Deep Learning Unveiled: Theory, Mathematics, and Programming.

Session 1: Fundamentals of Deep Learning and Perceptron

Hi, I’m Amine Nasri

www.linkedin.com/in/amine-nasri19

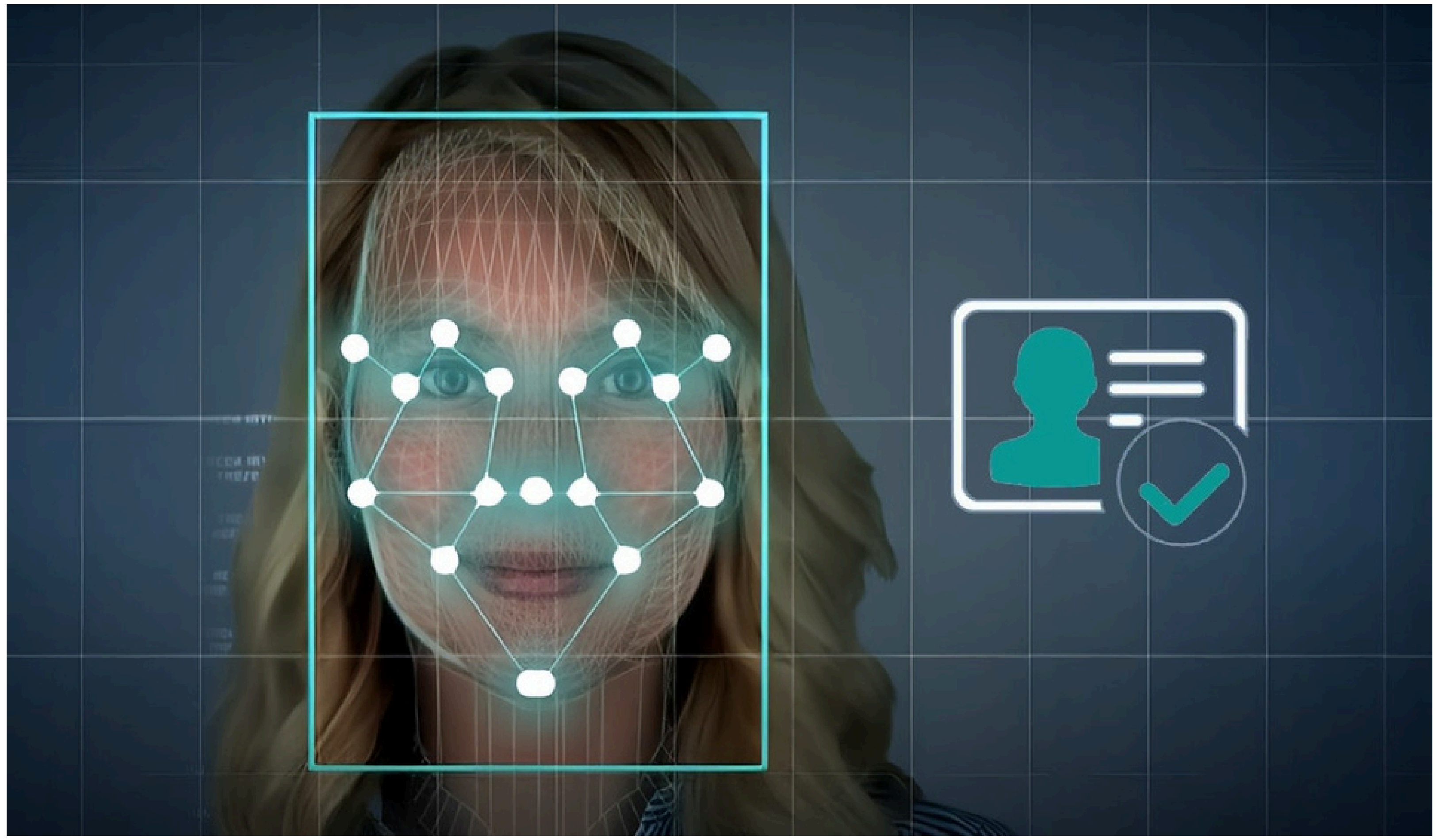
https://github.com/steakmeatdev

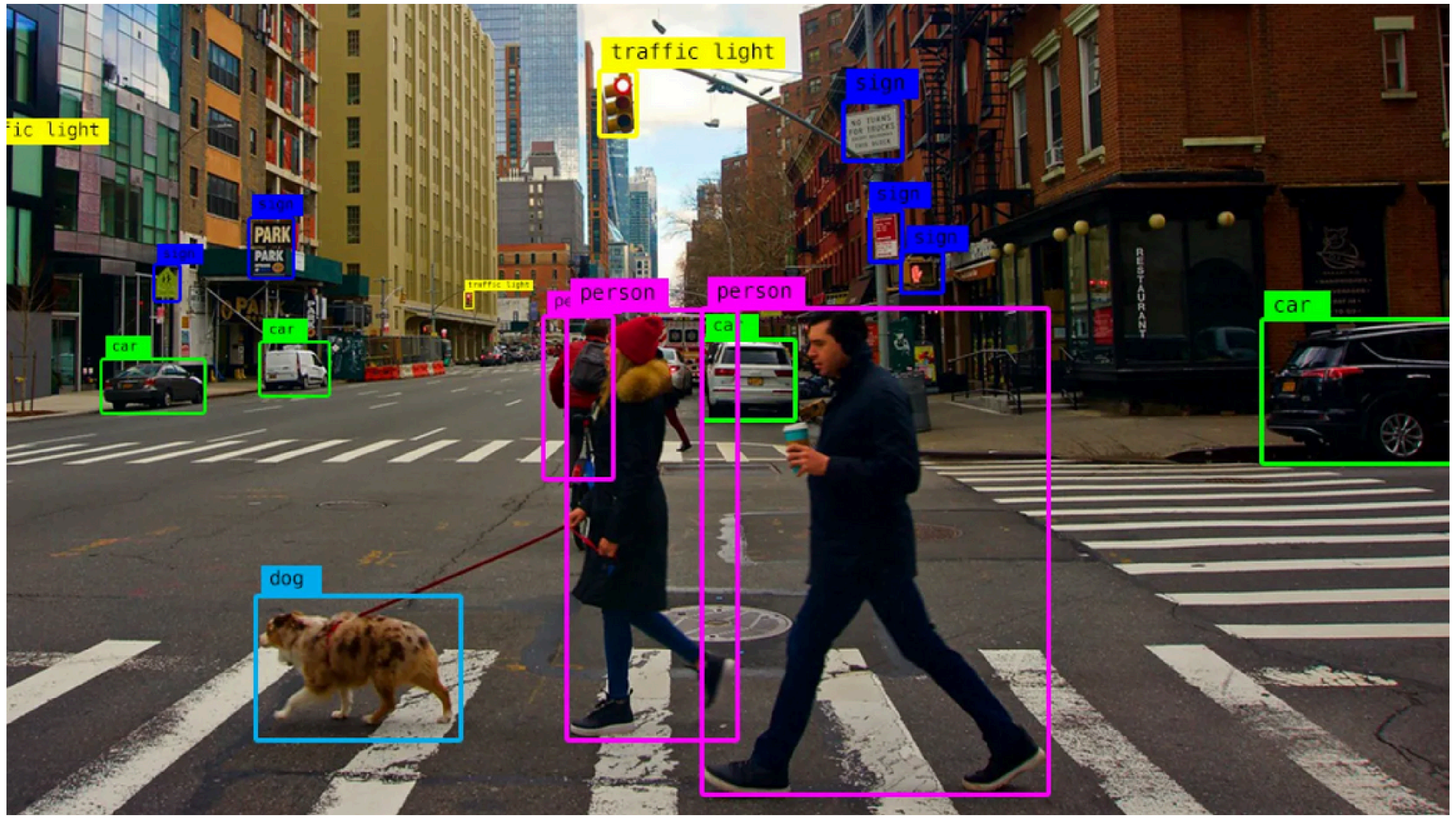
Deep Learning

Deep learning is a method in artificial intelligence (AI) that teaches computers to process data in a way that is inspired by the **human brain.**

Deep learning models can recognize **complex patterns** in pictures, text, sounds, and other data to produce accurate insights and predictions.

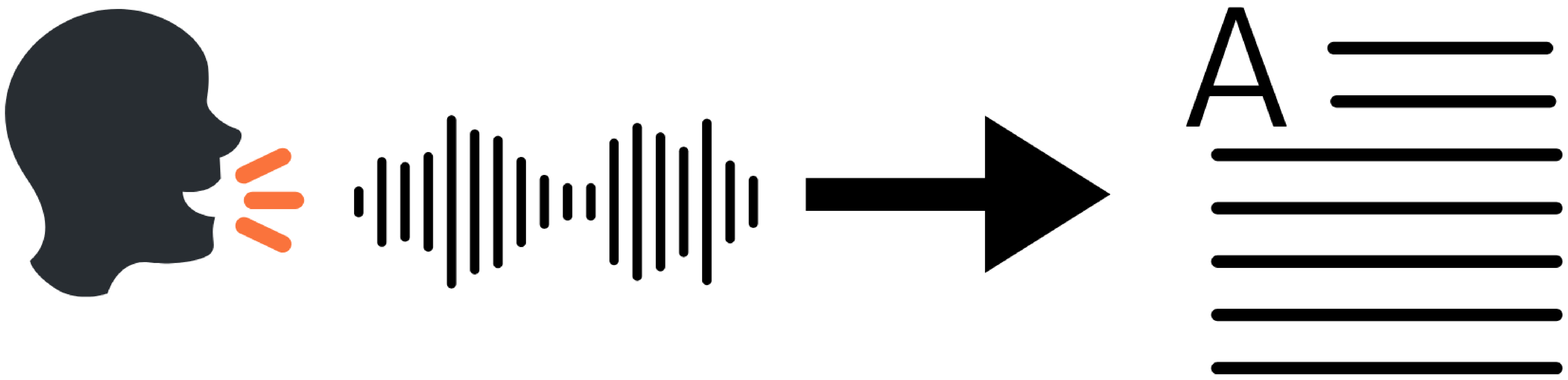
What are the uses of deep learning?





Computer Vision

What are the uses of deep learning?



Speech recognition

What are the uses of deep learning?

ChatGPT

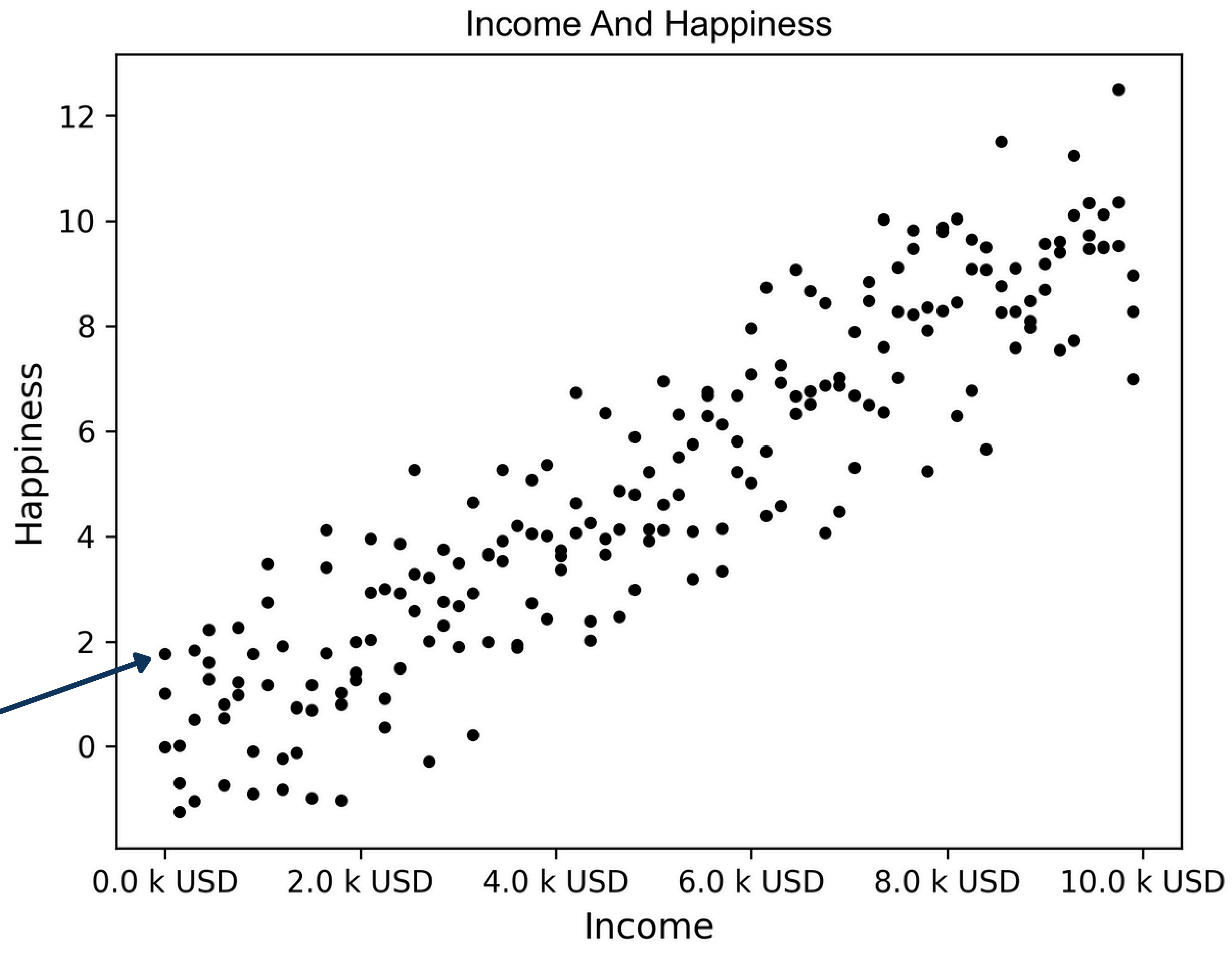
Gemini

Natural Language Processing

We surveyed 198 people about their happiness levels (on a scale from 0 to 10) and their income, and collected the following data.

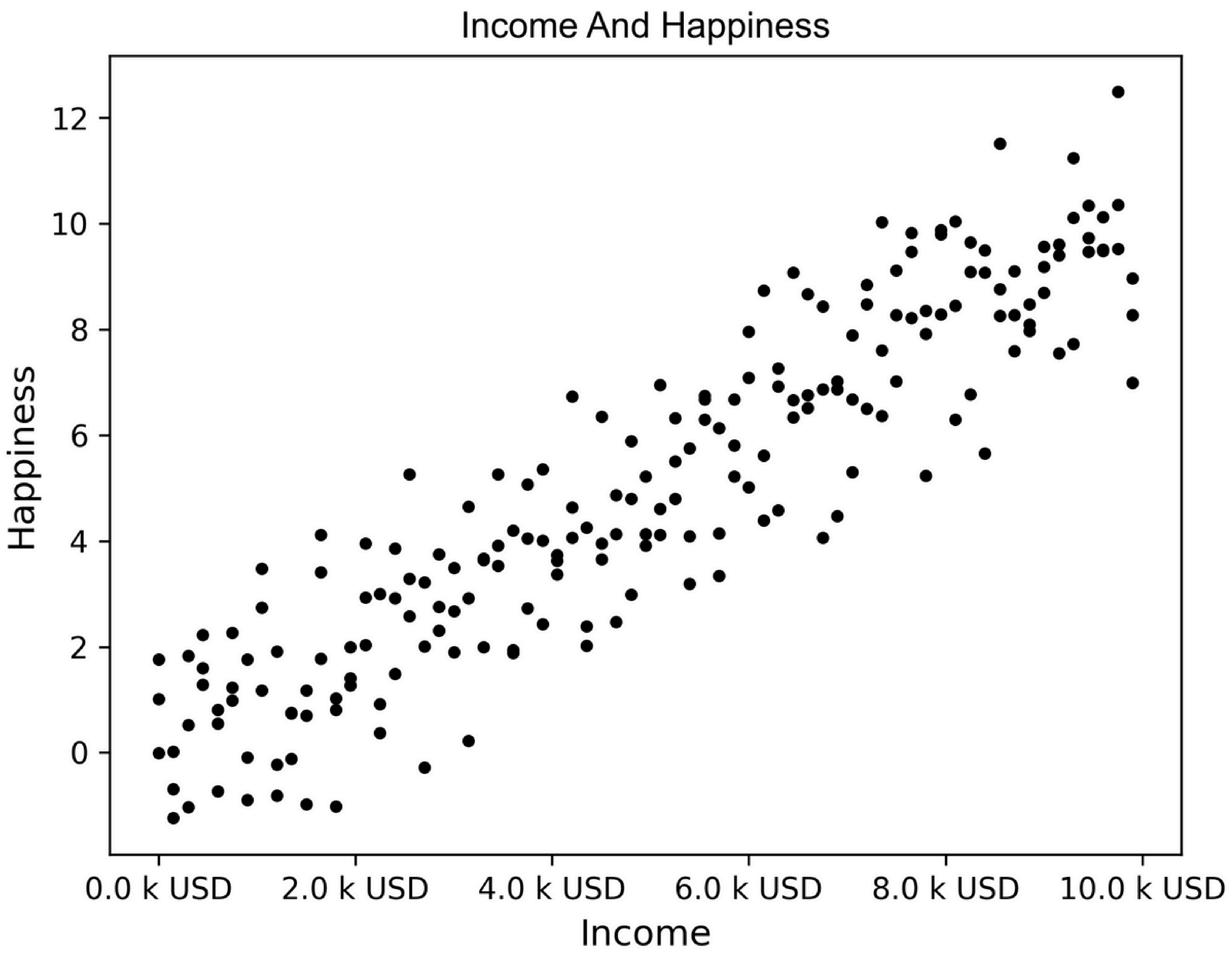
(THIS IS NOT REAL DATA)

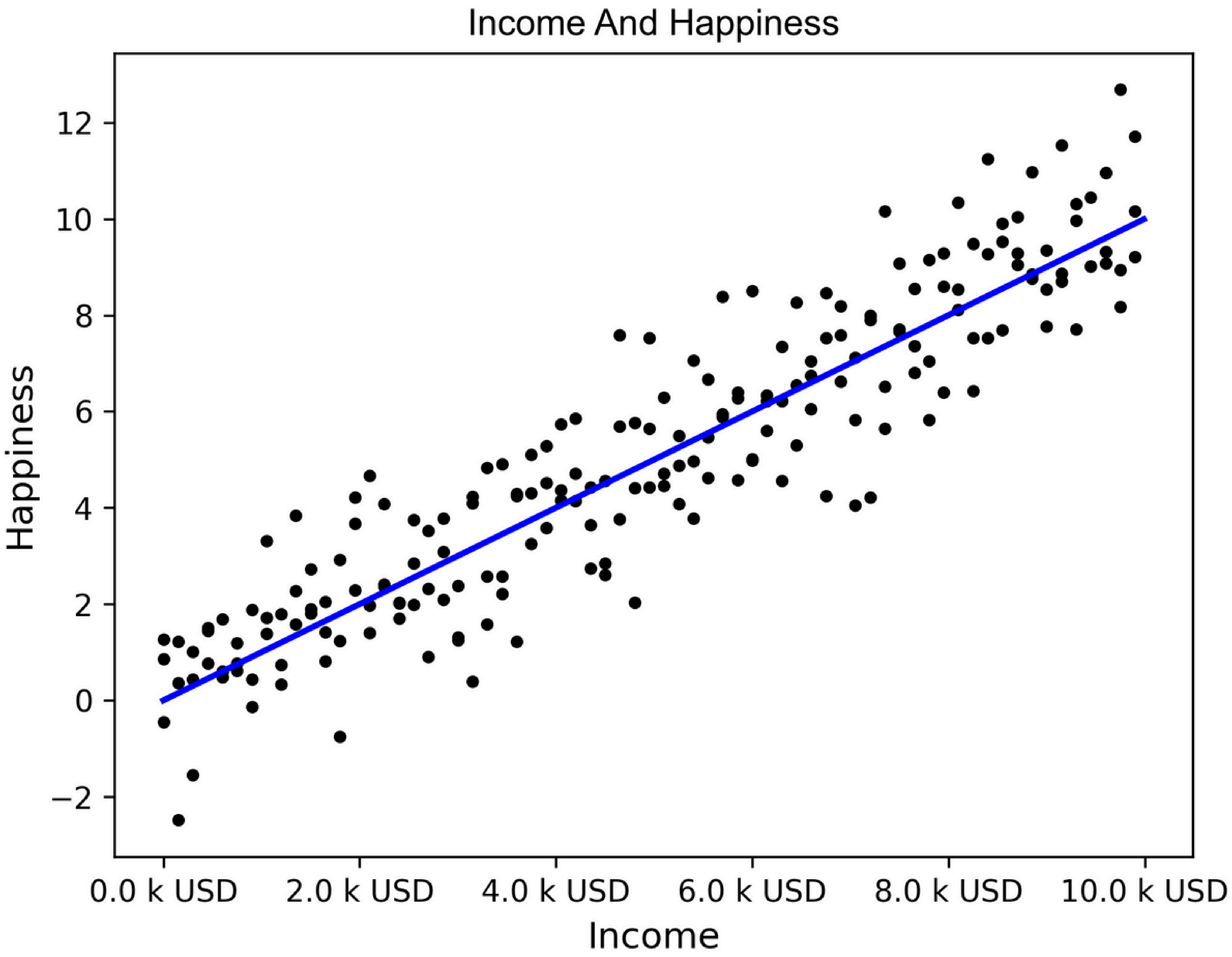
Example: one person



Prediction

What if we wanted to know David's happiness level, given only his income of 7300 USD?

