

Aprendizado de Máquina e Reconhecimento de Padrões



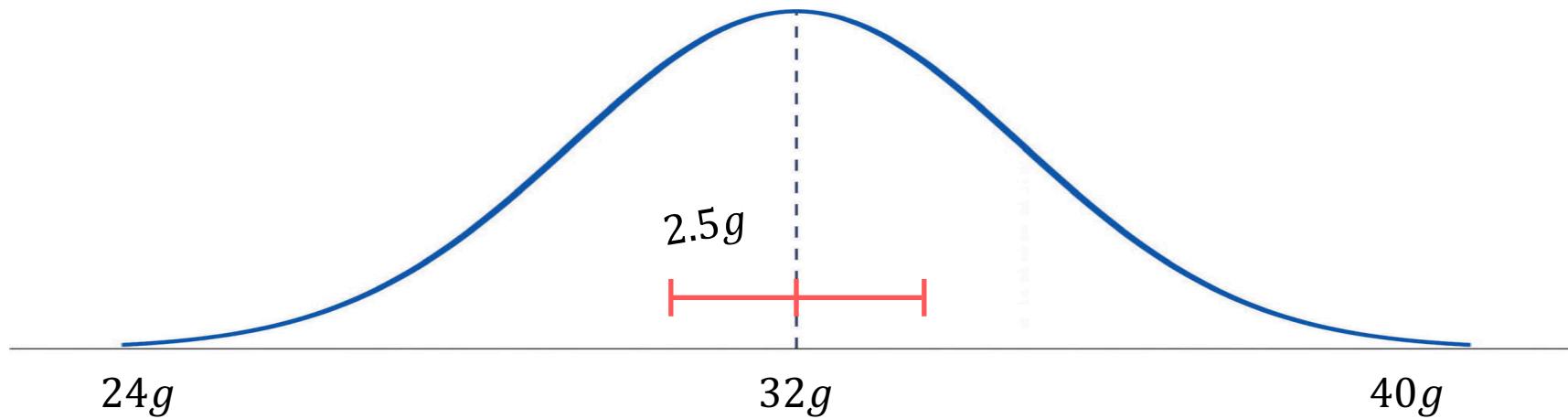
Naive Bayes

Prof. Dr. Samuel Martins (Samuka)
samuel.martins@ifsp.edu.br



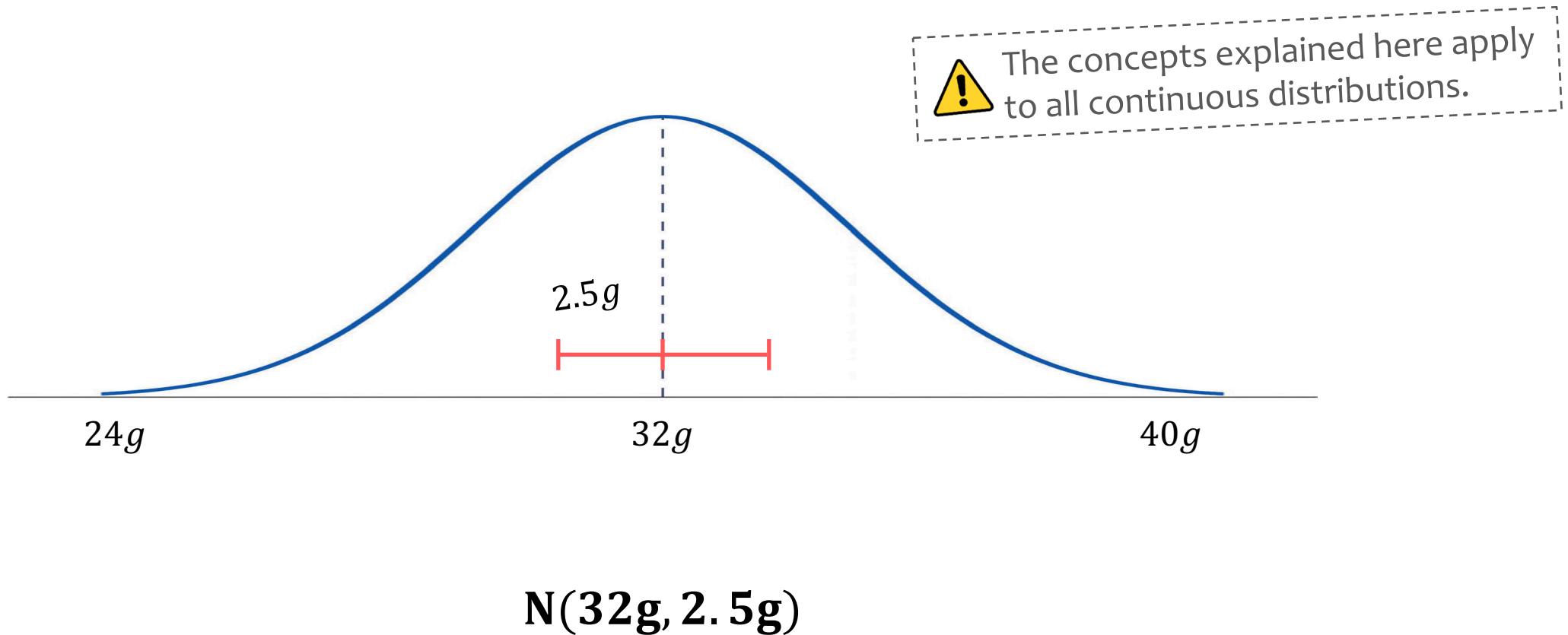
Preliminaries: Probability vs Likelihood

Let's assume that the weights of a mouse population follow a **Normal Distribution** $N(\mu, \sigma) = N(32g, 2.5g)$

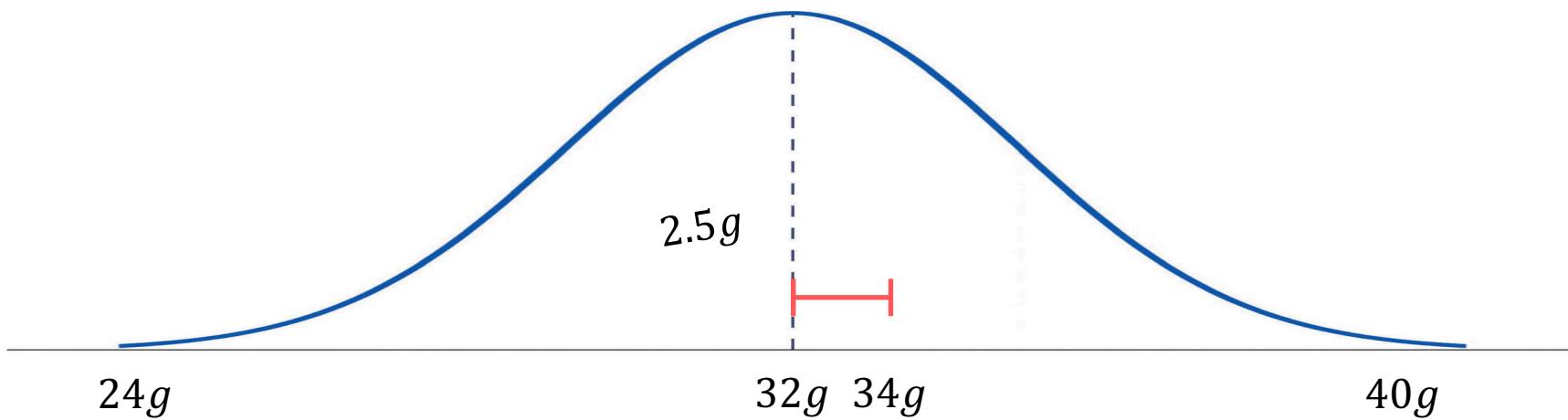


$$N(32g, 2.5g)$$

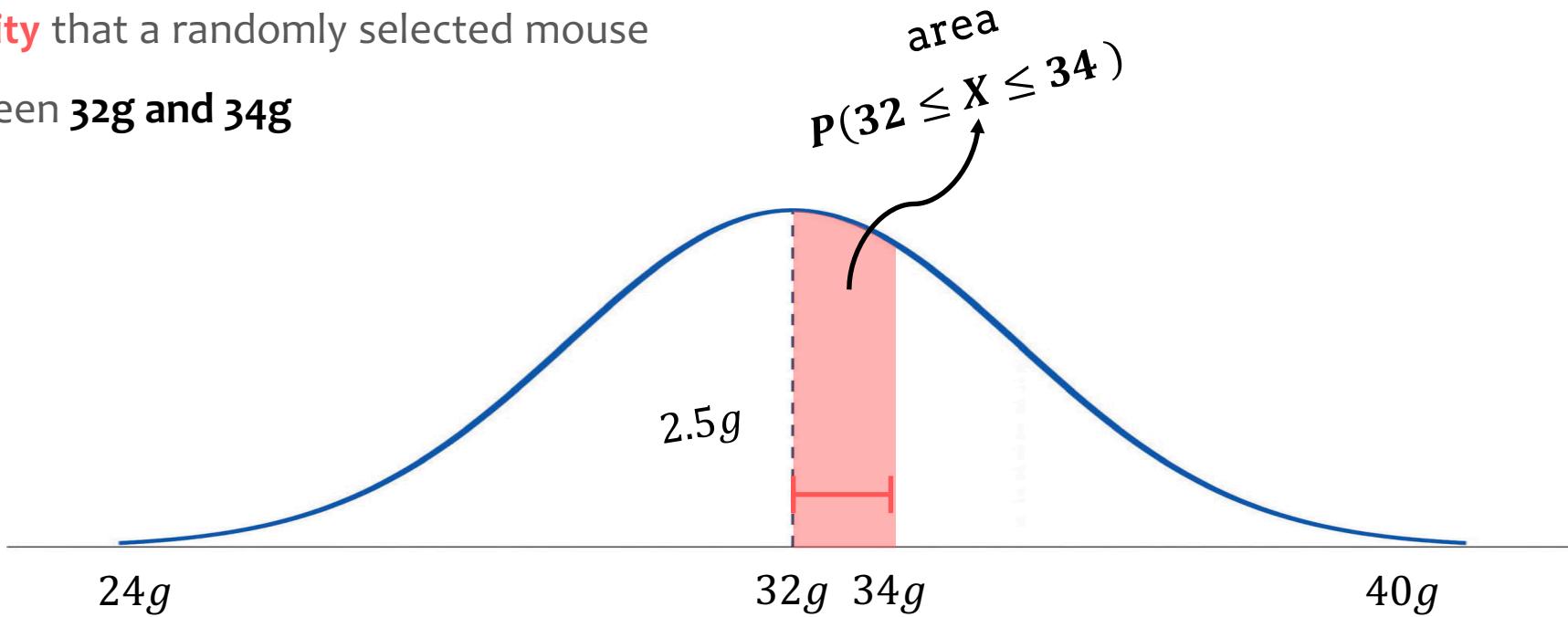
Let's assume that the weights of a mouse population follow a **Normal Distribution** $N(\mu, \sigma) = N(32g, 2.5g)$



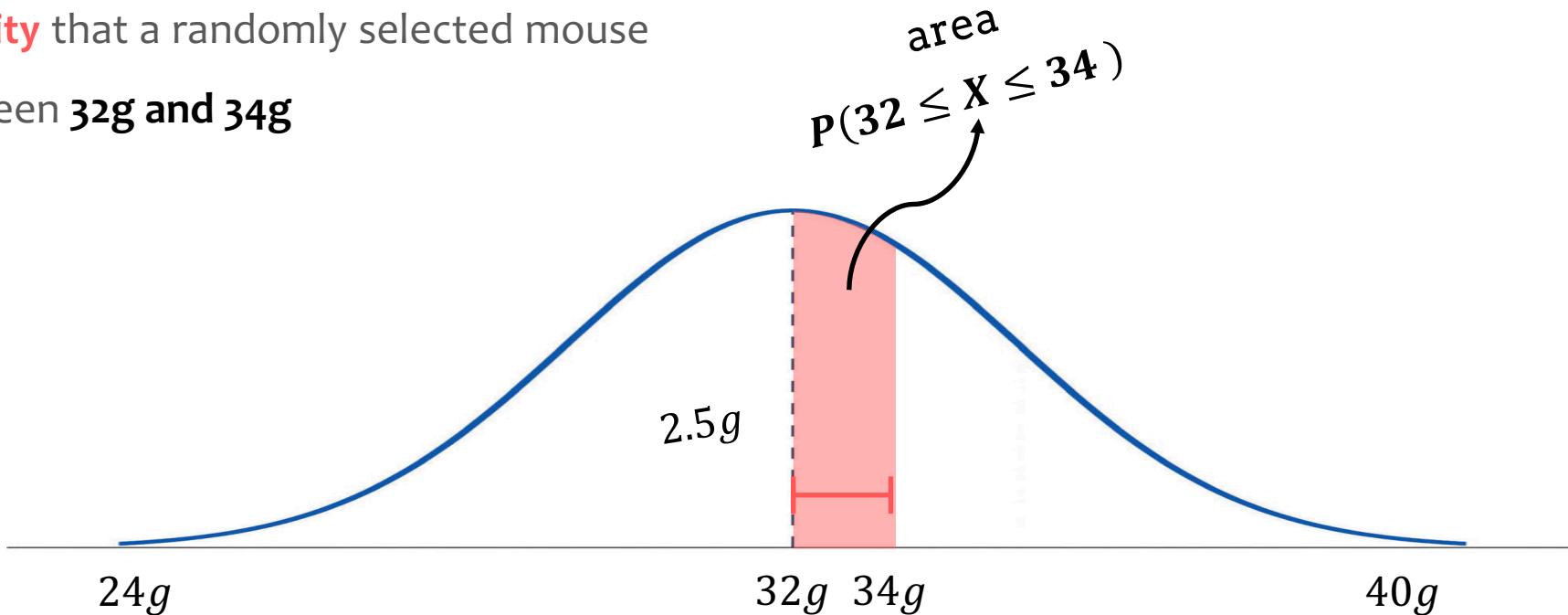
The **probability** that a randomly selected mouse
weighs between **32g and 34g**



The **probability** that a randomly selected mouse weighs between **32g and 34g**

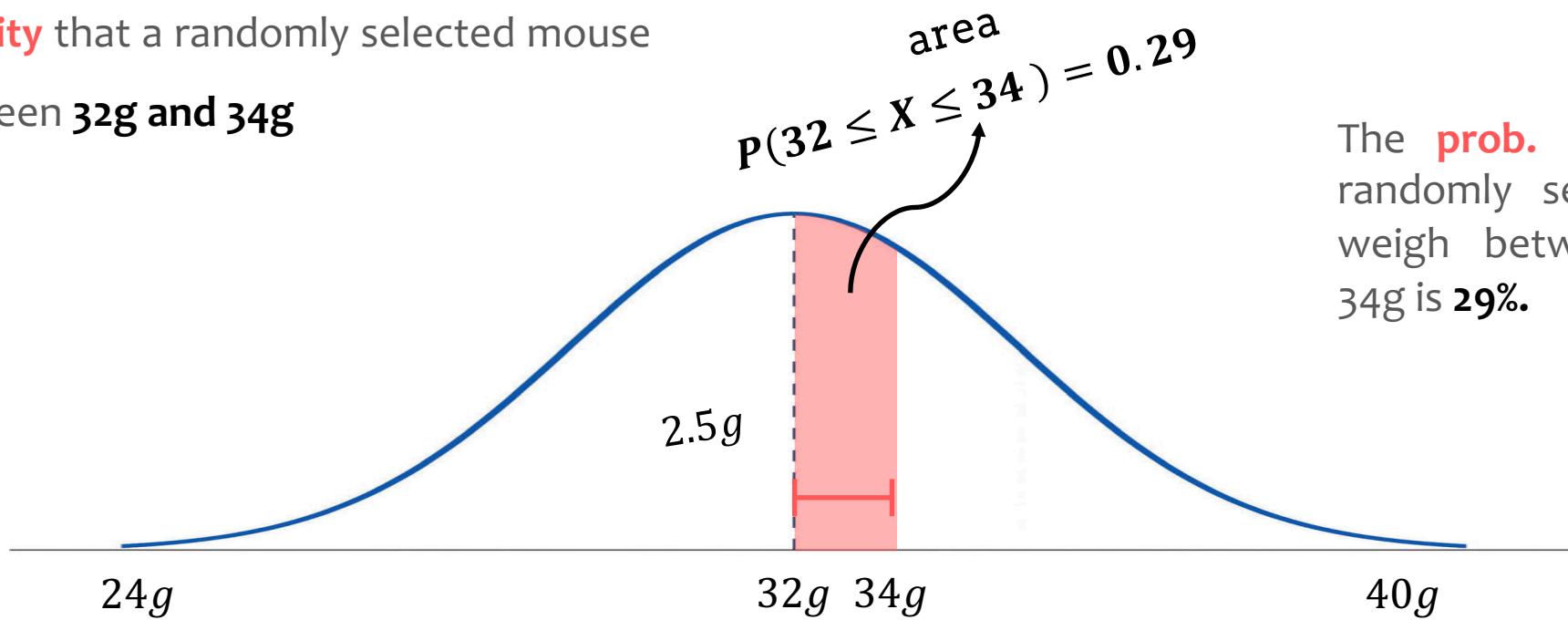


The **probability** that a randomly selected mouse weighs between **32g and 34g**



$$P(a \leq X \leq b) = \int_a^b f(x)dx = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

The **probability** that a randomly selected mouse weighs between **32g and 34g**



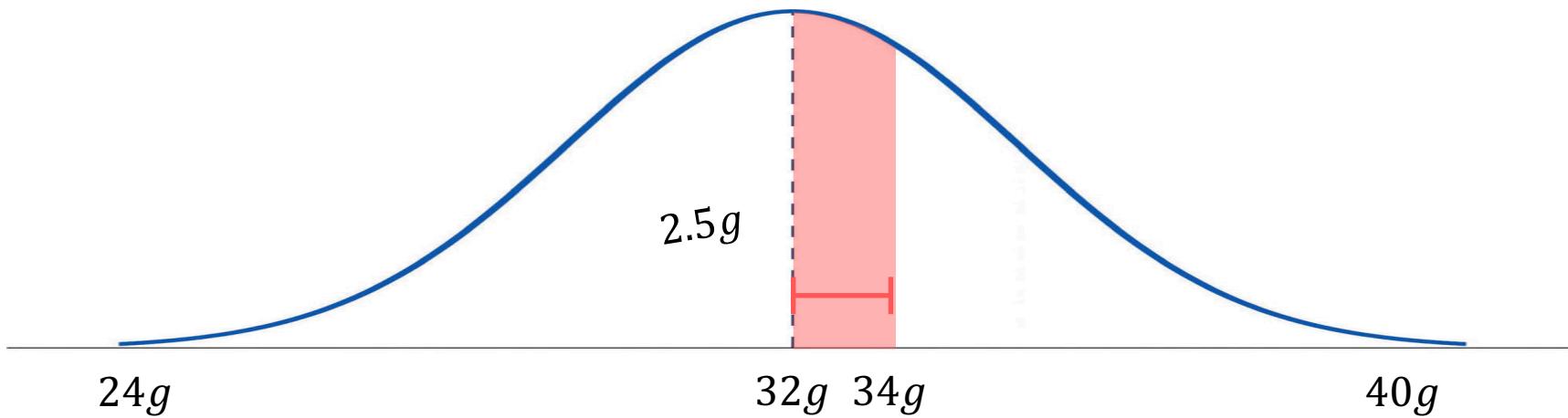
In this case, the **area under the curve** is **0.29**

The **prob. (chance)** of a randomly selected mouse weigh between 32g and 34g is **29%**.

$$P(a \leq X \leq b) = \int_a^b f(x)dx = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

Mathematically, we use the following notation:

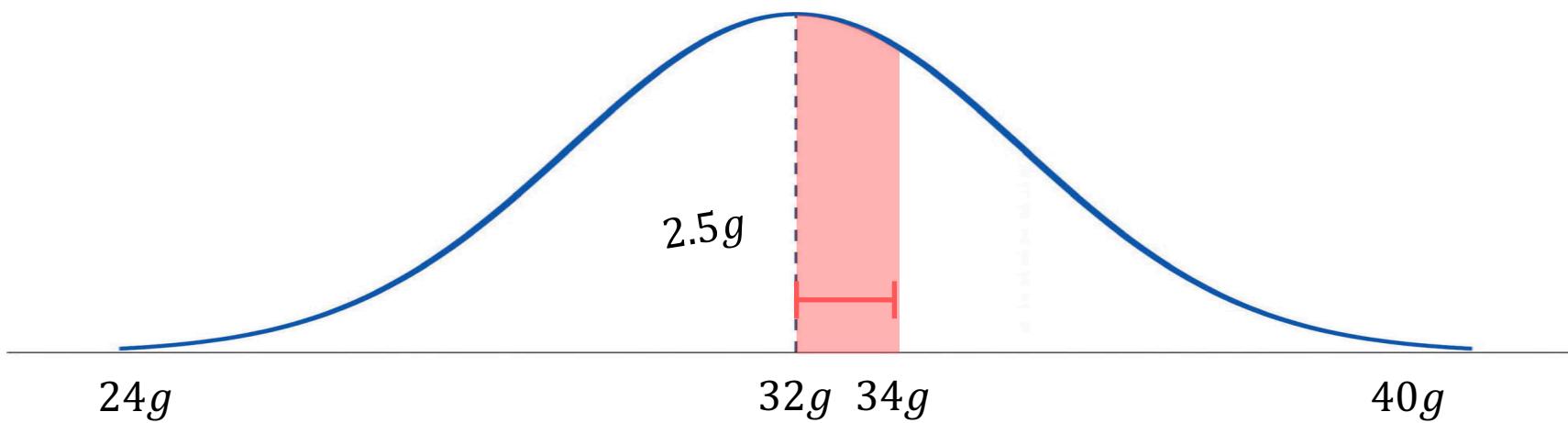
$$P(\text{weight between } 32g \text{ and } 34g \mid \text{mean} = 32g \text{ and std} = 2.5g) = 0.29$$



Mathematically, we use the following notation:

$$P(\text{weight between } 32g \text{ and } 34g \mid \text{mean} = 32g \text{ and std} = 2.5g) = 0.29$$

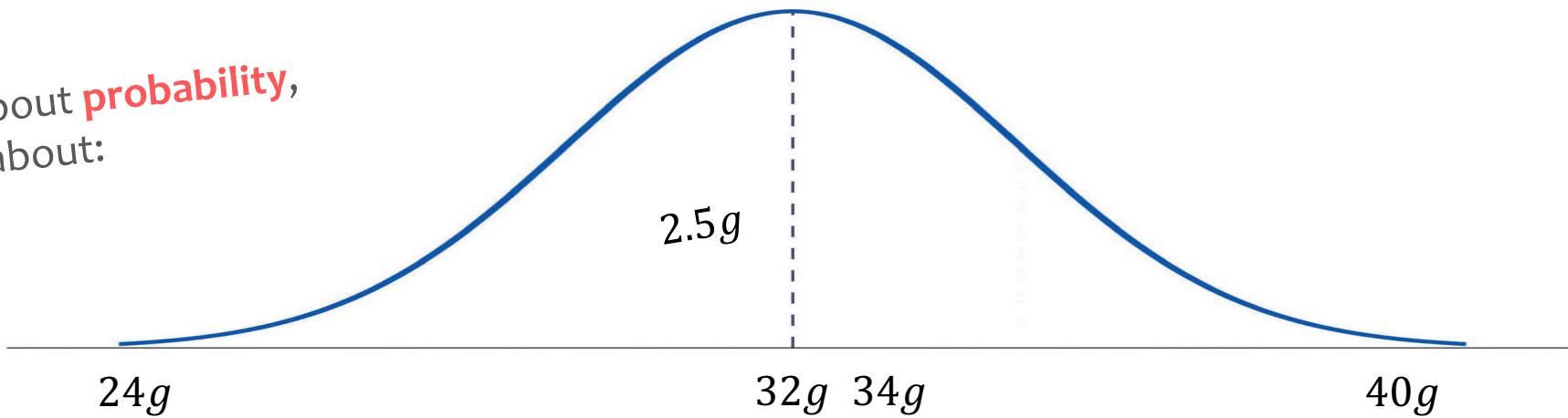
$$P(32 \leq X \leq 34 \mid \mu = 32, \sigma = 2.5) = 0.29$$



$$P(\text{weight between } 32g \text{ and } 34g \mid \text{mean} = 32g \text{ and std} = 2.5g) = 0.29$$

$$P(32 \leq X \leq 34 \mid \mu = 32, \sigma = 2.5) = 0.29$$

When talking about **probability**,
we are talking about:

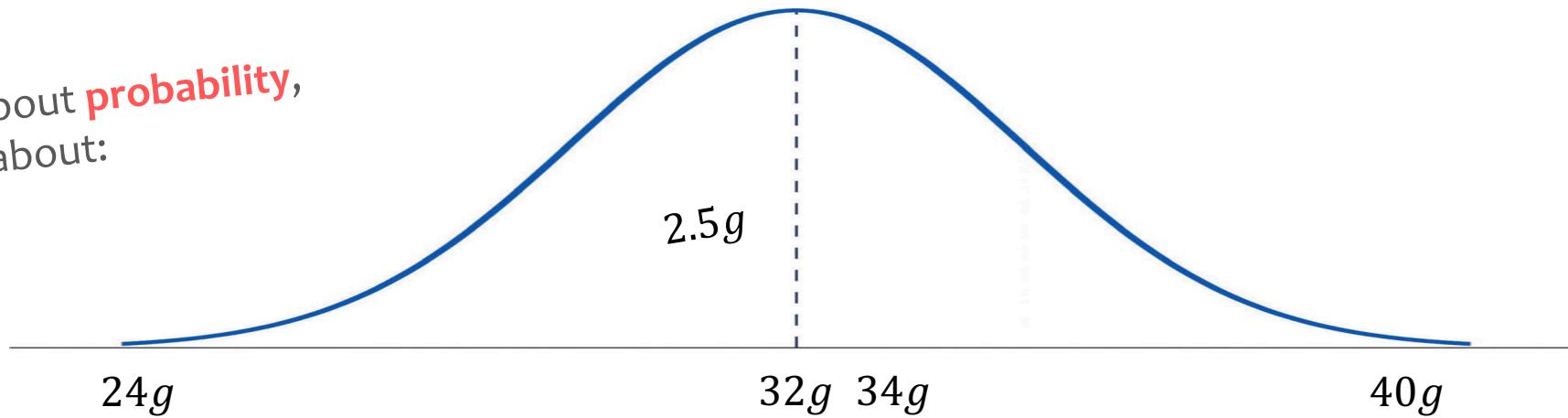


A **distribution** described by
this side of the equation.

$$P(\text{weight between } 32g \text{ and } 34g \mid \text{mean} = 32g \text{ and std} = 2.5g) = 0.29$$

$$P(32 \leq X \leq 34 \mid \mu = 32, \sigma = 2.5) = 0.29$$

When talking about **probability**,
we are talking about:

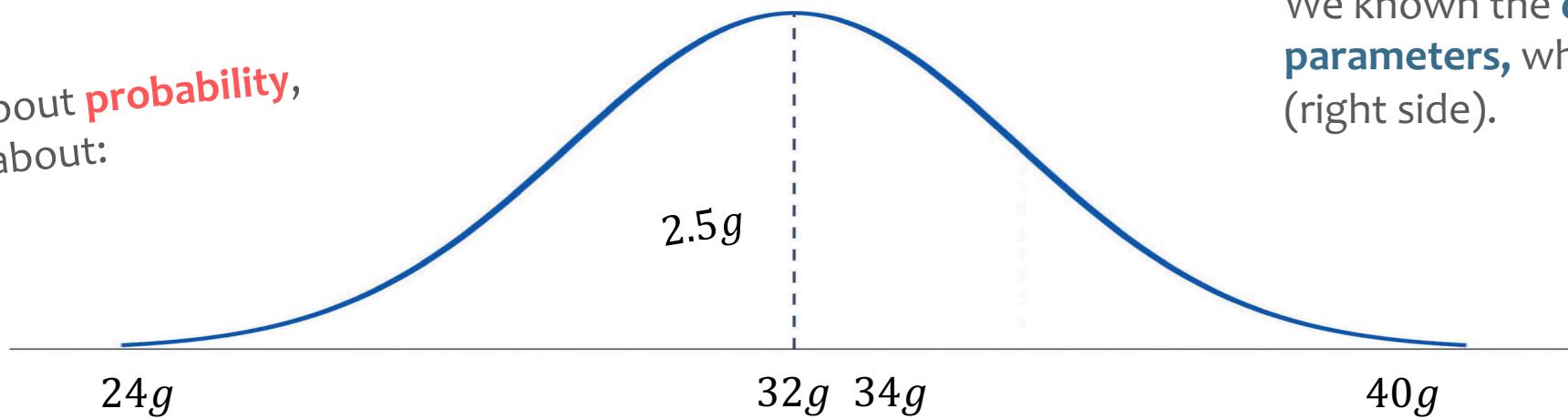


A **distribution** described by this side of the equation.

