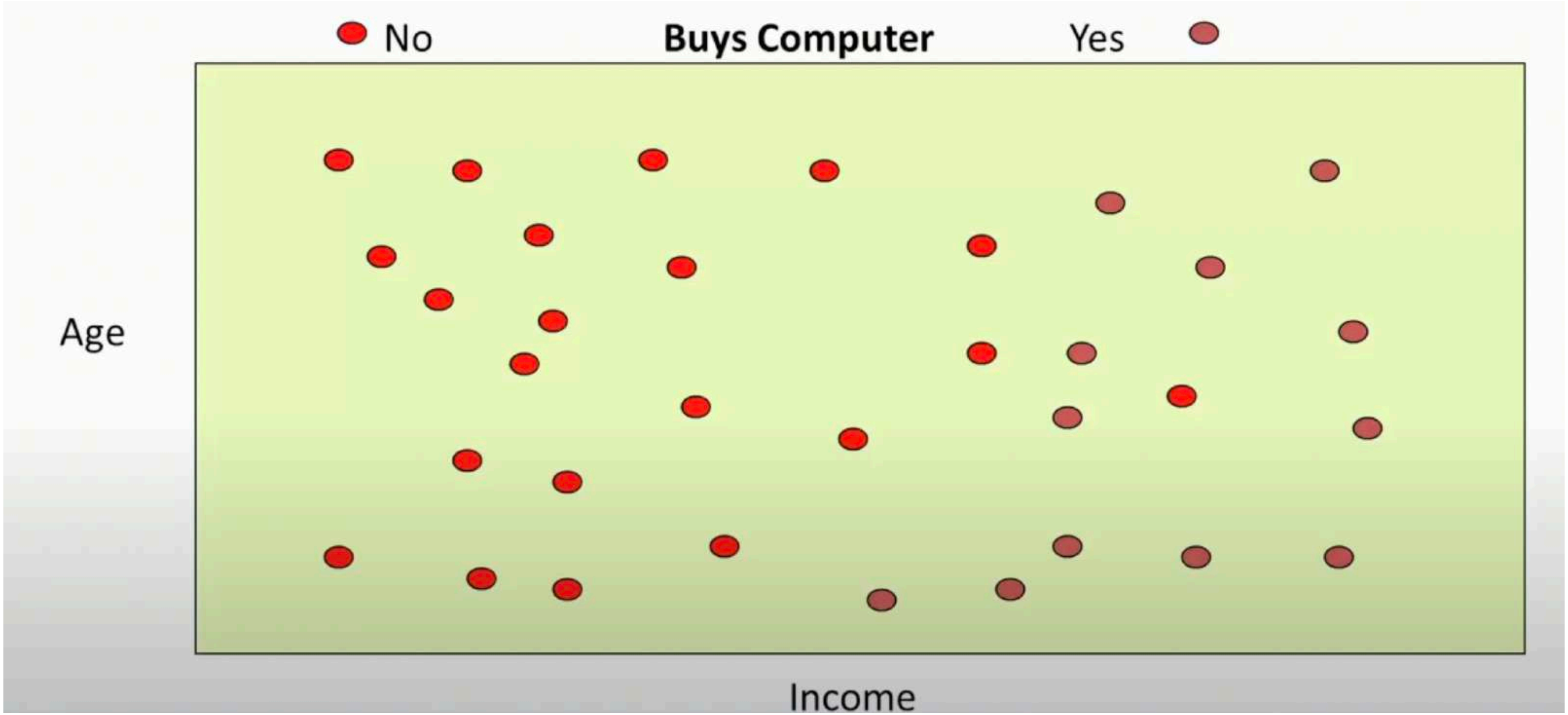
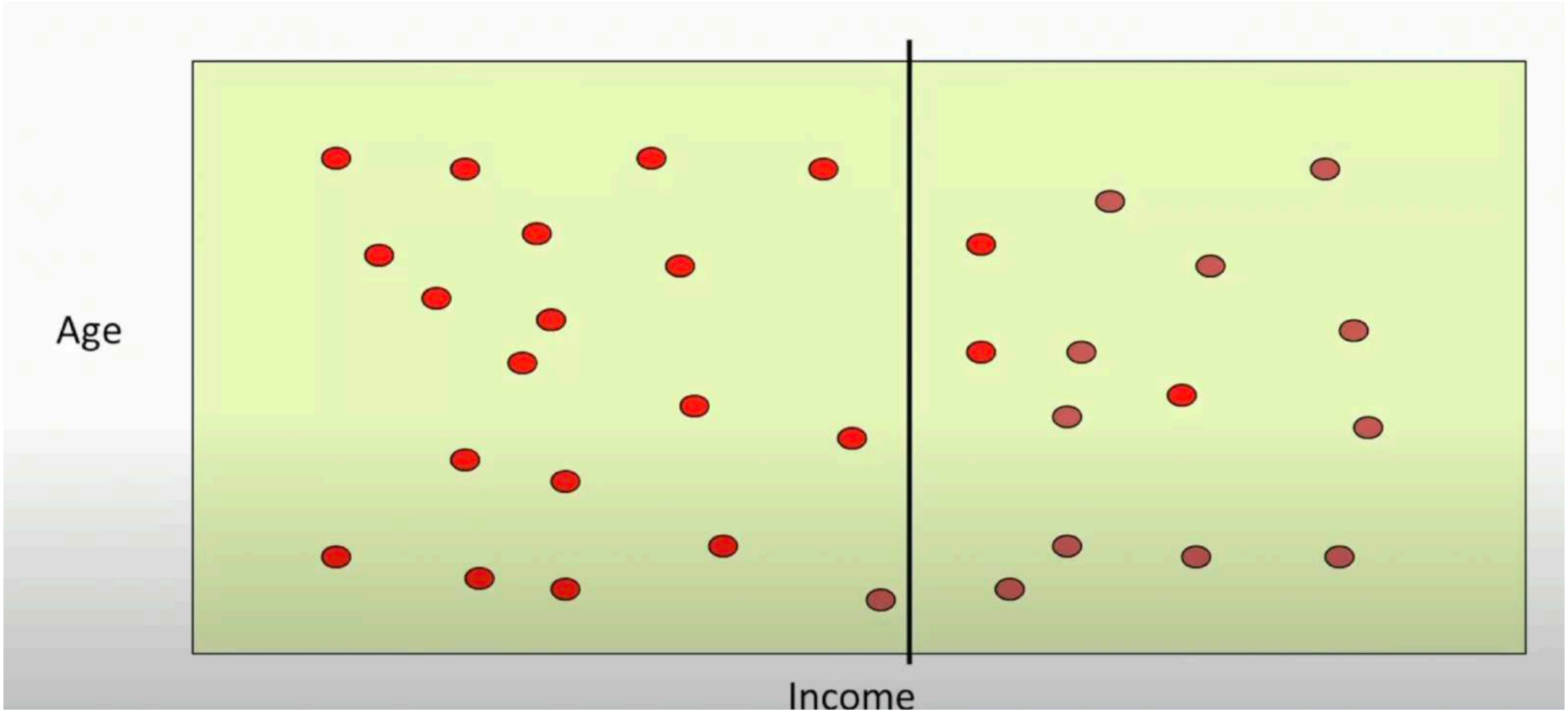
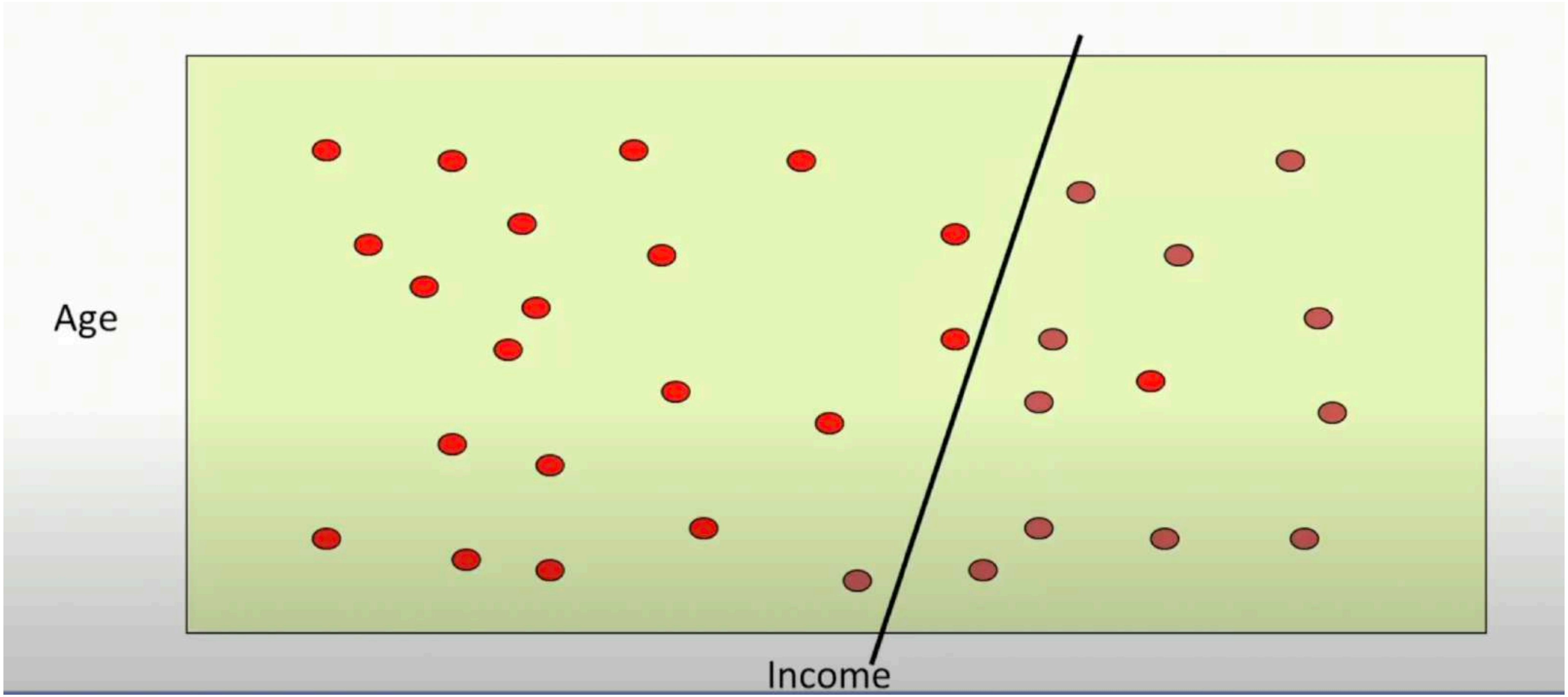
Logistic Regression

