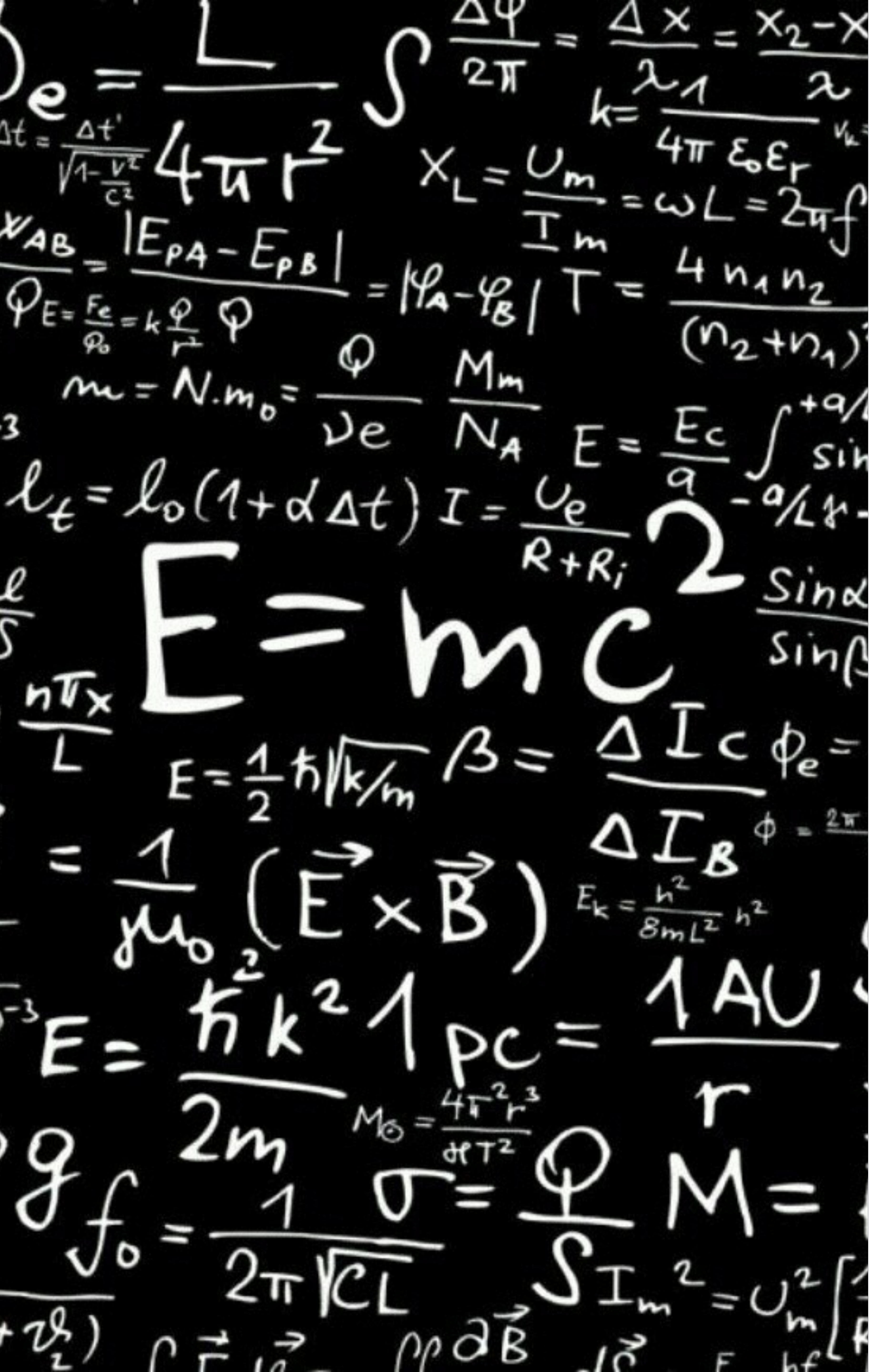
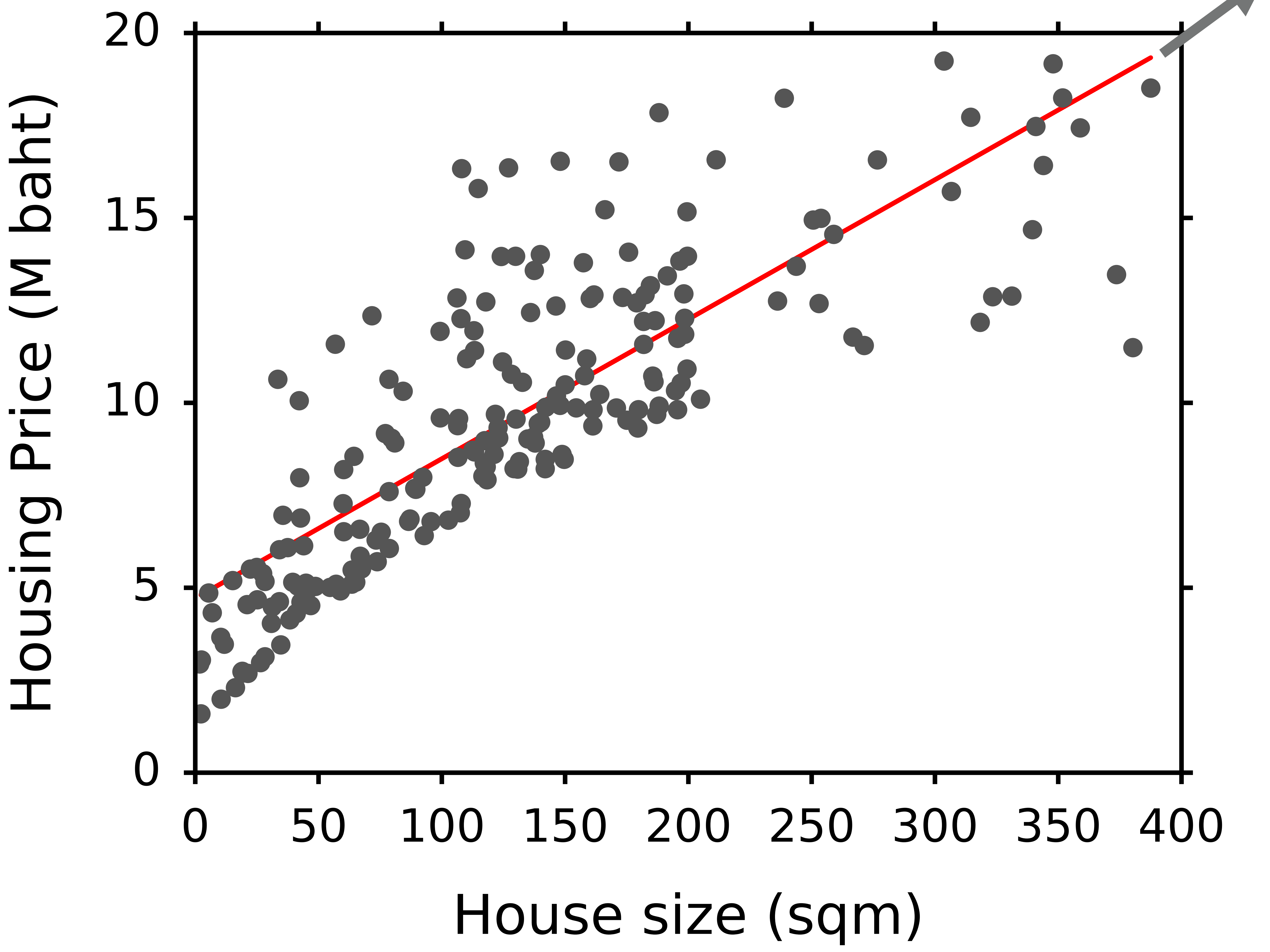


