From Linear Regression to Logistic Regression

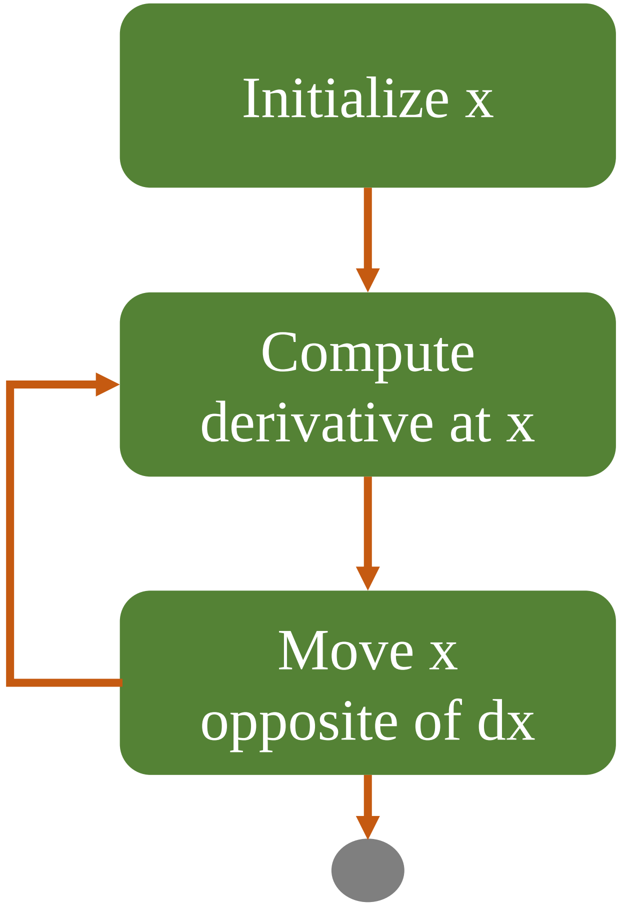
Quang-Vinh Dinh
PhD in Computer Science

Year 2024

code&data

Objectives

Review

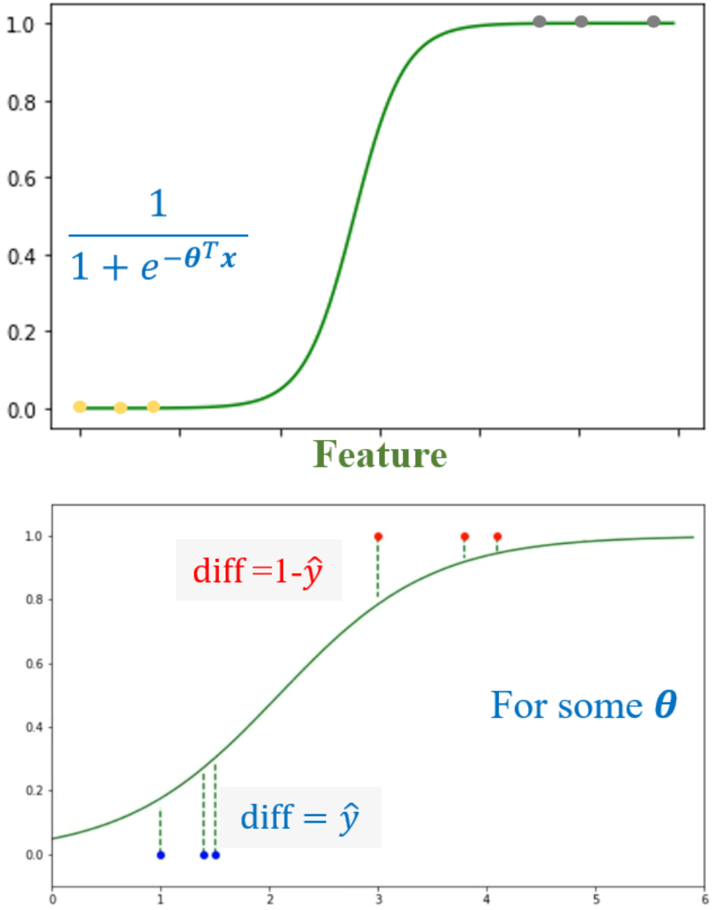


Initialize x

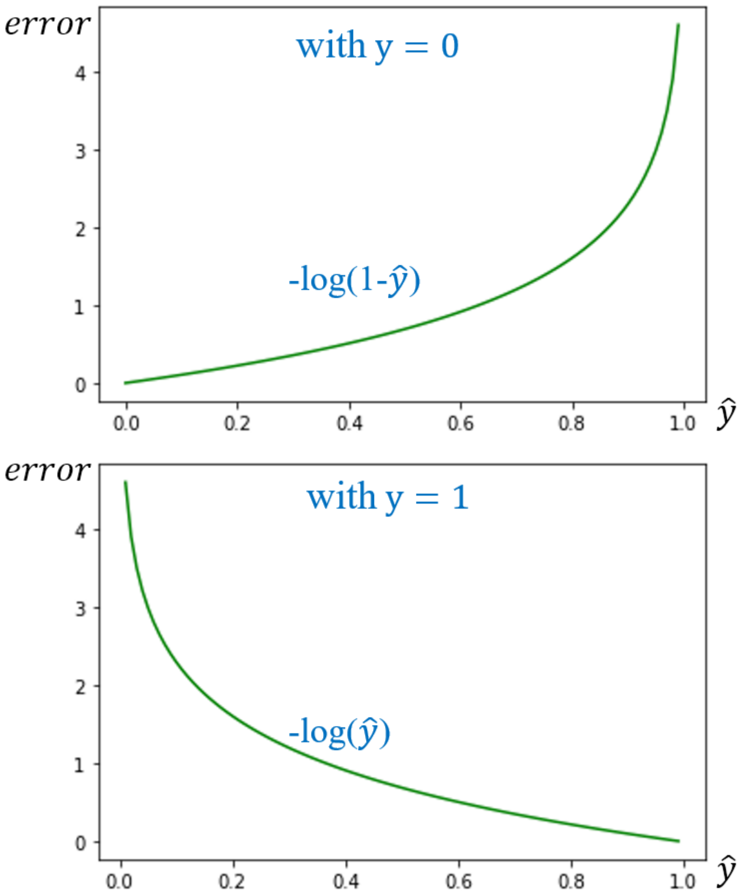
Compute derivative at x

Move x opposite of dx

Lo. R. Using MSE



Lo. R. Using BCE



Outline

SECTION 1

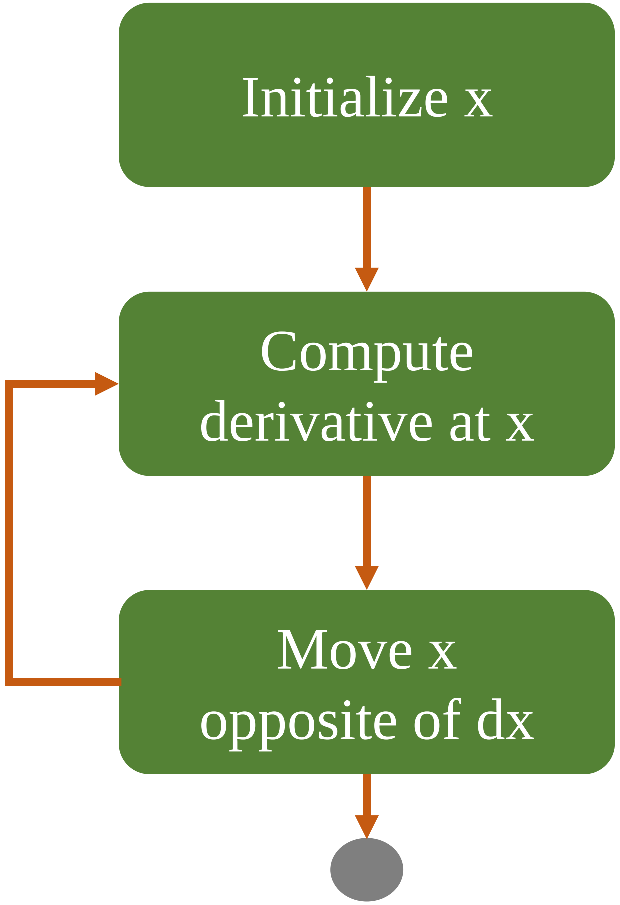
Review Optimization and Linear Regression

SECTION 2

Lo.R. Using MSE

SECTION 3

Lo.R. Using BCE



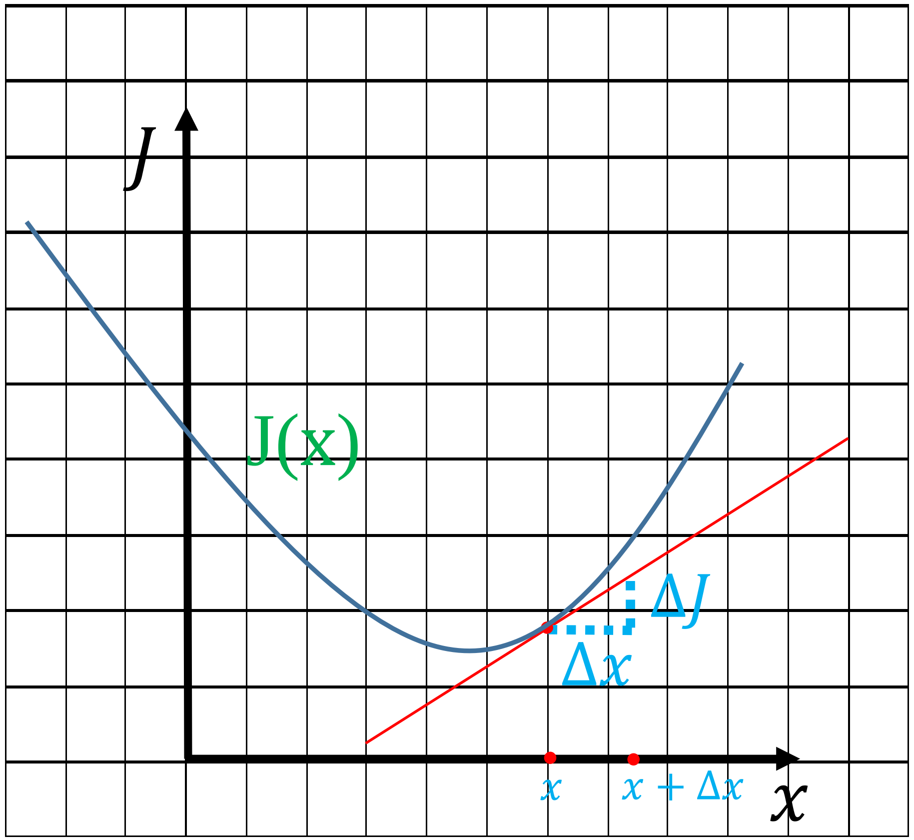
Initialize x

Compute derivative at x

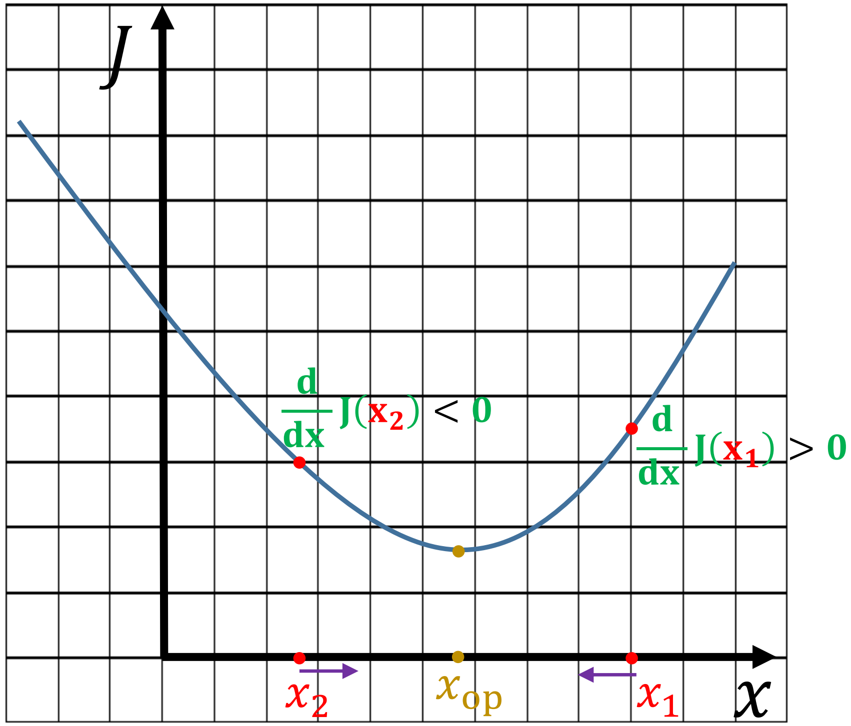
Move x opposite of dx

Optimization

❖ Gradient descent



$$\frac{d}{dx}J(x) = \lim_{\Delta x \to 0} \frac{J(x + \Delta x) - J(x)}{\Delta x}$$



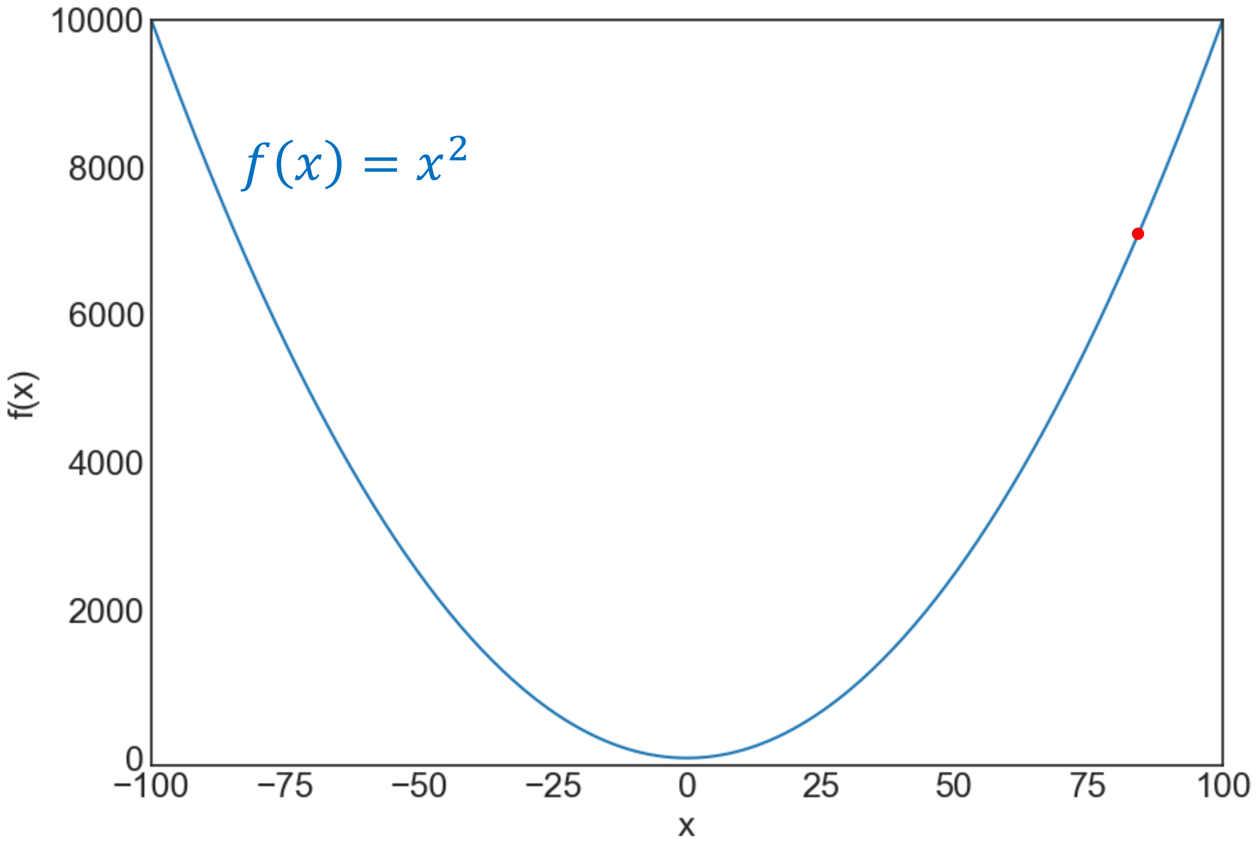
$$x_{new} = x_{old} - \eta \frac{d}{dx}J(x_{old})$$

Derivate at x_{old}

learning rate

Optimization

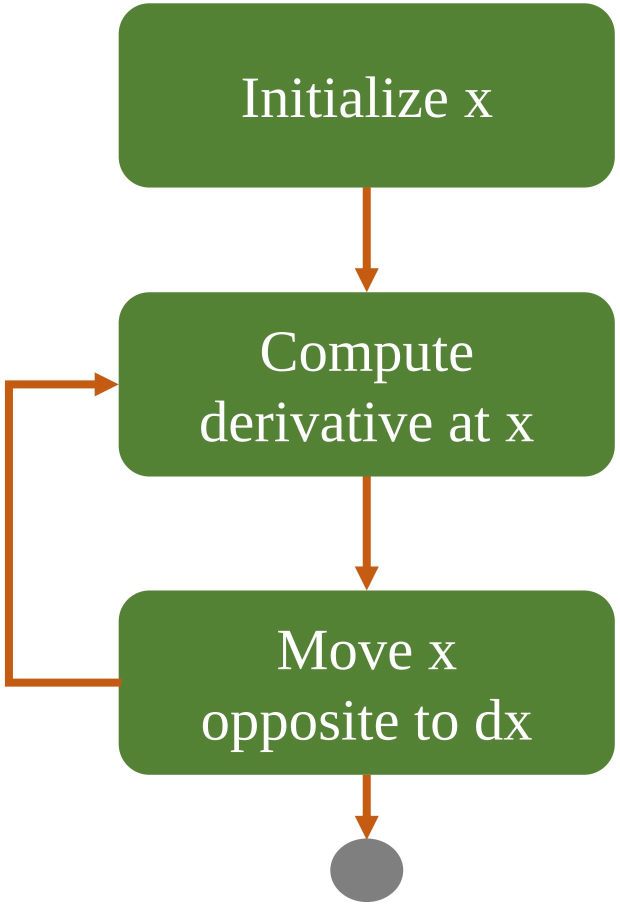
❖ Square function



$$-100 \leq x \leq 100$$

$$x \in \mathbb{N}$$

$$x_t = x_{t-1} - \eta f'(x_{t-1})$$



Initialize x

Compute derivative at x

Move x opposite to dx

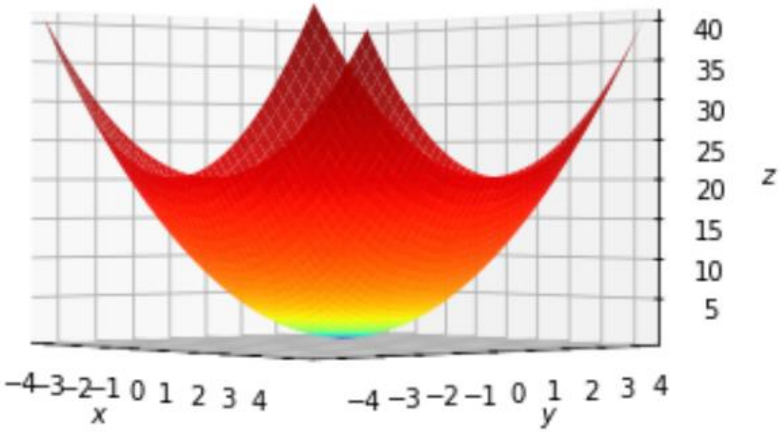
Optimization

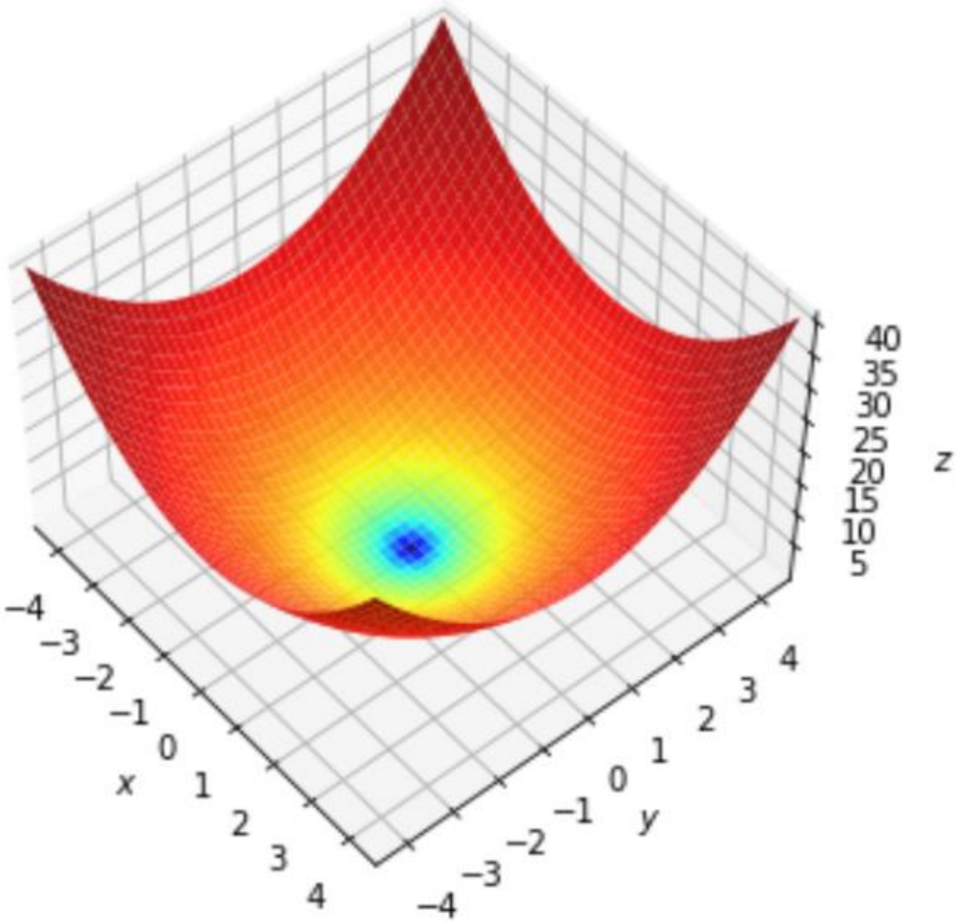
❖ Optimization: 2D function

$$f(x, y) = x^2 + y^2$$

$$-100 \leq x, y \leq 100$$

$$x, y \in \mathbb{N}$$





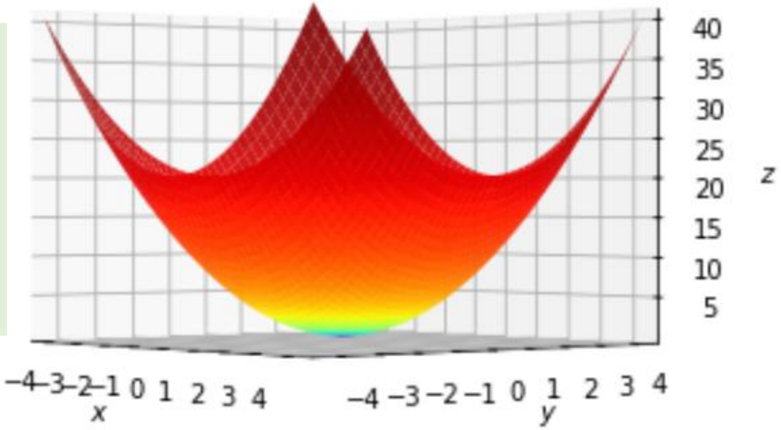
Derivative

❖ Optimization: 2D function

$$f(x, y) = x^2 + y^2$$

$$-100 \leq x, y \leq 100$$

$$x, y \in \mathbb{N}$$



$$x = x - \eta \frac{\partial f(x, y)}{\partial x}$$

$$y = y - \eta \frac{\partial f(x, y)}{\partial y}$$

$$\eta = 1.0$$

$$x_0 = 3.0 \qquad y_0 = 4.0$$

$$\frac{\partial f(x_0, y_0)}{\partial x} = 6.0 \qquad \frac{\partial f(x_0, y_0)}{\partial y} = 8.0$$

$$x_1 = 2.0 \qquad y_1 = 3.0$$

$$\frac{\partial f(x_1, y_1)}{\partial x} = 4.0 \qquad \frac{\partial f(x_1, y_1)}{\partial y} = 6.0$$

$$x_2 = 1.0 \qquad y_2 = 2.0$$

$$\frac{\partial f(x_2, y_2)}{\partial x} = 2.0 \qquad \frac{\partial f(x_2, y_2)}{\partial y} = 4.0$$

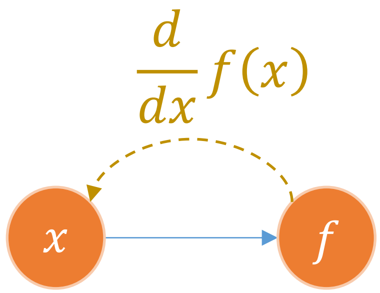
$$x_3 = 0.0 \qquad y_3 = 1.0$$

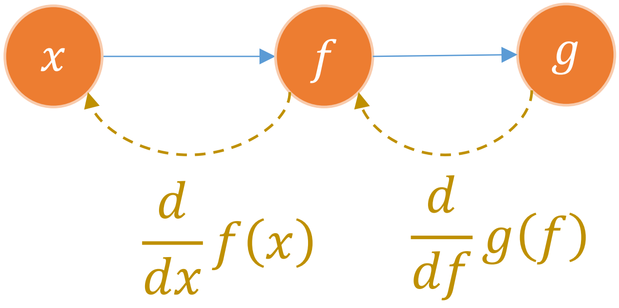
$$\frac{\partial f(x_3, y_3)}{\partial x} = 0.0 \qquad \frac{\partial f(x_3, y_3)}{\partial y} = 0.0$$

$$x_4 = 0.0 \qquad y_4 = 0.0$$

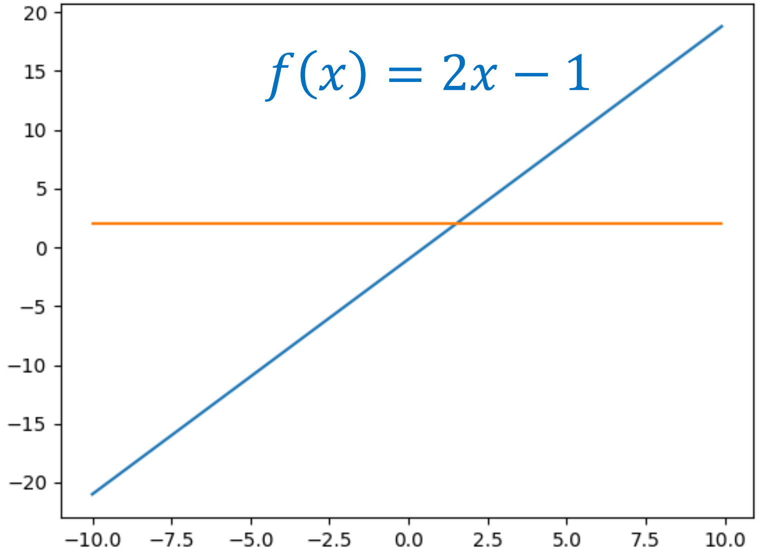
Optimization

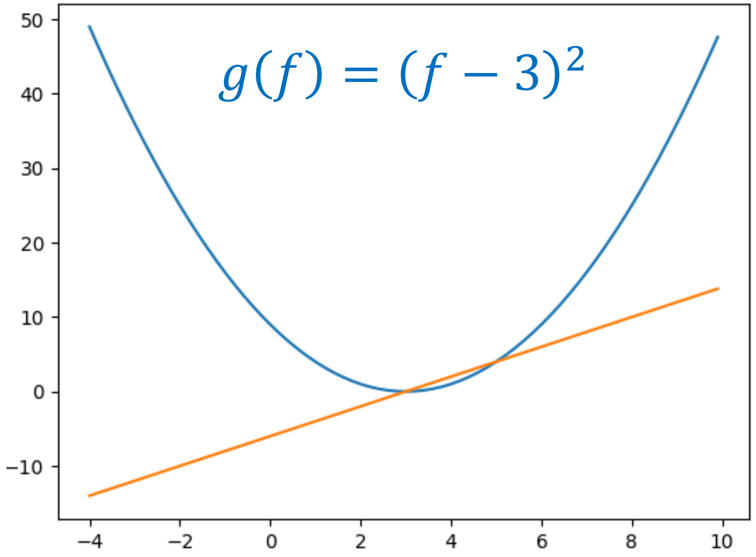
❖ For composite function

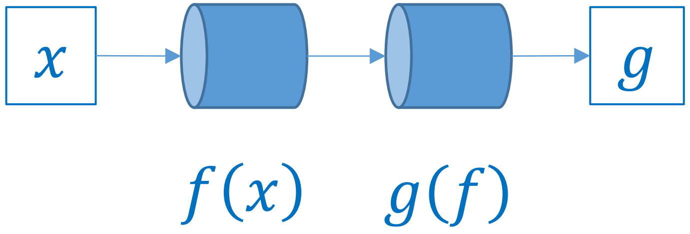


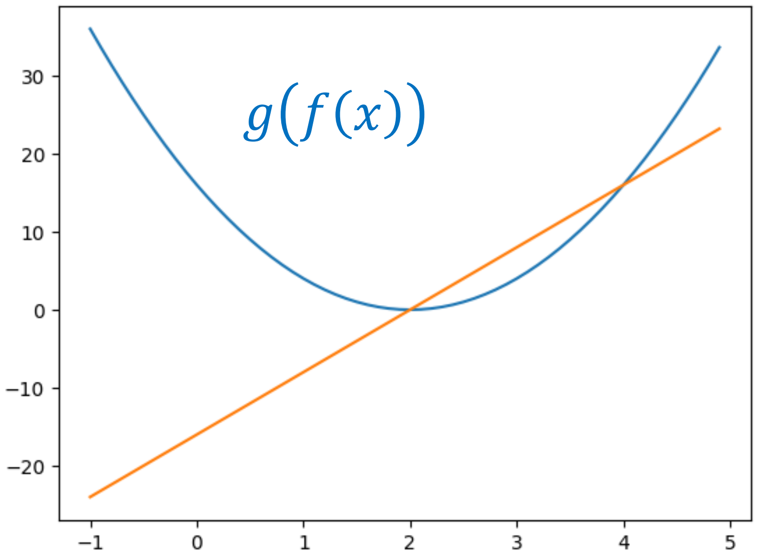


$$\frac{d}{dx}g(f(x)) = \left[\frac{d}{df}g(f)\right] * \left[\frac{d}{dx}f(x)\right]$$



