AI VIETNAM
Free AI Course 2020

Logistic Regression
(Draft)

Quang-Vinh Dinh
Ph.D. in Computer Science

Year 2020

Outline

- **Sigmoid function**
- From Linear to Logistic Regression
- Logistic Regression – Stochastic
- Logistic Regression – Mini-batch
- Logistic Regression – Batch

Sigmoid Function

Sigmoid function

$$y = \sigma(x) = \frac{1}{1 + e^{-x}}$$

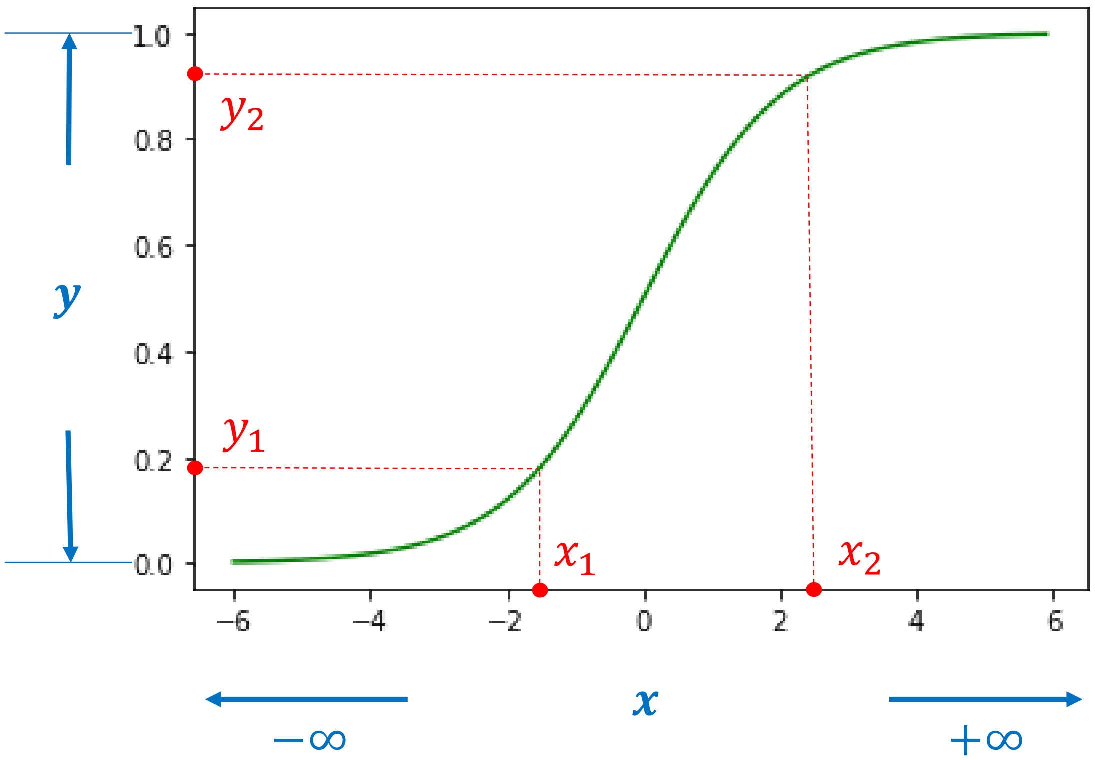
$$x \in (-\infty \quad +\infty)$$

$$y \in (0 \quad 1)$$

Property

$$\forall x_1 x_2 \in [a \ \ b] \text{ và } x_1 \leq x_2$$

$$\rightarrow \sigma(x_1) \leq \sigma(x_1)$$



Sigmoid Function

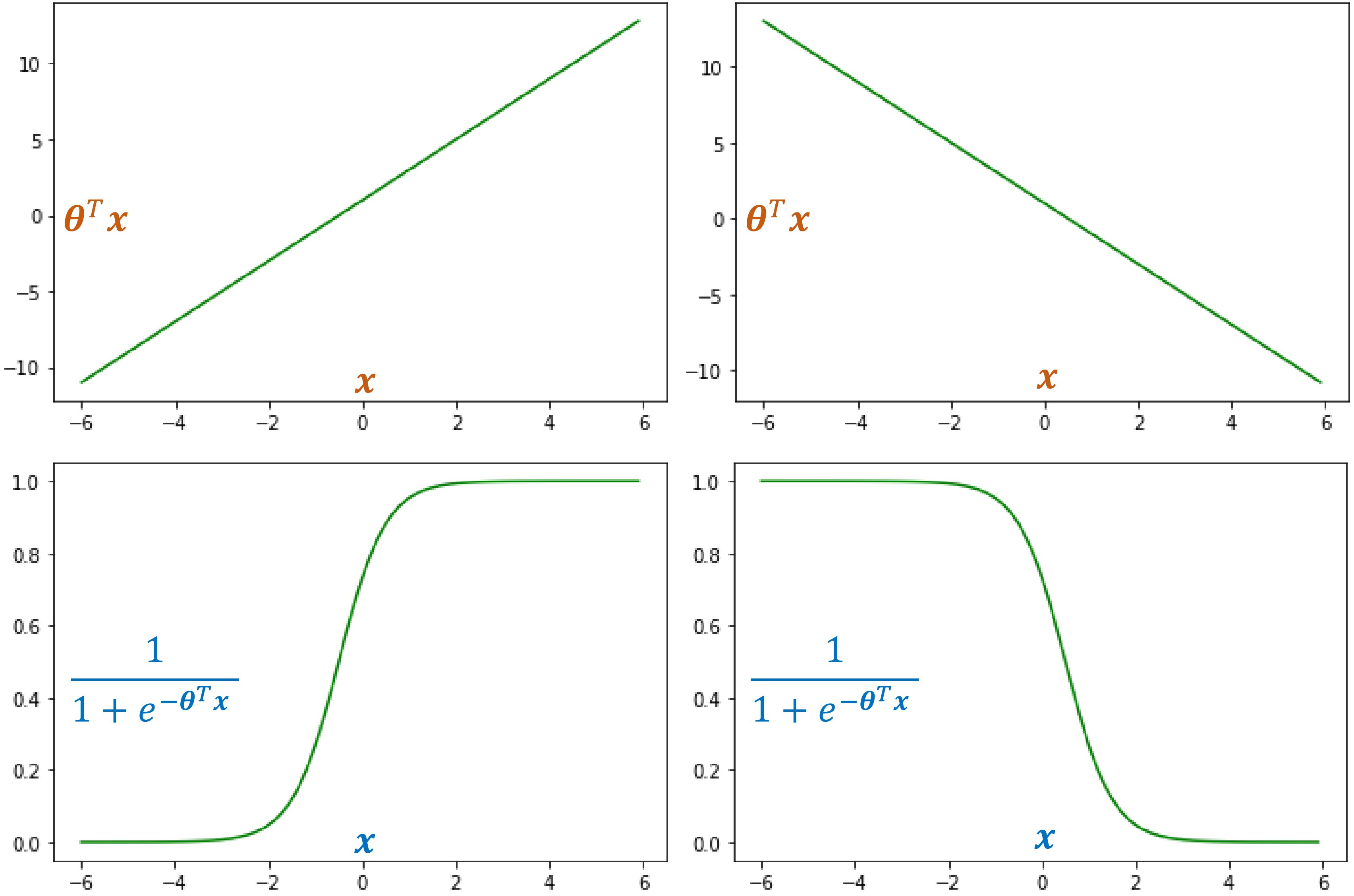
$y = \boldsymbol{\theta}^T \boldsymbol{x}$

$y \in (-\infty \quad +\infty)$

$z = \boldsymbol{\theta}^T \boldsymbol{x}$

$y = \sigma(z) = \dfrac{1}{1 + e^{-z}}$

$y \in (0 \quad 1)$



Sigmoid Function

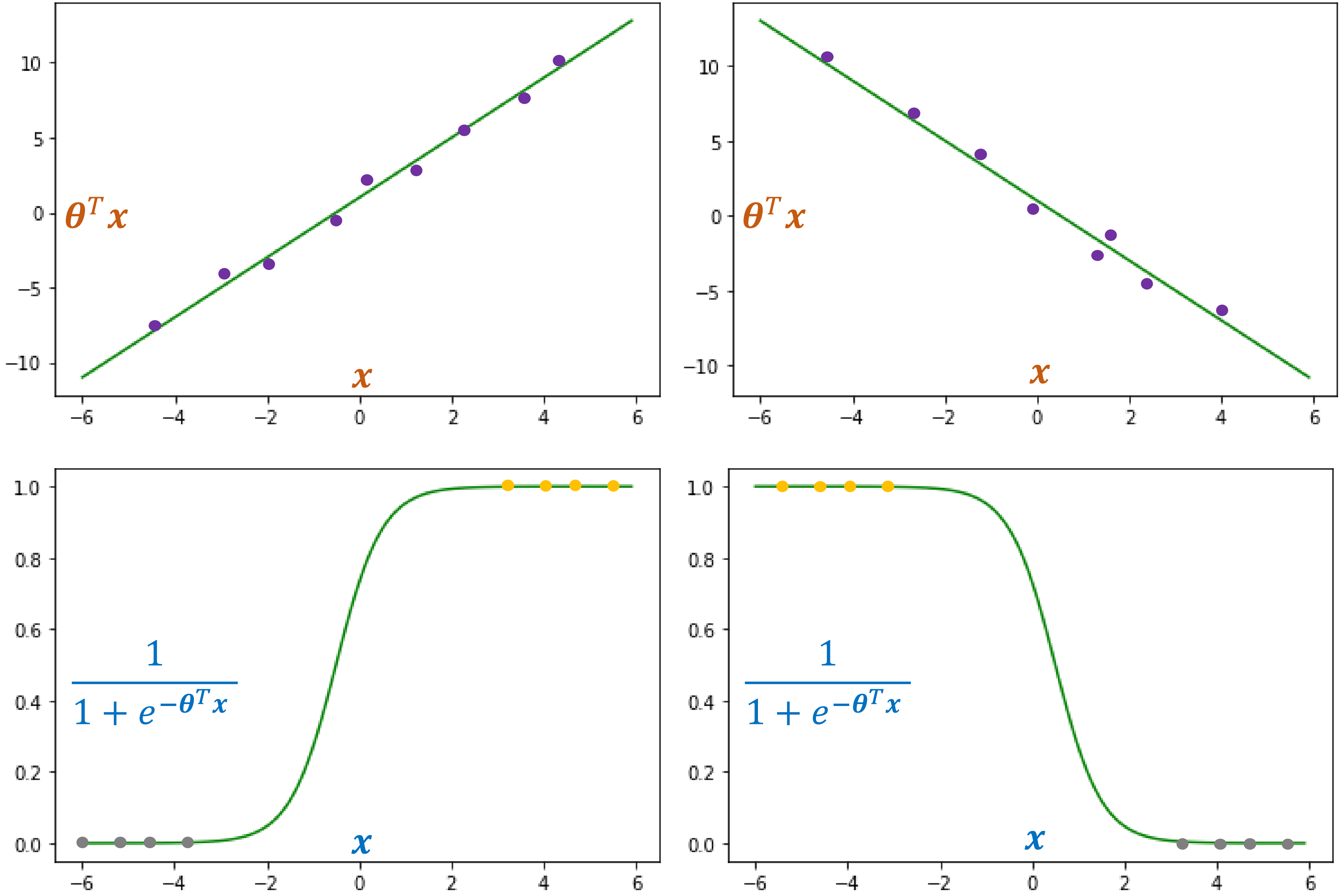
$$y = \boldsymbol{\theta}^T \boldsymbol{x}$$

$$y \in (-\infty \quad +\infty)$$

$$z = \boldsymbol{\theta}^T \boldsymbol{x}$$

$$y = \sigma(z) = \frac{1}{1 + e^{-z}}$$

$$y \in (0 \quad 1)$$



Outline

- Sigmoid function
- **From Linear to Logistic Regression**
- Logistic Regression – Stochastic
- Logistic Regression – Mini-batch
- Logistic Regression – Batch

