Machine Learning

Multi-class Classifier

Dr. Mehran Safayani

safayani@iut.ac.ir

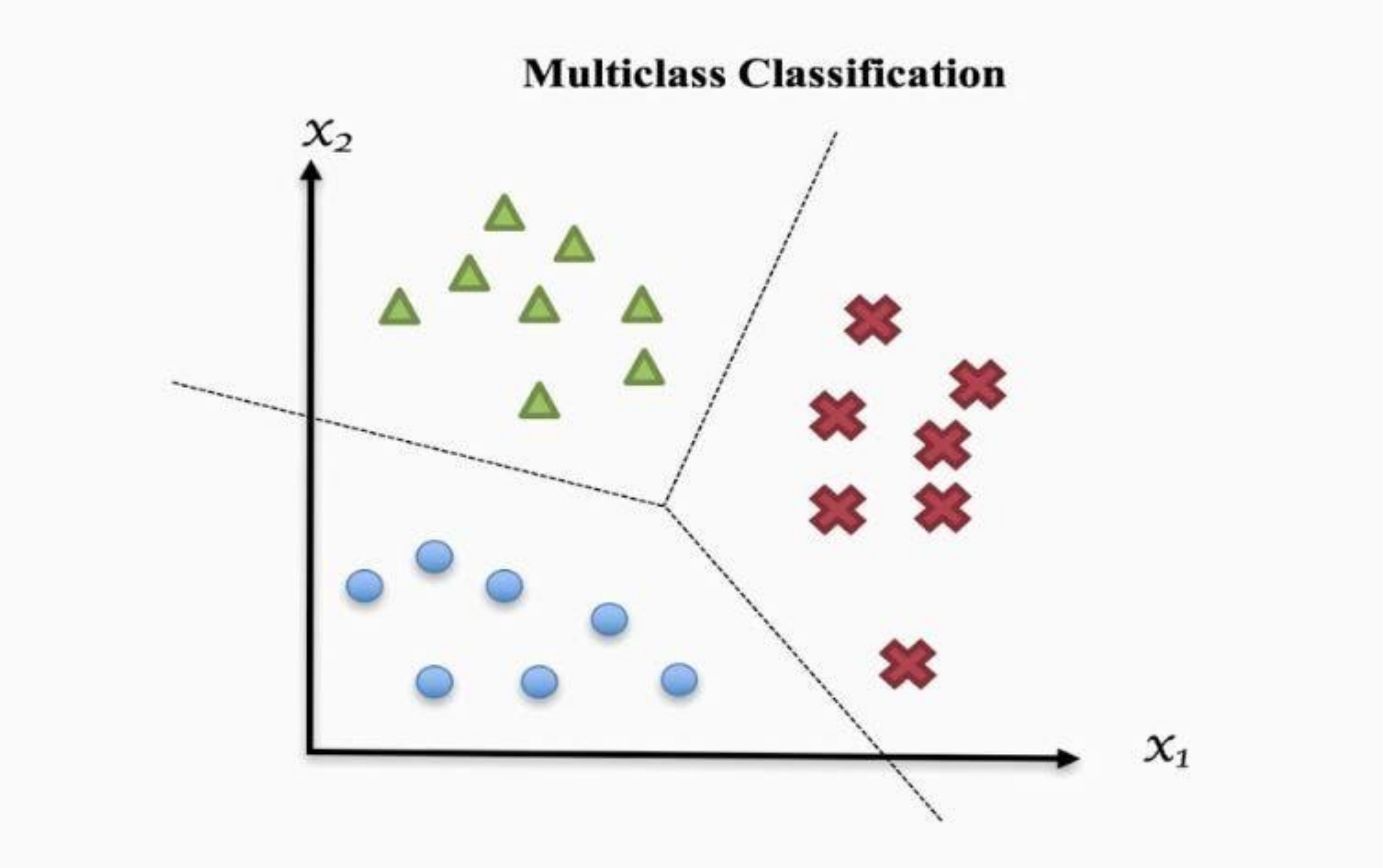
safayani.iut.ac.ir

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

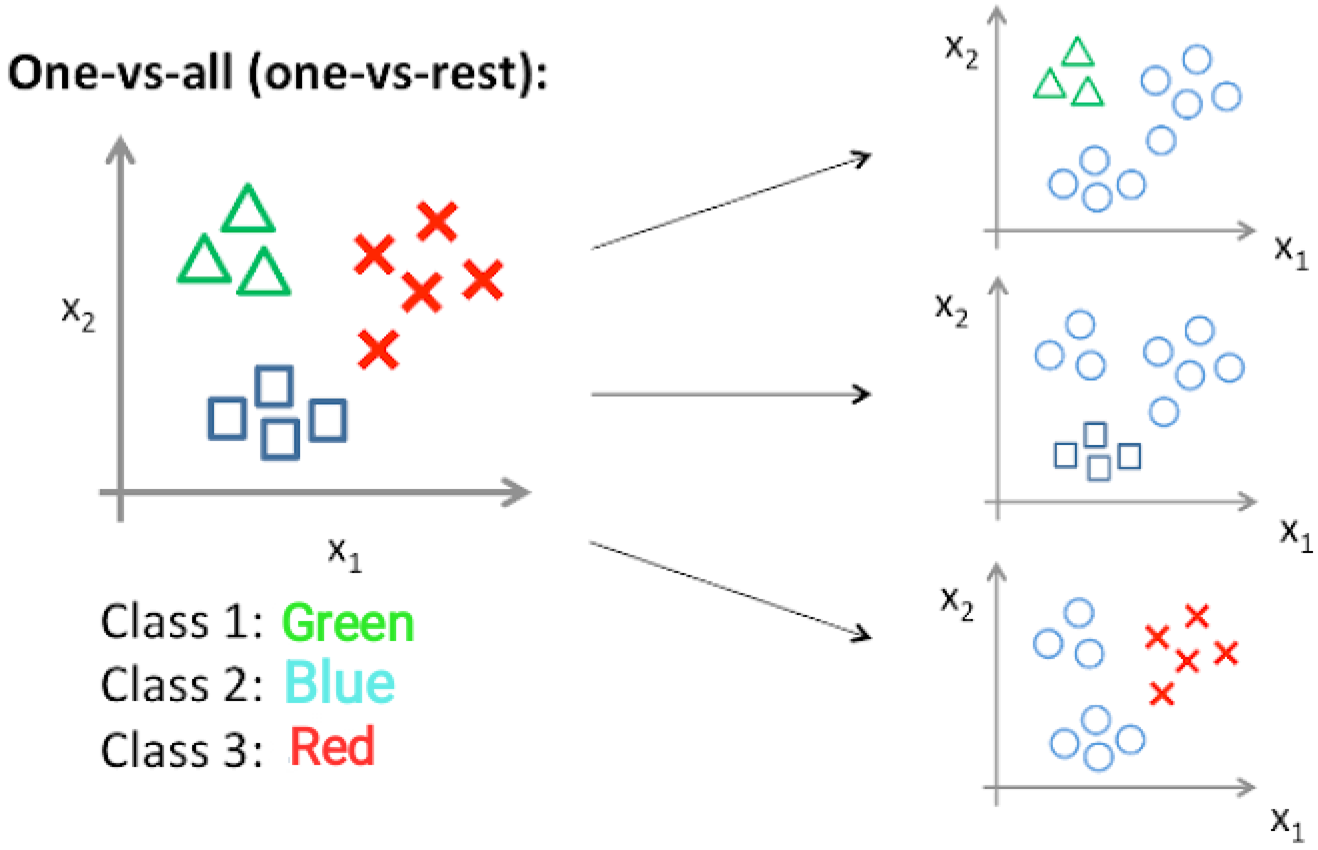
https://github.com/safayani/machine_learning_course

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

Multi-class Classification



Multiclass Classification: One-Vs-Rest(All)



One-hot-vector

- **Classifier 1:- [Green] vs [Red, Blue]**
- **Classifier 2:- [Blue] vs [Green, Red]**
- **Classifier 3:- [Red] vs [Blue, Green]**

Multiclass Classification: One-Vs-Rest(All)

Main Dataset

Features			Classes
x1	x2	x3	G
x4	x5	x6	B
x7	x8	x9	R
x10	x11	x12	G
x13	x14	x15	B
x16	x17	x18	R

Class 1 :- Green Class 2 :- Blue Class 3 :- Red

Training Dataset 1
Class :- Green

Features			Green
x1	x2	x3	**+1**
x4	x5	x6	-1
x7	x8	x9	-1
x10	x11	x12	**+1**
x13	x14	x15	-1
x16	x17	x18	-1

Training Dataset 2
Class :- Blue

Features			Blue
x1	x2	x3	-1
x4	x5	x6	**+1**
x7	x8	x9	-1
x10	x11	x12	-1
x13	x14	x15	**+1**
x16	x17	x18	-1