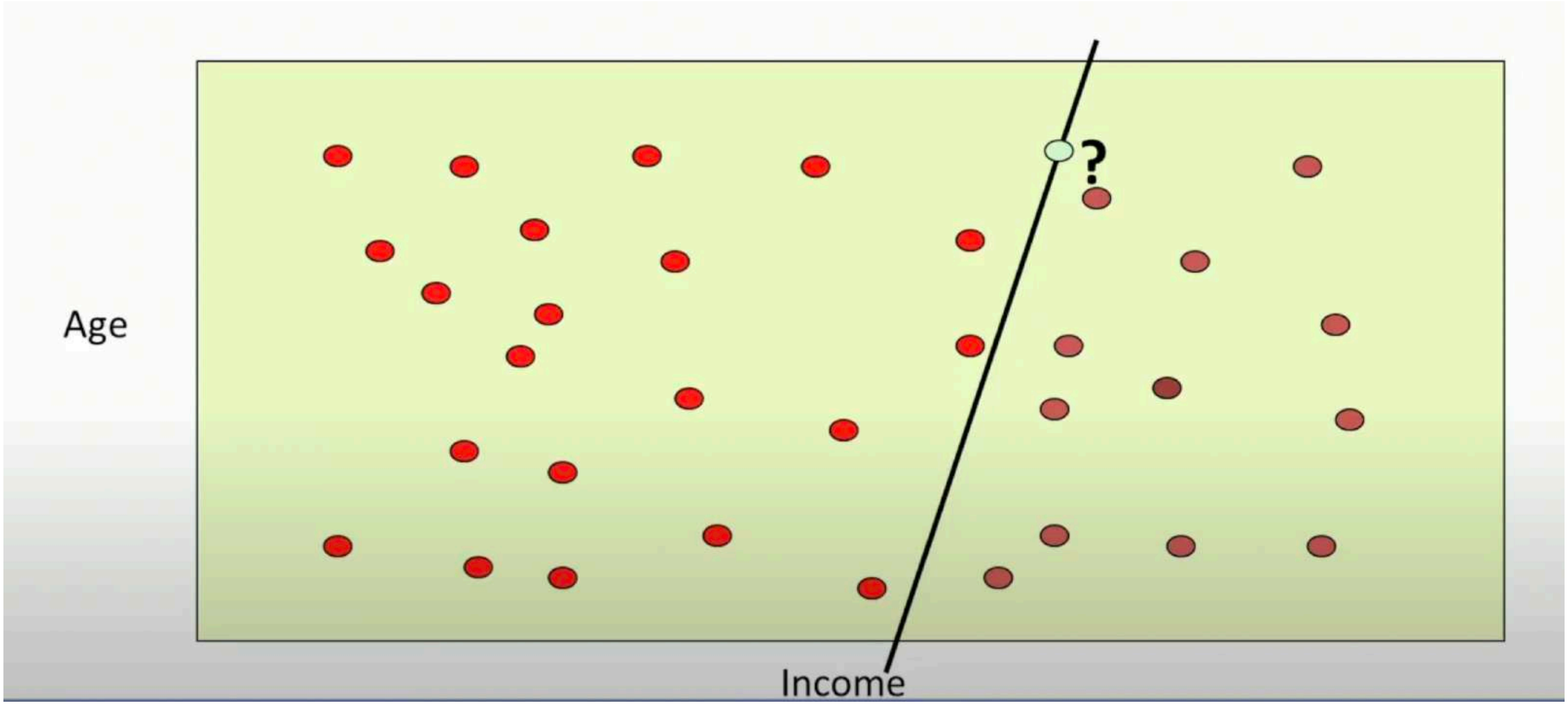
Logistic Regression: What And Why?

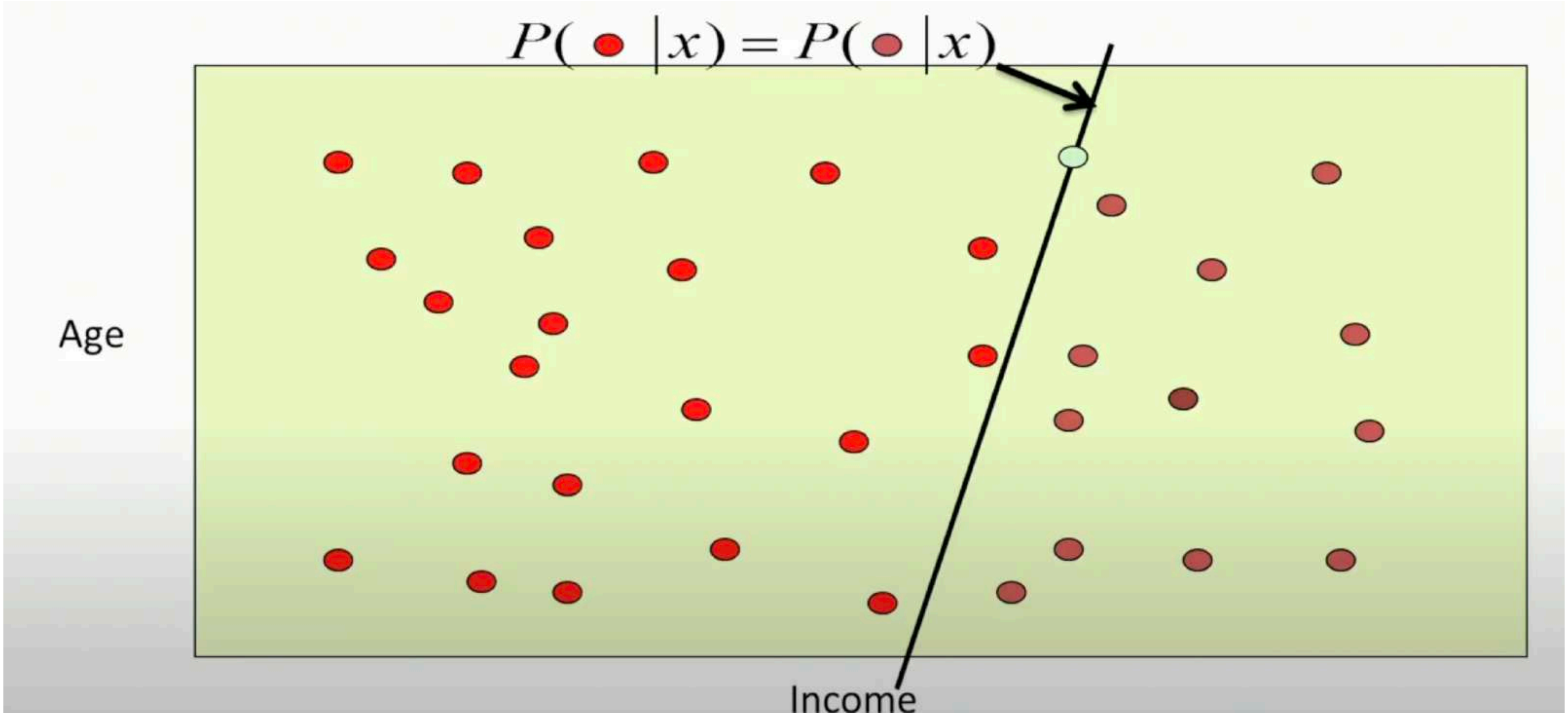
Logistic Regression produces results in a binary format which is used to predict the outcome of a categorical dependent variable. So the outcome should be discrete/ categorical such as:


