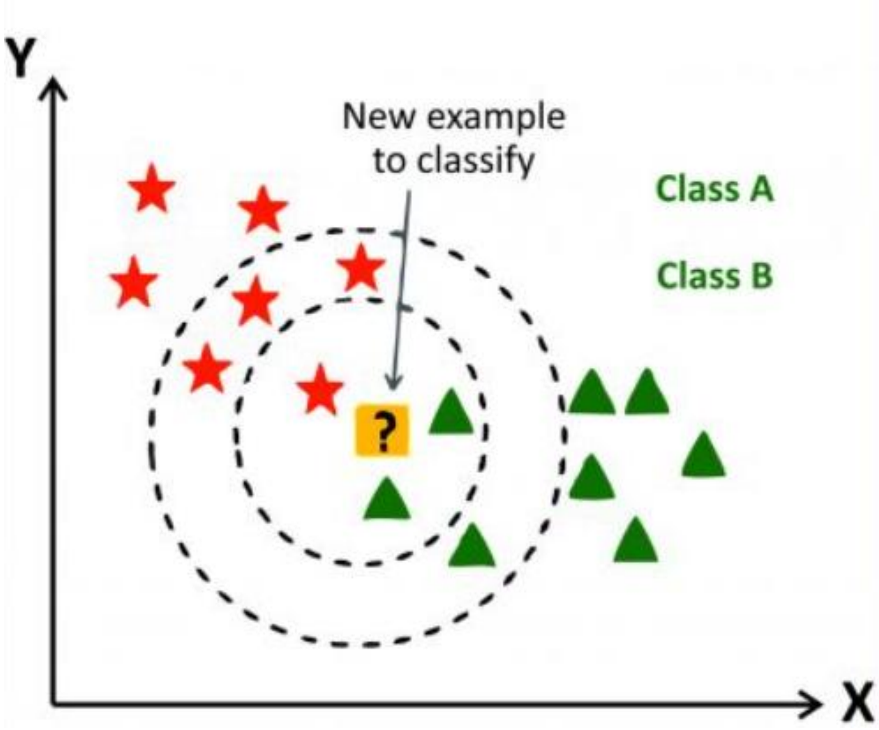
KNN



KNN Basic Concept

- **Supervised learning algorithm** used for classification.
- Classifies based on closest neighbors
- Similarity measured using distance metrics
- Common metrics: Euclidean, Manhattan
- Decision depends on majority class vote
- Simple, intuitive, but computationally expensive

Why KNN is Lazy Learner

- It doesn’t learn a model during training.
- It stores all training data as-is.
- Computation happens at prediction time, not before.
- It waits (is “lazy”) until a query is made.
- This makes it simple but slower for large datasets.
- It trades no training time for high prediction time.
- So, it’s called lazy because it **learns only when needed**.