$$P(\text{weight between } 32g \text{ and } 34g \mid \text{mean} = 32g \text{ and std} = 2.5g) = 0.29$$

$$P(32 \leq X \leq 34 \mid \mu = 32, \sigma = 2.5) = 0.29$$

When talking about **probability**, we are talking about:



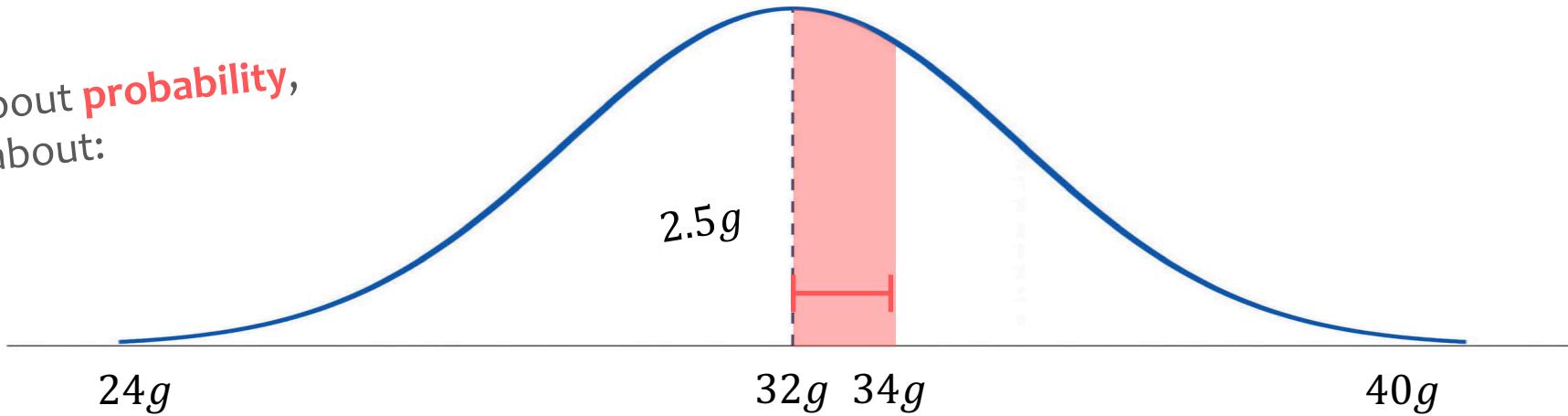
We know the **distribution parameters**, which are **fixed** (right side).

and the **area under the curve**,
described by this side of the equation.

$$P(\text{weight between } 32g \text{ and } 34g \mid \text{mean} = 32g \text{ and std} = 2.5g) = 0.29$$

$$P(32 \leq X \leq 34 \mid \mu = 32, \sigma = 2.5) = 0.29$$

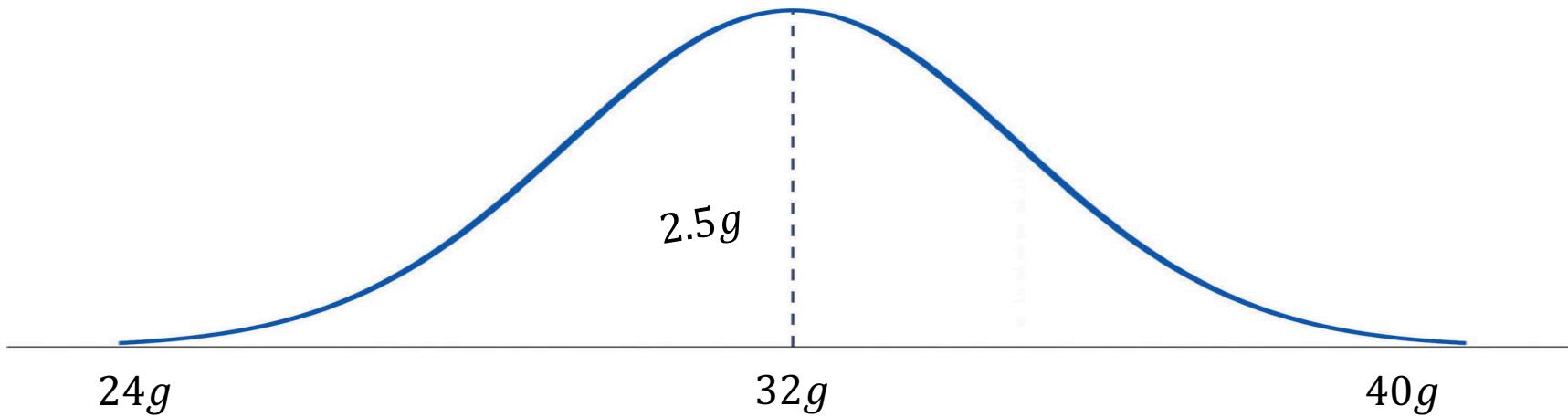
When talking about **probability**,
we are talking about:



Using the same **distribution**...

$$P(\text{weight between } 32g \text{ and } 34g \mid \text{mean} = 32g \text{ and std} = 2.5g) = 0.29$$

$$P(32 \leq X \leq 34 \mid \mu = 32, \sigma = 2.5) = 0.29$$

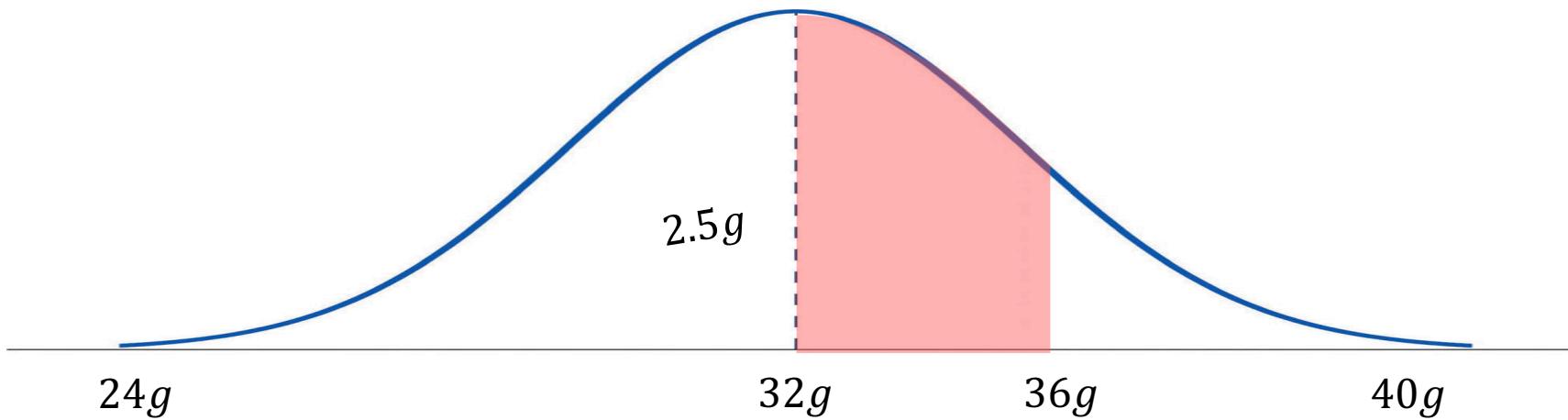


...we can change the left side of the equation to get a **new probability**!

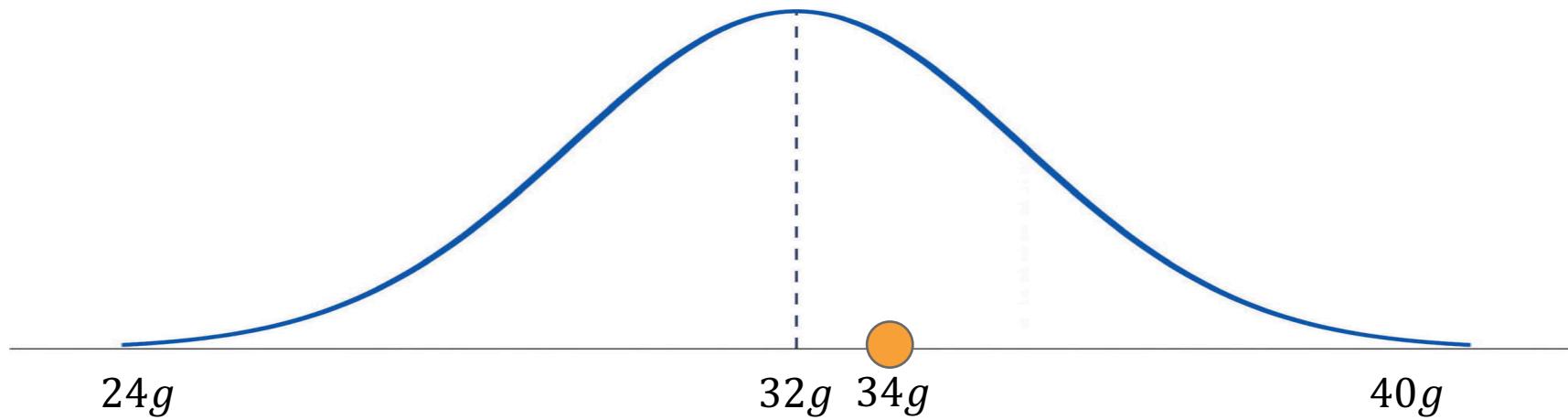
Using the same **distribution**...

$$P(\text{weight between } 32g \text{ and } 36g \mid \text{mean} = 32g \text{ and std} = 2.5g) = 0.44$$

$$P(32 \leq X \leq 36 \mid \mu = 32, \sigma = 2.5) = 0.44$$

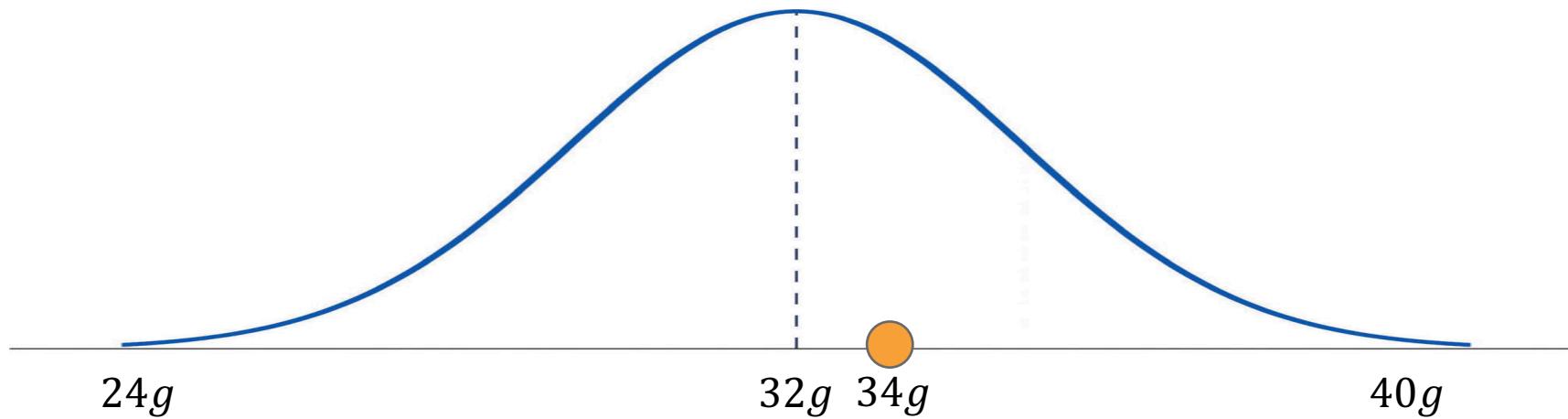


In case of **likelihood**, we assume that we already have selected the mouse (e.g., weighing 34g).

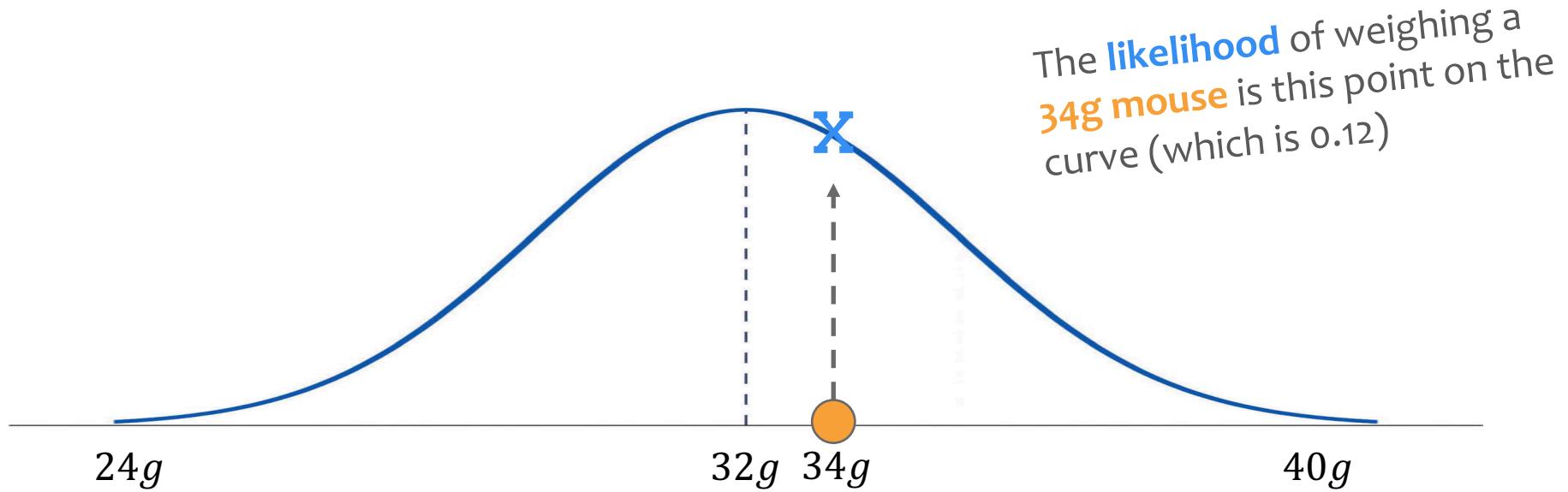


In case of **likelihood**, we assume that we already have selected the mouse (e.g., weighing 34g).

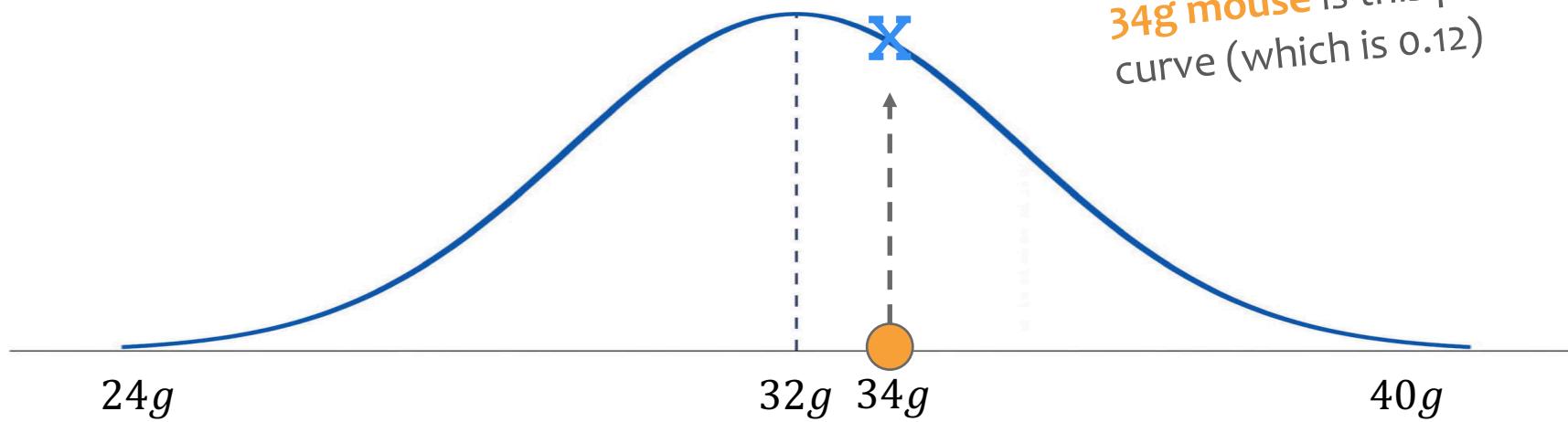
sample



In case of **likelihood**, we assume that we already have **selected the mouse** (e.g., weighing 34g).



In case of **likelihood**, we assume that we already have **selected the mouse** (e.g., weighing 34g).

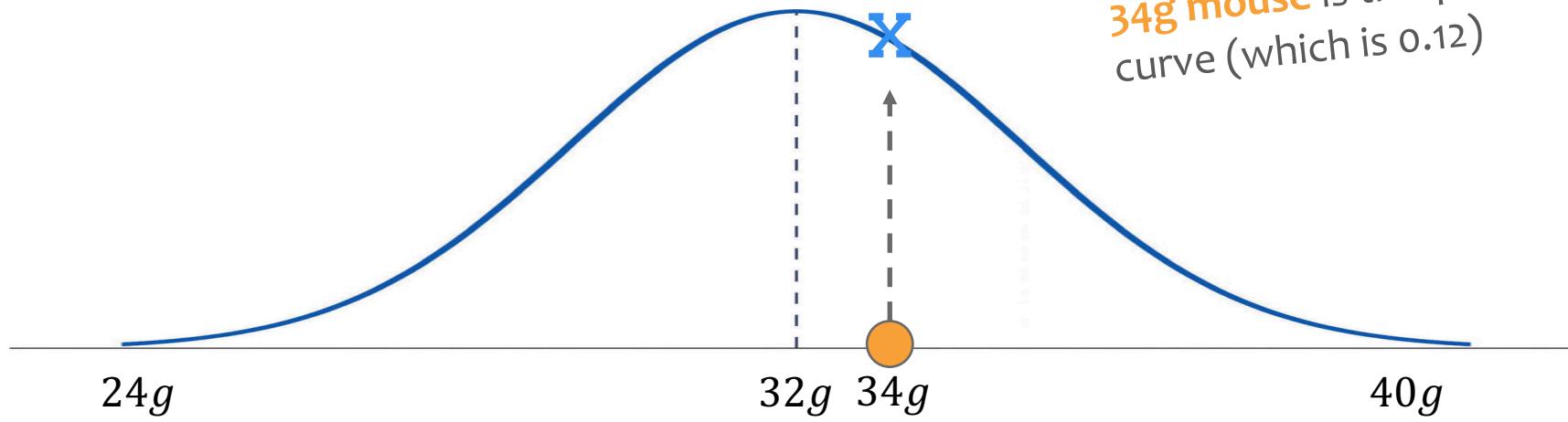


The **likelihood** of weighing a **34g mouse** is this point on the curve (which is 0.12)

We can think **likelihood** as:
“How **likely** (what is the probability) a **particular population**, described by a given **distribution (parameters)**, is to produce an **observed sample**? ”



In case of **likelihood**, we assume that we already have **selected the mouse** (e.g., weighing 34g).



The **likelihood** of weighing a **34g mouse** is this point on the curve (which is 0.12)

Note we **do not** know the **population/distribution**, only a **sample**.

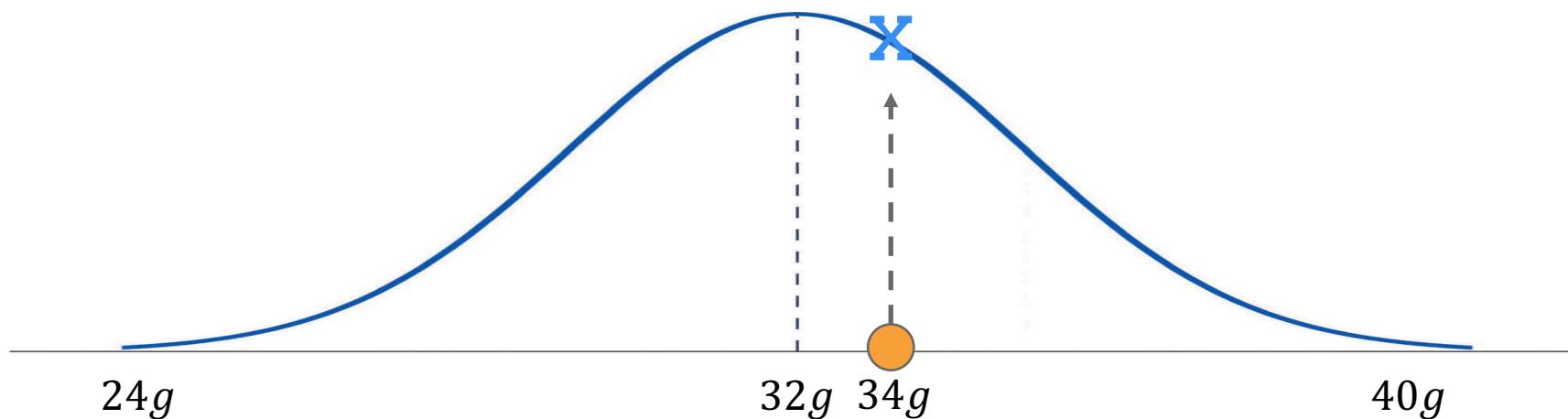
We can think **likelihood** as:
“How **likely** (what is the probability) a **particular population**, described by a given **distribution (parameters)**, is to produce an **observed sample**? ”

Mathematically, we use the following notation:

$$L(\text{mean} = 32g \text{ and std} = 2.5g \mid \text{mouse weighs } 34g) = 0.12$$

$$L(\mu = 32, \sigma = 2.5 \mid x = 34) = 0.12$$

$$\theta = \mu, \sigma$$



likelihood function

$$L(\theta|x_i) = f_\theta(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$L(\mu, \sigma|x_i) = f_{\mu, \sigma}(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$



The **likelihood** is equal to the **probability density** of the observed outcome given a distribution.

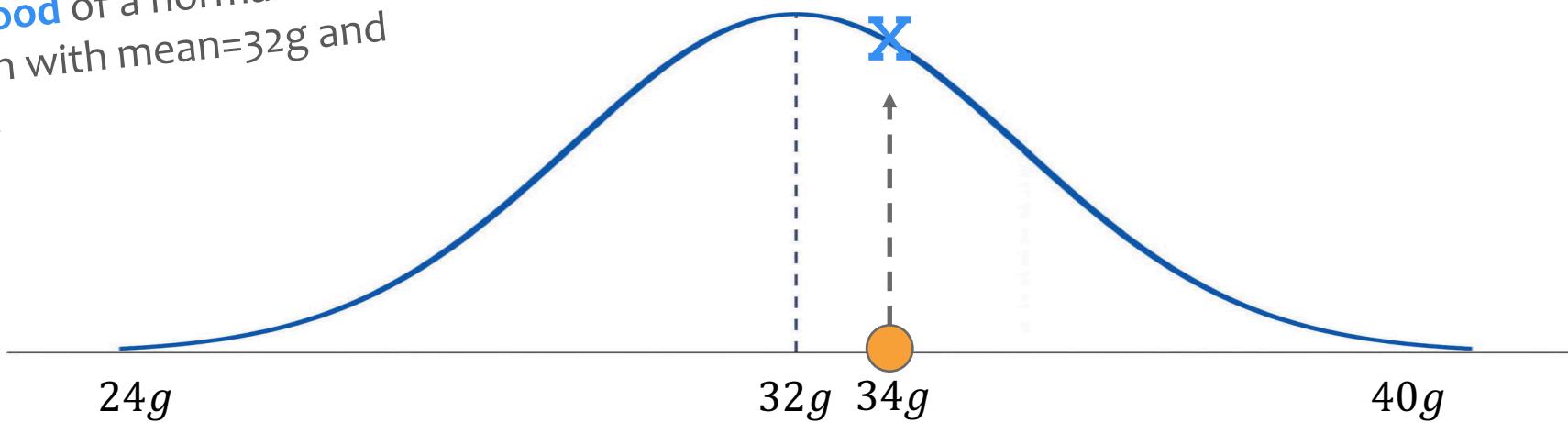
Mathematically, we use the following notation:

$$L(\text{mean} = 32g \text{ and std} = 2.5g | \text{mouse weighs } 34g) = 0.12$$

$$L(\mu = 32, \sigma = 2.5 | x = 34) = 0.12$$

$$\theta = \mu, \sigma$$

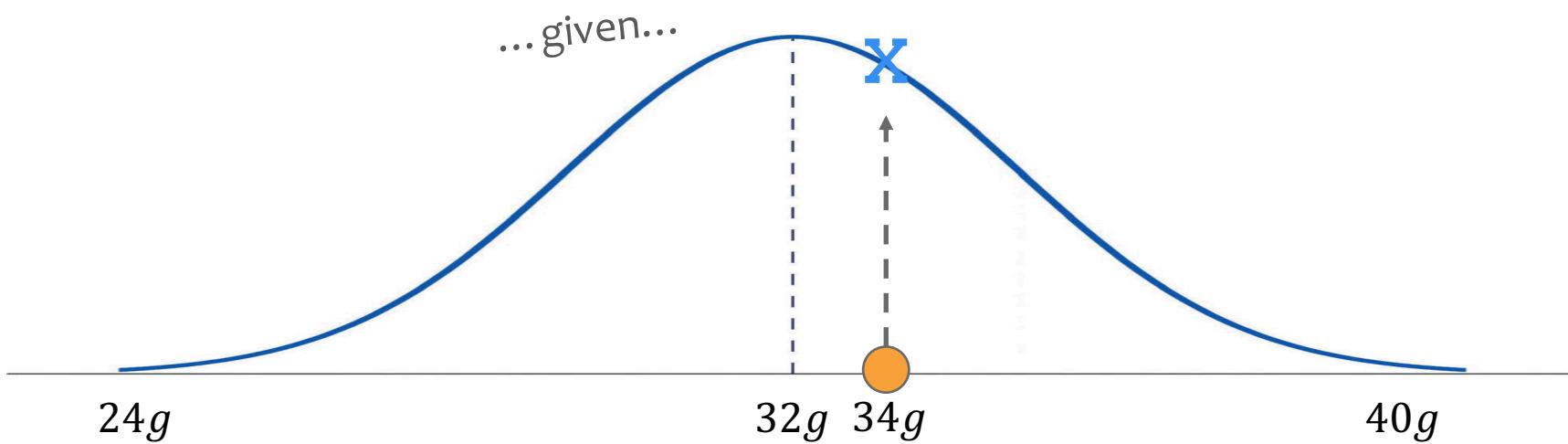
The **likelihood** of a normal distribution with mean=32g and std=2.5g...



