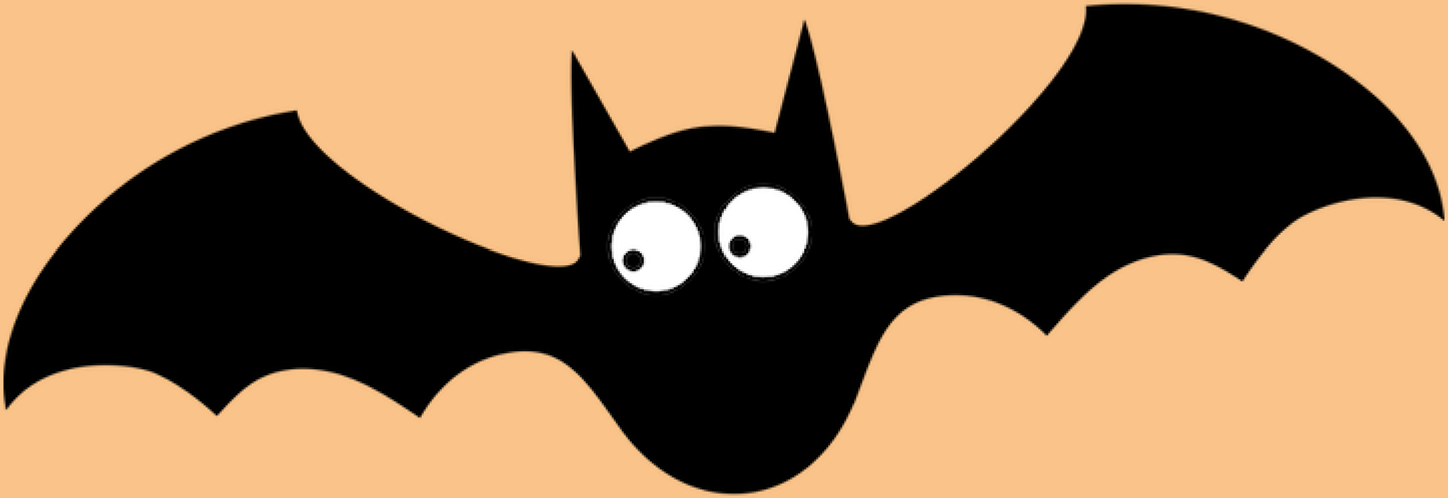


Linear Model

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 $\boldsymbol{y}$ and $\boldsymbol{x}$ 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

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

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?

- Define the loss function: We will use the **square loss**

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

How to find the coefficients?

- Define the loss function: We will use 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	$w_0 + w_1 \cdot 1 + w_2 \cdot 0$	$(w_0 + w_1 \cdot 1 + w_2 \cdot 0 + 2)^2$
2	1	0	$w_0 + w_1 \cdot 2 + w_2 \cdot 1$	$(w_0 + w_1 \cdot 2 + w_2 \cdot 1 - 0)^2$
3	-2	-1	$w_0 + w_1 \cdot 3 + w_2 \cdot -2$	$(w_0 + w_1 \cdot 3 + w_2 \cdot -2 + 1)^2$
4	3	1	$w_0 + w_1 \cdot 4 + w_2 \cdot 3$	$(w_0 + w_1 \cdot 4 + w_2 \cdot 3 - 1)^2$

How to find the coefficients?

- Define the loss function: We will use 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	$w_0 + w_1 \cdot 1 + w_2 \cdot 0$	$(w_0 + w_1 \cdot 1 + w_2 \cdot 0 + 2)^2$
2	1	0	$w_0 + w_1 \cdot 2 + w_2 \cdot 1$	$(w_0 + w_1 \cdot 2 + w_2 \cdot 1 - 0)^2$
3	-2	-1	$w_0 + w_1 \cdot 3 + w_2 \cdot -2$	$(w_0 + w_1 \cdot 3 + w_2 \cdot -2 + 1)^2$
4	3	1	$w_0 + w_1 \cdot 4 + w_2 \cdot 3$	$(w_0 + w_1 \cdot 4 + w_2 \cdot 3 - 1)^2$

- The total loss function:

