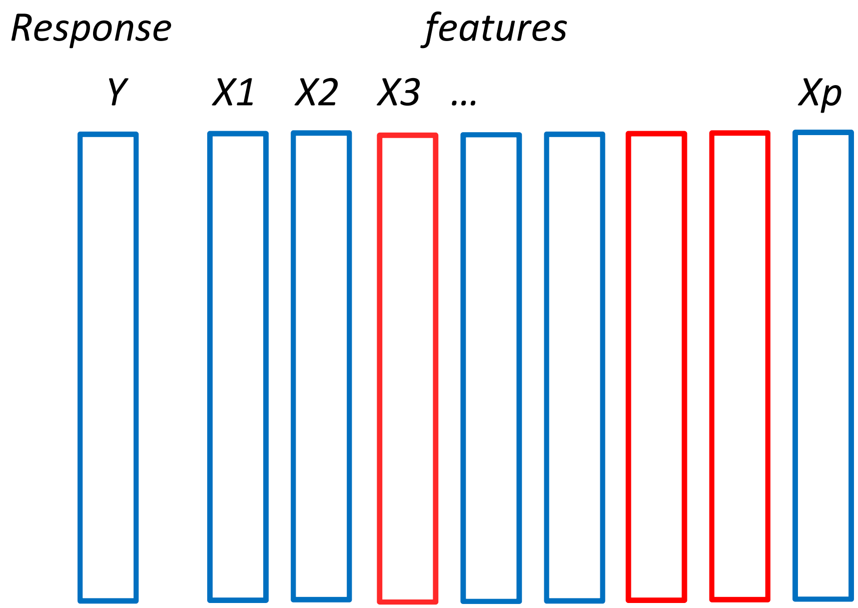
Logistic Regression

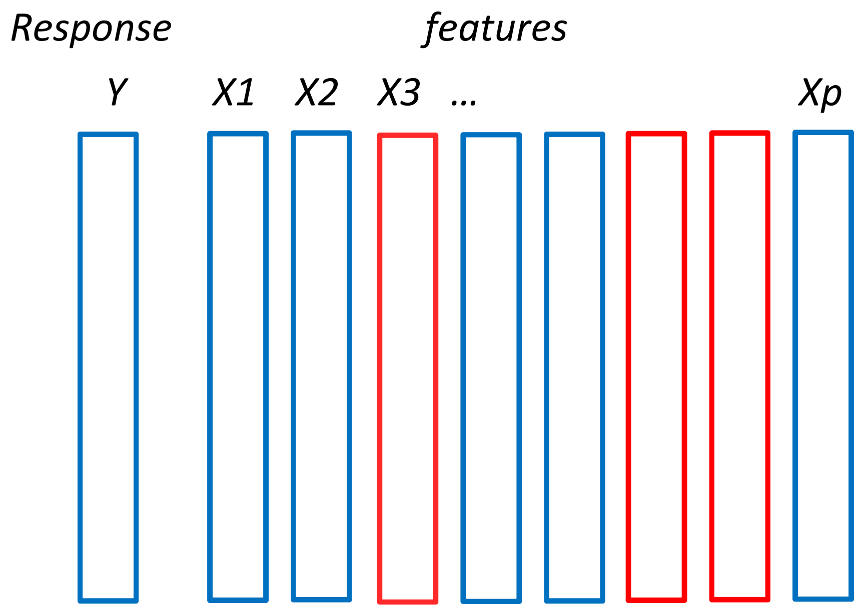
Cesar Acosta Ph.D.

Department of Industrial and Systems Engineering
University of Southern California

Regression vs Classification

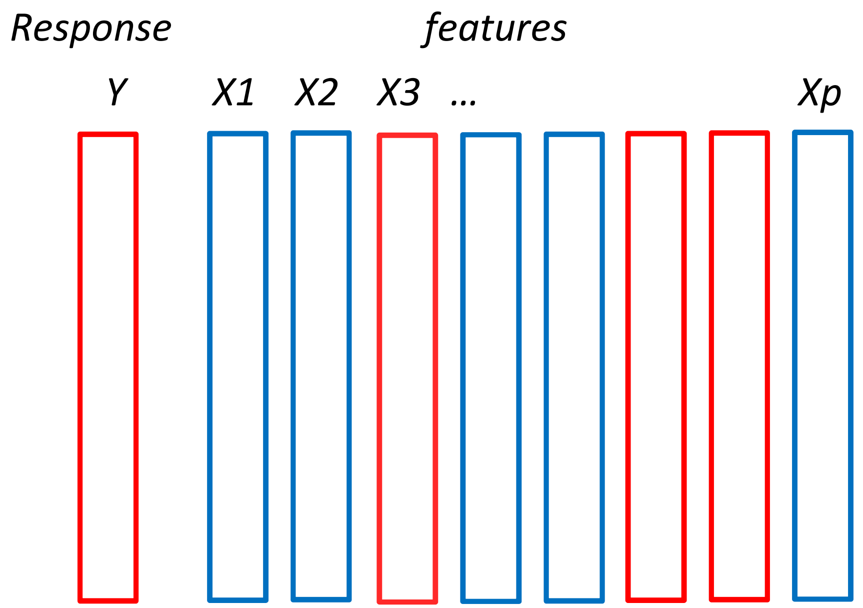


Regression vs Classification



Response is numeric in Regression problems

Regression vs Classification



Response is categorical in Classification problems

Classification problems - Example

- *Response* *product choice (product A, product B, product C)*
- *Predictors*
  - *age group*
  - *gender*
  - *location*
  - *ses*
  - *student*
  - *married*

Classification problems - Example

- *Response* *product choice (product A, product B, product C)*
- *Predictors*
    - *age group*
    - *gender*
    - *location*
    - *ses*
    - *student*
    - *married*

Response is categorical with 3 categories

Logistic Regression

Logistic Regression models
are used in classification problems
where the response has two categories

Logistic Regression

Predict if an English citizen agrees with Brexit

X: years of working experience

Y: Agrees (A)

Disagrees (D)