Mathematically, we use the following notation:

$$L(\text{mean} = 32g \text{ and } \text{std} = 2.5g \mid \text{mouse weighs } 34g) = 0.12$$

$$L(\mu = 32, \sigma = 2.5 \mid x = 34) = 0.12$$

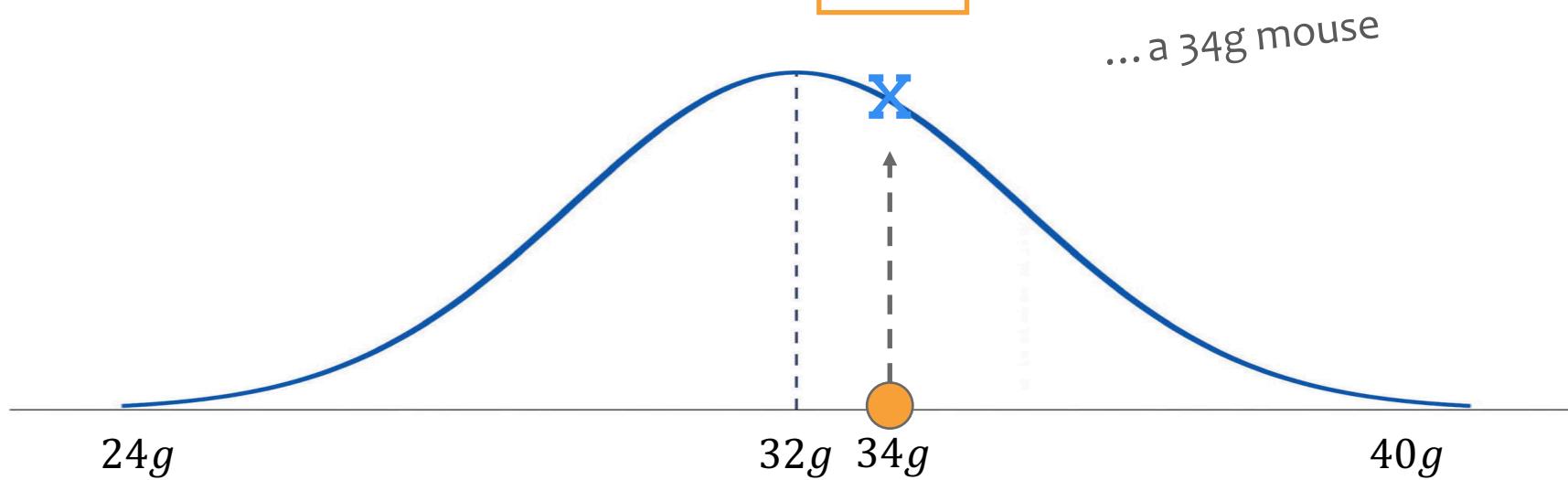


Mathematically, we use the following notation:

$$L(\text{mean} = 32g \text{ and std} = 2.5g \mid \text{mouse weighs } 34g) = 0.12$$

$$L(\mu = 32, \sigma = 2.5 \mid x = 34) = 0.12$$

$$\theta = \mu, \sigma$$

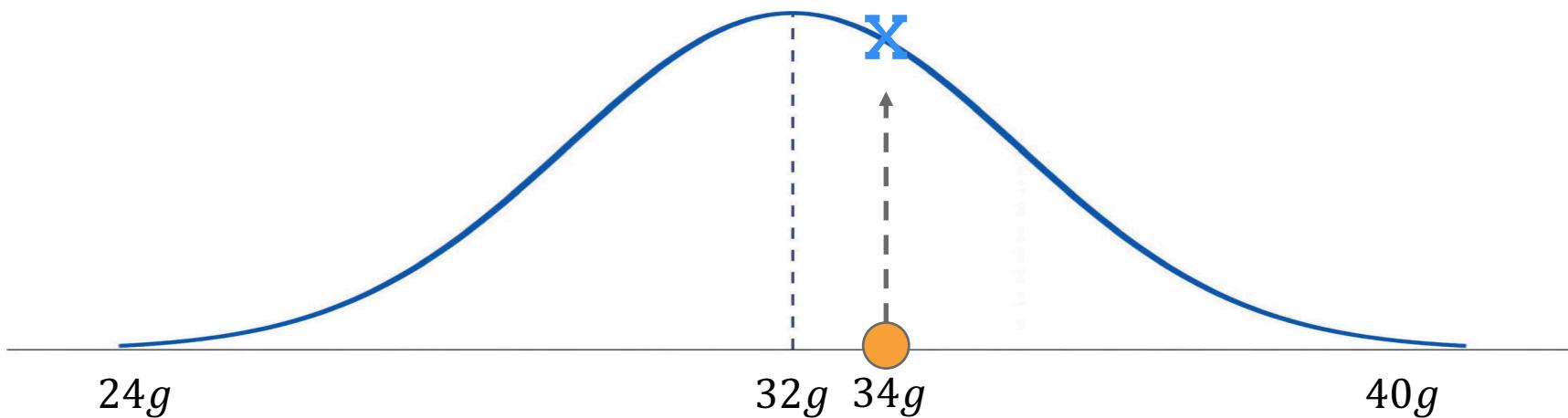


Mathematically, we use the following notation:

$$L(\text{mean} = 32g \text{ and std} = 2.5g \mid \text{mouse weighs } 34g) = 0.12$$

$$L(\mu = 32, \sigma = 2.5 \mid x = 34) = 0.12$$

$$\theta = \mu, \sigma$$

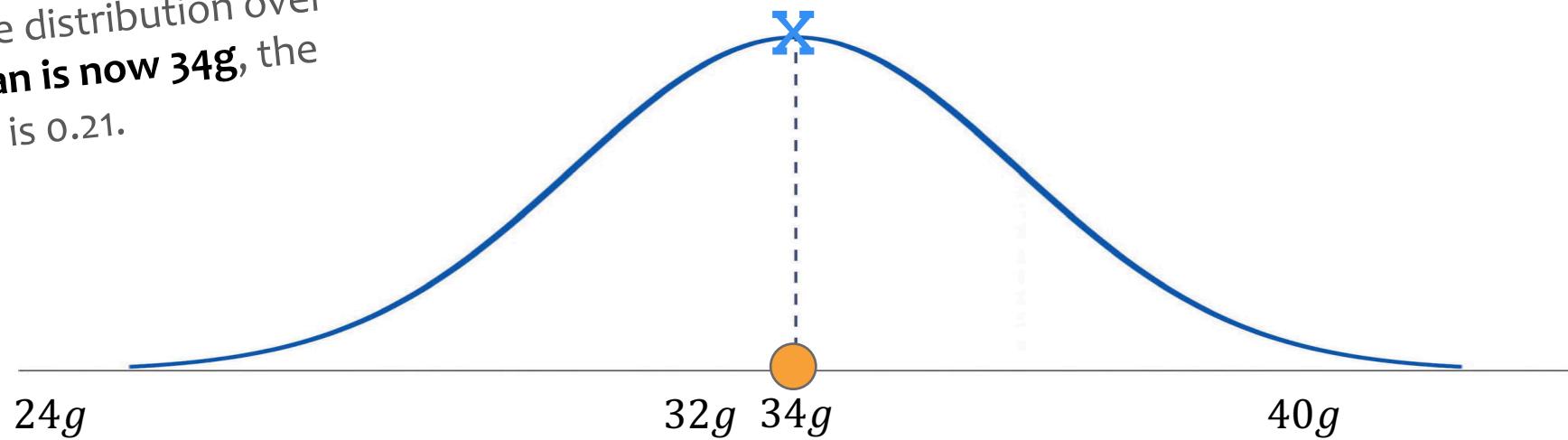


$$L(\text{mean} = 34g \text{ and std} = 2.5g | \text{mouse weighs } 34g) = 0.21$$

$$L(\mu = 34, \sigma = 2.5 | x = 34) = 0.21$$

$$\theta = \mu, \sigma$$

If we shifted the distribution over so that the **mean is now 34g**, the **new likelihood** is 0.21.

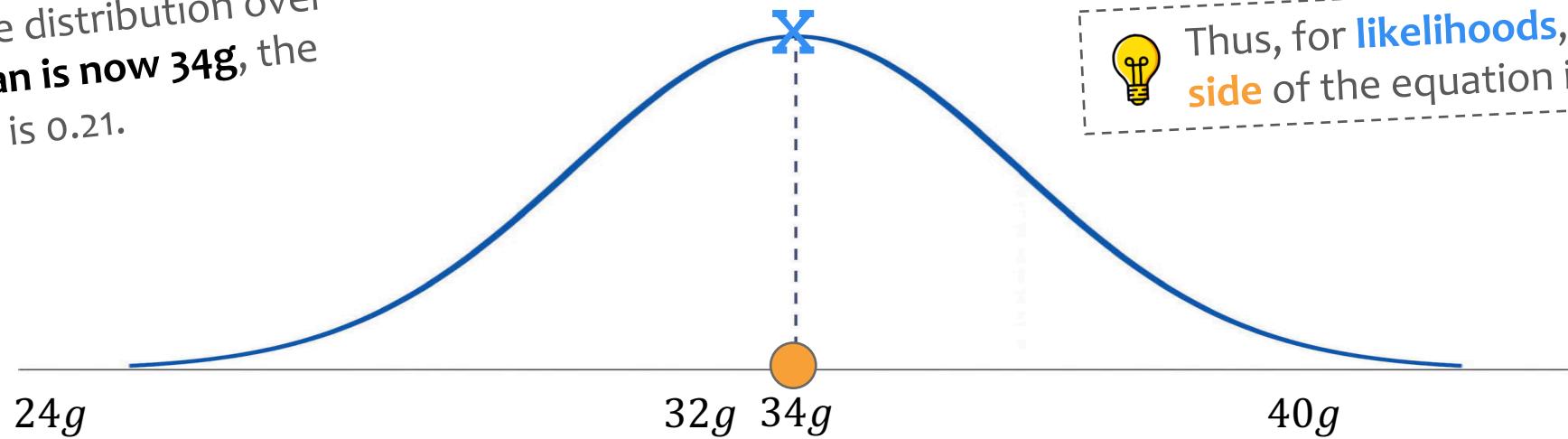


$$L(\text{mean} = 34g \text{ and std} = 2.5g \mid \text{mouse weighs } 34g) = 0.21$$

$$L(\mu = 34, \sigma = 2.5 \mid x = 34) = 0.21$$

$$\theta = \mu, \sigma$$

If we shifted the distribution over so that the **mean is now 34g**, the **new likelihood** is 0.21.



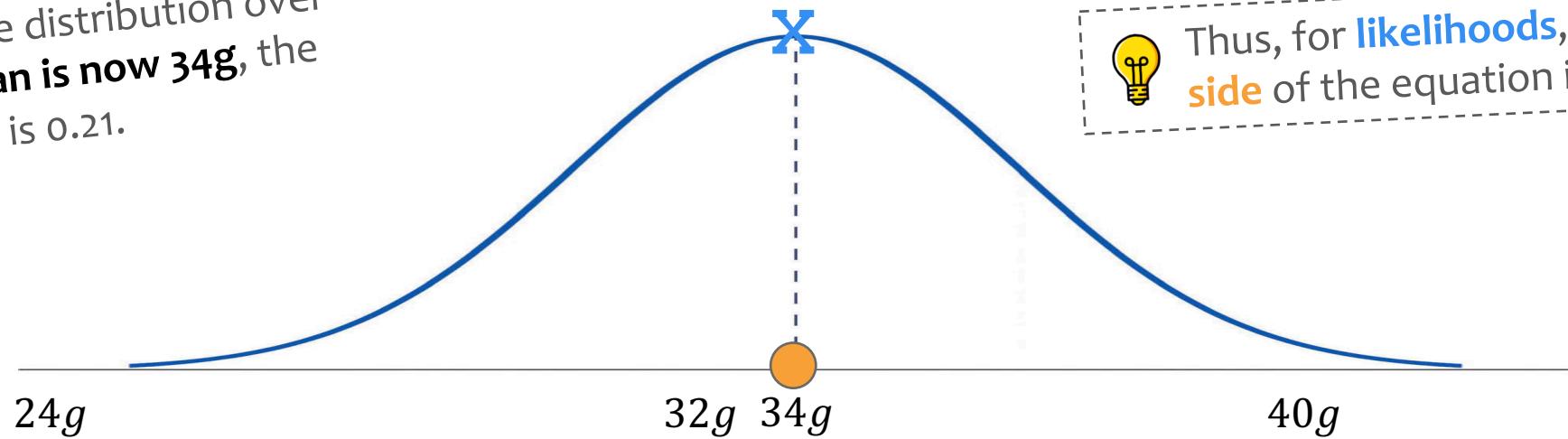
Thus, for **likelihoods, the right side** of the equation is **fixed!**

$$L(\text{mean} = 34g \text{ and std} = 2.5g \mid \text{mouse weighs } 34g) = 0.21$$

$$L(\mu = 34, \sigma = 2.5 \mid x = 34) = 0.21$$

$$\theta = \mu, \sigma$$

If we shifted the distribution over so that the **mean is now 34g**, the **new likelihood** is 0.21.



Thus, for **likelihoods, the right side** of the equation is **fixed!**

Recalling...



We can think **likelihood** as:

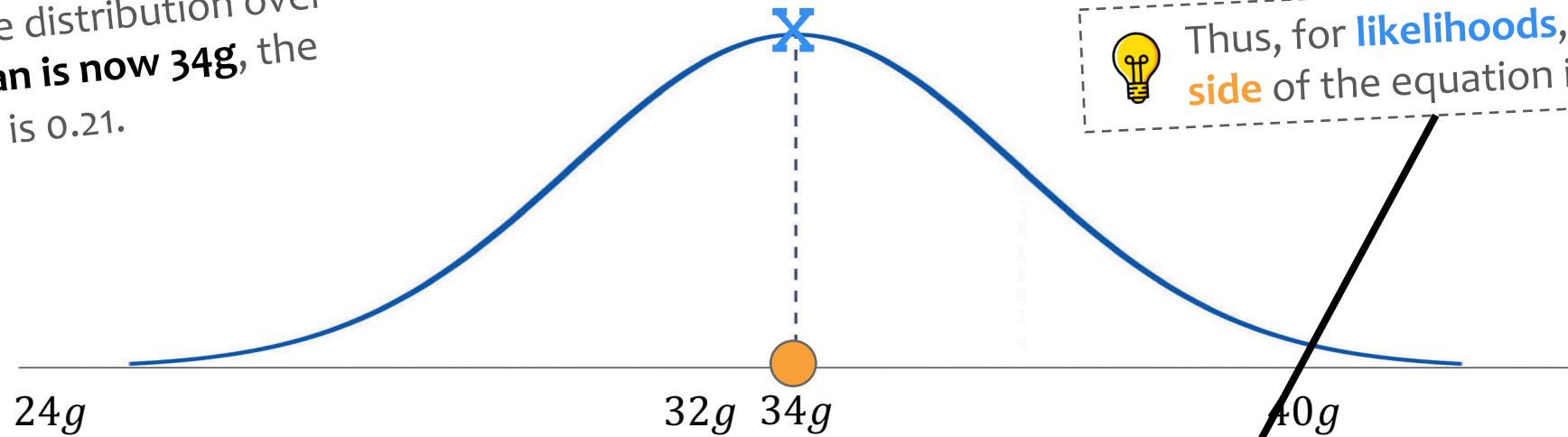
“How **likely** a particular **population** is to produce an **observe sample**.”

$$L(\text{mean} = 34g \text{ and std} = 2.5g \mid \text{mouse weighs } 34g) = 0.21$$

$$L(\mu = 34, \sigma = 2.5 \mid x = 34) = 0.21$$

$$\theta = \mu, \sigma$$

If we shifted the distribution over so that the **mean is now 34g**, the **new likelihood** is 0.21.



Recalling...



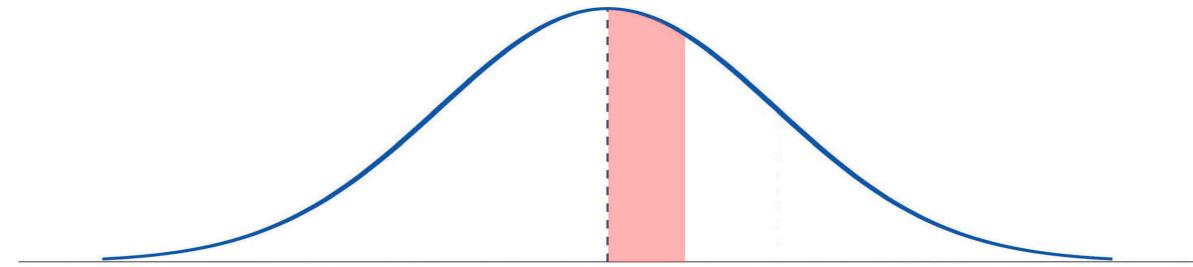
We can think **likelihood** as:
“How **likely** a particular **population** is to produce an **observe sample**.”

We can then check **likelihood** for **different populations** (represented by different data distributions) for a given **observe sample**.

Summary

Probabilities are the **areas under** a fixed distribution.

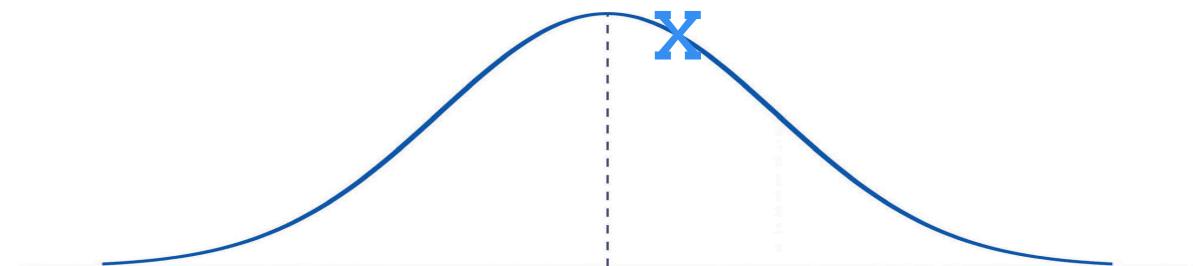
$$P(\text{data} \mid \text{distribution})$$



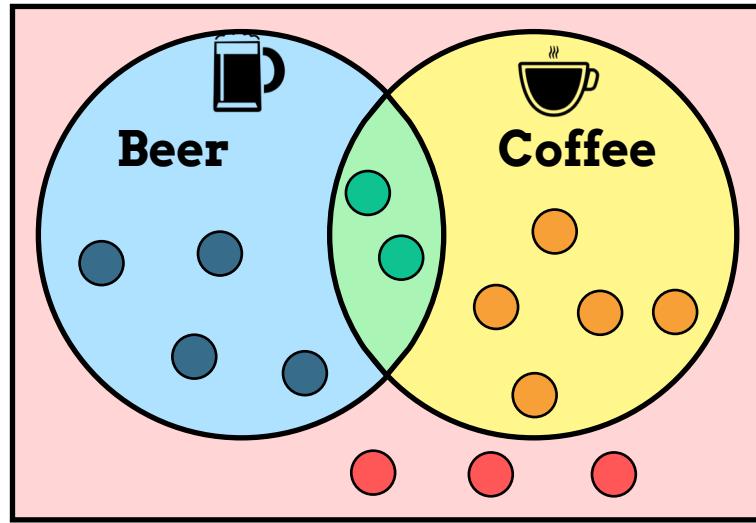
Likelihoods are the **y-axis values** for fixed data points

with distributions that can be moved.

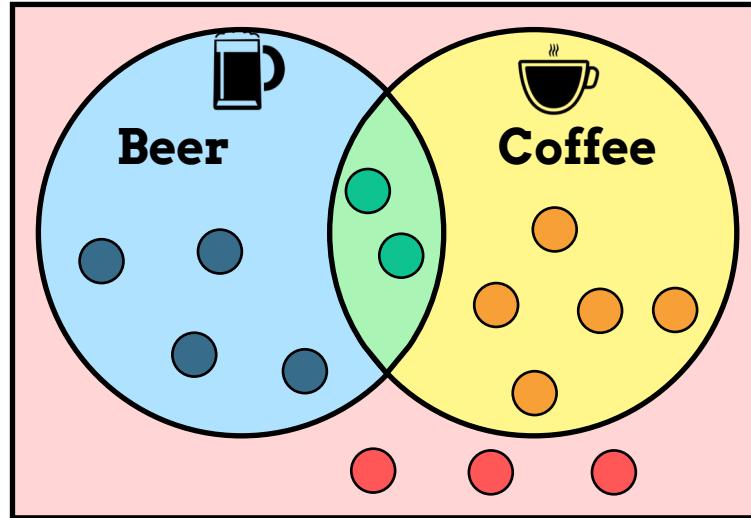
$$L(\text{distribution} \mid \text{data})$$



Bayes' Theorem



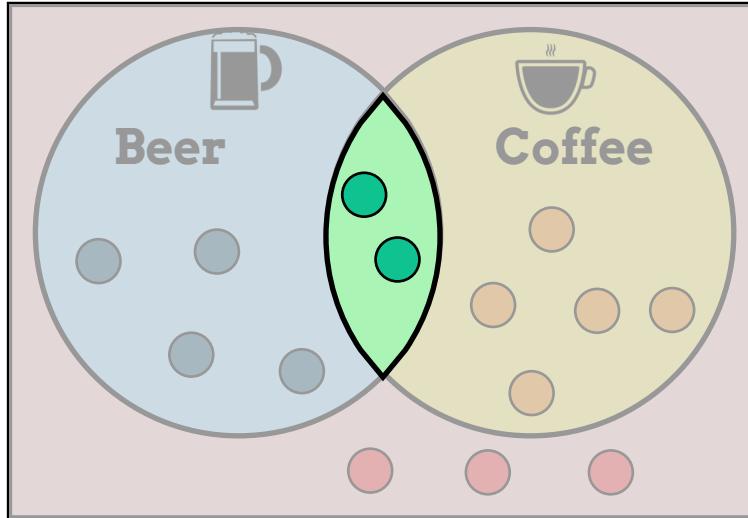
Contingency Table



	Likes Beer	Doesn't Like Beer
Likes Coffee		
Doesn't Like Coffee		

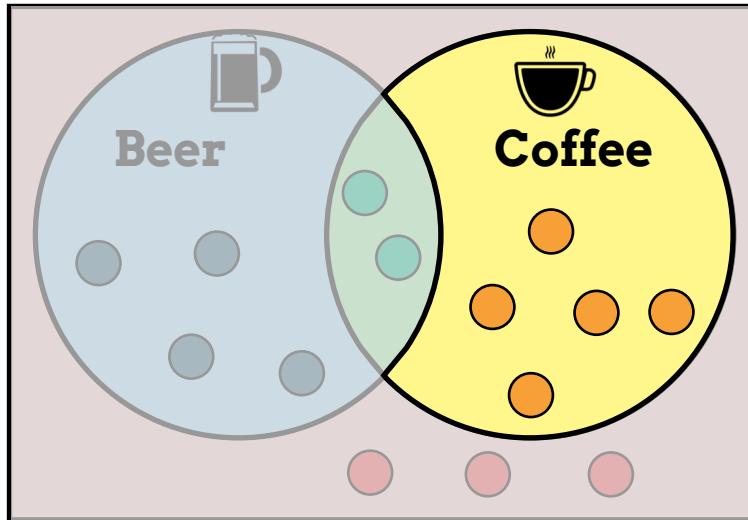
Contingency Table

	Likes Beer	Doesn't Like Beer
Likes Coffee	2 $P=2/14$	
Doesn't Like Coffee		



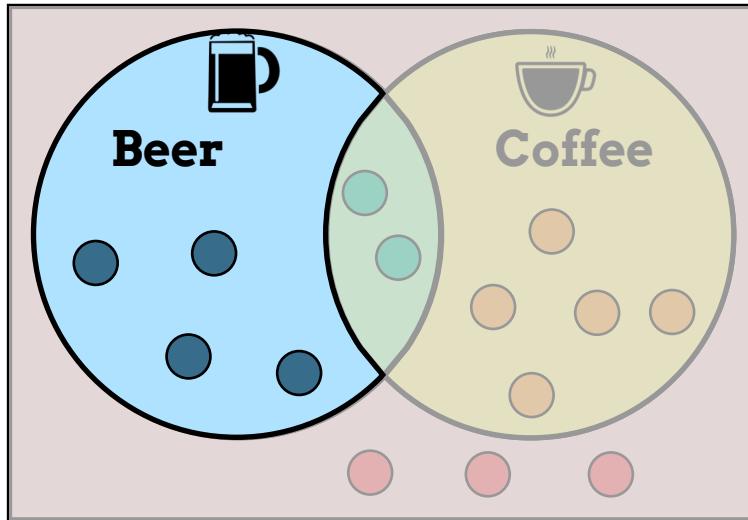
Contingency Table

	Likes Beer	Doesn't Like Beer
Likes Coffee	2 $P=2/14$	5 $P=5/14$
Doesn't Like Coffee		



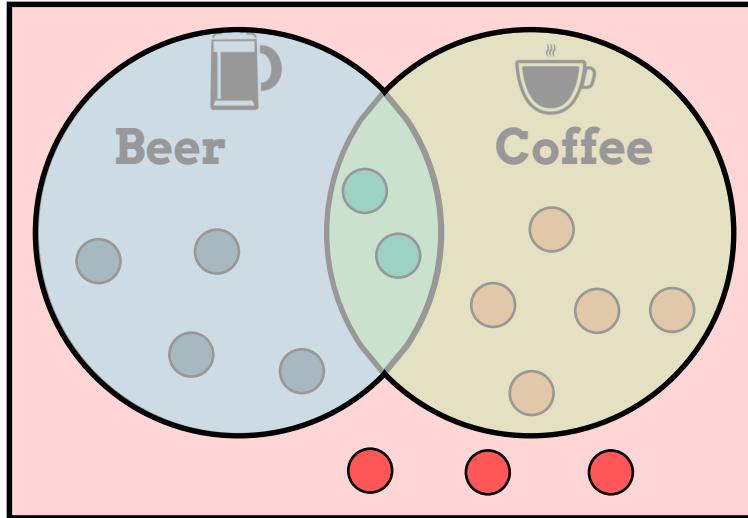
Contingency Table

	Likes Beer	Doesn't Like Beer
Likes Coffee	2 $P=2/14$	5 $P=5/14$
Doesn't Like Coffee	4 $P=4/14$	

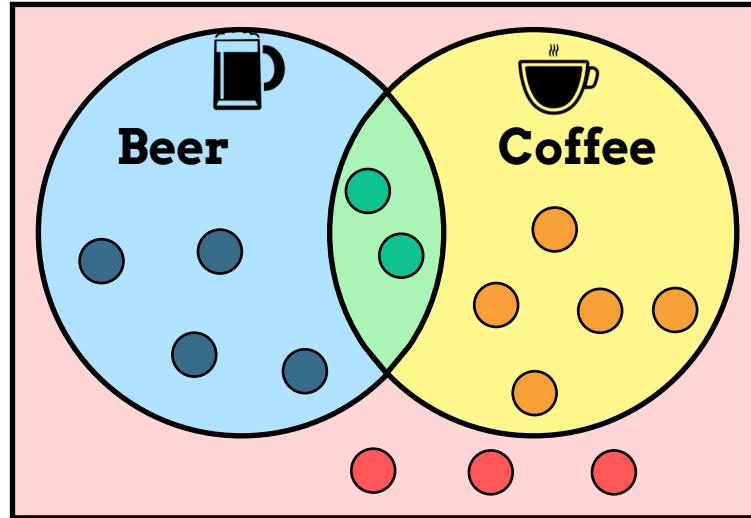


Contingency Table

	Likes Beer	Doesn't Like Beer
Likes Coffee	2 $P=2/14$	5 $P=5/14$
Doesn't Like Coffee	4 $P=4/14$	3 $P=3/14$

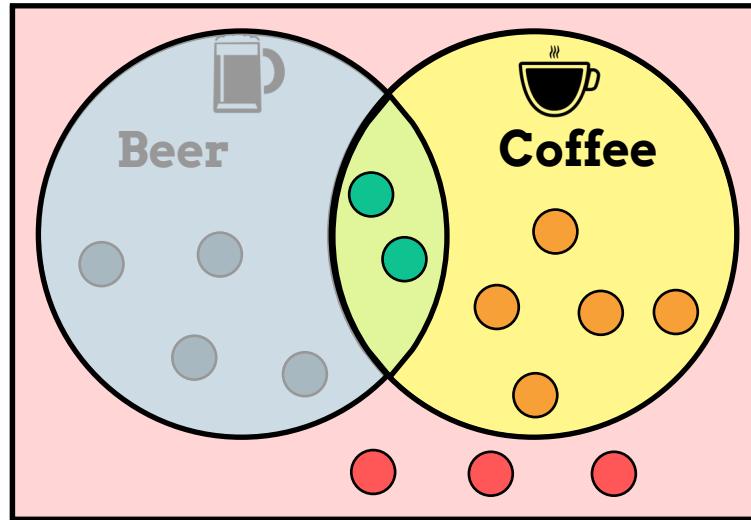


Contingency Table



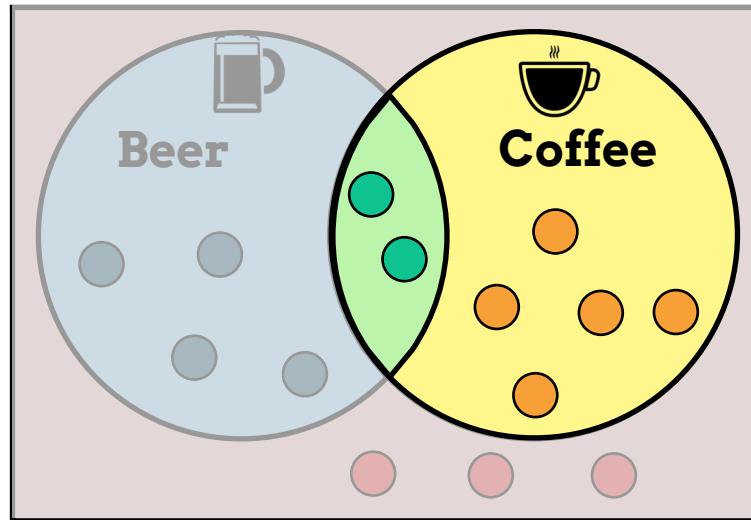
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	
Doesn't Like Coffee	4 P=4/14	3 P=3/14	

Contingency Table



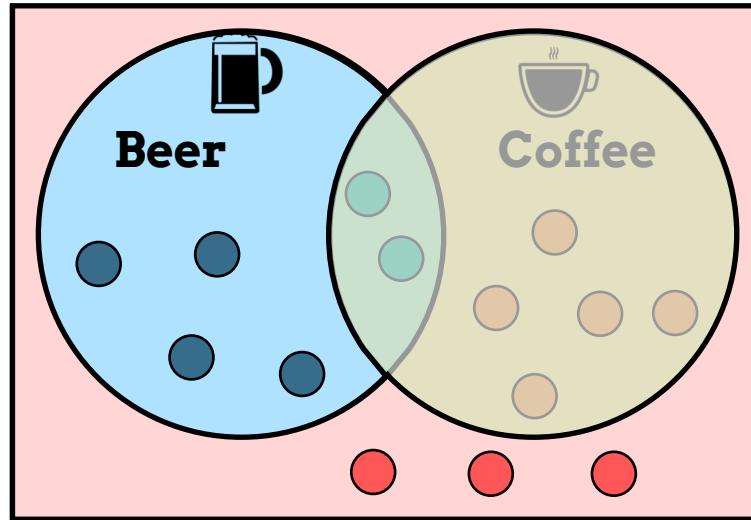
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	

