Naive Bayes

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

What Is Bayes Therom About?

- P(A | B) → Probability of A happening given B is true.
- P(B | A) → Probability of B happening given A is true.
- P(A) → Prior probability of A happening.
- P(B) → Total probability of B happening.

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

Example: Diagnosing a Disease:

- P(Disease | Symptom) → Probability of having a disease given the symptom.
- P(Symptom | Disease) → Probability of the symptom if the person has the disease.

Why is Bayes' Theorem Important?

- Updates beliefs as new evidence comes in.
- Forms the foundation of Naïve Bayes Classifier for making predictions.
- Used in spam filtering, medical diagnosis, and machine learning.

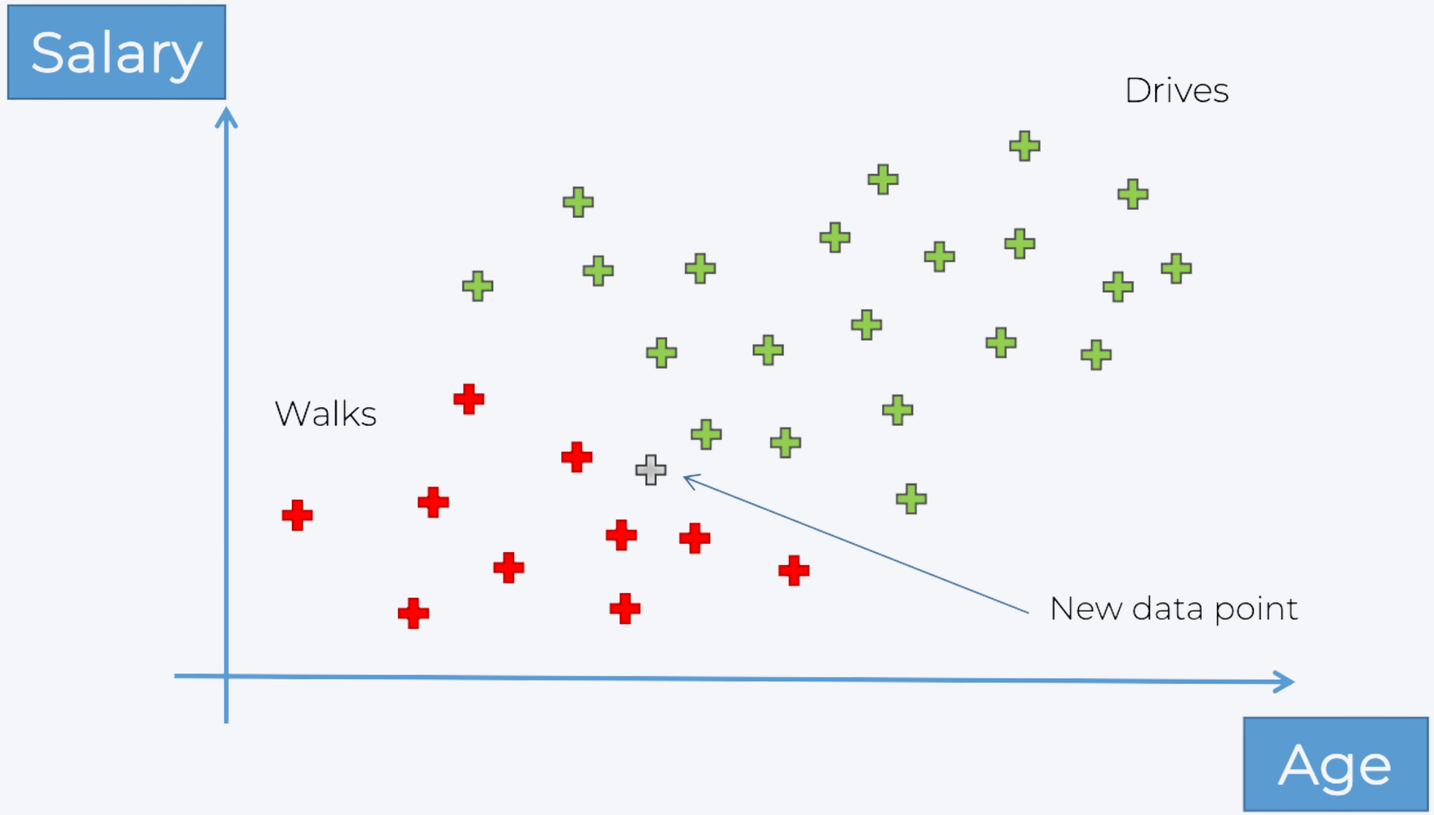
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

How does it work ?

When a new data point is introduced, we need to classify whether this new individual is more likely to walk or drive based on given data.