X	Y
33	A
27	A
12	D
41	A
	.
	.
19	D

Logistic Regression

Predict if an English citizen agrees with Brexit

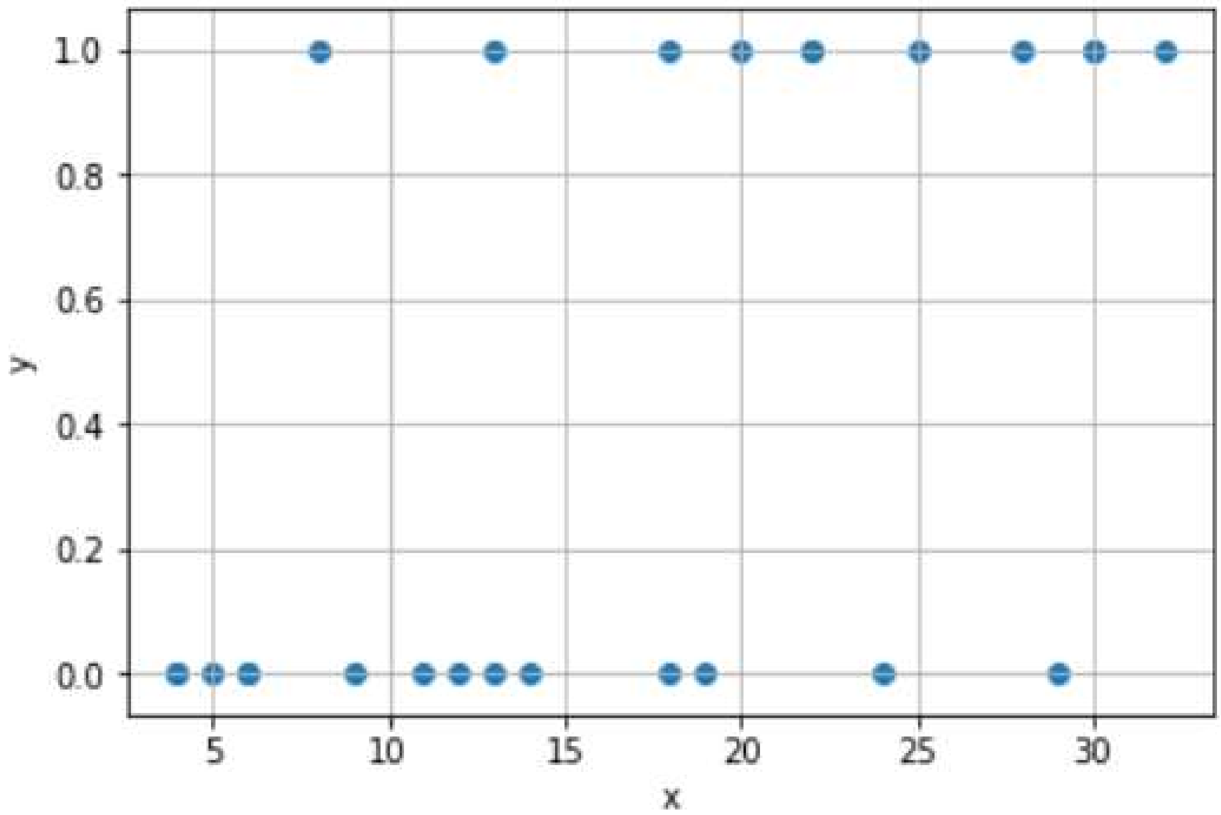
X: years of working experience

Y: Agrees (1)

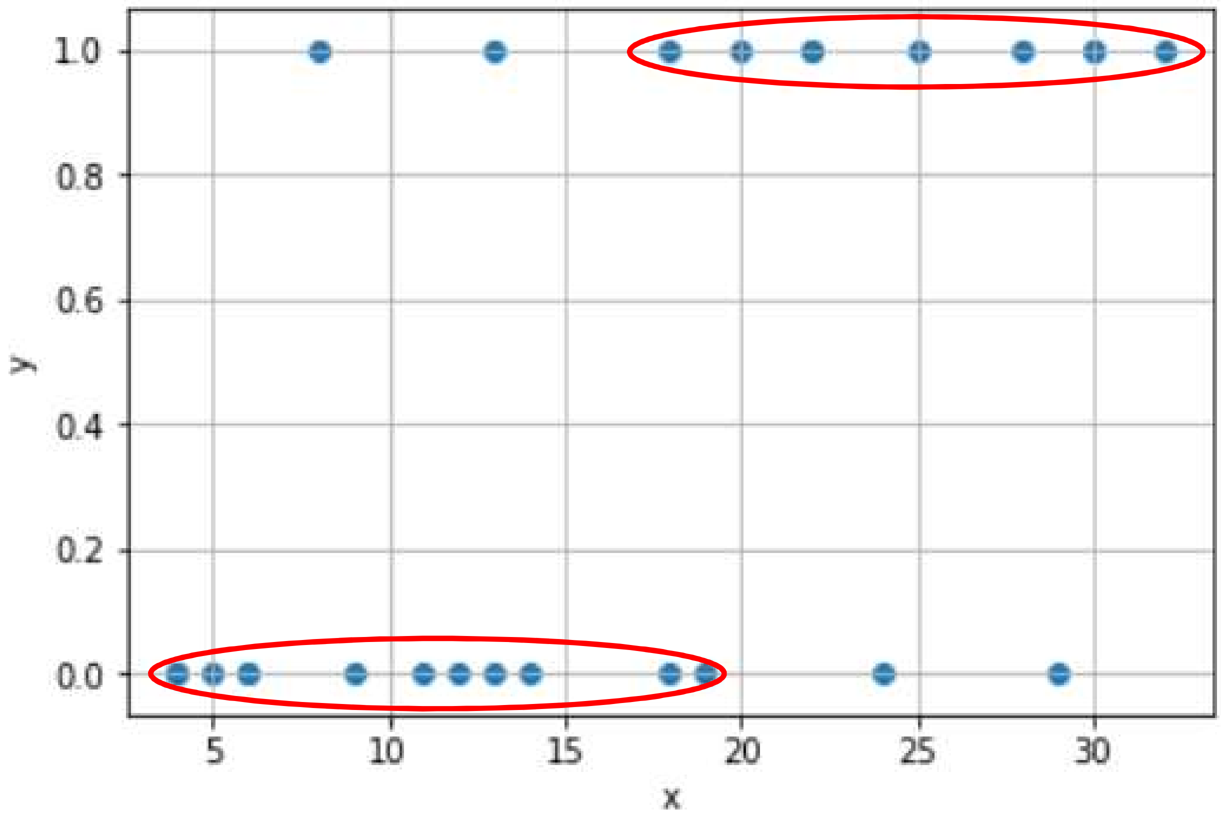
Disagrees (0)

