{i2}$$

$$\boxed{\frac{\partial \hat{y}}{\partial w_{22}} = v_2 \times n \times 1 - n \times x_{i2}}$$

$$\frac{\partial \hat{y}}{\partial w_{30}} = \frac{\partial \hat{y}}{\partial r} \times \frac{\partial r}{\partial o} \times \frac{\partial o}{\partial k} \times \frac{\partial k}{\partial g} \times \frac{\partial g}{\partial w_{30}}$$

$$= 1 \times v_3 \times o \;(1-o) \times 1 \times 1$$

$$\boxed{\frac{\partial \hat{y}}{\partial w_{30}} = v_3 \times o \times (1-o)}$$

$$\frac{\partial \hat{y}}{\partial w_{31}} = \frac{\partial \hat{y}}{\partial r} \times \frac{\partial r}{\partial o} \times \frac{\partial o}{\partial k} \times \frac{\partial k}{\partial h} \times \frac{\partial h}{\partial w_{31}}$$

$$= 1 \times v_3 \times o \times (1-o) \times 1 \times x_{i1}$$

$$\boxed{\frac{\partial \hat{y}}{\partial w_{31}} = v_3 \times o \times (1-o \times x_{i1}}$$

$$\frac{\partial \hat{y}}{\partial w_{32}} = \frac{\partial \hat{y}}{\partial r} \times \frac{\partial r}{\partial o} \times \frac{\partial o}{\partial l} \times \frac{\partial l}{\partial i} \times \frac{\partial i}{\partial w_{32}}$$

$$\boxed{\frac{\partial \hat{y}}{\partial w_{32}} = v_3 \times o \times (1-o) \times x_{i2}}$$

★ Gradient of cost function, first training example = $\boxed{(1,2),8}$

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y}-y) = 2(0.0568-8)$$

$$= -15.886.$$

$$\frac{\partial \hat{y}}{\partial v} = \begin{bmatrix} \frac{\partial \hat{y}}{\partial v_0} & \frac{\partial \hat{y}}{\partial v_1} & \frac{\partial \hat{y}}{\partial v_2} & \frac{\partial \hat{y}}{\partial v_3} \end{bmatrix} = \begin{bmatrix} 1 & 0.52 & 0.52 & 0.52 \end{bmatrix}$$

$$\frac{\partial \hat{y}}{\partial w} = \begin{bmatrix} \partial \hat{y}/\partial w_{10} & \partial \hat{y}/\partial w_{11} & \partial \hat{y}/\partial w_{12} \\ \partial \hat{y}/\partial w_{20} & \partial \hat{y}/\partial w_{21} & \partial \hat{y}/\partial w_{22} \\ \partial \hat{y}/\partial w_{30} & \partial \hat{y}/\partial w_{31} & \partial \hat{y}/\partial_{32} \end{bmatrix}$$

$$= \begin{bmatrix} v_1 \times m(1-m) & v_1 \times m(1-m) x_{i1} & v_1 \times m(1-m) \times x_{i2} \\ v_2 \times n(1-n) & v_2 \times n(1-n) x_{i1} & v_2 \times n(1-n) x_{i2} \\ v_3 \times o(1-o) & v_3 \times o(1-o) x_{i1} & v_3 \times o(1-o) x_{i2} \end{bmatrix}$$

$$\frac{\partial \hat{y}}{\partial w} = \begin{bmatrix} 0.00499 & 0.00499 & 0.00998 \\ 0.00748 & 0.00748 & 0.0149 \\ 0.00998 & 0.00998 & 0.0199 \end{bmatrix}$$

$$\Delta v = \frac{\partial K}{\partial v} = \frac{\partial K}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial v}, \quad \Delta v = \begin{bmatrix} -15.888 & -8.262 & -8.262 & -8.262 \end{bmatrix}$$

$$\Delta w = \frac{\partial l}{\partial w} = \frac{\partial l}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w}$$

$$\Delta w = \begin{bmatrix} -0.0792 & -0.0792 & -0.1585 \\ -0.1188 & -0.1187 & -0.2375 \\ -0.1585 & -0.1585 & -0.3168 \end{bmatrix} \quad (I)$$