No
Buys Computer
Yes
Age
Income

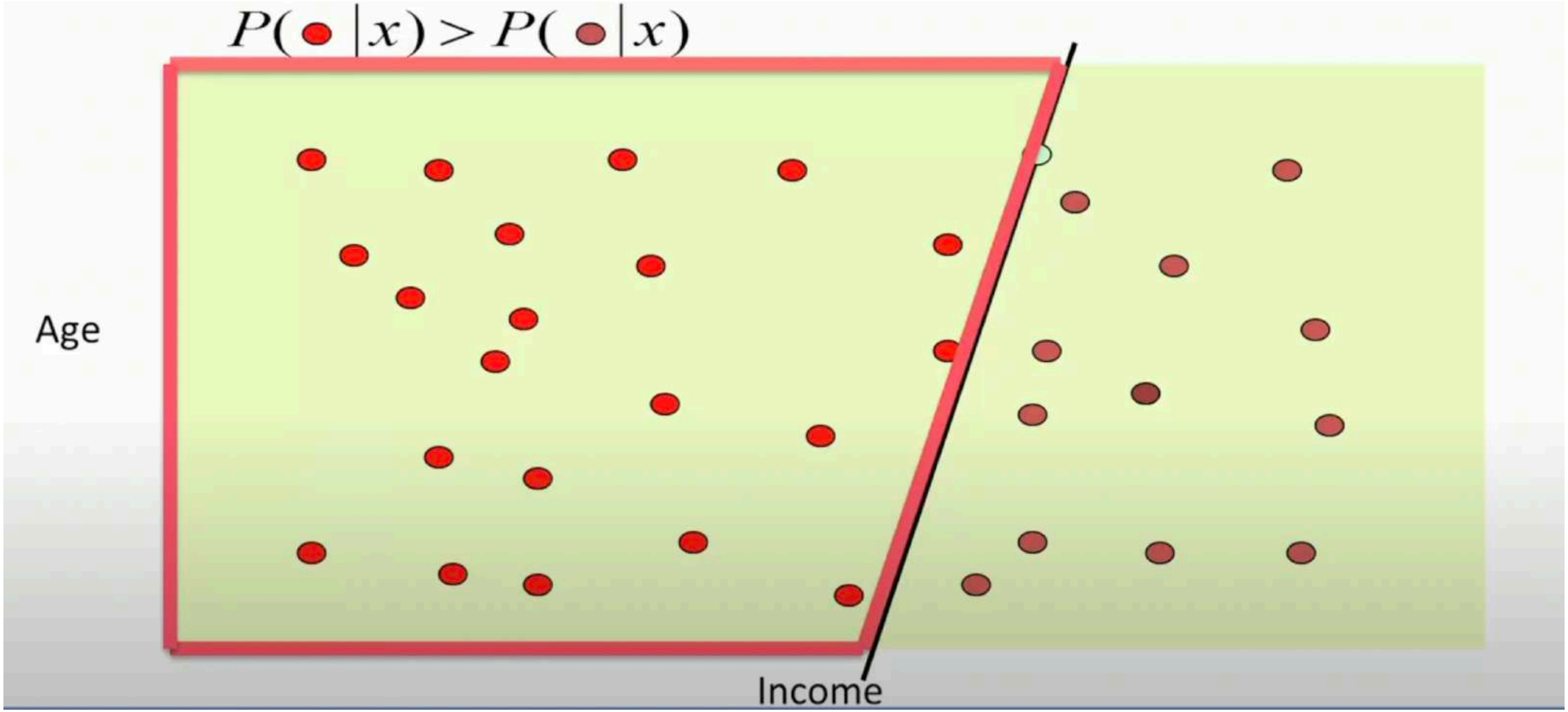

Age
Income


Age
Income


?
Age
Income



$P(\bullet|x) = P(\bullet|x)$
Age
Income



P(| x) > P(| x)
Age
Income

- Interested in knowing $p(c|x)$
  - Not just in the *right* classification
  - E.g. Medical domain
  - Confidence of classification
- Treat as a regression problem?

- Use an indicator variable for class
  - 1 for *buys* and 0 for *does not buy*

$$X_1 = \langle 30000, 25 \rangle, Y_1 = DoesnotBuyComputer$$
$$X_2 = \langle 80000, 45 \rangle, Y_2 = BuysComputer$$
$$\vdots$$

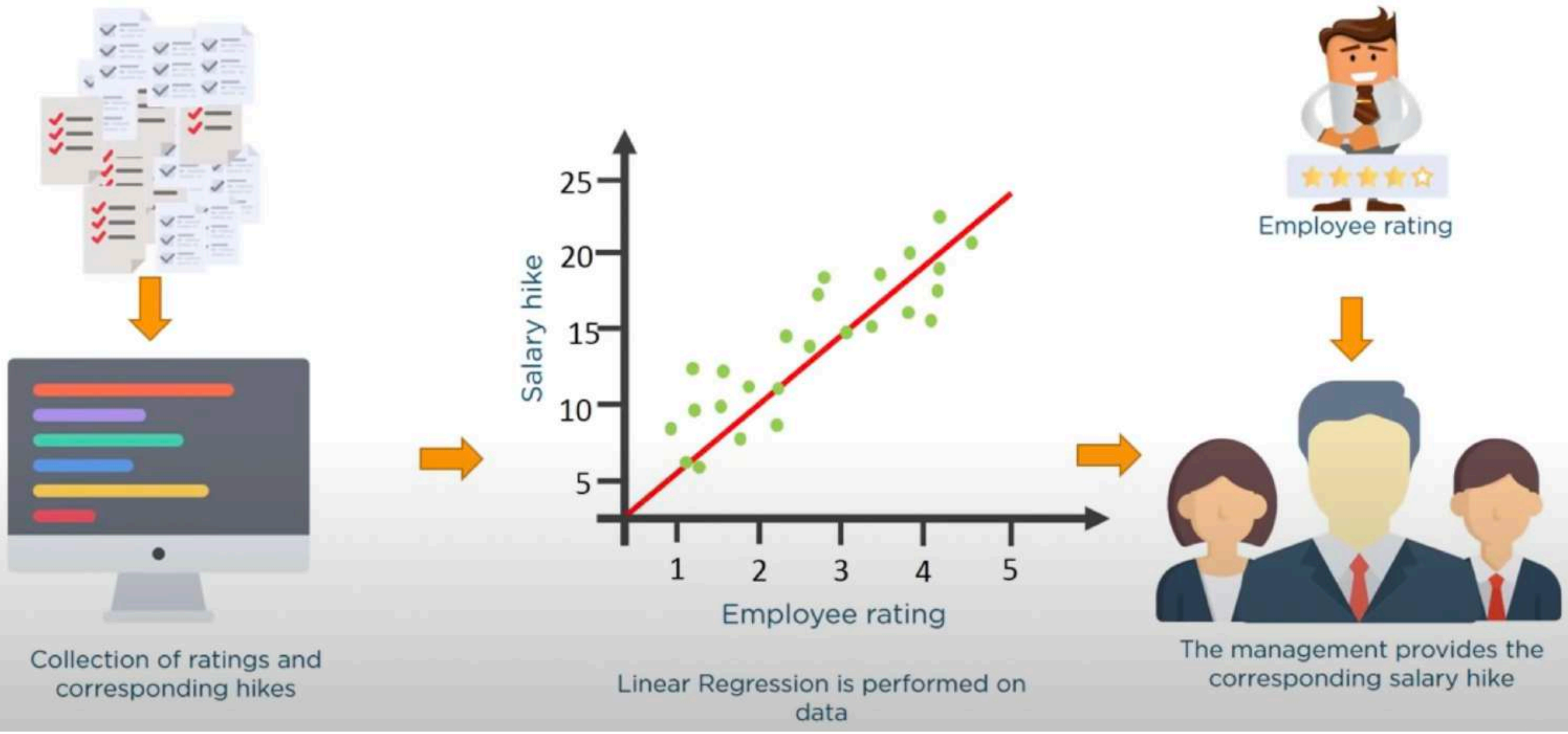
$$\longrightarrow$$

$$X_1 = \langle 0.15, 0.25 \rangle, Y_1 = 0$$
$$X_2 = \langle 0.4, 0.45 \rangle, Y_2 = +1$$
$$\vdots$$

  - Use linear regression!
  - $f(x)$ can be interpreted as $p(y=1|x)$

- What are the problems?
  - Linear regression is not limited in range
  - Output cannot be interpreted as a probability
    - Can be negative!
  - Works in practice, but not that well

