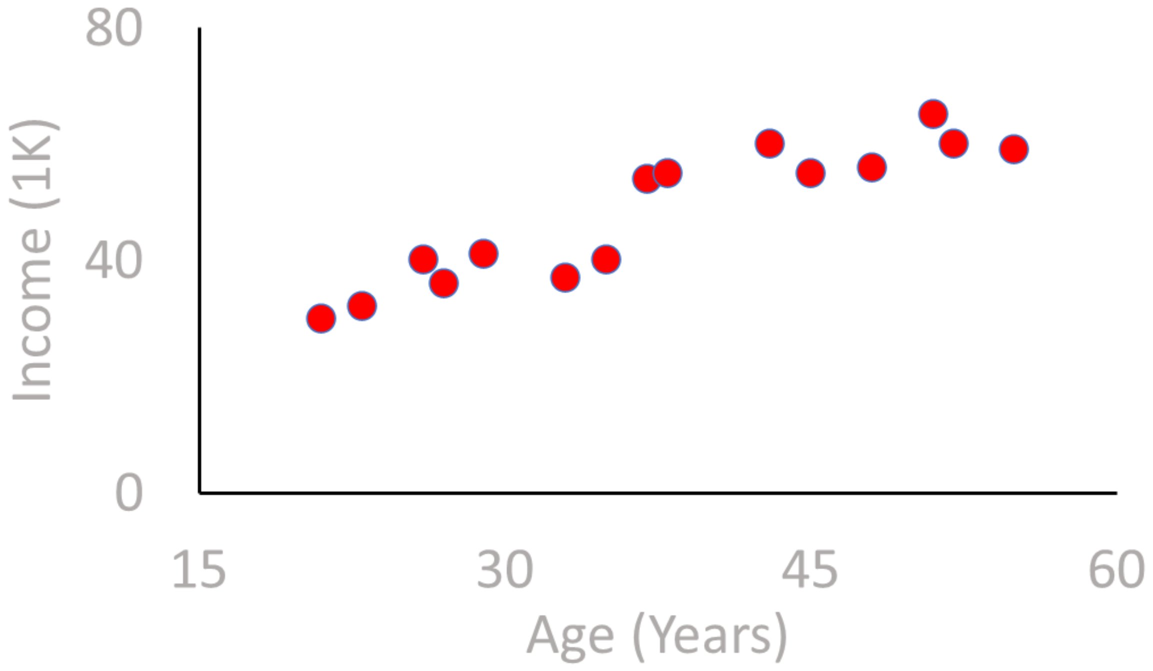
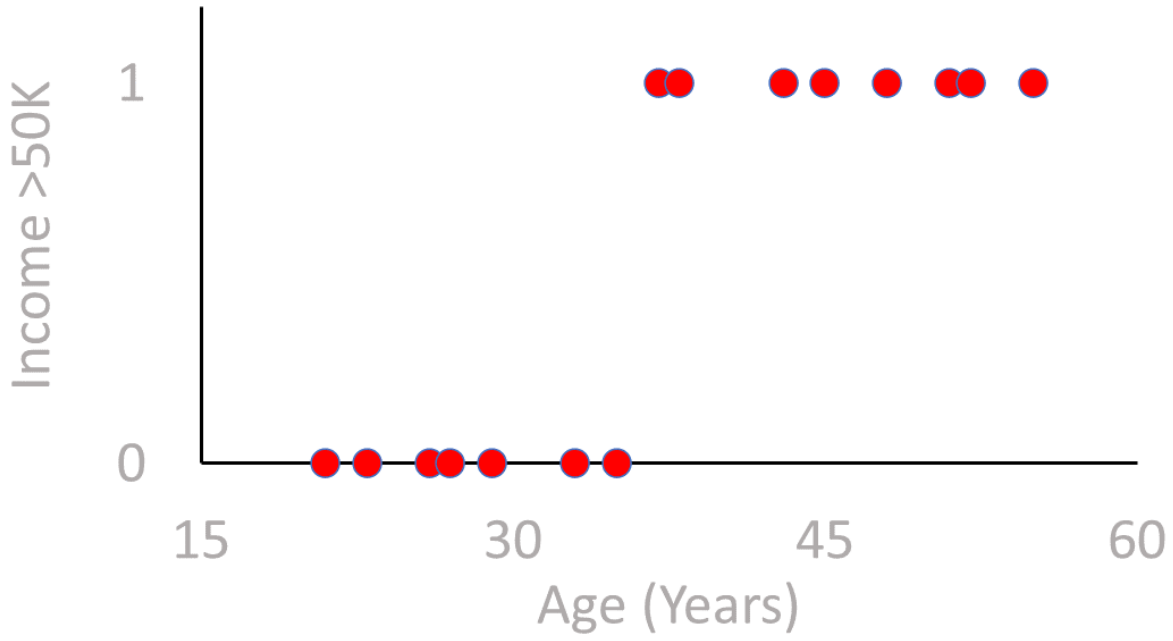
Logistic Regression