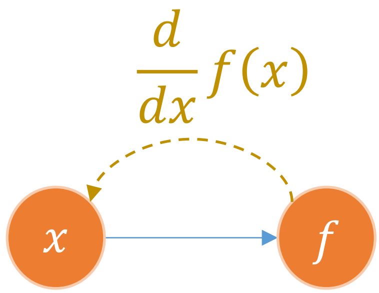


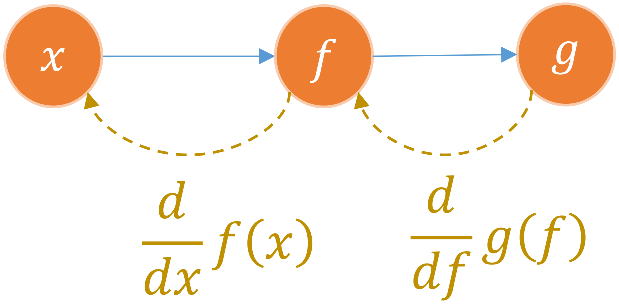




Optimization

❖ For composite function and chain rule





$$\frac{d}{dx}g(f(x)) = \left[\frac{d}{df}g(f)\right] * \left[\frac{d}{dx}f(x)\right]$$

$$f(x) = 2x - 1$$

$$g(f) = (f - 3)^2$$

⇒

$$f'(x) = 2$$

$$g'(f) = 2(f - 3)$$

⇒

$$\frac{dg}{dx} = \frac{dg}{df}\frac{df}{dx}$$

$$= 2(f - 3)2$$

$$= 4(2x - 1 - 3)$$

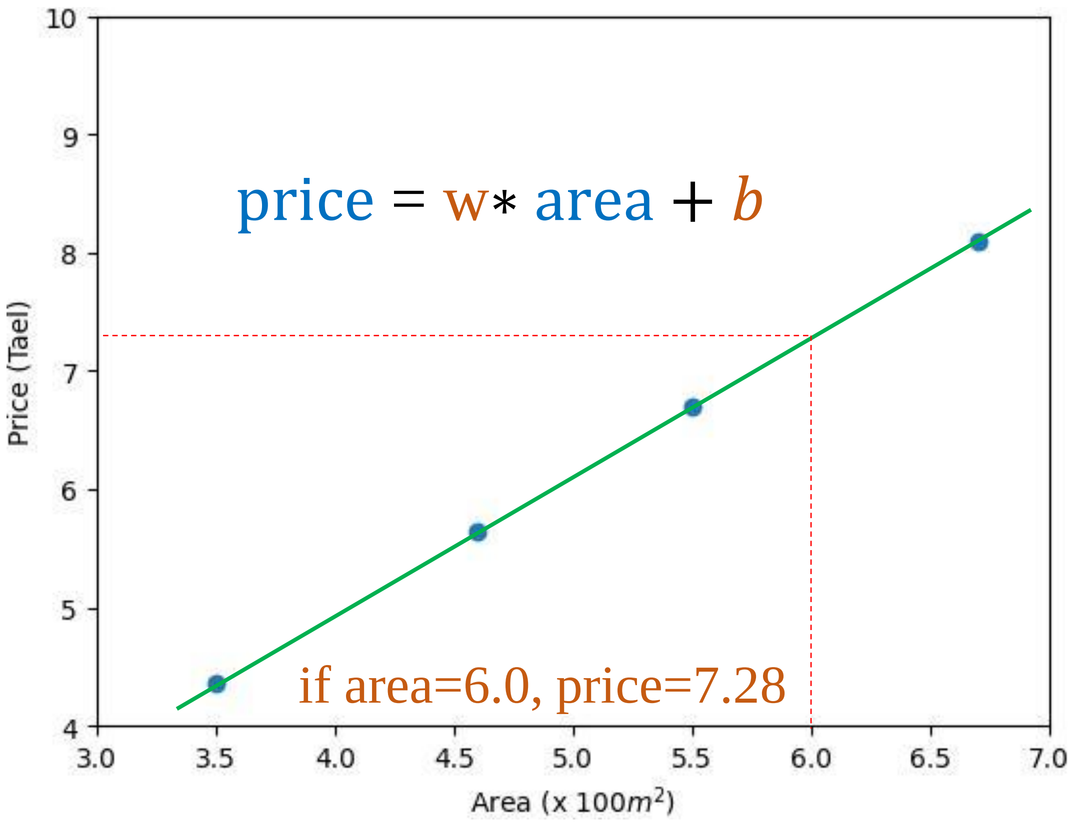
$$= 8x - 16$$

Review Linear Regression

House Price Prediction

House price data

Feature	Label
area	**price**
6.7	8.1
4.6	5.6
3.5	4.3
5.5	6.7

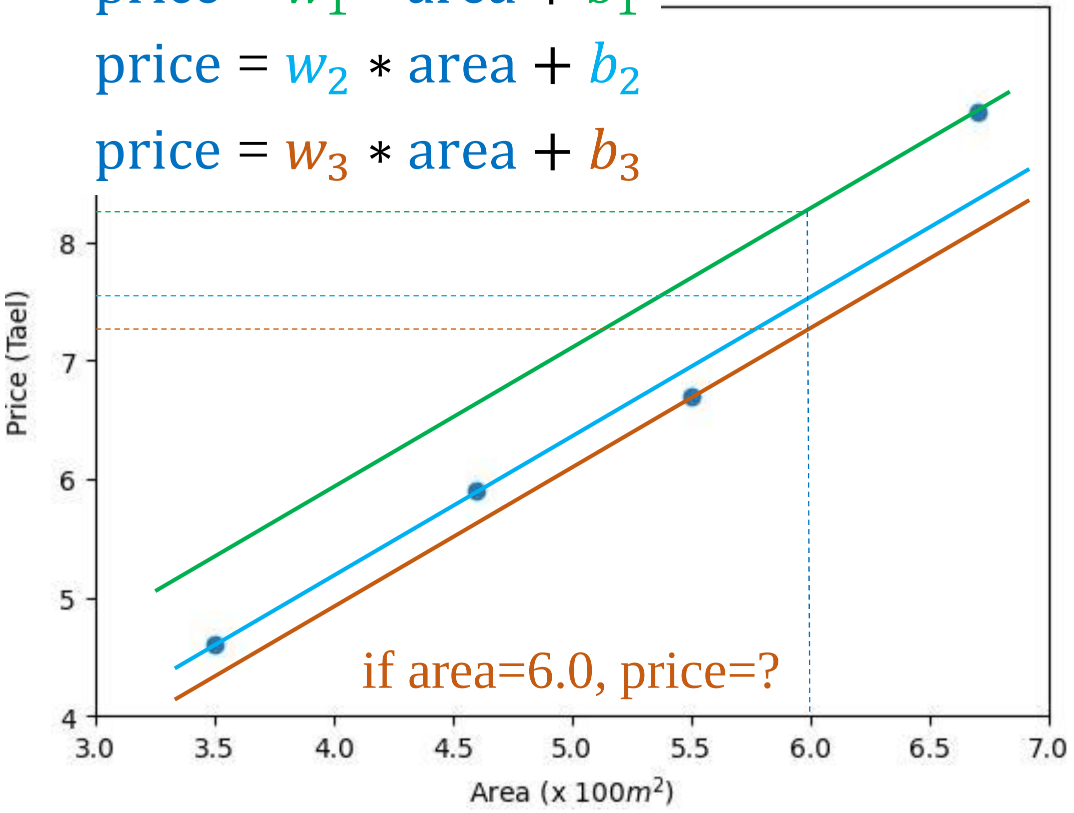


$$\text{price} = w * \text{area} + b$$

if area=6.0, price=7.28

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

House price data



$$\text{price} = w_1 * \text{area} + b_1$$

$$\text{price} = w_2 * \text{area} + b_2$$

$$\text{price} = w_3 * \text{area} + b_3$$

if area=6.0, price=?

Linear Regression

❖ Area-based house price prediction

w = -0.34

b = 0.04

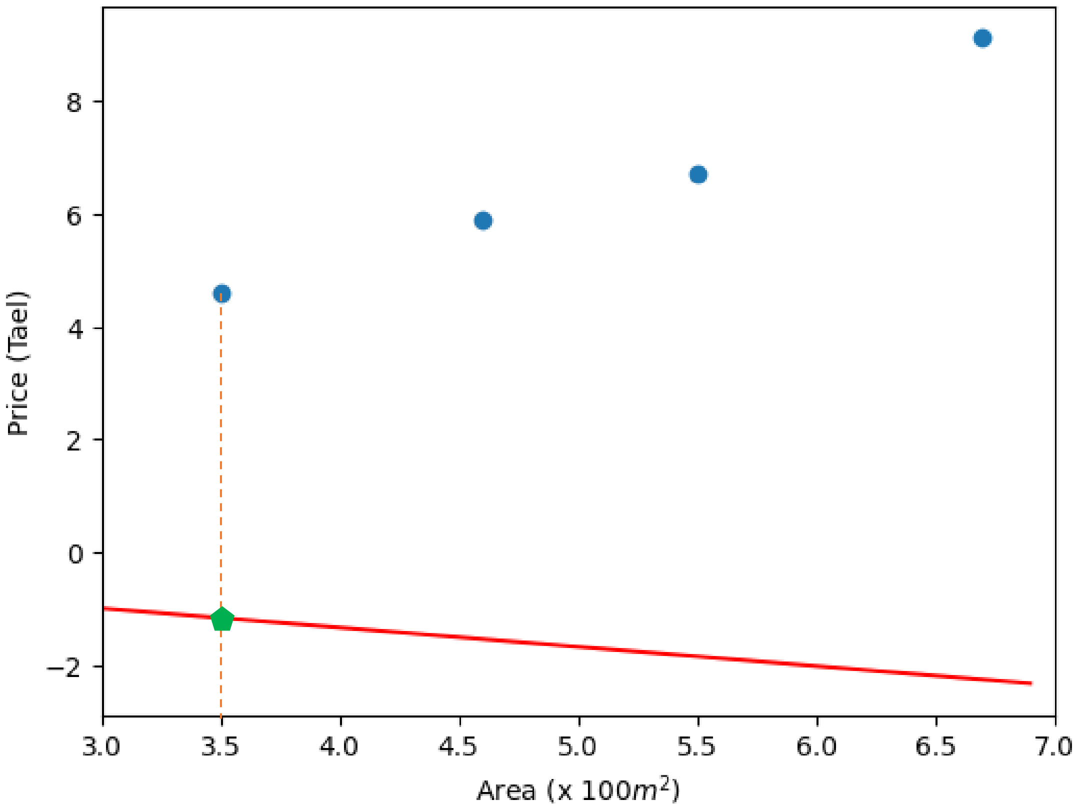
$$\text{predicted\_price} = w * \text{area} + b$$

$$\text{error} = (\text{predicted\_price} - \text{real\_price})^2$$

$$\hat{y} = wx + b$$

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

area	price	predicted	error
6.7	9.1	-2.238	128.55
4.6	5.9	-1.524	55.11
3.5	4.6	-1.15	33.06
5.5	6.7	-1.83	72.76



Linear Regression

❖ Area-based house price prediction

$$w = 1.17$$

$$b = 0.26$$

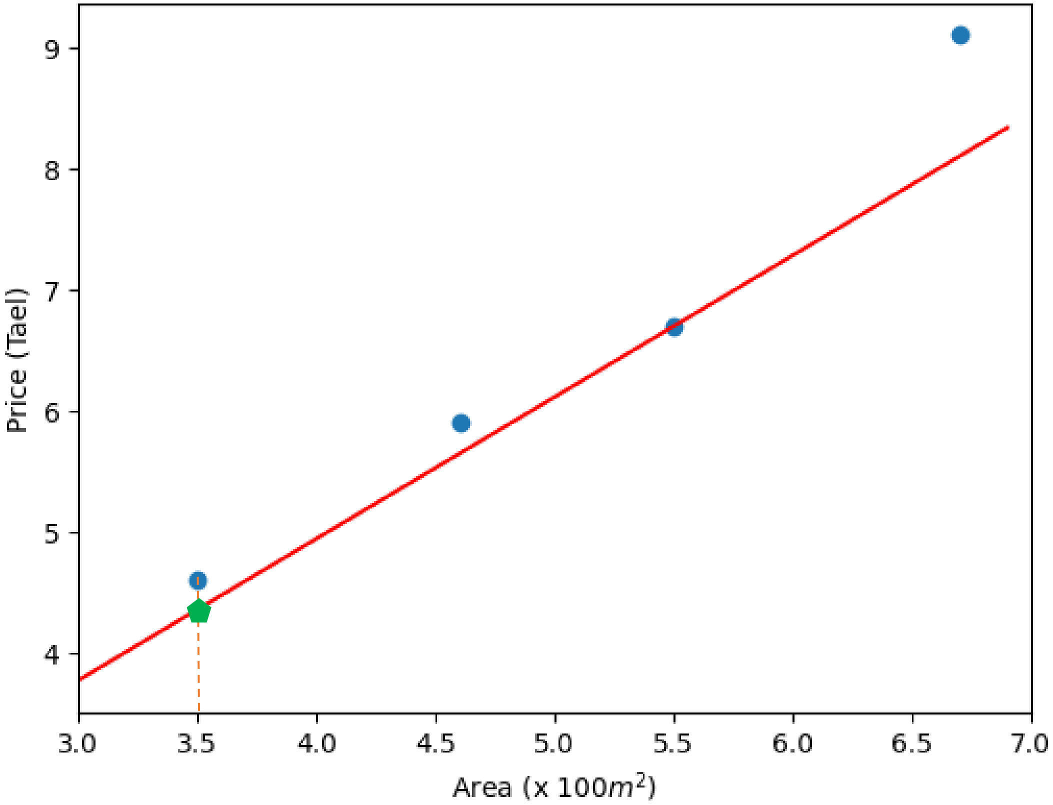
$$\text{predicted\_price} = w * \text{area} + b$$

$$\text{error} = (\text{predicted\_price} - \text{real\_price})^2$$

$$\hat{y} = wx + b$$

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

area	price	predicted	error
6.7	9.1	8.099	1.002
4.6	5.9	5.642	0.066
3.5	4.6	4.355	0.06
5.5	6.7	6.695	0.00002



Linear Regression

Linear equation

$$\hat{y} = wx + b$$

where $\hat{y}$ is a predicted value,

w and b are parameters

and x is input feature

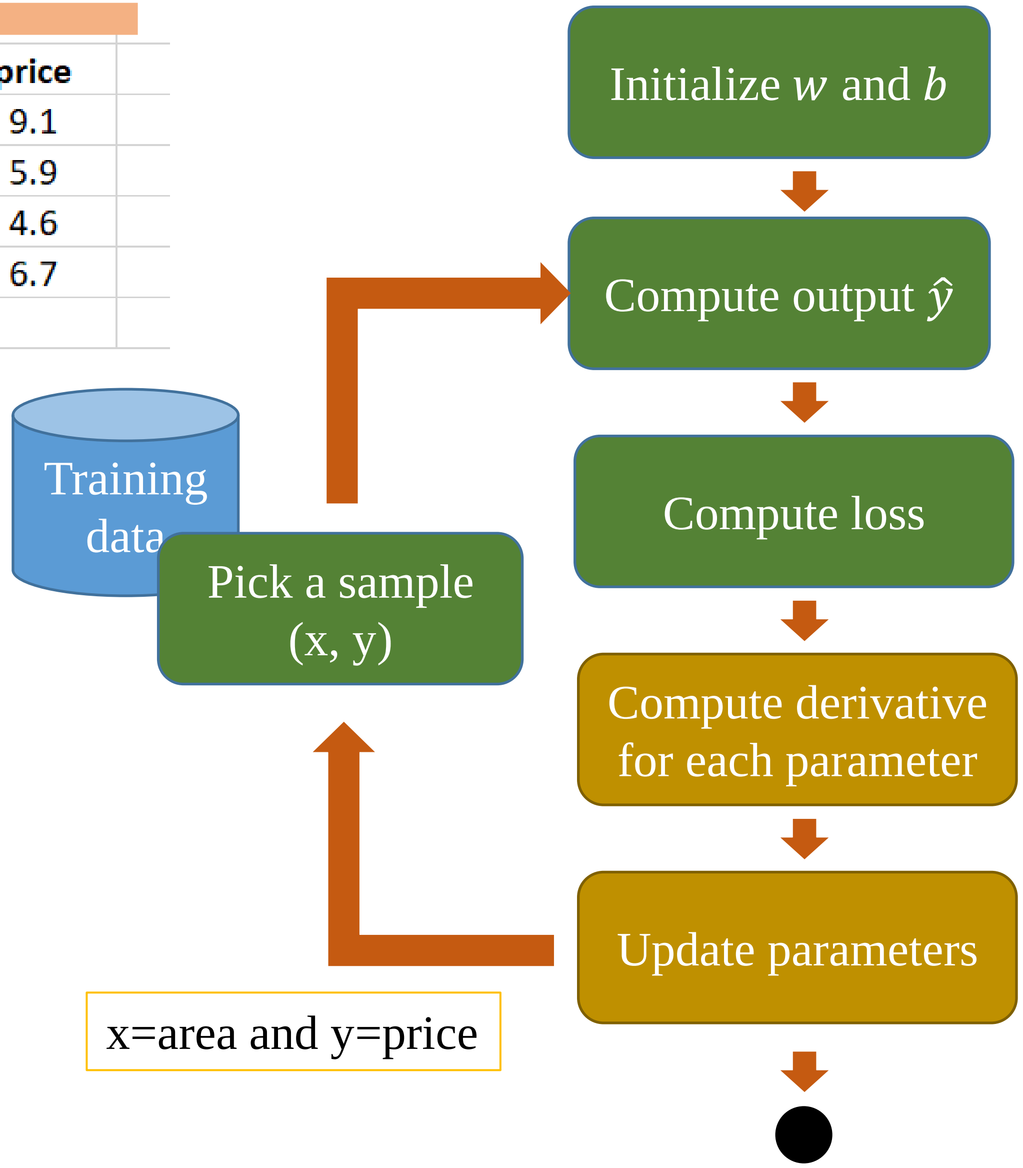
Error (loss) computation

Idea: compare predicted values $\hat{y}$ and label values y

Squared loss

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

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



Training data

Pick a sample (x, y)

Initialize w and b

Compute output $\hat{y}$

Compute loss

Compute derivative for each parameter

Update parameters

x=area and y=price

Linear Regression

Linear equation

$$\hat{y} = wx + b$$

where $\hat{y}$ is a predicted value,

w and b are parameters

and x is input feature

Error (loss) computation

Idea: compare predicted values $\hat{y}$ and label values y

Squared loss

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

Find better w and b

Use gradient descent to minimize the loss function

Compute derivate for each parameter

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w} = 2x(\hat{y} - y)$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial b} = 2(\hat{y} - y)$$

Update parameters

$$w = w - \eta \frac{\partial L}{\partial w} \qquad b = b - \eta \frac{\partial L}{\partial b}$$

η is learning rate

Outline

SECTION 1

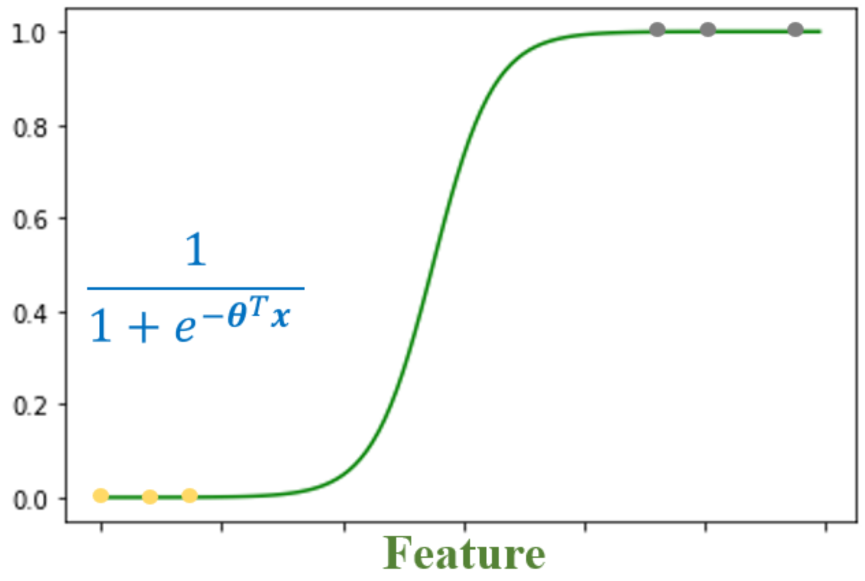
Review

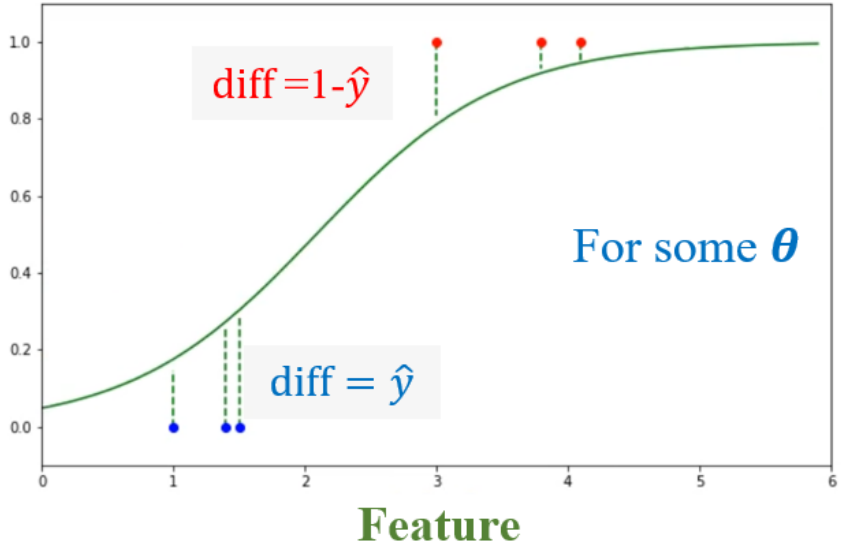
SECTION 2

Lo.R. Using MSE

SECTION 3

Lo.R. Using BCE





Idea of Logistic Regression

❖ Using the approach of 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

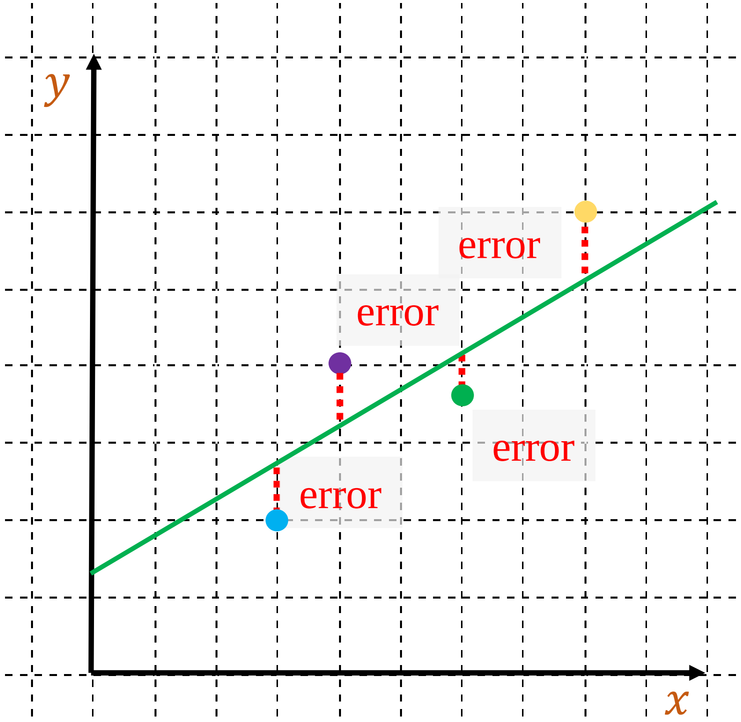
Training data

construct →

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

$$\hat{y} \in (-\infty \quad +\infty)$$

Model



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

Idea of Logistic Regression

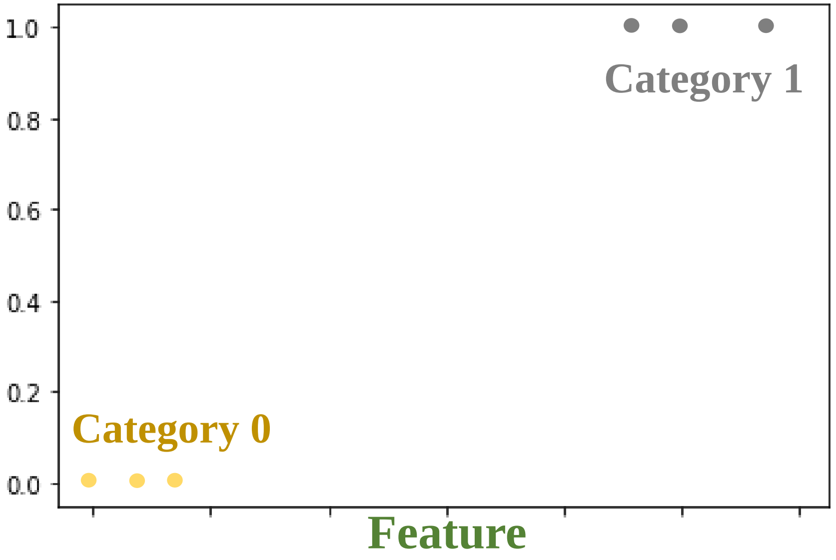
❖ Given a new kind of data

Feature	**Label**	
Petal_Length	Category	
1.4	Flower A	**Category 0**
1	Flower A	
1.5	Flower A	
3	Flower B	**Category 1**
3.8	Flower B	
4.1	Flower B	

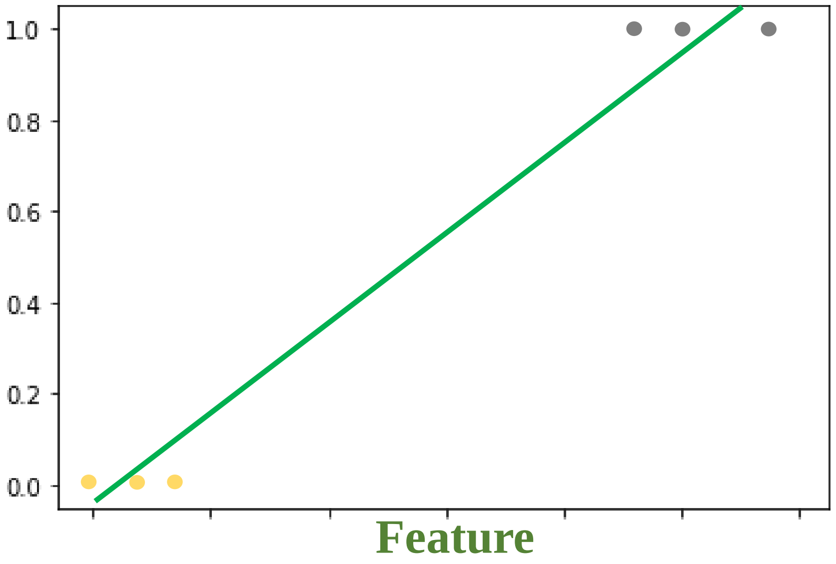
Assign numbers to categories

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

Plot data

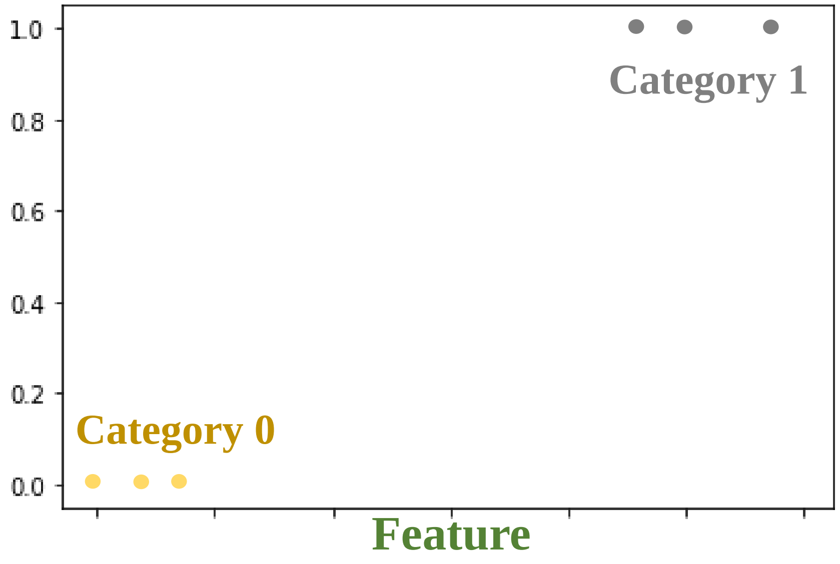


A line is not suitable for this data



What function?

Discussion



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

$$\hat{y} \in (-\infty \quad +\infty)$$

Mô hình Logistic Growth

❖ Look up what science has done!

Phương trình vi phân của mô hình logistic growth có dạng:

$$\frac{dy}{dt} = ry\left(1 - \frac{y}{K}\right)$$

- ✓ $y^{(t)}$: *Kích thước quần thể tại thời gian* t
- ✓ r: Tốc độ tăng trưởng nội tại
- ✓ K: Sức chứa tối đa

Biểu diễn rằng sự thay đổi của quần thể theo thời gian phụ thuộc vào hai yếu tố:

- ✓ Tốc độ tăng trưởng tỉ lệ thuận với kích thước quần thể hiện tại.
- ✓ Hiệu ứng giảm tốc khi kích thước quần thể tiến gần tới giới hạn sức chứa K.