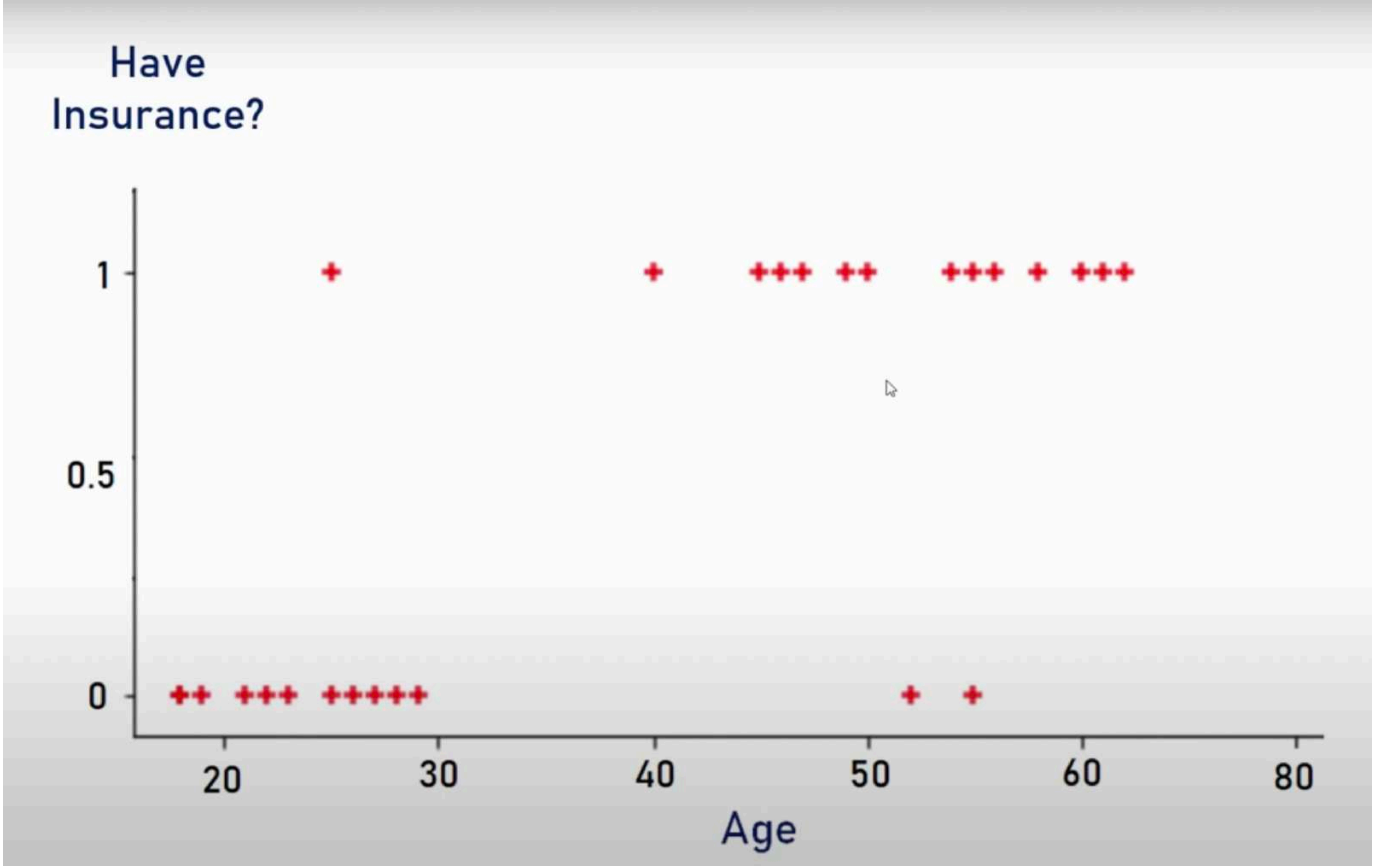
Linear Regression

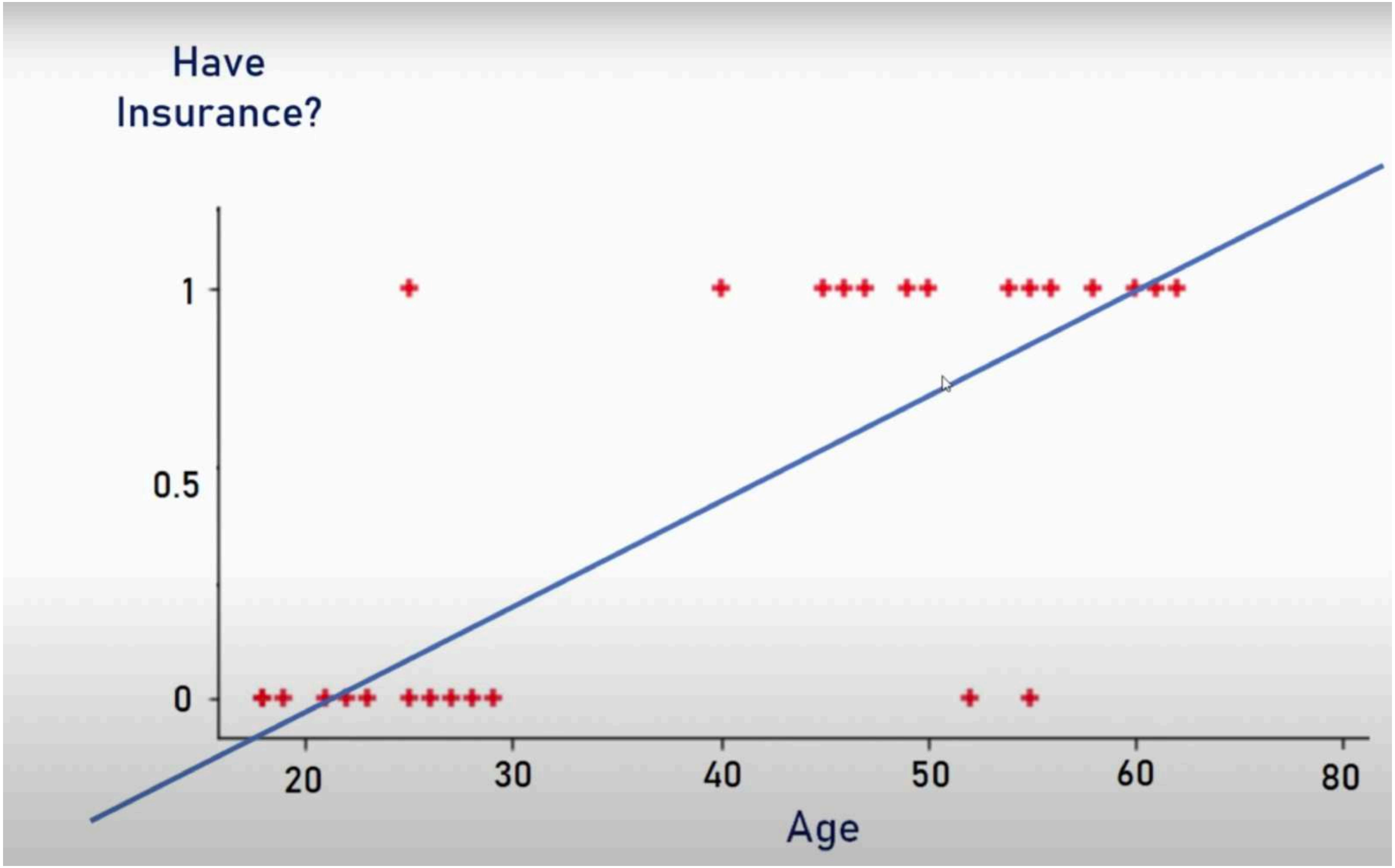


Linear and Logistic Regression

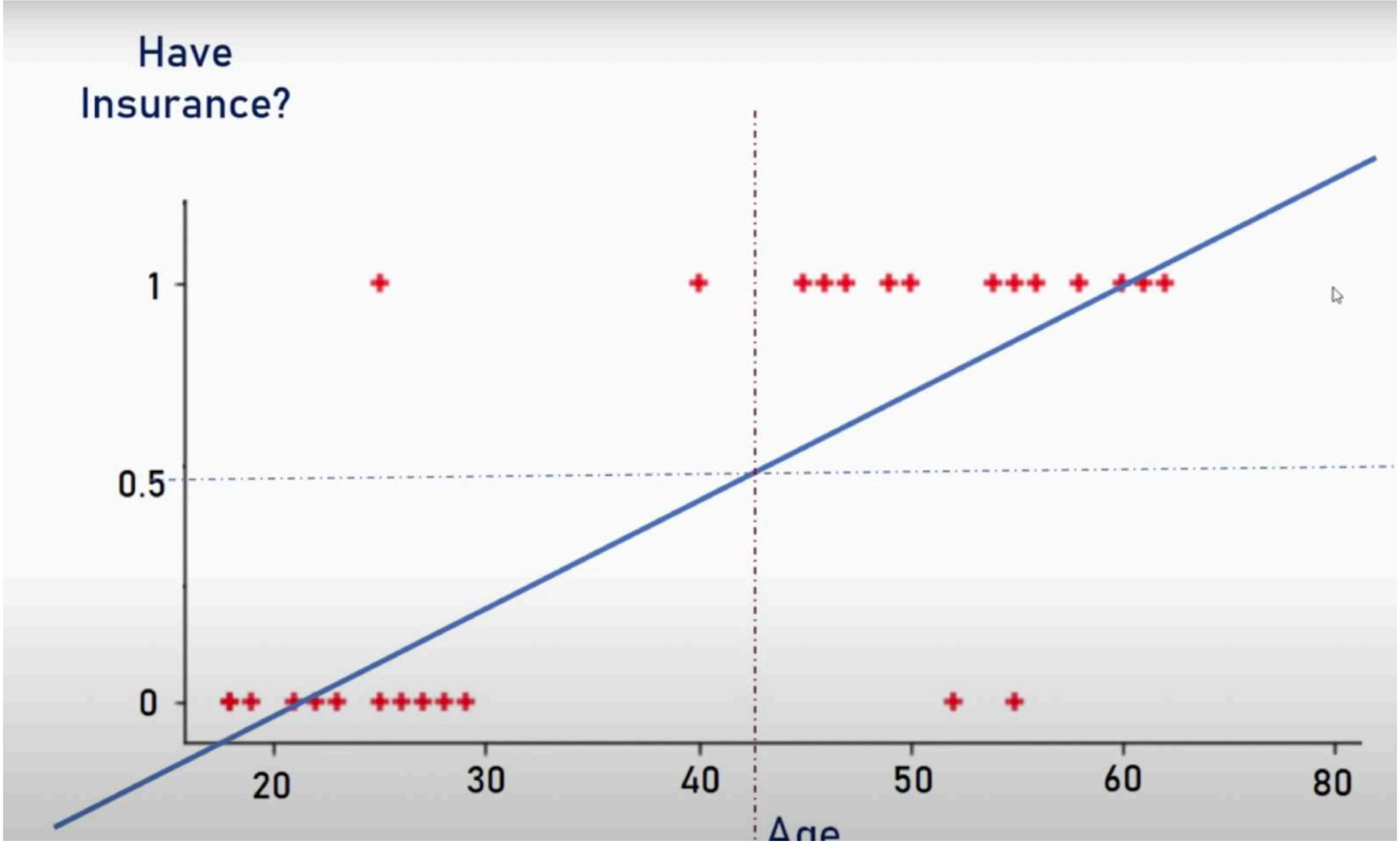
What if you wanted to know whether the employee would get a promotion or not based on their rating