Idea of Logistic Regression

❖ Linear regression

Area-based House Price Data

Feature	Label
area	price
6.7	9.1
4.6	5.9
3.5	4.6
5.5	6.7

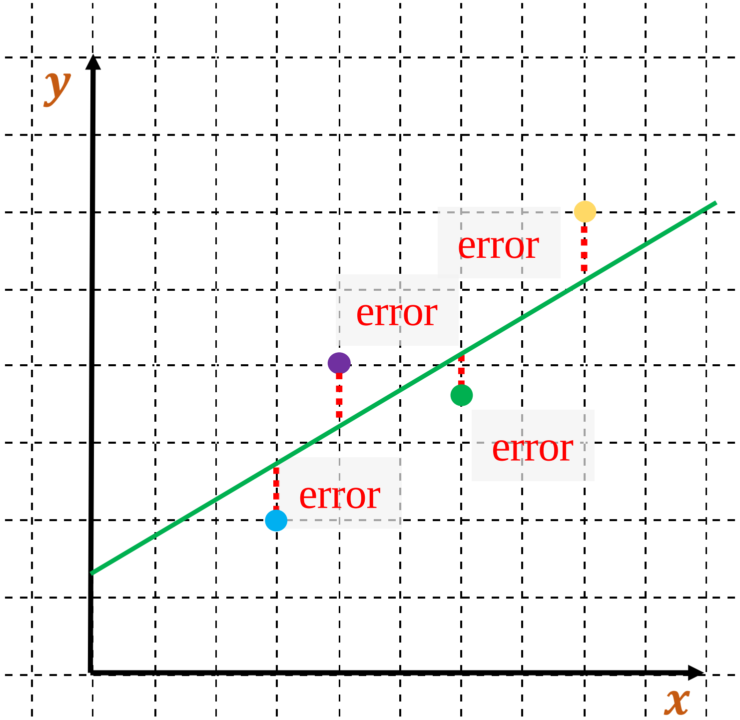
Training data

construct →

$y = \boldsymbol{\theta}^T \boldsymbol{x} = ax + b$

$y \in (-\infty \quad +\infty)$

Model



Find the line $y = \boldsymbol{\theta}^T \boldsymbol{x}$ that is best fit a given data, then use y to predict for new data

Idea of Logistic Regression

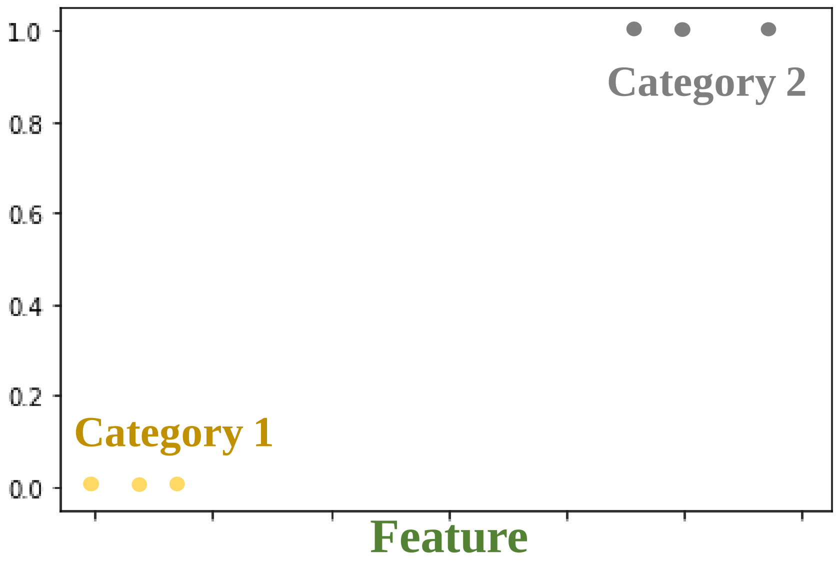
❖ Given a new kind of data

Petal_Length (Feature)	Category (Label)	
1.4	Flower A	Category 1
1	Flower A	Category 1
1.5	Flower A	Category 1
3	Flower B	Category 2
3.8	Flower B	Category 2
4.1	Flower B	Category 2

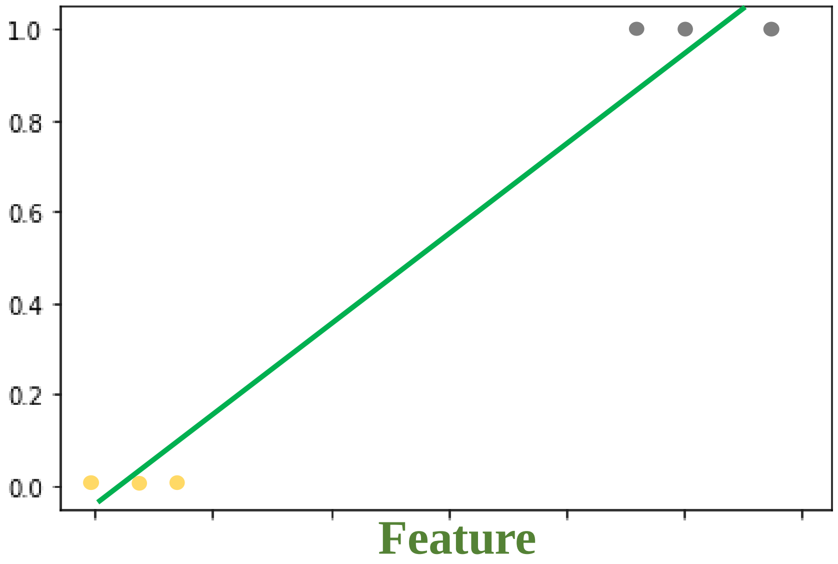
Assign numbers to categories

Petal_Length (Feature)	Category (Label)	
1.4	0	Category 1
1	0	Category 1
1.5	0	Category 1
3	1	Category 2
3.8	1	Category 2
4.1	1	Category 2

Plot data



A line is not suitable for this data



Idea of Logistic Regression

❖ Given a new kind of data

Feature | **Label**

Petal_Length	Category	
1.4	Flower A	Category 1
1	Flower A	Category 1
1.5	Flower A	Category 1
3	Flower B	Category 2
3.8	Flower B	Category 2
4.1	Flower B	Category 2

Assign numbers to categories

Feature | **Label**

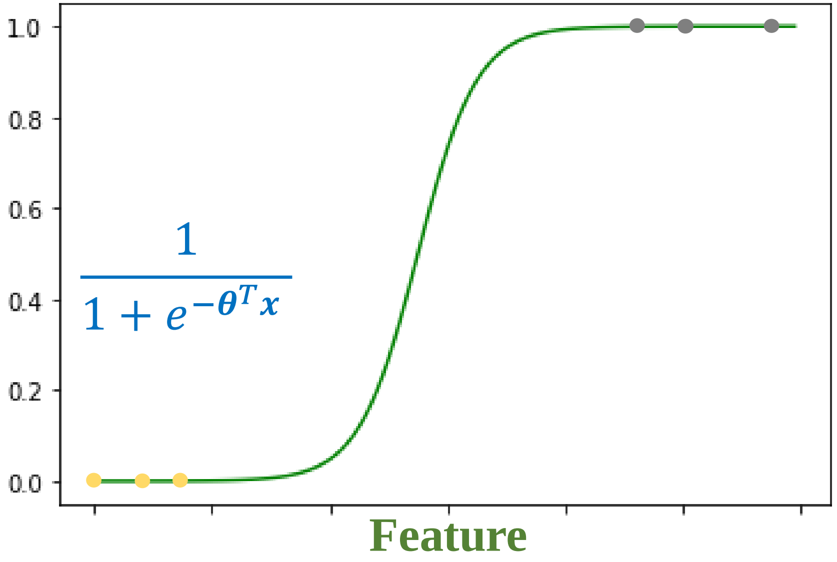
Petal_Length	Category	
1.4	0	Category 1
1	0	Category 1
1.5	0	Category 1
3	1	Category 2
3.8	1	Category 2
4.1	1	Category 2

Sigmoid function could fit the data

$$z = \boldsymbol{\theta}^T \boldsymbol{x}$$

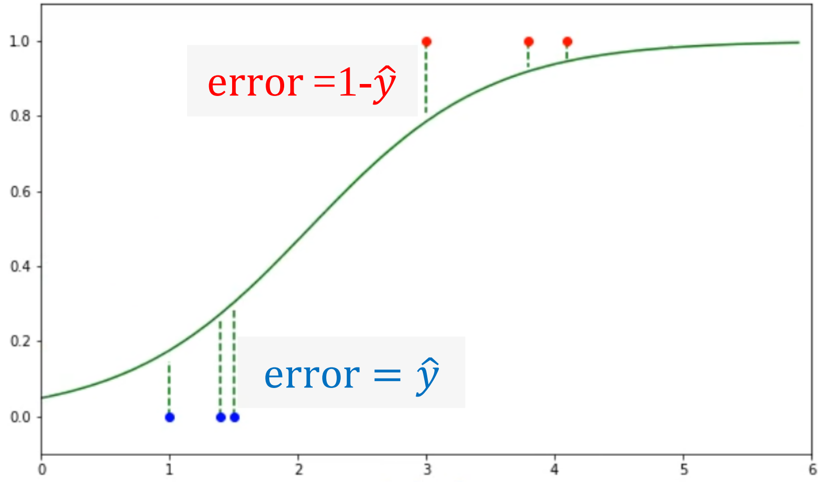
$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

$$\hat{y} \in (0 \quad 1)$$



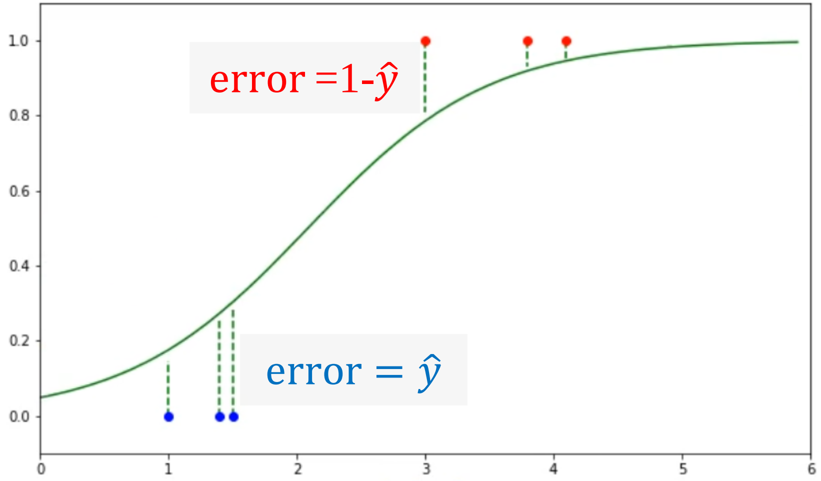
Error

$$\text{if } y = 1$$
$$\text{error} = 1 - \hat{y}$$
$$\text{if } y = 0$$
$$\text{error} = \hat{y}$$