Contingency Table



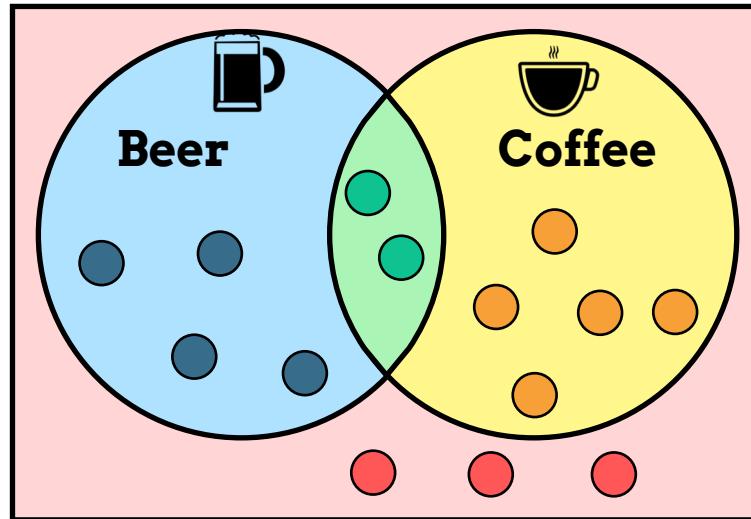
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	

Contingency Table



	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14

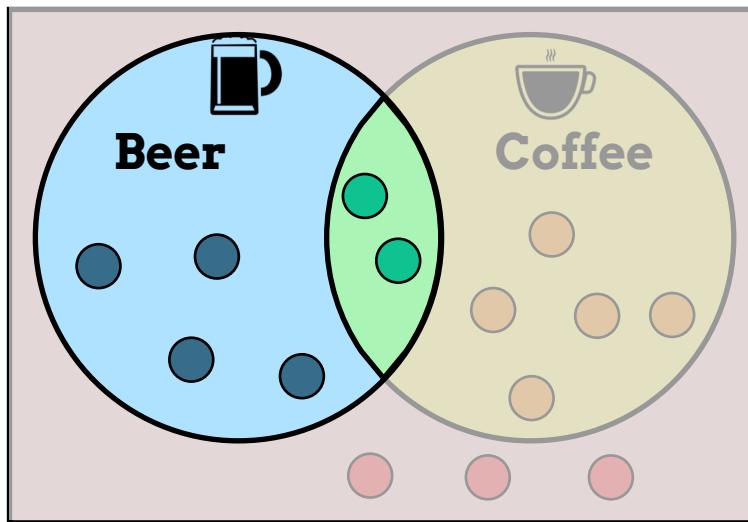
Contingency Table



	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 $P=2/14$	5 $P=5/14$	$2+5=7$ $P=7/14$
Doesn't Like Coffee	4 $P=4/14$	3 $P=3/14$	$4+3=7$ $P=7/14$
Column Total			

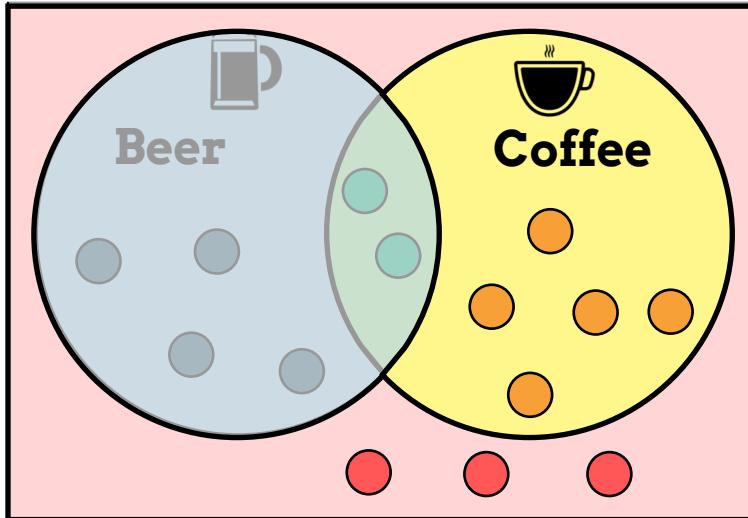
Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14		

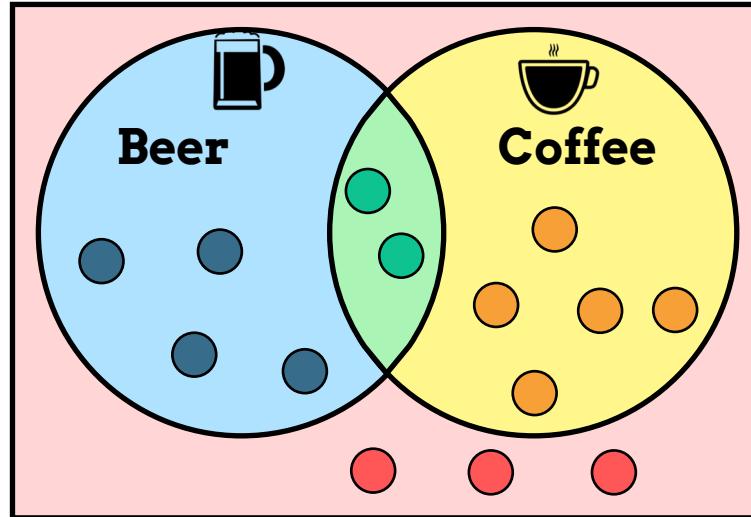


Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 $P=2/14$	5 $P=5/14$	$2+5=7$ $P=7/14$
Doesn't Like Coffee	4 $P=4/14$	3 $P=3/14$	$4+3=7$ $P=7/14$
Column Total	$2+4=6$ $P=6/14$	$5+3=8$ $P=8/14$	



Contingency Table

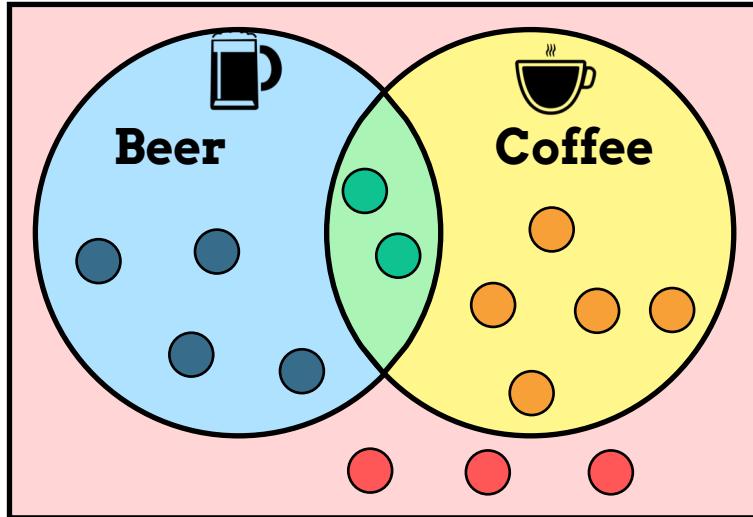


	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

Contingency Table

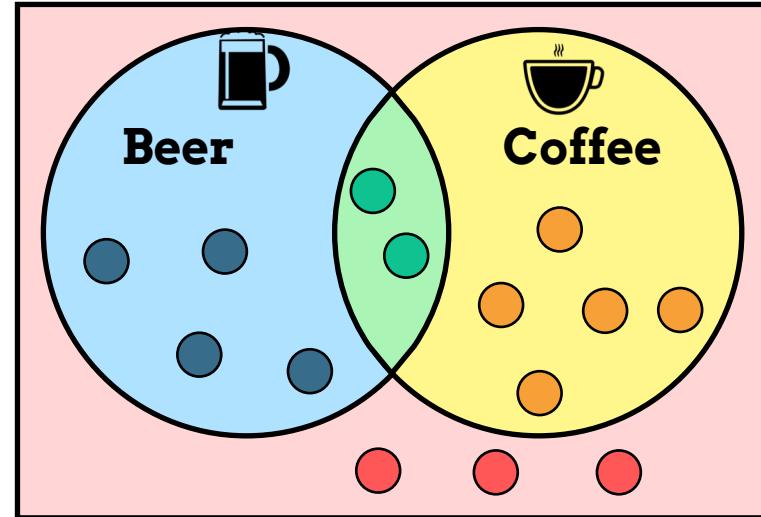
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 $P=2/14$	5 $P=5/14$	$2+5=7$ $P=7/14$
Doesn't Like Coffee	4 $P=4/14$	3 $P=3/14$	$4+3=7$ $P=7/14$
Column Total	$2+4=6$ $P=6/14$	$5+3=8$ $P=8/14$	

What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?



Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	



What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

$$P(\text{no like } b \mid \text{like } c) = ?$$

Contingency Table

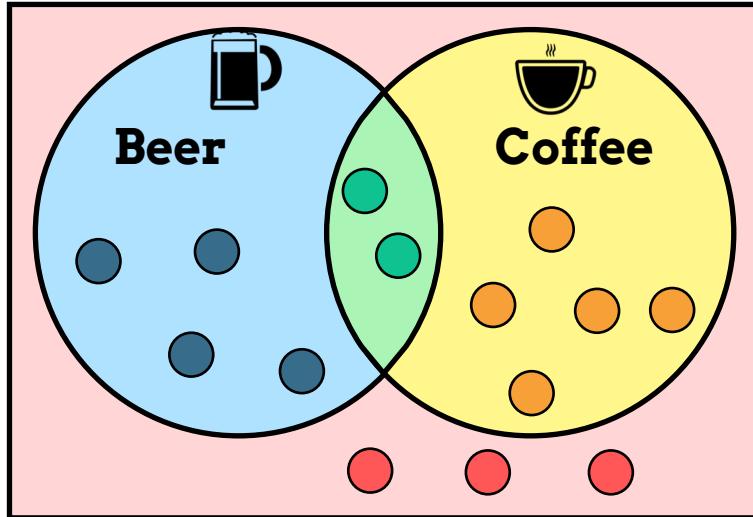
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

$$P(\text{no like } b \mid \text{like } c) = ?$$

We can rewrite that, being **redundant**, just to simplify our interpretation.

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = ?$$



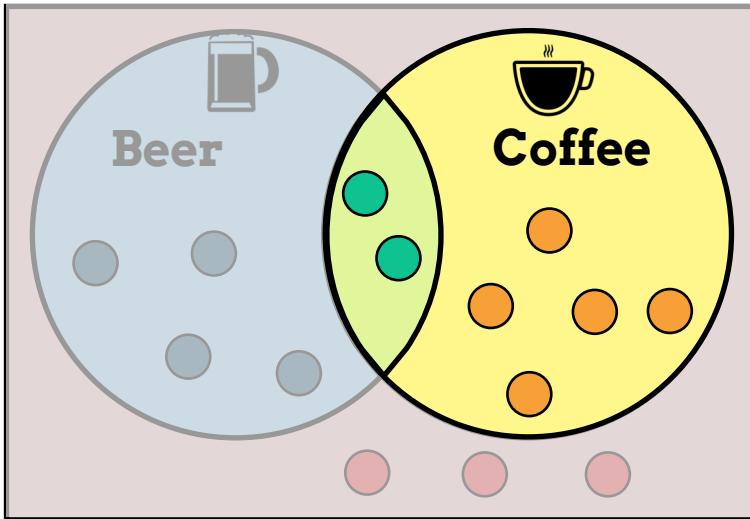
Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

$$P(\text{no like } b \text{ AND like } c)$$

$$\text{like } c) = P(\text{no like } b \text{ AND like } c)$$

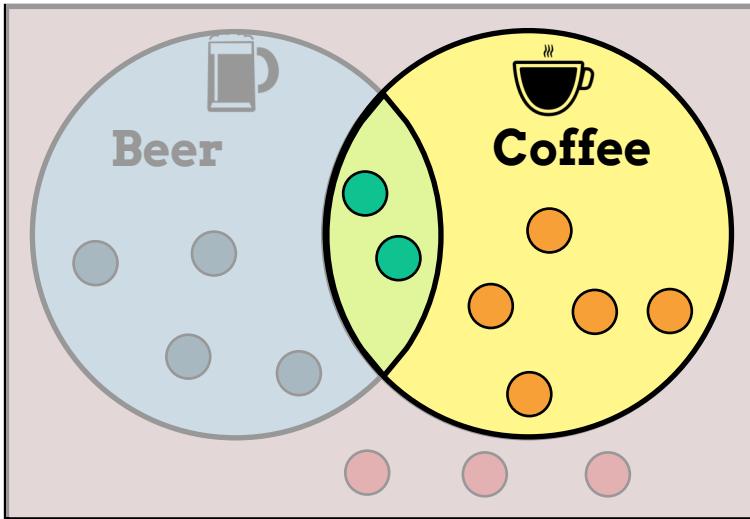


Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)}$$

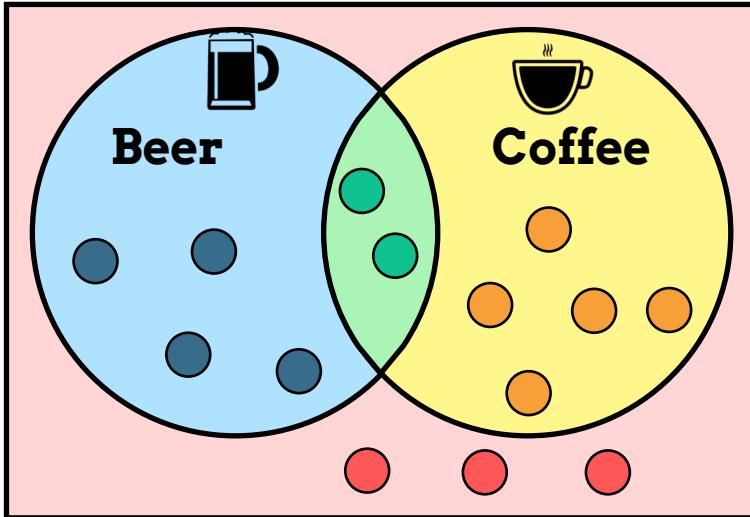


Contingency Table

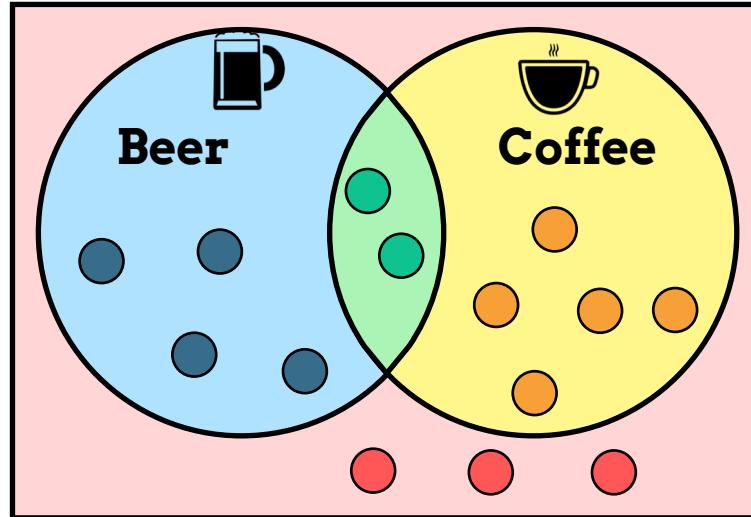
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{5}{P(\text{like } c)}$$



Contingency Table



	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

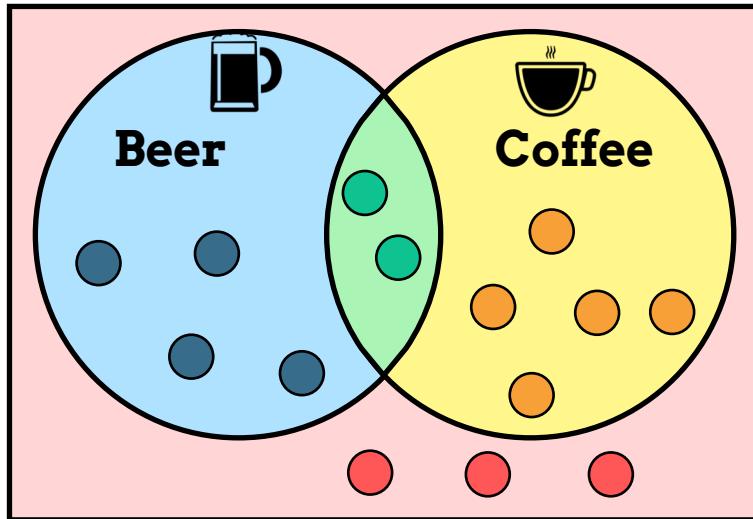
$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{5}{7}$$

Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{5}{7} = \boxed{0.71}$$

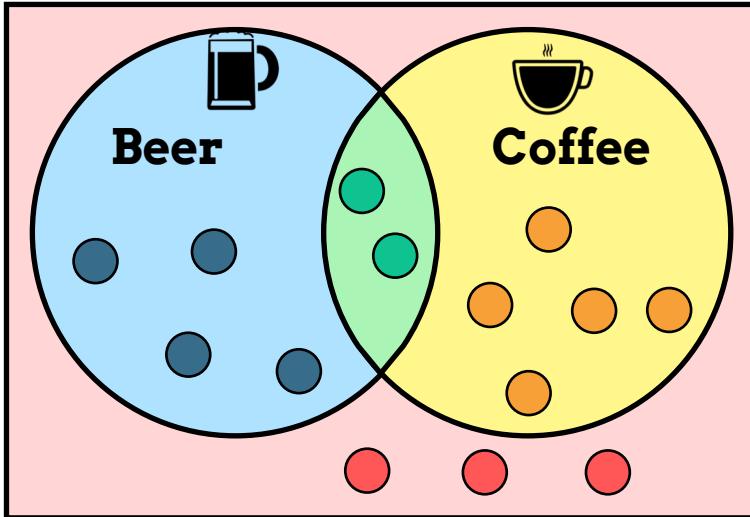


Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{\frac{5}{14}}{\frac{7}{14}} = 0.71$$

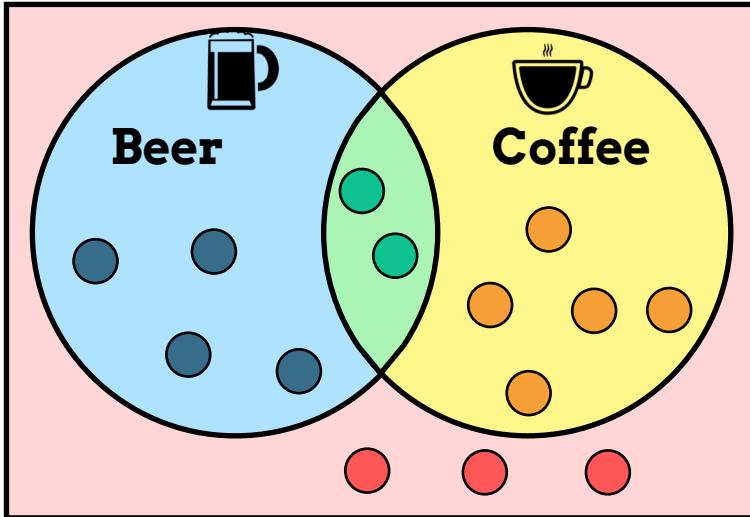


Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{\frac{5}{14}}{\frac{7}{14}} = \boxed{0.71}$$



Contingency Table

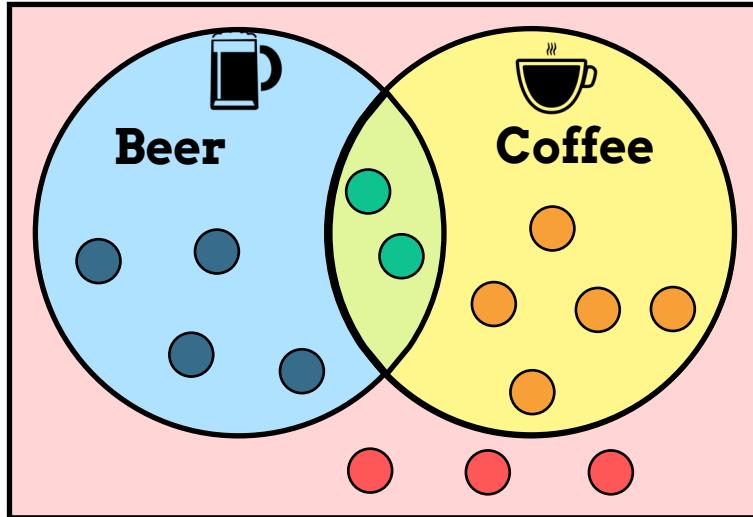
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)} = 0.71$$

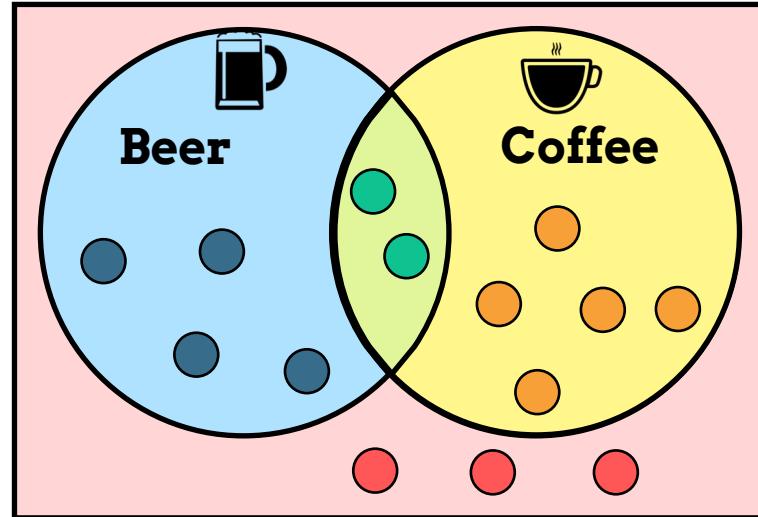


We have just derived the formula of **conditional probability**.



Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	



What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

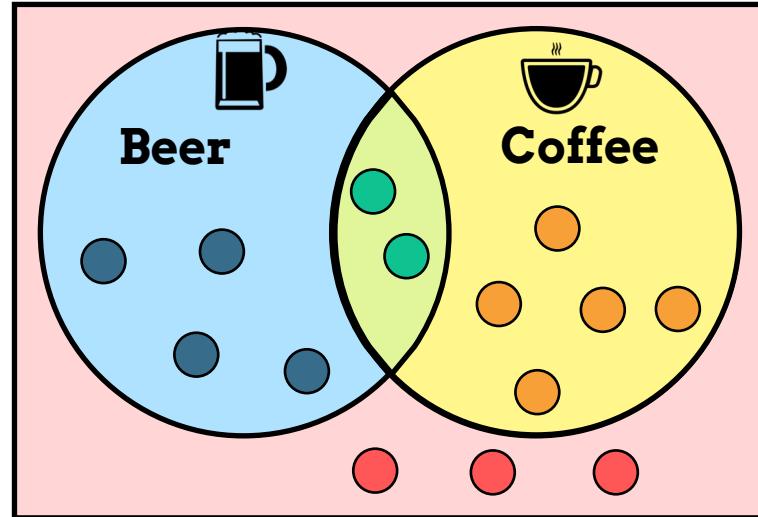
$$P(\text{no like } b \text{ AND like } c | \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)} = 0.71$$

We have just derived the formula of **conditional probability**.

probability that an event will happen
(that someone does not like beer, but likes coffee)...

Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	



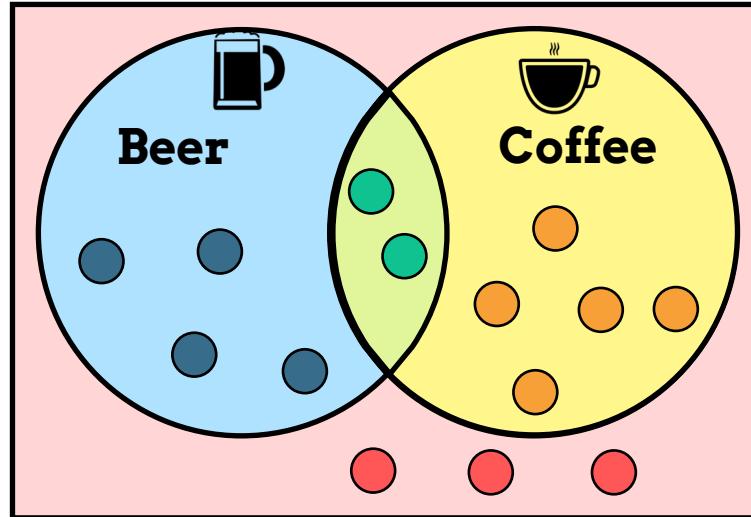
What is the **(conditional) probability** that someone does **not like beer** given that (s)he **likes coffee**?

$$P(\text{no like } b \text{ AND like } c | \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)} = 0.71$$

We have just derived the formula of **conditional probability**.

probability that an event will happen
(that someone does not like beer, but likes coffee)...
scaled by the knowledge we already know about the event
(prob. that someone likes coffee)...

Contingency Table



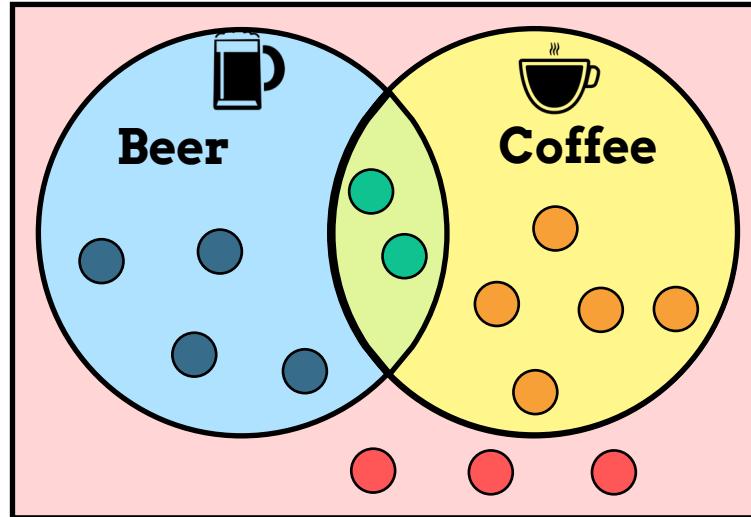
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

$$P(\text{no like } b \text{ AND like } c | \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)} = 0.71$$

Remember this part is **redundant**:

“Someone who *does not like beer and likes coffee*, given that (s)he *likes coffee*”.

Contingency Table



	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

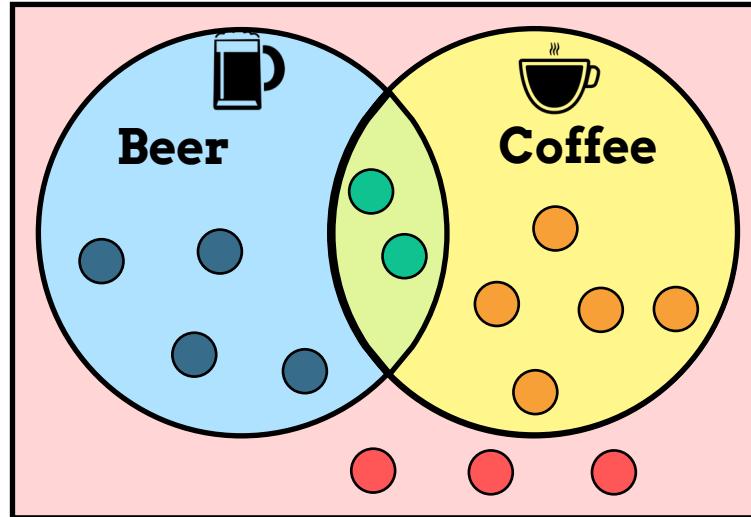
$$P(\text{no like } b \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)} = 0.71$$

Remember this part is **redundant**:

“Someone who *does not like beer and likes coffee*, given that (s)he *likes coffee*”.

The common and shorten statement does not have that part.