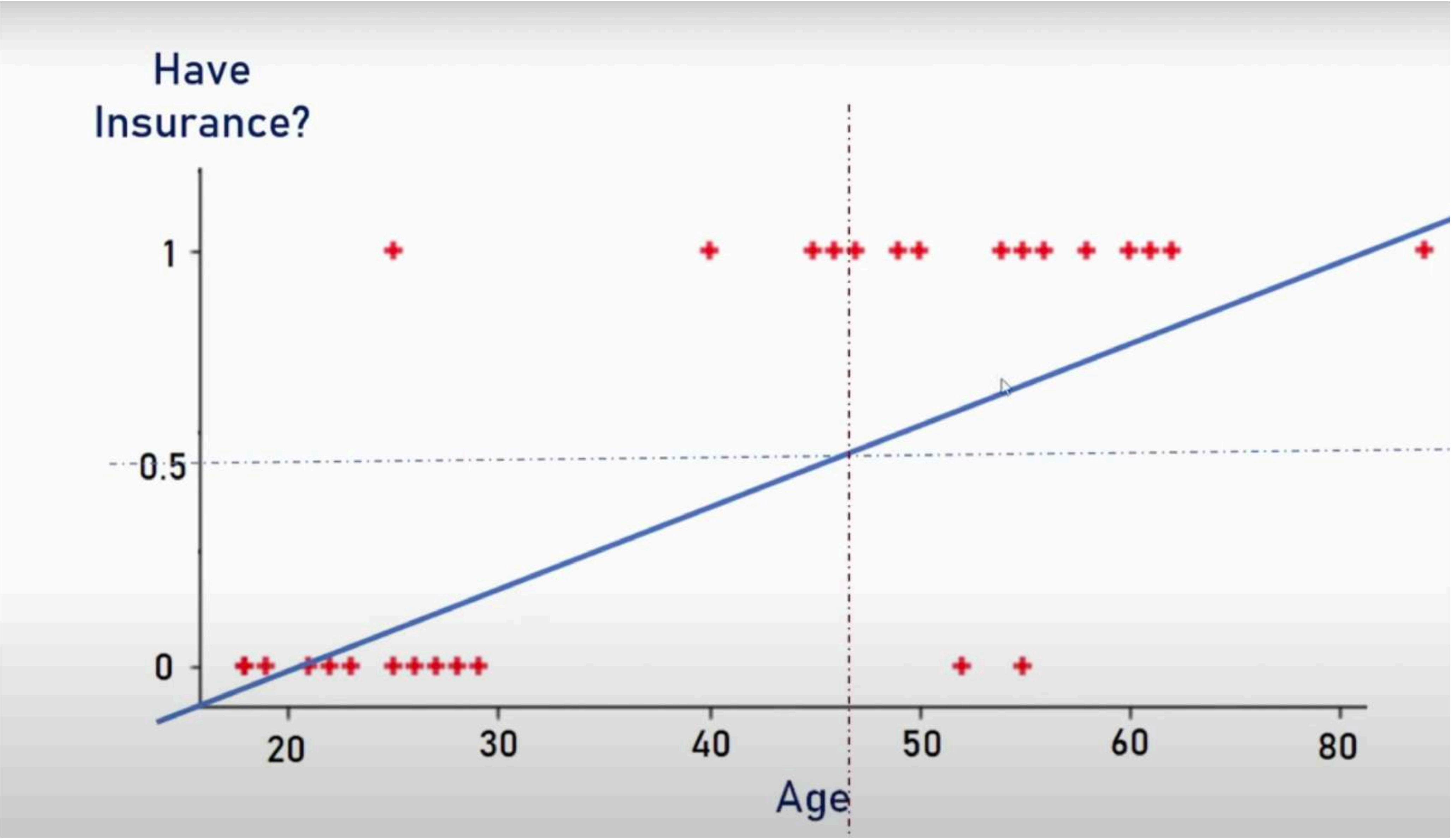
age	have_insurance
22	0
25	0
47	1
52	0
46	1
56	1
55	0
60	1
62	1
61	1
18	0
28	0
27	0
29	0
49	1