Pierre-François Verhulst (1804–1849)

https://en.wikipedia.org/wiki/Logistic_function

Mô hình Logistic Growth

❖ Giải phương trình vi phân

Đầu tiên, chúng ta sẽ tách biến:

$$\frac{1}{y\left(1-\frac{y}{K}\right)}dy = r\,dt$$

Để giải phương trình này, chúng ta sẽ phân tích mẫu số của tích phân:

$$\frac{1}{y\left(1-\frac{y}{K}\right)} = \frac{1}{y} + \frac{1}{K-y}$$

Nên phương trình trở thành:

$$\left(\frac{1}{y} + \frac{1}{K-y}\right)\mathrm{dy} = r\,\mathrm{dt}$$

Giờ tích phân cả hai vế:

$$\int\left(\frac{1}{y} + \frac{1}{K-y}\right)\mathrm{dy} = \int r\,\mathrm{dt}$$

Tích phân của vế trái là:

$$\ln|y| - \ln|K-y| = rt + C$$

Ở đây, C là hằng số tích phân.

Mô hình Logistic Growth

❖ Đưa về dạng hàm sigmoid

Bây giờ, chúng ta sẽ sắp xếp lại phương trình này để biểu diễn y:

$$\ln\left(\frac{y}{K-y}\right) = rt + C$$

Lấy số mũ của cả hai vế:

$$\frac{y}{K-y} = \mathrm{e}^{rt+c}$$

Đặt $A = \mathrm{e}^{C}$, ta có:

$$\frac{y}{K-y} = A\mathrm{e}^{rt}$$

Từ đây, chúng ta có thể giải y:

$$y = \frac{AK\mathrm{e}^{rt}}{1 + A\mathrm{e}^{rt}}$$

Đặt $A = \frac{y_0}{K-y_0}$, trong đó y_0 là giá trị ban đầu của y khi $t = 0$.
Sau đó, ta có lời giải hoàn chỉnh của phương trình logistic:

$$y(t) = \frac{K}{1 + \left(\frac{K-y_0}{y_0}\right)\mathrm{e}^{-rt}}$$

Mô hình Logistic Growth

❖ Liên hệ với hàm sigmoid

Hàm sigmoid có dạng:

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

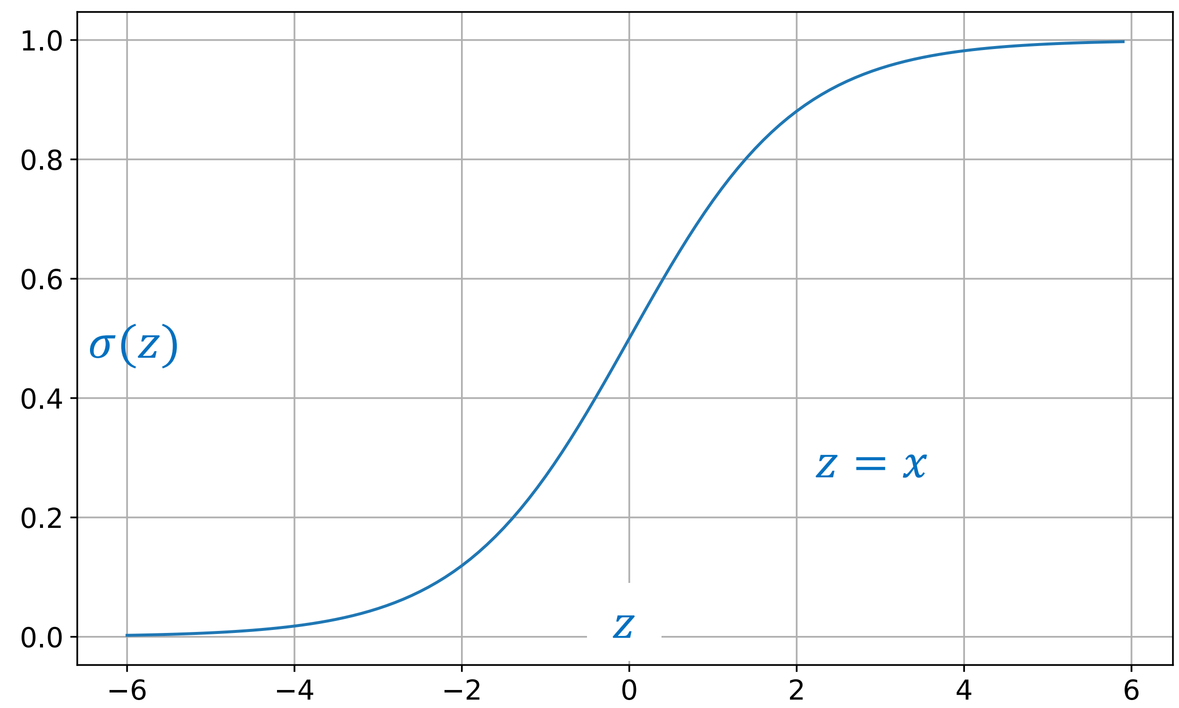
Để thấy sự tương đồng, chúng ta có thể viết lại lời giải của phương trình logistic như sau:

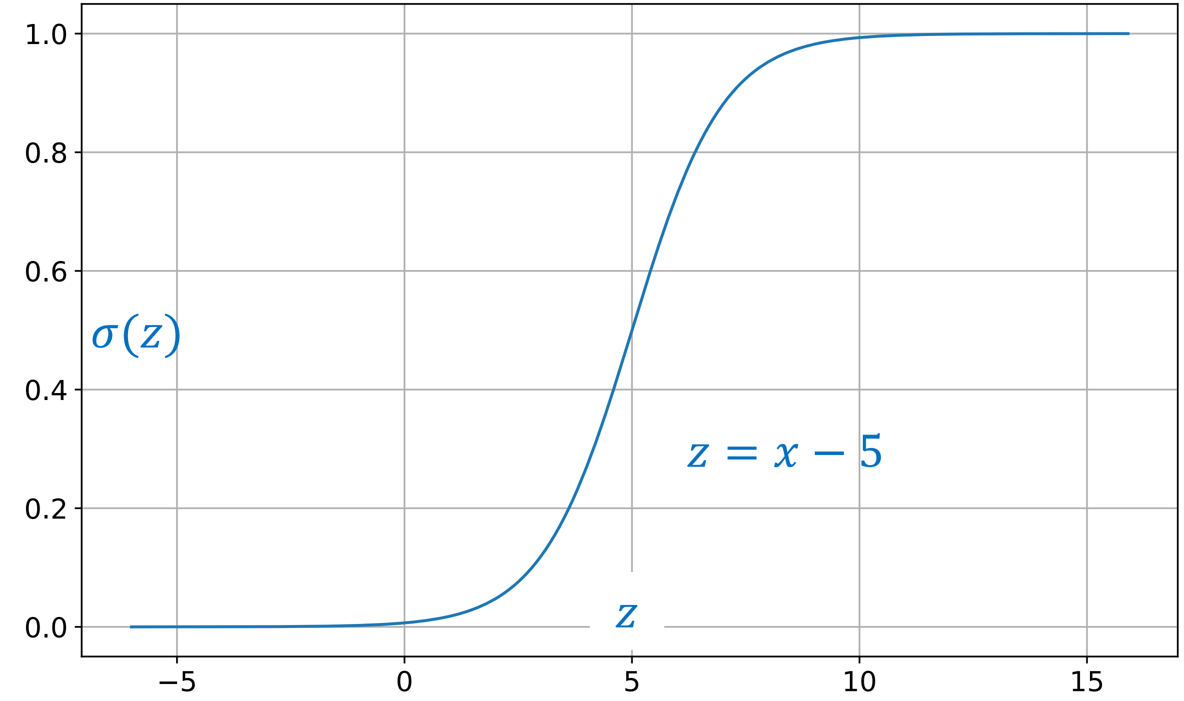
$$y(t) = \frac{K}{1 + \mathrm{e}^{-rt}\left(\frac{K - y_0}{y_0}\right)}$$

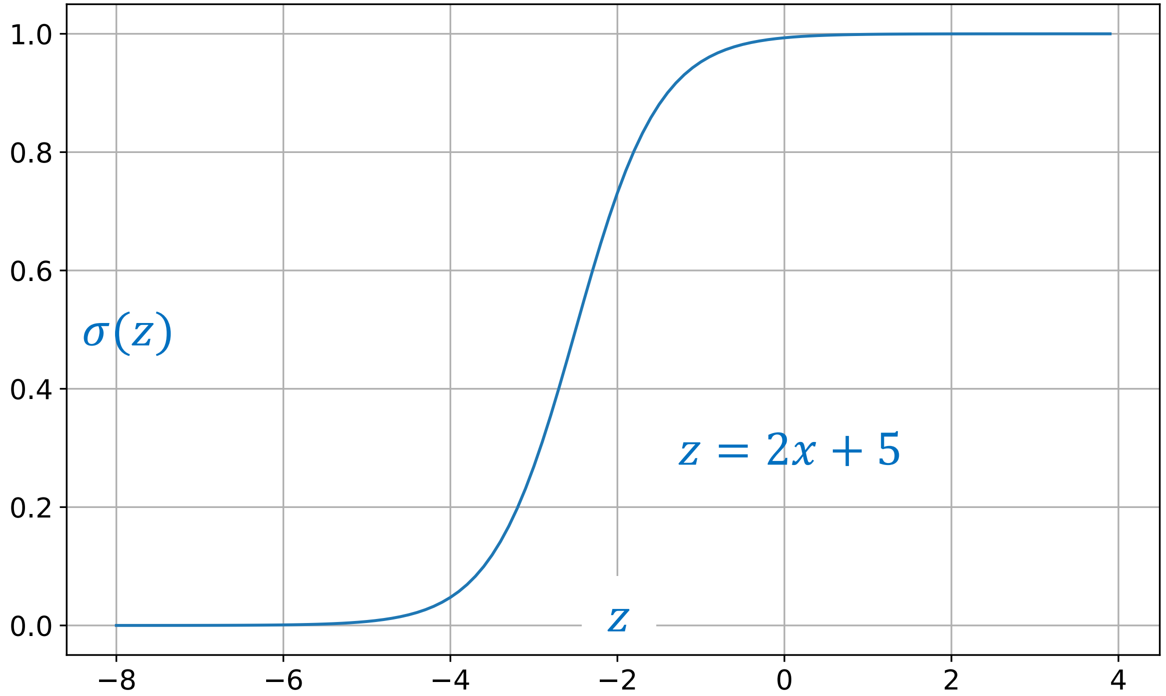
Điều này cho thấy rằng hàm logistic thực chất là một dạng tổng quát của hàm sigmoid với giá trị K là giới hạn trên và rt điều khiển độ dốc.

