Aprendizado de Máquina e Reconhecimento de Padrões

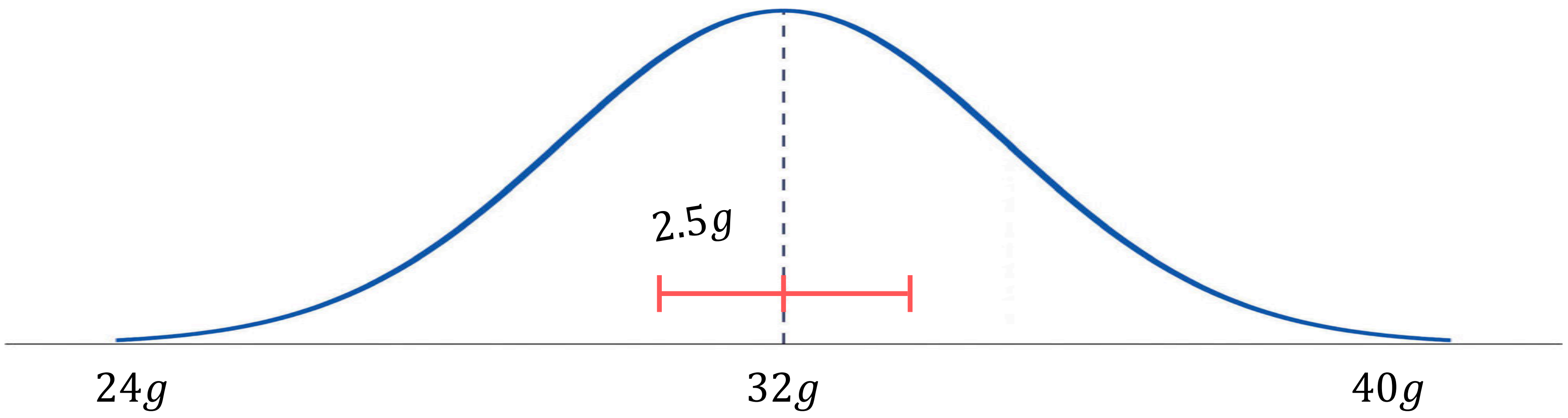
Naive Bayes

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

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
São Paulo
Campus Campinas

Preliminaries: Probability vs Likelihood

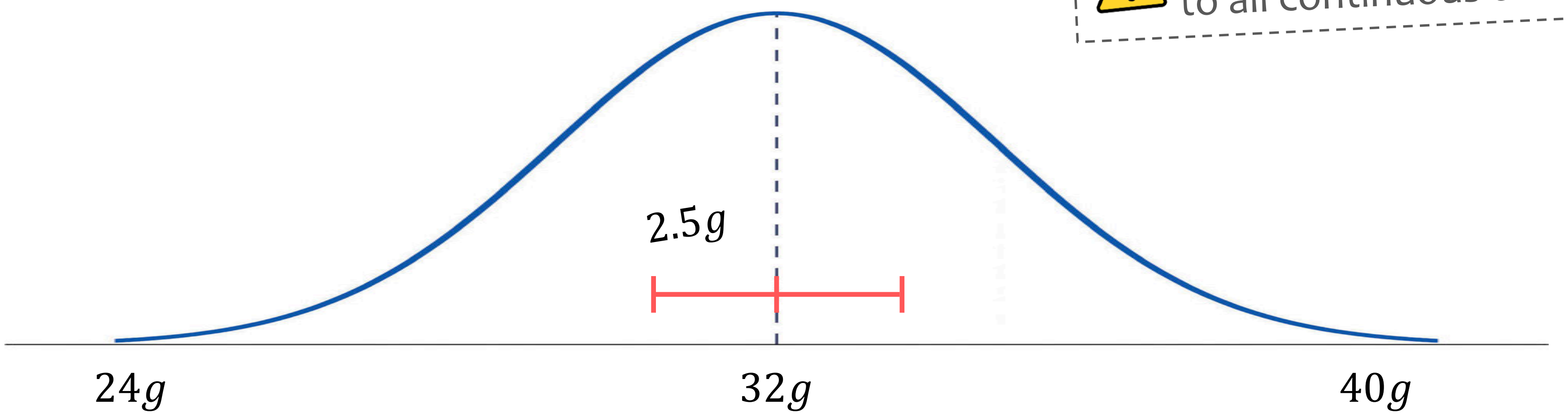
Let's assume that the weights of a mouse population follow a **Normal Distribution** $\mathbf{N(\mu, \sigma) = N(32g, 2.5g)}$



$$\mathbf{N(32g, 2.5g)}$$

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

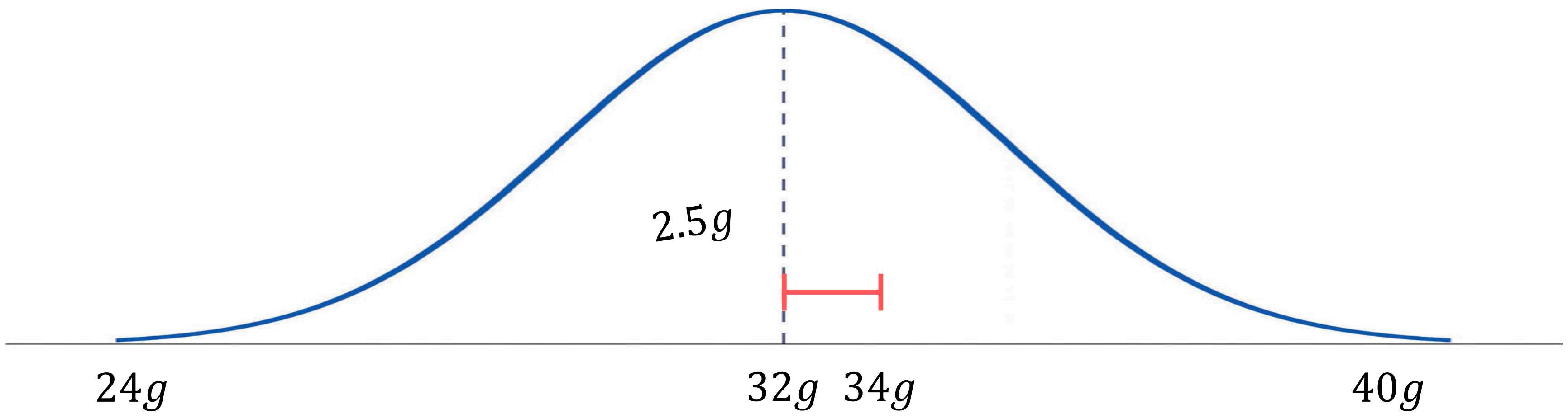
The concepts explained here apply to all continuous distributions.



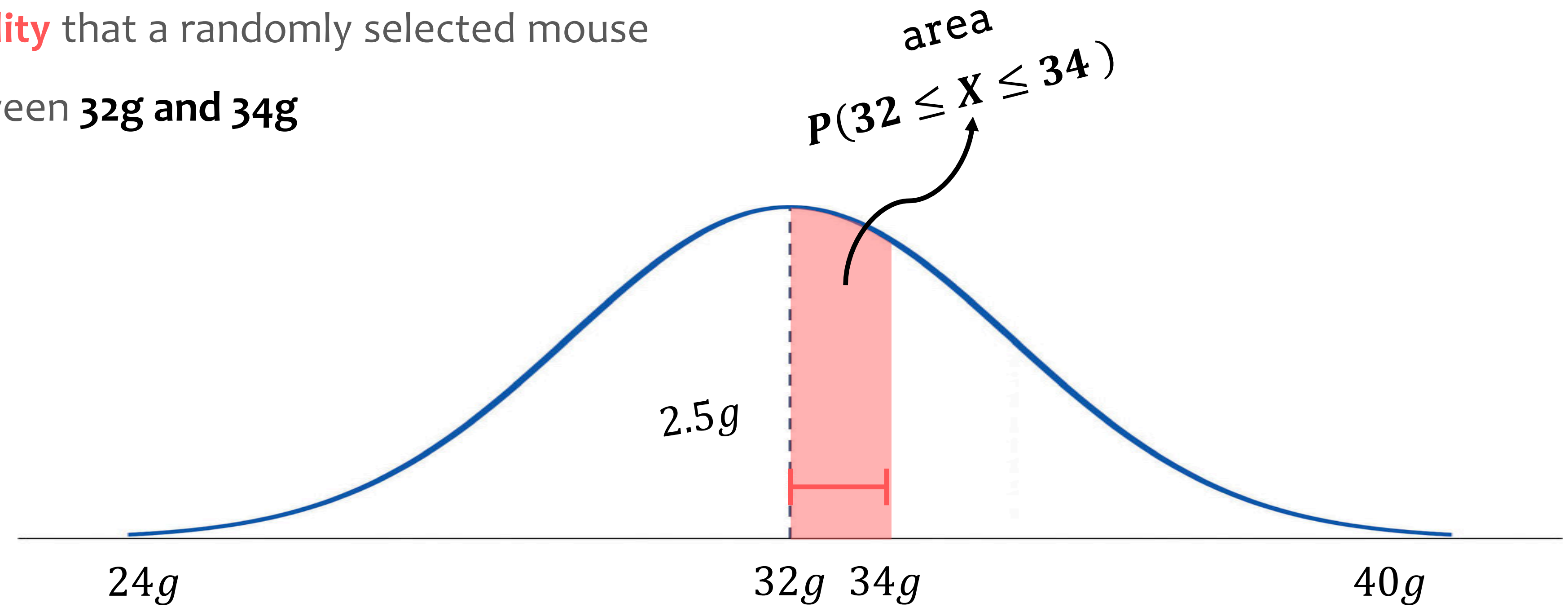
$$\mathbf{N(32g, 2.5g)}$$

The **probability** that a randomly selected mouse

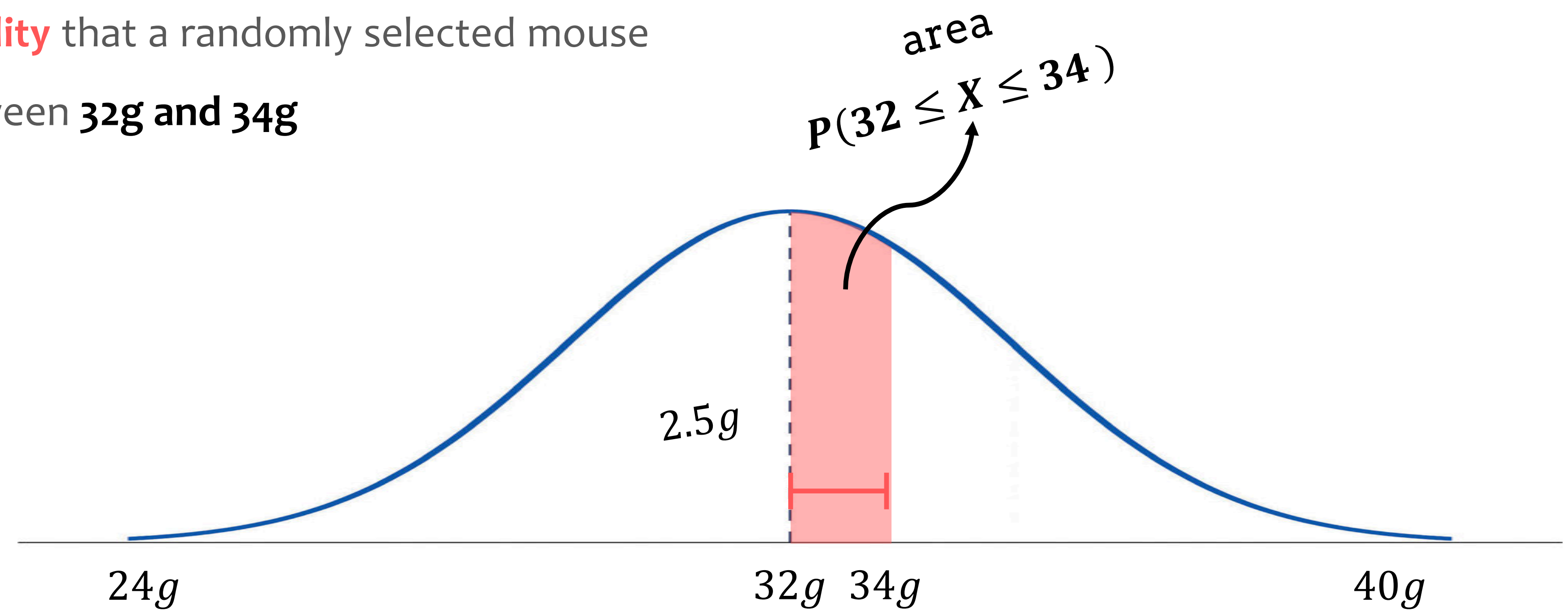
weighs between **32g and 34g**



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



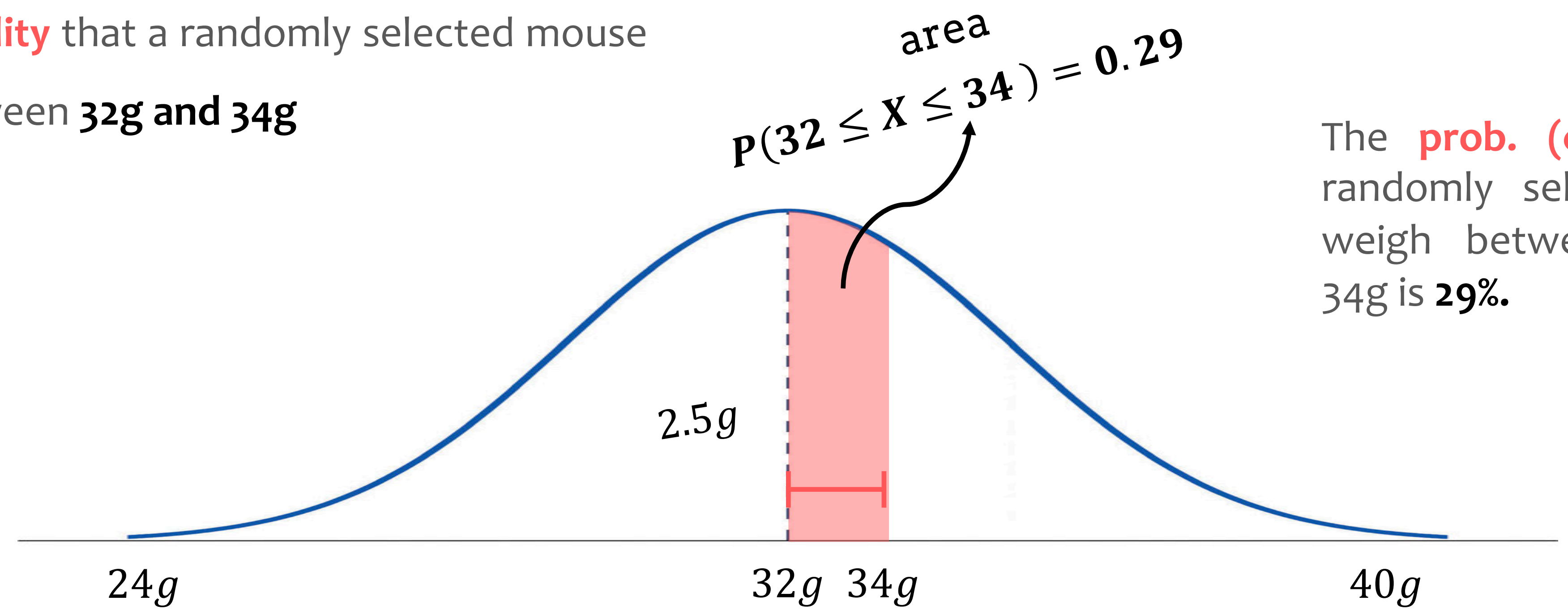
The **probability** that a randomly selected mouse weighs between **32g and 34g**



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

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

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

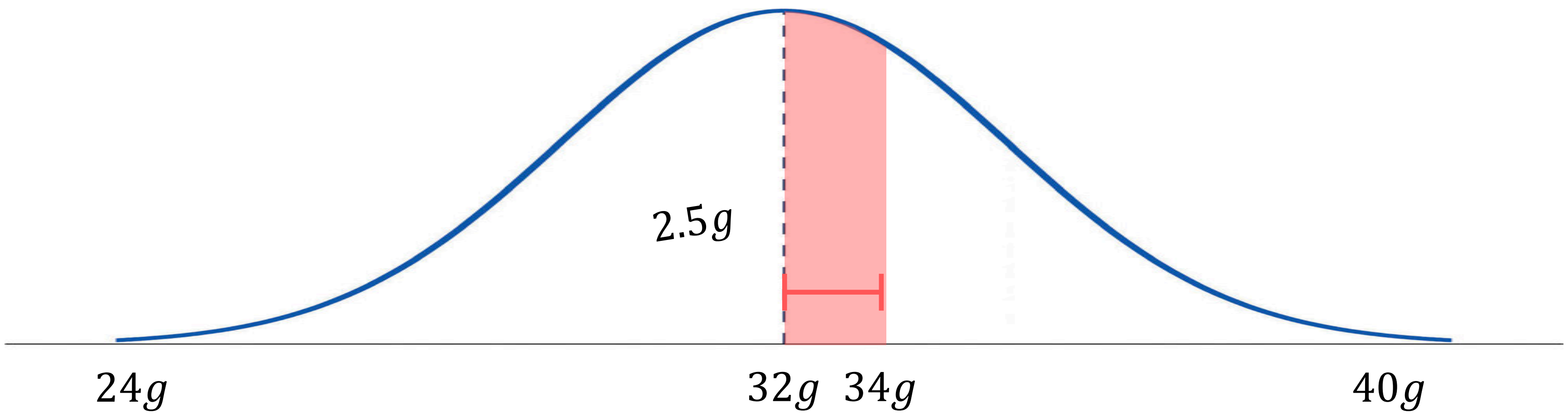


The **prob. (chance)** of a randomly selected mouse weigh between 32g and 34g is **29%.**

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

Mathematically, we use the following notation:

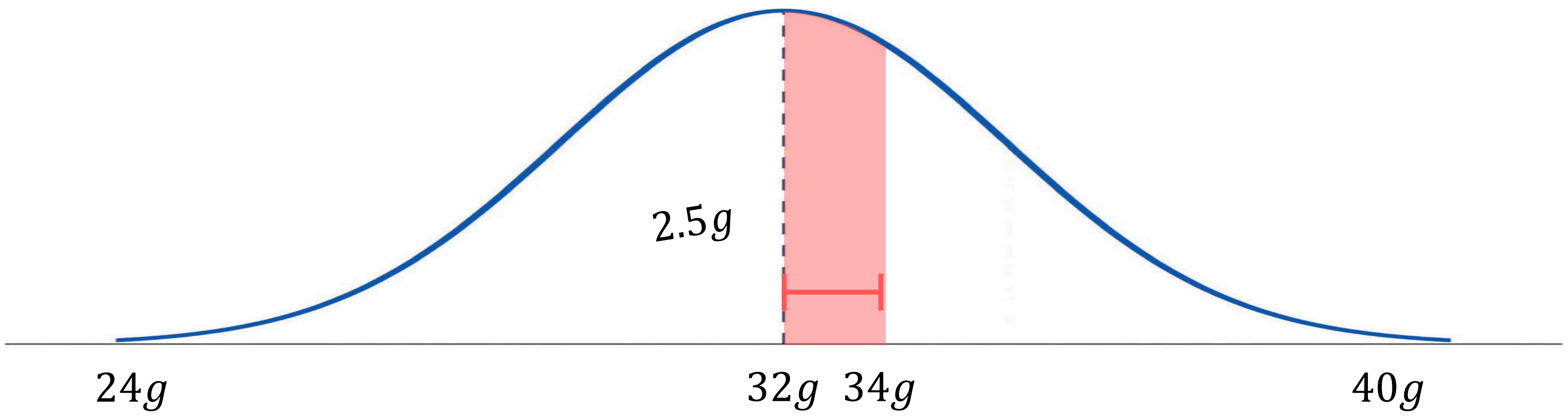
$$P(weight\ between\ 32g\ and\ 34g \mid mean = 32g\ and\ std = 2.5g) = 0.29$$



Mathematically, we use the following notation:

$$P(weight\ between\ 32g\ and\ 34g\ |\ mean = 32g\ and\ std = 2.5g) = 0.29$$

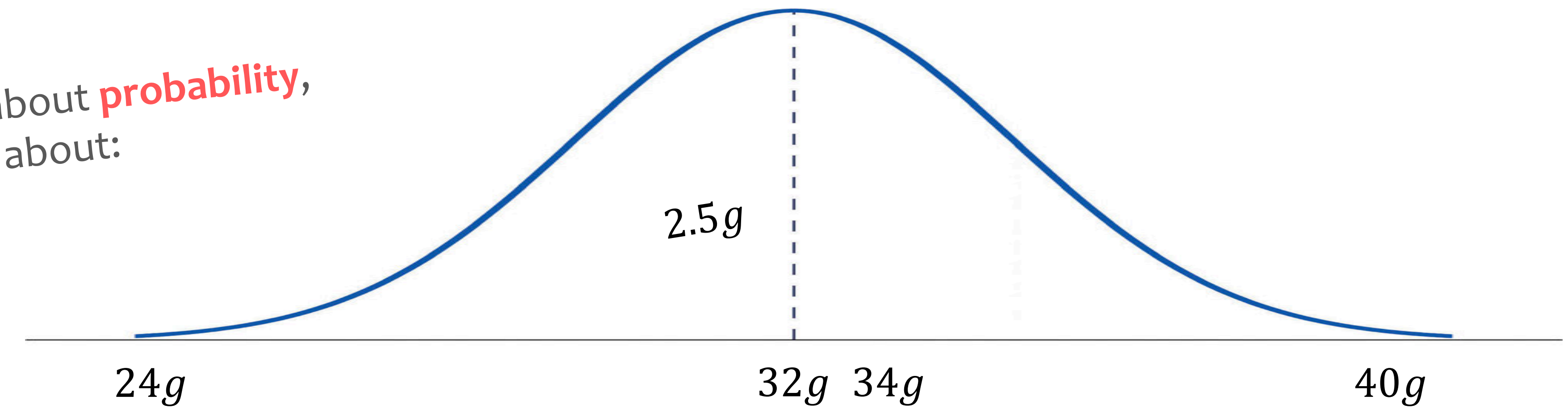
$$P(32 \leq X \leq 34\ |\ \mu = 32, \sigma = 2.5) = 0.29$$



$$P(weight\ between\ 32g\ and\ 34g\ |\ mean = 32g\ and\ std = 2.5g) = 0.29$$

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

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

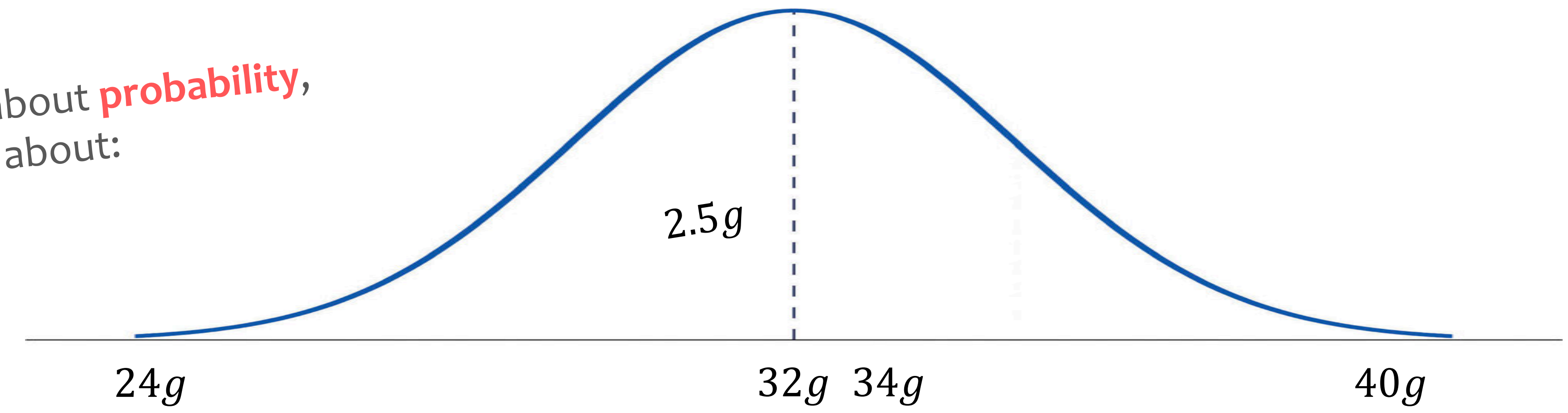


A **distribution** described by this side of the equation.

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

$$P(32 \leq X \leq 34\ |\ \boxed{\mu = 32, \sigma = 2.5}) = 0.29$$

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

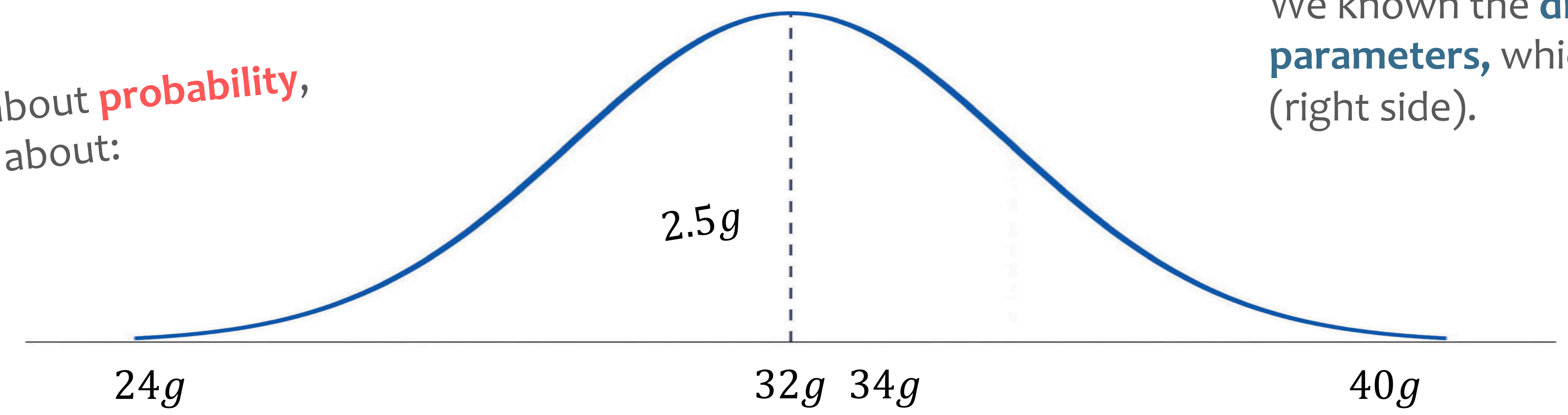


A **distribution** described by this side of the equation.