Have
Insurance?
1
0.5
0
20
30
40
50
60
80
Age

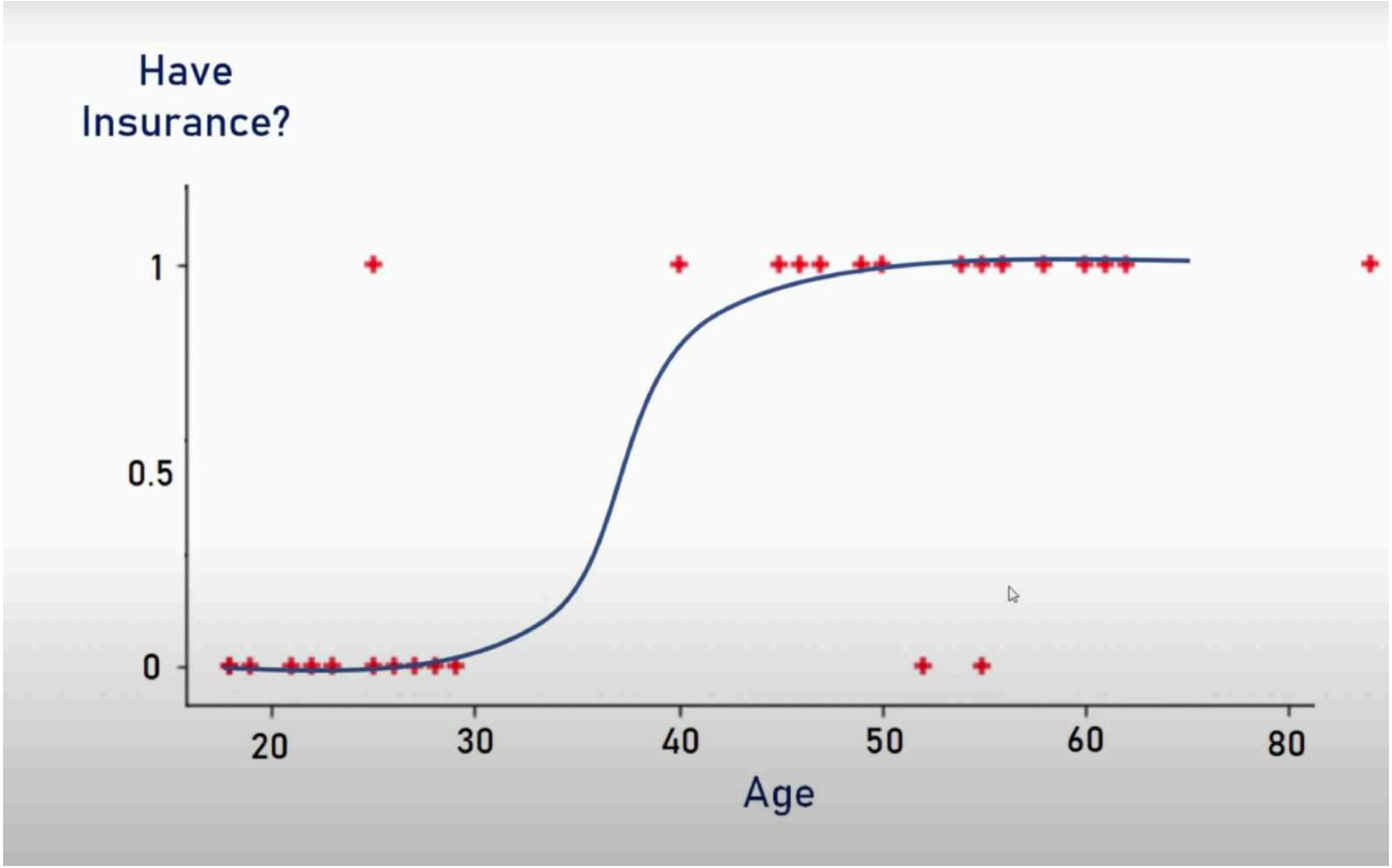


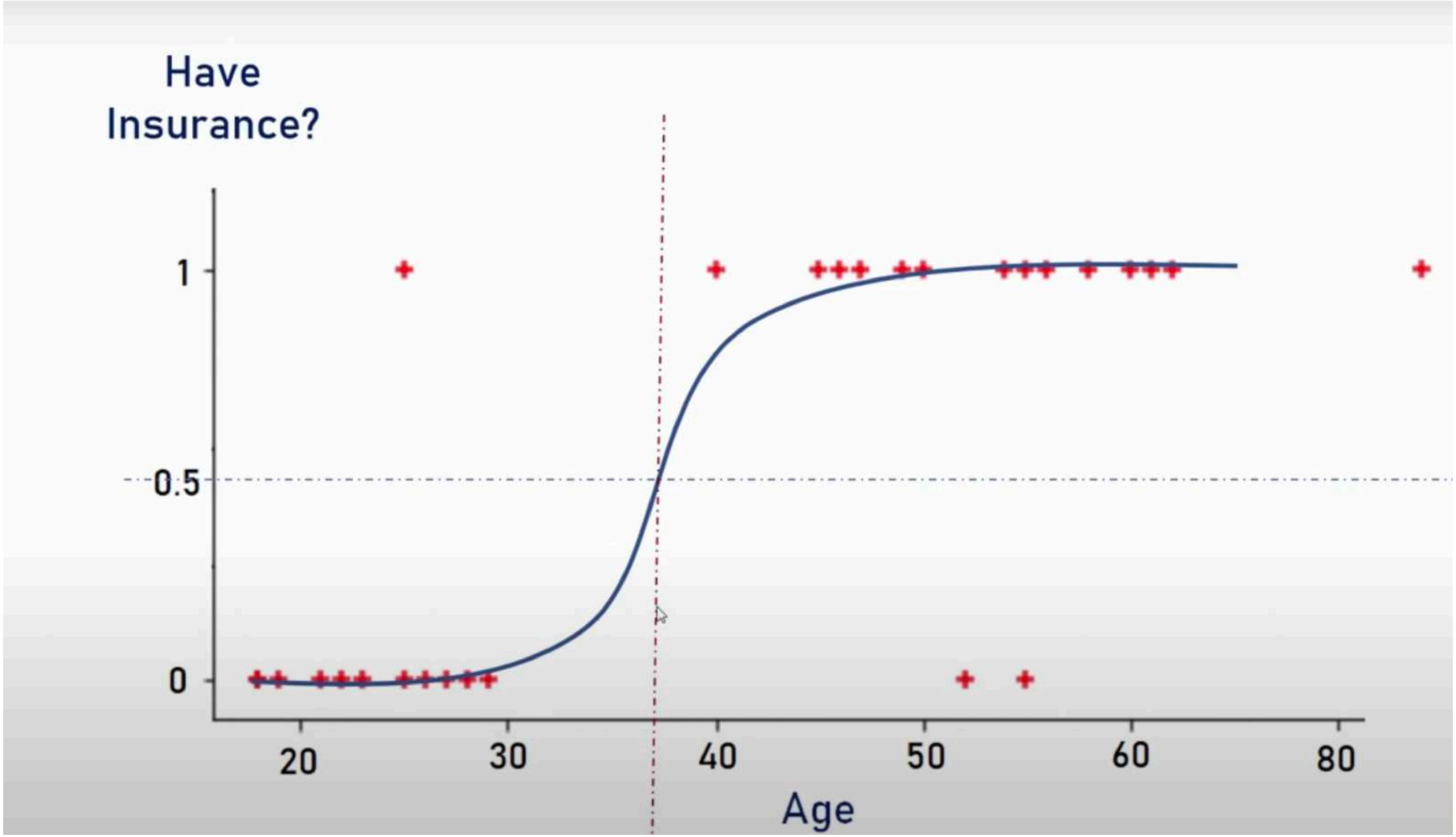
Have
Insurance?
1
0.5
0
20
30
40
50
60
80
Age


Have
Insurance?
1
0.5
0
20
30
40
50
60
80
Age



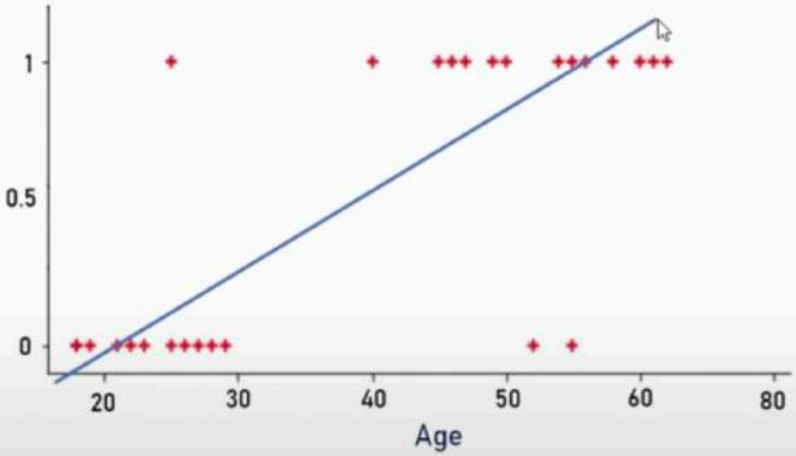
Have Insurance?
1
0.5
0
20
30
40
50
60
80
Age

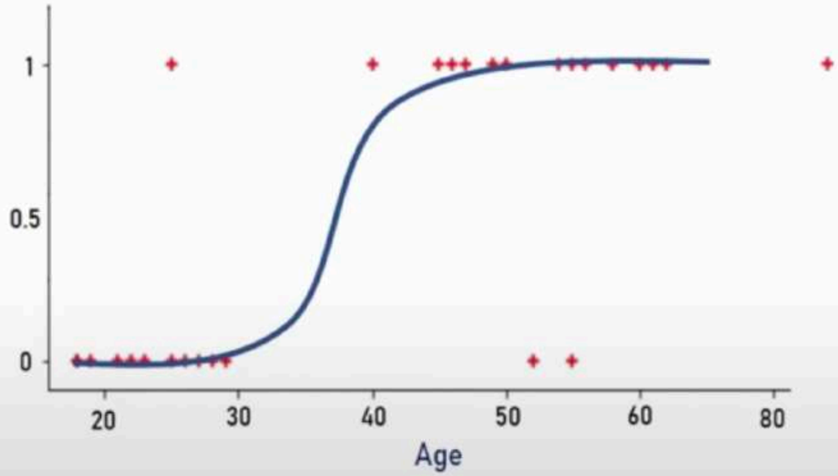

Have
Insurance?
1
0.5
0
20
30
40
50
60
80
Age


Have
Insurance?
1
0.5
0
20
30
40
50
60
80
Age

$$y = m * x + b$$

$$y = \frac{1}{1 + e^{-(m*x+b)}}$$





Logistic Regression

Use linear regression still?

On a transformed function

Logistic or Logit function

- Log-Odds
- Let $p(x)$ denote the $p(y=1|x)$
- Logit transformation is given by:

$$\log\left(\frac{p(x)}{1-p(x)}\right)$$

Logistic Regression

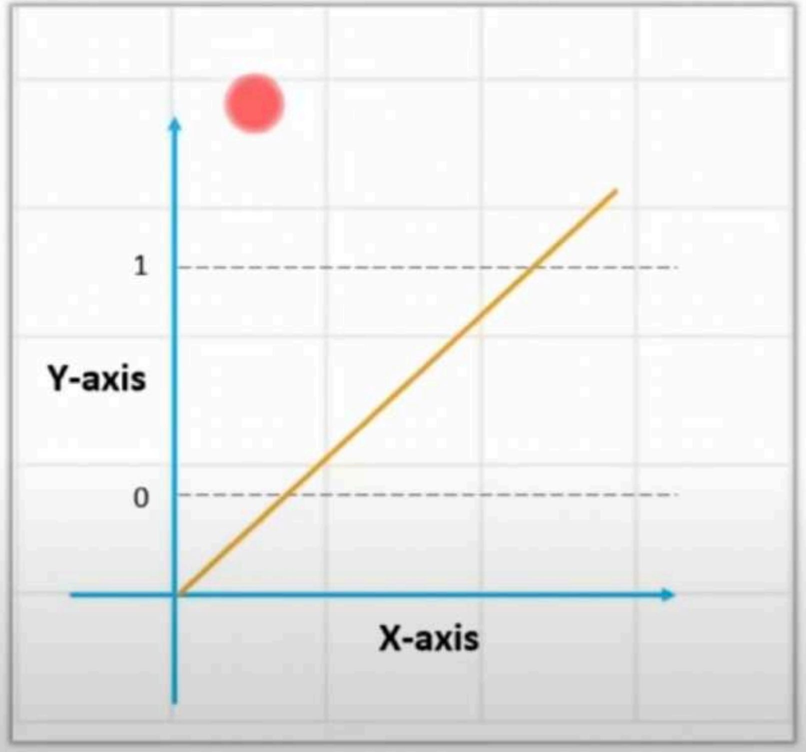
Formally a logistic regression model tries to fit:

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + x \cdot \beta_1$$

Solving for *p(x)*

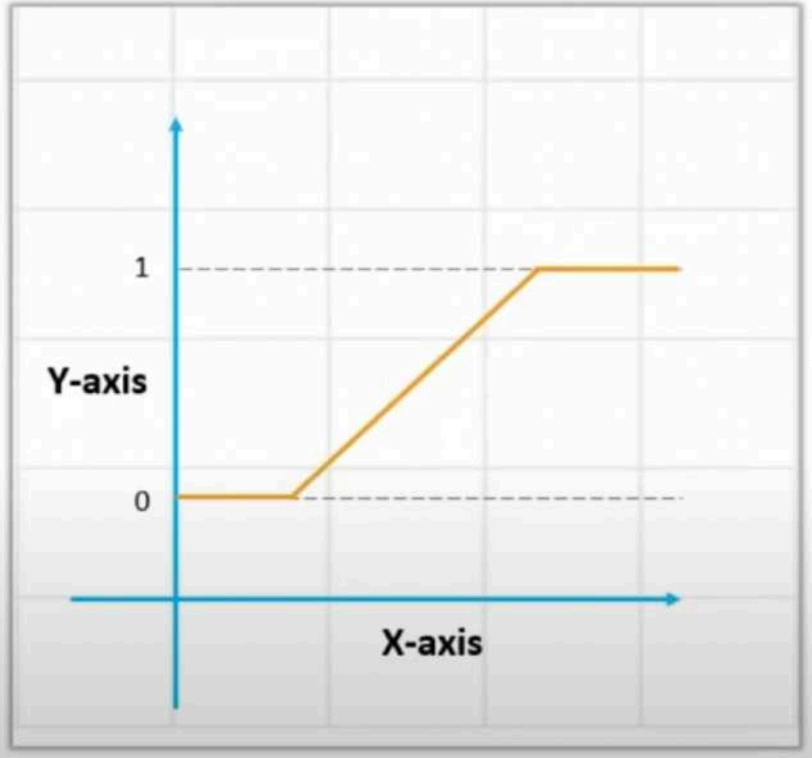
$$p(x) = \frac{e^{\beta_0 + x\cdot\beta}}{1+e^{\beta_0 + x\cdot\beta}} = \frac{1}{1+e^{-(\beta_0 + x\cdot\beta)}}$$

Why Not Linear Regression?



Since our value of Y will be between 0 and 1, the linear line has to be clipped at 0 and 1.

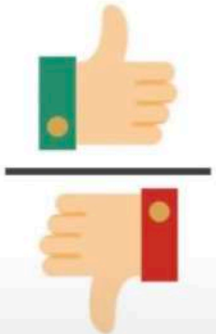
Why Not Linear Regression?



With this, our resulting curve cannot be formulated into a single formula. Hence we came up with **Logistic**!

The Math behind Logistic Regression

To understand Logistic Regression, let's talk about the odds of success

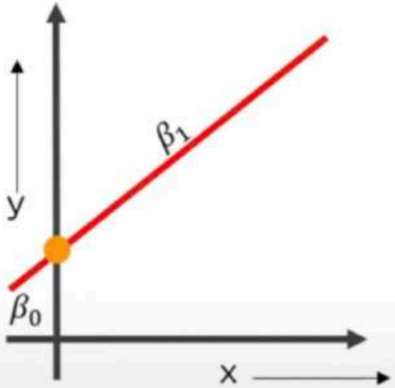
Odds (θ) = 

$$\frac{\text{Probability of an event happening}}{\text{Probability of an event not happening}} \quad or \quad \theta = \frac{p}{1-p}$$

The values of odds range from 0 to ∞
The values of probability change from 0 to 1

The Math behind Logistic Regression

Take the equation of the straight line



Here, β_0 is the y-intercept
β_1 is the slope of the line
x is the value of the x co-ordinate
y is the value of the prediction

The equation would be: $y = \beta_0 + \beta_1 x$

Logistic Regression Equation

The Logistic Regression Equation is derived from the Straight Line Equation

Equation of a straight line

Y = C + BIX1 + B2X2 +.... —— Range is from –(infinity) to (infinity)

Let's try to reduce the Logistic Regression Equation from Straight Line Equation

Y = C + BIX1 + B2X2 +.... —— In Logistic equation Y can be only from 0 to 1

Now , to get the range of Y between 0 and infinity, let's transform Y

$\frac{Y}{1-Y}$ Y= 0 then 0; Y= 1 then infinity —— Now, the range is between 0 to infinity

Let us transform it further, to get range between –(infinity) and (infinity)

$\log\left[\frac{Y}{1-Y}\right] \Rightarrow$ Y = C + BIX1 + B2X2 +.... —— **Final Logistic Regression Equation**

The Math behind Logistic Regression

Now, we predict the odds of success

$$\log\left(\frac{p(x)}{1-P(x)}\right) = \beta_0 + \beta_1 x$$

Exponentiating both sides:

$$e^{ln\left(\frac{p(x)}{1-p(x)}\right)} = e^{\beta_0+\beta_1 x}$$

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

Let $Y = e^{\beta_0+\beta_1 x}$

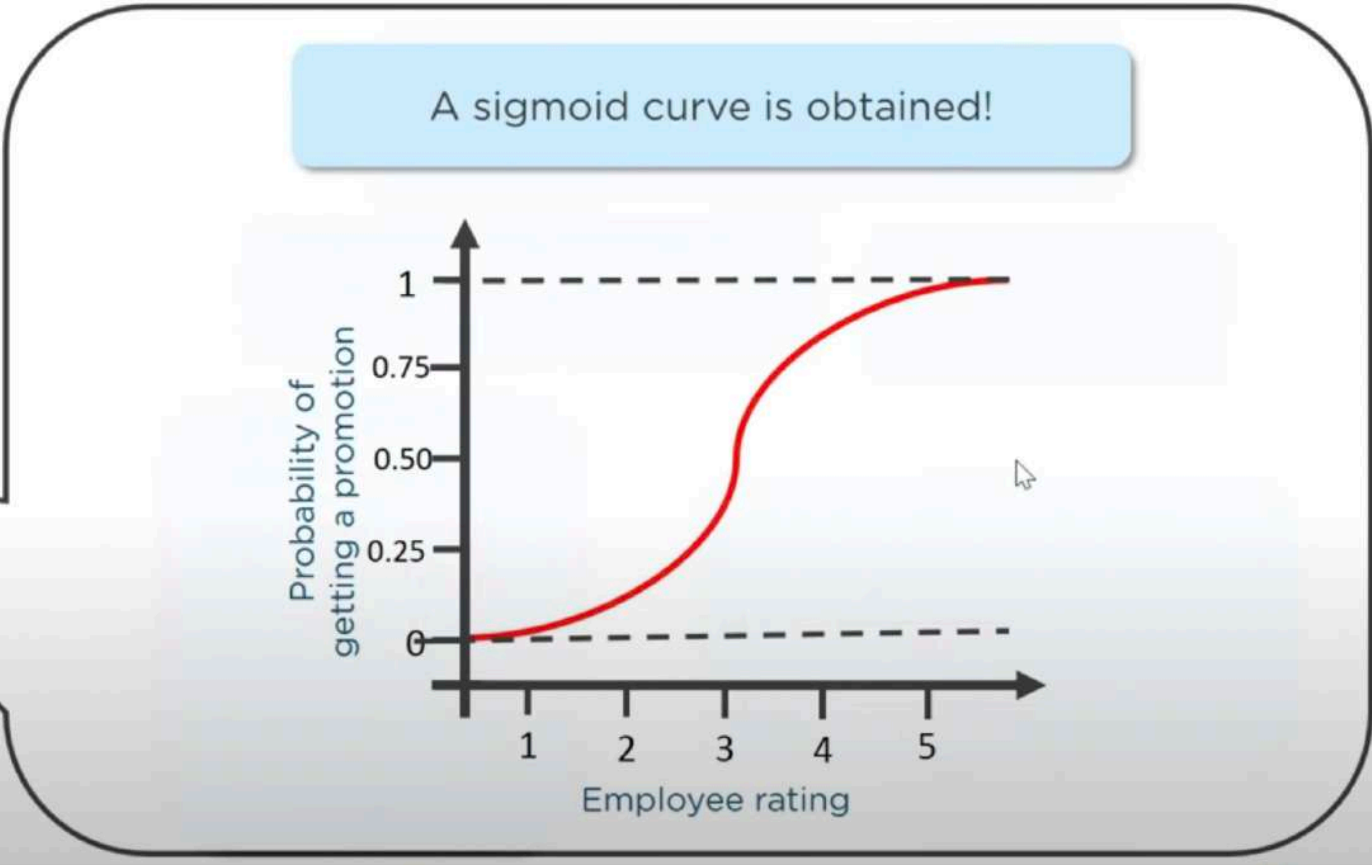
Then $\frac{p(x)}{1-p(x)} = Y$

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

The equation of a sigmoid function:

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

The Math behind Logistic Regression



Logistic Regression

- Predict class is 1 when $p(x) > 0.5$ and 0 otherwise
  - Minimizes misclassification rate
- Linear classifier
  - Decision boundary is: $\beta_0 + x \cdot \beta_1 = 0$
- Powerful
  - Works well in practice

Logistic Regression

Linear regression with a logistic transformation

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + x \cdot \beta_1$$

We optimize the likelihood of the training data with respect to the parameters β

Likelihood

It is the probability of the training data D, given a parameter setting

– It is a function of the parameter, since the training data D is fixed

$$L(\beta_0, \beta) = \prod_{i=1}^{n} p(x_i)^{y_i} \left(1 - p(x_i)\right)^{(1-y_i)}$$

Log Likelihood

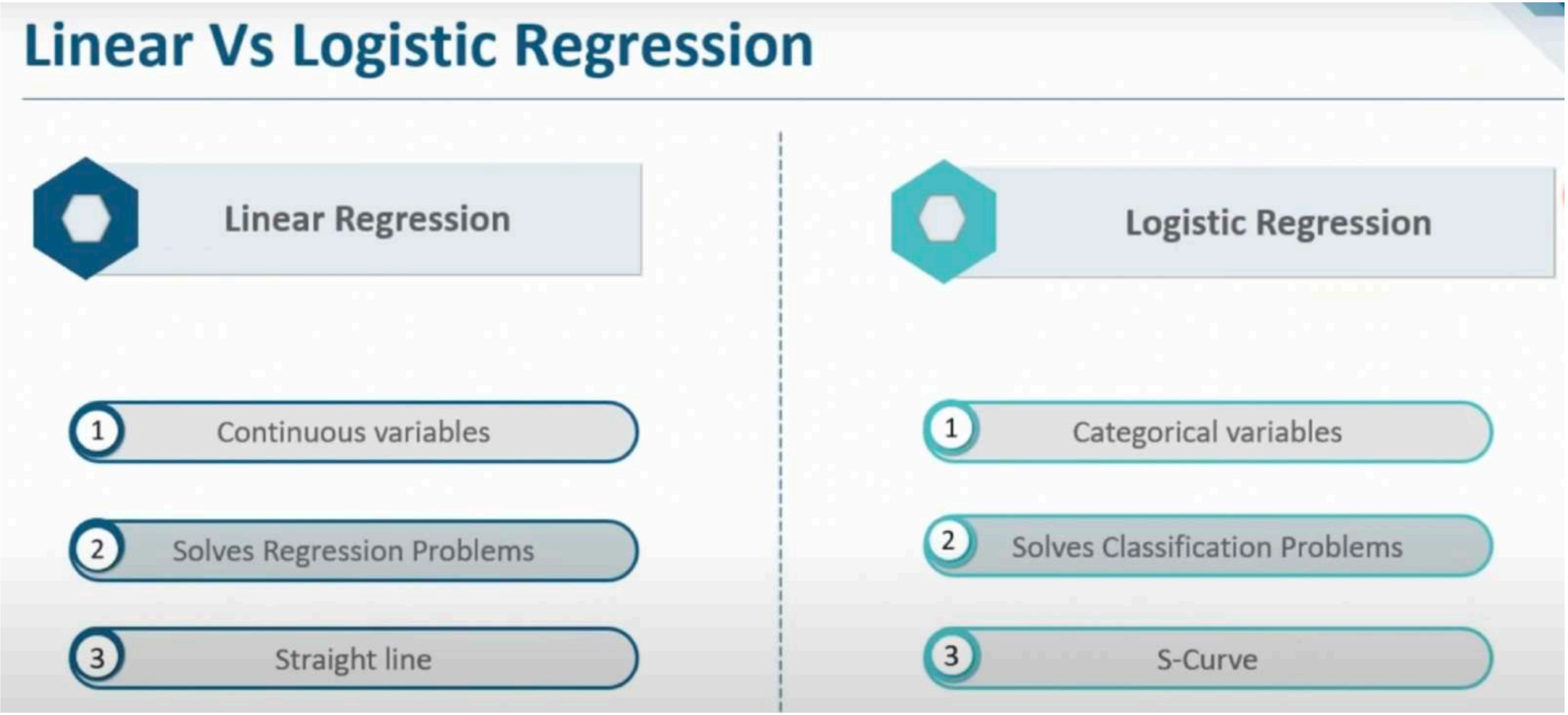
Usually operate with logarithms to simplify

$$\ell(\beta_0, \beta) = \sum_{i=1}^{n} \left(y_i \log p(x_i) + (1 - y_i) \log (1 - p(x_i)) \right)$$

$$= \sum_{i=1}^{n} \log (1 - p(x_i)) + \sum_{i=1}^{n} y_i \log \frac{p(x_i)}{1 - p(x_i)}$$

$$= \sum_{i=1}^{n} \log (1 - p(x_i)) + \sum_{i=1}^{n} y_i (\beta_0 + x_i \cdot \beta)$$

$$= \sum_{i=1}^{n} - \log \left(1 + e^{\beta_0 + x_i \cdot \beta}\right) + \sum_{i=1}^{n} y_i (\beta_0 + x_i \cdot \beta)$$


Linear Vs Logistic Regression
Linear Regression
1 Continuous variables
2 Solves Regression Problems
3 Straight line
Logistic Regression
1 Categorical variables
2 Solves Classification Problems
3 S-Curve

Logistic Regression Applications

Weather Prediction

Helps determine the kind of weather that can be expected

Logistic Regression Applications

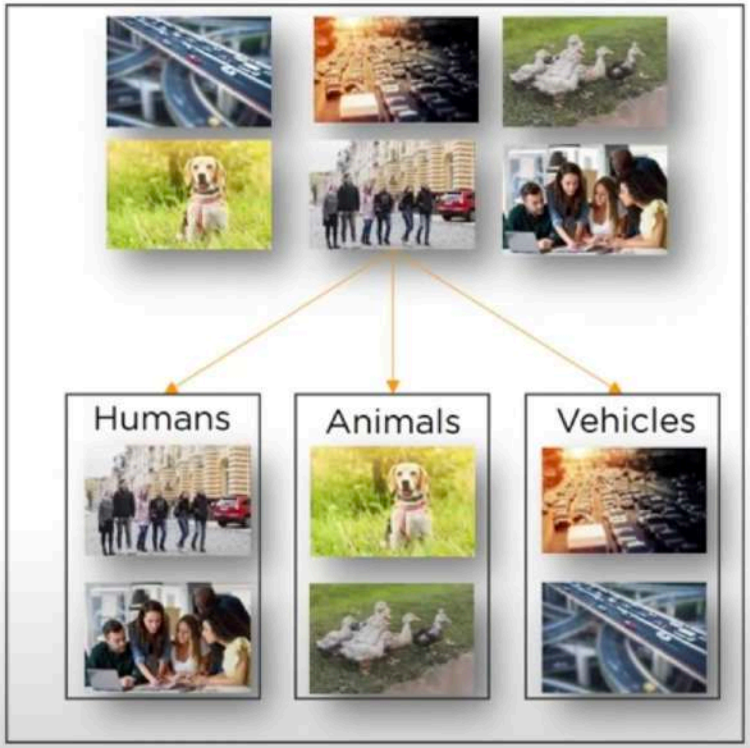
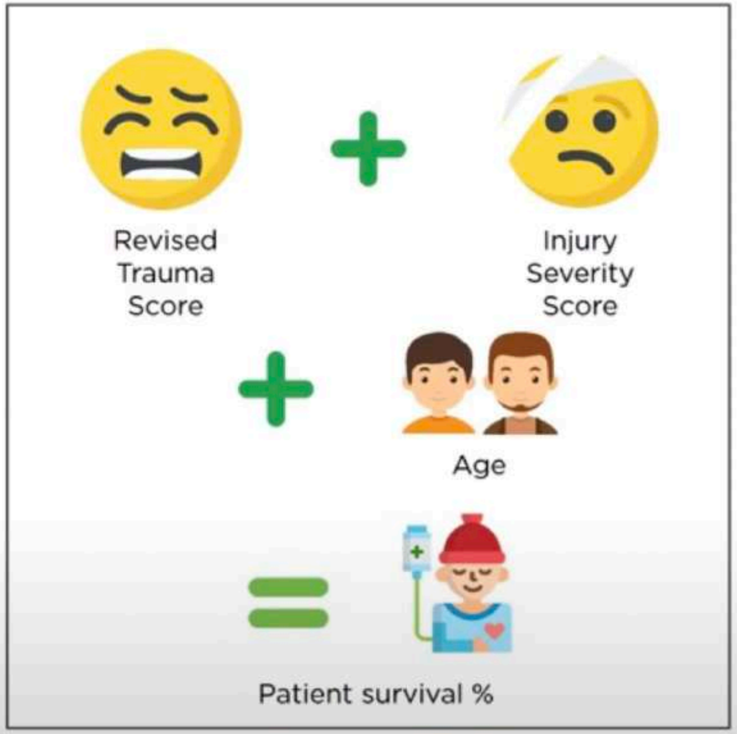


Image Categorization

Identifies the different components that are present in the image, and helps categorize them

Logistic Regression Applications



Healthcare (TRISS)

Determines the possibility of patient survival, taking age, ISS and RTS into consideration