

Logistic Regression

Quang-Vinh Dinh
Ph.D. in Computer Science

Outline

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

Sigmoid Function

Sigmoid function

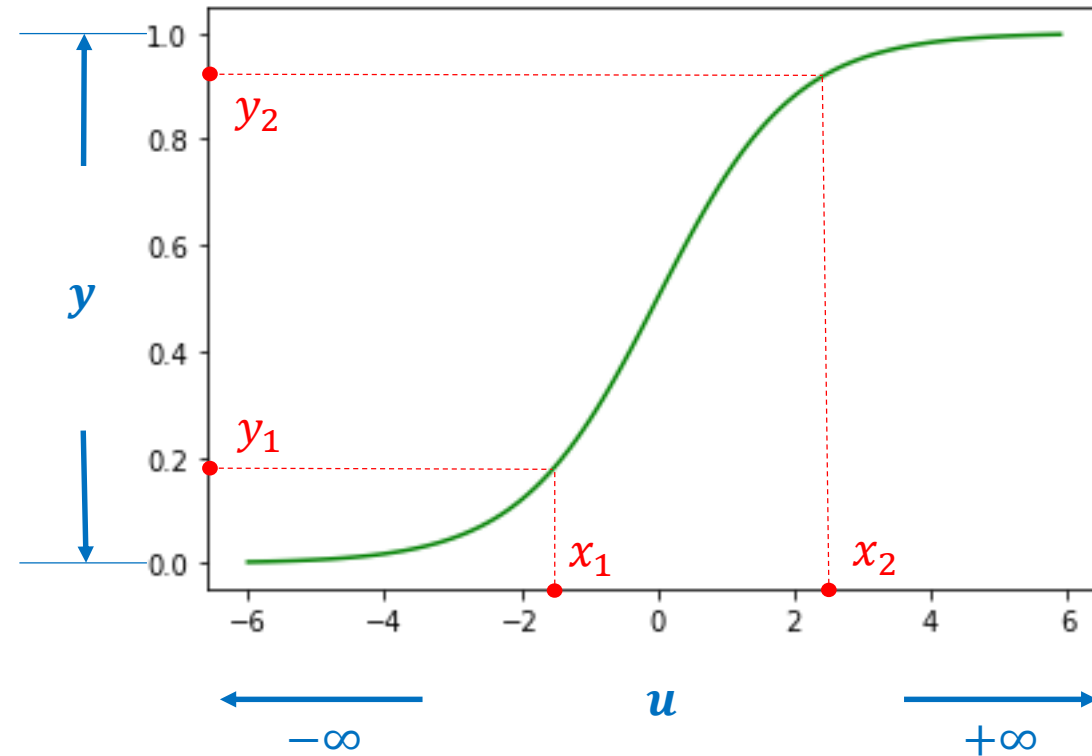
$$y = \sigma(u) = \frac{1}{1 + e^{-u}}$$

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

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

Property

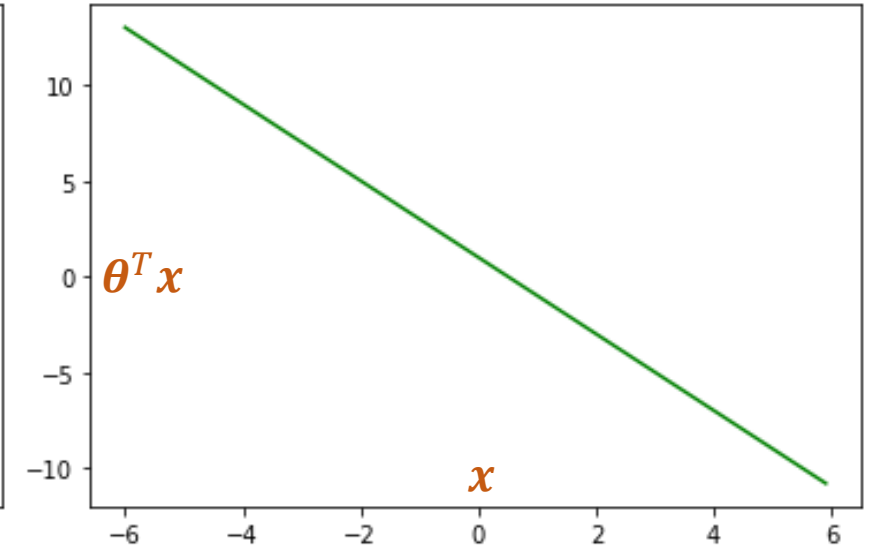
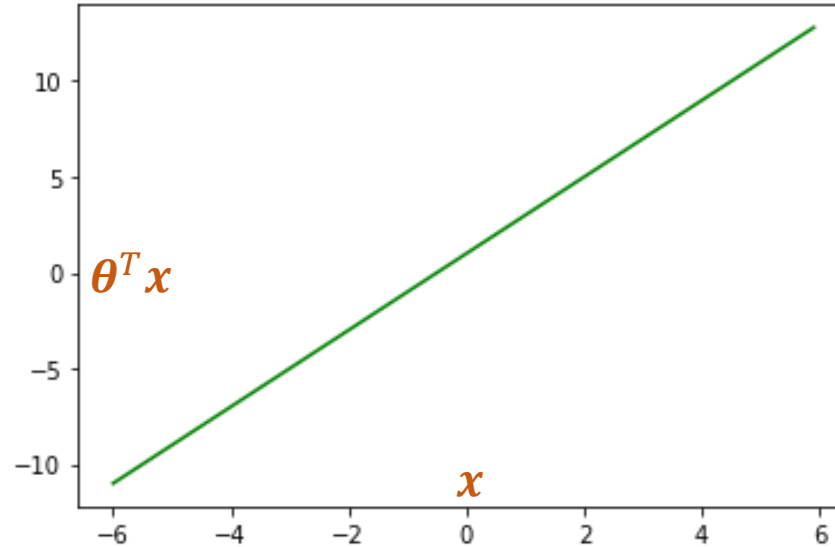
$$\forall u_1 u_2 \in [a \quad b] \text{ và } u_1 \leq u_2 \\ \rightarrow \sigma(u_1) \leq \sigma(u_2)$$



Sigmoid Function

$$y = \theta^T x$$

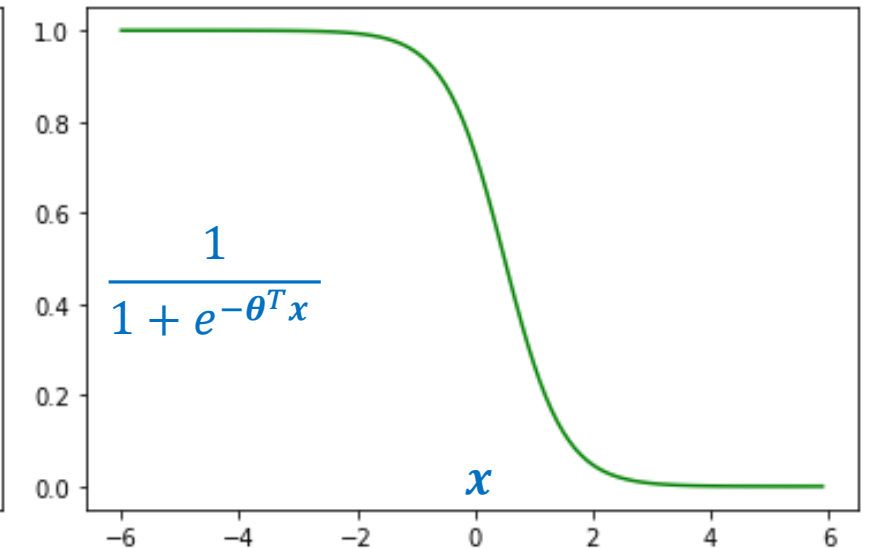
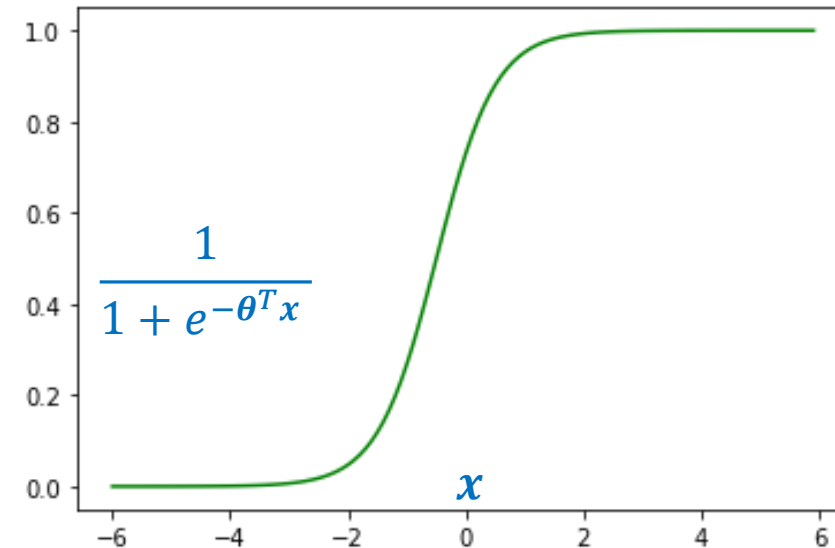
$$y \in (-\infty + \infty)$$



$$z = \theta^T x$$

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

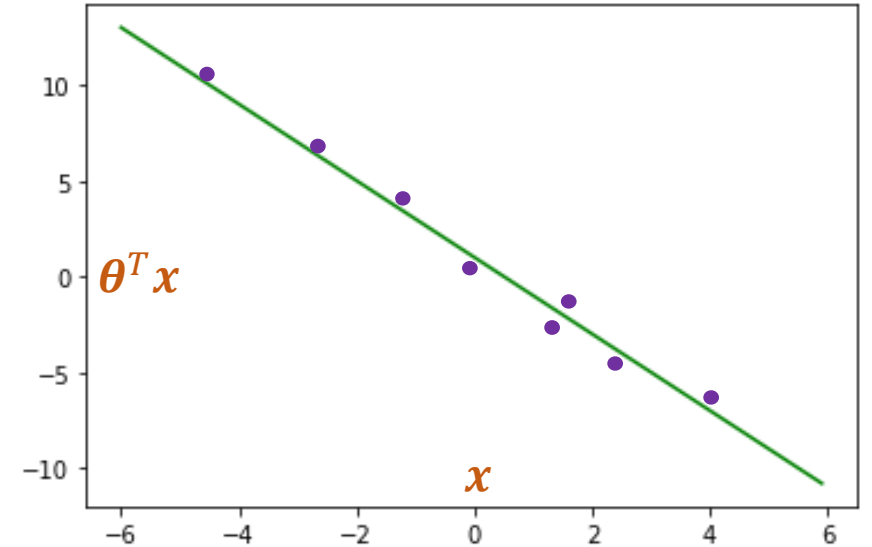
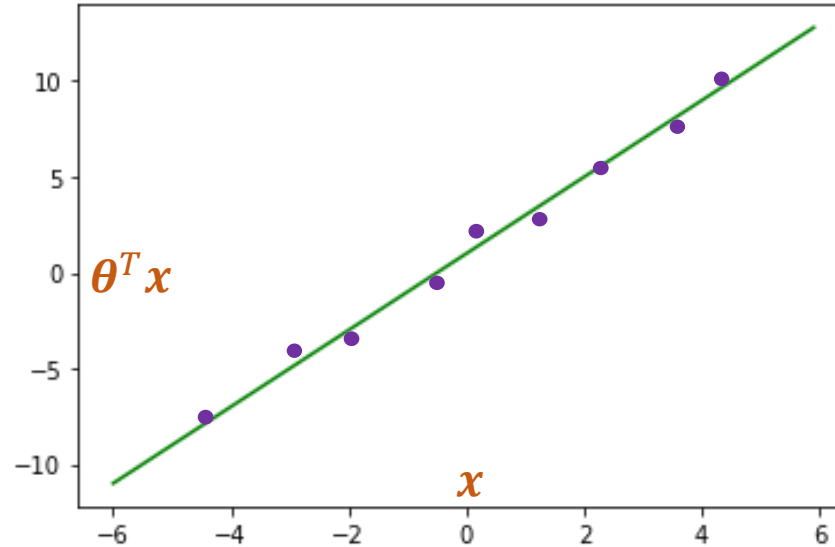
$$y \in (0 \ 1)$$



