Module 04 – Extra Class

LOGISTIC REGRESSION

Nguyen Quoc Thai

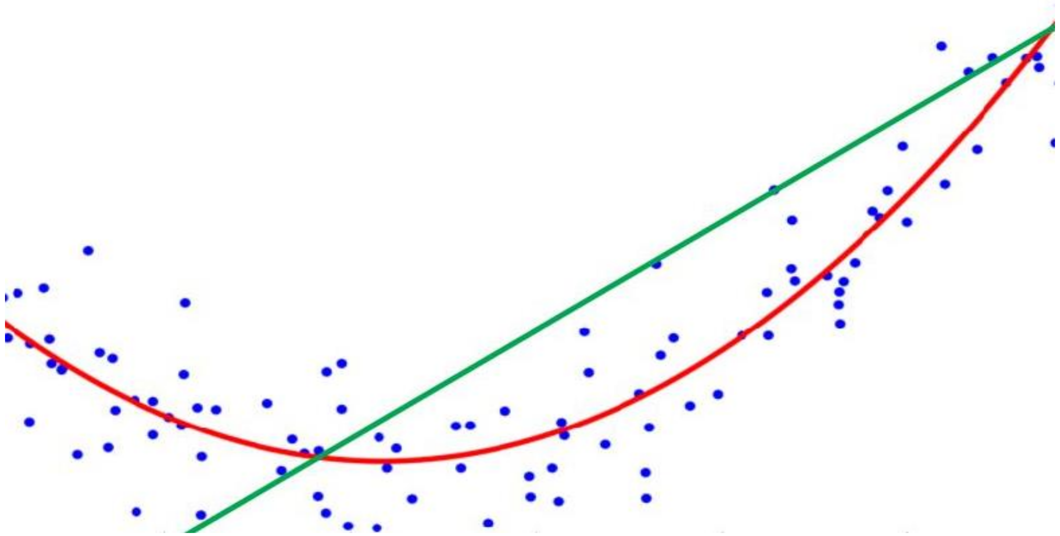
Code & Data

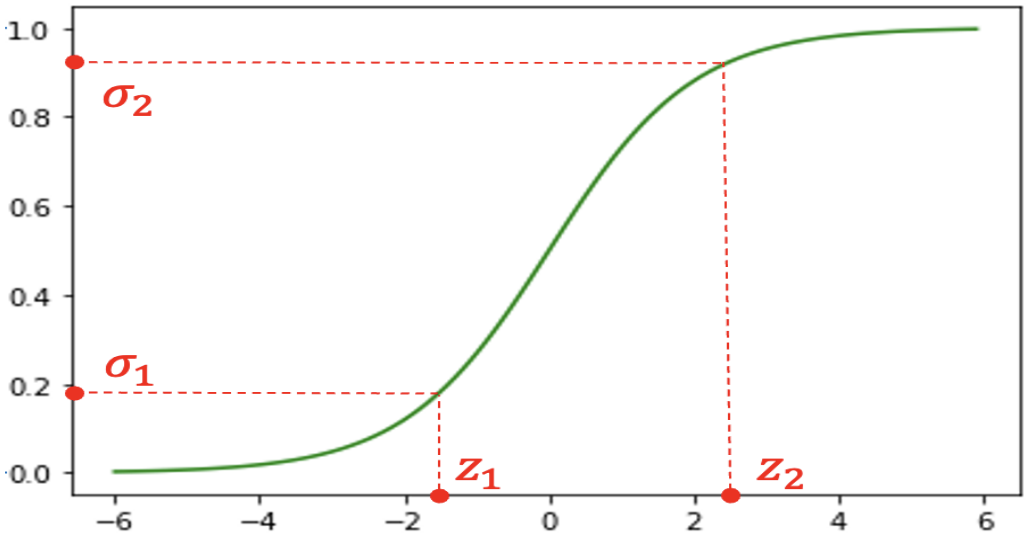
AI VIET NAM
@aivietnam.edu.vn

Objectives

Linear Regression (Review)

- Linear Regression
- Gradient Descent
- Optimal Learning Rate





Logistic Regression

- Logistic Regression
- Sigmoid Function
- One Sample
- N Sample

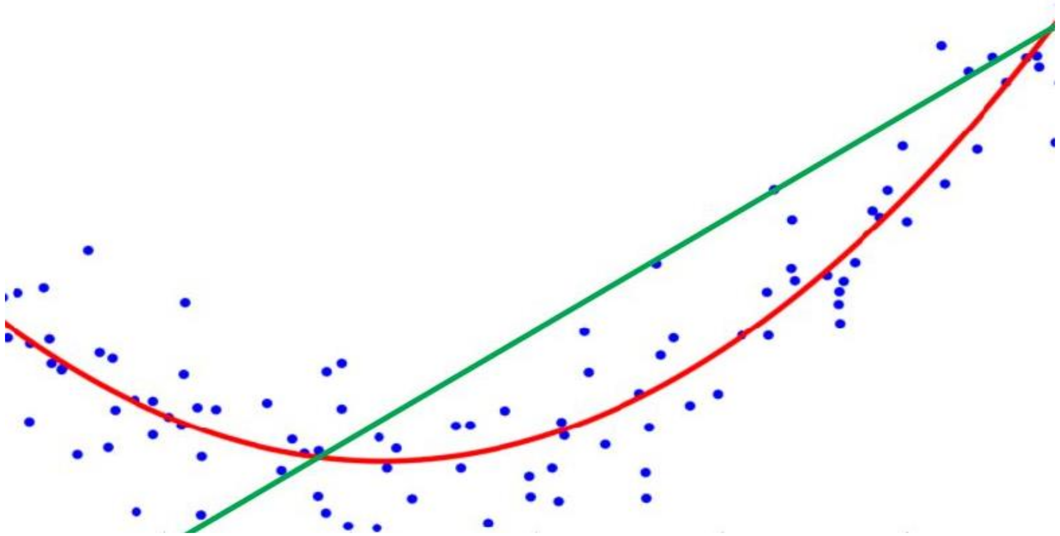
Outline

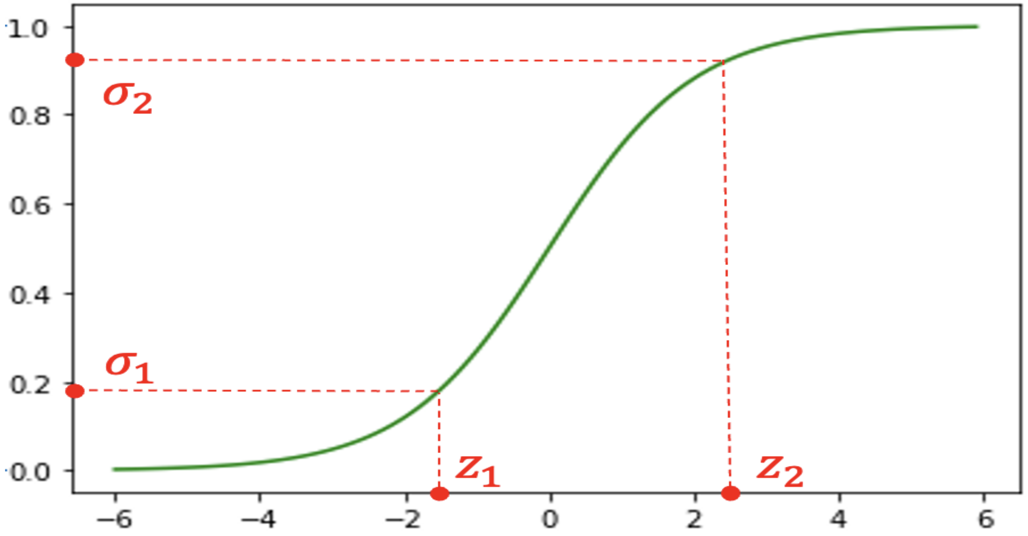
SECTION 1

Linear Regression

SECTION 2

Logistic Regression





Linear Regression

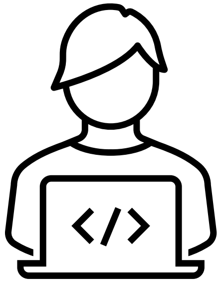
Introduction

Data

Level	Salary
0	8
1	15
2	18
3	22
4	26
5	30
6	38
7	47

Level	Salary
3,5	???
10	???