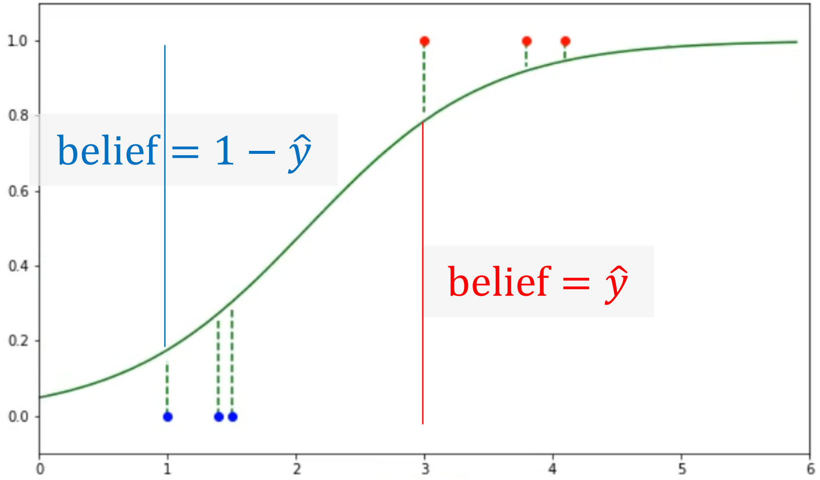
Idea of Logistic Regression

❖ Construct loss



Error

if $y = 1$
  error $= 1 - \hat{y}$
if $y = 0$
  error $= \hat{y}$



Belief

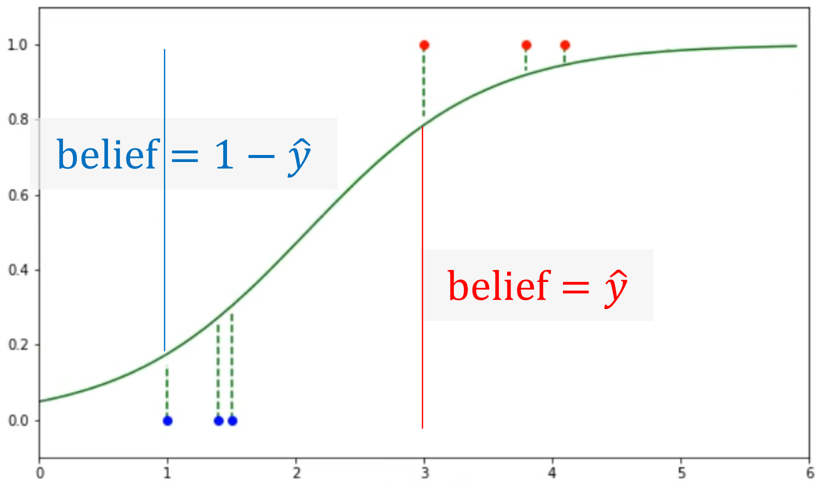
if $y = 1$
  belief $= \hat{y}$
if $y = 0$
  belief $= 1 - \hat{y}$

$$P = \hat{y}_i^{y_i}(1 - \hat{y}_i)^{1-y_i}$$

Minimize error ~ maximize belief ~ Minimize (-belief)

Idea of Logistic Regression

❖ Construct loss



Belief

if $y = 1$

$\quad$ $\text{belief} = \hat{y}$

if $y = 0$

$\quad$ $\text{belief} = 1 - \hat{y}$

$$P_i = \hat{y}_i^{y_i}(1 - \hat{y}_i)^{1-y_i}$$

$$\text{belief} = \prod_{i=1}^{n} P_i \qquad \text{since iid}$$

$$\text{log\_belief} = \sum_{i=1}^{n} \log P_i$$

$$\text{log\_belief} = \sum_{i=1}^{n} [y_i \log\hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)]$$

$$\text{loss} = -\text{log\_belief}$$

$$= -\sum_{i=1}^{n} [y_i \log\hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)]$$

$$\text{L} = \frac{1}{N}\left(-y^T log(\hat{y}) - (1 - y^T) log(1 - \hat{y})\right)$$

Binary cross-entropy

Idea of Logistic Regression

❖ Construct loss

Model and Loss

$$z = \boldsymbol{\theta}^T \boldsymbol{x}$$

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

$$\mathrm{L} = \frac{1}{N}[-y^T \log(\hat{y}) - (1 - y^T)\log(1 - \hat{y})]$$

Derivative

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z} \frac{\partial z}{\partial \theta}$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{1}{N}\left(-\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right) = \frac{1}{N} \frac{\hat{y} - y}{\hat{y}(1 - \hat{y})}$$

$$\frac{\partial \hat{y}}{\partial z} = \hat{y}(1 - \hat{y})$$

$$\frac{\partial z}{\partial \theta} = x$$

$$\frac{\partial L}{\partial \theta} = \frac{1}{N} x^T (\hat{y} - y)$$

Add simple derivative

Idea of Logistic Regression

❖ Construct loss

Model and Loss

$$z = \boldsymbol{\theta}^T \boldsymbol{x}$$

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

$$\mathrm{L} = \frac{1}{N}[-y^T \log(\hat{y}) - (1 - y^T)\log(1 - \hat{y})]$$

Derivative

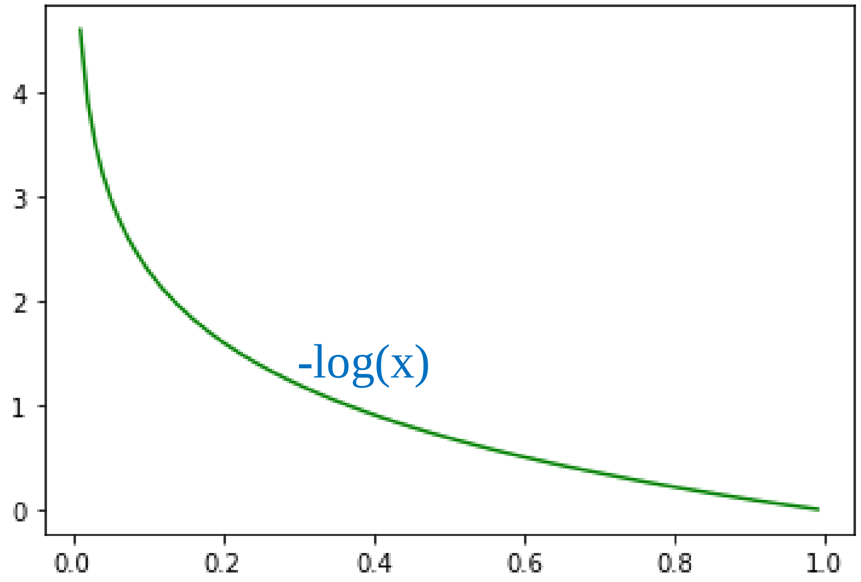
$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z} \frac{\partial z}{\partial \theta}$$

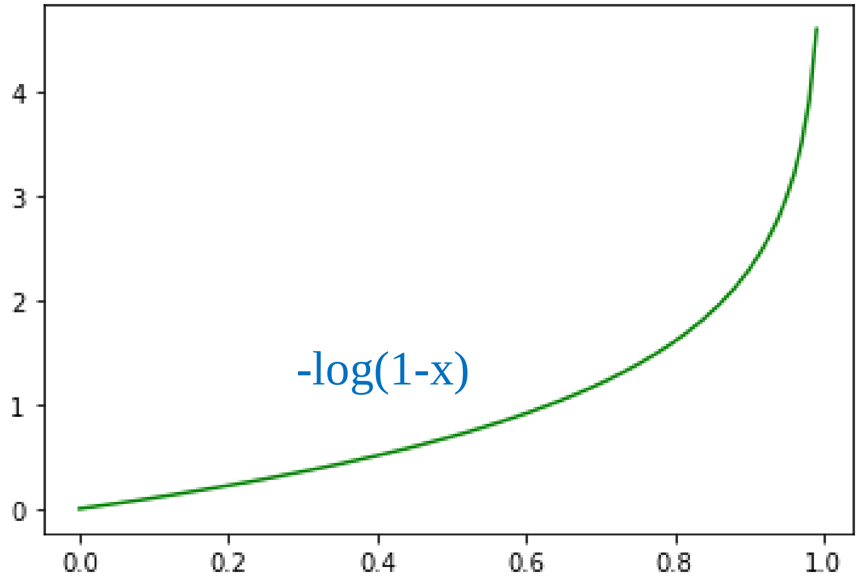
$$\frac{\partial L}{\partial \hat{y}} = \frac{1}{N}\left(-\frac{y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}}\right) = \frac{1}{N} \frac{\hat{y} - y}{\hat{y}(1 - \hat{y})}$$

$$\frac{\partial \hat{y}}{\partial z} = \hat{y}(1 - \hat{y})$$

$$\frac{\partial z}{\partial \theta} = x$$

$$\frac{\partial L}{\partial \theta} = \frac{1}{N} x^T (\hat{y} - y)$$





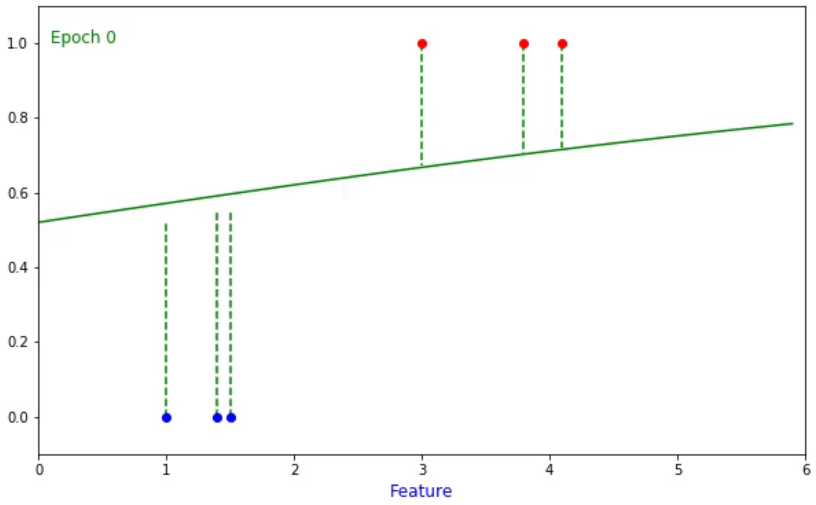
Idea of Logistic Regression

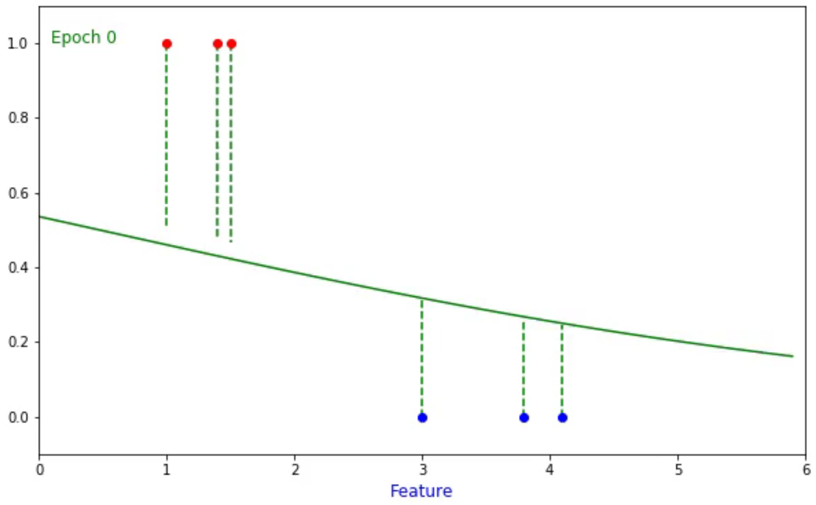
Feature	Label	
Petal_Length	Category	
1.4	0	Category 1
1	0	
1.5	0	
3	1	Category 2
3.8	1	
4.1	1	

$$z = \boldsymbol{\theta}^T \boldsymbol{x}$$

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

Feature	Label	
Petal_Length	Category	
1.4	1	Category 1
1	1	
1.5	1	
3	0	Category 2
3.8	0	
4.1	0	