Contingency Table



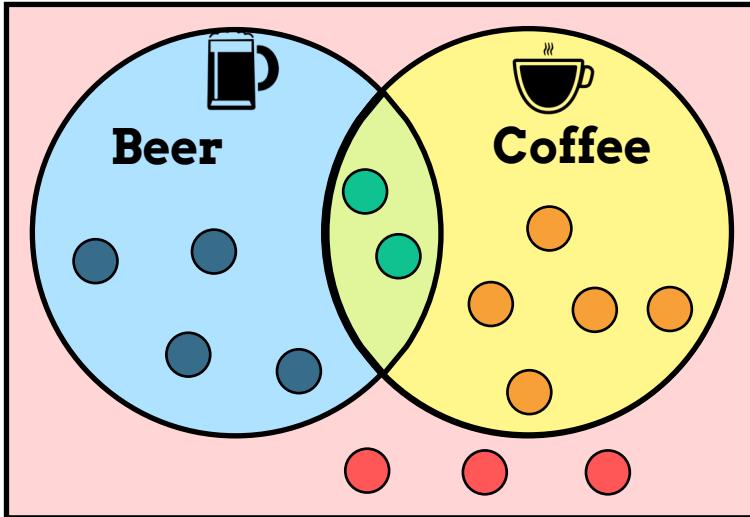
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

$$P(\text{no like } b \text{ AND like } c | \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)} = 0.71$$

However, the **redundant notation** for **conditional probability** makes it obvious that we want the prob. that an event happens...

Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	



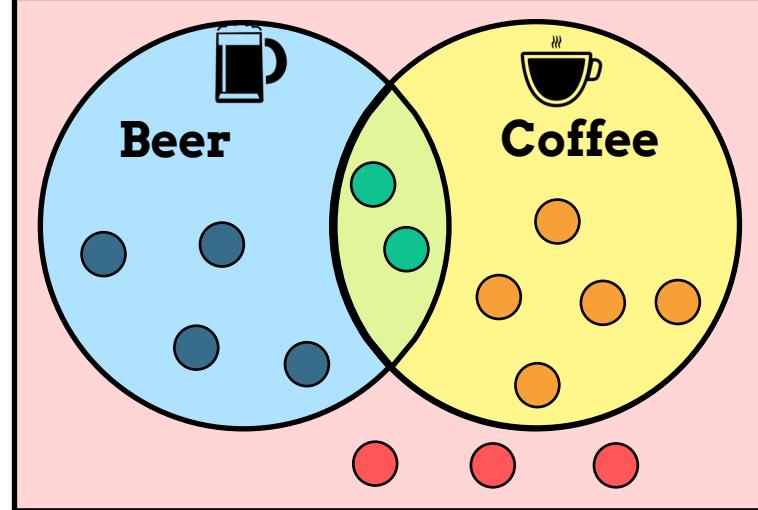
$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)} = 0.71$$

However, the **redundant notation** for **conditional probability** makes it obvious that we want the prob. that an event happens...

... scaled by the knowledge we already know about the event.

Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

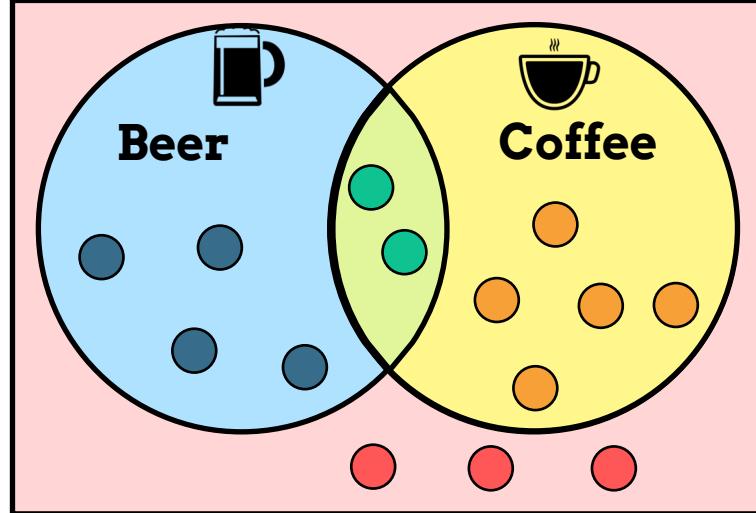


Let's see what happens when we **change** our **a priori** knowledge about the event.

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)}$$

Contingency Table

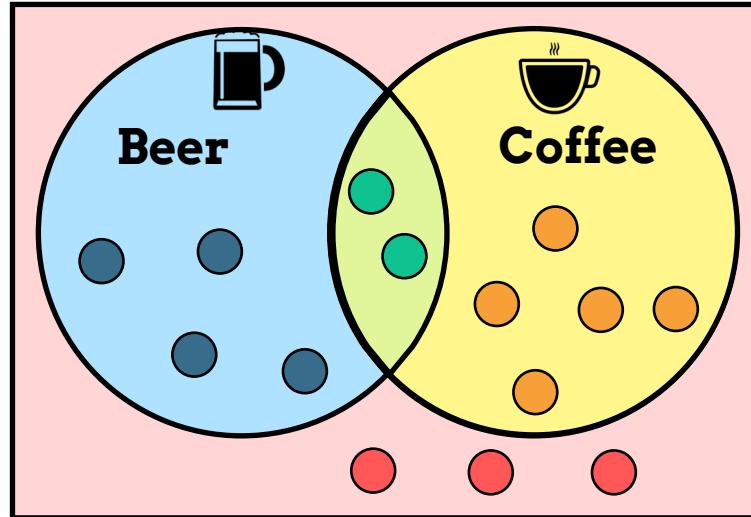
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	



Let's see what happens when we **change** our **a priori** knowledge about the event.

$$P(\text{no like } b \text{ AND like } c \mid \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)}$$

Contingency Table



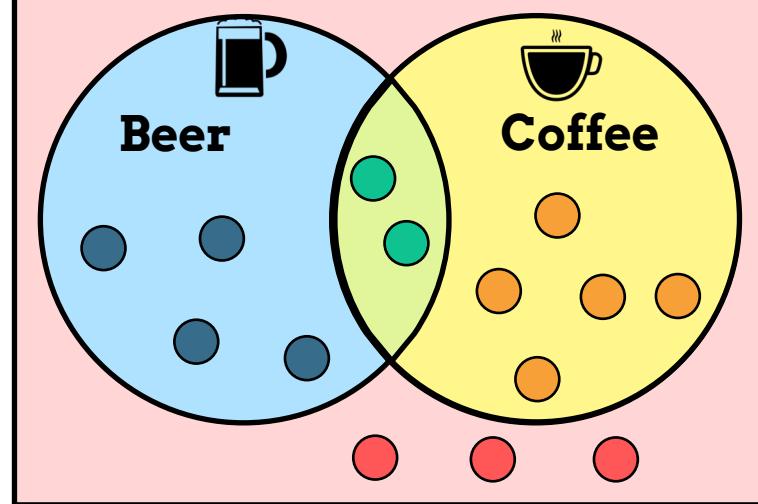
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

probability that an event will happen
(that someone does not like beer, but likes coffee)...

$$P(\text{no like } b \text{ AND like } c \mid \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)}$$

Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	



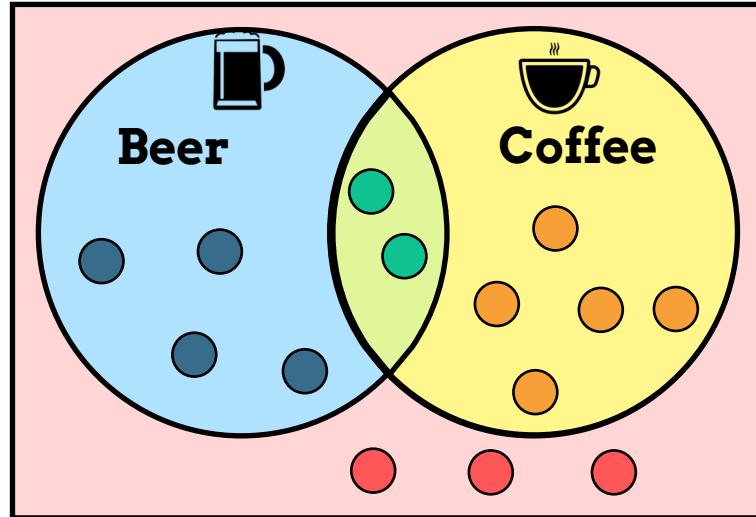
probability that an event will happen
(that someone does not like beer, but likes coffee)...

... scaled by the knowledge we already know about the event
(prob. that someone do not like beer)...

$$P(\text{no like } b \text{ AND like } c \mid \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)}$$

Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

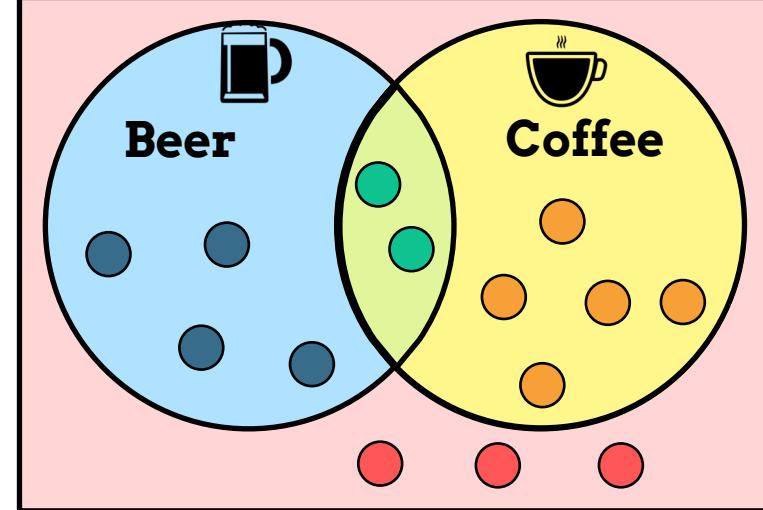


After plugging in the numbers...

$$P(\text{no like } b \text{ AND like } c \mid \text{no like } c) = \frac{\frac{5}{14}}{\frac{8}{14}} = 0.64$$

Contingency Table

	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	



Now, let's compare both **conditional probabilities**...

$$P(\text{no like } b \text{ AND like } c \mid \text{no like } c) = \frac{\frac{5}{14}}{\frac{8}{14}} = 0.64$$

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)}$$

$$P(\text{no like } b \text{ AND like } c \mid \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)}$$

In both cases, we want know the
probability of the same event...

$$\boxed{P(\text{no like } b \text{ AND like } c \mid \text{like } c)} = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)}$$

$$\boxed{P(\text{no like } b \text{ AND like } c \mid \text{no like } b)} = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)}$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)}$$

In both cases, we want know the **probability of the same event...**
which means **the numerators are the same**

$$P(\text{no like } b \text{ AND like } c | \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)}$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)}$$

But, since we have **different knowledge** in each case...

$$P(\text{no like } b \text{ AND like } c | \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)}$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)}$$

$$P(\text{no like } b \text{ AND like } c | \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)}$$

But, since we have **different knowledge** in each case...
we scale the probabilities of the events differently...

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)} = \boxed{0.71}$$

$$P(\text{no like } b \text{ AND like } c \mid \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)} = \boxed{0.63}$$

But, since we have **different knowledge** in each case...
we scale the probabilities of the events differently...
getting different **conditional probabilities**

Note that, **the numerator is equal** in both equations...

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)}$$

$$P(\text{no like } b \text{ AND like } c \mid \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)}$$

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{like } c)}$$

$$P(\text{no like } b \text{ AND like } c \mid \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c)}{P(\text{no like } b)}$$

Note that, **the numerator is equal** in both equations...
so that even if we did not know about such a prob.,
we still could solve the **conditional probability**...

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) * P(\text{like } c) = P(\text{no like } b \text{ AND like } c)$$

$$P(\text{no like } b \text{ AND like } c \mid \text{no like } b) * P(\text{no like } b) = P(\text{no like } b \text{ AND like } c)$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) * P(\text{like } c) = P(\text{no like } b \text{ AND like } c)$$

$$P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b) = P(\text{no like } b \text{ AND like } c)$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) * P(\text{like } c) = P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b)$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) * P(\text{like } c) = P(\text{no like } b \text{ AND like } c)$$

$$P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b) = P(\text{no like } b \text{ AND like } c)$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) * P(\text{like } c) = P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b)$$



$$P(\text{no like } b \text{ AND like } c | \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b)}{P(\text{like } c)}$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) * P(\text{like } c) = P(\text{no like } b \text{ AND like } c)$$

$$P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b) = P(\text{no like } b \text{ AND like } c)$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) * P(\text{like } c) = P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b)$$



$$P(\text{no like } b \text{ AND like } c | \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b)}{P(\text{like } c)}$$

Or

$$P(\text{no like } b \text{ AND like } c | \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c | \text{like } c) * P(\text{like } c)}{P(\text{no like } b)}$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) * P(\text{like } c) = P(\text{no like } b \text{ AND like } c)$$

$$P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b) = P(\text{no like } b \text{ AND like } c)$$

$$P(\text{no like } b \text{ AND like } c | \text{like } c) * P(\text{like } c) = P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b)$$



We have just derived the formula
Bayes' Theorem.

$$P(\text{no like } b \text{ AND like } c | \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c | \text{no like } b) * P(\text{no like } b)}{P(\text{like } c)}$$

Or

$$P(\text{no like } b \text{ AND like } c | \text{no like } b) = \frac{P(\text{no like } b \text{ AND like } c | \text{like } c) * P(\text{like } c)}{P(\text{no like } b)}$$

B = likes coffee

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c \mid \text{no like } b) * P(\text{no like } b)}{P(\text{like } c)}$$

Bayes' Theorem tells us that this **conditional probability**, which is based on the knowledge about the person **likes coffee (event B)**...

A = does not like beer

B = likes coffee

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c \mid \text{no like } b) * P(\text{no like } b)}{P(\text{like } c)}$$

Bayes' Theorem tells us that this **conditional probability**, which is based on the knowledge about the person **likes coffee (event B)**...

... can be derived from this **conditional probability**, which is based on which is based on the knowledge about the person **does not like beer (event A)**.

A = does not like beer

B = likes coffee

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c \mid \text{no like } b) * P(\text{no like } b)}{P(\text{like } c)}$$

A = does not like beer

B = likes coffee

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c \mid \text{no like } b) * P(\text{no like } b)}{P(\text{like } c)}$$



$$P(A \text{ AND } B \mid B) = \frac{P(A \text{ AND } B \mid A) * P(A)}{P(B)}$$

A = does not like beer

B = likes coffee

$$P(\text{no like } b \text{ AND like } c \mid \text{like } c) = \frac{P(\text{no like } b \text{ AND like } c \mid \text{no like } b) * P(\text{no like } b)}{P(\text{like } c)}$$



$$P(A \text{ AND } B \mid B) = \frac{P(A \text{ AND } B \mid A) * P(A)}{P(B)}$$



Removing
redundancy

$$P(A|B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

Bayes' Theorem

Bayes' Theorem

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Likelihood
The prob. of **B** being True, given **A**.

Prior Probability
The prob. of **A** being True.

Posterior (*a posteriori* prob.)
The prob. of **A** being True, given **B**.

Marginal Likelihood
The prob. of **B** being True.

The diagram illustrates the components of Bayes' Theorem. It features a central equation $P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$. Four arrows point from text labels to specific terms in the equation: an arrow from 'Likelihood' points to $P(B|A)$; an arrow from 'Prior Probability' points to $P(A)$; an arrow from 'Posterior (*a posteriori* prob.)' points to $P(A|B)$; and an arrow from 'Marginal Likelihood' points to $P(B)$.

Bayes' Theorem

a different way
of seeing that

H: Hypothesis
E: Evidence

Likelihood

- The prob. of seeing the **evidence** if the **hypothesis** is true.
- It updates our initial guess through the **evidence** that comes from **data**.

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Posterior (*a posteriori* prob.)

The prob. of the **hypothesis** is true given the observed **evidence**.

Prior Probability

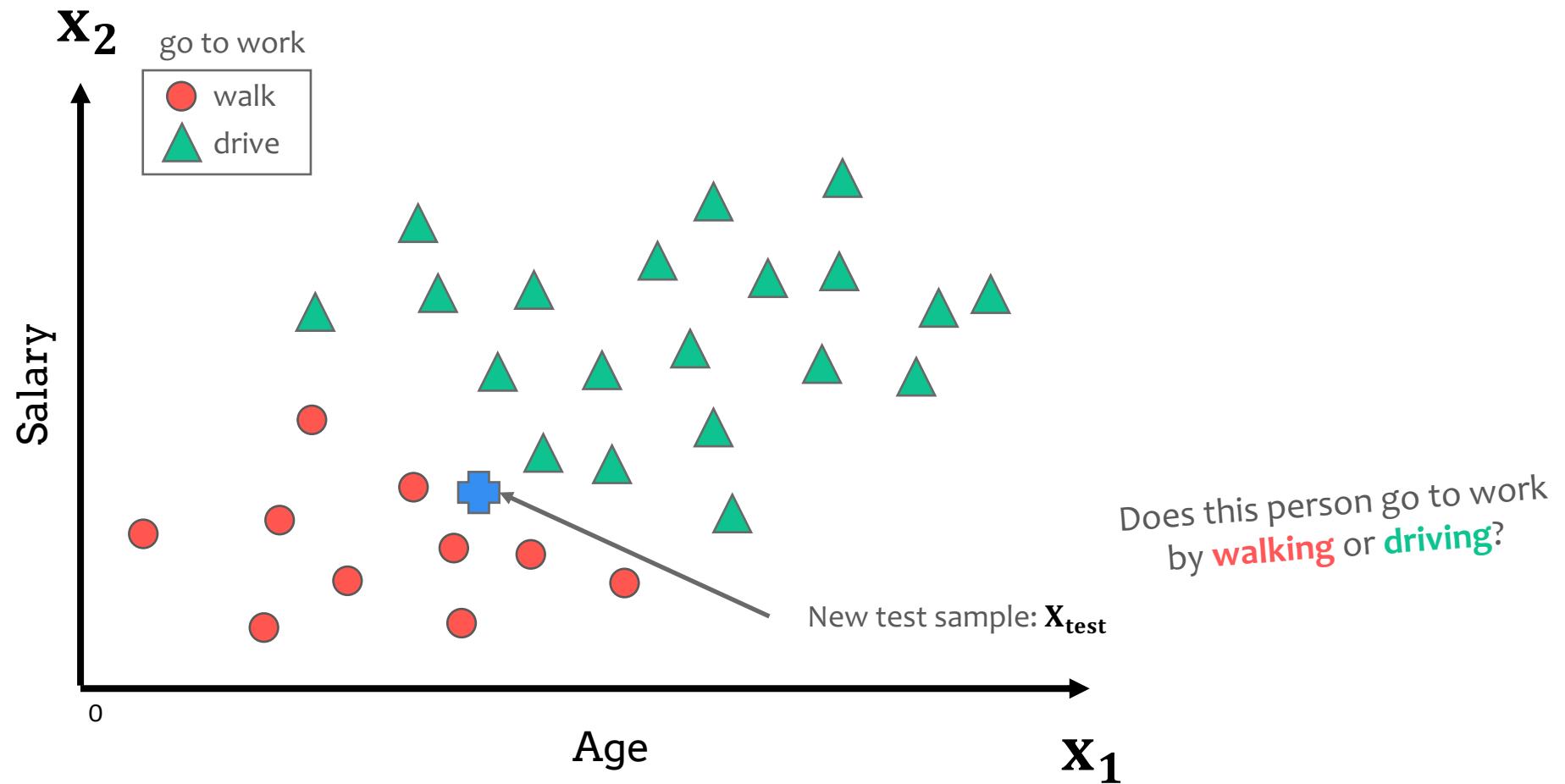
- The prob. of the **hypothesis** is true.
- Our **initial guess** about the **hypothesis** before any **evidence** is present.
- How probable was our **hypothesis** **before** observing the **evidence**?

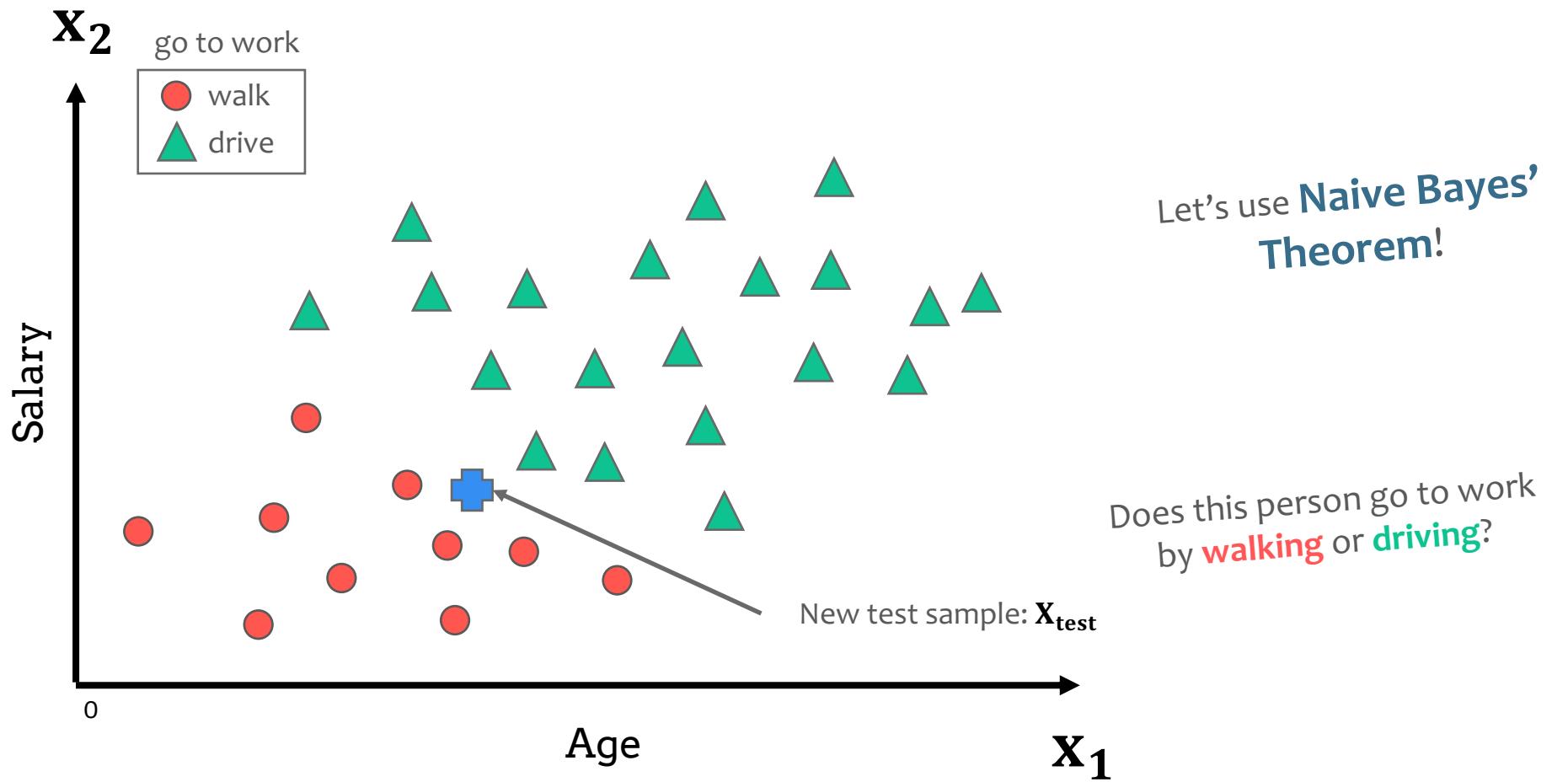
Marginal Likelihood

- The prob. of **B** being true.
- The prob. of observing the **evidence**.
- How probable is the new **evidence** under **all** possible **hypothesis**?

Naive Bayes Classifier







Naive Bayes Classifier

Bayes' Theorem

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

The diagram illustrates the components of Bayes' Theorem. The formula is shown as:

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Annotations with arrows point to each part:

- An arrow points from the term $P(H)$ to the label "Prior probability".
- An arrow points from the term $P(E|H)$ to the label "Likelihood".
- An arrow points from the term $P(E)$ to the label "Marginal Likelihood".
- An arrow points from the term $P(H|E)$ to the label "Posterior probability".

Naive Bayes Classifier

Bayes' Theorem

$$P(\text{class}|\text{data}) = \frac{P(\text{data}|\text{class}) * P(\text{class})}{P(\text{data})}$$

Diagram illustrating the components of Bayes' Theorem:

- Likelihood**: Points to the term $P(\text{data}|\text{class})$.
- Prior probability**: Points to the term $P(\text{class})$.
- Posterior probability**: Points to the term $P(\text{class}|\text{data})$.
- Marginal Likelihood**: Points to the term $P(\text{data})$.

Naive Bayes Classifier

Step 1

$$P(Walks|x_{test}) = \frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})}$$

Diagram illustrating the components of the Naive Bayes formula:

- #1 Prior probability**: $P(Walks)$, indicated by an upward arrow.
- #2 Marginal Likelihood**: $P(x_{test})$, indicated by a downward arrow.
- #3 Likelihood**: $P(x_{test}|Walks)$, indicated by a diagonal arrow pointing from the features of the new sample to the likelihood term.
- #4 Posterior probability**: $P(Walks|x_{test})$, indicated by a downward arrow.
- Labels: "class" above $P(Walks)$, "feats of the new sample" above x_{test} .

Naive Bayes Classifier

Step 2

$$P(\text{Drives} | \mathbf{x}_{test}) = \frac{P(\mathbf{x}_{test} | \text{Drives}) * P(\text{Drives})}{P(\mathbf{x}_{test})}$$

Diagram illustrating the components of the Naive Bayes formula:

- #1 Prior probability**: $P(\text{Drives})$, indicated by an upward arrow.
- #2 Marginal Likelihood**: $P(\mathbf{x}_{test})$, indicated by a downward arrow.
- #3 Likelihood**: $P(\mathbf{x}_{test} | \text{Drives})$, indicated by a diagonal arrow pointing from the features of the new sample to the likelihood term.
- #4 Posterior probability**: $P(\text{Drives} | \mathbf{x}_{test})$, indicated by a downward arrow pointing to the final result.
- class**: Drives (the class being predicted).
- feats of the new sample**: \mathbf{x}_{test} (the input features).

Naive Bayes Classifier

Step 3

$$P(\text{Walks} | \mathbf{x}_{test}) \text{ vs } P(\text{Drives} | \mathbf{x}_{test})$$

We assign the **new test instance** to the class with **the highest posteriori probability**.

Naive Bayes Classifier

Step 3

Let's do this

$$P(\text{Walks} | \mathbf{x}_{test}) \text{ vs } P(\text{Drives} | \mathbf{x}_{test})$$

We assign the **new test instance** to the class with **the highest posteriori probability**.

Naive Bayes

$$P(Walks|x_{test}) = \frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})}$$

Diagram illustrating the components of the Naive Bayes formula:

- #1 Prior probability**: $P(Walks)$, indicated by an upward arrow.
- #2 Marginal Likelihood**: $P(x_{test})$, indicated by a downward arrow.
- #3 Likelihood**: $P(x_{test}|Walks)$, indicated by a diagonal arrow pointing from the features of the new sample to the likelihood term.
- #4 Posterior probability**: $P(Walks|x_{test})$, indicated by a downward arrow.
- Labels: "class" above $P(Walks)$, "feats of the new sample" above x_{test} .