$x_2 = x_{i2} =$ Second training example : $\boxed{((1,3,1)\ 11)}$

$\hat{y}_2 = 0.567$

$\cancel{\partial x}\ \frac{dL}{\partial \hat{y}} = 2(\hat{y}-y) = 2\ (0.0567 - 11) = 2\ (-10.94)$

$$\boxed{\frac{\partial L}{\partial \hat{y}} = -21.88}$$

$$\frac{\partial \hat{y}}{\partial v} = \begin{bmatrix} \frac{\partial \hat{y}}{\partial v_0} & \frac{\partial \hat{y}}{\partial v_1} & \frac{\partial \hat{y}}{\partial v_2} & \frac{\partial \hat{y}}{\partial v_3} \end{bmatrix} = \begin{bmatrix} 1 & 0.52 & 0.52 & 0.587 \end{bmatrix}$$

$$\frac{\partial \hat{y}}{\partial w} = \begin{bmatrix} 0.0049 & 0.00492 & 0.0149 \\ 0.00748 & 0.0748 & 0.2246 \\ 0.00998 & 0.00998 & 0.0299 \end{bmatrix}$$

$$\Delta v = \frac{\partial L}{\partial v} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial v} \Rightarrow \Delta v = -21.88 \times \begin{bmatrix} 1 & 0.52 & 0.52 & 0.52 \end{bmatrix}$$

$$\Delta v = \begin{bmatrix} 21.88 & -11.380 & -11.380 & -11.315 \end{bmatrix}$$

$$\Delta w = \frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w} = -21.88 \begin{bmatrix} 0.0049 & 0.00492 & 0.0149 \\ 0.00748 & 0.00748 & 0.02246 \\ 0.00998 & 0.00998 & 0.0299 \end{bmatrix}$$

$$\Delta w = \begin{bmatrix} -0.1072 & -0.10767 & -0.3261 \\ -0.1637 & -0.1637 & -0.04915 \\ -0.21842 & -0.21842 & -0.65429 \end{bmatrix} \quad \text{— (II)}$$

* for $x_3 : x_{i3}$: Third training example : $\boxed{(2,2),\ 10.}$

$$\Delta \hat{y}_2 = 0.0567$$

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y) = 2(0.0567 - 10) = -19.886.$$

$$\frac{\partial \hat{y}}{\partial v} = \begin{bmatrix} \frac{\partial \hat{y}}{\partial v_0} & \frac{\partial \hat{y}}{\partial v_1} & \frac{\partial \hat{y}}{\partial v_2} & \frac{\partial \hat{y}}{\partial v_3} \end{bmatrix}$$

$$\frac{\partial \hat{y}}{\partial v} = \begin{bmatrix} 1 & 0.52 & 0.52 & 0.51 \end{bmatrix}$$

$$\frac{\partial \hat{y}}{\partial w} = \begin{bmatrix} v_1 (1-m) m & v_1 m (1-m) 2 & v_1 m (1-m) 2 \\ v_2 (n)(1-n) & v_2 (n)(1-n) 2 & v_2 (n)(1-n) 2 \\ v_3 (o)(1-o) & v_3 (o)(1-o) 2 & v_3 (o)(1-o) 2 \end{bmatrix}$$

$$\frac{\partial \hat{y}}{\partial w} = \begin{bmatrix} 0.00499 & 0.00998 & 0.00998 \\ 0.00748 & 0.0149 & 0.0149 \\ 0.00998 & 0.01996 & 0.01996 \end{bmatrix}$$

$$\Delta v = \frac{\partial L}{\partial v} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial v} = -19.886 \times \begin{bmatrix} 1 & 0.52 & 0.52 & 0.51 \end{bmatrix}$$

$$\Delta v = \begin{bmatrix} -19.886 & -10.340 & -10.340 & -10.281 \end{bmatrix}$$

$$\Delta w = \frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w} = -19.886 \begin{bmatrix} 0.00499 & 0.00998 & 0.00998 \\ 0.00748 & 0.0149 & 0.0149 \\ 0.00998 & 0.01996 & 0.01996 \end{bmatrix}$$

Answer

$$\boxed{\Delta w = \begin{bmatrix} -0.0992 & -0.1984 & -0.1984 \\ -0.1487 & -0.2963 & -0.2963 \\ -0.1984 & -0.3969 & -0.3969 \end{bmatrix}}$$

(III)