Sigmoid function

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



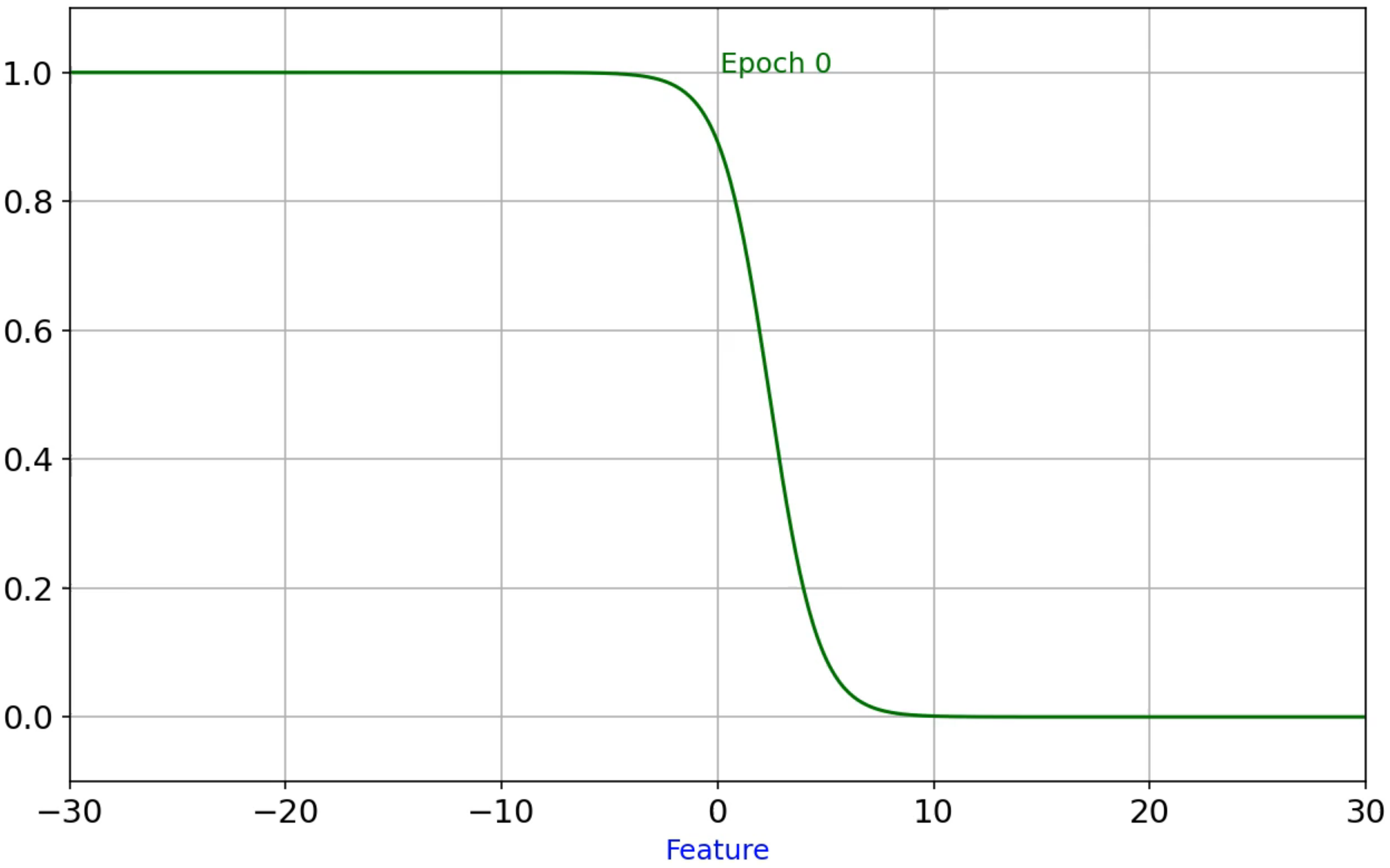




Sigmoid function

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

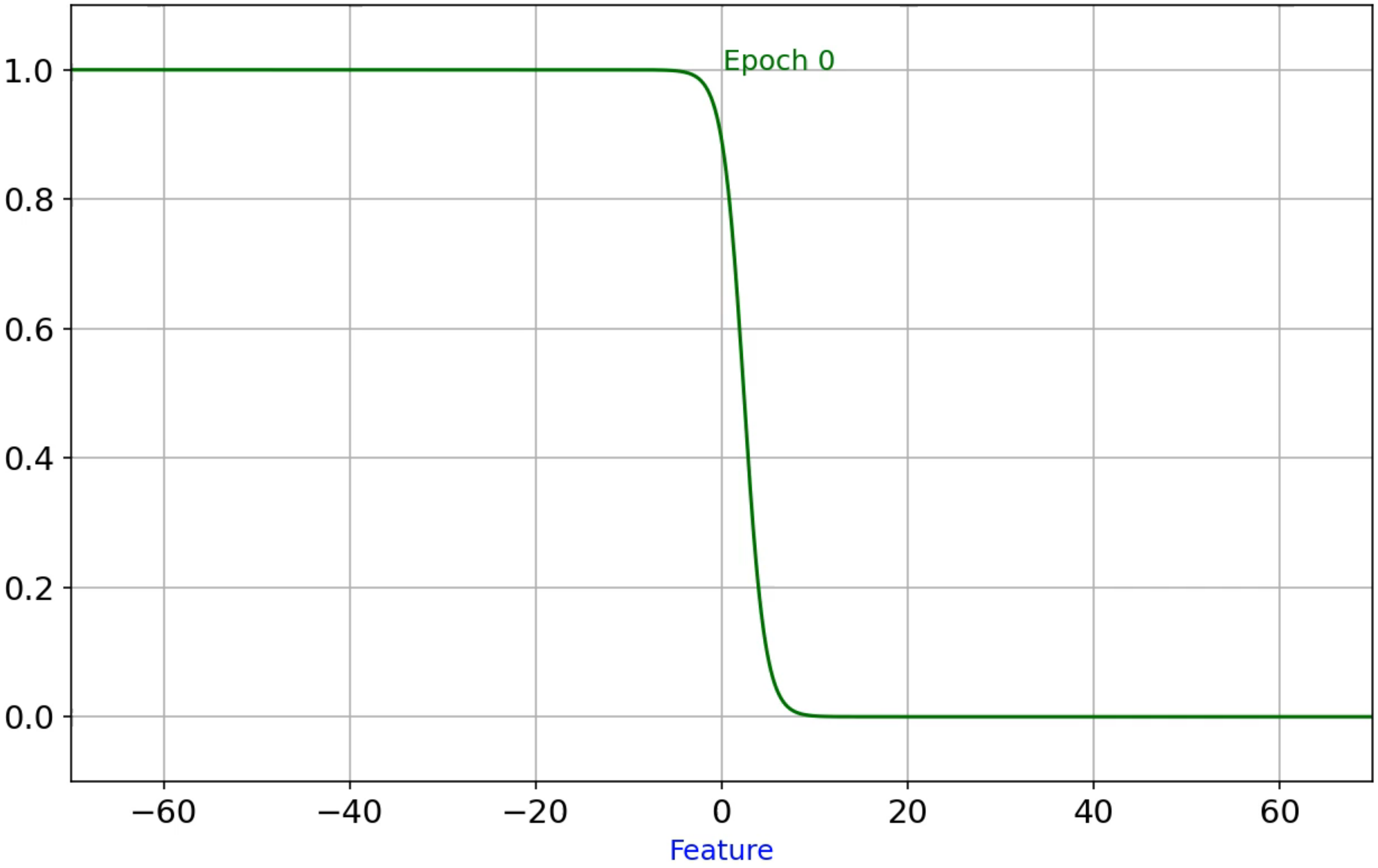
$$z = wx + b$$



Sigmoid function

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

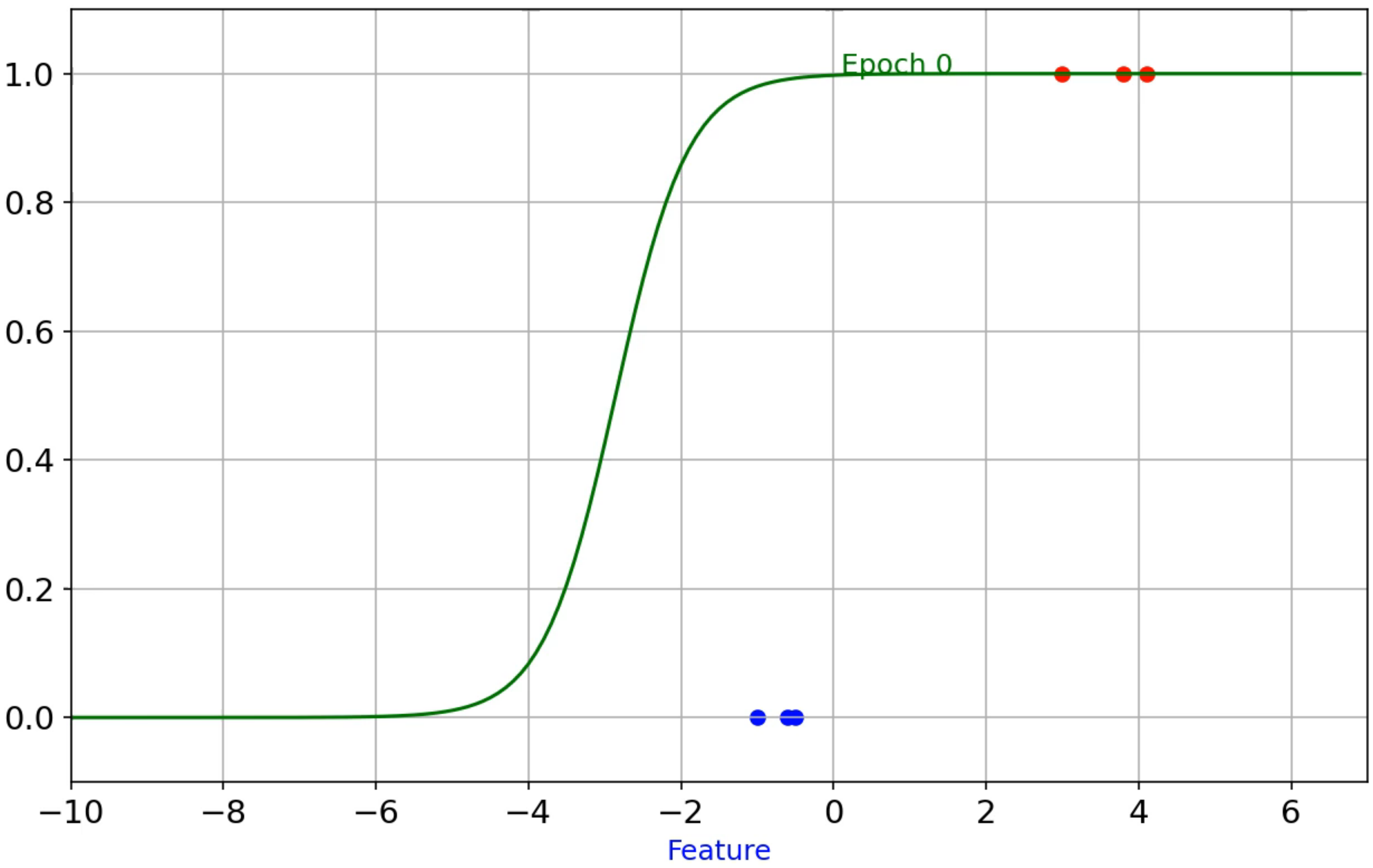
$$z = wx + b$$



Sigmoid function

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

$$z = wx + b$$



Logistic Regression

❖ Using the Sigmoid function

Sigmoid function

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

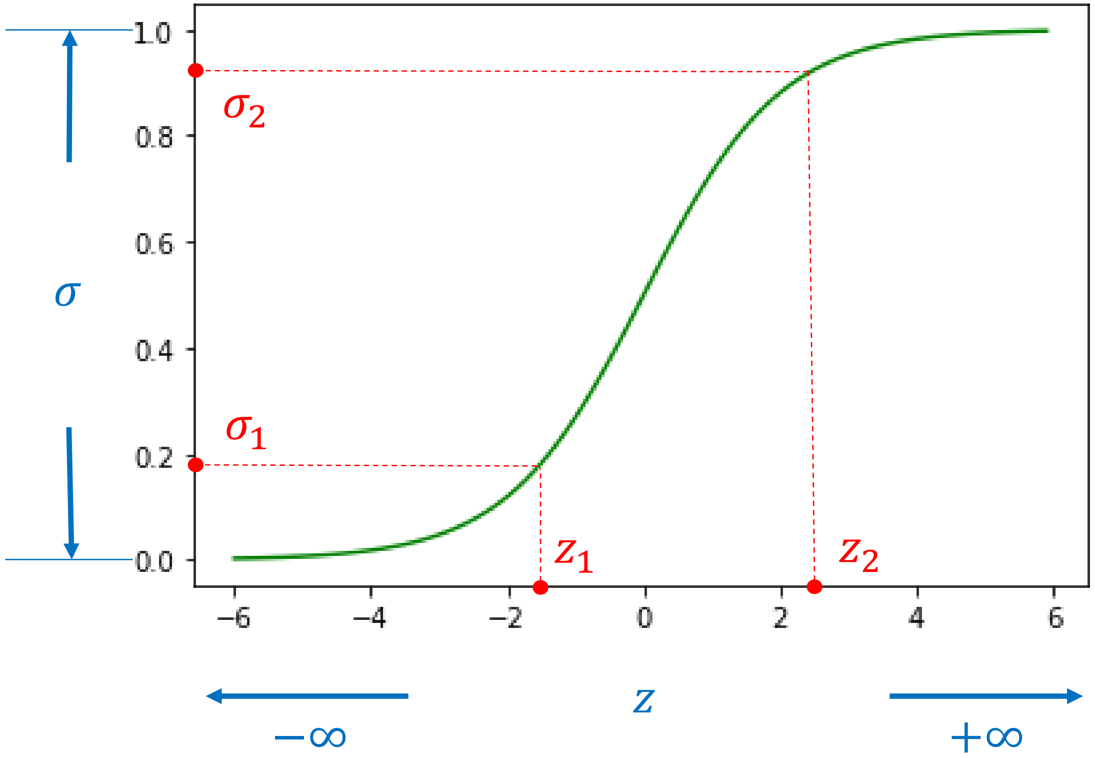
$$z \in (-\infty \quad +\infty)$$

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

Property

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

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



Logistic Regression

❖ Using the Sigmoid function

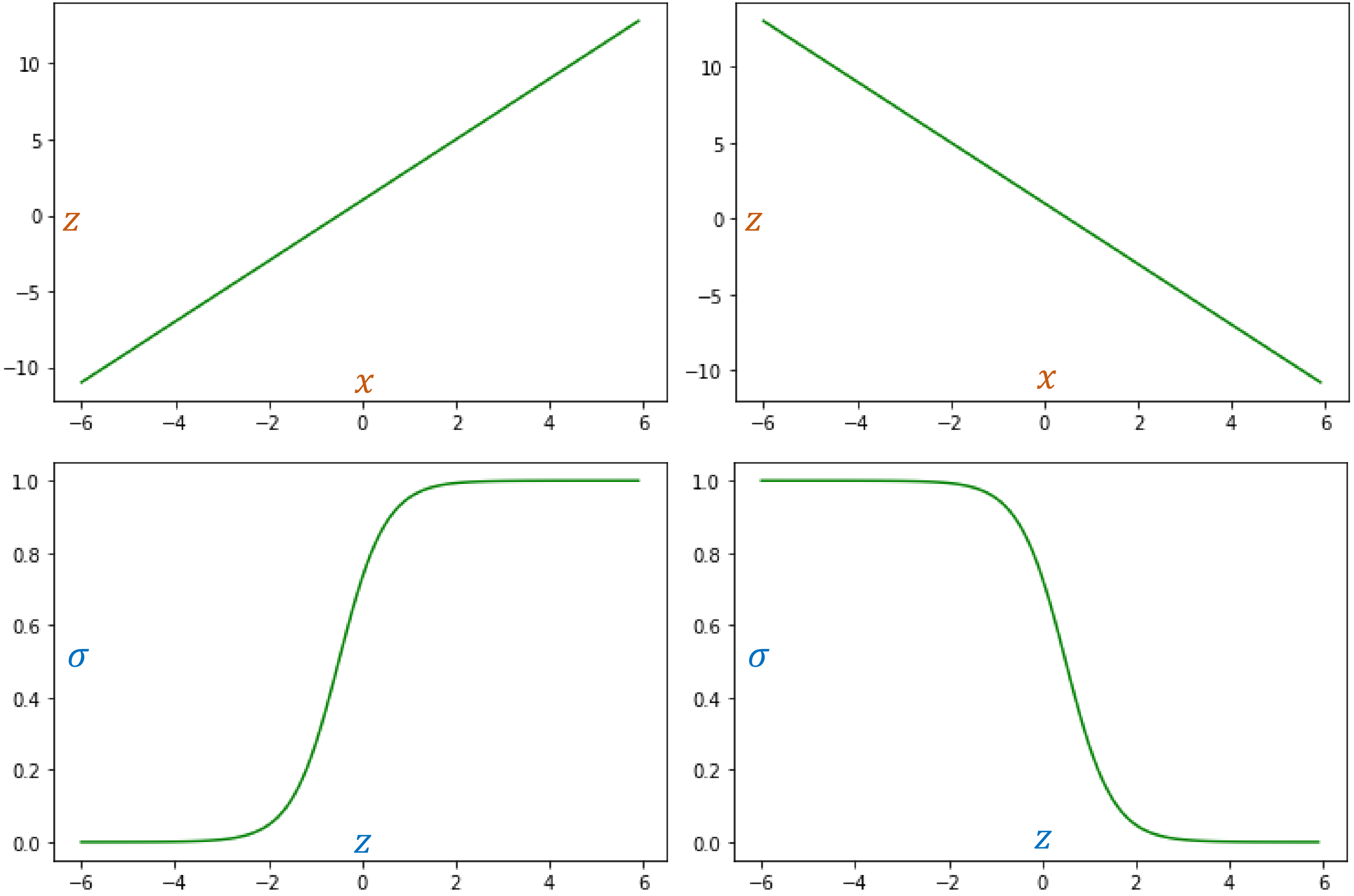
$$z = wx + b$$

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

$$z = wx + b$$

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

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



Logistic Regression

❖ Using the Sigmoid function

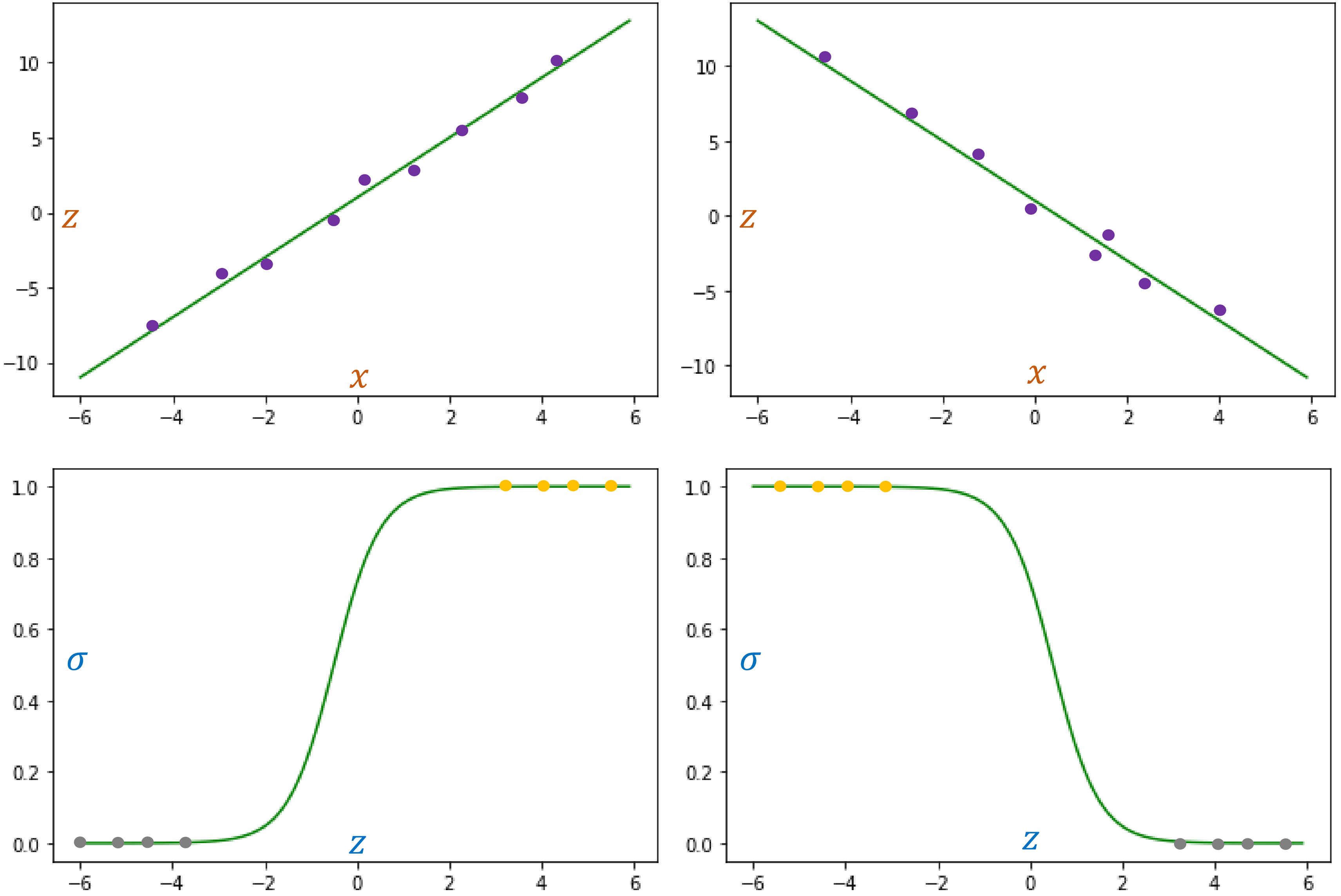
$z = wx + b$

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

$z = wx + b$

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

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



Logistic Regression

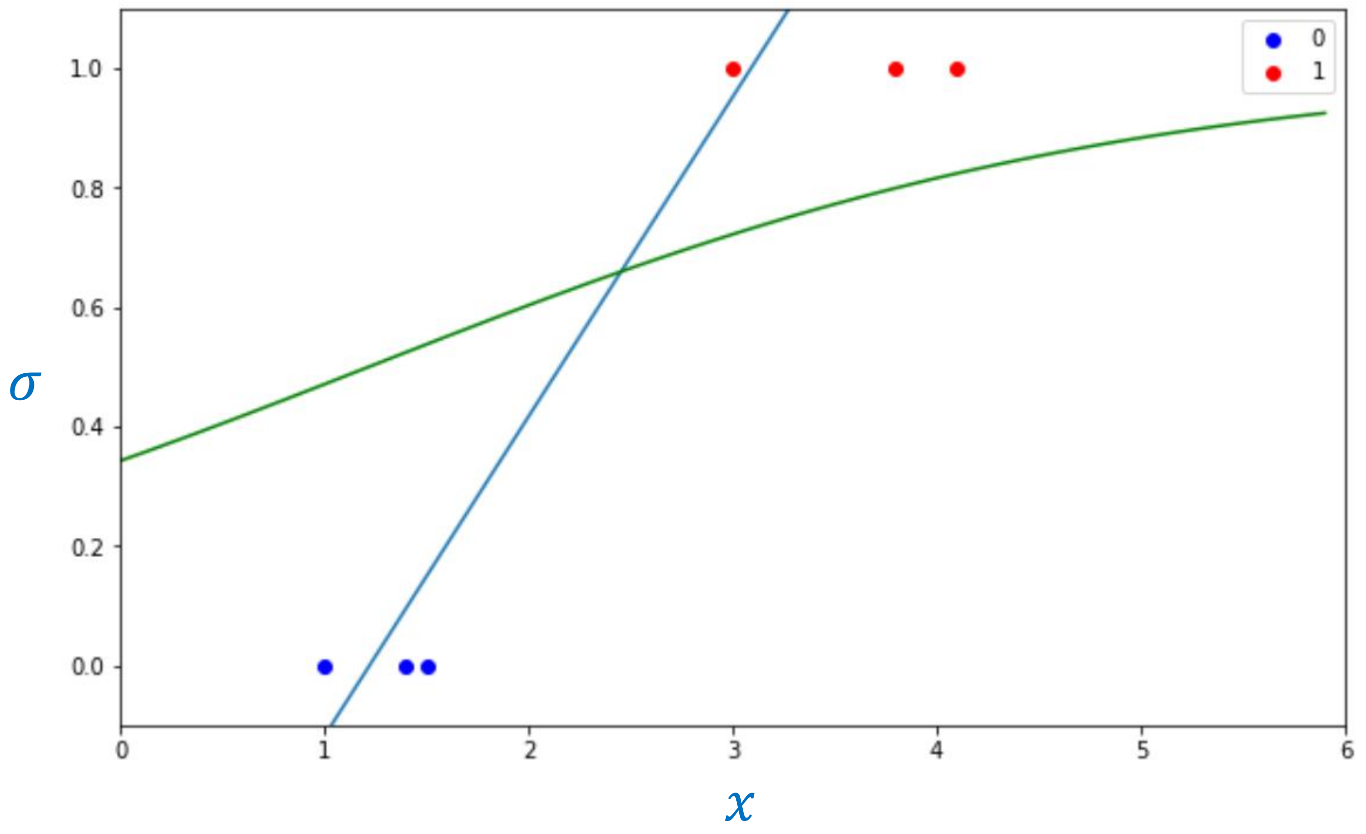
❖ Using the Sigmoid function

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

Category 0

Category 1

z	$\sigma(z)$
0.095	0.52
-0.119	0.47
0.1485	0.53
0.951	0.72
1.379	0.79
1.5395	0.82



$$z = wx + b$$

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

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

$$z = 0.535 * x - 0.654$$

Idea of Logistic Regression

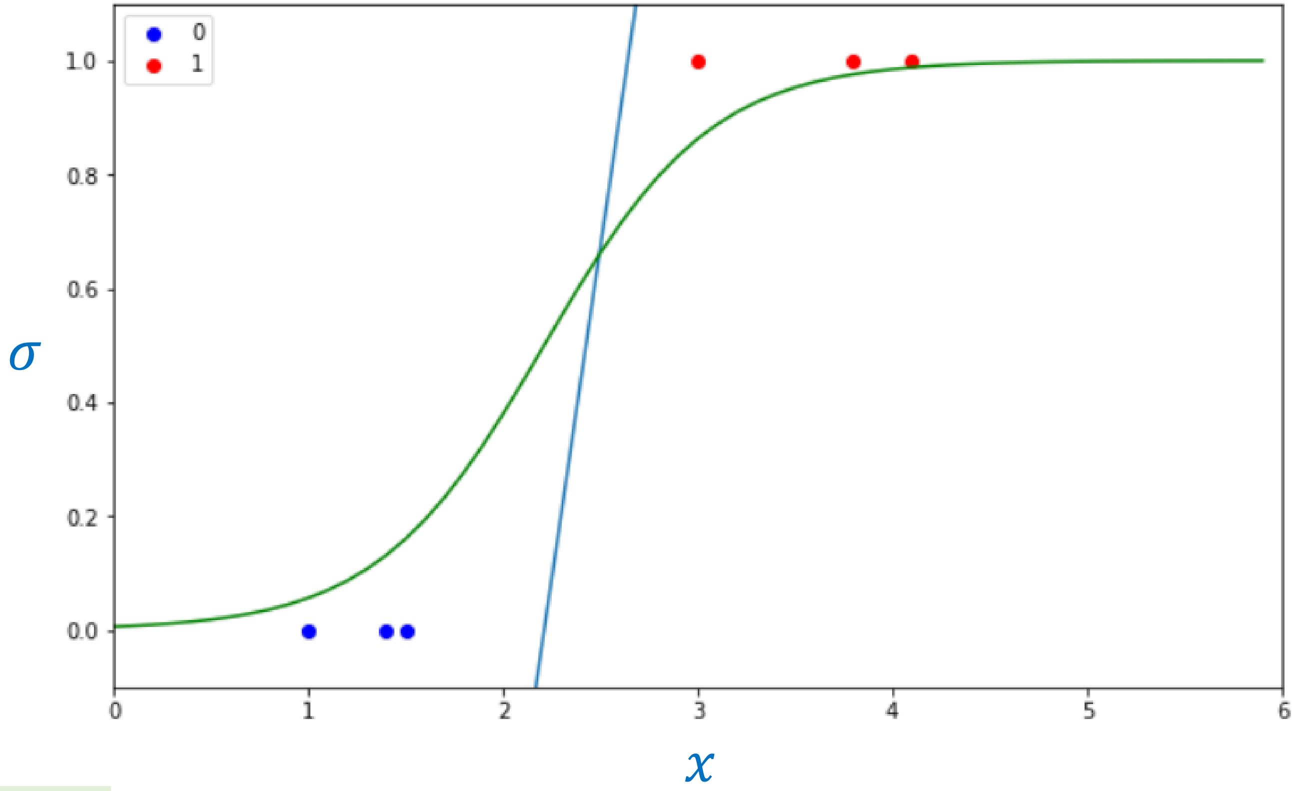
❖ Using the Sigmoid function

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

Category 0

Category 1

z	$\sigma(z)$
-1.89	0.1309
-2.82	0.0559
-1.65	0.1598
1.837	0.8625
3.701	0.9759
4.401	0.9878



$$z = wx + b$$

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

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

$$z = 2.331 * x - 5.156$$

Idea of Logistic Regression

❖ What about the loss function?

Feature	Label	
Petal_Length	Category	
1.4	Flower A	Category 0
1	Flower A	
1.5	Flower A	
3	Flower B	Category 1
3.8	Flower B	
4.1	Flower B	

Assign numbers to categories

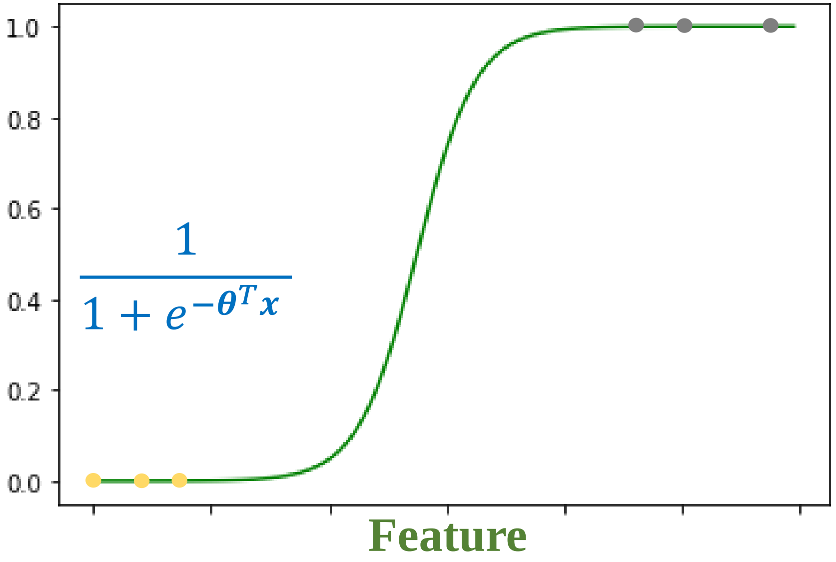
Feature	Label	
Petal_Length	Category	
1.4	0	Category 0
1	0	
1.5	0	
3	1	Category 1
3.8	1	
4.1	1	

Sigmoid function could fit the data

$$z = wx + b$$

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

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



Error

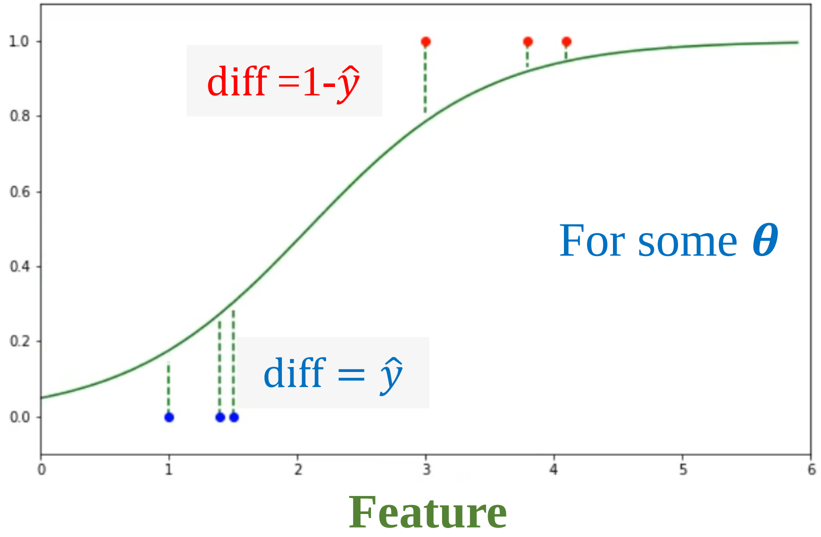
if $y = 1$

$$\text{diff} = 1 - \hat{y}$$

if $y = 0$

$$\text{diff} = \hat{y}$$

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



Logistic Regression-MSE

❖ Construct loss

Model and Loss

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

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

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

Derivative

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

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

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

$$\frac{\partial z}{\partial \theta_i} = x_i$$

$$\frac{\partial L}{\partial \theta_i} = 2x_i(\hat{y} - y)\hat{y}(1 - \hat{y})$$

Logistic Regression-MSE

❖ Implement

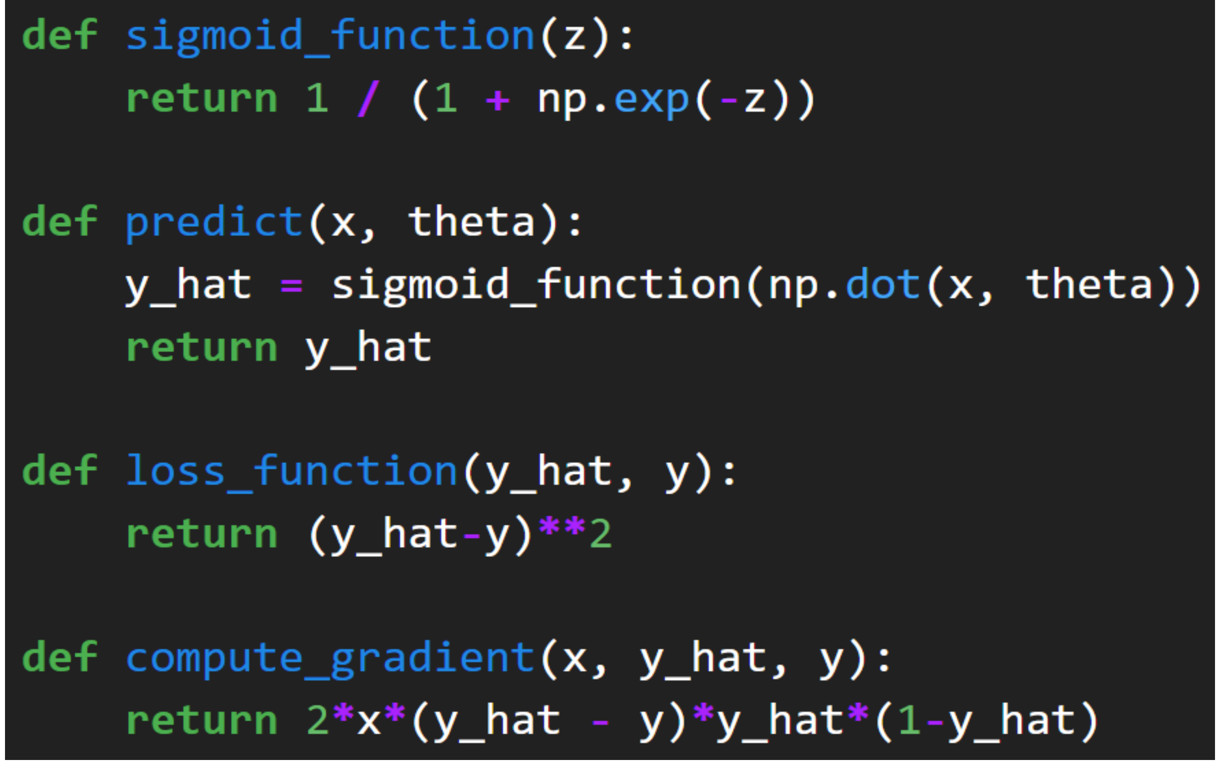
Model and Loss

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

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

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

$$\frac{\partial L}{\partial \theta_i} = 2x_i(\hat{y} - y)\hat{y}(1 - \hat{y})$$

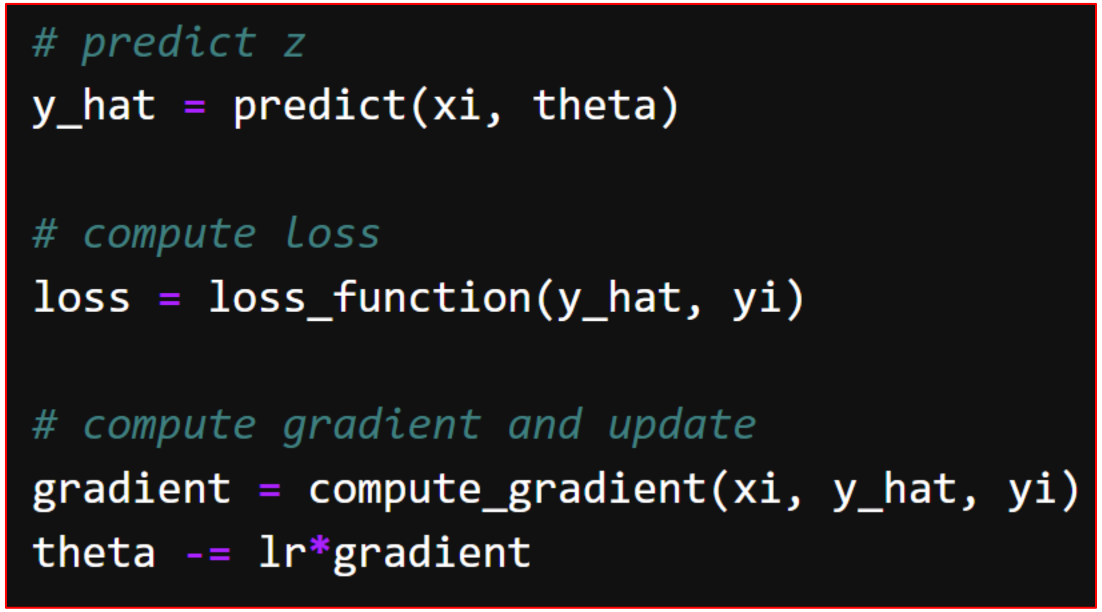


```python
def sigmoid_function(z):
    return 1 / (1 + np.exp(-z))

def predict(x, theta):
    y_hat = sigmoid_function(np.dot(x, theta))
    return y_hat

def loss_function(y_hat, y):
    return (y_hat-y)**2

def compute_gradient(x, y_hat, y):
    return 2*x*(y_hat - y)*y_hat*(1-y_hat)
```



```python
# predict z
y_hat = predict(xi, theta)

# compute loss
loss = loss_function(y_hat, yi)

# compute gradient and update
gradient = compute_gradient(xi, y_hat, yi)
theta -= lr*gradient
```

Logistic Regression-MSE

❖ Result

Done?

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

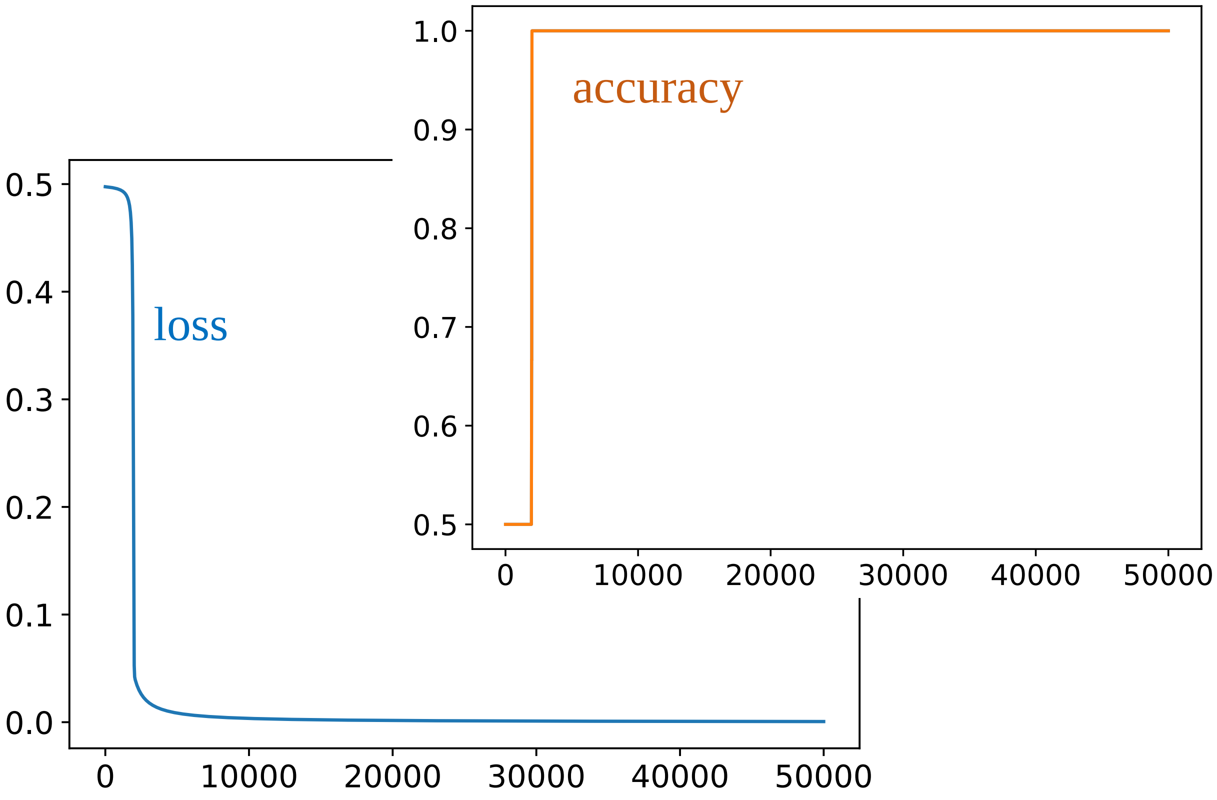
Model and Loss

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

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

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

$$\frac{\partial L}{\partial \theta_i} = 2x_i(\hat{y} - y)\hat{y}(1 - \hat{y})$$



Discussion

Hessian Matrices

❖ Definition

The Hessian matrix or Hessian is a square matrix of second-order partial derivatives of a scalar-valued function

https://en.wikipedia.org/wiki/Hessian_matrix

Given $f(x, y)$

$f: R^2 \rightarrow R$

$$H_f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

Given $f(x, y) = x^2 + 2x^2 y + y^3$

$$\frac{\partial f}{\partial x} = 2x + 4xy$$

$$\frac{\partial f}{\partial y} = 2x^2 + 3y^2$$

$$H_f = \begin{bmatrix} 2 + 4y & 4x \\ 4x & 6y \end{bmatrix}$$

Logistic Regression-MSE

❖ Formulae

Model and Loss

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

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

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

Derivative

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

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y) \qquad \frac{\partial z}{\partial \theta_i} = x_i$$

First-order partial derivatives

$$\frac{\partial L}{\partial \theta_i} = 2x_i(\hat{y} - y)\hat{y}(1 - \hat{y})$$

Mean Squared Error

Model and Loss

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

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

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

Derivative

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

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y) \qquad \frac{\partial z}{\partial \theta_i} = x_i$$

$$\frac{\partial L}{\partial \theta_i} = 2x_i(\hat{y} - y)\hat{y}(1 - \hat{y})$$

$$\frac{\partial L}{\partial \theta_i} = 2x_i(\hat{y} - y)\hat{y}(1 - \hat{y}) = 2x_i\left(-\hat{y}^3 + \hat{y}^2 - y\hat{y} + y\hat{y}^2\right)$$

$$\frac{\partial^2 L}{\partial \theta_i^2} = \frac{\partial}{\partial \theta_i}[2x_i(-\hat{y}^3 + \hat{y}^2 - y\hat{y} + y\hat{y}^2)]$$

$$= 2x_i\left[-3\hat{y}^2 x_i \hat{y}(1 - \hat{y}) + 2x_i \hat{y}\hat{y}(1 - \hat{y}) - y x_i \hat{y}(1 - \hat{y}) + 2x_i y \hat{y}\hat{y}(1 - \hat{y})\right]$$

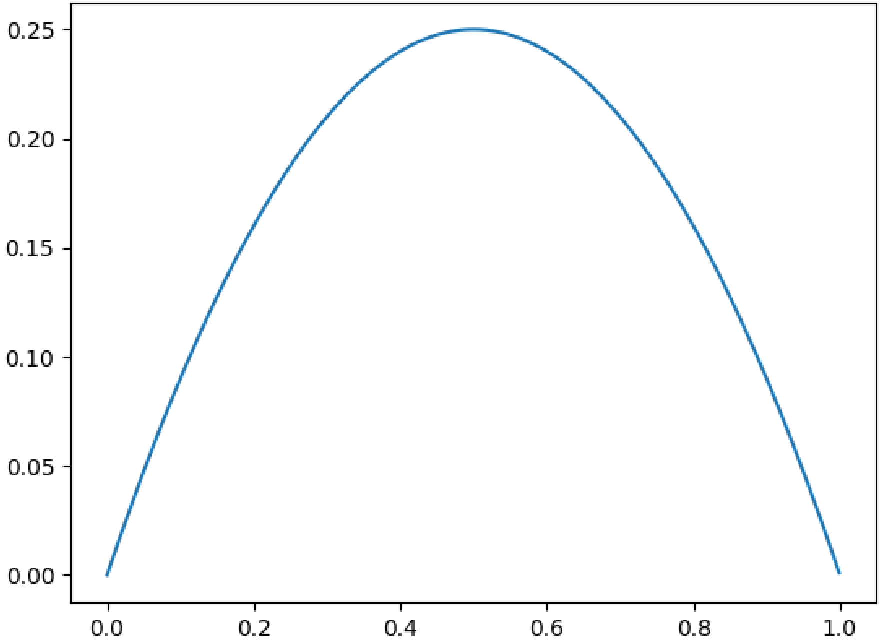
$$= 2x_i^2 \hat{y}(1 - \hat{y})\left[-3\hat{y}^2 + 2\hat{y} - y + 2y\hat{y}\right]$$

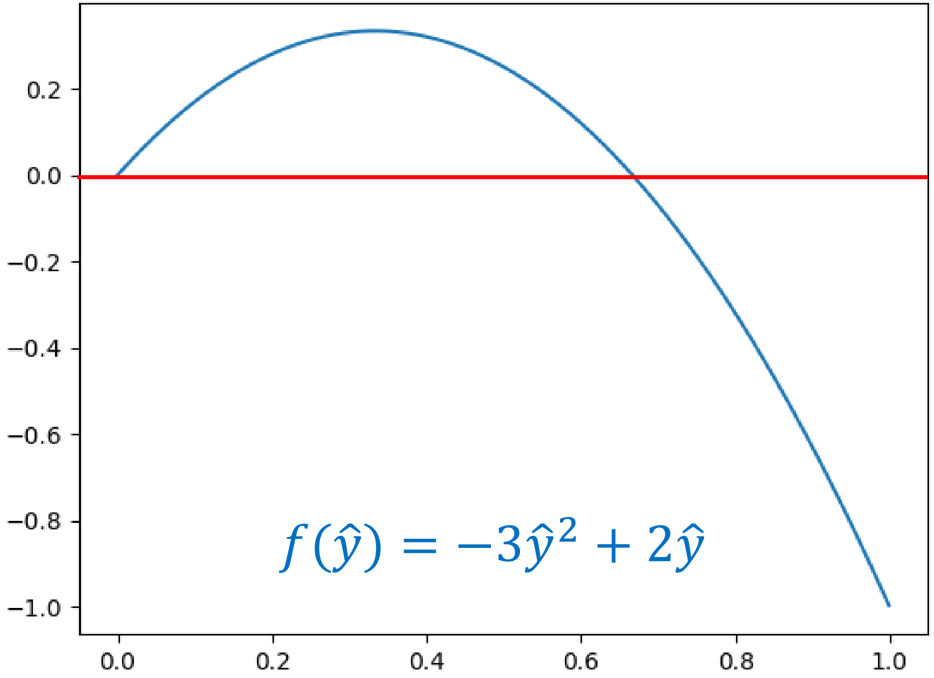
Mean Squared Error

$$\frac{\partial^2 L}{\partial \theta_i^2} = 2x_i^2 \hat{y}(1-\hat{y})[-3\hat{y}^2 + 2\hat{y} - y + 2y\hat{y}]$$

