CS 156:Introduction to Artificial Intelligence

Instructor: Dr. Sayma Akther

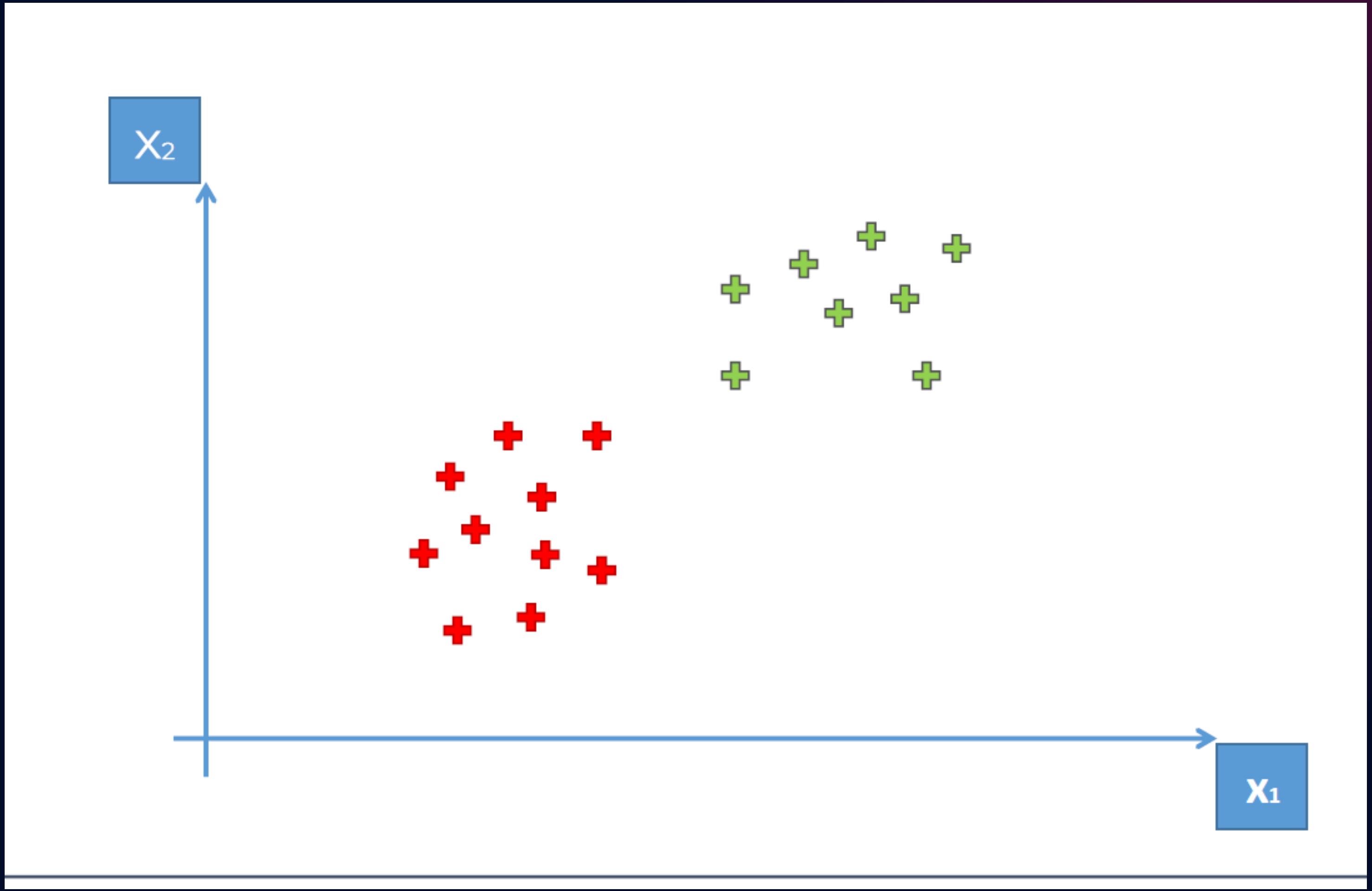
San José State University

SVM (Support Vector Machine) Algorithm

The SVM algorithm is a classification process in which raw data is shown as points in an n-dimensional space (n being the number of features you have)

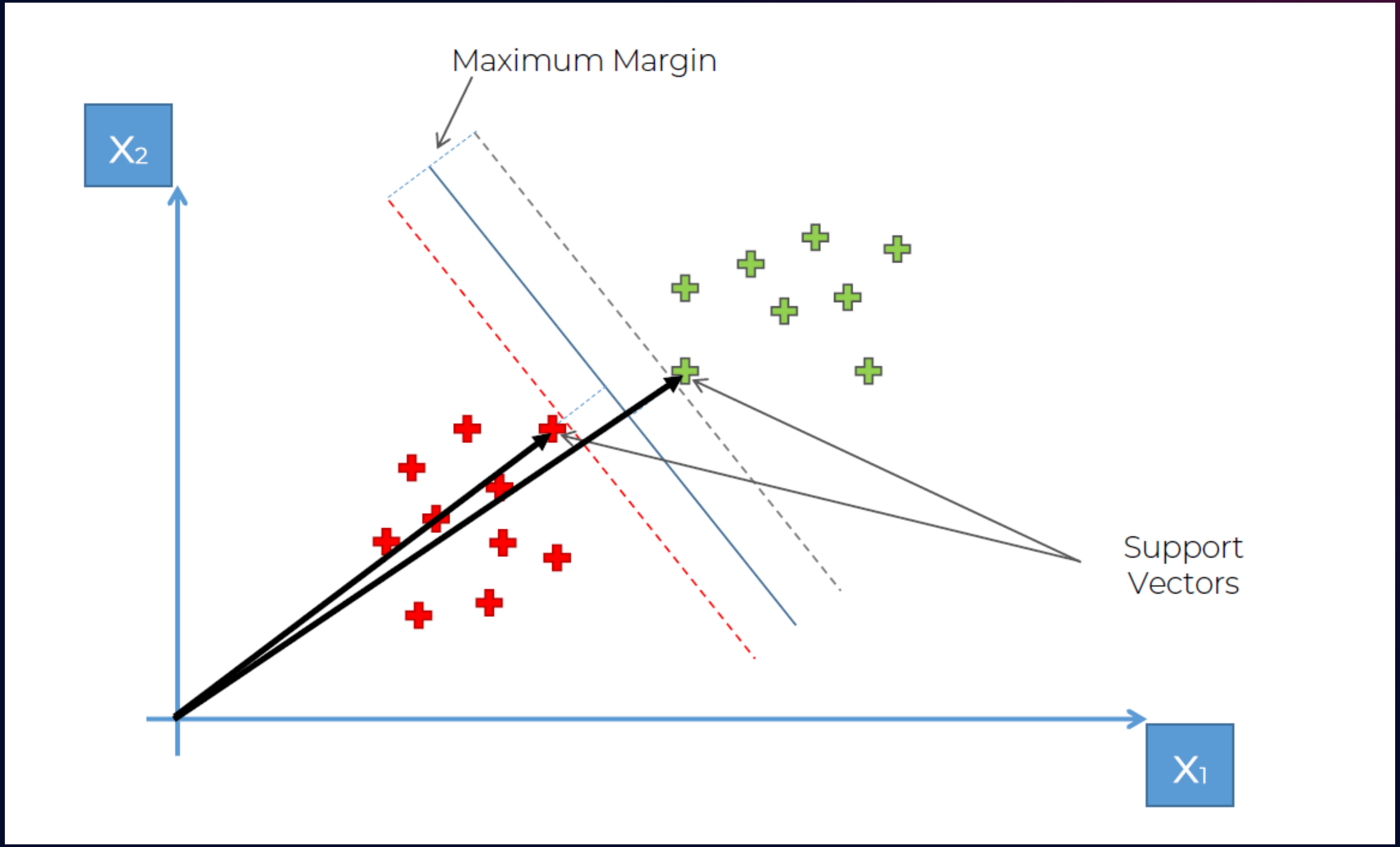
The value of each characteristic is then assigned to a specific location, making it simple to categorize the data. Classifier lines can divide data and plot it on a graph