Training Dataset 3
Class :- Red

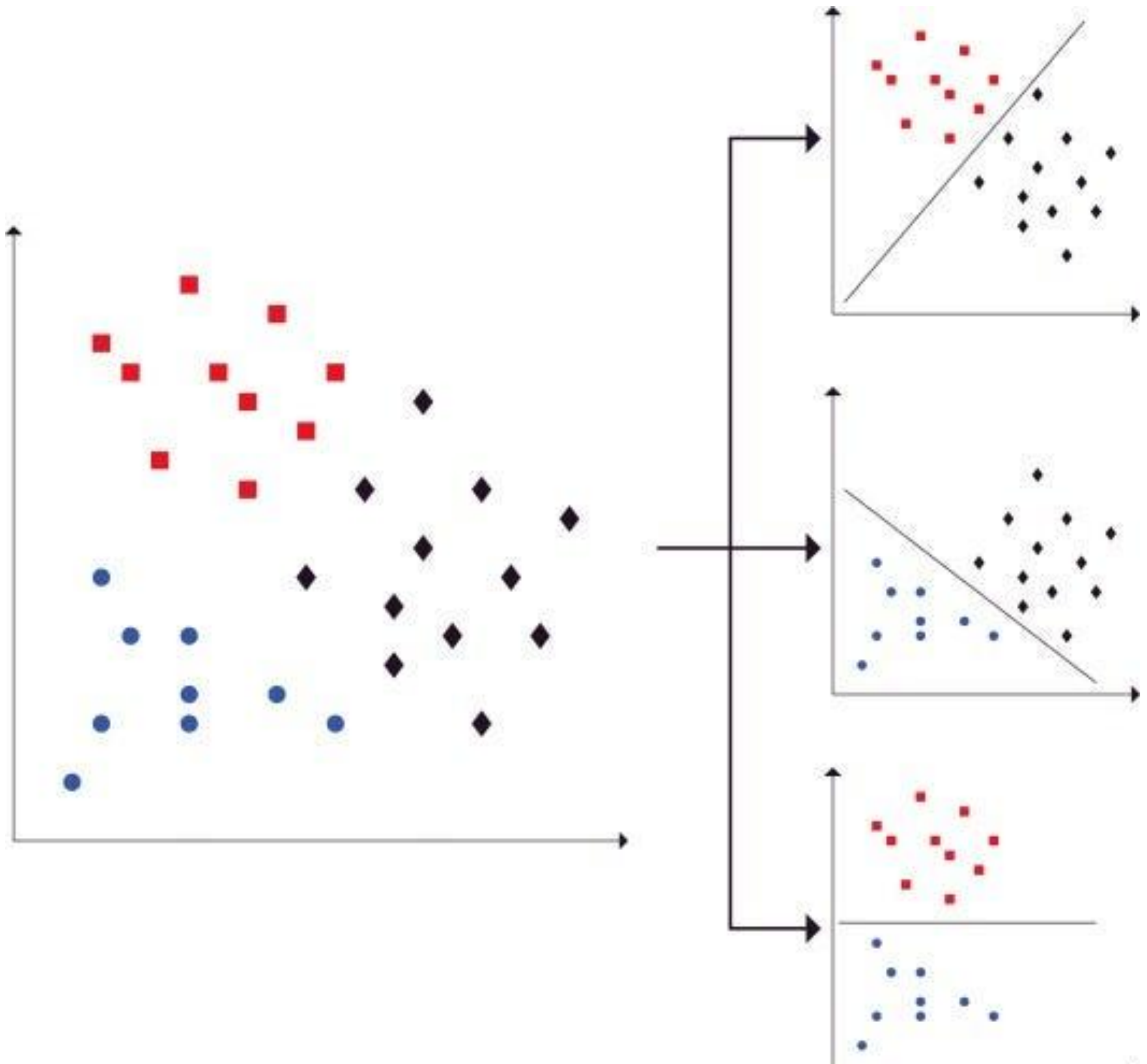
Features			Red
x1	x2	x3	-1
x4	x5	x6	-1
x7	x8	x9	**+1**
x10	x11	x12	-1
x13	x14	x15	-1
x16	x17	x18	**+1**

Test time:

$$h_\theta^i(x) = P(y = i|x)$$

$$\underset{i}{\operatorname{argmax}} h_\theta^i(x)$$

Multiclass Classification: One-Vs-One



N* (N-1)/2 binary classifier models

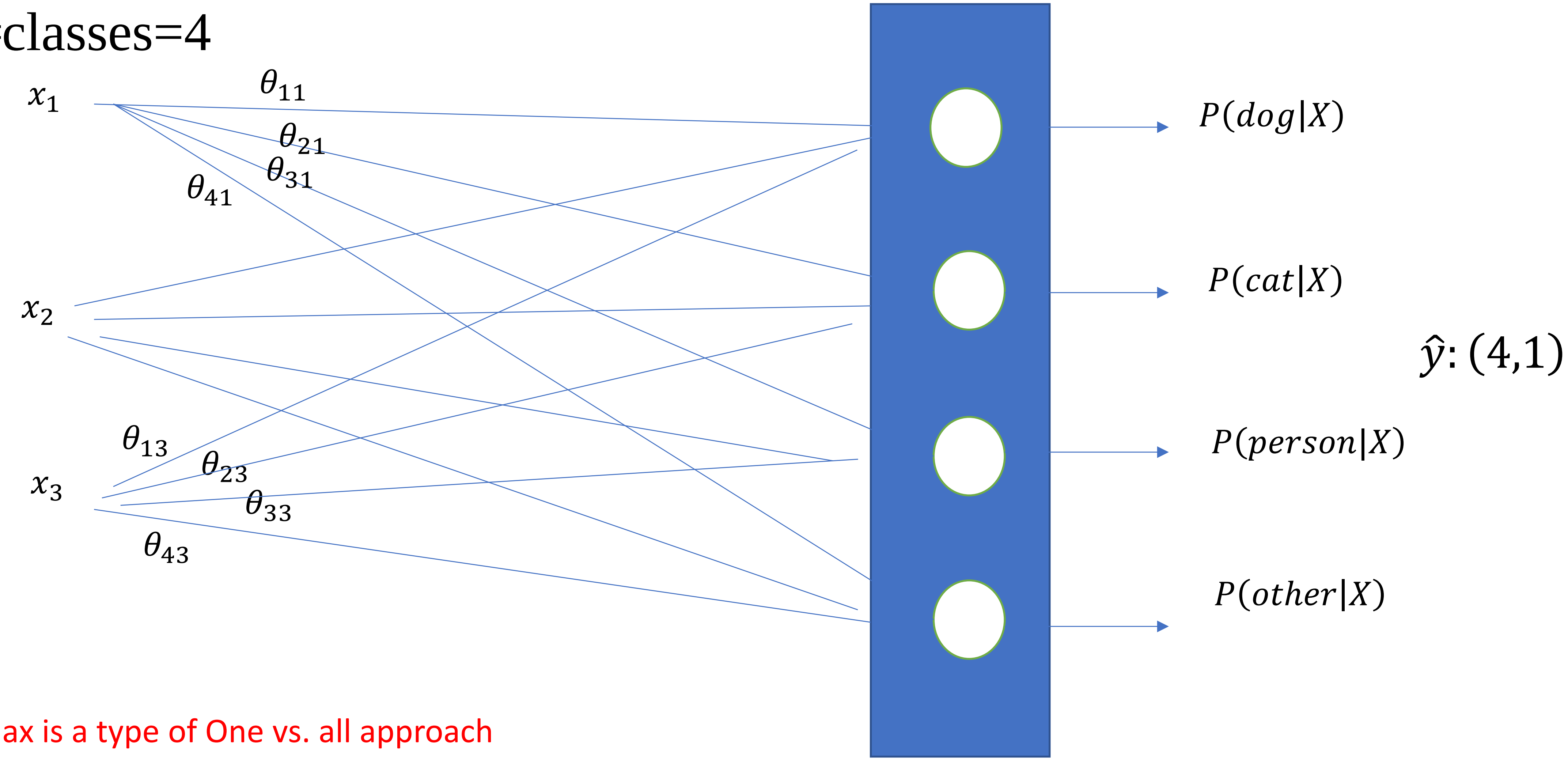
- Classifier 1: Green vs. Blue
- Classifier 2: Green vs. Red
- Classifier 3: Blue vs. Red

Test time:

Each binary classifier predicts one class label. When we input the test data to the classifier, then the model with the majority counts is concluded as a result.