LINEAR AND LOGISTIC REGRESSION

Linear Regression Model



$$h(x) = \underbrace{\theta_0}_{} + \underbrace{\theta_1}_{} x$$

parameters of the model

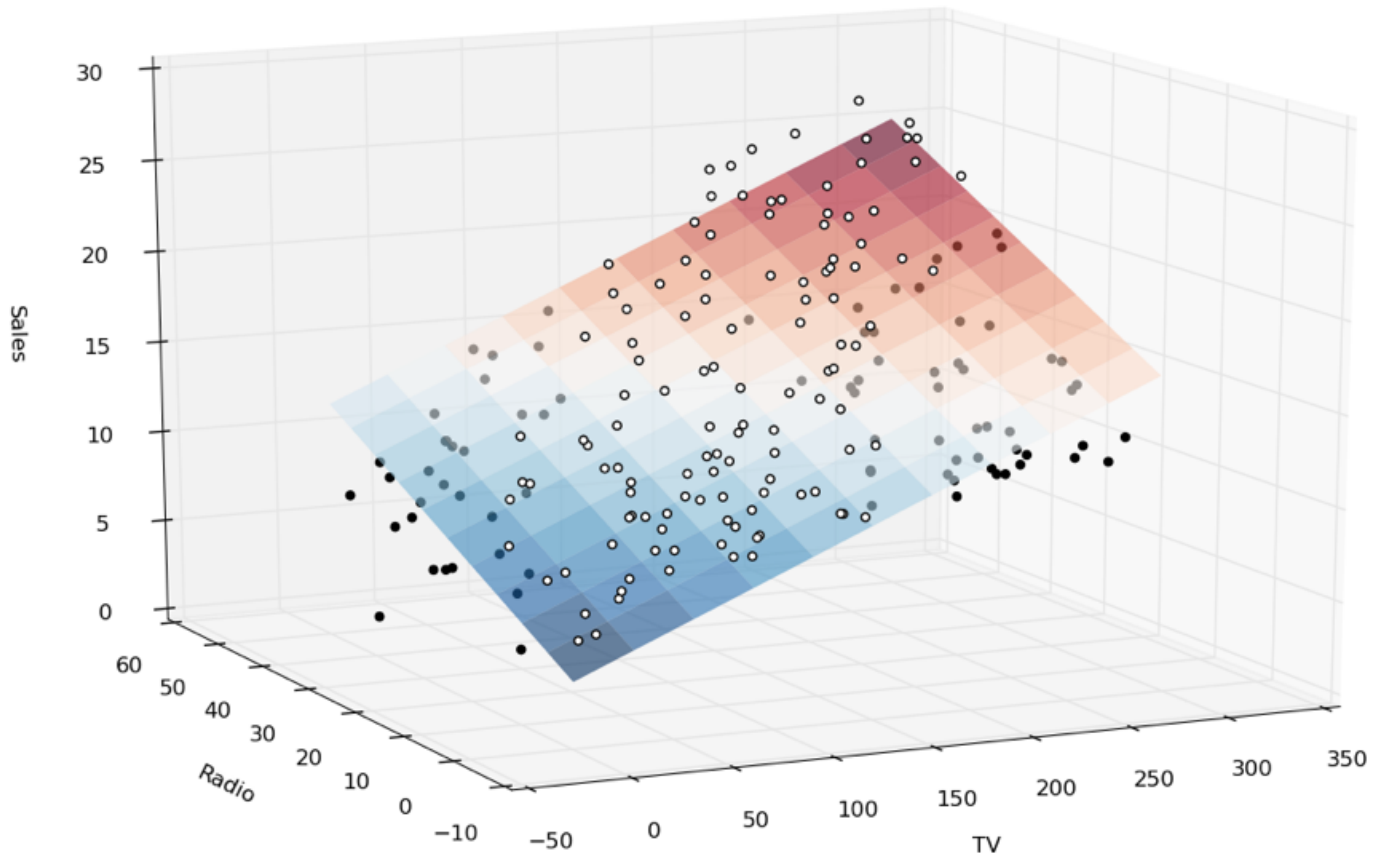
$$\theta_0 = 0.038$$

$$\theta_1 = 4.717$$

Linear Regression Model

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \ldots + \theta_n x_n$$

$$h(x) = \Sigma_i^n \theta_i x_i$$



source: https://goo.gl/qFBRts

Linear Regression Model

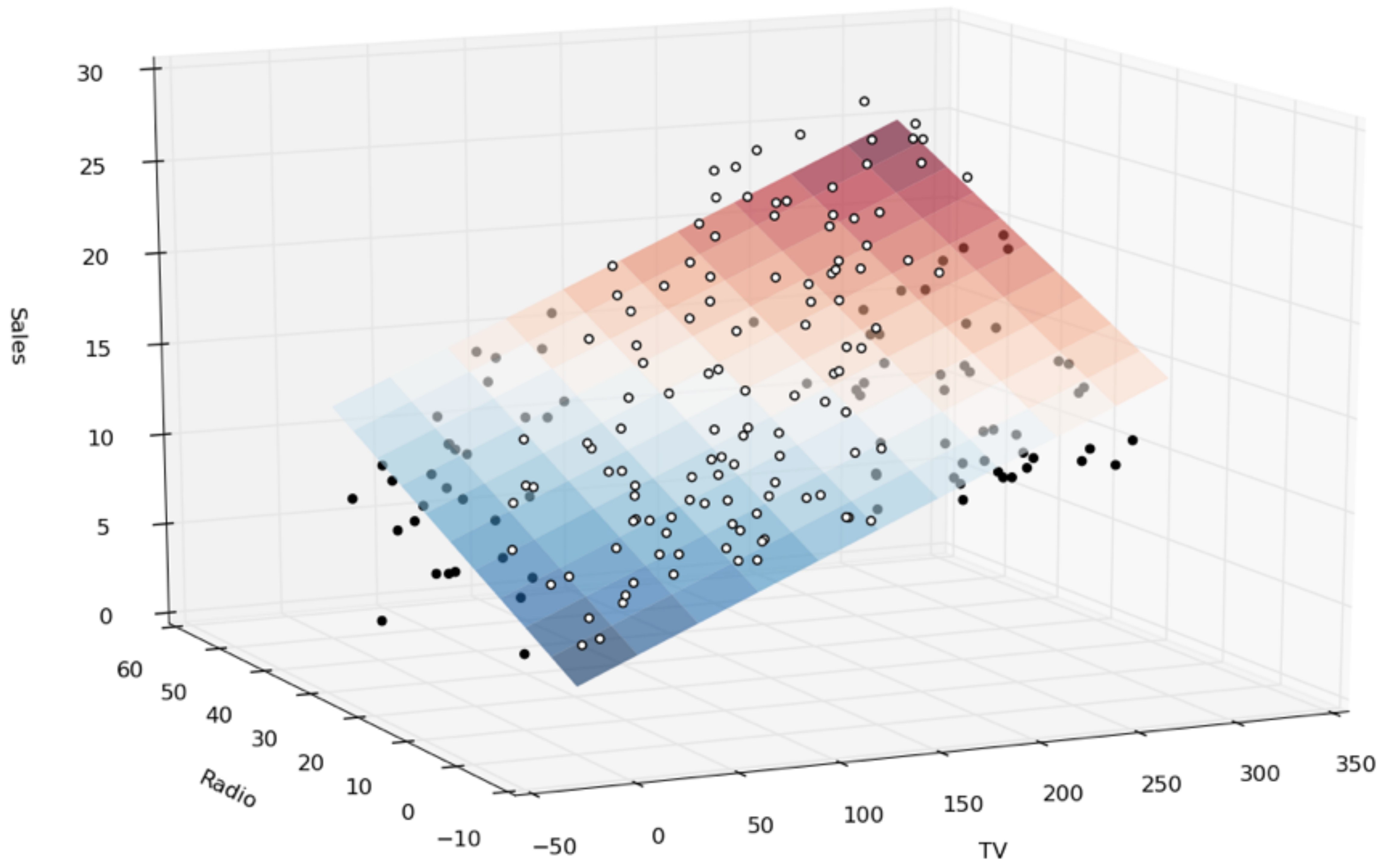
$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \ldots + \theta_n x_n$$

$$h(x) = \Sigma_i^n \theta_i x_i$$

Vector Form of Linear Regression

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \ldots \\ x_n \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \ldots \\ \theta_n \end{bmatrix}$$

$$\mathrm{h(x)} = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 & \ldots & \theta_n \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix} = \theta^T \mathbf{x}$$



source: https://goo.gl/qFBRts

Regression Performance

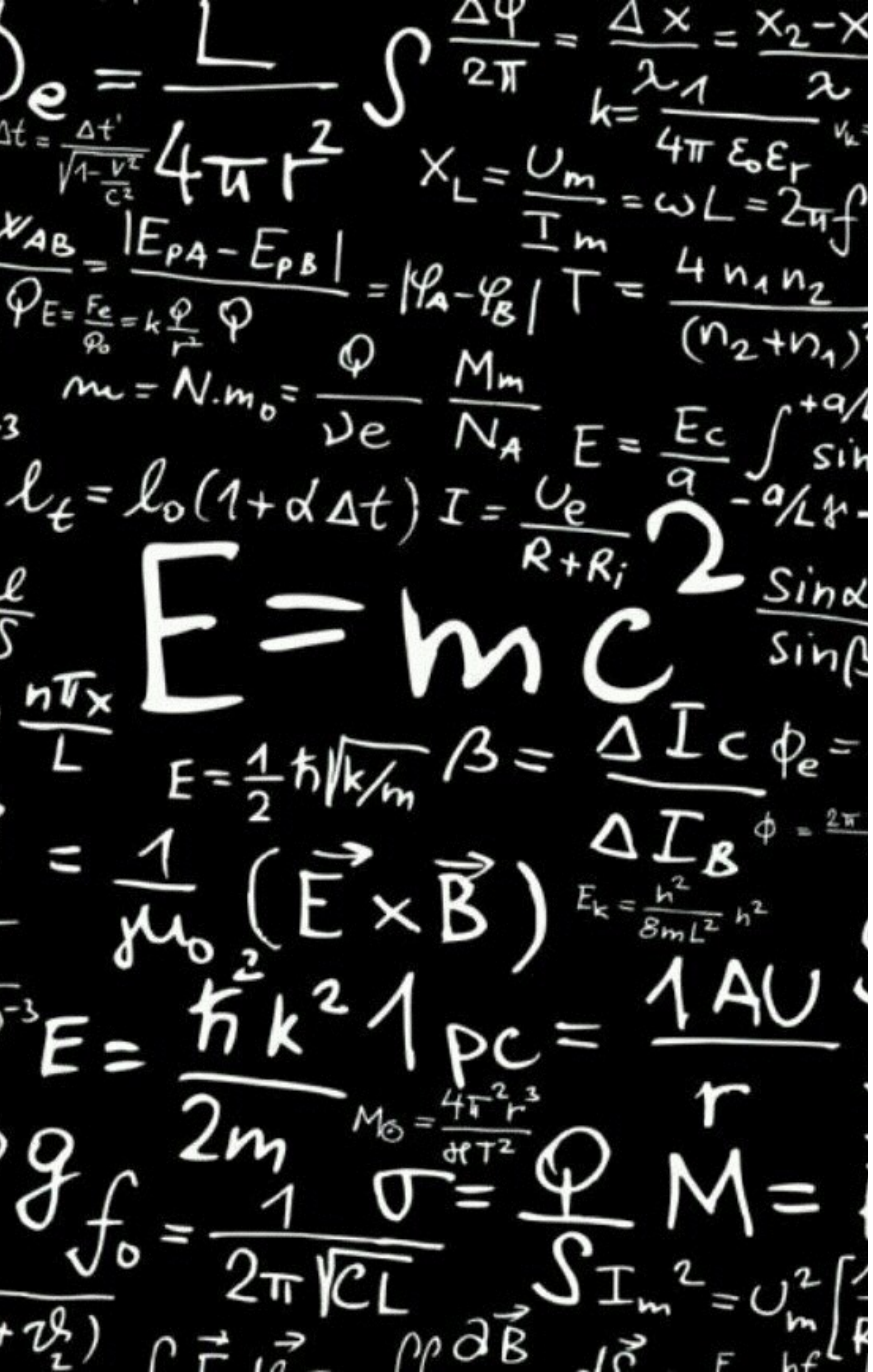
- R-squared: how close the data are to the fitted regression line.
- R-squared = the percentage of the response variable variation that is explained by a linear model.