Outline

- Sigmoid function
- From Linear to Logistic Regression
- **Logistic Regression – Stochastic**
- Logistic Regression – Mini-batch
- Logistic Regression – Batch

Logistic Regression-Stochastic

1) Pick a sample (x, y) from training data

2) Tính output $\hat{y}$

$$z = \boldsymbol{\theta}^T \boldsymbol{x}$$

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

3) Tính loss

$$L(\boldsymbol{\theta}) = \left(-\mathrm{y}^{\mathrm{T}}\log\hat{\mathrm{y}} - (1-\mathrm{y})^{\mathrm{T}}\log(1-\hat{\mathrm{y}})\right)$$

4) Tính đạo hàm

$$L'_{\boldsymbol{\theta}} = \mathrm{x}^{\mathrm{T}}(\hat{\mathrm{y}} - y)$$

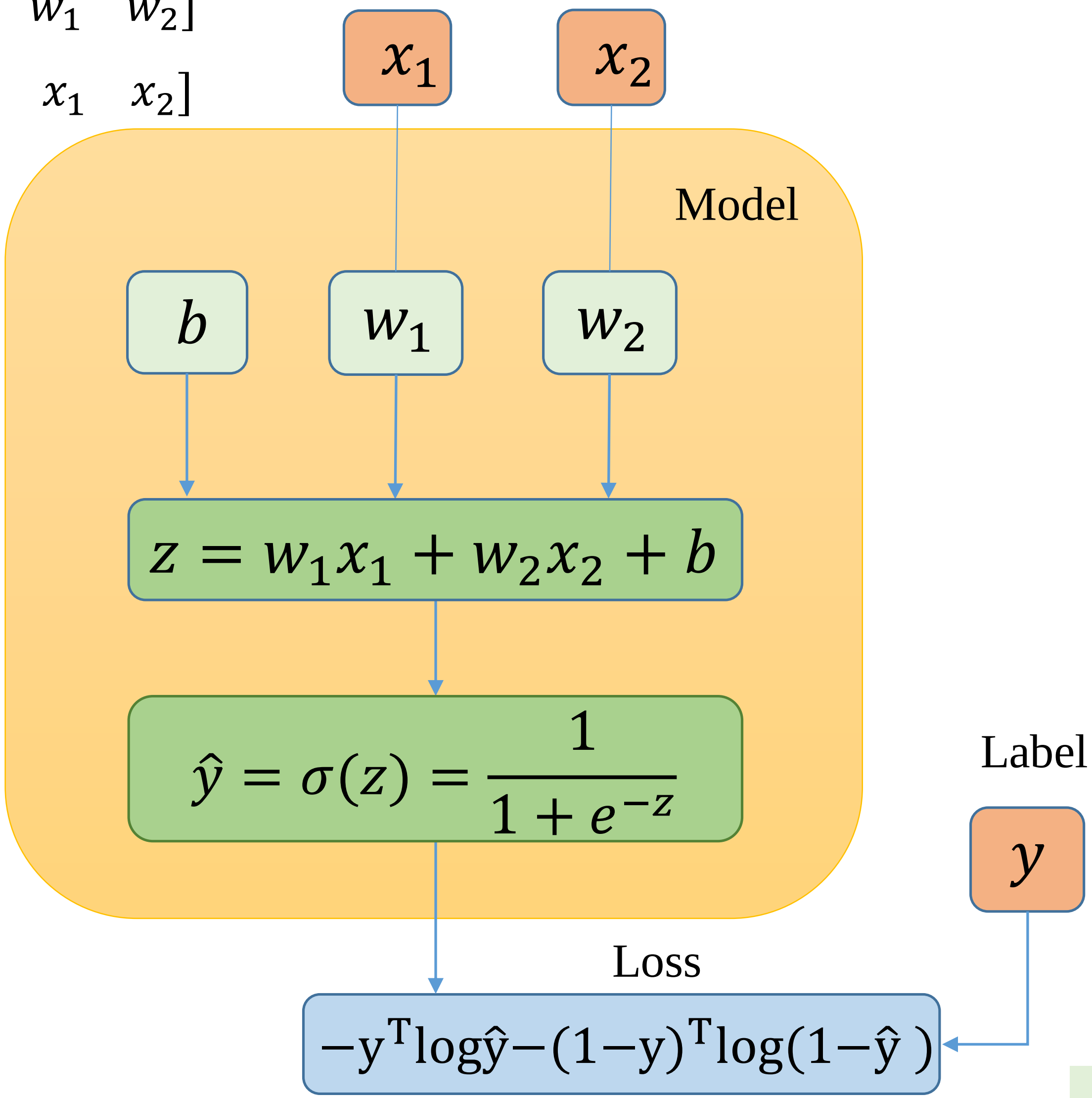
5) Cập nhật tham số

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta L'_{\boldsymbol{\theta}}$$

η is learning rate

$$\boldsymbol{\theta}^T = [b \quad w_1 \quad w_2]$$

$$\boldsymbol{x}^T = [1 \quad x_1 \quad x_2]$$

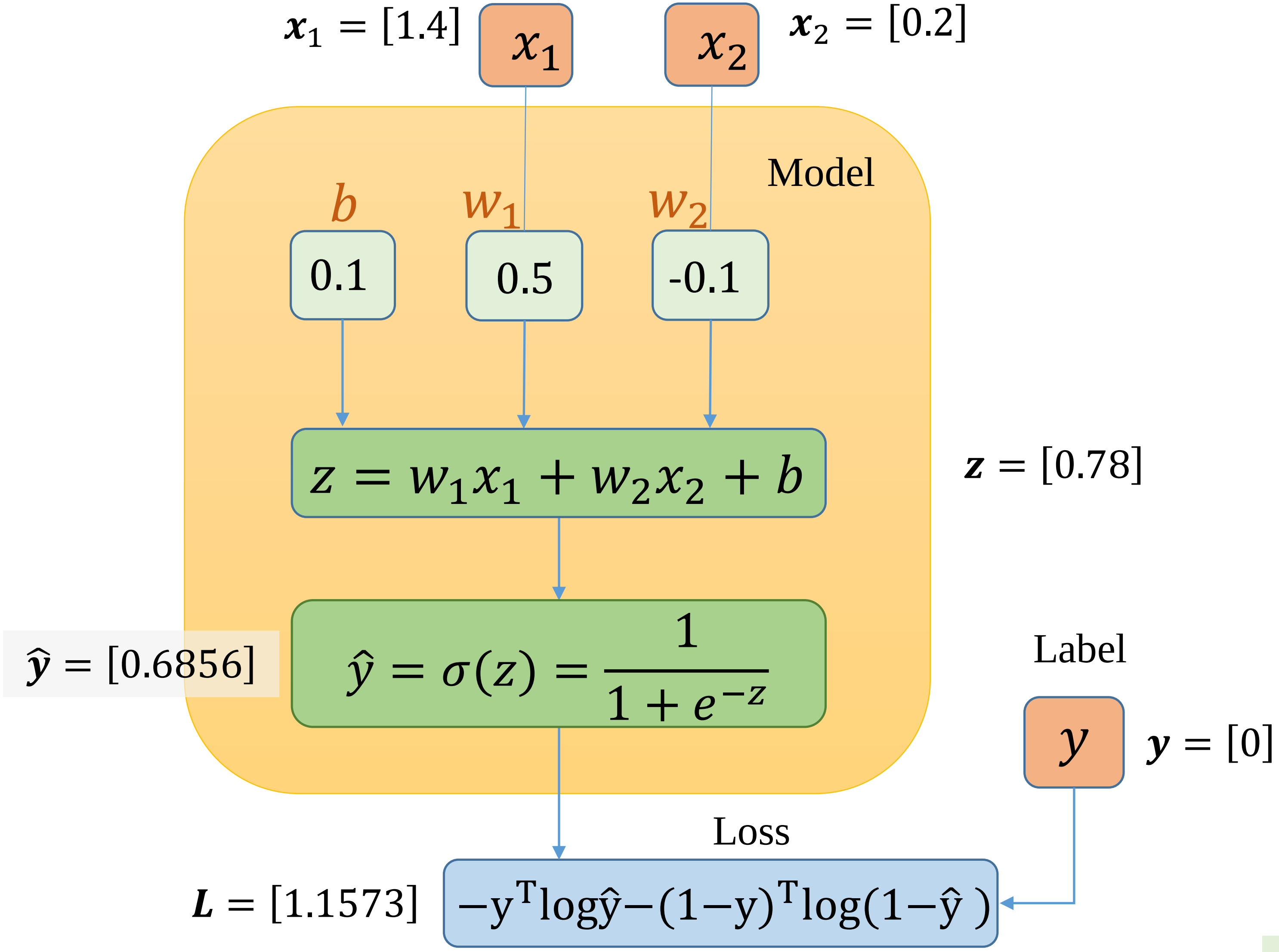


Logistic Regression-Stochastic

Dataset

Petal_Length	Petal_Width	Label
1.4	0.2	0
1.5	0.2	0
3	1.1	1
4.1	1.3	1

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 1.4 \\ 0.2 \end{bmatrix} \qquad \boldsymbol{y} = [0]$$



Logistic Regression-Stochastic

Dataset

Petal_Length	Petal_Width	Label
1.4	0.2	0
1.5	0.2	0
3	1.1	1
4.1	1.3	1

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 1.4 \\ 0.2 \end{bmatrix} \qquad \boldsymbol{y} = [0]$$