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

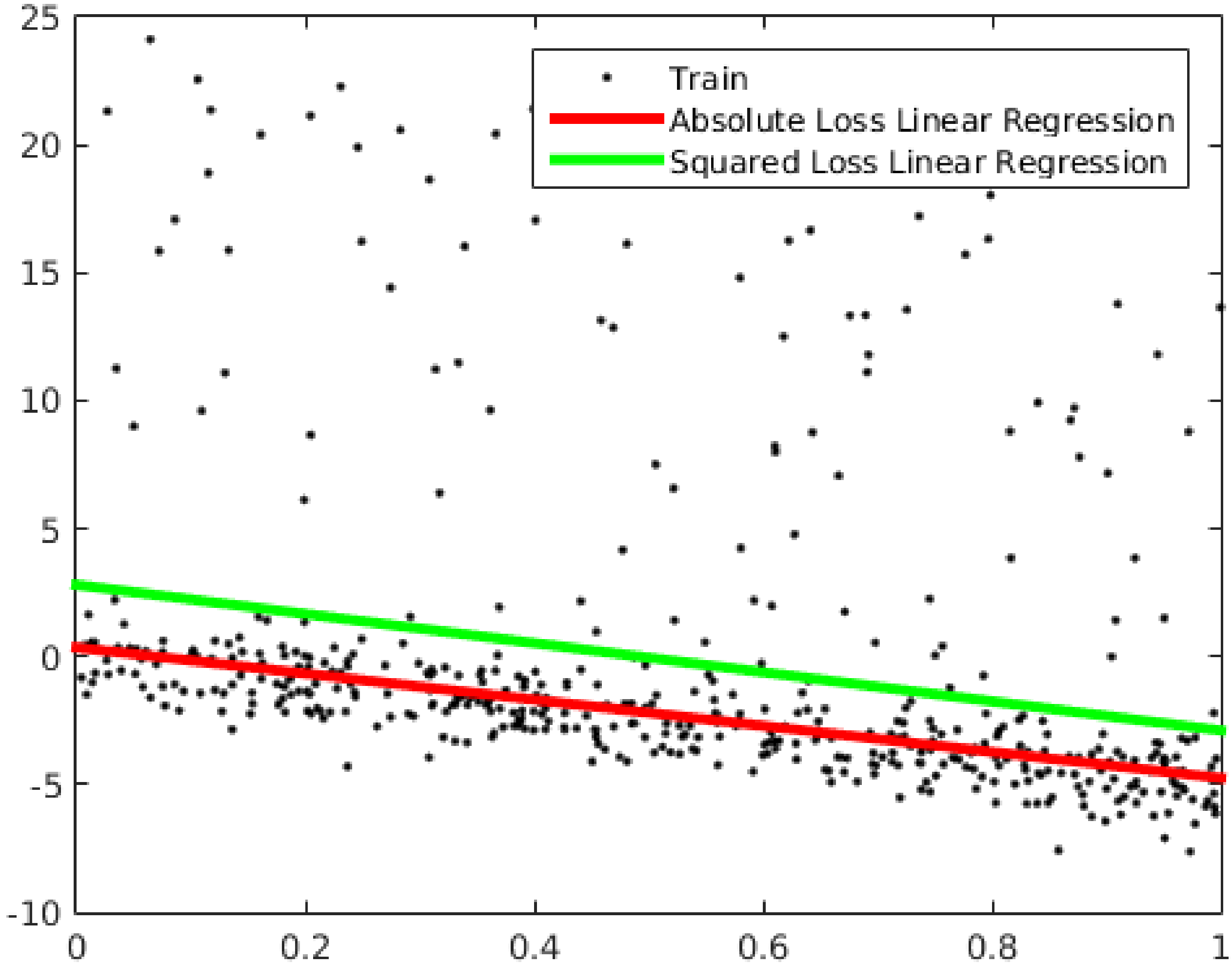
How about other loss functions?

- Absolute loss:

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

- 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

Logistic Regression

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

- 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 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

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.

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

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

- Total Loss:

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

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

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

LASSO - Principle

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

LASSO - Demonstration

- SAS
- Python