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

$$P(32 \le X \le 34 \mid \boxed{\mu = 32, \sigma = 2.5}) = 0.29$$

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



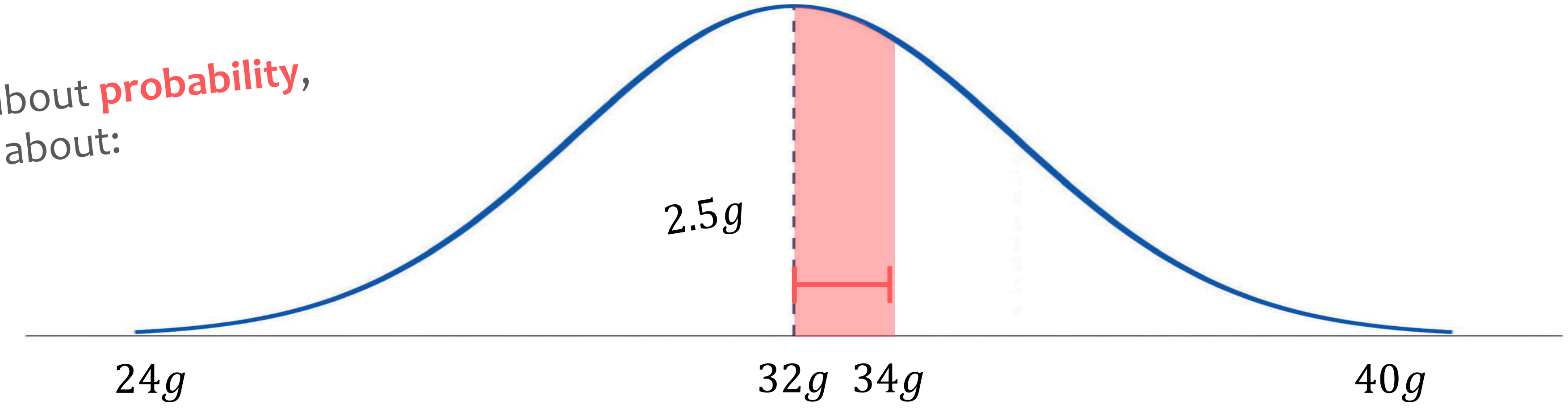
We known the **distribution parameters,** which are **fixed** (right side).

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

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

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

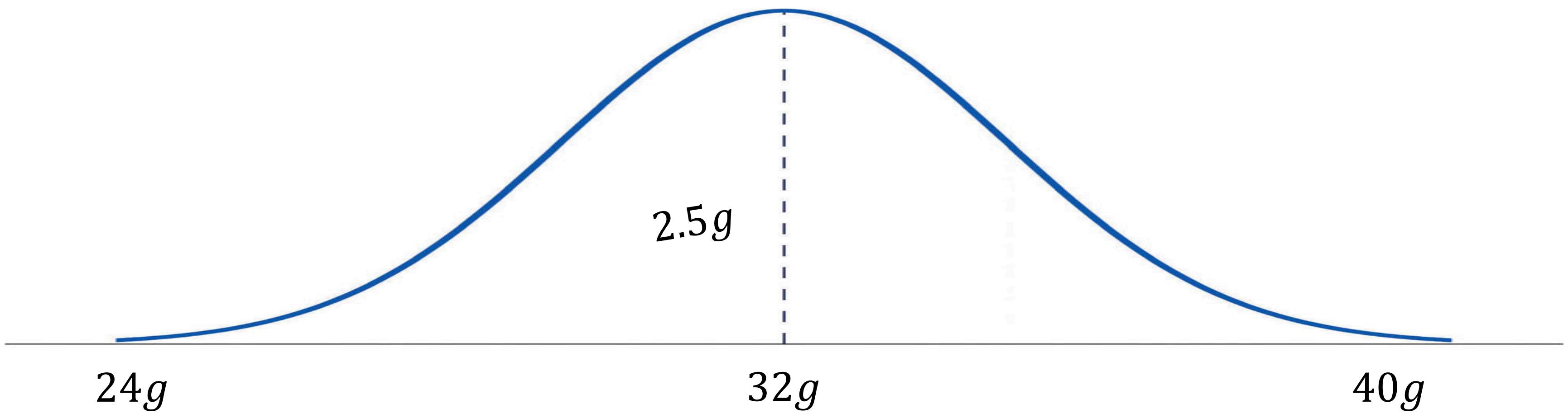
When talking about **probability**,
we are talking about:



Using the same **distribution...**

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

$$P(32 \leq X \leq 34\ |\ \boxed{\mu = 32, \sigma = 2.5}) = 0.29$$

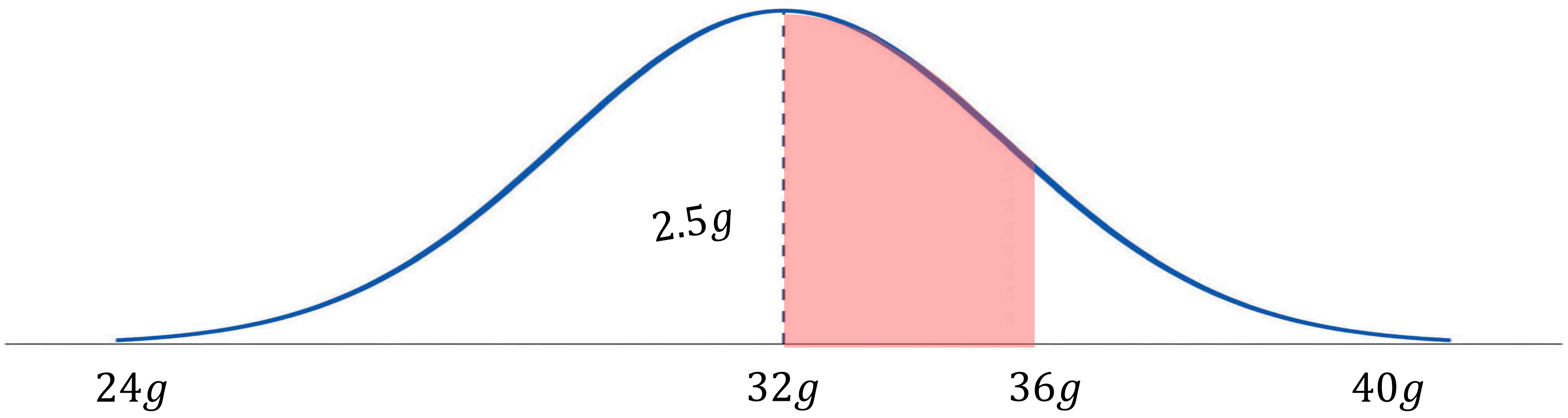


… we can change the left side of the equation to get a **new probability**!

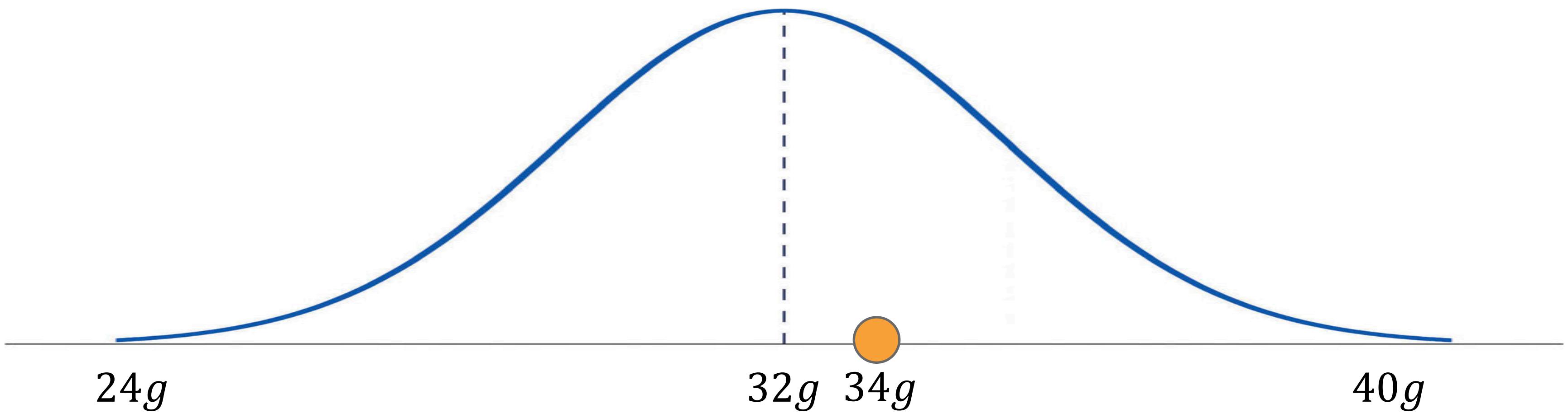
Using the same **distribution…**

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

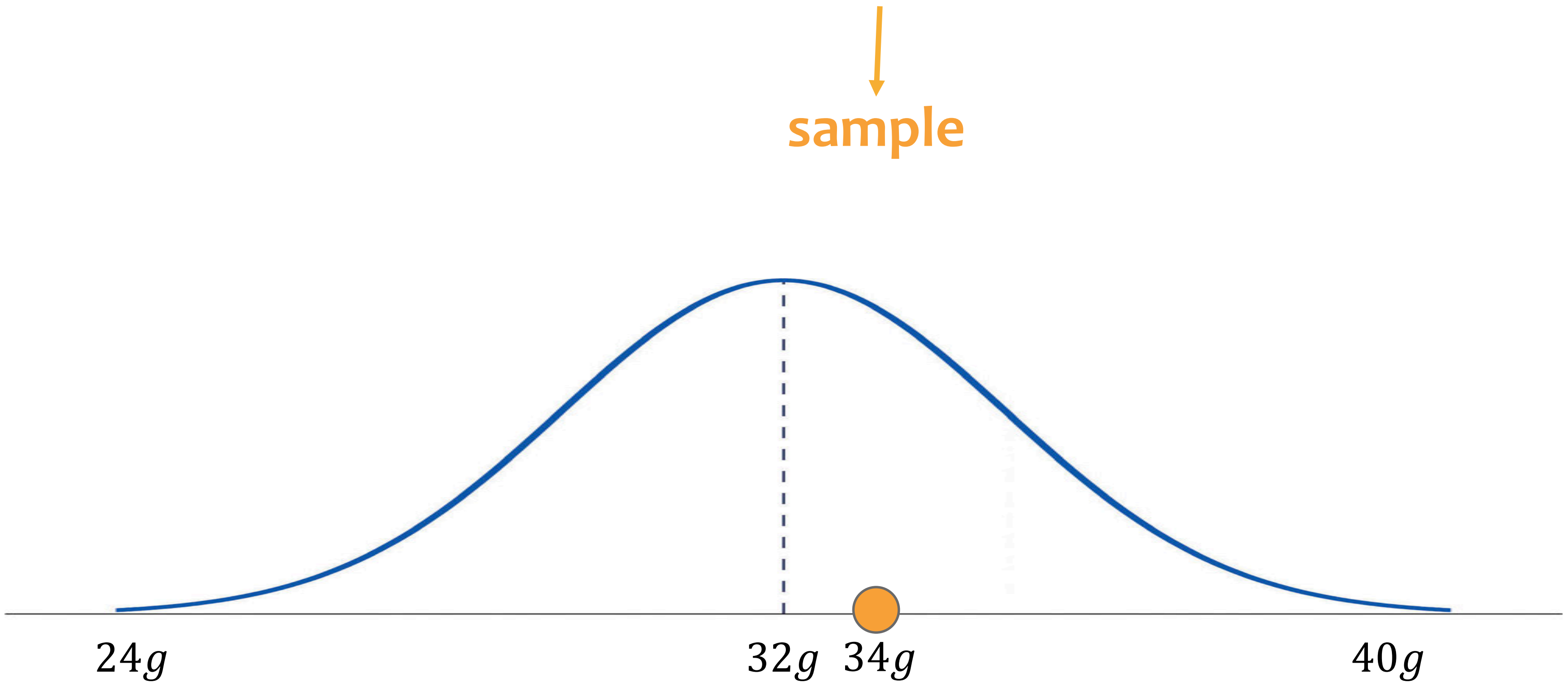
$$P(\boxed{32 \leq X \leq 36} \mid \boxed{\mu = 32, \sigma = 2.5}) = 0.44$$



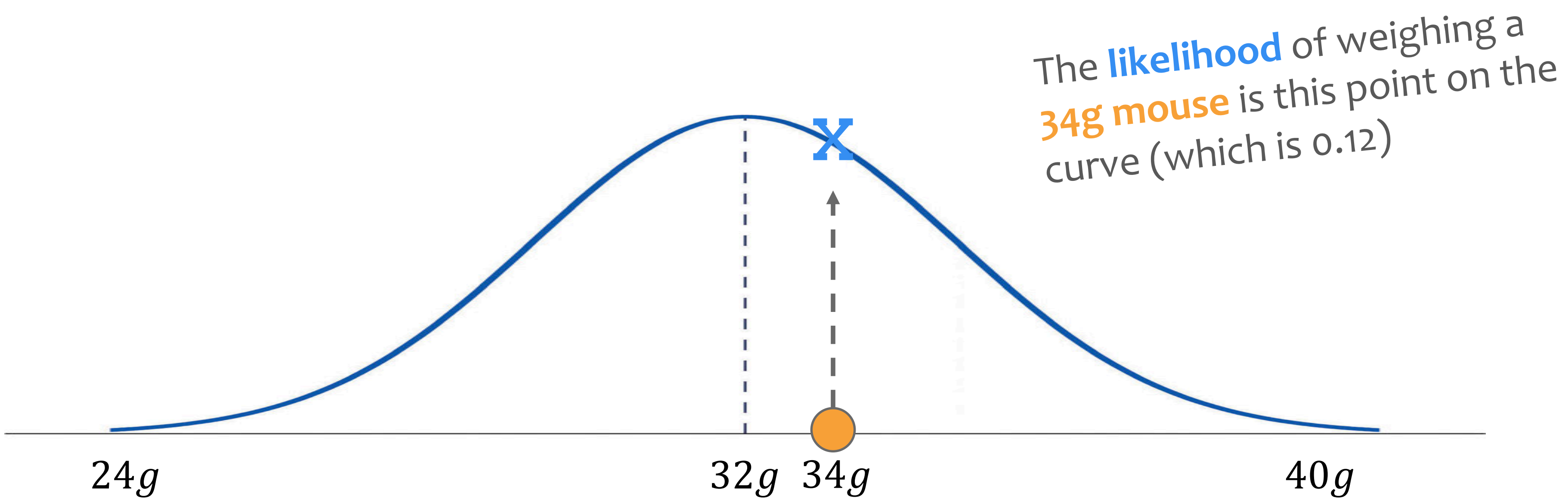
In case of **likelihood**, we assume that we already have selected the mouse (e.g., weighing 34g).



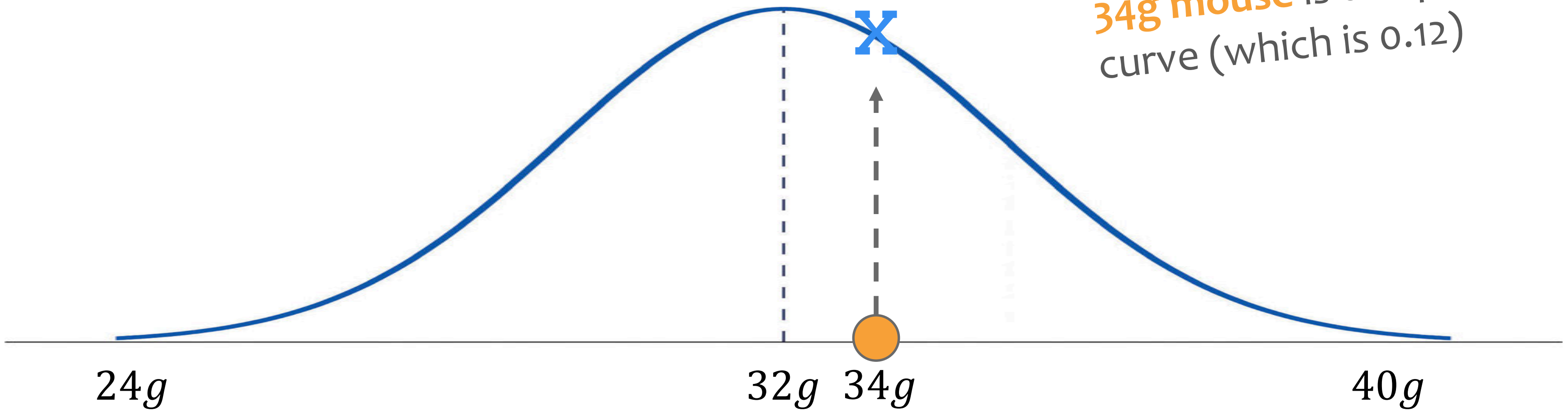
In case of **likelihood**, we assume that we already have selected the mouse (e.g., weighing 34g).



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



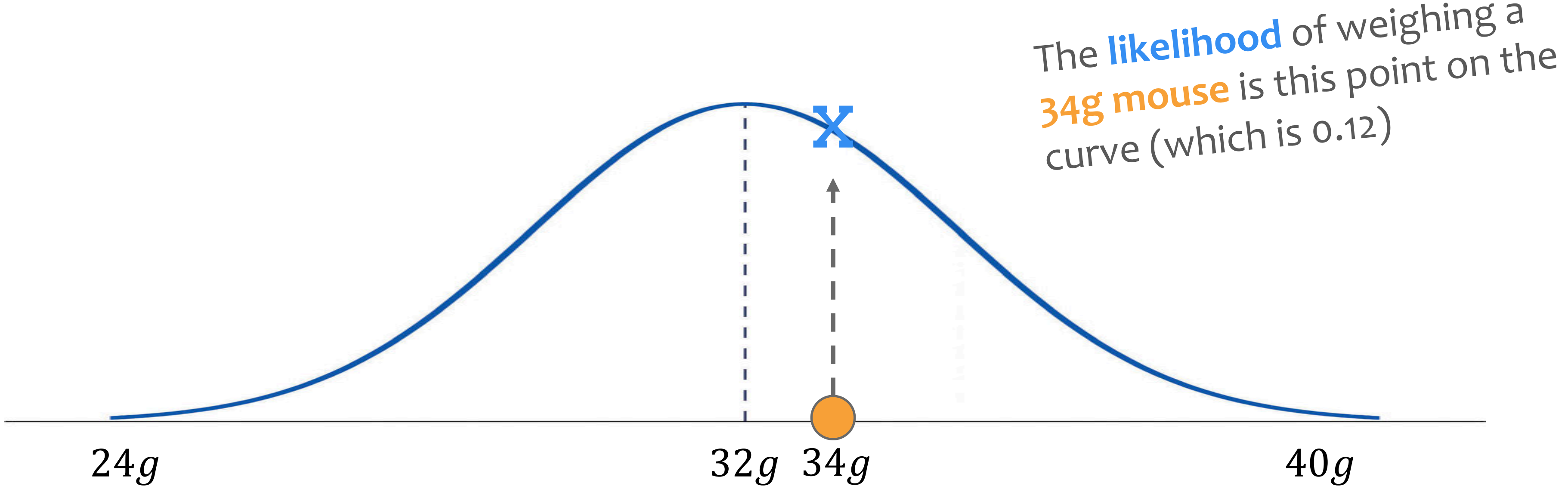
In case of **likelihood**, we assume that we already have selected the mouse (e.g., weighing 34g).



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).



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

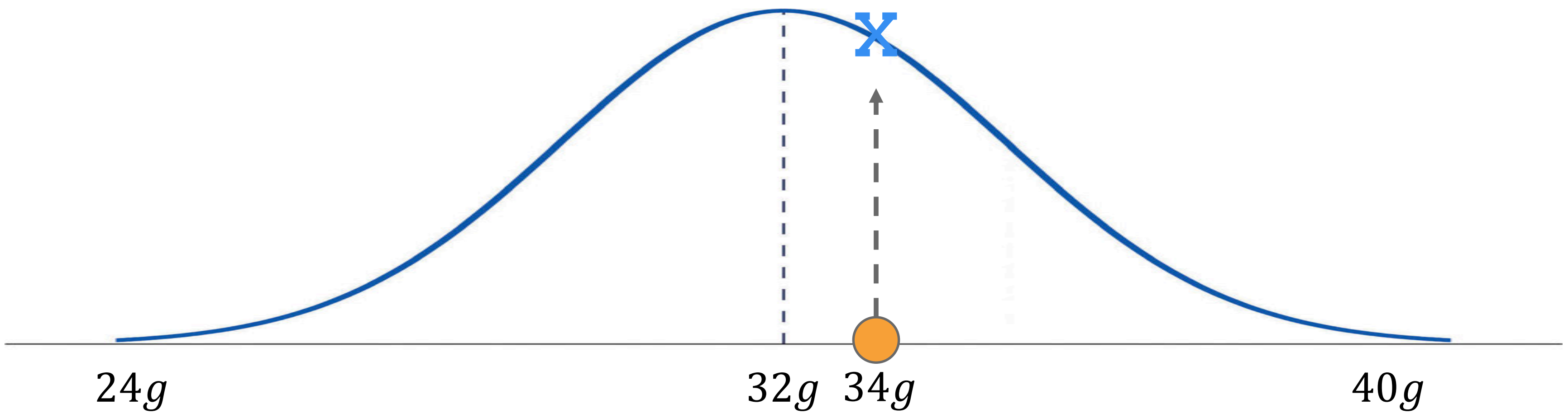
We can think **likelihood** as:
"How **likely** (what is the probability) a ***particular*** **population**, described by a given **distribution (parameters),** is to produce an **observed sample**?"

Mathematically, we use the following notation:

$$L(mean = 32g \; and \; std = 2.5g \mid mouse \; weighs \; 34g\,) = 0.12$$

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

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



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

likelihood function

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

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

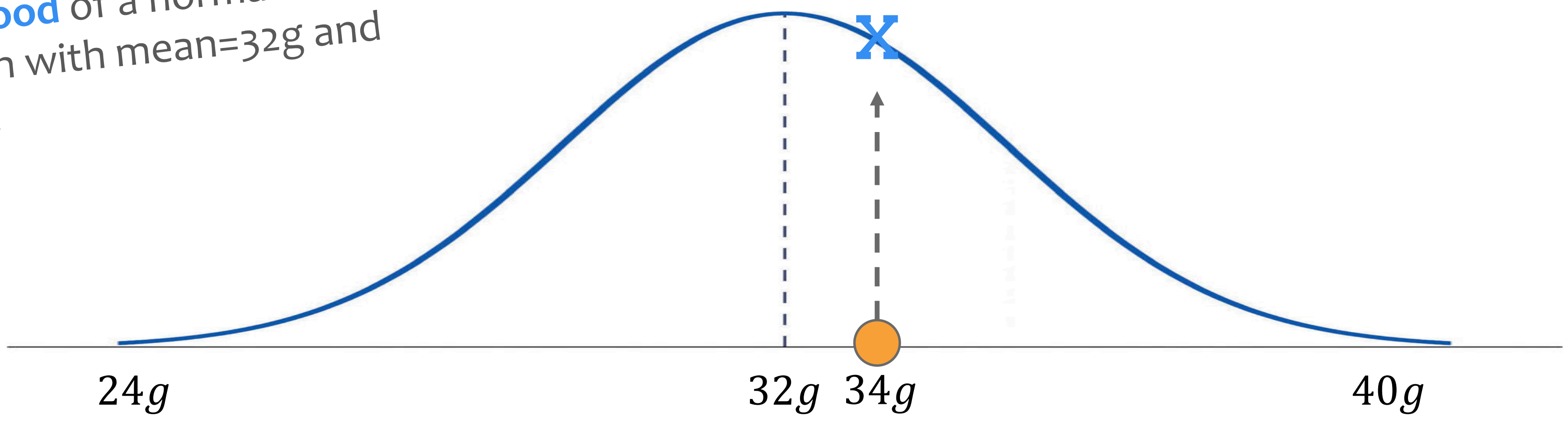
Mathematically, we use the following notation:

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

$$L(\boxed{\mu = 32, \sigma = 2.5} \mid \mathrm{x} = 34) = 0.12$$

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

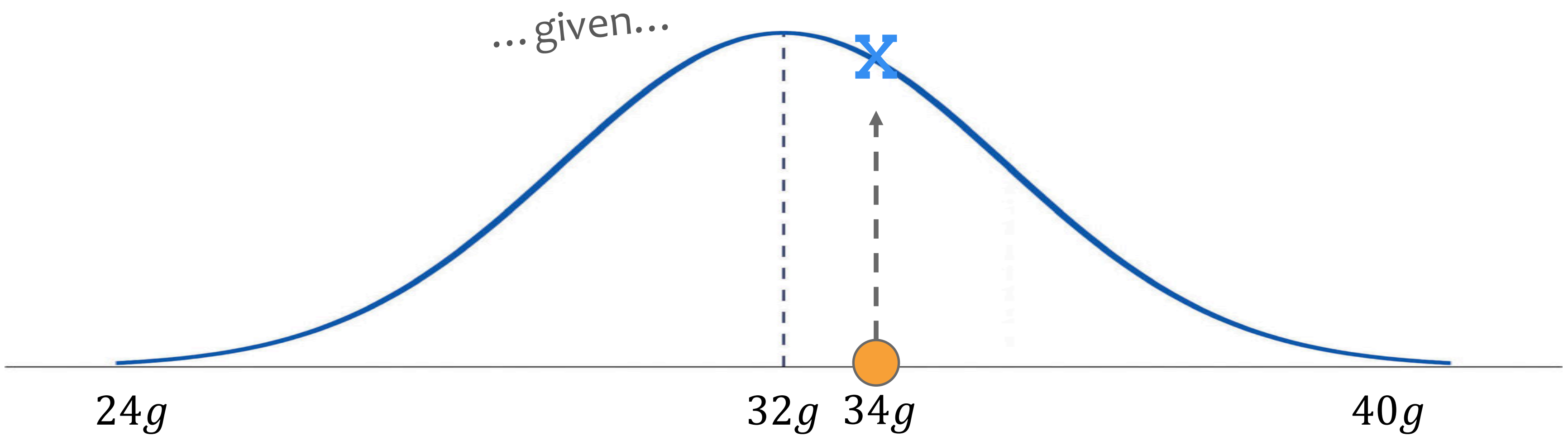
The **likelihood** of a normal distribution with mean=32g and std=2.5g…



Mathematically, we use the following notation:

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

$$L(\mu = 32, \sigma = 2.5 \boxed{|} \mathrm{x} = 34) = 0.12$$



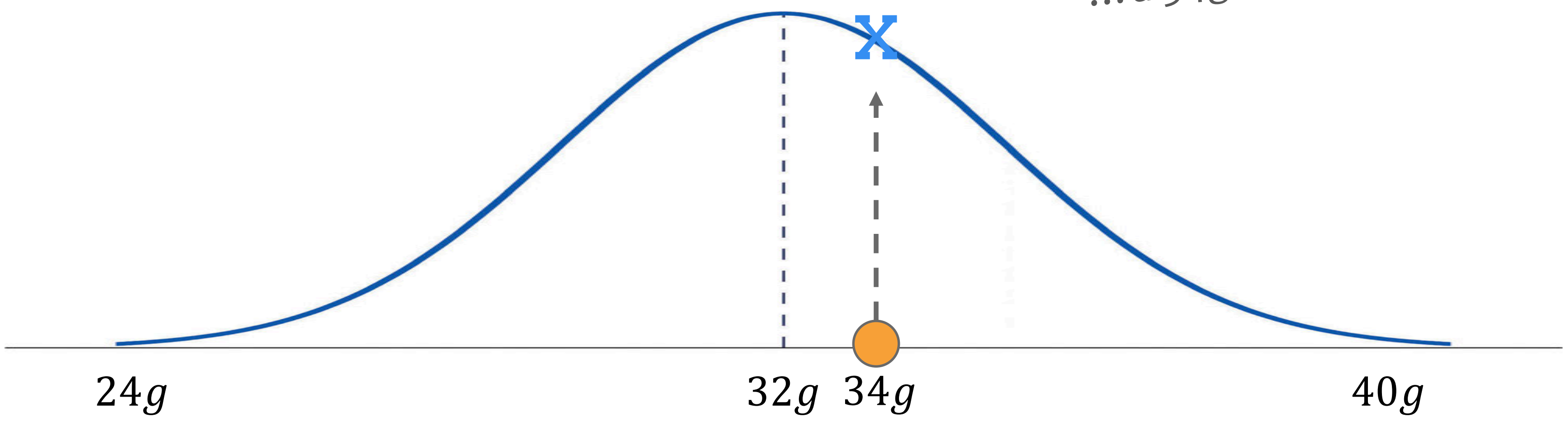
Mathematically, we use the following notation:

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

$$L(\mu = 32, \sigma = 2.5 \mid \boxed{\mathrm{x} = 34)} = 0.12$$

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

…a 34g mouse

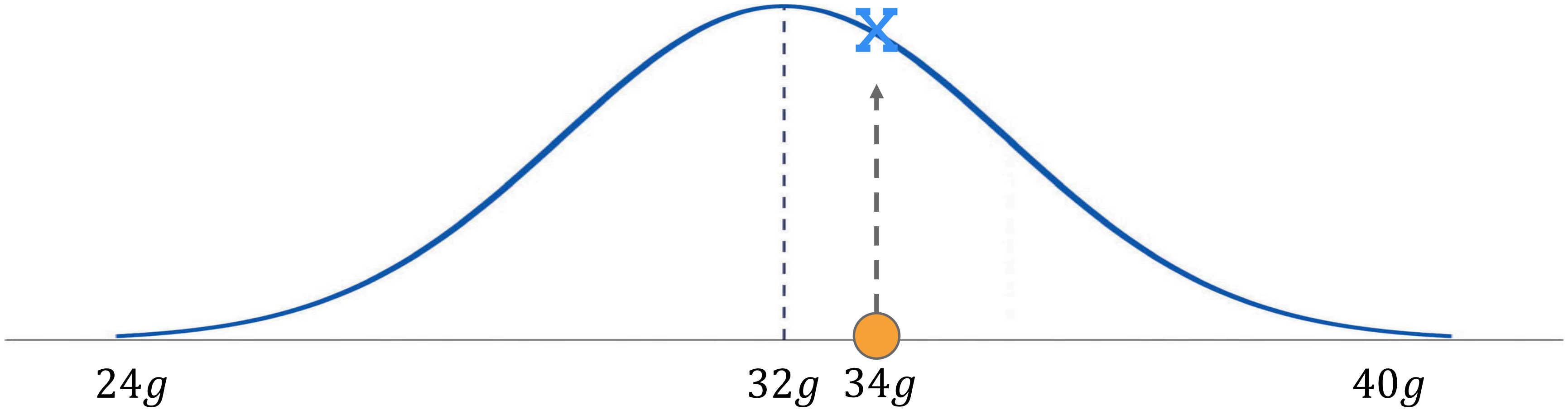


Mathematically, we use the following notation:

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

$$L(\mu = 32, \sigma = 2.5 \mid \mathrm{x} = 34) \boxed{= 0.12}$$

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

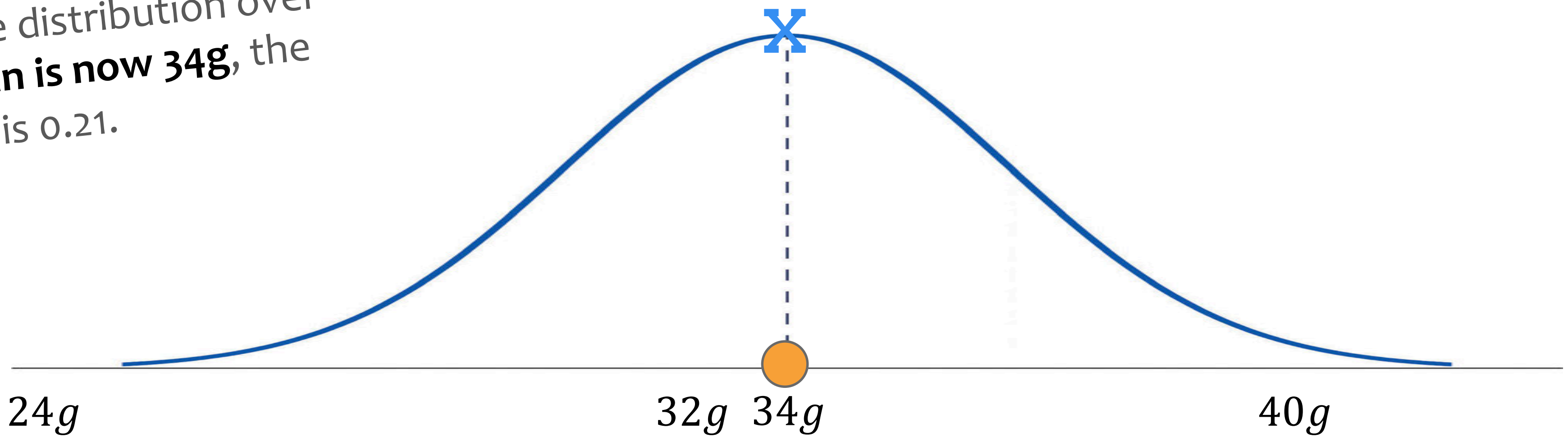


$$L(\boxed{mean = 34g\ and\ std = 2.5g}\ |\ mouse\ weighs\ 34g\ ) = 0.21$$

$$L(\boxed{\mu = 34, \sigma = 2.5}\ |\ \mathrm{x} = 34) = 0.21$$

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

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

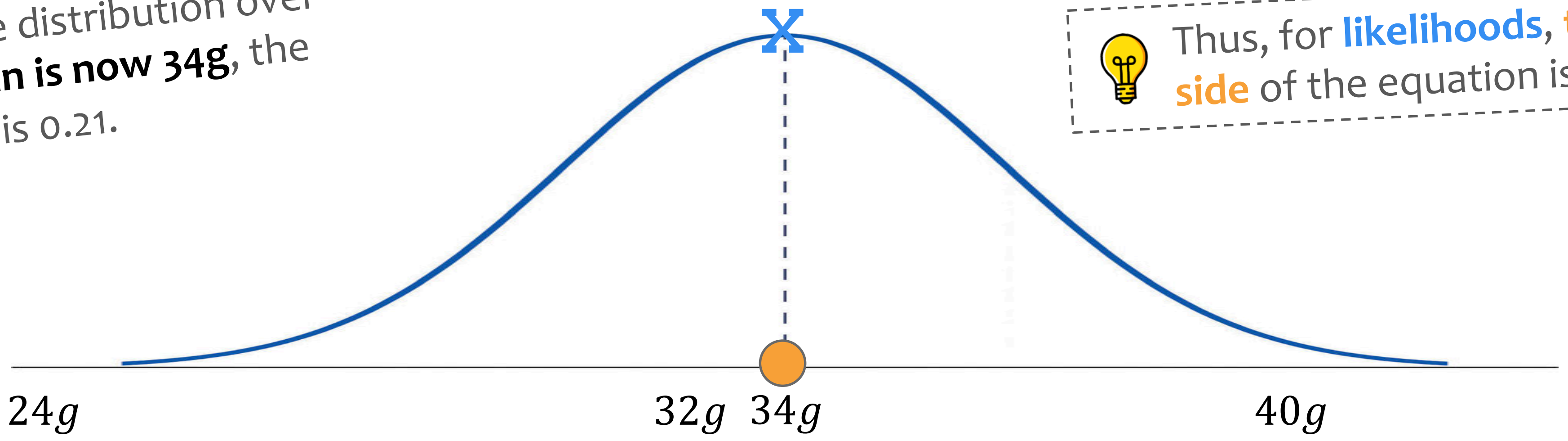


$$L(mean = 34g \text{ and } std = 2.5g \mid \boxed{mouse\ weighs\ 34g}) = 0.21$$

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

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

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



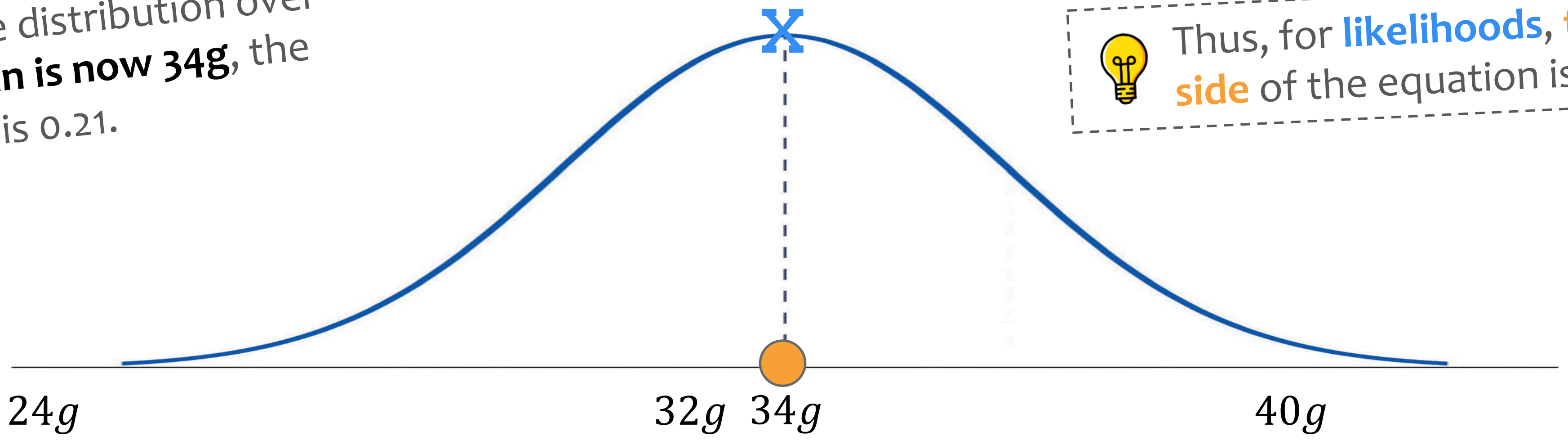
Thus, for **likelihoods**, **the right side** of the equation is **fixed**!

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

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

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

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**."

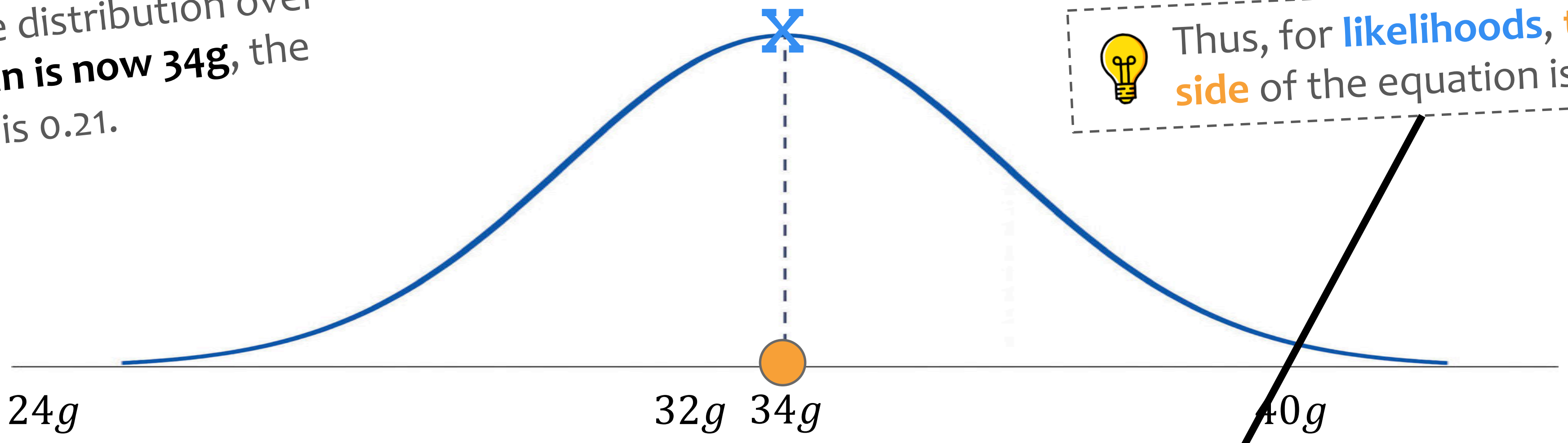
$$L(mean = 34g \; and \; std = 2.5g \mid \boxed{mouse \; weighs \; 34g}) = 0.21$$

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

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

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…

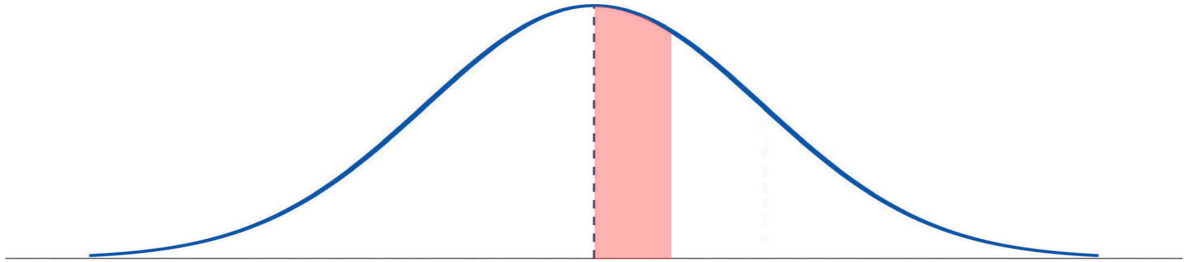
We can think **likelihood** as:
"How **likely** a particular **population** is to produce an **observe sample**."

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

Summary

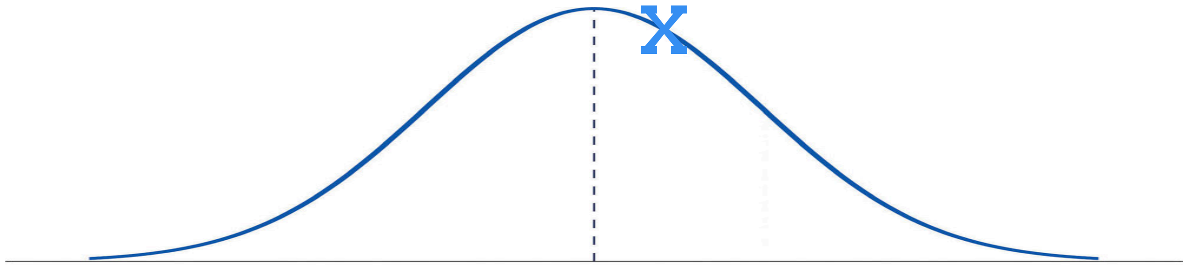
Probabilities are the **areas under** a *fixed distribution.*

$$P(data \mid distribution)$$

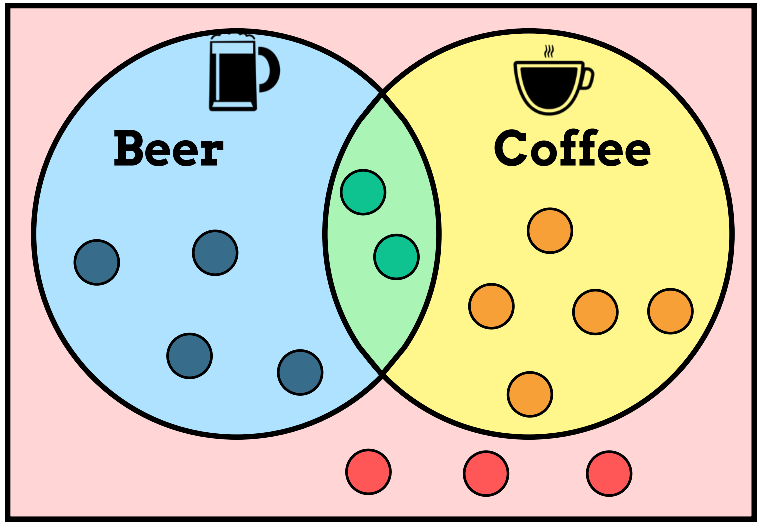


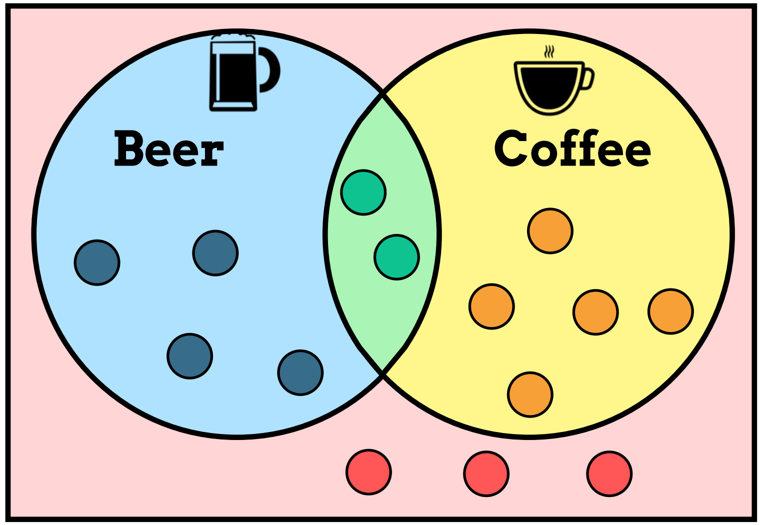
Likelihoods are the **y-axis values** for *fixed data points* with distributions that can be moved.