Learning



Prediction

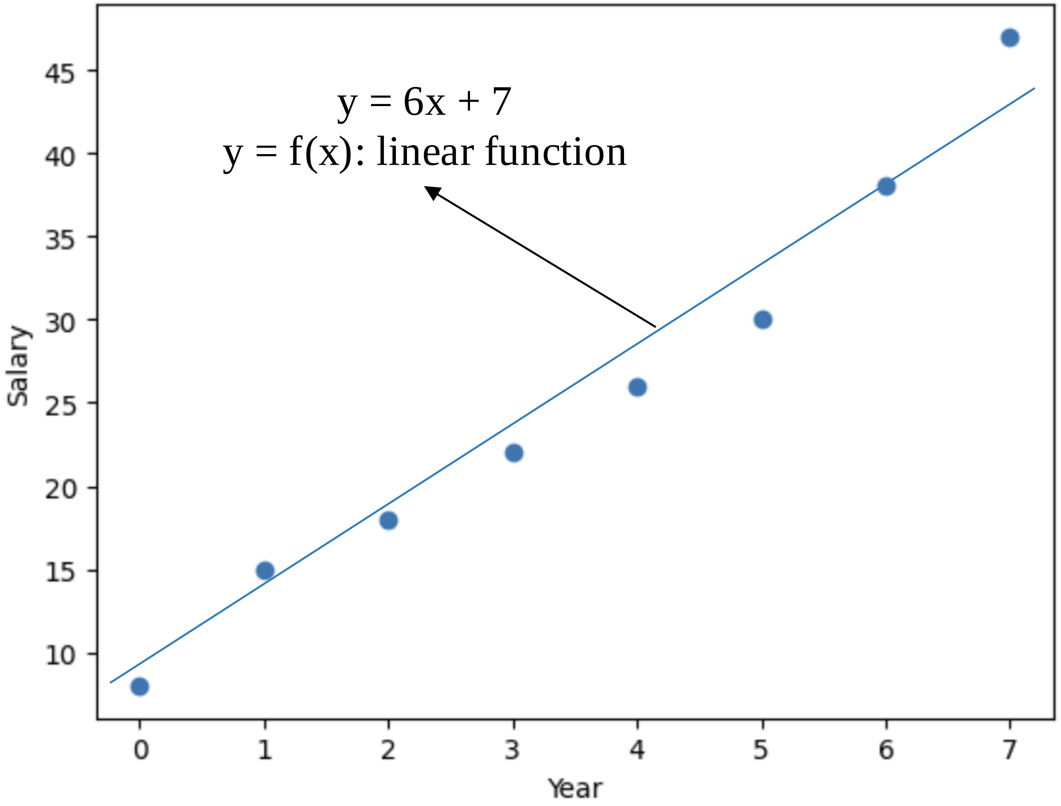
Linear Regression

Linear Regression

Data

Level	Salary
0	8
1	15
2	18
3	22
4	26
5	30
6	38
7	47

Visualization



Linear Regression

Linear Regression

Data

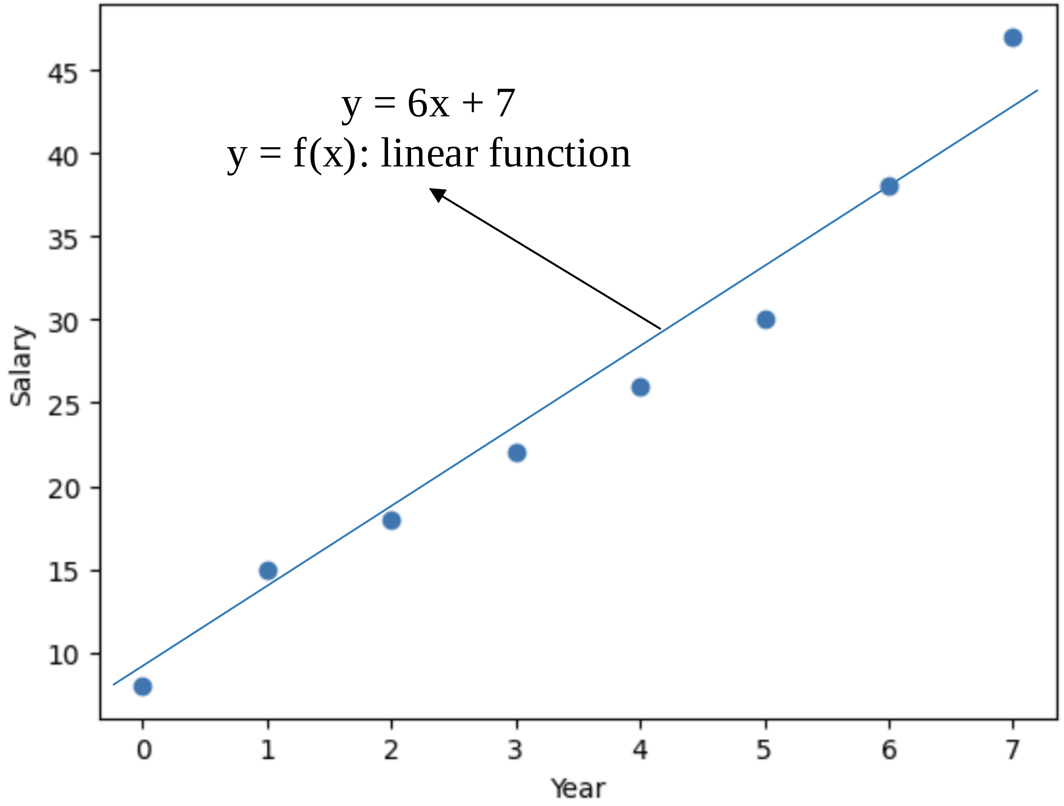
Level	Salary
0	8
1	15
2	18
3	22
4	26
5	30
6	38
7	47

