



Deep Learning

Dr. Mehran Safayani

safayani@iut.ac.ir

safayani.iut.ac.ir



<https://www.aparat.com/mehran.safayani>



https://github.com/safayani/deep_learning_course

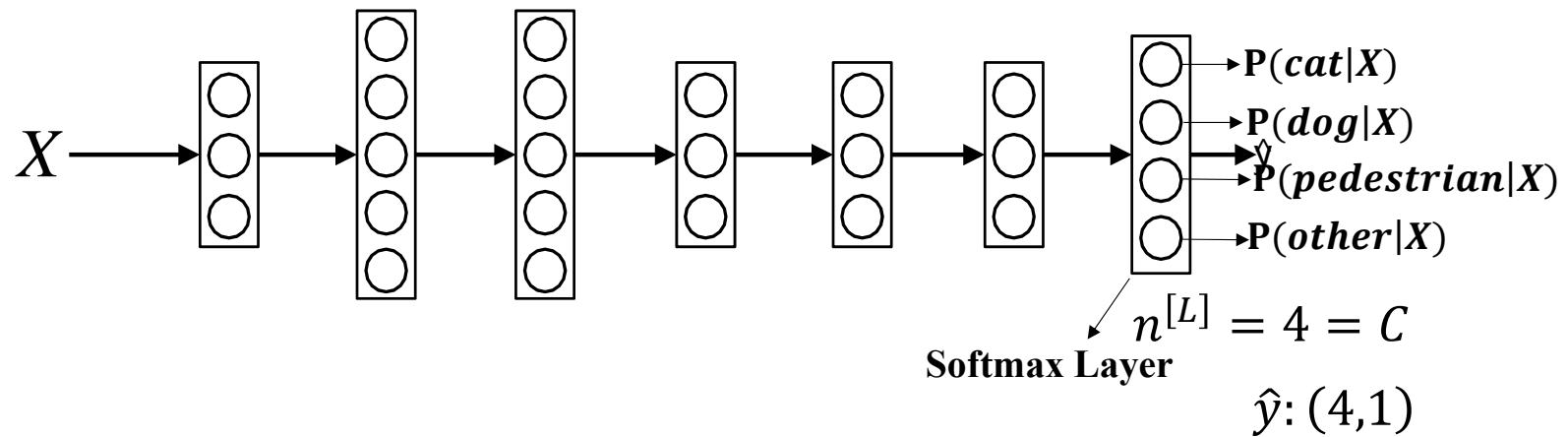


Department of Electrical and computer engineering, Isfahan university of technology, Isfahan, Iran

