Question - NN parameters

(a) # of parameters in untied CNN.

Convolution Layer

(i) Total convolutions = 13×13 Params in a kernel $= 4 \times 4$

no bias

Each convolution has different parameters

$\Rightarrow (13 \times 13) \times (4 \times 4)$

$\Rightarrow 169 \times 16 = 2704$

FCN layer

The size of the intermediate image is (13×13)

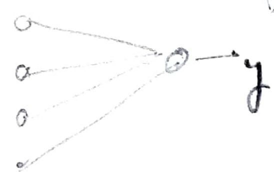
Each convolution will give a value.

The FCN layer will connect each of these nodes with the output node.

$\Rightarrow$ No. of param

$= 13 \times 13$

$= 169$



output of convolutions

Total = 2873

(ii) Tied weights

Total convolutions $= 13 \times 13$ Params in a kernel $= 4 \times 4$

Since the kernel parameters are shared hence the CNN layer only has 16 parameters.

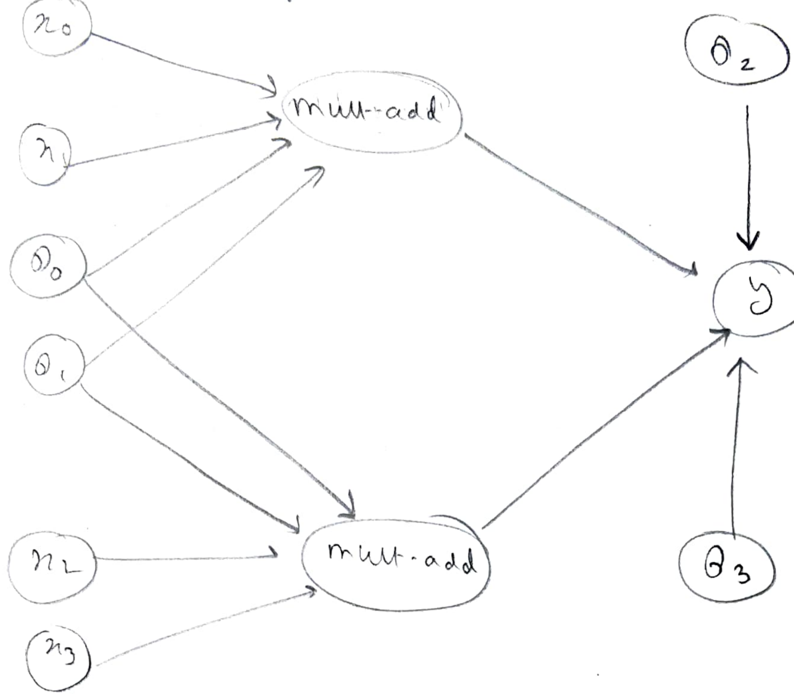
FCN layer has $13 \times 13 = 169$ parameters

Total $= 169 + 16 = 185$ parameters

Computation Graph

(b)

(i)



(ii) The equations are: $\frac{\partial L}{\partial \theta_2}$ $\frac{\partial L}{\partial y}$

$$z_1 = \theta_0 x_0 + \theta_1 x_1 \qquad L = (y-\hat{y})^2$$

$$z_2 = \theta_0 \cdot x_2 + \theta_1 \cdot x_3$$

$$y = \theta_2 \cdot z_1 + \theta_3 \cdot z_2$$

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

For θ_2

$$\frac{\partial L}{\partial \theta_2} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial \theta_2} = 2(y-\hat{y})\left[\frac{\partial(\theta_2 \cdot z_1 + \theta_3 \cdot z_2)}{\partial \theta_2}\right] = \boxed{2(y-\hat{y})z_1}$$

update rule $\boxed{\theta_2^{i+1} \leftarrow \theta_2^{i} - \alpha \times [2(y-\hat{y})z_1]}$

For θ_3

$$\frac{\partial L}{\partial \theta_3} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial \theta_3} = 2(y-\hat{y})\left[\frac{\partial(\theta_2 \cdot z_1 + \theta_3 \cdot z_2)}{\partial \theta_3}\right] = \boxed{2(y-\hat{y})z_2}$$

$$\boxed{\theta_3^{i+1} \leftarrow \theta_3^{i} - \alpha(2)(y-\hat{y})z_2}$$

For θ_0

$$\frac{\partial L}{\partial \theta_0} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial \theta_0} + \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial \theta_0}$$