Naive Bayes

$$P(Walks|x_{test}) = \frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})}$$

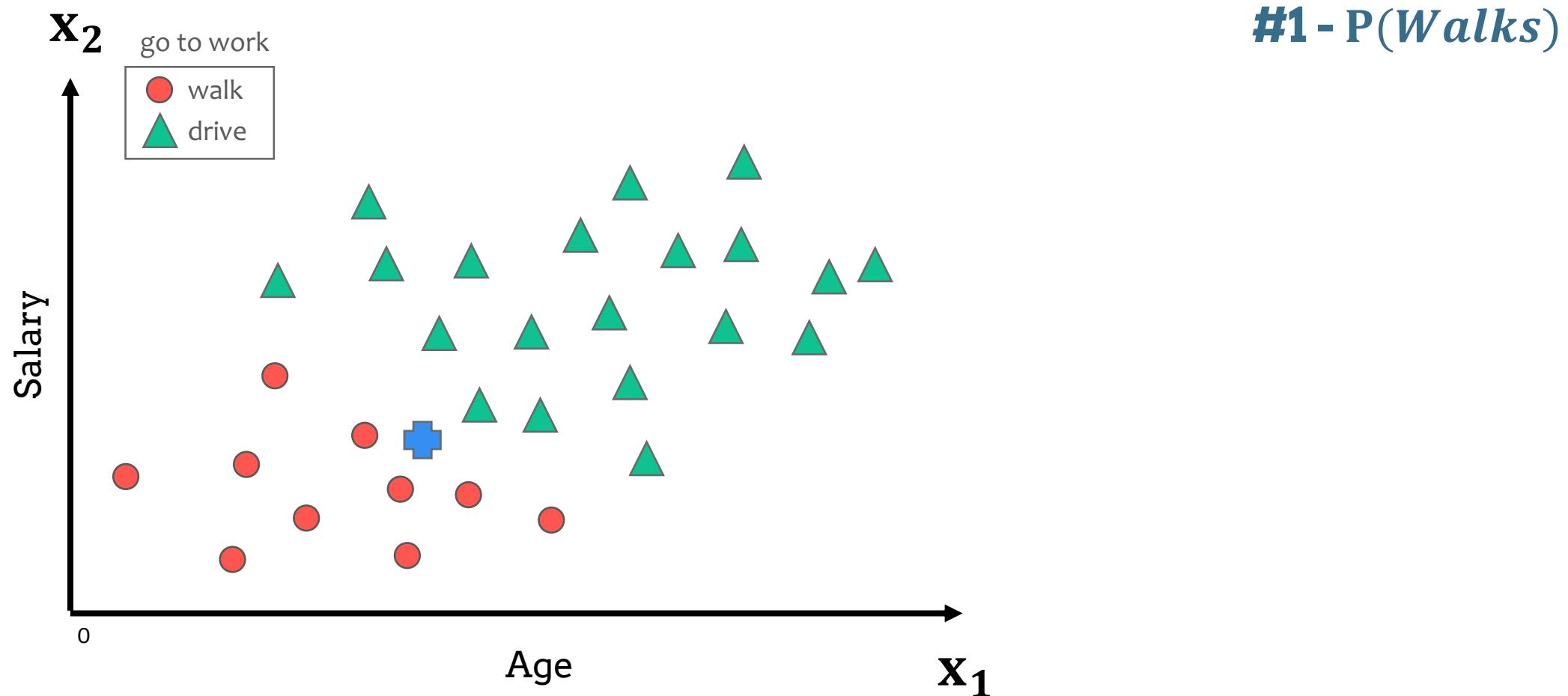
Diagram illustrating the components of the Naive Bayes formula:

- #1 Prior probability** (red arrow pointing to $P(Walks)$)
- #2 Marginal Likelihood** (black arrow pointing to $P(x_{test})$)
- #3 Likelihood** (black arrow pointing to $P(x_{test}|Walks)$)
- #4 Posterior probability** (black arrow pointing to $P(Walks|x_{test})$)

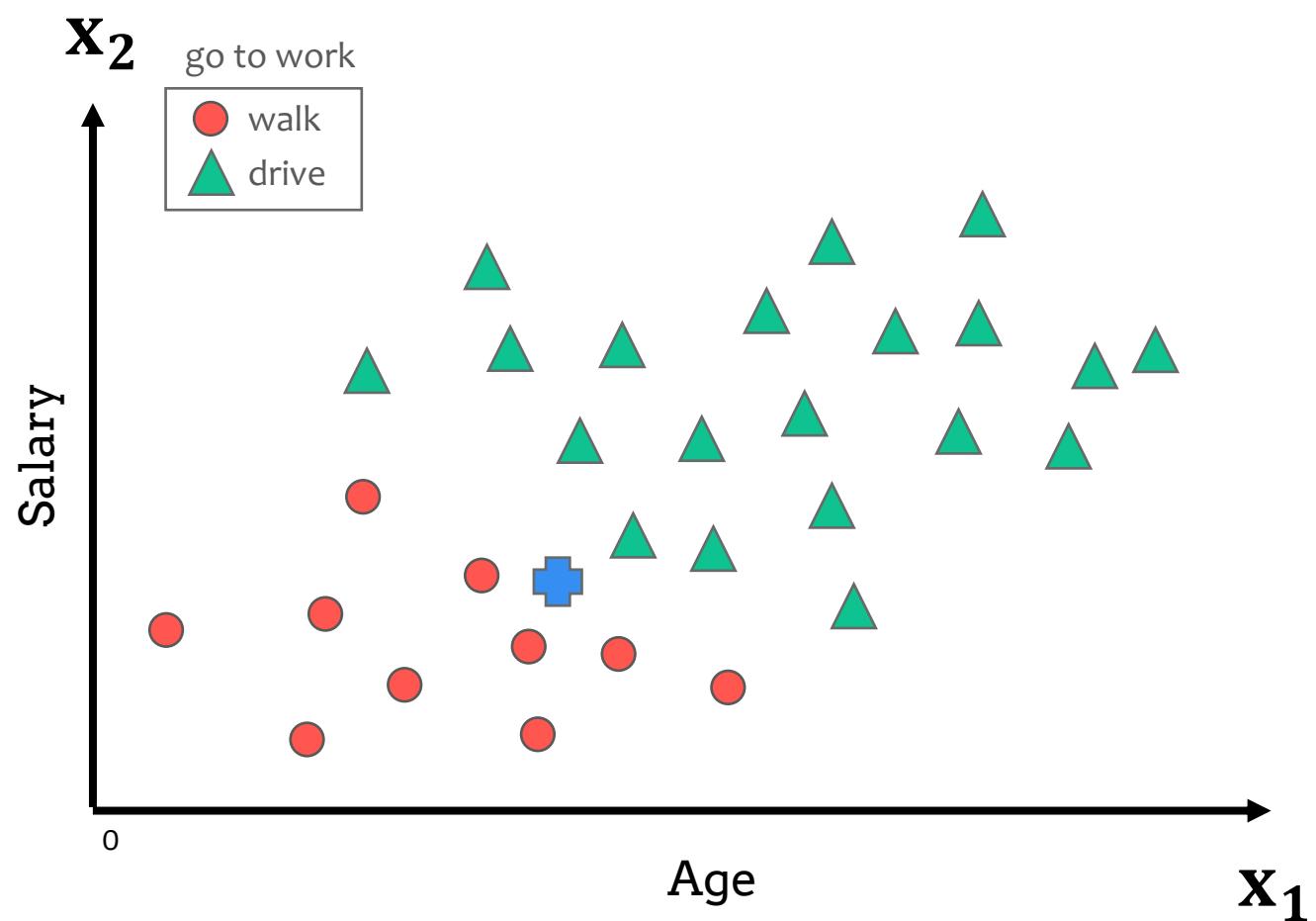
Annotations:

- class: above $P(Walks|x_{test})$
- feats of the new sample: above x_{test}

Naive Bayes: Step 1



Naive Bayes: Step 1



#1 - $P(Walks)$

This is our **initial guess** that people go to work by **walking**.

Naive Bayes: Step 1



#1 - $P(Walks)$

This is our **initial guess** that people go to work by **walking**.

This can be any **probability**.

Naive Bayes: Step 1



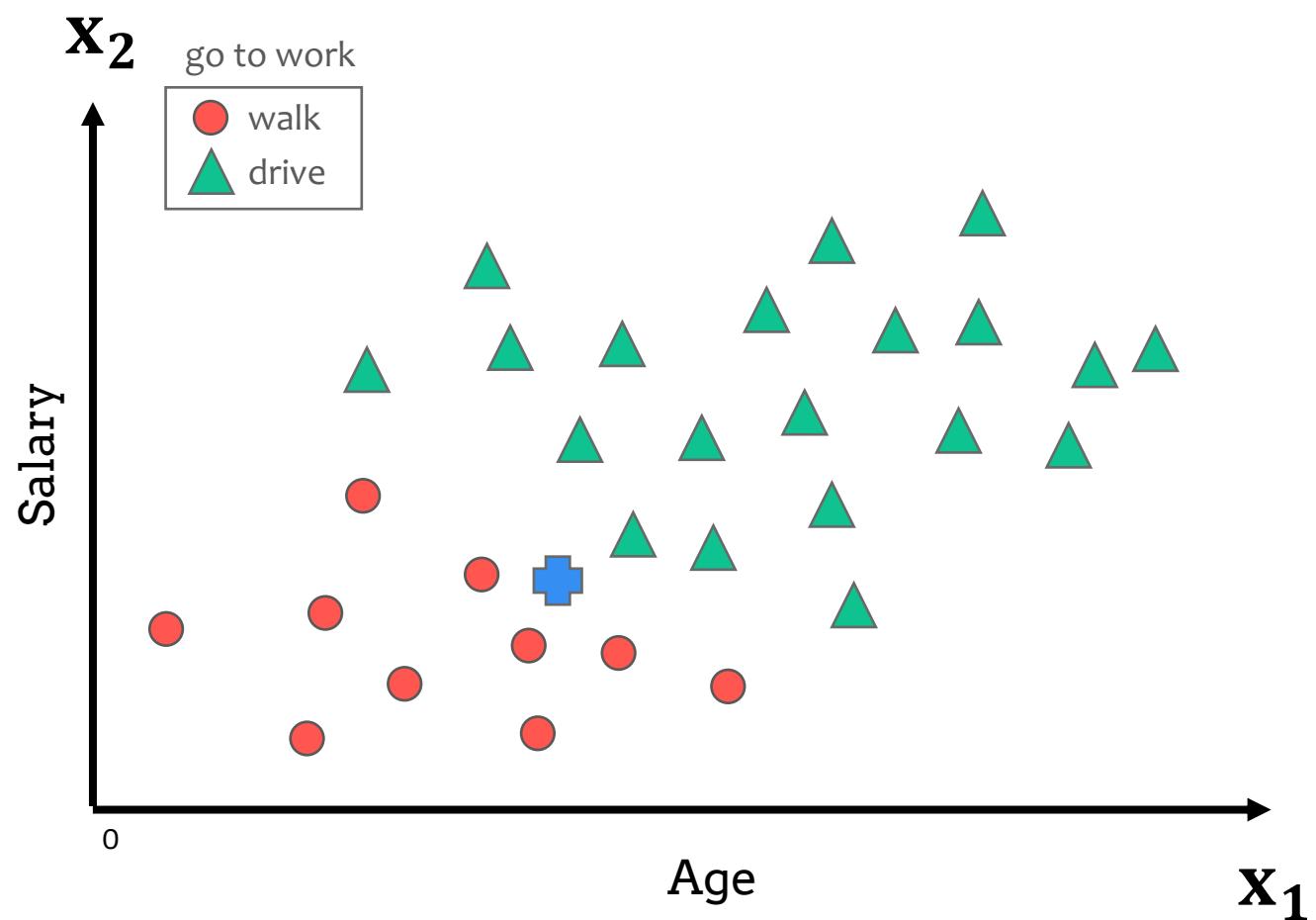
#1 - $P(Walks)$

This is our **initial guess** that people go to work by **walking**.

This can be any **probability**.

However, a **common guess** is estimated from the **training data**.

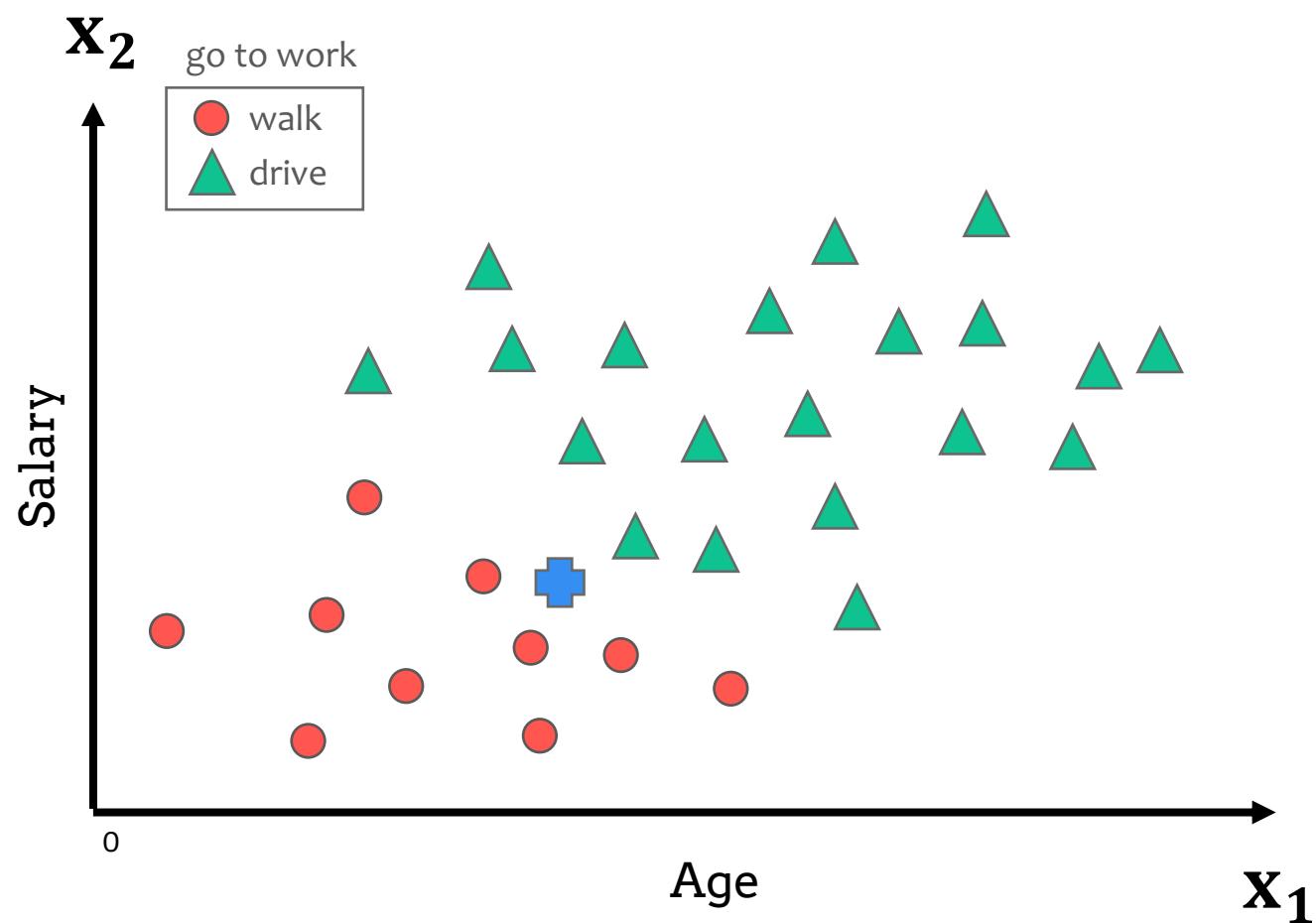
Naive Bayes: Step 1



#1 - $P(Walks)$

$$P(Walks) = \frac{\text{Number of Walkers}}{\text{Total Train. Samples}}$$

Naive Bayes: Step 1



#1 - P(Walks)

$$P(\text{Walks}) = \frac{\text{Number of Walkers}}{\text{Total Train. Samples}}$$

$$P(\text{Walks}) = \frac{10}{30}$$

Naive Bayes

$$P(Walks|x_{test}) = \frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})}$$

Diagram illustrating the components of the Naive Bayes formula:

- #1 Prior probability** (red arrow pointing to $P(Walks)$)
- #2 Marginal Likelihood** (black arrow pointing to $P(x_{test})$)
- #3 Likelihood** (black arrow pointing to $P(x_{test}|Walks)$)
- #4 Posterior probability** (black arrow pointing to $P(Walks|x_{test})$)

Annotations:

- class: above $P(Walks)$
- feats of the new sample: above x_{test}

Naive Bayes

$$P(Walks|x_{test}) = \frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})}$$

class feats of the new sample

#4 Posterior probability

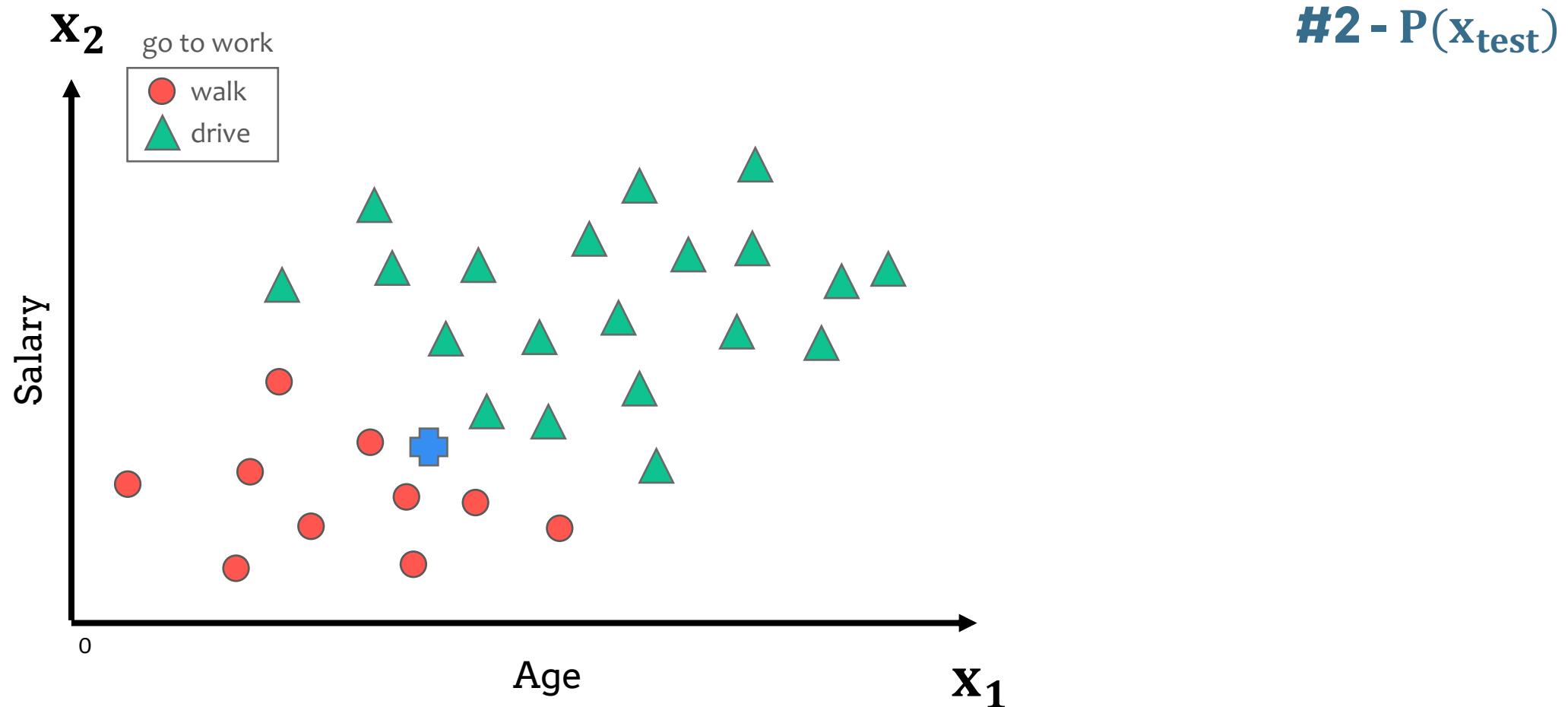
#3 Likelihood

Prior probability #1 (OK!)

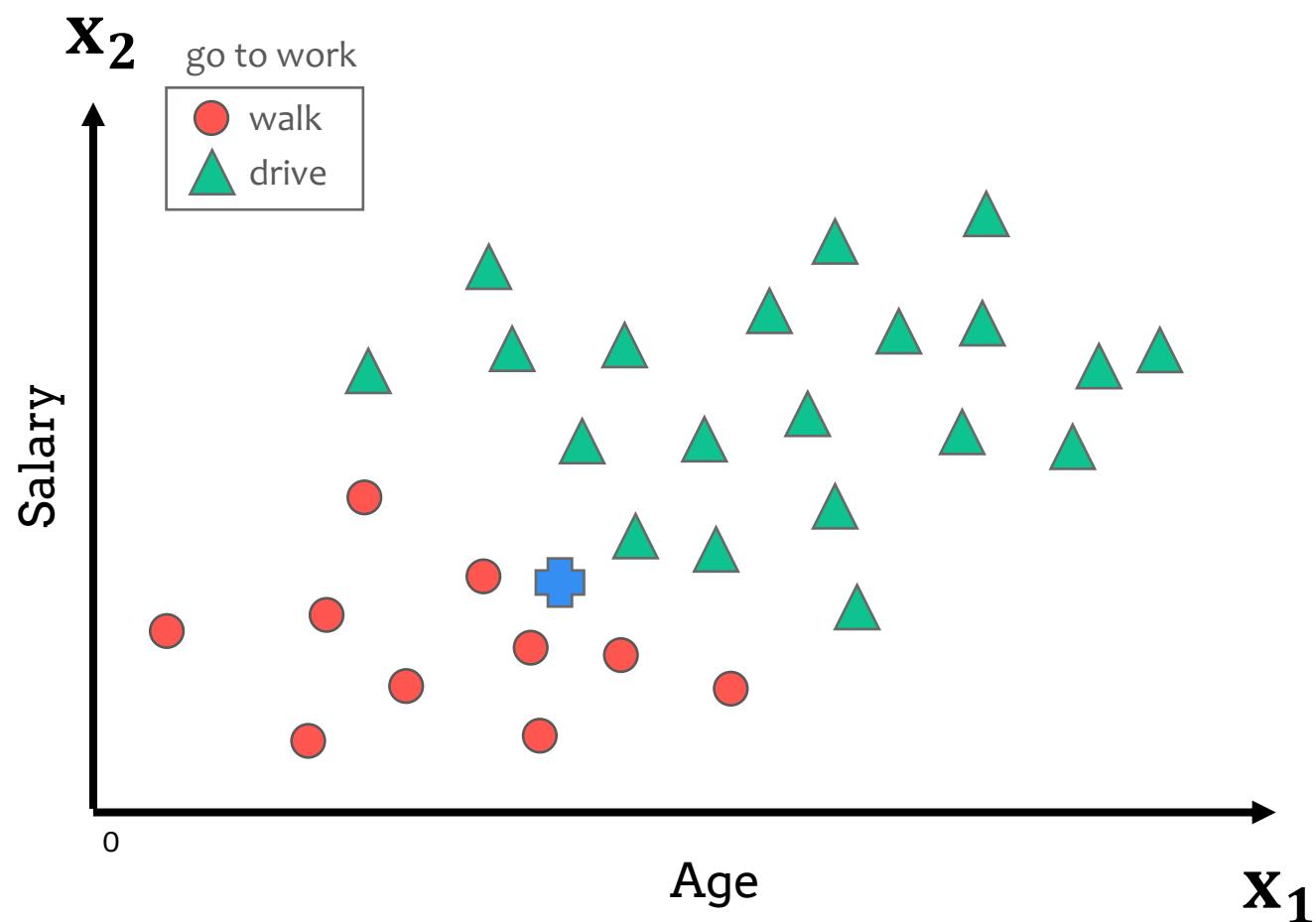
Marginal Likelihood #2

The diagram illustrates the components of the Naive Bayes formula. The numerator $P(x_{test}|Walks)$ is labeled '#3 Likelihood' with an arrow pointing to it. The denominator $P(x_{test})$ is labeled '#2 Marginal Likelihood' with an arrow pointing to it. The term $P(Walks)$ is labeled '#1 (OK!)' with an arrow pointing to it. A large red arrow points upwards from the bottom right towards the Marginal Likelihood component.

Naive Bayes: Step 2



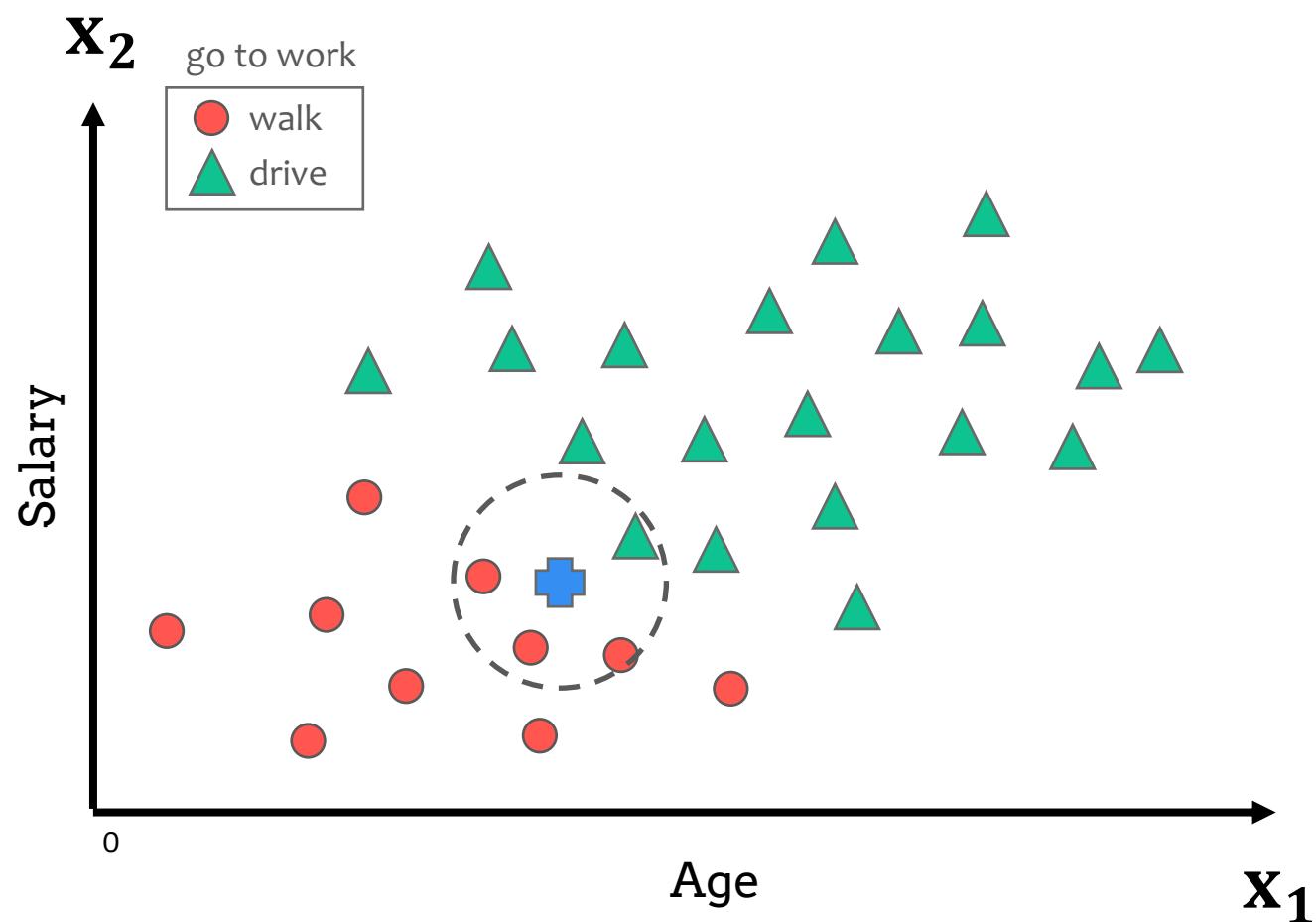
Naive Bayes: Step 2



#2 - $P(x_{\text{test}})$

What is the **(prior) probability** of a randomly selected **training sample** from our data will be similar to the **test sample**?

Naive Bayes: Step 2

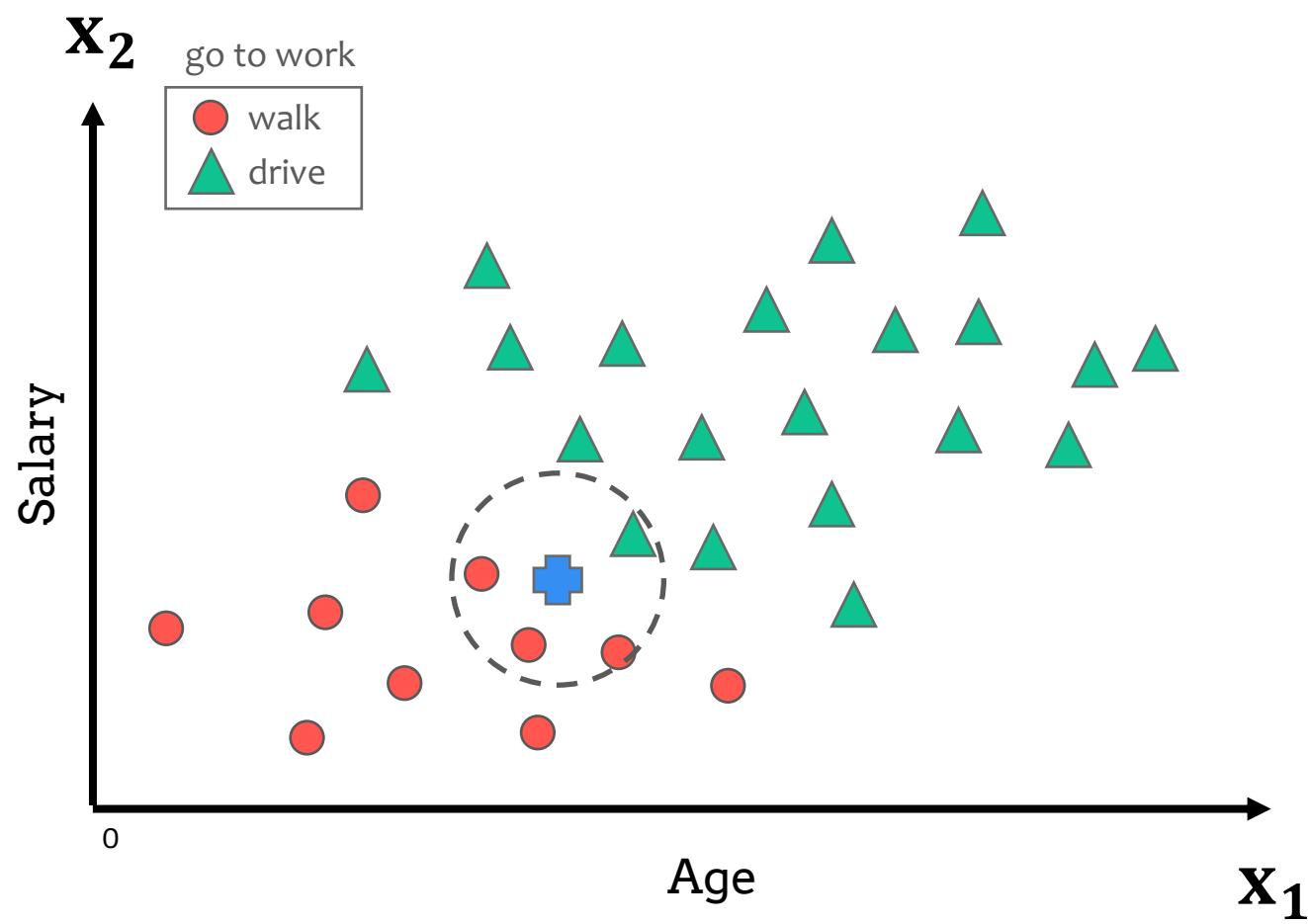


#2 - $P(x_{\text{test}})$

What is the **(prior) probability** of a randomly selected **training sample** from our data will be similar to the **test sample**?