Supervised Learning

Linear vs. Logistic Regression





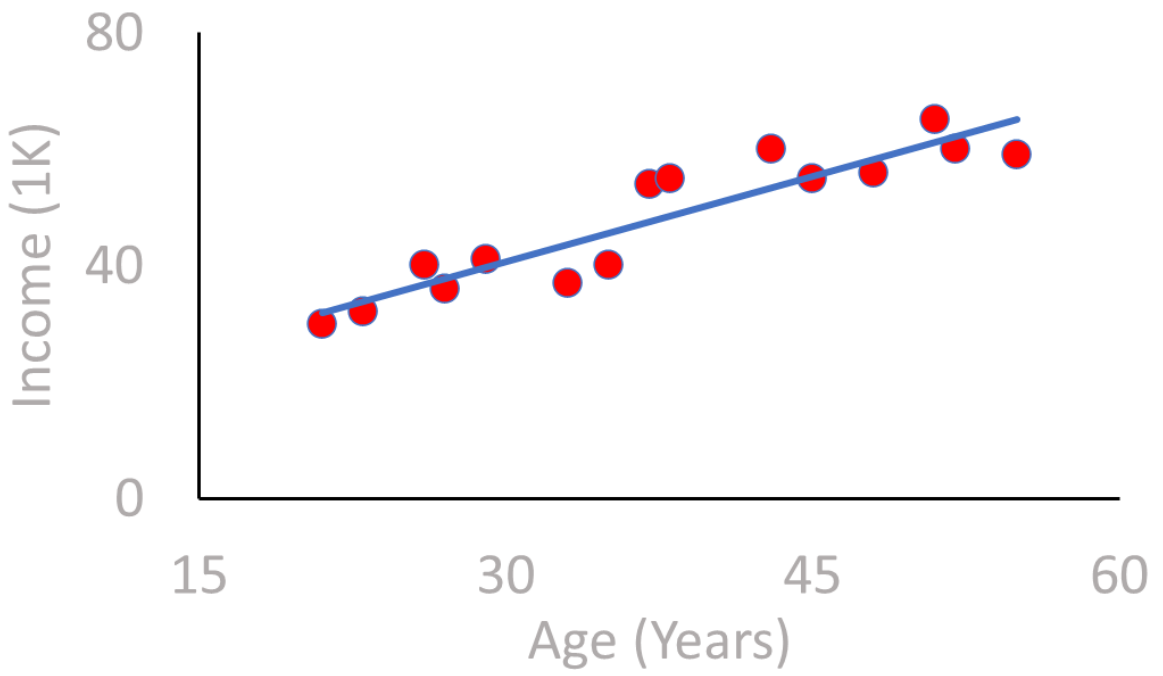
Linear Regression

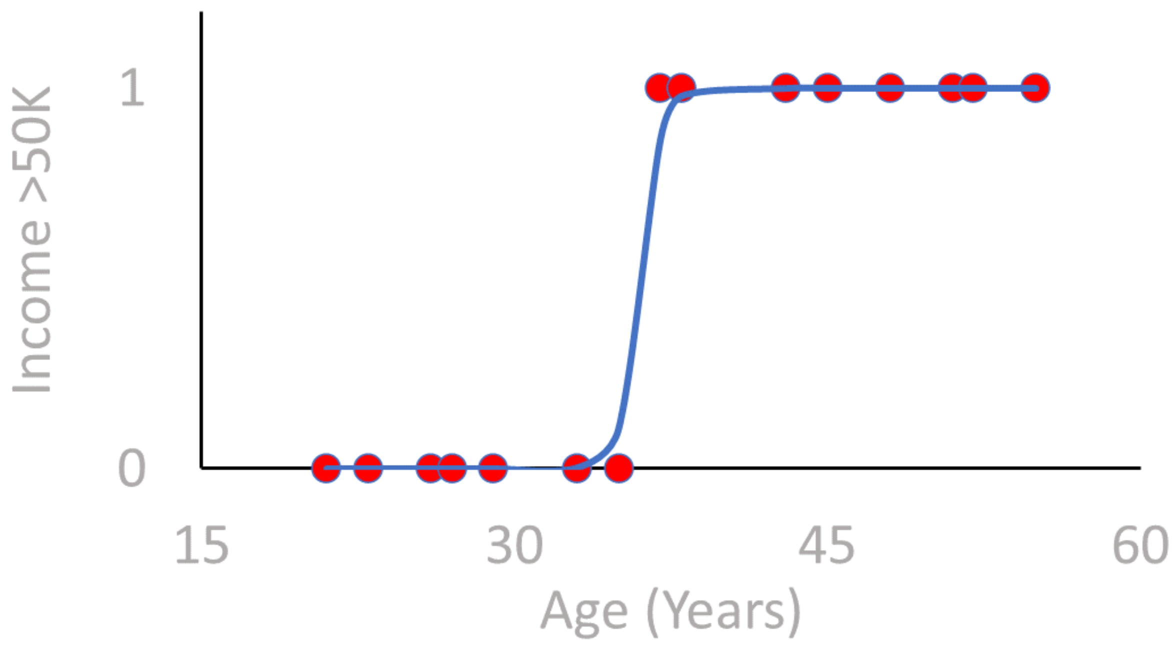
- Target variable is continuous.

Logistic Regression

- Target variable is categorical (binary).