Modeling

$$y = ax + b$$

Find a and b to fit the data

Visualization



Linear Regression

Linear Regression using Gradient Descent

Modeling

$y = ax + b$

Init θ

lr = 0.1

$y = 2x + 2$

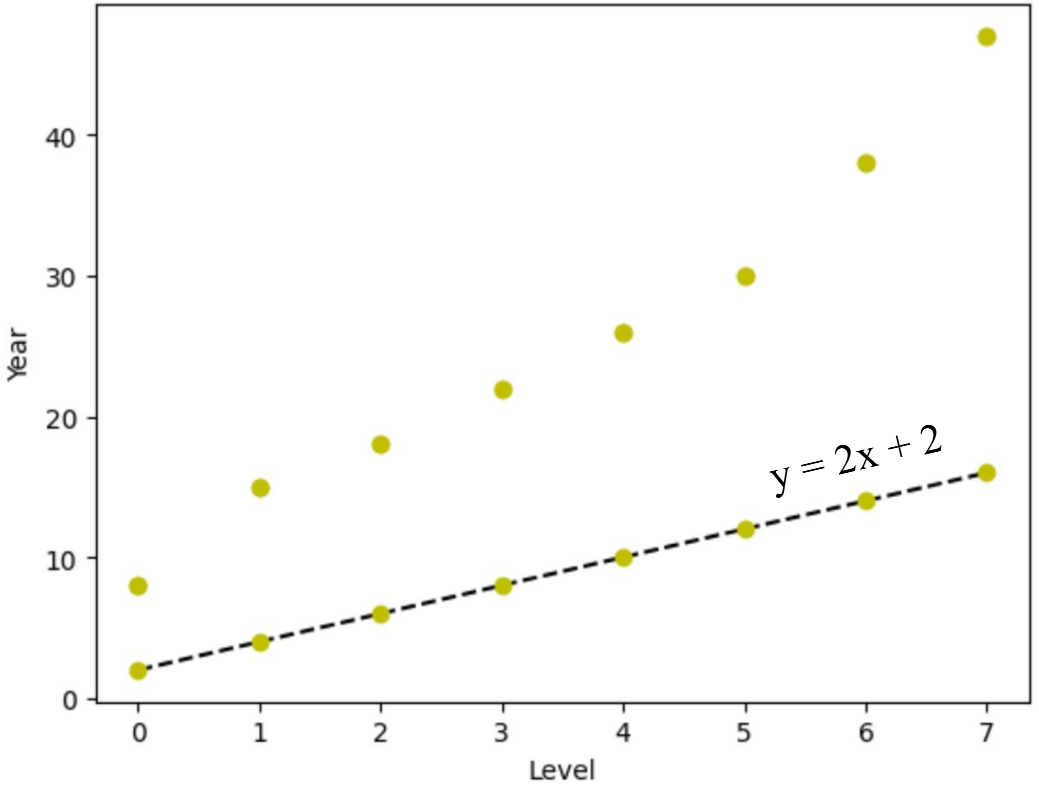
Data

$x = [1\ \ 2]$

$y = 18$

$y = 2x + 2 = 6$

Visualization



Linear Regression

Linear Regression using Gradient Descent

Modeling

$y = ax + b$

Init θ
lr = 0.1

$y = 2x + 2$

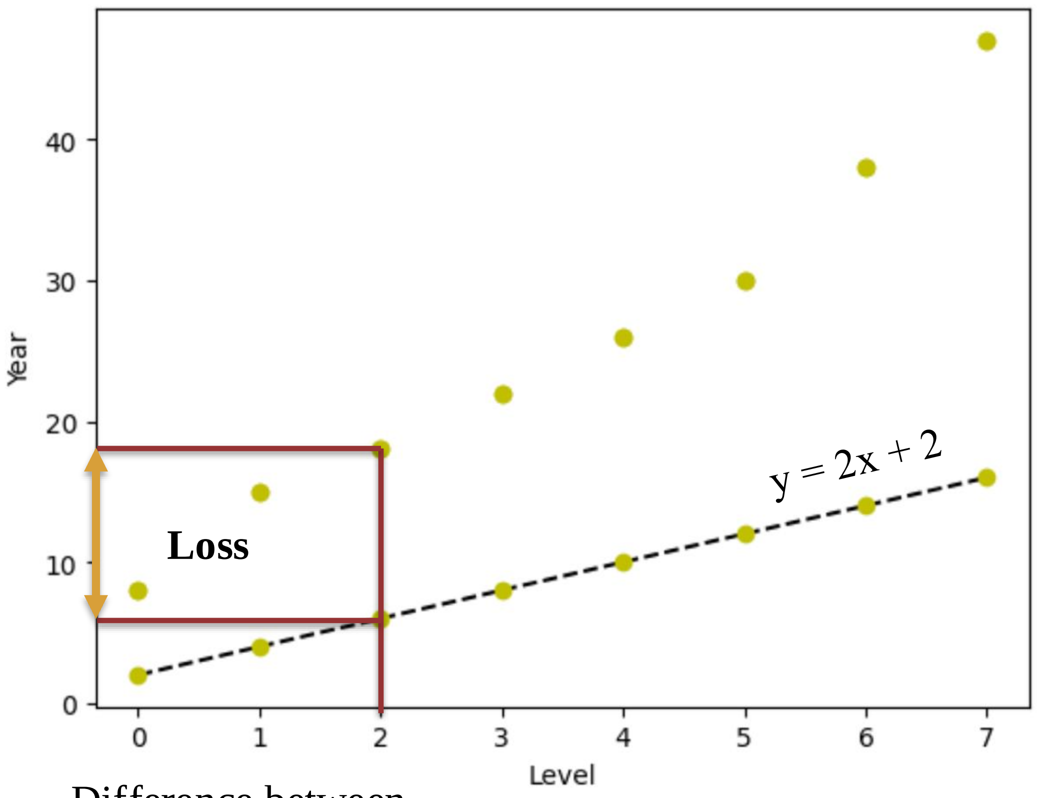
Data

$x = [1\ 2]$
$y = 18$

$y = 2x + 2 = 6$

$L = (6 - 18)^2 = 144$

Visualization



Difference between predicted and actual value

Linear Regression

Linear Regression using Gradient Descent

Modeling

$y = ax + b$ → Init θ, lr = 0.1 → $y = 2x + 2$

Data: $x = [1 \;\; 2]$, $y = 18$

$y = 2x + 2 = 6$

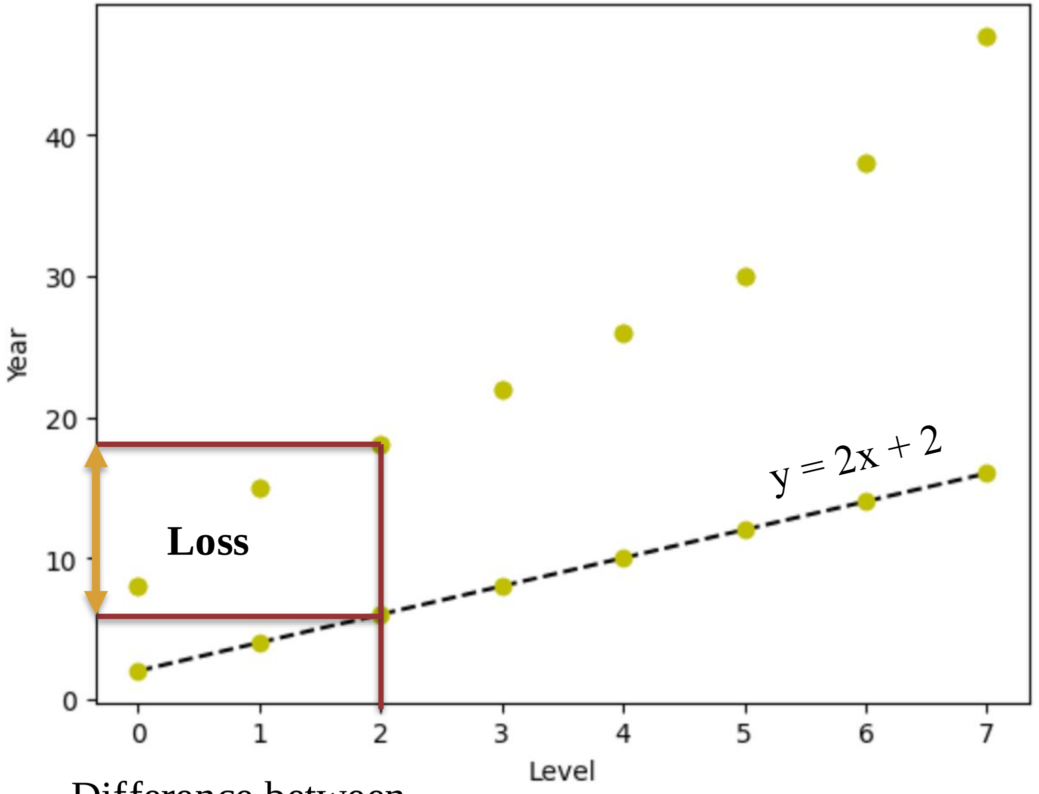
$L = (6 - 18)^2 = 144$

k = -24

$L' = \begin{bmatrix} -24 \\ -48 \end{bmatrix}$

$\theta = \begin{bmatrix} 4.4 \\ 6.8 \end{bmatrix}$

Visualization



Difference between predicted and actual value

Linear Regression

Linear Regression using Gradient Descent

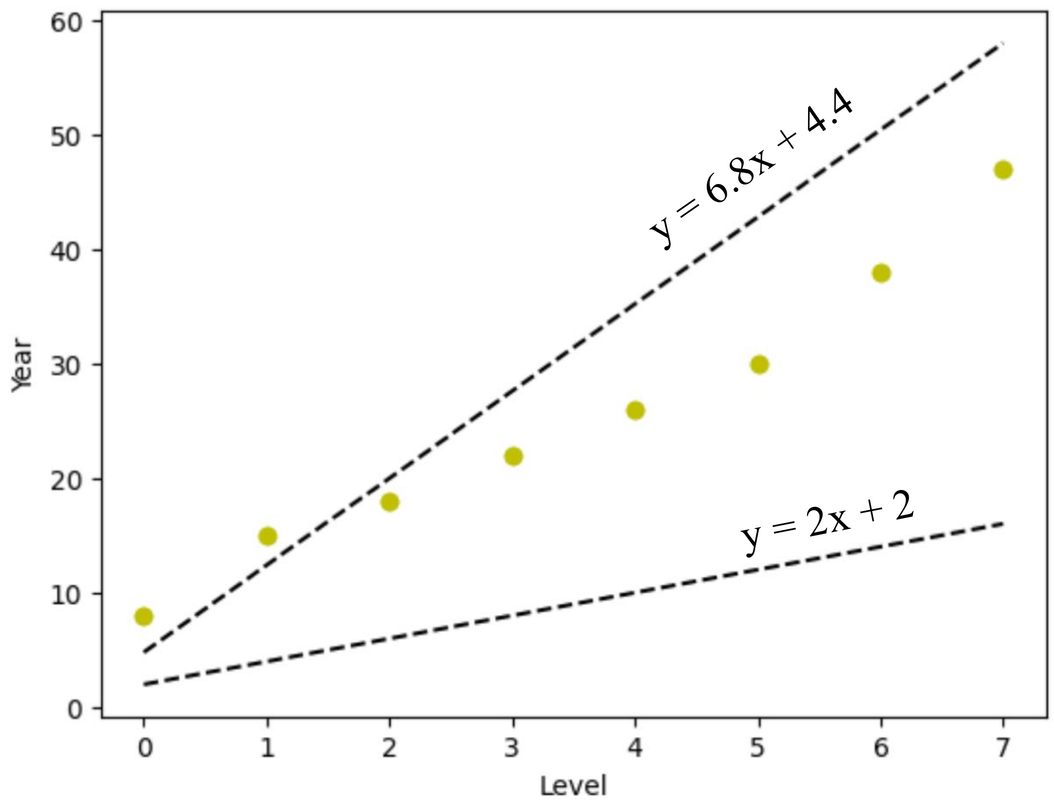
Modeling

$$y = ax + b \quad \dashrightarrow \quad y = 2x + 2$$

Updated

$$y = 6.8x + 4.4$$

Visualization



Linear Regression

Linear Regression using Gradient Descent

Level	Salary
0	8
1	15
2	18
3	22
4	26
5	30
6	38
7	47

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

2) Compute output $\hat{y}$

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

3) Compute loss

$$L = (\hat{y} - y)^2$$

4) Compute derivative

$$\nabla_{\theta} \mathrm{L} = 2\boldsymbol{x}(\hat{y} - y)$$

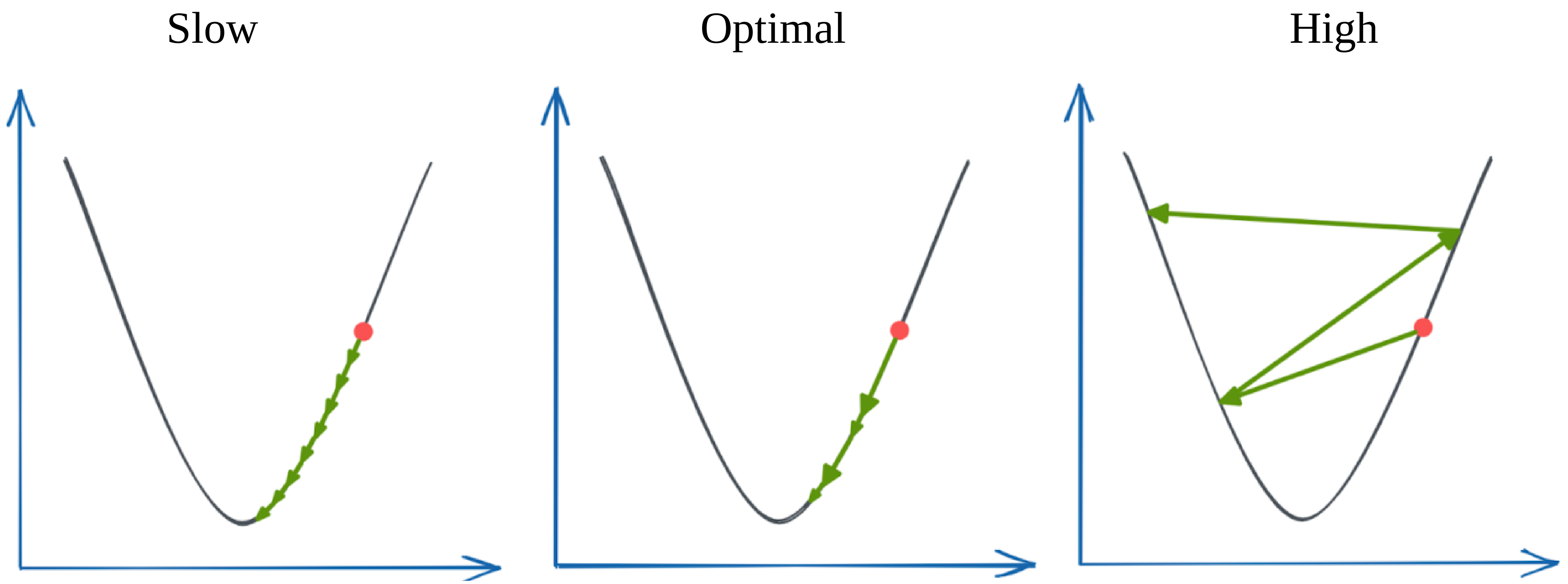
5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\theta} \mathrm{L}$$