Sigmoid Function

$$y = \theta^T x$$

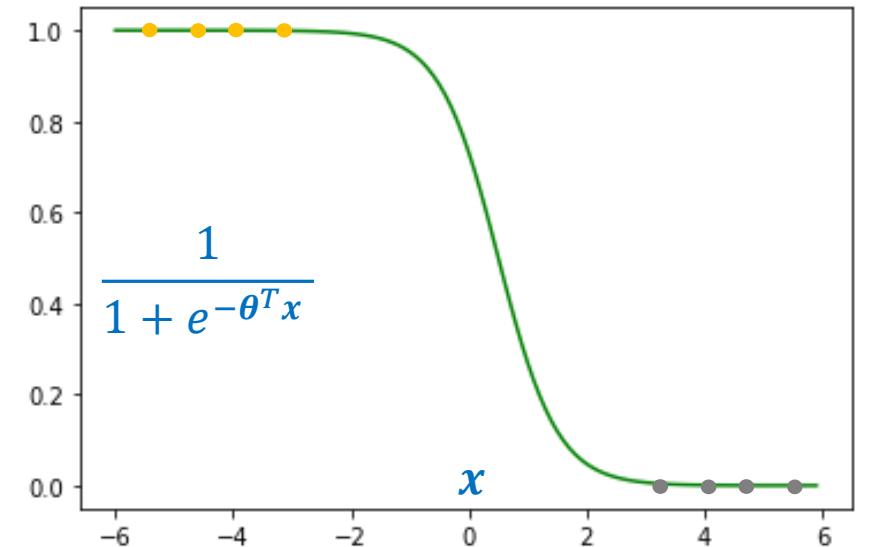
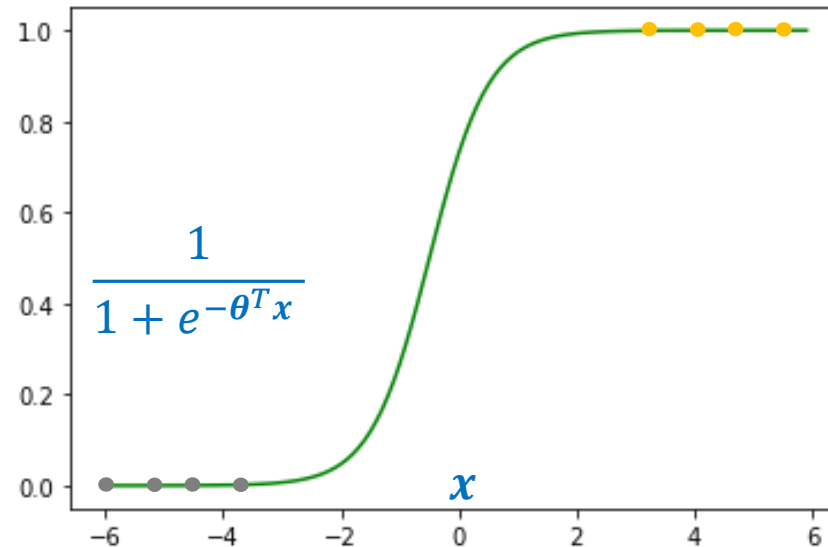
$$y \in (-\infty + \infty)$$



$$z = \theta^T x$$

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

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



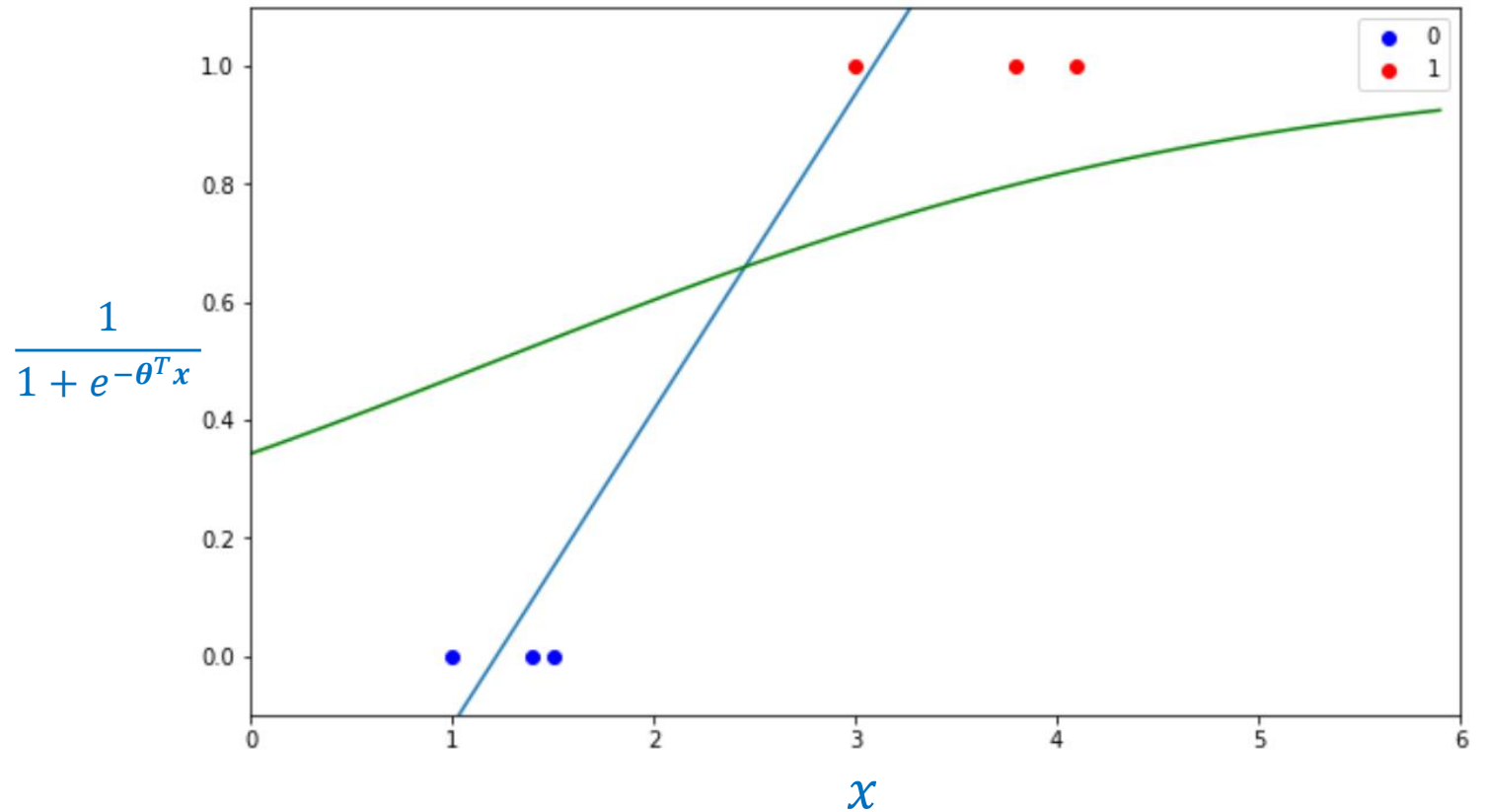
Sigmoid Function

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	

$$z = \theta^T x$$

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

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



$$z = 0.535 * x - 0.654$$

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

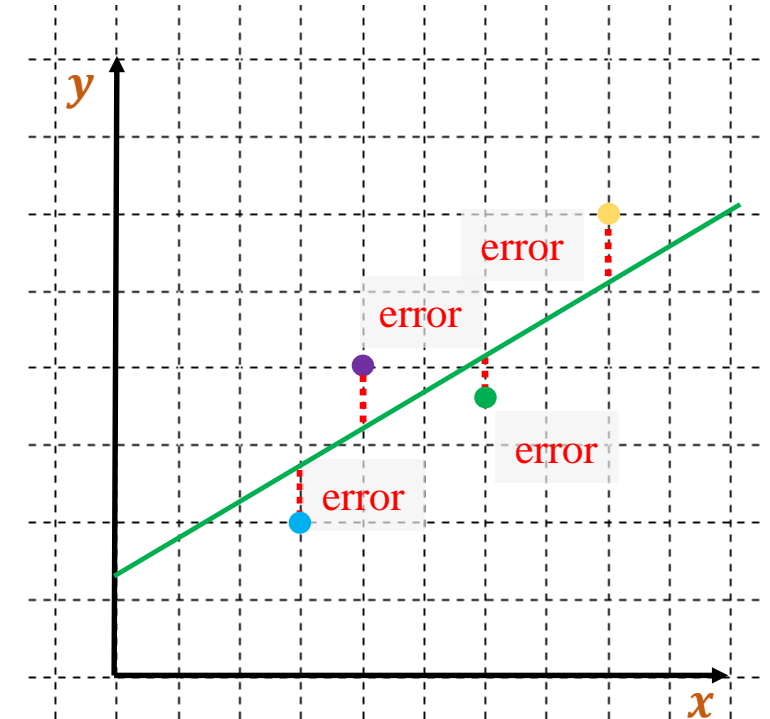
Feature	Label
area	price
6.7	9.1
4.6	5.9
3.5	4.6
5.5	6.7

Training data

construct

$$y = \theta^T x = ax + b$$
$$y \in (-\infty + \infty)$$

Model



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

Idea of Logistic Regression

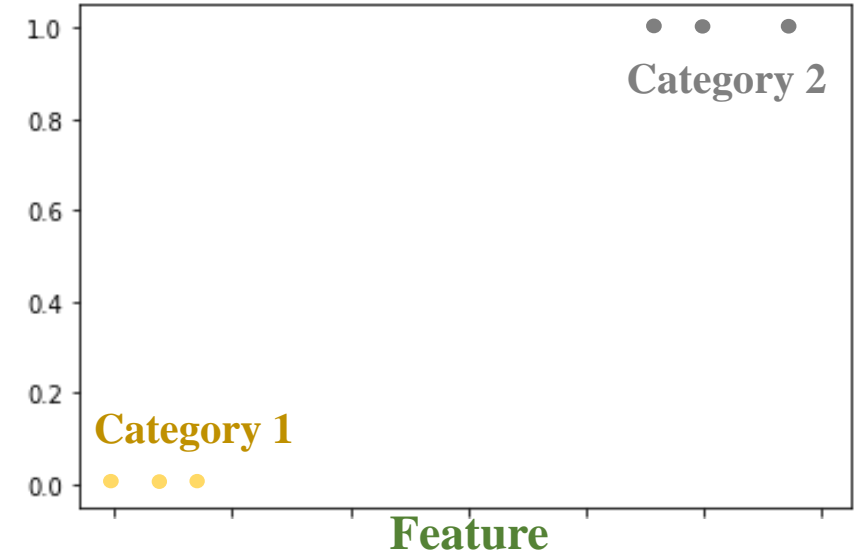
❖ Given a new kind of data

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

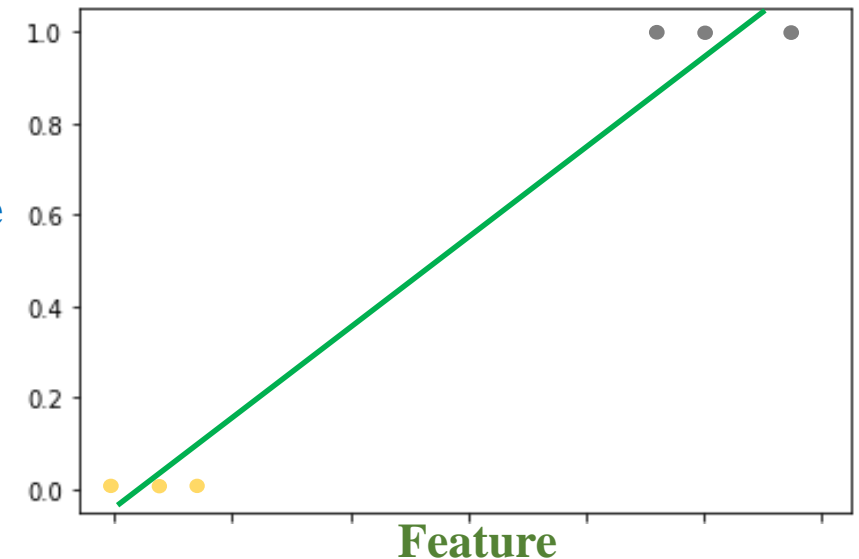
↓ Assign numbers to categories

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	

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	
1.5	Flower A	
3	Flower B	Category 2
3.8	Flower B	
4.1	Flower B	