Draw a circle (hypersphere) with a given radius around the **test sample**. The **training samples** inside the circle are **the most similar** to the **test sample**.

Naive Bayes: Step 2



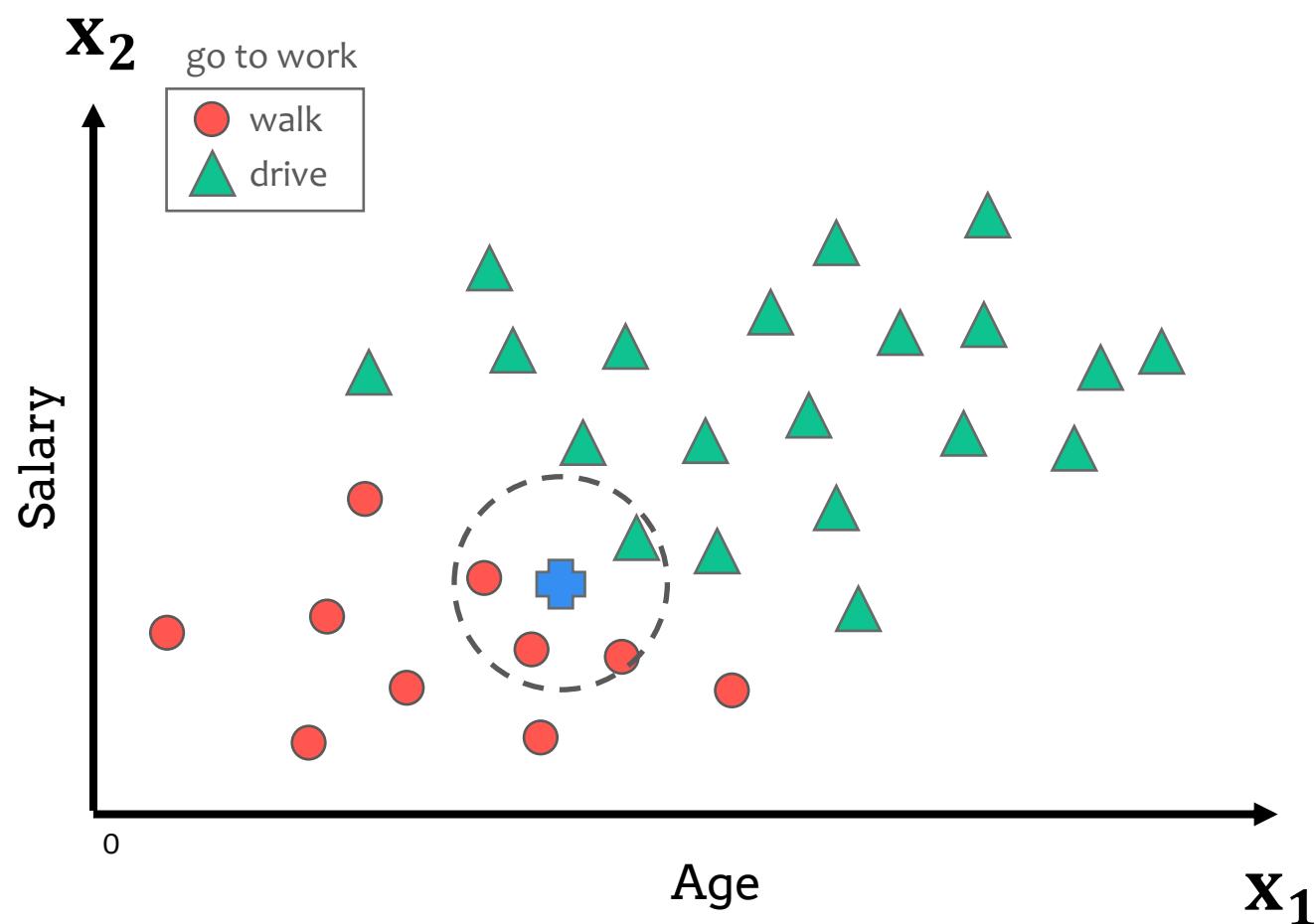
#2 - $P(x_{\text{test}})$

What is the **(prior) probability** of a randomly selected **training sample** from our data will be similar to the **test sample**?

Draw a circle (hypersphere) with a given radius around the **test sample**. The **training samples** inside the circle are **the most similar** to the **test sample**.

Now, count the number of training samples **inside this circle** and divide it by **the total number of training samples**.

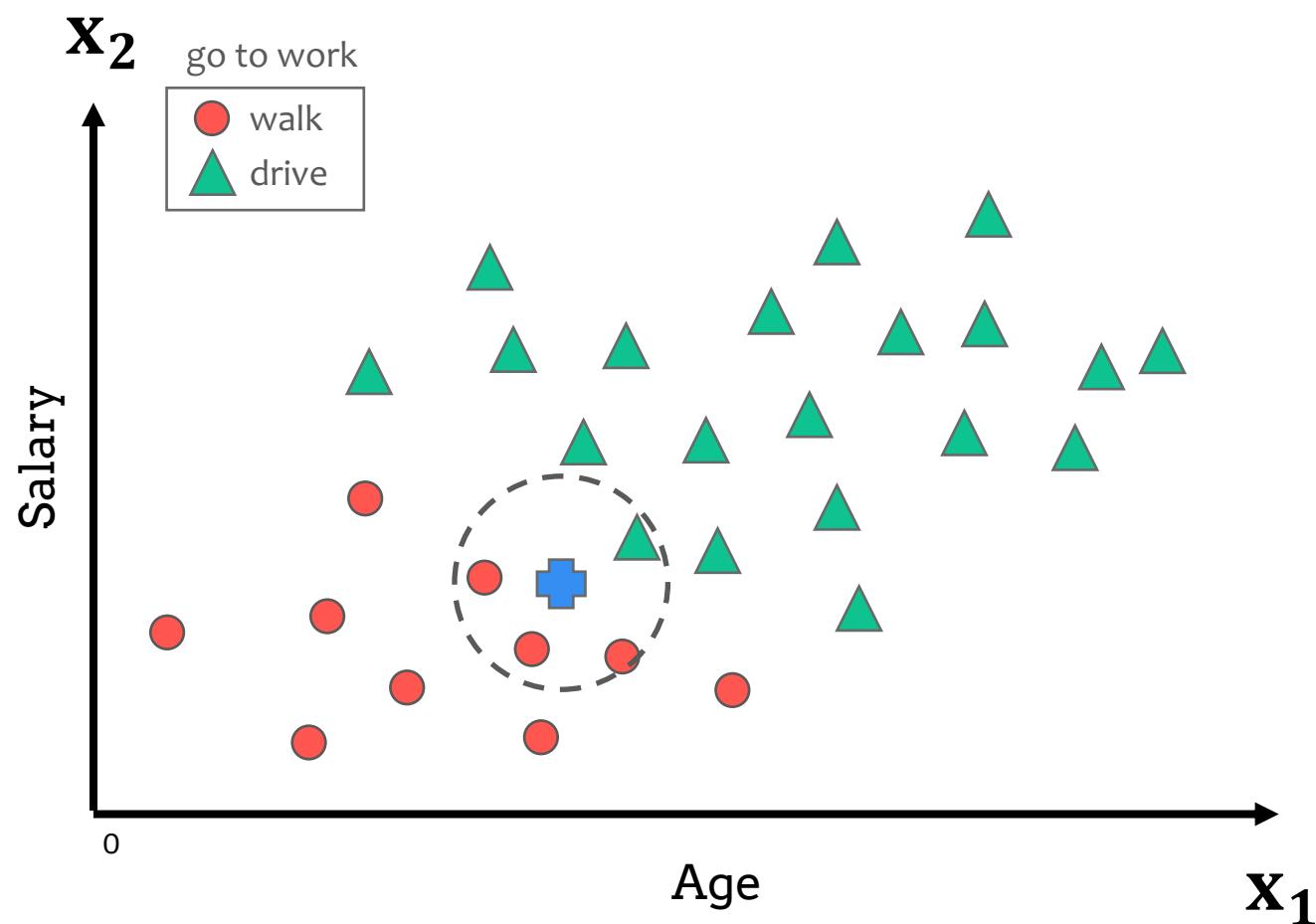
Naive Bayes: Step 2



#2 - $P(x_{\text{test}})$

$$P(x_{\text{test}}) = \frac{\text{Number of Similar Train. Samples}}{\text{Total Train. Samples}}$$

Naive Bayes: Step 2

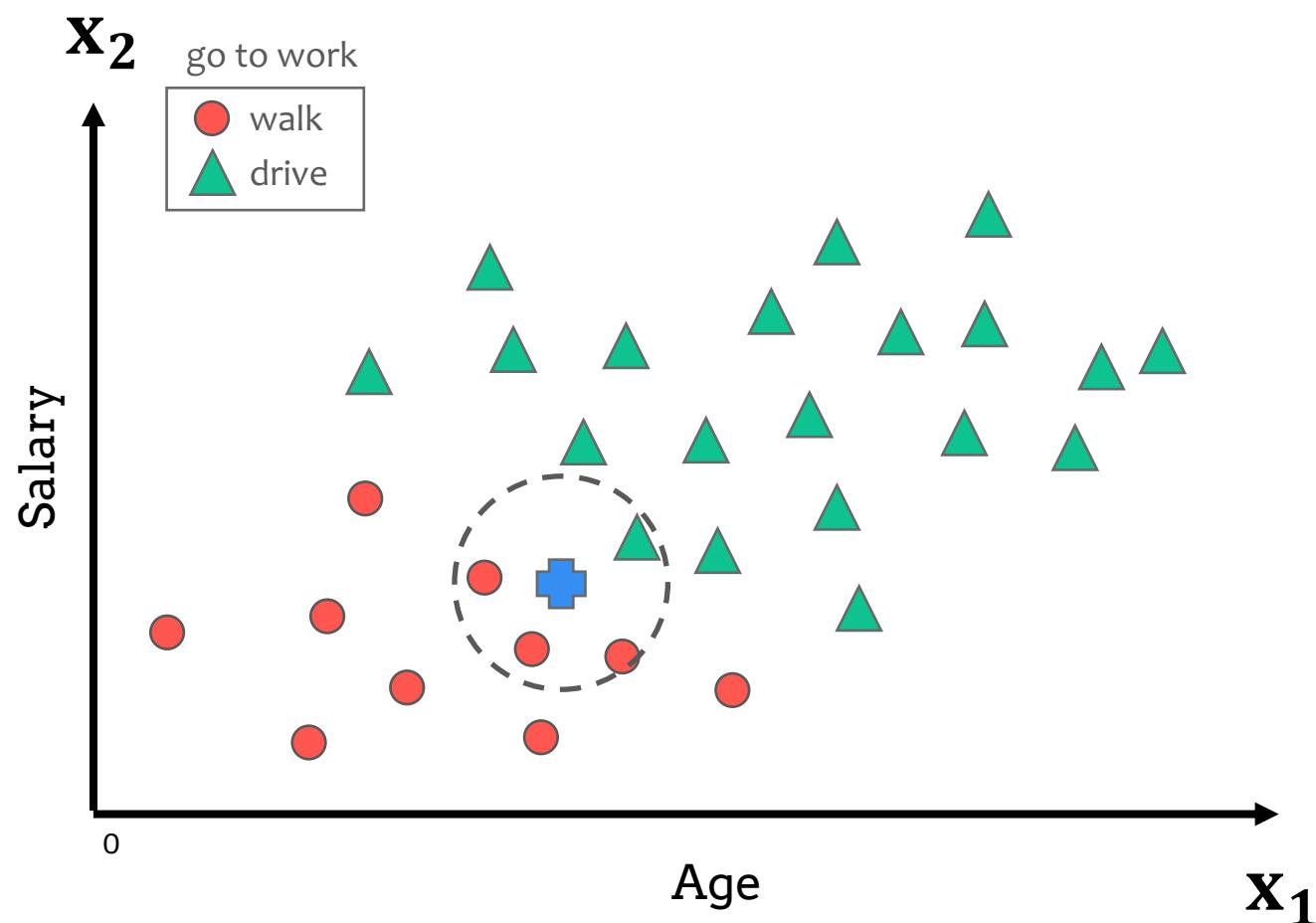


#2 - $P(x_{\text{test}})$

$$P(x_{\text{test}}) = \frac{\text{Number of Similar Train. Samples}}{\text{Total Train. Samples}}$$

$$P(x_{\text{test}}) = \frac{4}{30}$$

Naive Bayes: Step 2



#2 - $P(x_{\text{test}})$

$$P(x_{\text{test}}) = \frac{\text{Number of Similar Train. Samples}}{\text{Total Train. Samples}}$$

$$P(x_{\text{test}}) = \frac{4}{30}$$

We'll see soon we do not need to compute this probability for classification.

Naive Bayes

$$P(Walks|x_{test}) = \frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})}$$

class feats of the new sample

#4 Posterior probability

#3 Likelihood

Prior probability #1 (OK!)

Marginal Likelihood #2

The diagram illustrates the components of the Naive Bayes formula. The numerator $P(x_{test}|Walks)$ is labeled '#3 Likelihood' with an arrow pointing to it. The denominator $P(x_{test})$ is labeled '#2 Marginal Likelihood' with an arrow pointing to it. The term $P(Walks)$ is labeled '#1 (OK!)' with an arrow pointing to it. A large red arrow points upwards from the bottom right towards the Marginal Likelihood component.

Naive Bayes

$$P(Walks|x_{test}) = \frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})}$$

Diagram illustrating the components of the Naive Bayes formula:

- #1 (OK!)** Prior probability $P(Walks)$ (green text)
- #2 (OK!)** Marginal Likelihood $P(x_{test})$ (green text)
- #3 Likelihood** $P(x_{test}|Walks)$ (black text)
- #4 Posterior probability** $P(Walks|x_{test})$ (black text)

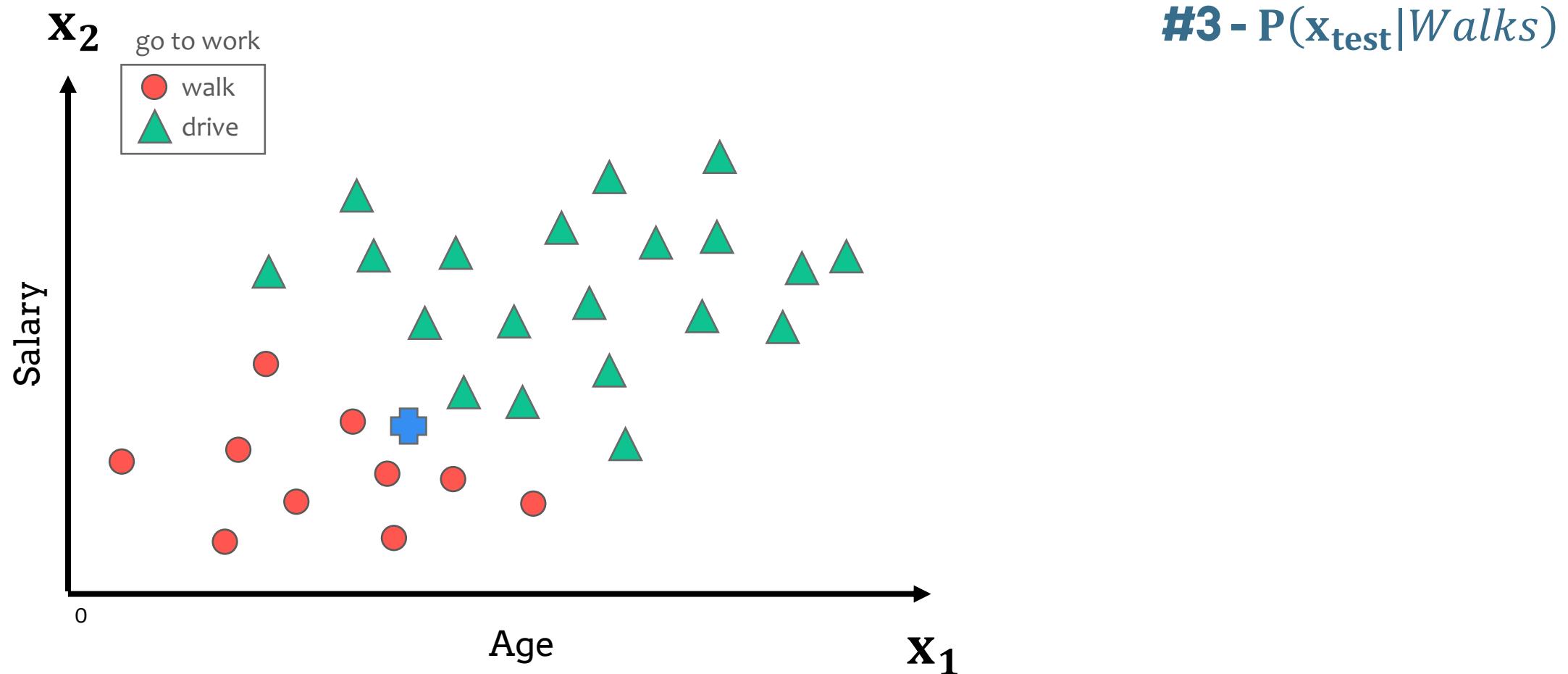
Annotations:

- A red arrow points down to the term $P(x_{test}|Walks)$, labeled **#3 Likelihood**.
- A green arrow points up to the term $P(Walks)$, labeled **#1 (OK!)**.
- A green arrow points down to the term $P(x_{test})$, labeled **#2 (OK!)**.
- A black arrow points up to the term $P(x_{test}|Walks)$, labeled **#3 Likelihood**.

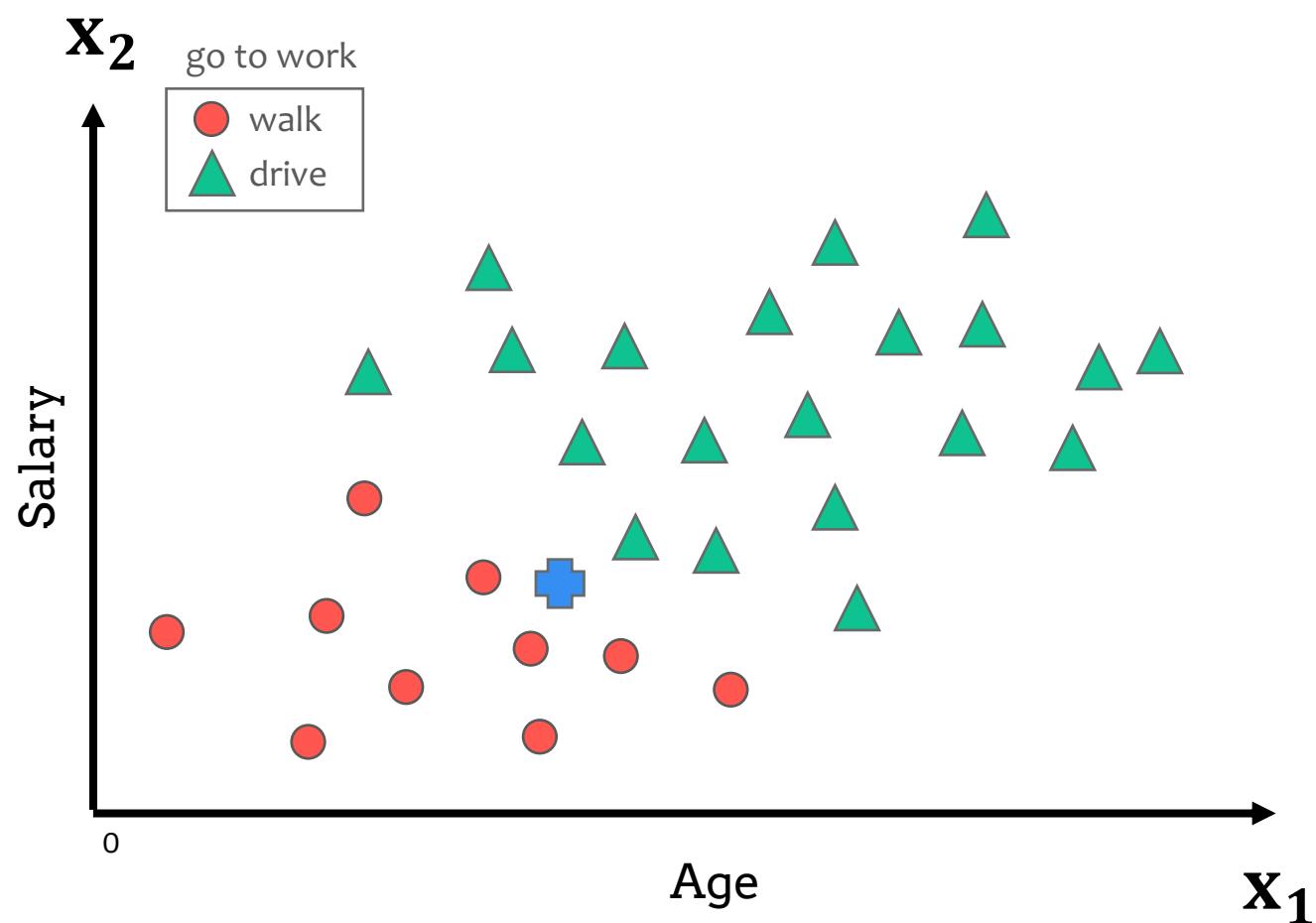
Labels for terms:

- class
- feats of the new sample

Naive Bayes: Step 3



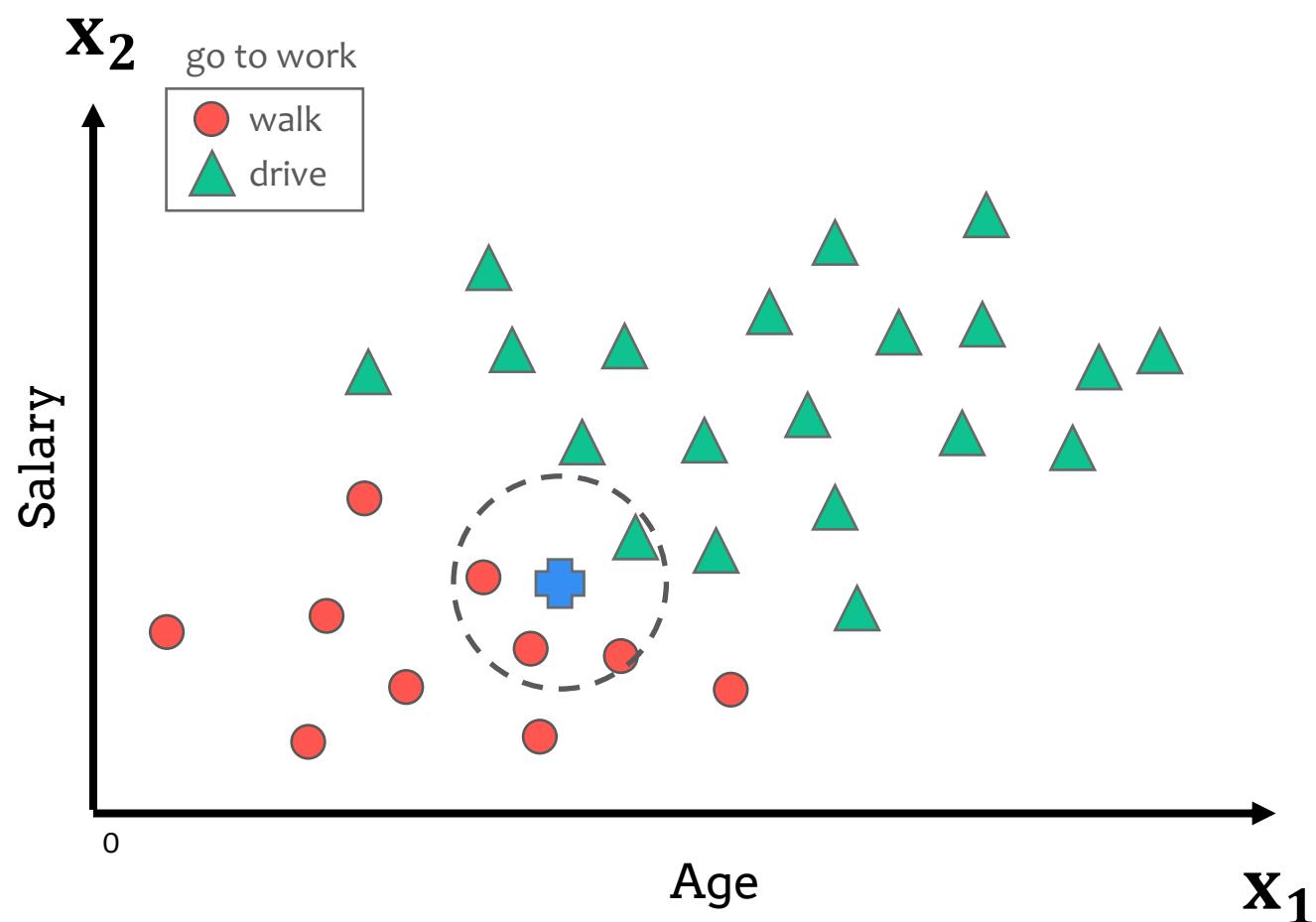
Naive Bayes: Step 3



#3 - $P(x_{\text{test}}|Walks)$

What is the **(prior) probability** of a randomly selected **training sample** from our data will be similar to the **test sample**, given it has the class **Walk**?

