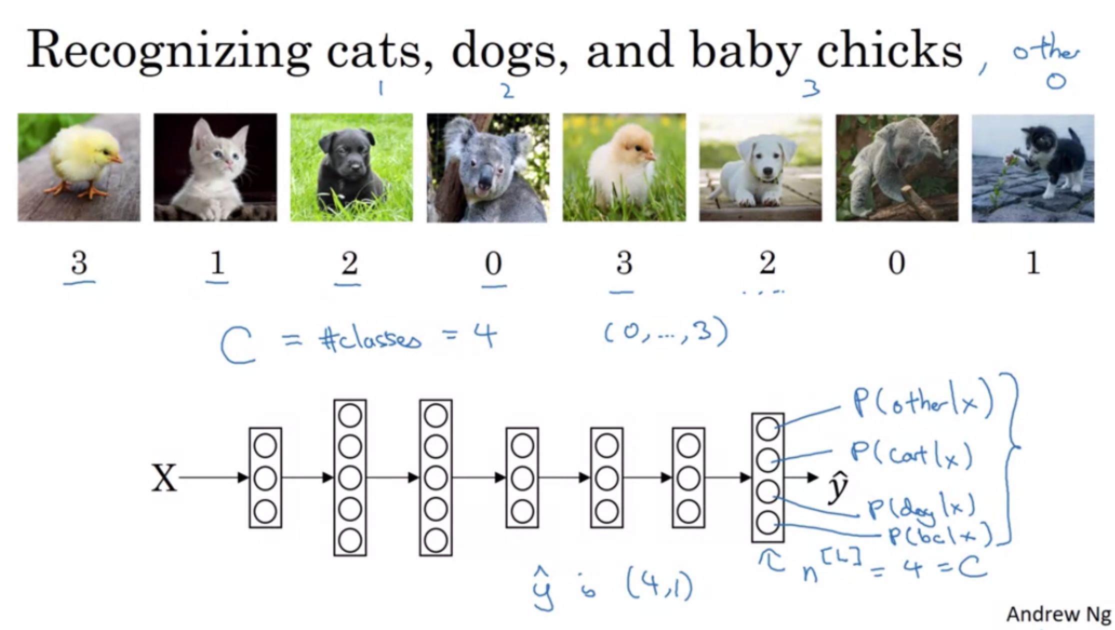
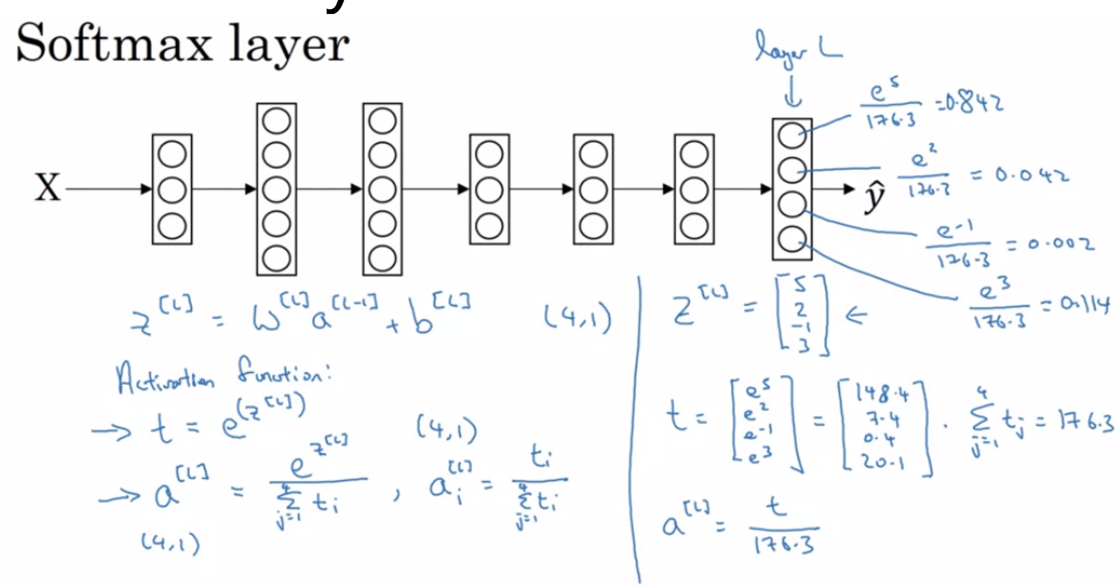
Week 3-3 Softmax Regression (Multi-class Classification)