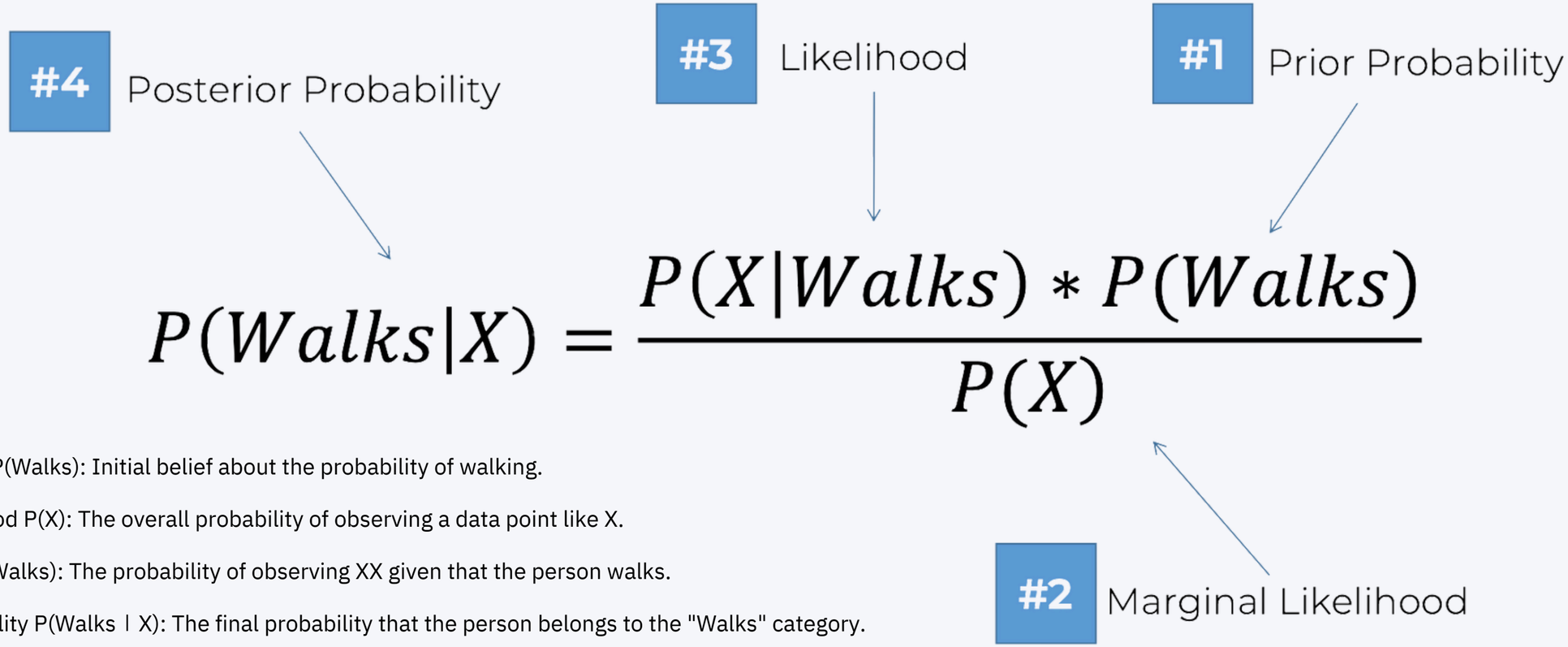


https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

Step #1



- Prior Probability P(Walks): Initial belief about the probability of walking.
- Marginal Likelihood P(X): The overall probability of observing a data point like X.
- Likelihood P(X | Walks): The probability of observing XX given that the person walks.
- Posterior Probability P(Walks | X): The final probability that the person belongs to the "Walks" category.

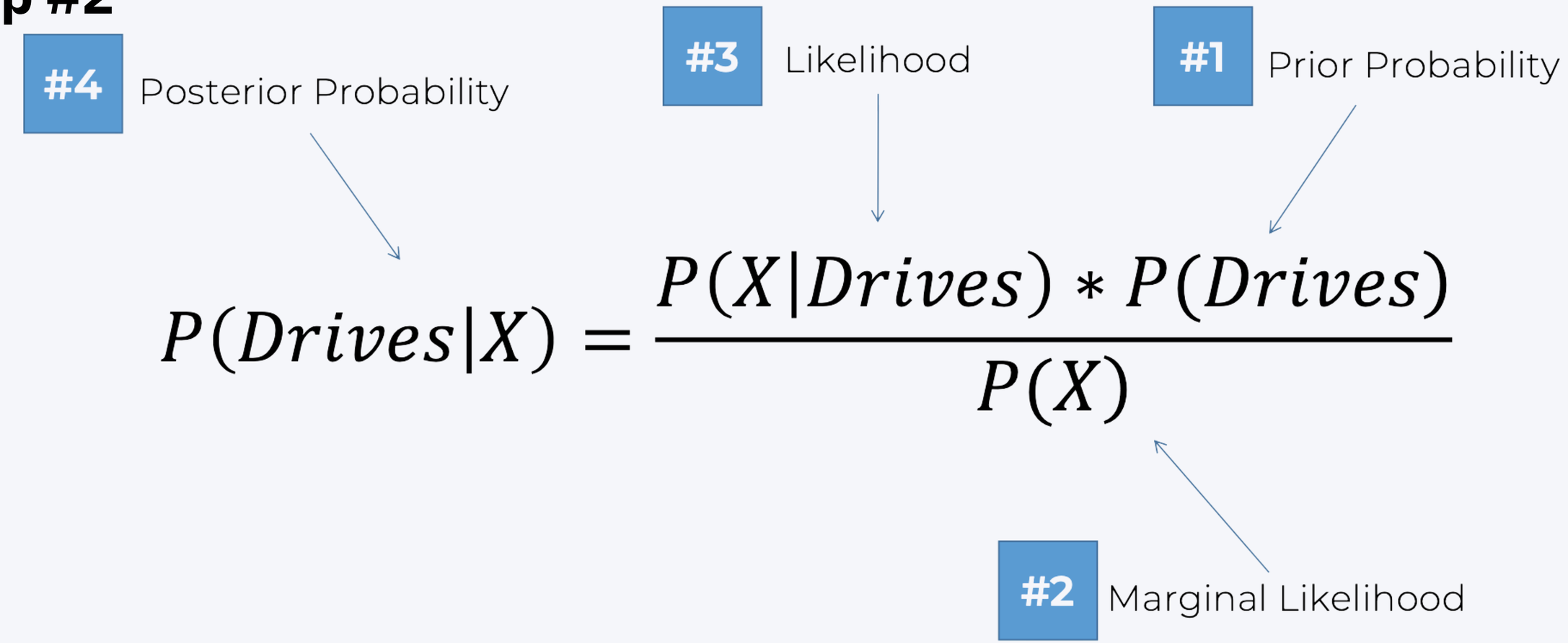
The probability of a person walking given their data point **P(Walks|X)** is computed using Bayes' Theorem.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