Note that there are two paths from loss back to the parameter θ_0, as the weights are tied.

$$\frac{\partial L}{\partial \theta_0} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z_1} \cdot \frac{\partial z_1}{\partial \theta_0} + \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z_2} \cdot \frac{\partial z_2}{\partial \theta_0}$$

$$= \frac{\partial L}{\partial y} \times \left[\frac{\partial(\theta_2 \cdot z_1 + \theta_3 \cdot z_2)}{\partial z_1} \cdot \frac{\partial(\theta_0 x_0 + \theta_1 x_1)}{\partial \theta_0}\right] + \frac{\partial L}{\partial y} \times \left[\frac{\partial(\theta_2 \cdot z_1 + \theta_3 \cdot z_2)}{\partial z_2} \times \frac{\partial(\theta_0 x_2 + \theta_1 x_3)}{\partial \theta_0}\right]$$

$$= 2(y-\hat{y})\left[\theta_2 \cdot x_0 + \theta_3 \cdot x_2\right]$$

For θ_1

Note that there are two paths from Loss to θ_1 as the weights are tied.

$$\frac{\partial L}{\partial \theta_1} = \frac{\partial L}{\partial y}\times\frac{\partial y}{\partial z_1}\times\frac{\partial z_1}{\partial\theta_1} + \frac{\partial L}{\partial y}\times\frac{\partial y}{\partial z_2}\times\frac{\partial z_2}{\partial\theta_1}$$

$$= \frac{\partial L}{\partial y}\times\frac{\partial[\theta_2 . z_1+\theta_3 . z_2]}{\partial z_1}\times\frac{\partial(\theta_0 x_0+\theta_1 x_1)}{\partial\theta_1} +$$

$$\frac{\partial L}{\partial y}\times\frac{\partial[\theta_2 . z_1+\theta_3 . z_2]}{\partial z_2}\times\frac{\partial[\theta_0 x_2+\theta_1 . x_3]}{\partial\theta_1}$$

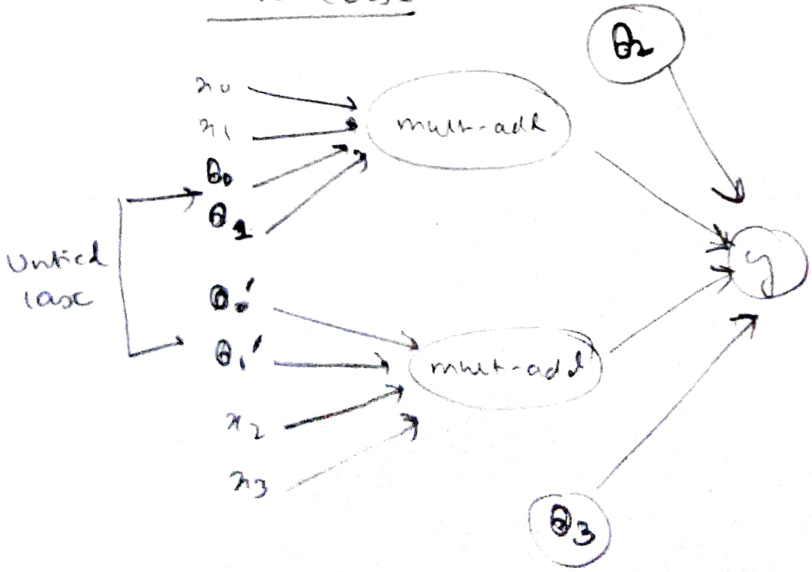
$$= \frac{\partial L}{\partial y}\left[\theta_2 \cdot x_1 + \theta_3 \cdot x_3\right] = 2(y-\hat{y})(x_1 . \theta_2 + x_3\theta_3)$$

$$\theta_1^{i+1} \leftarrow \theta^i - \alpha\left[2(y-\hat{y})(x_1 . \theta_2 + x_3\theta_3)\right]$$

(iii) Back propagating the gradient through any edge in the computation graph requires application of a single chain rule step.

No. of edges will be the same for both cases, hence same amt. of compn happens.

untied case



still twelve edges.