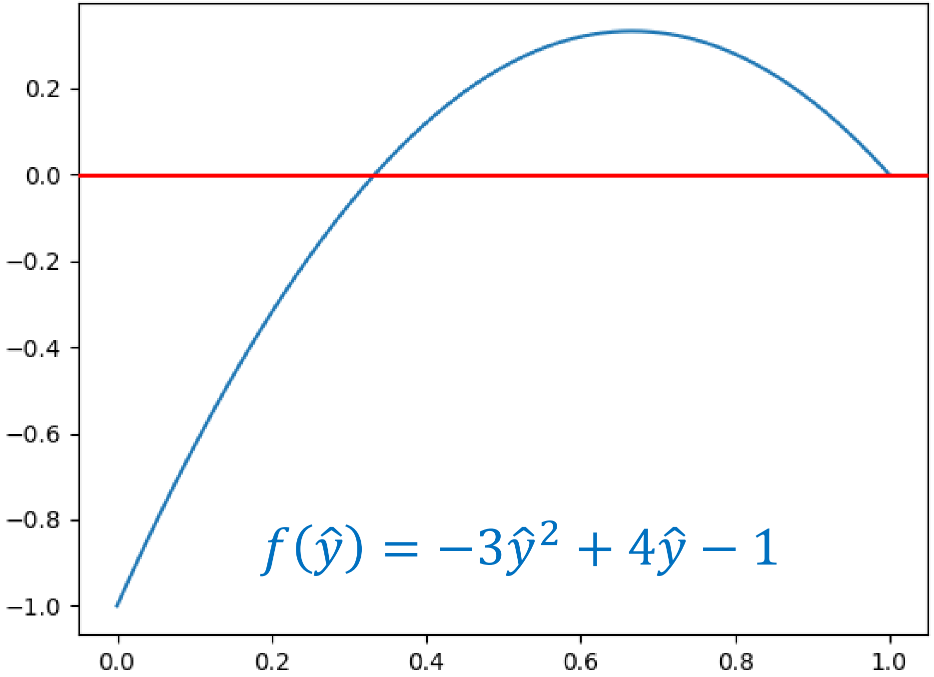
$x_i^2 \geq 0$

$\hat{y}(1-\hat{y}) \in \left[0, \frac{1}{4}\right]$





$f(\hat{y}) = -3\hat{y}^2 + 2\hat{y}$



$f(\hat{y}) = -3\hat{y}^2 + 4\hat{y} - 1$

$y = 0$

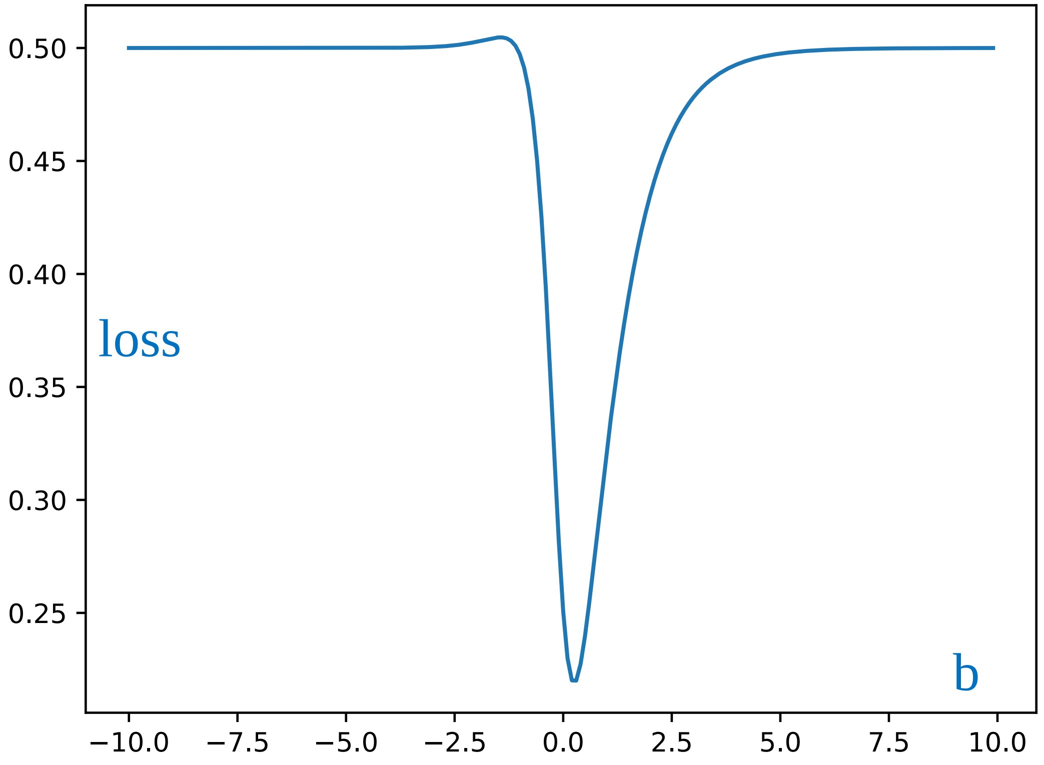
$f(\hat{y}) = -3\hat{y}^2 + 2\hat{y}$

$y = 1$

$f(\hat{y}) = -3\hat{y}^2 + 4\hat{y} - 1$

Mean Squared Error

❖ Visualization



Mean Squared Error

Model and Loss

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

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

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

$$\frac{\partial L}{\partial \theta_i} = 2x_i(\hat{y} - y)\hat{y}(1 - \hat{y})$$

Outline

SECTION 1

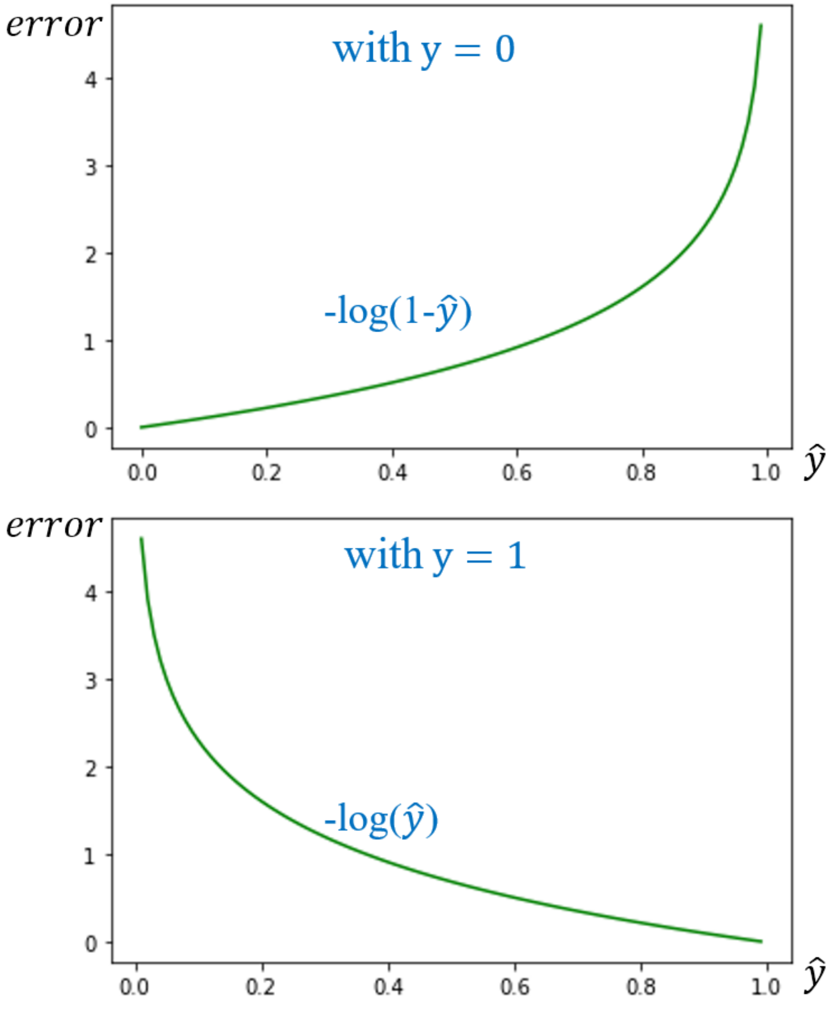
Review

SECTION 2

Lo.R. Using MSE

SECTION 3

Lo.R. Using BCE



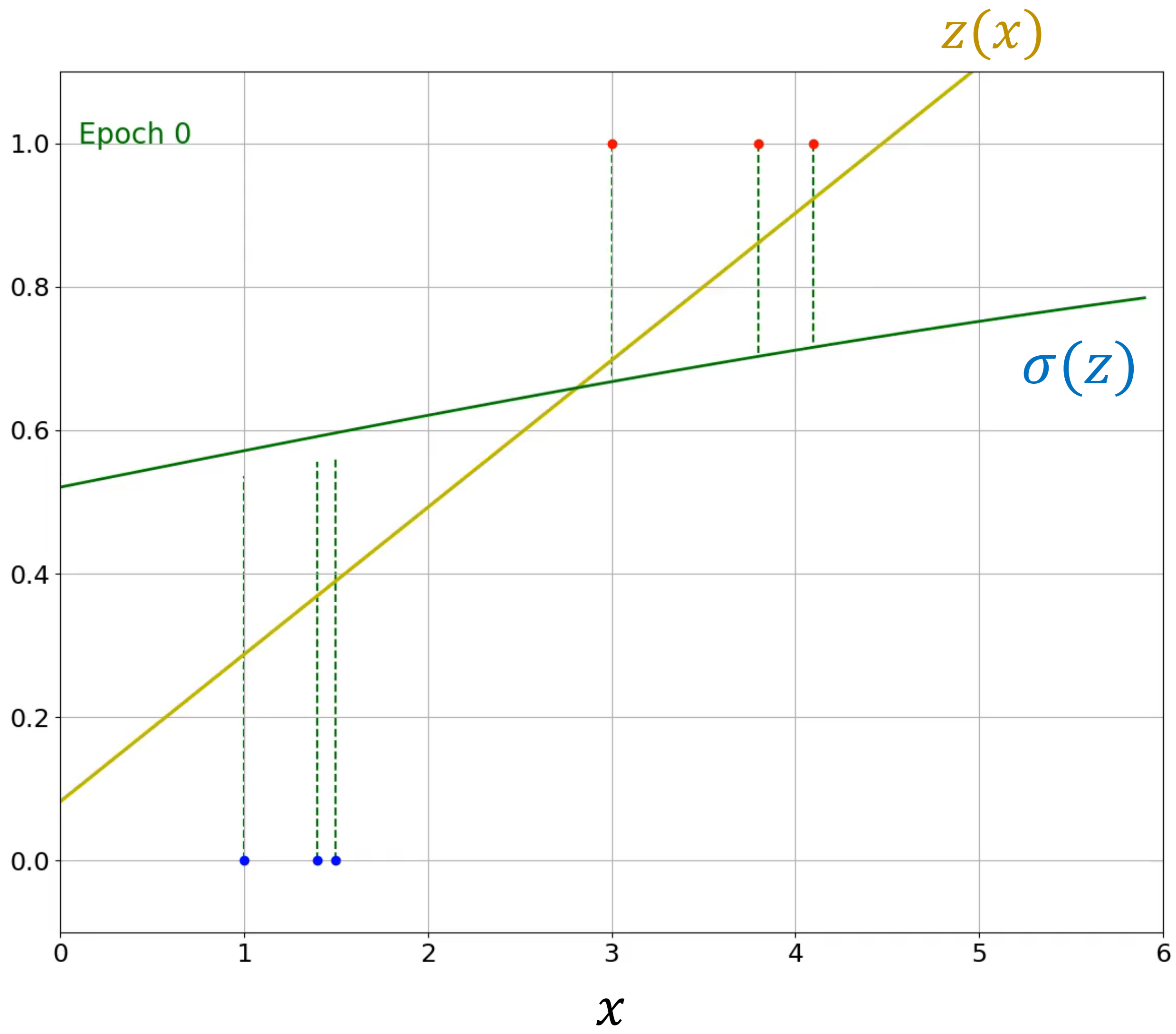
Rethinking Logistic Regression

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

$$z = wx + b$$

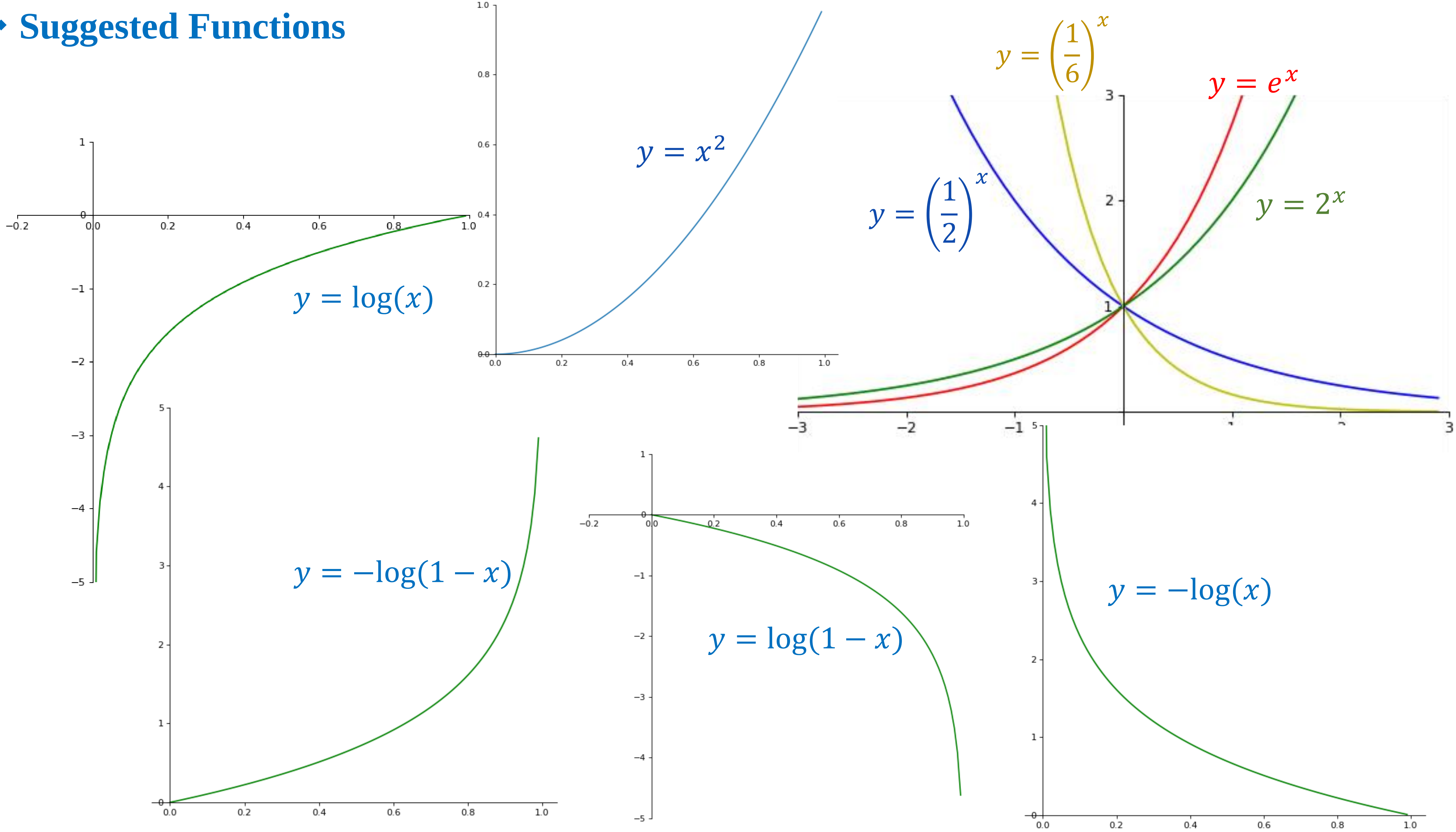
$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

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



How to evaluate the performance of a model?

❖ Suggested Functions



Rethinking Logistic Regression

❖ Loss function

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

Category 0

Category 1

$$z = wx + b$$

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

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

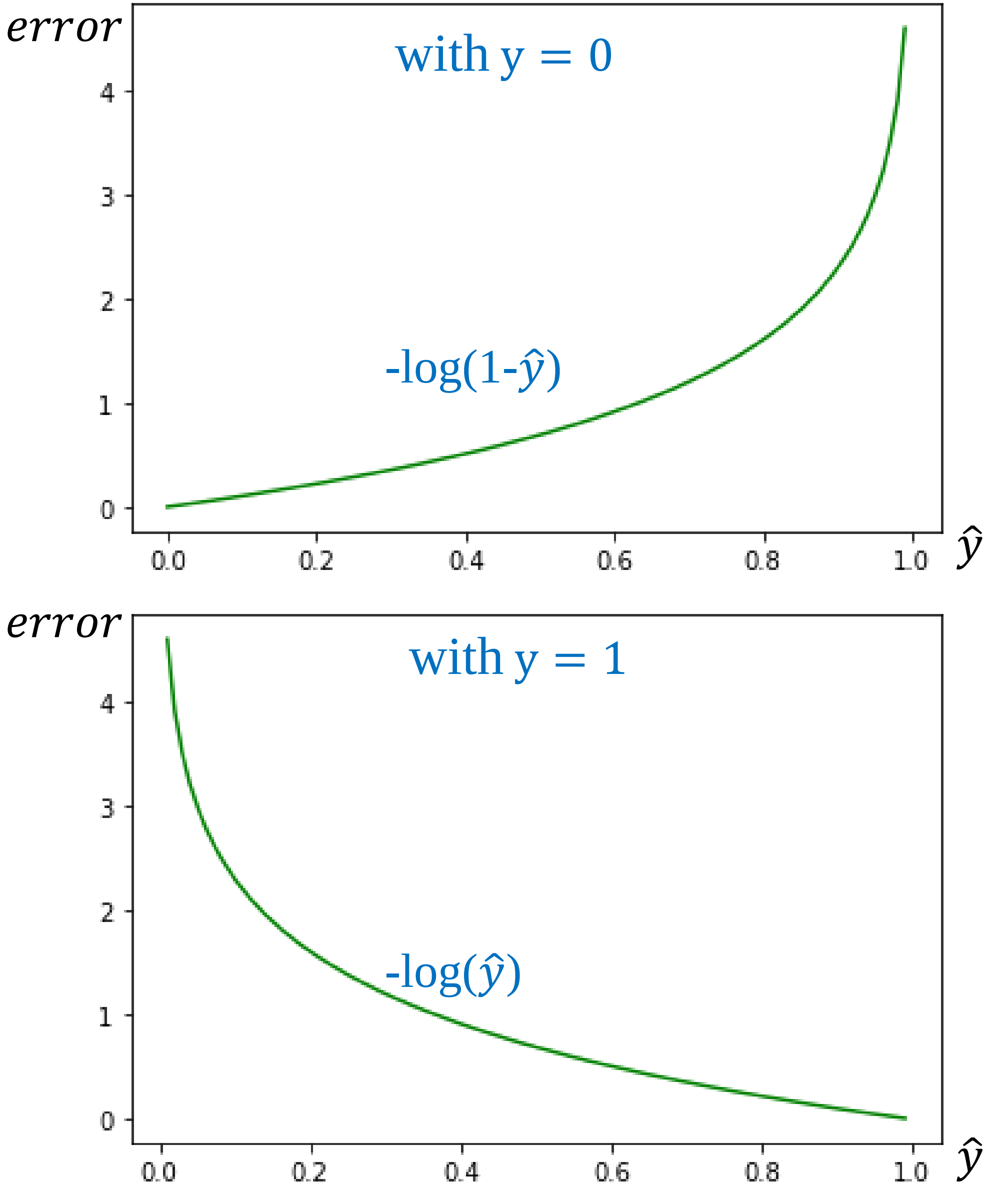
if y = 1

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

if y = 0

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

How to remove if?



Idea of Logistic Regression

❖ Loss function

Feature	Output	Label
Input	**Output**	**Label**
…	0.3	0
…	0.8	0
…	0.7	0
…	0.4	0
…	0.6	1
…	0.8	1
…	0.9	1
…	0.2	1

if $y = 0$

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

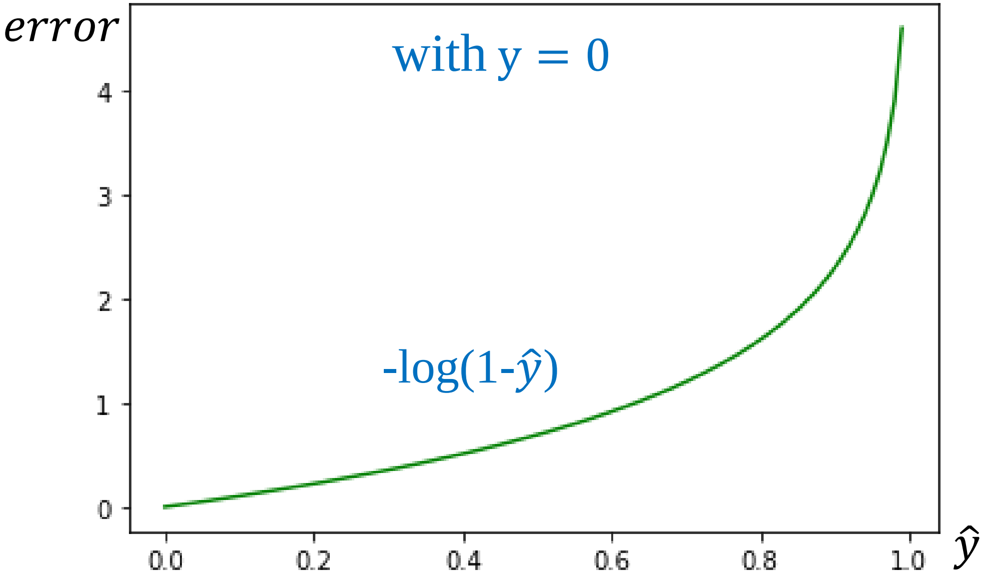
if $y = 1$

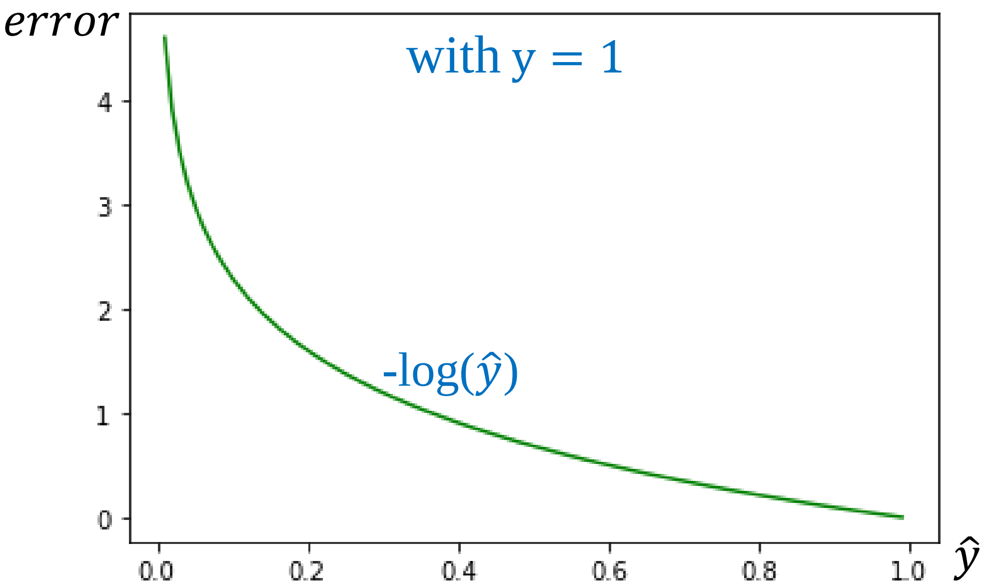
$$L(\hat{y}) = -\log(\hat{y})$$

one sample

Binary cross-entropy

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





Idea of Logistic Regression

❖ Construct loss

Model and Loss

$$z = wx + b$$

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

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

Derivative

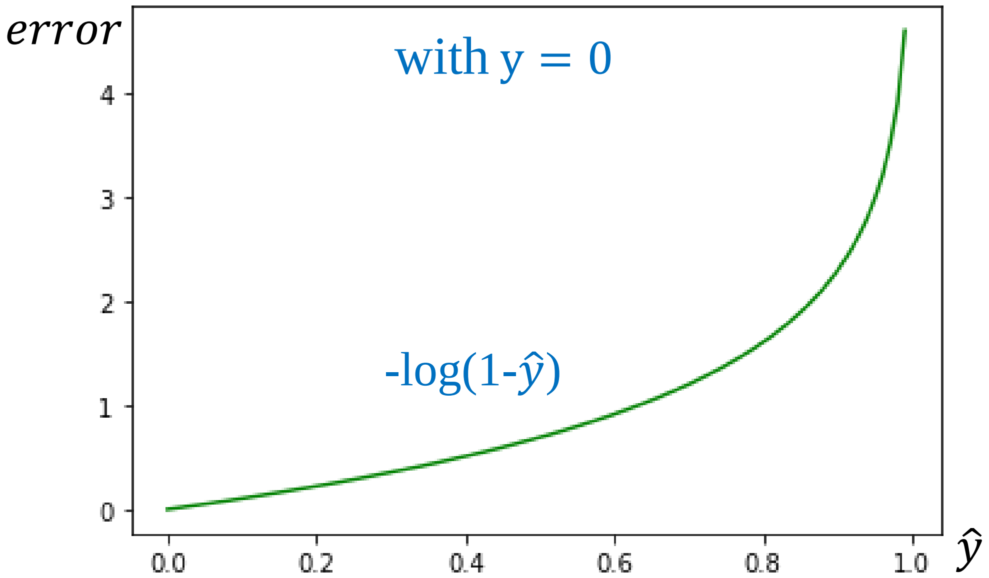
$$\frac{\partial L}{\partial \theta_i} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z} \frac{\partial z}{\partial \theta_i}$$

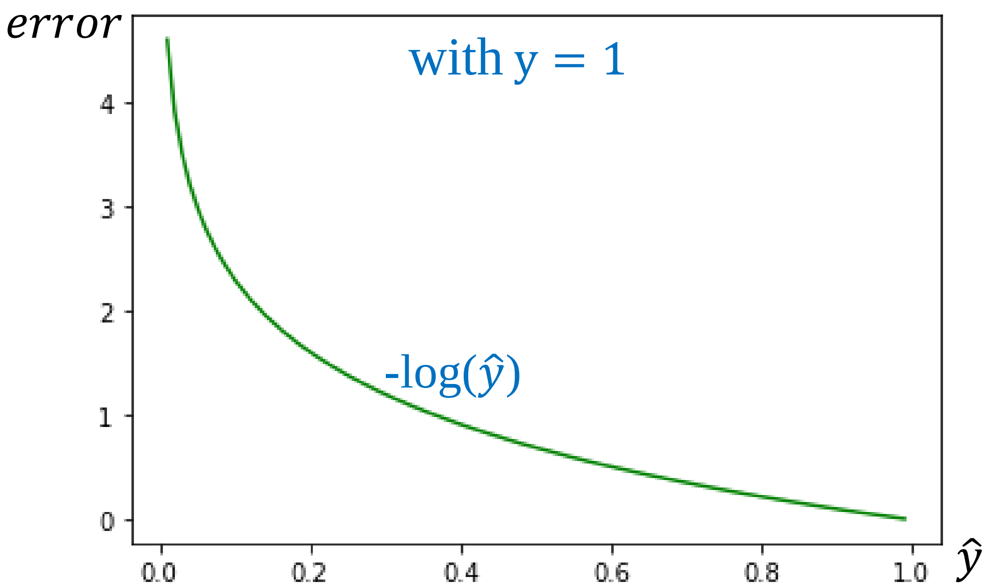
$$\frac{\partial L}{\partial \hat{y}} = -\frac{y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}} = \frac{\hat{y} - y}{\hat{y}(1 - \hat{y})}$$

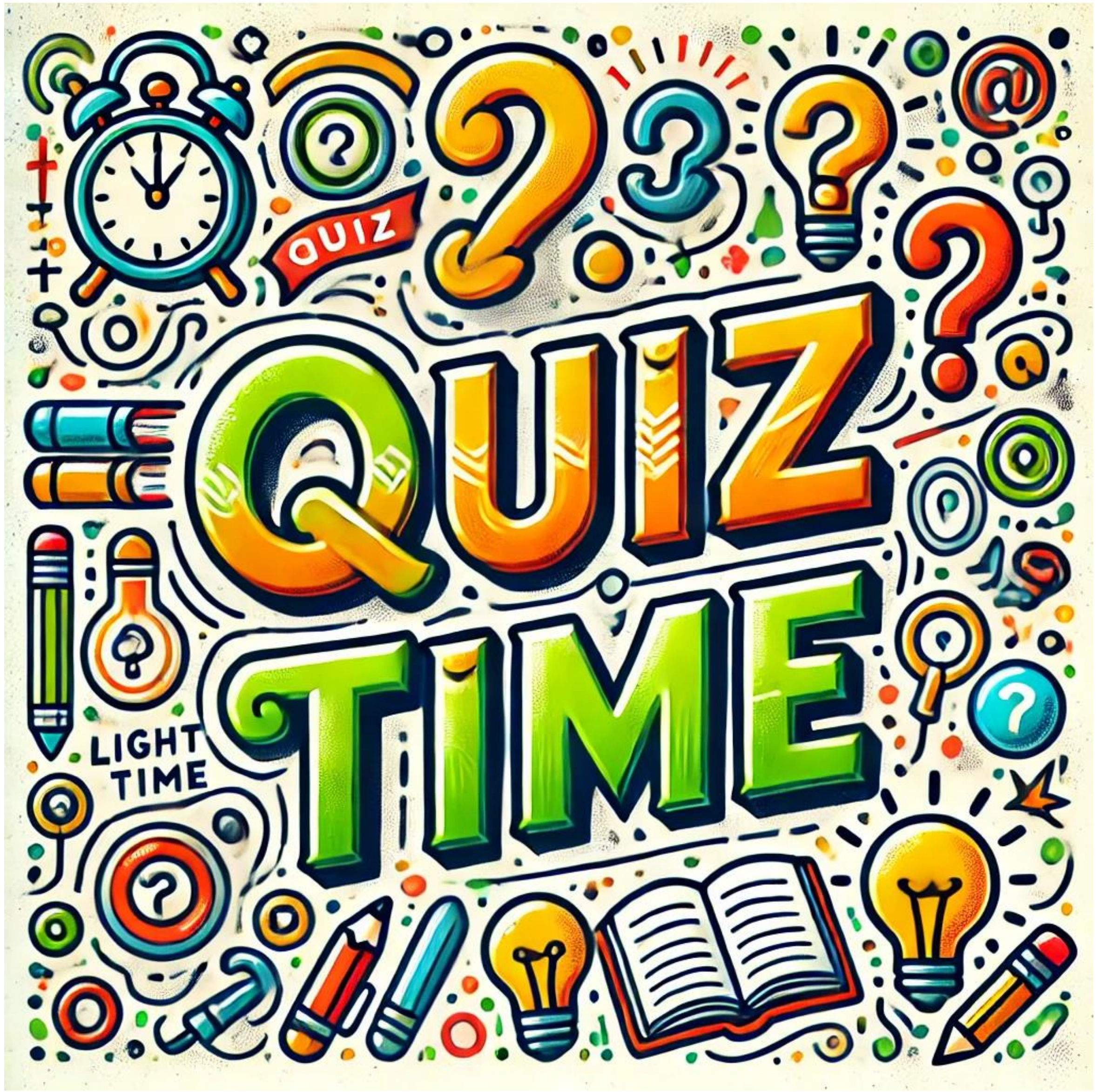
$$\frac{\partial \hat{y}}{\partial z} = \hat{y}(1 - \hat{y})$$

$$\frac{\partial z}{\partial \theta_i} = x_i$$

$$\frac{\partial L}{\partial \theta_i} = x_i(\hat{y} - y)$$






QUIZ
QUIZ
TIME
LIGHT
TIME

Question 1

❖ Hãy sắp xếp thứ tự đúng của các bước để thực hiện thuật toán Logistic Regression