Step #2



Similarly, the probability of a person driving given their data point **P(Drives|X)** is computed using the same formula.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

Step #3

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

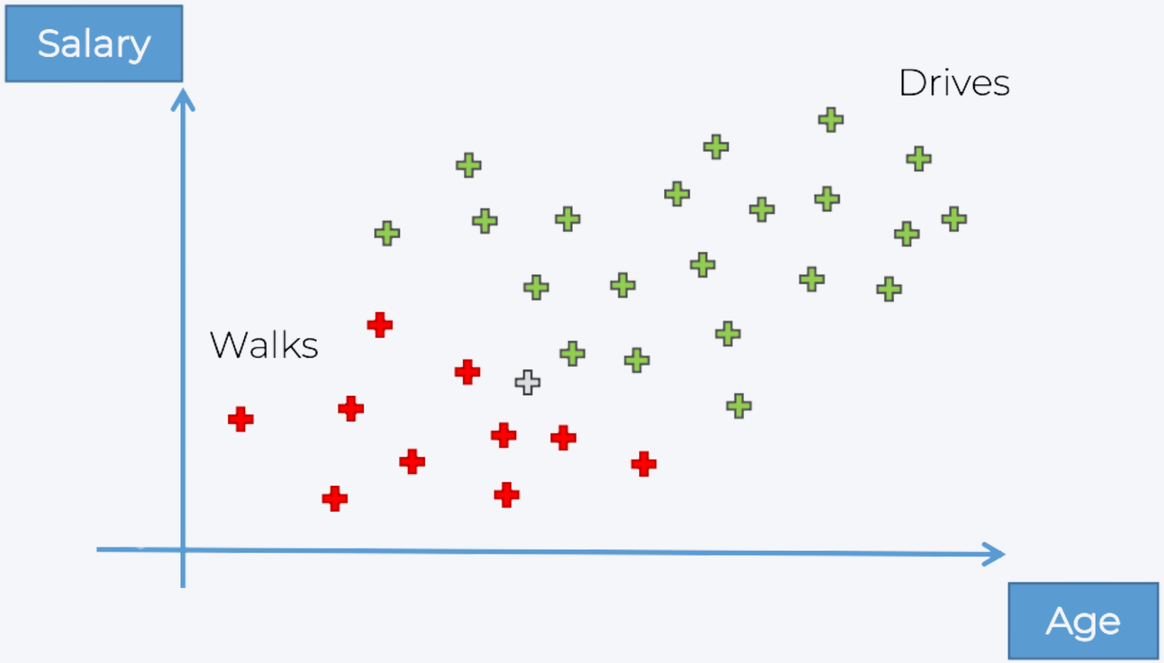
- The classification decision is made by comparing **P(Walks | X) vs. P(Drives | X).**
- The category with the higher probability is chosen.

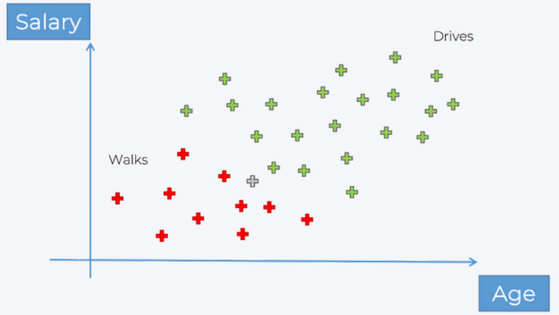
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

Naive bayes | In Action >> Step #1





#1. P(Walks)