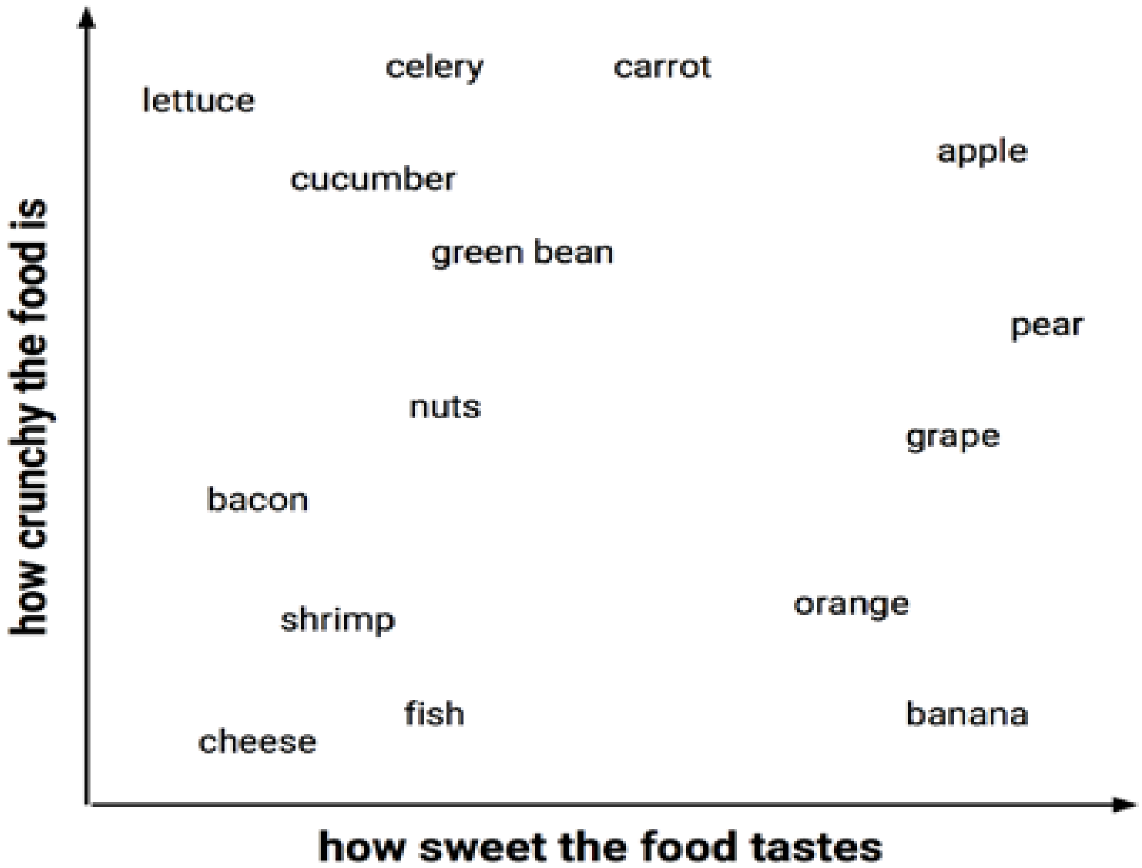
Steps in KNN Algorithm

•**Select the value of K** (number of neighbors).
•**Compute the distance** between the query point and all training points.
•**Find the K nearest neighbors** to the query point.
•**Count the class labels** of these nearest neighbors.
•**Assign the majority class** to the query point.