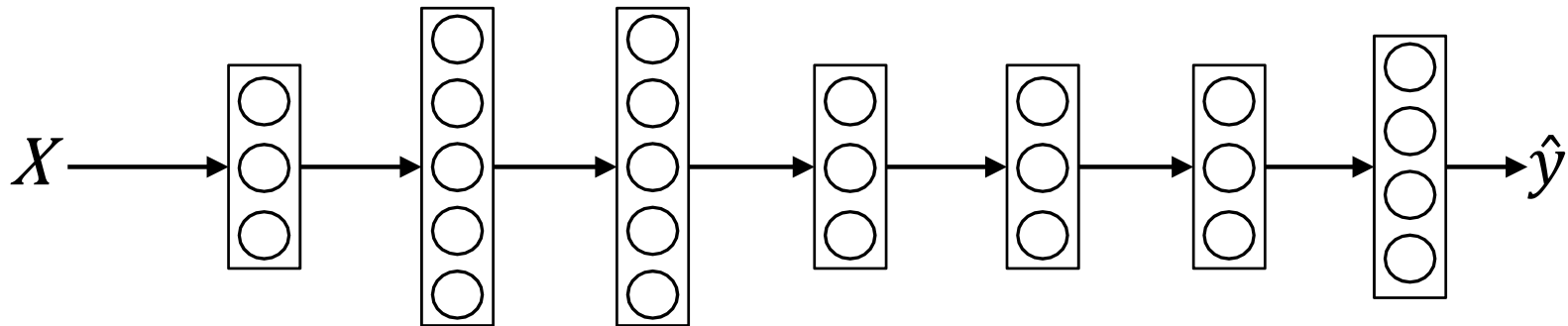
Multi-class classification

Softmax regression

- $C = \text{\#classes} = 4$



Softmax layer



- $z^{[L]} = w^{[L]}a^{[L-1]} + b^{[L]}$
- Softmax Activation function

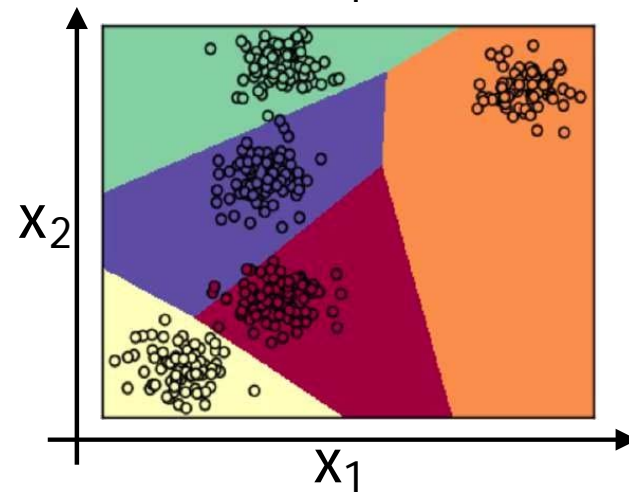
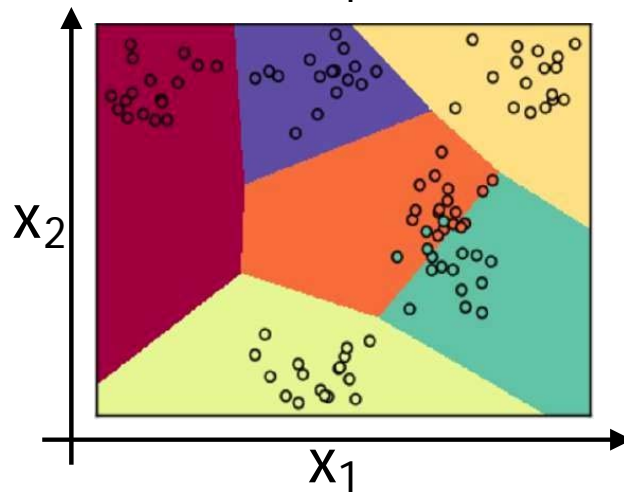
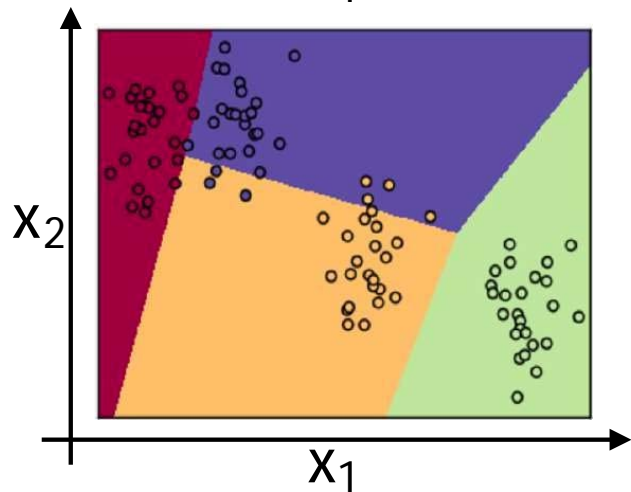
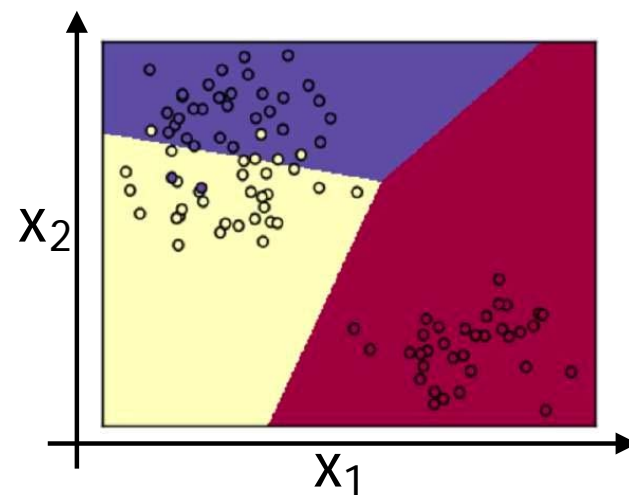
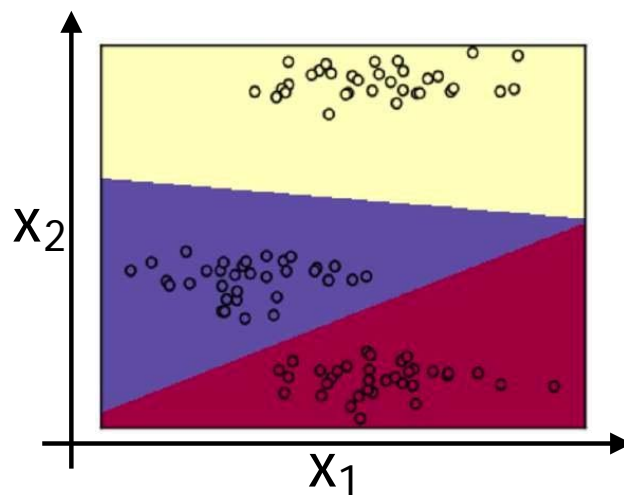
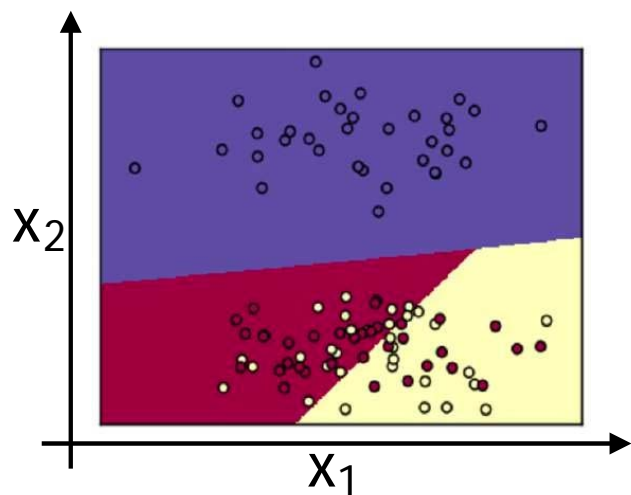
$$t = e^{(z^{[L]})} \quad (4,1)$$

