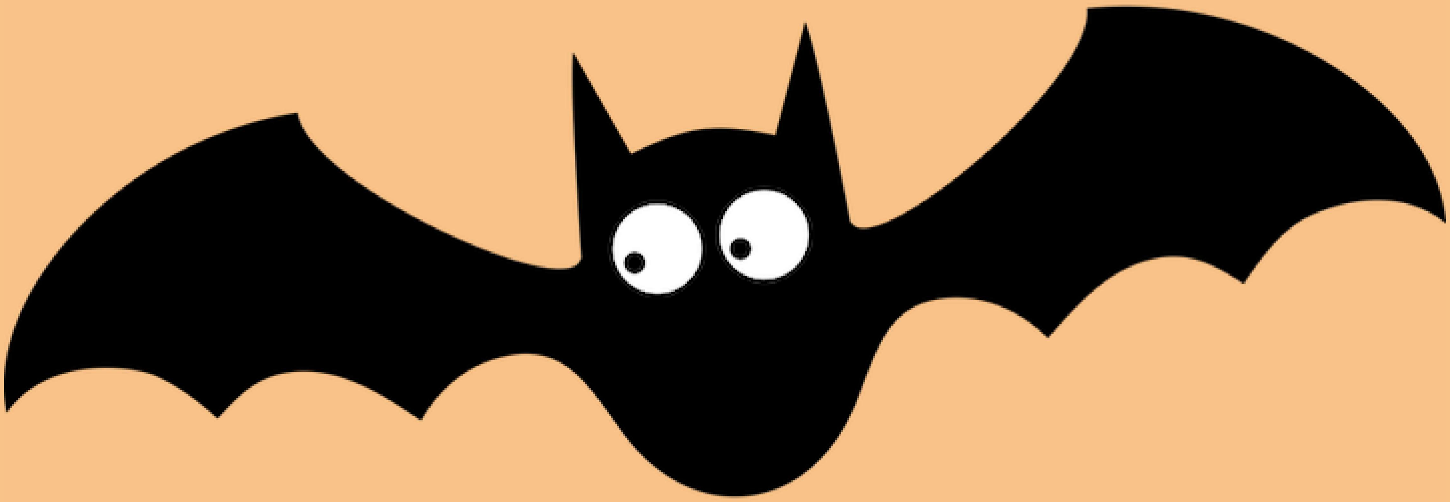


Linear Model

Son Nguyen

A review of Linear Model for Regression

- Given the data

x_1	x_2	y
1	0	-2
2	1	0
3	-2	-1
4	3	1

- How are y and x related?

A review of Linear Model for Regression

- Given the data

x_1	x_2	y
1	0	-2
2	1	0
3	-2	-1
4	3	1

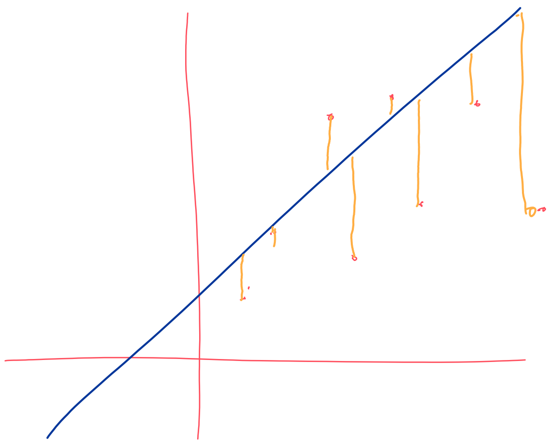
- Linear model predicts y is a linear combination of x_1, x_2

$$\hat{y} = \underbrace{w_0 + w_1 x_1 + w_2 x_2}_{\text{linear relation}}$$

A review of Linear Model for Regression

- Given the data

x_1	x_2	y
1	0	-2
2	1	0
3	-2	-1
4	3	1



- Linear model predicts y is a linear combination of x_1, x_2

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2$$

- The goal of linear model is to solve for w_0, w_1 and w_2
- To **train** a linear model is to find w_0, w_1 and w_2

How to find the coefficients?

How to find the coefficients?

- **Step 1**: Define the loss function $l(y, \hat{y})$

How to find the coefficients?

- **Step 1**: Define the loss function $l(y, \hat{y})$
- **Step 2**: Find w that minimizes the total loss function.

How to find the coefficients?

- Least Squared Method uses the **square loss**

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

How to find the coefficients?

- Least Squared Method uses the **square loss**

error^2

loss

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

- We want to find w_0, w_1 and w_2 that minimizes a **loss function**.

distance

x_1	x_2	y	$\hat{y} = w_0 + w_1x_1 + w_2x_2$	$(\hat{y} - y)^2$
1	0	-2	$w_0 + w_1 \cdot 1 + w_2 \cdot 0$	$(w_0 + w_1 \cdot 1 + w_2 \cdot 0 + 2)^2$ 1
2	1	0	$w_0 + w_1 \cdot 2 + w_2 \cdot 1$	$(w_0 + w_1 \cdot 2 + w_2 \cdot 1 - 0)^2$ 2
3	-2	-1	$w_0 + w_1 \cdot 3 + w_2 \cdot -2$	$(w_0 + w_1 \cdot 3 + w_2 \cdot -2 + 1)^2$ 3
4	3	1	$w_0 + w_1 \cdot 4 + w_2 \cdot 3$	$(w_0 + w_1 \cdot 4 + w_2 \cdot 3 - 1)^2$ 4

How to find the coefficients?

$w = 0$

$\boxed{l(w) = w^2 + w + 2}$

$l'(w) = 2w + 1$

$= 0 \Rightarrow 2w = -1/2$

- Least Squared Method uses the **square loss**

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

- We want to find w_0, w_1 and w_2 that minimizes a **loss function**.

x_1	x_2	y	$\hat{y} = w_0 + w_1 x_1 + w_2 x_2$	$(\hat{y} - y)^2$	
1	0	-2	$w_0 + w_1 \cdot 1 + w_2 \cdot 0$	$(w_0 + w_1 \cdot 1 + w_2 \cdot 0 + 2)^2$	1
2	1	0	$w_0 + w_1 \cdot 2 + w_2 \cdot 1$	$(w_0 + w_1 \cdot 2 + w_2 \cdot 1 - 0)^2$	2
3	-2	-1	$w_0 + w_1 \cdot 3 + w_2 \cdot -2$	$(w_0 + w_1 \cdot 3 + w_2 \cdot -2 + 1)^2$	3
4	3	1	$w_0 + w_1 \cdot 4 + w_2 \cdot 3$	$(w_0 + w_1 \cdot 4 + w_2 \cdot 3 - 1)^2$	4

- The total loss function:

$$L = L(w_0, w_1, w_2) = (w_0 + w_1 + 2)^2 + (w_0 + 2w_1 + w_2)^2 + (w_0 + 3w_1 - 2w_2 + 1)^2 + (w_0 + 4w_1 + 3w_2 - 1)^2$$

$$\begin{cases} \frac{\partial L}{\partial w_0} = 0 \\ \frac{\partial L}{\partial w_1} = 0 \\ \frac{\partial L}{\partial w_2} = 0 \end{cases}$$

How to find the coefficients?

- Least Squared Method uses the **square loss**

$|w| + w + 2$

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

- We want to find w_0, w_1 and w_2 that minimizes a **loss function**.

- The total loss function:

$$\begin{aligned} L = L(w_0, w_1, w_2) &= (w_0 + w_1 + 2)^2 + (w_0 + 2w_1 + w_2)^2 \\ &+ (w_0 + 3w_1 - 2w_2 + 1)^2 + (w_0 + 4w_1 + 3w_2 - 1)^2 \end{aligned}$$

- Solve for the partial derivatives equaling 0 to find w_0, w_1 and w_2.

How about other loss functions?

- Absolute loss:

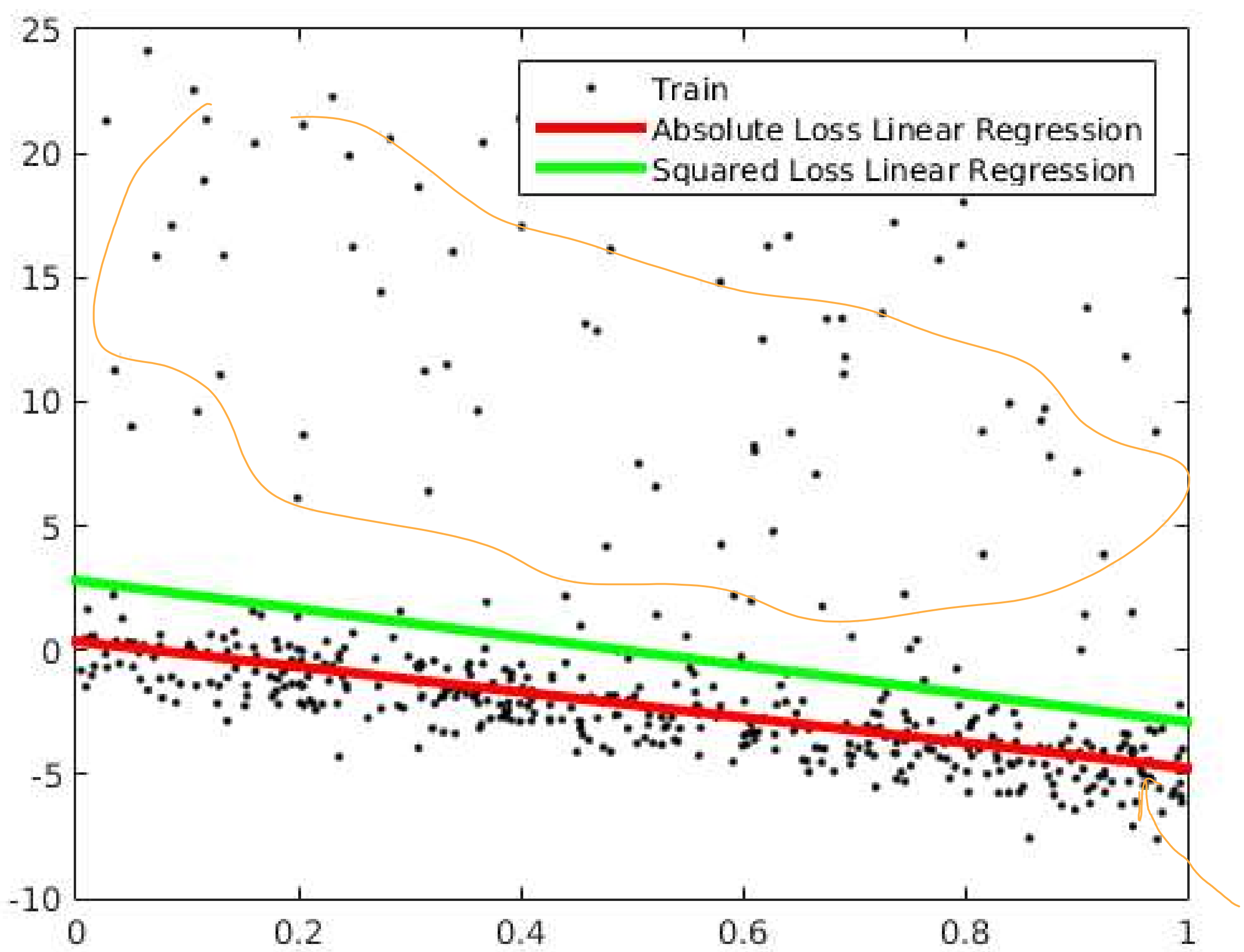
$$L(\hat{y}, y) = |\hat{y} - y|$$

- The total loss function:

$$\begin{aligned} L = L(w_0, w_1, w_2) &= |w_0 + w_1 + 2| + |w_0 + 2w_1 + w_2| \\ &+ |w_0 + 3w_1 - 2w_2 + 1| + |w_0 + 4w_1 + 3w_2 - 1| \end{aligned}$$

- Use Linear Programming to find w_0, w_1 and w_2 that minimizes the total loss.
- Least absolute deviations regression

Linear Models



How about other loss functions?

