$y' = 32.783 + 0.2001 * x$

$R^2 = 1 - \frac{SSE}{SST}$

$x \rightarrow$ weight
$y \rightarrow$ height.

① $y = 32.783 + 0.2001 \times 140$

$= 60.797$

$y' = 60$

$(y' - y)^2 = 0.635209$

② $y = 32.783 + 0.2001 \times 155$

$= 63.7985$

$y' = 62$

mean $= 68.14$

weight	Height	$(y-y')^2$	SSR	Predicted value
140	60	0.635	53.919	60.797
155	62	3.235		63.7985
159	67	5.765		64.5989
179	70	0.0408		68.6009
192	71	0.041		71.2022
200	72	0.645		72.803
212	75	0.42		75.2092
		SSE = 12.69		

$$SSR = \sum (y' - \bar{y})^2$$

predicted value

mean of height.

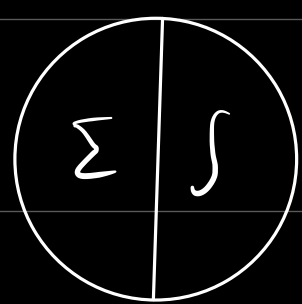
$$\bar{y} = 68.1428$$

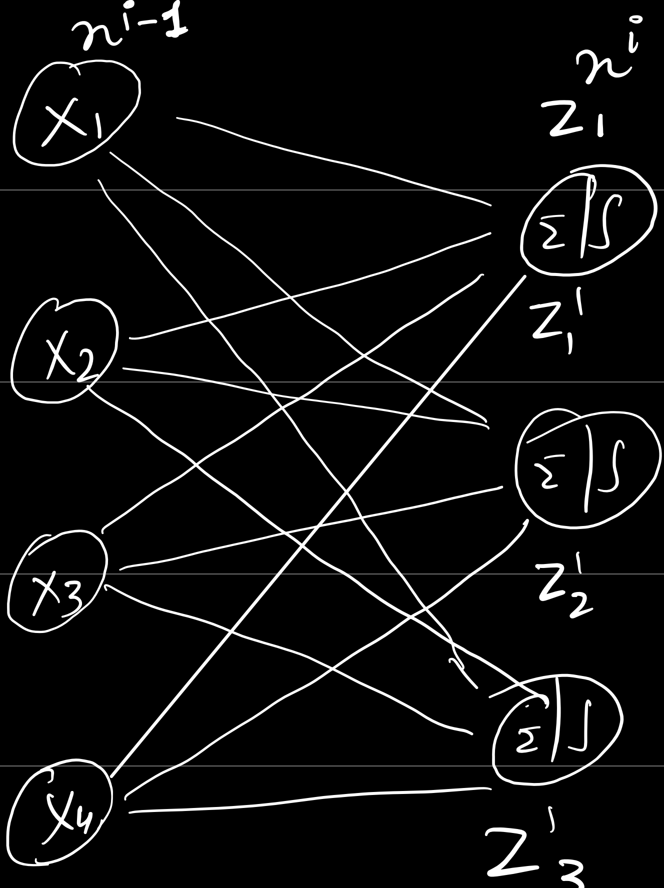
$$r^2 = \frac{SSR}{SST}$$

$$SST = SSR + SSE$$

Each neuron is doing

1. Aggregation
2. Applying some non-linear function to it. [Activation]





$$Z_2' = w_{21} X_1 + w_{22} X_2 + w_{23} X_3 + w_{24} X_4 + b$$