R-squared = Explained variation / Total variation

- R-squared = 0 (model explains none of the variability of data).
- R-squared = 1 (model explains all of the variability of data).

Linear Regression

- We use mean square error or root mean square error as a cost function of linear regression.
- Remember that linear regression model assume all features and labels are normally distributed.
- All features should have equal variance and are not correlated to each other.

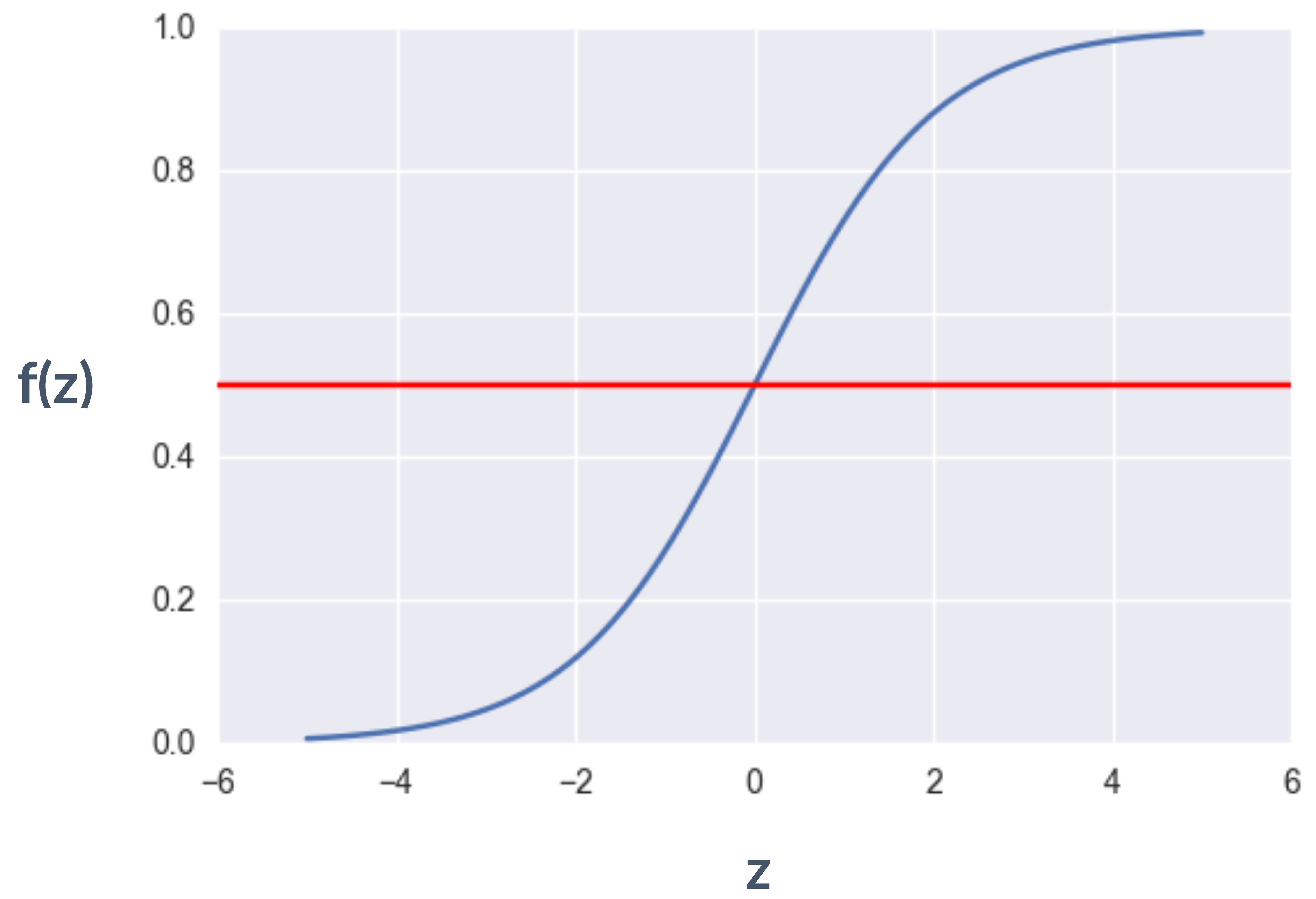


LOGISTIC REGRESSION MODEL

Logistic Regression Model

Logistic regression is similar to linear regression but used for classification problem…