X	Y
33	1
27	1
12	0
41	1
	.
	.
19	0

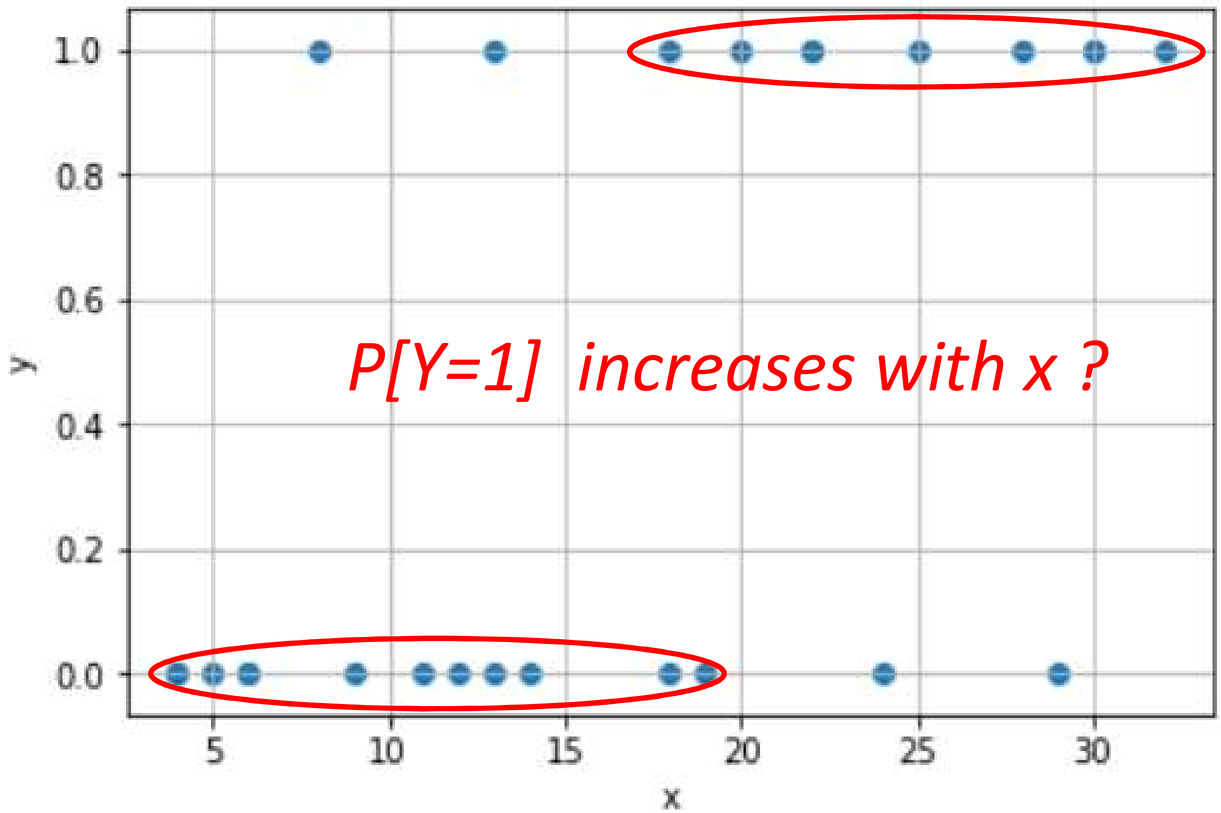
Logistic regression



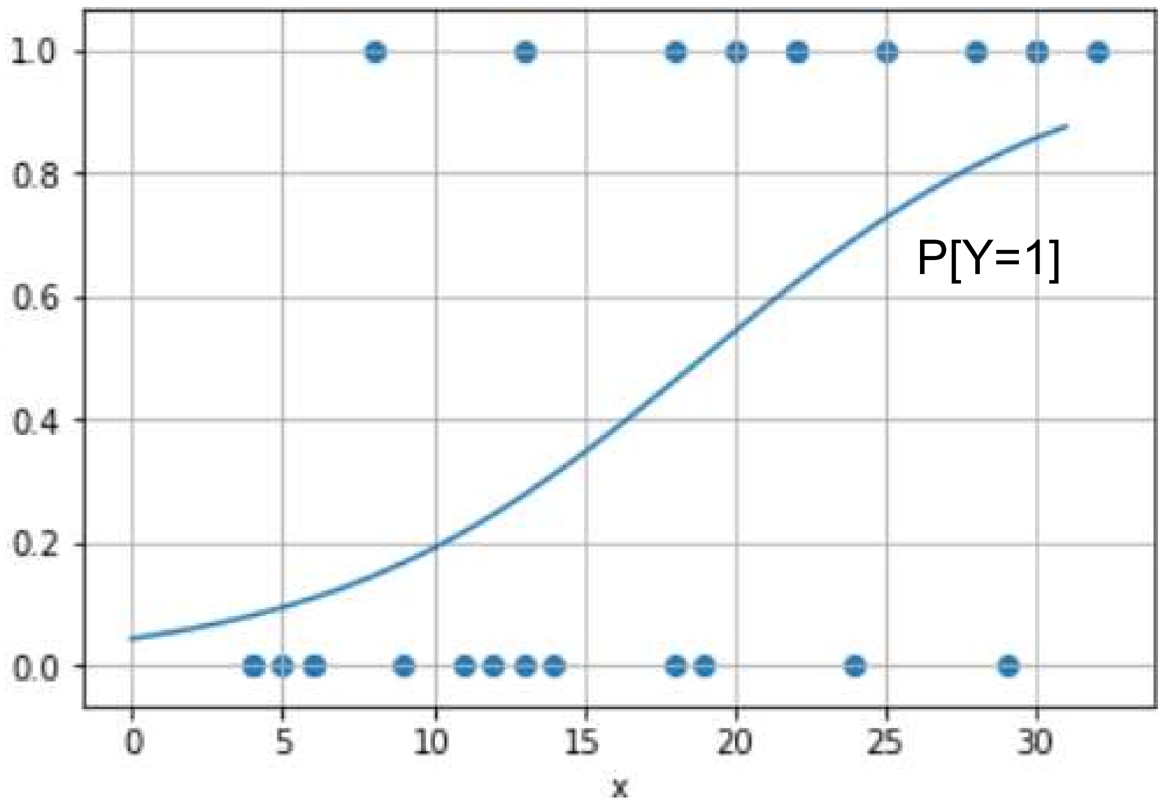
Logistic regression



Logistic regression



Logistic regression



Logistic regression

- *Odds of random event*
- *Indicator random variable*
- *Bernoulli random variable*
- *Logistic distribution function*