η is learning rate

Linear Regression

Optimal Learning Rate



AI VIET NAM
@aivietnam.edu.vn

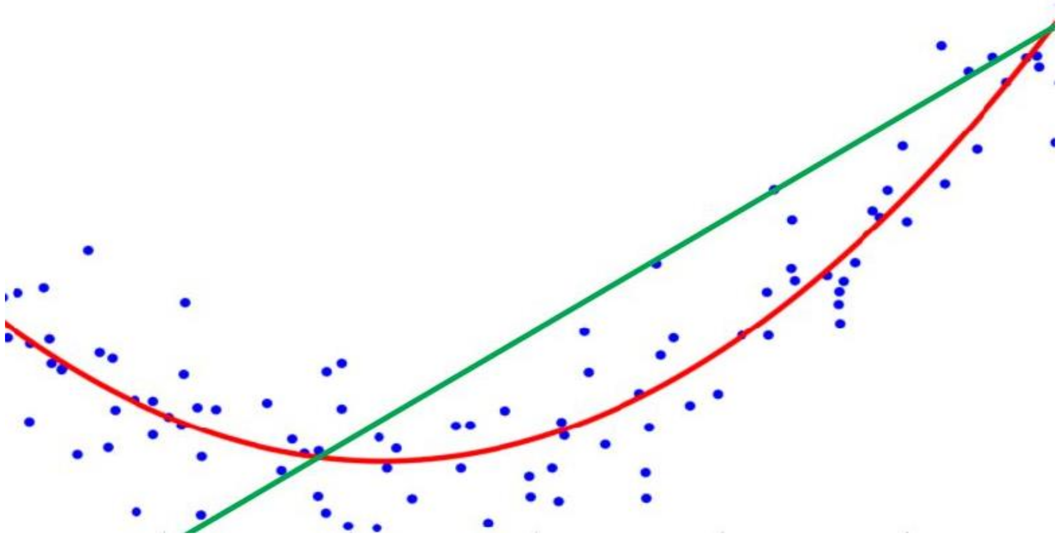
Outline

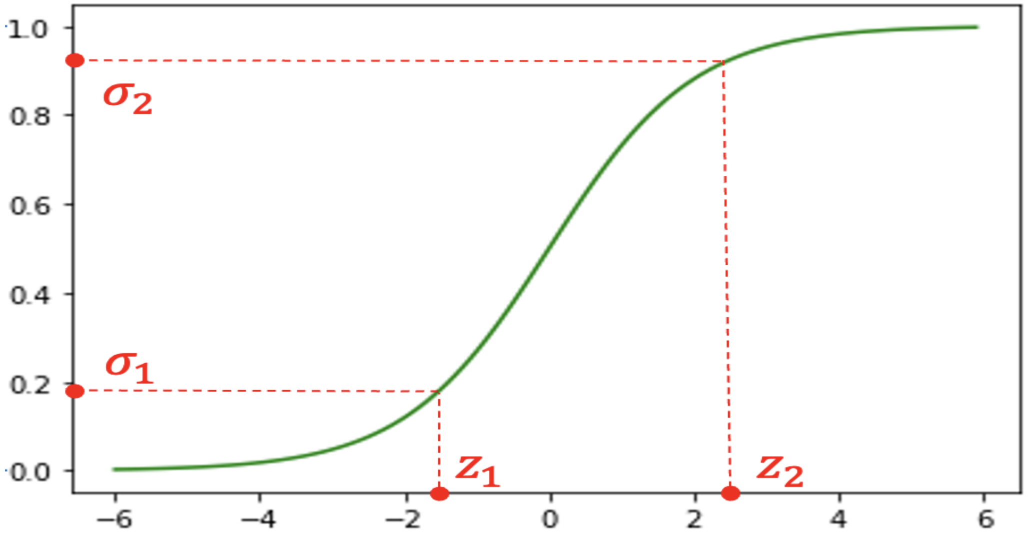
SECTION 1

Linear Regression

SECTION 2

Logistic Regression





Logistic Regression

Classification Task

Data

Level	Salary
0	8
1	15
2	18
3	22
4	26
5	30
6	38
7	47

Continuous value

$y = ax + b$

Values: [0, 1]
Discrete values:{0, 1}

Data #1

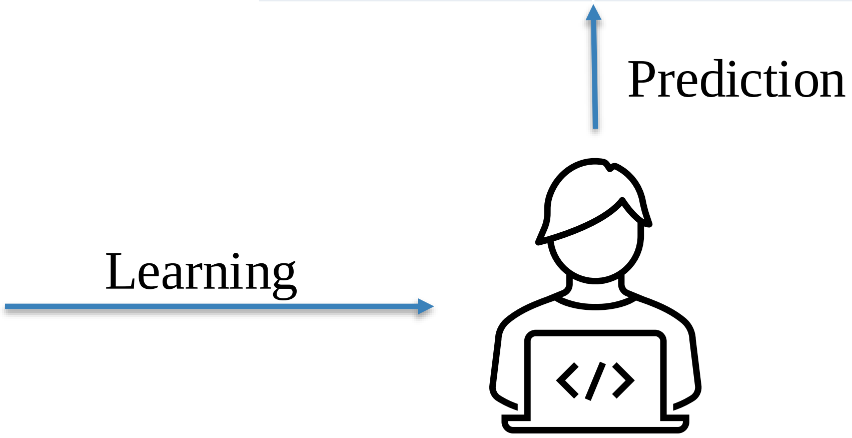
Hours	Pass
0.5	0
1.0	0
1.5	0
2.0	0
2.5	1
3.0	1
3.5	1
4.0	1

Logistic Regression

Classification Task

Hours	Pass
0.5	0
1.0	0
1.5	0
2.0	0
2.5	1
3.0	1
3.5	1
4.0	1

Hours	Pass
0.25	?
4.5	?



Logistic Regression

Classification Task

Hours	Pass
0.5	0
1.0	0
1.5	0
2.0	0
2.5	1
3.0	1
3.5	1
4.0	1

y = f(x)