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

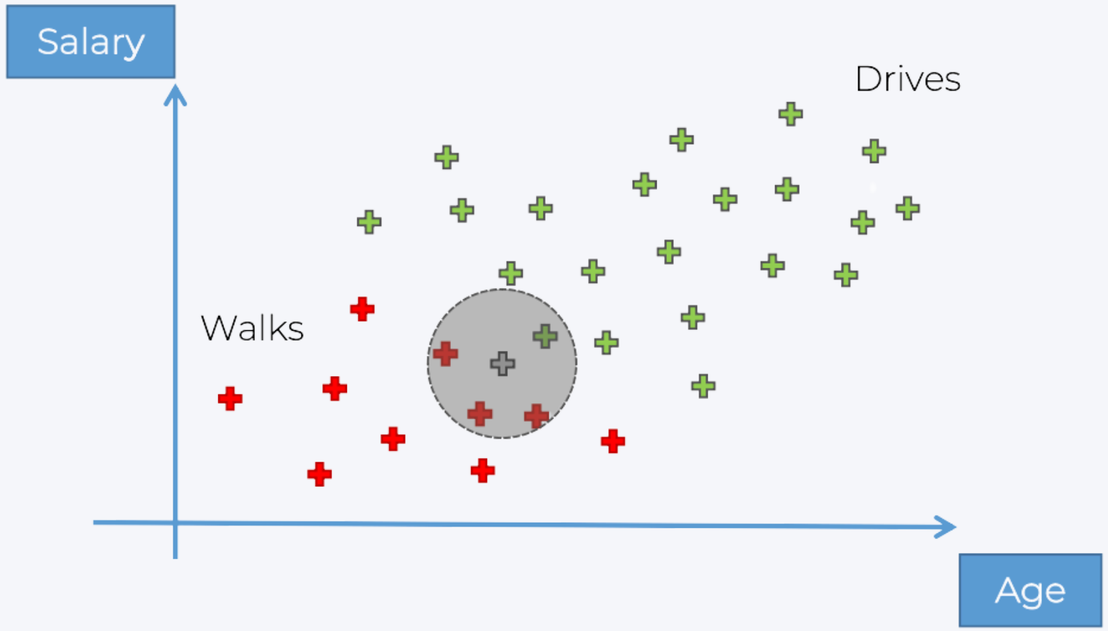
$$P(Walks) = \frac{10}{30}$$

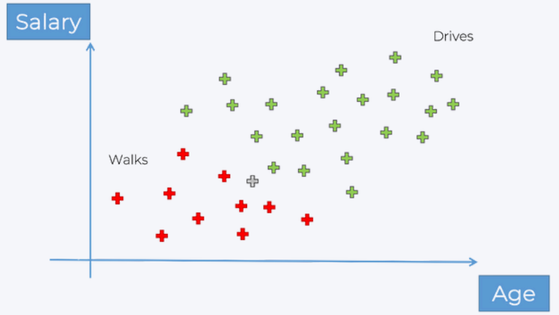
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

Naive bayes | In Action >> Step #1





#2. P(X)