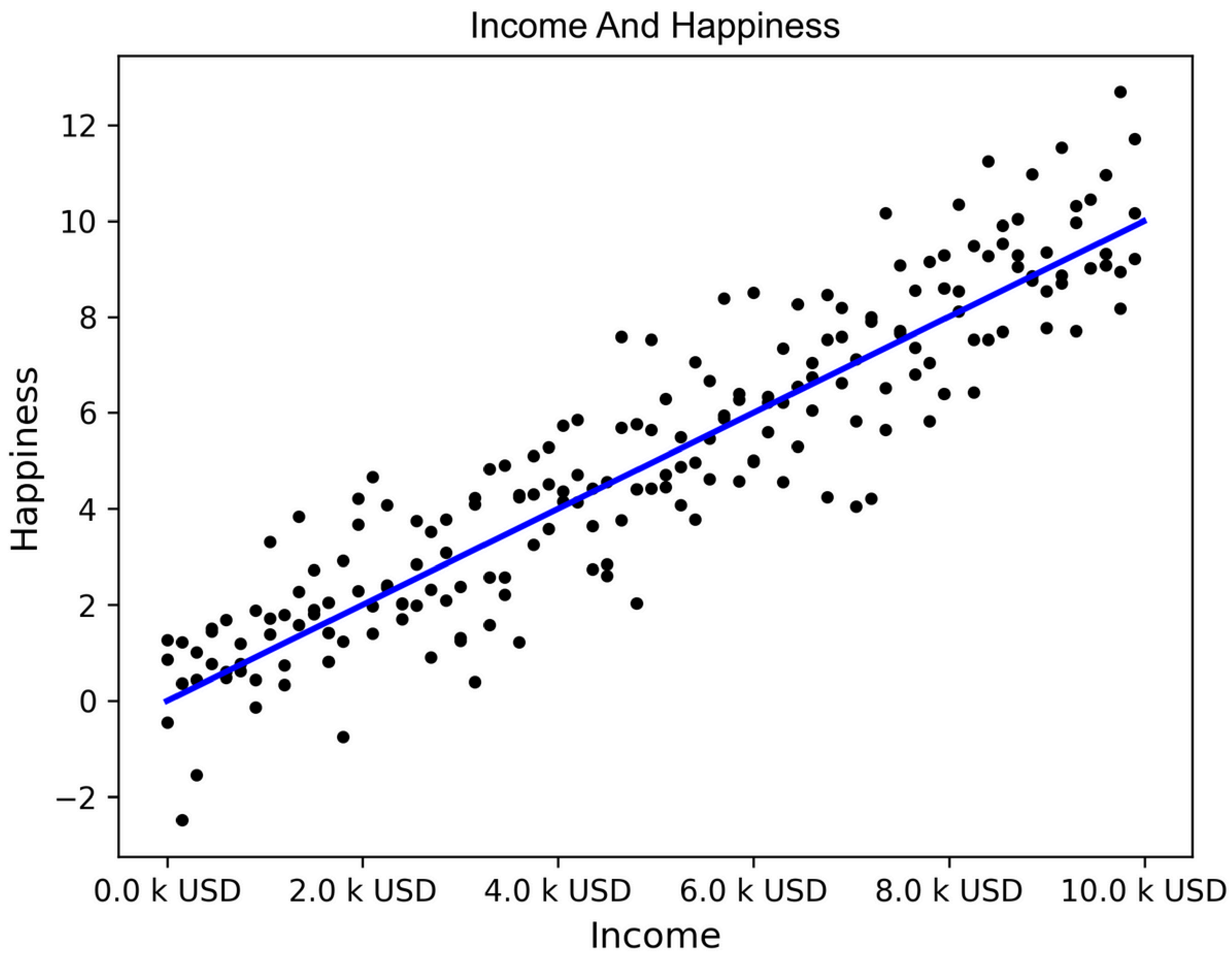

Income And Happiness
Happiness
Income
0
2
4
6
8
10
12
0.0 k USD
2.0 k USD
4.0 k USD
6.0 k USD
8.0 k USD
10.0 k USD



Income And Happiness
12
10
8
6
4
2
0
−2
Happiness
0.0 k USD
2.0 k USD
4.0 k USD
6.0 k USD
8.0 k USD
10.0 k USD
Income

Prediction

$$f(x) = ax + b$$



Prediction

$$f(x) = ax + b$$

$$a = \frac{1}{1000}$$

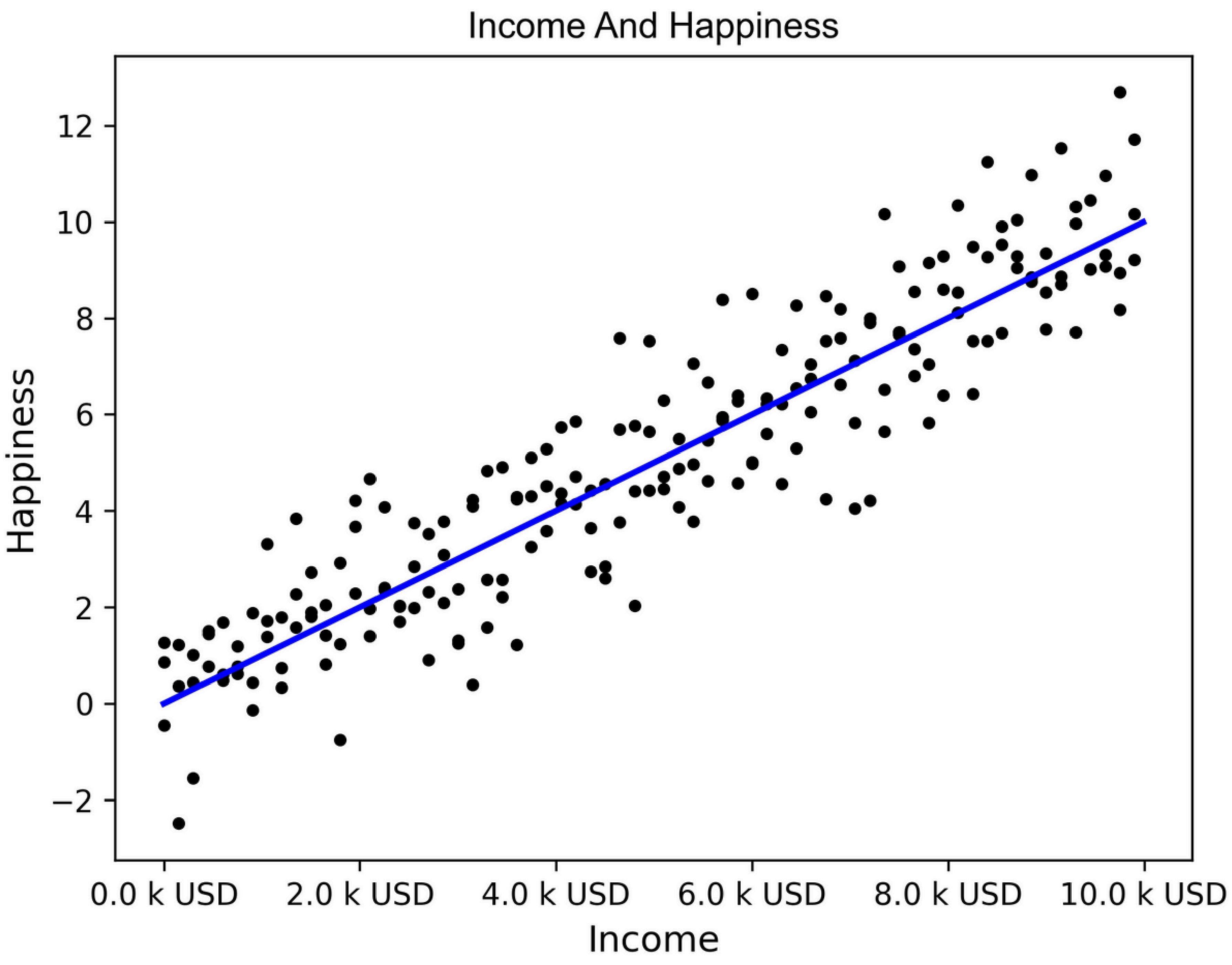
$$b = 0$$

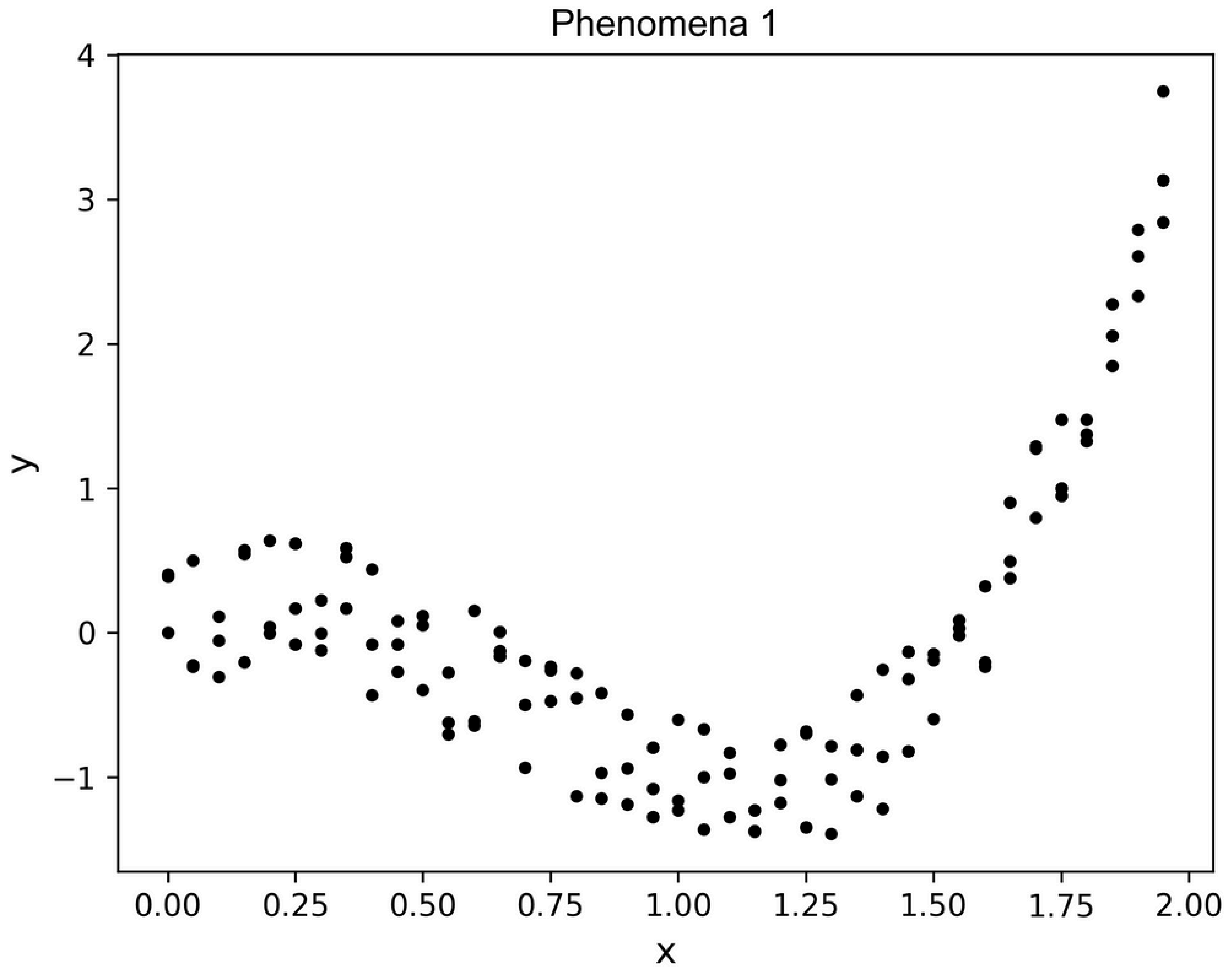
Prediction

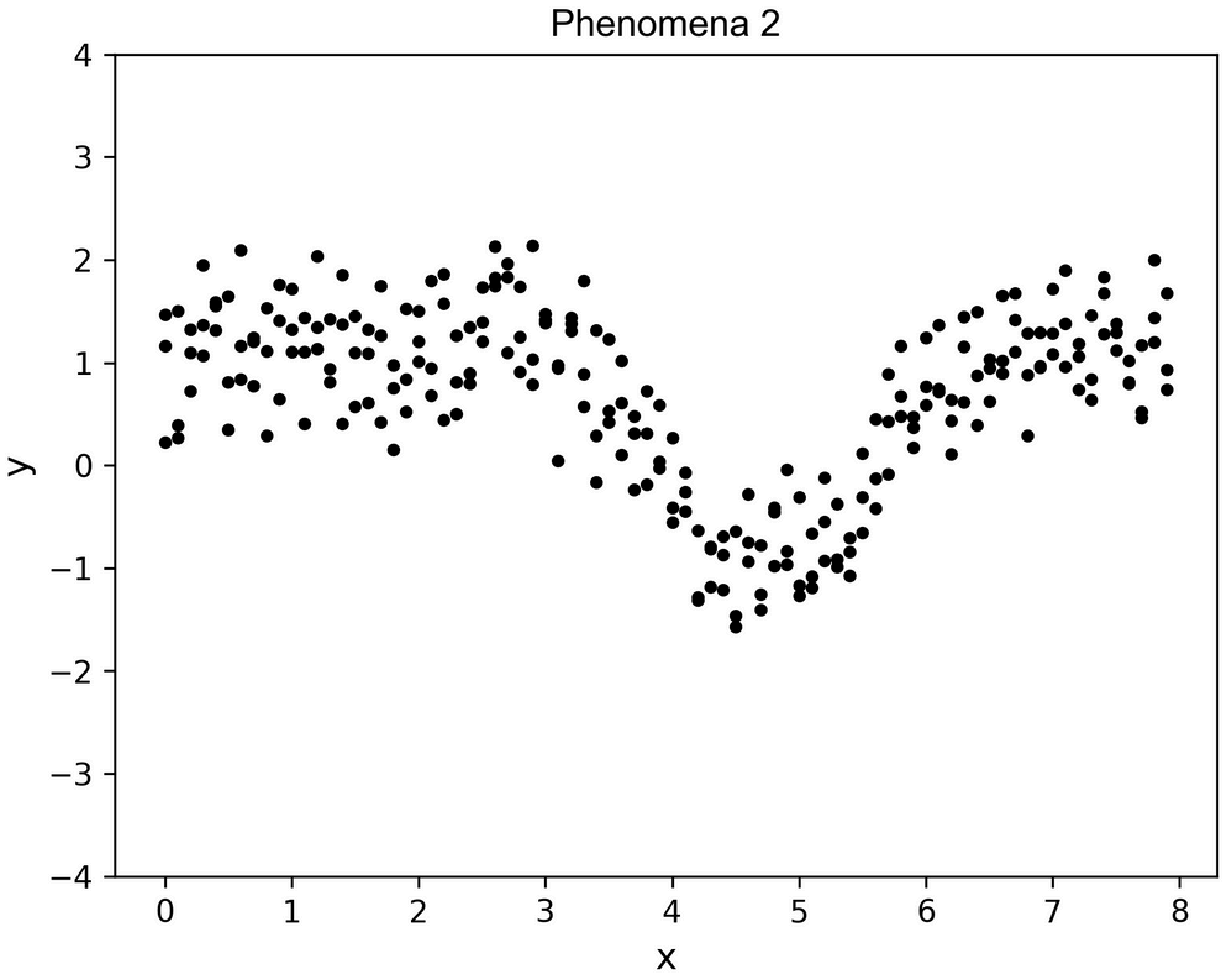
$$f\left(x\right)=\frac{1}{1000}x$$

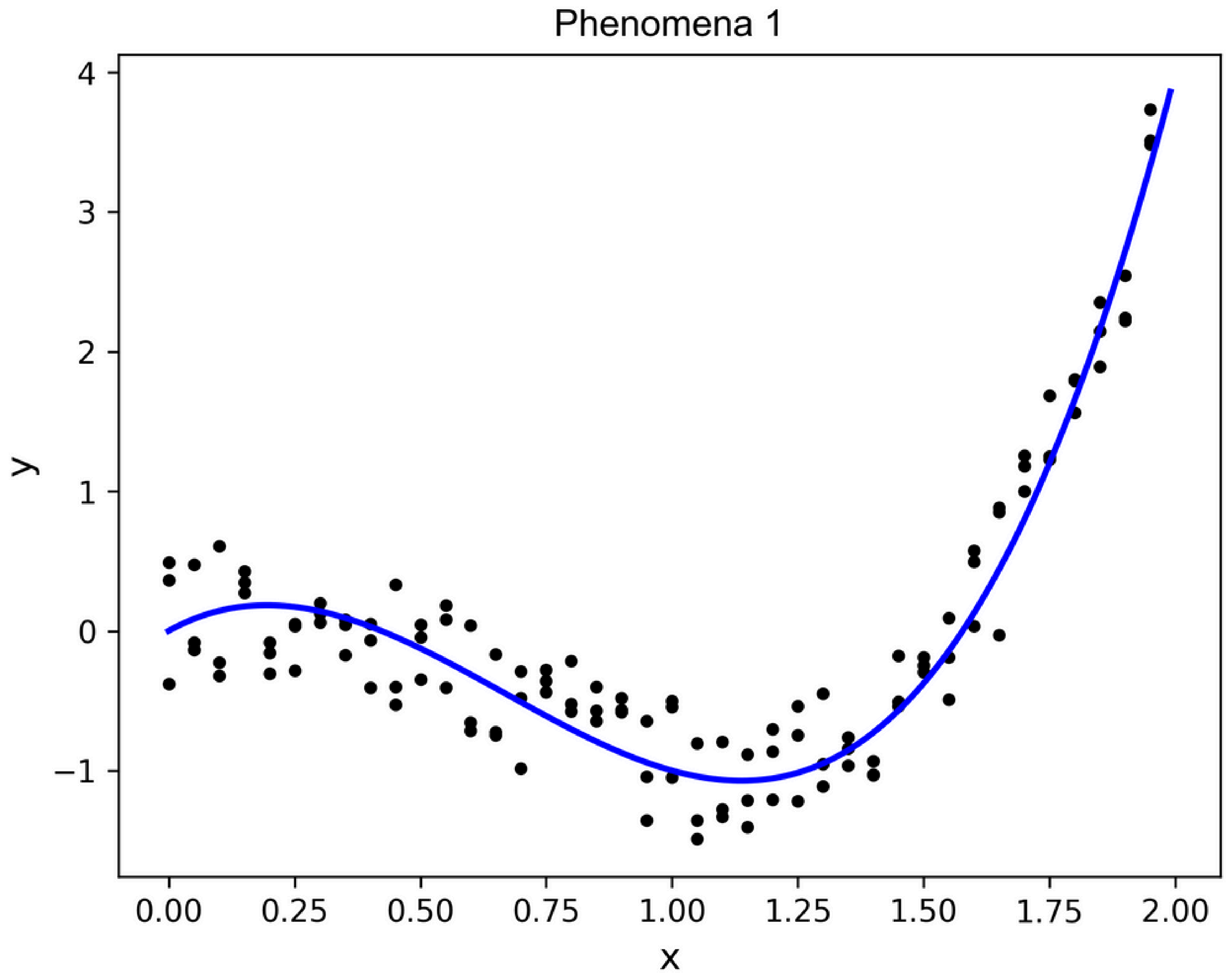
$$f\left(7300\right)=\frac{1}{1000}\times\left(7300\right)$$

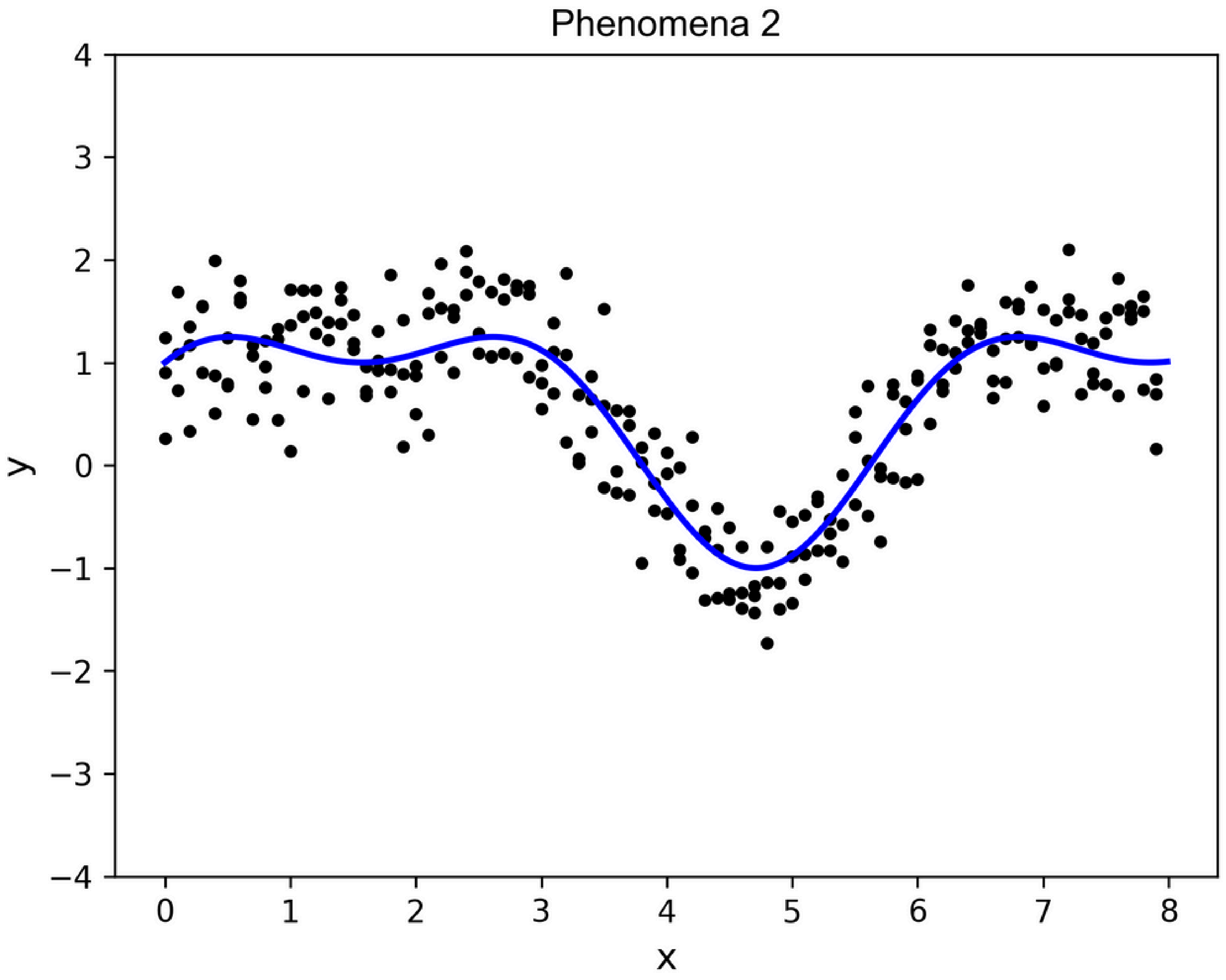
$$\approx \mathbf{7.3}$$




Phenomena 1
y
x
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
4
3
2
1
0
−1


Phenomena 2
y
x
0
1
2
3
4
5
6
7
8
4
3
2
1
0
−1
−2
−3
−4

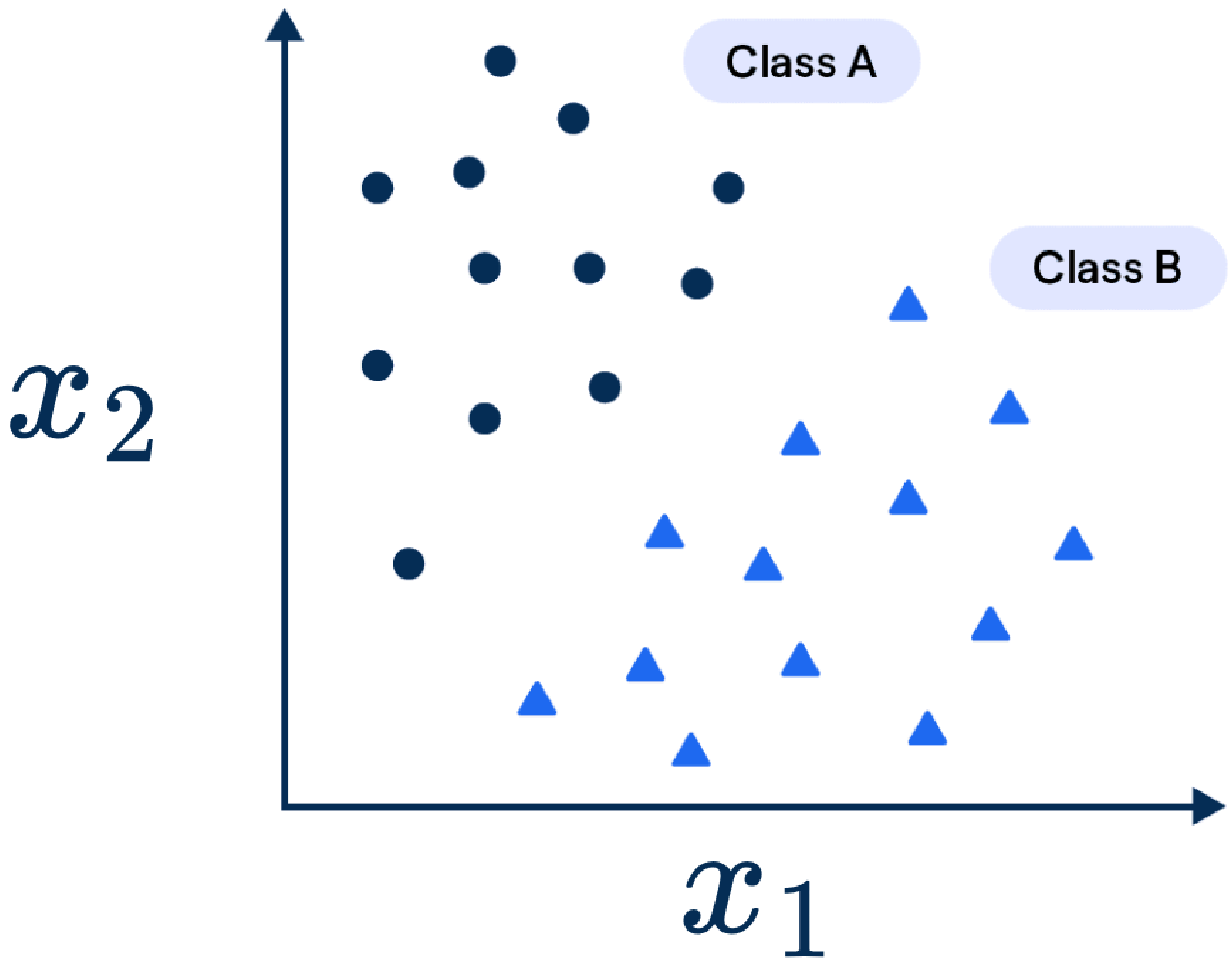

Phenomena 1
y
x
4
3
2
1
0
−1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00


Phenomena 2
y
x
4
3
2
1
0
−1
−2
−3
−4
0
1
2
3
4
5
6
7
8

Prediction

Prediction in AI is when a computer uses information it has learned to make guesses about what might happen next. For example, it can look at patterns in data, to suggest what might occur in the future.