Ordinary least squares regression	Least absolute deviations regression
Not very robust	Robust
Stable solution	Unstable solution
Always one solution	Possibly multiple solutions

A general framework

- **Problem**: Given the data of $x_1, x_2, \ldots, x_d, y$, establish the *best* relation between y and $x = [x_1, x_2, \ldots, x_d]$.
- A solution framework:
    - Step 1: Assume the model function $\hat{y} = f(x, w)$, where w is a parameter vector.
    - Step 2: Define the loss function $l(y, \hat{y})$
    - Step 3: Find w that minimizes the loss function using gradient descent

LASSO

- Consider a linear model

$$y = 100x_1 + 0.01x_2 + 50x_3 - 0.002x_4$$

- x_2 and x_4 are not important because the coefficients are too small.
- We want to get rid of x_2 and x_4

(1) Linear model: minimize the loss function : $\boxed{L(w)}$

(2) LASSO

minimize $\boxed{L(w)}$

the magnitude w

$|w| \leq \beta$

LASSO - Principle

- LASSO forces the sum of the absolute value of the coefficients to be less than a fixed value.
- which forces certain coefficients (slopes) to be set to zero
- effectively making the model simpler

Linear Model vs. LASSO - Principle

- Linear Model minimizes

$$L(w_0, w_1, w_2) \quad \omega$$

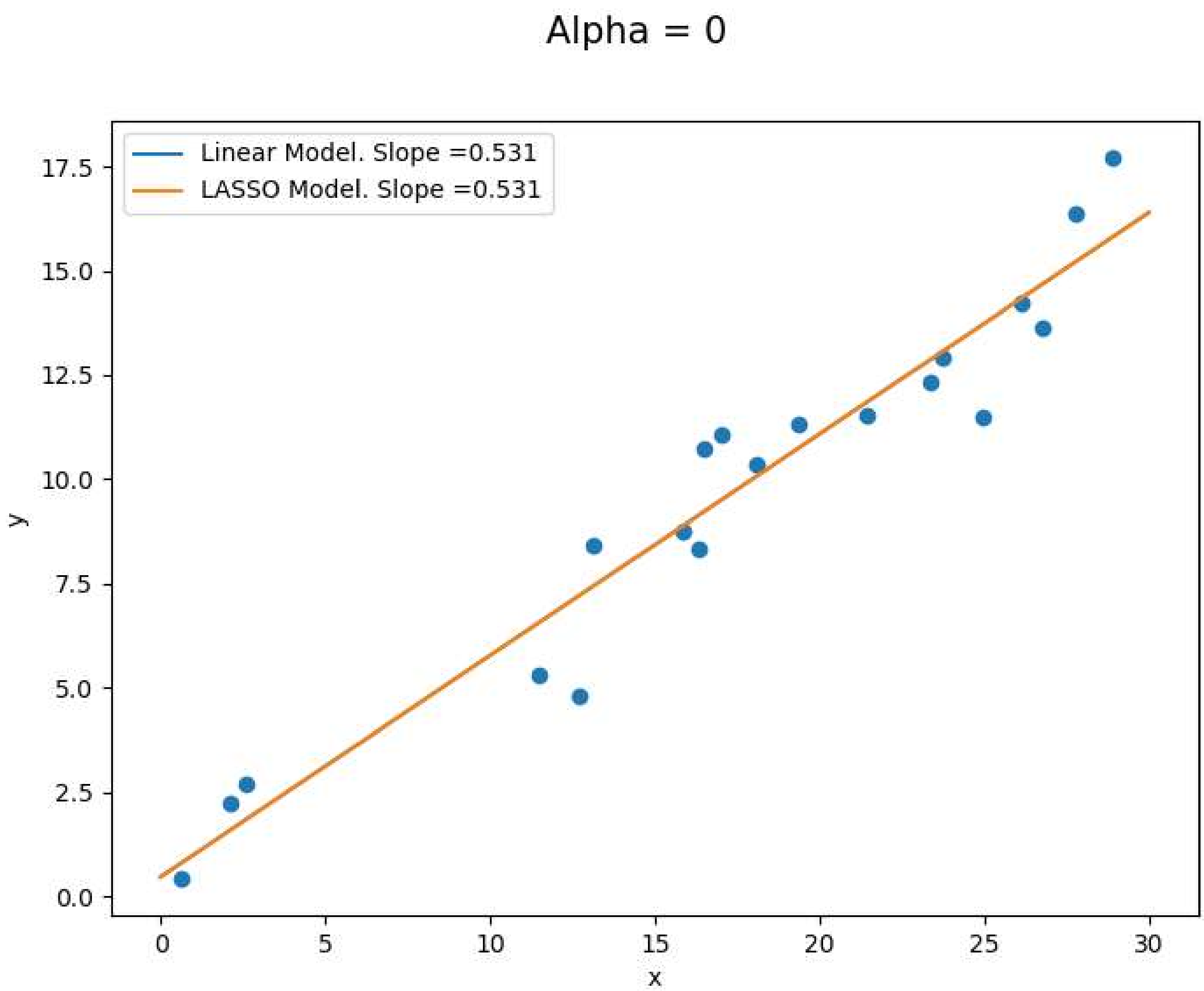
- LASSO minimizes

$$L(w_0, w_1, w_2) + \alpha \underbrace{\Big(|w_1| + |w_2| \Big)}_{\to 0}$$

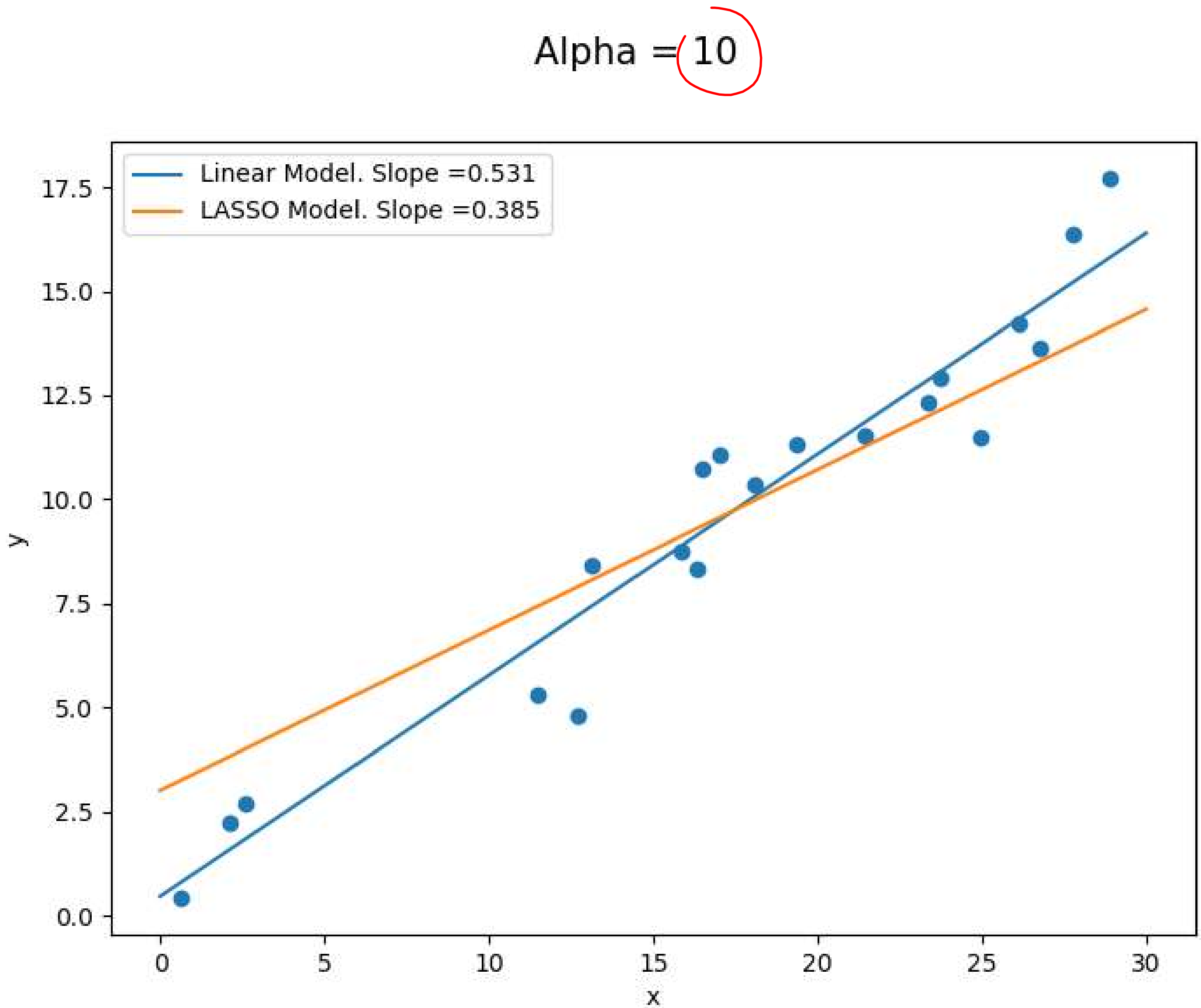
- The greater α, the easier w_1 and w_2 will be zeros.

- When $\alpha = 0$, LASSO is the linear model.