Linear vs. Logistic Regression





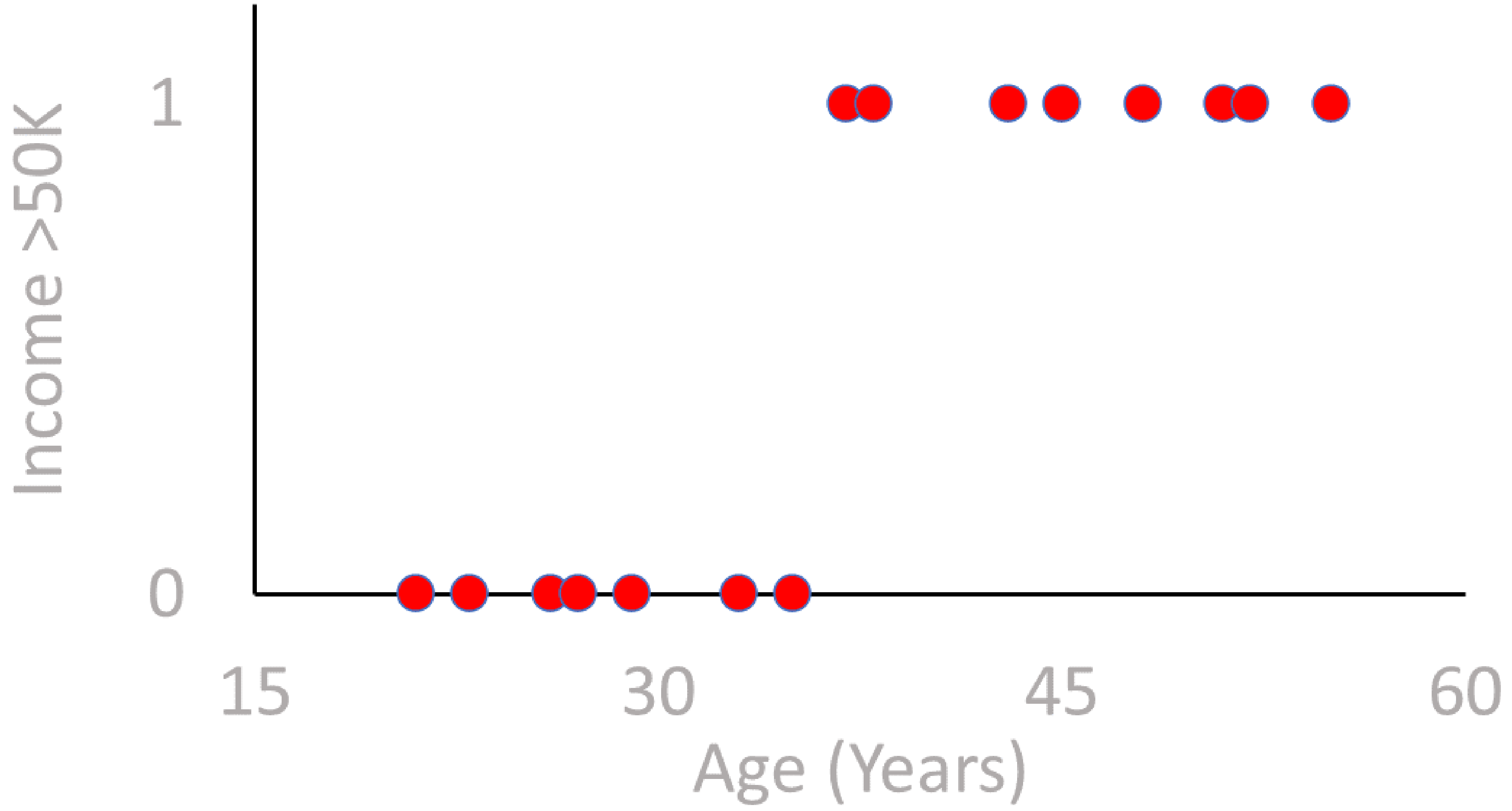
Linear Regression

- Target variable is continuous.
- Straight linear line to fit the data.
- Ordinary least square (OLS) to find the best fit line.

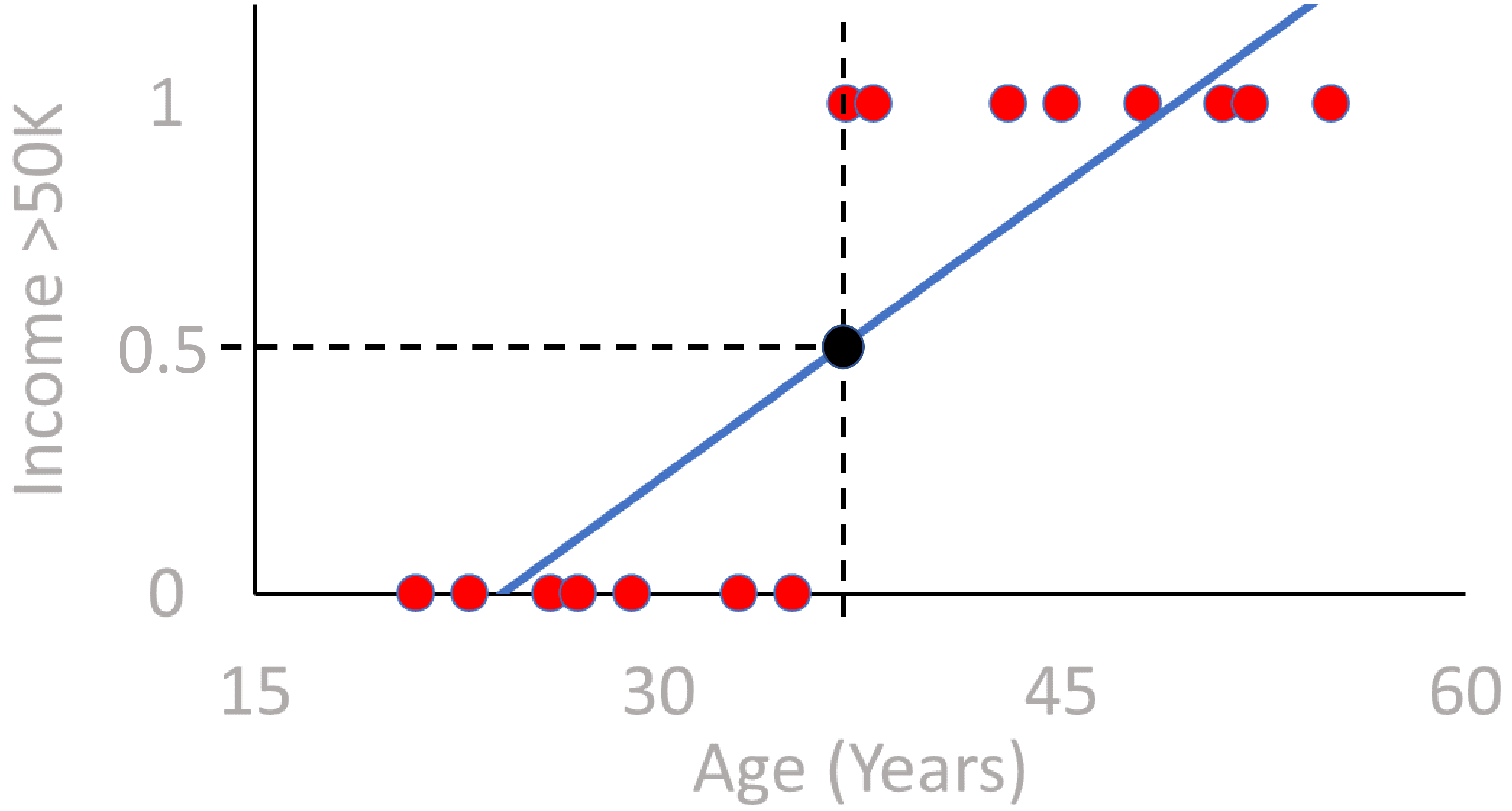
Logistic Regression

- Target variable is categorical (binary).
- S-shaped curve (sigmoid function) to fit the data.
- Maximum likelihood estimation (MLE) to find the best fit curve.