$$Z^1 = \begin{bmatrix} z_1^1 \\ z_2^1 \\ z_3^1 \end{bmatrix}$$

Total no. of weights req. $= n^{i-1} \times n^i$

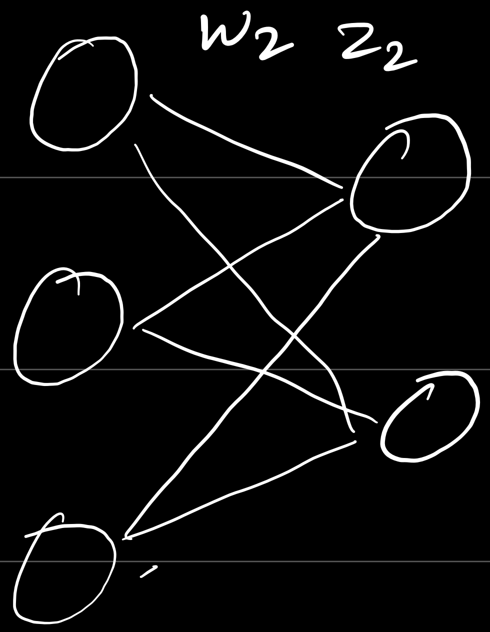
→ Making the weight matrix either $n^{i-1} \times n^i$ or $n^i \times n^{i-1}$

→ no. of bias $= n^i$

$$b = \begin{bmatrix} b_1^1 \\ b_2^1 \\ b_3^1 \end{bmatrix}$$

→ Binary class ⇒ 1 neuron

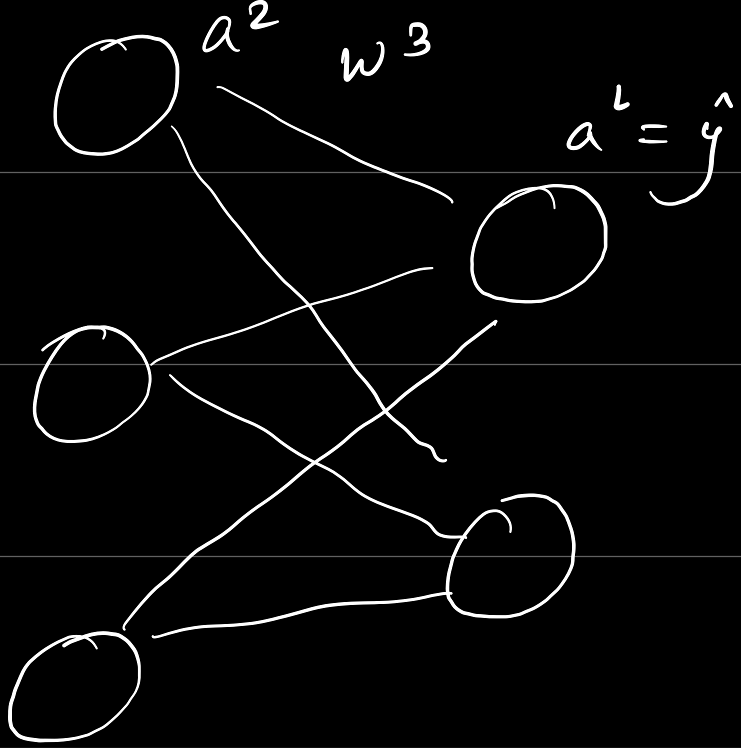
Multi class → k neurons for k class.



$$Z_2 = \begin{bmatrix} z_1^2 \\ z_2^2 \end{bmatrix} = \begin{bmatrix} w_{11}^2 & w_{12}^2 & w_{13}^2 \\ w_{21}^2 & w_{22}^2 & w_{23}^2 \end{bmatrix} \begin{bmatrix} a_1^1 \\ a_2^1 \\ a_3^1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$[2 \times 3] \; [3 \times]$$

$$= \begin{bmatrix} \sum_{i=1}^{3} w_{1i}^2 a_i^1 + b_1 \\ \sum_{i=1}^{3} w_{2i}^2 a_i^1 + b_2 \end{bmatrix}$$