Scanned by CamScanner

d) Formula given in the class

updated $v_j \leftarrow v_j - n \sum_{i=1}^{m} (\hat{y}_{(j)}^{(i)} - y_{(j)}^{(i)}) z^{(i)}$

$$\Delta v_j = \sum_{i=1}^{m} (\hat{y}_j^{(i)} - y_{(j)}^{(i)}) z^{(i)}$$

$$z^{(i)} = \text{sigmoid}(w_j^T x).$$

updated $w_j \leftarrow w_j - n \sum_{i=1}^{m} \sum_{e=1}^{K} (\hat{y}_e^{(i)} - y_e^{(i)}) v_j z_j^{(i)} (1 - z_j^{(i)}) x^{(i)}$

$$\Delta w_j = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) v_j \cdot z_j^{(i)} (1 - z_j^{(i)}) x^{(i)}$$

$\cancel{z}$ $z^{(i)}$ is vector which contains sigmoid applied values.

$z = [m \ n \ o]$.

for $x = [1, 2]$, $y = [8]$

$$\Delta v_j = (\hat{y}_j^{(i)} - y_j^{(i)}) [1 \ m \ n \ o]$$

$$\Delta v_j = -15.886 [1 \ 0.52 \ 0.52 \ 0.517]$$

$$\Delta v_j = [-15.886 \ -8.261 \ -8.261 \ -8.214]$$

$$\Delta w_1 = (-15.886)(0.02)(0.52)(1 - 0.52)[1 \ 1 \ 2]$$

$$\Delta w_1 = [-0.0793 \ -0.07930 \ -0.1586].$$

$$\Delta w_2 = (-15.886)(0.03)(0.52)(1 - 0.52)[1 \ 1 \ 2]$$

Scanned by CamScanner

$= [-0.1189 \quad -0.1189 \quad -0.2379]$.

$\Delta w_3 = (-15.886)(0.04)\cancel{0.0}(0.517)(1-0.517)[1\ 1\ 2]$
$= [-0.1585 \quad -0.1585 \quad -0.3168]$

$$\Delta w = \begin{bmatrix} -0.793 & -0.793 & -0.1586 \\ -0.1189 & -0.1189 & -0.2379 \\ -0.1585 & -0.1585 & -0.3168 \end{bmatrix} \quad -(IV)$$

$\Delta v = [-15.886 \quad -8.261 \quad -8.261 \quad -8.215]$

As you can see (IV) results matches with ~~ex~~ result (I).

★ for $x_{i2} = [1, 3]$

$\Delta v = -21.88 [1 \quad 0.52 \quad 0.52 \quad 0.517]$
$\Delta v = [-21.88 \quad -11.399 \quad -11.393 \quad -11.311]$

$\Delta w = \Delta w_1 = -21.88 (0.02)(0.52)(1-0.52)[1\ 1\ 3]$
$\Delta w_1 = [-0.1092 \quad -0.1092 \quad -0.3276]$,

$\Delta w_2 = -21.88 (0.03)(0.52)(1-0.52) [1\ 1\ 3]$
$= [-0.1638 \quad -0.1638 \quad -0.4915]$

$\Delta w_3 = -21.88 (0.04)(0.517)(1-0.517)[1\ 1\ 3]$
$\Delta w_3 = [\cancel{-21.}\ -0.2185 \quad -0.2185 \quad -0.6556]$.

$$\Delta w = \begin{bmatrix} -0.1092 & -0.1092 & -0.3276 \\ -0.1638 & -0.1638 & -0.4915 \\ -0.2185 & -0.2185 & -0.6556 \end{bmatrix} \quad -(V).$$

~~Δw~~ $\Delta v = [-21.88 \quad -11.399 \quad -11.399 \quad -11.311]$

Results (V) matches with (II) results

for $x_{i3} = (2,2), 10$

$$\Delta V = -19.886 \; [1 \;\; 0.52 \;\; 0.52 \;\; 0.519]$$

$$\Delta V = [-19.88 \;\; -10.340 \;\; -10.340 \;\; -10.281]$$

$$\Delta W = \Delta W_1 = -19.886 \; (0.02)(0.52)(1-0.52) \; [1 \;\; 2 \;\; 2]$$

$$= [-0.09927 \;\; -0.19853 \;\; -0.19854]$$

$$\Delta W_2 = -19.886 \; (0.03)(0.52)(1-0.52) \; [1 \;\; 2 \;\; 2]$$

$$= [-0.1483 \;\; -0.2978 \;\; -0.2978]$$

$$\Delta W_3 = -19.886 \; (0.04)(0.52)(1-0.52) \; [1 \;\; 2 \;\; 2]$$

$$= [-0.1985 \;\; -0.3970 \;\; -0.3970]$$

$$\Delta W = \begin{bmatrix} -0.09927 & -0.19854 & -0.19855 \\ -0.01469 & -0.2978 & -0.2978 \\ -0.1985 & -0.3970 & -0.3970 \end{bmatrix}$$

$$\Delta V = [-19.886 \;\; -10.340 \;\; -10.340 \;\; -10.281] \; \text{— (VI)}$$

Results of (VI) matches with result of (III).