$$h(x) = f(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots)$$

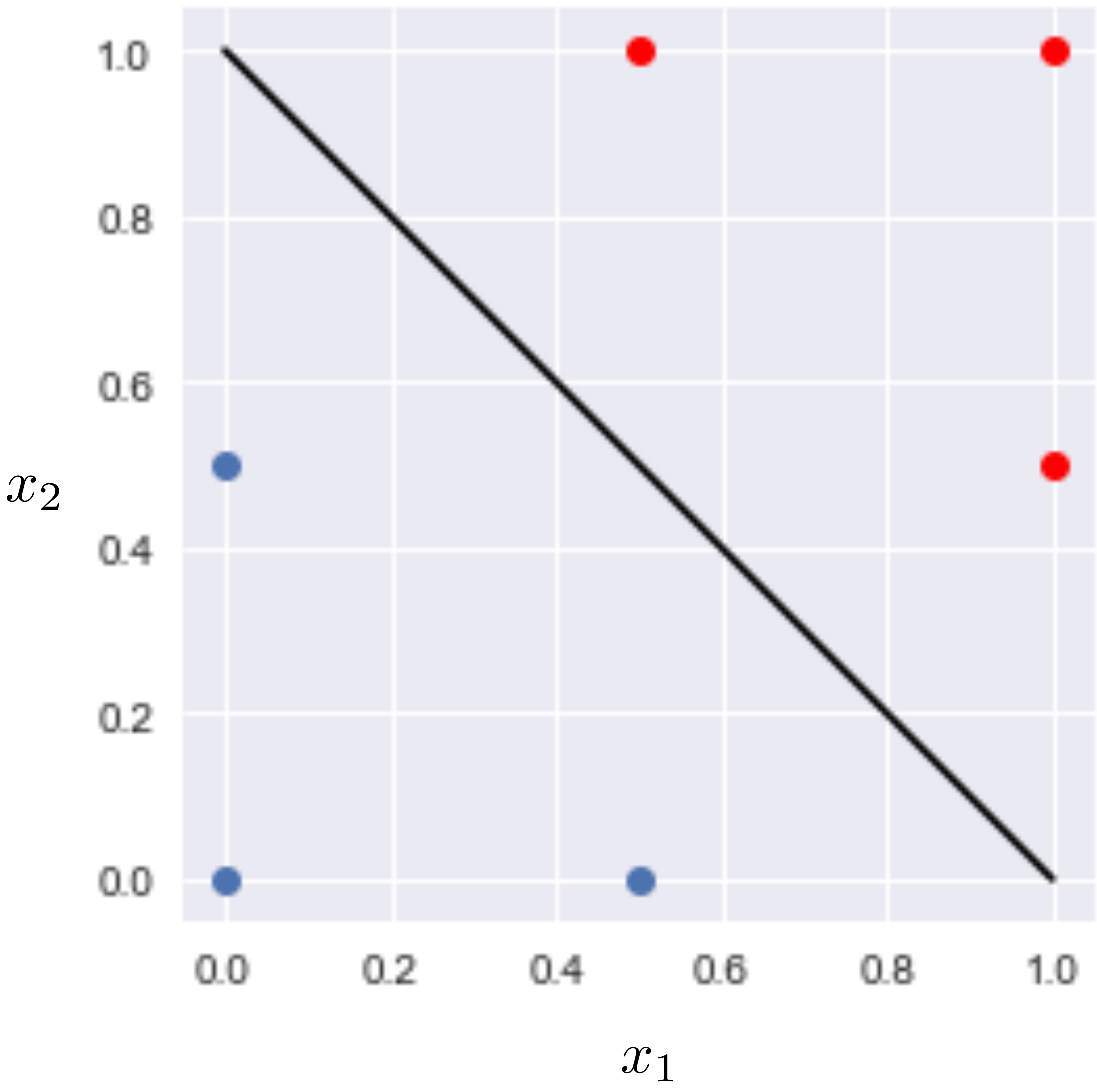


$$f(z) = \frac{1}{1 + e^{-z}}$$

y’ = 1, if f(z) > 0.5 or z > 0

y’ = 0, if f(z) < 0.5 or z < 0

Logistic Regression Model Quiz



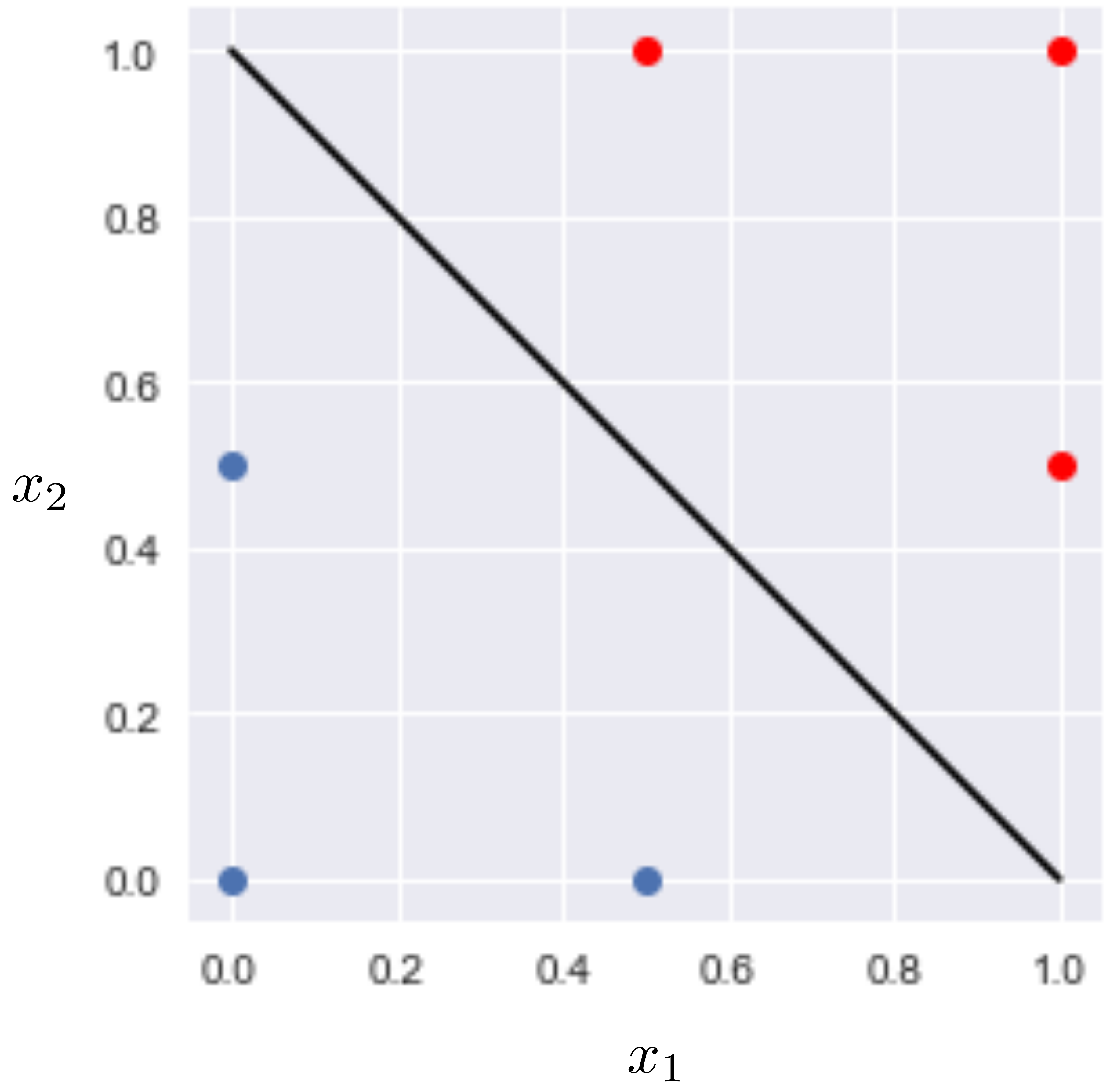
model: $h(x) = f(\theta_0 + \theta_1 x_1 + \theta_2 x_2) = f(z)$

parameter:

$$\theta_0 = -1$$
$$\theta_1 = 1$$
$$\theta_2 = 1$$

x1	x2	z	y′
1	1		
0	0		
0.5	0		
0	0.5		

Logistic Regression Model



model: $h(x) = f(\theta_0 + \theta_1 x_1 + \theta_2 x_2) = f(z)$

parameter:
$$\theta_0 = -1$$
$$\theta_1 = 1$$
$$\theta_2 = 1$$

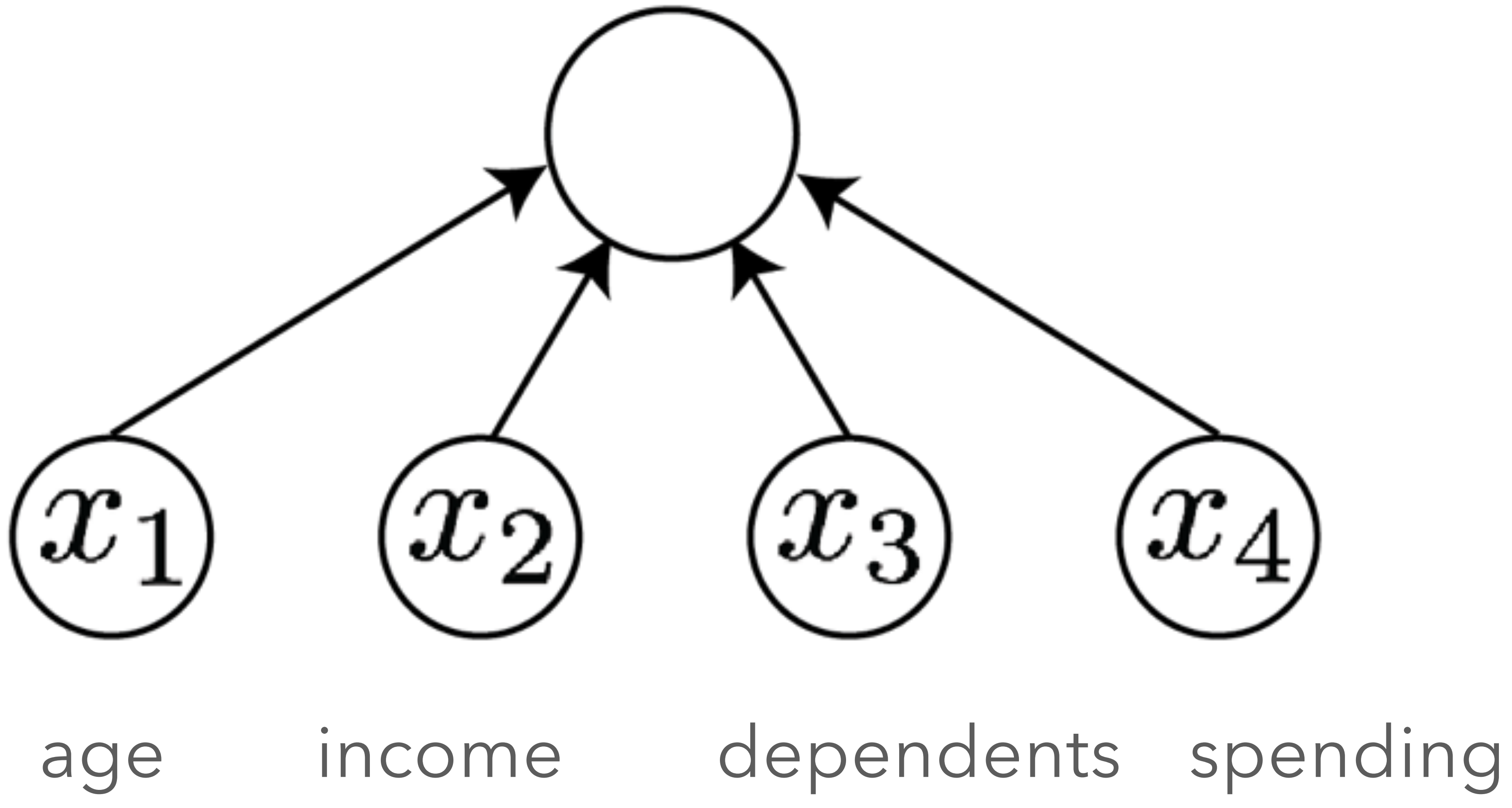
Logistic regression model draws a boundary at z=0

$$1 * x_1 + 1 * x_2 - 1 = 0$$
$$x_2 = -x_1 + 1$$

A simple example

Let's look at a simple logistic regression procedure.

$$h = f(\Sigma_j w_j x_j)$$



x1	x2	x3	x4	history
40	50	0	30	1
25	40	2	35	1
18	10	0	12	0
34	22	1	10	1

A simple example

We first need to do preprocessing, such as normalization and standardization.