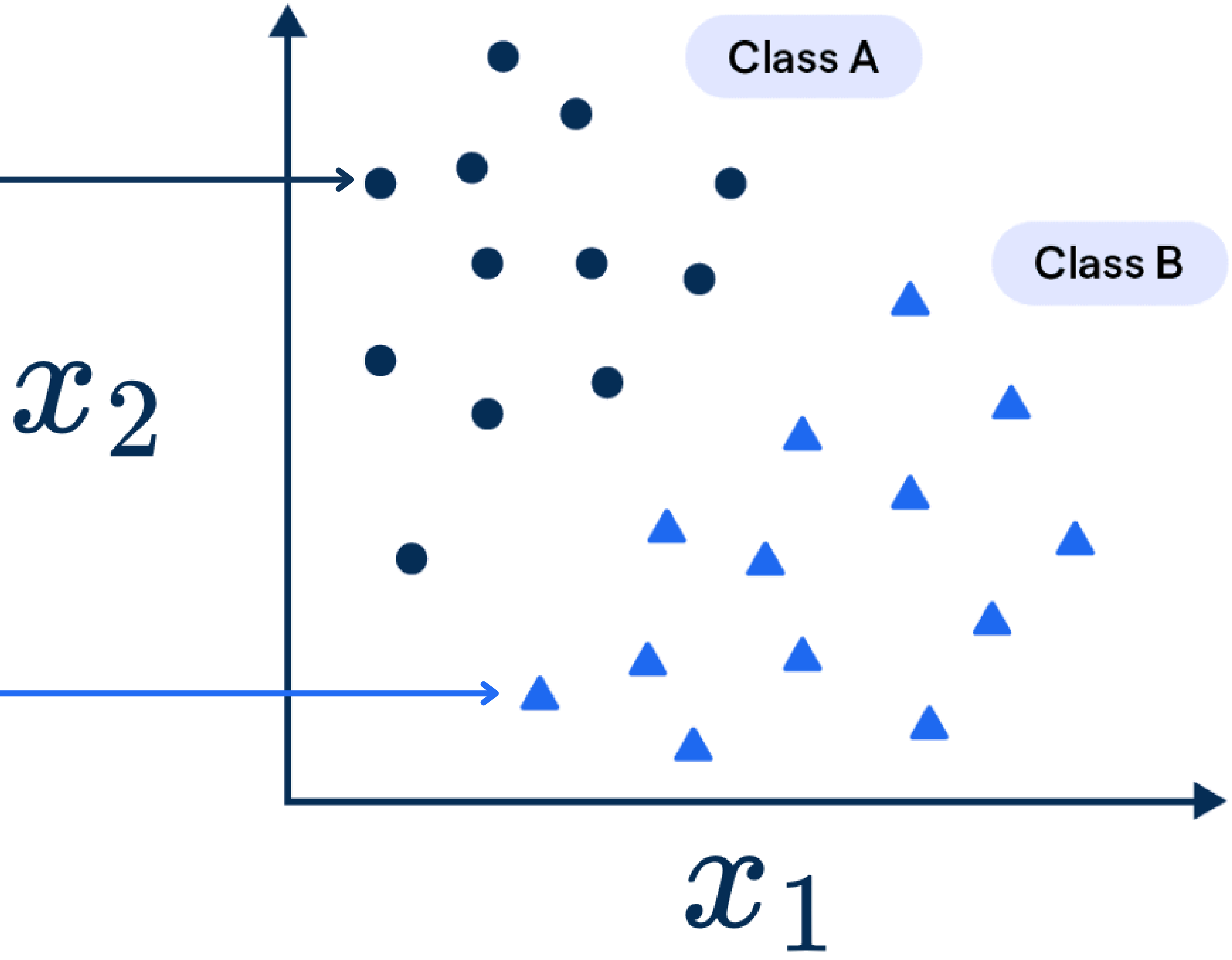
Classification



Classification

example 1:

$x_1 = 1$
$x_2 = 5$
$y = 0\,(class A)$



example 0:

$x_1 = 2$
$x_2 = 1$
$y = 1\,(class B)$

Previous example (prediction)

$x_1 =$ **Income**

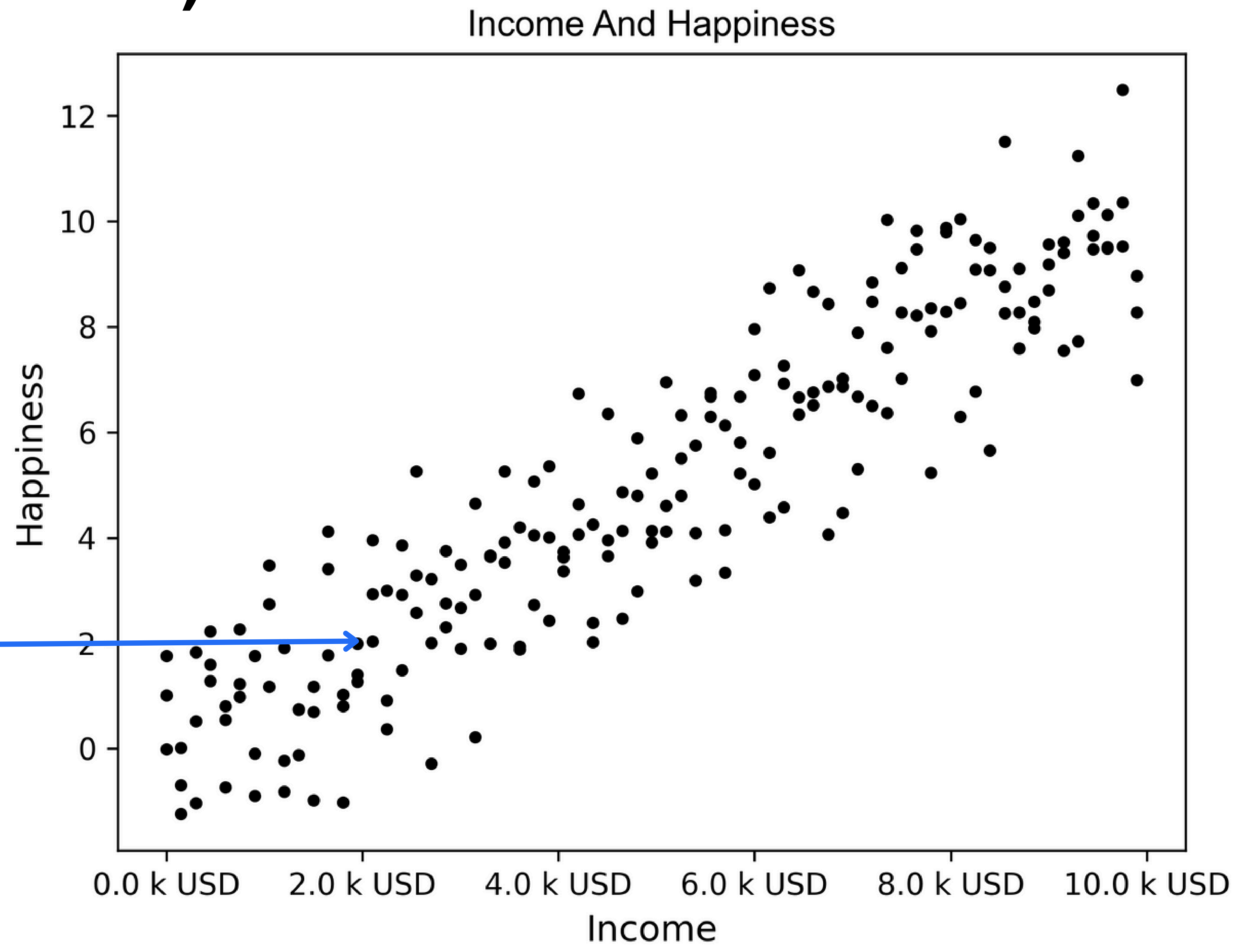
$y =$ **Happiness**

example 0:

$x_1 = 2000$

$x_2 =$ ❌

$y = 2.3$



Classification

Classification is when a computer learns to sort things into different groups based on their features. For example, it can help identify if a plant is toxic or not based on the width and height of a leaf.

Classification vs Prediction

In deep learning:

<u>Classification</u> is the task of categorizing data into <u>predefined classes</u> based on learned patterns, such as identifying an image as a "cat" or "dog."

<u>Prediction</u> involves estimating a <u>continuous</u> or future value based on learned trends, such as forecasting temperature or stock prices.

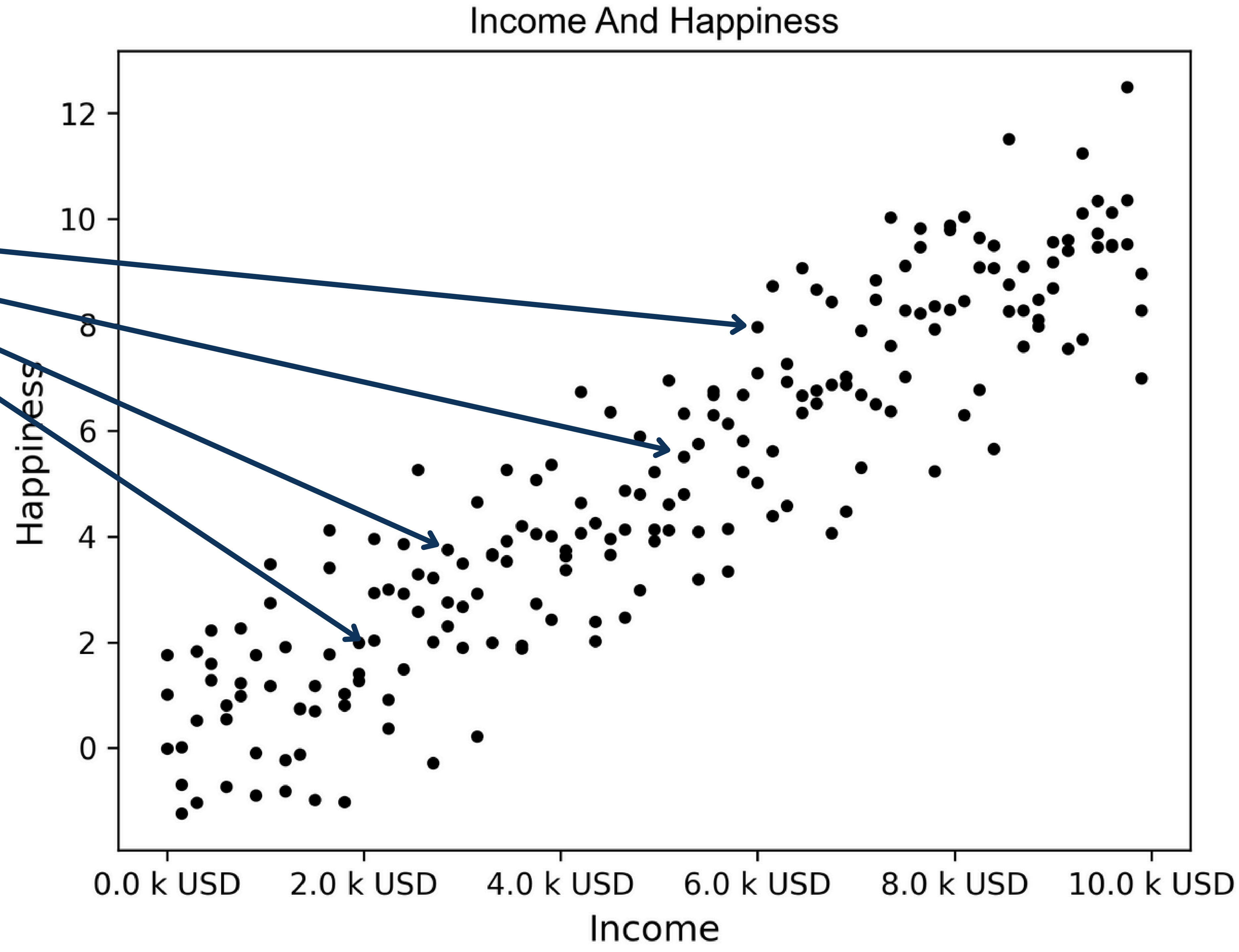
Both rely on learning from labeled data but differ in <u>output type</u>, <u>discrete</u> for classification, <u>continuous</u> for prediction.

Let’s Unveil Deep Learning!

Essential Terms

These points are called <u>**Instances**</u> or <u>**Examples**</u>, they belong to the <u>**Dataset.**</u>

The <u>**Dataset**</u> is essentially a collection of <u>**Examples**</u>.



Essential Terms

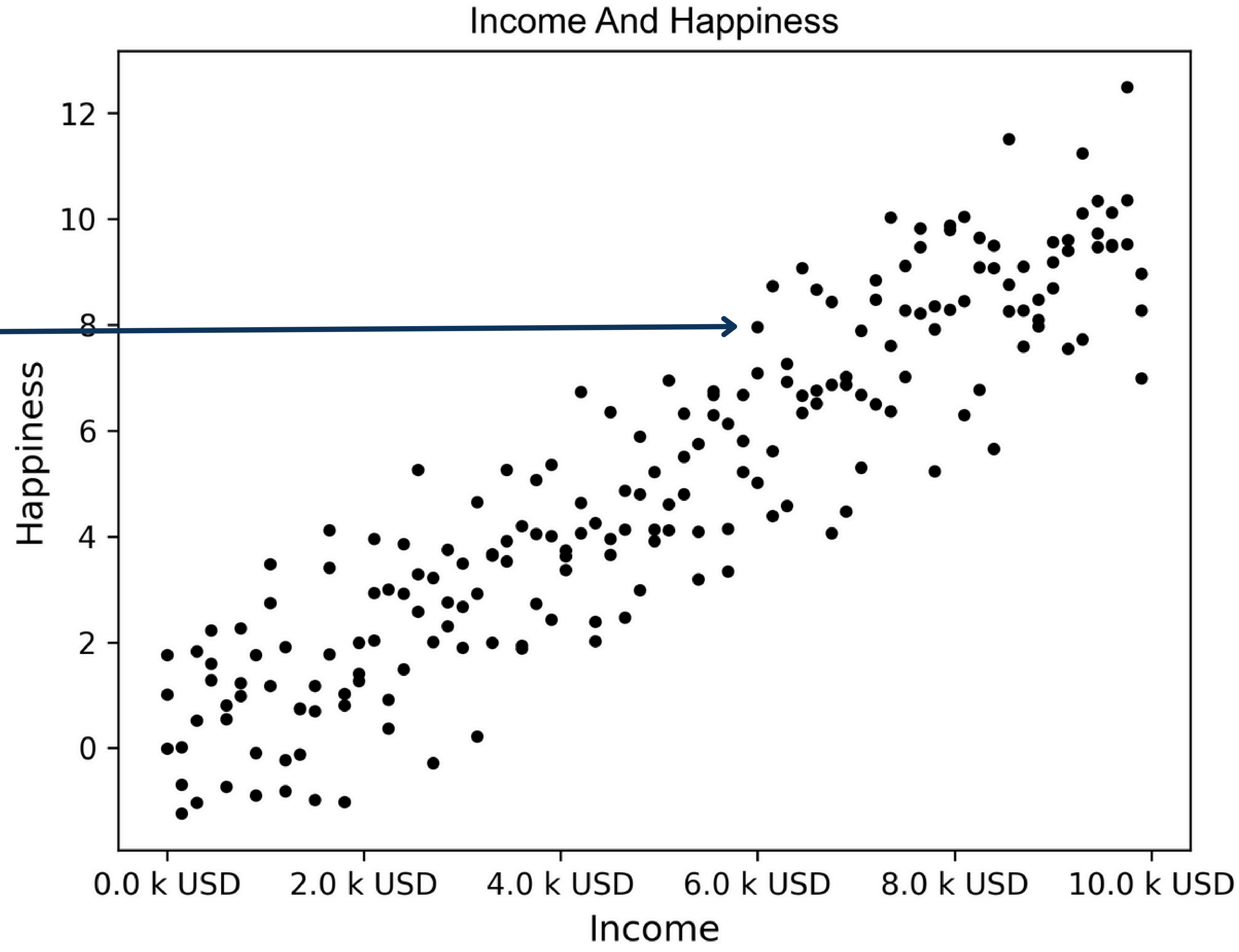
Feature ⟶

Target variable ⟶

example:

$$x_1 = 6000$$

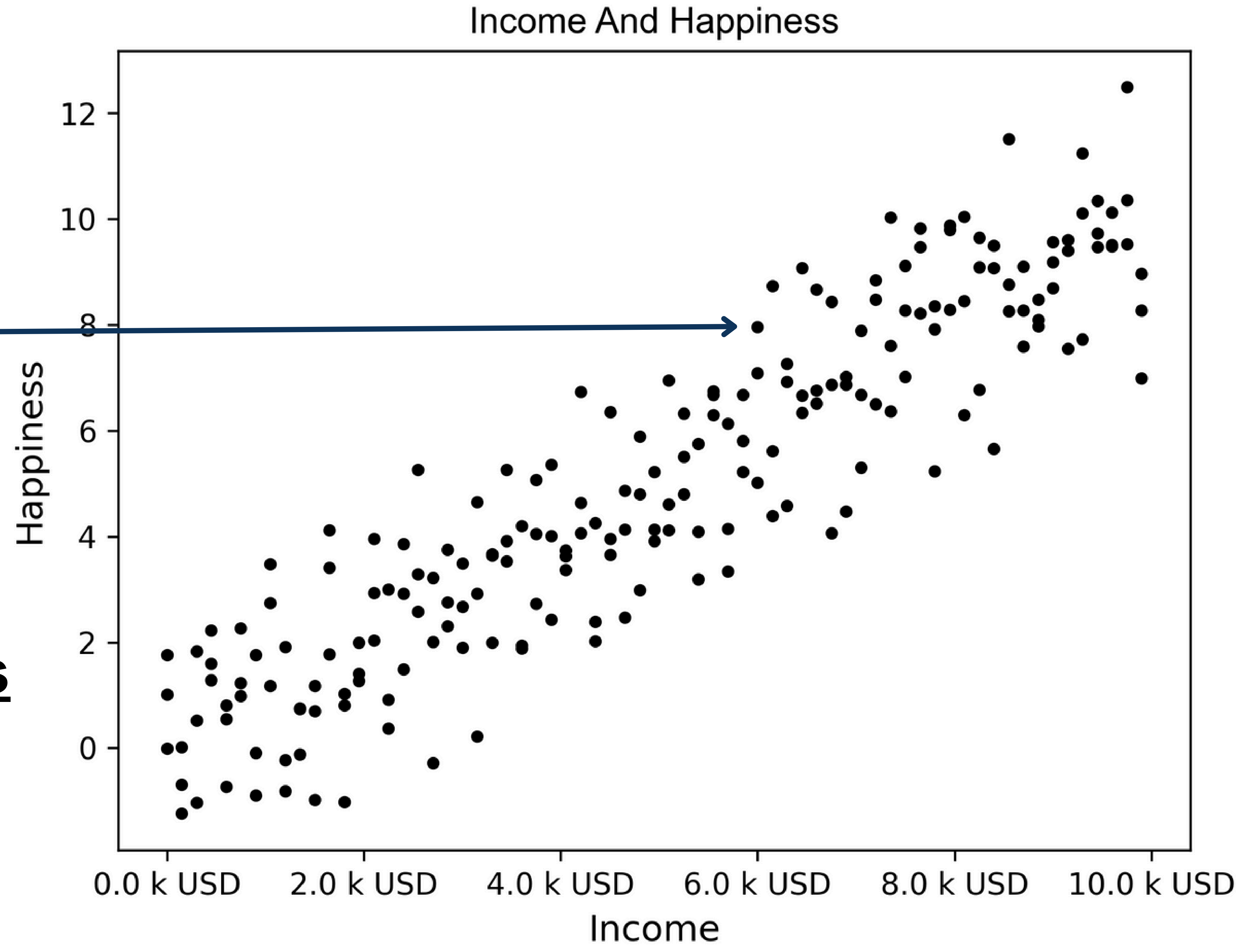
$$y = 7.9$$



Essential Terms

Feature → **example:** $x_1 = 6000$

Target variable → $y = 7.9$



Each example consists of features and, in supervised learning, a target value.