Odds of a random event

A random event 'A' may be observed with probability π

Odds of a random event

A random event 'A' may be observed with probability π

The odds of event A

$$\text{Odds}\,[A] = \frac{\pi}{1-\pi}$$

Odds of a random event

A random event 'A' may be observed with probability π

The odds of event A

$$\text{Odds}\,[A] = \frac{\pi}{1-\pi}$$

how much likely is that A occurs

than it is that A does not occur

Odds of a random event - Example

Assume that 2/3 of voters are in favor of candidate A and 1/3 in favor of candidate B

Odds of a random event - Example

Assume that 2/3 of voters are in favor of candidate A

and 1/3 in favor of candidate B

The odds of candidate A

$$\textit{Odds [A]} = \frac{\pi}{1-\pi} = \frac{2/3}{1-2/3} = \frac{2}{1}$$

Odds of a random event - Example

Assume that 2/3 of voters are in favor of candidate A and 1/3 in favor of candidate B

The odds of candidate A

$$\text{Odds [A]} = \frac{\pi}{1-\pi} = \frac{2/3}{1-2/3} = \frac{2}{1}$$

The probability of voting for A is twice the probability of voting for candidate B

Indicator random variable

Definition: The indicator r.v. of event A has pdf

$$y \begin{cases} 1 & \textit{if event A occurs} \\ 0 & \textit{otherwise} \end{cases}$$

where P[A] = π

Bernoulli random variable

Definition: A r.v. Y is called Bernoulli if its pdf is

$$y \begin{cases} 1 & \textit{with probability} \quad \pi \\ 0 & \textit{with probability} \quad 1-\pi \end{cases}$$

Bernoulli random variable

Definition: A r.v. Y is called Bernoulli if its pdf is

$$y \begin{cases} 1 & \text{with probability} \quad \pi \\ 0 & \text{with probability} \quad 1-\pi \end{cases}$$

$$\begin{aligned} E[Y] &= 1\,P[Y=1] + 0\,P[Y=0] \\ &= 1 \quad \pi \;+ 0 \;\; (1-\pi) \\ &= \pi \end{aligned}$$

Bernoulli random variable

Definition: A r.v. Y is called Bernoulli if its pdf is

$$y \begin{cases} 1 & \text{with probability } \pi \\ 0 & \text{with probability } 1-\pi \end{cases}$$

$$\boxed{E[Y] = P[Y = 1]}$$

Bernoulli random variable

Definition: A r.v. Y is called Bernoulli if its pdf is

$$y \begin{cases} 1 & \textit{with probability} \quad \pi \\ 0 & \textit{with probability} \quad 1-\pi \end{cases}$$

with pdf

$$P[Y = y] = \pi^{y} (1-\pi)^{1-y}$$

Bernoulli random variable - Example

A Bernoulli r.v. is defined for customer gender as

$$y \begin{cases} 1 & \textit{if category male} \quad \textit{wp.} \quad \pi \\ 0 & \textit{if category female} \quad \textit{wp.}\ 1-\pi \end{cases}$$

$$\frac{P[Y=1]}{P[Y=0]} = \frac{\pi}{1-\pi}$$

Bernoulli random variable - Example

A Bernoulli r.v. is defined for customer gender as

$$y \begin{cases} 1 & \textit{if category male} \quad \textit{wp.} \quad \pi \\ 0 & \textit{if category female} \quad \textit{wp.} \; 1-\pi \end{cases}$$

$$\frac{P[Y=1]}{P[Y=0]} = \frac{\pi}{1-\pi}$$

the odds of a male customer

Bernoulli random variable - Example

A Bernoulli r.v. is defined for customer gender as

$$y \begin{cases} 1 & \textit{if category male} \quad \textit{wp.} \quad \pi \\ 0 & \textit{if category female} \quad \textit{wp.}\ 1\text{-}\pi \end{cases}$$

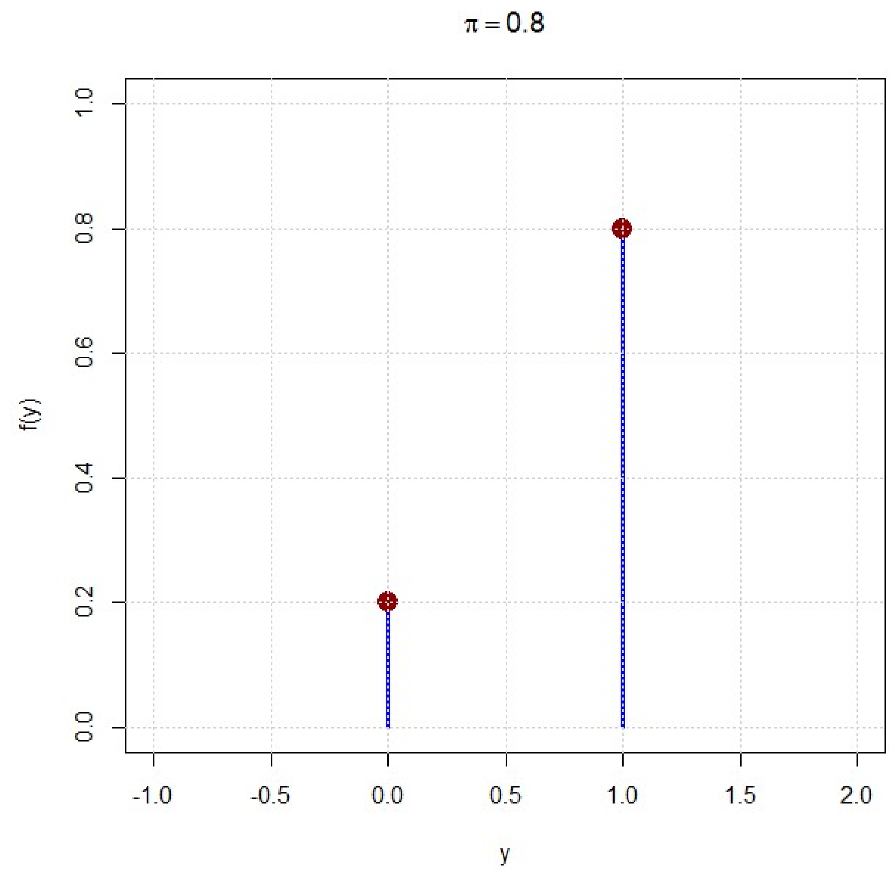
$$\frac{P[Y=1]}{P[Y=0]} = \frac{\pi}{1-\pi}$$