LASSO for Linear Model

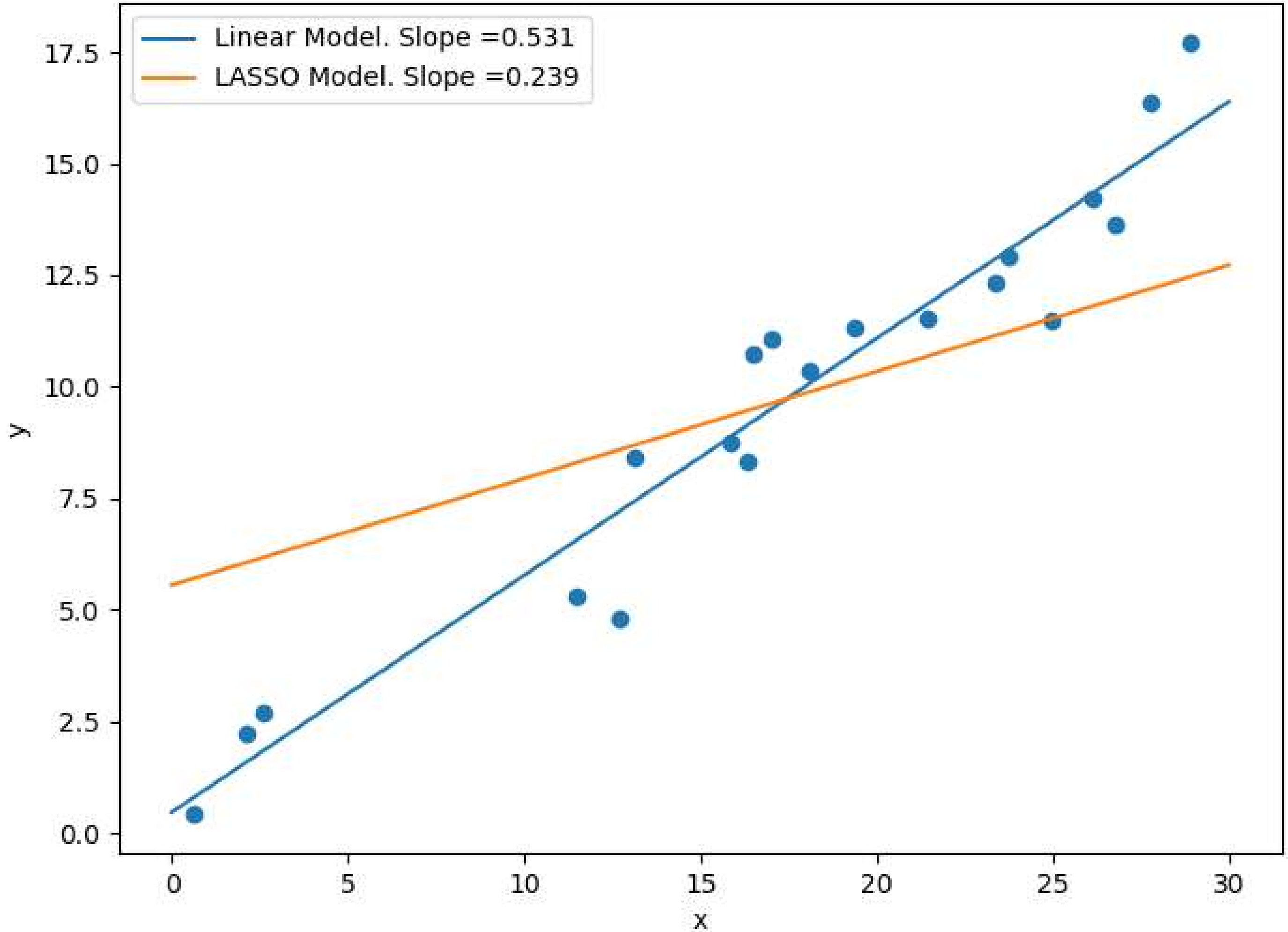


LASSO for Linear Model

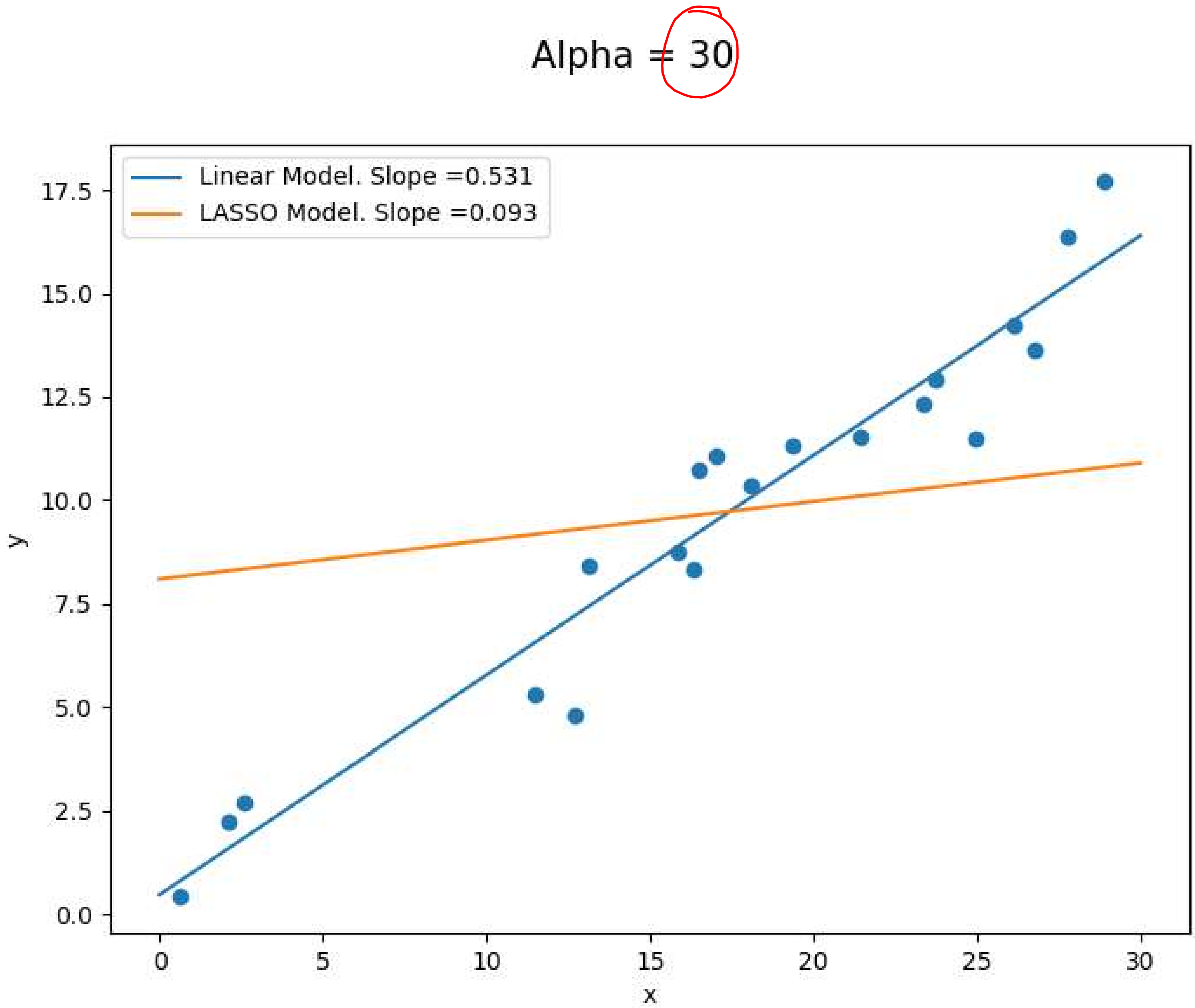


LASSO for Linear Model

Alpha = 20

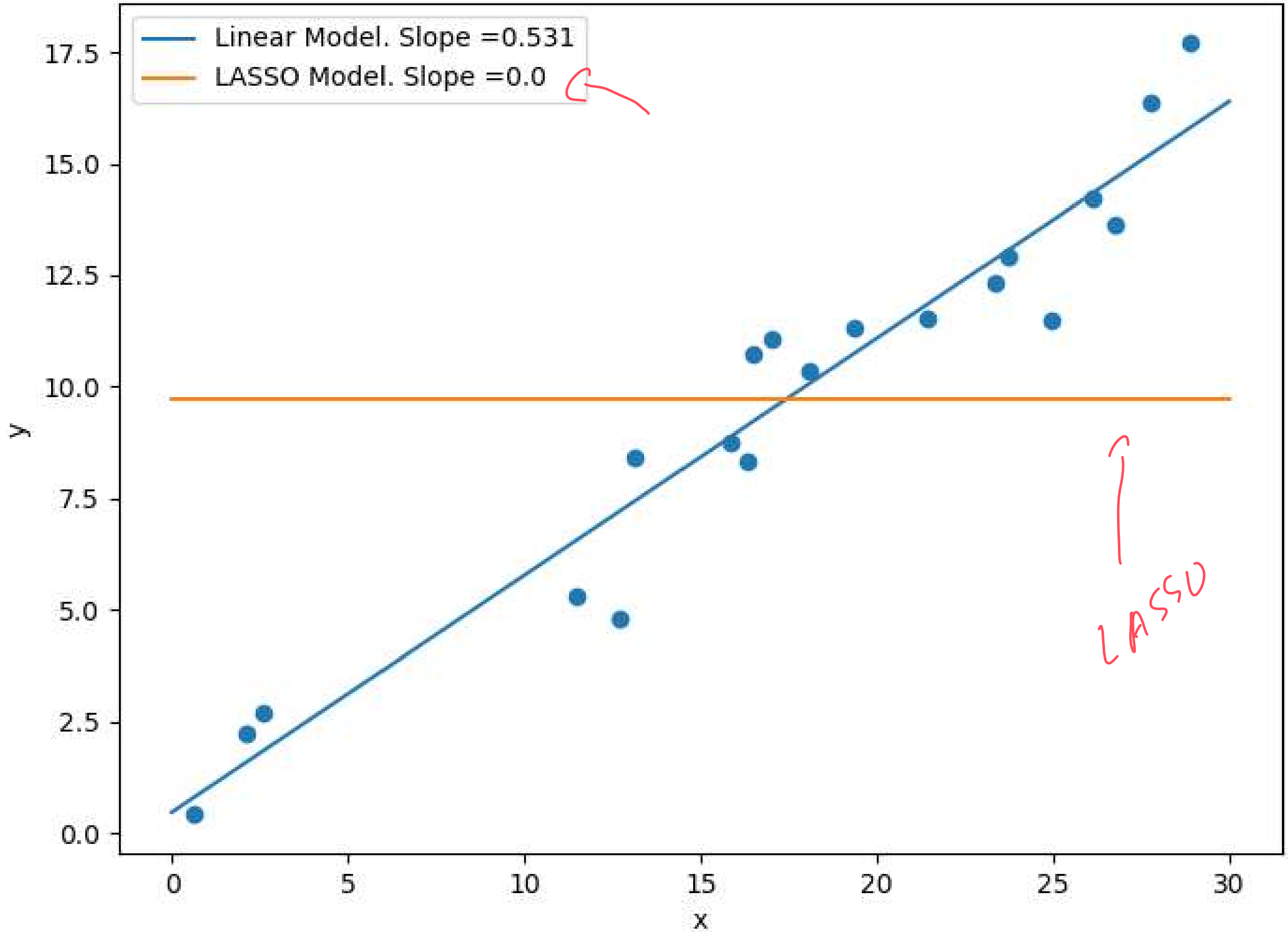


LASSO for Linear Model



LASSO for Linear Model

Alpha = 40



LASSO for Variables Selection

- Data

x_1	x_2	x_3	x_4	x_5	x_6	y
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...

- Assume that the truth relation between the input $x_1, x_2, x_3, x_4, x_5, x_6$ and the output y is

$$y = 4x_2 + 3x_4 + 7x_6$$

- We see that only x_2, x_4 and x_6 impact y

- LASSO can help to identify variables that have effect on y

LASSO for Variables Selection

- The result when training the linear model and the LASSO

	w_1	w_2	w_3	w_4	w_5	w_6
Truth	0	4	0	3	0	7
Linear Model	-0.244061	3.54013	0.221939	2.6042	0.0982158	6.83617
LASSO	-0	2.65623	0	1.84839	0	5.80624

- In Linear Model, x_1, x_3 and x_5 have effect on y (which is WRONG!)
- In LASSO, x_1, x_3 and x_5 have no effect on y (CORRECT!)
- LASSO can also be applied before another model.

Logistic Regression

x_1	x_2	y
1	0	1
2	1	0
3	-2	0
4	3	1

$y = w_0 + w_1 x_1 + w_2 x_2$

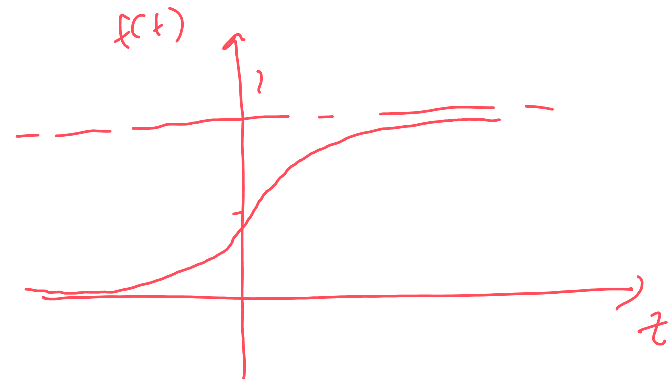
- How are y and x related?

Logistic Regression

x_1	x_2	y
1	0	1
2	1	0
3	-2	0
4	3	1

Logistic function

$$\boxed{f(t) = \frac{1}{1 + e^{-t}}}$$



- Logistic Regression models $P(y = 1|x) = \hat{y}$ as:

(1)
$$\hat{y} = \frac{1}{1 + e^{-\overbrace{(w_0 + w_1 \cdot x_1 + w_2 \cdot x_2)}^{t}}}$$

- OR,

$\Leftrightarrow$

(2)
$$\log\left(\frac{\hat{y}}{1 - \hat{y}}\right) = w_0 + w_1 \cdot x_1 + w_2 \cdot x_2$$

where $\hat{y}$ is the predicted value of the probability of $y = 1$ given x_1 and x_2.

Logistic Regression

x_1	x_2	y
1	0	1
2	1	0
3	-2	0
4	3	1

- Logistic Regression models $P(y=1|x)=\hat{y}$ as:

$$\hat{y}=\frac{1}{1+e^{-(w_0+w_1\cdot x_1+w_2\cdot x_2)}}$$

- OR,

$$\log\left(\frac{\hat{y}}{1-\hat{y}}\right)=w_0+w_1\cdot x_1+w_2\cdot x_2$$

where $\hat{y}$ is the predicted value of the probability of $y=1$ given x_1 and x_2.

Logistic Regression

$$\log\left(\frac{\hat{y}}{1-\hat{y}}\right) = w_0 + w_1 \cdot x_1 + w_2 \cdot x_2$$

- $\left(\frac{\hat{y}}{1-\hat{y}}\right)$ is also called odd-ratio.
- Logistic Regression assumes that the log of the odd ratio is linear.

How to find w_0, w_1, w_2?

- **Step 1**: Define the loss function $l(\hat{y}, y)$
- **Step 2**: Find w that minimizes the total loss function

Logistic Regression

- Define the loss function: We use the log-loss or cross-entropy loss function

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

not the square loss

- Total Loss:

$$\begin{aligned} L(w_0, w_1, w_2) = &- \log\left(\frac{1}{1 + e^{-w_0 - w_1}}\right) \\ &- \log\left(1 - \frac{1}{1 + e^{-w_0 - 2w_1 - w_2}}\right) \\ &- \log\left(1 - \frac{1}{1 + e^{-w_0 - 3w_1 + w_2}}\right) \\ &- \log\left(\frac{1}{1 + e^{-w_0 - 4w_1 - 3w_2}}\right) \end{aligned}$$

$\frac{\partial}{\partial w_0} = 0$

$\frac{\partial}{\partial w_1} = 0$

$\frac{\partial}{\partial w_2} = 0$

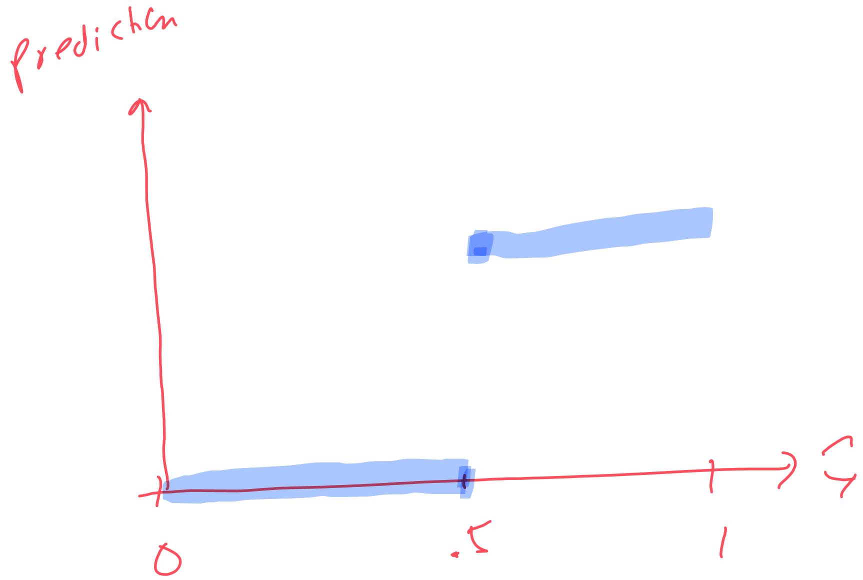
- We need to find w_0, w_1, w_2 that minimizes the total loss