How to separate these points?



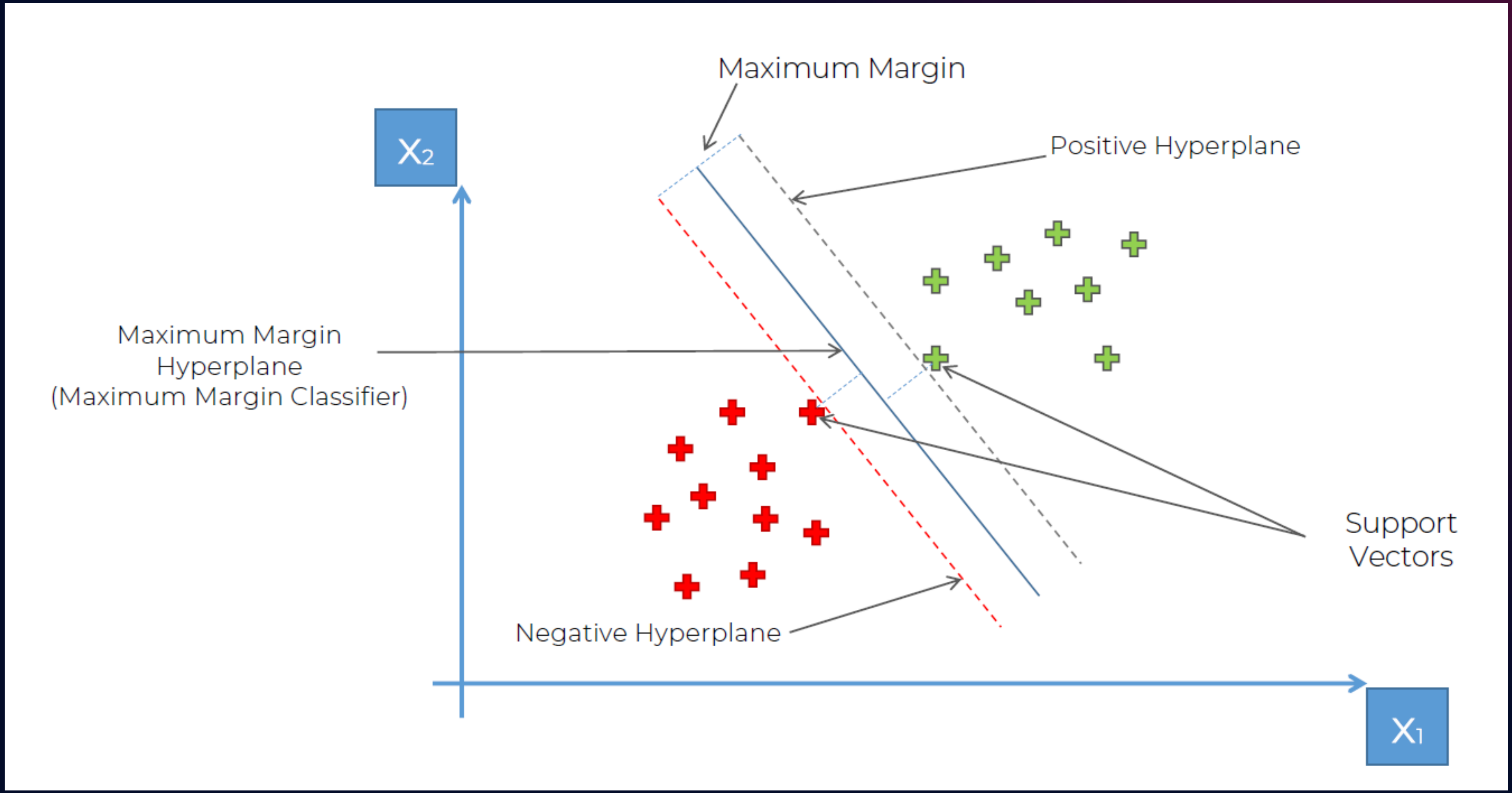
slide~superdatascience

Support Vectors



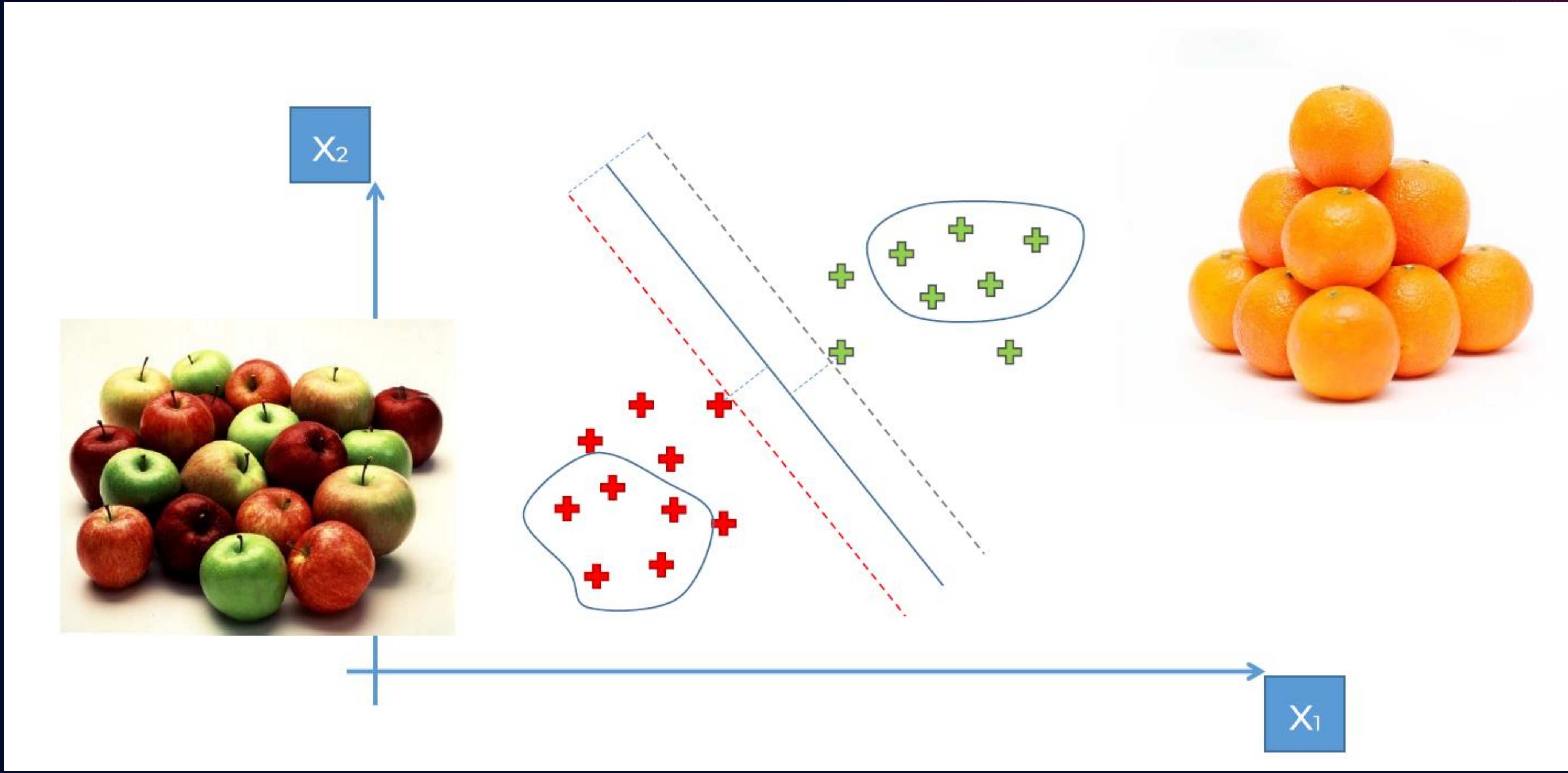
slide~superdatascience

Hyperplanes

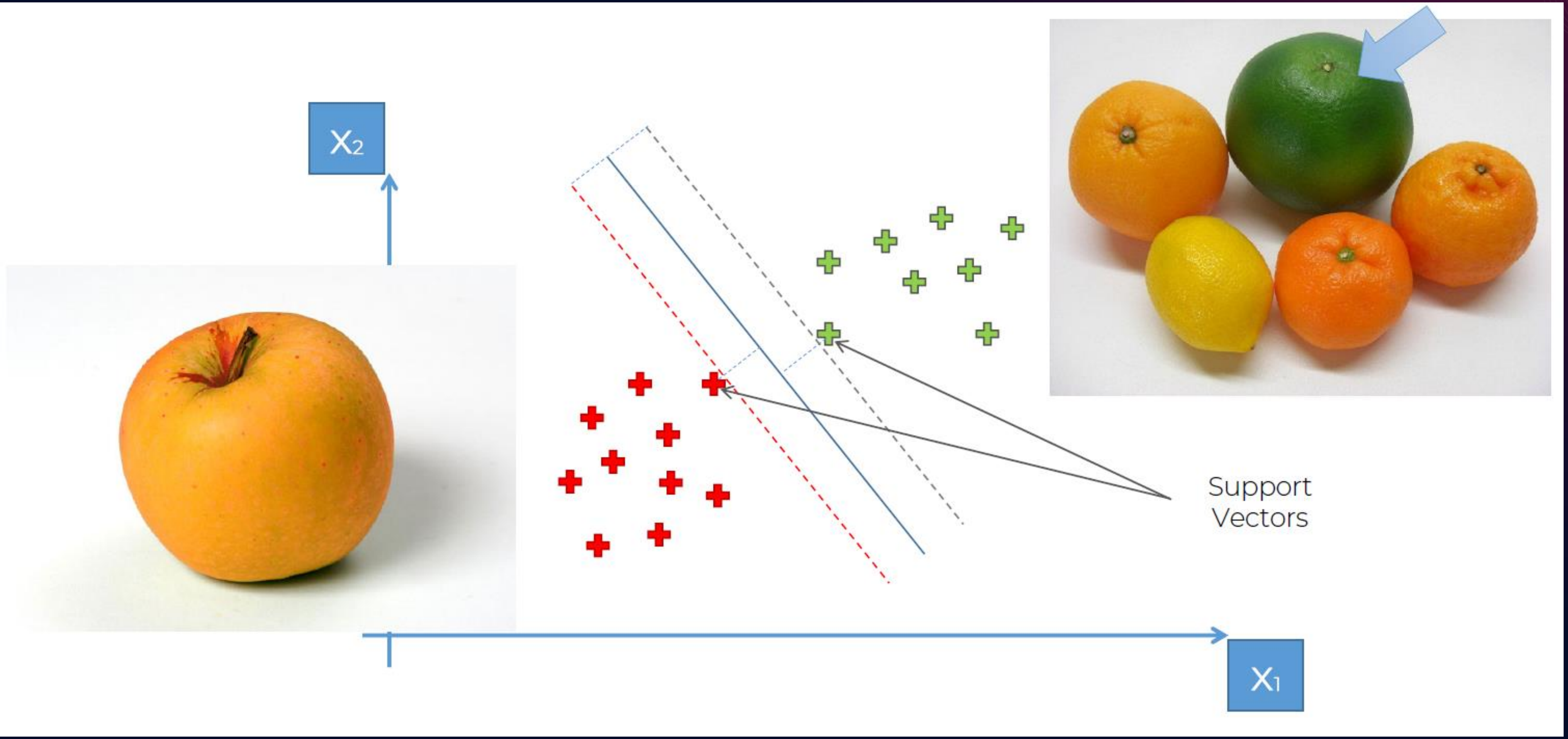


What’s so special about support vector machine

What’s so special about support vector machine



What’s so special about support vector machine



Bayes' Rule

- Exactly the process we just used
- The most important formula in probabilistic machine learning

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

(Super Easy) Derivation:

$$\boxed{\begin{aligned} P(A \land B) \\ P(B \land A) \end{aligned}} \begin{aligned} &= P(A|B) \times P(B) \\ &= P(B|A) \times P(A) \end{aligned}$$

Just set equal...

$$P(A|B) \times P(B) = P(B|A) \times P(A)$$

and solve...