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

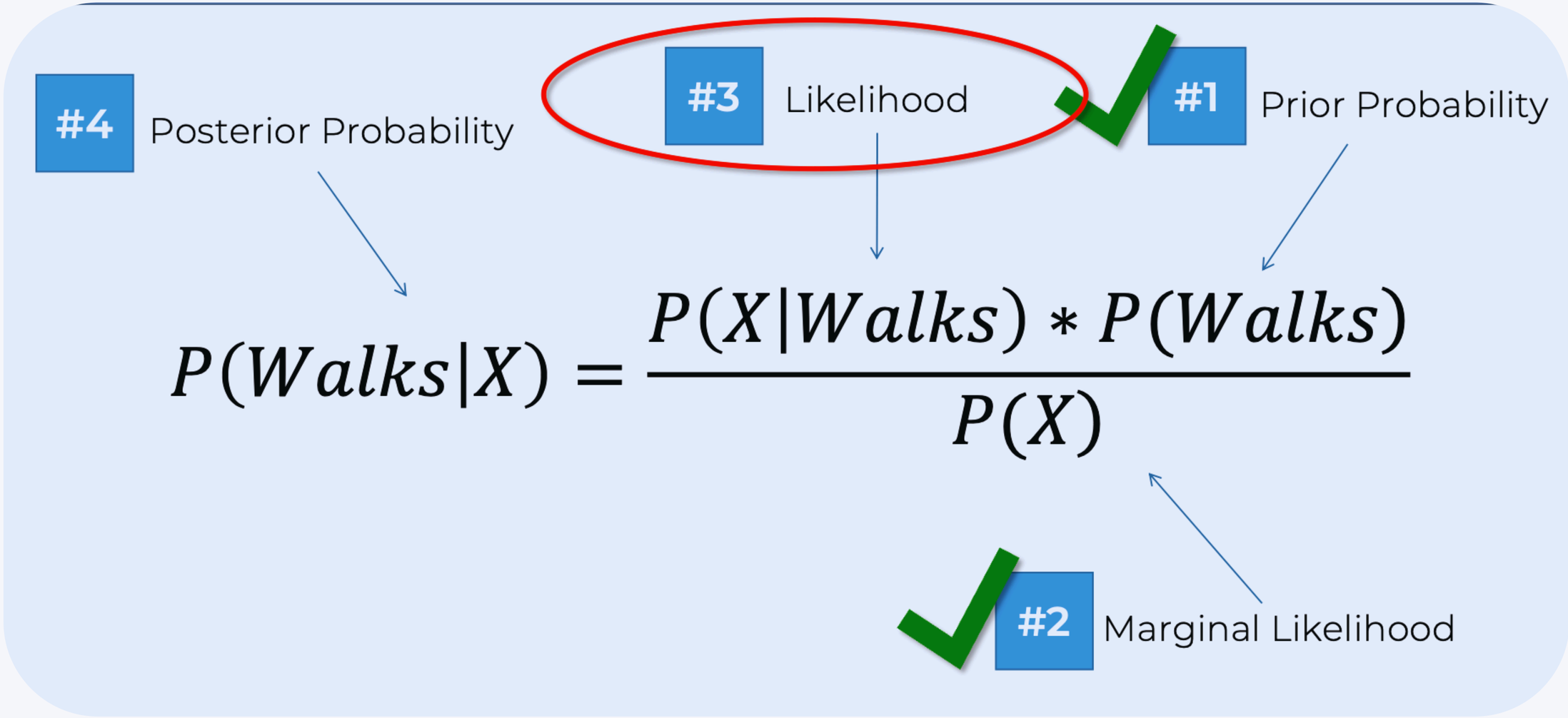
$$P(X) = \frac{4}{30}$$

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

Naive bayes | In Action >> Step #1

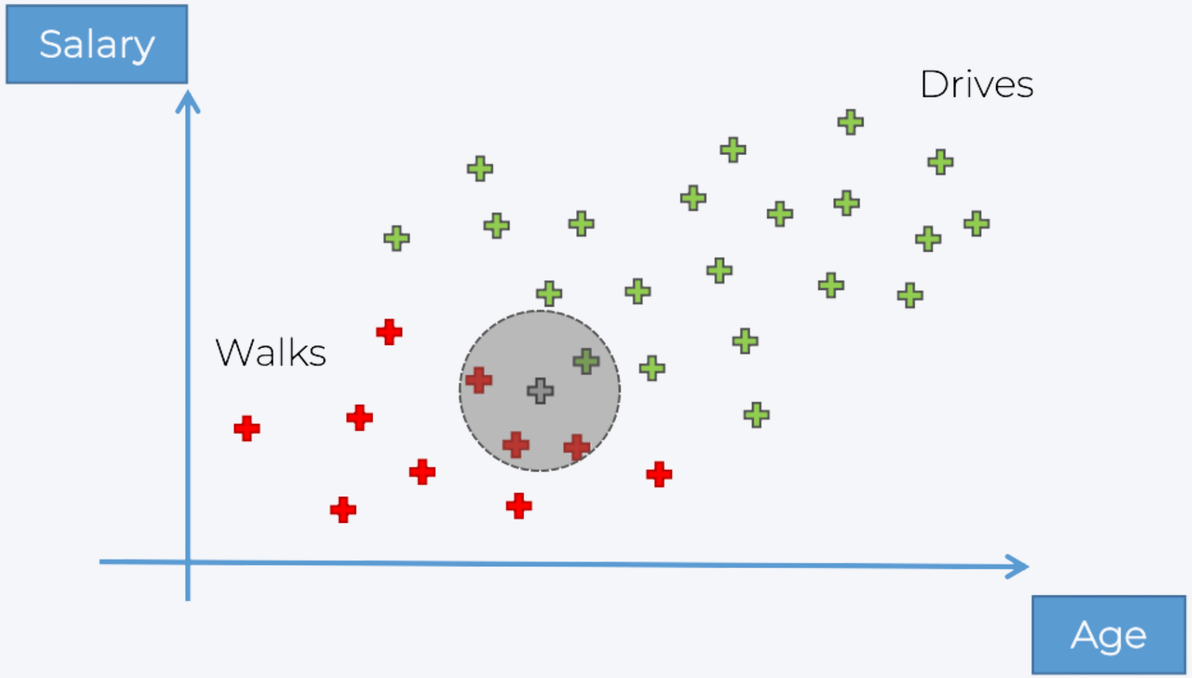


https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

Naive bayes | In Action >> Step #1



#3. P(X|Walks)

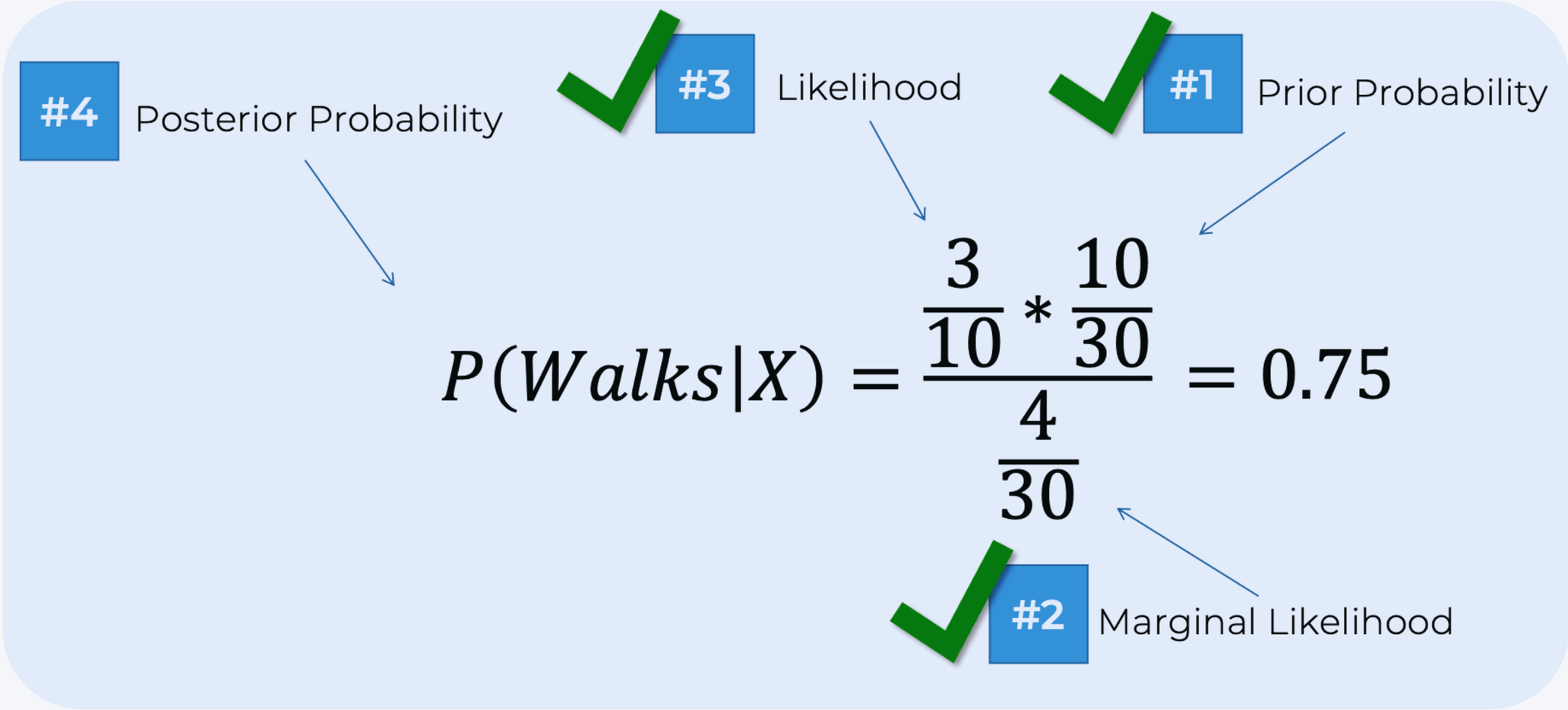
$$P(X|Walks) = \frac{Number\ of\ Similar\ Observations\ Among\ those\ who\ Walk}{Total\ number\ of\ Walkers}$$