Assign numbers
to categories

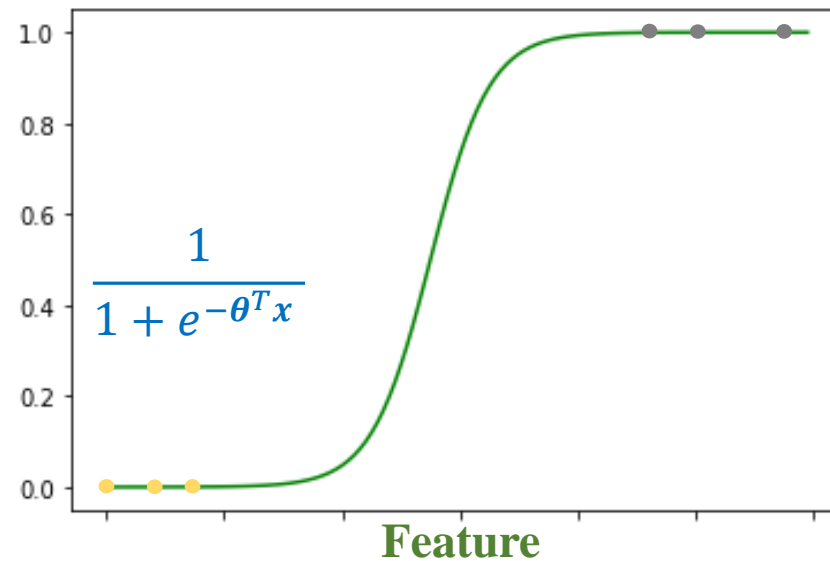
Feature	Label	
Petal_Length	Category	
1.4	0	Category 1
1	0	
1.5	0	
3	1	Category 2
3.8	1	
4.1	1	

Sigmoid function
could fit the data

$$z = \theta^T x$$

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

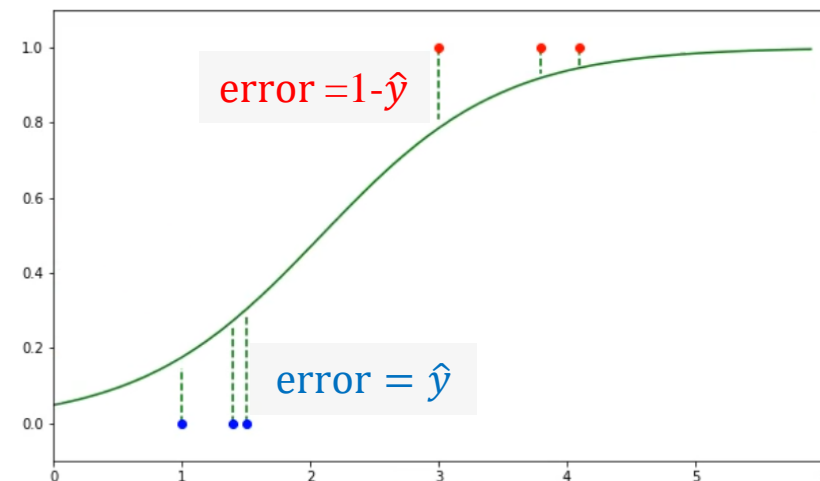
$$\hat{y} \in (0 \ 1)$$



Error

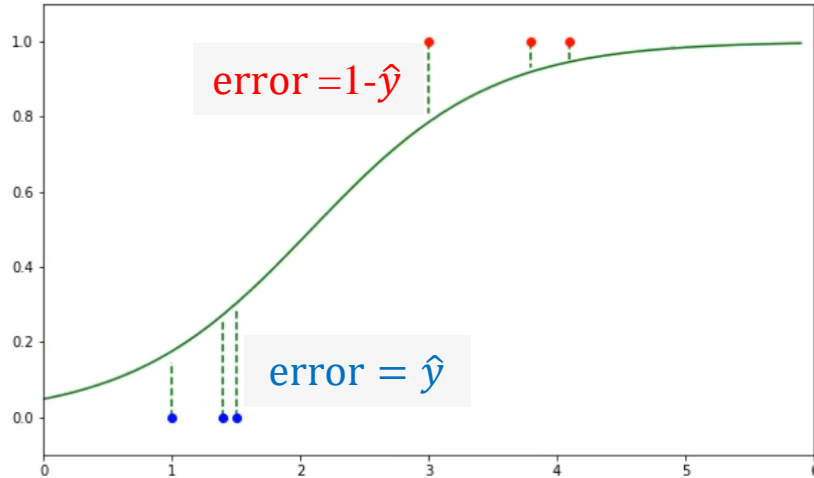
$$\text{if } y = 1 \\ \text{error} = 1 - \hat{y}$$

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



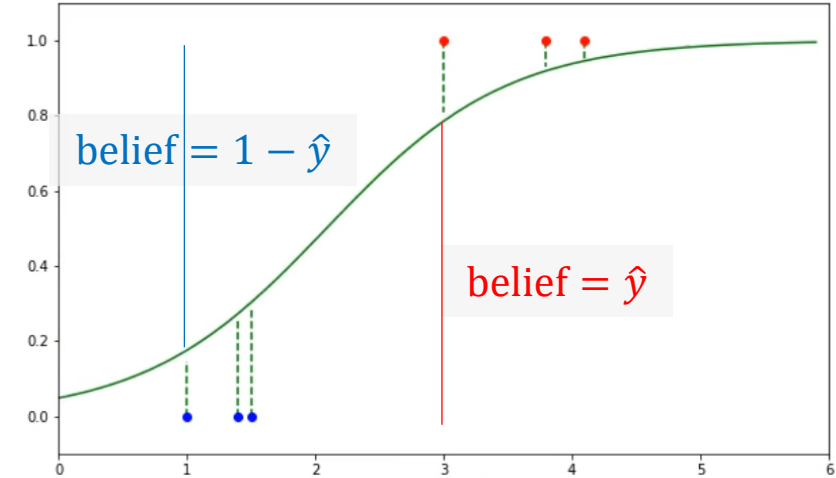
Idea of Logistic Regression

❖ Construct loss



Error

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



Belief

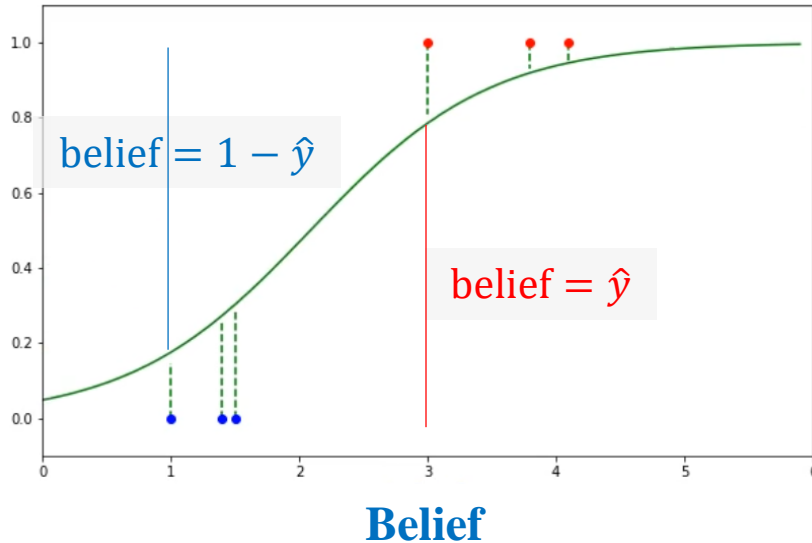
if $y = 1$
belief = \hat{y}
if $y = 0$
belief = $1 - \hat{y}$

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

Minimize error ~ maximize belief ~ Minimize (-belief)

Idea of Logistic Regression

❖ Construct loss



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

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

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

$$\begin{aligned} \text{belief} &= \prod_{i=1}^n P_i && \text{since iid} \\ \log_belief &= \sum_{i=1}^n \log P_i \\ \log_belief &= \sum_{i=1}^n [y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)] \\ \text{loss} &= -\log_belief \\ &= -\sum_{i=1}^n [y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)] \end{aligned}$$

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

Binary cross-entropy

Idea of Logistic Regression

❖ Construct loss

$$\begin{aligned} z &= \theta^T x \\ \hat{y} &= \sigma(z) = \frac{1}{1 + e^{-z}} \\ L &= \frac{1}{N} [-y^T \log(\hat{y}) - (1 - y^T) \log(1 - \hat{y})] \end{aligned}$$

Model and Loss

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

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

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

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

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

Idea of Logistic Regression

❖ Construct loss

Model and Loss

$$z = \theta^T x$$
$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$
$$L = \frac{1}{N} [-y^T \log(\hat{y}) - (1 - y^T) \log(1 - \hat{y})]$$

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

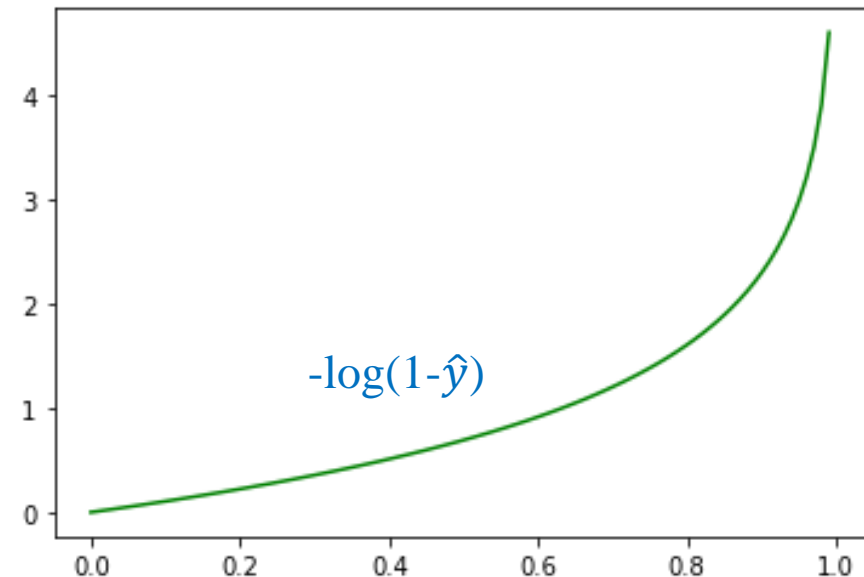
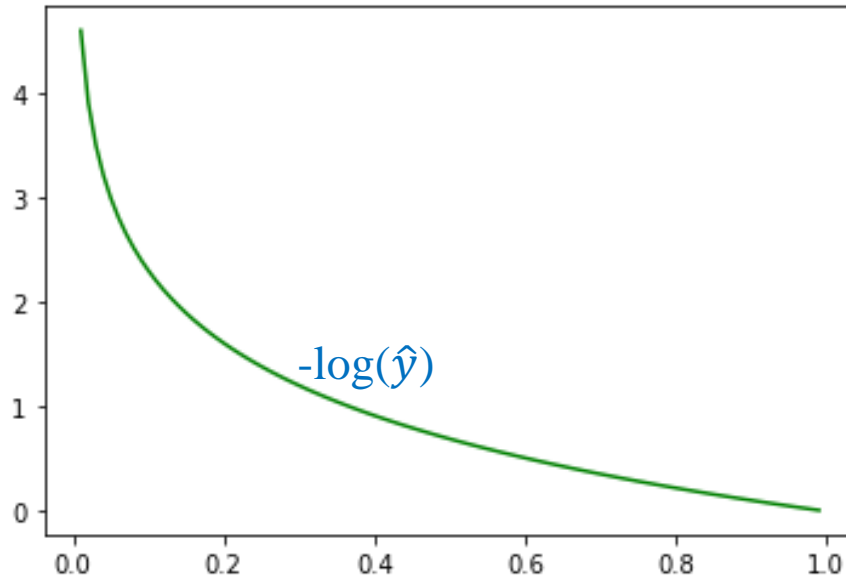
Derivative