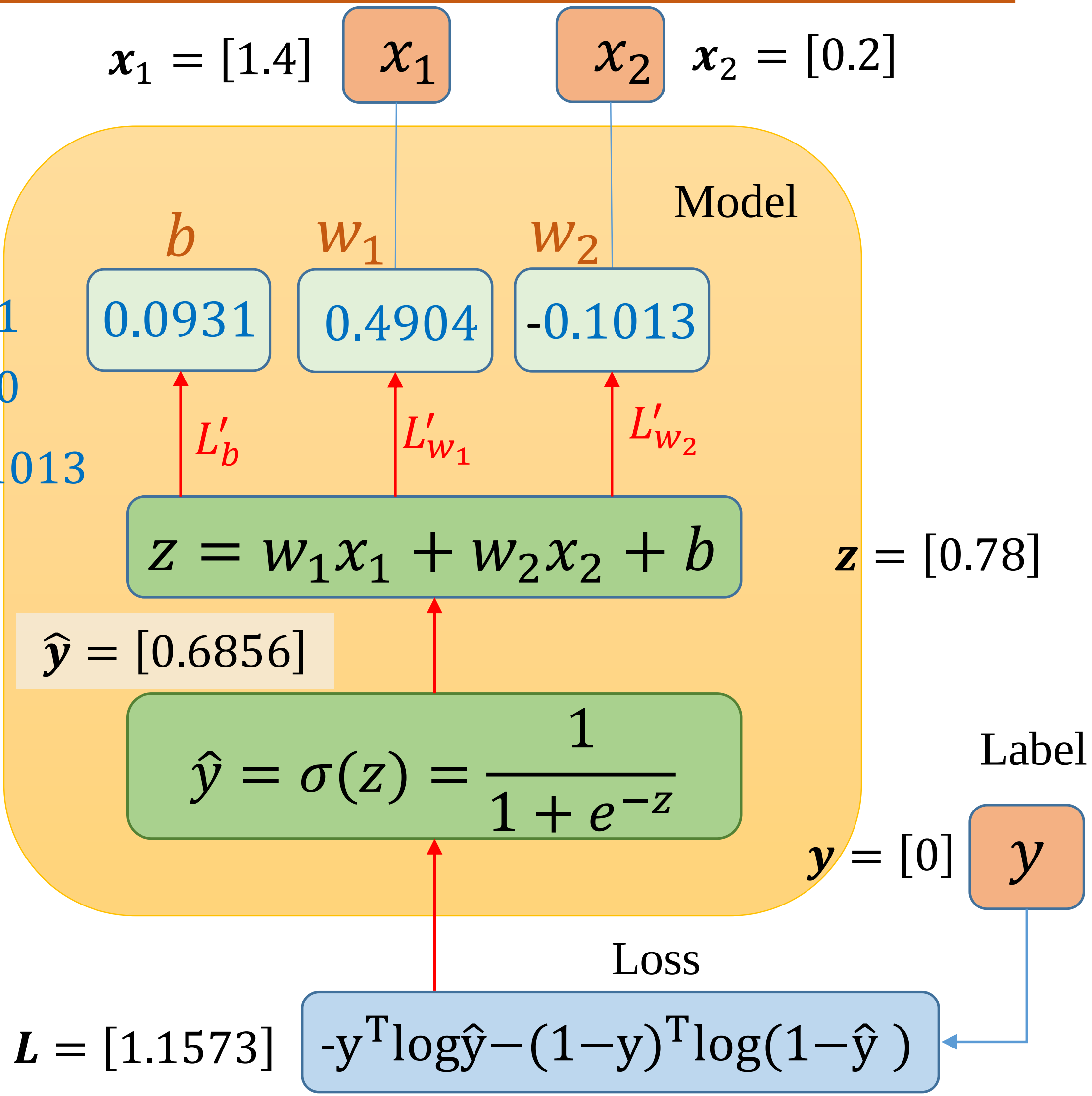
$\eta = 0.01$

$b = 0.1 - \eta 0.6856 = 0.0931$

$w_1 = 0.5 - \eta 0.9598 = 0.4990$

$w_2 = -0.1 + \eta 0.1371 = -0.1013$

$$L'_{\boldsymbol{\theta}} = \mathrm{x}^{\mathrm{T}}(\hat{\mathrm{y}} - y)$$
$$= [1 \;\; 1.4 \;\; 0.2][0.6856]$$
$$= \begin{bmatrix} 0.6856 \\ 0.9599 \\ -0.1371 \end{bmatrix} = \begin{bmatrix} L'_b \\ L'_{w_1} \\ L'_{w_2} \end{bmatrix}$$



Outline

- Sigmoid function
- From Linear to Logistic Regression
- Logistic Regression – Stochastic
- **Logistic Regression – Mini-batch**
- Logistic Regression – Batch

Logistic Regression - Minibatch

1) Pick m samples from training data

2) Tính output $\hat{y}$

$$z = \boldsymbol{\theta}^T \boldsymbol{x}$$

$$\hat{\boldsymbol{y}} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

3) Tính loss

$$L(\boldsymbol{\theta}) = \frac{1}{\mathrm{m}}\left(-\mathrm{y}^{\mathrm{T}}\log\hat{\mathrm{y}} - (1-\mathrm{y})^{\mathrm{T}}\log(1-\hat{\mathrm{y}})\right)$$

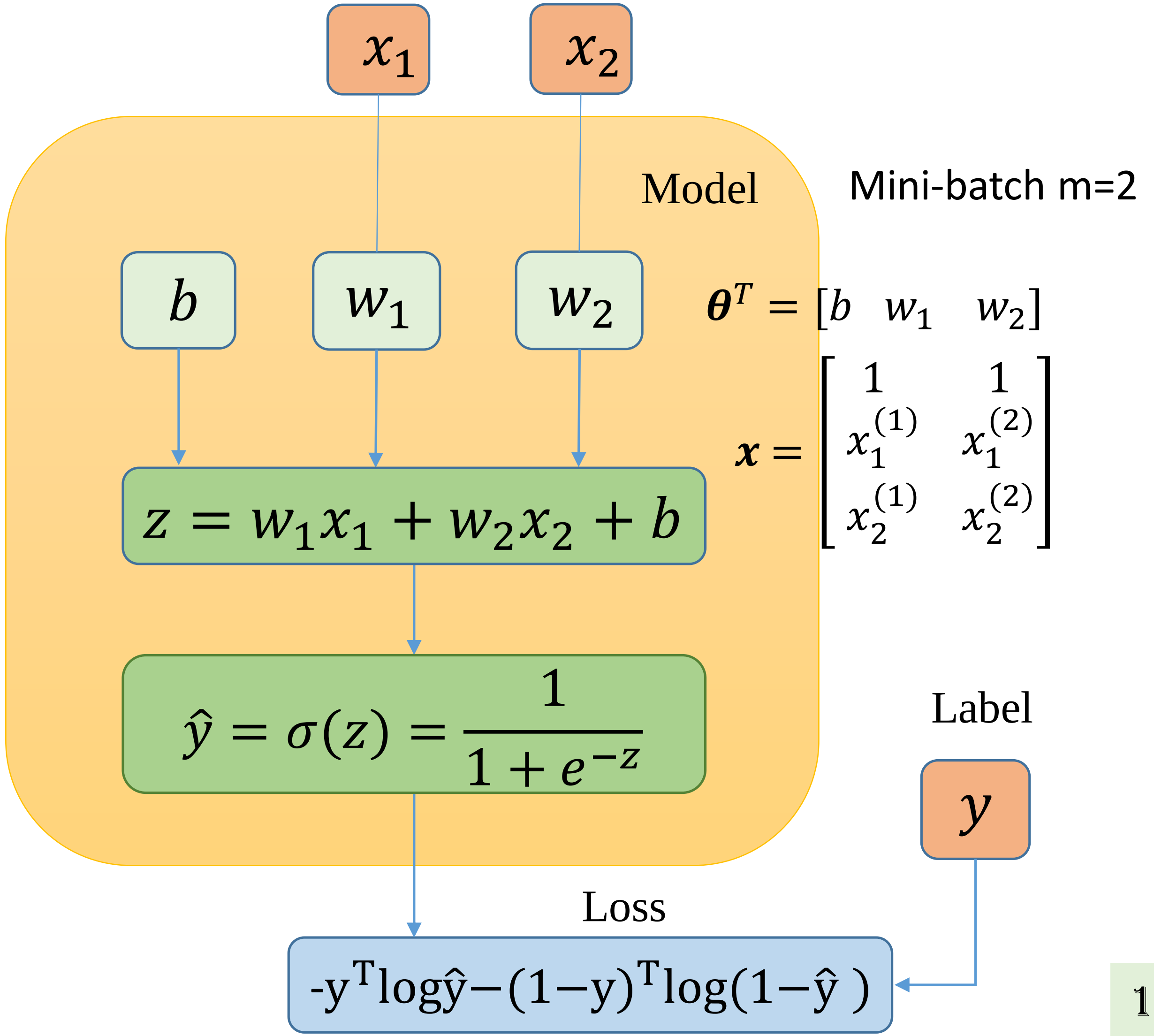
4) Tính đạo hàm

$$L'_{\boldsymbol{\theta}} = \frac{1}{\mathrm{m}}\mathrm{x}^{\mathrm{T}}(\hat{\mathrm{y}} - y)$$

5) Cập nhật tham số

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta L'_{\boldsymbol{\theta}}$$

η is learning rate

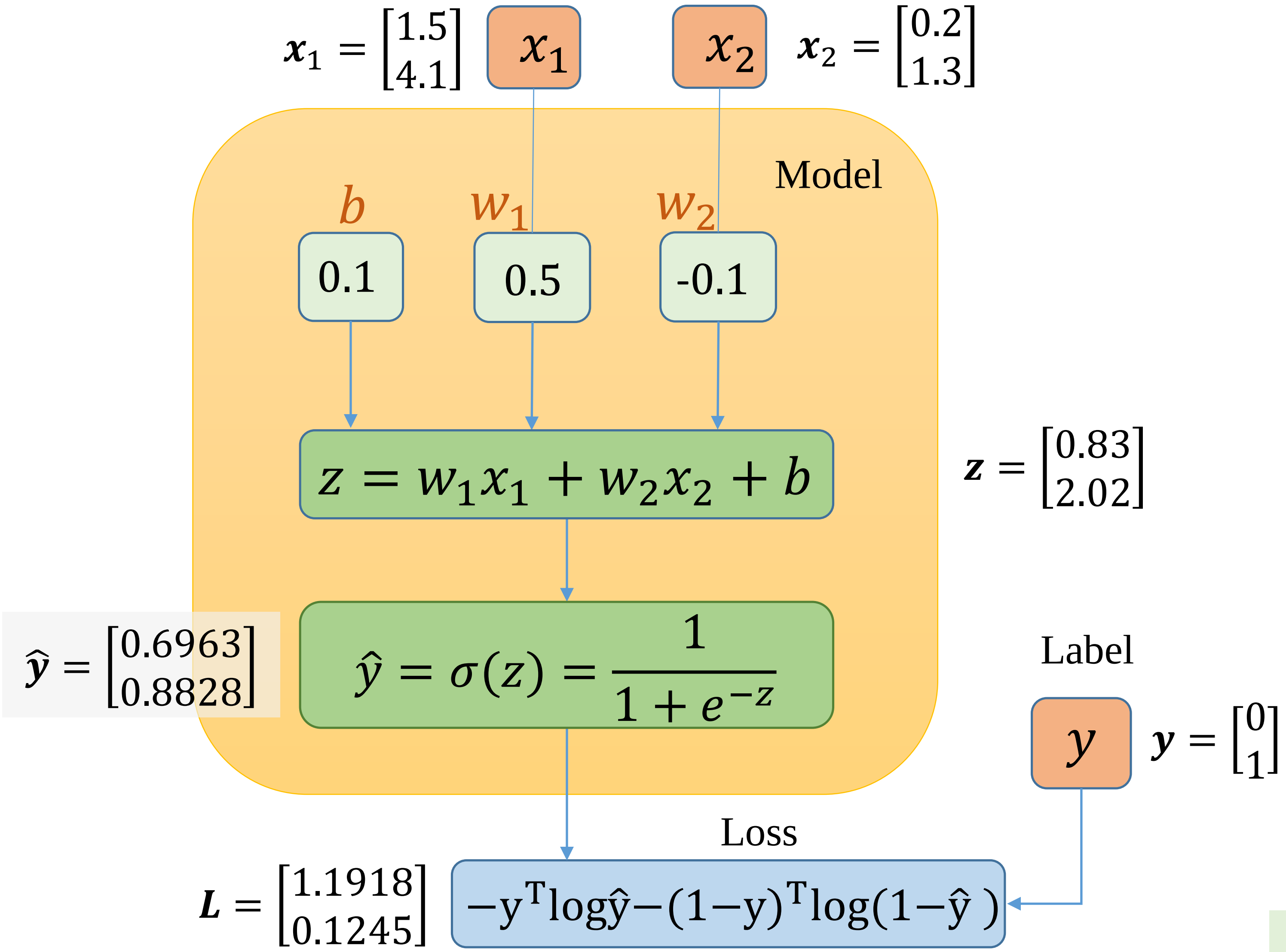


Logistic Regression - Minibatch

Dataset

Petal_Length	Petal_Width	Label
1.4	0.2	0
1.5	0.2	0
3	1.1	1
4.1	1.3	1

$$\boldsymbol{x} = \begin{bmatrix} 1 & 1 \\ 1.5 & 4.1 \\ 0.2 & 1.3 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$