Bayes, Thomas (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53:370-418**

Bayes' Rule for Machine Learning

- Allows us to reason from evidence to hypotheses
- Another way of thinking about Bayes' rule:

$$\boxed{P(\text{hypothesis} \mid \text{evidence}) = \frac{P(\text{evidence} \mid \text{hypothesis}) \times P(\text{hypothesis})}{P(\text{evidence})}}$$

Naive Bayes Algorithm

Naive Bayes is a probabilistic Machine Learning technique based on the Bayes Theorem and is used for a wide range of classification problems

According to a Naive Bayes classifier, the presence of one feature in a class does not influence the existence of any other feature

A Naive Bayesian model is straightforward to build and works well with massive datasets. It is simple to use and has been demonstrated to outperform even the most sophisticated classification algorithms

Naïve Bayes

Step 1:

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

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

#2 Marginal Likelihood

Step 2:

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

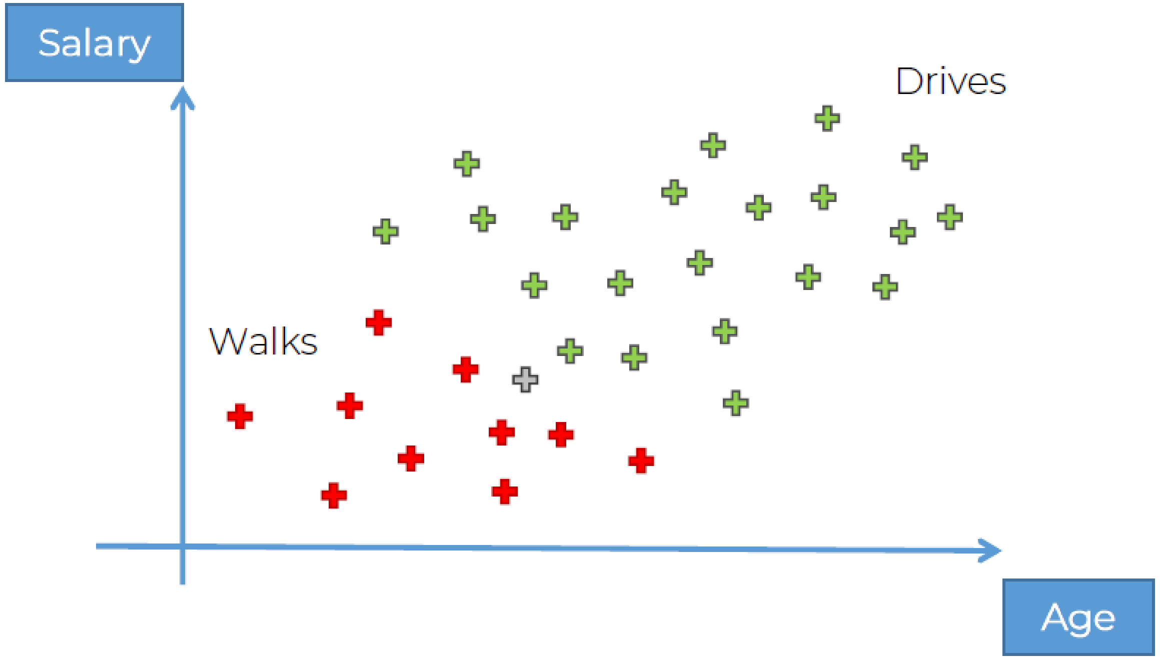
$$P(Drives|X) = \frac{P(X|Drives) * P(Drives)}{P(X)}$$

#2 Marginal Likelihood

Step 3:

$$P(Walks|X)\ v.s.\ P(Drives|X)$$

Naïve Bayes: Step 1



#1. P(Walks)

$$P(Walks) = \frac{Number\ of\ Walkers}{Total\ Observations}$$

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

Naïve Bayes: Step 1

#4 Posterior Probability

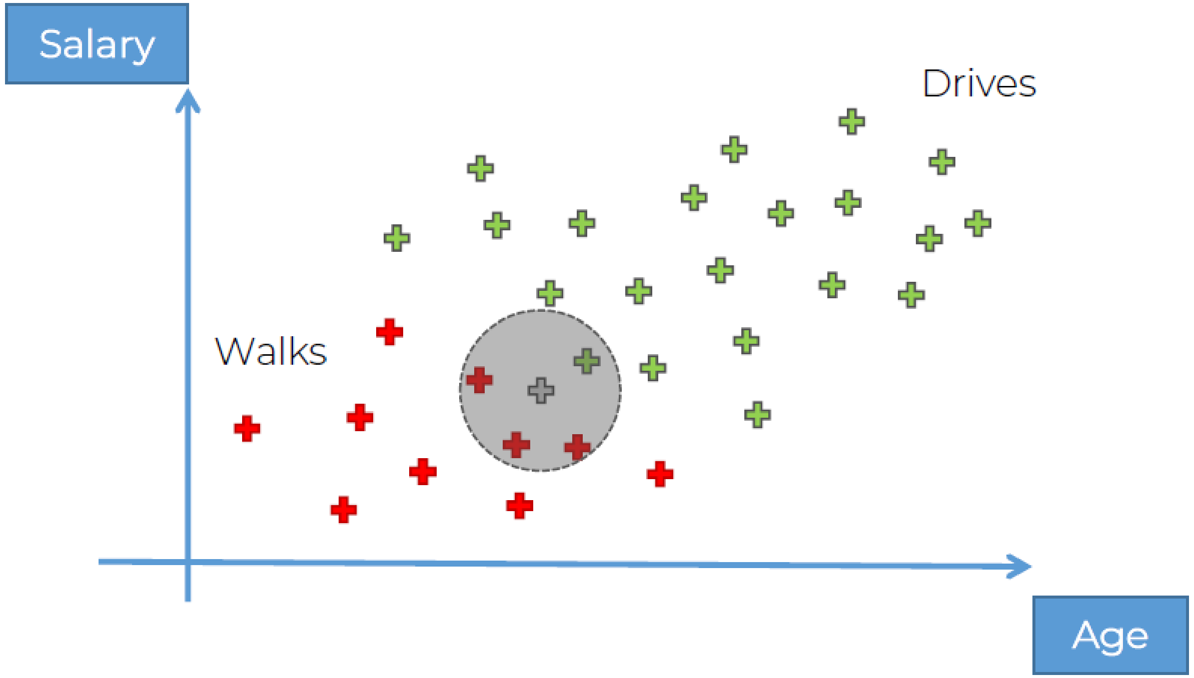
#3 Likelihood

#1 Prior Probability

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

#2 Marginal Likelihood

Naïve Bayes: Step 1



#2. P(X)