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

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

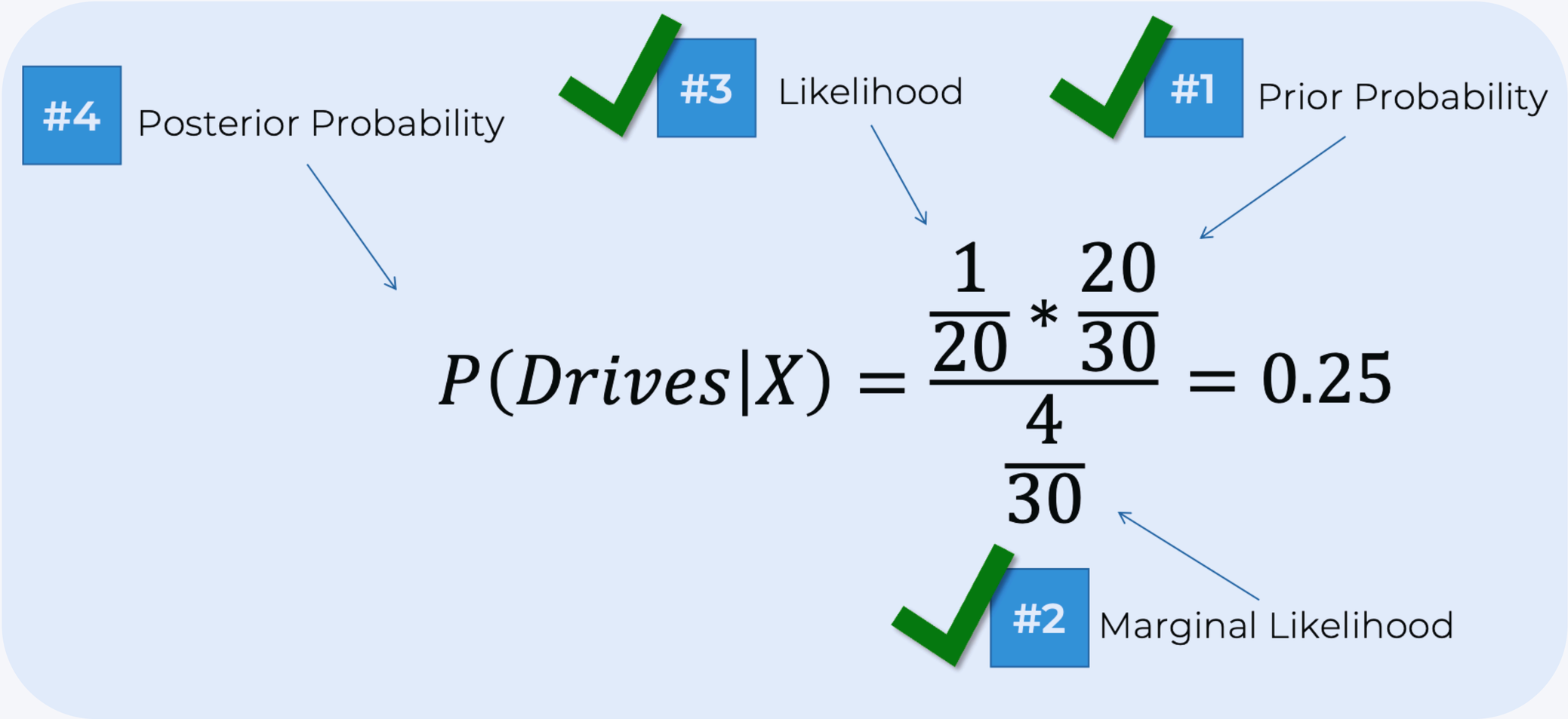
Naive bayes | In Action >> Step #1



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

Naive bayes | In Action >> Step #2



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

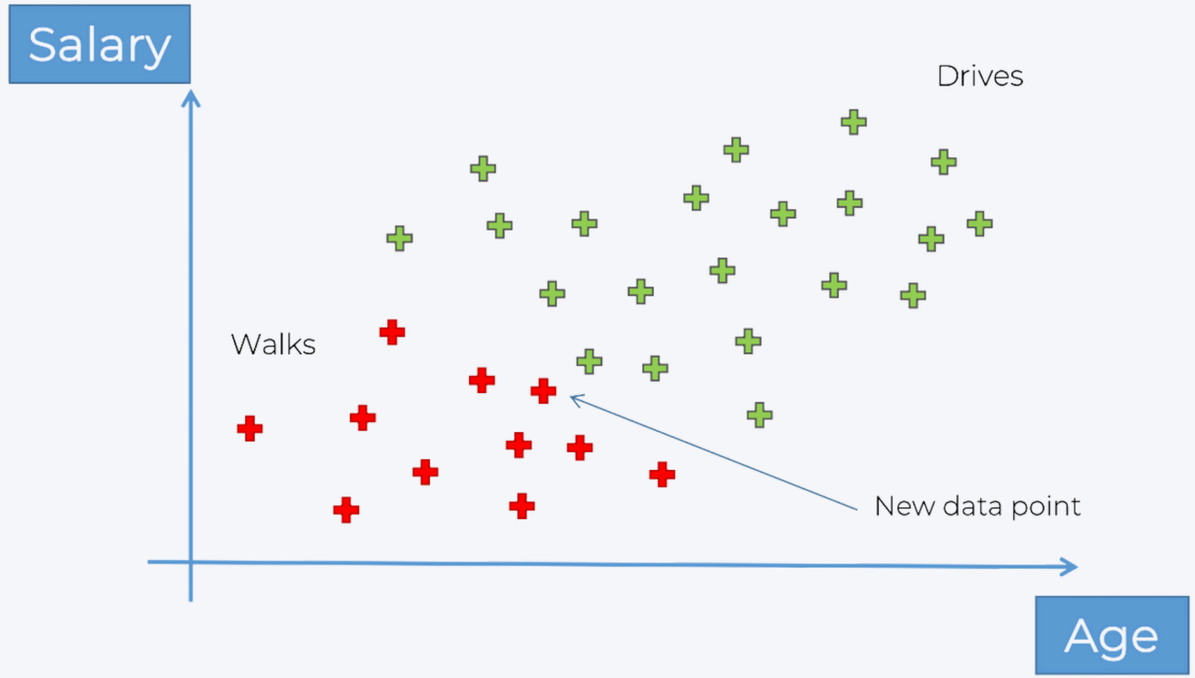
Naive bayes | In Action >> Step #2

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

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

$$0.75 > 0.25$$

$$P(Walks|X) > P(Drives|X)$$



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

Why is it called "Naive"?