$$L(distribution \mid data)$$



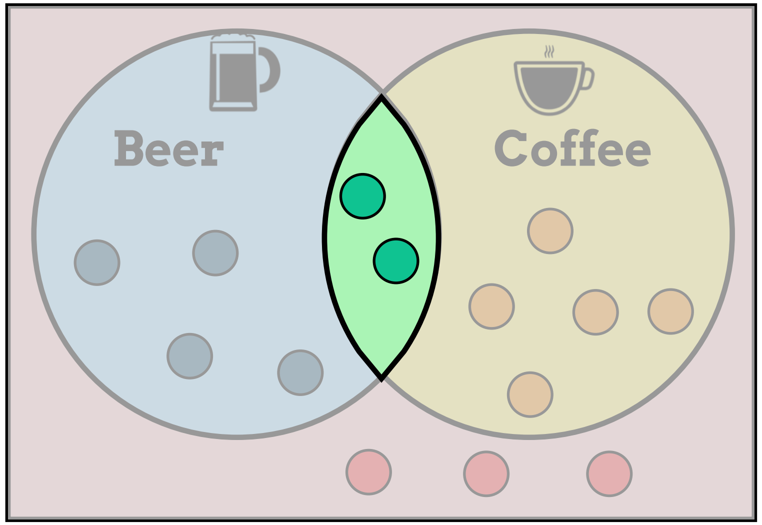
Bayes' Theorem


Beer
Coffee



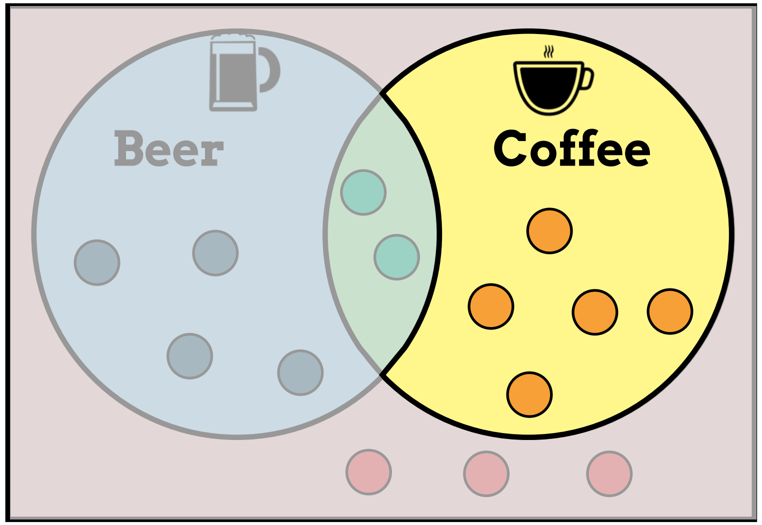
Contingency Table

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



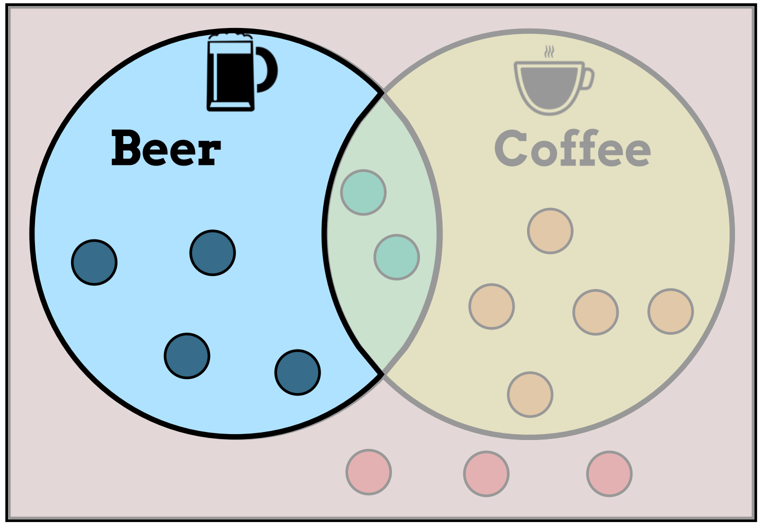
Contingency Table

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



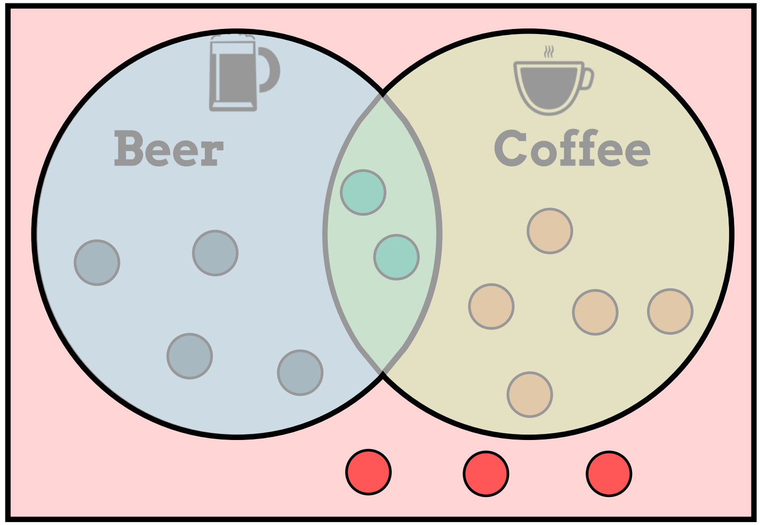
Contingency Table

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



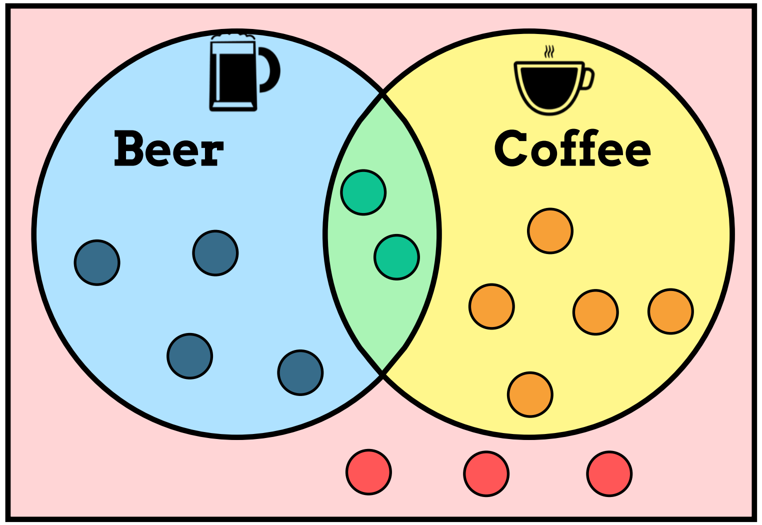
Contingency Table

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



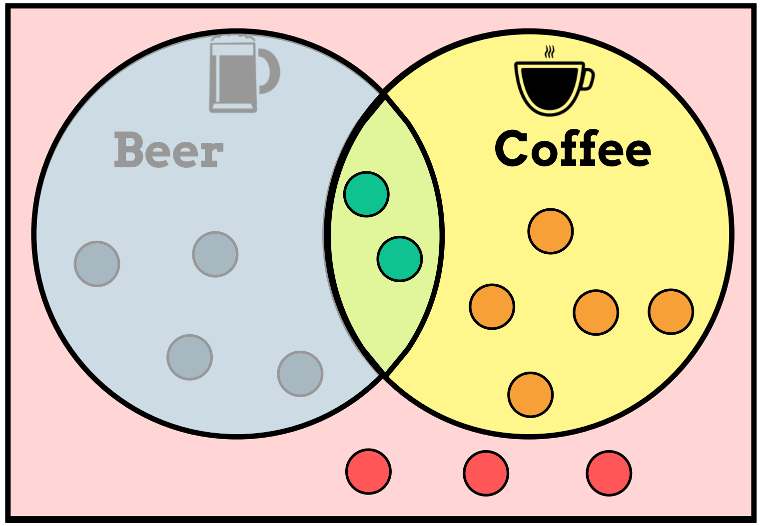
Contingency Table

	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



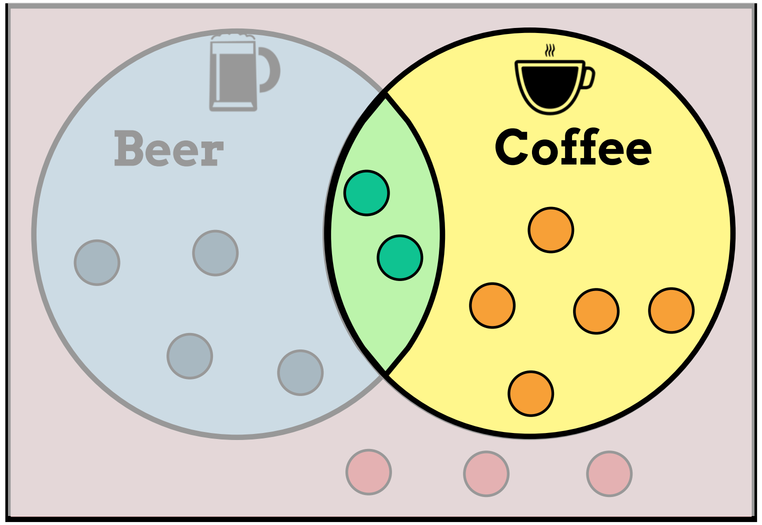
Contingency Table

	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	



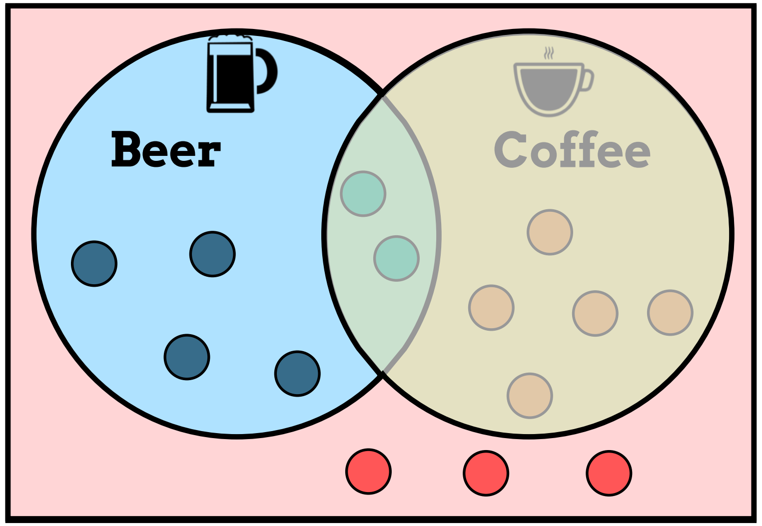
Contingency Table

	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	



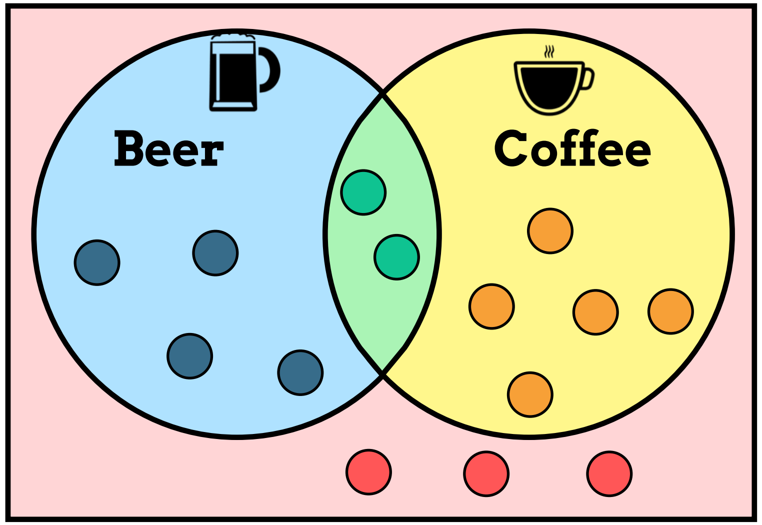
Contingency Table

	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	



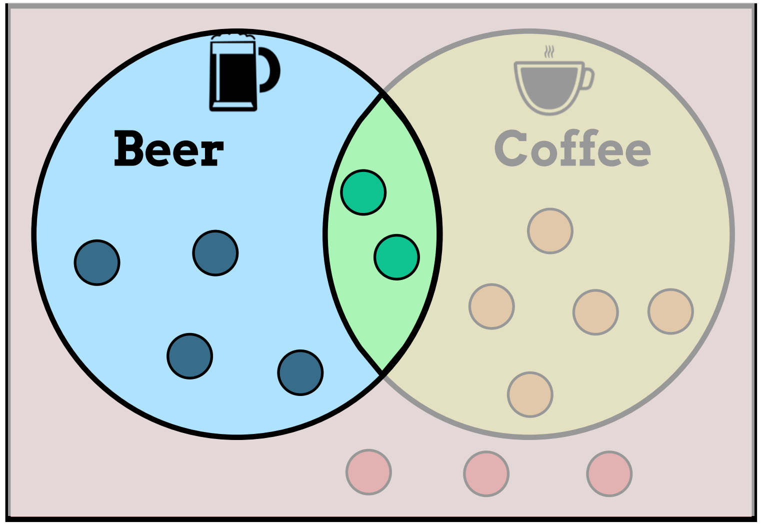
Contingency Table

	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



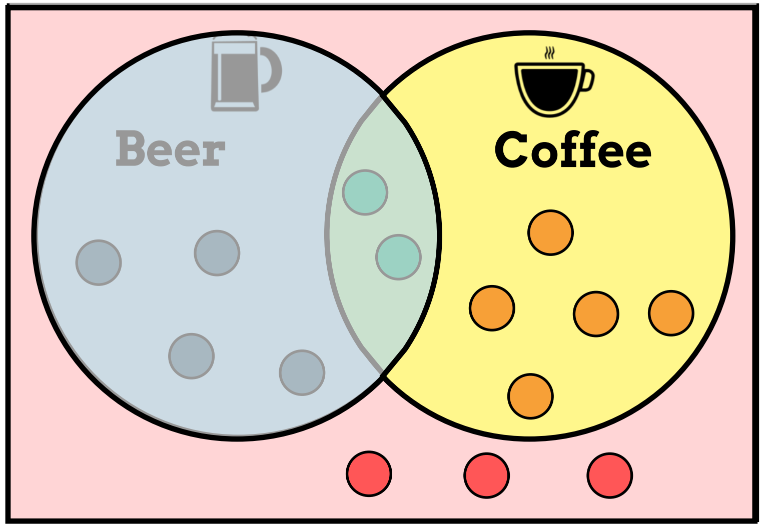
Contingency Table

	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			



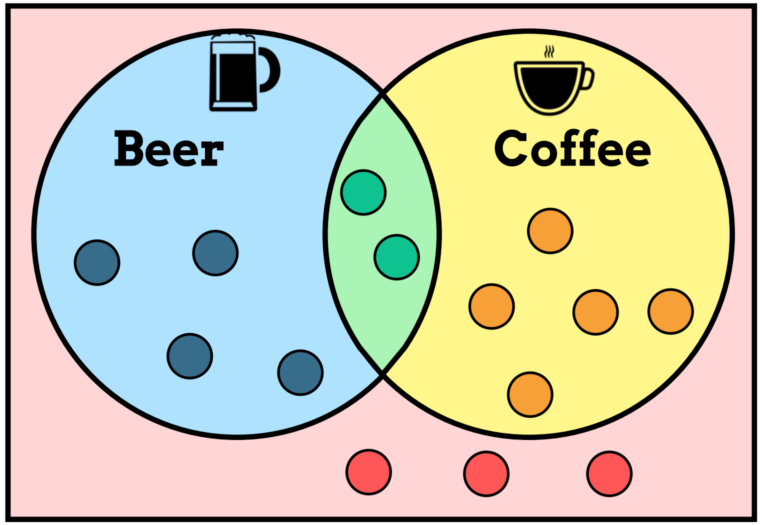
Contingency Table

	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		



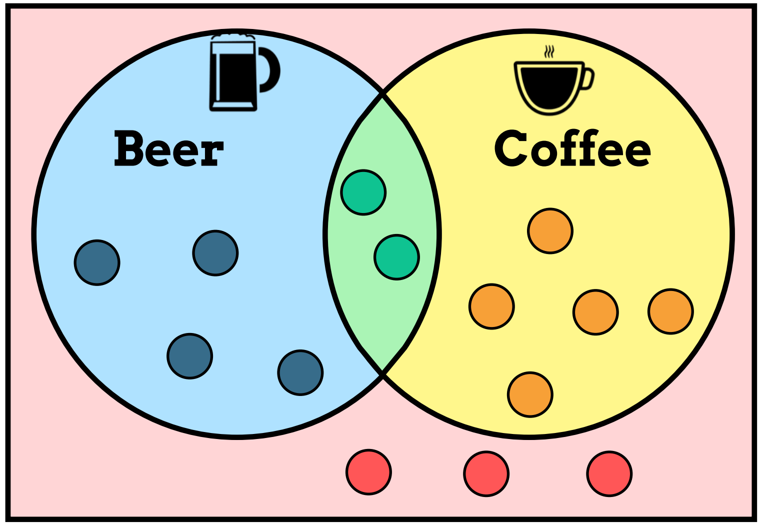
Contingency Table

	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

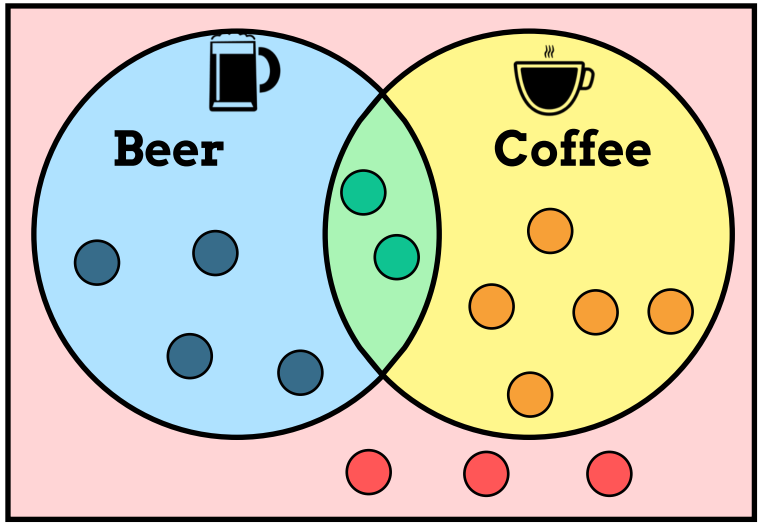
	Likes Beer	Doesn’t Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn’t Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	



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?**

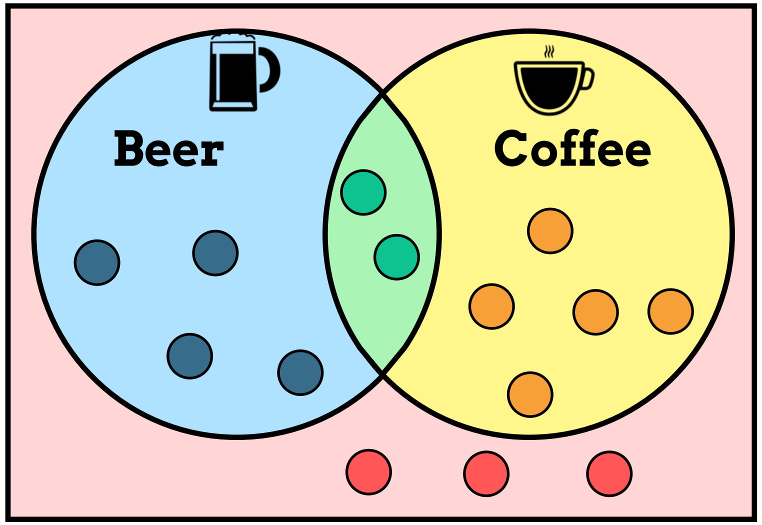


Contingency Table

	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?**

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



Contingency Table

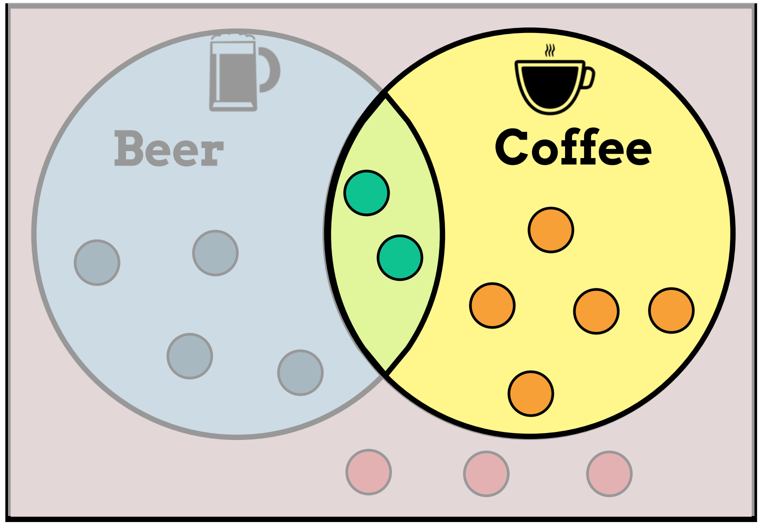
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

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

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

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

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

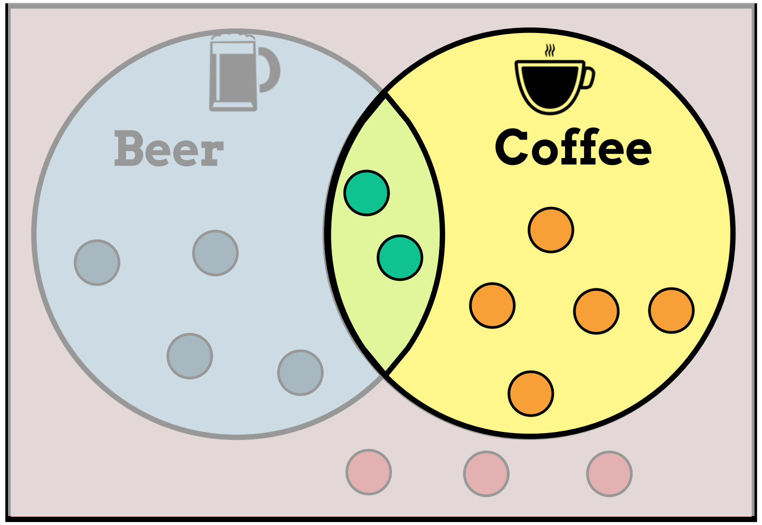


Contingency Table

	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?**

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

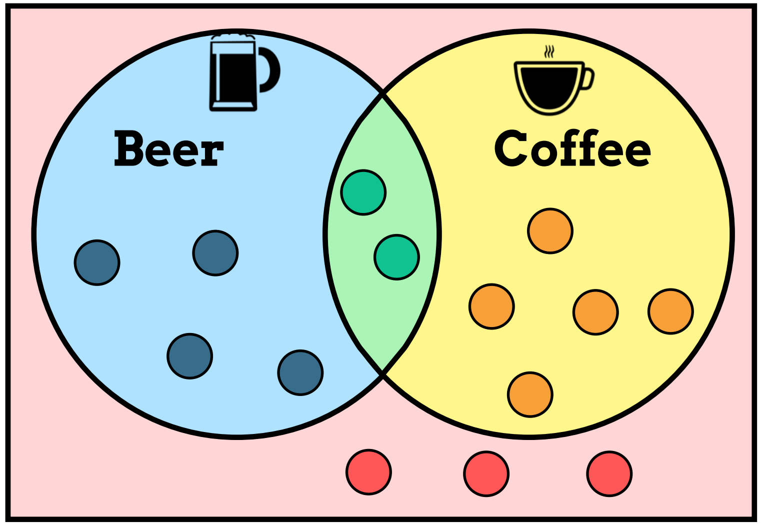


Contingency Table

	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?**

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

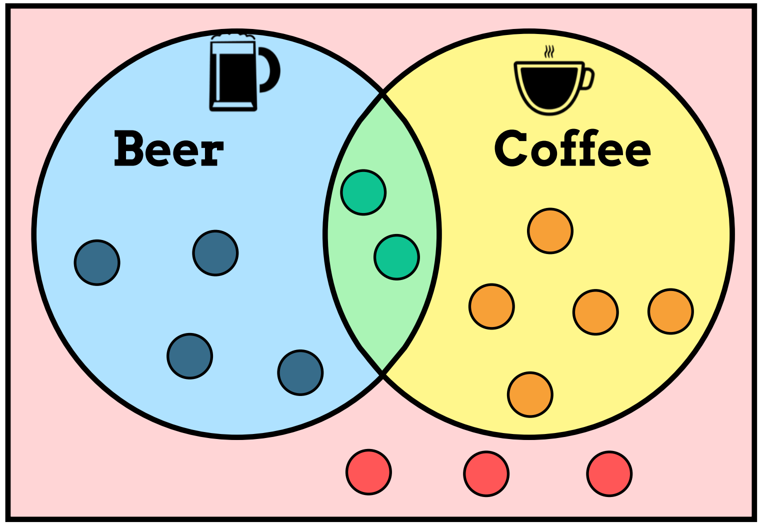


Contingency Table

	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?**

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

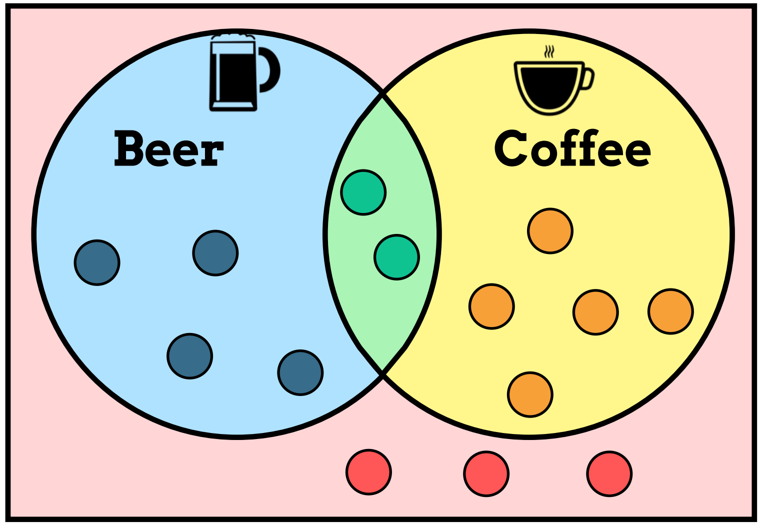


Contingency Table

	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?**

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

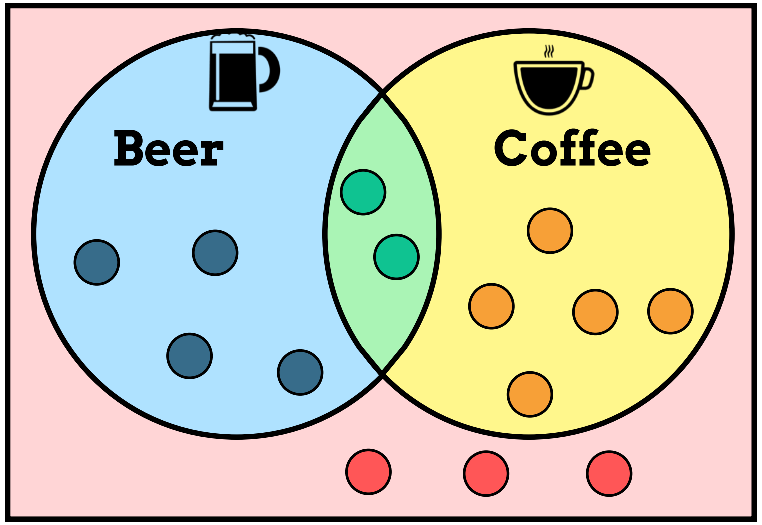


Contingency Table

	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?**

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

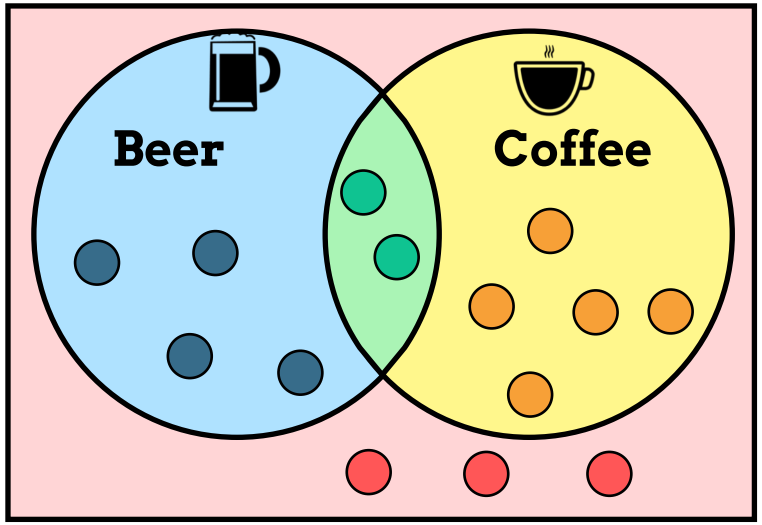


Contingency Table

	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?**

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

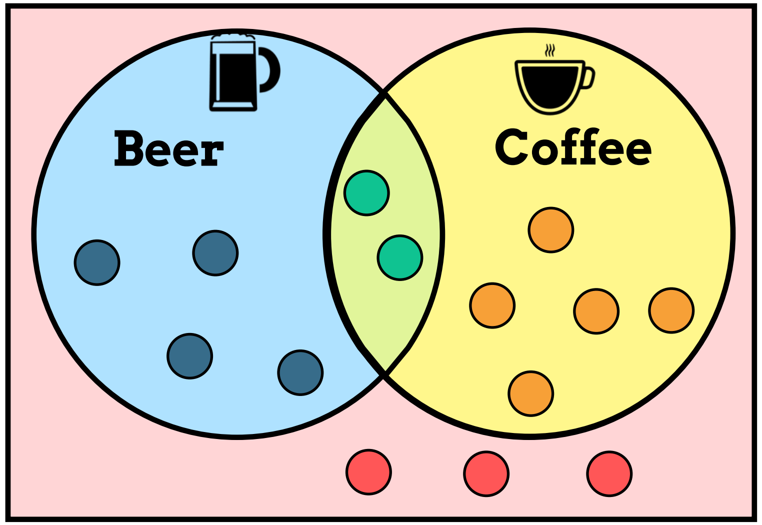


Contingency Table

	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?**

$$\mathrm{P(no\ like}\ \boldsymbol{b}\ \mathrm{AND\ like}\ \boldsymbol{c}\ |\ \mathrm{like}\ \boldsymbol{c}) = \frac{\frac{5}{14}}{\frac{7}{14}} = \boxed{\mathbf{0.71}}$$



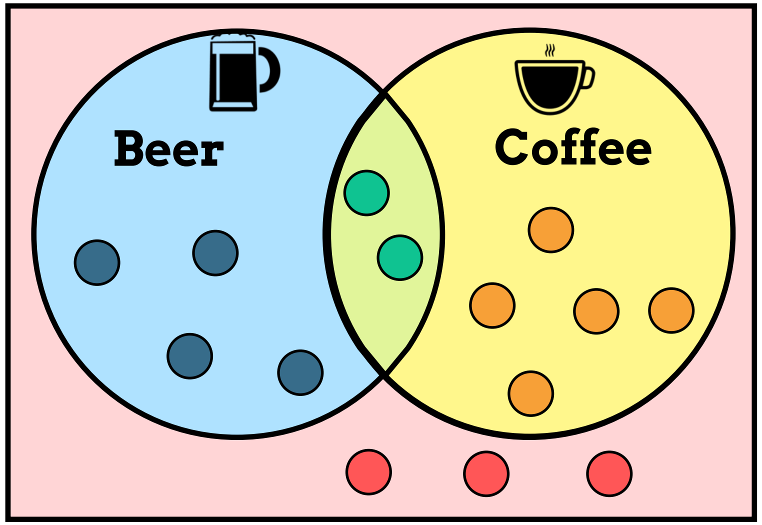
Contingency Table

	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?**

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

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



Contingency Table

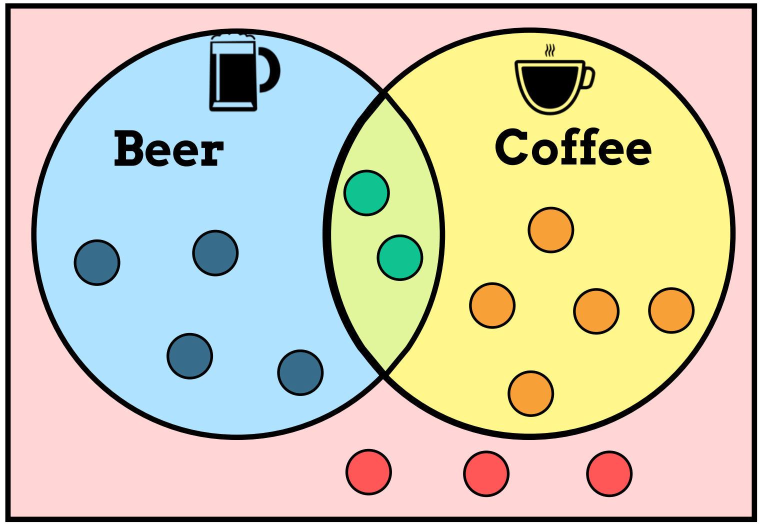
	Likes Beer	Doesn't Like Beer	Row Total
Likes Coffee	2 P=2/14	5 P=5/14	2+5=7 P=7/14
Doesn't Like Coffee	4 P=4/14	3 P=3/14	4+3=7 P=7/14
Column Total	2+4=6 P=6/14	5+3=8 P=8/14	

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

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

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	

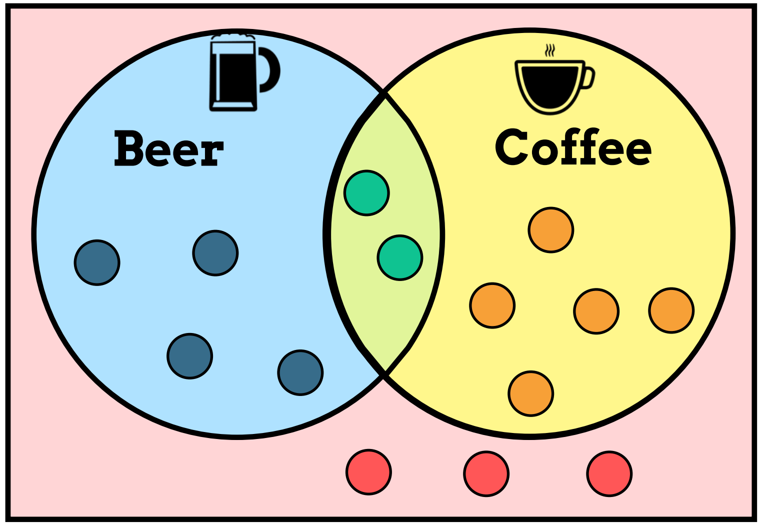
What is the (conditional) probability that someone does **not like beer** given that (s)he **likes coffee?**

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

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**)...

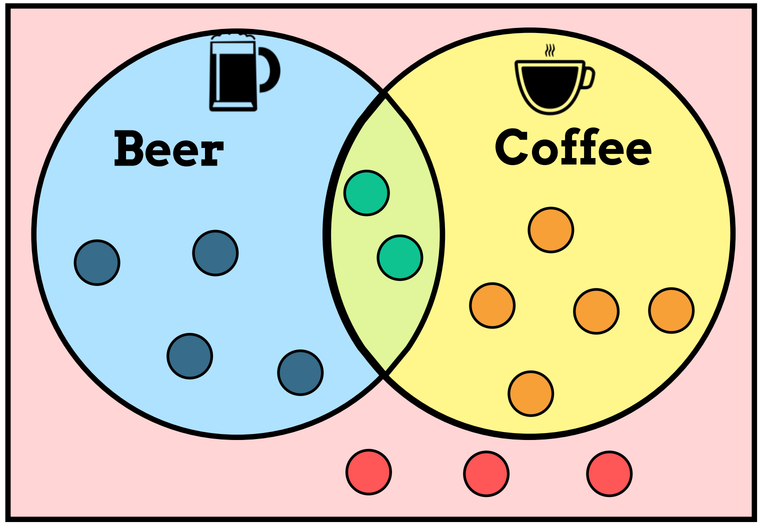


Contingency Table

	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	

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

Remember this part is ***redundant***:
"Someone who *does not like beer* and *likes coffee*, given that (s)he *likes coffee*".