$$P(X) = \frac{Number\ of\ Similar\ Observations}{Total\ Observations}$$

$$P(X) = \frac{4}{30}$$

Naïve Bayes: Step 1

#4 Posterior Probability

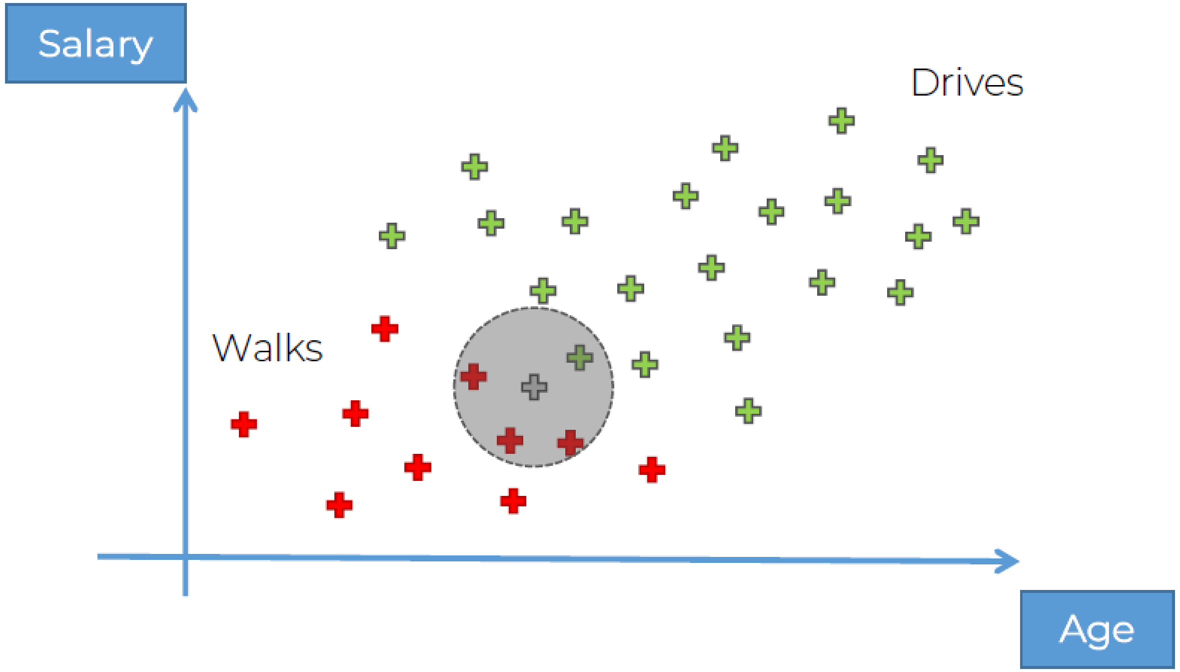
#3 Likelihood

#1 Prior Probability

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

#2 Marginal Likelihood

Naïve Bayes: Step 1



#3. P(X|Walks)

$$P(X|Walks) = \frac{\begin{array}{c}Number\ of\ Similar\\ Observations\\ Among\ those\ who\ Walk\end{array}}{Total\ number\ of\ Walkers}$$

$$P(X|Walks) = \frac{3}{10}$$

Naïve Bayes: Step 1

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

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

#2 Marginal Likelihood

Naïve Bayes: Step 1

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

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

#2 Marginal Likelihood

Naïve Bayes: Step 2

#4 Posterior Probability

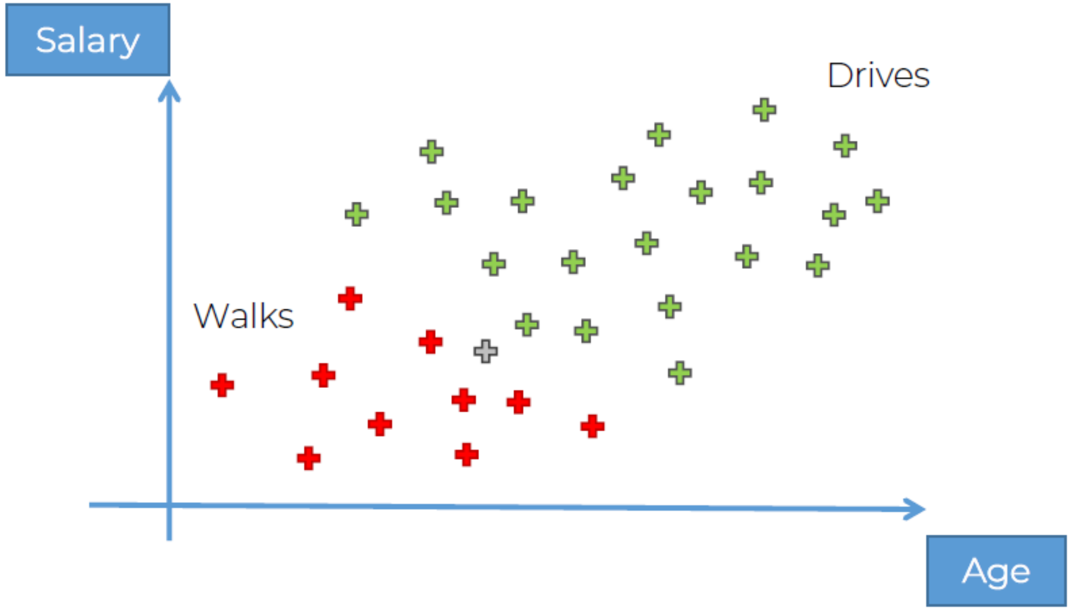
#3 Likelihood

#1 Prior Probability

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

#2 Marginal Likelihood

Naïve Bayes: Step 2



#1. P(Drives)

$$P(Drives) = \frac{Number\ of\ Drivers}{Total\ Observations}$$

$$P(Drives) = \frac{20}{30}$$

Naïve Bayes: Step 2

#4 Posterior Probability

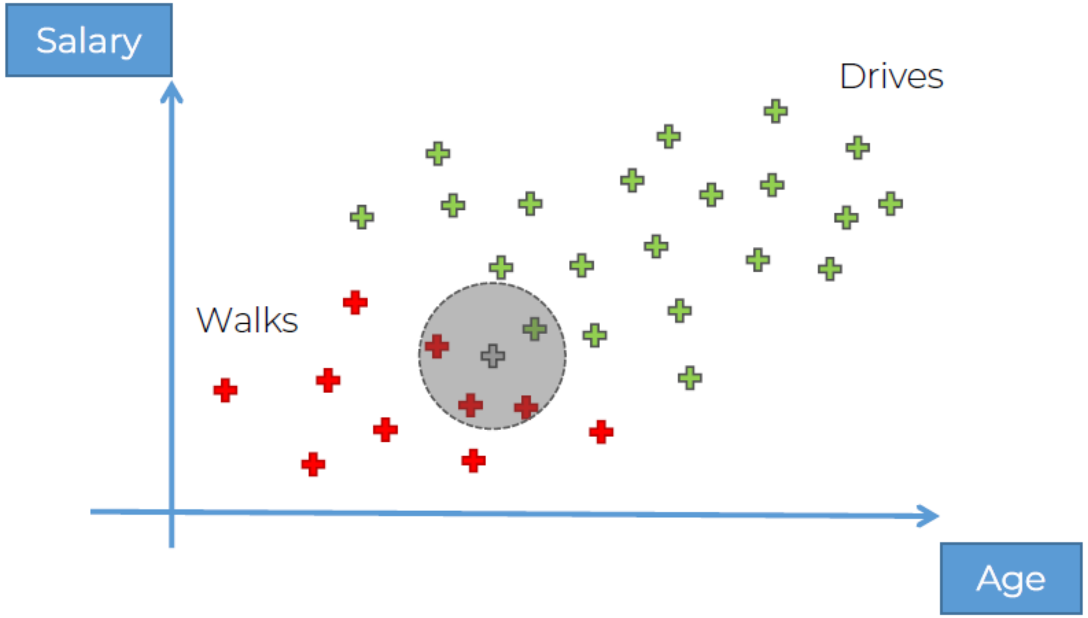
#3 Likelihood

#1 Prior Probability

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

#2 Marginal Likelihood

Naïve Bayes: Step 2



#2. P(X)