x1	x2	x3	x4	history
40	50	0	30	1
25	40	2	35	1
18	10	0	12	0
34	22	1	10	1

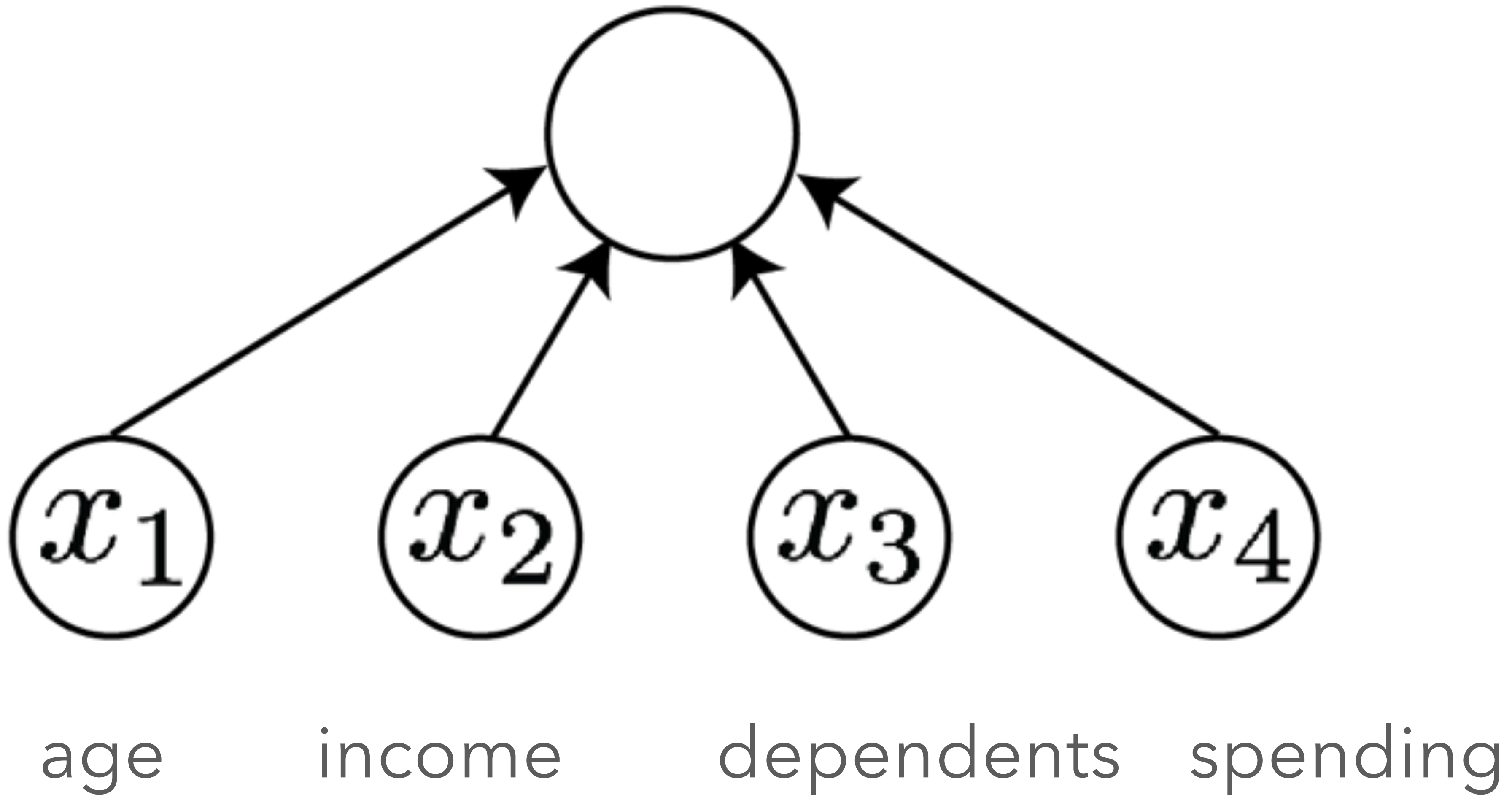
x1	x2	x3	x4	history
0.44	0.63	0	0.6	1
0.28	0.50	0.5	0.7	1
0.20	0.13	0	0.24	0
0.38	0.28	0.25	0.2	1

A simple example

Then fit the logistic regression to the data.

Suppose after fitting, here are the weight numbers.

$$h = f(\Sigma_j w_j x_j)$$

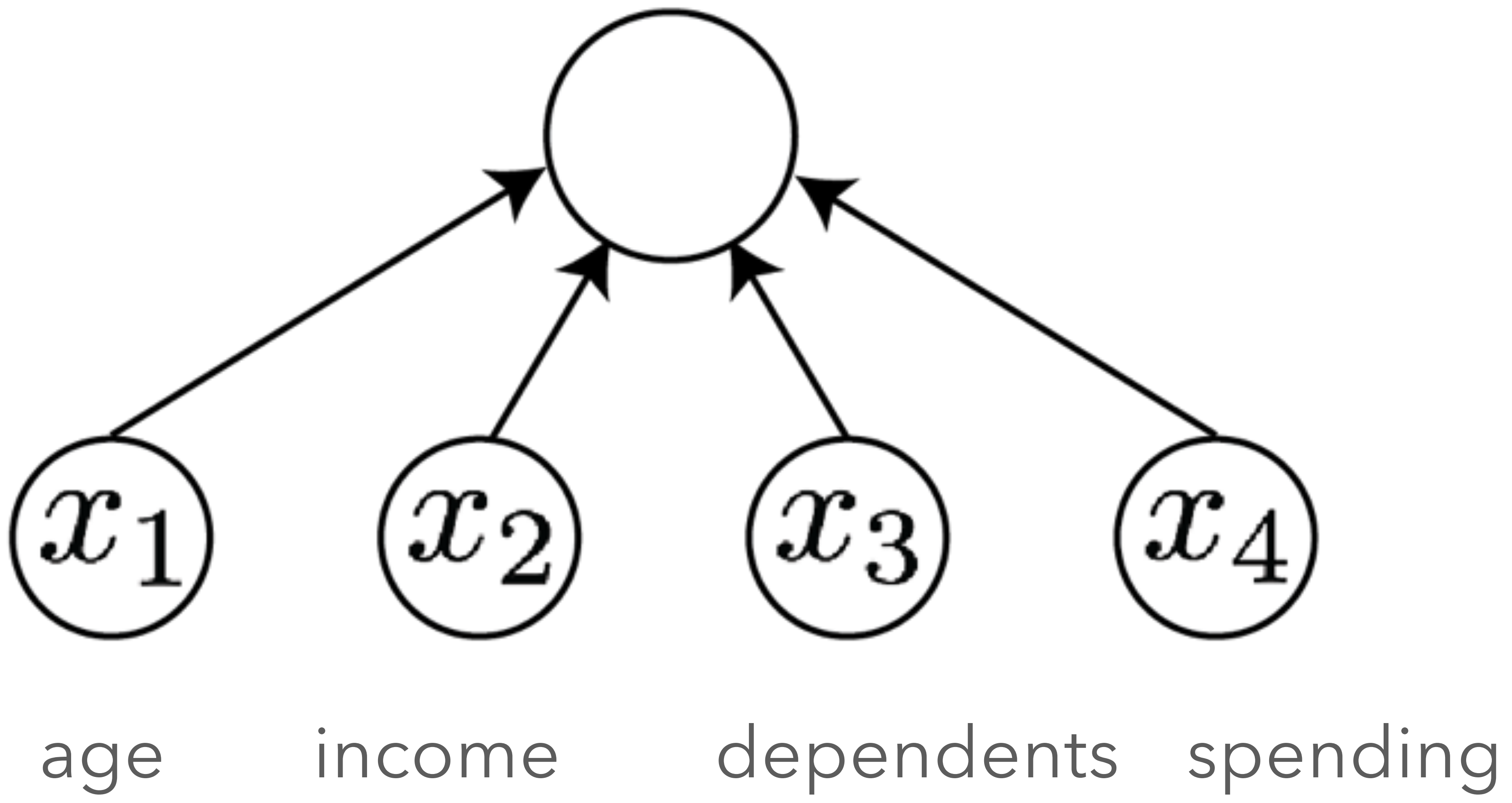


w1	0.7
w2	0.6
w3	-0.1
w4	-0.2

A simple example

Let us make prediction for a single customer…

$$h = f(\Sigma_j w_j x_j)$$



	X	W	X*W
age	0.44	0.7	0.31
income	0.63	0.6	0.38
dependent	0.00	-0.1	0.00
spending	0.60	-0.2	-0.12
		sum:	0.57
		h:	0.64

What the output means

Classification results

	X	W	X*W
age	0.44	0.7	0.31
income	0.63	0.6	0.38
dependent	0.00	-0.1	0.00
spending	0.60	-0.2	-0.12
		sum:	0.57
		h:	0.64

h indicates the probability of customer being good

h = 0.64

64% chance that he will be good

36% chance that he will be bad

Classification Cost Function

Cost function: sum of errors from classifying sample i $\sum_i (E_i)$

$$E_i = -(y^i log(h(x^i)) + (1 - y^i) log(1 - h(x^i)))$$

This cost function is 'cross entropy' error, defining how actual classes match your class predictions.