1. Tính đạo hàm của hàm mất mát theo các tham số mô hình.
2. Khởi tạo các tham số ban đầu của mô hình (w và b).
3. Cập nhật các tham số mô hình sử dụng Gradient Descent.
4. Dự đoán đầu ra $\hat{y}$ bằng cách tính giá trị $z = \theta^T x$ và áp dụng hàm sigmoid.
5. Tính giá trị hàm mất mát dựa trên đầu ra dự đoán $\hat{y}$ và giá trị thực y.

a) 4, 5, 1, 2, 3

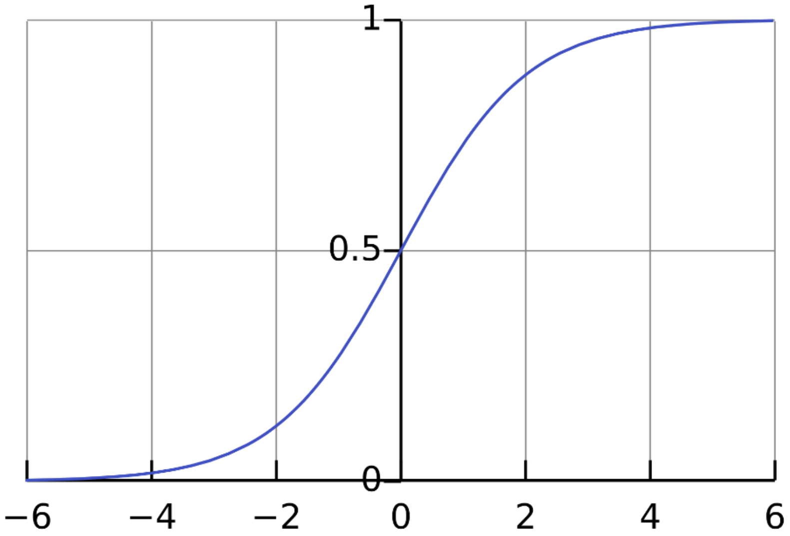
b) 2, 4, 5, 1, 3

c) 2, 5, 4, 1, 3

d) 2, 4, 3, 5, 1

Question 2

❖ Cho hàm sigmoid $\sigma(x) = \frac{1}{1+e^{-x}}$; Với $x_1 = 2$ và $x_2 = 10$ thì khẳng định nào sau đây đúng?



a) $\sigma(x_1) > \sigma(x_2)$

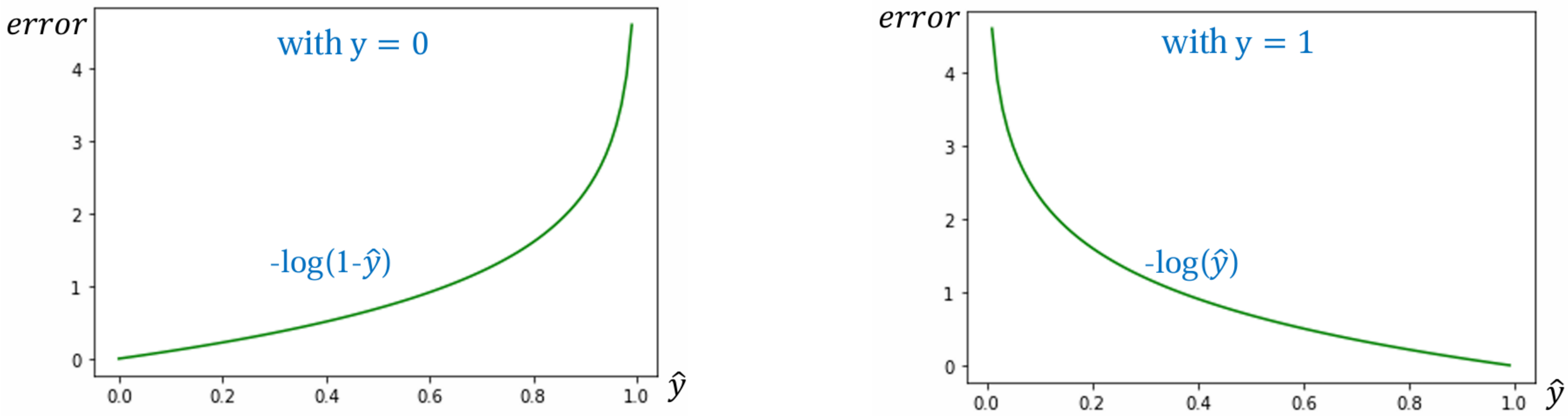
b) $\boldsymbol{\sigma(x_1) < \sigma(x_2)}$

c) $\sigma(x_1) = \sigma(x_2)$

d) Chưa thể khẳng định được điều gì

Question 3

❖ Công thức của hàm mất mát Binary Cross-Entropy (BCE) trong Logistic Regression là gì?



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

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

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

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

Question 4

- Hãy tính đạo hàm bậc hai theo x của hàm $f(x, y) = x^2 + 3y^2 + 4xy$

a) $\frac{\partial^2 f}{\partial x^2} = 2x + 4y$

b) $\frac{\partial^2 f}{\partial x^2} = 2$

c) $\frac{\partial^2 f}{\partial x^2} = 6y + 4x$

d) $\frac{\partial^2 f}{\partial x^2} = 4$

Question 5

- Hãy tính đạo hàm bậc hai theo y của hàm $f(x, y) = x^2 + 3y^2 + 4xy$

a) $\frac{\partial^2 f}{\partial y^2} = 6y + 4x$

b) $\frac{\partial^2 f}{\partial y^2} = 2x + 4y$

c) $\frac{\partial^2 f}{\partial y^2} = 4$

d) $\frac{\partial^2 f}{\partial y^2} = 6$

Question 6

- Hãy tính đạo hàm hỗn hợp của hàm $f(x, y) = x^2 + 3y^2 + 4xy$

a) $\frac{\partial^2 f}{\partial x \partial y} = 6, \ \frac{\partial^2 f}{\partial y \partial x} = 2$

b) $\frac{\partial^2 f}{\partial x \partial y} = 2, \ \frac{\partial^2 f}{\partial y \partial x} = 6$

c) $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 4$

d) $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 2$

Idea of Logistic Regression

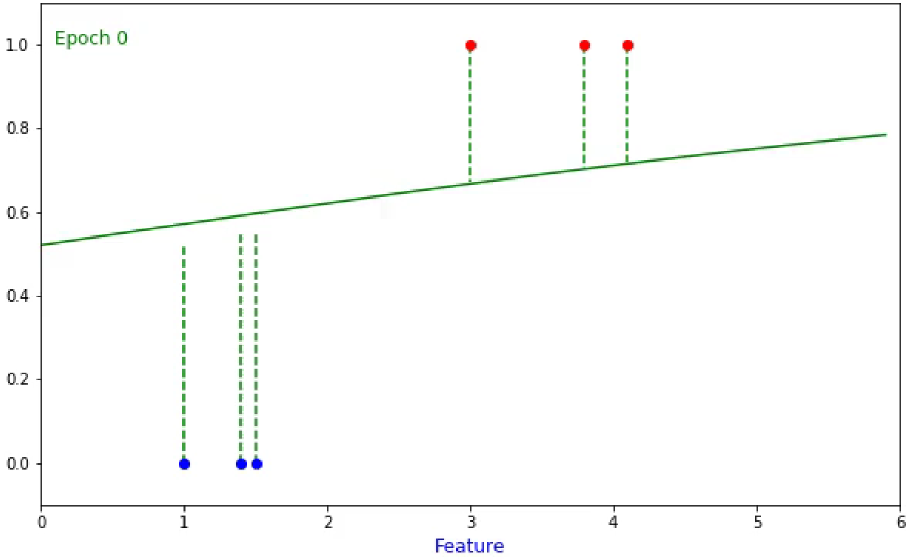
Feature	**Label**	
Petal_Length	Category	
1.4	0	**Category 0**
1	0	
1.5	0	
3	1	**Category 1**
3.8	1	
4.1	1	

$$z = wx + b$$

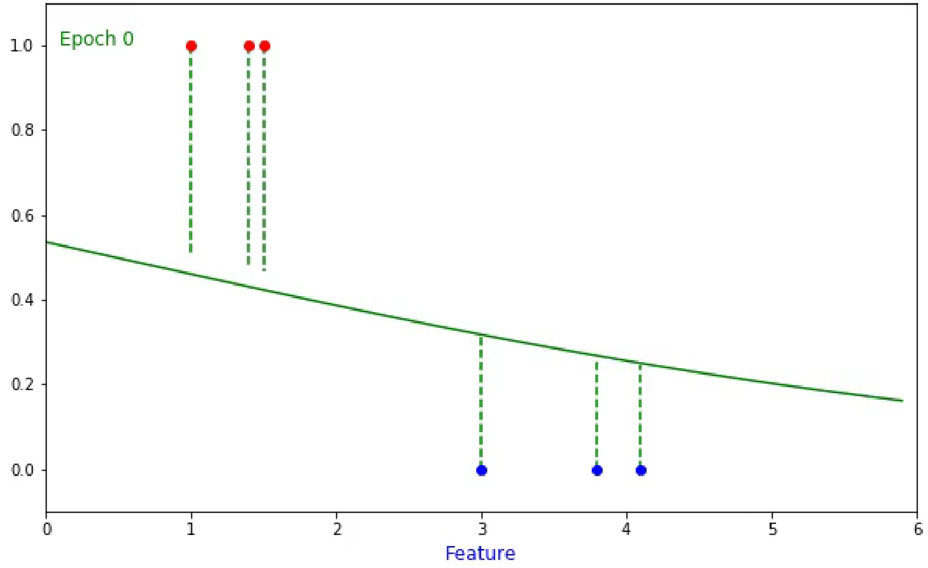
$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

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

$$\frac{1}{1 + e^{-z}}$$



$$\frac{1}{1 + e^{-z}}$$



Binary Cross-entropy

❖ Convex function

Model and Loss

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

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

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

Derivative

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

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

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

$$\frac{\partial z}{\partial \theta_i} = x_i$$

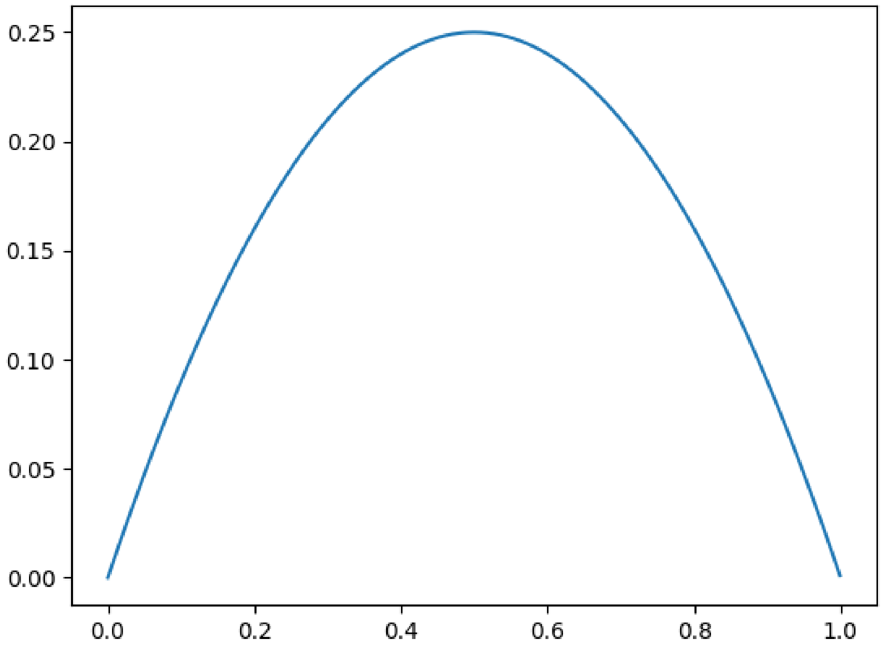
$$\frac{\partial L}{\partial \theta_i} = x_i(\hat{y} - y)$$

simplified version

$$\frac{\partial L}{\partial \theta_i} = x_i(\hat{y} - y)$$

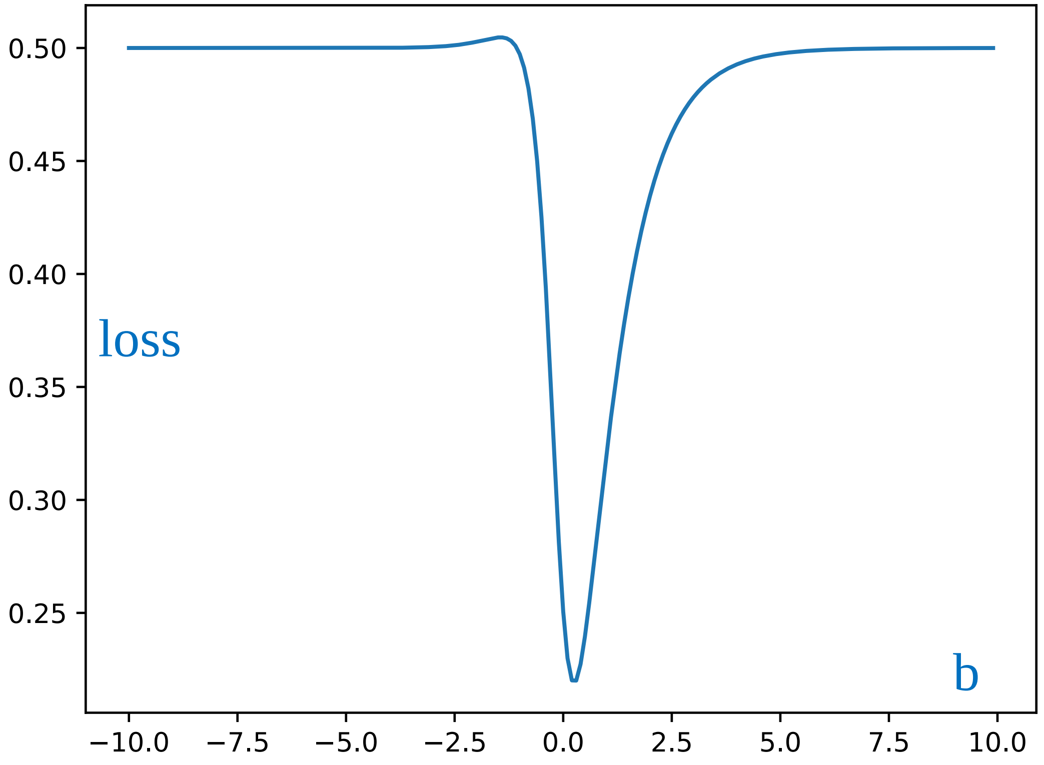
$$\frac{\partial^2 L}{\partial \theta_i^2} = \frac{\partial}{\partial \theta_i}[x_i(\hat{y} - y)] = x_i^2(\hat{y} - \hat{y}^2) \geq 0$$

$$x_i^2 \geq 0 \qquad \hat{y} - \hat{y}^2 \in \left[0, \frac{1}{4}\right]$$

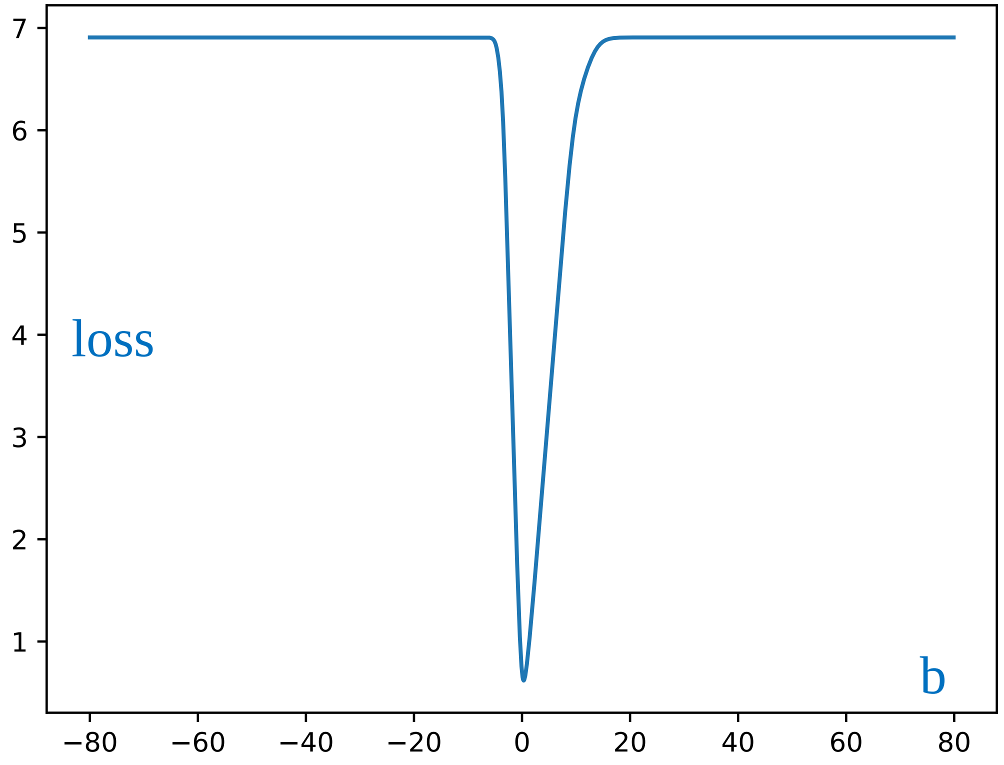


MSE and BCE

❖ Visualization



Mean Squared Error



Binary Cross-Entropy

Example

Logistic Regression-Stochastic

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

2) Compute output $\hat{y}$

$$z = wx + b$$

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

3) Compute loss

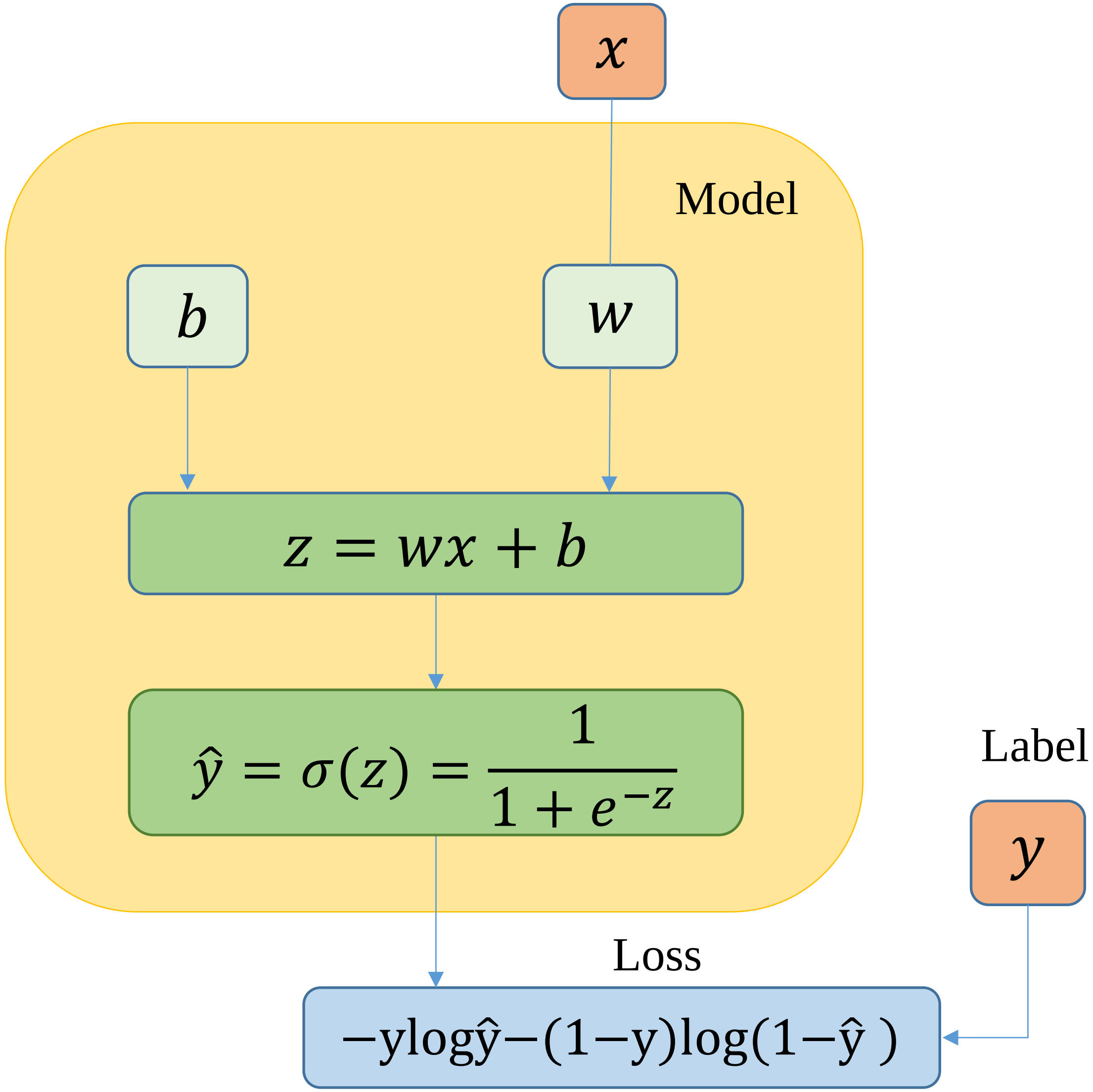
$$L(\hat{y}, y) = (-\text{y}\log\hat{\text{y}} - (1-\text{y})\log(1-\hat{\text{y}}\ ))$$

4) Compute derivative

$$\frac{\partial L}{\partial w} = x(\hat{y} - y) \qquad \frac{\partial L}{\partial b} = (\hat{y} - y)$$

5) Update parameters

$$w = w - \eta \frac{\partial L}{\partial w} \qquad b = b - \eta \frac{\partial L}{\partial b}$$



x

Model

b

w

$$z = wx + b$$

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

Label

y

Loss

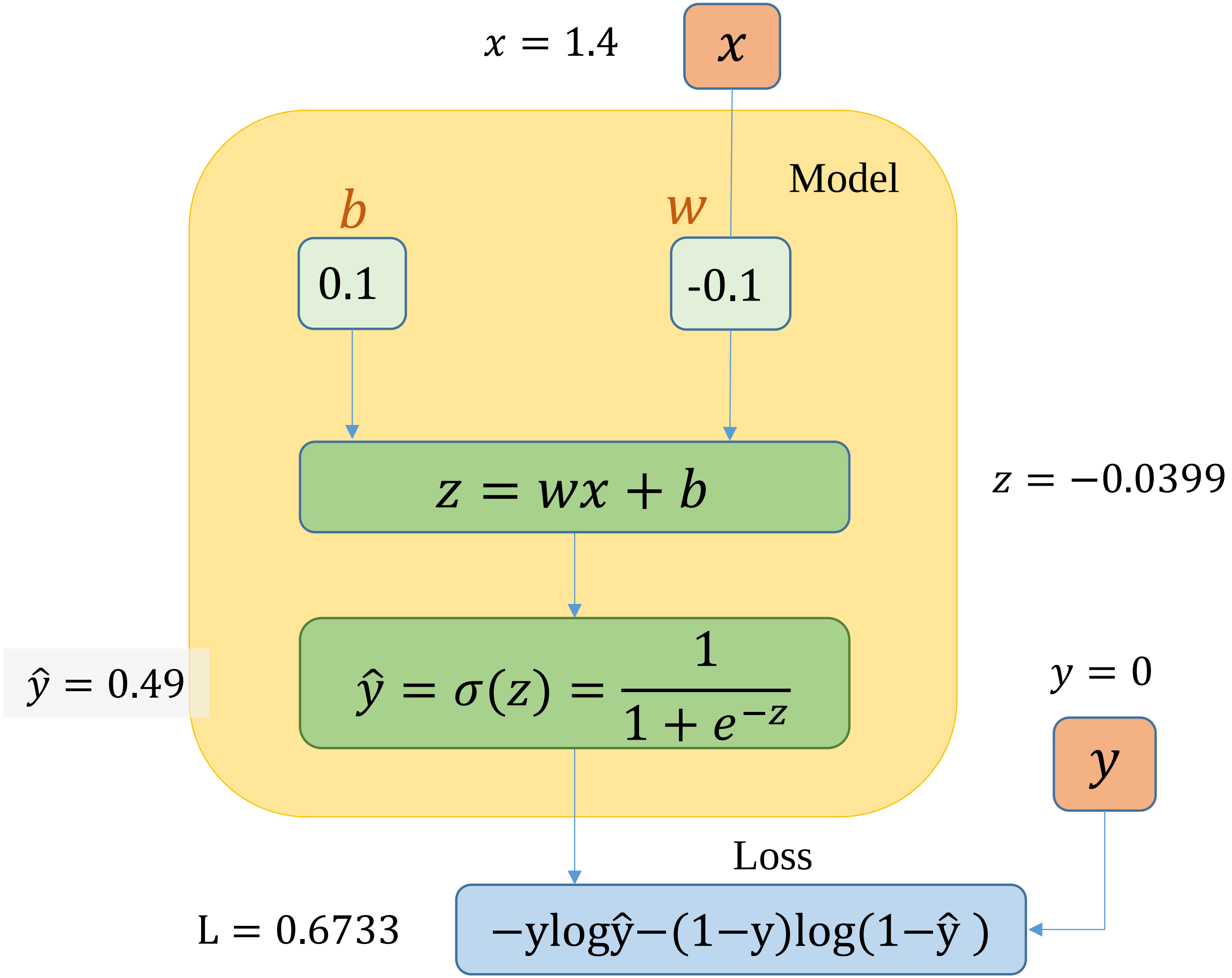
$$-\text{y}\log\hat{\text{y}} - (1-\text{y})\log(1-\hat{\text{y}}\ )$$

Logistic Regression-Stochastic

Dataset

Petal_Length	Label
1.4	0
1.5	0
3	1
4.1	1

$x = 1.4$ $y = 0$



Logistic Regression-Stochastic

$\eta = 0.01$

Dataset

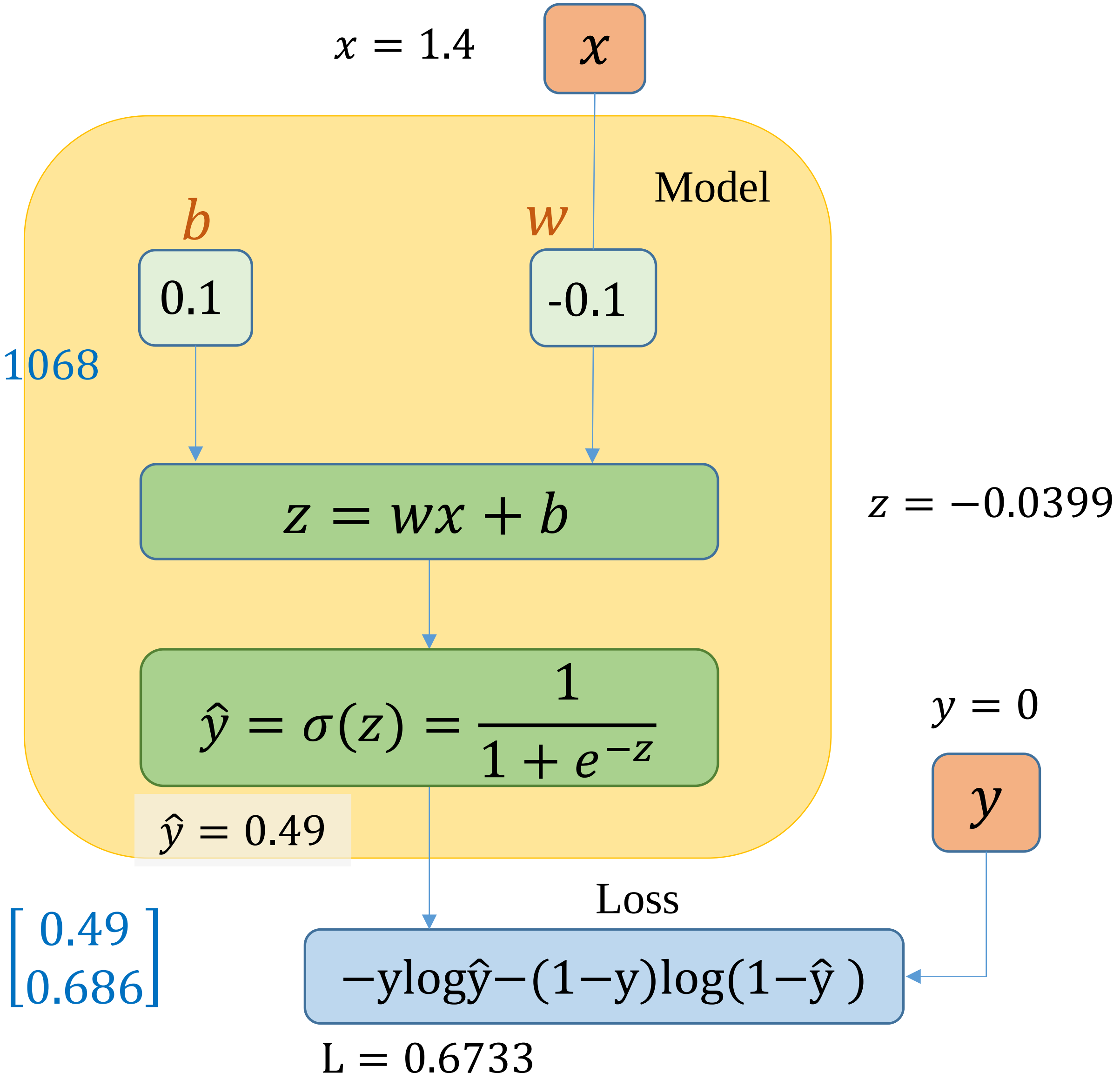
Petal_Length	Label
1.4	0
1.5	0
3	1
4.1	1