$$b = 0.1 - \eta 0.28961 = 0.097103$$

$$w_1 = 0.5 - \eta 0.28217 = 0.49717$$

$$w_2 = -0.1 + \eta 0.0064 = -0.09993$$

Dataset

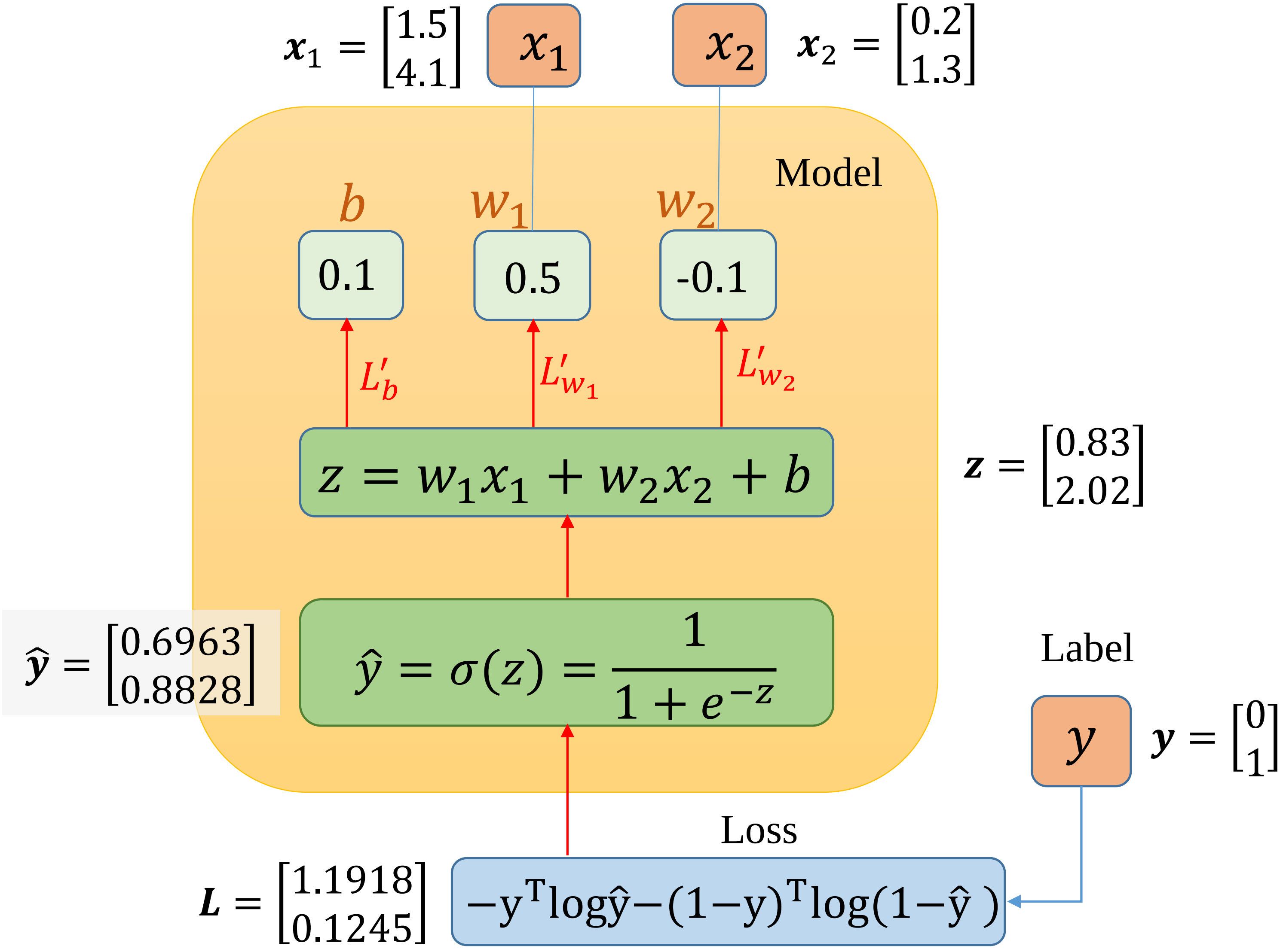
Petal_Length	Petal_Width	Label
1.4	0.2	0
1.5	0.2	0
3	1.1	1
4.1	1.3	1

$$\boldsymbol{x} = \begin{bmatrix} 1 & 1.5 & 0.2 \\ 1 & 4.1 & 1.3 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$L'_{\boldsymbol{\theta}} = \frac{1}{\mathrm{N}} \mathrm{x}^{\mathrm{T}} (\hat{\mathrm{y}} - y)$$

$$= \frac{1}{4} \begin{bmatrix} 1.0 & 1.0 \\ 1.5 & 4.1 \\ 0.2 & 1.3 \end{bmatrix} \begin{bmatrix} 0.6963 \\ -0.1171 \end{bmatrix}$$

$$= \begin{bmatrix} 0.28961 \\ 0.28217 \\ -0.0064 \end{bmatrix} = \begin{bmatrix} L'_b \\ L'_{w_1} \\ L'_{w_2} \end{bmatrix}$$



Average loss =

Outline

- Sigmoid function
- From Linear to Logistic Regression
- Logistic Regression – Stochastic
- Logistic Regression – Mini-batch
- **Logistic Regression – Batch**

Logistic Regression - Batch

1) Pick all the samples from training data

2) Tính output $\hat{y}$

$$z = \boldsymbol{\theta}^T \boldsymbol{x}$$

$$\hat{\boldsymbol{y}} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

3) Tính loss (binary cross-entropy)

$$L(\boldsymbol{\theta}) = \frac{1}{\mathrm{N}}\left(-\mathrm{y}^{\mathrm{T}}\log\hat{\mathrm{y}} - (1-\mathrm{y})^{\mathrm{T}}\log(1-\hat{\mathrm{y}})\right)$$

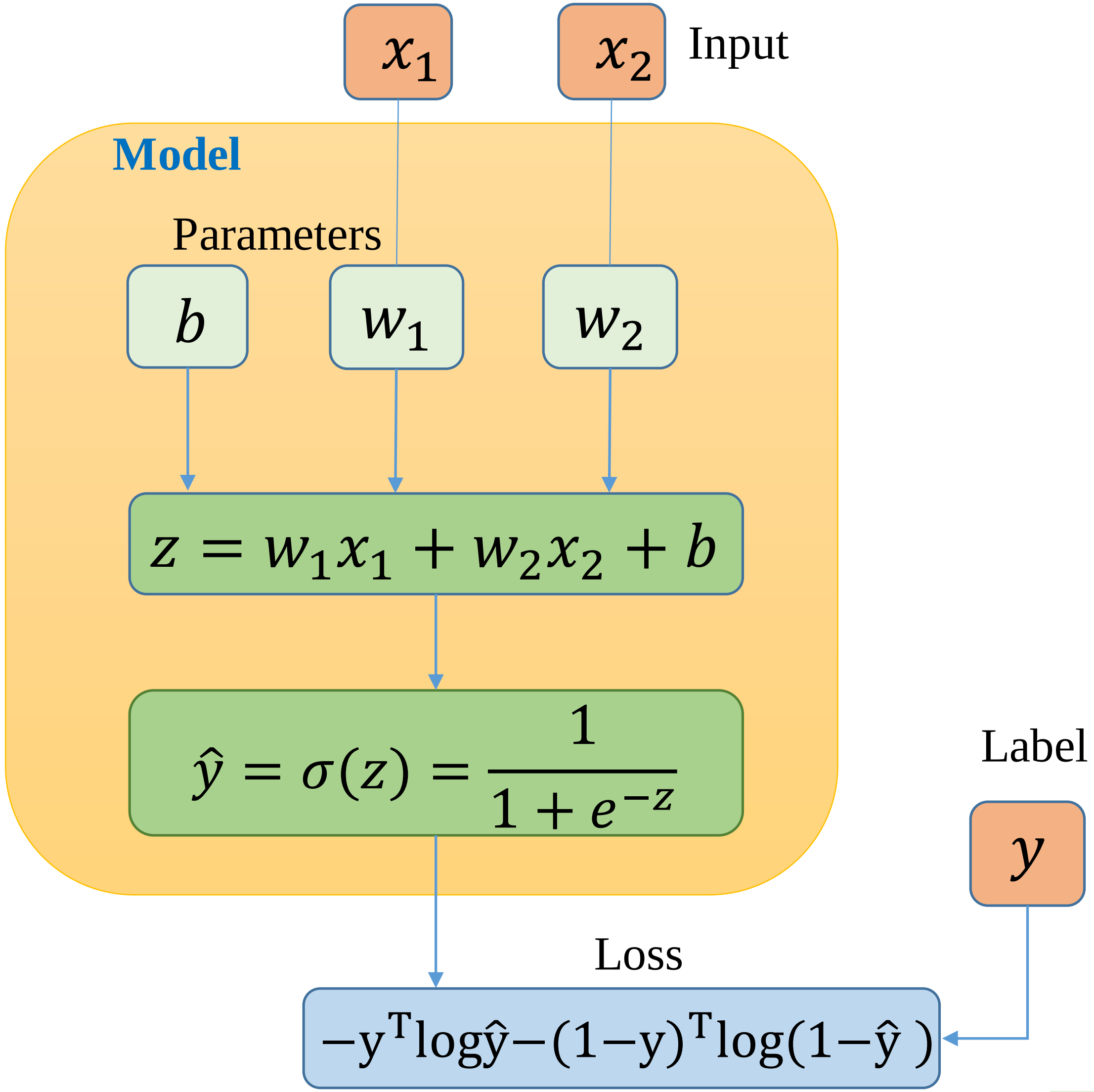
4) Tính đạo hàm

$$L'_{\boldsymbol{\theta}} = \frac{1}{\mathrm{N}}\mathrm{x}^{\mathrm{T}}(\hat{\mathrm{y}} - y)$$

5) Cập nhật tham số

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta L'_{\boldsymbol{\theta}}$$

