Sparse categorical Cross entropy

In sparse categorical output column is ~~used~~ converted → Integer encoding.

All is same like ~~spar~~ categorical Cross entropy

when Y=1 predication → [0.1 0.4 0.5] $\hat{y}_{=1}$ ← 1/2/3

$$L = -Y_1 \log(\hat{Y}_1) = -1 \log(0.1)$$

fast then categorical cross

when Y=2

$$L = -Y_2 (\log \hat{Y}_2) = -1 \log(0.4)$$

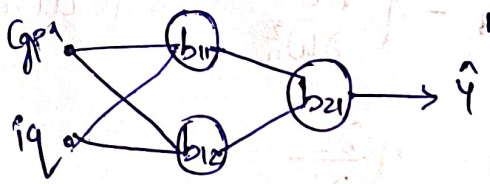
Used in large categorical.

→ Backpropogation :-

It is a algo use to train neural network.

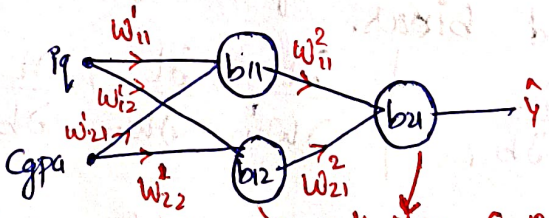
training

cgpa	iq	lpa
8	80	8
7	70	7



weights → bias → correct value

Back prop



[W']

Step 0: Initilize W, b

Step 1:- W→1, b→0

You select a point (row)

Step 2) Predict (lpa) → forward prop [Dot product] (galti ho rhi hai)

W→1, b→0. let → 8lpa.

Step 3) Choose a loss function..

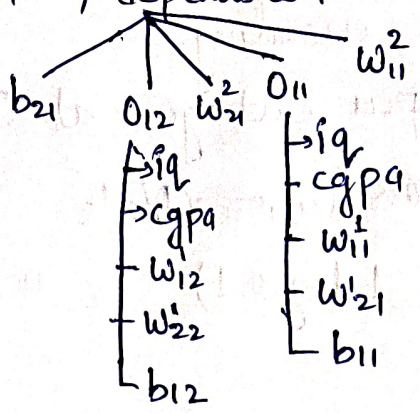
(Here MSE) $L = (3-8)^2 = 225$ error.

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

To reduce error you change $\hat{Y}$.

$$\hat{y} = O_{21} = W_{11}^2 O_{11} + W_{21}^2 O_{12} + b_{21}$$

↳ $O_{21} = \hat{y}$ depends on



loss ko reduce karne ke liya picha ja kr value change karna parega. This is called backpropgration

Step 4) Weights & bias → update using Grandient Descent

$$W_{new} = W_{old} - \eta \frac{\partial L}{\partial W_{old}}$$

$$b_{new} = b_{old} - \eta \frac{\partial L}{\partial b_{old}}$$

like

$$W_{11}^2{}_{new} = W_{11}^2{}_{old} - \eta \frac{\partial L}{\partial W_{11}^2}$$ → derivative of loss wrt W_{11}^2

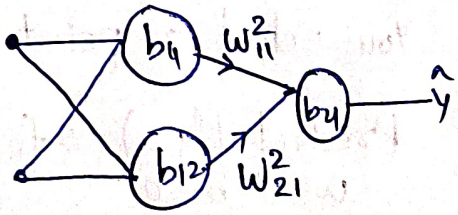
$$b_{21\,new} = b_{21\,old} - \eta \frac{\partial L}{\partial b_{21}}$$

calculating derivative of loss fun^n for each of their weights and biases.

$$\frac{\partial L}{\partial W_{11}^2}, \frac{\partial L}{\partial W_{21}^2}, \frac{\partial L}{\partial b_{21}} \;\Big|\; \frac{\partial L}{\partial W'_{11}}, \frac{\partial L}{\partial W'_{12}}, \frac{\partial L}{\partial b_{11}} \;\Big|\; \frac{\partial L}{\partial W'_{12}}, \frac{\partial L}{\partial W'_{2}}, \frac{\partial L}{\partial b_{12}}$$

derivative means jab hum weights or baises mai change kr raha hai tho loss mai kitna change aya gai.