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

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

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

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



Idea of Logistic Regression

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

Category 1

Category 2

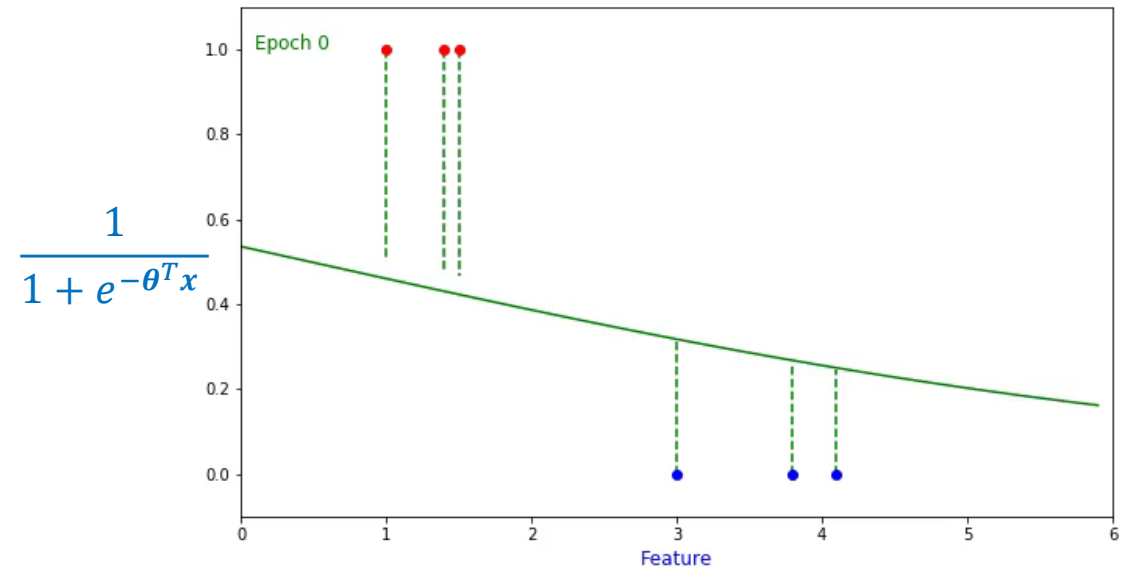
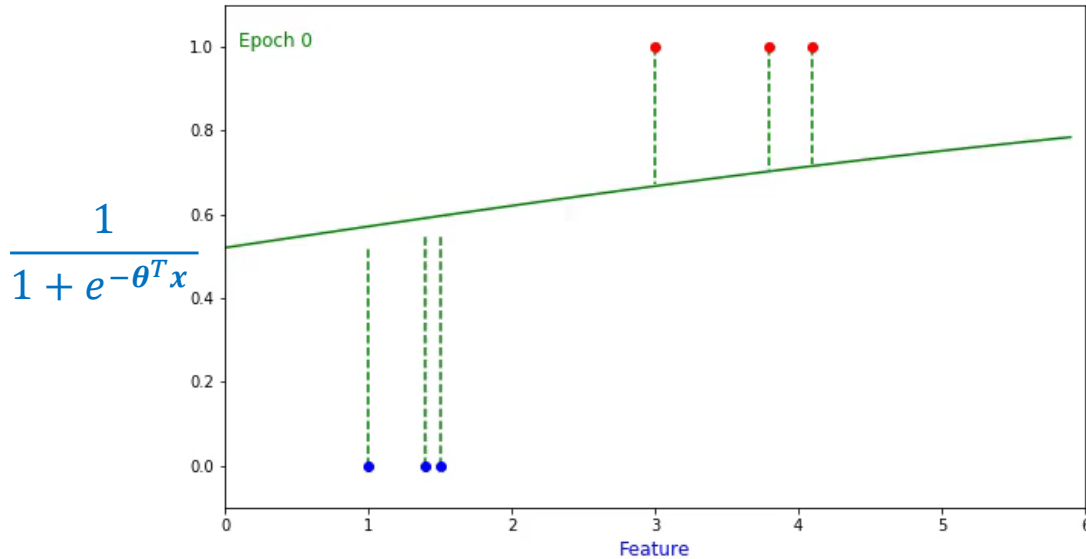
$$z = \theta^T x$$