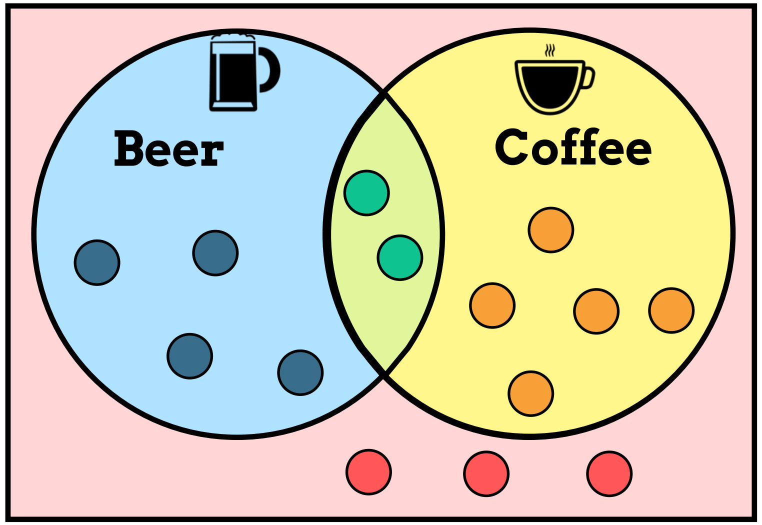
Contingency Table

	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	

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

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.

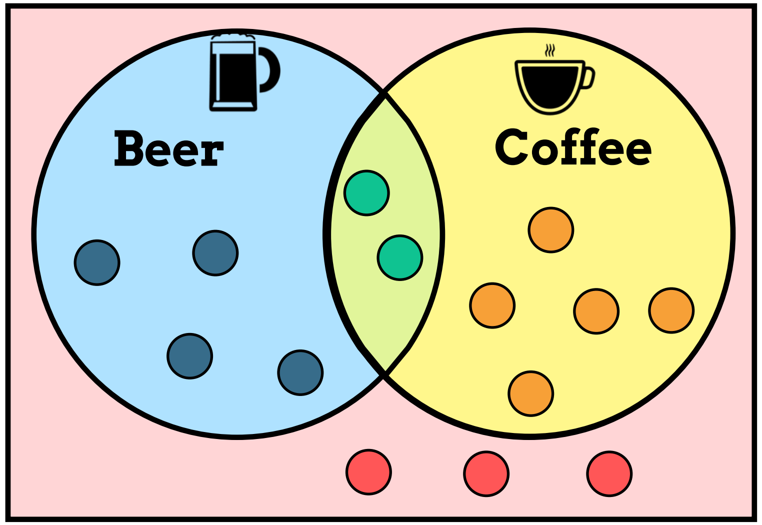


Contingency Table

	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	

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

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



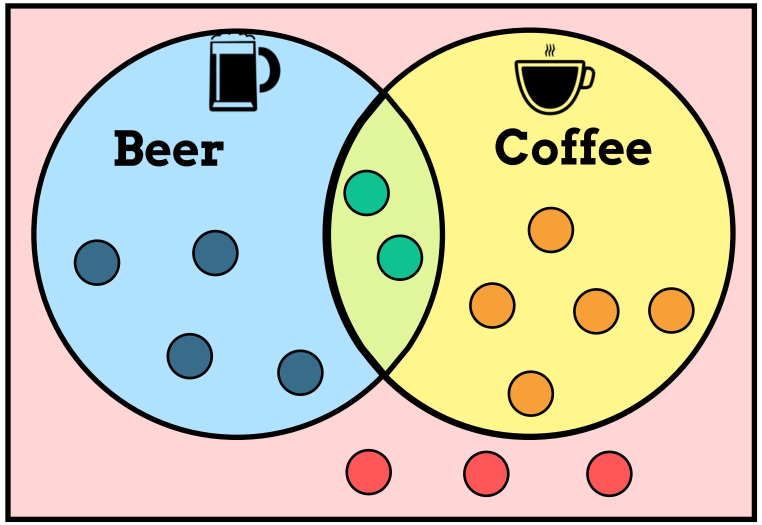
Contingency Table

	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	

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

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.

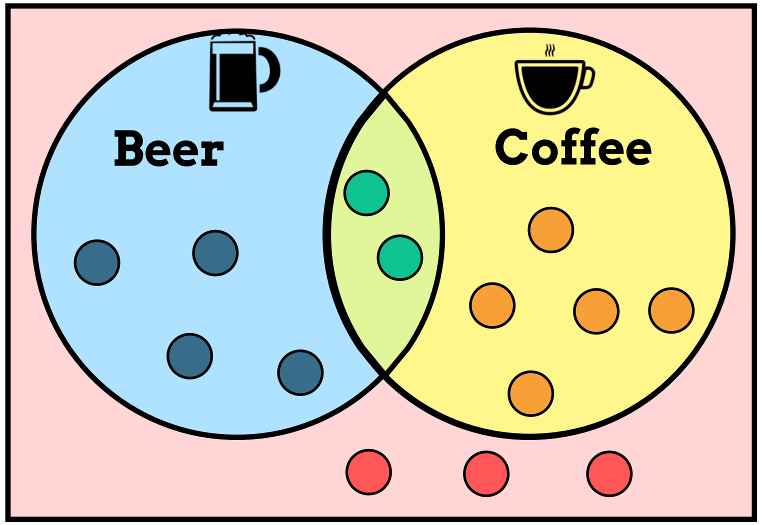


Contingency Table

	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.

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

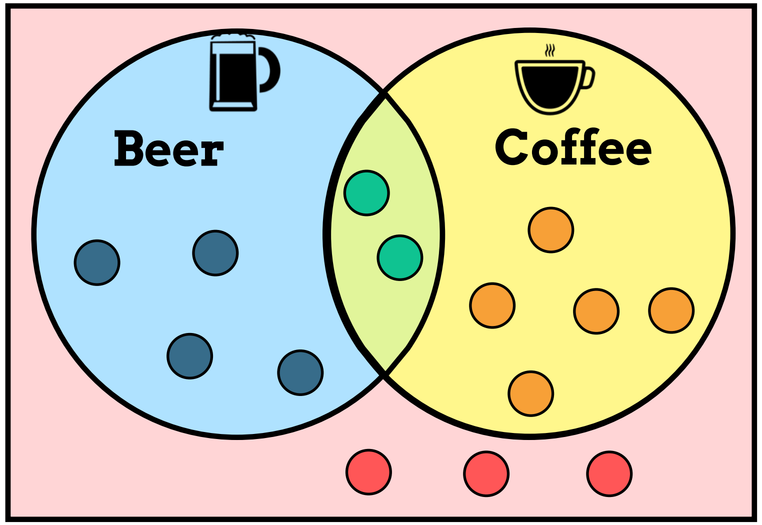


Contingency Table

	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.

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

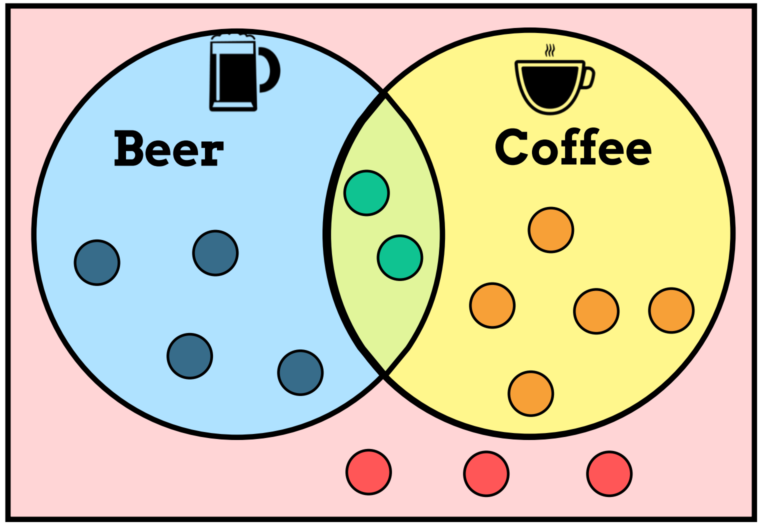


Contingency Table

	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)…

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



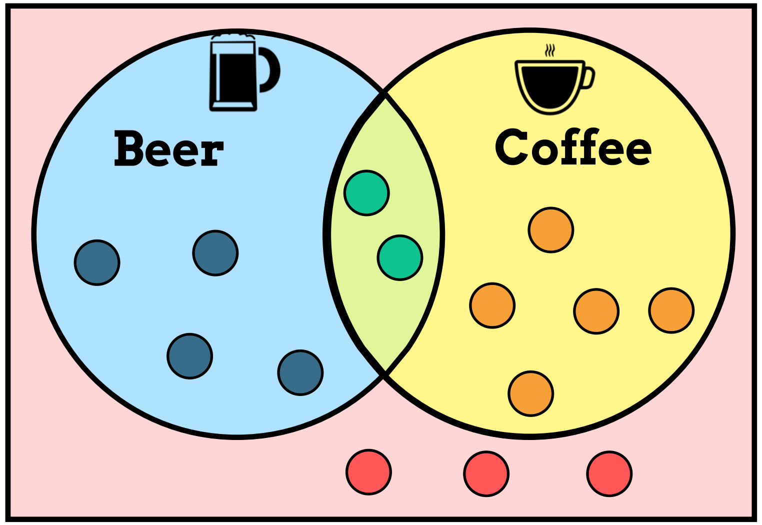
Contingency Table

	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)…

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

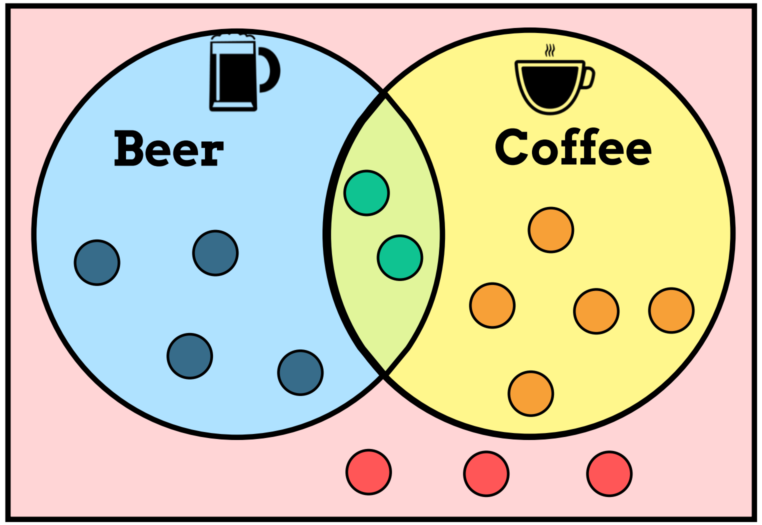


Contingency Table

	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…

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



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**…

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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…

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

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

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

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

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**…

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

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

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

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

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

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

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

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

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

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

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

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

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

Or

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

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

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

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

We have just derived the formula **Bayes' Theorem.**

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

Or

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

B = likes coffee

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

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

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

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

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

A = does not like beer
B = likes coffee

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

$$\downarrow$$

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

A = does not like beer
B = likes coffee

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

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

Removing redundancy

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

Bayes' Theorem

Bayes' Theorem

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

Prior Probability
The prob. of A being True.

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

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

Marginal Likelihood
The prob. of B being True.

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**.

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**?

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

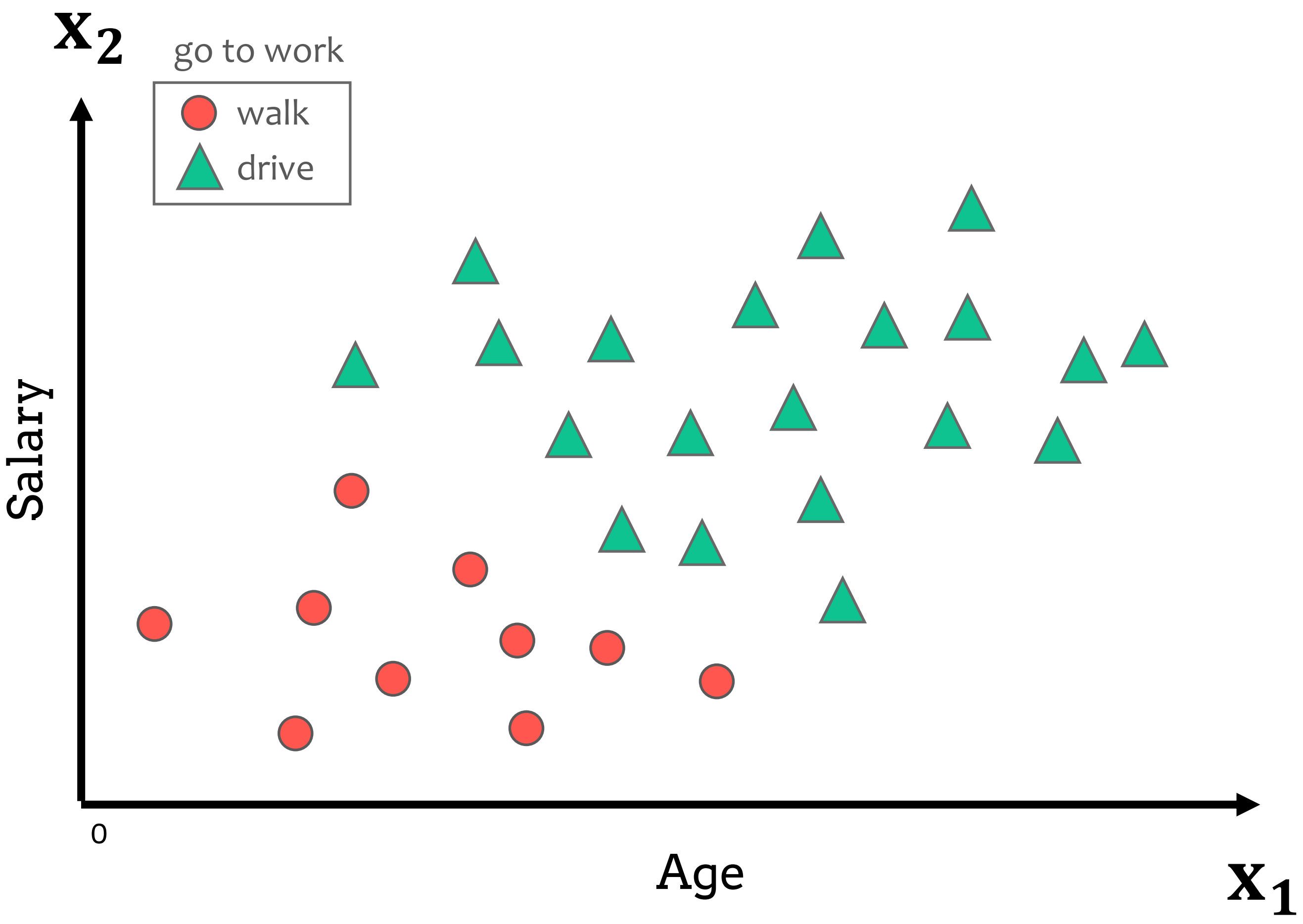
Posterior (*a posteriori prob.*)

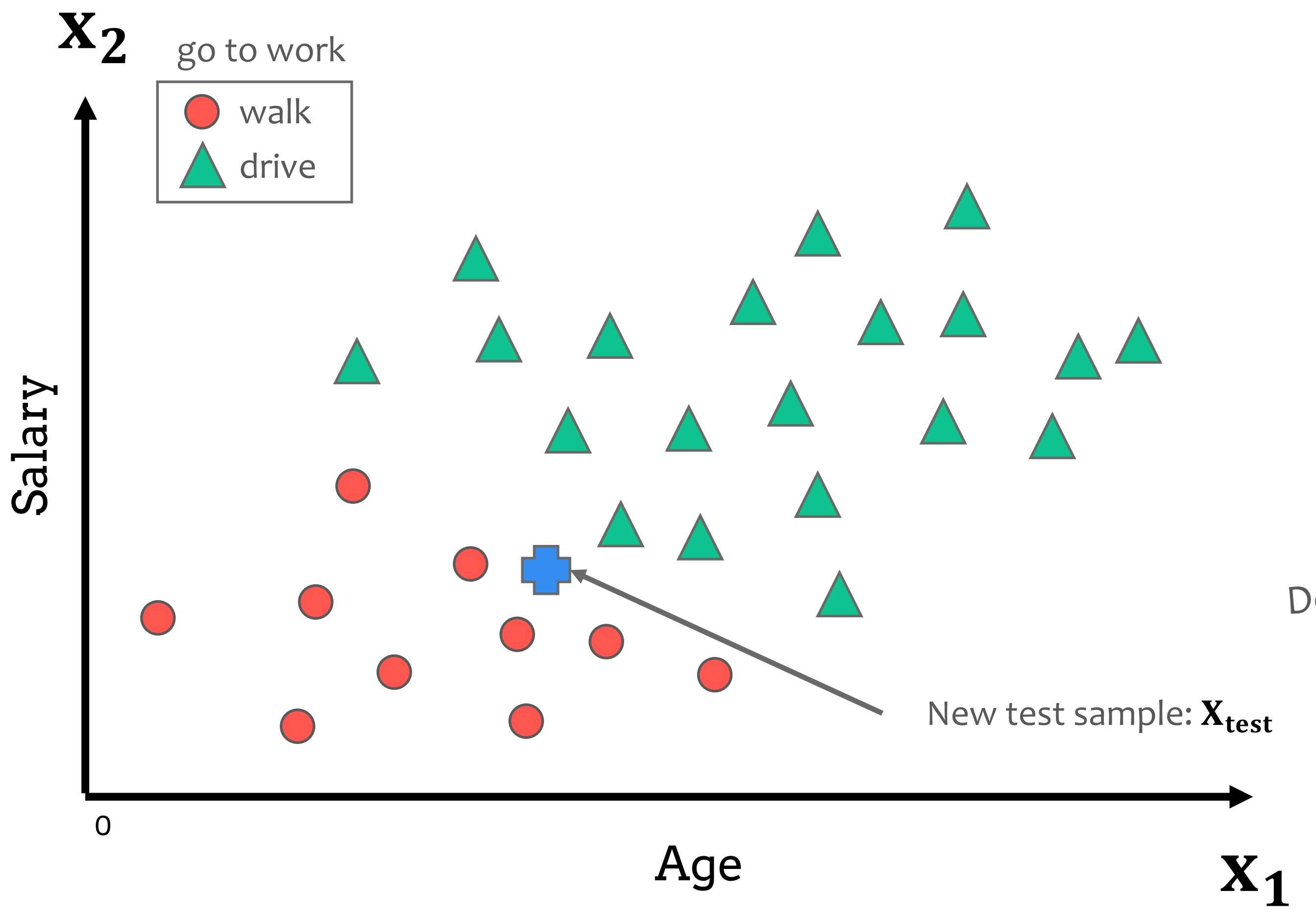
The prob. of the **hypothesis** is true given the observed **evidence**.

Marginal Likelihood

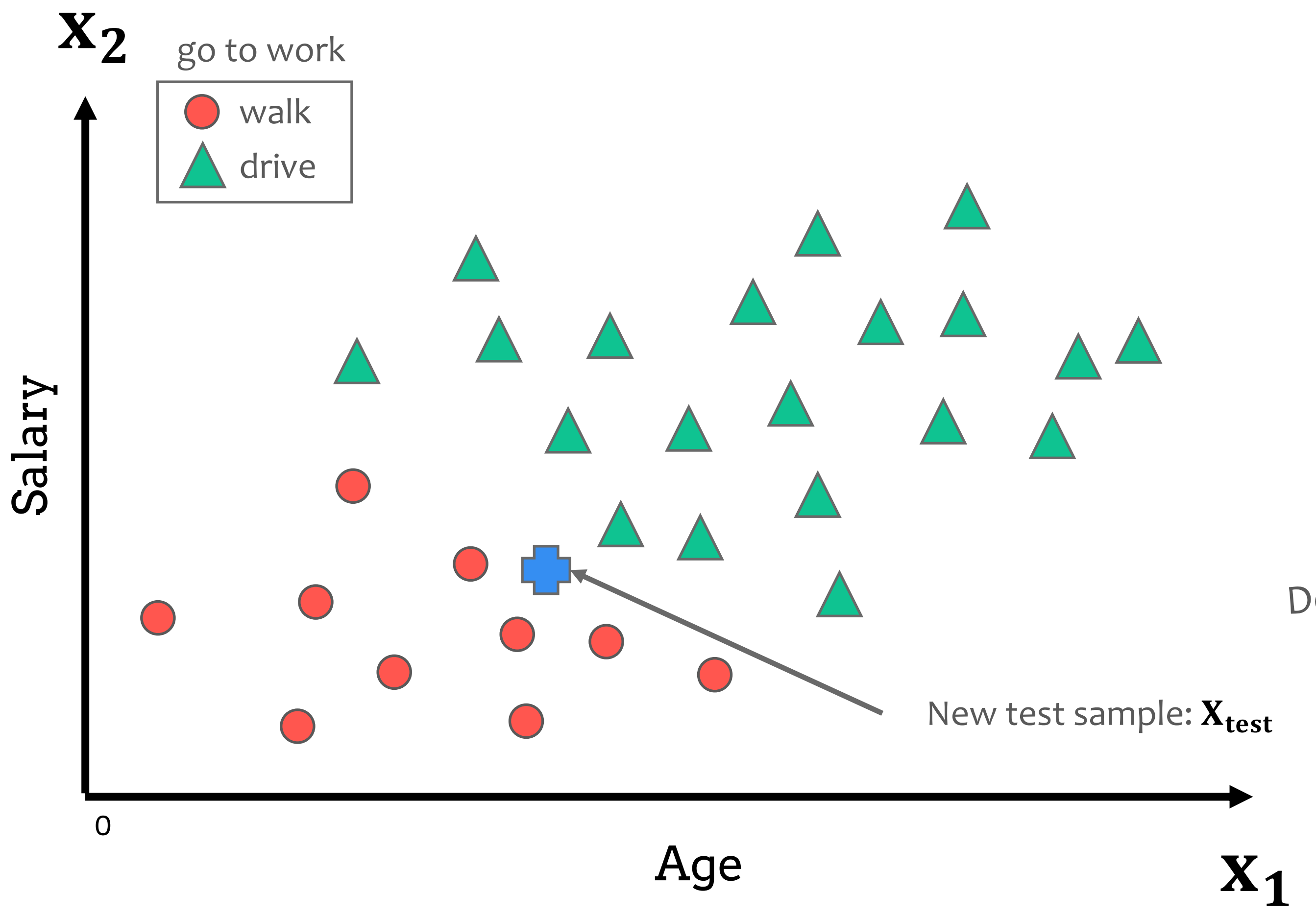
- The prob. of **B** being true.
- The prob. of observing the **evidence**.
- How probable is the new **evidence** under ***all*** possible **hypothesis**?

Naive Bayes Classifier


$\mathbf{x_2}$
go to work
walk
drive
Salary
0
Age
$\mathbf{x_1}$



Does this person go to work by walking or driving?



Let's use **Naive Bayes' Theorem**!

Does this person go to work by **walking** or **driving**?