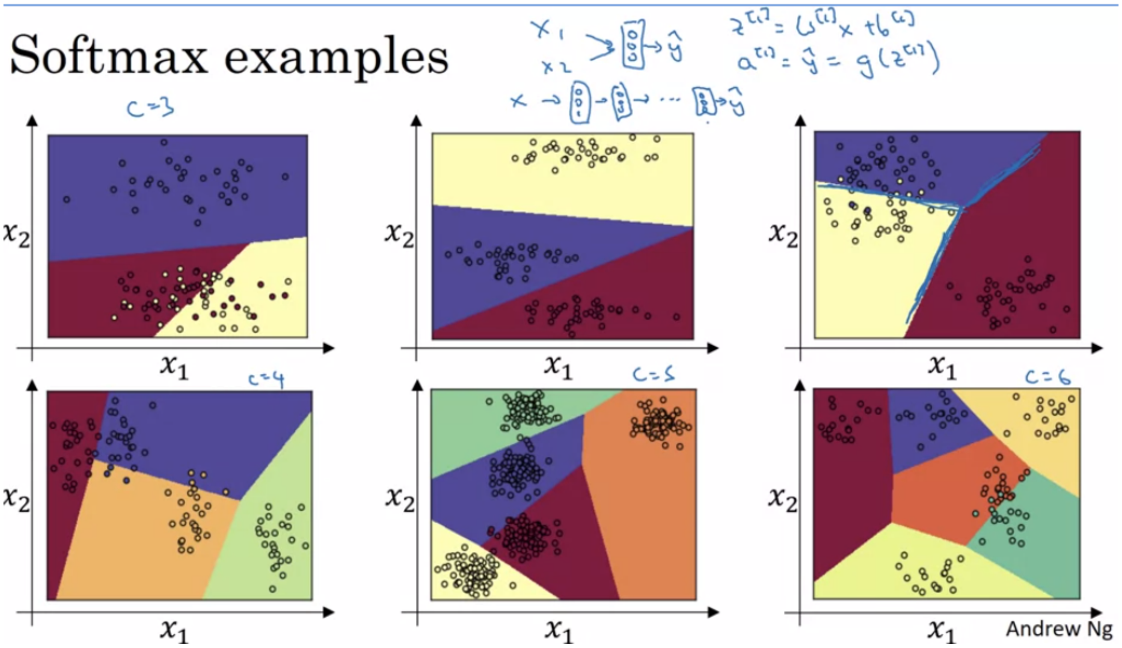
笔记本:	DL 2 - Deep NN Hyperparameter Tuning, Regularization & Optimization
创建时间:	2021/1/9 14:33
更新时间:	2021/1/9 14:43



Softmax layer



So we can maybe we kind of see that this is like a generalization of logistic regression with sort of linear decision boundaries, but with more than two classes



Understanding softmax

(4,1)

$$z^{[L]} = \begin{bmatrix} 5 \\ 2 \\ -1 \\ 3 \end{bmatrix} \qquad t = \begin{bmatrix} e^5 \\ e^2 \\ e^{-1} \\ e^3 \end{bmatrix}$$

C=4 $\qquad g^{[L]}(\cdot)$

"soft max"

$$a^{[L]} = g^{[L]}(z^{[L]}) = \begin{bmatrix} e^5/(e^5+e^2+e^{-1}+e^3) \\ e^2/(e^5+e^2+e^{-1}+e^3) \\ e^{-1}/(e^5+e^2+e^{-1}+e^3) \\ e^3/(e^5+e^2+e^{-1}+e^3) \end{bmatrix} = \begin{bmatrix} 0.842 \\ 0.042 \\ 0.002 \\ 0.114 \end{bmatrix}$$

"hard max"

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Softmax regression generalizes logistic regression to C classes.

If C=2, softmax reduces to logistic regression. $a^{[L]} = \begin{bmatrix} 0.842 \\ \cancel{0.158} \end{bmatrix}$

Loss:

Loss function

$$y^{(i)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad (4,1) \quad \text{— cat}, \quad y_2 = 1, \quad y_1 = y_3 = y_4 = 0$$

$$a^{[L](i)} = \hat{y}^{(i)} = \begin{bmatrix} 0.3 \\ 0.2 \\ 0.1 \\ 0.4 \end{bmatrix} \quad (4,1) \qquad C = 4$$

$$\underbrace{\mathcal{L}(\hat{y}, y)}_{\text{small}} = \underbrace{-\sum_{j=1}^{4} y_j \log \hat{y}_j}$$

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

$$-y_2 \log \hat{y}_2 = -\log \hat{y}_2 \qquad \text{Make } \hat{y}_2 \text{ big.}$$

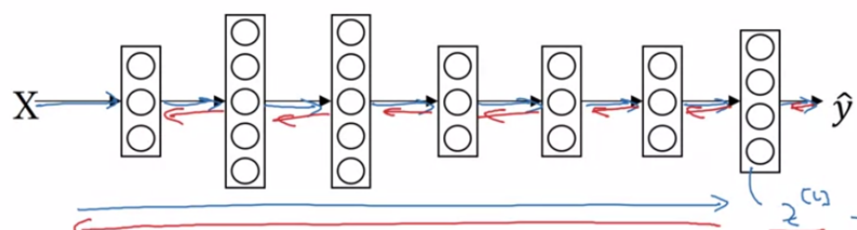
$$Y = [y^{(1)}\ y^{(2)} \dots, y^{(m)}] = \begin{bmatrix} 0 & 0 & 1 & \dots \\ 1 & 0 & 0 & \\ 0 & 1 & 0 & \\ 0 & 0 & 0 & \end{bmatrix} \quad (4,m)$$

$$\hat{Y} = [\hat{y}^{(1)}, \dots, \hat{y}^{(m)}] = \begin{bmatrix} 0.3 & \dots \\ 0.2 & \\ 0.1 & \\ 0.4 & \end{bmatrix} \quad (4,m)$$

Andrew Ng

GD

Gradient descent with softmax



$$z^{[L]} \ (4,1) \longrightarrow a^{[L]} = \hat{y} \to \mathcal{L}(\hat{y}, y)$$

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

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