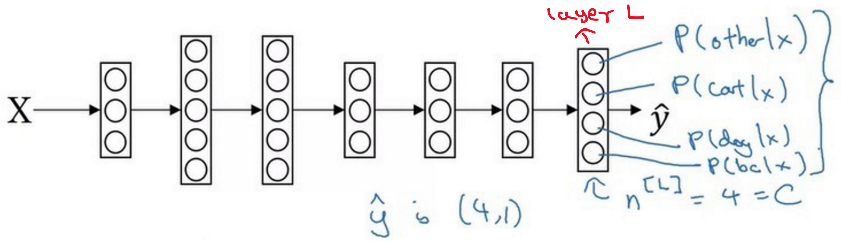
Week 3c - Multi-Class classification

Friday, August 14, 2020 11:30 PM

① Softmax Regression

→ If we want to recognise cats (1), dogs (2) & baby chicks (3) (+ others (0))

C = no. of classes in output = 4 (0...3)



→ Softmax Layer

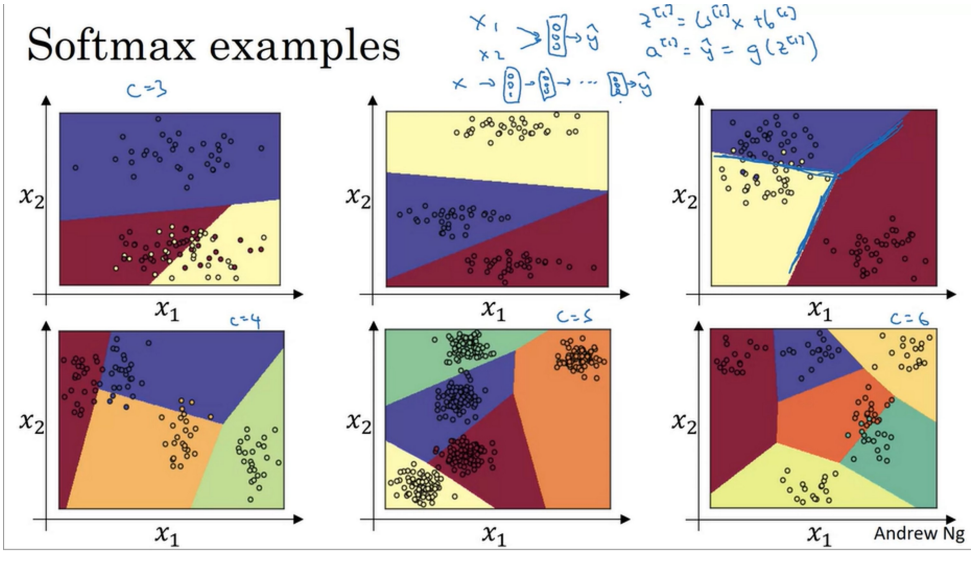
- Softmax function is used as activation function in output layer

$$z^{[L]} = w^{[L]} a^{[L-1]} + b^{[L]} \quad (4,1)$$

Activation function:

Let $t = e^{(z^{[L]})}$

$$a^{[L]} = \frac{e^{z^{[L]}}}{\sum_{i=1}^{4} t_i} \Rightarrow \boxed{a_i^{[L]} = \frac{t_i}{\sum_{i=1}^{4} t_i}}$$



② Training a Softmax Classifier

→ Softmax regression generalizes logistic regression to C classes
(ie) if C=2, softmax reduces to logistic reg.

→ Loss function

$$y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \rightarrow \text{cat} \quad a^{[L]} = \hat{y} = \begin{bmatrix} 0.3 \\ 0.2 \\ 0.1 \\ 0.4 \end{bmatrix} \quad C = 4$$

$$\boxed{\mathcal{L}(\hat{y}, y) = -\sum_{j=1}^{C} y_j \log \hat{y}_j}$$

$$= -y_2 \log \hat{y}_2 \quad \text{(Since other values are 0)}$$

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

∴ To reduce loss, make $\hat{y}_2$ big

→ Cost function

$$J(w^{[L]}, b^{[L]}, \ldots) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(\hat{y}^{(i)}, y^{(i)})$$

→ $Y \rightarrow (C, m)$
$\hat{Y} \rightarrow (L, m)$ [dimensions]

→ Gradient descent with softmax

- Forward prop is finished with
$$z^{[L]}_{(4,1)} \rightarrow a^{[L]} = \hat{y} \rightarrow \mathcal{L}(\hat{y}, y)$$
- Backward prop is initialized with
$$dz^{[L]}_{(C,1)} = \hat{y} - y$$
$$\rightarrow \frac{\partial J}{\partial z^{[L]}}$$