Q.2

$f(x) = (2x+3y)^2$

a) $\nabla f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$

$\frac{\partial f}{\partial x} = 2(2x+3y) \times 2 = 4(2x+3y)$

$\frac{\partial f}{\partial y} = 2(2x+3y) \times 3 = 6(2x+3y)$.

$\nabla f(x,y) = \begin{bmatrix} 4(2x+3y) \\ 6(2x+3y) \end{bmatrix}$

b) Jacobian Matrix

$F(x,y) = \begin{bmatrix} x^2+2y \\ 3x+4y^2 \end{bmatrix}$

Compute Jacobian Matrix $DF(1,2)$.

$DF(x,y) = \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix} = \begin{bmatrix} x^2+2y \\ 3x+4y^2 \end{bmatrix}$

$= \begin{bmatrix} \partial f_1/\partial x & \partial f_1/\partial y \\ \partial f_2/\partial x & \partial f_2/\partial y \end{bmatrix}$

$= \begin{bmatrix} 2x & 2 \\ 3 & 8y \end{bmatrix}$

$DF(1,2) = \begin{bmatrix} 2 & 2 \\ 3 & 16 \end{bmatrix}$

$16_2 = 125^3$ $4 \times 2x \; x^2$.

$2x \times$

$$G(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}, \quad F(x,y) = \begin{bmatrix} x^2 + 2y \\ 3x + 4y^2 \end{bmatrix}.$$

$$F \circ G(x) = \begin{bmatrix} x^2 + 2x^2 \\ 3x + 4x^4 \end{bmatrix}$$

☆ $\boxed{\text{using chain rule :}}$ $D(F \circ G)(x) = DF(g(x))\, DG(x)$

$$D(F \cdot G)(x) = \begin{bmatrix} 2x + 4x \\ 3 + 16x^3 \end{bmatrix} = \begin{bmatrix} 6x \\ 3 + 16x^3 \end{bmatrix}$$

$$D(F \circ G)(2) = \begin{bmatrix} 6 \times 2 \\ 3 + 16(2)^3 \end{bmatrix} = \begin{bmatrix} 12 \\ 3 + 128 \end{bmatrix} = \begin{bmatrix} 12 \\ 131 \end{bmatrix}.$$

☆ ~~Without chain rule.~~ $\boxed{\text{Using chain rule.}}$

$$F(x,y) = \begin{bmatrix} x^2 + 2y \\ 3x + 4y^2 \end{bmatrix}$$

$$DF'(x,y) = \begin{bmatrix} 2x & 2 \\ 3 & 8y \end{bmatrix}.$$

$$DF(G(x)) = \begin{bmatrix} 2x & 2 \\ 3 & 8x^2 \end{bmatrix}$$

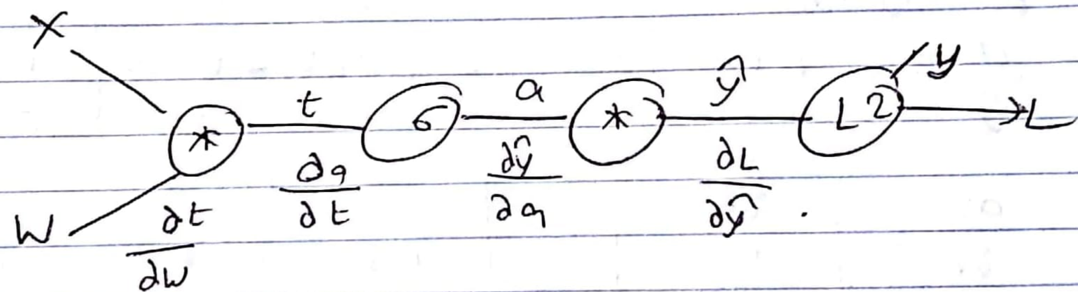
$$DG(x) = \begin{bmatrix} 1 \\ 2x \end{bmatrix}$$