$x = 1.4$ $y = 0$

$b = 0.1 - \eta 0.49 = 0.095$

$w = -0.1 - \eta 0.686 = -0.1068$

$$\begin{bmatrix} L'_b \\ L'_w \end{bmatrix} = \begin{bmatrix} 1 * 0.49 \\ 1.4 * 0.49 \end{bmatrix} = \begin{bmatrix} 0.49 \\ 0.686 \end{bmatrix}$$



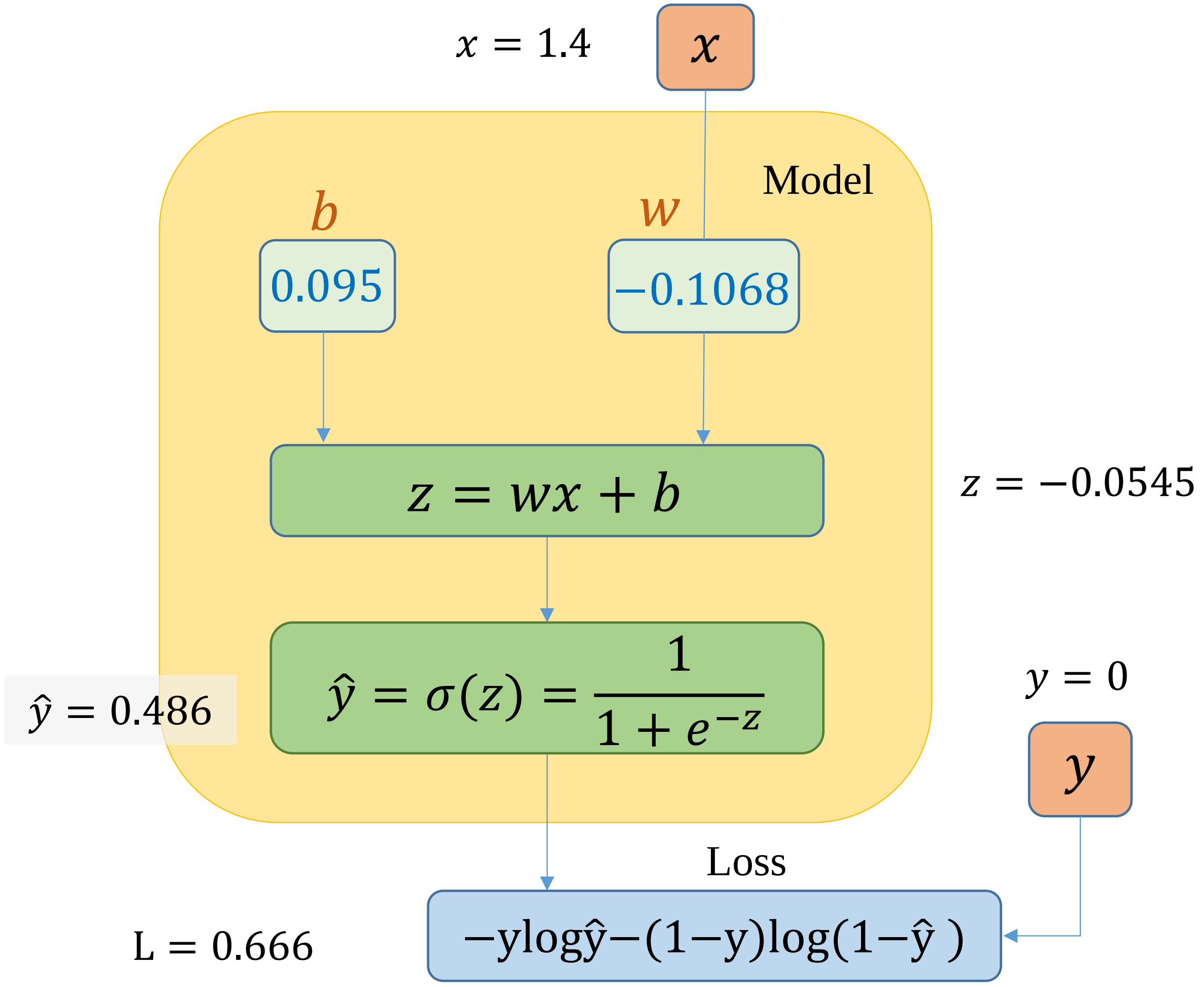
Logistic Regression-Stochastic

Dataset

Petal_Length	Label
1.4	0
1.5	0
3	1
4.1	1

$x = 1.4$ $y = 0$

previous L = 0.6733



$x = 1.4$

x

Model

b

0.095

w

-0.1068

$z = wx + b$

$z = -0.0545$

$\hat{y} = 0.486$

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

$y = 0$

y

Loss

L = 0.666

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

Another Example

Logistic Regression-Stochastic

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

2) Compute output $\hat{y}$

$$z = w_1x_1 + w_2x_2 + b$$

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

3) Compute loss

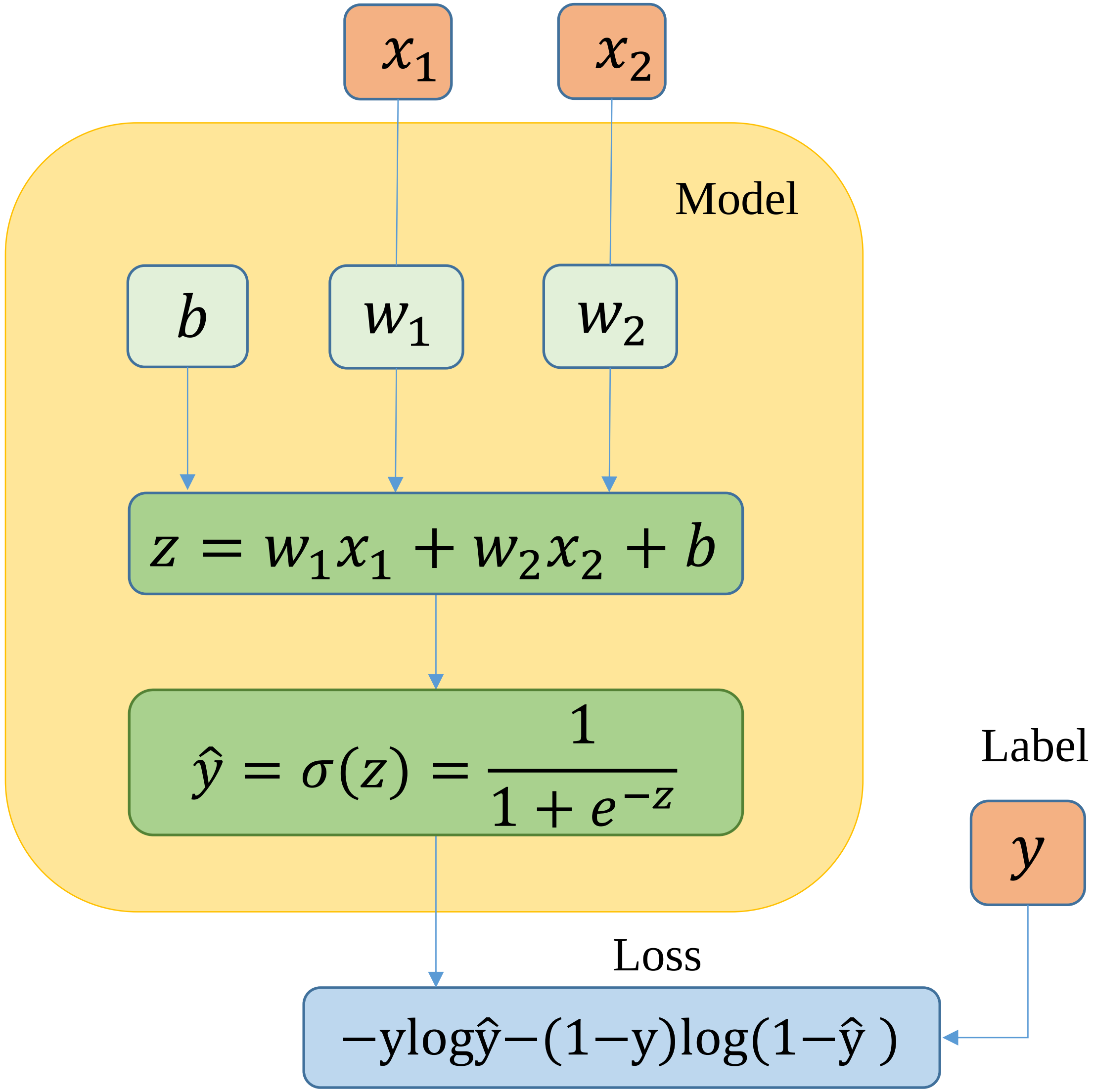
$$L(\hat{y}, y) = (-\mathrm{ylog}\hat{y} - (1-\mathrm{y})\log(1-\hat{y}\ ))$$

4) Compute derivative

$$\frac{\partial L}{\partial w_i} = x_i(\hat{y} - y) \qquad \frac{\partial L}{\partial b} = \ (\hat{y} - y)$$

5) Update parameters

$$w_i = w_i - \eta\frac{\partial L}{\partial w_i} \qquad b = b - \eta\frac{\partial L}{\partial b}$$



Logistic Regression-Stochastic

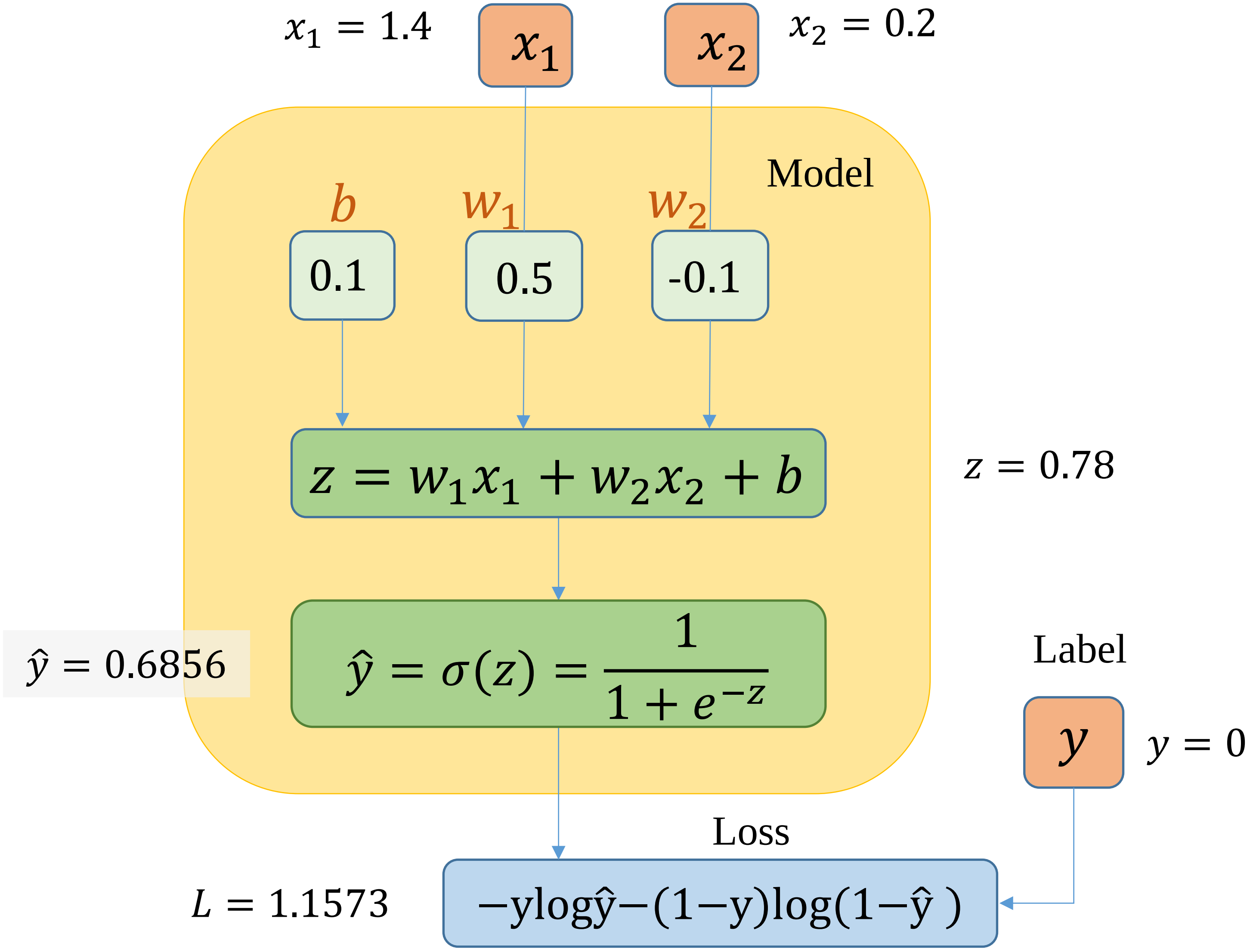
Dataset

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

$x_1 = 1.4$

$y = 0$

$x_2 = 0.2$



$x_1 = 1.4$ $x_2 = 0.2$

Model

b w_1 w_2

0.1 0.5 -0.1

$z = w_1x_1 + w_2x_2 + b$

$z = 0.78$

$\hat{y} = 0.6856$

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

Label

$y = 0$

Loss

$L = 1.1573$

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

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

$x_1 = 1.4$

$y = 0$

$x_2 = 0.2$

$$\begin{bmatrix} L'_b \\ L'_{w_1} \\ L'_{w_2} \end{bmatrix} = \begin{bmatrix} 1 * 0.6856 \\ 1.4 * 0.6856 \\ 0.2 * 0.6856 \end{bmatrix} = \begin{bmatrix} 0.6856 \\ 0.9599 \\ 0.1371 \end{bmatrix}$$

$\eta = 0.01$

$b = 0.1 - \eta 0.6856$

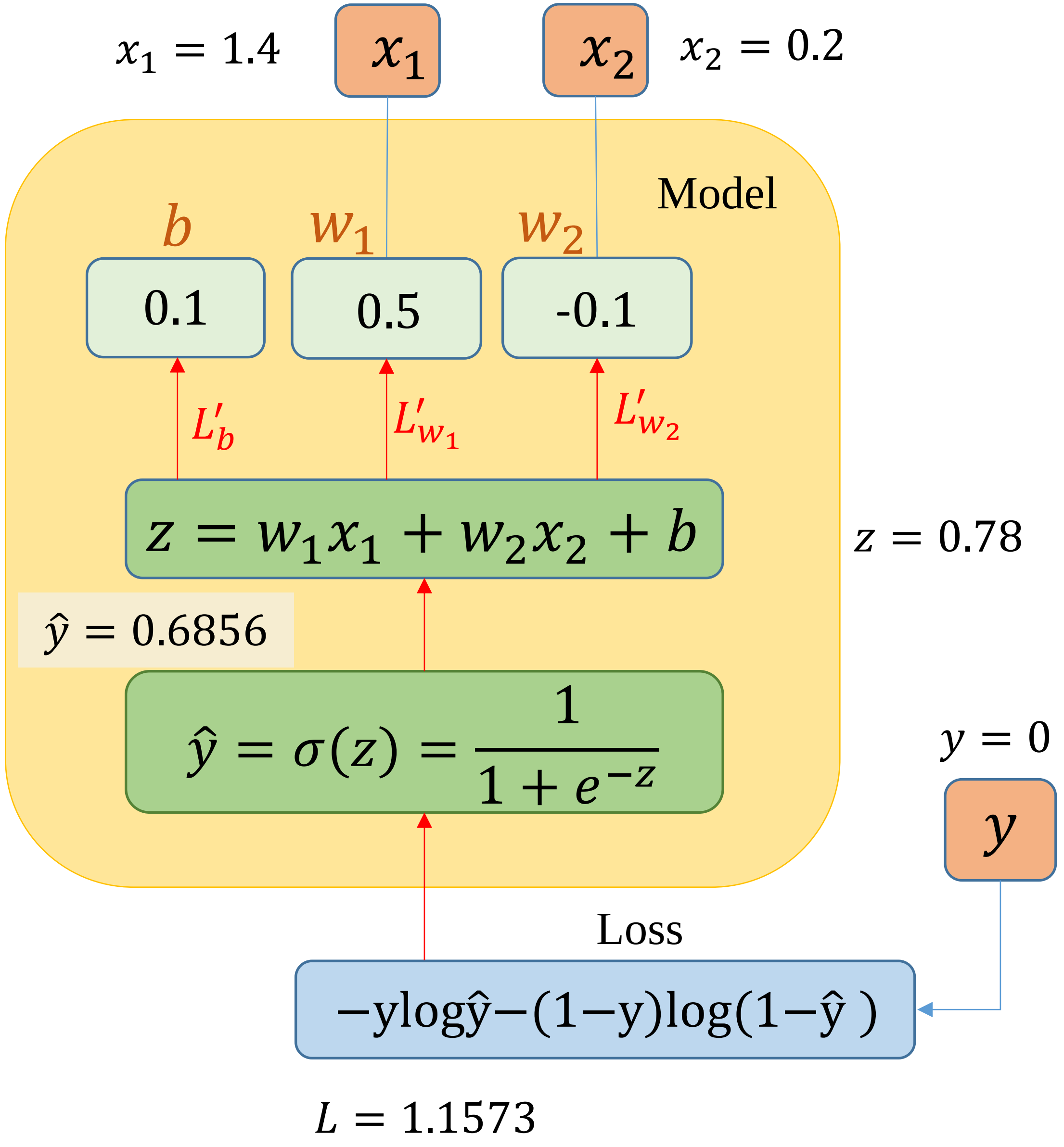
$= 0.0931$

$w_1 = 0.5 - \eta 0.9598$

$= 0.4990$

$w_2 = -0.1 + \eta 0.1371$

$= -0.1013$



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

$x_1 = 1.4$

$y = 0$

$x_2 = 0.2$

$$\begin{bmatrix} L'_b \\ L'_{w_1} \\ L'_{w_2} \end{bmatrix} = \begin{bmatrix} 1 * 0.6856 \\ 1.4 * 0.6856 \\ 0.2 * 0.6856 \end{bmatrix} = \begin{bmatrix} 0.6856 \\ 0.9599 \\ 0.1371 \end{bmatrix}$$

$\eta = 0.01$

$b = 0.1 - \eta 0.6856$

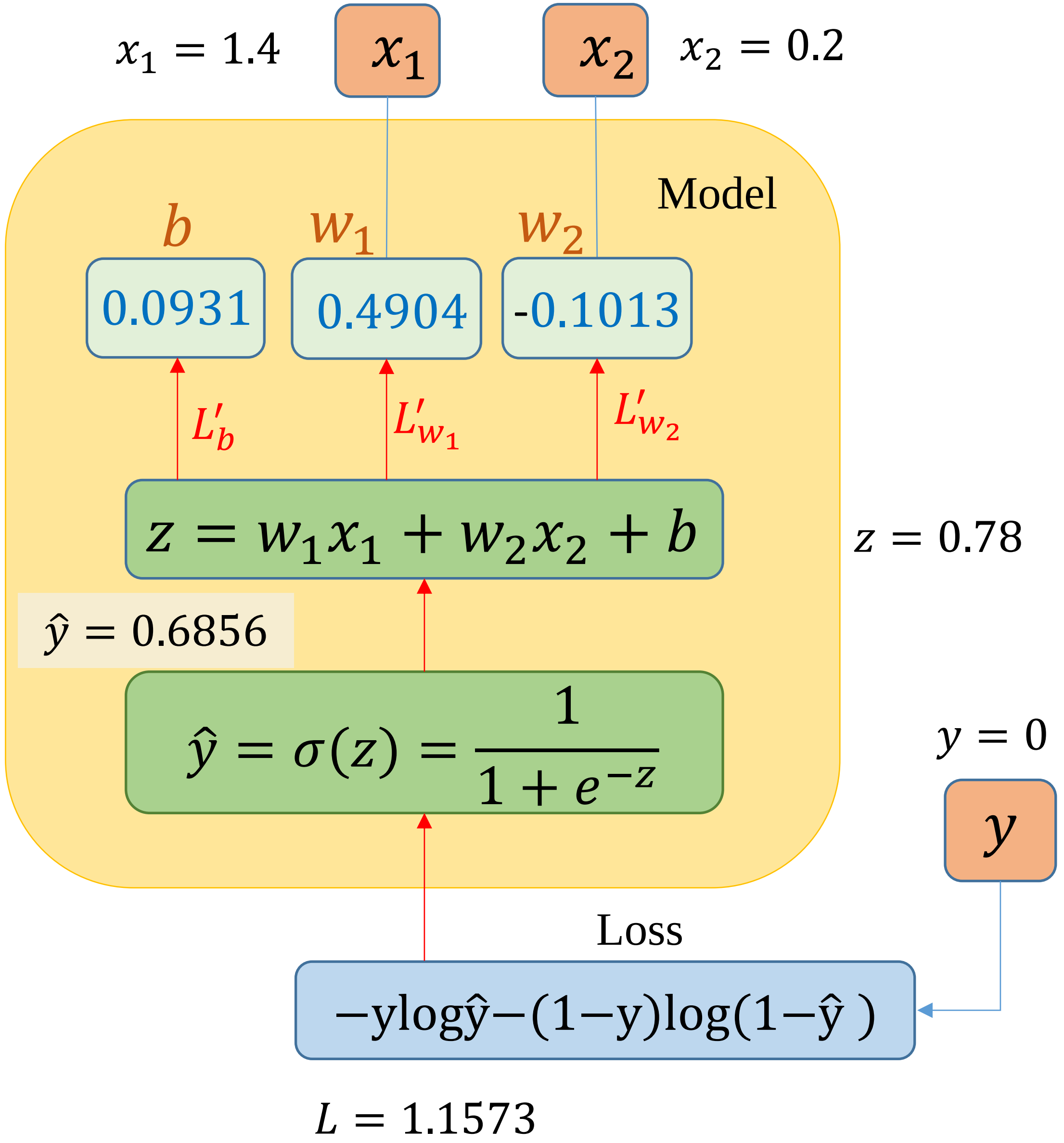
$= 0.0931$

$w_1 = 0.5 - \eta 0.9598$

$= 0.4990$

$w_2 = -0.1 + \eta 0.1371$

$= -0.1013$



$x_1 = 1.4$ x_1 x_2 $x_2 = 0.2$

Model

b w_1 w_2

0.0931 0.4904 -0.1013

L'_b L'_{w_1} L'_{w_2}

$z = w_1x_1 + w_2x_2 + b$

$z = 0.78$

$\hat{y} = 0.6856$

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

$y = 0$

y

Loss

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

$L = 1.1573$

Logistic Regression-Stochastic

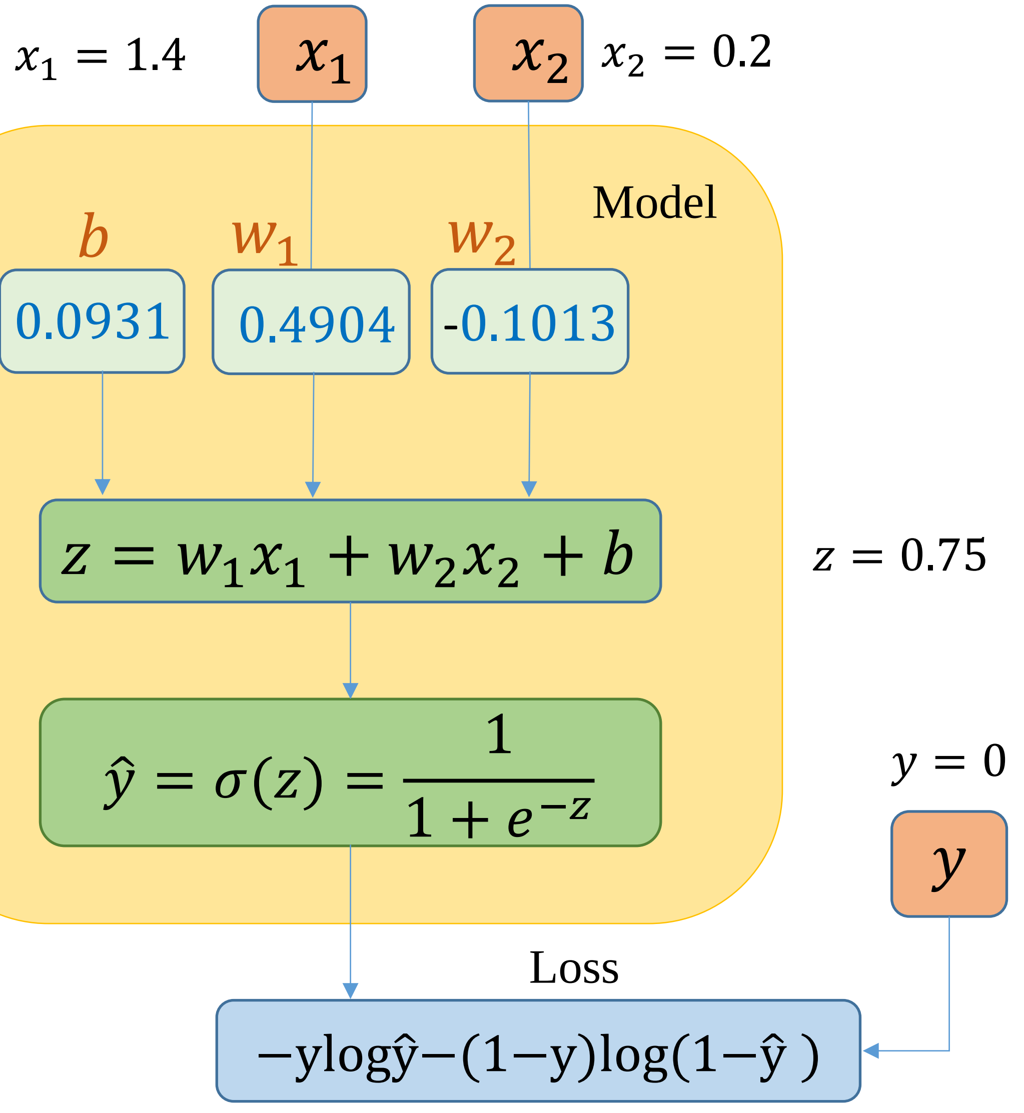
Dataset

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

$x_1 = 1.4$

$y = 0$

$x_2 = 0.2$



$x_1 = 1.4$ $x_2 = 0.2$

Model

b w_1 w_2

0.0931 0.4904 -0.1013

$$z = w_1 x_1 + w_2 x_2 + b$$

$z = 0.75$

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

$\hat{y} = 0.6812$

$y = 0$

Loss

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

previous $\boldsymbol{L} = [1.1573]$

$L = 1.1432$

Summary

Traditional

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

2) Compute the output $\hat{y}$

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

3) Compute loss

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

4) Compute derivative

$$\frac{\partial L}{\partial w} = x(\hat{y} - y) \qquad \frac{\partial L}{\partial b} = (\hat{y} - y)$$

5) Update parameters

$$w = w - \eta \frac{\partial L}{\partial w} \qquad b = b - \eta \frac{\partial L}{\partial b}$$

η is learning rate

Vectorized

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

2) Compute output $\hat{y}$

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

3) Compute loss

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

4) Compute derivative

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

5) Update parameters

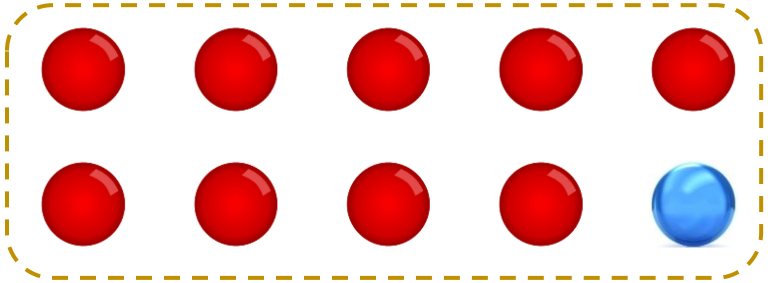
$$\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\boldsymbol{\theta}} L$$

η is learning rate

Construct Logistic Regression in a different way (optional)

Entropy

❖ Motivation



A: Get a red ball

B: Get a blue ball

$$p(A) = \frac{9}{10} = 0.9$$

$$p(B) = \frac{1}{10} = 0.1$$

E: Pick a ball from the basket

Experiment 1

Got a red ball

Experiment 2

Got a blue ball

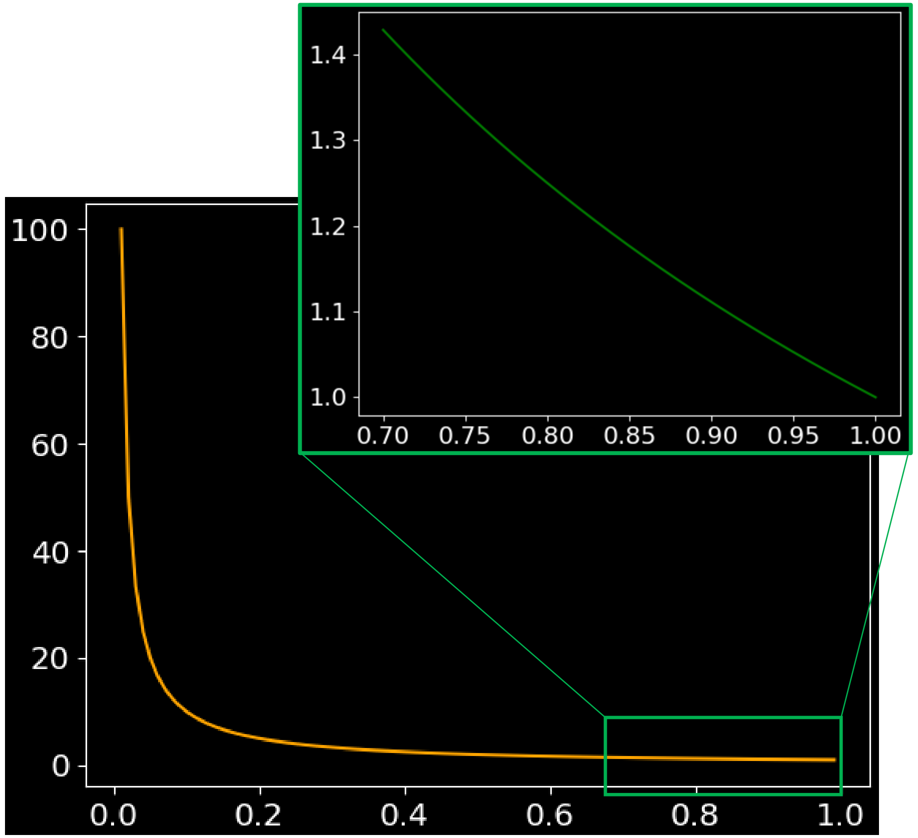
Which experiment makes you more surprised?

