deeplearning.ai

Multi-class classification

Trying a softmax classifier

Understanding softmax

$C = 4$

$$z^{[L]} = \begin{bmatrix} 5 \\ 2 \\ -1 \\ 3 \end{bmatrix}, \quad temp = \begin{bmatrix} e^5 \\ e^2 \\ e^{-1} \\ e^3 \end{bmatrix}$$

$g^{[L]} = \text{softmax}$

$sum = e^5 + e^2 + e^{-1} + e^3$

$$\Rightarrow g^{[L]}(z^{[L]}) = \begin{bmatrix} e^5/sum \\ e^2/sum \\ e^{-1}/sum \\ e^3/sum \end{bmatrix} = \begin{bmatrix} .842 \\ .042 \\ .002 \\ .114 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Name of softmax comes from Hardmax which would have converted Result matrix to $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

- Softmax regression generalizes logistic Regression to "c" classes
- If $c = 2 \Rightarrow$ we have logistic Regression (Binary classification)

Andrew Ng

Loss function

If last layer = softmax, How would you train your NN

$$y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \text{Ground truth} \qquad \hat{y} \text{ or } a^{[L]} = \begin{bmatrix} .3 \\ .2 \\ .1 \\ .4 \end{bmatrix}$$

$$L(\hat{y}, y) = -\sum_{j=1}^{4} y_j \log \hat{y}_j \qquad (\text{Now } y_1 = y_3 = y_4 = 0)$$

$$= -y_2 \cdot \log \hat{y}_2 = -\log \hat{y}_2$$

our Aim is to $\downarrow L(\hat{y}, y) \Rightarrow \downarrow -\log \hat{y}_2 \Rightarrow$ make $\hat{y}_2$ as big as possible

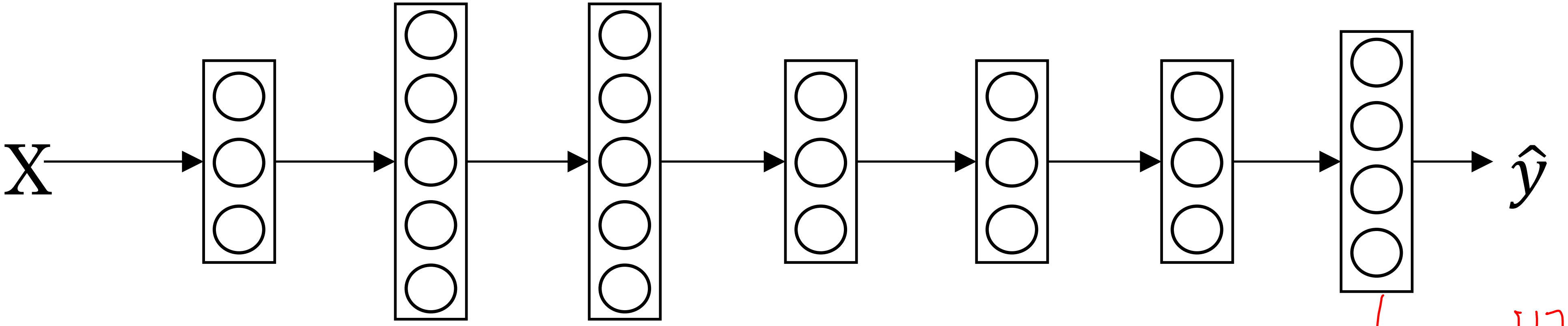
$\Rightarrow$ If $y_2 = 1$, then our prob which was 0.2 corresponding to $\hat{y}_2$ implies we are 20% certain the Image is a cat, which is bad, would have been better if we were 90% sure it was a cat $\Rightarrow$ we have a bad model

Btw, this was Loss over 1 example, to find overall model perf

$$J(W^{[1]}, b^{[1]} \ldots W^{[L]}, b^{[L]}) = \frac{1}{M} \sum_{i=1}^{M} L(\underbrace{y^{(i)}}_{(4 \times 1)}, \underbrace{\hat{y}^{(i)}}_{(4 \times 1)})$$

Andrew Ng

Summary of softmax classifier



$z^{[L]} \Rightarrow a^{[L]} = \hat{y}$

(4×1) (4×1)

then find loss $L(\hat{y}, y)$

Backprop $dz^{[L]} = \hat{y} - y$

$\frac{\partial J}{\partial z^{[L]}}$ can be derived given loss $= -\sum_{j=1}^{c} y_j \cdot \log \hat{y}_j$

$c = \#$ classes

From $dz^{[L]}$ we can find $dz^{[L-1]}$, $dz^{[L-2]}$ & $dw^{[L-1]}$ etc.