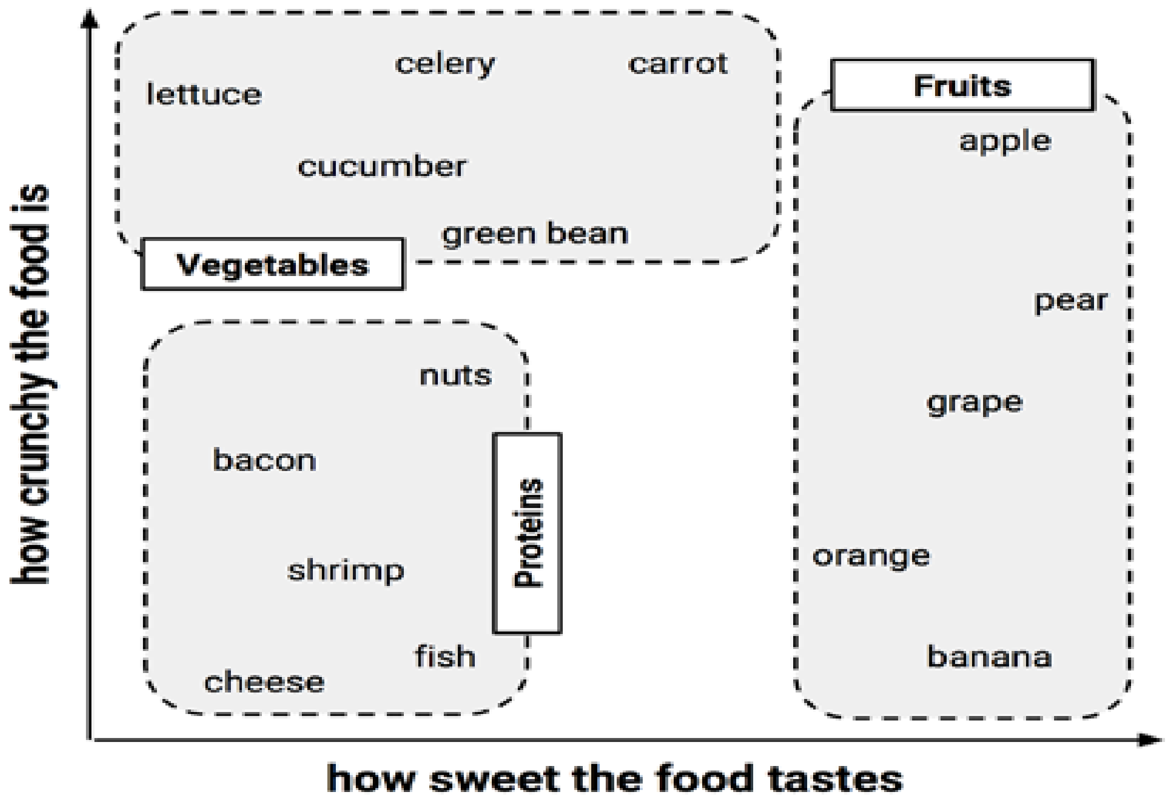
Example

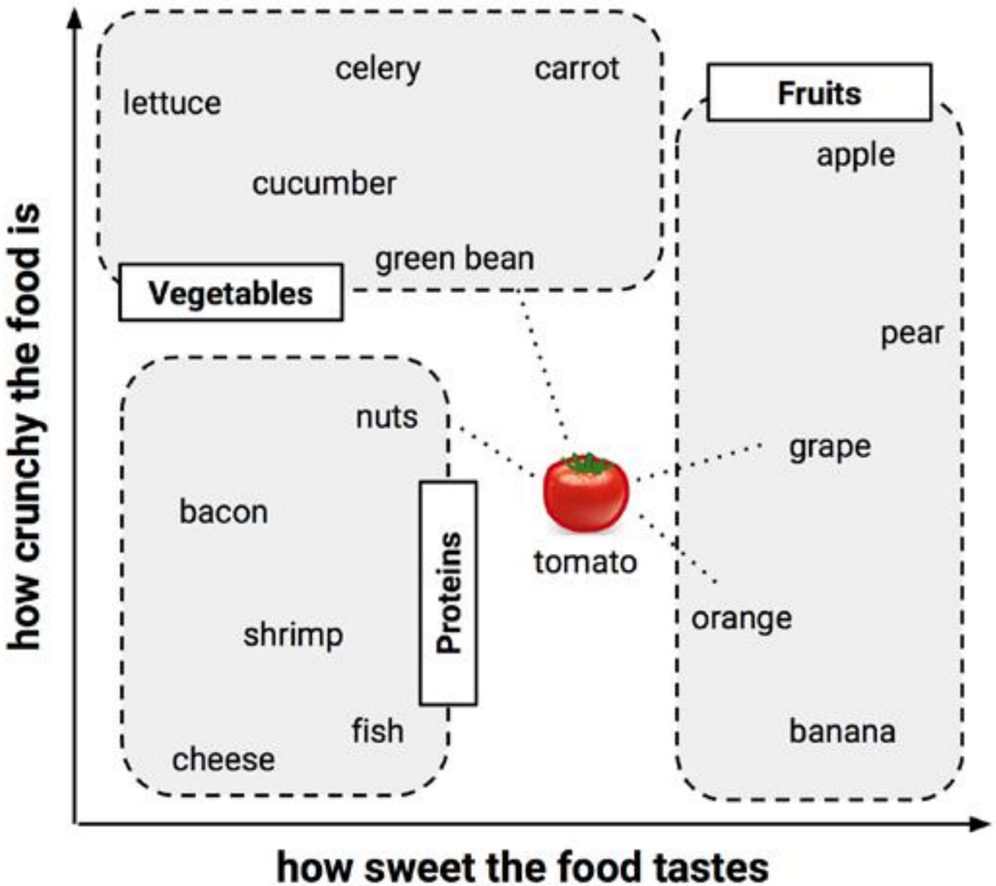
Ingredient	Sweetness	Crunchiness	Food type
apple	10	9	fruit
bacon	1	4	protein
banana	10	1	fruit
carrot	7	10	vegetable
celery	3	10	vegetable
cheese	1	1	protein