$$DF(G(x)) \cdot G(x) = \begin{bmatrix} 2x & 2 \\ 3 & 8x^2 \end{bmatrix} \begin{bmatrix} 1 \\ 2x \end{bmatrix}$$

$$= \begin{bmatrix} 2x + 4x \\ 3 + 16x^3 \end{bmatrix}$$

$$DF(G \circ G)(2) = \begin{bmatrix} 2(2) + 4(2) \\ 3 + 16(2)^3 \end{bmatrix} = \begin{bmatrix} 12 \\ 131 \end{bmatrix}$$

d)



$x^T = [1 \quad 1 \quad 2]$ → Input vector.

$$W = \begin{bmatrix} w_1^T \\ w_2^T \\ w_3^T \end{bmatrix} = \begin{bmatrix} 0.01 & 0.02 & 0.03 \\ 0.03 & 0.01 & 0.02 \\ 0.02 & 0.03 & 0.01 \end{bmatrix}$$ } weight matrix.

$$v = [0.01 \quad 0.02 \quad 0.03 \quad 0.04]$$

Forward propagation

$$a = \text{Sigmoid}(t) = \text{Sigmoid}(wx).$$

$$wx = \begin{bmatrix} 0.01 & 0.02 & 0.03 \\ 0.03 & 0.01 & 0.02 \\ 0.02 & 0.03 & 0.01 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.09 \\ 0.08 \\ 0.07 \end{bmatrix}$$

$$a = \text{sigmoid}(wx) = \begin{bmatrix} 0.52 \\ 0.519 \\ 0.517 \end{bmatrix}$$

$$\hat{y} = va$$

$$= \begin{bmatrix} 0.01 & 0.02 & 0.03 & 0.04 \end{bmatrix} \begin{bmatrix} 1 \\ 0.52 \\ 0.519 \\ 0.517 \end{bmatrix}$$

$$\hat{y} = 0.056$$

☆ Backward propagation:

$$L = (\hat{y} - y)^2$$

$$\frac{dL}{d\hat{y}} = 2(\hat{y} - y) = 2(0.056 - 8) = \boxed{-15.886}$$

$$\frac{\partial \hat{y}}{\partial v} = \begin{bmatrix} \frac{\partial \hat{y}}{\partial v_0} \\ \frac{\partial \hat{y}}{\partial v_1} \\ \frac{\partial \hat{y}}{\partial v_2} \\ \frac{\partial \hat{y}}{\partial v_3} \end{bmatrix} = \begin{bmatrix} 1 \\ 0.52 \\ 0.519 \\ 0.517 \end{bmatrix}$$

$$\frac{d\hat{y}}{\partial a} = \begin{bmatrix} \partial \hat{y}/\partial a_1 \\ \partial \hat{y}/\partial a_2 \\ \partial \hat{y}/\partial a_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0.02 \\ 0.03 \\ 0.04 \end{bmatrix}$$

$$\frac{\partial a}{\partial w} = \begin{bmatrix} \partial a_1/\partial w_1 \\ \partial a_2/\partial w_2 \\ \partial a_3/\partial w_3 \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial w_1} \text{sigmoid}(w_1 x) \\ \frac{\partial}{\partial w_2} \text{sigmoid}(w_2 x) \\ \frac{\partial}{\partial w_3} \text{sigmoid}(w_3 x) \end{bmatrix}$$

$$\frac{\partial a}{\partial W} = \begin{bmatrix} a_1(1-a_1) \\ a_2(1-a_2) \\ a_3(1-a_3) \end{bmatrix} \begin{bmatrix} x_1^T \\ x_1^T \\ x_1^T \end{bmatrix}$$

$$= \begin{bmatrix} 0.52(1-0.52) \\ 0.519(1-0.519) \\ 0.517(1-0.517) \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.249 & 0.249 & 0.499 \\ 0.249 & 0.249 & 0.499 \\ 0.249 & 0.249 & 0.499 \end{bmatrix}$$

$$\frac{\partial L}{\partial v} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial v} = \begin{bmatrix} -15.888 \\ -8.260 \\ -8.244 \\ -8.213 \end{bmatrix}$$