$$P(X) = \frac{Number\ of\ Similar\ Observations}{Total\ Observations}$$

$$P(X) = \frac{4}{30}$$

Naïve Bayes: Step 2

#4 Posterior Probability

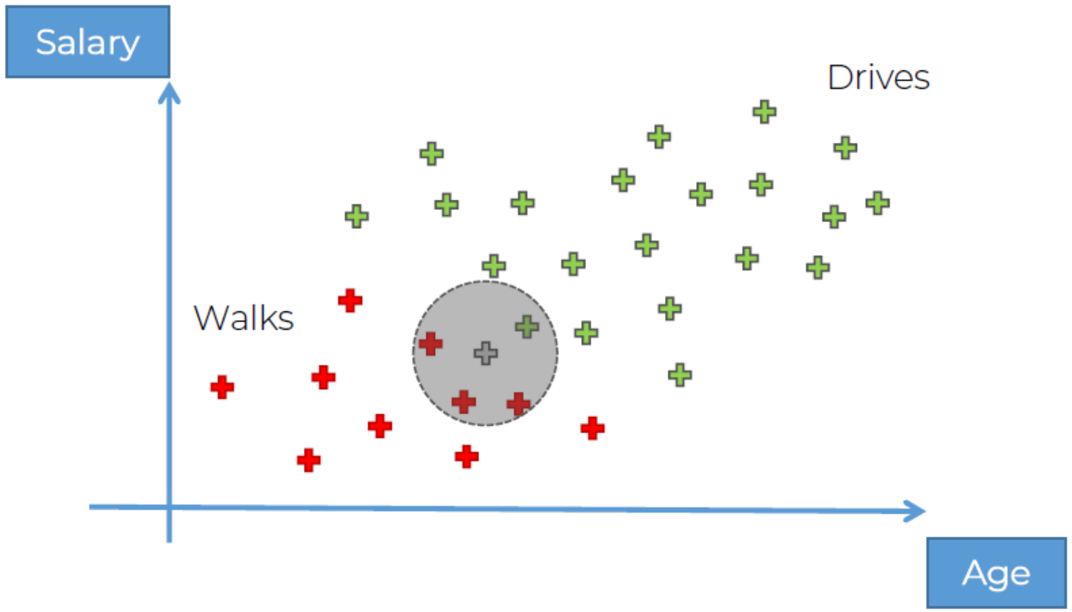
#3 Likelihood

#1 Prior Probability

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

#2 Marginal Likelihood

Naïve Bayes: Step 2



#3. P(X|Drives)

$$P(X|Drives) = \frac{\begin{array}{c}Number\ of\ Similar\\ Observations\\ Among\ those\ who\ Walk\end{array}}{Total\ number\ of\ Walkers}$$

$$P(X|Drives) = \frac{1}{20}$$

Naïve Bayes: Step 2

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

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

#2 Marginal Likelihood

Naïve Bayes: Step 2

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

$$P(Drives|X) = \frac{\frac{1}{20} * \frac{20}{30}}{\frac{4}{30}} = 0.25$$

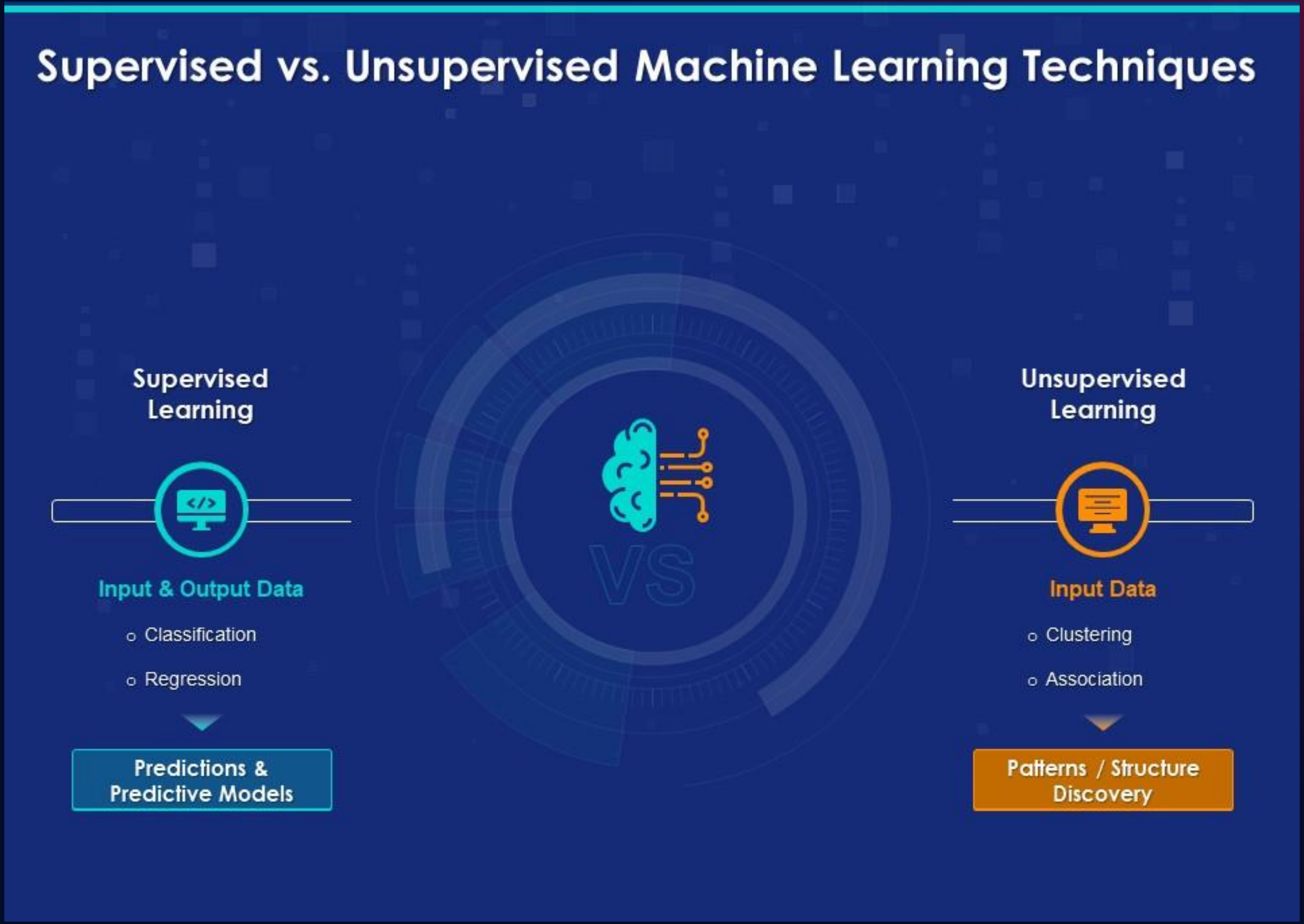
#2 Marginal Likelihood

Naïve Bayes: Step 3

$$P(Walks|X) \; v.s. \; P(Drives|X)$$

Naïve Bayes: Step 3

$$0.75 \; v.s. \; 0.25$$


Supervised vs. Unsupervised Machine Learning Techniques
Supervised Learning
Input & Output Data
Classification
Regression
Predictions & Predictive Models
VS
Unsupervised Learning
Input Data
Clustering
Association
Patterns / Structure Discovery