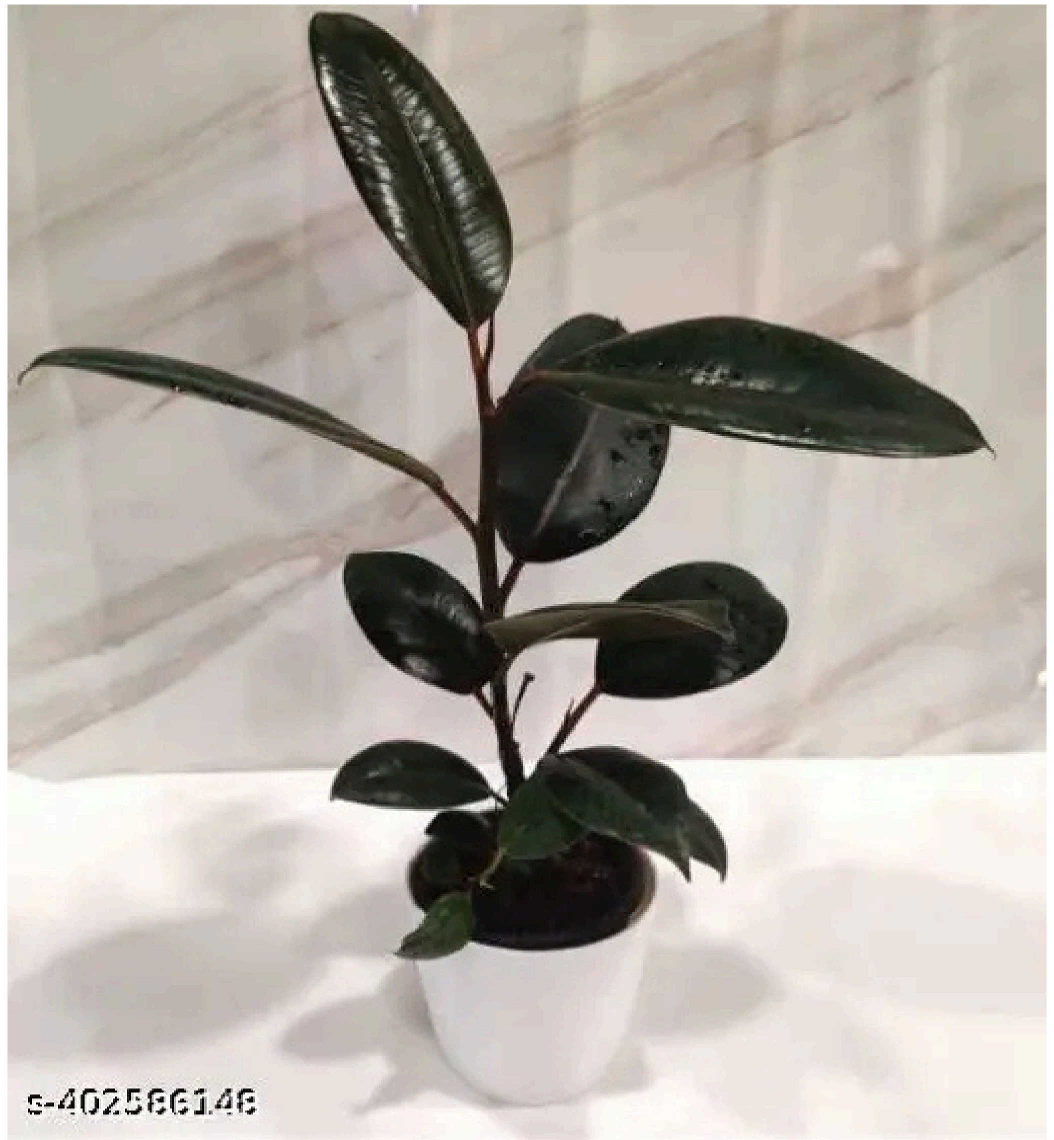
The toxic and non-toxic plant

We want to create a program that can classify a plant as toxic or non-toxic.

The toxic and non-toxic plant

Toxic Plants

The toxic and non-toxic plant



Non-toxic plant

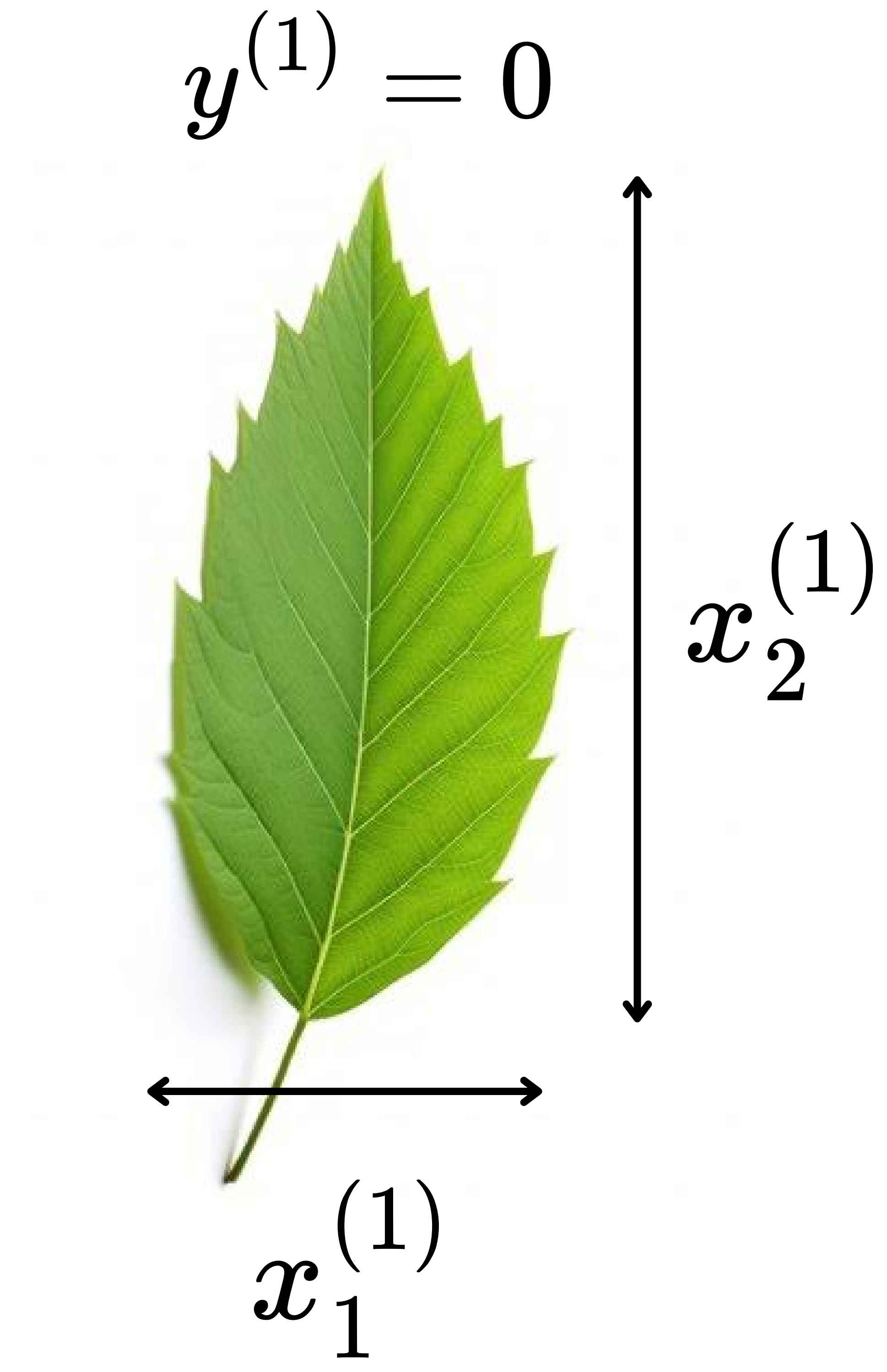
The toxic and non-toxic plant

(n) Represents the <u>number</u> of the example

x_1 Represents the <u>width</u> of the leaf

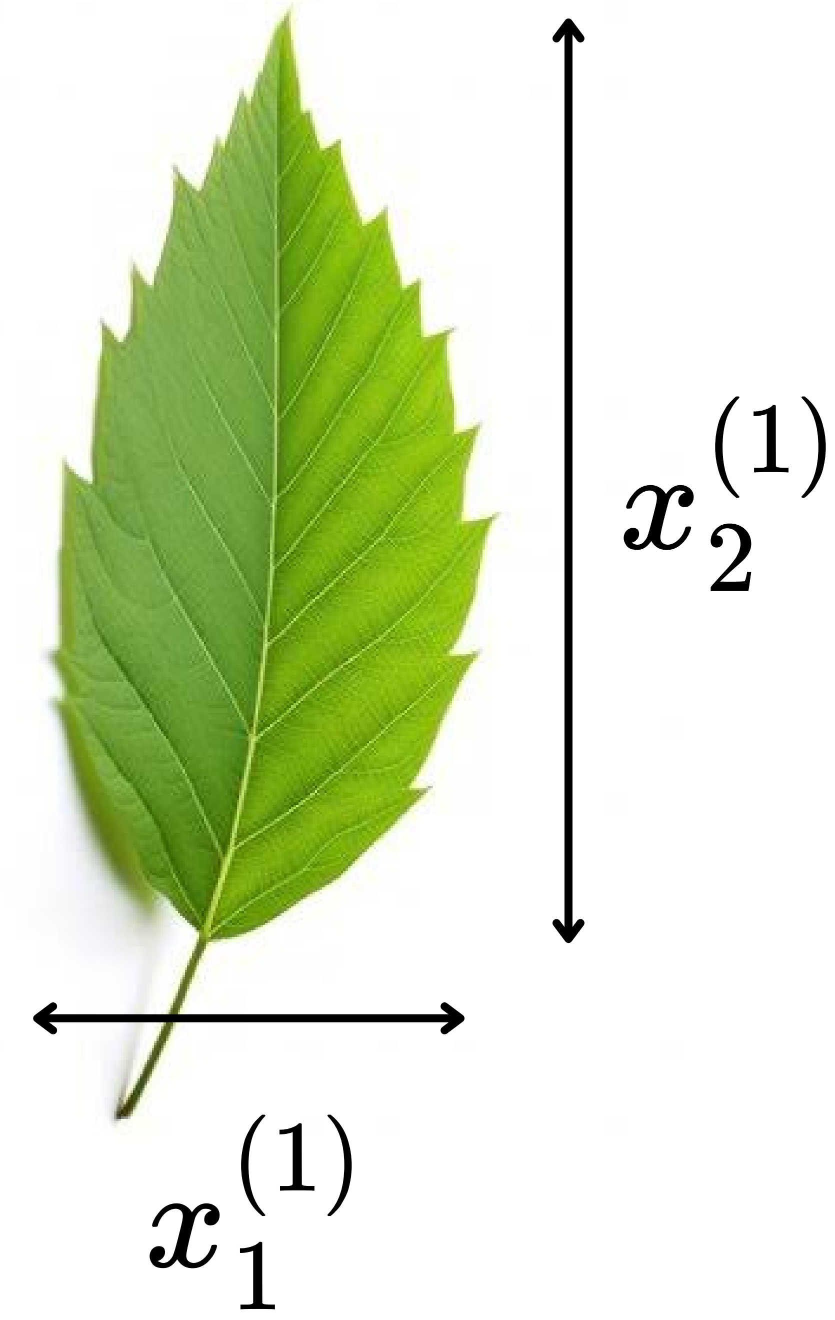
x_2 Represents the <u>length</u> of the leaf

y Represents the <u>class</u> of the leaf

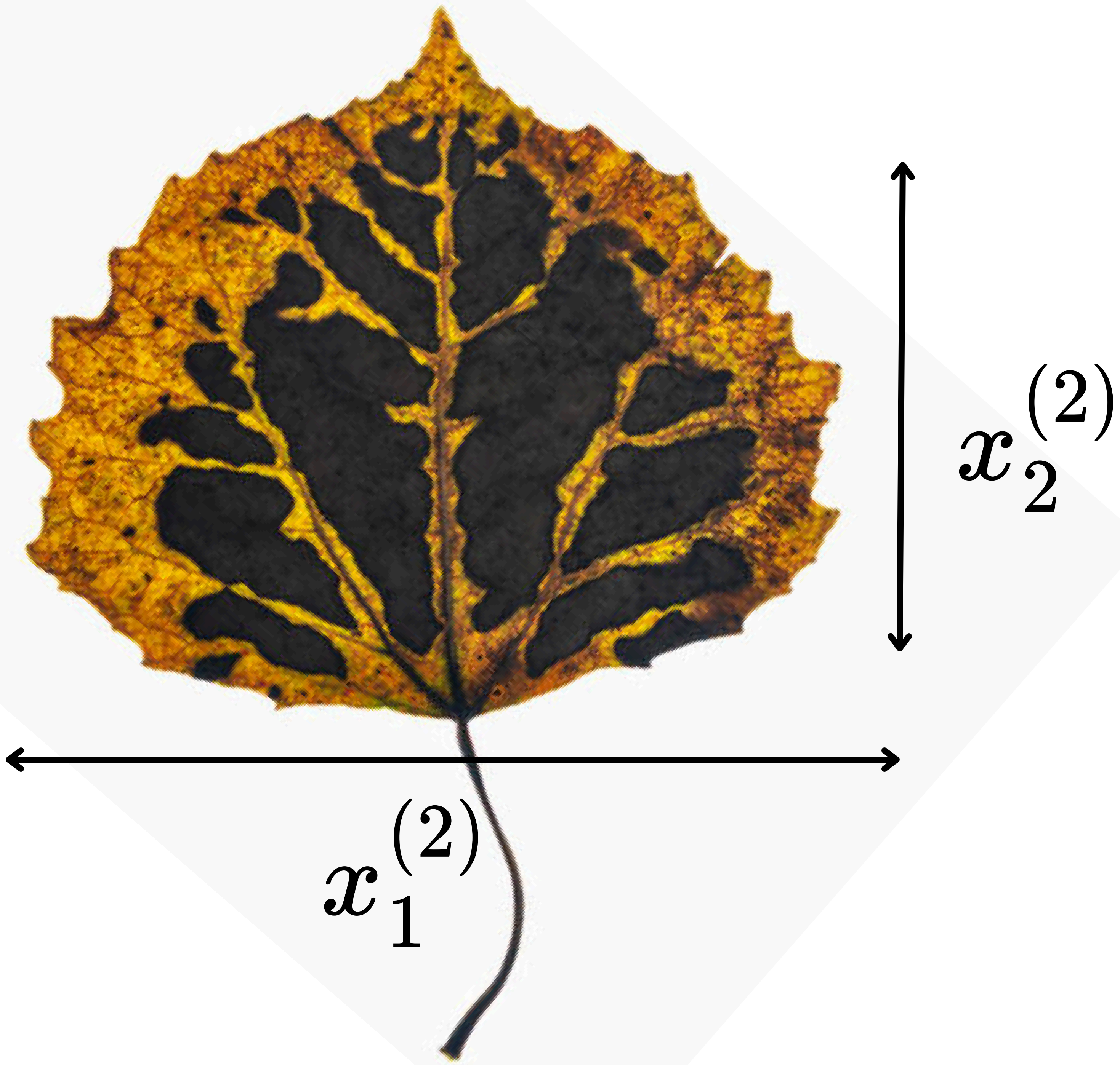


Example 1

The toxic and non-toxic plant



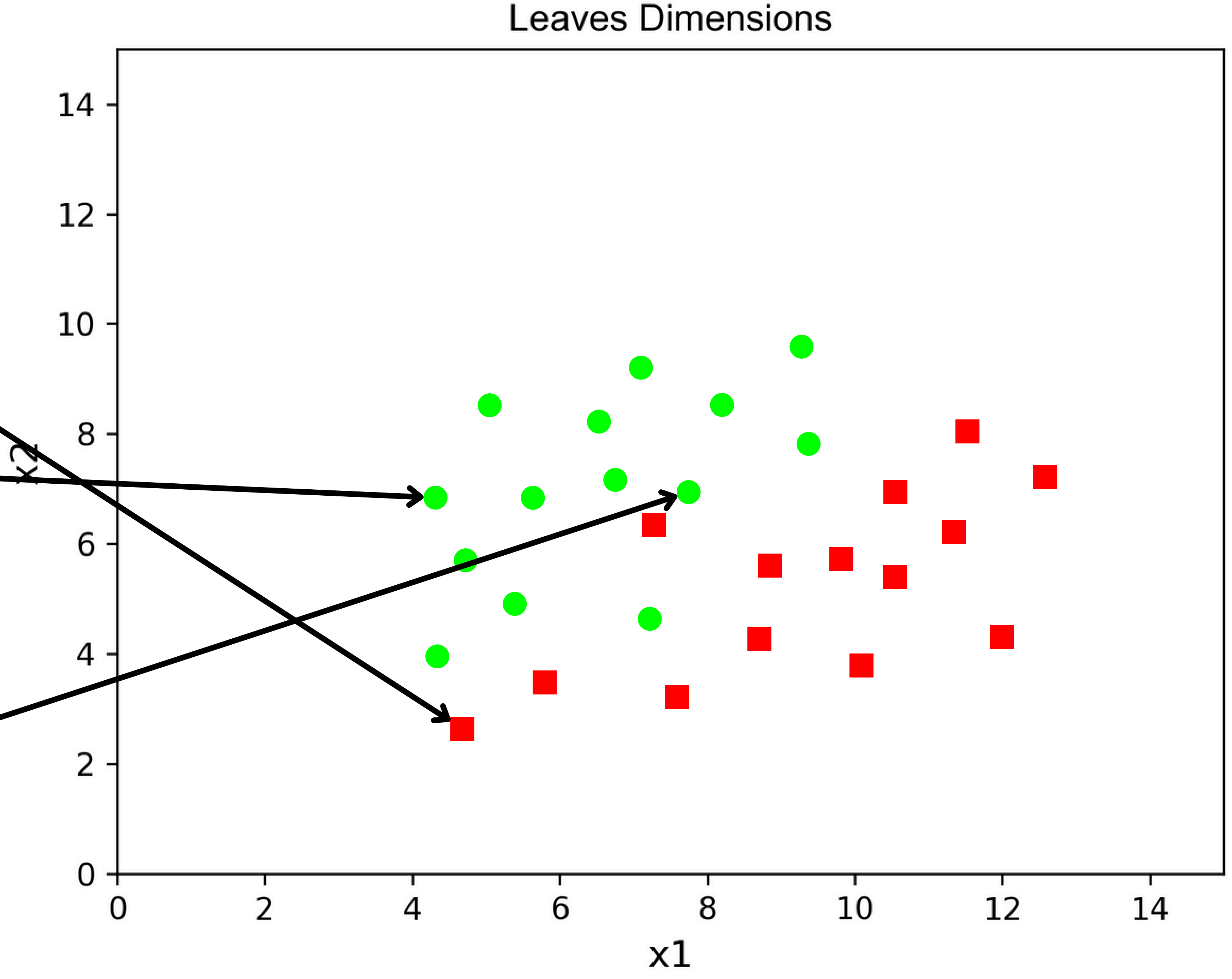
Example 1



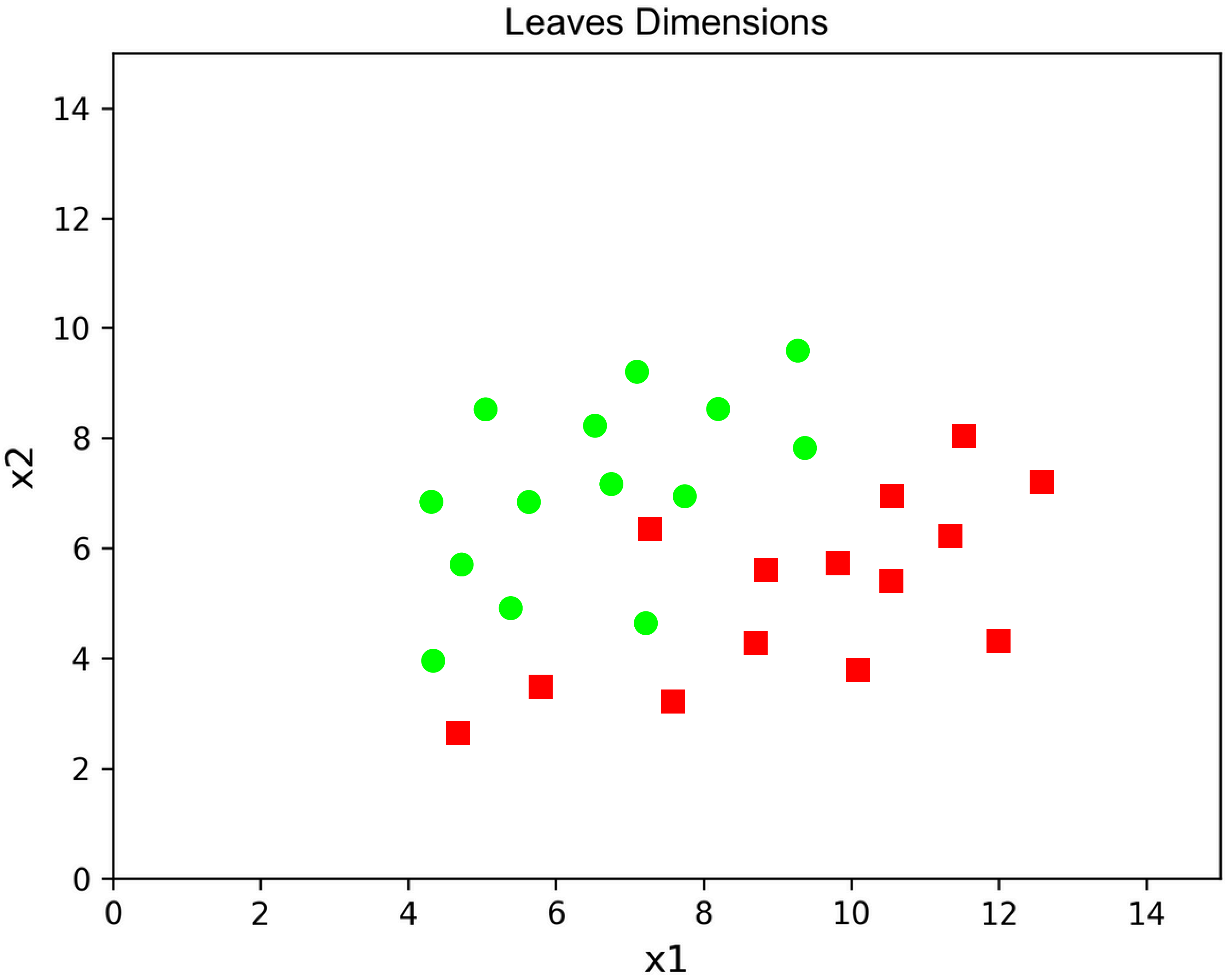
Example 2

The toxic and non-toxic plant

ID	x_1	x_2	y
1	4.7	2.6	1
2	4.2	6.9	0
⋮	⋮	⋮	⋮
28	7.8	6.8	0



Separating the two classes of points



From 2 to 3 dimensions

To understand how to classify something with two features (or variables), it's helpful to first see how classification works with a single variable.