how much likely is it a customer male, than is it a female

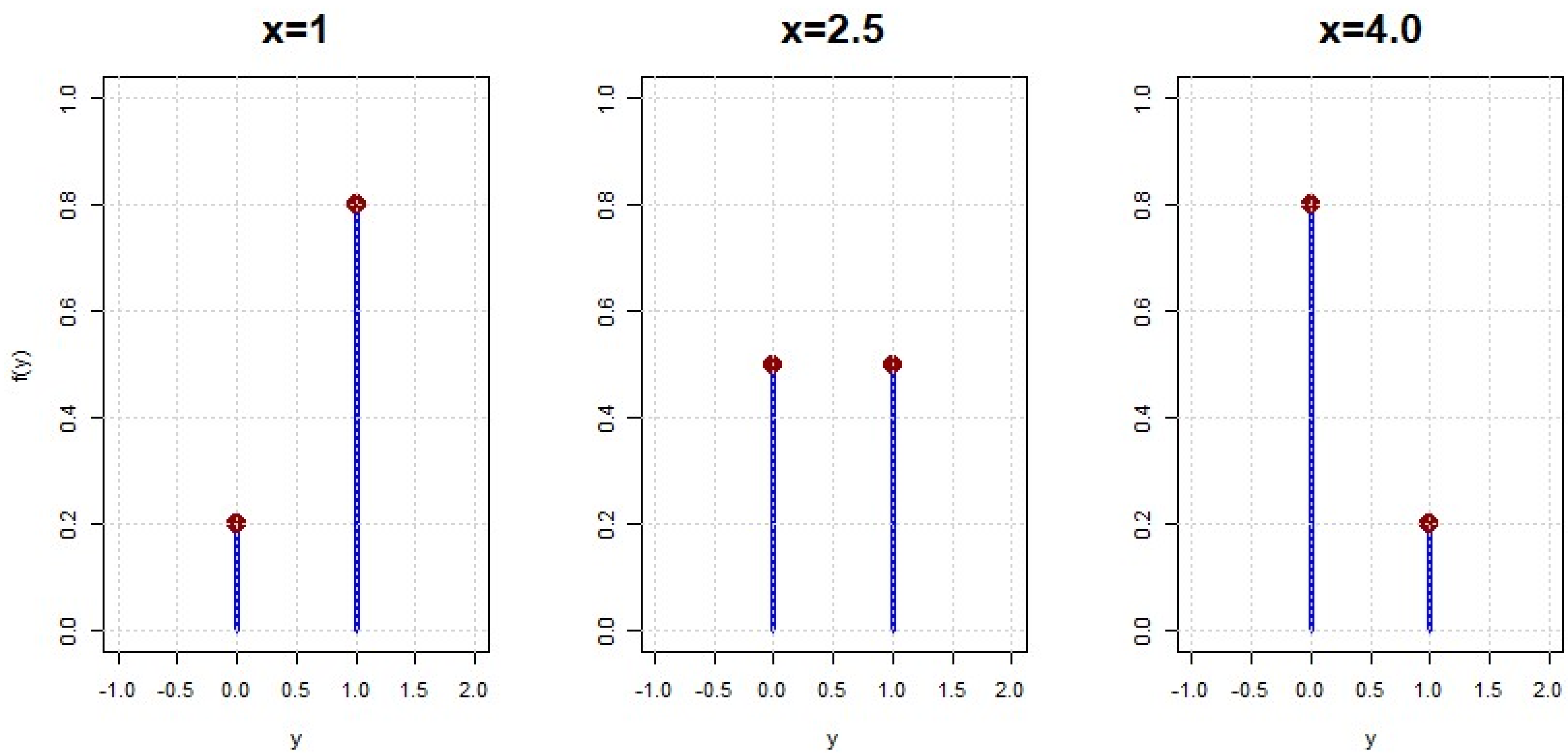
Bernoulli probability function

$$f(y) = P[Y = y]$$
$$= \pi^{y} (1-\pi)^{1-y}$$