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

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

Category 1

Category 2



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}

$$z = \theta^T x$$

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

3) Tính loss

$$L(\theta) = (-y \log \hat{y} - (1-y) \log(1-\hat{y}))$$

4) Tính đạo hàm

$$L'_{\theta} = x(\hat{y} - y)$$

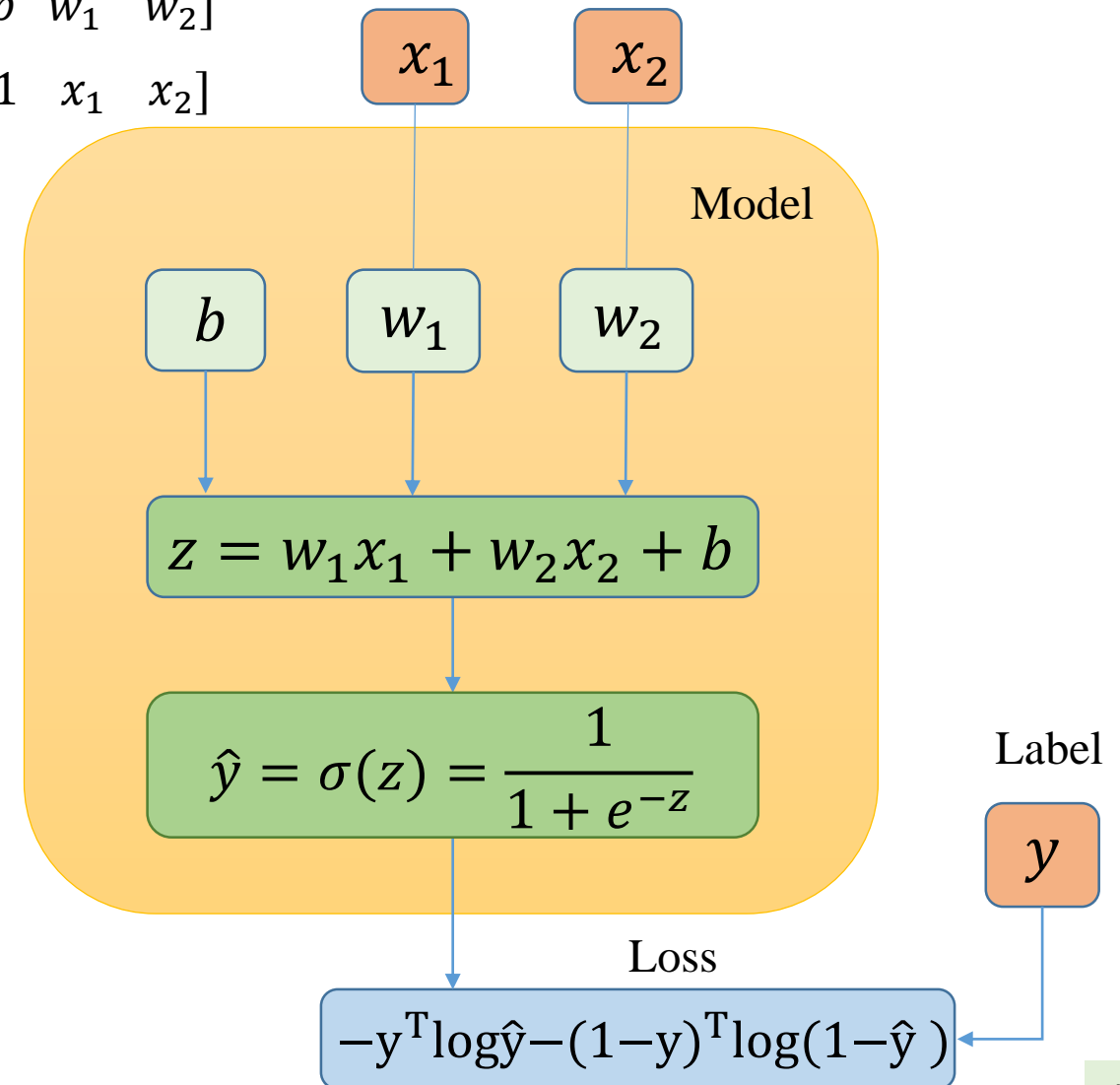
5) Cập nhật tham số

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

η is learning rate

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

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



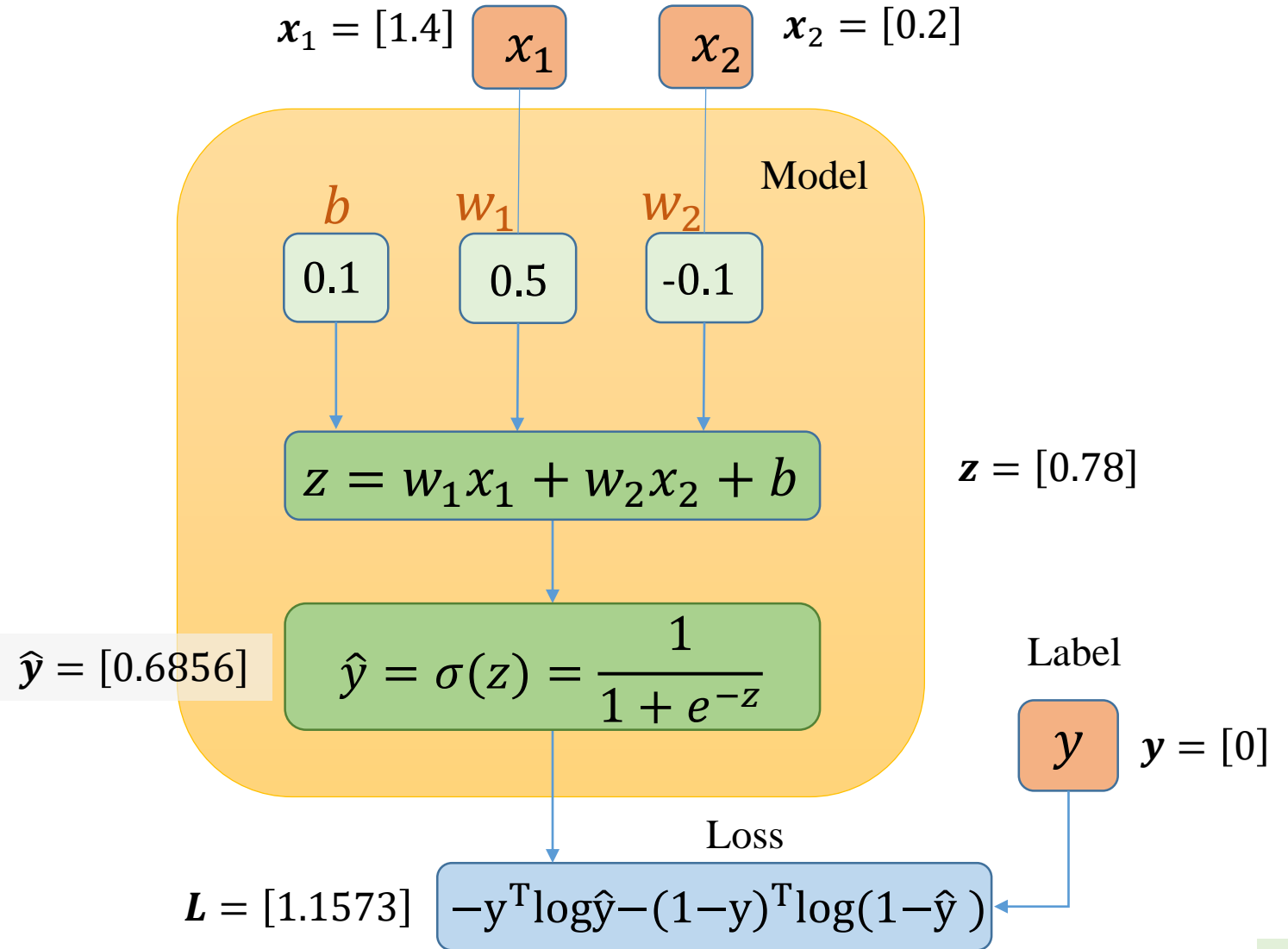
Logistic Regression-Stochastic

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

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1.4 \\ 0.2 \end{bmatrix}$$

$$\mathbf{y} = [0]$$



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

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

$$\eta = 0.01$$

$$b = 0.1 - \eta 0.6856 = 0.0931$$

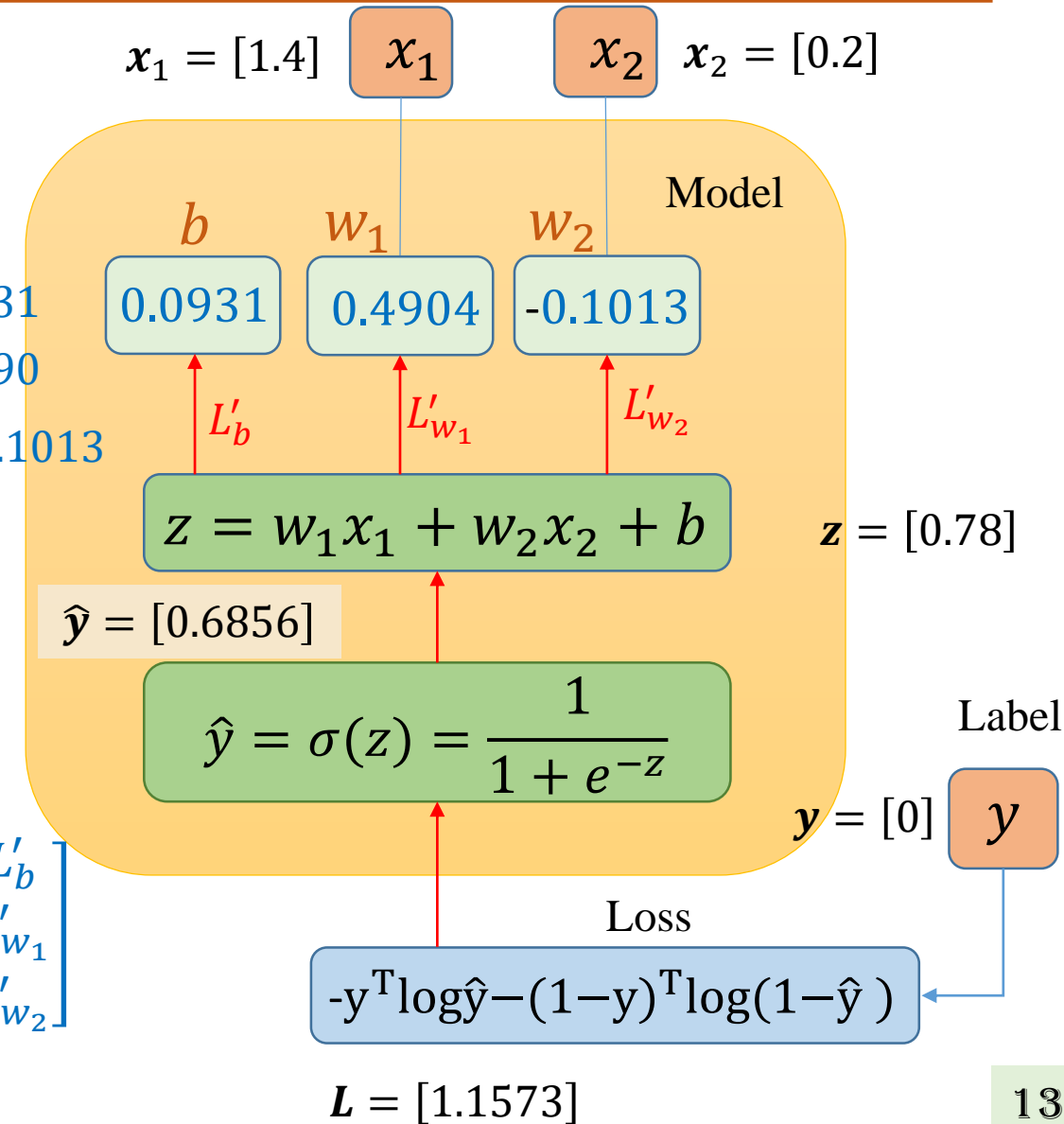
$$w_1 = 0.5 - \eta 0.9598 = 0.4904$$

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

$$L'_{\theta} = \mathbf{x}(\hat{\mathbf{y}} - \mathbf{y})$$

$$= \begin{bmatrix} 1 \\ 1.4 \\ 0.2 \end{bmatrix} [0.6856]$$

$$= \begin{bmatrix} 0.6856 \\ 0.9599 \\ -0.1371 \end{bmatrix} = \begin{bmatrix} L'_b \\ L'_{w_1} \\ L'_{w_2} \end{bmatrix}$$



