C2W3-multiclass classification

Softmax regression:-

C = # classes

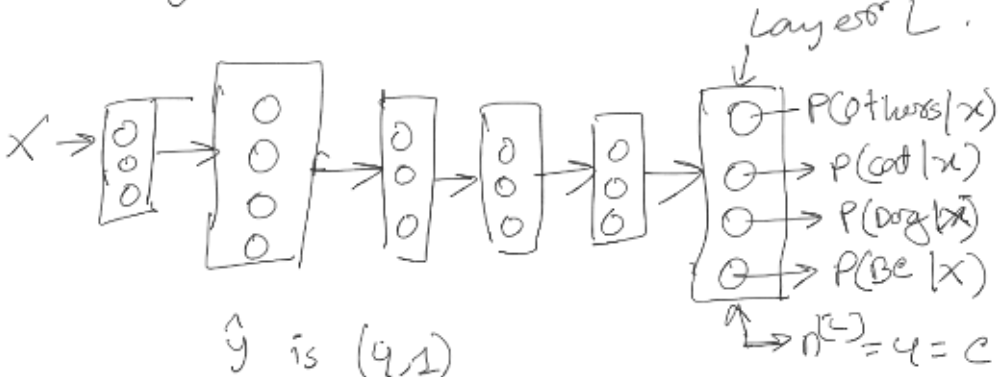
Let we have a model that recognizes 4 classes

0 → others.

1 → cat

2 → Dog.

3 → baby chicken



$\hat{y}$ is (4,1)

$\sum \hat{y}_i = 1.$

$z^{[L]} = w^{[L]} \cdot a^{[L-1]} + b^{[L]}$

Activation function

$t = e^{z^{[L]}}$

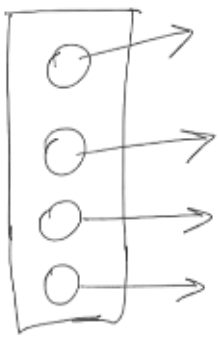
$a^{[L]} = \frac{e^{z^{[L]}}}{\sum_{j=1}^{4} t_i}$, $a_i^{[L]} = \frac{t_i}{\sum_{j=1}^{4} t_i}$

Example:-

$z^{[L]} = \begin{bmatrix} 5 \\ 2 \\ -1 \\ 3 \end{bmatrix}$, $t = \begin{bmatrix} e^5 \\ e^2 \\ e^{-1} \\ e^3 \end{bmatrix} = \begin{bmatrix} 148.4 \\ 7.4 \\ 0.4 \\ 20.1 \end{bmatrix}$ $\sum_{j=1}^{4} t_j = 176.3$

$a^{[L]} = \frac{t}{176.3}$

e^5



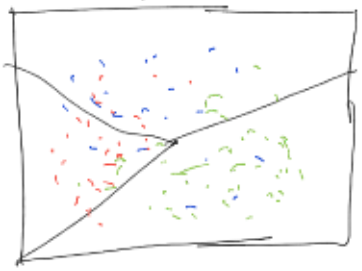
$\frac{e^5}{176.3} = 0.842 \rightarrow 84\%$ chance that this class 0.

$\frac{e^2}{176.3} = 0.042 \rightarrow 4\%$ " " " " 1

$\frac{e^{-1}}{176.3} = 0.002 \rightarrow 0.2\%$ " " " " 2

$\frac{e^3}{176.3} = 0.114 \rightarrow 11\%$ " " " " 3

Decision boundary of any two classes are linear if there is no hidden layers.



Softmax generalize the logistic regression in C classes instead of two classes.

if $C = 2$ & we apply softmax it becomes logistic regression

Loss function:-

$y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ ← cat $\quad \hat{y} = \begin{bmatrix} 0.3 \\ 0.2 \\ 0.1 \\ 0.4 \end{bmatrix}$ ← 20% chance it's a cat. not good.

$$\underbrace{\mathcal{L}(\hat{y}, y)}_{\text{small}} = -\sum_{j=1}^{4} y_j \log \hat{y}_j \quad \Big| \; y_0 = y_2 = y_3 = 0$$

$$= -y_2 \log \hat{y}_2 = \underbrace{-\log \hat{y}_2}_{\text{we need to make this small so that } \hat{y}_2 \text{ has to be as big as possible}}$$

$$\text{Cost, } J = \frac{1}{m} \sum \mathcal{L}(\hat{y}^{(i)}, y^{(i)})$$

$$Y = [y^{(1)} \; y^{(2)} \; y^{(3)} \ldots y^{(m)}] \;, \; \hat{y} = [\hat{y}^{(1)}, \hat{y}^{(2)}, \hat{y}^{(3)} \ldots \hat{y}^{(m)}]$$

$$= \begin{bmatrix} 0 & 0 & 1 & \\ 1 & 0 & 0 & \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & \end{bmatrix}$$
$(4, m)$

$$= \begin{bmatrix} 0.3 \\ 0.3 \\ 0.1 \\ 0.3 \end{bmatrix} \cdots$$
$(4, m)$

Gradient descent with softmax:-

Backprop:-

$$dz^{[L]} = \hat{y} - y$$
$(4,1)$

$$\frac{\partial J}{\partial z^{[L]}}$$