η is learning rate



Logistic Regression

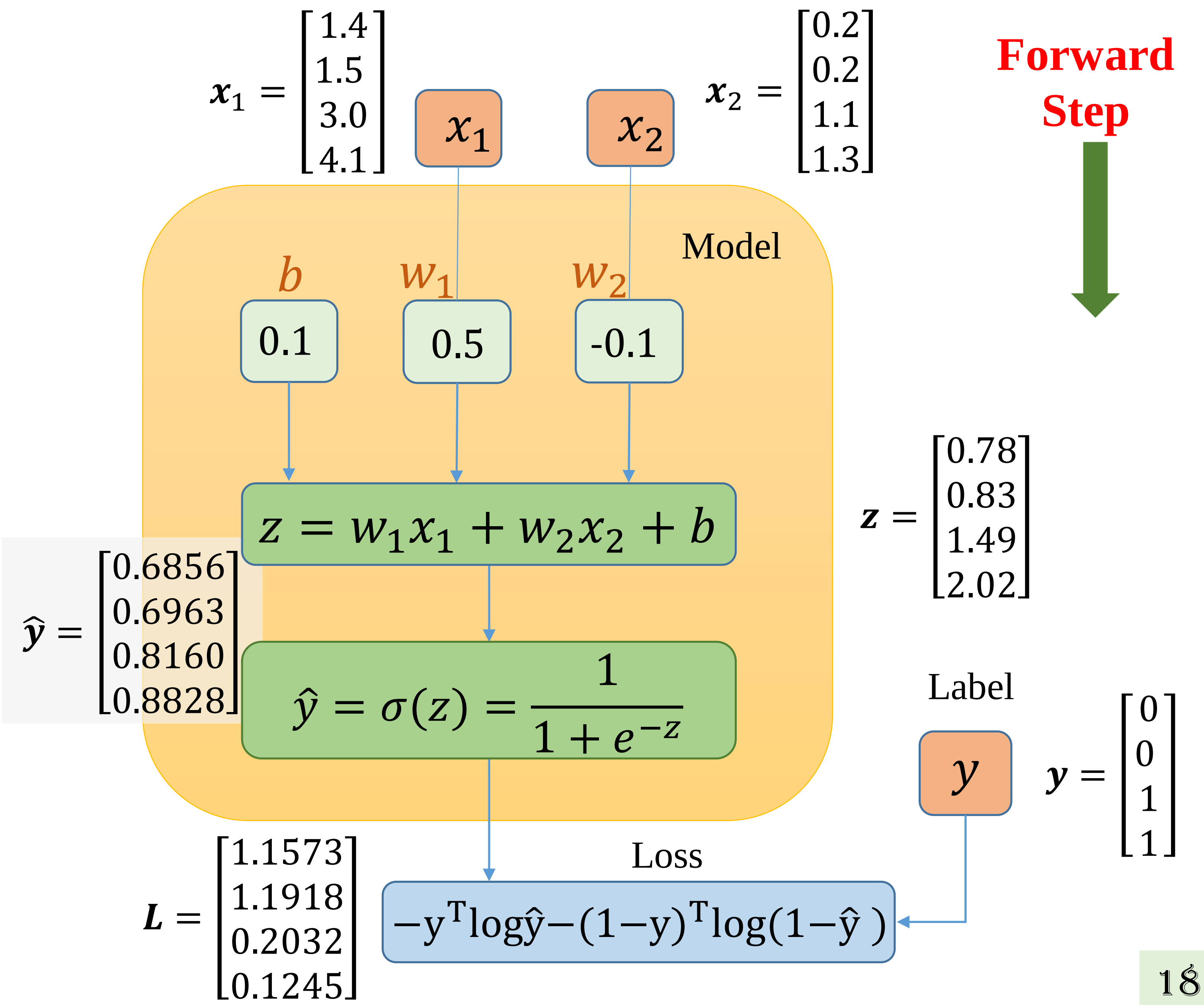
Phân loại hoa Iris dựa vào chiều dài và chiều rộng của cánh hoa

Dataset

Petal_Length	Petal_Width	Label
1.4	0.2	0
1.5	0.2	0
3	1.1	1
4.1	1.3	1

$$\boldsymbol{x} = \begin{bmatrix} 1 & 1.4 & 0.2 \\ 1 & 1.5 & 0.2 \\ 1 & 3.0 & 1.1 \\ 1 & 4.1 & 1.3 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Average loss = 0.6692



Phân loại hoa Iris dựa vào chiều dài và chiều rộng của cánh hoa

Backward Step

Dataset

Petal_Length	Petal_Width	Label
1.4	0.2	0
1.5	0.2	0
3	1.1	1
4.1	1.3	1

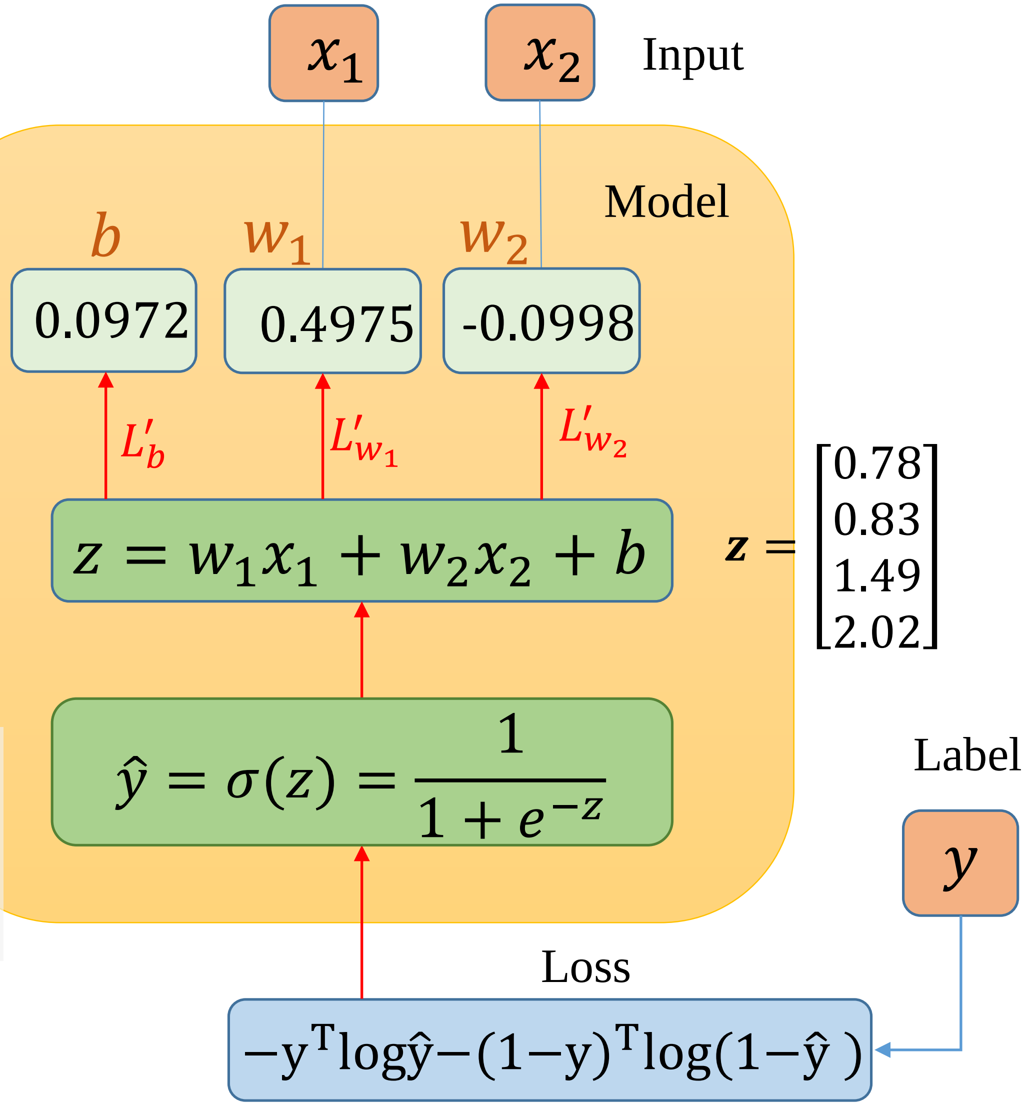
$$\boldsymbol{x} = \begin{bmatrix} 1 & 1.4 & 0.2 \\ 1 & 1.5 & 0.2 \\ 1 & 3.0 & 1.1 \\ 1 & 4.1 & 1.3 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{x}^{\mathrm{T}} = \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 \\ 1.4 & 1.5 & 3.0 & 4.1 \\ 0.2 & 0.2 & 1.1 & 1.3 \end{bmatrix}$$

$\eta = 0.01$

$$b = 0.1 - \eta 0.2702 = 0.0972$$
$$w_1 = 0.5 - \eta 0.2431 = 0.4975$$
$$w_2 = -0.1 + \eta 0.0195 = -0.0998$$

$$\hat{\boldsymbol{y}} = \begin{bmatrix} 0.6856 \\ 0.6963 \\ 0.8160 \\ 0.8828 \end{bmatrix}$$



$$L'_{\boldsymbol{\theta}} = \frac{1}{\mathrm{N}} \mathrm{x}^{\mathrm{T}}(\hat{\mathrm{y}} - y)$$

$$= \frac{1}{4} \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 \\ 1.4 & 1.5 & 3.0 & 4.1 \\ 0.2 & 0.2 & 1.1 & 1.3 \end{bmatrix} \begin{bmatrix} 0.6856 \\ 0.6963 \\ -0.1839 \\ -0.1171 \end{bmatrix} = \begin{bmatrix} 0.2702 \\ 0.2431 \\ -0.019 \end{bmatrix} = \begin{bmatrix} L'_b \\ L'_{w_1} \\ L'_{w_2} \end{bmatrix}$$