Advantages of Supervised Learning

Advantages

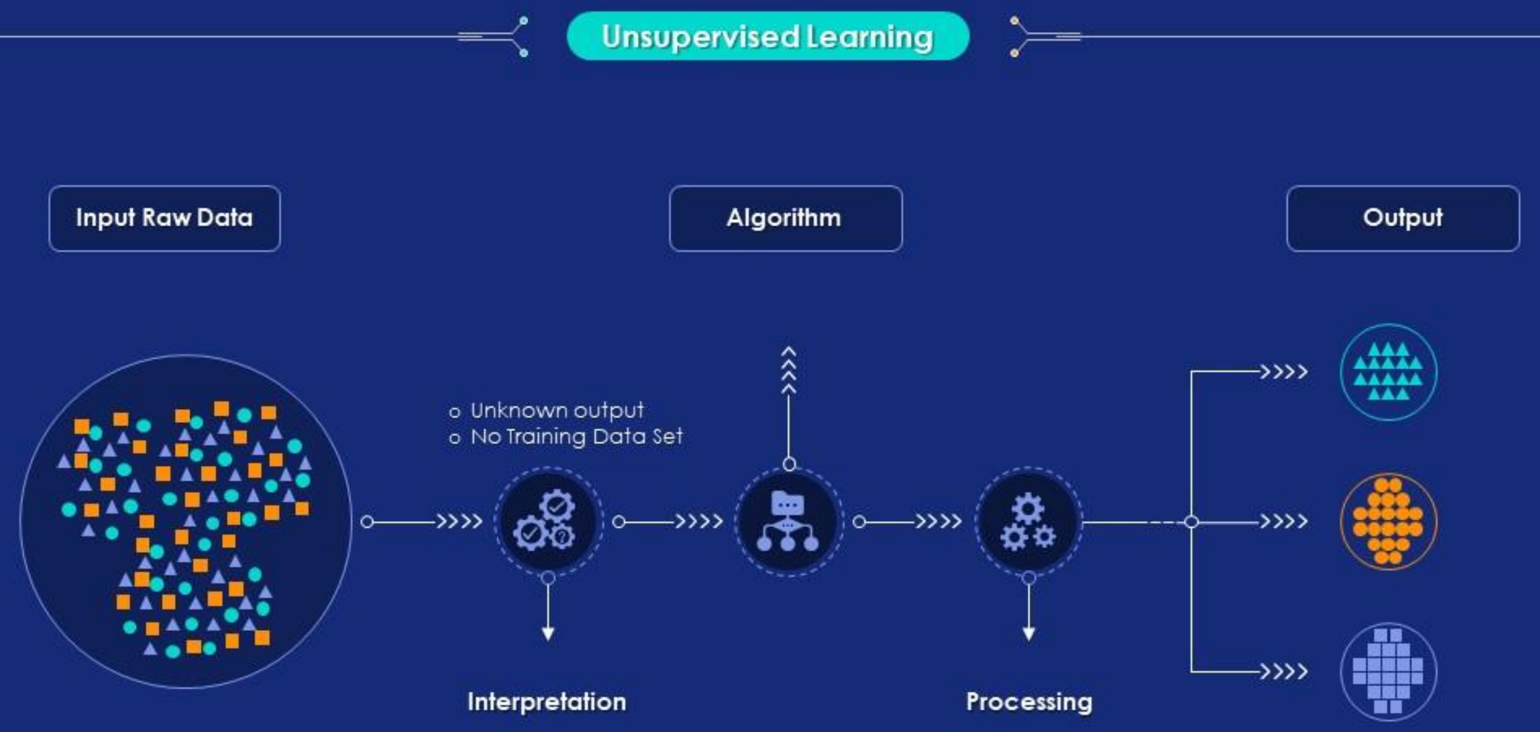
- It allows you to be very specific about the definition of the labels. In other words, you'll train the algorithm to differentiate different classes where you'll set a perfect decision boundary.
- You are ready to determine the amount of classes you would like to possess.
- The input file is extremely documented and is labeled.
- The results produced by the supervised method are more accurate and reliable as compared to the results produced by the unsupervised techniques of machine learning. this is often mainly because the input file within the supervised algorithm is documented and labeled. this is often a key difference between supervised and unsupervised learning.
- The answers within the analysis and therefore the output of your algorithm are likely to be known thanks to that each one the classes used are known.