$\hat{y}$	y
.6	1
.3	1
.2	0

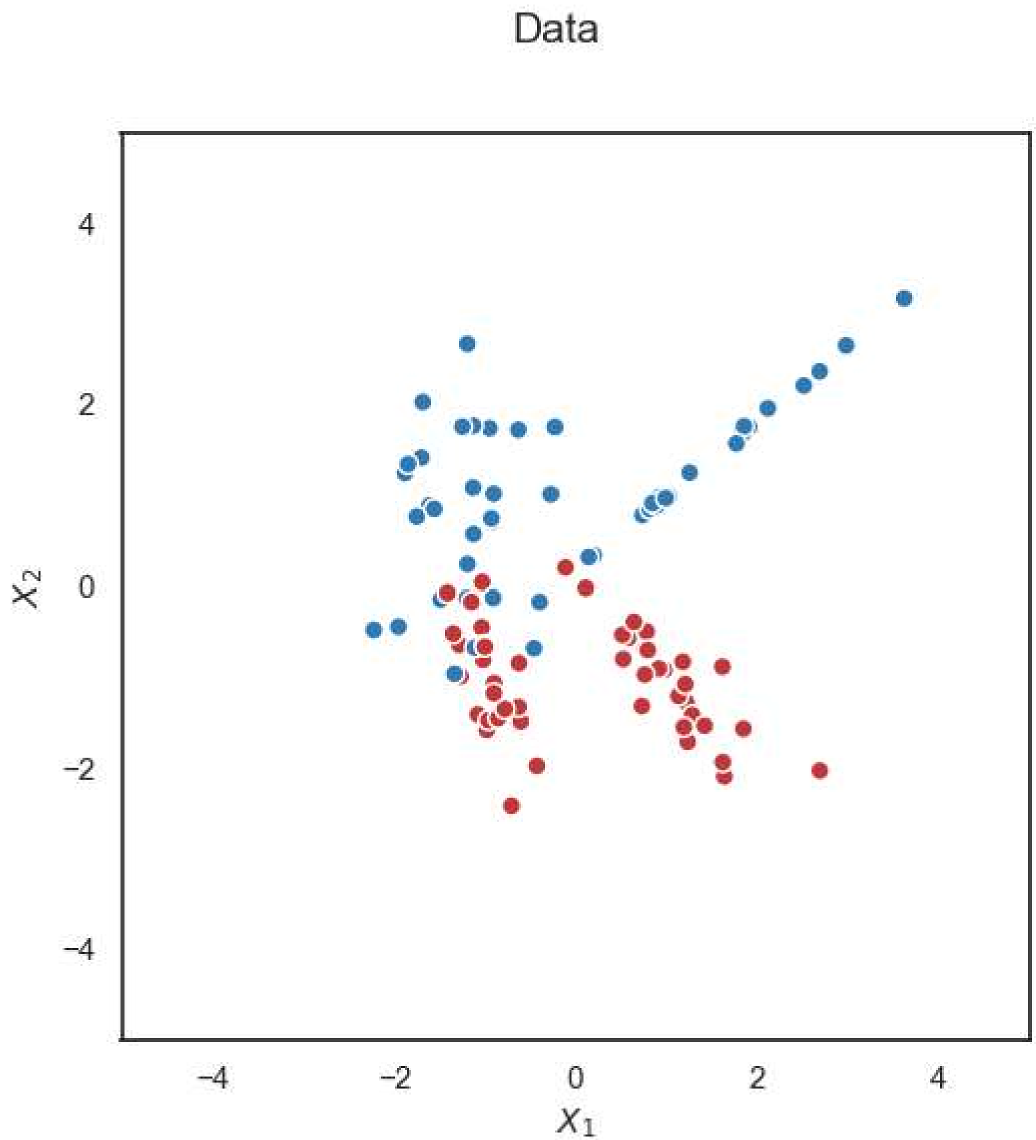
$\rightarrow$

predicted	true
1	1
0	1
0	0

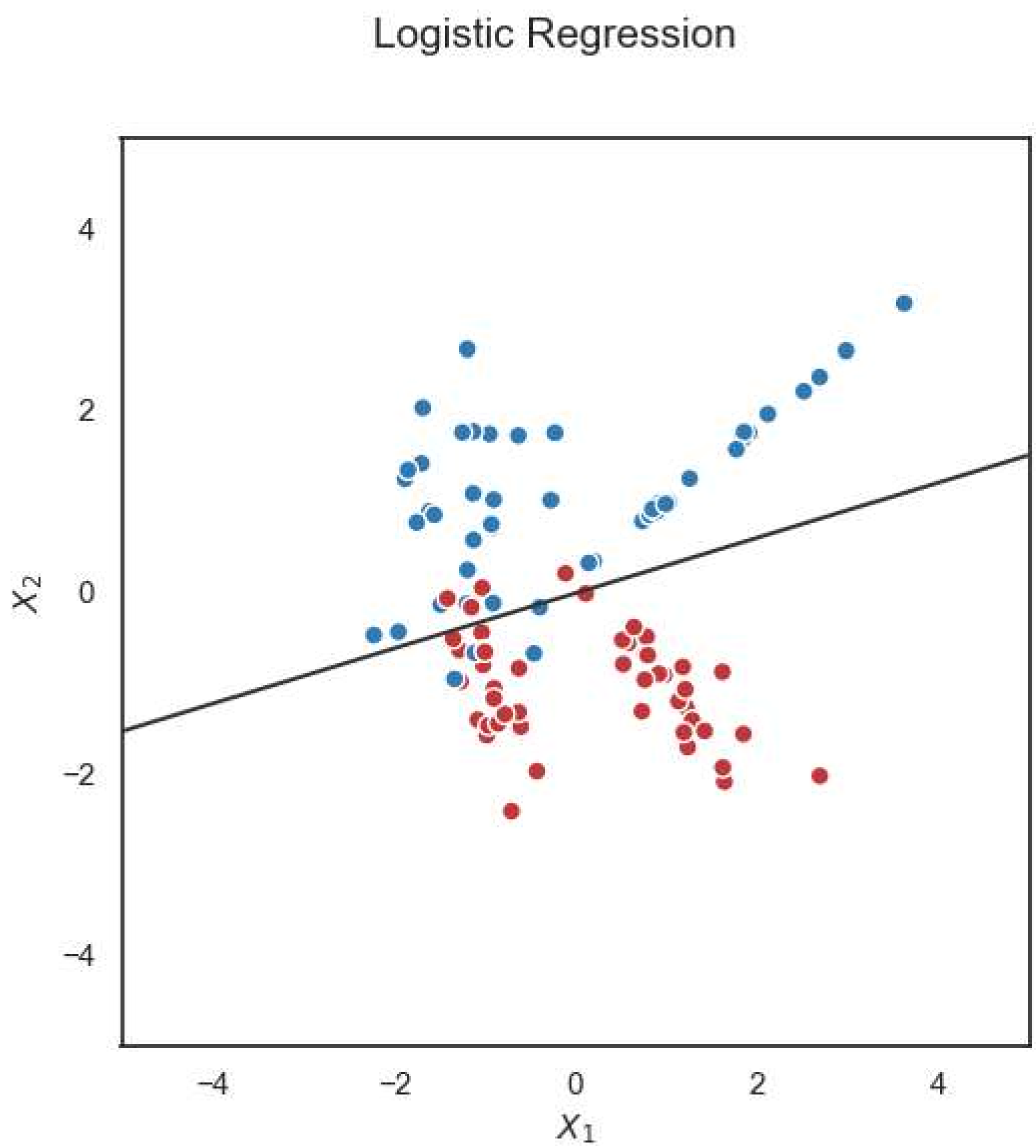
$\rightarrow$ mis classification = 1/3


prediction
0
.5
1
y

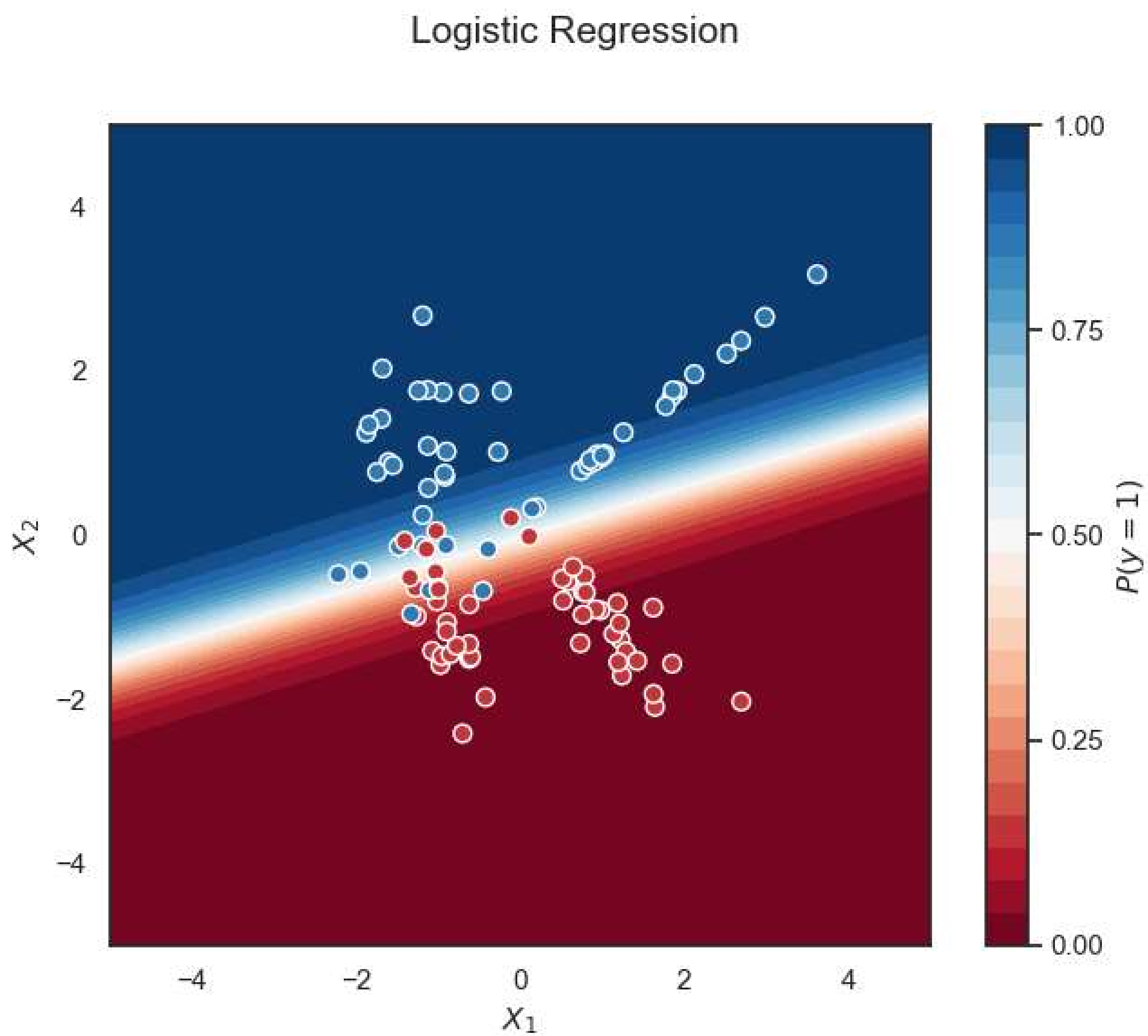
Logistic Regression



Logistic Regression