Example



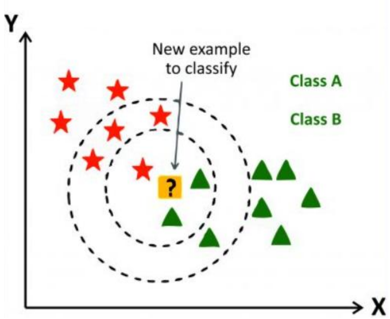
celery
carrot
lettuce
apple
cucumber
green bean
pear
nuts
grape
bacon
orange
shrimp
fish
banana
cheese
how crunchy the food is
how sweet the food tastes


celery
carrot
lettuce
cucumber
green bean
Vegetables
Fruits
apple
pear
grape
orange
banana
nuts
bacon
shrimp
fish
cheese
Proteins
how crunchy the food is
how sweet the food tastes



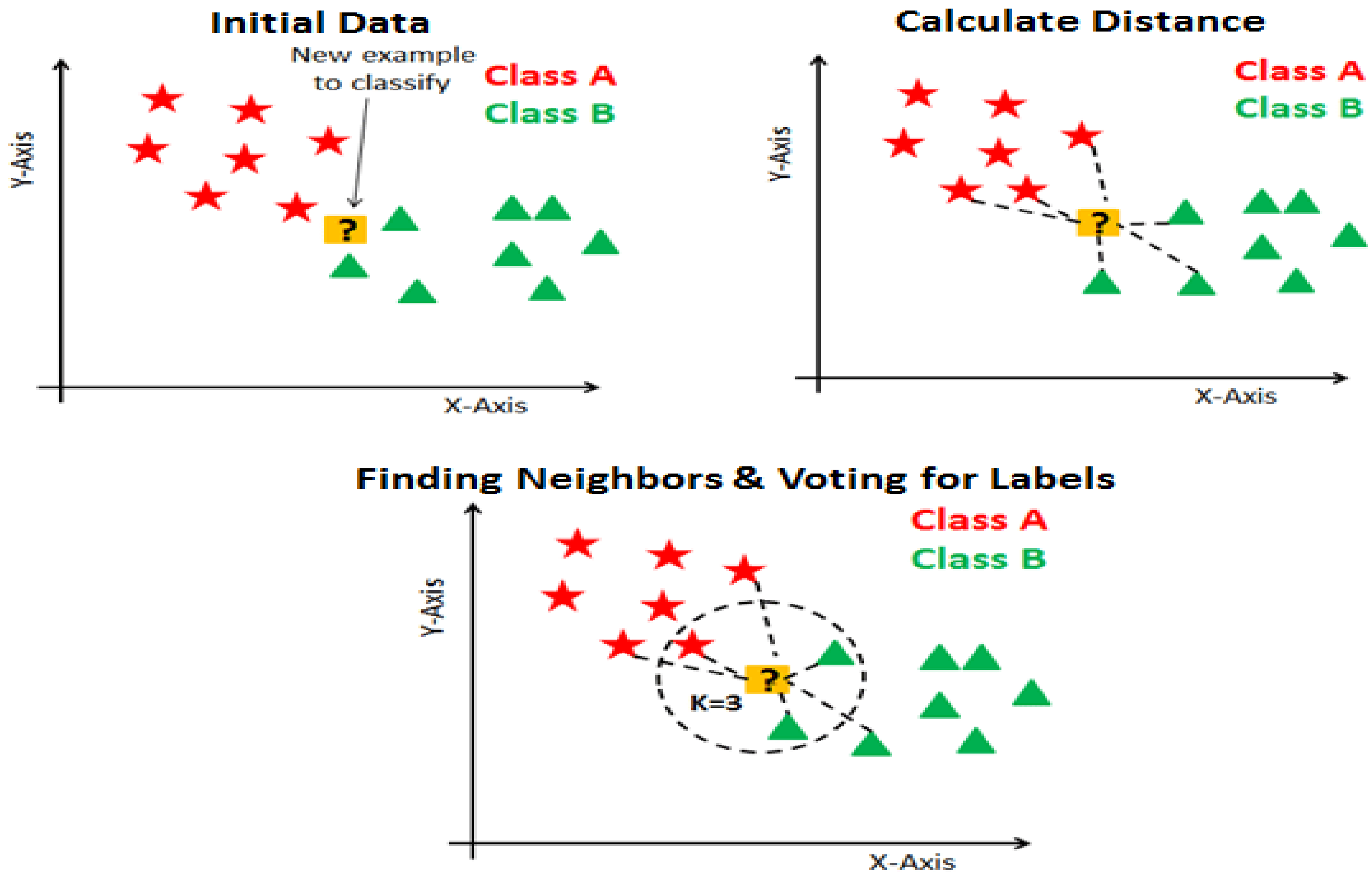
Vegetables
lettuce
celery
carrot
cucumber
green bean
Fruits
apple
pear
grape
orange
banana
Proteins
nuts
bacon
shrimp
fish
cheese
tomato
how crunchy the food is
how sweet the food tastes