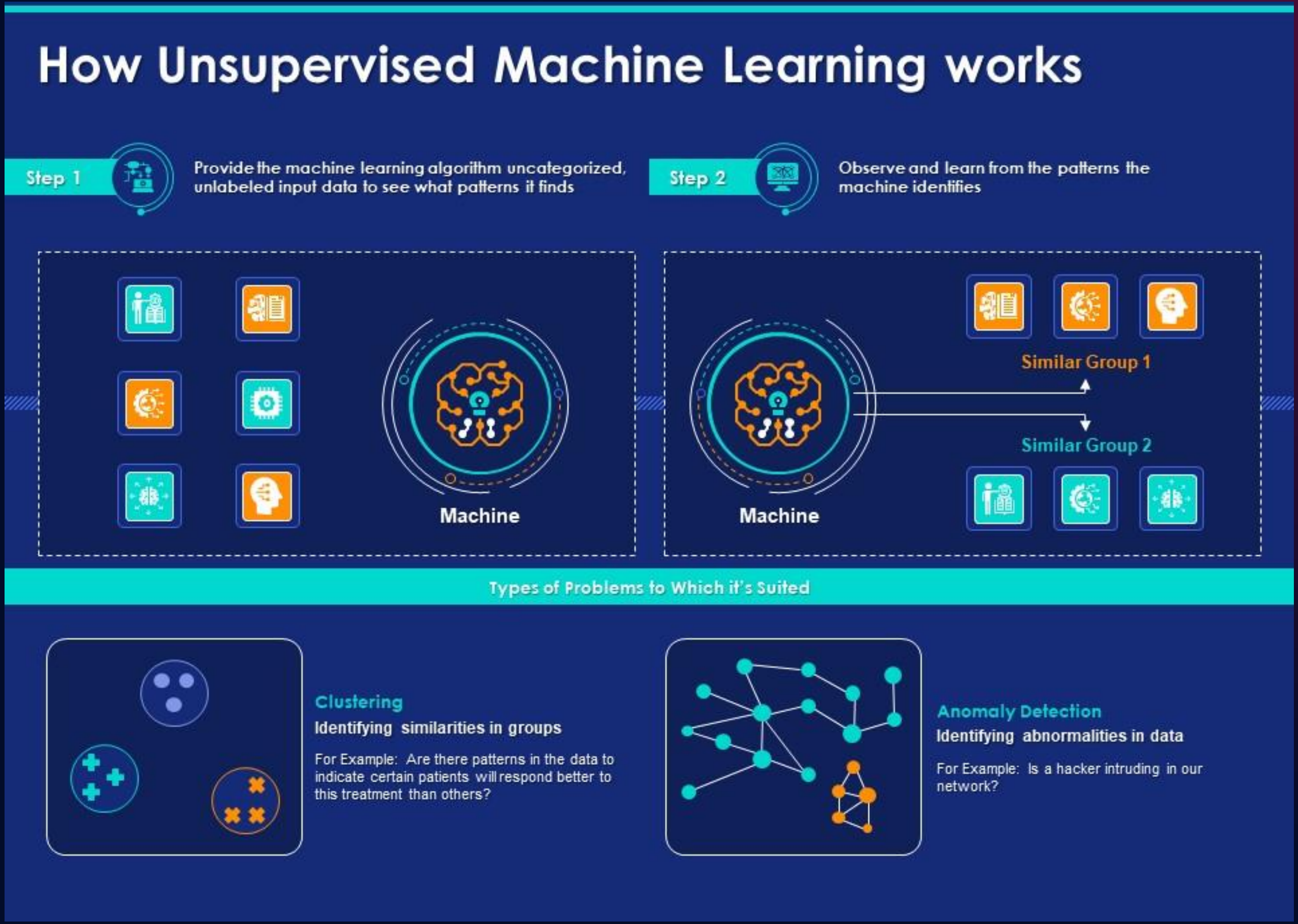
Disadvantages of Supervised Learning

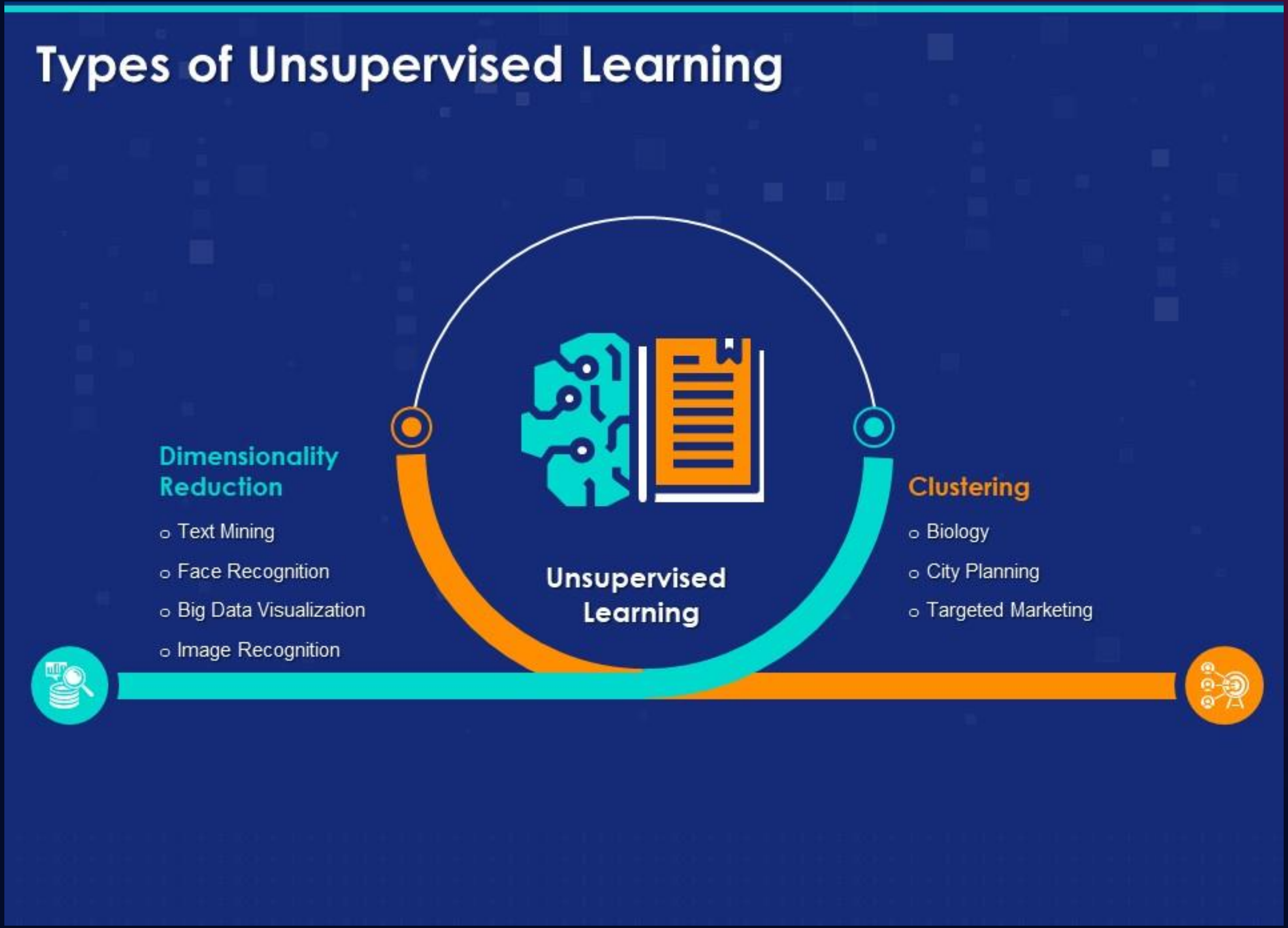
- It doesn't happen in real time while the unsupervised learning is about the important time. this is often also a serious difference between supervised and unsupervised learning. Supervised machine learning uses of-line analysis.
- It is needed tons of computation time for training.
- Supervised learning are often a posh method as compared with the unsupervised method. The key reason is that you simply need to understand alright and label the inputs in supervised learning.
- If you've got a dynamic big and growing data, you're unsure of the labels to predefine the principles. this will be a true challenge.

Disadvantages

What is Unsupervised Learning?




How Unsupervised Machine Learning works
Step 1
Provide the machine learning algorithm uncategorized, unlabeled input data to see what patterns it finds
Step 2
Observe and learn from the patterns the machine identifies
Machine
Similar Group 1
Similar Group 2
Machine
Types of Problems to Which it's Suited
Clustering
Identifying similarities in groups
For Example: Are there patterns in the data to indicate certain patients will respond better to this treatment than others?
Anomaly Detection
Identifying abnormalities in data
For Example: Is a hacker intruding in our network?


Types of Unsupervised Learning
Dimensionality Reduction
Text Mining
Face Recognition
Big Data Visualization
Image Recognition
Unsupervised Learning
Clustering
Biology
City Planning
Targeted Marketing

Disadvantages of Unsupervised Learning

You cannot get very specific about the definition of the info sorting and therefore the output. This is often because the info utilized in unsupervised learning is labeled and not known. It's employment of the machine to label and group the data before determining the hidden patterns.

Less accuracy of the results. This is often also because the input file isn't known and not labeled by people beforehand , which suggests that the machine will got to do that alone.

The results of the analysis can't be ascertained. there's no prior knowledge within the unsupervised method of machine learning. Additionally, the numbers of classes also are not known. It results in the lack to determine the results generated by the analysis.