$$\frac{\partial L}{\partial W_{11}^2} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial W_{11}^2}$$



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

$$\frac{\partial \hat{y}}{\partial W_{11}^2} = \frac{\partial}{\partial W_{11}^2} (O_{11} W_{11}^2 + O_{12} W_{21}^2 + b_{21}) = O_{11}$$

$$\boxed{\frac{\partial L}{\partial W_{11}^2} = -2(y - \hat{y}) O_{11}}$$

$$\frac{\partial L}{\partial w_{21}^2} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_{21}^2}$$

$$\frac{\partial \hat{y}}{\partial w_{21}^2} = \frac{\partial}{\partial w_{21}^2}\left(o_{11} w_{11}^2 + o_{12} w_{21}^2 + b_{21}\right) = o_{12}$$

$$\boxed{\frac{\partial L}{\partial w_{21}^2} = -2(y-\hat{y}) \cdot o_{12}}$$

$$\boxed{\frac{\partial L}{\partial b_{21}} = -2(y-\hat{y})}$$

$$\frac{\partial L}{\partial w_{11}^1} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial o_{11}} \times \frac{\partial o_{11}}{\partial w_{11}^1} \quad \frac{\partial L}{\partial w_{21}^1} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_{11}} \frac{\partial o_{11}}{\partial w_{21}^1} = -2(y-\hat{y}) w_{11}^2 x_{i2}$$

$-2(y-\hat{y}) w_{11}^2 x_{i1}$

$$\frac{\partial L}{\partial b_{11}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial o_{11}} \frac{\partial o_{11}}{\partial b_{11}} = -2(y-\hat{y}) w_{21}^2$$

$$\frac{\partial L}{\partial w_{12}^1} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_{12}} \frac{\partial o_{12}}{\partial w_{12}^1} \quad \frac{\partial L}{\partial w_{22}^1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial o_{12}} \frac{\partial o_{12}}{\partial w_{22}^1} = -2(y-\hat{y}) w_{21}^2 x_{i2}$$

$-2(y-\hat{y}) w_{21}^2 x_{i1}$

$$\frac{\partial L}{\partial b_{12}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_{12}} \frac{\partial o_{12}}{\partial b_{12}} = -2(y-\hat{y}) w_{21}^2$$

$$\frac{\partial \hat{y}}{\partial o_{11}} = \frac{\partial}{\partial o_{11}}\left(w_{11}^2 o_{11} + w_{21}^2 o_{12} + b_{21}\right) = w_{11}^2$$

$$\frac{\partial \hat{y}}{\partial o_{12}} = \frac{\partial}{\partial o_{12}}\left(w_{11}^2 o_{11} + w_{21}^2 o_{12} + b_{21}\right) = w_{21}^2$$

$$\frac{\partial o_{11}}{\partial w_{11}^1} = \frac{\partial}{\partial w_{11}^1}\left(iq\, w_{11}^1 + cgpa\, w_{21}^1 + b_{11}\right) = iq = x_{i1}$$

(value of iq for Student)

$$\frac{\partial o_{11}}{\partial w_{21}^1} = \cdot cgpa = \cdot x_{i2}$$

$$\frac{\partial o_{11}}{\partial b_{11}} = 1$$

$$\frac{\partial O_{12}}{\partial w'_{12}} = \frac{\partial}{\partial w'_{12}}(iq\,w'_{12} + cgpa\,w'_{22} + b_{12}) = iq = x_{i1}$$

$$\frac{\partial O_{12}}{\partial w'_{22}} = x_{i2} \qquad \frac{\partial O_{12}}{\partial b_{12}} = 1$$

Step 5) once again

a) weights/bias → initialize $w = 0.1$, $b = 0$ OR random value

epoch ← for i in range.

1) for i in range(4): ← no. of rows in data

1a) 1 student → forward prop → predict(lpa)

1b) calculate loss (MSE)

1c) Adjust all the weights and bias

$$w_{new} = w_{old} - \eta \frac{\partial L}{\partial w_{old}}$$ } → 9 time → total weight & biases

Do this multiple time for better result/ loss minimize nahi hota tab tak loop chalate rahi.