h	y	cost
0.8	1	0.10
0.1	1	1.00
0.1	0	0.05

Classification Performance

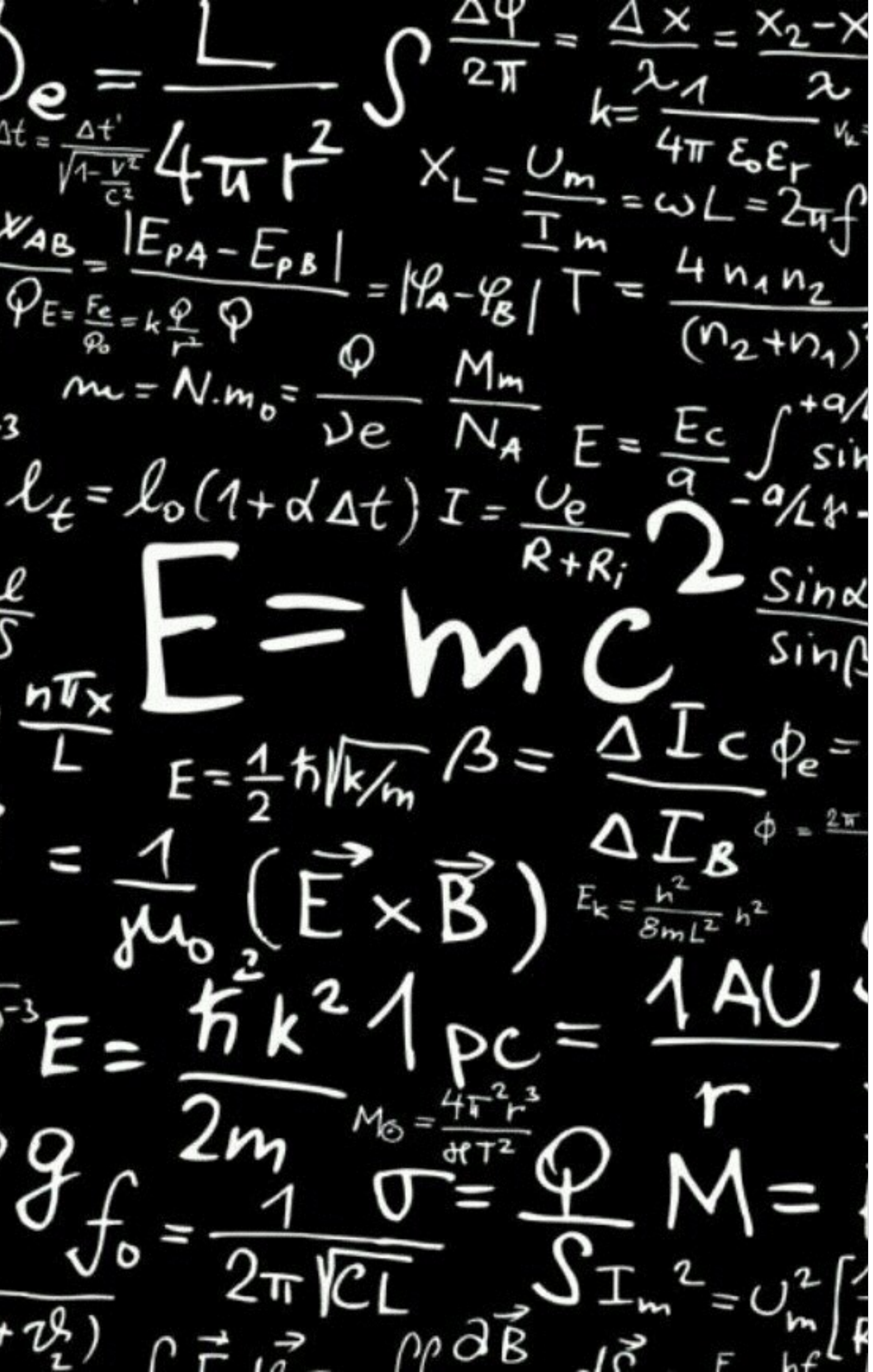
- Log Loss
- Accuracy
- Precision
- Recall
- F1-Measure

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

$$PRE = \frac{TP}{TP + FP}$$

$$REC = TPR = \frac{TP}{P} = \frac{TP}{FN + TP}$$

$$F_1 = 2 \cdot \frac{PRE \cdot REC}{PRE + REC}$$



GENERALIZED LINEAR MODEL

Generalized Linear Model

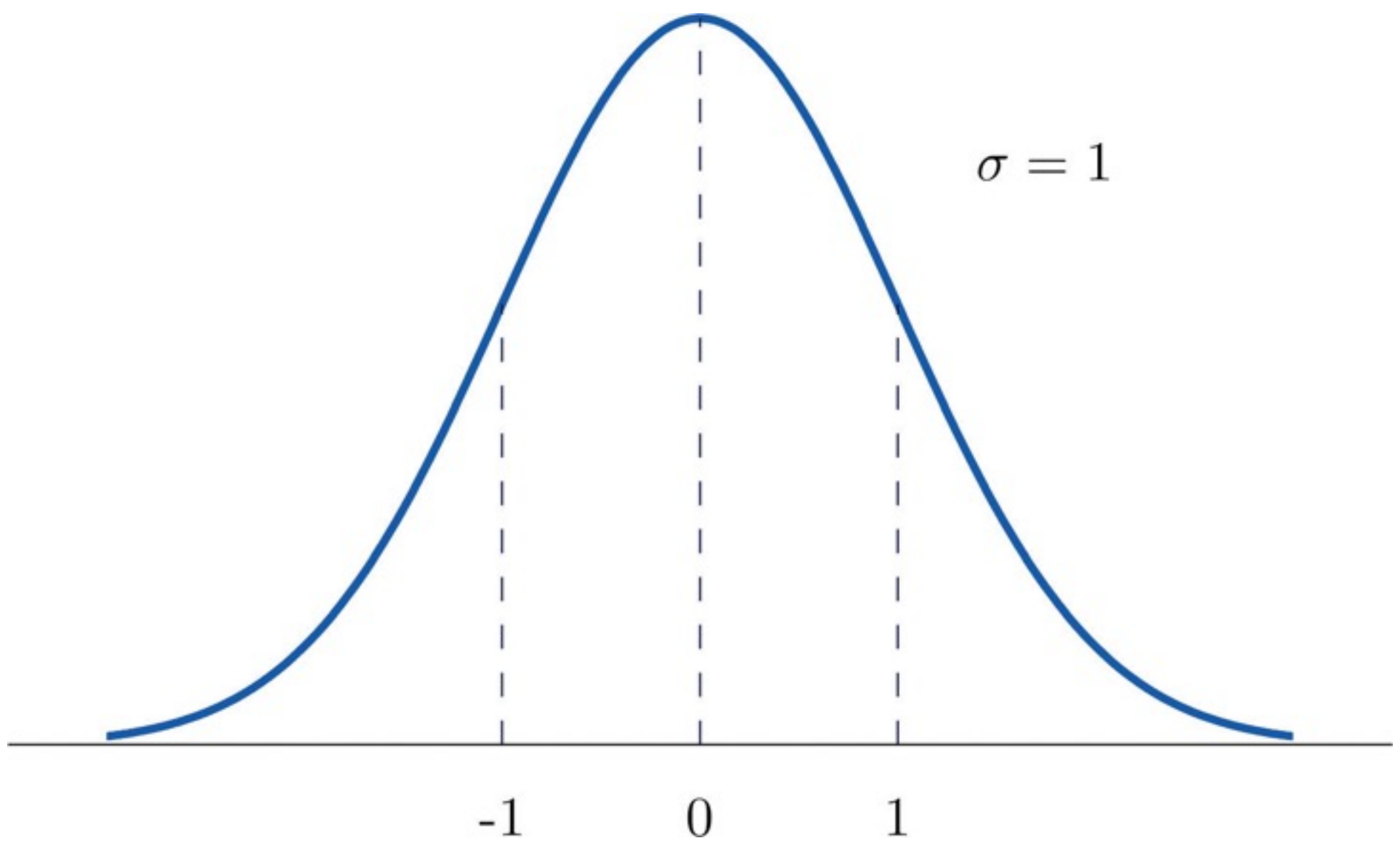
The functional form is the same as logistic regression.

$$h(x) = f(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + ...)$$

But the difference lies in the ASSUMPTIONS about the distribution of the response variables h(x).

Linear Model

$$h(x) = f(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + ...)$$

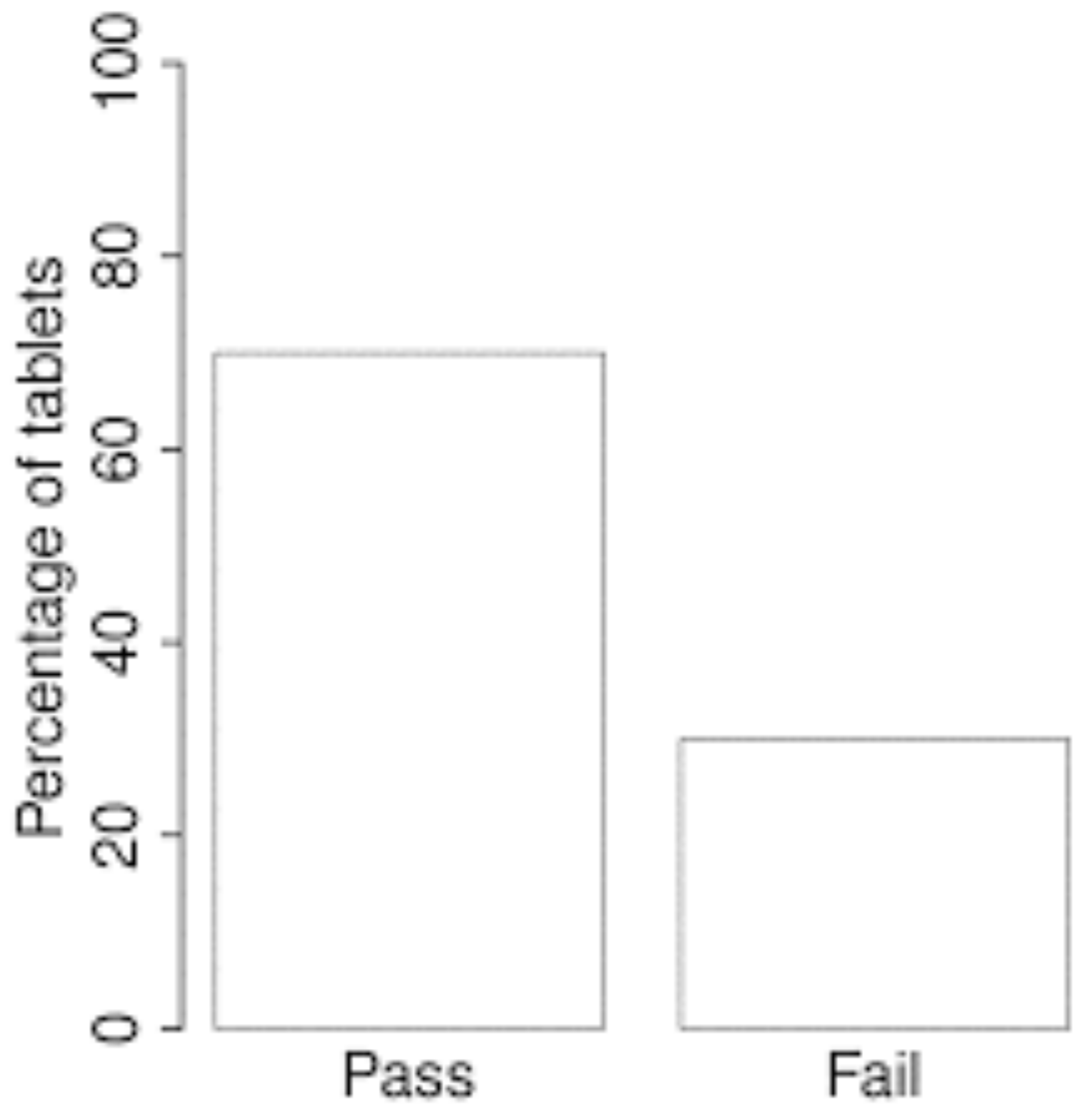


$$f(z) = z$$

In a linear model, f(x) is an identity function and h(x) has a normal distribution.