$$y = 0,1$$



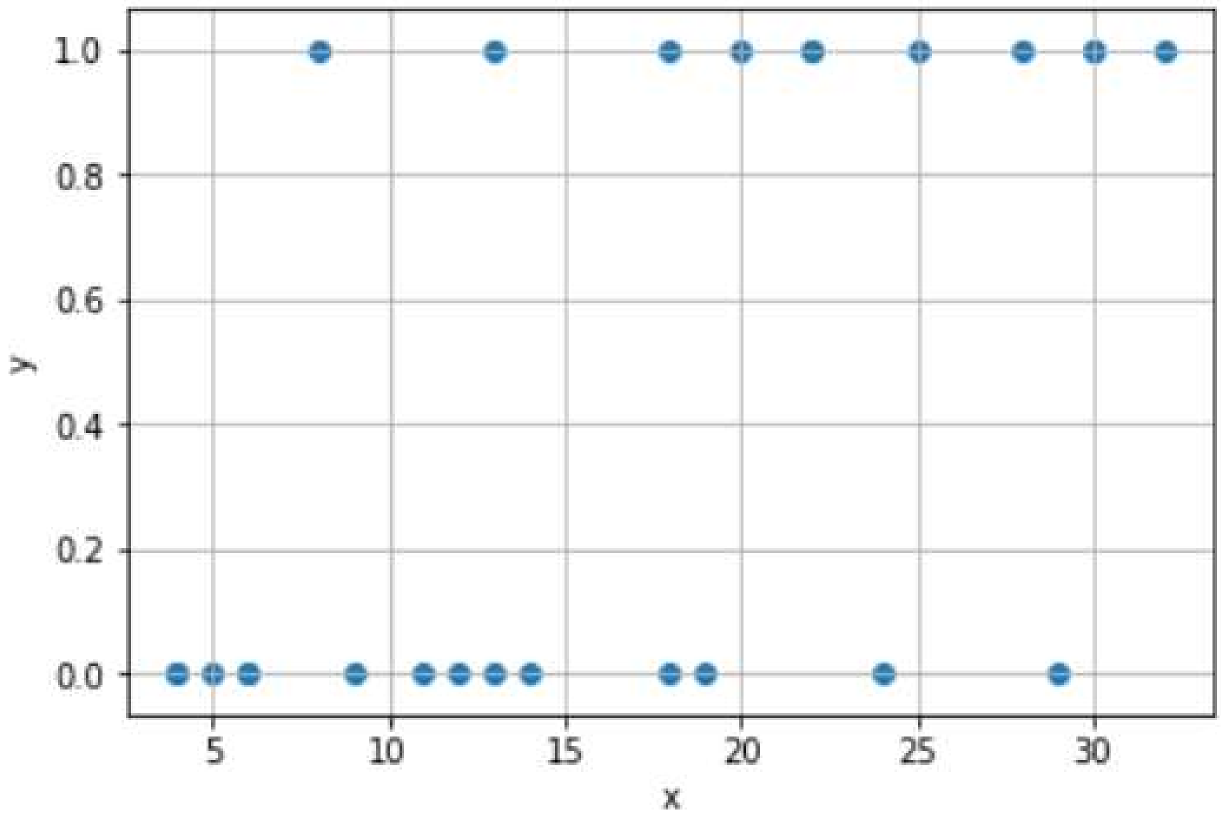
Three Bernoulli probability functions



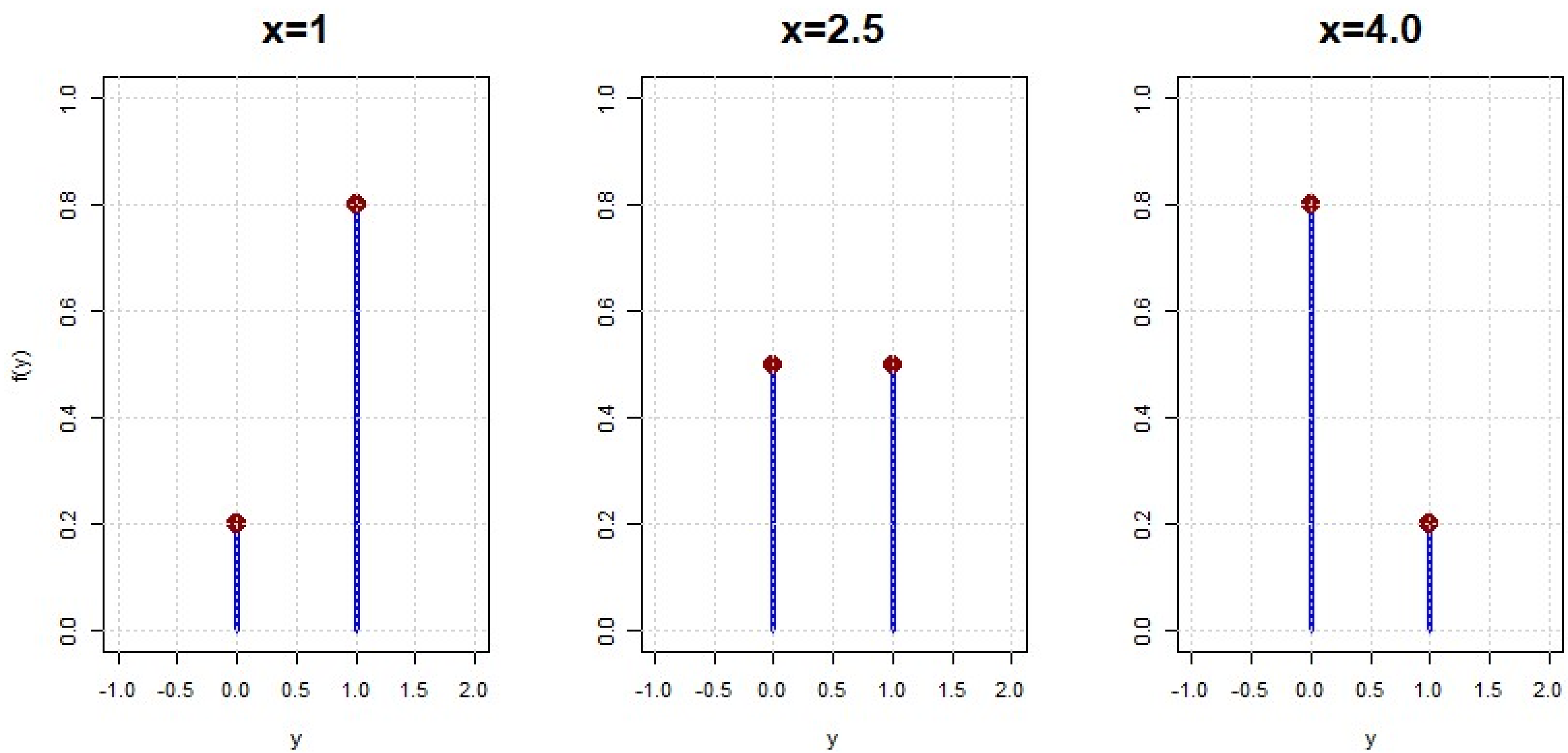
Logistic regression

Is there a relation between Y and X?