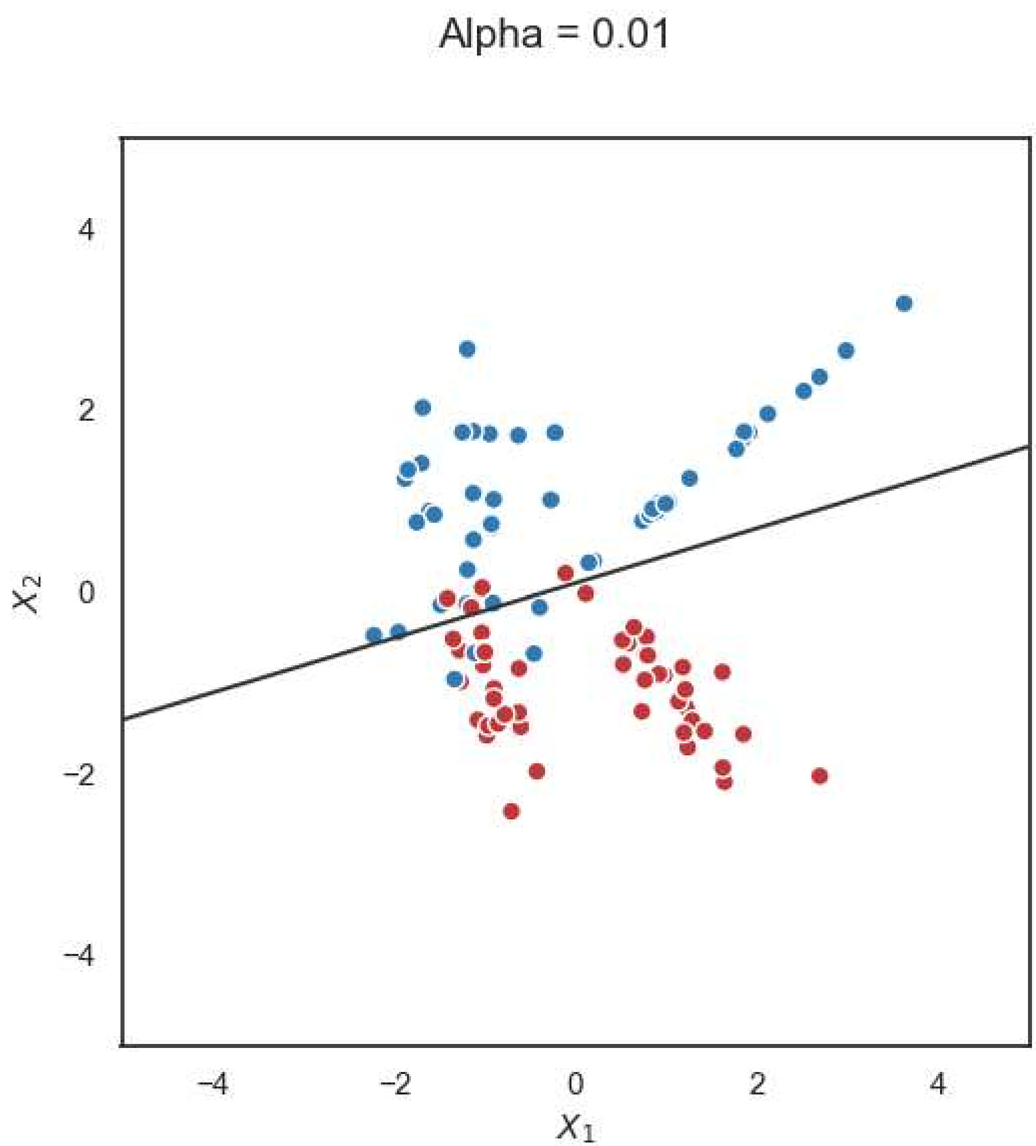
Logistic Regression



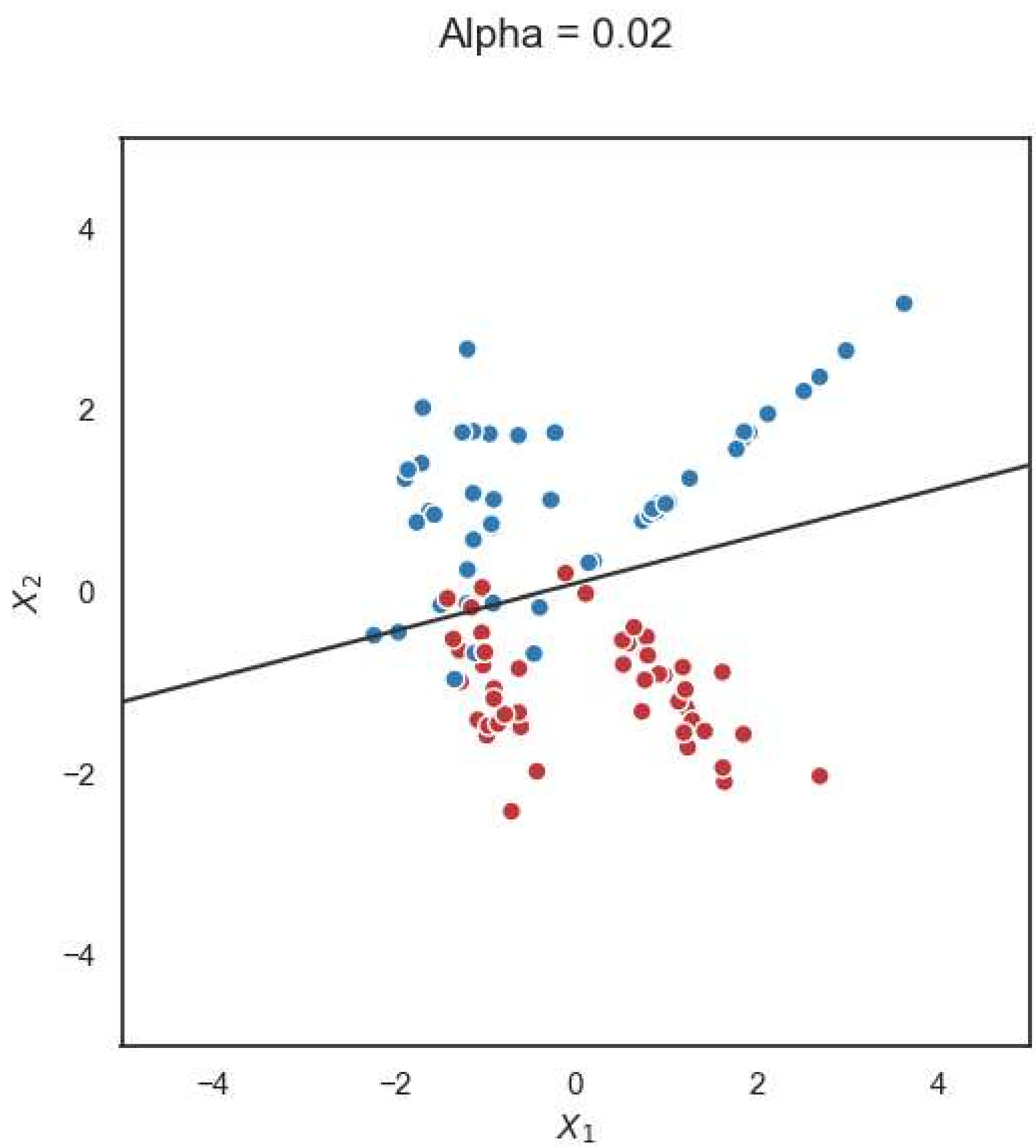
LASSO for Logistic Regression

- The idea is the same as for linear model

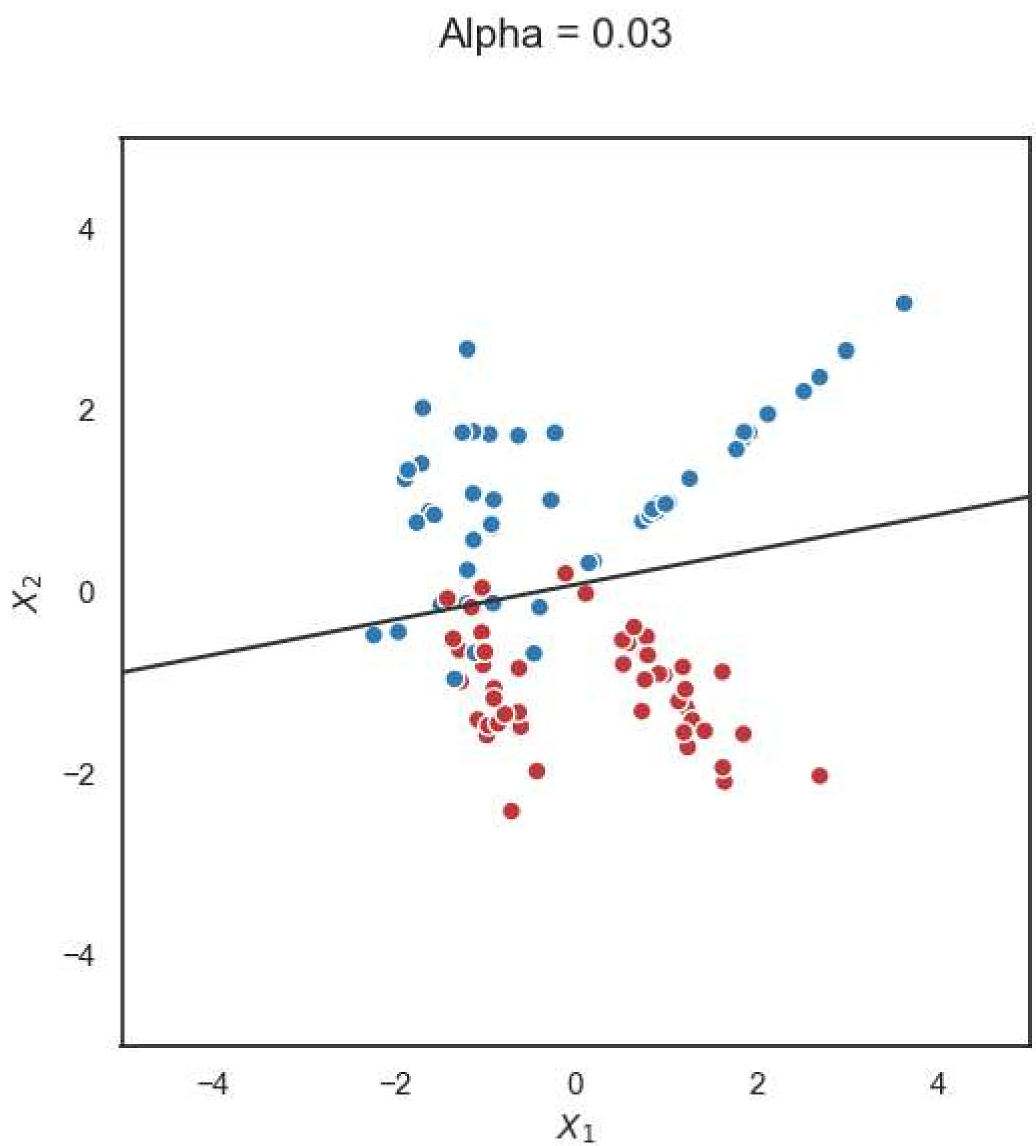
LASSO for Logistic Regression



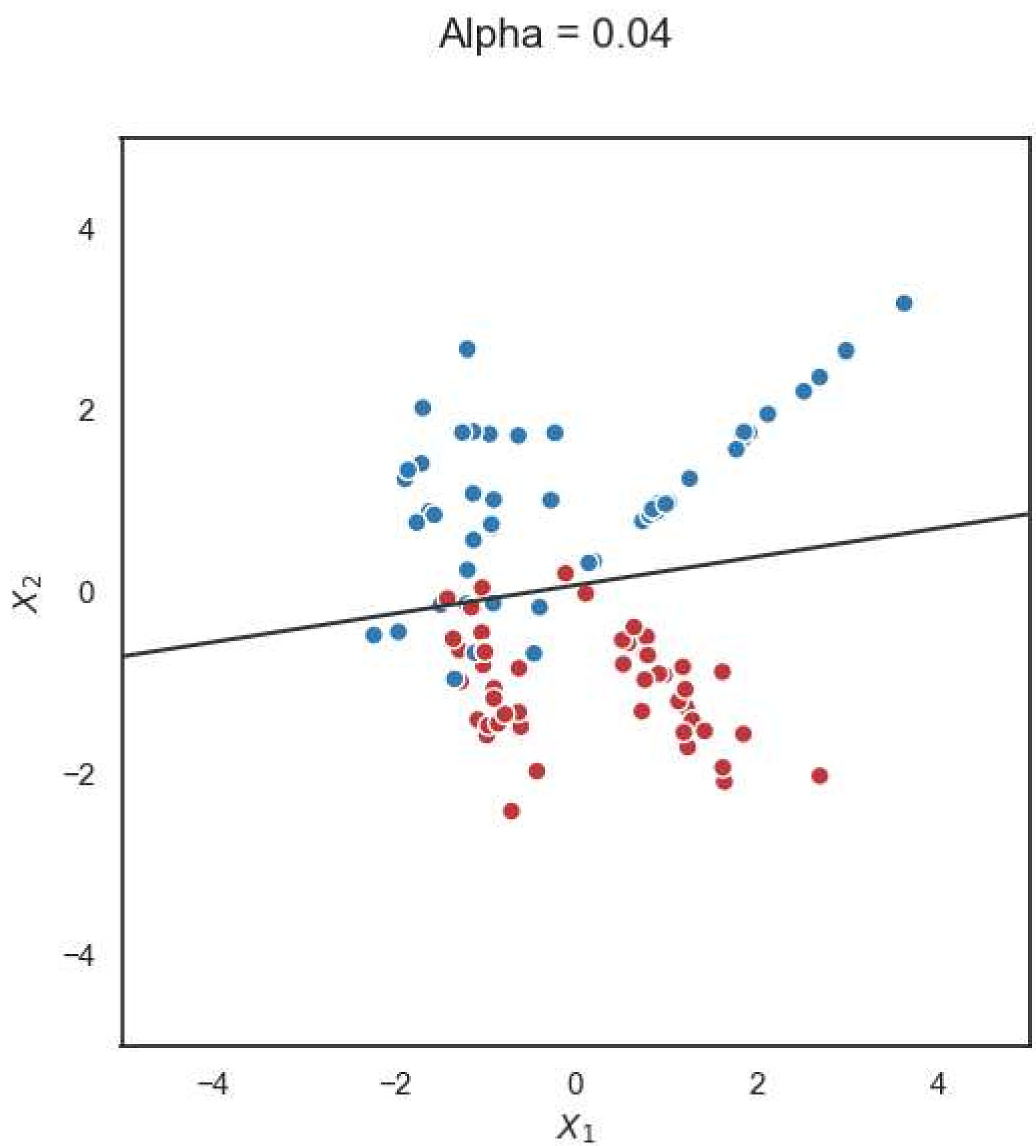
LASSO for Logistic Regression