For this purpose, we'll study the weather to determine if it's classified as hot or cold.

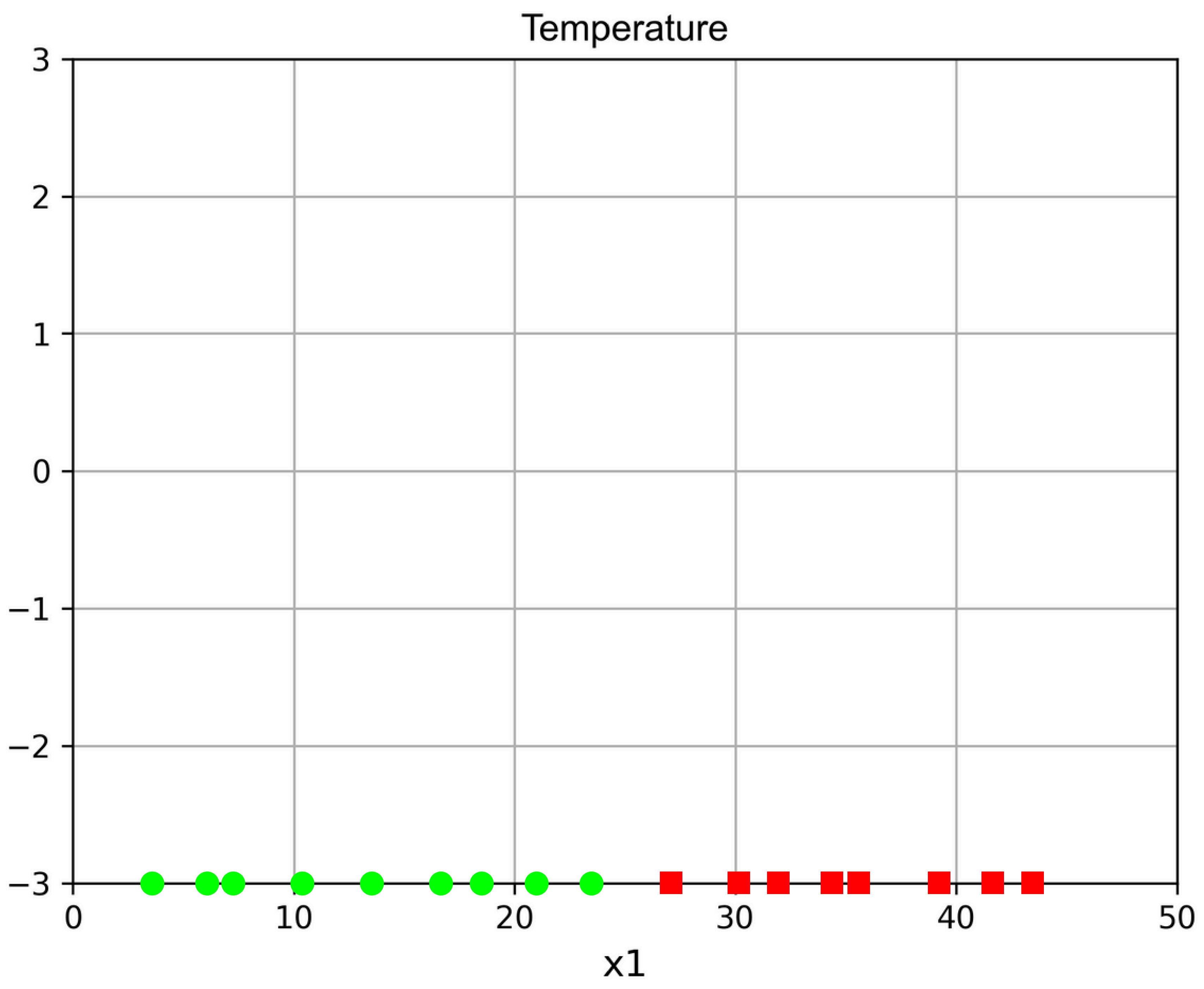
If the temperature is $\geqslant 25^{\circ}$C, we consider the weather to be hot, otherwise it is cold.

From 2 to 3 dimensions

Here, each example has <u>one</u> feature and belongs to one of two classes.

x_1 **Represents the <u>temperature</u>.**

y **Represents the <u>class</u> of the weather. 0 for cold and 1 for hot.**



The Perceptron

A perceptron is a basic unit of a neural network that mimics a single neuron.

It takes multiple inputs, applies weights to each, sums them, and passes the result through an activation function to produce an output.

Perceptrons are the foundation of binary classifiers, determining if input data belongs to one of two classes.

How does a Perceptron work?

In our previous case:

x_1 **Represented the <u>temperature</u>.**

y **Represented the <u>class</u> of the weather. 0 for cold and 1 for hot.**

How does a Perceptron work?

Let's define the function z, For our previous case:

$$z\left(x_1\right) = w_1 \times x_1 + b$$

w_1 **Is called <u>Weight</u>**

b **Is called <u>Bias</u>**

How does a Perceptron work?

$$z(x_1) = w_1 x_1 + b \begin{cases} \geqslant 0 \longrightarrow y_{pred} = 1 \\ < 0 \longrightarrow y_{pred} = 0 \end{cases}$$

y_{pred} **is called <u>Predicted Output</u> or <u>Predicted Value</u>**

How does a Perceptron work?

Note:

y_{pred} **is the <u>Predicted Output</u> from the <u>Perceptron.</u> It is different than** y

y **This is the <u>actual</u> or true label value of the <u>example</u> from the <u>dataset</u>.**

How does a Perceptron work?

$$z\left(x_1\right) = w_1 x_1 + b$$

$$\begin{cases} y_{pred} = 1 & if \quad z\left(x_1\right) \geqslant 0 \\ y_{pred} = 0 & else \end{cases}$$

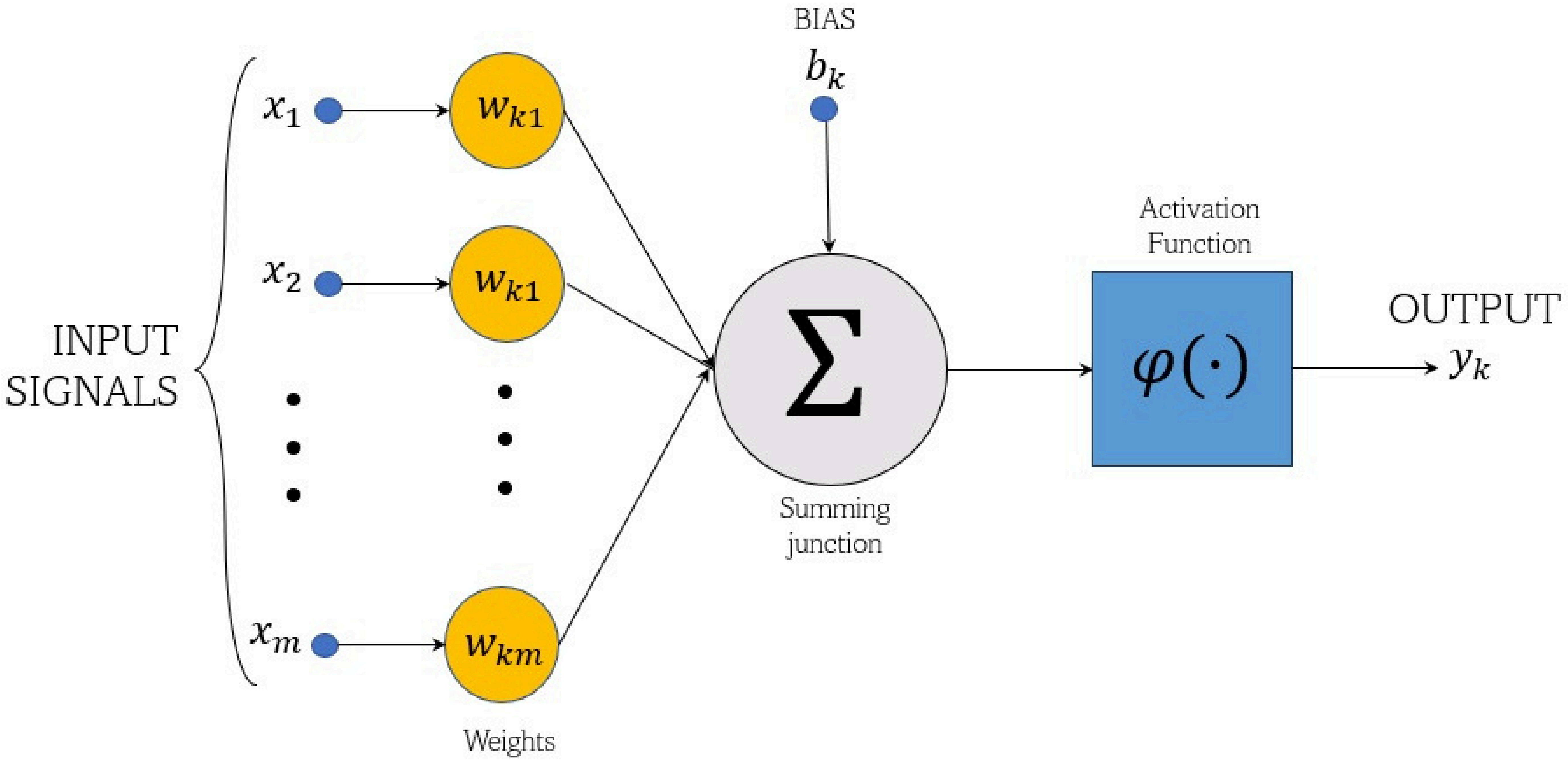
The Perceptron

Z formula if we had:

2 features ⟶ $$z\left(x_1, x_2\right) = w_1 x_1 + w_2 x_2 + b$$

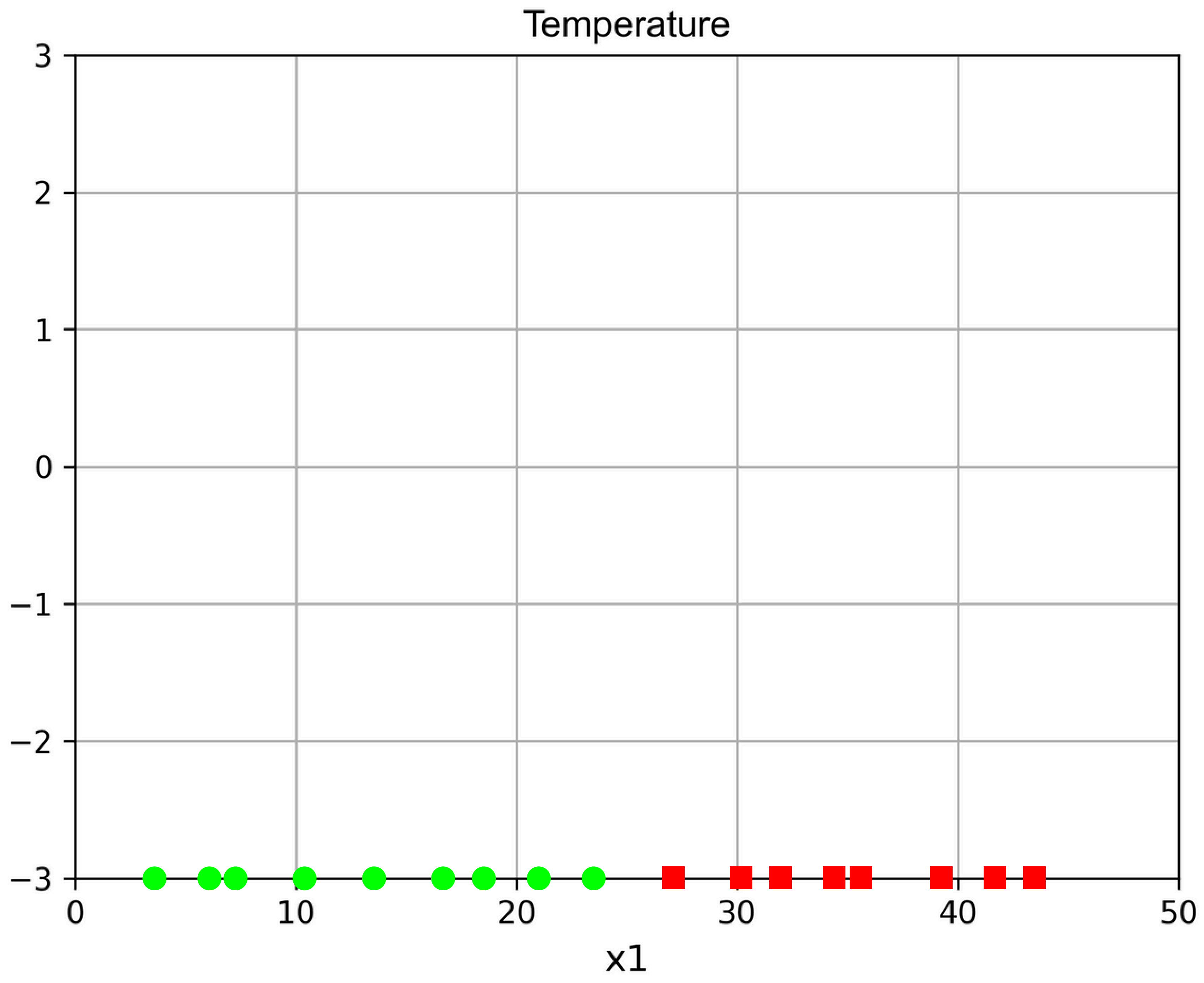
n features ⟶ $$z\left(x_1, x_2, \ldots, x_n\right) = \sum_{i=1}^{n} w_i x_i + b$$

The Perceptron



Weather scenario

Returning to the weather scenario, we will focus on separating the two classes of data points.



Weather scenario

Remember that if $z(x_1) \geqslant 0$ **then** $y_{pred} = 1$

So what we really want is to have is:

$$z(x_1) \geqslant 0 \quad \text{if} \quad x_1 \geqslant 25$$

$$z(x_1) < 0 \quad \text{if} \quad x_1 < 25$$

Weather scenario

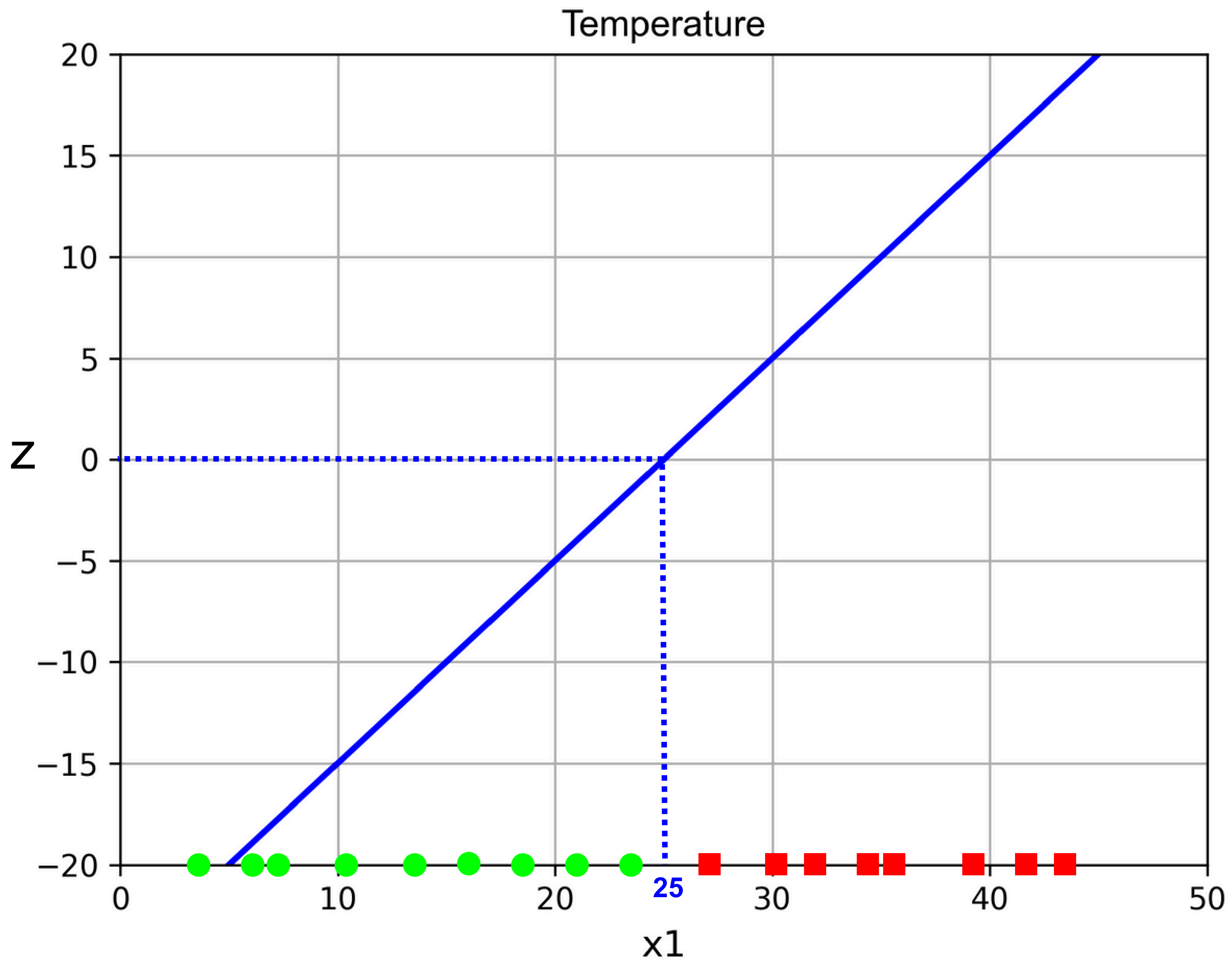
We want a $z\left(x\right)$ **that has the following graph:**

$$z\left(x_1\right) = x_1 - 25$$

So that:

$z\left(x_1\right) \geqslant 0$ **if** $x_1 \geqslant 25$

$z\left(x_1\right) < 0$ **if** $x_1 < 25$



Weather scenario

$$z\left(x_1\right) = x_1 - 25$$

Meaning:

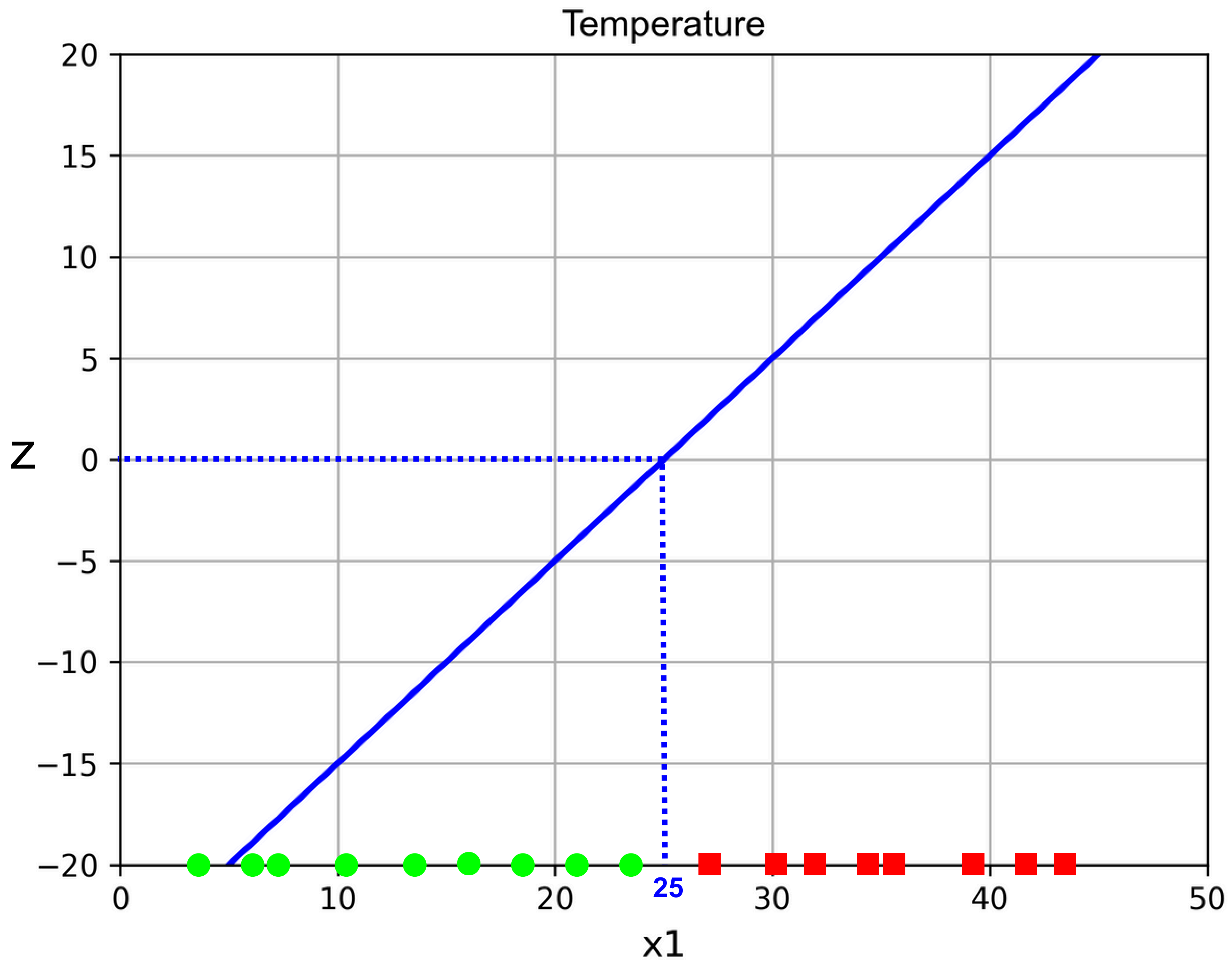
$z\left(x_1\right) \geqslant 0$ **if** $x_1 \geqslant 25$

$z\left(x_1\right) < 0$ **if** $x_1 < 25$

Therefore:

$y_{pred} = 1$ **if** $x_1 \geqslant 25$

$y_{pred} = 0$ **if** $x_1 < 25$

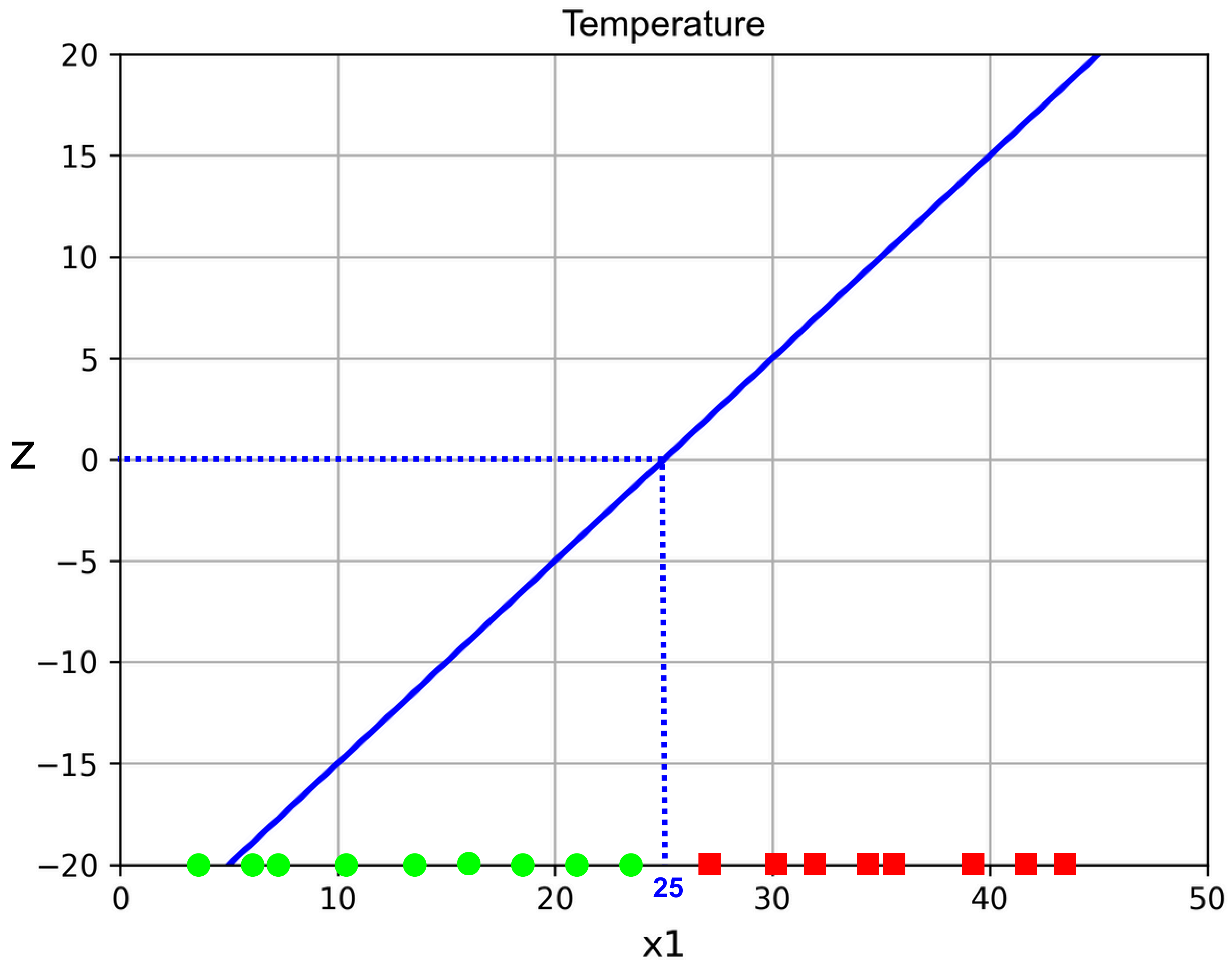


Weather scenario

But the question remains: how did we derive this expression?

$$z\left(x_1\right) = x_1 - 25$$

We create a model and train it!