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

$$a^{[L]} = g^{[L]}(z^{[L]})$$

Softmax Activation

Softmax examples



Softmax examples

$$\bullet z^{[L]} = \begin{bmatrix} 5 \\ 2 \\ -1 \\ 3 \end{bmatrix} \quad t = \begin{bmatrix} e^5 \\ e^2 \\ e^{-1} \\ e^3 \end{bmatrix} = \begin{bmatrix} 148.4 \\ 7.4 \\ 0.4 \\ 20.1 \end{bmatrix}$$

$$\sum_{j=1}^4 t_j = 176.3$$

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

$$\frac{1}{176.3} \begin{bmatrix} 148.4 \\ 7.4 \\ 0.4 \\ 20.1 \end{bmatrix} = \begin{bmatrix} 0.842 \\ 0.042 \\ 0.002 \\ 0.114 \end{bmatrix} = \hat{y}^{(i)}$$

$$y_{(4,1)}^{(i)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

one hot vector

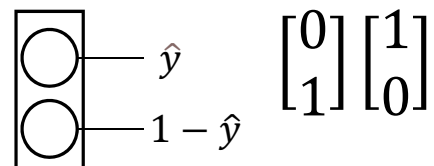
Loss function

- $-[y \log y + (1 - y) \log(1 - y)]$

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

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

$$\bigcirc \quad \hat{y}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

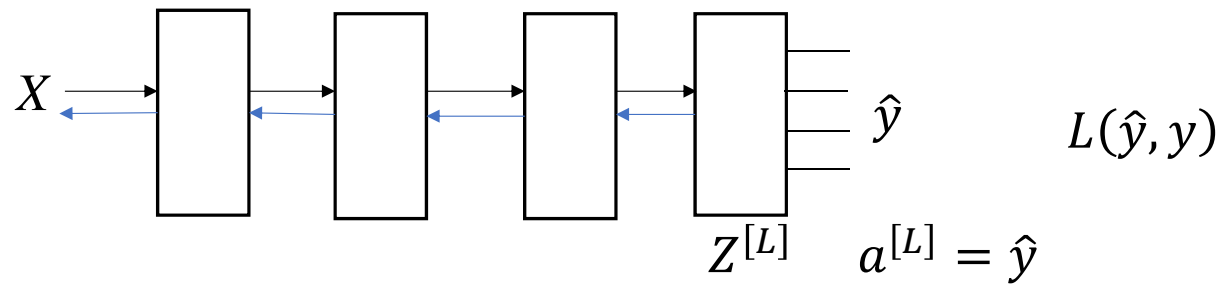


$$Y = [y^{(1)}, y^{(2)}, \dots, y^{(m)}] \quad \hat{Y} = [\hat{y}^{(1)}, \hat{y}^{(2)}, \dots, \hat{y}^{(m)}]$$

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

$$\begin{bmatrix} 0.3 \\ 0.2 \\ 0.3 \\ 0.2 \end{bmatrix} \dots$$

Gradient descent with softmax



• اثبات کنید:

$$dz_{4 \times 1}^{[L]} = \hat{y}_{4 \times 1} - y_{4 \times 1}$$