Naive Bayes: Step 3

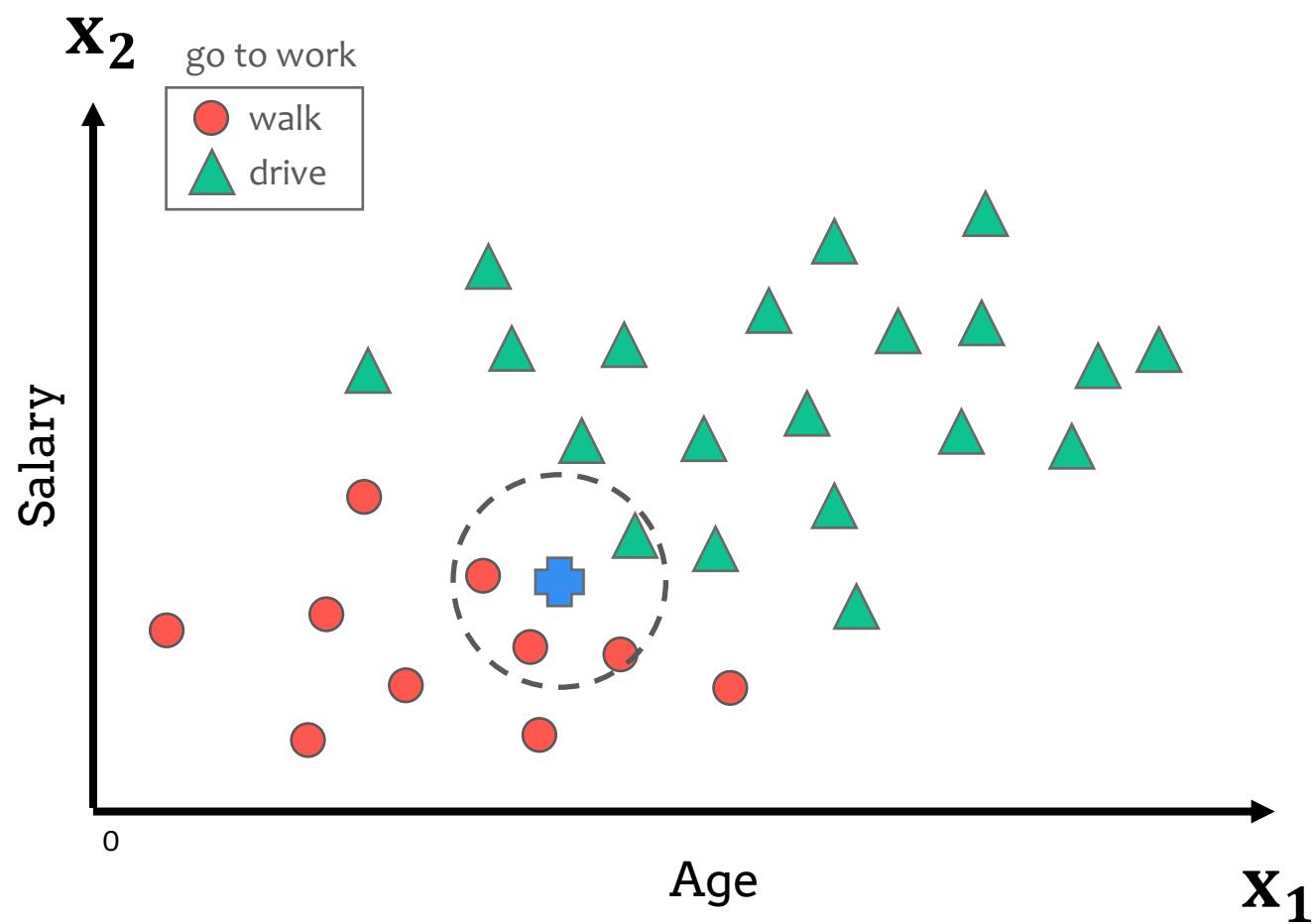


#3 - $P(x_{\text{test}}|Walks)$

What is the **(prior) probability** of a randomly selected **training sample** from our data will be similar to the **test sample**, given it has the class **Walk**?

Once again, we draw a circle (hypersphere) with a given radius around the **test sample**.

Naive Bayes: Step 3



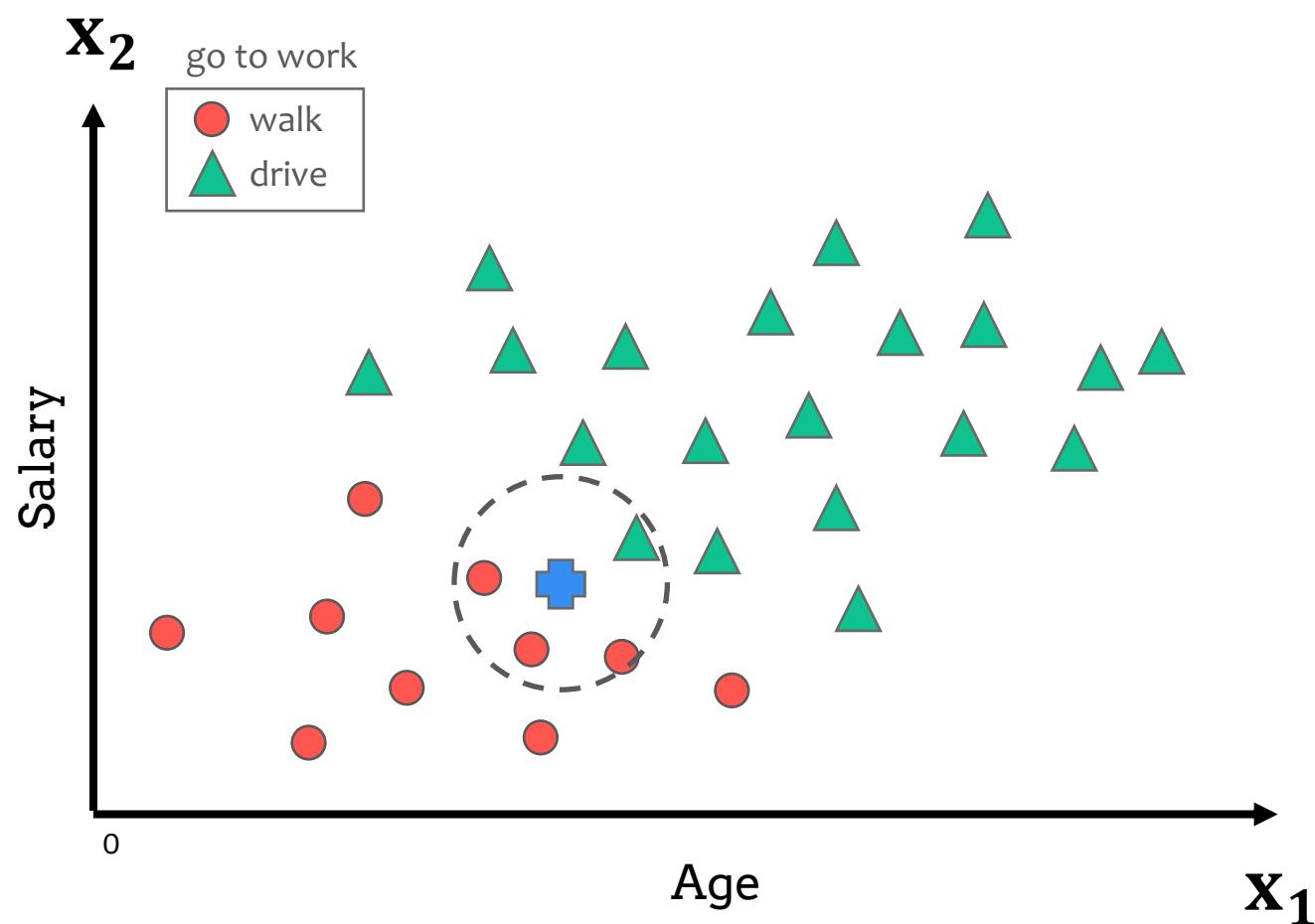
#3 - $P(x_{\text{test}}|Walks)$

What is the **(prior) probability** of a randomly selected **training sample** from our data will be similar to the **test sample**, given it has the class **Walk**?

Once again, we draw a circle (hypersphere) with a given radius around the **test sample**.

We then count the number of training samples **inside this circle** with class **Walk** and divide it by **the total number of training samples** with that class.

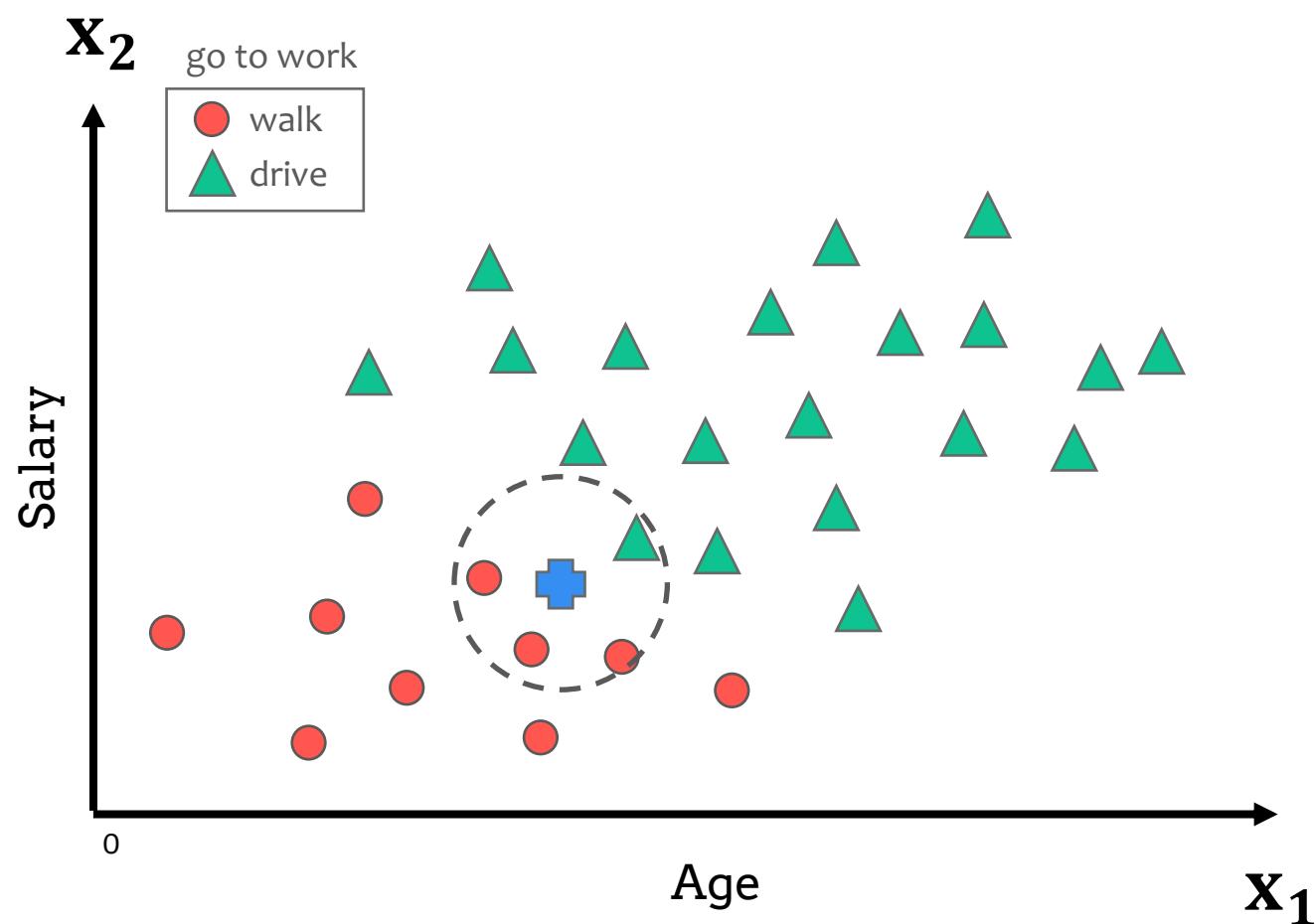
Naive Bayes: Step 3



#3 - $P(x_{\text{test}}|Walks)$

$$P(x_{\text{test}}|Walks) = \frac{\text{Number of similar train. samples with class Walk}}{\text{Total train. samples with class Walk}}$$

Naive Bayes: Step 3

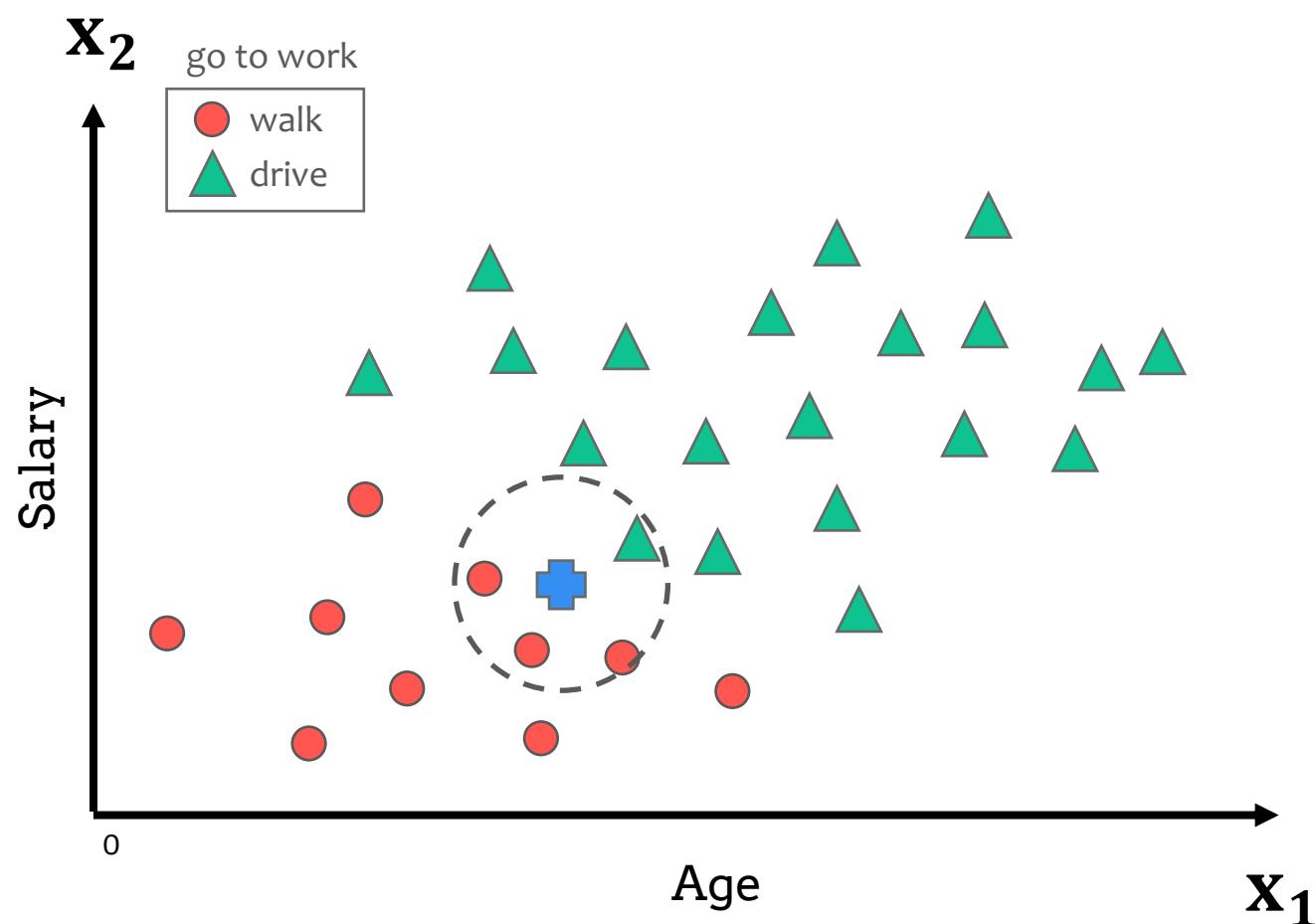


#3 - $P(x_{\text{test}}|Walks)$

$$P(x_{\text{test}}|Walks) = \frac{\text{Number of similar train. samples with class Walk}}{\text{Total train. samples with class Walk}}$$

$$P(x_{\text{test}}|Walks) = \frac{3}{10}$$

Naive Bayes: Step 3



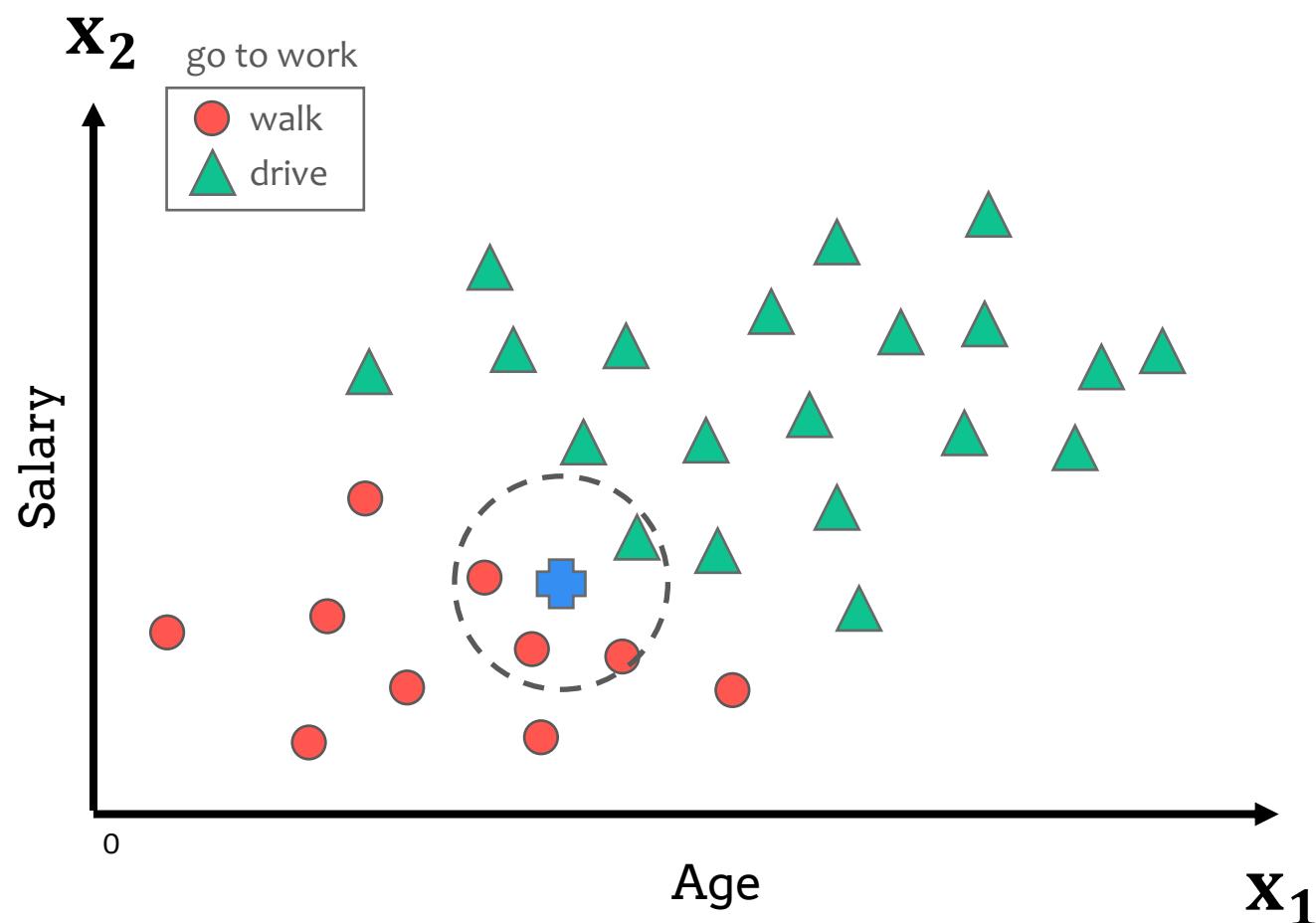
#3 - $P(x_{\text{test}}|Walks)$

$$P(x_{\text{test}}|Walks) = \frac{\text{Number of similar train. samples with class Walk}}{\text{Total train. samples with class Walk}}$$

$$P(x_{\text{test}}|Walks) = \frac{3}{10}$$

This is an **oversimplified** way to compute the **likelihood** for such a problem, being very sensitive to **outliers**.
The common way is to assume that **our data follows a given data distribution** (e.g., normal) and then use it to compute the **likelihood**.

Naive Bayes: Step 3



#3 - $P(x_{\text{test}}|Walks)$

$$P(x_{\text{test}}|Walks) = \frac{\text{Number of similar train. samples with class Walk}}{\text{Total train. samples with class Walk}}$$

$$P(x_{\text{test}}|Walks) = \frac{3}{10}$$

This is an **oversimplified** way to compute the **likelihood** for such a problem, being very sensitive to **outliers**.

The common way is to assume that **our data follows a given data distribution** (e.g., normal) and then use it to compute the **likelihood**.



We'll see soon another way of computing this **likelihood** through Gaussian Distribution.

Naive Bayes

$$P(Walks|x_{test}) = \frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})}$$

Diagram illustrating the components of the Naive Bayes formula:

- #3 Likelihood**: $P(x_{test}|Walks)$ (red text). An arrow points from this term to the numerator.
- Prior probability #1 (OK!)**: $P(Walks)$ (green text). An arrow points from this term to the numerator.
- Marginal Likelihood #2 (OK!)**: $P(x_{test})$ (green text). An arrow points from this term to the denominator.
- #4 Posterior probability**: $P(Walks|x_{test})$ (black text). An arrow points from the entire formula to this result.
- class**: x_{test} (black text). An arrow points from this term to the first argument of the formula.
- feats of the new sample**: x_{test} (black text). An arrow points from this term to the first argument of the formula.

Naive Bayes

$$P(Walks|x_{test}) = \frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})}$$

class feats of the new sample

#4 Posterior probability

#3 (OK!) Likelihood

Prior probability #1 (OK!)

Marginal Likelihood #2 (OK!)

The diagram illustrates the components of the Naive Bayes formula. At the top right, the term "Likelihood" is labeled with "#3 (OK!)" in green. Below it, the term "Prior probability" is labeled with "#1 (OK!)" in green. To the left of the fraction, the term "Marginal Likelihood" is labeled with "#2 (OK!)" in green. The fraction itself is labeled with "feats of the new sample" above the numerator and "class" above the denominator. A large red arrow points upwards from the bottom left towards the fraction, indicating the flow of computation from the features of the new sample through the likelihood and prior probability to the final posterior probability.

Naive Bayes: Step 4

$$P(Walks|x_{test}) = \frac{\frac{3}{10} * \frac{10}{30}}{\frac{4}{30}} = 0.75$$

Naive Bayes

Step 2

Now, repeat the same calculation for the other class: *Drives*

$$P(Drives|x_{test}) = \frac{P(x_{test}|Drives) * P(Drives)}{P(x_{test})}$$

The diagram illustrates the components of the Bayes' theorem formula:

- #1** Prior probability $P(Drives)$
- #2** Marginal Likelihood $P(x_{test})$
- #3** Likelihood $P(x_{test}|Drives)$
- #4** Posterior probability $P(Drives|x_{test})$

Annotations provide context for each term:

- x_{test} is labeled as "feats of the new sample".
- "class" is labeled under $P(Drives)$.

Naive Bayes

$$P(\text{Drives} | \mathbf{x}_{test}) = \frac{\frac{1}{20} * \frac{20}{30}}{\frac{4}{30}} = 0.25$$

Naive Bayes

Step 3

$P(Walks|x_{test})$ vs $P(Drives|x_{test})$

Naive Bayes

Step 3

$P(Walks|x_{test})$ vs $P(Drives|x_{test})$

0.75

0.25

Naive Bayes

Step 3

$P(Walks|x_{test})$ vs $P(Drives|x_{test})$

0.75 0.25

one is complement of the other

Naive Bayes

Step 3

$$P(\text{Walks} | \mathbf{x}_{test}) \text{ vs } P(\text{Drives} | \mathbf{x}_{test})$$

0.75 0.25

one is complement of the other

The **test sample** is classified as **Walker**.

Why Naive?

Why Naive?

It considers that **all variables/features** are **mutually independent**, which is not always true.

Why Naive?

It considers that **all variables/features** are **mutually independent**, which is not always true.

$$P(x_{test} | \text{Walks}) = P(x_{1_{test}} | \text{Walks}) \cdot P(x_{2_{test}} | \text{Walks}) \cdot \dots \cdot P(x_{n_{test}} | \text{Walks})$$

Why Naive?

It considers that **all variables/features** are **mutually independent**, which is not always true.

$$P(x_{test} | \text{Walks}) = P(x_{1_{test}} | \text{Walks}) \cdot P(x_{2_{test}} | \text{Walks}) \cdot \dots \cdot P(x_{n_{test}} | \text{Walks})$$

$$P(x_{test} | \text{Walks}) = \prod_{i=1}^n P(x_{i_{test}} | \text{Walks})$$

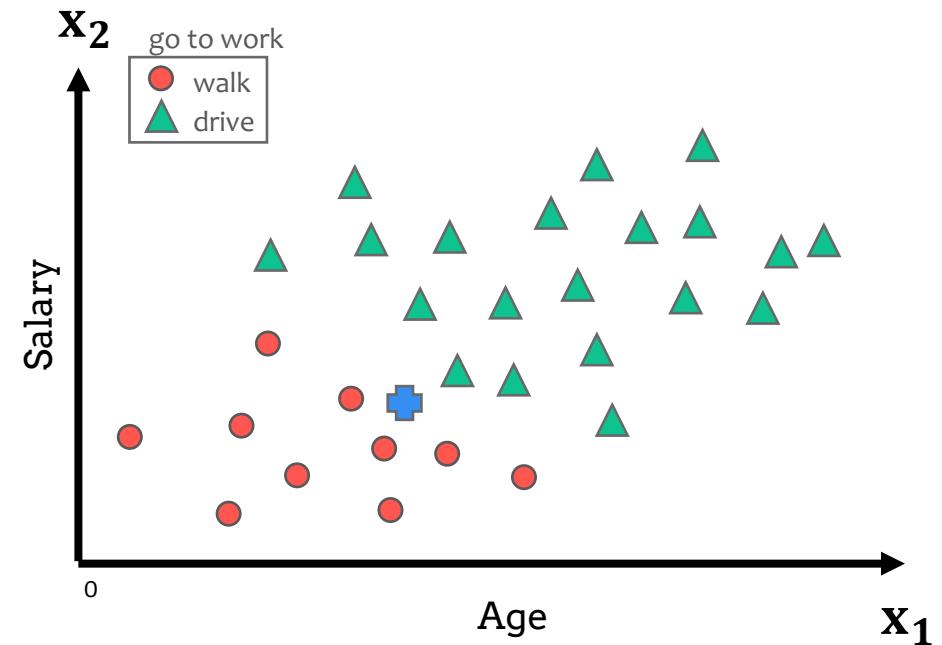
Why Naive?

It considers that **all variables/features** are **mutually independent**, which is not always true.

$$P(x_{test} | \text{Walks}) = P(x_{1test} | \text{Walks}) \cdot P(x_{2test} | \text{Walks}) \cdot \dots \cdot P(x_{n_{test}} | \text{Walks})$$

$$P(x_{test} | \text{Walks}) = \prod_{i=1}^n P(x_{i_{test}} | \text{Walks})$$

For example, in our problem, both features seems **to be correlated**.



Why Naive?

It considers that **all variables/features** are **mutually independent**, which is not always true.

$$P(x_{test} | \text{Walks}) = P(x_{1_{test}} | \text{Walks}) \cdot P(x_{2_{test}} | \text{Walks}) \cdots P(x_{n_{test}} | \text{Walks})$$

Idiot's Bayes—Not So Stupid After All?

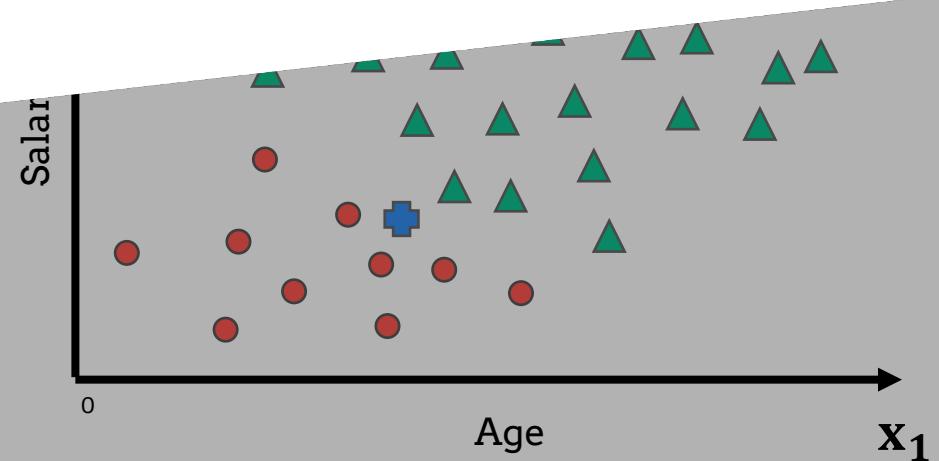
$$P(x_{te...})$$

David J. Hand¹ and Keming Yu²

¹Department of Mathematics, Imperial College, London, UK. E-mail: d.j.hand@ic.ac.uk

²University of Plymouth, UK

For example, in our problem, both features seems **to be correlated**.



Do we need to compute $P(x_{\text{test}})$?

Do we need to compute $P(x_{test})$?

$P(Walks|x_{test})$ vs $P(Drives|x_{test})$

Do we need to compute $P(x_{test})$?

$P(Walks|x_{test})$ vs $P(Drives|x_{test})$

$$\frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})} \text{ vs } \frac{P(x_{test}|Drives) * P(Drives)}{P(x_{test})}$$

Do we need to compute $P(x_{test})$?

$P(Walks|x_{test})$ vs $P(Drives|x_{test})$

$$\frac{P(x_{test}|Walks) * P(Walks)}{\cancel{P(x_{test})}} \text{ vs } \frac{P(x_{test}|Drives) * P(Drives)}{\cancel{P(x_{test})}}$$

Aprendizado de Máquina e Reconhecimento de Padrões



Naive Bayes

Prof. Dr. Samuel Martins (Samuka)
samuel.martins@ifsp.edu.br