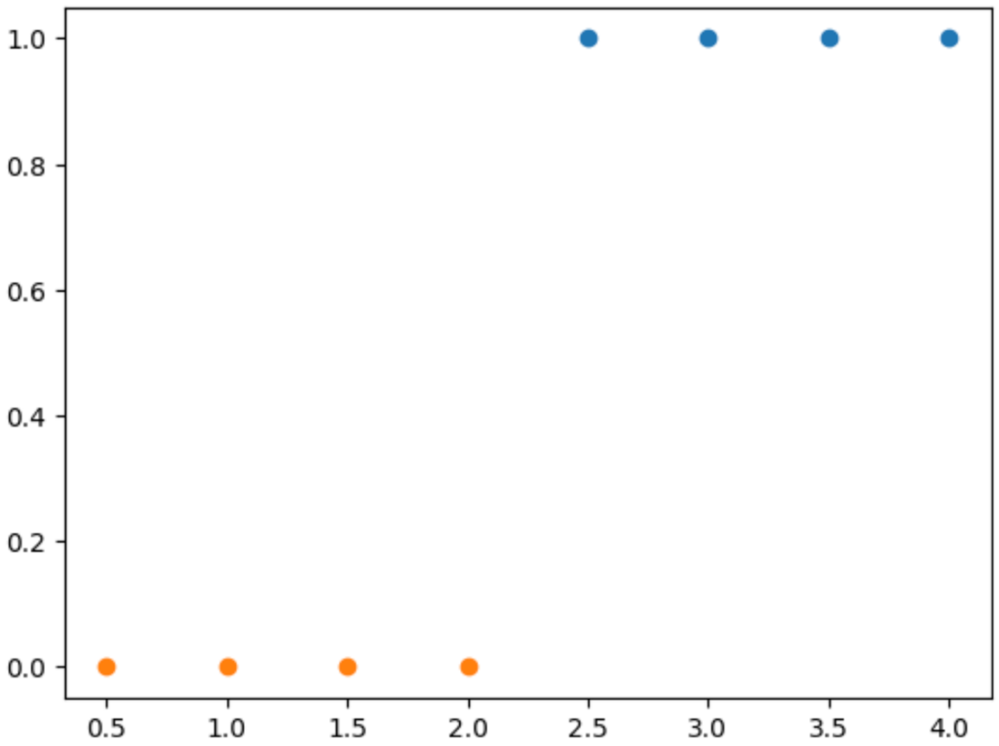
↓

Find a function to fit the data

↓

Sigmoid function

Visualization



Logistic Regression

Sigmoid Function

Sigmoid function

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

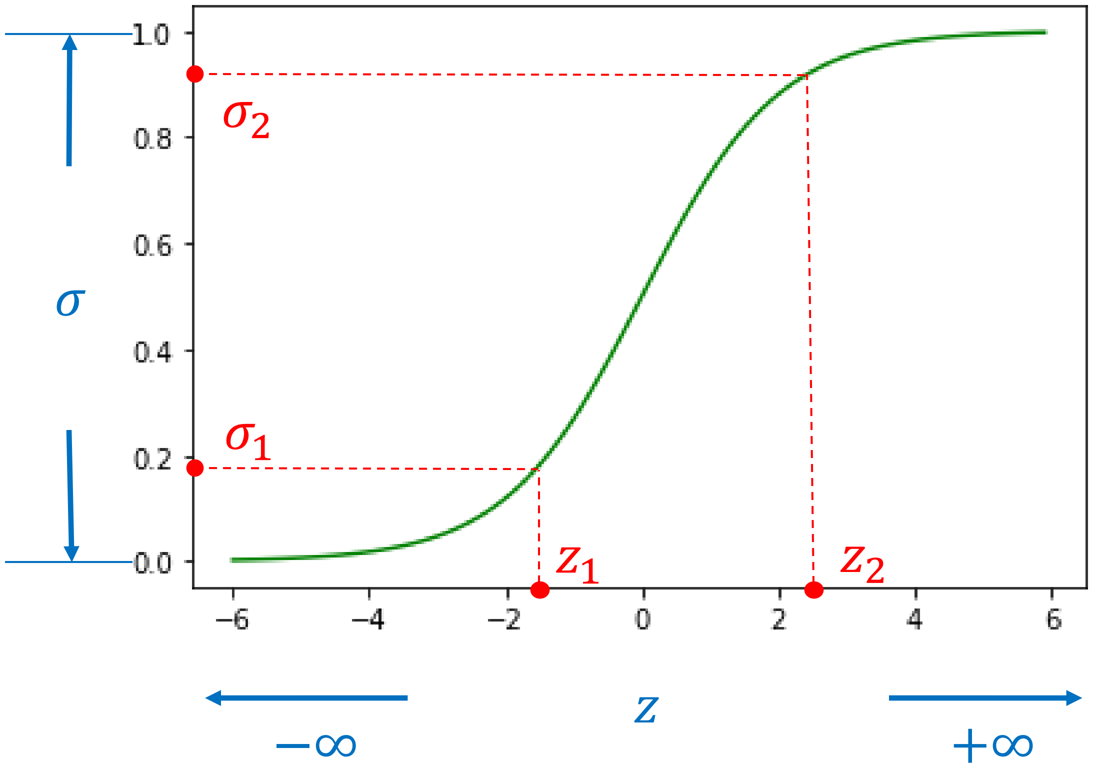
$$z \in (-\infty \quad +\infty)$$

$$\sigma(z) \in (0 \quad 1)$$

Property

$$\forall z_1 z_2 \in [a \ \ b] \text{ and } z_1 \leq z_2$$

$$\rightarrow \sigma(z_1) \leq z(u_1)$$



Logistic Regression

Sigmoid Function

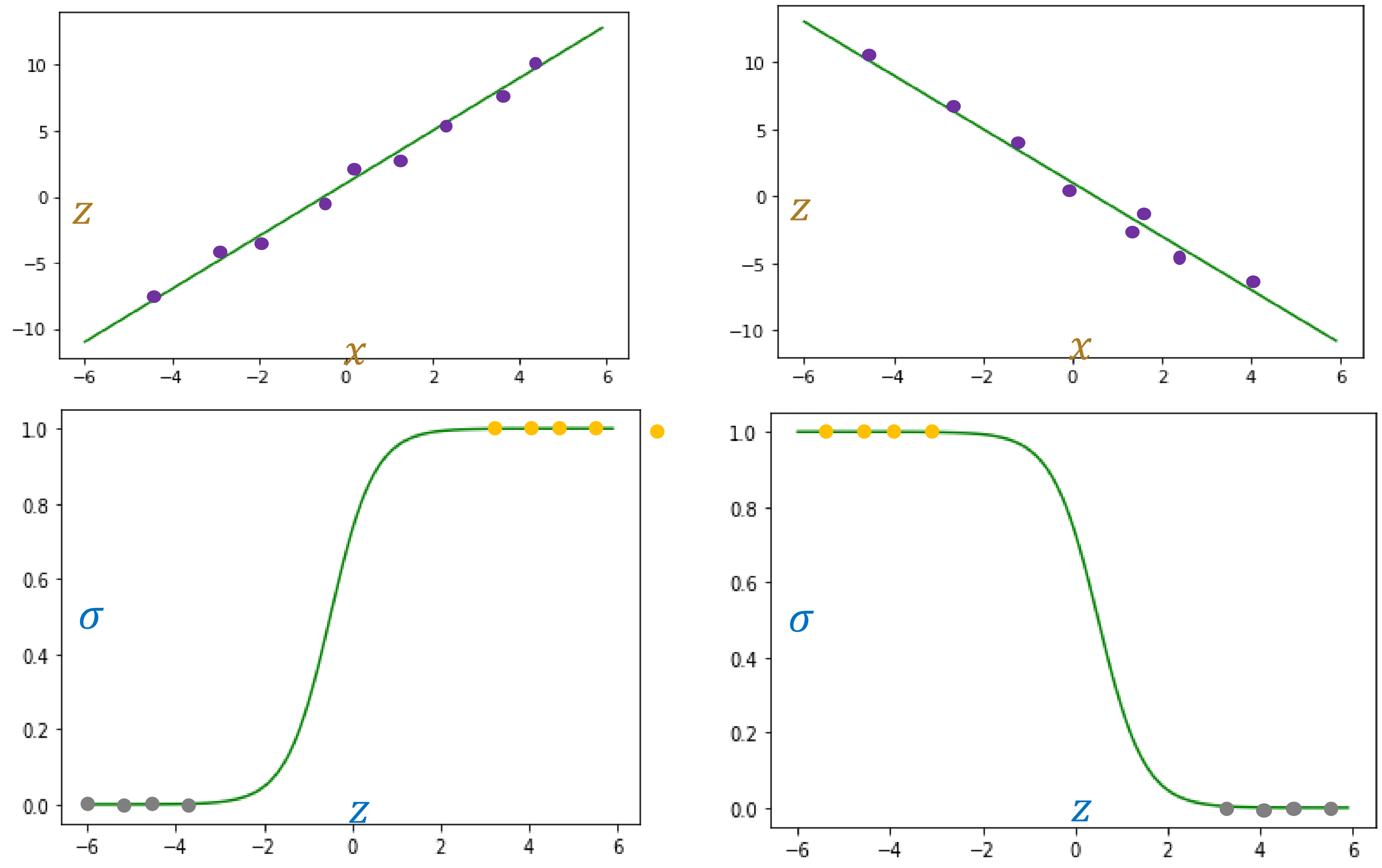
$$z = \boldsymbol{\theta}^T \boldsymbol{x}$$

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

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

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

$$\sigma(z) \in (0 \quad 1)$$



AI VIET NAM
@aivietnam.edu.vn

Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

$$\nabla_{\theta} L = \boldsymbol{x}(\hat{y} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\theta} L$$

η is learning rate

$$\boldsymbol{\theta} = \begin{bmatrix} b \\ w \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}$$

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

$$\boldsymbol{\eta} = 0.1$$

Hours	Pass
0.5	0
1.0	0
1.5	0
2.0	0
2.5	1
3.0	1
3.5	1
4.0	1

Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

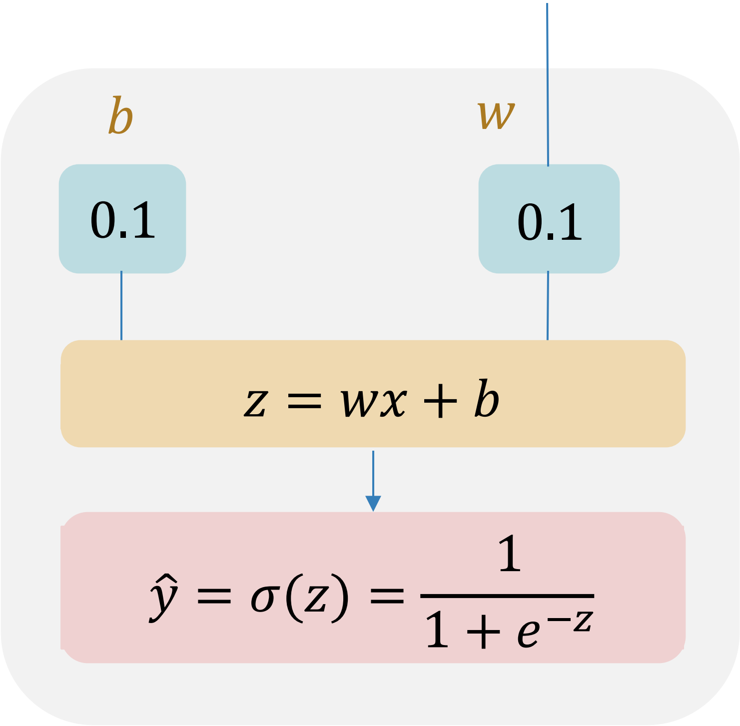
$$\nabla_{\theta} L = \boldsymbol{x}(\hat{y} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\theta} L$$

η is learning rate

$\boldsymbol{\eta} = 0.1$ $\quad$ $\boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}$ $\quad$ $\boldsymbol{x} = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}$ $\quad$ $\boldsymbol{y} = [0]$