Model training

Let's go through a few steps to build and train a model that can classify the weather as either hot or cold based on the variable x_1 (temperature).

We start with our <u>prediction function</u>:

$$z(x_1) = w_1 x_1 + b$$

Model training

$$z\left(x_1\right) = w_1 x_1 + b$$

We have no prior knowledge of the ideal values for w_1 and b

Model training

We start with:

$$z\left(x_1\right) = w_1 x_1 + b$$

We have no prior knowledge of the ideal values for w_1 and b

So we just assign completely <u>random</u> values to <u>Weight</u> and <u>Bias</u>.

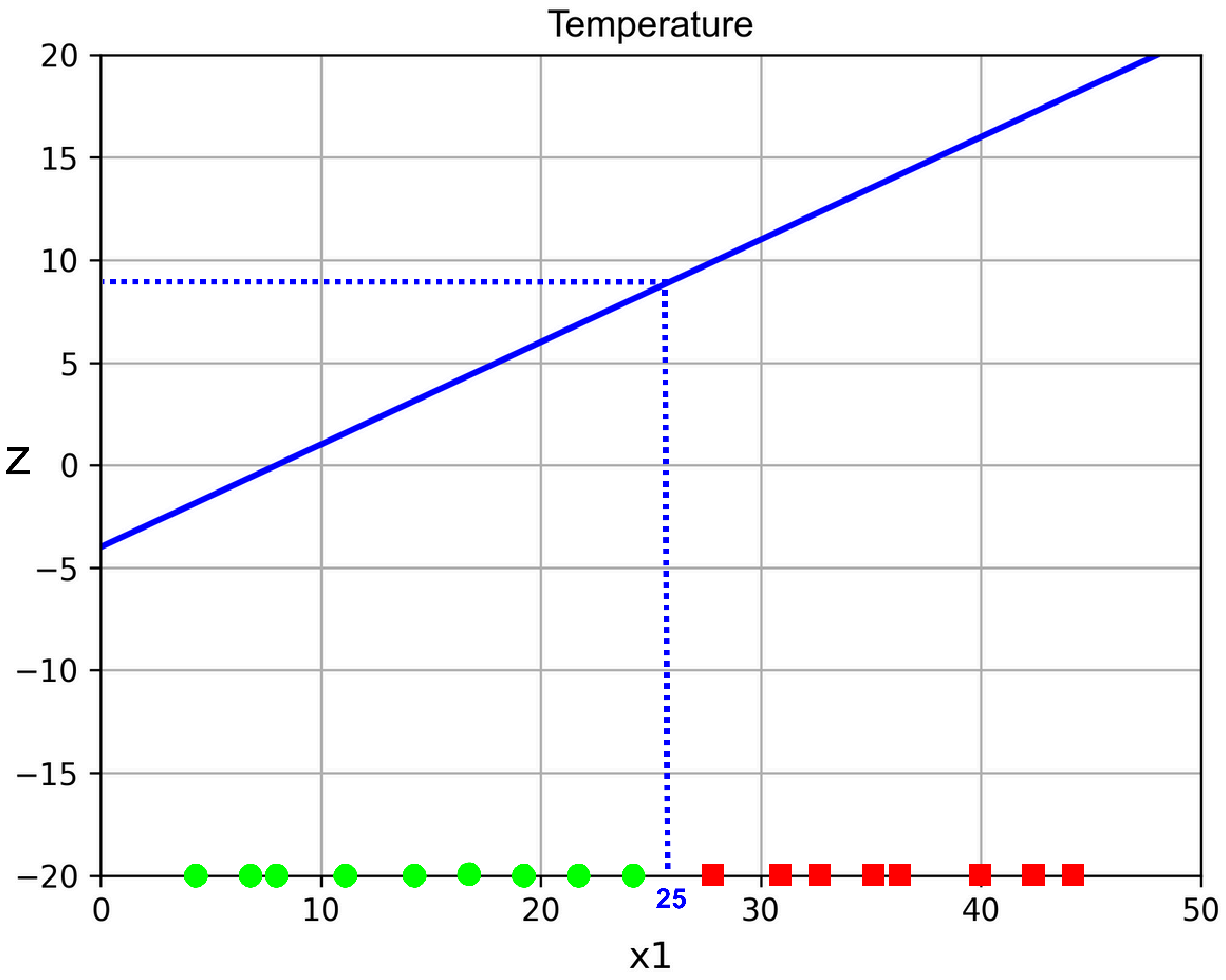
Model training

We start with:

$$w_1 = 0.5$$

$$b = -4$$

$$z\left(x_1\right) = (0.5)x_1 - 4$$



Sigmoid Function

Then, we need to use an activation function.

An activation function helps determine the probability that a data point belongs to a specific class.

It helps to determine the probability that a data point belongs to the 'Hot' class and also the probability that it belongs to the 'Cold' class.

Sigmoid Function

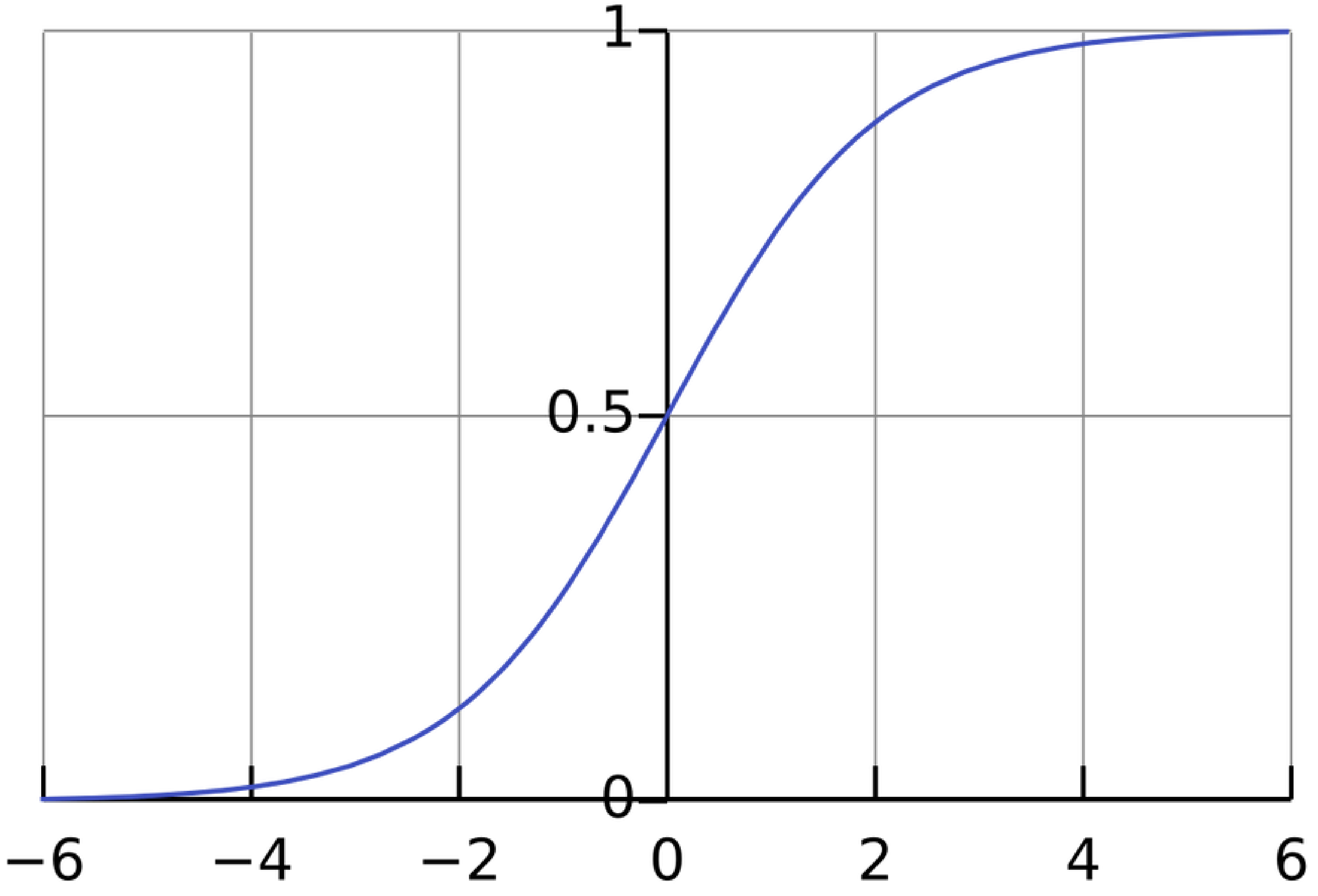
For our model, we are going to use the sigmoid function a **as an activation function.**

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

$$\lim_{x\to+\propto} a(x)=1$$

$$\lim_{x\to-\propto} a(x)=0$$

$$a\left(0\right)=\frac{1}{2}$$



Bernoulli’s law

To calculate the probability of an example belonging to class y we <u>need</u> to apply Bernoulli’s law.

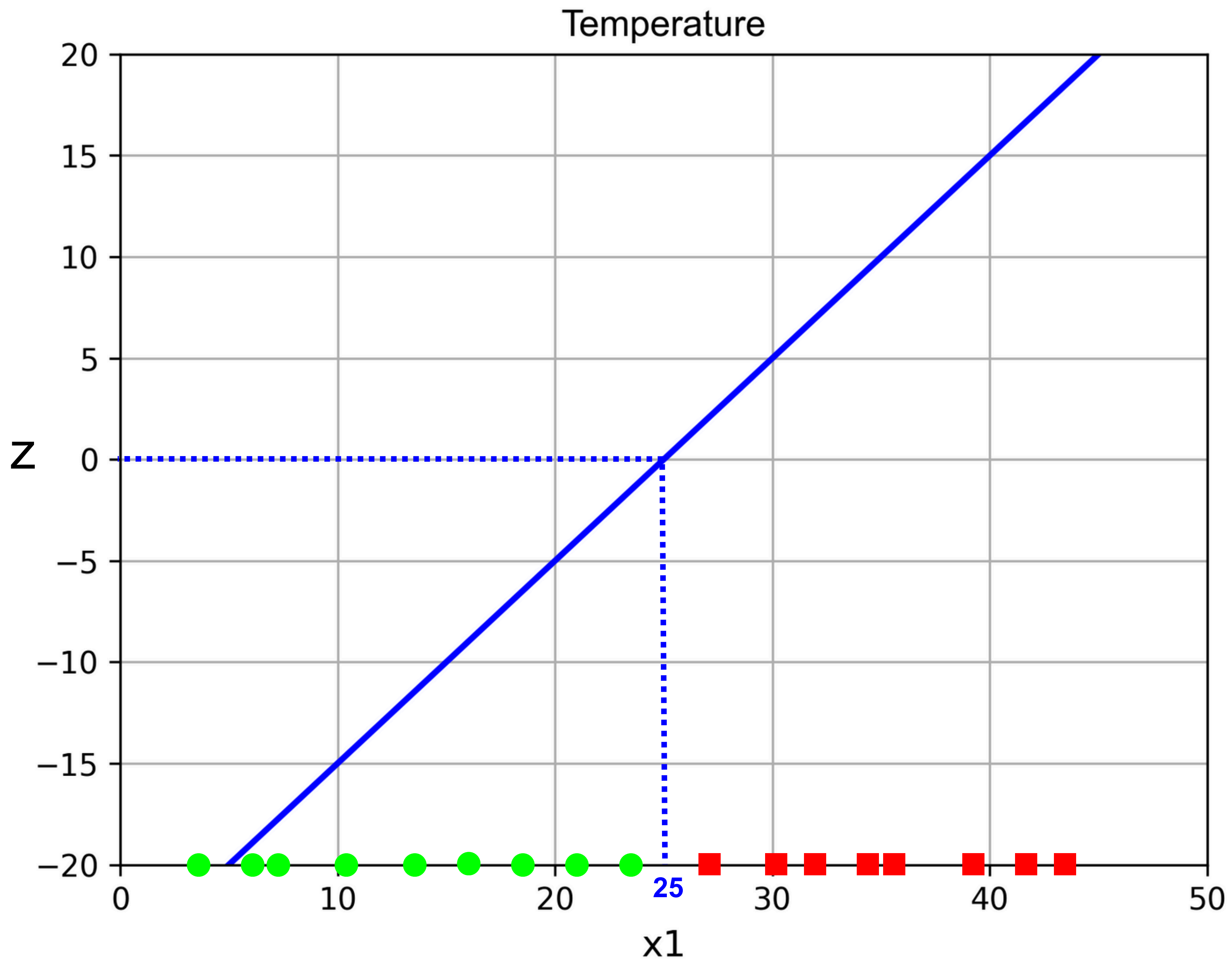
$$P\left(Y = y\right) = a\left(z\right)^{y} \times \left(1 - a\left(z\right)\right)^{1-y}$$

Probabilities

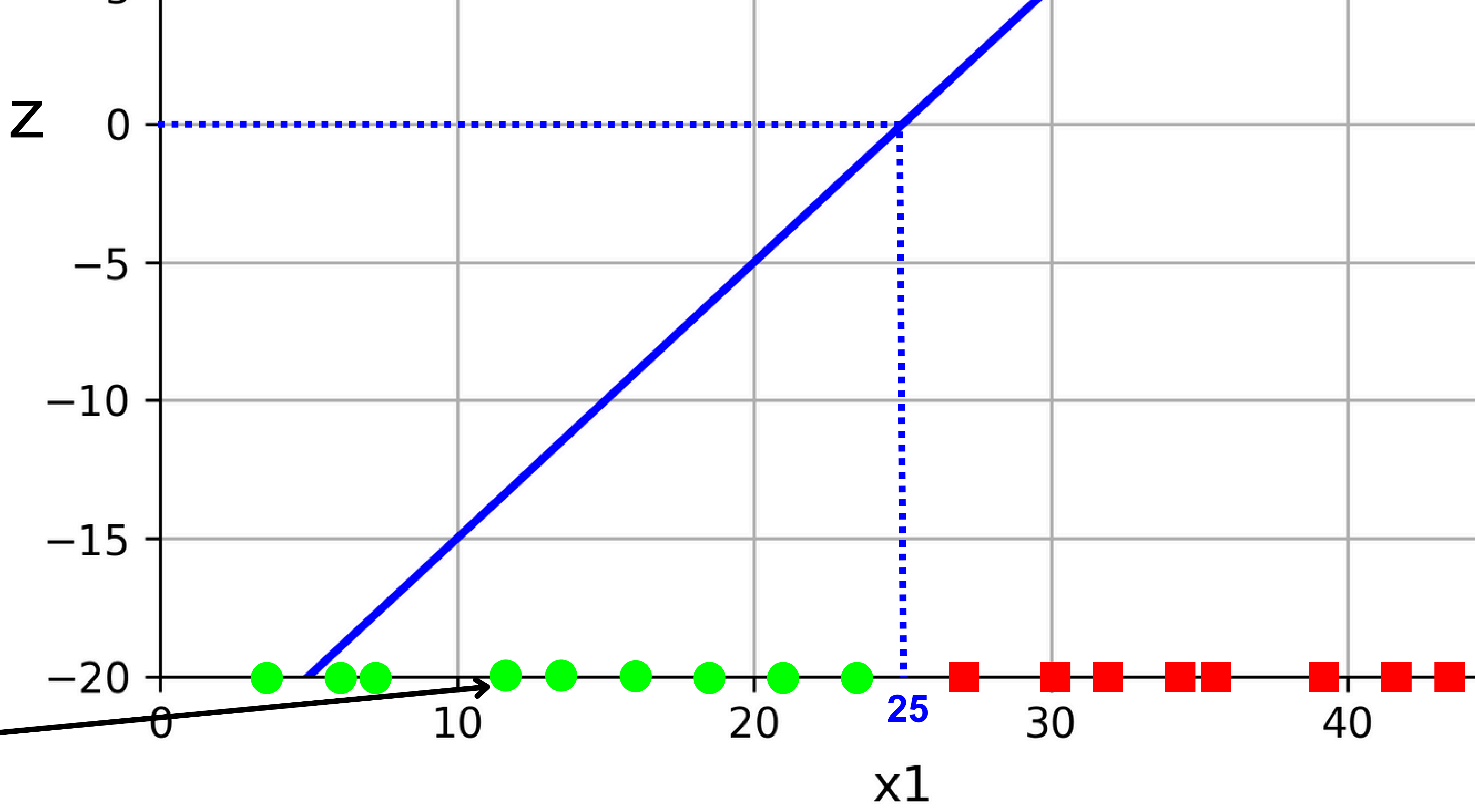
Let's calculate some probabilities for a more solid understanding.

Let's take our previous, accurate model, where:

$$z\left(x_1\right) = x_1 - 25$$



$$z(x_1) = x_1 - 25$$



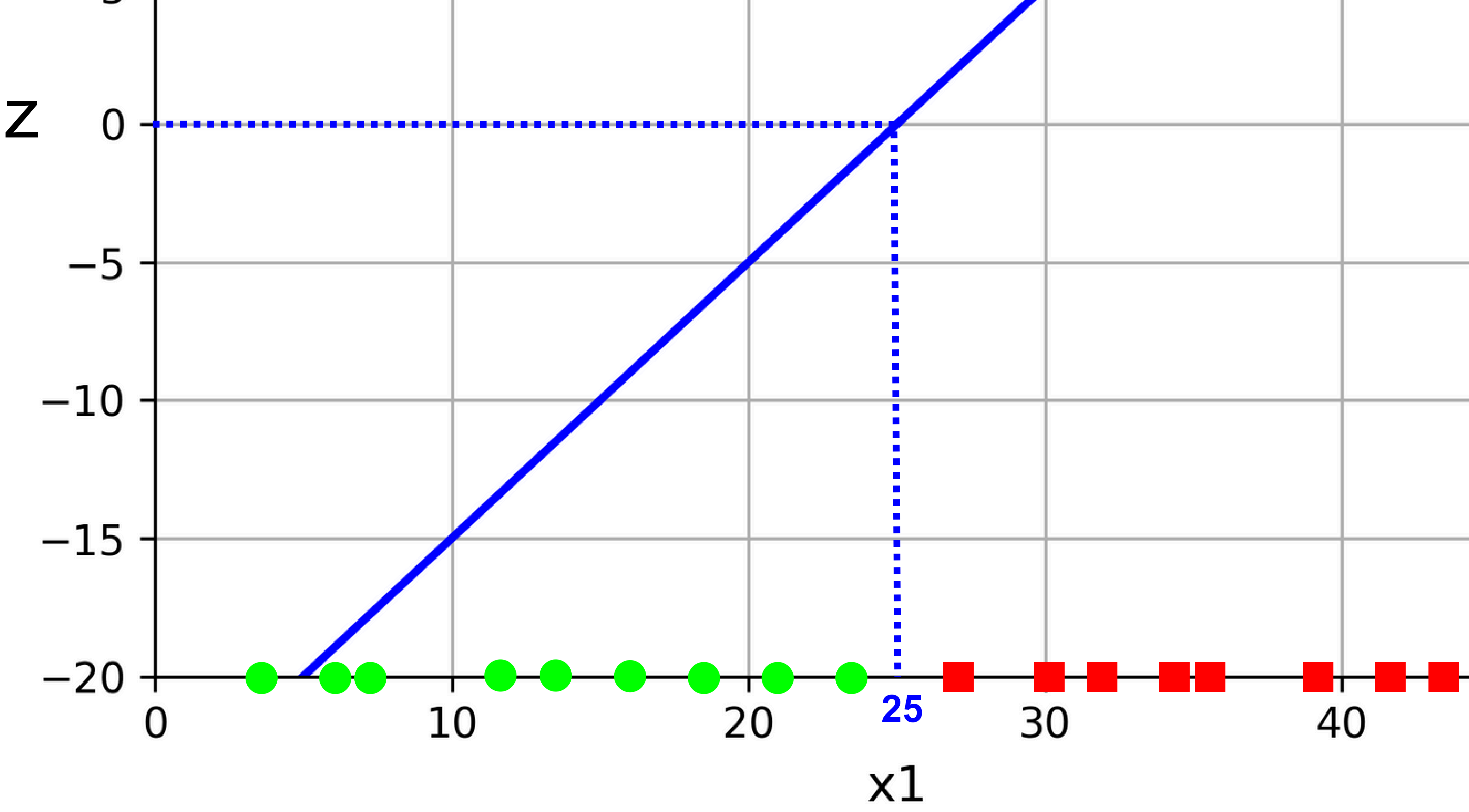
example (3)