Logistic regression



Three Bernoulli probability functions



E[Y] = 0.8 *E[Y] = 0.5* *E[Y] = 0.2*

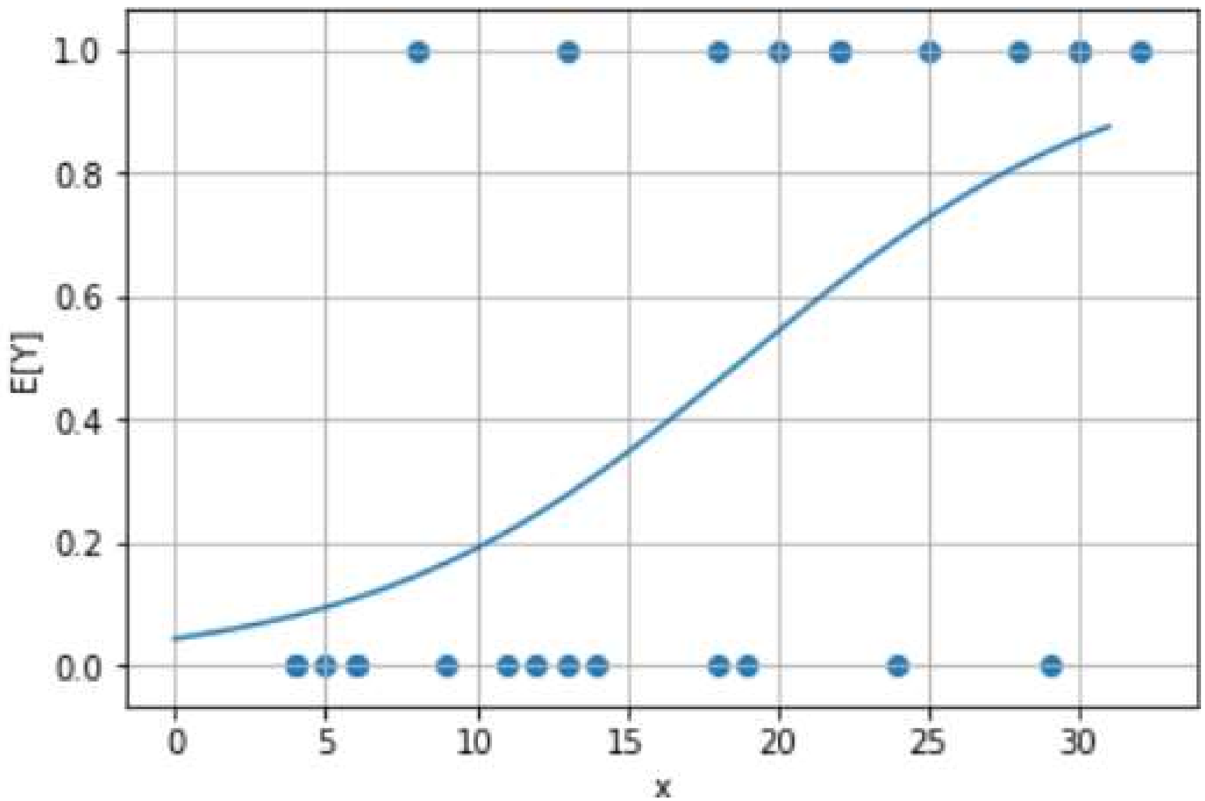
Linear regression

Is there a relation between E[Y] and X?

Logistic regression

Is there a relation between π and X?

Logistic regression



Logistic random variable

A r.v. X is called Logistic with mean μ

pdf $$f(x) = \frac{1}{\beta}\frac{e^{-\left(\frac{x-\mu}{\beta}\right)}}{\left[1+e^{-\left(\frac{x-\mu}{\beta}\right)}\right]^2} \qquad -\infty < x < \infty$$

cdf $$F(x) = \frac{e^{\left(\frac{x-\mu}{\beta}\right)}}{1+e^{\left(\frac{x-\mu}{\beta}\right)}} = \frac{1}{1+e^{-\left(\frac{x-\mu}{\beta}\right)}}$$