$z = 0.15$

$\hat{y} = 0.54$

Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

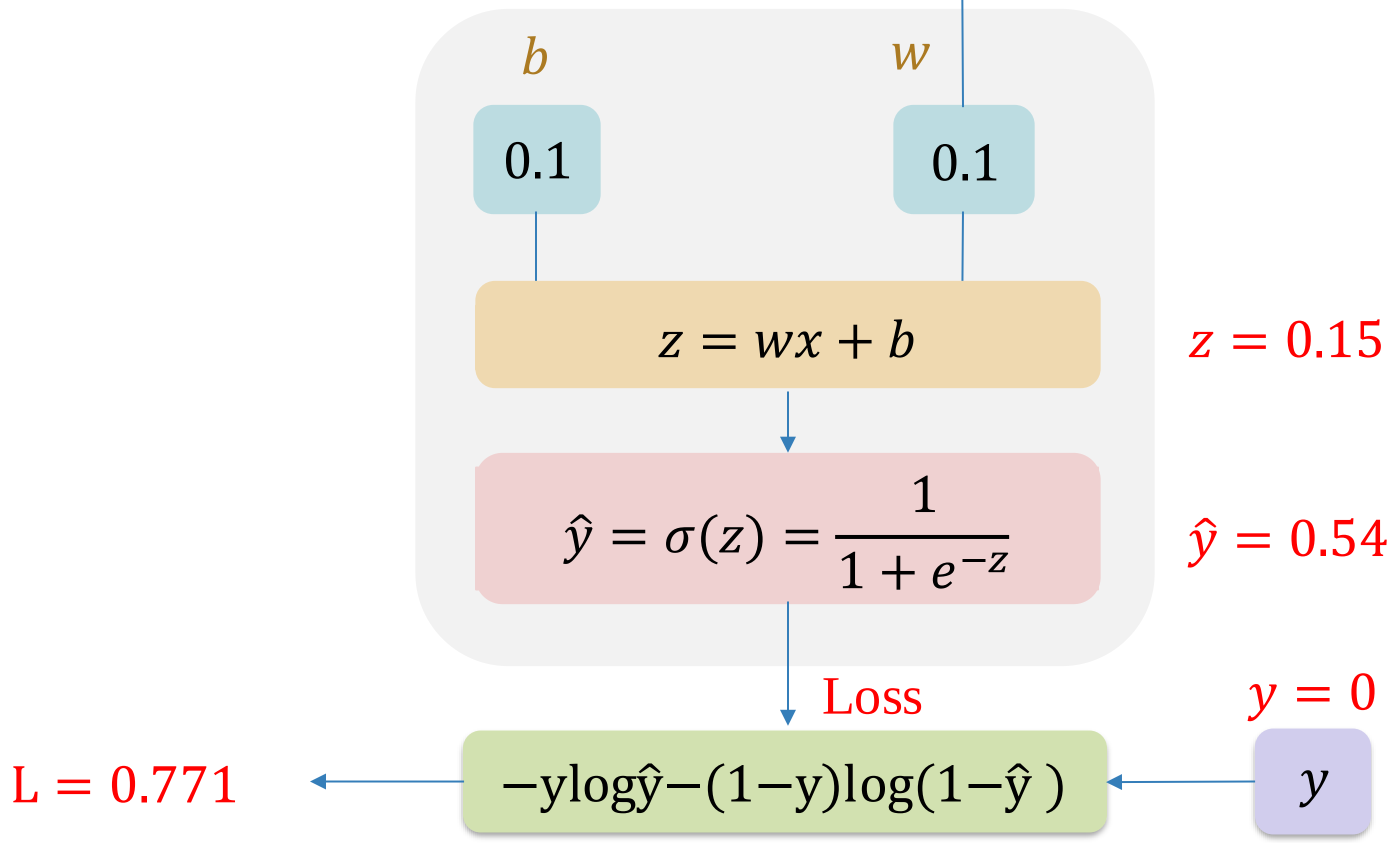
$$\nabla_{\theta}\text{L} = \boldsymbol{x}(\hat{\text{y}} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta\nabla_{\theta}\text{L}$$

η is learning rate

$\boldsymbol{\eta} = 0.1$ $\quad$ $\boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}$ $\quad$ $\boldsymbol{x} = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}$ $\quad$ $\boldsymbol{y} = [0]$



Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

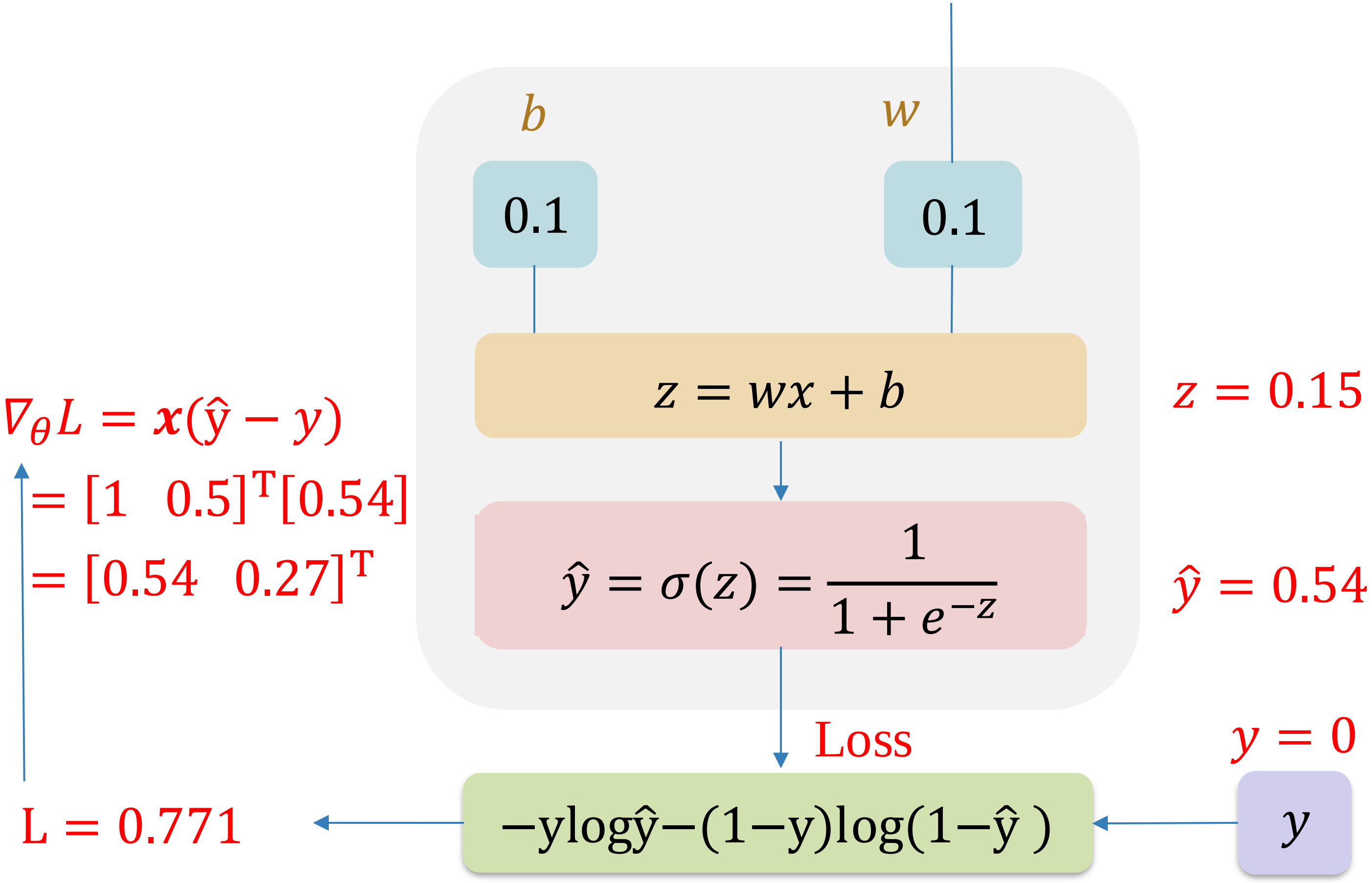
$$\nabla_\theta \text{L} = \boldsymbol{x}(\hat{\text{y}} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_\theta \text{L}$$

η is learning rate

$\boldsymbol{\eta} = 0.1$ $\boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}$ $\boldsymbol{x} = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}$ $\boldsymbol{y} = [0]$



Logistic Regression

Logistic Regression using Gradient Descent

1) Pick a sample (x, y) from training
2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

$$\nabla_\theta \mathrm{L} = \boldsymbol{x}(\hat{\mathrm{y}} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_\theta \mathrm{L}$$

η is learning rate

$\boldsymbol{\eta} = 0.1$ $\quad$ $\boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}$ $\quad$ $\boldsymbol{x} = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}$ $\quad$ $\boldsymbol{y} = [0]$

b $\quad$ w

0.1 $\quad$ 0.1

$z = wx + b$ $\quad$ $z = 0.15$

$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$ $\quad$ $\hat{y} = 0.54$

Loss $\quad$ $y = 0$

$-\mathrm{y}\log\hat{\mathrm{y}} - (1-\mathrm{y})\log(1-\hat{\mathrm{y}})$ $\quad$ y

$\mathrm{L} = 0.771$

$\nabla_\theta L = \boldsymbol{x}(\hat{\mathrm{y}} - y)$

$= [1 \quad 0.5]^{\mathrm{T}}[0.54]$

$= [0.54 \quad 0.27]^{\mathrm{T}}$

$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_\theta \mathrm{L}$

$= [0.046 \quad 0.073]^{\mathrm{T}}$

Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

$$\nabla_{\theta}\mathrm{L} = \boldsymbol{x}(\hat{\mathrm{y}} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\theta}\mathrm{L}$$

η is learning rate

$$\boldsymbol{\theta} = \begin{bmatrix} 0.046 \\ 0.073 \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}$$

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

$$\boldsymbol{\eta} = 0.1$$

Hours	Pass
0.5	0
1.0	0
1.5	0
2.0	0
2.5	1
3.0	1
3.5	1
4.0	1

Logistic Regression

Prediction

Hours	Pass
0.5	0
1.0	0
1.5	0
2.0	0
2.5	1
3.0	1
3.5	1
4.0	1

Learning

Hours	Pass
0.25	?
4.5	?

$$z = [0.046 \quad 0.073]^{\mathrm{T}} \boldsymbol{x}$$

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

Logistic Regression

Prediction

Hours	Pass
0.25	**?**
4.5	?

$$z = [0.046 \quad 0.073]^{\mathrm{T}}\boldsymbol{x}$$

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

Thresholds = 0.5

$y_{pred}: 0$

Hours	Pass
0.25	?
4.5	**?**

$$z = [0.046 \quad 0.073]^{\mathrm{T}}\boldsymbol{x}$$

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

Thresholds = 0.5

$y_{pred}: 1$

QUIZ TIME

Logistic Regression

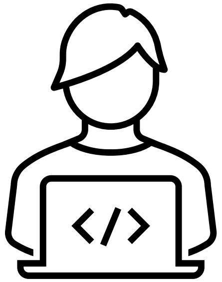
Multivariable Logistic Regression

Day	Hours	Pass
1	0.5	0
2	1.0	0
3	1.5	1
2	2.0	0
1	2.5	0
2	3.0	1
1	3.5	1
2	4.0	1

Day	Hours	Pass
2	0.25	?
1	4.5	?