$$x_1^{(3)} = 12$$

$$y^{(3)} = 0$$

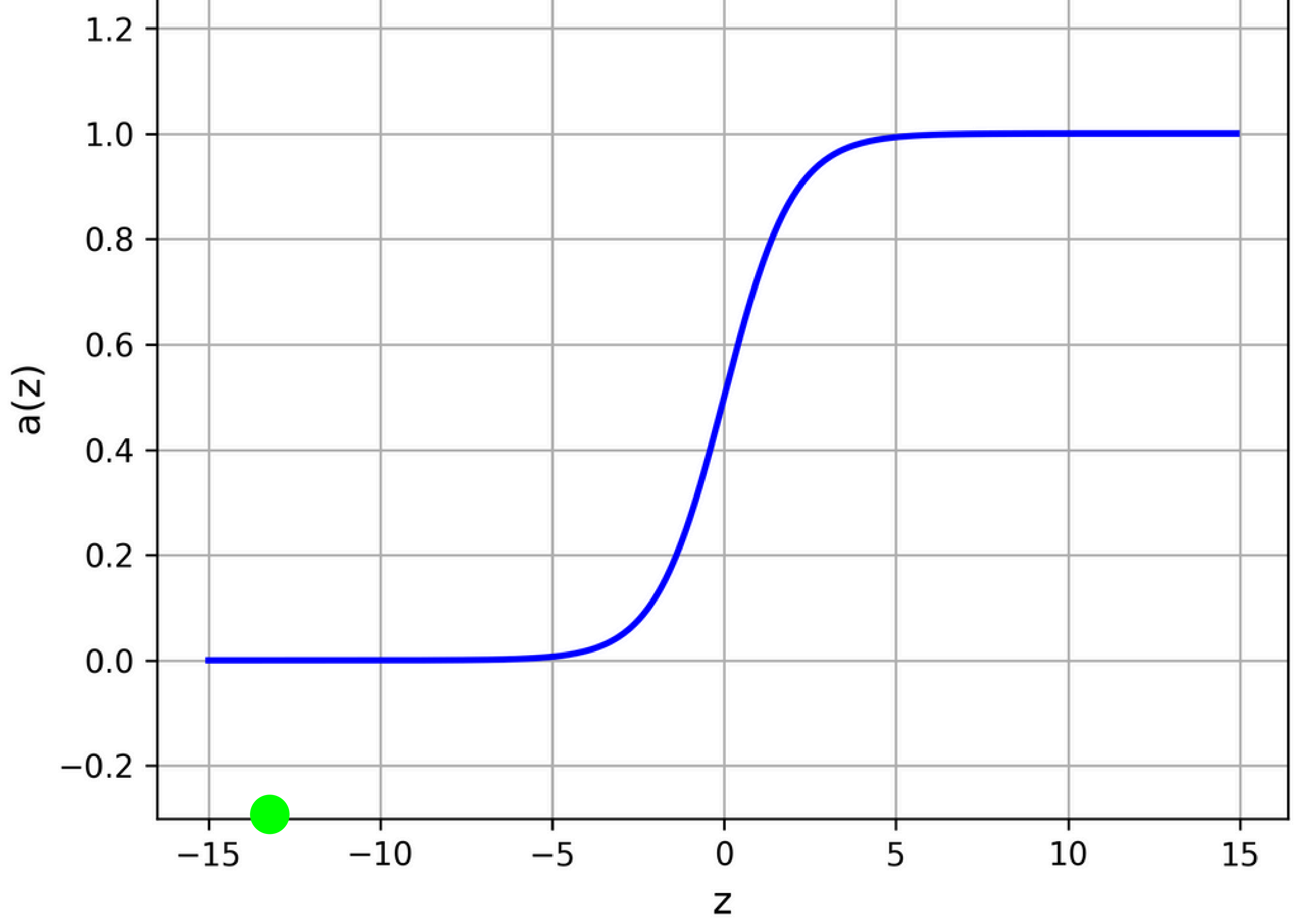
$$z(12) = 12 - 25 = -13 \longrightarrow a(z(12)) = \frac{1}{1 + e^{-z(12)}}$$

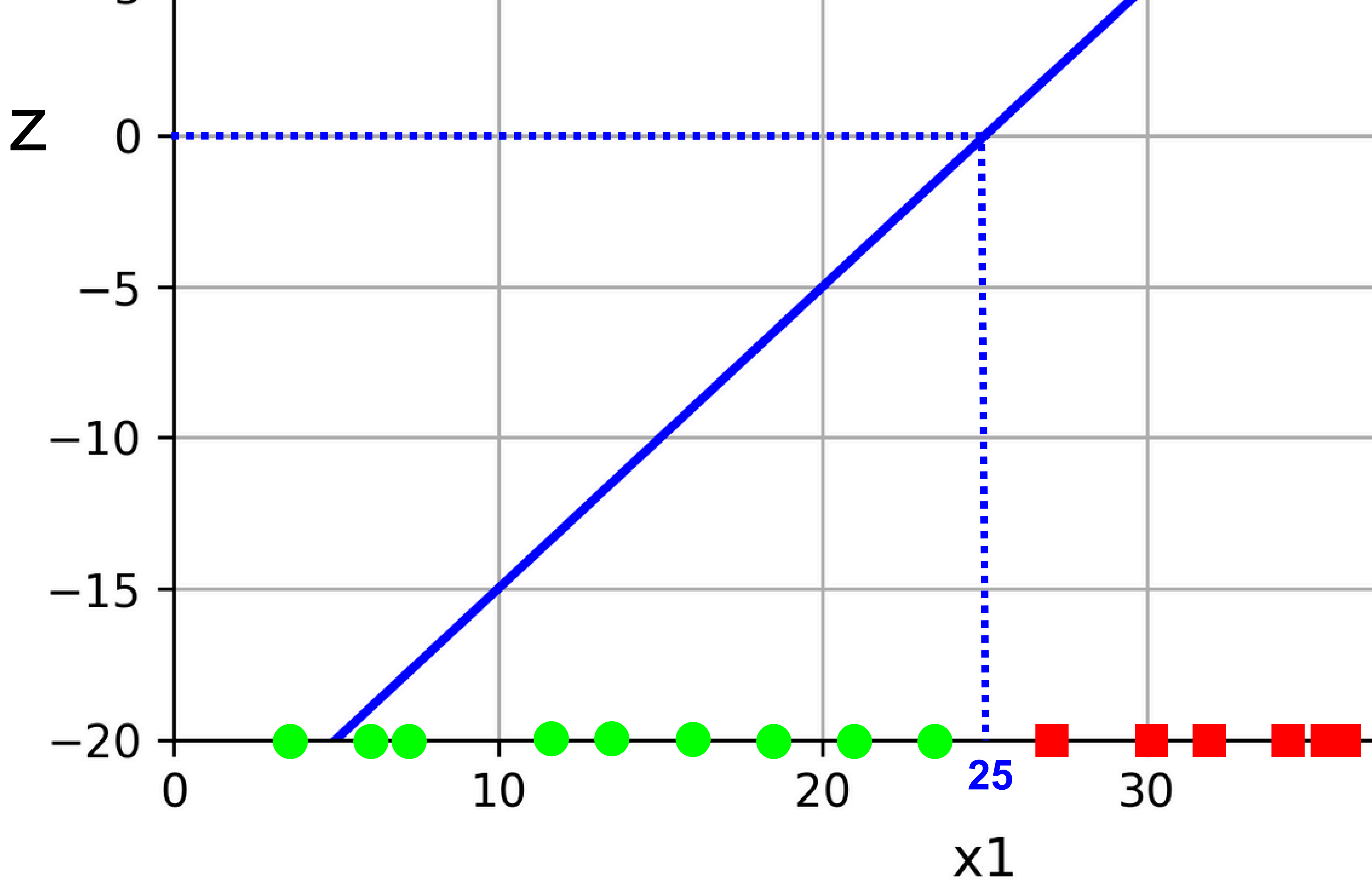
$$z(x_1) = x_1 - 25$$



$$z(12) = 12 - 25 = -13$$

$$a(z(12)) = \frac{1}{1+e^{-z(12)}} = \frac{1}{1+e^{13}} \approx 2.26 \times 10^{-6}$$





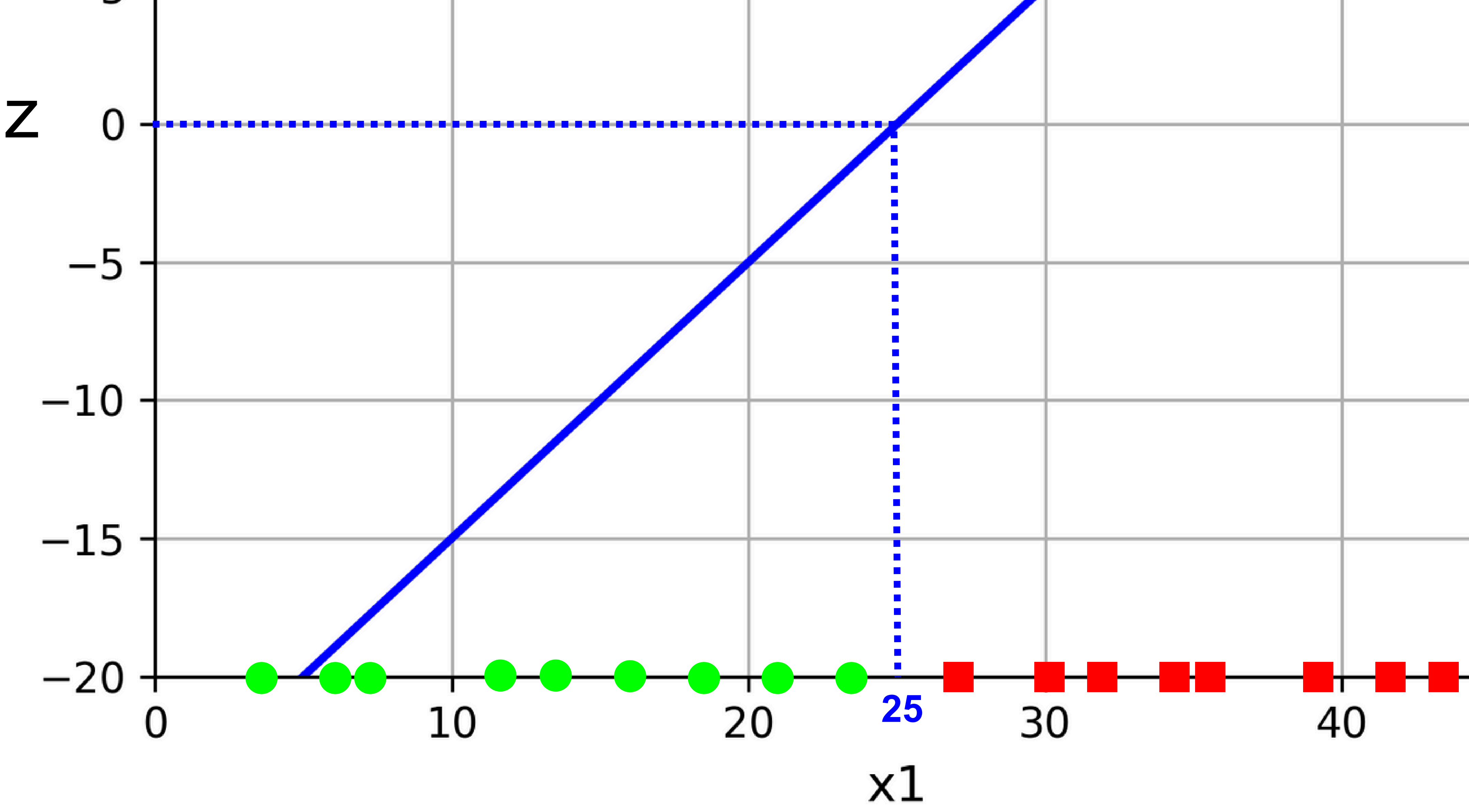
$$z\left(12\right) = 12 - 25 = -13$$

$$a\left(z\left(12\right)\right) = \frac{1}{1 + e^{-z(12)}} = \frac{1}{1 + e^{13}} \approx 2.26 \times 10^{-6}$$

$$z(x_1) = x_1 - 25$$

$$z(12) = 12 - 25 = -13$$

$$a(z(12)) \approx 2.26 \times 10^{-6}$$



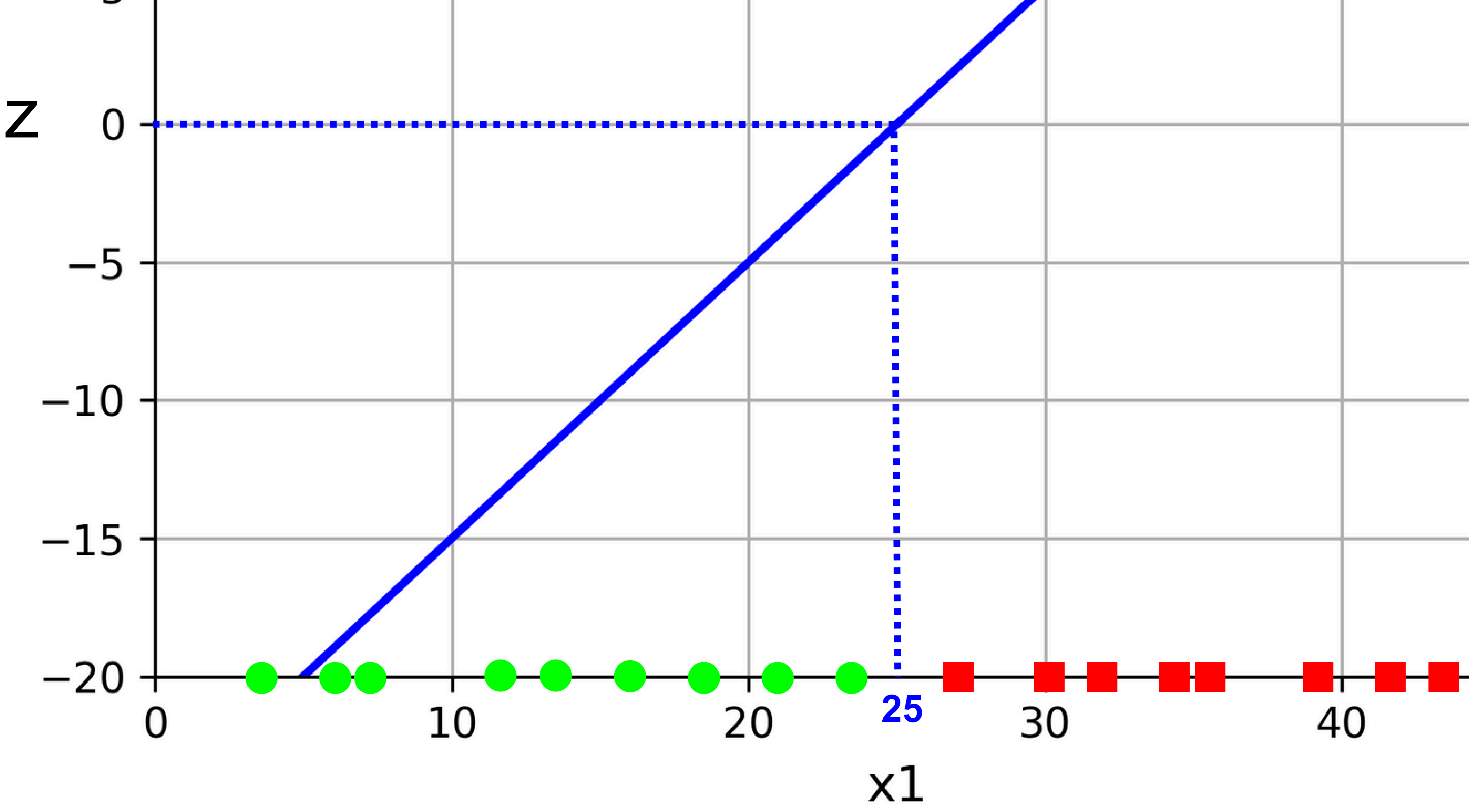
Since example (3) belongs to class 0, its probability calculated with Bernoulli's law is:

$$P(Y = 0) = a(12)^0 \times (1 - a(12))^{1-0}$$

$$z(x_1) = x_1 - 25$$

$$z(12) = 12 - 25 = -13$$

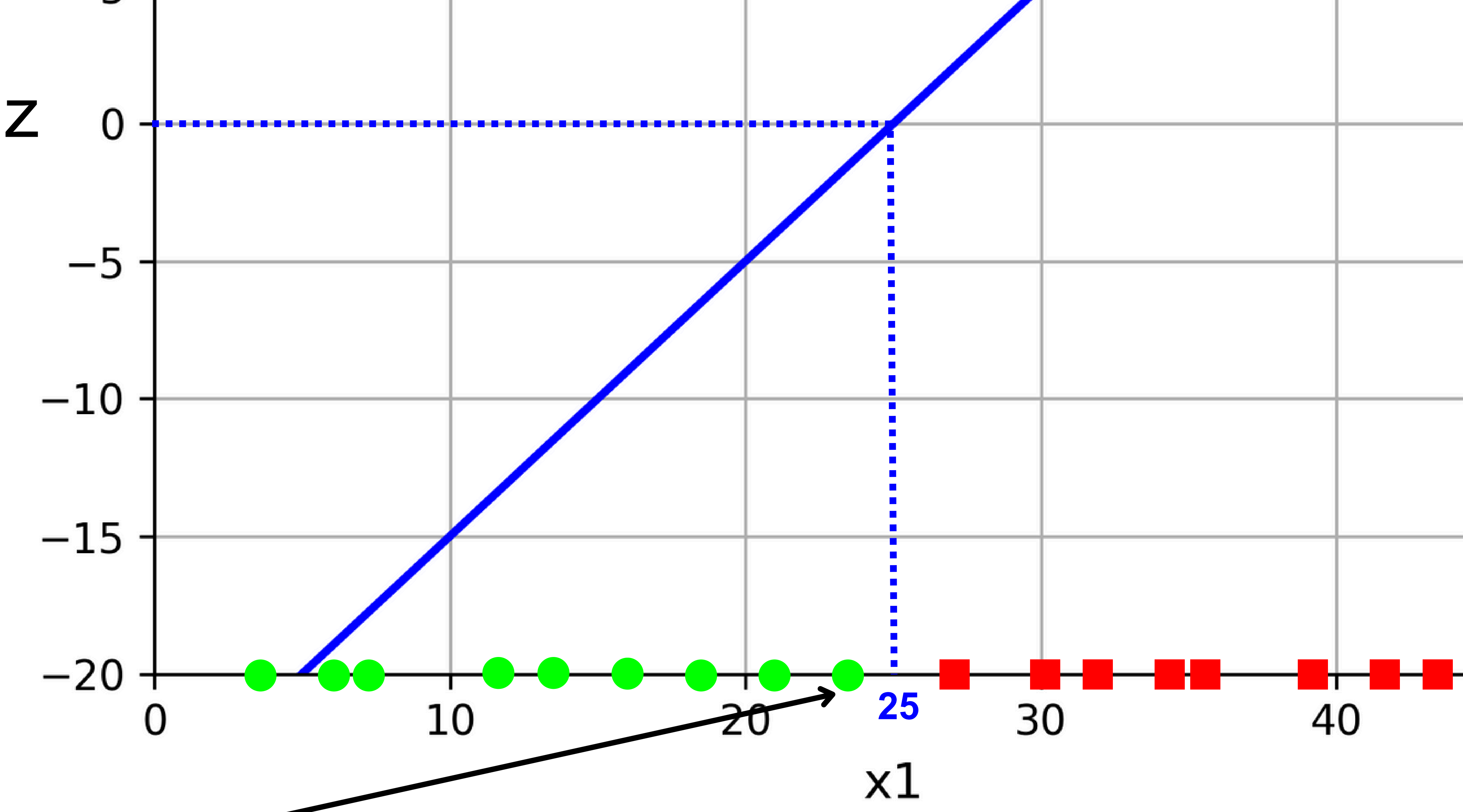
$$a(z(12)) \approx 2.26 \times 10^{-6}$$



$$P(Y=0) = a(12)^0 \times (1 - a(12))^{1-0}$$

$$P(Y=0) = (1 - a(12))^{1-0} = 1 - a(12) \approx 0.99$$

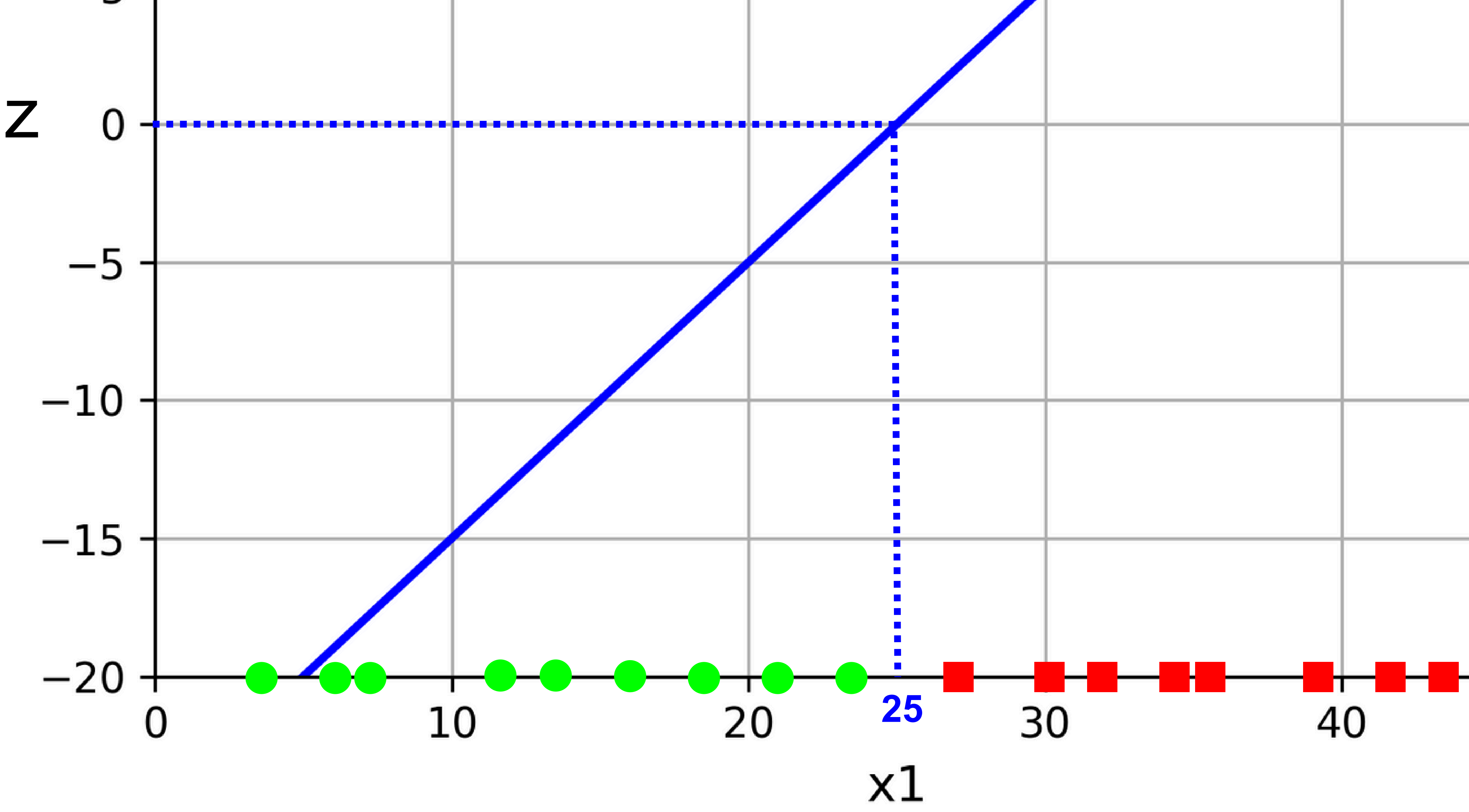
$$z(x_1) = x_1 - 25$$



example (8)

$$\begin{cases} \text{example (8)} \\ x_1^{(8)} = 24 \\ y^{(8)} = 0 \end{cases}$$

$$\begin{cases} \text{example (8)} \\ x_1^{(8)} = 24 \\ y^{(8)} = 0 \end{cases}$$