Because it assumes independence between features, meaning each feature contributes to the classification independently.

Why Can We Ignore P(X)?

- Since P(X) is the same for all classes, it cancels out when comparing probabilities.

More than Two Classes?

- The method extends to multiple classes by computing the probability for each and picking the highest.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

أكاديمية طويق
Tuwaiq Academy

QUIZ

Naive Bayes

أكاديمية طويق
Tuwaiq Academy

Hands-On Code

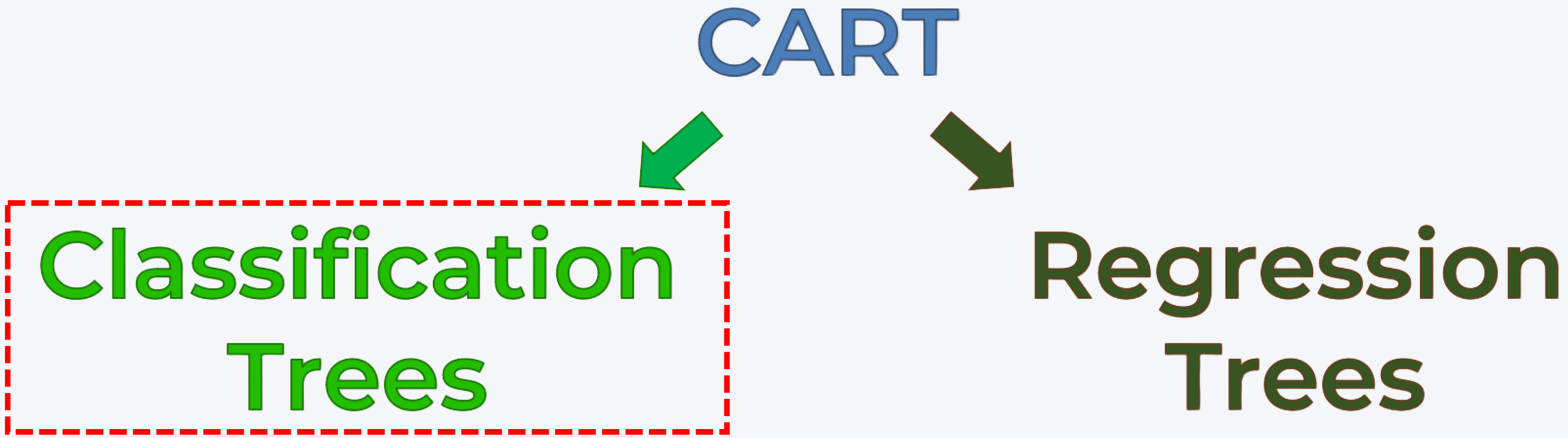
Naive Bayes

Decision Tree Classification

Decision Tree Classification

أكاديمية طويق
Tuwaiq Academy

CART



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Decision Tree Classification

أكاديمية طويق
Tuwaiq Academy

Difference Between Decision Tree Regression and Classification:

- **Decision Tree Regression:** Predicts a continuous numerical value based on input features. The tree is built using splits that minimize variance within groups.

- **Decision Tree Classification:** Assigns an input to a specific category (class) by recursively splitting the data based on features that best separate different classes.