How to measure the surprises?

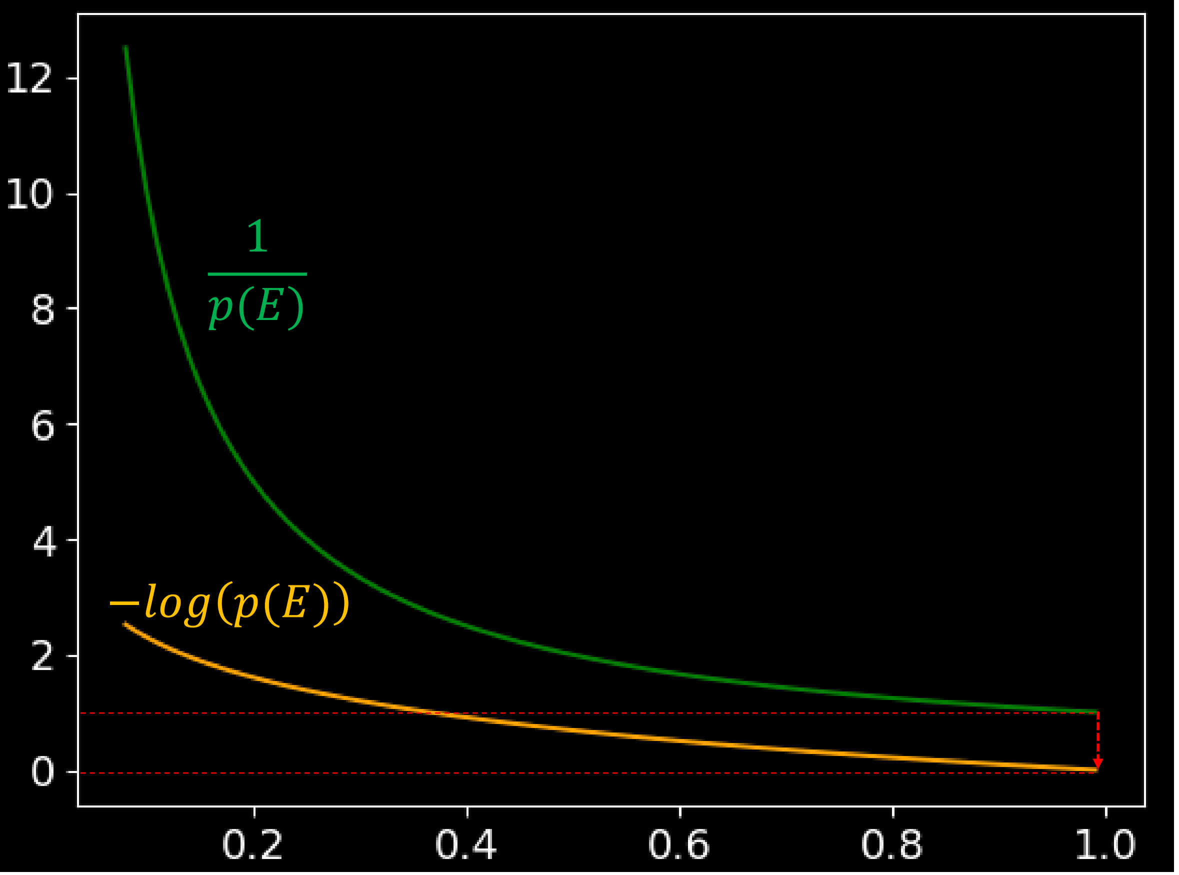
Observation

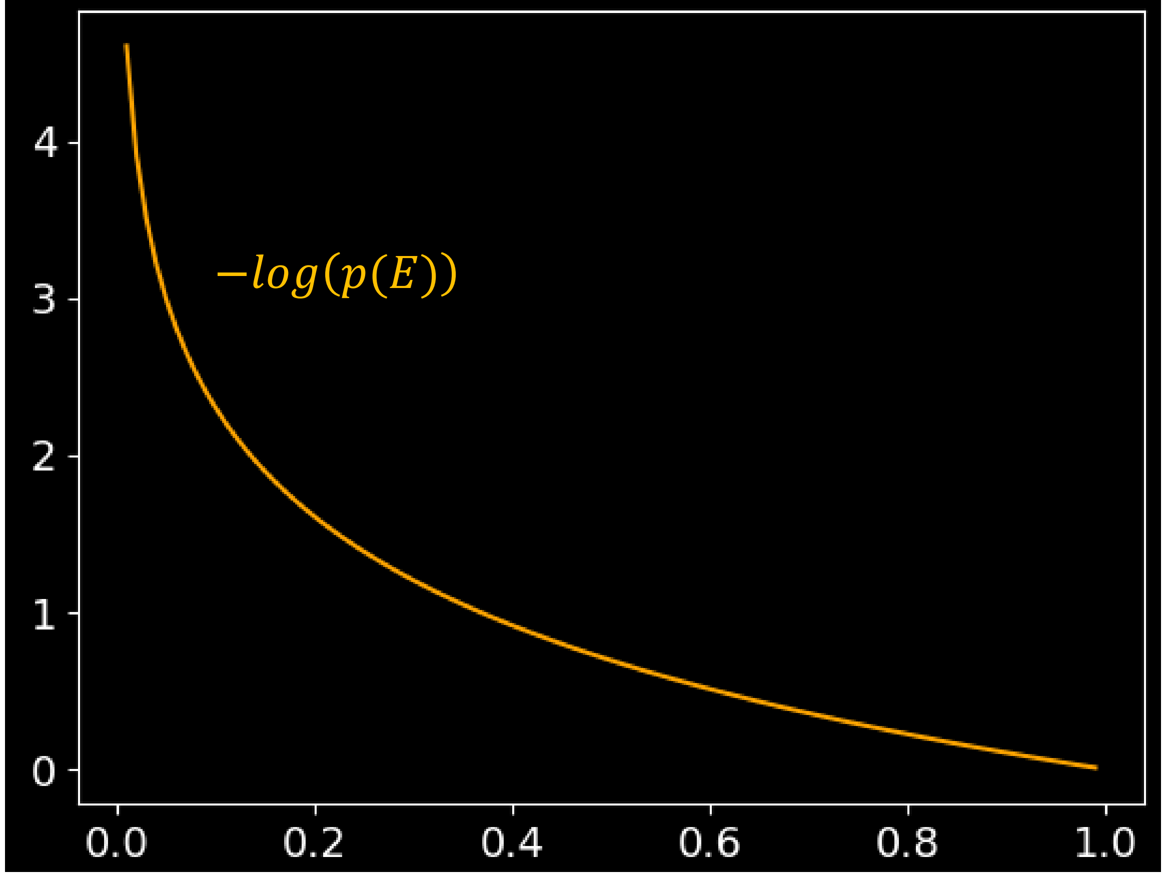
$$Surprise(E) \downarrow \uparrow p(E)$$

$$\Rightarrow Surprise(E) = \frac{1}{p(E)}$$



Problem?





Monotonic decrease of the function surprise(E)

$$log(Surprise(E)) = log\left(\frac{1}{p(E)}\right)$$

$$= -log(p(E))$$

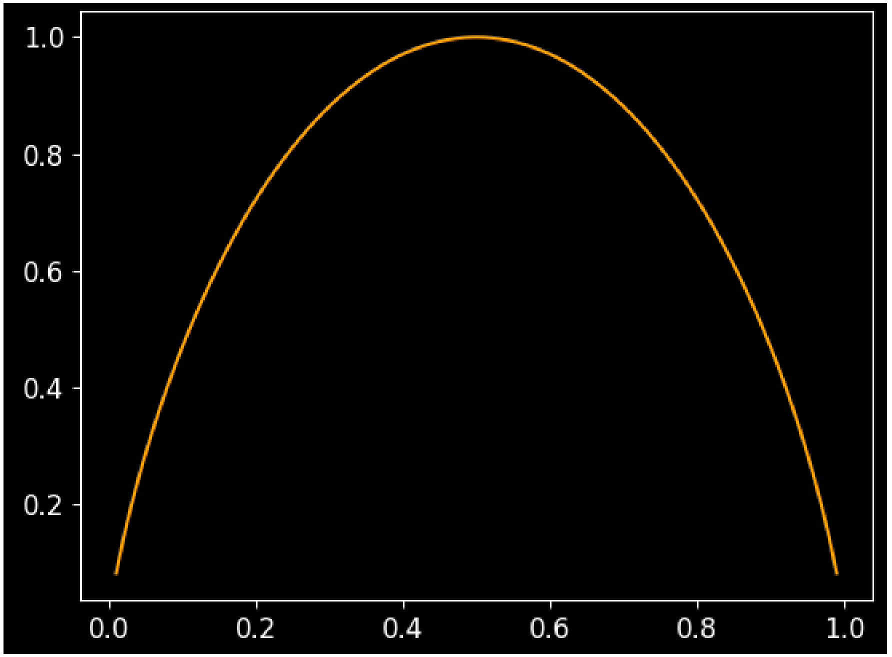
In information theory

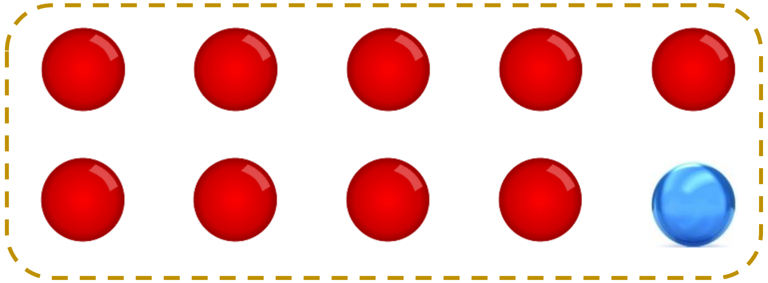
$$\text{Information}(x) = -log(p(x))$$

Entropy

Entropy: Average of information

$$H(X) := -\sum_{x \in X} p(x)\, log(p(x))$$





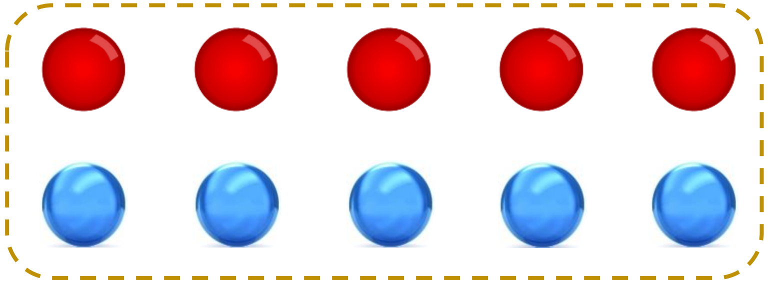
$$p(X = 0) = \frac{9}{10} = 0.9$$

$$p(X = 1) = \frac{1}{10} = 0.1$$

$$H(X) = -\sum_{x \in X} p(x)\, log(p(x))$$

$$= -0.9log(0.9) - 0.1log(0.1)$$

$$= 0.468$$



$$p(X = 0) = \frac{5}{10} = 0.5$$

$$p(X = 1) = \frac{5}{10} = 0.5$$

$$H(X) = -\sum_{x \in X} p(x)\, log(p(x))$$

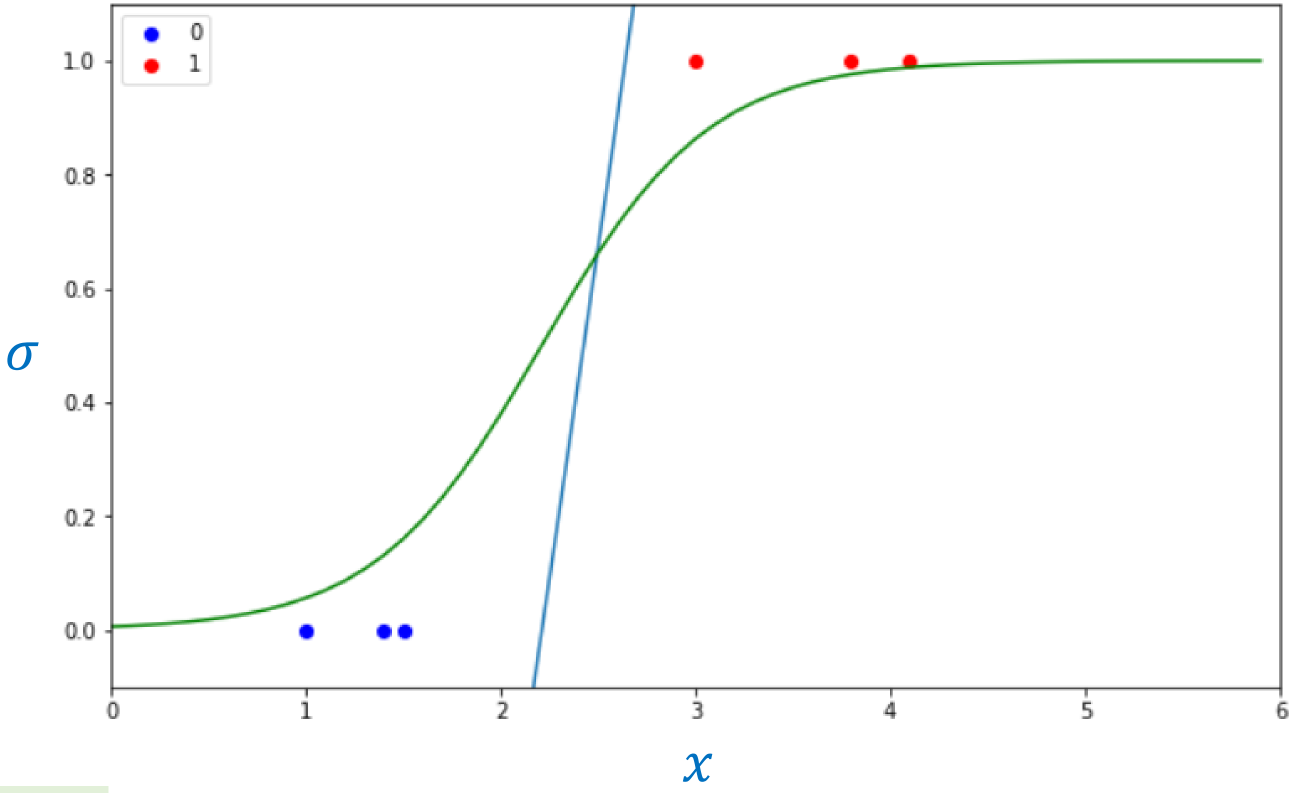
$$= -0.5log(0.5) - 0.5log(0.5)$$

$$= 1.0$$

Logistic Regression

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

z	$\sigma(z)$
-1.89	0.1309
-2.82	0.0559
-1.65	0.1598
1.837	0.8625
3.701	0.9759
4.401	0.9878



$$z = wx + b$$

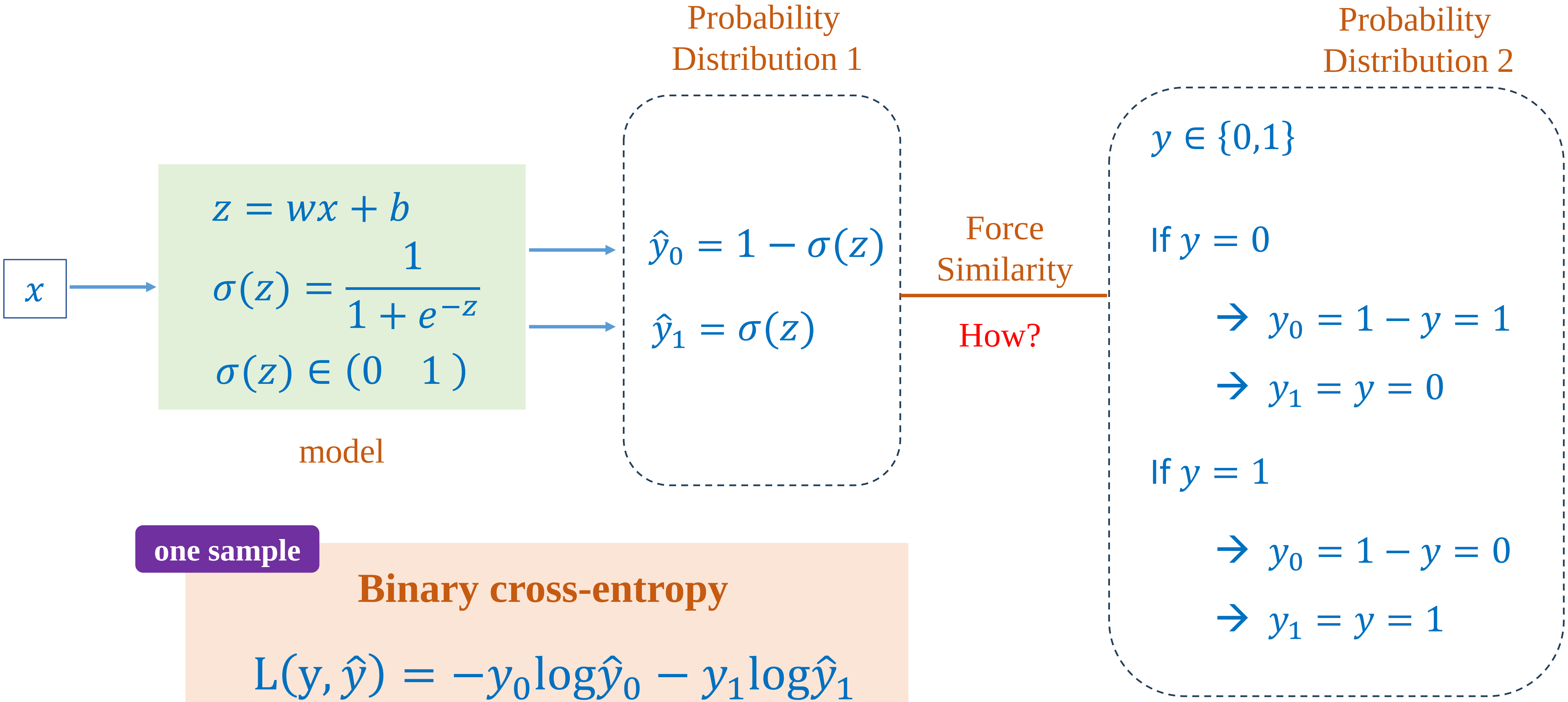
$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

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

How to interpret $\sigma(z)$?

(Binary) Cross-Entropy

Given a sample (x, y)



x

$z = wx + b$

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

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

model

Probability Distribution 1

$\hat{y}_0 = 1 - \sigma(z)$

$\hat{y}_1 = \sigma(z)$

Force Similarity

How?

Probability Distribution 2

$y \in \{0,1\}$

If $y = 0$

$\rightarrow y_0 = 1 - y = 1$

$\rightarrow y_1 = y = 0$

If $y = 1$

$\rightarrow y_0 = 1 - y = 0$

$\rightarrow y_1 = y = 1$

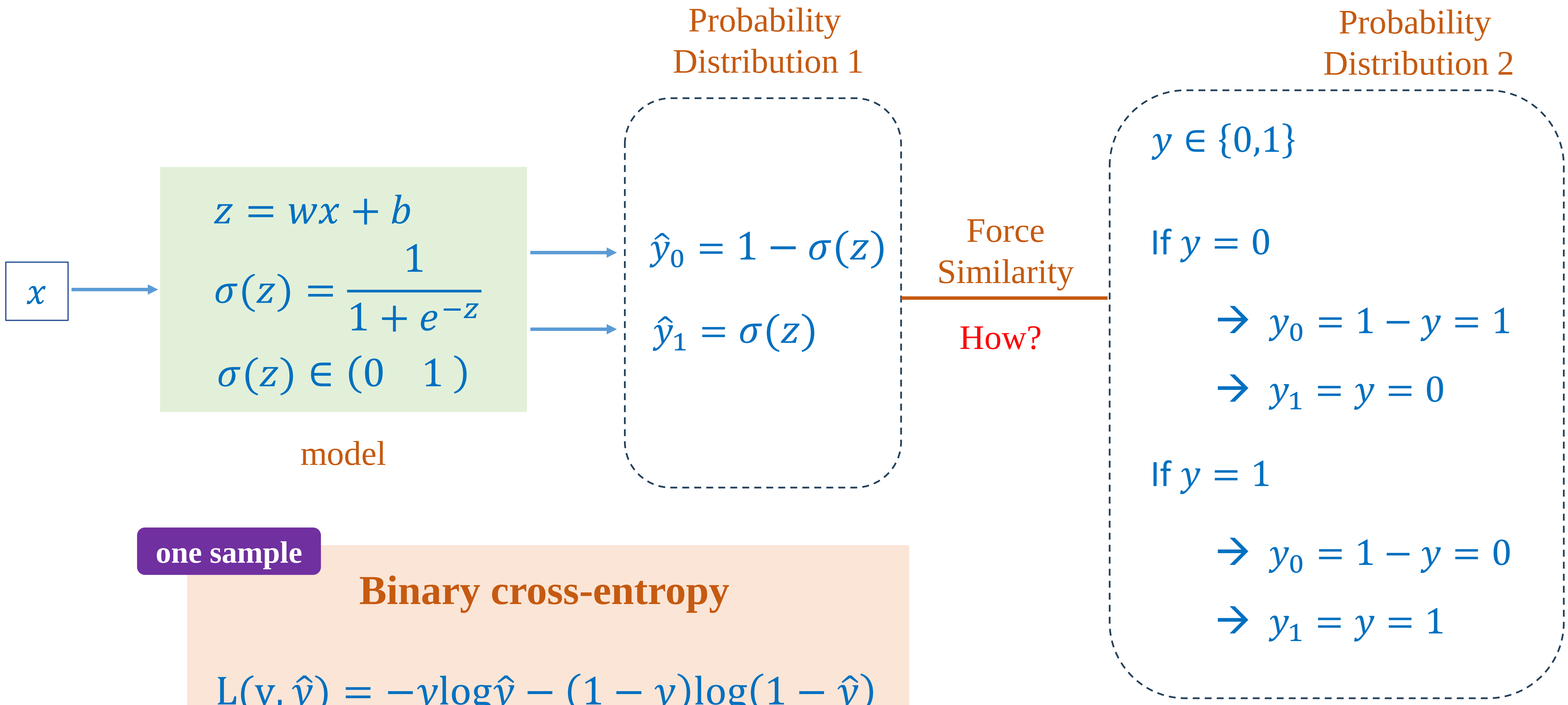
one sample

Binary cross-entropy

$$\text{L}(y, \hat{y}) = -y_0 \log \hat{y}_0 - y_1 \log \hat{y}_1$$

(Binary) Cross-Entropy

Given a sample (x, y)



x

$z = wx + b$

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

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

model

Probability Distribution 1

$\hat{y}_0 = 1 - \sigma(z)$

$\hat{y}_1 = \sigma(z)$

Force Similarity

How?

Probability Distribution 2

$y \in \{0,1\}$

If $y = 0$

→ $y_0 = 1 - y = 1$

→ $y_1 = y = 0$

If $y = 1$

→ $y_0 = 1 - y = 0$

→ $y_1 = y = 1$

one sample

Binary cross-entropy

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

(Binary) Cross-Entropy

Prove convexity

Suppose we have:

- A single feature input x, and a target $y \in \{0, 1\}$.
- A logistic model with one feature, so $z = wx + b$, where w is the weight and b is the bias.

The model outputs the probability:

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

The binary cross-entropy loss for this single point is:

$$L(w, b) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y}))$$

Step 1: Compute the Gradient of $L(w, b)$

1. Rewrite the loss in terms of $z = wx + b$:

$$L(z) = -\big(y \log\big(\sigma(z)\big) + (1 - y) \log\big(1 - \sigma(z)\big)\big)$$

2. First Derivative with respect to z:

Since $\hat{y} = \sigma(z)$, the derivative of L with respect to z is:

$$\frac{dL}{dz} = \hat{y} - y$$

3. Gradient with respect to w and b:

Using the chain rule:

- $\frac{\partial L}{\partial w} = \frac{dL}{dz} \cdot \frac{dz}{dw} = (\hat{y} - y)x$
- $\frac{\partial L}{\partial b} = \frac{dL}{dz} \cdot \frac{dz}{db} = (\hat{y} - y)$

Step 2: Compute the Hessian Matrix of $L(w, b)$

Now, let's find the second derivatives to get the Hessian matrix of L with respect to w and b. The Hessian matrix H is:

$$H = \begin{bmatrix} \frac{\partial^2 L}{\partial w^2} & \frac{\partial^2 L}{\partial w \partial b} \\ \frac{\partial^2 L}{\partial b \partial w} & \frac{\partial^2 L}{\partial b^2} \end{bmatrix}$$

To compute each element:

1. Second derivative with respect to w:

$$\frac{\partial^2 L}{\partial w^2} = \frac{\partial}{\partial w}\big((\hat{y} - y)x\big) = \hat{y}(1 - \hat{y})x^2$$

2. Second derivative with respect to b :

$$\frac{\partial^2 L}{\partial b^2} = \frac{\partial}{\partial b}(\hat{y} - y) = \hat{y}(1 - \hat{y})$$

3. Cross derivative with respect to w and b:

$$\frac{\partial^2 L}{\partial w \partial b} = \frac{\partial^2 L}{\partial b \partial w} = \frac{\partial}{\partial w}(\hat{y} - y) = \hat{y}(1 - \hat{y})x$$

So, the Hessian matrix H becomes:

$$H = \begin{bmatrix} \hat{y}(1 - \hat{y})x^2 & \hat{y}(1 - \hat{y})x \\ \hat{y}(1 - \hat{y})x & \hat{y}(1 - \hat{y}) \end{bmatrix}$$

Step 3: Check Positive Semi-Definiteness of H

To verify convexity, we need to confirm that H is positive semi-definite. For a 2×2 matrix, H is positive semi-definite if its determinant is non-negative and its diagonal entries are non-negative.

1. Determinant of H:

$$\det(H) = (\hat{y}(1 - \hat{y})x^2) \cdot \big(\hat{y}(1 - \hat{y})\big) - (\hat{y}(1 - \hat{y})x)^2$$
$$= \hat{y}^2(1 - \hat{y})^2(x^2 - x^2) = 0$$

Since the determinant is zero, H is semi-definite.

2. Diagonal Entries of H:

Both diagonal entries, $\hat{y}(1 - \hat{y})x^2$ and $\hat{y}(1 - \hat{y})$, are non-negative (since $0 \leq p \leq 1$)

Prove using the eigenvalues of Hessian matrix

To prove that the Hessian matrix H has non-negative eigenvalues, let’s explicitly analyze the eigenvalues of H for a 2×2 matrix.

The Hessian matrix we derived is:

$$H = \begin{bmatrix} \hat{y}(1-\hat{y})x^2 & \hat{y}(1-\hat{y})x \\ \hat{y}(1-\hat{y})x & \hat{y}(1-\hat{y}) \end{bmatrix}$$

Where:

- $\hat{y} = \sigma(z) = \frac{1}{1+e^{-z}}$,
- $0 \leq p \leq 1$, so $\hat{y}(1-\hat{y}) \geq 0$

For a 2×2 matrix of the form:

$$H = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$$

The eigenvalues λ are given by the solutions to the characteristic polynomial:

$$\det(H - \lambda I) = 0$$

Where I is the identity matrix. For our Hessian matrix H, this becomes:

$$\det \begin{bmatrix} a-\lambda & b \\ b & d-\lambda \end{bmatrix} = 0$$

Expanding the determinant, we get:

$$(a-\lambda)(d-\lambda) - b^2 = 0$$

The eigenvalues λ are then:

$$\lambda = \frac{(a+d) \pm \sqrt{(a-d)^2 + 4b^2}}{2}$$

(Binary) Cross-Entropy

Applying to Our Hessian Matrix

For our Hessian matrix:

- $a = \hat{y}(1-\hat{y})x^2$,
- $b = \hat{y}(1-\hat{y})x$,
- $d = \hat{y}(1-\hat{y})$.

Substitute these values into the formula for the eigenvalues:

1. Sum of the diagonal elements (trace of H):

$$a + d = \hat{y}(1-\hat{y})x^2 + \hat{y}(1-\hat{y}) = \hat{y}(1-\hat{y})(x^2+1)$$

2. Difference of the diagonal elements:

$$a - d = \hat{y}(1-\hat{y})(x^2-1)$$

3. Eigenvalues:

Using the eigenvalue formula:

$$\lambda = \frac{(a+d) \pm \sqrt{(a-d)^2 + 4b^2}}{2}$$

Substitute $a + d = \hat{y}(1-\hat{y})(x^2+1)$, $a - d = \hat{y}(1-\hat{y})(x^2-1)$, and $b = \hat{y}(1-\hat{y})x$:

$$\lambda = \frac{\hat{y}(1-\hat{y})(x^2+1) \ \pm \sqrt{\left(\hat{y}(1-\hat{y})(x^2-1)\right)^2 + 4(\hat{y}(1-\hat{y})x)^2}}{2}$$

4. Simplifying the Square Root:

Notice that:

$$\left(\hat{y}(1-\hat{y})(x^2-1)\right)^2 + 4(\hat{y}(1-\hat{y})x)^2 = \hat{y}^2(1-\hat{y})^2\left((x^2-1)^2 + 4x^2\right)$$

Expanding $(x^2-1)^2 + 4x^2$, we get:

$$(x^4 - 2x^2 + 1) + 4x^2 = x^4 + 2x^2 + 1 = (x^2+1)^2$$

So:

$$\sqrt{\left(\hat{y}(1-\hat{y})(x^2-1)\right)^2 + 4(\hat{y}(1-\hat{y})x)^2} = \hat{y}(1-\hat{y})(x^2+1)$$

5. Eigenvalues:

Therefore:

$$\lambda = \frac{\hat{y}(1-\hat{y})(\mathrm{x}^2+1) \ \pm \hat{y}(1-\hat{y})(x^2+1)}{2}$$

This yields two eigenvalues:

$\lambda_1 = \hat{y}(1-\hat{y})(x^2+1)$ and $\lambda_2 = 0$