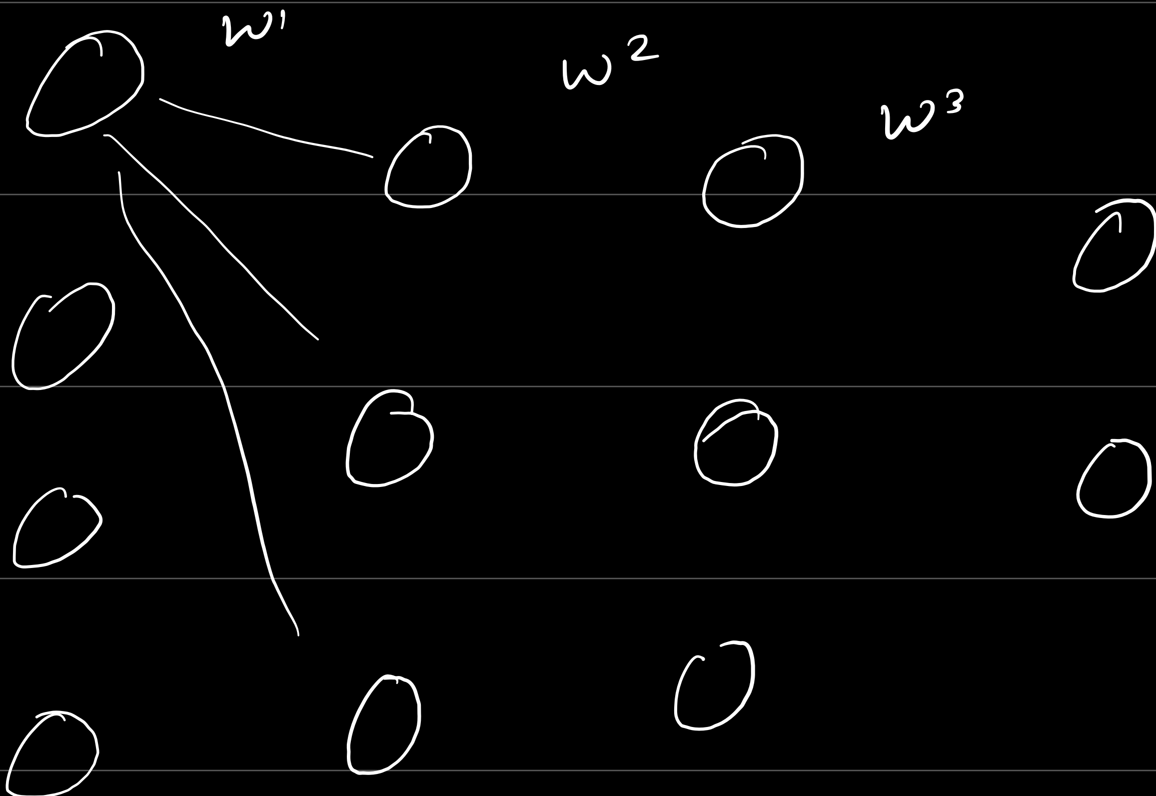


$$\hat{Z} = \begin{bmatrix} \hat{z}_1 \\ \hat{z}_2 \end{bmatrix} = \overset{w_3}{\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix}} \begin{bmatrix} a_1^2 \\ a_2^2 \\ a_3^2 \end{bmatrix} + \overset{b_3}{\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}$$

$$= \begin{bmatrix} \sum_{i=1}^{3} w_{1i} a_i^2 + b_1 \\ \sum_{j=1}^{3} w_{2i} a_i^2 + b_2 \end{bmatrix}$$

$$\begin{bmatrix} a_1^1 \\ a_2^1 \\ a_3^1 \end{bmatrix} = \begin{bmatrix} g(z_1^1) \\ g(z_2^1) \\ g(z_3^1) \end{bmatrix} \overset{\text{If sigmoid}}{=} \begin{bmatrix} \sigma(z_1^1) \\ \sigma(z_2^1) \\ \sigma(z_3^1) \end{bmatrix}$$

Q How many parameters to Learn



$w^1 = 12$

$b^1 = 1$

$w^2 = 9$

$b^2 = 1$

$w^L = 6$

$b^L = 1$

$$L = \frac{1}{2}(\hat{y} - y)^2$$

$$\hat{y} = \sigma(z)$$

$$z = wx + b$$

$$\sigma(wx+b)$$

$=$

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

$$(\hat{y} - y)\,\sigma(z)(1 - \sigma(z))\,x$$

$$(\hat{y} - y)(\hat{y})(1 - \hat{y})\,x.$$

$$\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z} \quad \frac{\partial z}{\partial w}$$

$=$