Naive Bayes Classifier

Bayes' Theorem

Likelihood

Prior probability

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

Posterior probability

Marginal Likelihood

Naive Bayes Classifier

Bayes' Theorem

Likelihood

Prior probability

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

Posterior probability

Marginal Likelihood

Naive Bayes Classifier

Step 1

#3 Likelihood

Prior probability #1

class

feats of the new sample

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

#4 Posterior probability

Marginal Likelihood #2

Naive Bayes Classifier

Step 2

#3 Likelihood

Prior probability #1

class

feats of the new sample

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

#4 Posterior probability

Marginal Likelihood #2

Naive Bayes Classifier

Step 3

$$P(Walks|\mathrm{x}_{test}) \ \textbf{vs} \ P(Drives|\mathrm{x}_{test})$$

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

Naive Bayes Classifier

Step 3

Let's do this

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

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

Naive Bayes

#3
Likelihood

Prior probability #1

class　　feats of the new sample

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

#4 **Posterior probability**

Marginal Likelihood
#2

Naive Bayes

#3
Likelihood

Prior probability #1

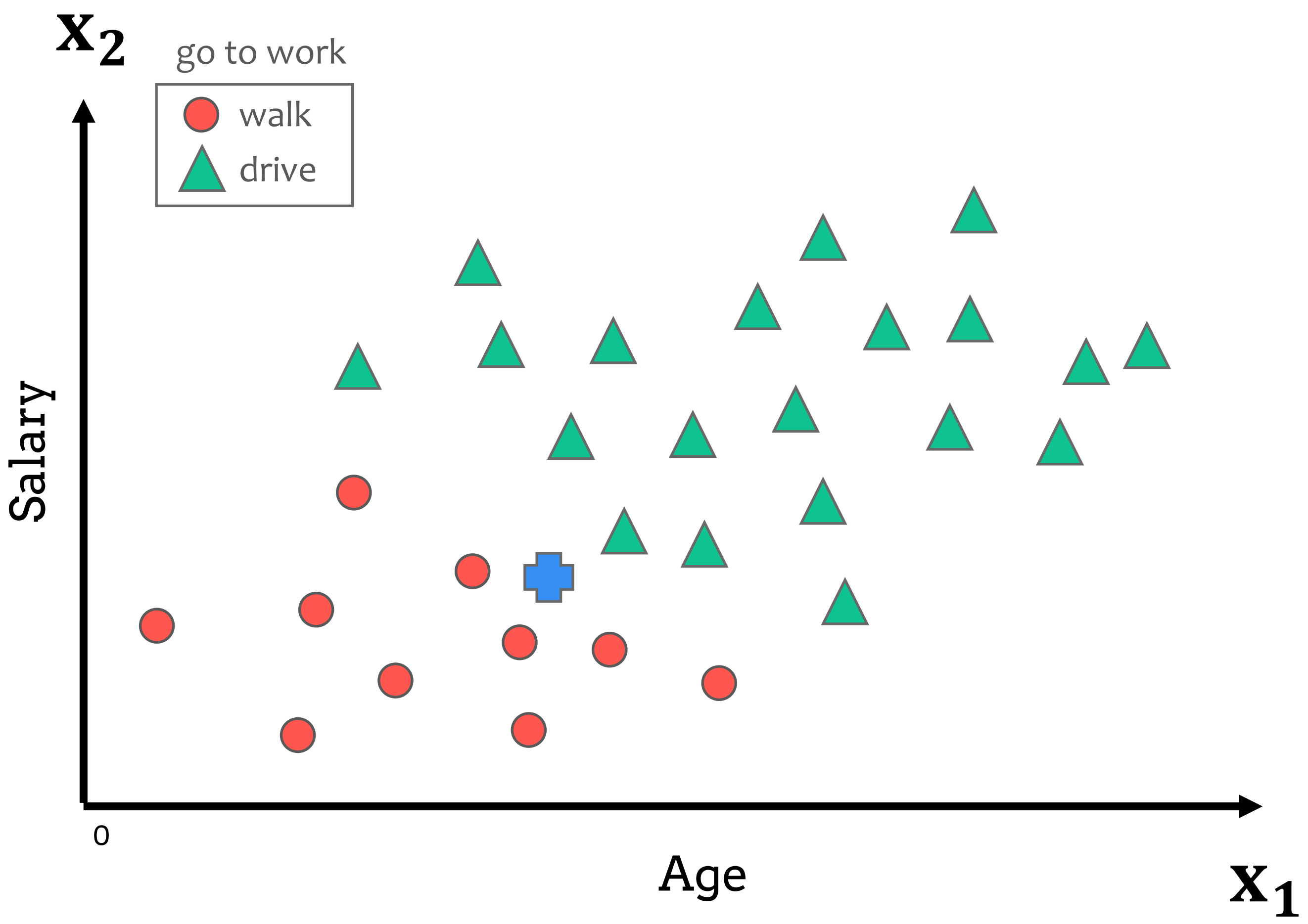
class
feats of the new sample

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

#4 **Posterior probability**

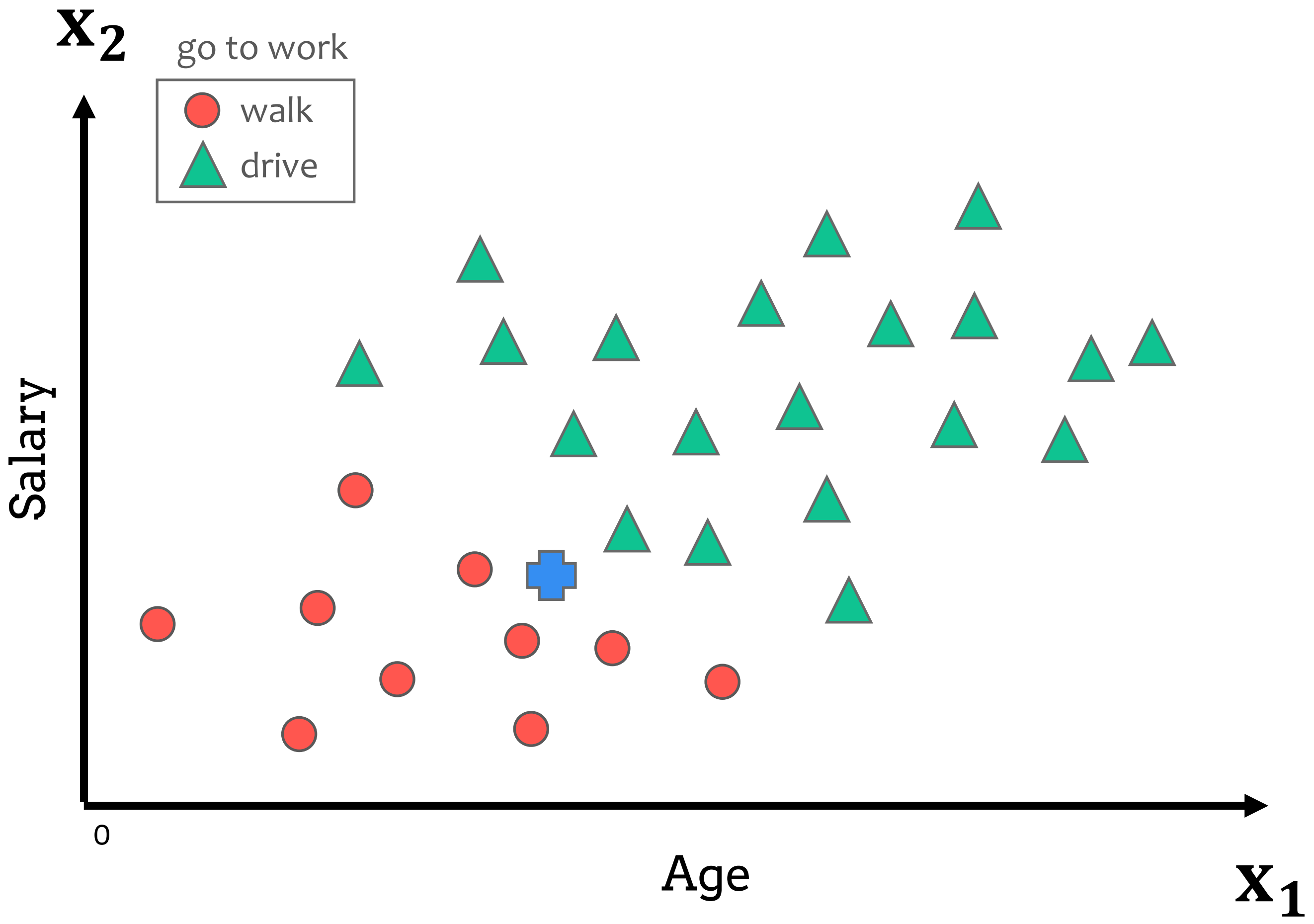
Marginal Likelihood
#2

Naive Bayes: Step 1



#1 - $P(Walks)$

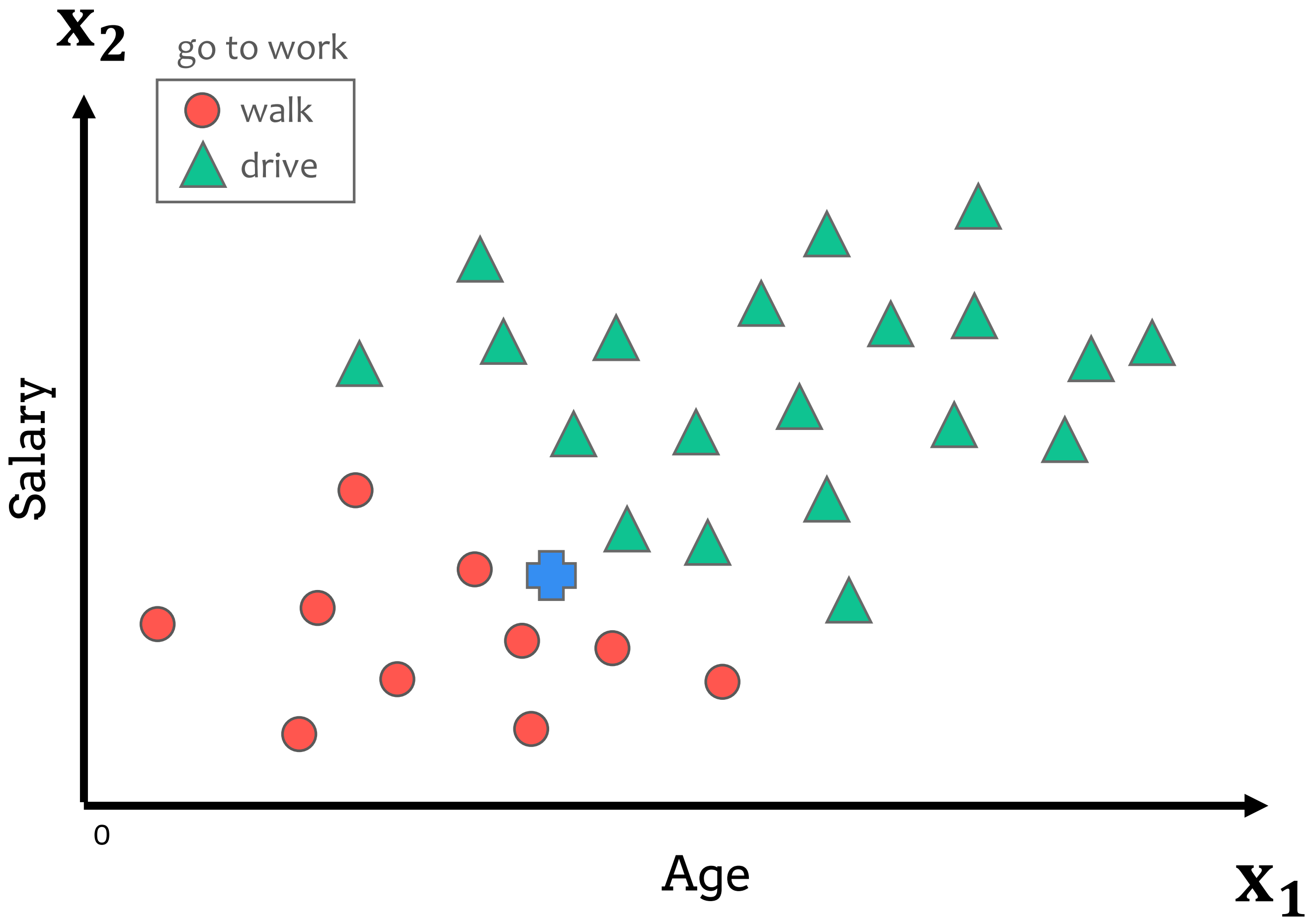
Naive Bayes: Step 1



#1 - $P(Walks)$

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

Naive Bayes: Step 1

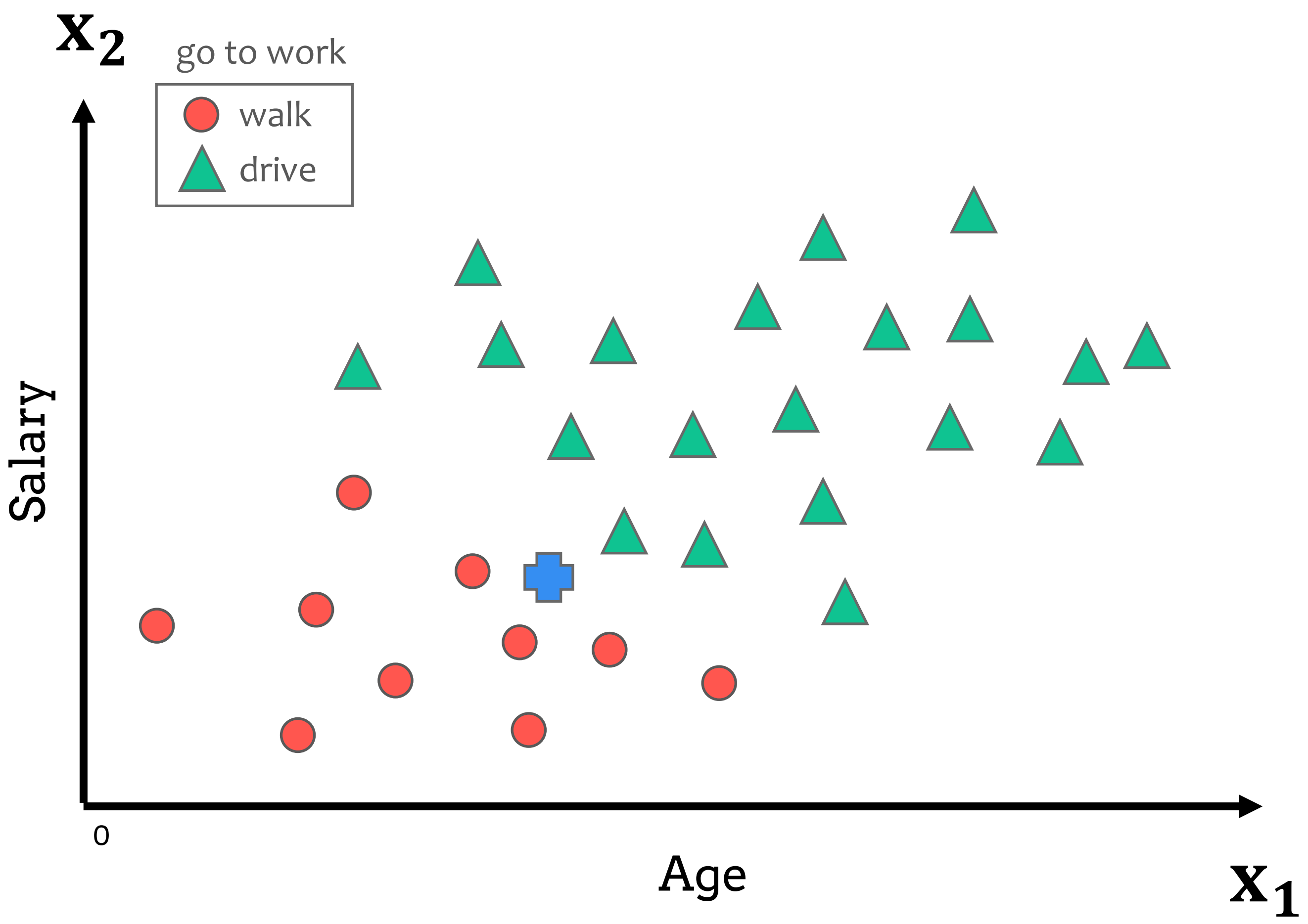


#1 - $P(Walks)$

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

This can be any **probability**.

Naive Bayes: Step 1



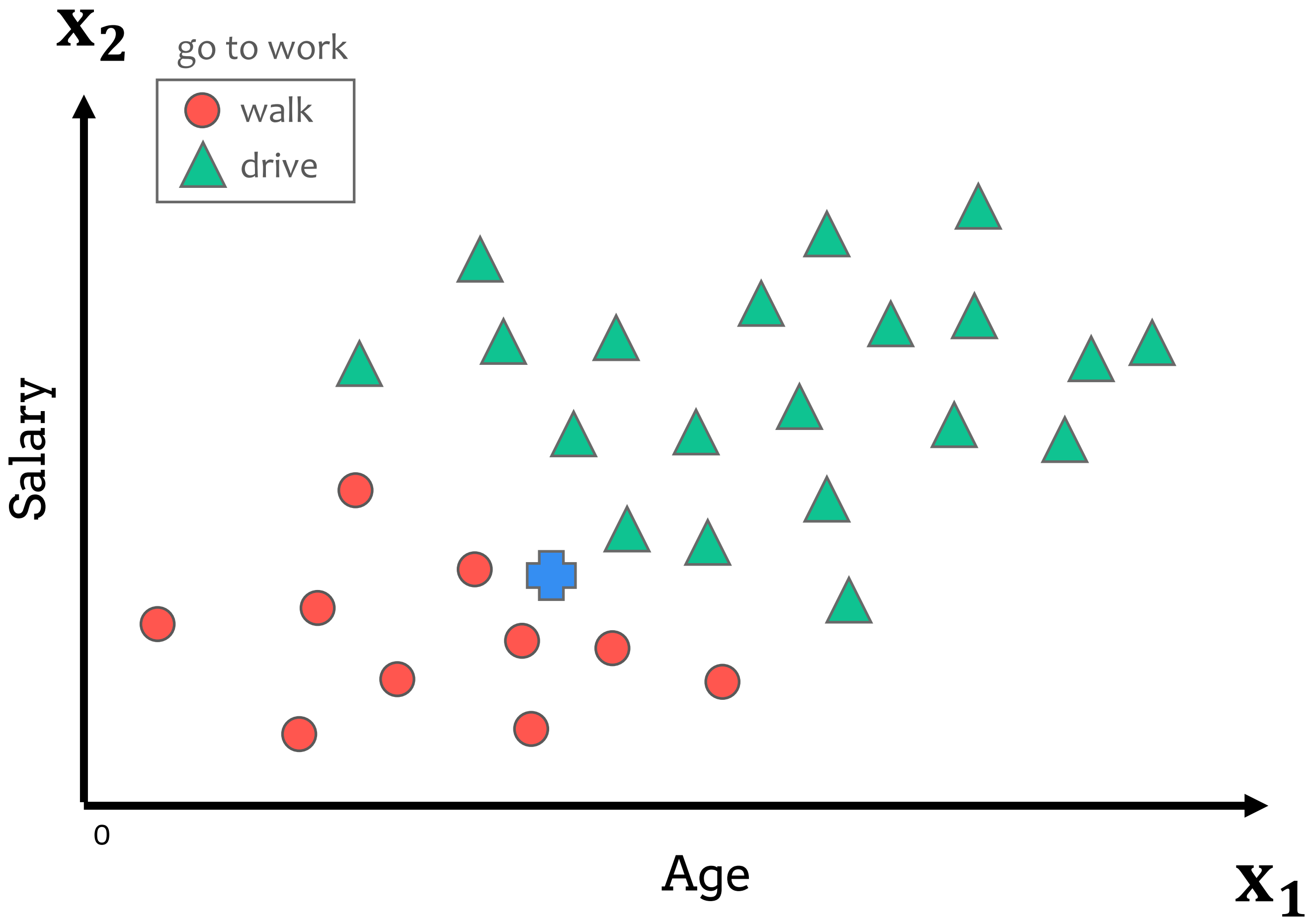
#1 - $P(Walks)$

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.**

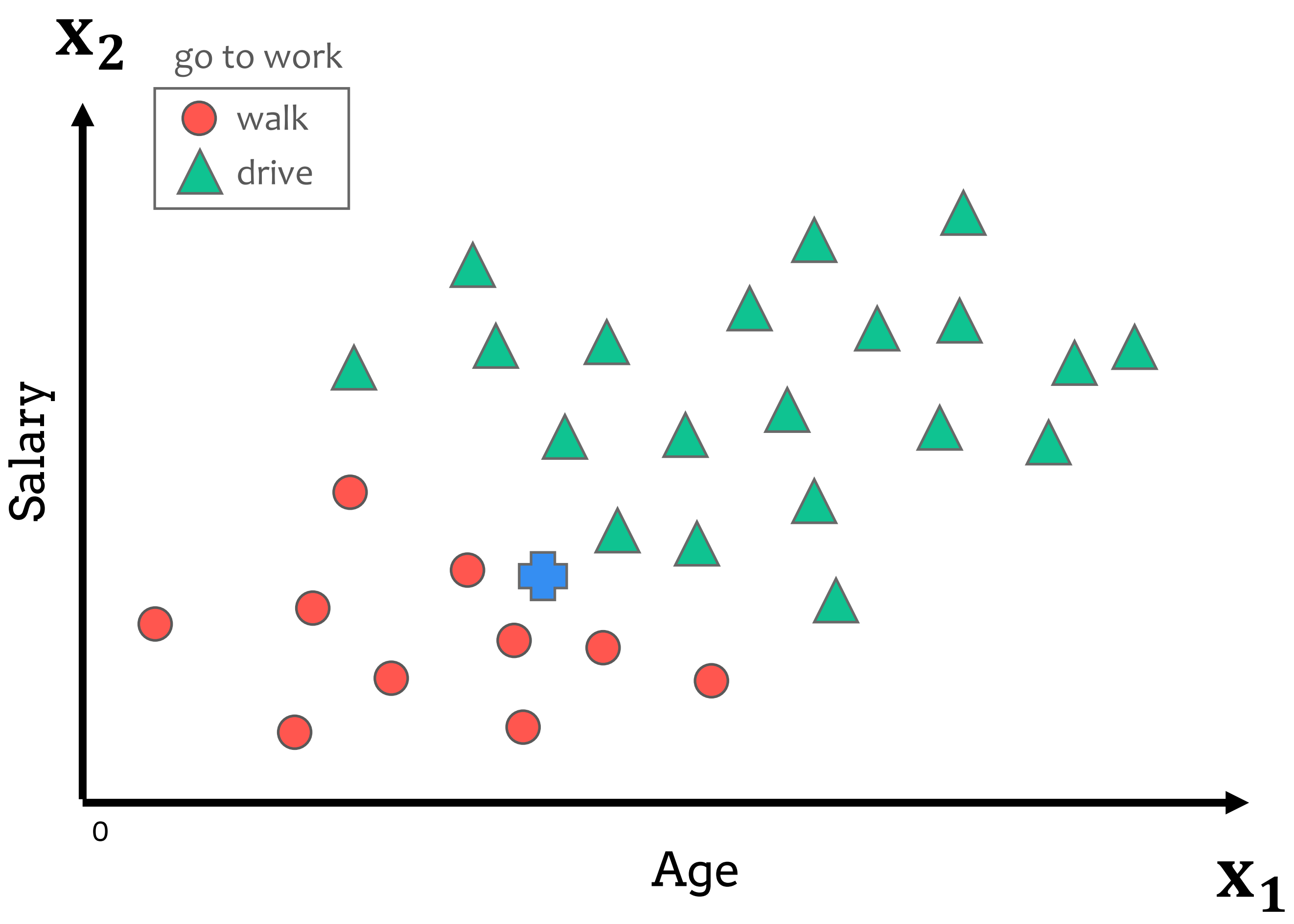
Naive Bayes: Step 1



#1 - $P(Walks)$

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

Naive Bayes: Step 1



#1 - $P(Walks)$

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

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

Naive Bayes

#3

Likelihood

Prior probability #1

class

feats of the new sample

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

#4 Posterior probability

Marginal Likelihood

#2

Naive Bayes

#3 Likelihood

Prior probability #1 (OK!)