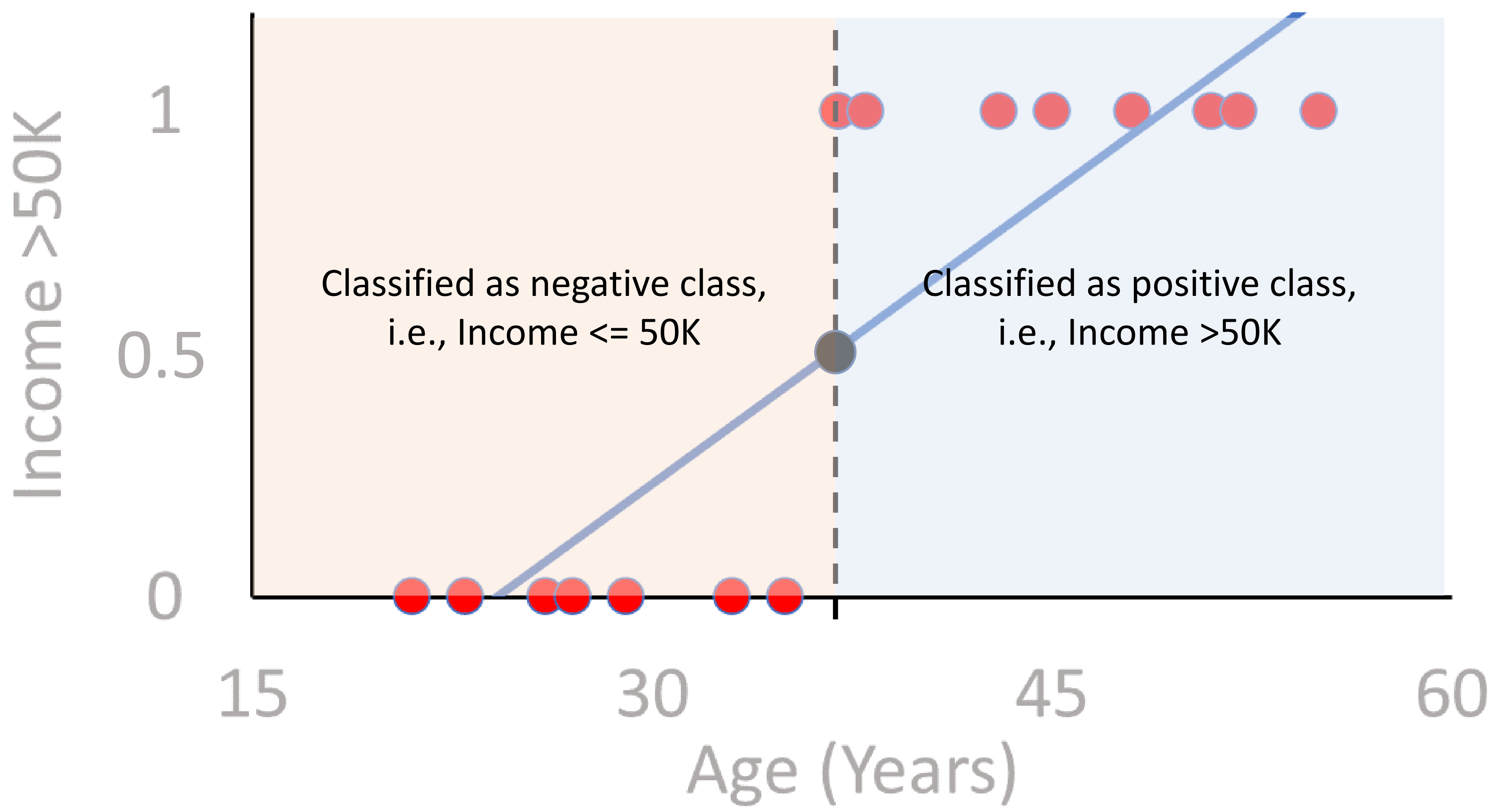
Logistic Regression



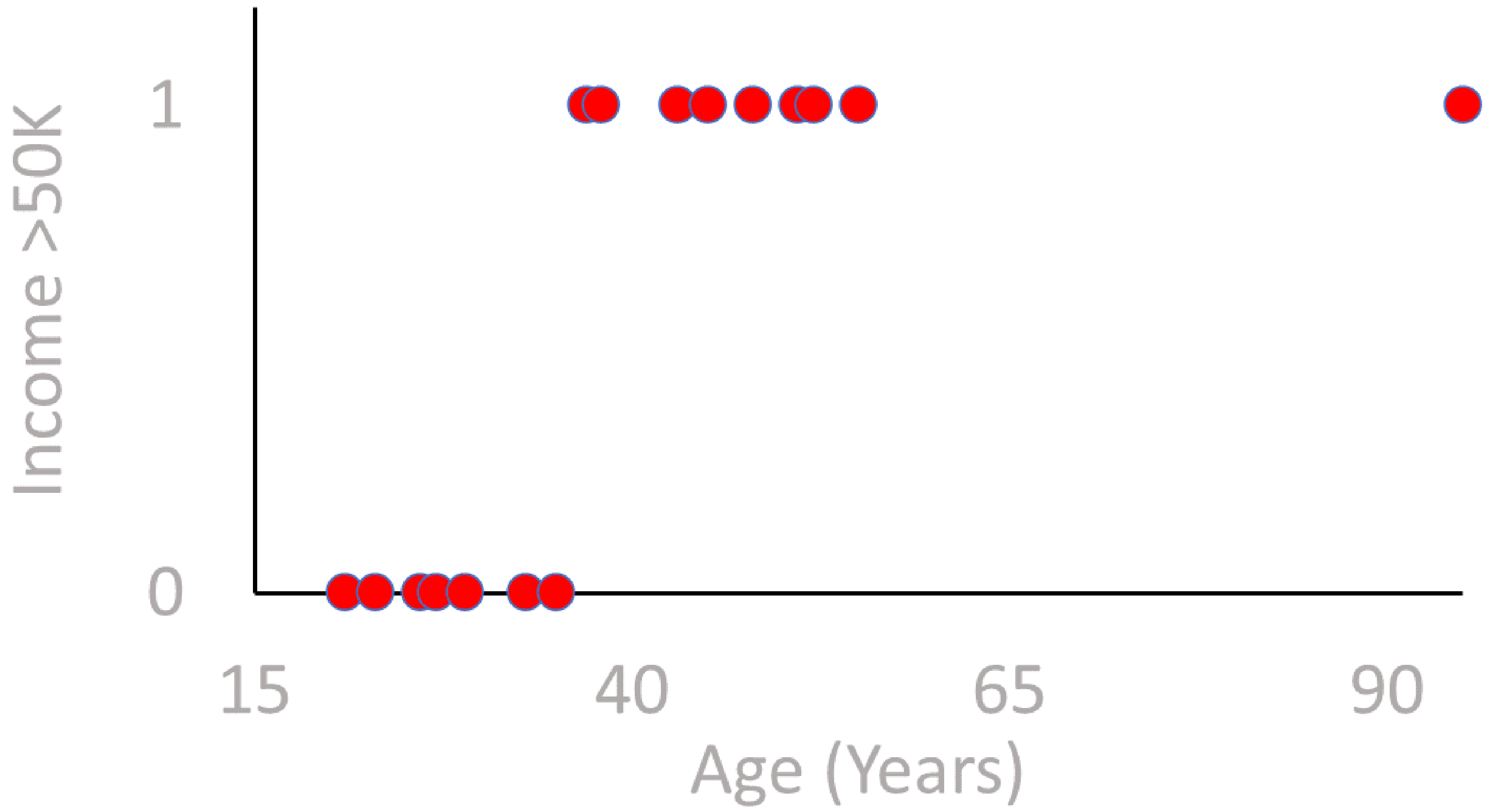
Logistic Regression



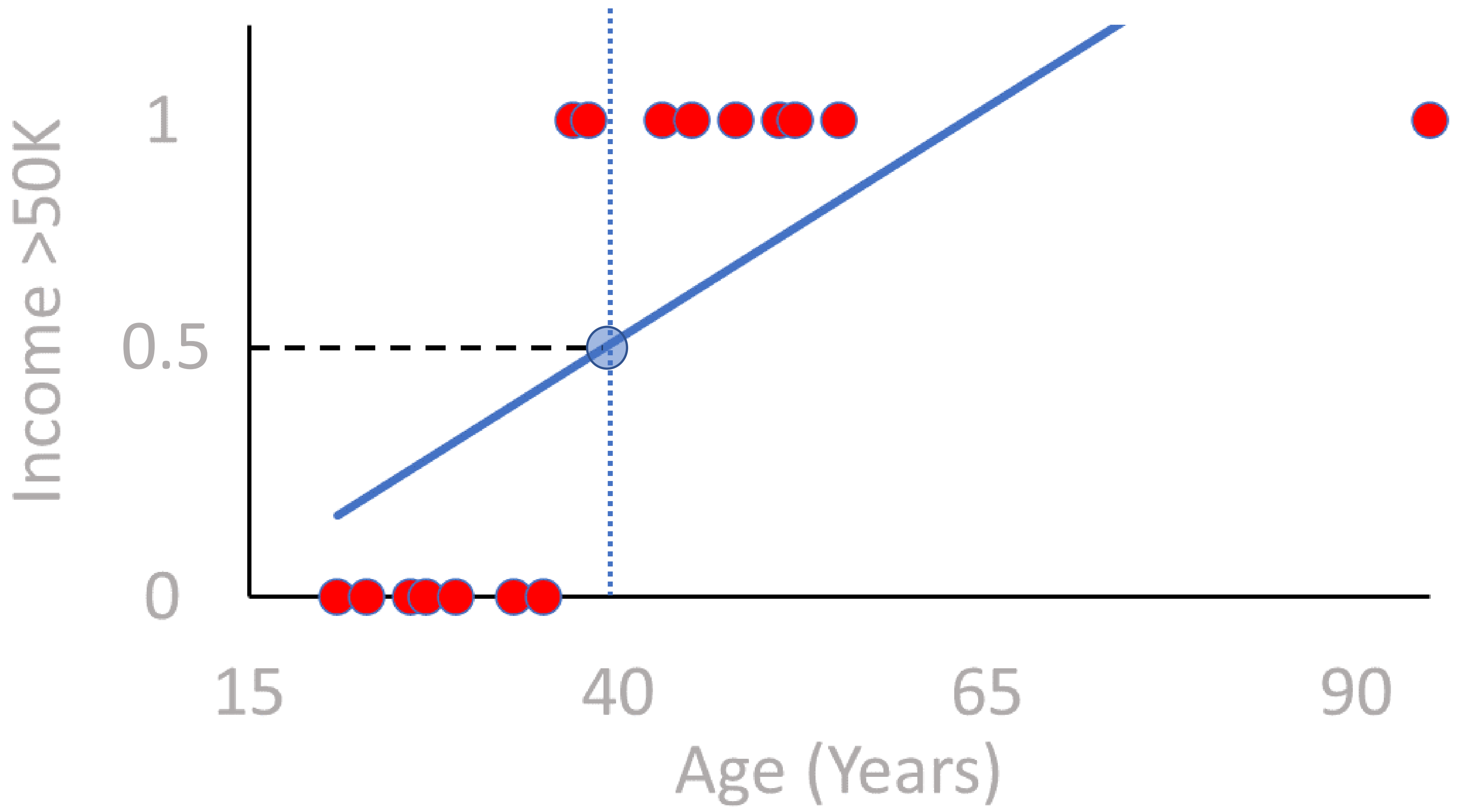
Logistic Regression



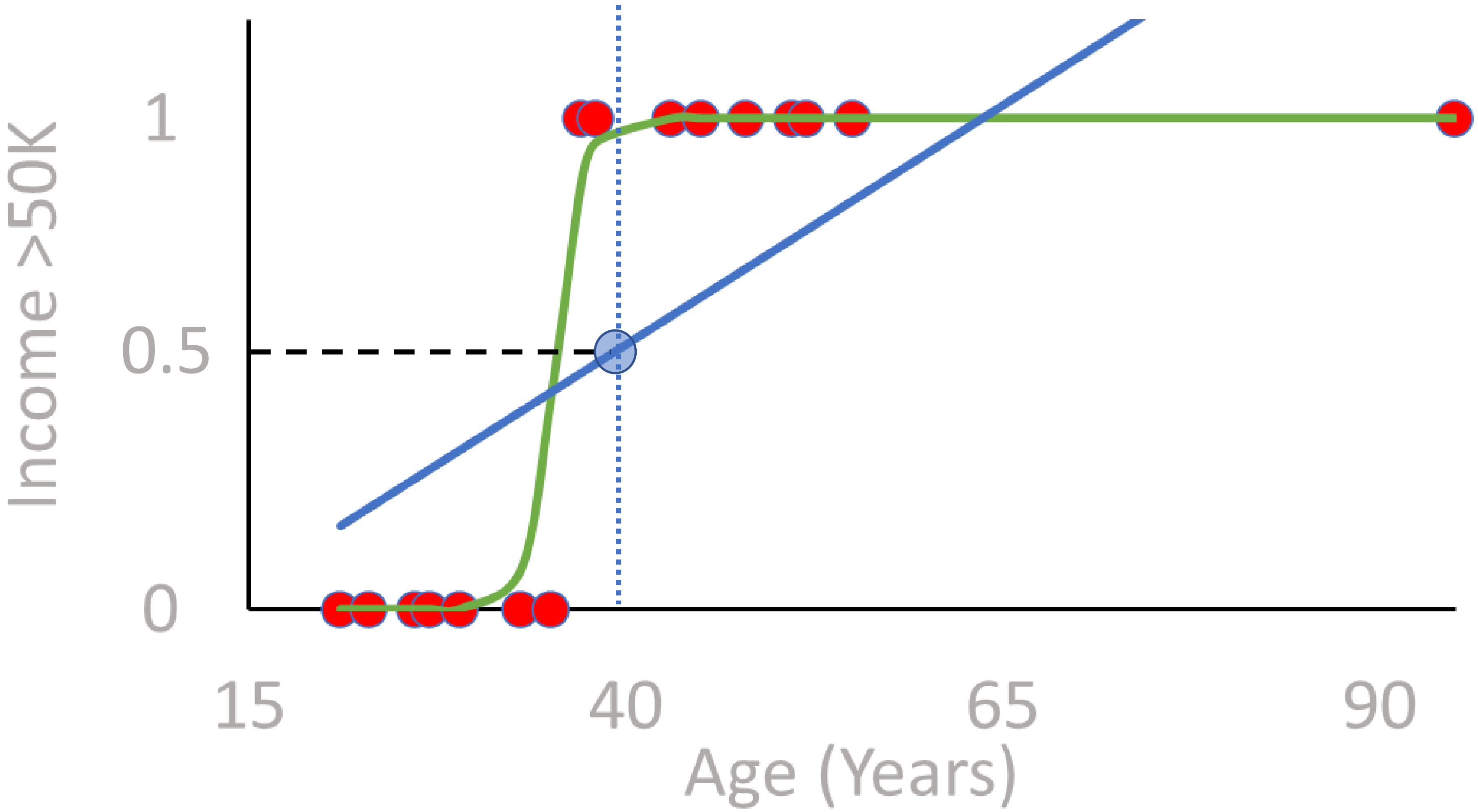
Logistic Regression



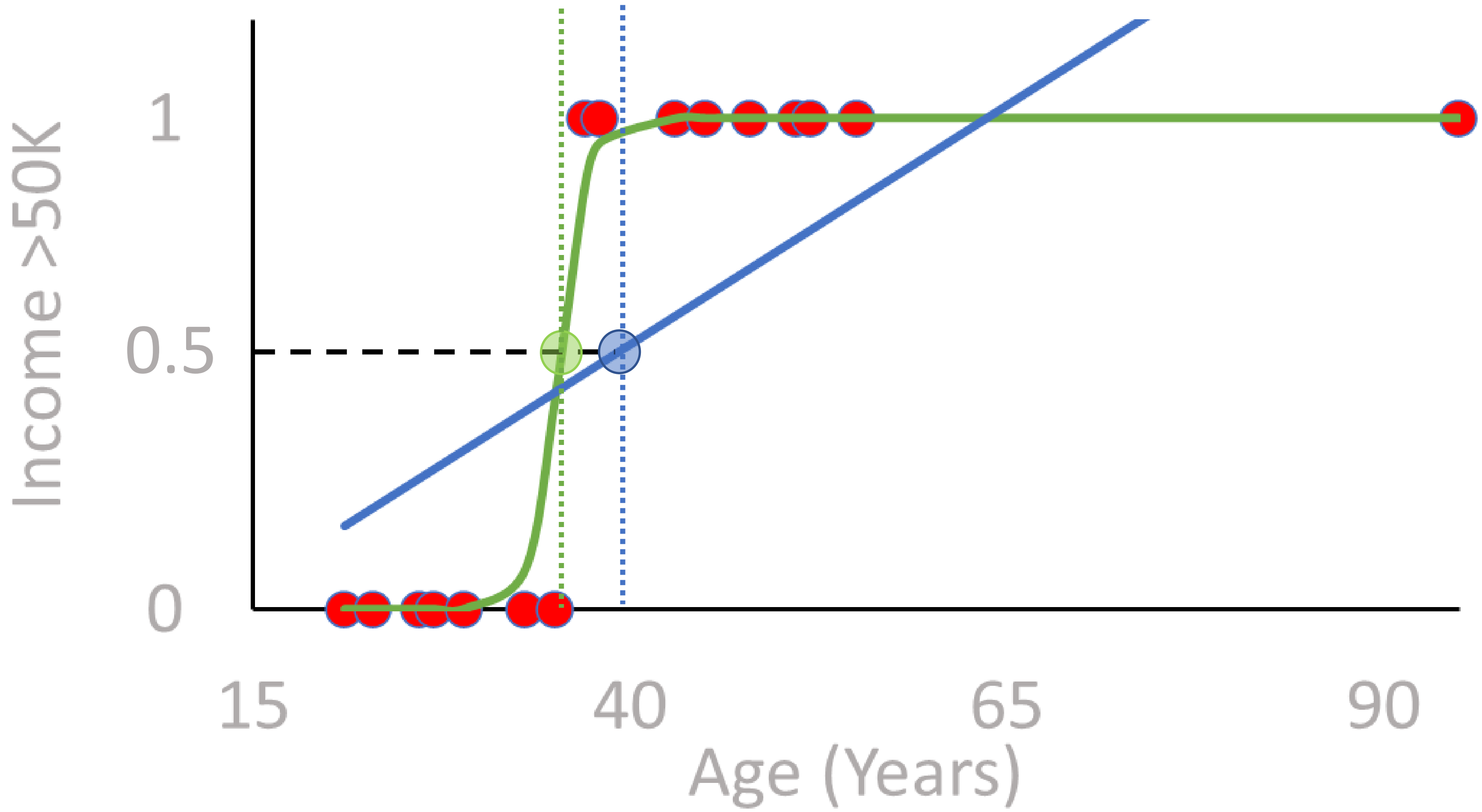
Logistic Regression



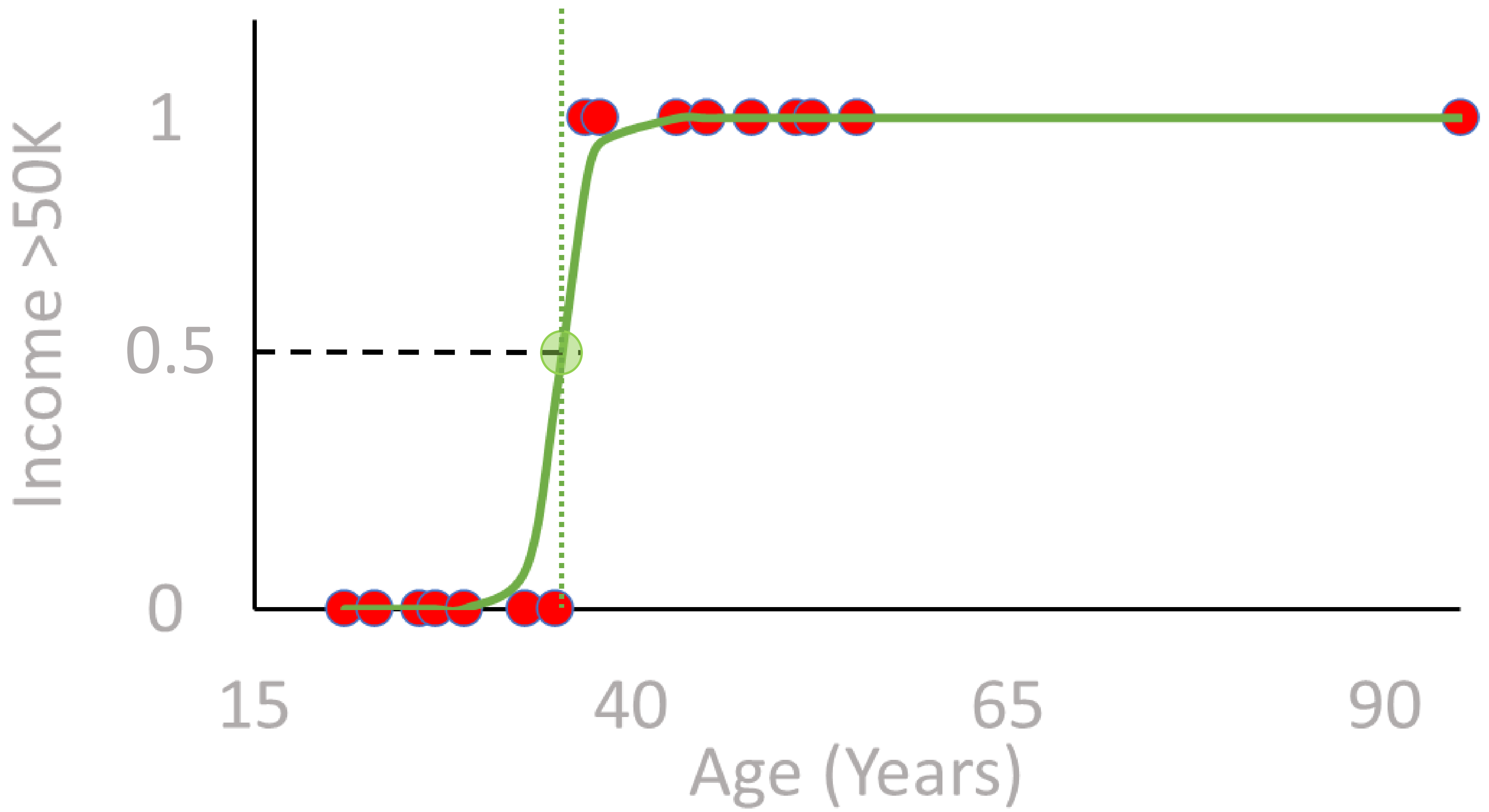
Logistic Regression



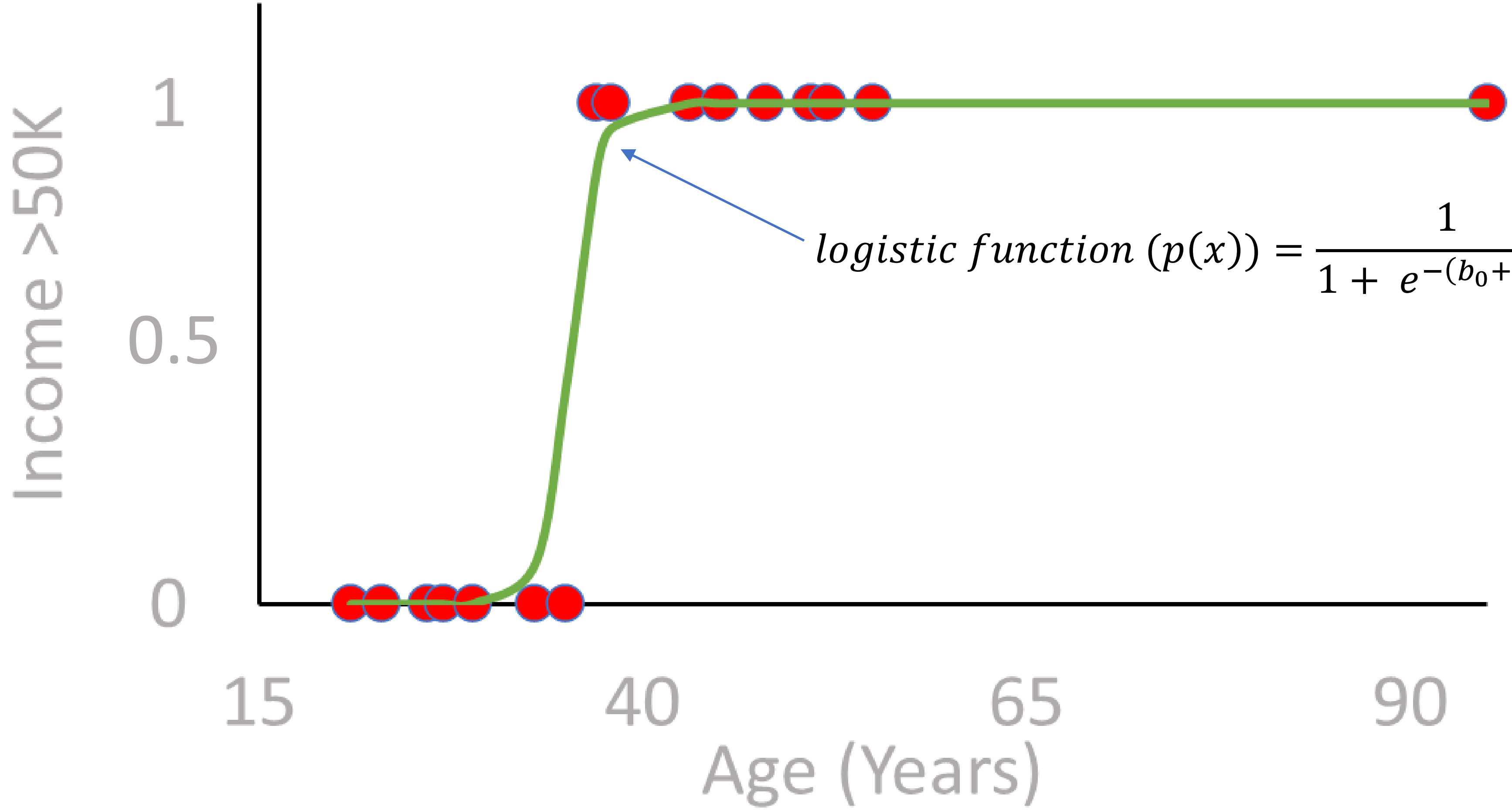
Logistic Regression



Logistic Regression

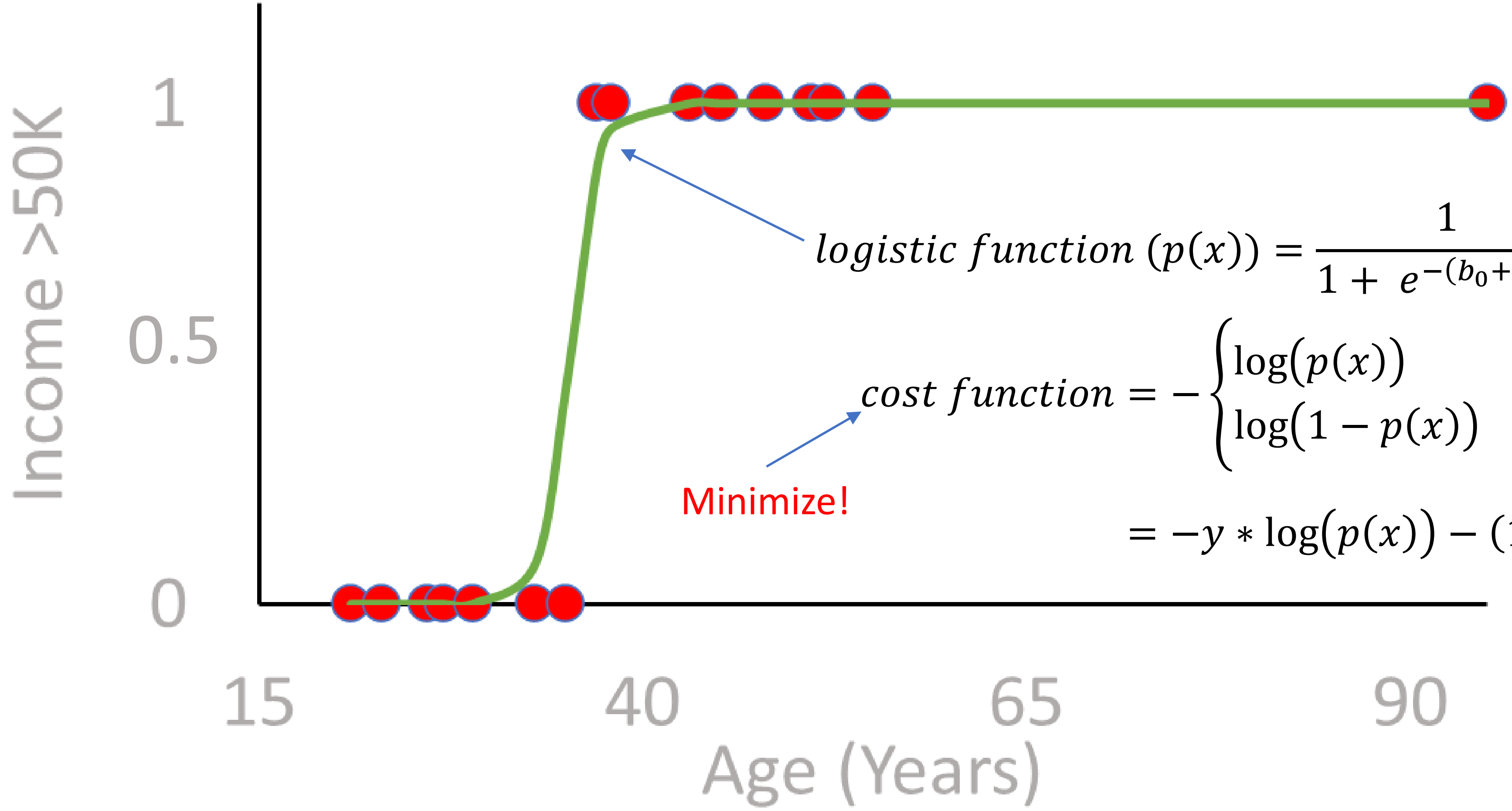


Logistic Regression