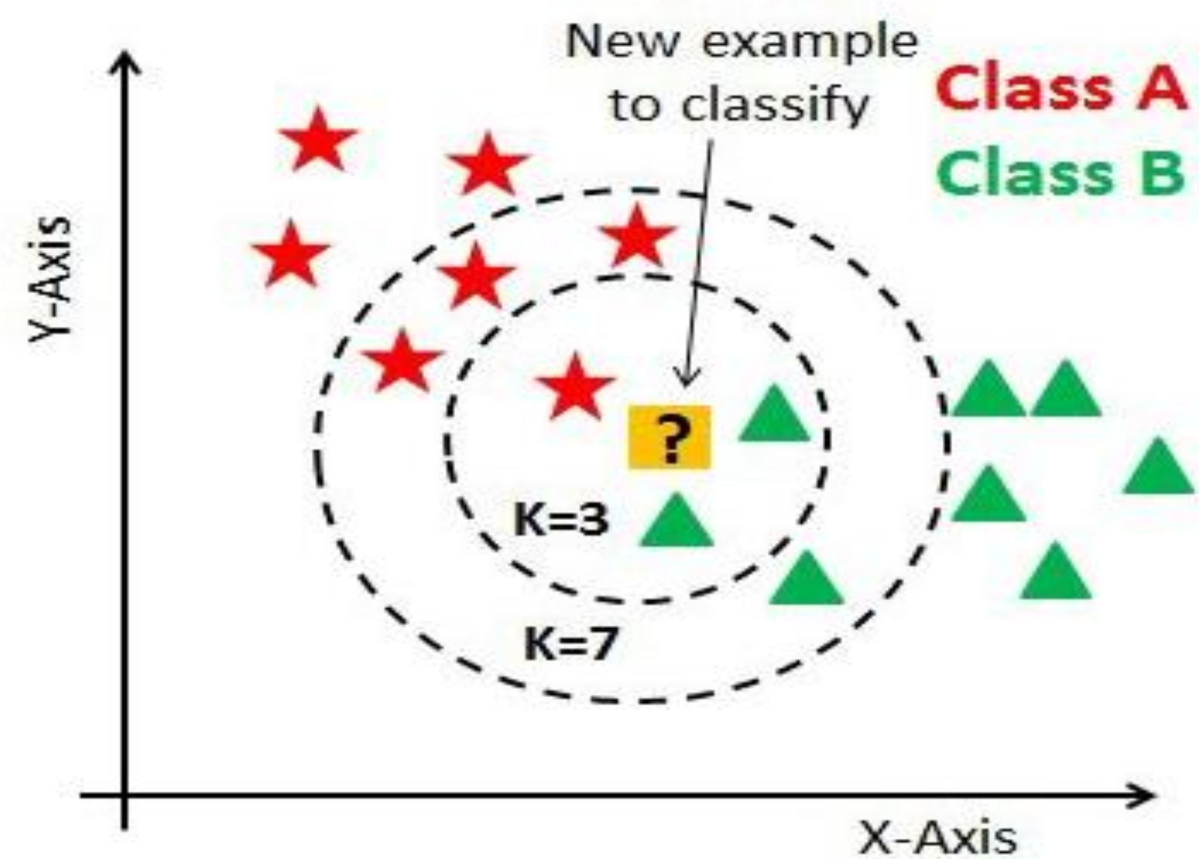
$$\text{dist}(\text{tomato, green bean}) = \sqrt{(6-3)^2 + (4-7)^2} = 4.2$$



Ingredient	Sweetness	Crunchiness	Food type	Distance to the tomato
grape	8	5	fruit	*sqrt((6 - 8)^2 + (4 - 5)^2) = 2.2*
green bean	3	7	vegetable	*sqrt((6 - 3)^2 + (4 - 7)^2) = 4.2*
nuts	3	6	protein	*sqrt((6 - 3)^2 + (4 - 6)^2) = 3.6*
orange	7	3	fruit	*sqrt((6 - 7)^2 + (4 - 3)^2) = 1.4*

Choosing an appropriate k




New example
to classify
Class A
Class B
Y-Axis
?
K=3
K=7
X-Axis

Choosing Value of K

- Small K → Model becomes sensitive to noise.
- Large K → Produces a smoother decision boundary.
- Odd K → Helps avoid ties in classification.
- Cross-validation → Used to find the best K value.
- Feature scaling → Essential for accurate distance measurement.