class

feats of the new sample

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

#4 Posterior probability

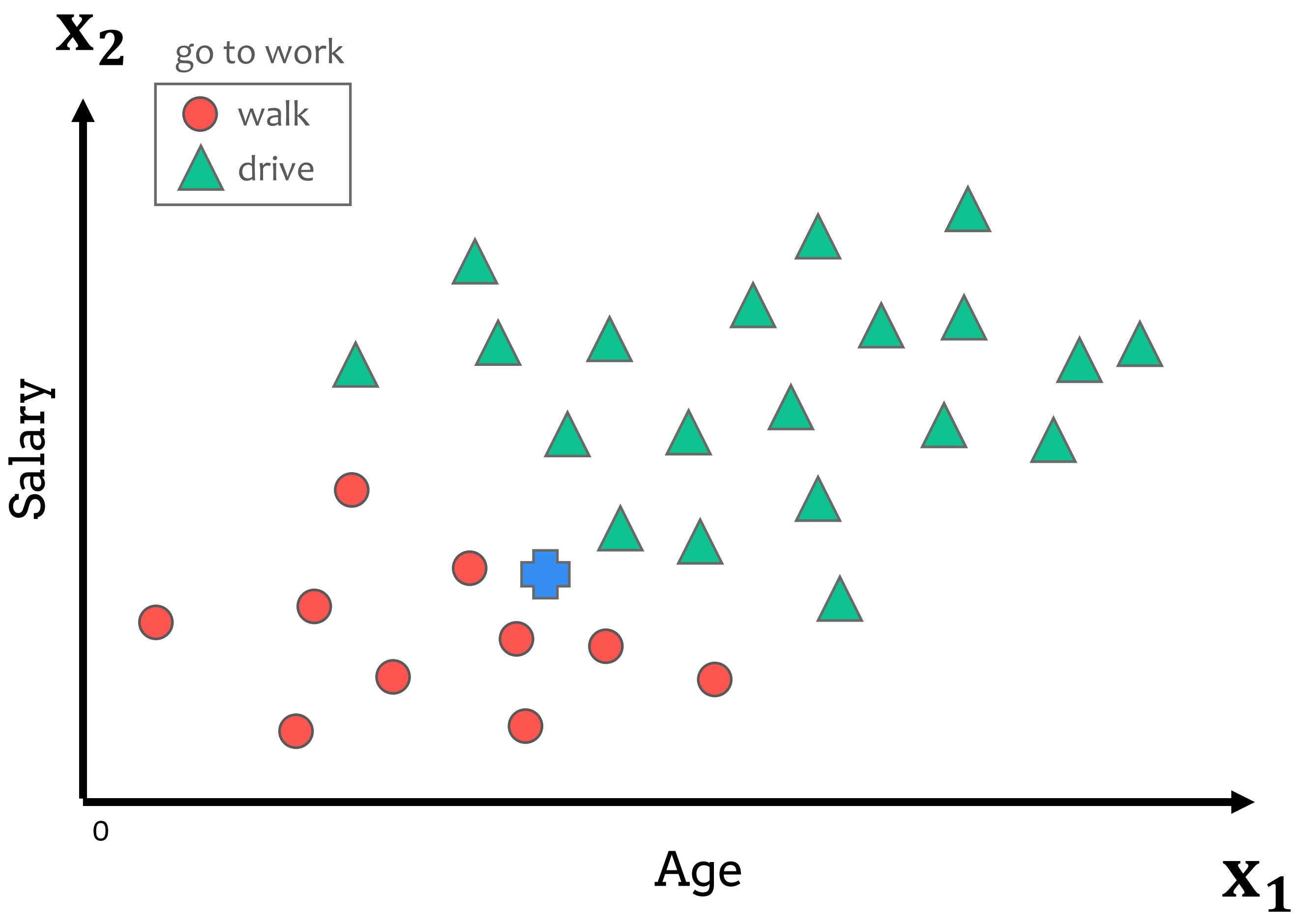
Marginal Likelihood

#2

Naive Bayes: Step 2

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

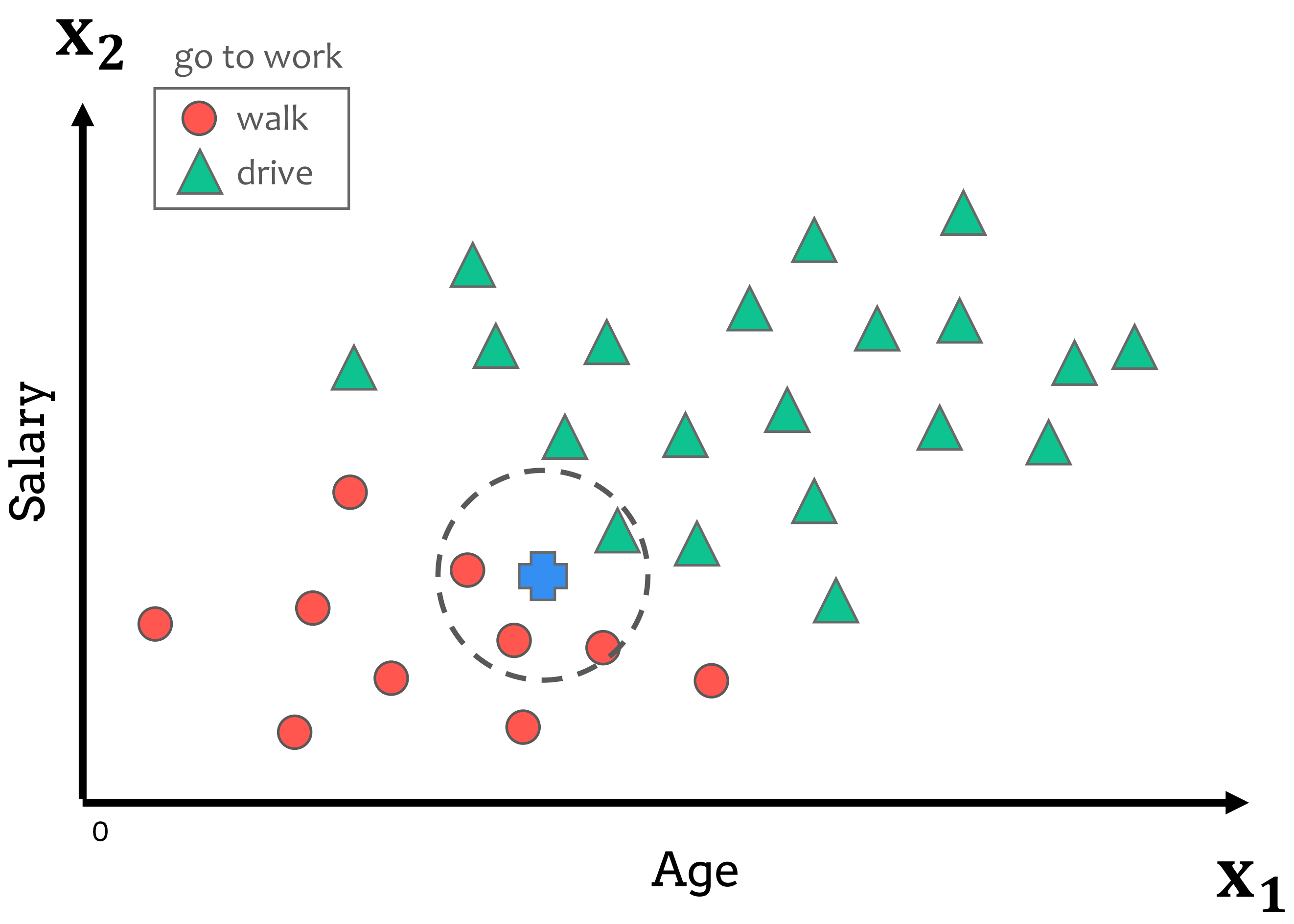
Naive Bayes: Step 2



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

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

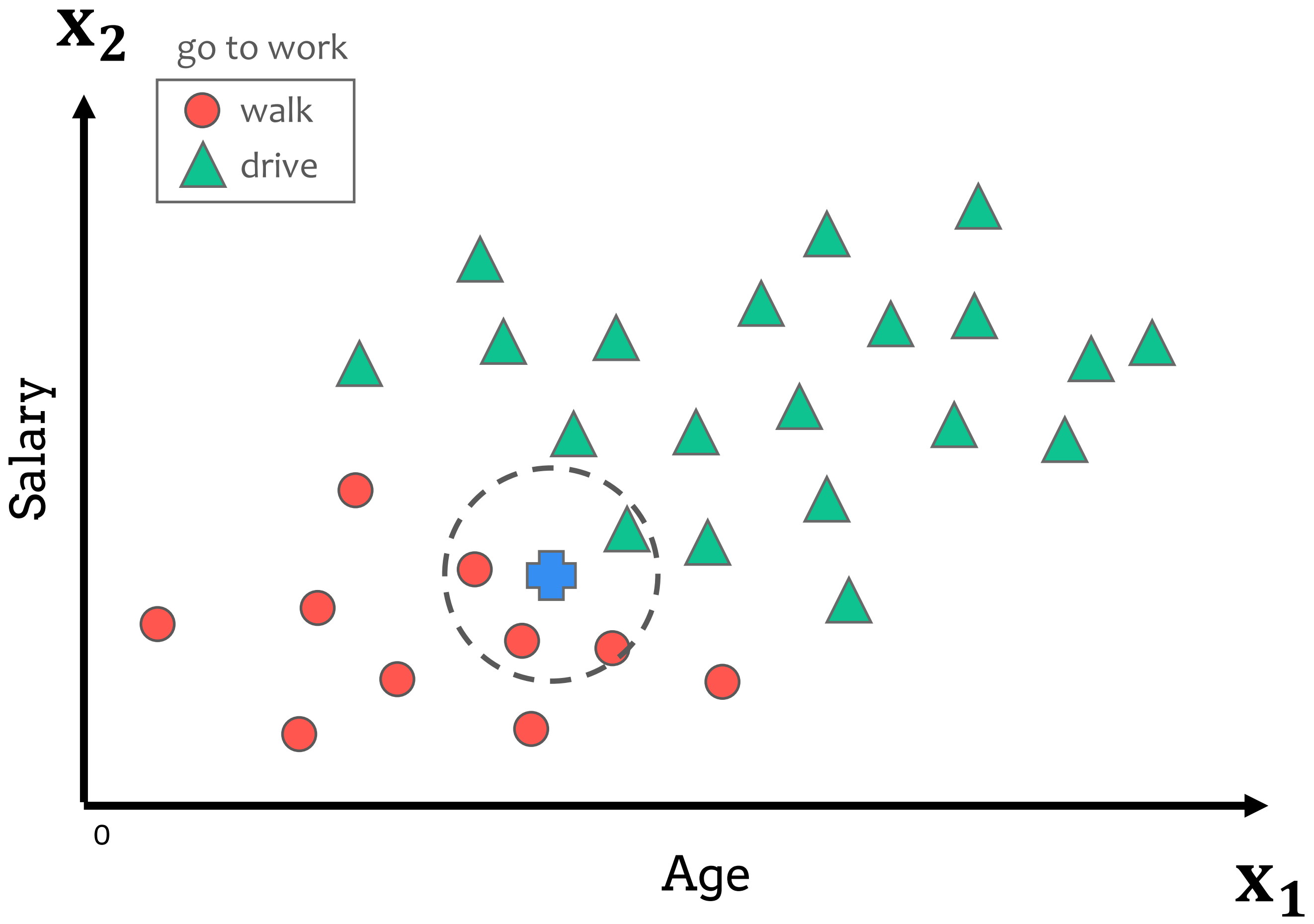


#2 - $P(x_{test})$

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



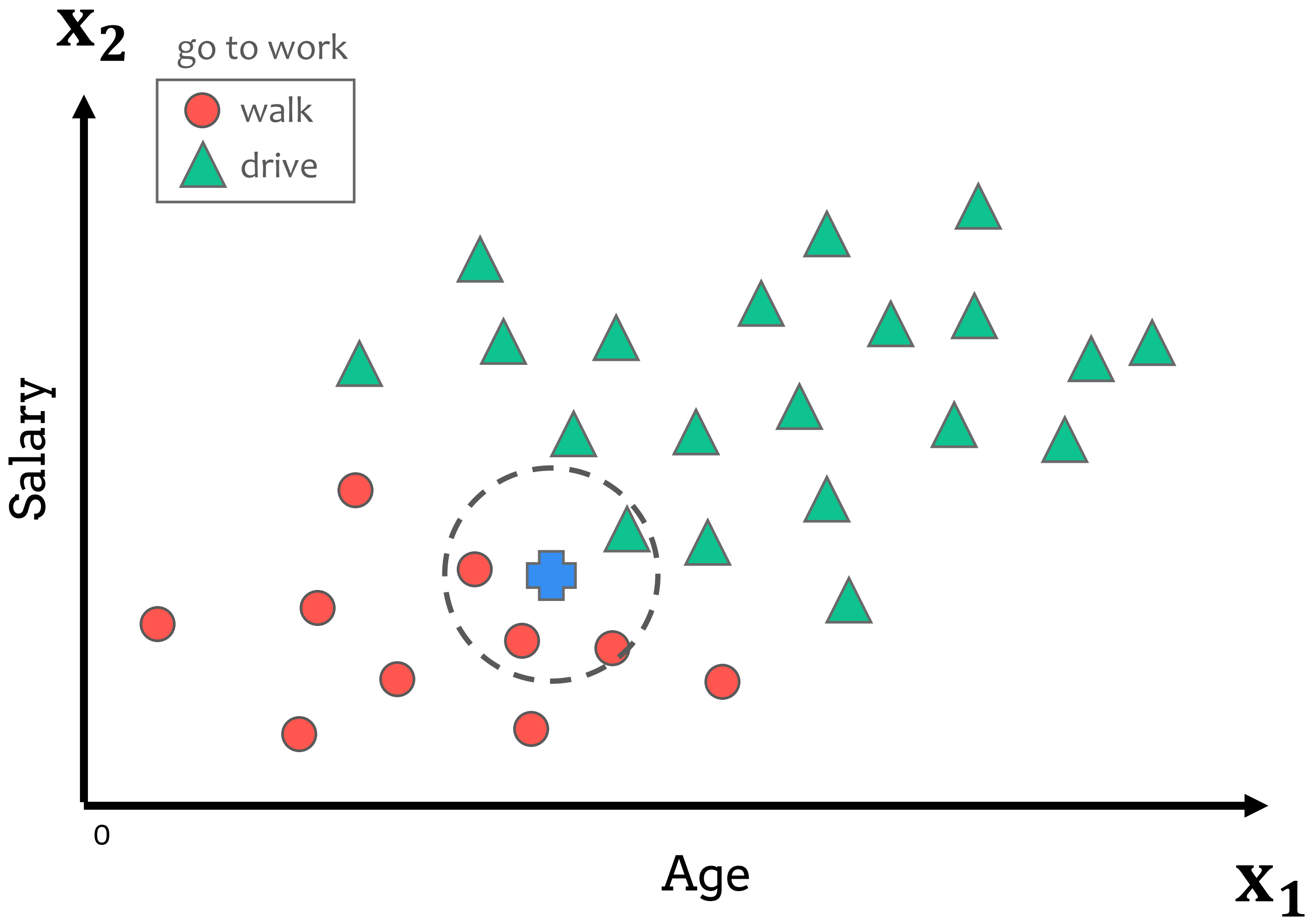
#2 - $P(x_{test})$

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**.

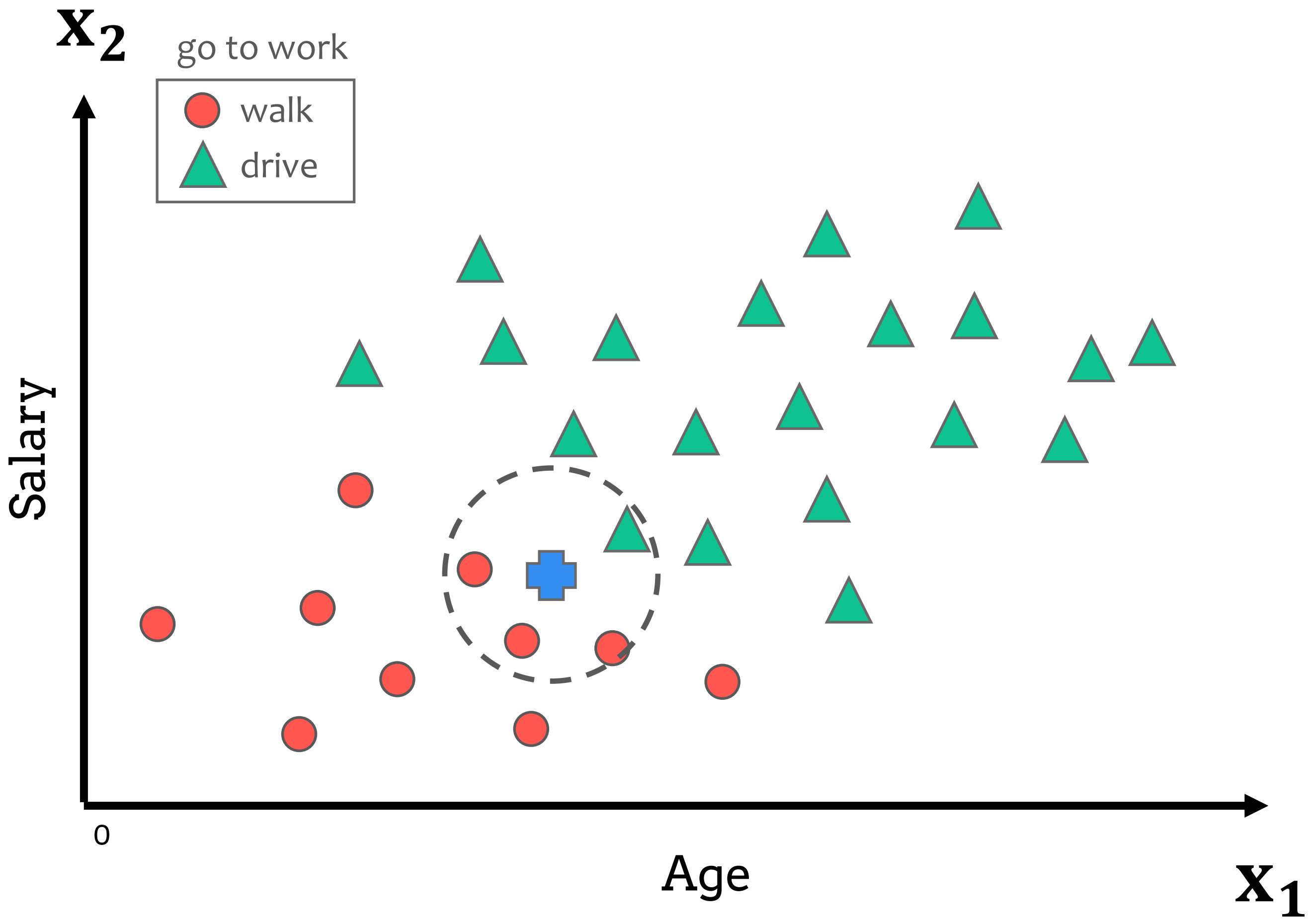
Naive Bayes: Step 2



#2 - $P(\mathbf{x}_{test})$

$$P(\mathbf{x}_{test}) = \frac{Number\ of\ Similar\ Train.Samples}{Total\ Train.Samples}$$

Naive Bayes: Step 2

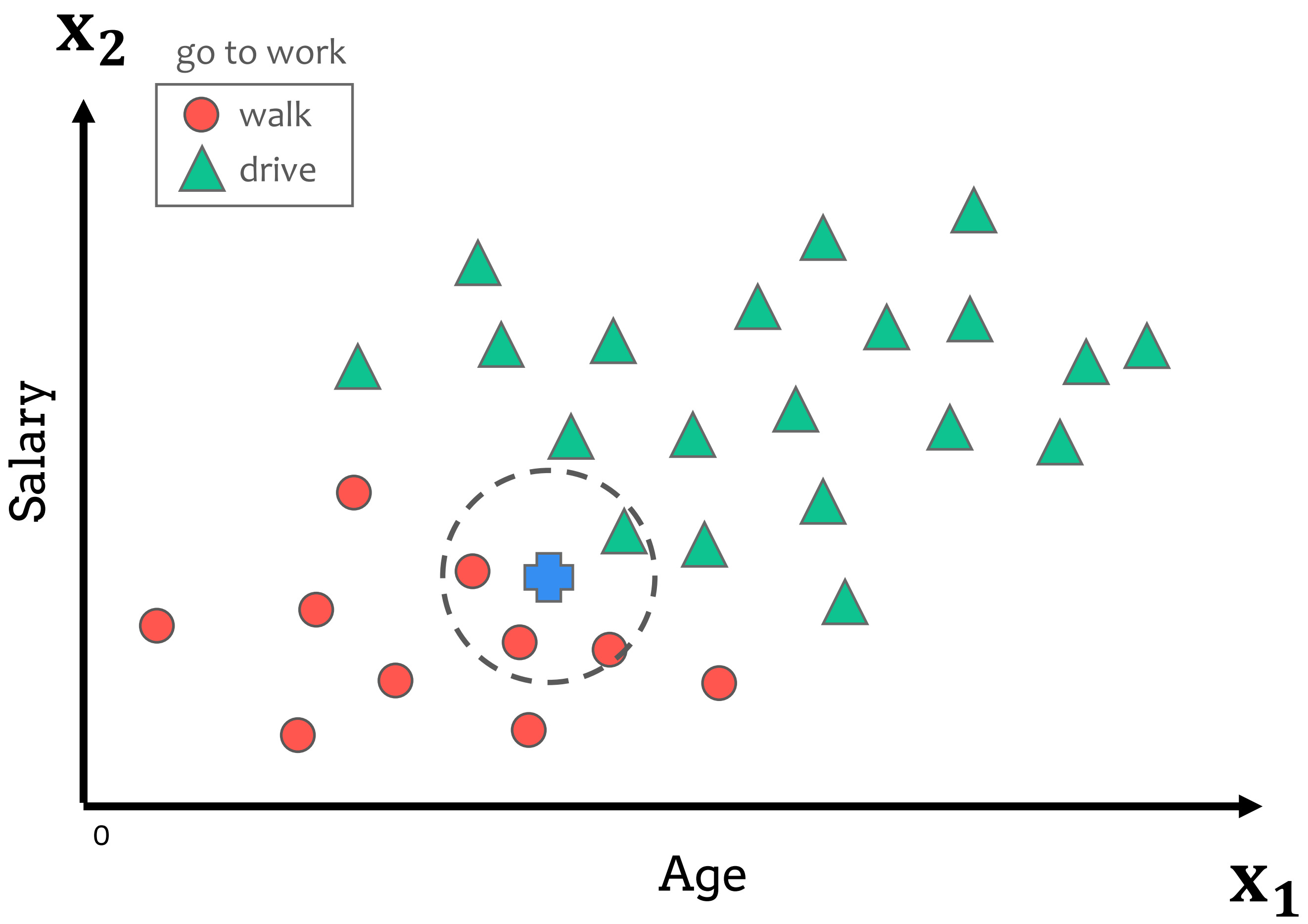


#2 - $P(\mathbf{x}_{\mathbf{test}})$

$$P(\mathbf{x}_{\mathbf{test}}) = \frac{Number\ of\ Similar\ Train.Samples}{Total\ Train.Samples}$$

$$P(\mathbf{x}_{\mathbf{test}}) = \frac{4}{30}$$

Naive Bayes: Step 2



#2 - $\mathrm{P}(\mathbf{x}_{\mathbf{test}})$

$$\mathrm{P}(\mathbf{x}_{\mathbf{test}}) = \frac{Number\ of\ Similar\ Train.Samples}{Total\ Train.Samples}$$

$$\mathrm{P}(\mathbf{x}_{\mathbf{test}}) = \frac{4}{30}$$

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

Naive Bayes

#3
Likelihood

Prior probability #1 (OK!)

class

feats of the new sample

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

#4 Posterior probability

Marginal Likelihood
#2

Naive Bayes

#3

Likelihood

Prior probability #1 (OK!)

class

feats of the new sample

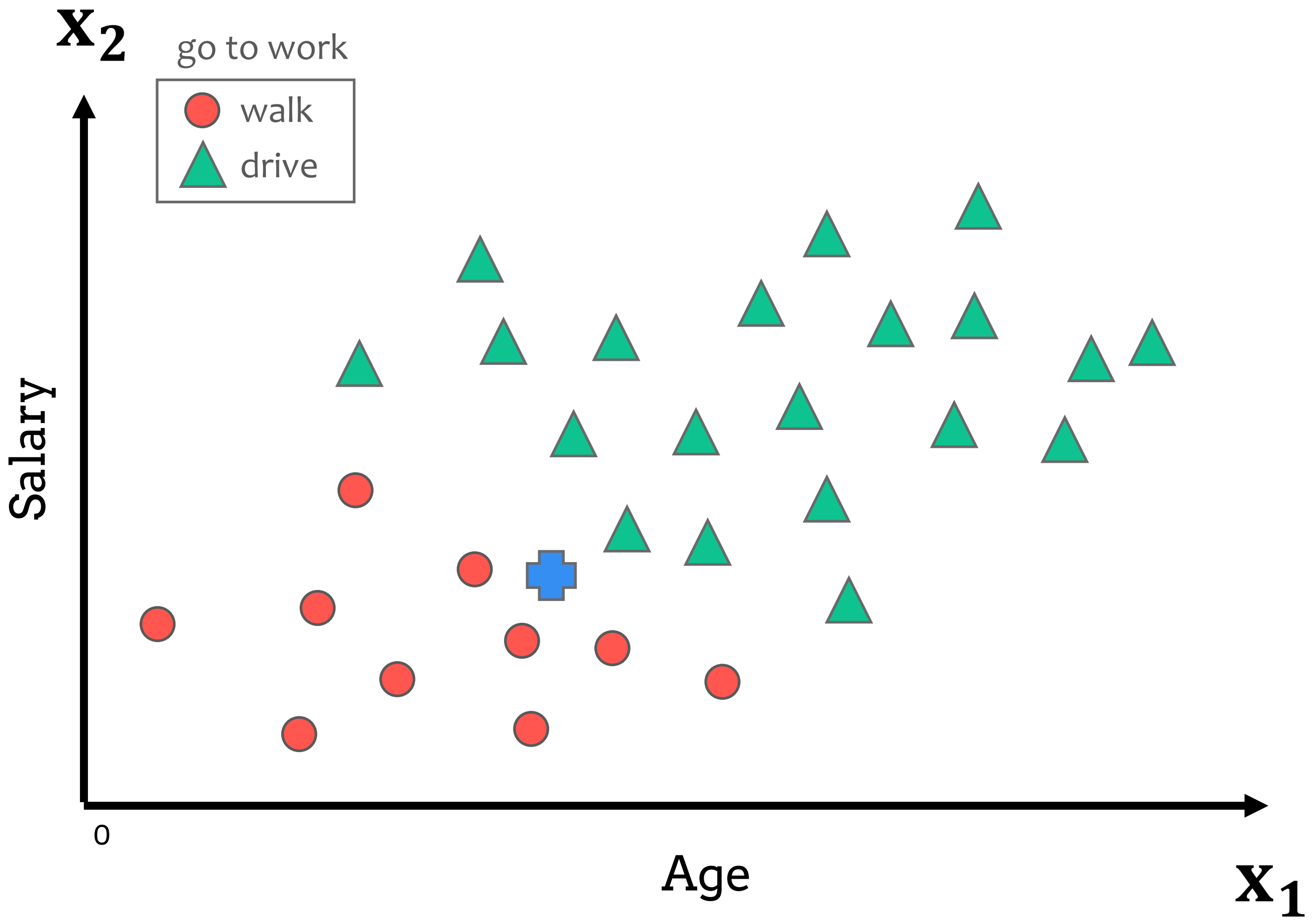
$$P(Walks|x_{test}) = \frac{P(x_{test}|Walks) * P(Walks)}{P(x_{test})}$$

#4 **Posterior probability**

Marginal Likelihood

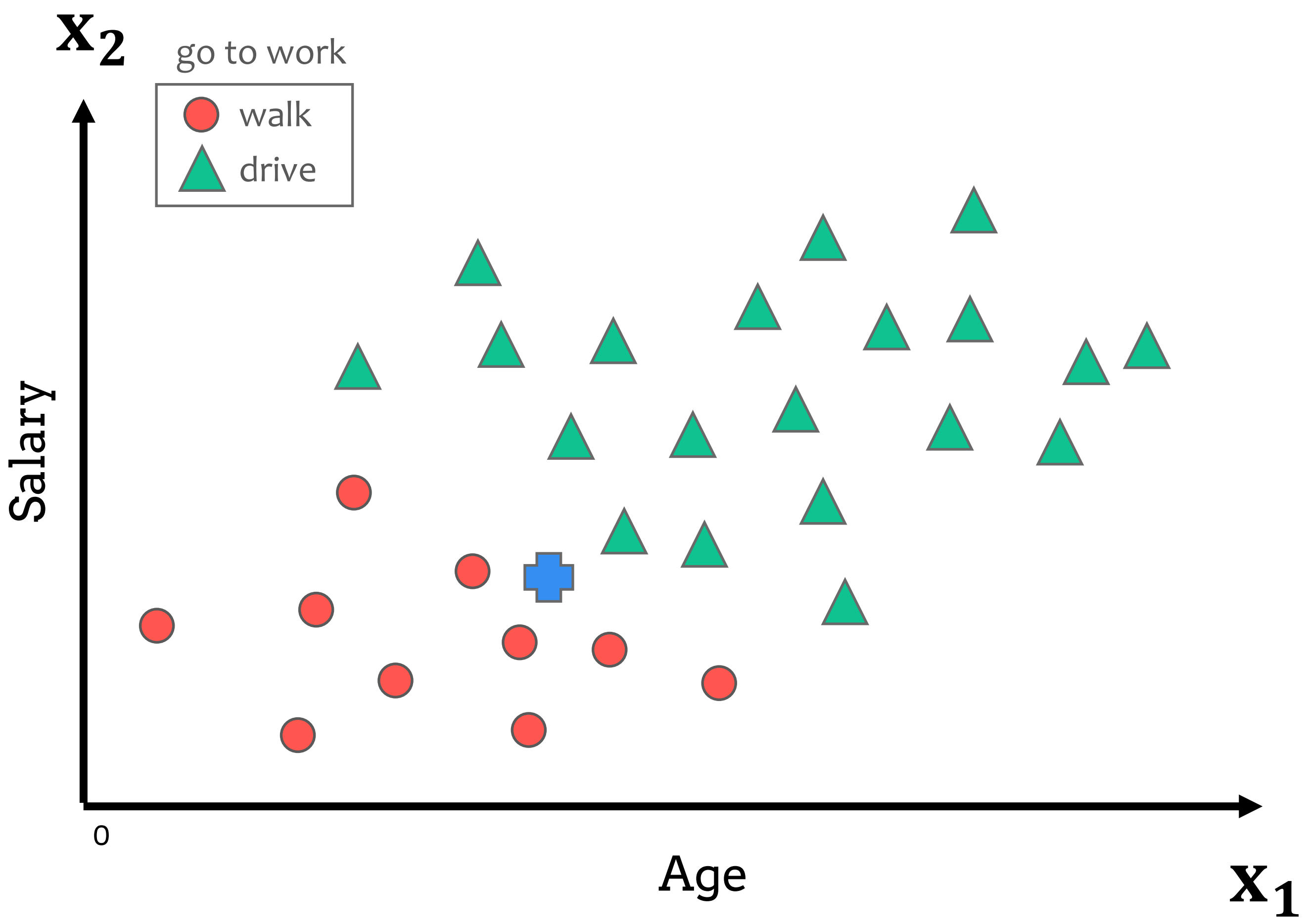
#2 (OK!)