Logistic Linear Model

$$h(x) = f(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + ...)$$



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

In a logistic linear model, f(x) is a logistic function and h(x) has a binomial distribution.

Bernoulli Trials

- Definition: Suppose an experiment can have only two possible outcomes, e.g., the customer churns.
  - Each performance of the experiment is called a Bernoulli trial.
  - One outcome is called a success and the other a failure.
  - If p is the probability of success and q the probability of failure, then $p + q = 1$.
- Many problems involve determining the probability of k successes when an experiment consists of n mutually independent Bernoulli trials.

Poisson Process

- Series of events (= trials), each producing a result
- The events occur randomly but at a specific rate (mean number of events per unit time, space, volume, etc.)
- The outcome (result) of each event can only have discrete values (for example, these results can be 0/1)
- Each possible result has a fixed probability
- The outcome of each event is random
- The outcome of each event is independent from the others (nor affected by the others).

Poisson Process

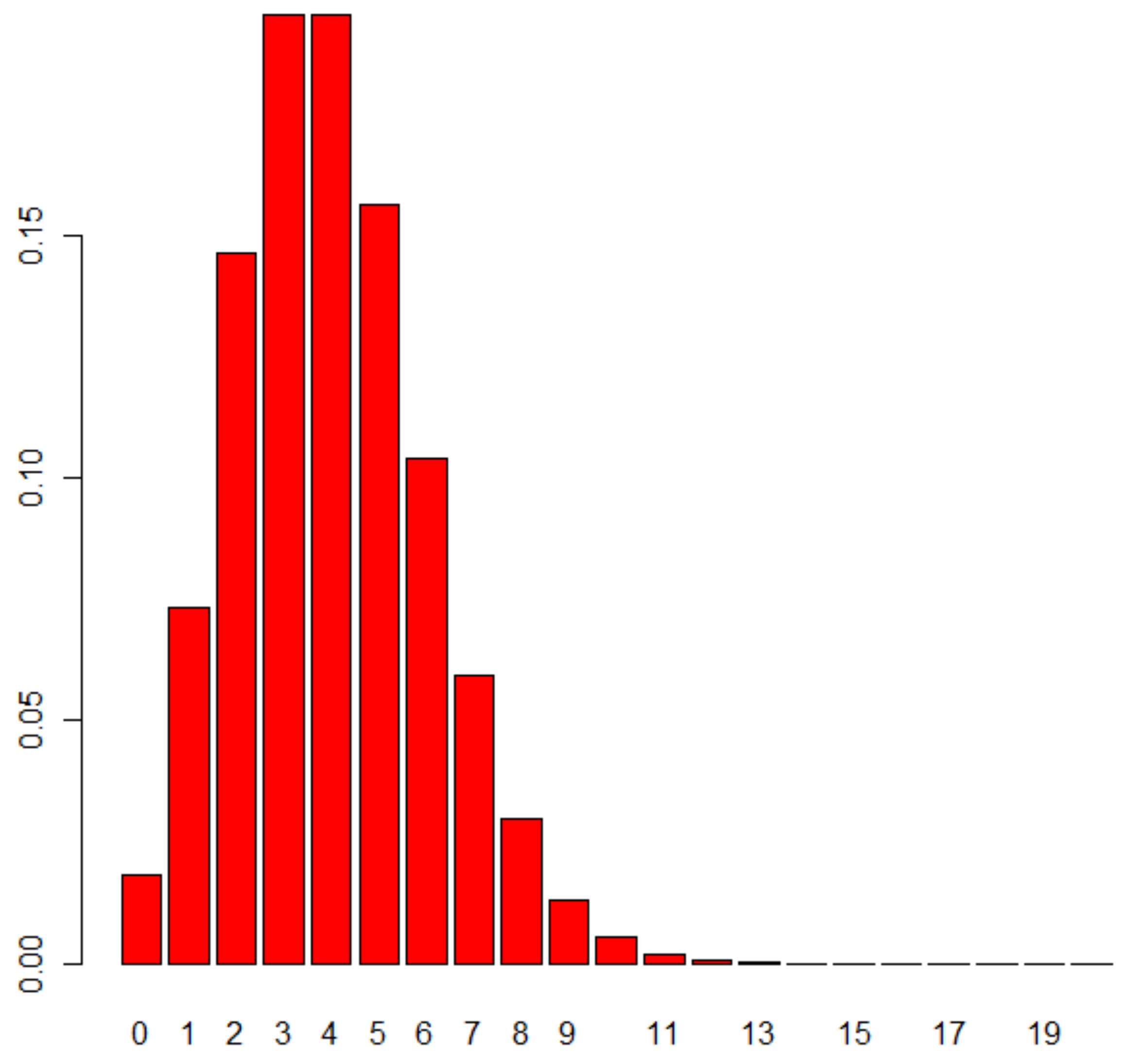
- Poisson process is very useful model because it describes the probability of events that
  - are rare
  - but do occur at a certain mean rate over a given interval of time
  - one consequence is that the probability of such event increases with the time lapsed since the last occurred.

Poisson distribution

- Describes the probability distribution that a certain event, with a result (a), will occur x times within an interval of time t, given that the mean rate of occurrence (frequency) of such an event is λ.

$$P(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \quad \text{or} \quad P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \ (\text{if } t = 1)$$

$$\lambda > 0; x = 1, 2, 3....$$



Mean and Variance of Poisson Distribution

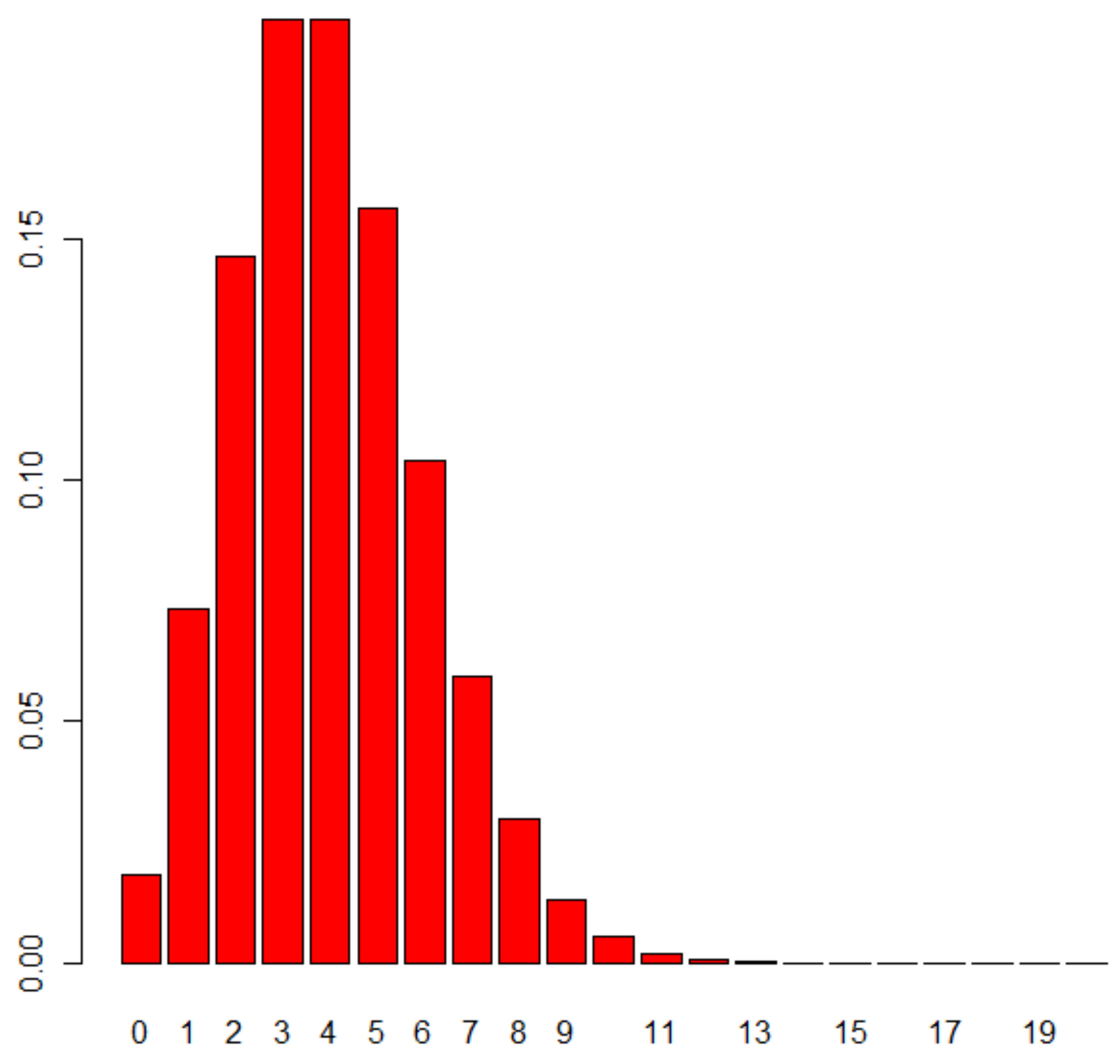
- The intrinsic property is λ (long-term frequency of the event, or its return period $1/\lambda$)
  - Mean (μ)= λ
  - Variance (σ^2) = λ

Poisson Regression

- Now we consider "Poisson regression"
  - Response is a count, assumed to have a Poisson distribution with (positive) mean μ
  - Assume that $\log \mu$ is a linear function of the covariates
- Alternatively,
  - μ = exp(linear function of covariates)
  - Poisson is a standard distribution when response is a count.