$$logistic\ function\ (p(x)) = \frac{1}{1 + e^{-(b_0+b_1x)}} = \frac{e^{(b_0+b_1x)}}{e^{(b_0+b_1x)} + 1}$$

Logistic Regression



$$logistic\ function\ (p(x)) = \frac{1}{1+\ e^{-(b_0+b_1x)}} = \frac{e^{(b_0+b_1x)}}{e^{(b_0+b_1x)}+1}$$

$$cost\ function = -\begin{cases} \log\big(p(x)\big) & ,when\ y = 1 \\ \log\big(1-p(x)\big) & ,when\ y = 0 \end{cases}$$

Minimize!

$$= -y * \log\big(p(x)\big) - (1-y) * \log(1-p(x))$$

Logistic Regression

Generalization: When there are multiple features (columns) in the data.

$$logistic\ function\ (p(X)) \ = \frac{1}{1+\ e^{-(b_0+b_1x_1+b_1x_1+\ \dots\ +\ b_nx_n)}}$$

$$= \frac{e^{(b_0+b_1x_1+b_2x_2+\ \dots\ +\ b_nx_n)}}{e^{(b_0+b_1x_1+b_2x_2+\ \dots\ +\ b_nx_n)}+1}$$

Logistic Regression

- When the outputs are binary, we usually use odds to describe their chances of occurrence.
- It is defined as the probability of the event occurring divided by the probability of the event not occurring.

$$Odds = \frac{Probability\ Event\ Occurs\ (p)}{Probability\ Event\ Does\ Not\ Occur\ (1-p)} = \frac{p}{1-p}$$

- Given the odds of an event, the probability of the event occurring can be computed by:

$$p = \frac{Odds}{1 + Odds}$$

- Comparing the above equation to the equation in the previous page, we get

$$Odds = e^{(b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n)}$$

$$\log(Odds) = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \qquad \text{(taking log on both sides)}$$

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

Logistic Regression

The Logit -> log(Odds)

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1 x_1 + b_2 x_2 + \ \ldots + b_n x_n$$

Interpreting the coefficients:

- The logit is a linear function of features $x_1, x_2, \ldots, x_n$
- Coefficients $b_0, b_1, b_2, \ldots, b_n$ are log of odds ratios.
- Taking their antilog, i.e., exp(coefficients) gives the odds ratios which can be interpreted easily.
- E.g. $\boxed{\log(Odds\ of\ > 50K\ Income) = 1.2 + 0.5 * CollegeDegree + 0.1 * Age}$
- Holding everything else constant, the odds of a person with a college degree having an income >50K is $\exp(0.5) = 1.65$ times more than the person without a college degree.
- Holding everything else constant ,a unit increase in Age will increase the odds of a person having an income >50K by $\exp(0.1) = 1.11$ times.