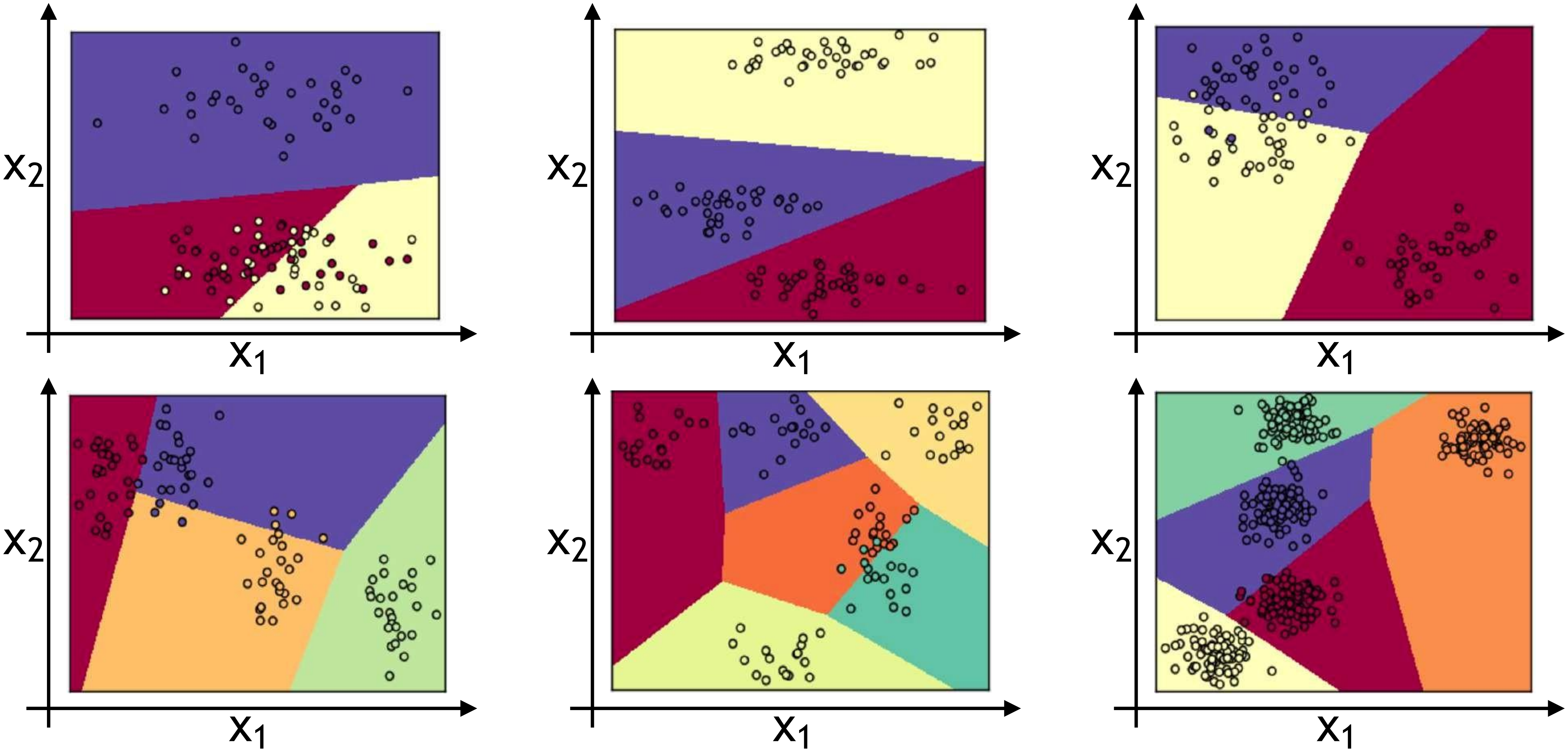
Multi-class classification: Softmax regression

- C=#classes=4



Softmax is a type of One vs. all approach

Softmax examples



Softmax

- $Z \quad = \theta \quad x \quad + b$
- Softmax Activation function

$$t = e^{(z \quad)} \qquad (4,1)$$

$$(4,1) \quad a \quad = \frac{e^{Z}}{\sum_{j=1}^{4} tj} \Rightarrow a_i \quad = \frac{t_i}{\sum_{j=1}^{4} tj}$$

$$a \quad = g \quad (z \quad)$$

Softmax Activation

Softmax examples

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

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

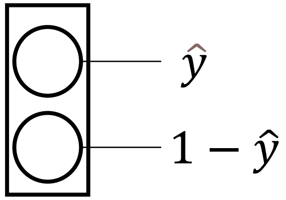
$$a = \frac{t}{176 \cdot 3}$$

$$\frac{1}{176 \cdot 3}\begin{bmatrix} 148.4 \\ 7.4 \\ 0.4 \\ 20.1 \end{bmatrix} = \begin{bmatrix} 0.842 \\ 0.042 \\ 0.002 \\ 0.114 \end{bmatrix} = \hat{y}^{(i)} \qquad \underset{(4,1)}{y^{(i)}} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

one hot vector

Loss function: Categorical Cross Entropy

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



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

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

$y_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ One-hot-vector

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

$\begin{bmatrix} 0 & 1 & 1 & 0 & \\ 1 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \end{bmatrix}$ $\begin{bmatrix} 0.3 & \\ 0.2 & \\ 0.3 & \cdots \\ 0.2 & \end{bmatrix}$

Gradient descent with softmax

• اثبات کنید:

$$dZ_{4\times1} = \hat{y}_{4\times1} - y_{4\times1}$$