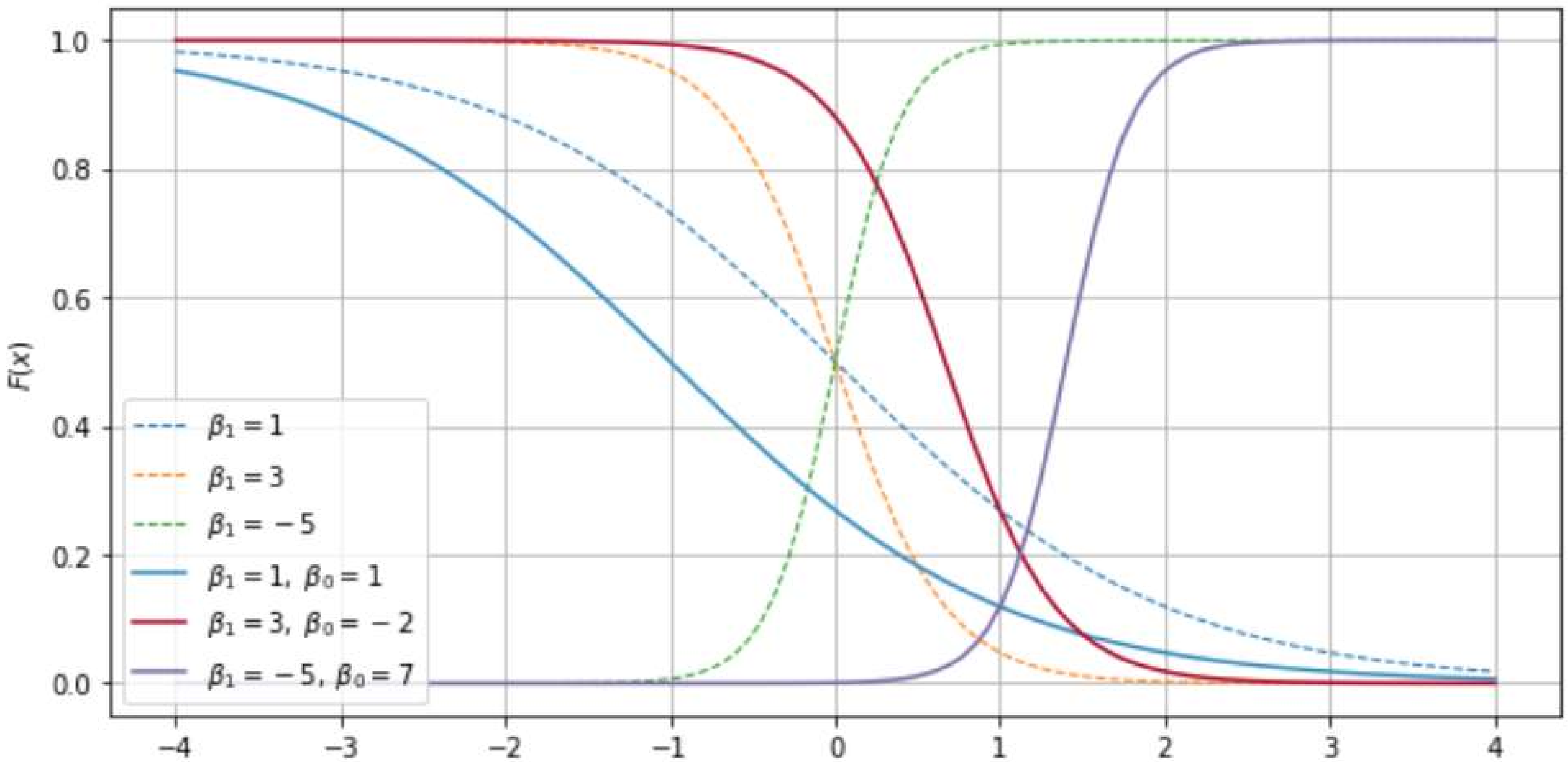
Logistic distribution

Logistic distribution

$$F(x) = \frac{e^{\left(\frac{x-\mu}{\beta}\right)}}{1+e^{\left(\frac{x-\mu}{\beta}\right)}} = \frac{1}{1+e^{-\left(\frac{x-\mu}{\beta}\right)}}$$

$$F(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \frac{1}{1+e^{-\beta_0-\beta_1 x}}$$

Logistic distributions



Logistic regression -relation

As x increases, π varies along the logistic cdf

$$\pi = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}}$$

Logistic regression -relation

As x increases, π varies along the logistic cdf

$$\pi = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}}$$

For arbitrary x_i

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_i}}$$

Logistic regression -assumption

This relation between π_i and x_i

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_i}}$$

Logistic regression -assumption

This relation between π_i and x_i

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_i}}$$

is estimated by

$$\hat{\pi}_i = \frac{1}{1 + e^{-b_0 - b_1 x_i}}$$

Logistic regression -predictions

- Probabilities are predicted by

$$\hat{\pi}_i = \frac{1}{1 + e^{-b_0 - b_1 x_i}}$$

- Category of the response is predicted by
  - if $\hat{\pi}_i \geq 0.5$ predict $\hat{y} = 1$
  - . $\hat{\pi}_i < 0.5$ predict $\hat{y} = 0$

Logistic regression -assumptions

- π_i changes with x_i (not linearly)
- As *x* increases π varies, moving along an *S* shape curve (logistic cdf is the shape curve)
- Standard regression assumptions do not apply
- For different *X*, the *Y* variables are independent

Logist regression -parameters

What is the meaning of β_1?

What is β_1 ?

Define for category 1

	probability	odds
when $X = x_1$	$\pi_1 = P[Y=1]$	$O_1 = \frac{\pi_1}{1-\pi_1}$
when $X = x_2$	$\pi_2 = P[Y=1]$	$O_2 = \frac{\pi_2}{1-\pi_2}$

Logistic regression -parameters

$$\pi = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}}$$

$$\frac{1}{\pi} = 1 + e^{-\beta_0 - \beta_1 x}$$

Logistic regression -parameters

$$\pi = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}}$$

$$\frac{1}{\pi} = 1 + e^{-\beta_0 - \beta_1 x}$$

$$\frac{1}{\pi} - 1 = e^{-\beta_0 - \beta_1 x}$$

Logistic regression -parameters

$$\pi = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}}$$

$$\frac{1}{\pi} = 1 + e^{-\beta_0 - \beta_1 x}$$

$$\frac{1}{\pi} - 1 = e^{-\beta_0 - \beta_1 x}$$

$$\frac{1 - \pi}{\pi} = e^{-\beta_0 - \beta_1 x}$$

odds

$$\frac{\pi}{1 - \pi} = e^{\beta_0 + \beta_1 x}$$

Logistic regression -parameters

$$\frac{\pi}{1-\pi} = e^{\beta_0 + \beta_1 x}$$

$$O = e^{\beta_0 + \beta_1 x}$$

the odds of category 1 as a function of x

$$\ln O = \beta_0 + \beta_1 x$$

the log odds of category 1 is a linear function of x

Logistic regression -parameters

Compare the odds of category 1,
when X changes from x_1 to x_2

$$O_1 = e^{\beta_0} e^{\beta_1 x_1}$$

$$O_2 = e^{\beta_0} e^{\beta_1 x_2}$$

Logistic regression -parameters

Compare the odds of category 1,

when X changes from x_1 to x_2

$$O_1 = e^{\beta_0}\, e^{\beta_1 x_1}$$

$$O_2 = e^{\beta_0}\, e^{\beta_1 x_2}$$

$$\frac{O_2}{O_1} = e^{\beta_1 (x_2 - x_1)}$$

Logistic regression -parameters

Compare the odds of category 1,

when X changes from x_1 to x_2

$$O_1 = e^{\beta_0} e^{\beta_1 x_1}$$

$$O_2 = e^{\beta_0} e^{\beta_1 x_2}$$

$$\frac{O_2}{O_1} = e^{\beta_1 (x_2 - x_1)}$$

If x_2-x_1 = 1

$$\ln\left(\frac{O_2}{O_1}\right) = \beta_1$$

Logistic regression -parameters

$$\ln\left(\frac{O_2}{O_1}\right) = \beta_1$$

$$\frac{O_2}{O_1} = e^{\beta_1}$$

β_1 is difference of log odds

e^{β_1} is the ratio of the odds

when x increases one unit

Logistic regression –cross validation

K-fold cross validation for classification problems

Logistic regression –cross validation

- There is a proportion for each category in the dataset

- Split the data into *k*-folds, such that the proportions between categories are similar across all folds, and as they are in the whole dataset

- kfold = Kfold(n_splits,shuffle=True,random_state=   )

Logistic regression –cross validation

- There is a proportion for each category in the dataset
- Split the data into *k*-folds, such that the proportions between categories are similar across all folds, and as they are in the whole dataset
- kfold = KFold(n_splits,shuffle=True,random_state=)

Logistic Regression

Multinomial Regression models
are used in classification problems
where the response has
more than two categories

Logistic regression –multiple categories

Use

LogisticRegression(multi_class='multinomial',...)