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}

$$z = \theta^T x$$

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

3) Tính loss

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

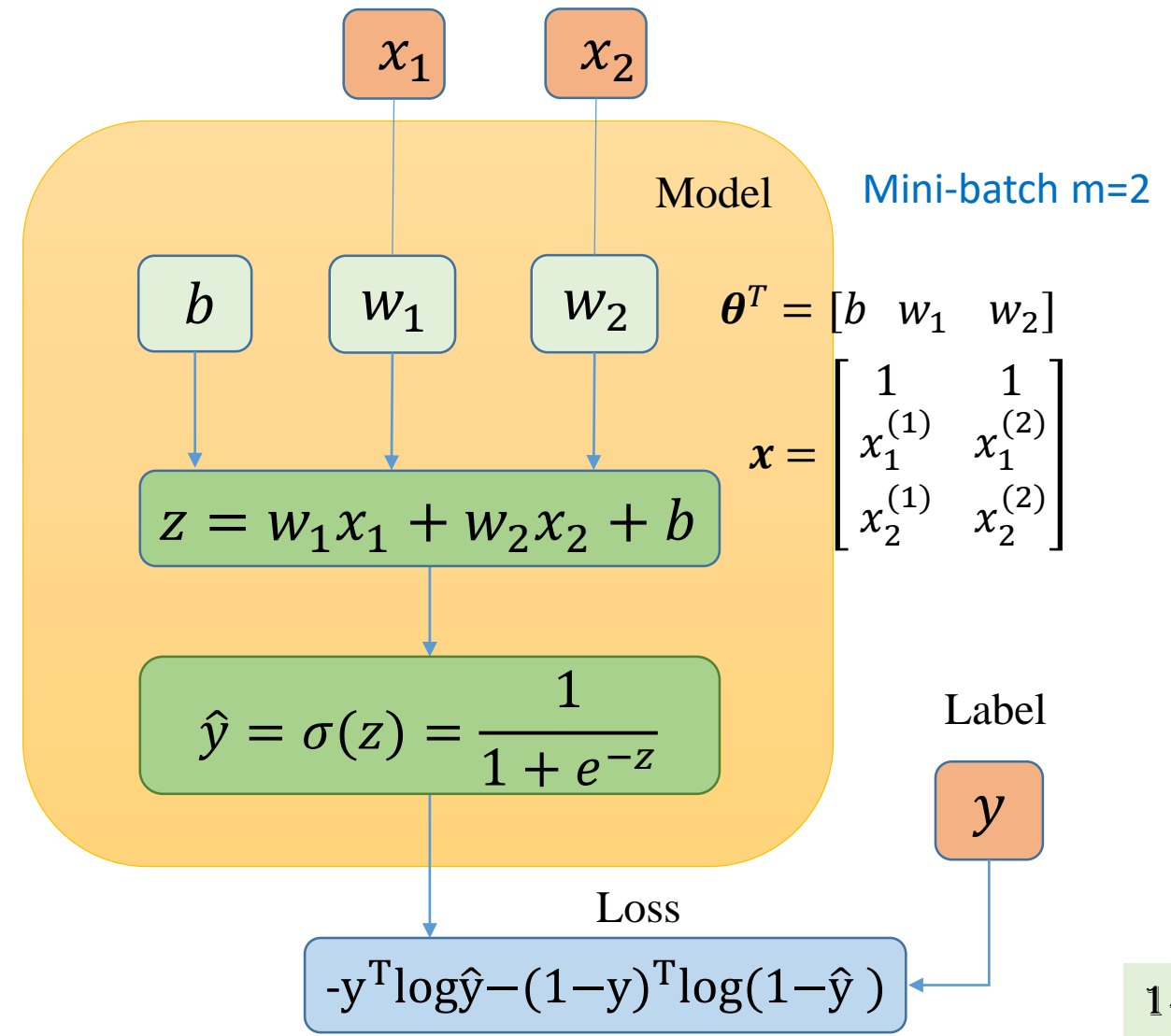
4) Tính đạo hàm

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

5) Cập nhật tham số

$$\theta = \theta - \eta L'_\theta$$

η is learning rate

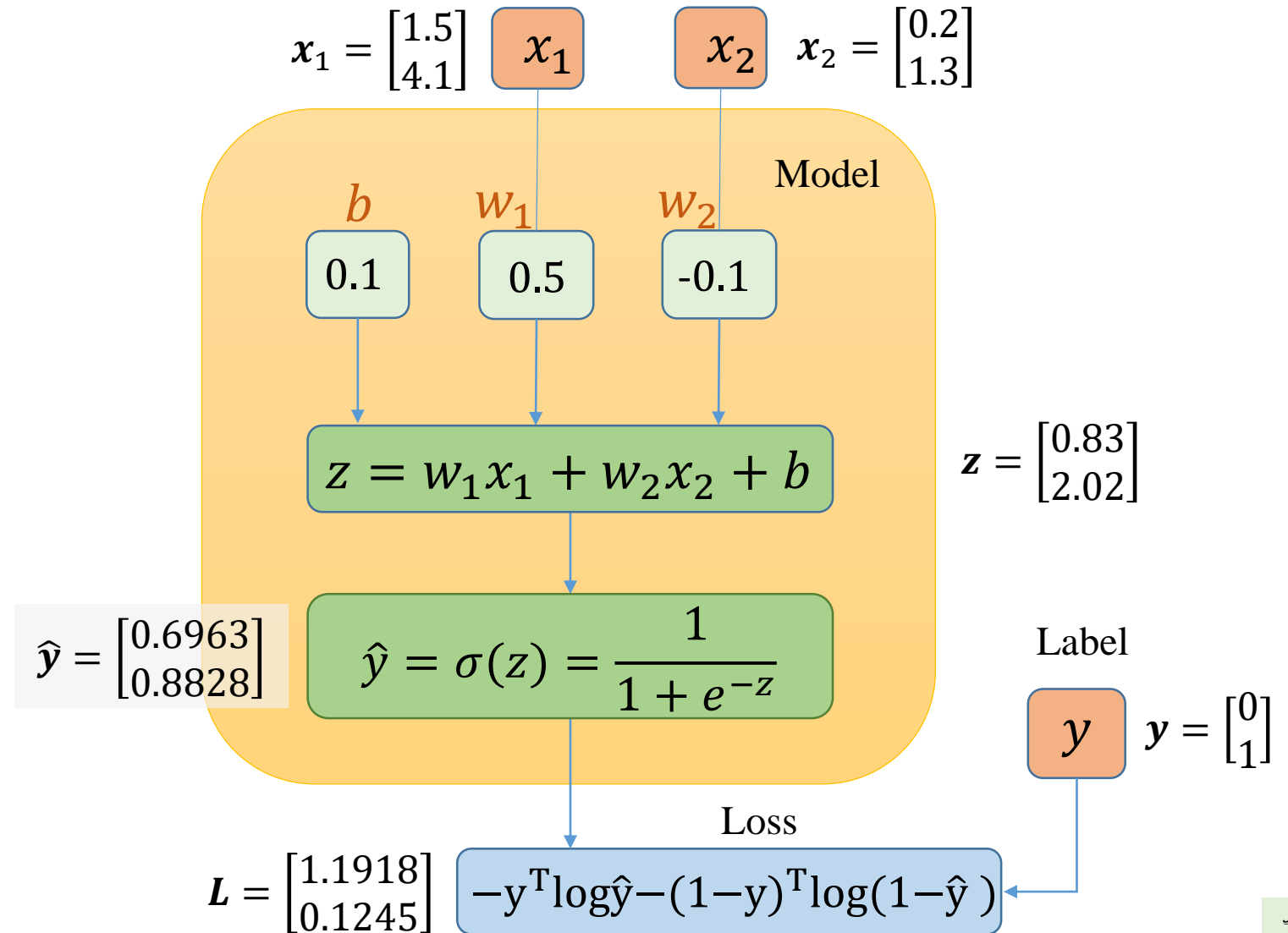


Logistic Regression - Minibatch

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

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



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

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

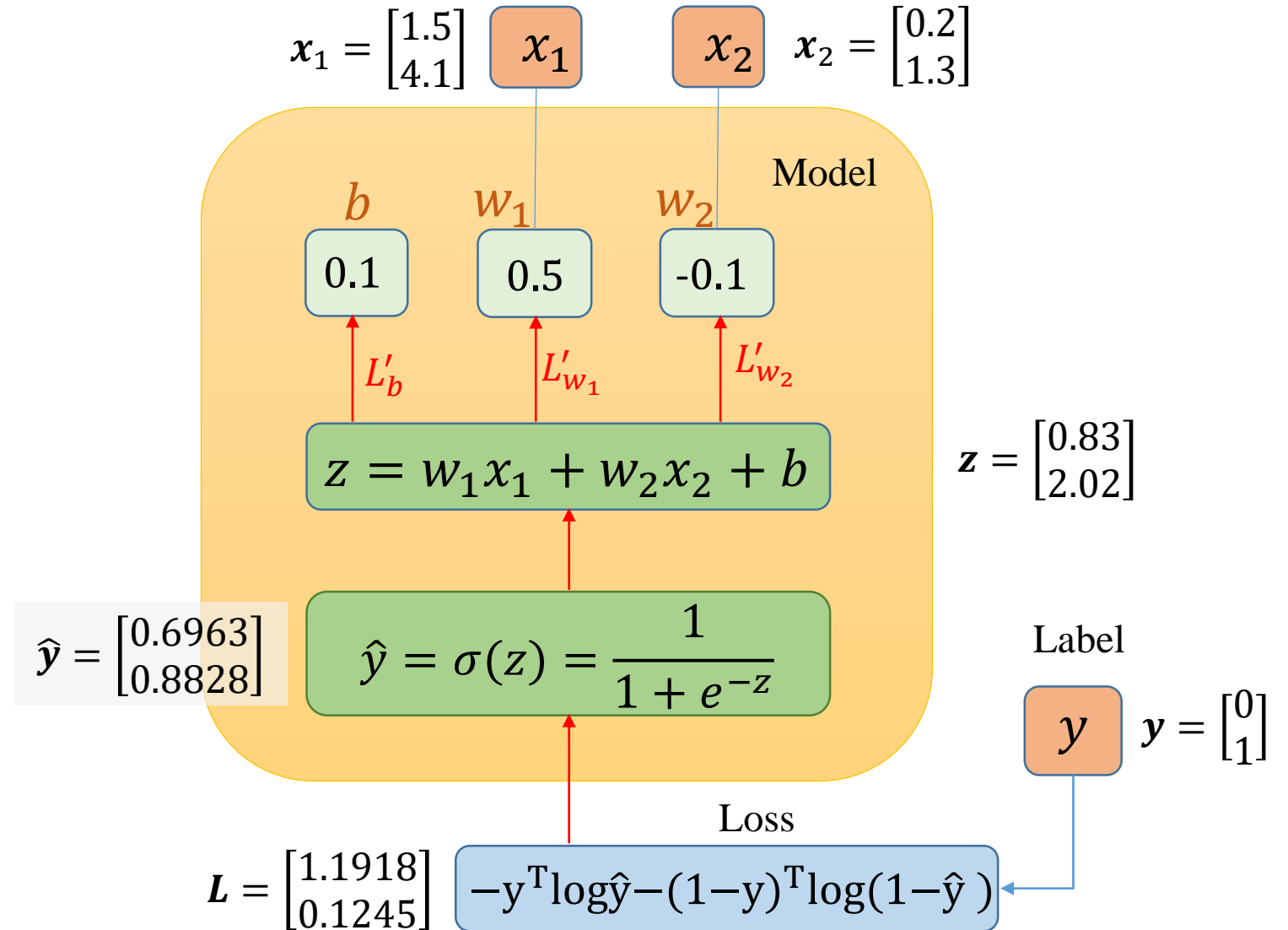
$$\begin{aligned} L'_{\theta} &= \frac{1}{m} \mathbf{x}^T (\hat{\mathbf{y}} - \mathbf{y}) \\ &= \frac{1}{2} \begin{bmatrix} 1.0 & 1.0 \\ 1.5 & 4.1 \\ 0.2 & 1.3 \end{bmatrix} \begin{bmatrix} 0.6963 \\ -0.1171 \end{bmatrix} \\ &= \begin{bmatrix} 0.28961 \\ 0.28217 \\ -0.0064 \end{bmatrix} = \begin{bmatrix} L'_b \\ L'_{w_1} \\ L'_{w_2} \end{bmatrix} \end{aligned}$$

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

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

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

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}

$$z = \theta^T x$$

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

3) Tính loss (binary cross-entropy)