Naive Bayes: Step 3



#3 - $\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks)$

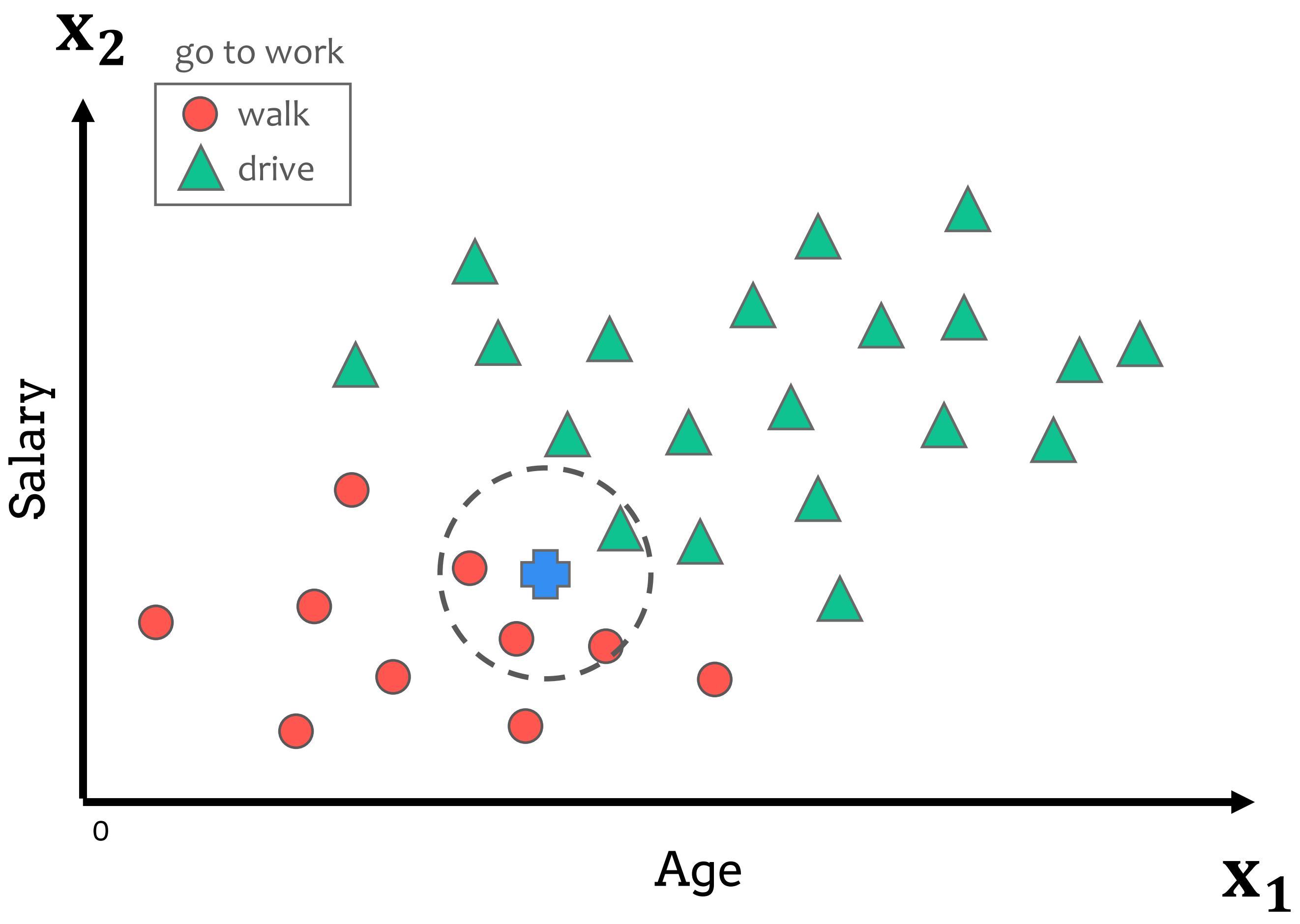
Naive Bayes: Step 3



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

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

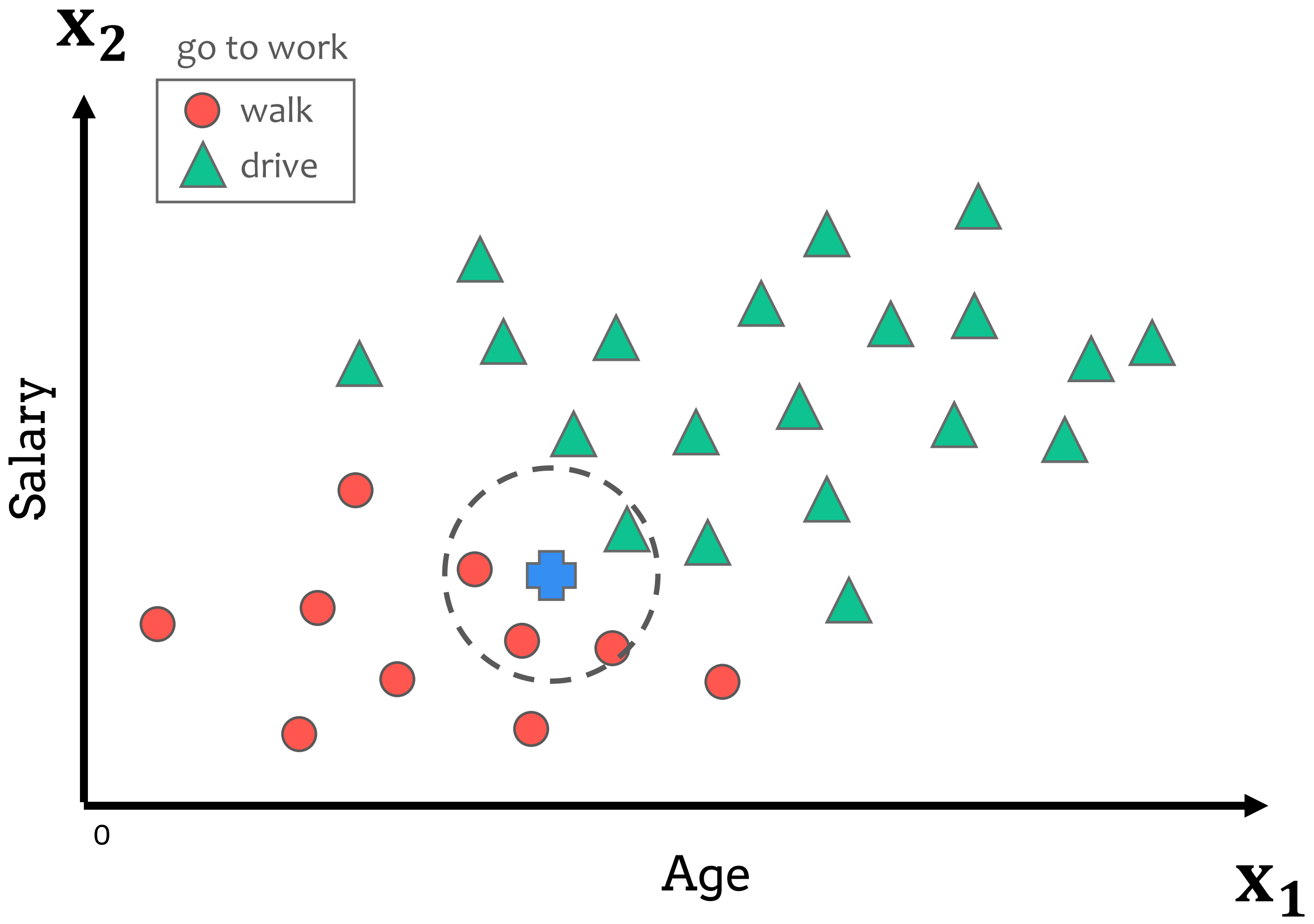


#3 - $P(\mathbf{x}_{\text{test}} | Walks)$

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



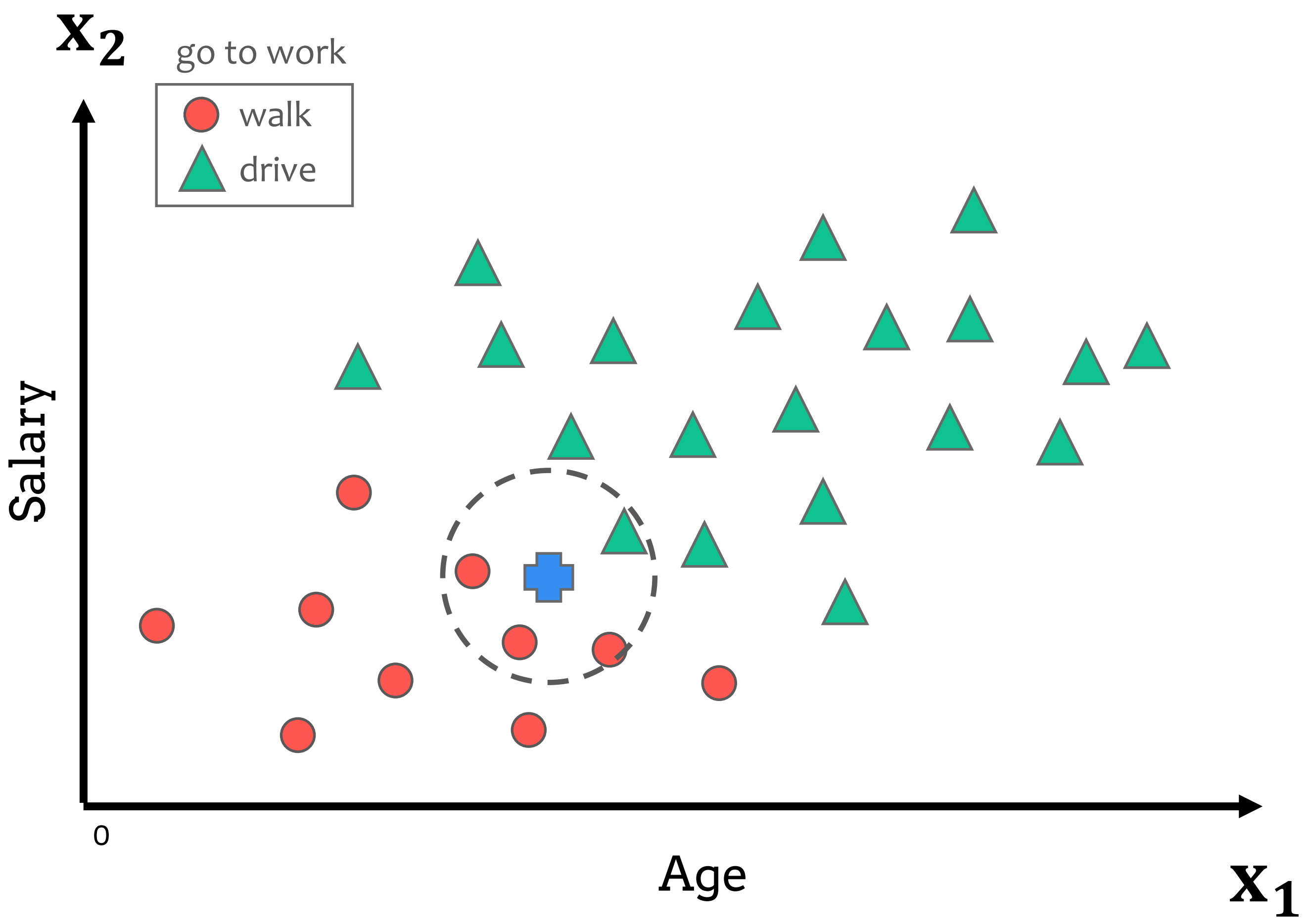
#3 - $P(x_{test}|Walks)$

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.

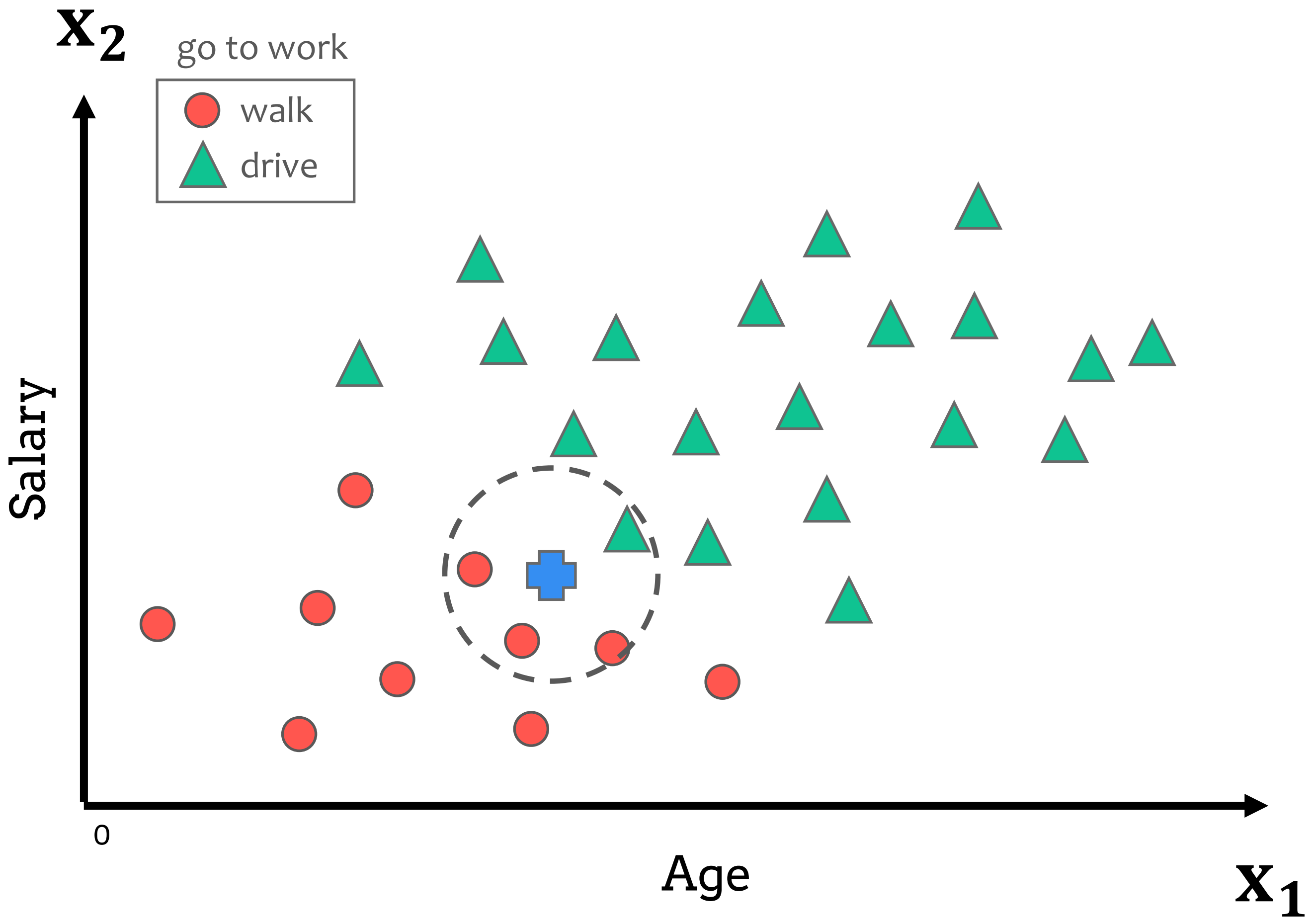
Naive Bayes: Step 3



#3 - $\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks)$

$$\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks) = \frac{Number\ of\ similar\ train.samples\ with\ class\ Walk}{Total\ train.\ samples\ with\ class\ Walk}$$

Naive Bayes: Step 3

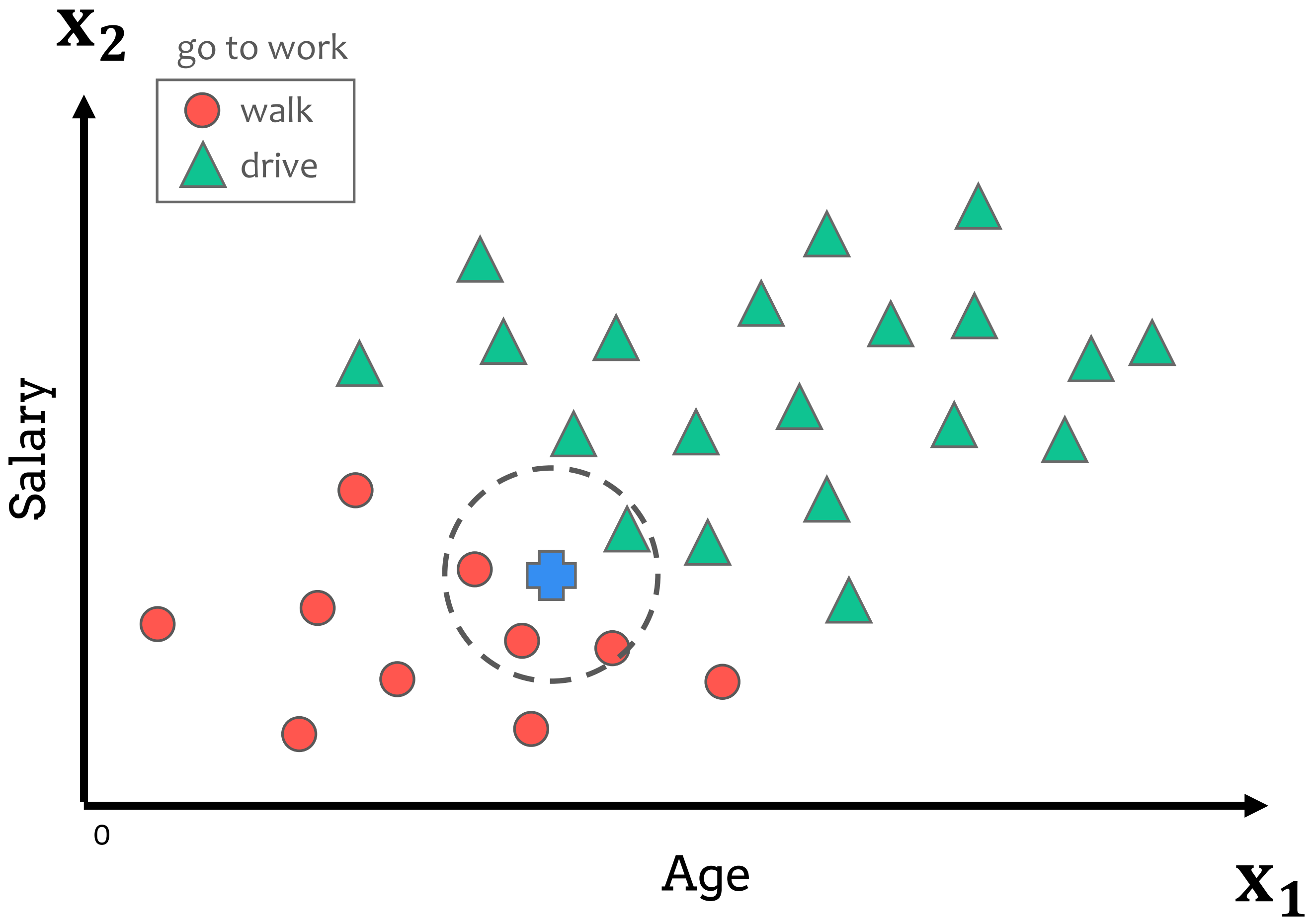


#3 - $\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks)$

$$\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks) = \frac{Number\ of\ similar\ train.samples\ with\ class\ Walk}{Total\ train.\ samples\ with\ class\ Walk}$$

$$\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks) = \frac{3}{10}$$

Naive Bayes: Step 3



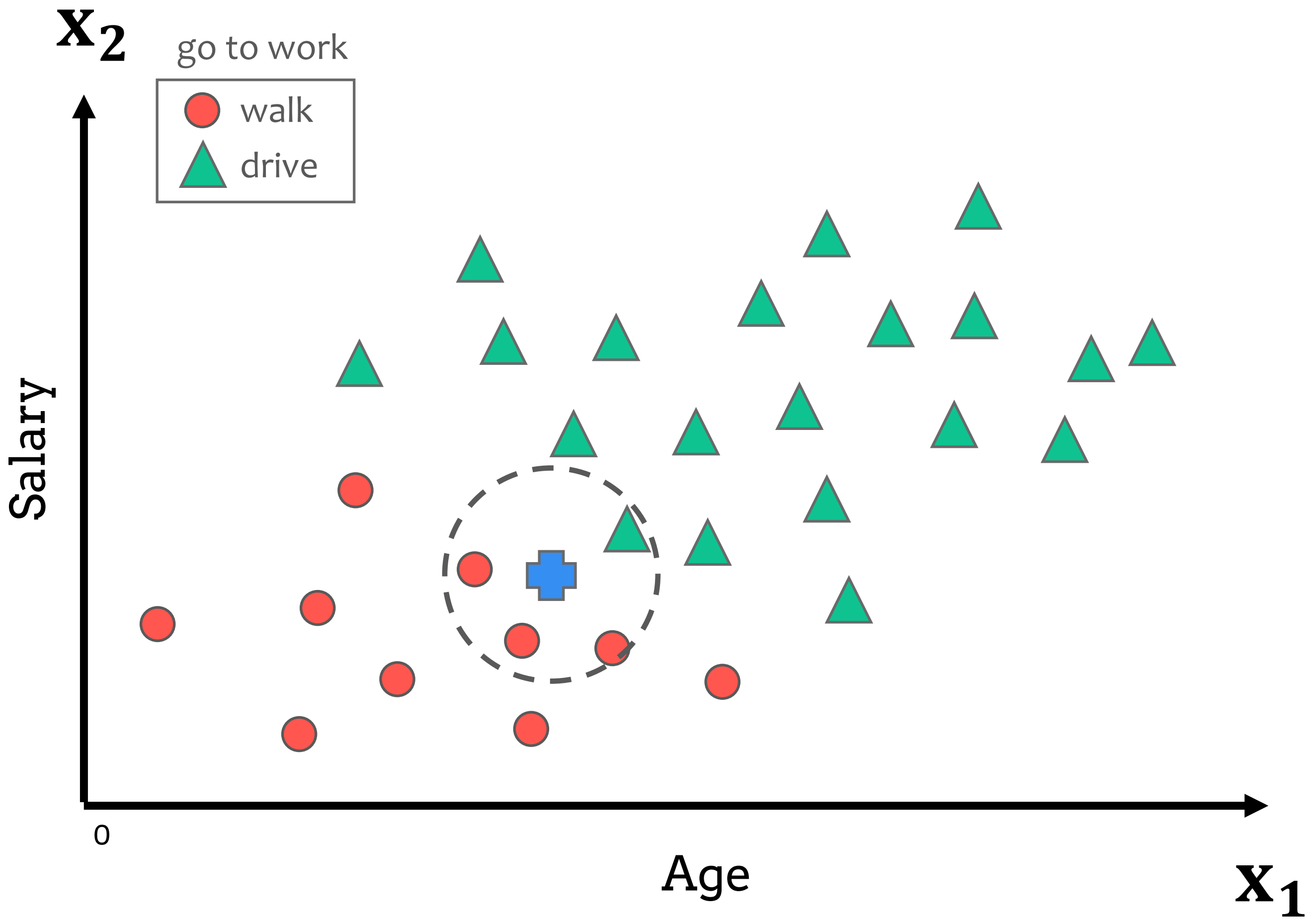
#3 - $\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks)$

$$\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks) = \frac{Number\ of\ similar\ train.samples\ with\ class\ Walk}{Total\ train.\ samples\ with\ class\ Walk}$$

$$\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks) = \frac{3}{10}$$

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 - $\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks)$

$$\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks) = \frac{Number\ of\ similar\ train.samples\ with\ class\ Walk}{Total\ train.\ samples\ with\ class\ Walk}$$

$$\mathrm{P}(\mathbf{x}_{\mathbf{test}}|Walks) = \frac{3}{10}$$

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

#3

Likelihood

Prior probability **#1 (OK!)**

class feats of the new sample

$$P(\textcolor{red}{Walks}|\mathrm{x}_{test}) = \frac{P(\mathrm{x}_{test}|\textcolor{red}{Walks}) * P(\textcolor{red}{Walks})}{P(\mathrm{x}_{test})}$$

#4 **Posterior probability**

Marginal Likelihood

#2 (OK!)

Naive Bayes

#3 (OK!)

Likelihood

Prior probability #1 (OK!)

class

feats of the new sample

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

#4 Posterior probability

Marginal Likelihood

#2 (OK!)

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***

class

feats of the new sample

#3 Likelihood

Prior probability #1

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

#4 Posterior probability

Marginal Likelihood

#2

Naive Bayes

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

Naive Bayes

Step 3

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

Naive Bayes

Step 3

$$P(Walks|\mathrm{x}_{test}) \ \textbf{vs} \ P(Drives|\mathrm{x}_{test})$$

$$\mathbf{0.75} \qquad \mathbf{0.25}$$

Naive Bayes

Step 3

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

$$\mathbf{0.75} \qquad \mathbf{0.25}$$

one is complement of the other

Naive Bayes

Step 3

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

$$\mathbf{0.75} \qquad \mathbf{0.25}$$

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.

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

Why Naive?

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

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

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

Why Naive?

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

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

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

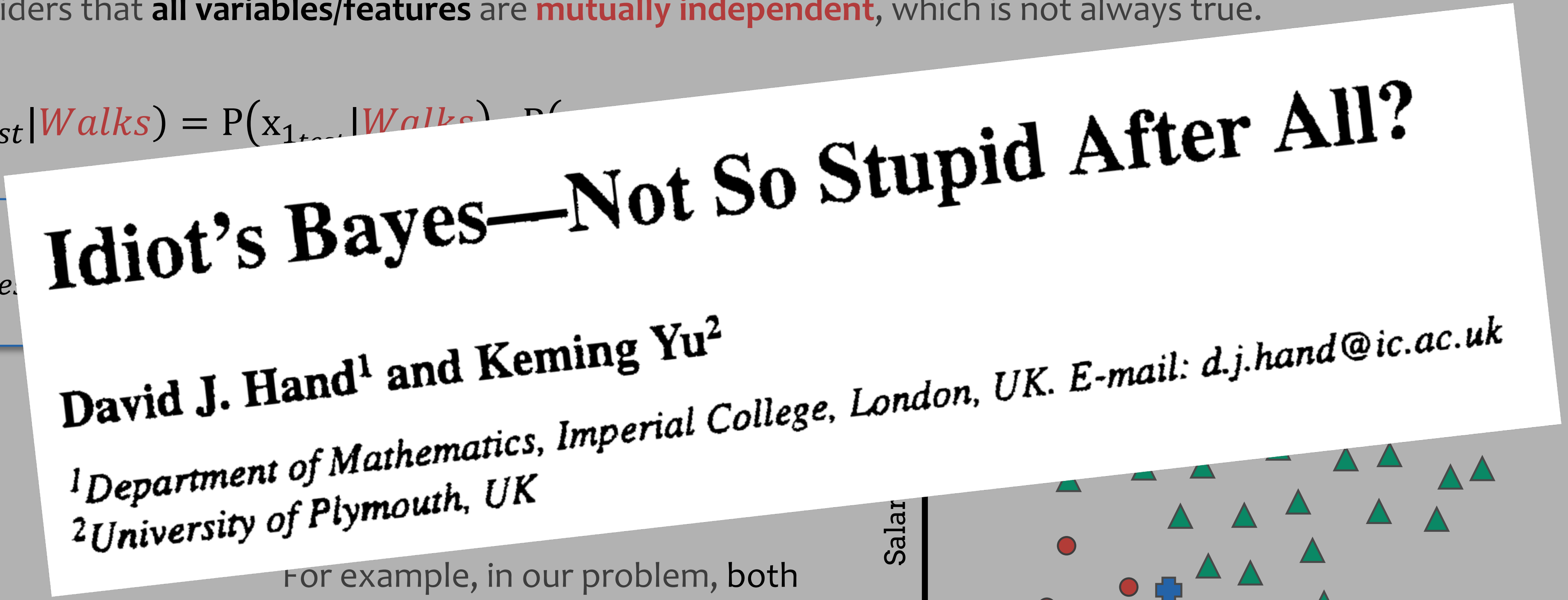
For example, in our problem, both features seems **to be correlated**.

Why Naive?

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

$P(x_{test}|Walks) = P(x_{1_{test}}|Walks) \cdot P($

$P(x_{te}$

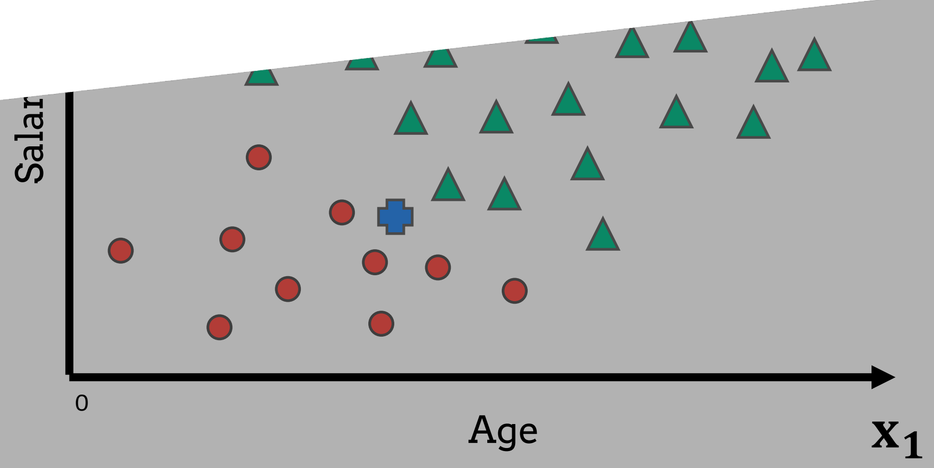

Idiot's Bayes—Not So Stupid After All?

David J. Hand[1] and Keming Yu[2]

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

[2]University of Plymouth, UK

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



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

Do we need to compute $\mathrm{P}(\mathrm{x}_{\mathrm{test}})$?

$\mathrm{P}(Walks|\mathrm{x}_{test})$ **vs** $\mathrm{P}(Drives|\mathrm{x}_{test})$

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

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

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

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

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

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

Aprendizado de Máquina e Reconhecimento de Padrões

Naive Bayes

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

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
São Paulo
Campus Campinas