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a} \cdot \frac{\partial a}{\partial W}$$

$$= -15.888 \cdot \begin{bmatrix} 0.02 \\ 0.03 \\ 0.04 \end{bmatrix} \cdot \begin{bmatrix} 0.249 & \cancel{0.0}\,0.249 & 0.4999 \\ 0.249 & 0.249 & 0.499 \\ 0.249 & 0.249 & 0.499 \end{bmatrix}$$

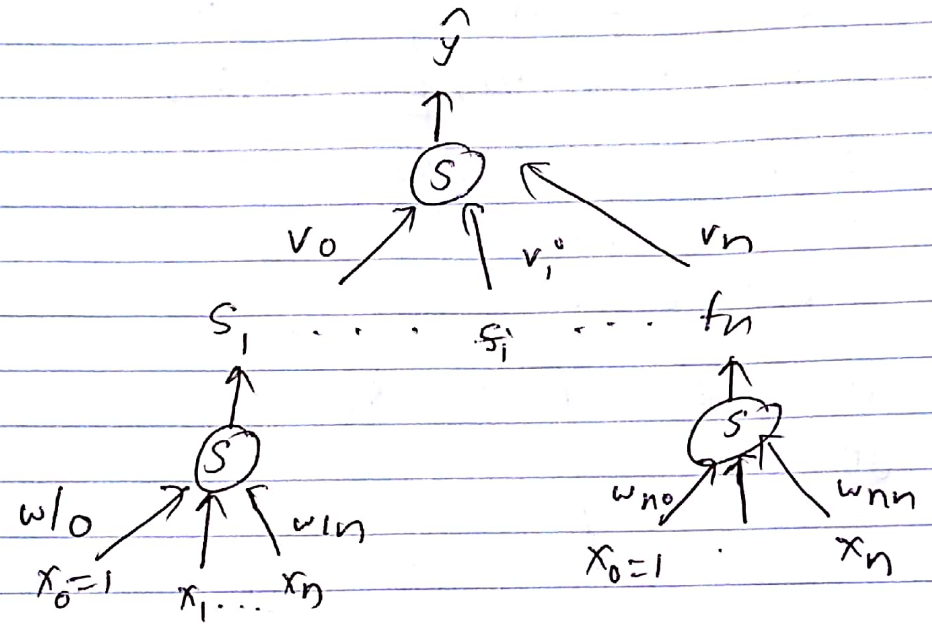
$$\frac{\partial L}{\partial W} = \begin{bmatrix} -0.0792 & -0.0792 & -0.1584 \\ -0.1187 & -0.8187 & -0.2375 \\ -0.1584 & -0.1584 & -0.3168 \end{bmatrix}$$

Scanned by CamScanner

Q.3

Training set is as follows :-

$$\{x^{(i)}, y^{(i)}\}_{i=1}^{m}, \quad x^{(i)} \in R^m, \quad y^{(i)} \in (0,1).$$



Loss function given = $E = \frac{1}{2}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \rightarrow$ L2 Loss function

★ output layer parameter updation :-

$$\frac{\partial E}{\partial v} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial v}$$

$$\frac{\partial E}{\partial v} = \underbrace{\frac{1}{2} \times 2 \sum_{i=1}^{m} (\hat{y}_i - y_i)}_{\partial E/\partial \hat{y}.} \underbrace{f^{(i)}}_{\partial \hat{y}/\partial v.}$$

$$\frac{\partial E}{\partial v} = \sum_{i=1}^{m} (\hat{y}_j - y_i) \cdot (\hat{y}^{(i)})(1 - \hat{y}^{(i)}) z^{(i)}.$$

$$v \leftarrow v - n \sum_{i=1}^{m} (\hat{y_i} - y_i)\, \hat{y}^{(i)} (1 - \hat{y}^{(i)})\, z^{(i)}.$$

☆ Hidden Layer Parameter Update :-

$$\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_i} \quad \frac{\partial z_j}{\partial w_j}$$

$$\frac{\partial E}{\partial w_j} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})\, v_j\, \underbrace{f_j^{(i)} (1 - f_j^{(i)})\, x^{(i)}}_{\text{Sigmoid activation derivation.}}$$

$$w_j \leftarrow w_j - n \sum_{i=1}^{n} (\hat{y}^{(i)} - y^{(i)})\, v_j \cdot z_j^{(i)} (1 - f_j^{(i)})\, x^{(i)}.$$