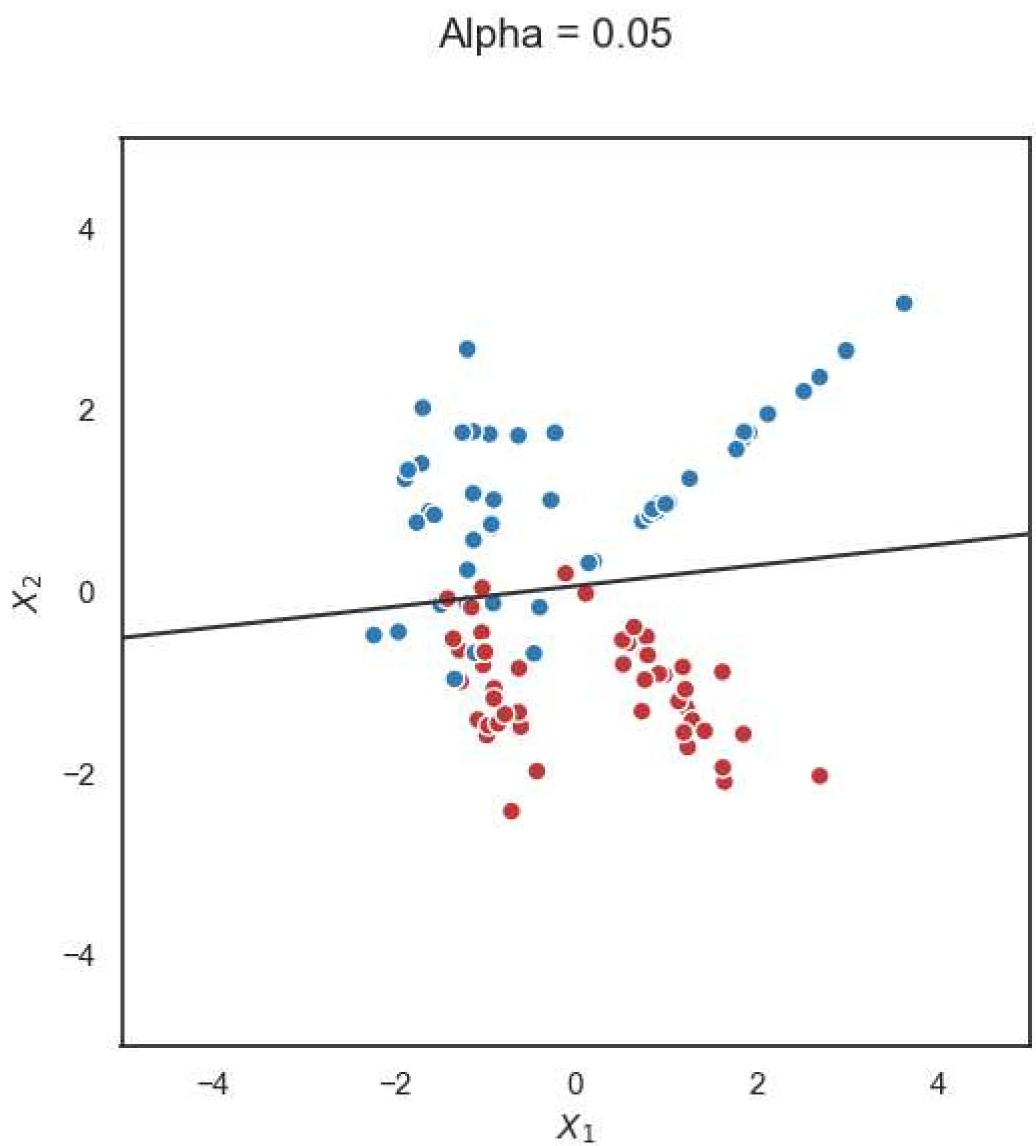
LASSO for Logistic Regression



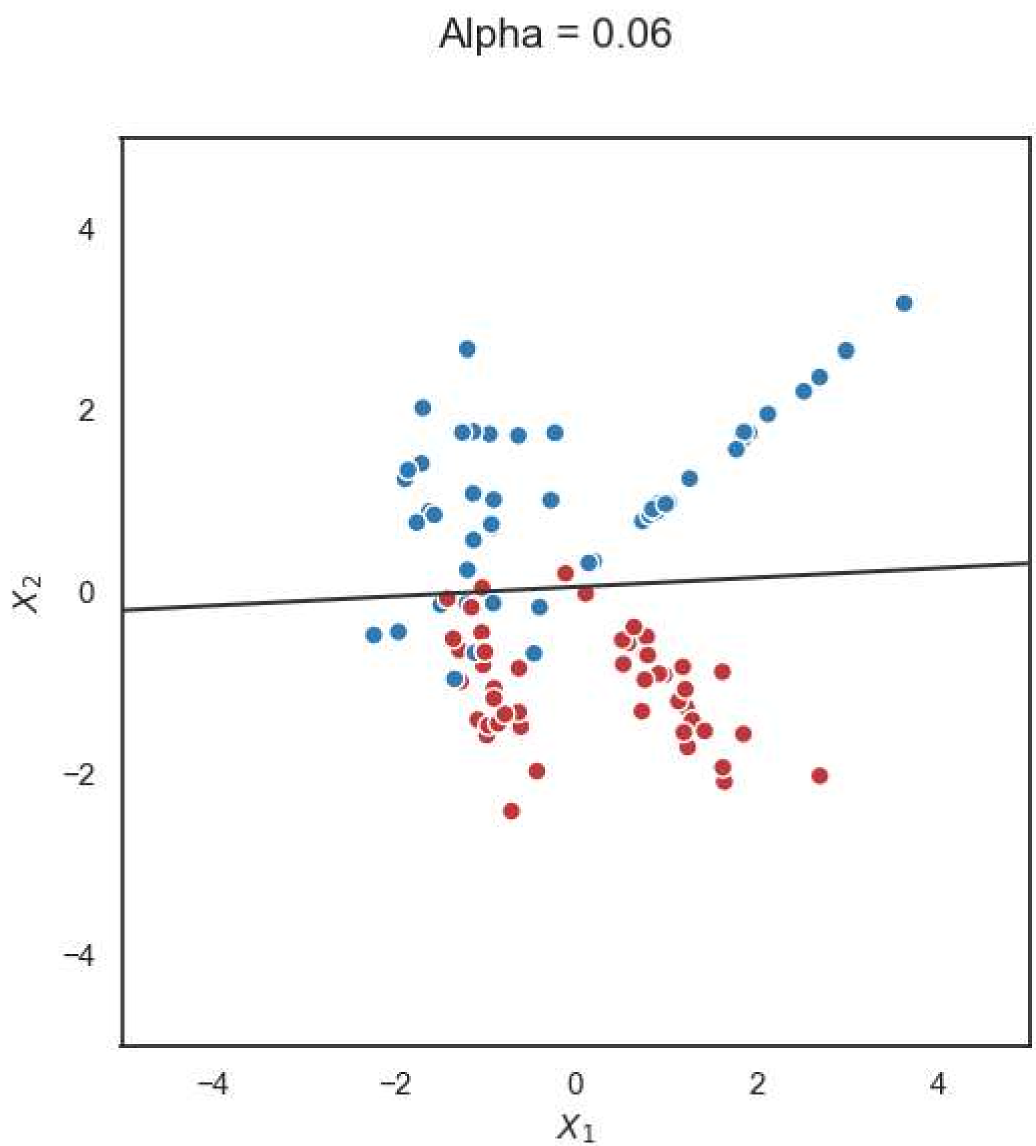
LASSO for Logistic Regression



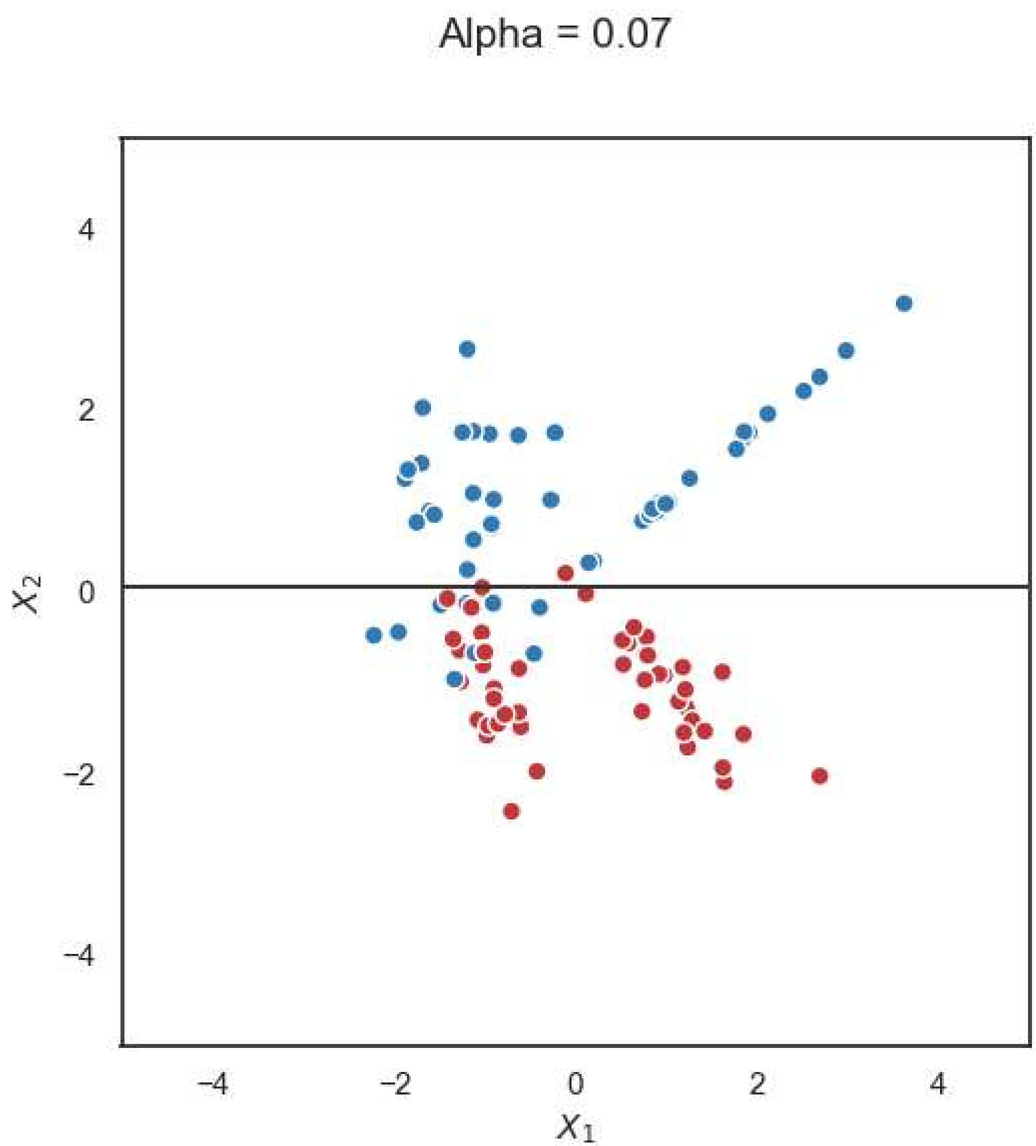
LASSO for Logistic Regression



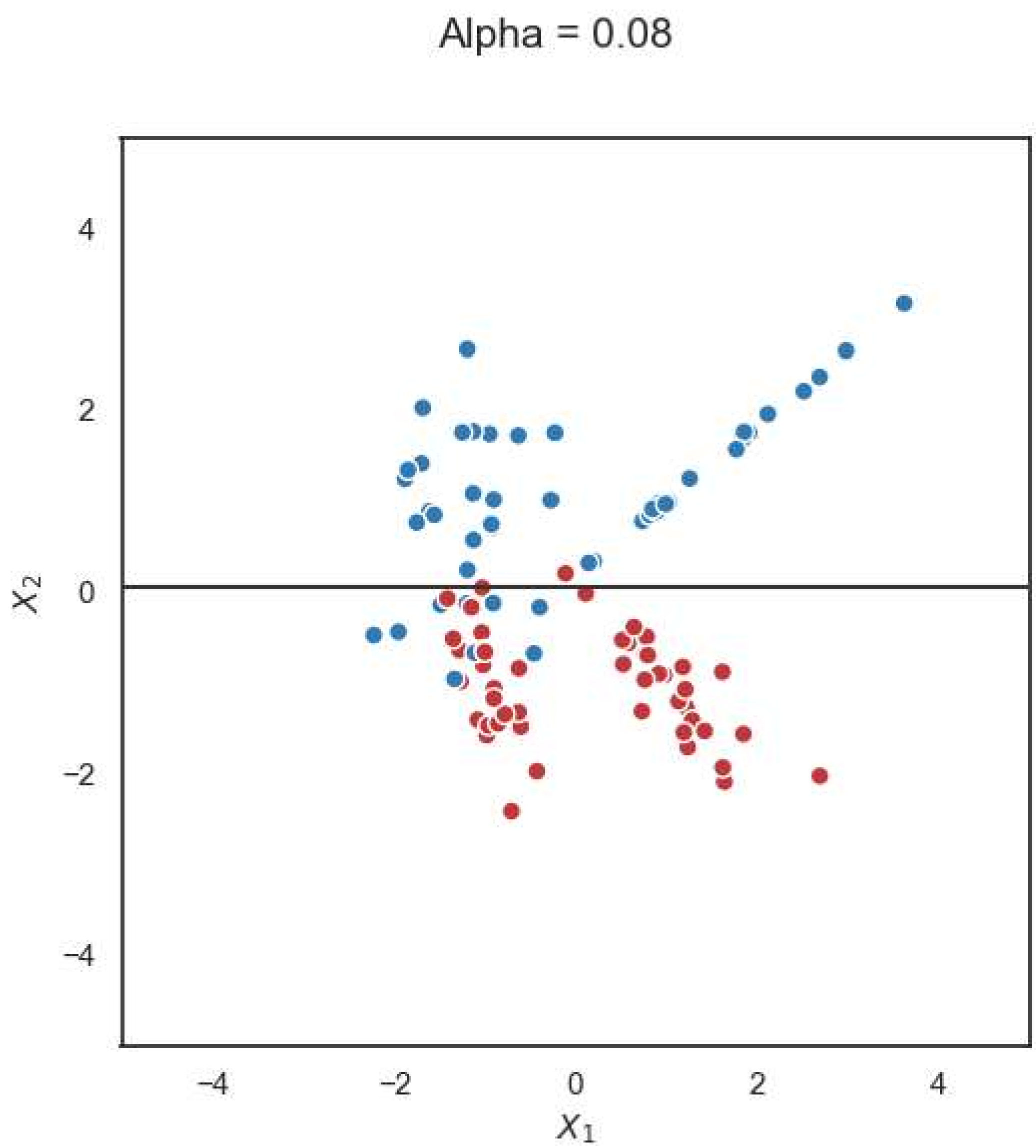
LASSO for Logistic Regression



LASSO for Logistic Regression