Phân loại hoa Iris dựa vào chiều dài và chiều rộng của cánh hoa

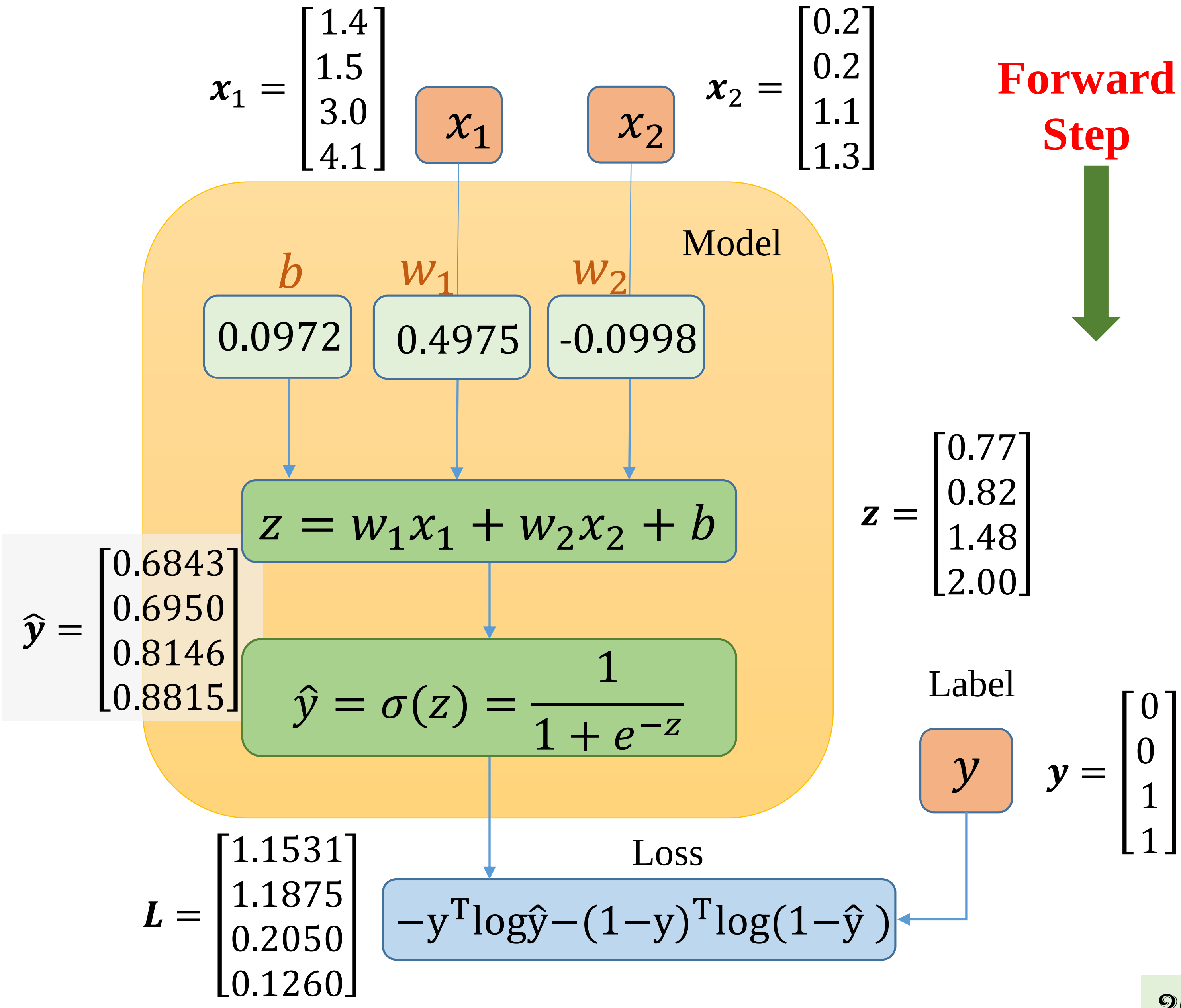
Dataset

Petal_Length	Petal_Width	Label
1.4	0.2	0
1.5	0.2	0
3	1.1	1
4.1	1.3	1

$$\boldsymbol{x} = \begin{bmatrix} 1 & 1.4 & 0.2 \\ 1 & 1.5 & 0.2 \\ 1 & 3.0 & 1.1 \\ 1 & 4.1 & 1.3 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Average loss = 0.6679

Loss giảm từ 0.6692 xuống 0.6679



Logistic Regression - Question

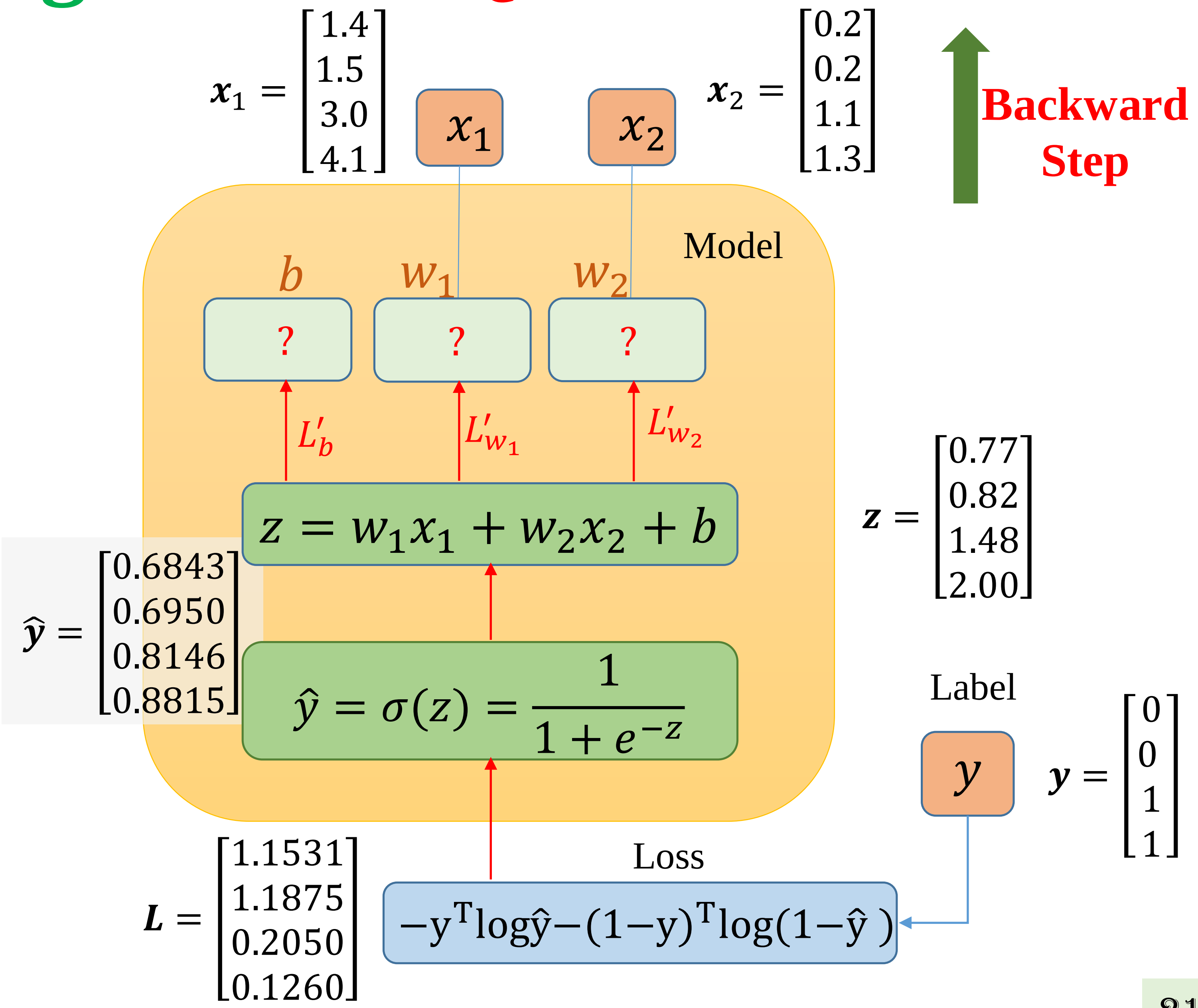
Phân loại hoa Iris dựa vào chiều dài và chiều rộng của cánh hoa

Dataset

Petal_Length	Petal_Width	Label
1.4	0.2	0
1.5	0.2	0
3	1.1	1
4.1	1.3	1

$$\boldsymbol{x} = \begin{bmatrix} 1 & 1.4 & 0.2 \\ 1 & 1.5 & 0.2 \\ 1 & 3.0 & 1.1 \\ 1 & 4.1 & 1.3 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{x}^{\mathrm{T}} = \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 \\ 1.4 & 1.5 & 3.0 & 4.1 \\ 0.2 & 0.2 & 1.1 & 1.3 \end{bmatrix}$$

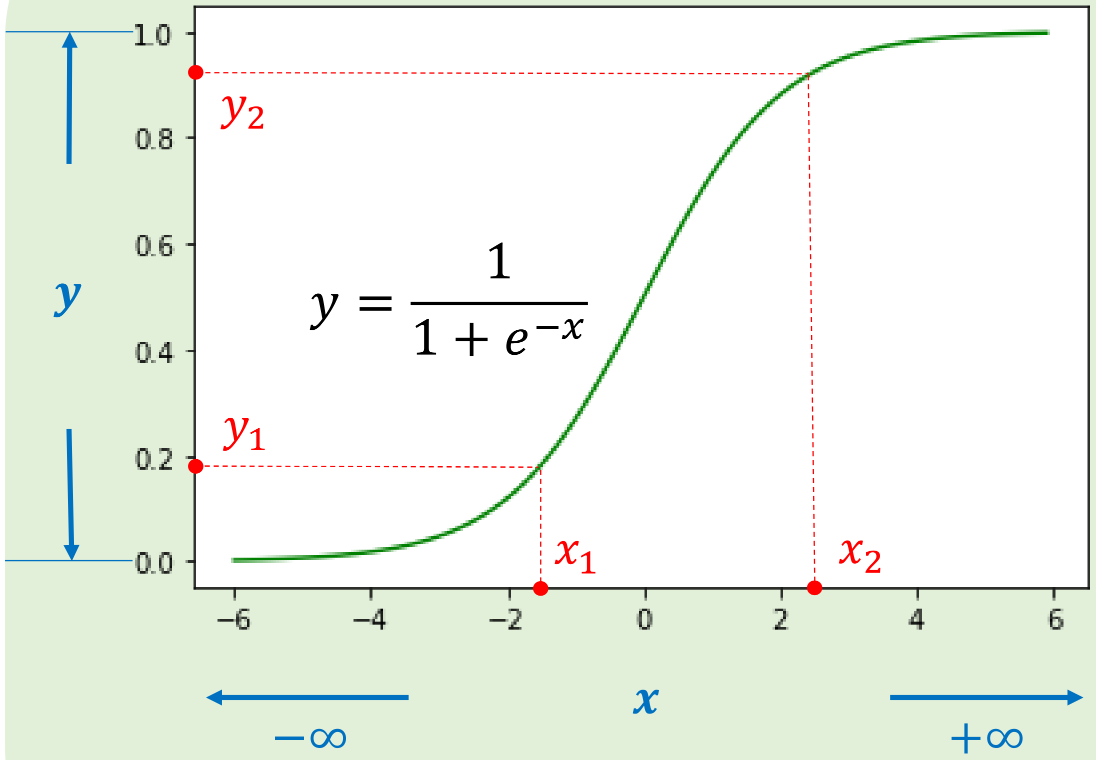


Logistic Regression

❖ Demo

```
Python 3.7.3 (default, Apr 24 2019, 15:29:51) [MSC v.1915 64 bit (AMD64)] ::
Type "help", "copyright", "credits" or "license" for more information.
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>> for epoch in range(n_epochs):
...     sum_of_losses = 0
...     gradients = np.zeros((2,1))
...
...     for index in range(4):
...         xi = X_b[index:index+1]
...         yi = y[index:index+1]
```

Summary



Sigmoid function

1) Pick all the samples from training data

2) Tính output $\hat{y}$

$$z = \boldsymbol{\theta}^T \boldsymbol{x}$$

$$\hat{\boldsymbol{y}} = \sigma(z) = \frac{1}{1+e^{-z}}$$

3) Tính loss (binary cross-entropy)

$$L(\boldsymbol{\theta}) = \frac{1}{\mathrm{N}}\left(-\mathrm{y}^{\mathrm{T}}\log\hat{\mathrm{y}} - (1-\mathrm{y})^{\mathrm{T}}\log(1-\hat{\mathrm{y}})\right)$$

4) Tính đạo hàm

$$L'_{\boldsymbol{\theta}} = \frac{1}{\mathrm{N}}\mathrm{x}^{\mathrm{T}}(\hat{\mathrm{y}} - y)$$

5) Cập nhật tham số

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta L'_{\boldsymbol{\theta}}$$

η is learning rate

Thank you!