Prediction

Learning



Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

$$\nabla_{\theta} L = \boldsymbol{x}(\hat{y} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\theta} L$$

η is learning rate

$$\boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.5 \end{bmatrix}$$

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

$$\boldsymbol{\eta} = 0.1$$

Day	Hours	Pass
1	0.5	0
2	1.0	0
3	1.5	1
2	2.0	0
1	2.5	0
2	3.0	1
1	3.5	1
2	4.0	1

AI VIET NAM
@aivietnam.edu.vn

Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

$$\nabla_{\theta}\mathrm{L} = \boldsymbol{x}(\hat{\mathrm{y}} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta\nabla_{\theta}\mathrm{L}$$

η is learning rate

```
x = np.array([1.0, 1.0, 0.5])
y = np.array([0])
x, y
```

```
(array([1. , 1. , 0.5]), array([0]))
```

```
theta = np.array([0.1, 0.2, 0.1])
theta
```

```
array([0.1, 0.2, 0.1])
```

Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

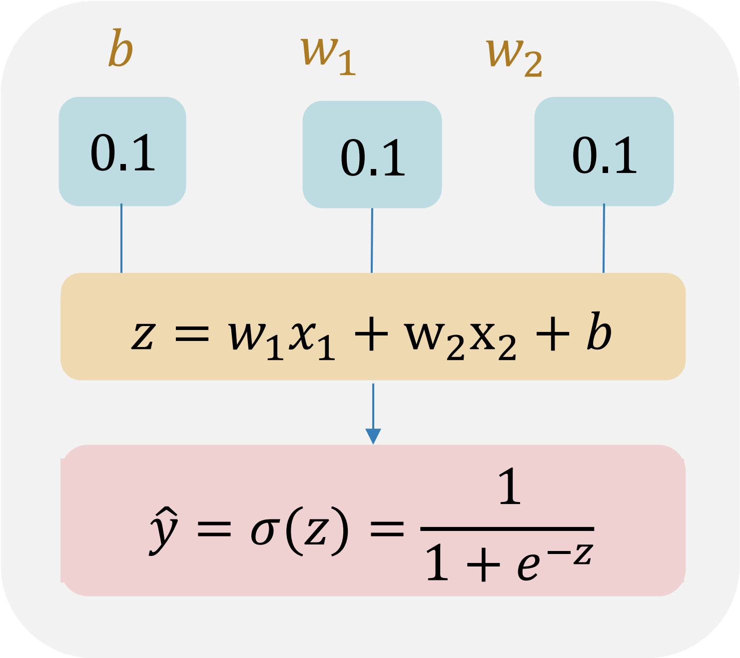
$$\nabla_\theta L = \boldsymbol{x}(\hat{y} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_\theta L$$

η is learning rate

$\boldsymbol{\eta} = 0.1$ $\quad \boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}$ $\quad \boldsymbol{x} = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.5 \end{bmatrix}$ $\quad \boldsymbol{y} = [0]$



$z = 0.35$

$\hat{y} = 0.59$

Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

$$\nabla_{\theta} L = \boldsymbol{x}(\hat{y} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\theta} L$$

η is learning rate

$$\boldsymbol{\eta} = 0.1 \quad \boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.5 \end{bmatrix} \quad \boldsymbol{y} = [0]$$

```python
# define logistic function
def logistic_function(x):
    return 1/(1 + np.exp(-x))
```

```python
# forward
def predict(x, theta):
    z = np.dot(x, theta)
    y_hat = logistic_function(z)
    return z, y_hat

z, y_hat = predict(x, theta)
z, y_hat
```

```
(0.35000000000000003, 0.5866175789173301)
```

Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

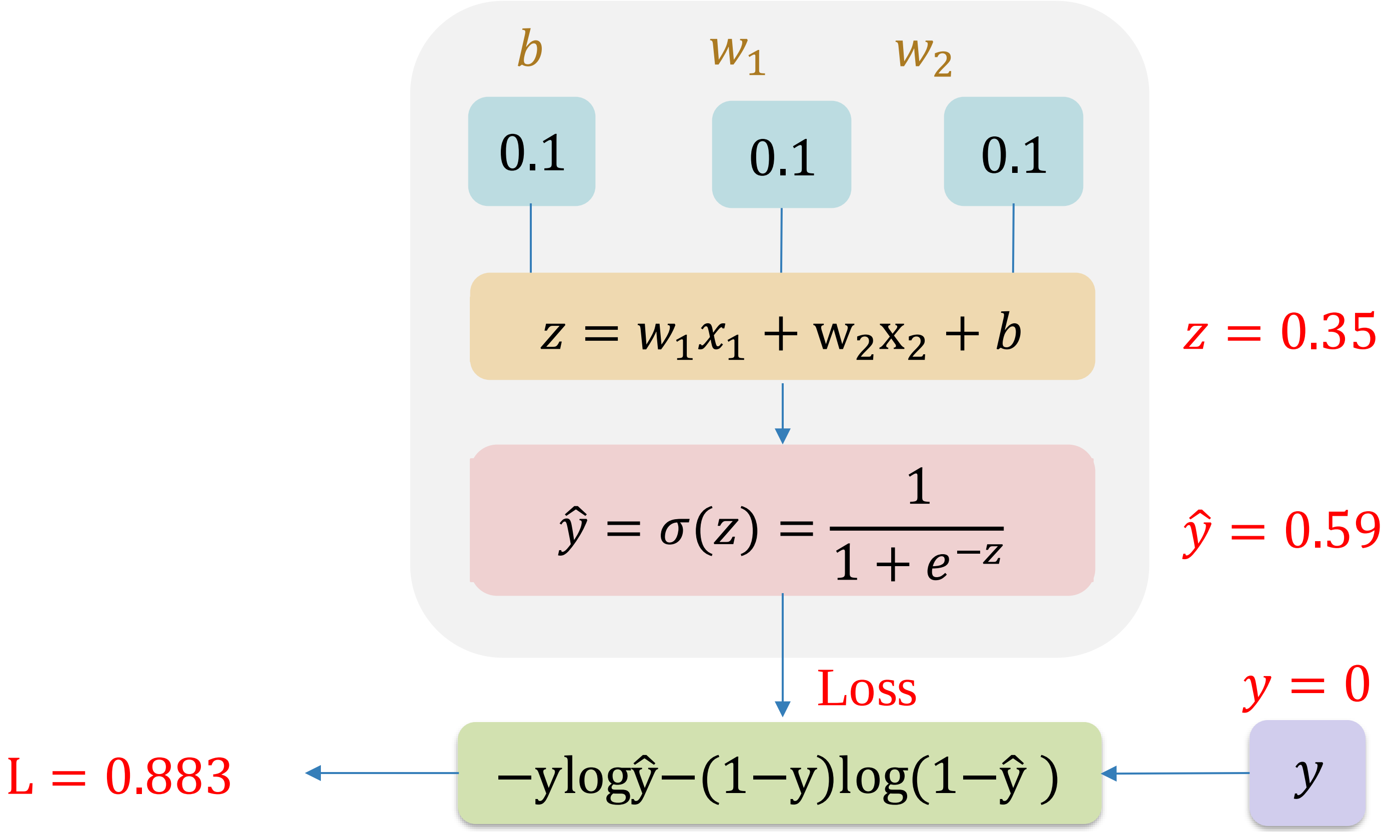
$$\nabla_{\theta} \mathrm{L} = \boldsymbol{x}(\hat{\mathrm{y}} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\theta} \mathrm{L}$$

η is learning rate

$\boldsymbol{\eta} = 0.1$ $\quad$ $\boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}$ $\quad$ $\boldsymbol{x} = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.5 \end{bmatrix}$ $\quad$ $\boldsymbol{y} = [0]$



Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

$$\nabla_{\theta}\mathrm{L} = \boldsymbol{x}(\hat{\mathrm{y}} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta\nabla_{\theta}\mathrm{L}$$

η is learning rate

$$\boldsymbol{\eta} = 0.1 \qquad \boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.5 \end{bmatrix} \qquad \boldsymbol{y} = [0]$$

```
# comput loss
def compute_loss(y_hat, y):
    loss = -1*((y * np.log(y_hat)) + ((1 - y) * np.log(1 - y_hat)))
    return loss

loss = compute_loss(y_hat, y)
loss
```

```
array([0.88338216])
```

Logistic Regression

Logistic Regression using Gradient Descent

$\boldsymbol{\eta} = 0.1 \qquad \boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.5 \end{bmatrix} \qquad \boldsymbol{y} = [0]$