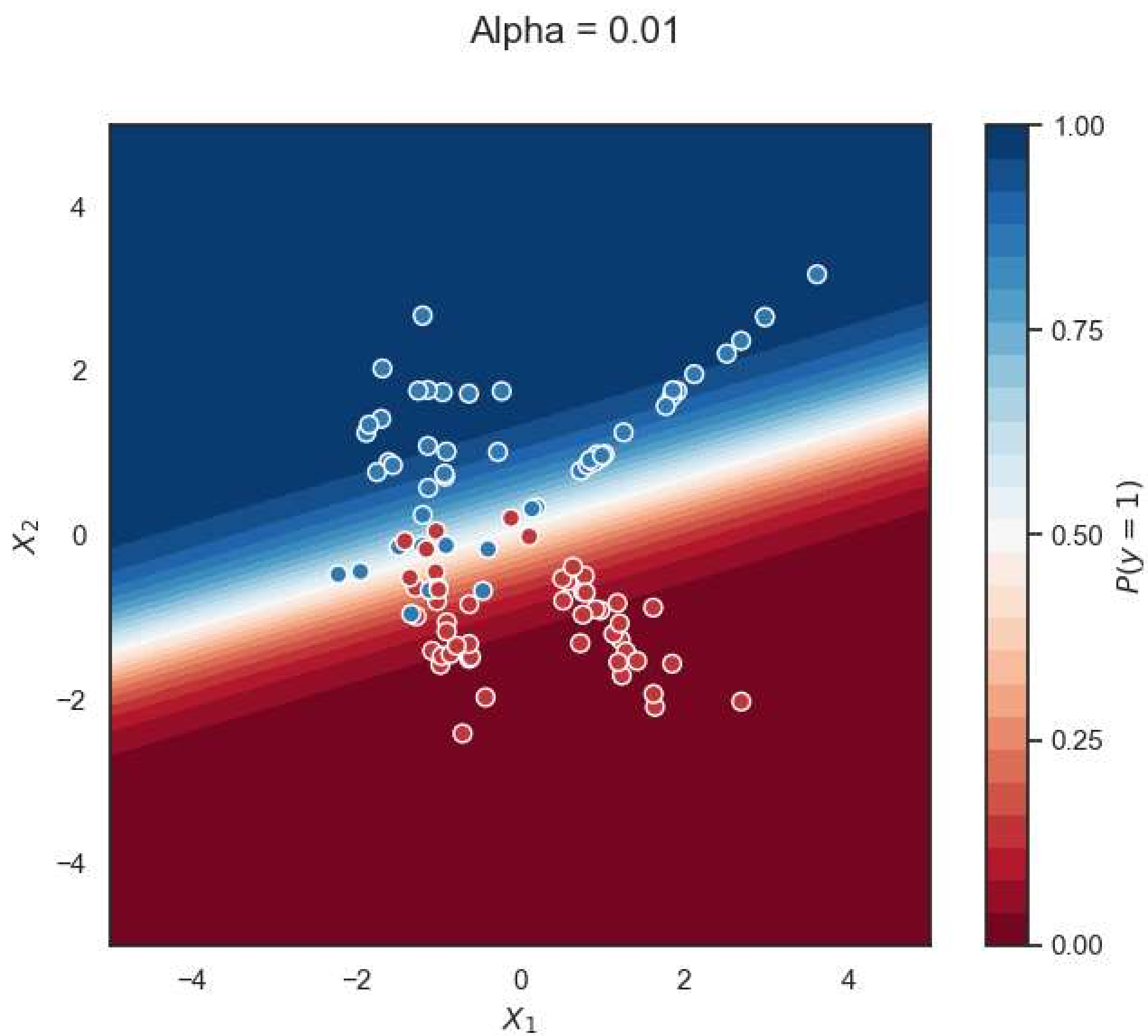


LASSO for Logistic Regression

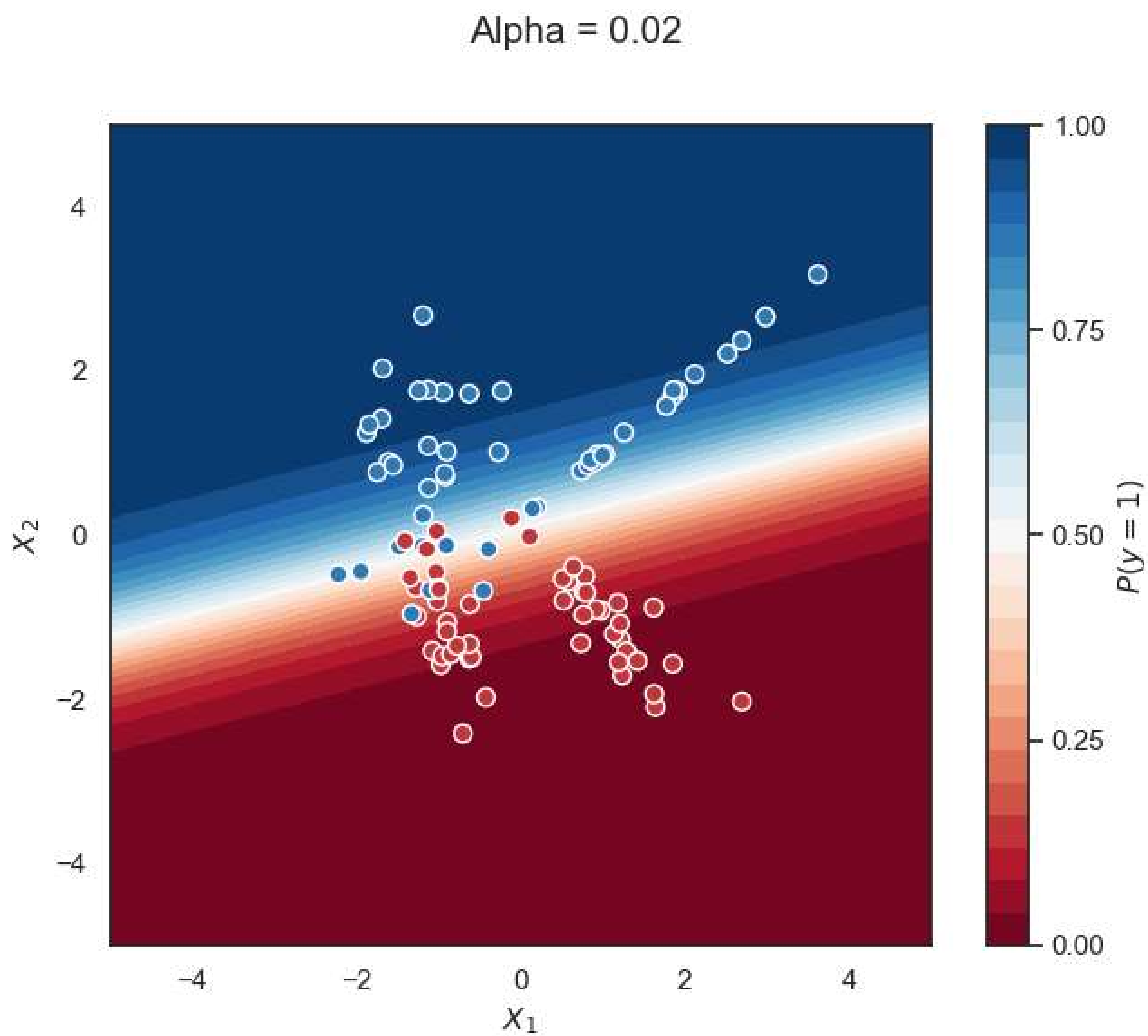


LASSO for Logistic Regression

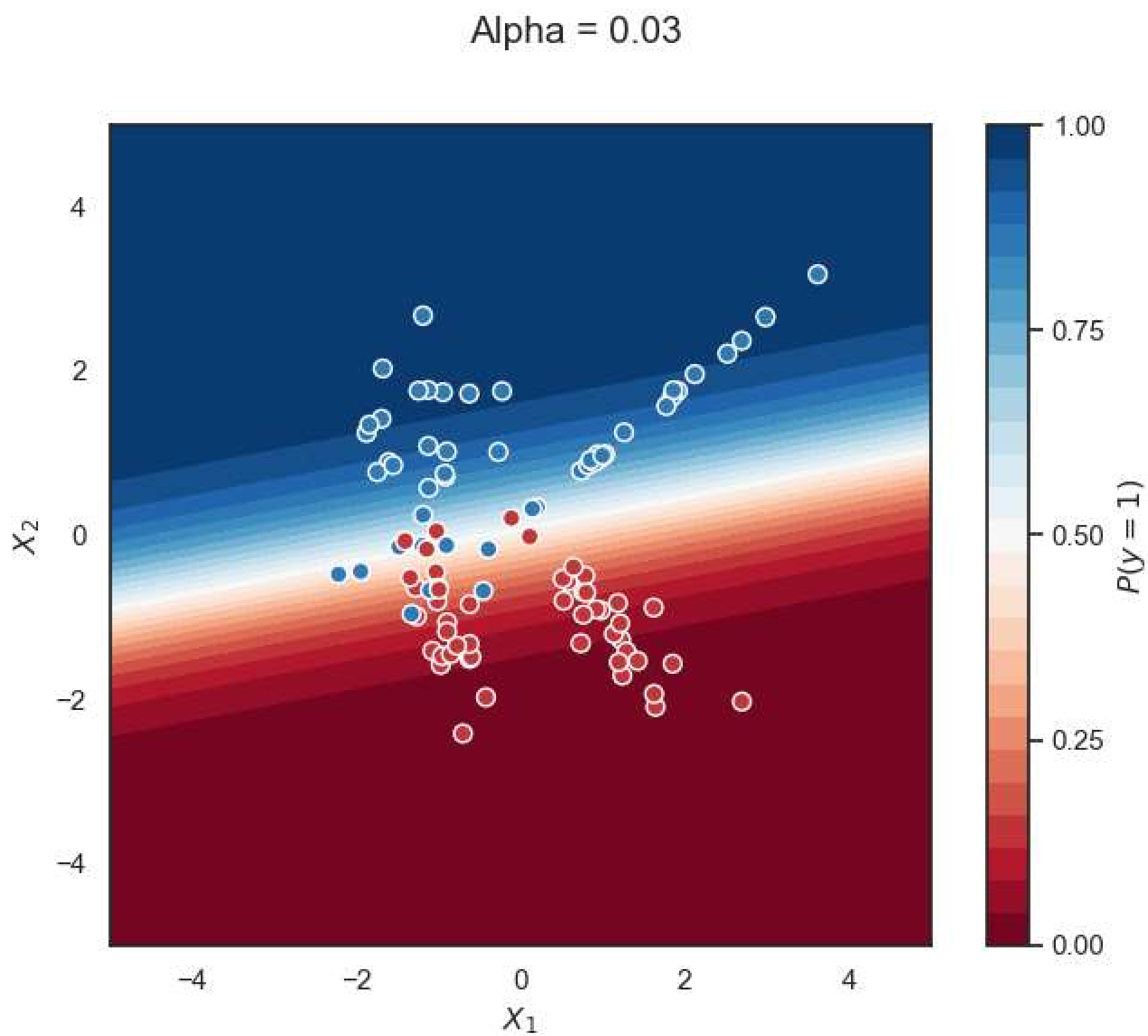
LASSO for Logistic Regression



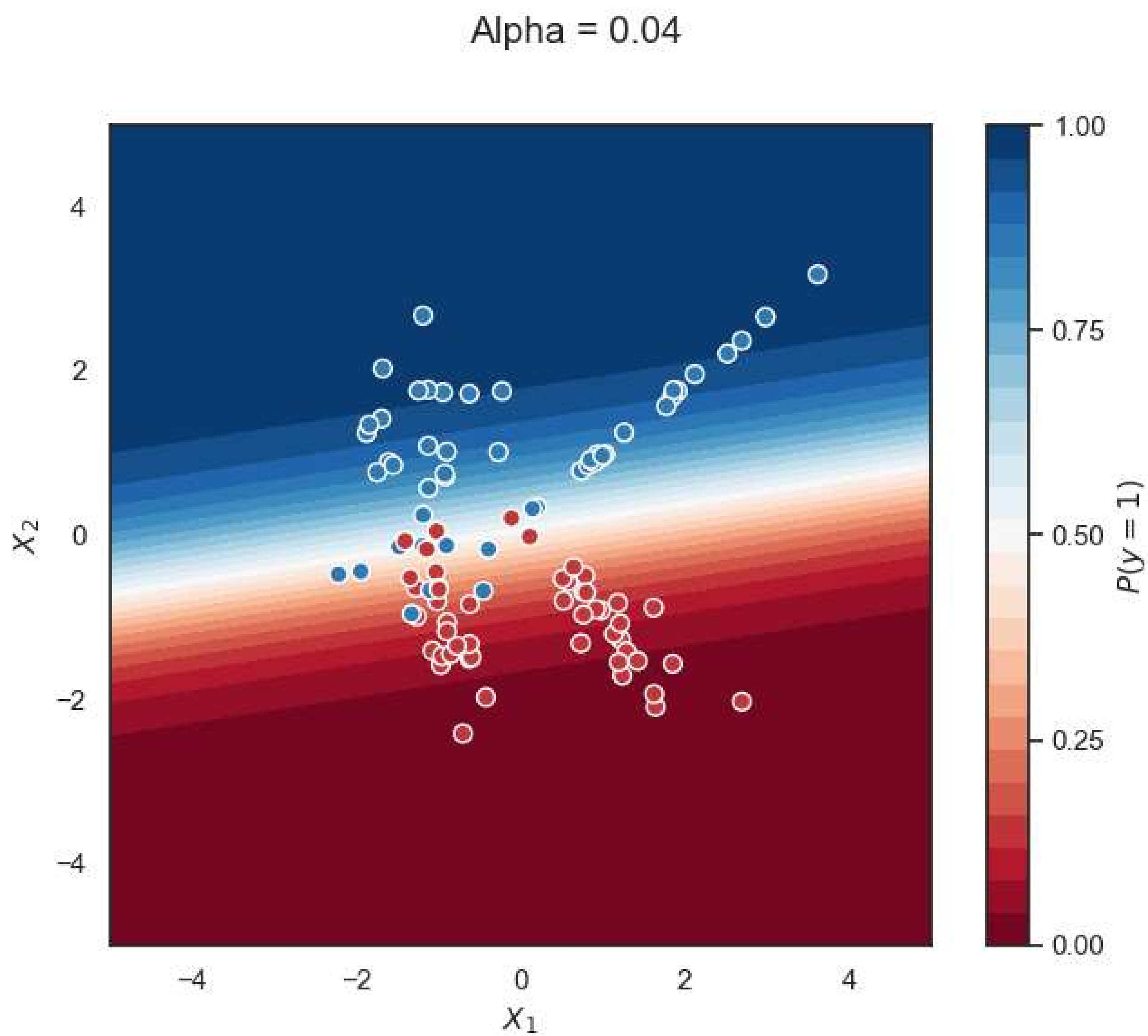
LASSO for Logistic Regression