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

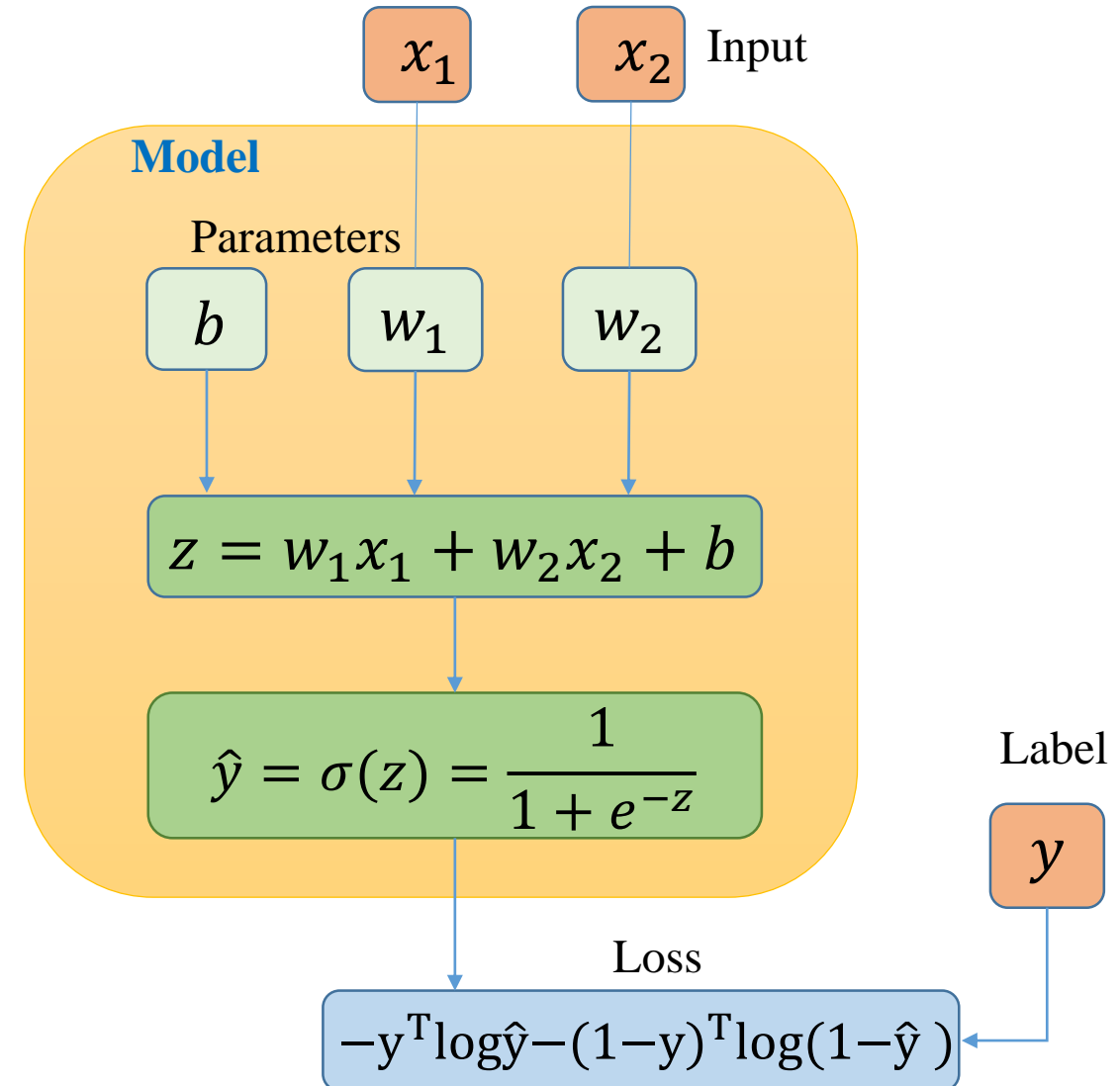
4) Tính đạo hàm

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

5) Cập nhật tham số

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

η is learning rate



Logistic Regression

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

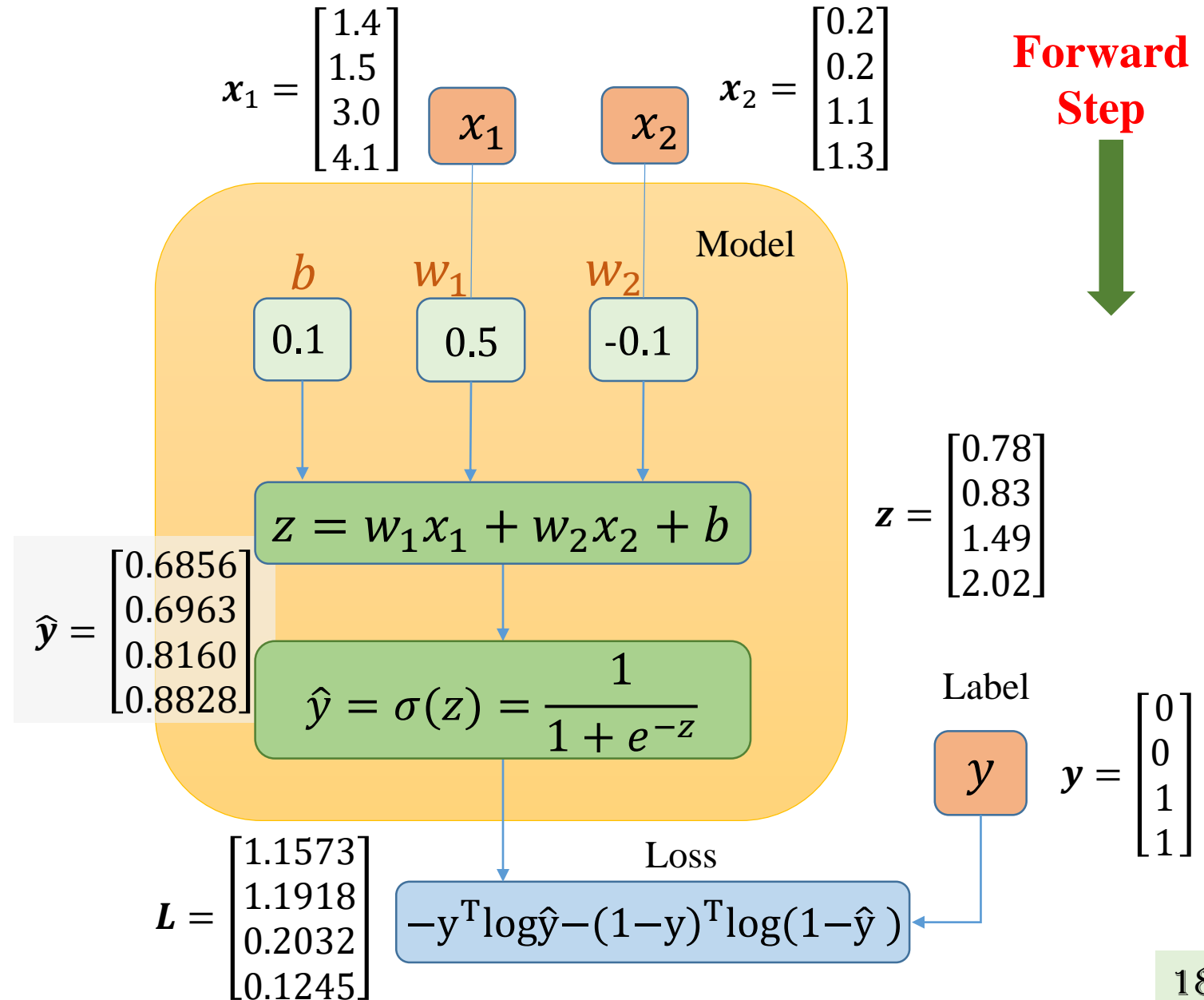
Dataset

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

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

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

Average loss = 0.6692



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

Dataset

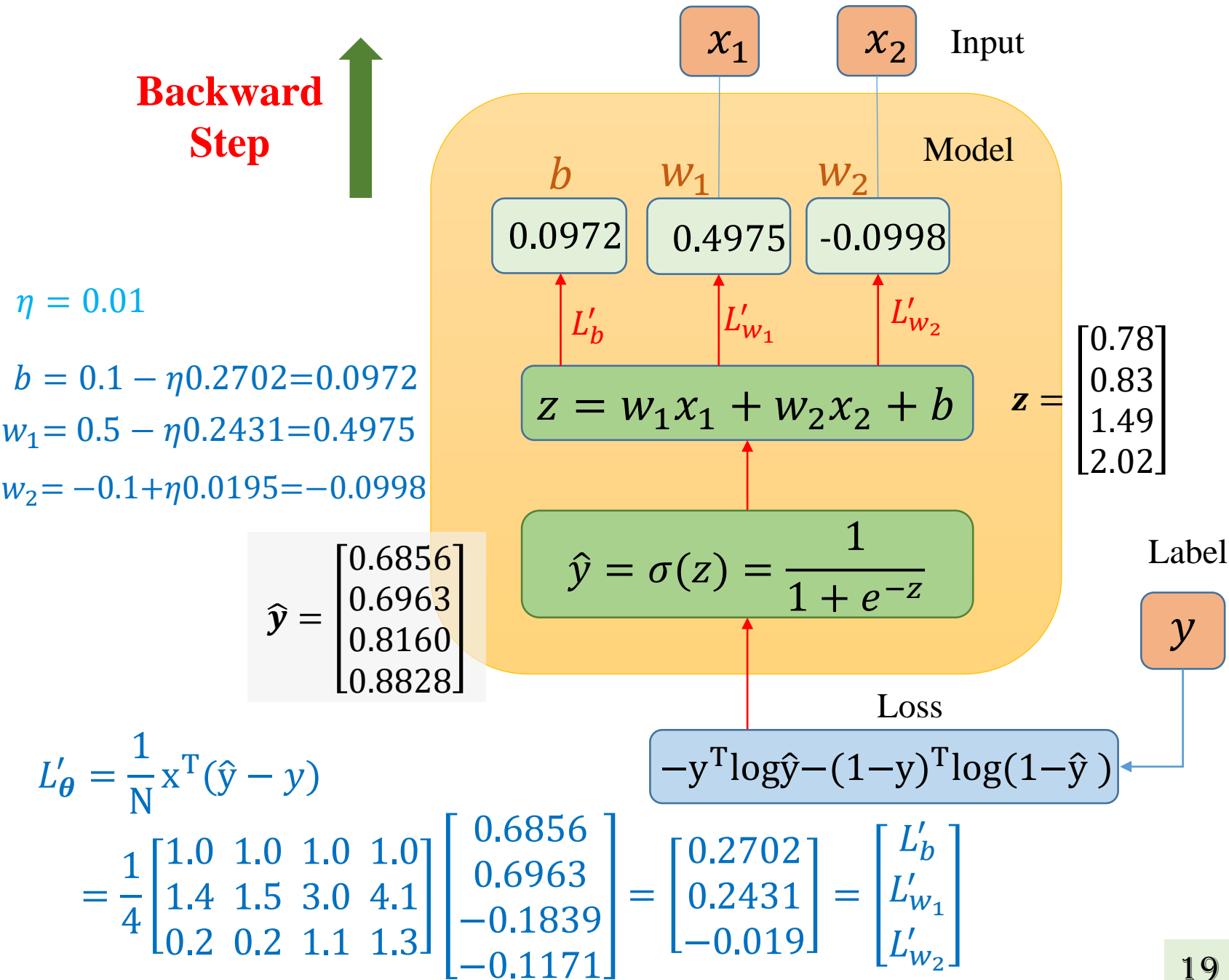
Petal_Length	Petal_Width	Label
1.4	0.2	0
1.5	0.2	0
3	1.1	1
4.1	1.3	1

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

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

$$\begin{aligned} \eta &= 0.01 \\ b &= 0.1 - \eta 0.2702 = 0.0972 \\ w_1 &= 0.5 - \eta 0.2431 = 0.4975 \\ w_2 &= -0.1 + \eta 0.0195 = -0.0998 \end{aligned}$$

$$\begin{aligned} L'_\theta &= \frac{1}{N} \mathbf{x}^T (\hat{\mathbf{y}} - \mathbf{y}) \\ &= \frac{1}{4} \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 \\ 1.4 & 1.5 & 3.0 & 4.1 \\ 0.2 & 0.2 & 1.1 & 1.3 \end{bmatrix} \begin{bmatrix} 0.6856 \\ 0.6963 \\ -0.1839 \\ -0.1171 \end{bmatrix} = \begin{bmatrix} 0.2702 \\ 0.2431 \\ -0.019 \end{bmatrix} = \begin{bmatrix} L'_b \\ L'_{w_1} \\ L'_{w_2} \end{bmatrix} \end{aligned}$$