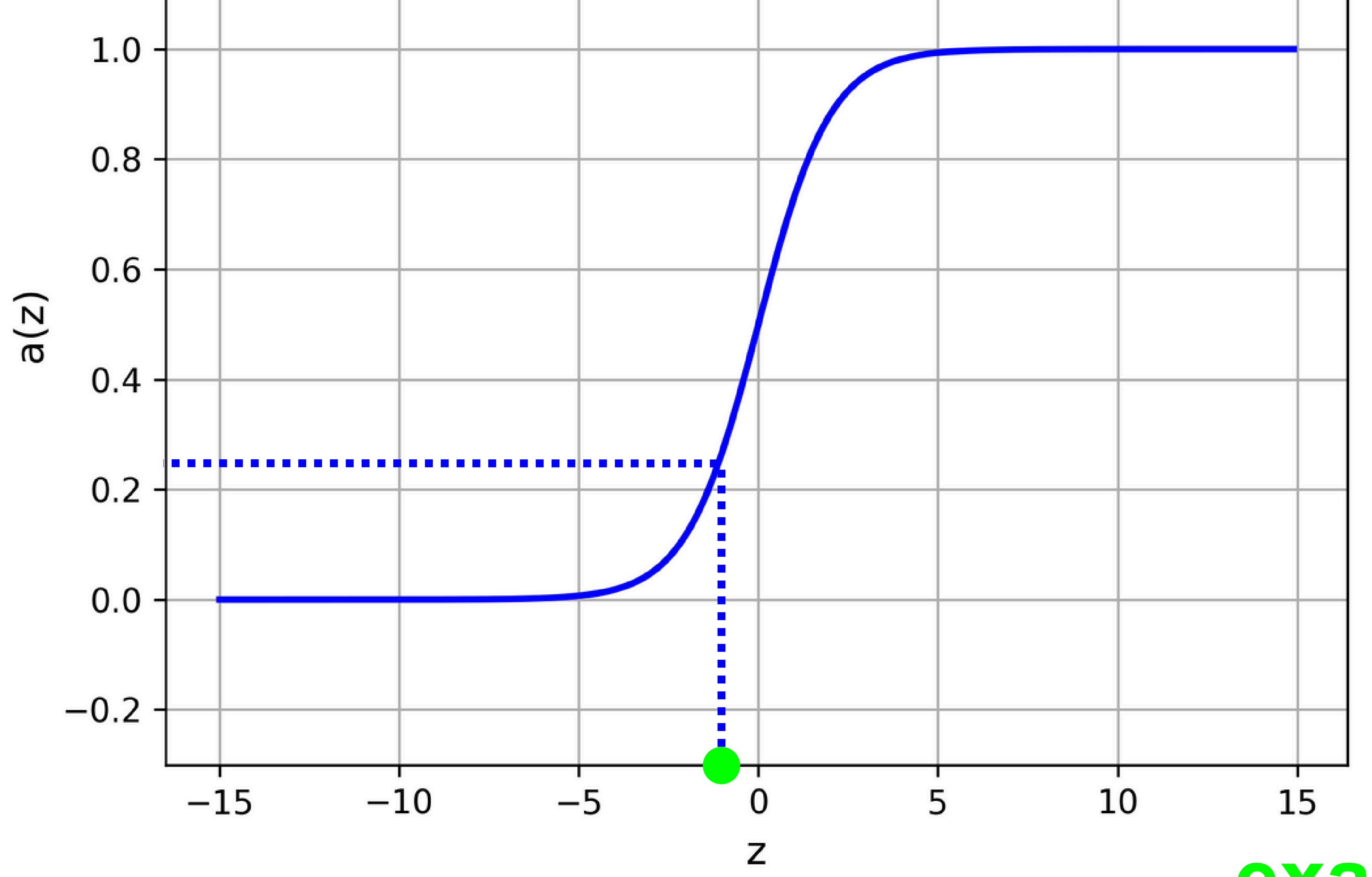
$$z(24) = 24 - 25 = -1$$

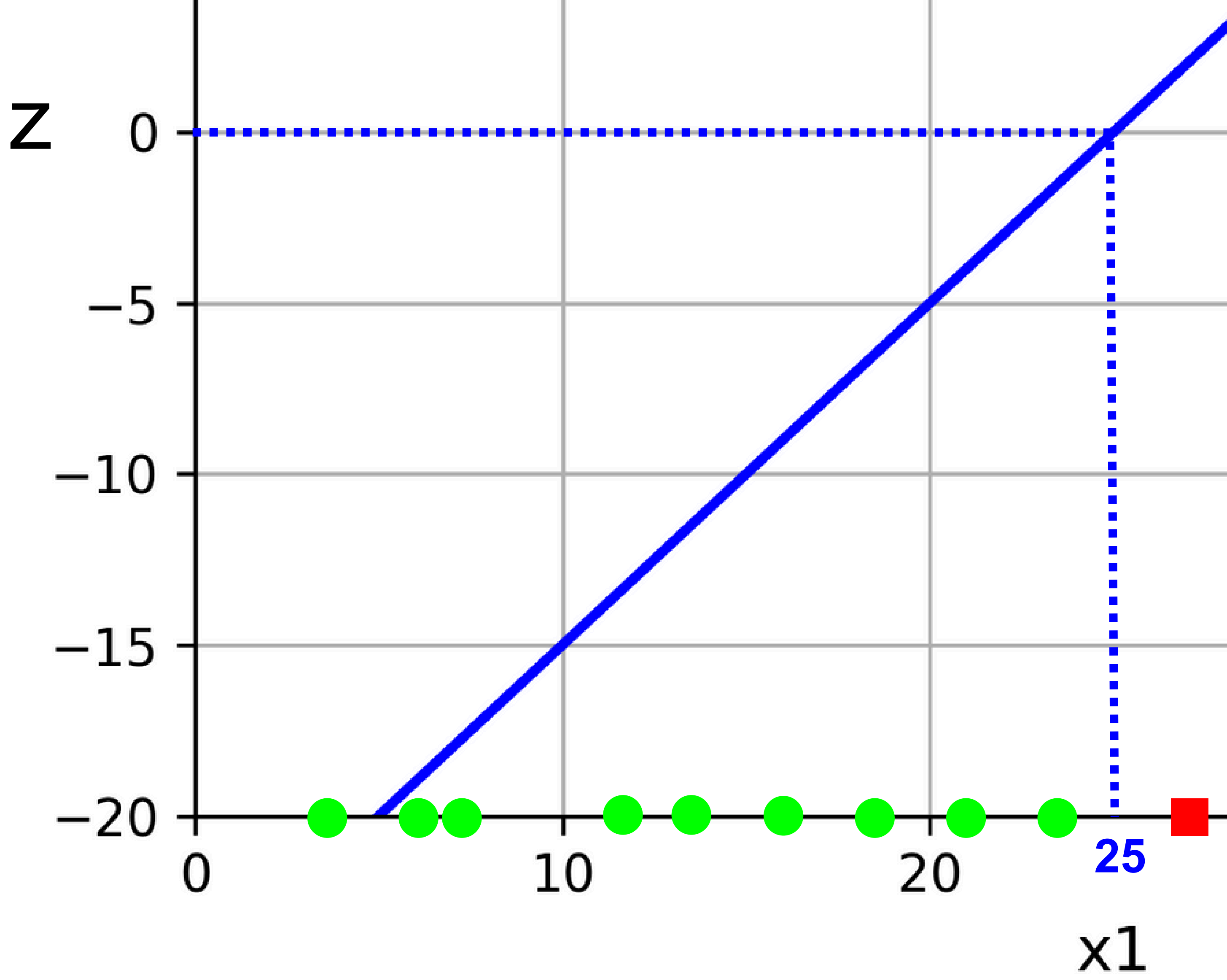
$$a(z(24)) \approx 0.27$$

$$P(Y=0) = a(24)^0 \times (1 - a(24))^{1-0}$$

The probabilty of example (8) belonging to class 0 is 0.73

$$P(Y=0) = (1 - a(24))^{1-0} = 1 - a(24) \approx 0.73$$





$$z\left(24\right) = 24 - 25 = -1$$

$$a\left(z\left(24\right)\right) \approx 0.27$$

example (8)

$$\begin{cases} x_1^{(8)} = 24 \\ y^{(8)} = 0 \end{cases}$$

The probabilty of example (8) belonging to class 0 is 0.73

$$P\left(Y = 0\right) = \left(1 - a\left(24\right)\right)^{1-0} = 1 - a\left(24\right) \approx 0.73$$

$$\begin{cases} \text{example (6)} \\ x_1^{(6)} = 24.8 \\ y^{(6)} = 0 \end{cases}$$

$$z(24.8) = 24.8 - 25 = -0.2$$

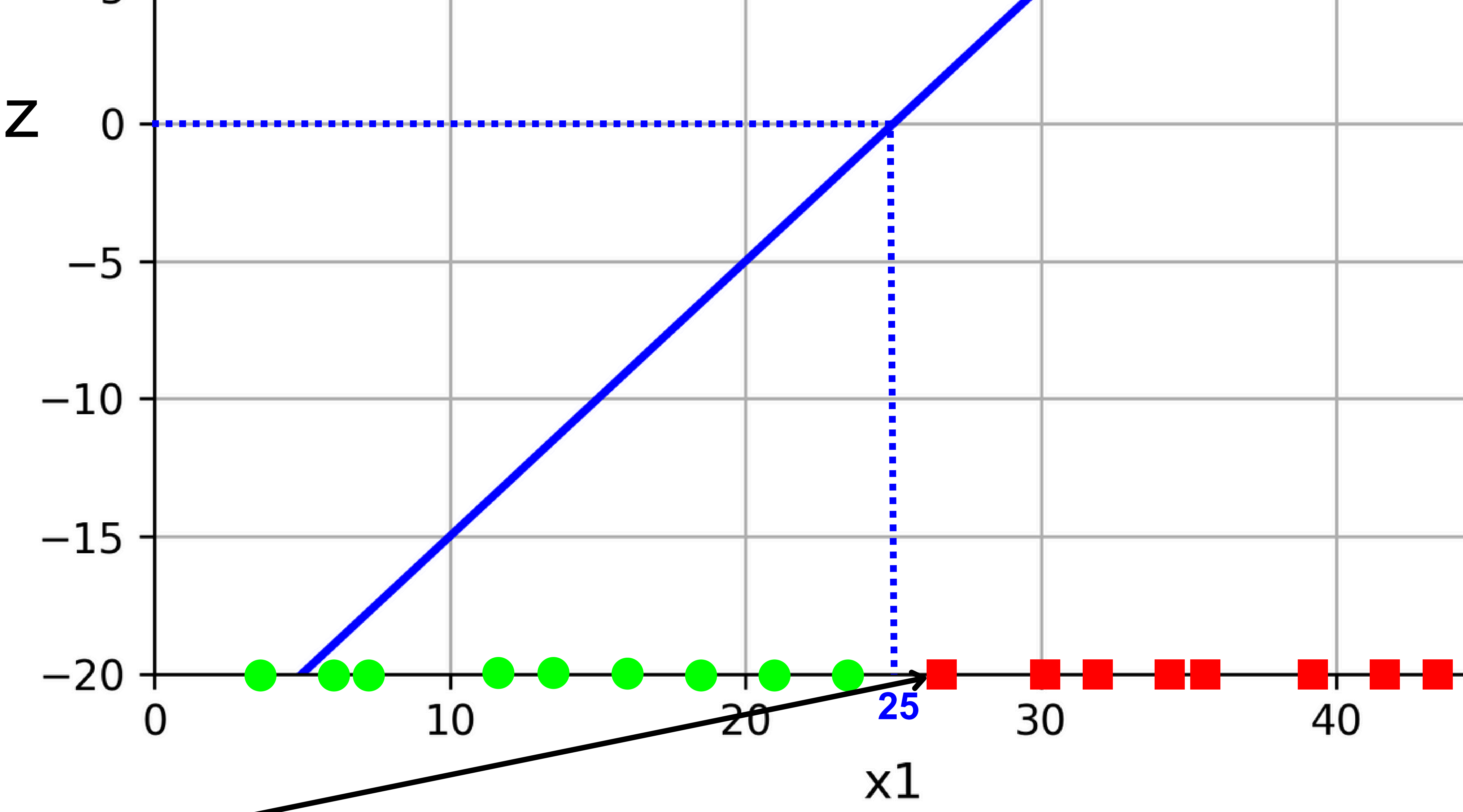
The probabilty of example (6) belonging to class 0 is 0.54

$$a(z(24.8)) \approx 0.45$$

$$P(Y=0) = a(24.8)^0 \times (1 - a(24.8))^{1-0}$$

$$P(Y=0) = (1 - a(24.8))^{1-0} = 1 - a(24.8) \approx 0.54$$

$$z\left(x_1\right) = x_1 - 25$$



$$\begin{cases} \textbf{example (11)} \\ x_1^{(11)} = 26 \\ y^{(11)} = 1 \end{cases}$$

$$\begin{cases} \text{example (11)} \\ x_1^{(11)} = 26 \\ y^{(11)} = 1 \end{cases}$$

$$z(26) = 26 - 25 = 1$$

$$a(z(26)) \approx 0.73$$

$$P(Y=1) = a(26)^1 \times (1 - a(26))^{1-1}$$

$$P(Y=1) = a(26)^1 \approx 0.73$$

The probabilty of example (11) belonging to class 1 is 0.73

Likelihood

Indicates the model's plausibility with respect to real data.

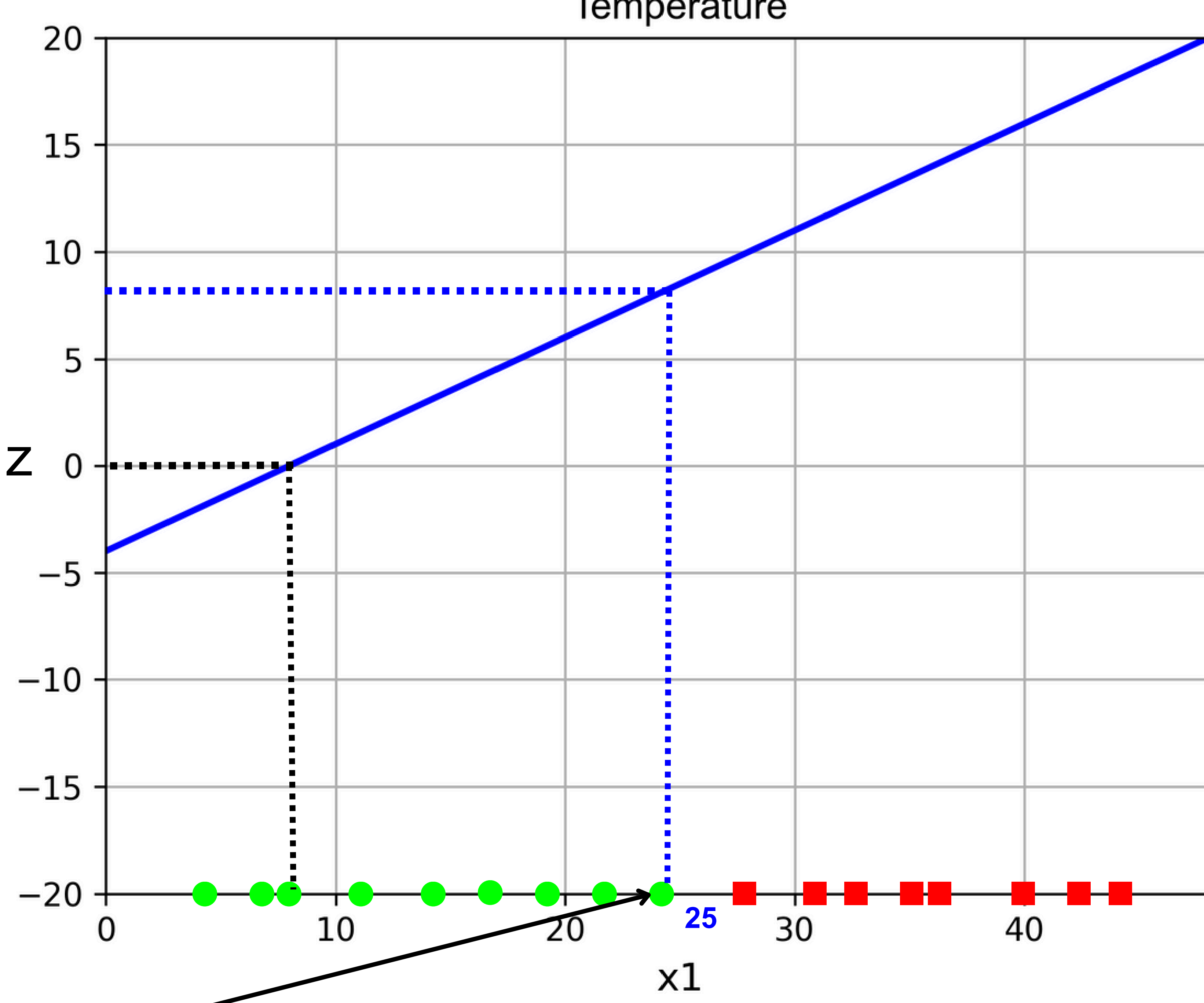
If the model predicts that the weather is hot with an 85% probability, then the model is considered 85% plausible for this prediction.

To calculate the Likelihood of our model, we will take the product of all predicted probabilities.

Likelihood

$$z(x_1) = (0.5)x_1 - 4$$

$$z(24) = 0.5(24) - 4 = 8$$



example (8)

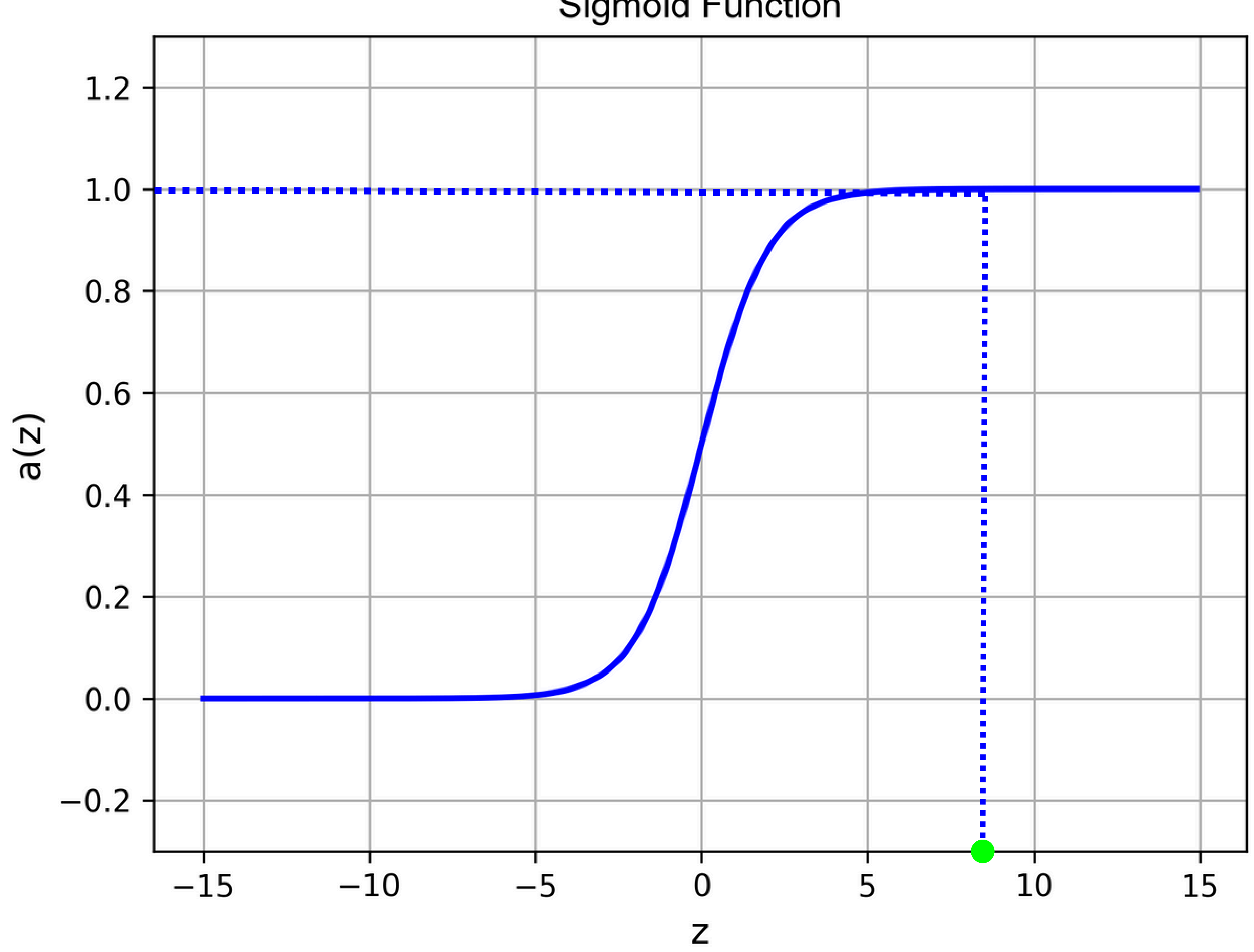
$$\begin{cases} x_1^{(8)} = 24 \\ y^{(8)} = 0 \end{cases}$$

Likelihood

$$z(x_1) = (0.5)x_1 - 4$$

$$z(24) = 0.5(24) - 4 = 8$$

$$a(z(24)) = a(8) \approx 0.99$$



$$P(Y=0) = (1 - a(24))^{1-0} = 1 - a(24) \approx 0.01$$

Likelihood

We calculate the Likelihood of our model with expression:

$$L = \prod_{i=1}^{m} P(Y = y_i)$$

$$L = \prod_{i=1}^{m} a_i^{y_i}(1 - a_i)^{1-y_i}$$