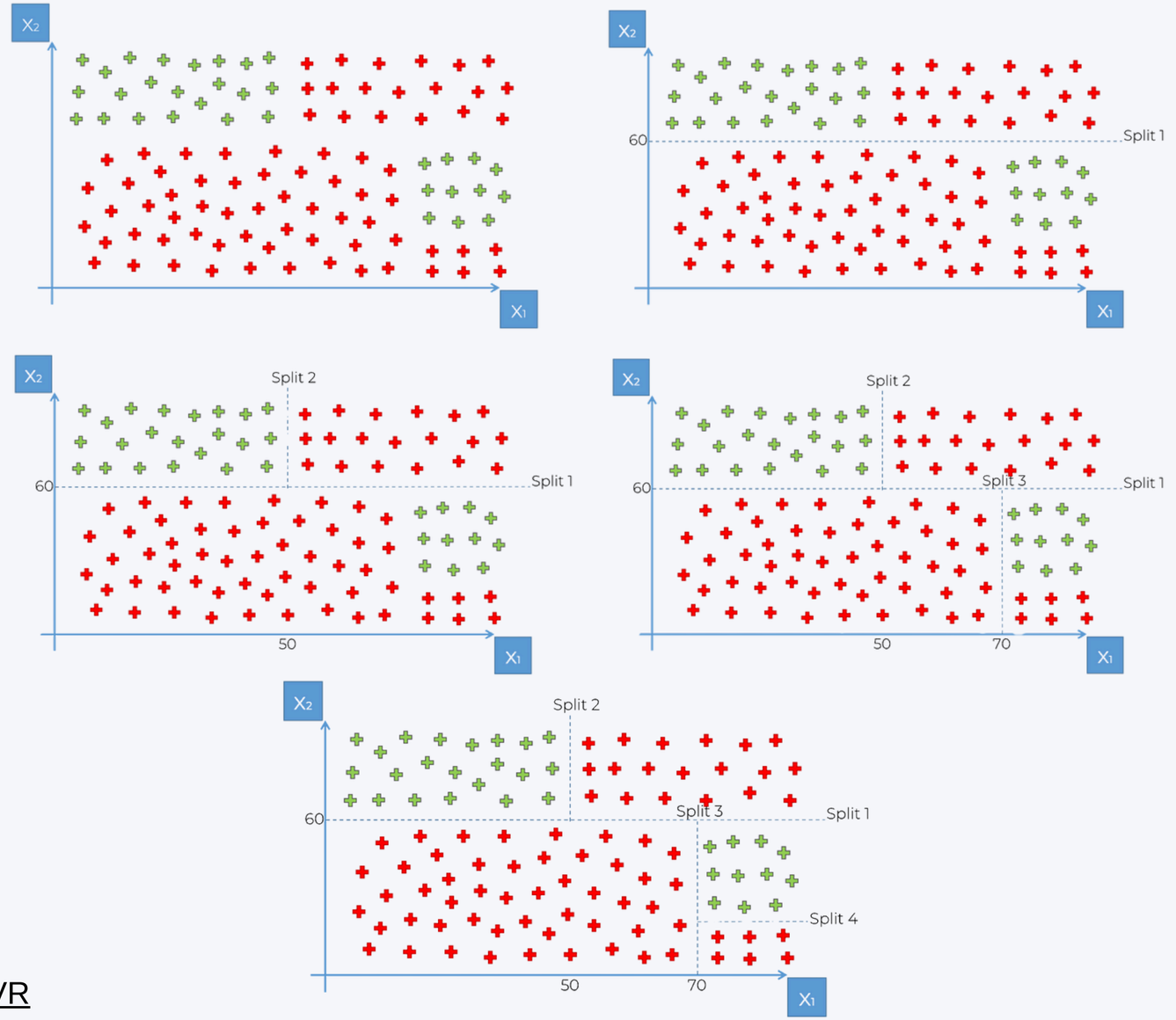
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Decision Tree Classification

أكاديمية طويق
Tuwaiq Academy

Explanation of Decision Tree Classification:

- The images illustrate how a decision tree classifier works by recursively splitting the data into smaller regions.
- Each split is made based on a feature that best separates the classes (red and green points).
- The first split (Split 1) separates based on $X_2 > 60$, dividing the data into two main groups.
- The second split (Split 2) further divides one of these groups based on $X_1 < 50$.
- Additional splits (Split 3, Split 4) continue until pure or nearly pure regions are formed.



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Decision Tree Classification

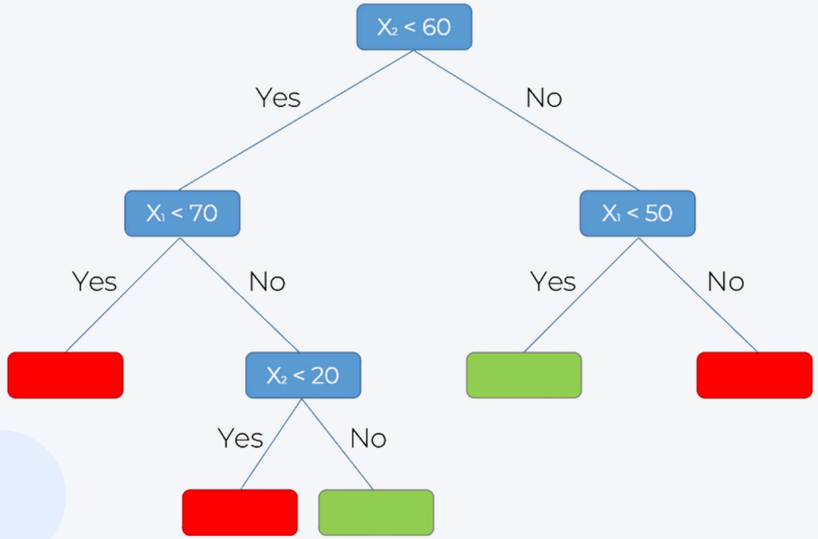
أكاديمية طويق
Tuwaiq Academy

Explanation of Decision Tree Classification:

The tree chooses splits based on feature values that maximize class separation. This is done using:

- **Gini Impurity (default in CART trees):**
  - Measures how mixed a group is.
  - A lower Gini value means a purer split.

$$Gini = 1 - \sum P(i)^2$$



- **Entropy & Information Gain**
  - Entropy measures uncertainty in data:

$$Entropy = -\sum P(i) \log_2 P(i)$$

- **Information Gain calculates how much uncertainty is reduced by a split:**
  - The split that maximizes Information Gain is chosen.

$$IG = Entropy_{parent} - \sum \left(\frac{|Subset|}{|Parent|} \times Entropy_{subset} \right)$$

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

أكاديمية طويق
Tuwaiq Academy

Hands-On Code

Decision Tree Classification

Random Forest Classification

Random Forest Classification

أكاديمية طويق
Tuwaiq Academy

How does Random Forest do Classification?

Ensemble Learning

- **STEP 1:** Pick at random **K** data points from the Training set.
- **STEP 2:** Build the Decision Tree associated to these **K** data points.
- **STEP 3:** Choose the number **Ntree** of trees you want to build and repeat STEPS 1 & 2
- **STEP 4:** For a new data point, make each one of your **Ntree** trees predict the category to which the data points belongs, and assign the new data point to the category that wins the majority vote.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

أكاديمية طويق
Tuwaiq Academy

Hands-On Code

Random Forest Classification