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

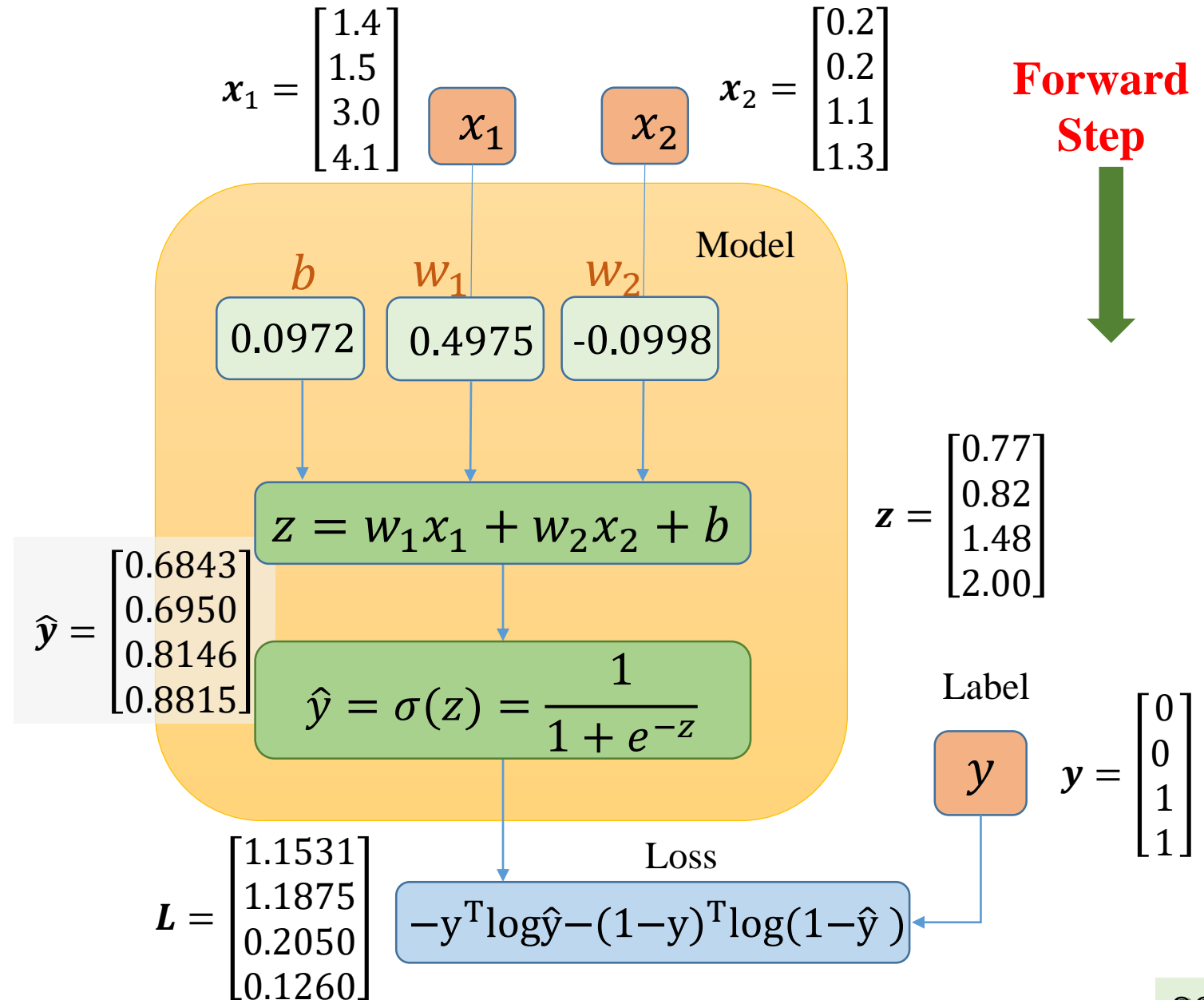
Dataset

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

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

Average loss = 0.6679

Loss giảm từ 0.6692 xuống 0.6679



Logistic Regression - Question

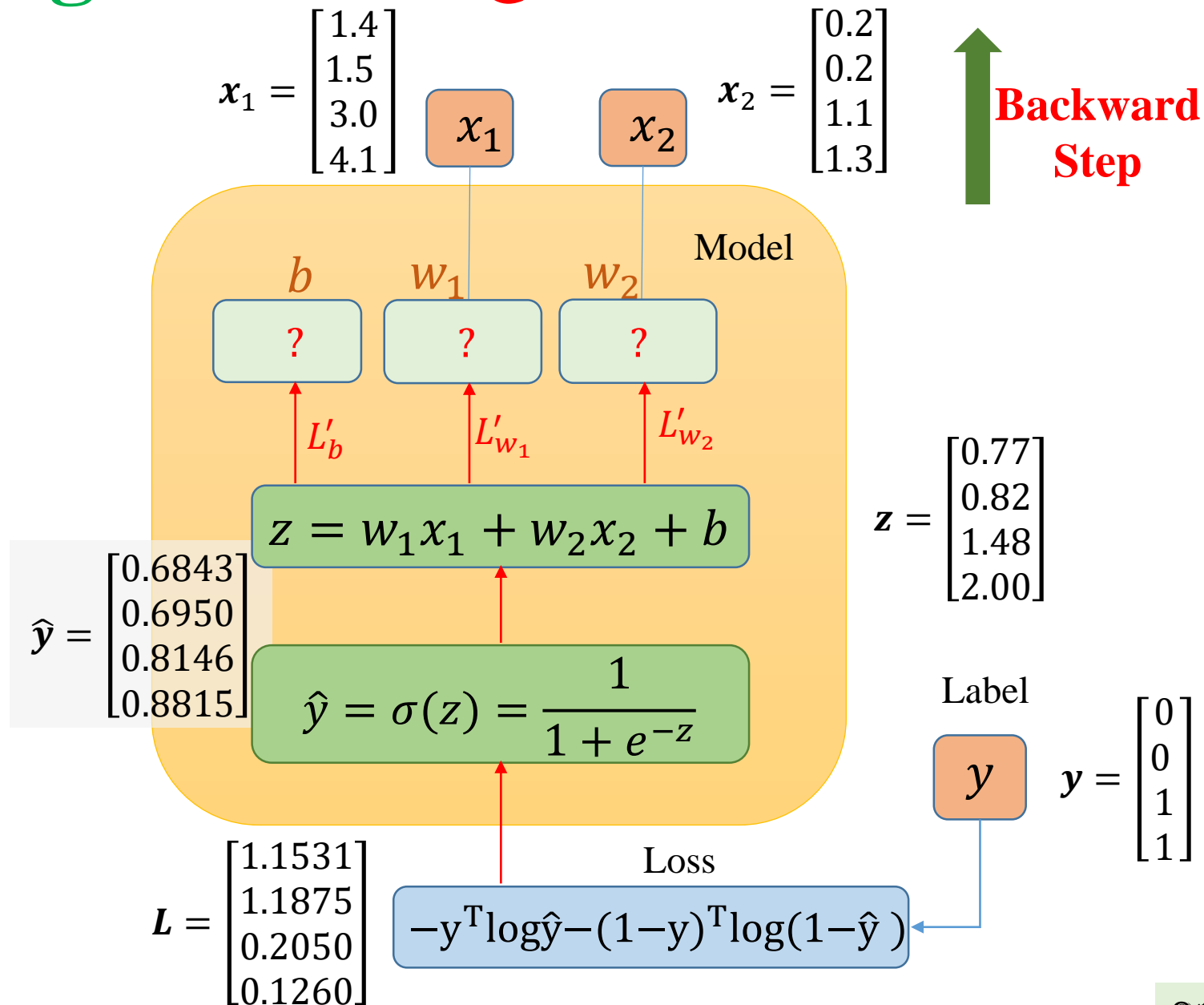
Phân loại hoa Iris dựa vào chiều dài và chiều rộng của cánh hoa

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

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

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



Logistic Regression

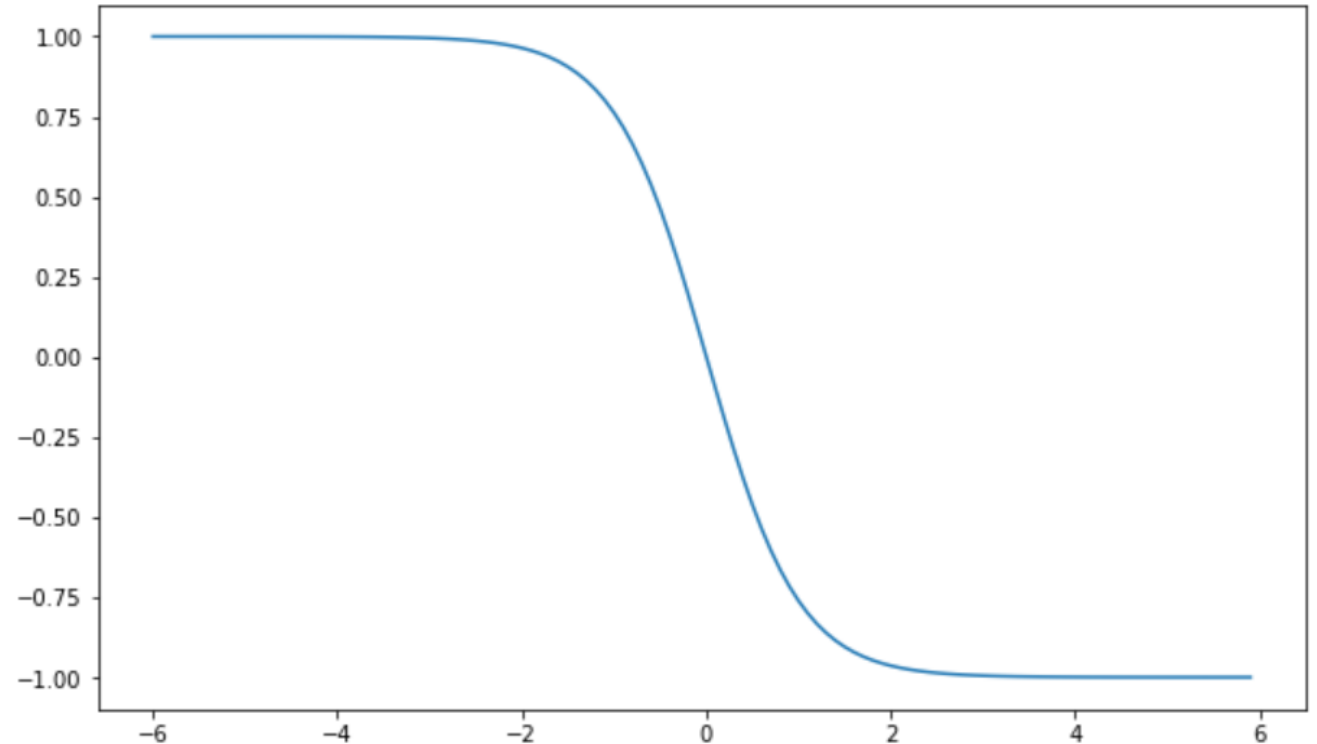
❖ Demo

```
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]
```

Tanh function

$$\begin{aligned}\tanh(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1} \\ &= 1 - \frac{2}{e^{2x} + 1}\end{aligned}$$

$$\begin{aligned}\tanh(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{-2x}} \\ &= -\frac{e^{-2x} - 1}{e^{-2x} + 1} = \frac{2}{e^{-2x} + 1} - 1\end{aligned}$$



Tanh function

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1} = 1 - \frac{2}{e^{2x} + 1}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{-2x}} = -\frac{e^{-2x} - 1}{e^{-2x} + 1} = \frac{2}{e^{-2x} + 1} - 1$$

$$\begin{aligned} \tanh'(x) &= \left(\frac{e^x - e^{-x}}{e^x + e^{-x}} \right)' = \frac{(e^x + e^{-x})(e^x + e^{-x}) - (e^x - e^{-x})(e^x - e^{-x})}{(e^x + e^{-x})^2} \\ &= \frac{(e^x + e^{-x})^2 - (e^x - e^{-x})^2}{(e^x + e^{-x})^2} \\ &= 1 - \left(\frac{e^x - e^{-x}}{e^x + e^{-x}} \right)^2 = 1 - \tanh^2(x) \end{aligned}$$

Tanh function

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1} = 1 - \frac{2}{e^{2x} + 1}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{-2x}} = -\frac{e^{-2x} - 1}{e^{-2x} + 1} = \frac{2}{e^{-2x} + 1} - 1$$

$$\begin{aligned} \tanh'(x) &= \left(\frac{2}{e^{-2x} + 1} - 1 \right)' = \frac{4e^{-2x}}{(e^{-2x} + 1)^2} = 4 \left(\frac{e^{-2x} + 1 - 1}{(e^{-2x} + 1)^2} \right) \\ &= 4 \left(\frac{1}{e^{-2x} + 1} - \frac{1}{(e^{-2x} + 1)^2} \right) = - \left(\frac{4}{(e^{-2x} + 1)^2} - \frac{4}{e^{-2x} + 1} \right) \\ &= - \left(\frac{4}{(e^{-2x} + 1)^2} - \frac{4}{e^{-2x} + 1} + 1 - 1 \right) = 1 - \left(\frac{2}{e^{-2x} + 1} - 1 \right)^2 = 1 - \tanh^2(x) \end{aligned}$$

Logistic Regression - Tanh

❖ Construct loss

$$\begin{aligned} z &= \theta^T x \\ \hat{y} &= \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \\ L &= \frac{1}{N} [-y^T \log(\hat{y}) - (1 - y^T) \log(1 - \hat{y})] \end{aligned}$$

Model and Loss

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

Derivative

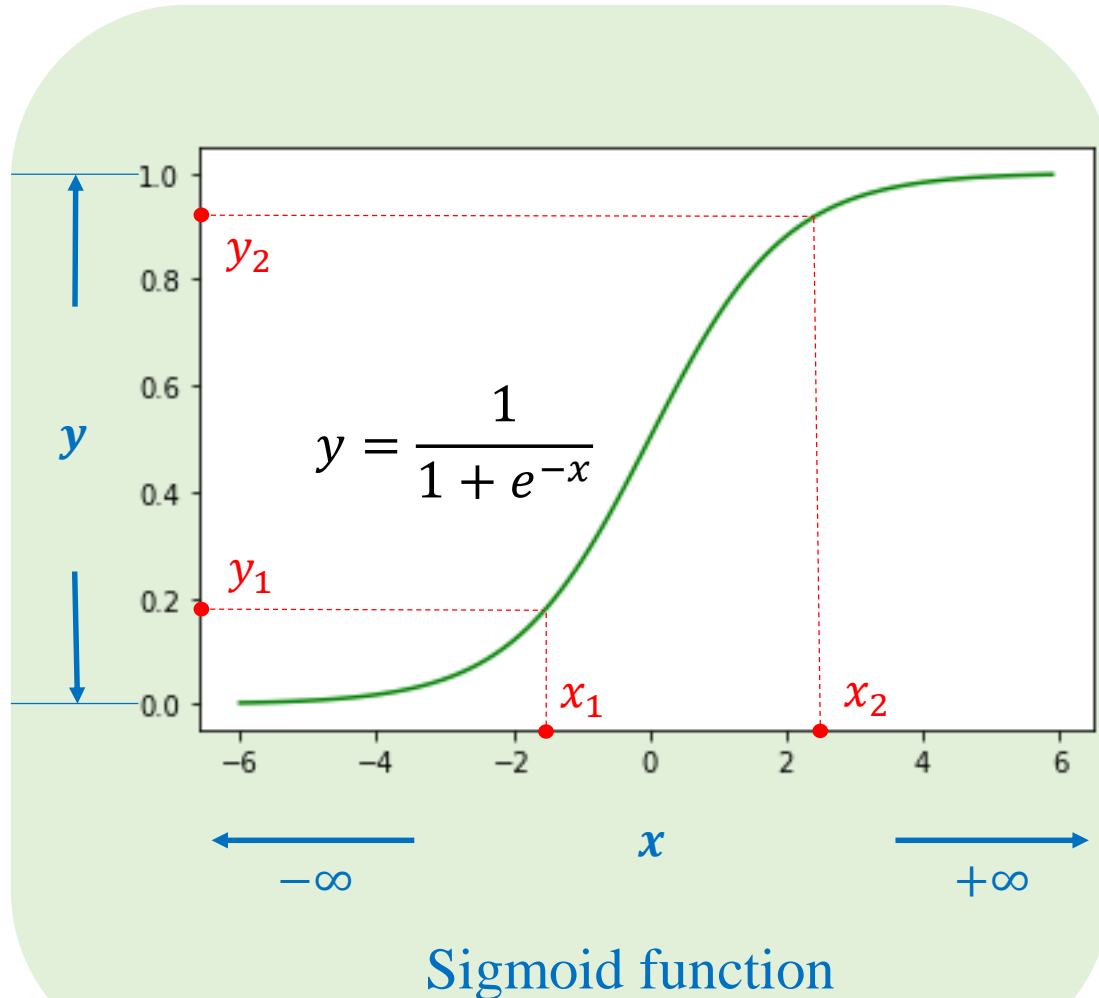
$$\frac{\partial L}{\partial \hat{y}} = \frac{1}{N} \left(-\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}} \right) = \frac{1}{N} \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}$$

$$\frac{\partial \hat{y}}{\partial z} = 1 - \hat{y}^2$$

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

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

Summary



1) Pick all the samples from training data

2) Tính output \hat{y}

$$z = \theta^T x$$

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

3) Tính loss (binary cross-entropy)

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

4) Tính đạo hàm

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

5) Cập nhật tham số

$$\theta = \theta - \eta L'_\theta$$

η is learning rate