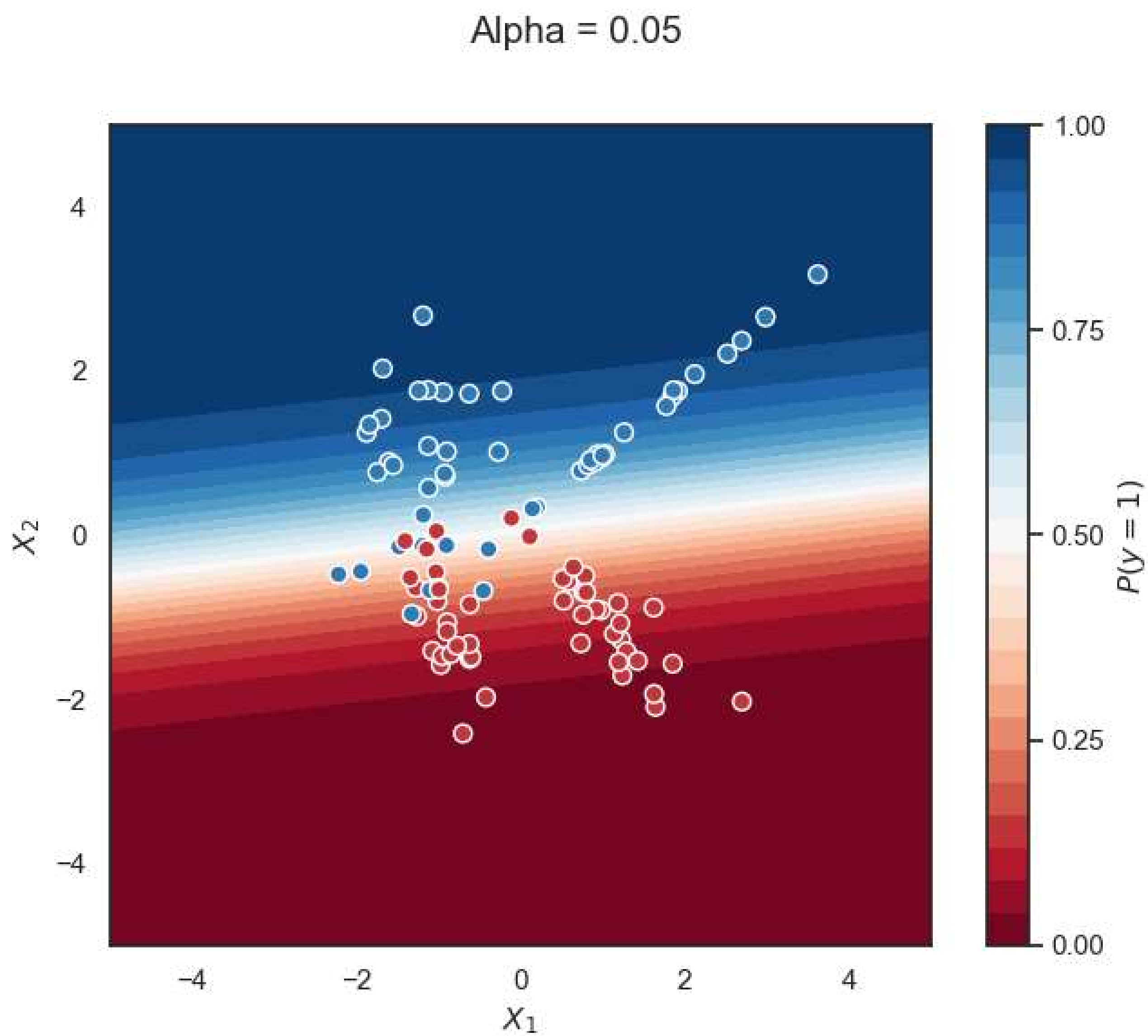
LASSO for Logistic Regression



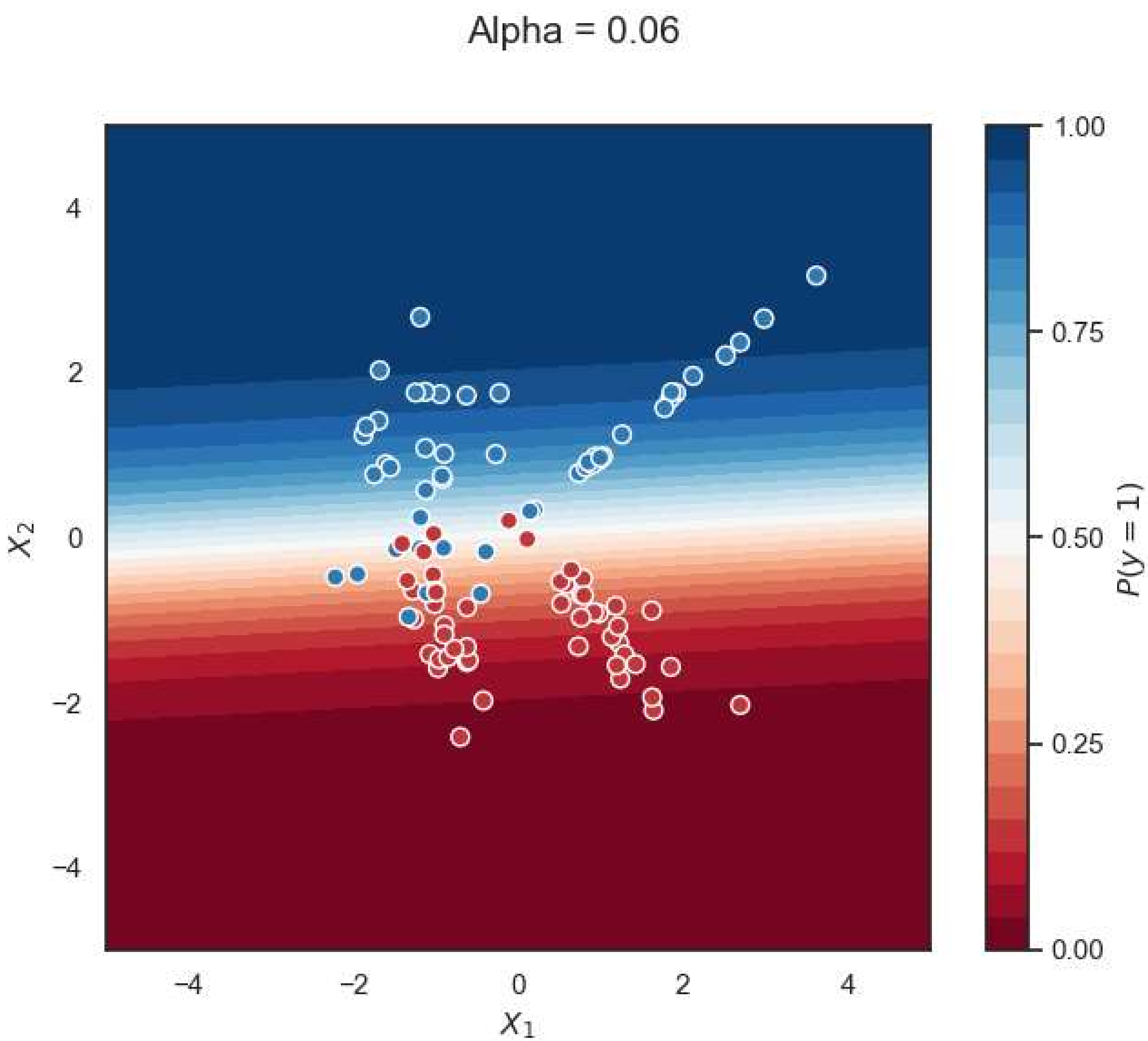
LASSO for Logistic Regression



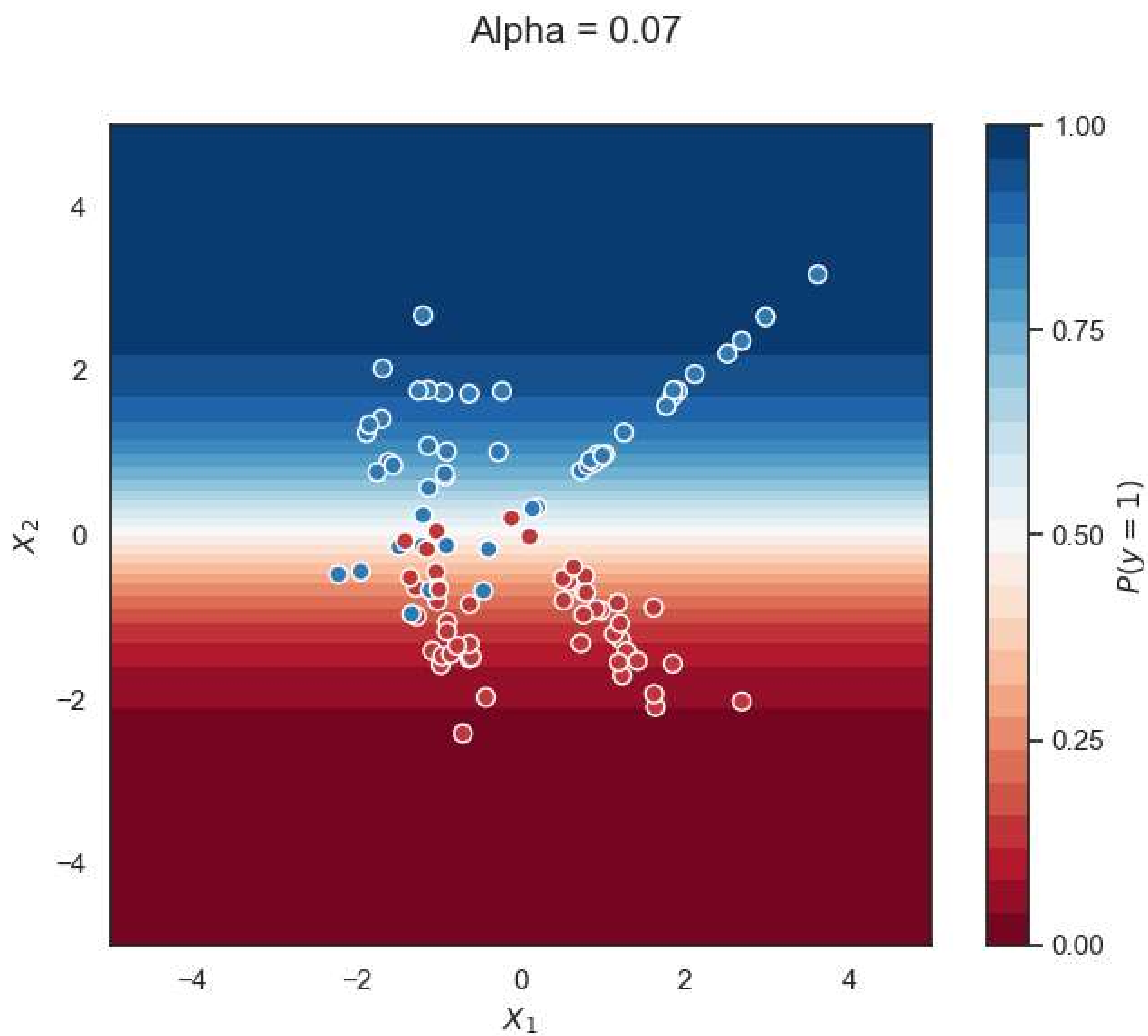
LASSO for Logistic Regression



LASSO for Logistic Regression



LASSO for Logistic Regression



LASSO for Logistic Regression