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

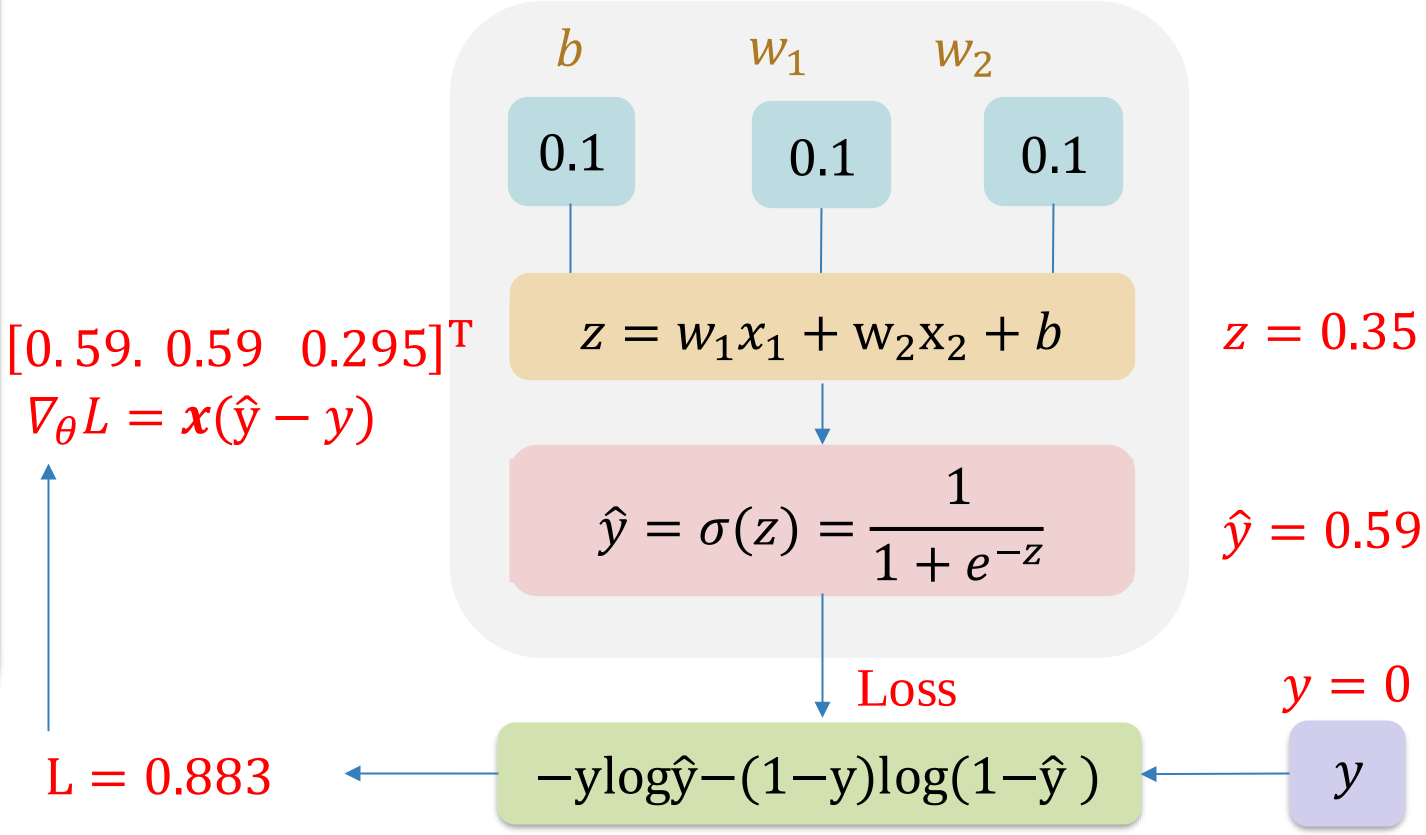
4) Compute derivative

$$\nabla_{\theta}\mathrm{L} = \boldsymbol{x}(\hat{\mathrm{y}} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta\nabla_{\theta}\mathrm{L}$$

η is learning rate



Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

$$\nabla_{\theta} L = \boldsymbol{x}(\hat{y} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\theta} L$$

η is learning rate

$$\boldsymbol{\eta} = 0.1 \qquad \boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.5 \end{bmatrix} \qquad \boldsymbol{y} = [0]$$

```
# compute gradient
def compute_gradient(x, y_hat, y):
    gradient = x*(y_hat - y)
    return gradient

gradient = compute_gradient(x, y_hat, y)
gradient
```

```
array([0.58661758, 0.58661758, 0.29330879])
```

Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

$$\nabla_{\theta} L = \boldsymbol{x}(\hat{y} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\theta} L$$

η is learning rate

$\boldsymbol{\eta} = 0.1$ $\quad \boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}$ $\quad \boldsymbol{x} = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.5 \end{bmatrix}$ $\quad \boldsymbol{y} = [0]$

$[0.04 \;\; 0.014 \;\; 0.07]^T$

$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\theta} L$

$[0.59 \;\; 0.59 \;\; 0.295]^T$

$\nabla_{\theta} L = \boldsymbol{x}(\hat{y} - y)$

b: 0.1 $\quad w_1$: 0.1 $\quad w_2$: 0.1

$z = w_1 x_1 + w_2 x_2 + b$ $\quad z = 0.35$

$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$ $\quad \hat{y} = 0.59$

Loss $\quad y = 0$

$L = 0.883$ $\leftarrow$ $-\mathrm{y}\log\hat{\mathrm{y}} - (1-\mathrm{y})\log(1-\hat{\mathrm{y}})$ $\leftarrow$ y

AI VIET NAM
@aivietnam.edu.vn

Logistic Regression

Logistic Regression using Gradient Descent

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

2) Compute output $\hat{y}$

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

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

3) Compute loss

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

4) Compute derivative

$$\nabla_{\theta}\mathrm{L} = \boldsymbol{x}(\hat{\mathrm{y}} - y)$$

5) Update parameters

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta\nabla_{\theta}\mathrm{L}$$

η is learning rate

$$\boldsymbol{\eta} = 0.1 \qquad \boldsymbol{\theta} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.5 \end{bmatrix} \qquad \boldsymbol{y} = [0]$$

```python
# update weights
learning_rate = 0.1
def update_weight(gradient, theta, learning_rate):
    theta -= (learning_rate * gradient)
    return theta

theta = update_weight(gradient, theta, learning_rate)
theta
```

```
array([0.04133824, 0.14133824, 0.07066912])
```

Logistic Regression

Prediction

Day	Hours	Pass
2	**0.25**	**?**
1	4.5	?

$$z = [0.04 \;\; 0.14 \;\; 0.07]^{\mathrm{T}} \boldsymbol{x}$$

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

Thresholds = 0.5

y_{pred}: 0

Day	Hours	Pass
2	0.25	?
1	**4.5**	**?**

$$z = [0.04 \;\; 0.14 \;\; 0.07]^{\mathrm{T}} \boldsymbol{x}$$

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

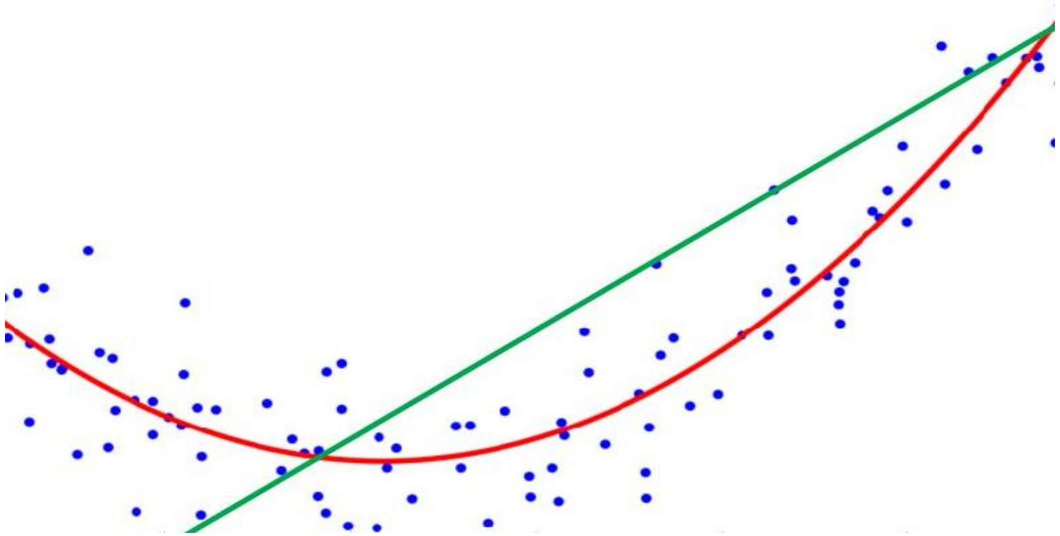
Thresholds = 0.5

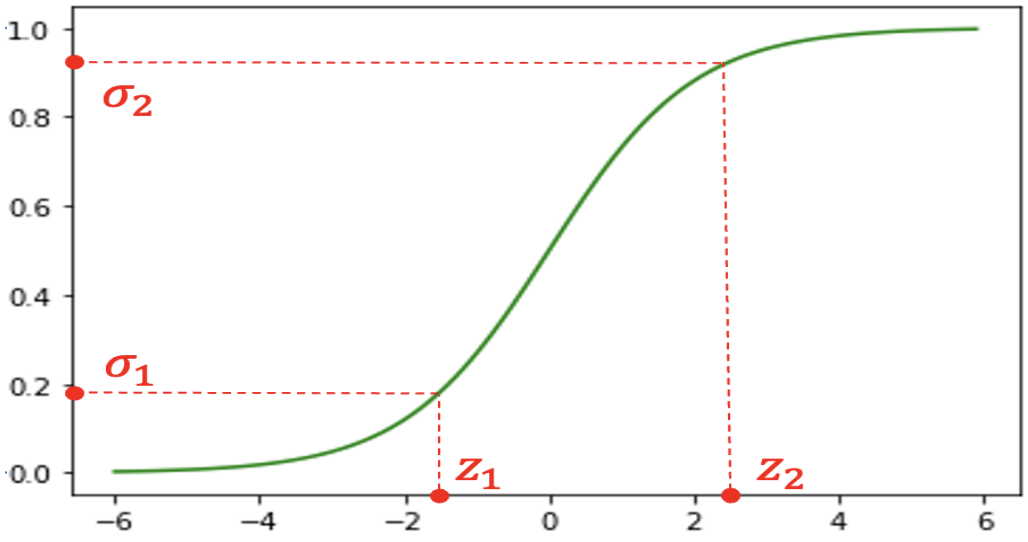
y_{pred}: 1

Objectives

Linear Regression (Review)

- Linear Regression
- Gradient Descent
- Optimal Learning Rate





Logistic Regression

- Logistic Regression
- Sigmoid Function
- One Sample
- N Sample

Thanks!

Any questions?