Poisson Regression

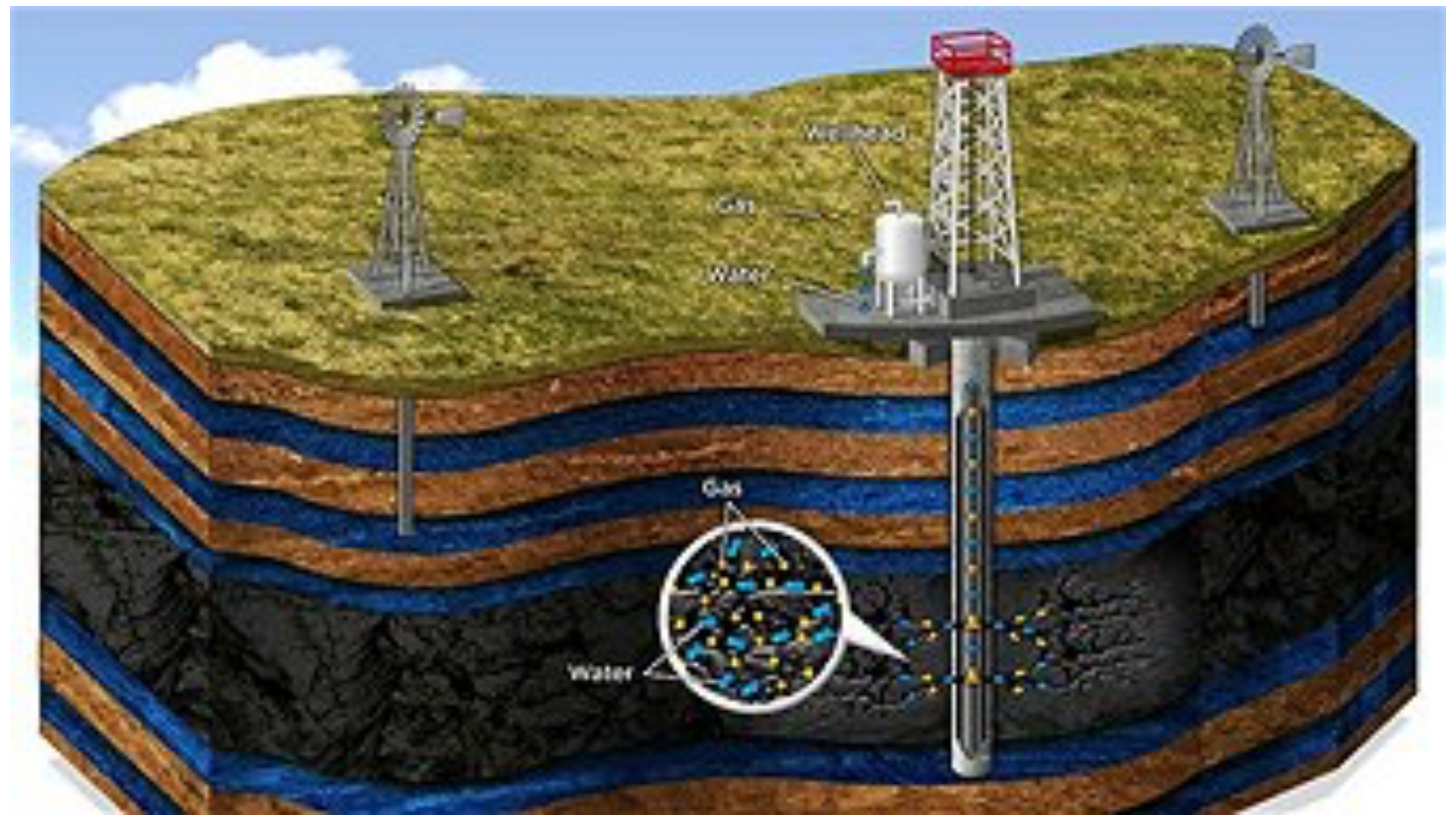
$$log(h(x)) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ...$$



In poisson regression, it's easier to talk about the inverse of function f(x) (a.k.a link function). This regression model would yield the predicted mean of poisson distribution μ or λ.

Poisson Regression

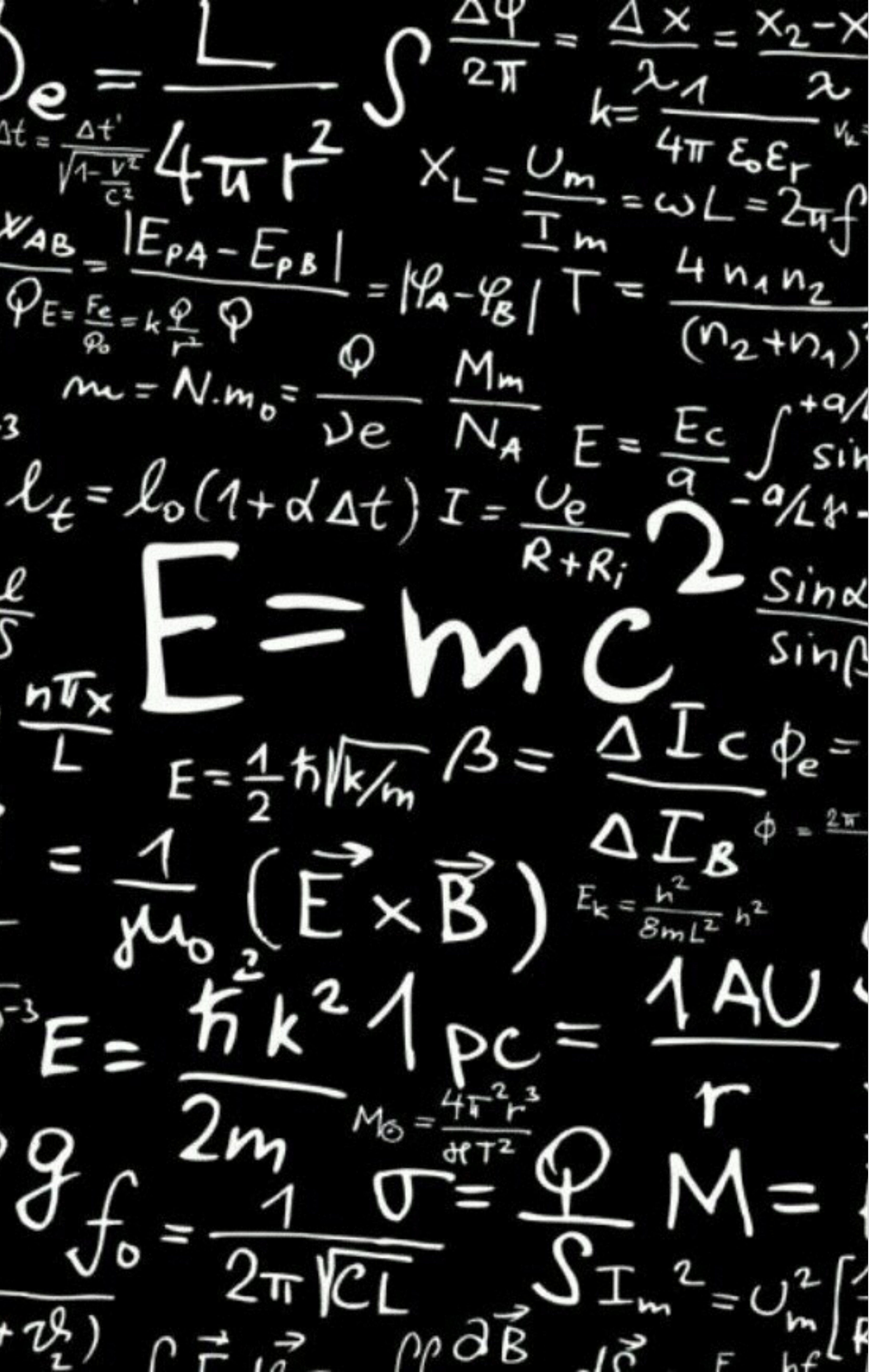
- This example features the number of accidents per mine in a 3 month period in 44 coal mines in West Virginia. The variables are
  - COUNT: the number of accidents (response)
  - INB: inner burden thickness
  - EXTRP: percentage of coal extracted from mine
  - AHS: the average height of the coal seam in the mine
  - AGE: the age of the mine



Poisson Regression

$$log(h(x)) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ...$$

- The fitted parameters would entail how rate of accidents vary with factors of x's.
  - If $\theta_j>0$, mean increases with x_j
  - If $\theta_j<0$, mean decreases with x_j
- Unit increase in x_j changes the mean by a factor of $exp(\theta_j)$



REGRESSION LAB