Advantages and Disadvantages of KNN

Advantages:

- **Handles multi-class problems** effectively.
- Works well with **nonlinear and complex decision boundaries**.

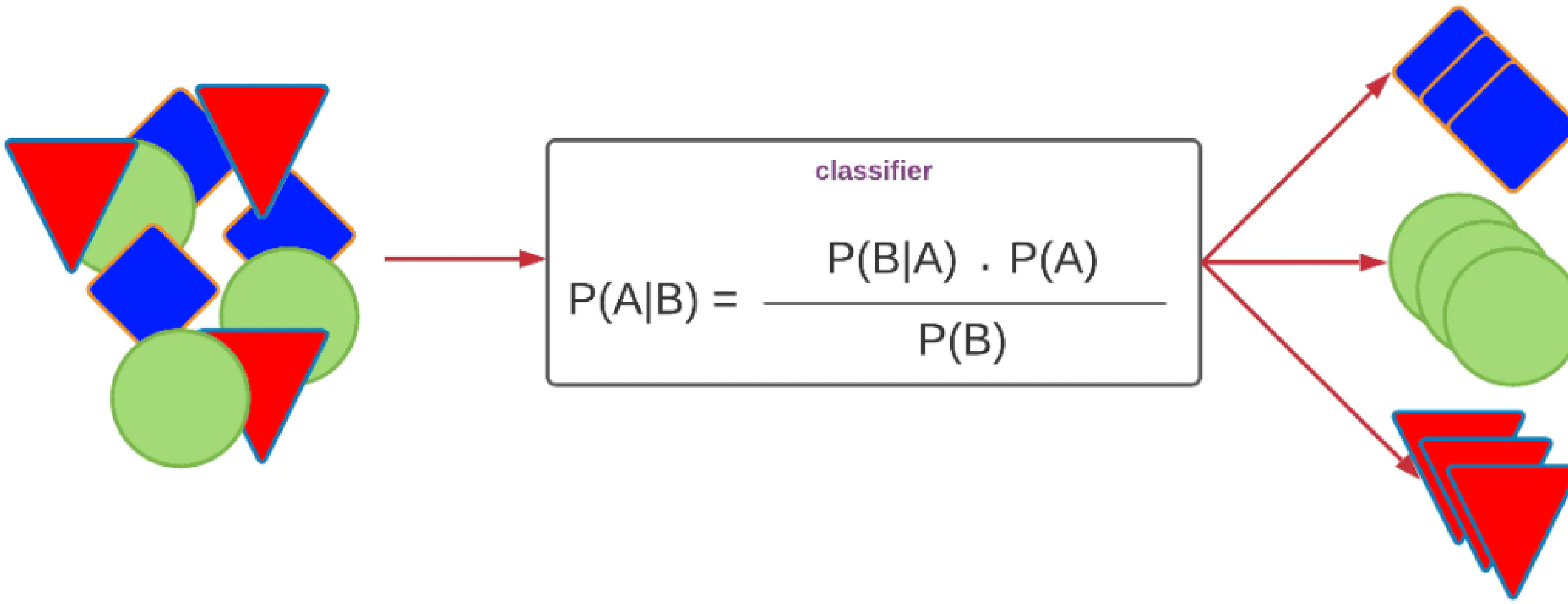
Disadvantages:

- **High computation and memory cost** during prediction.
- **Sensitive to irrelevant or noisy features.**
- **Requires normalization** for accurate distance calculation.

Introduction to Naive Bayes

- Based on Baye’s Theorem principles
- Assumes independence among predictor features
- Calculates probability of each class
- Chooses class with maximum posterior probability
- Works well for categorical input variables
- Widely used in text classification problems

Naive Bayes



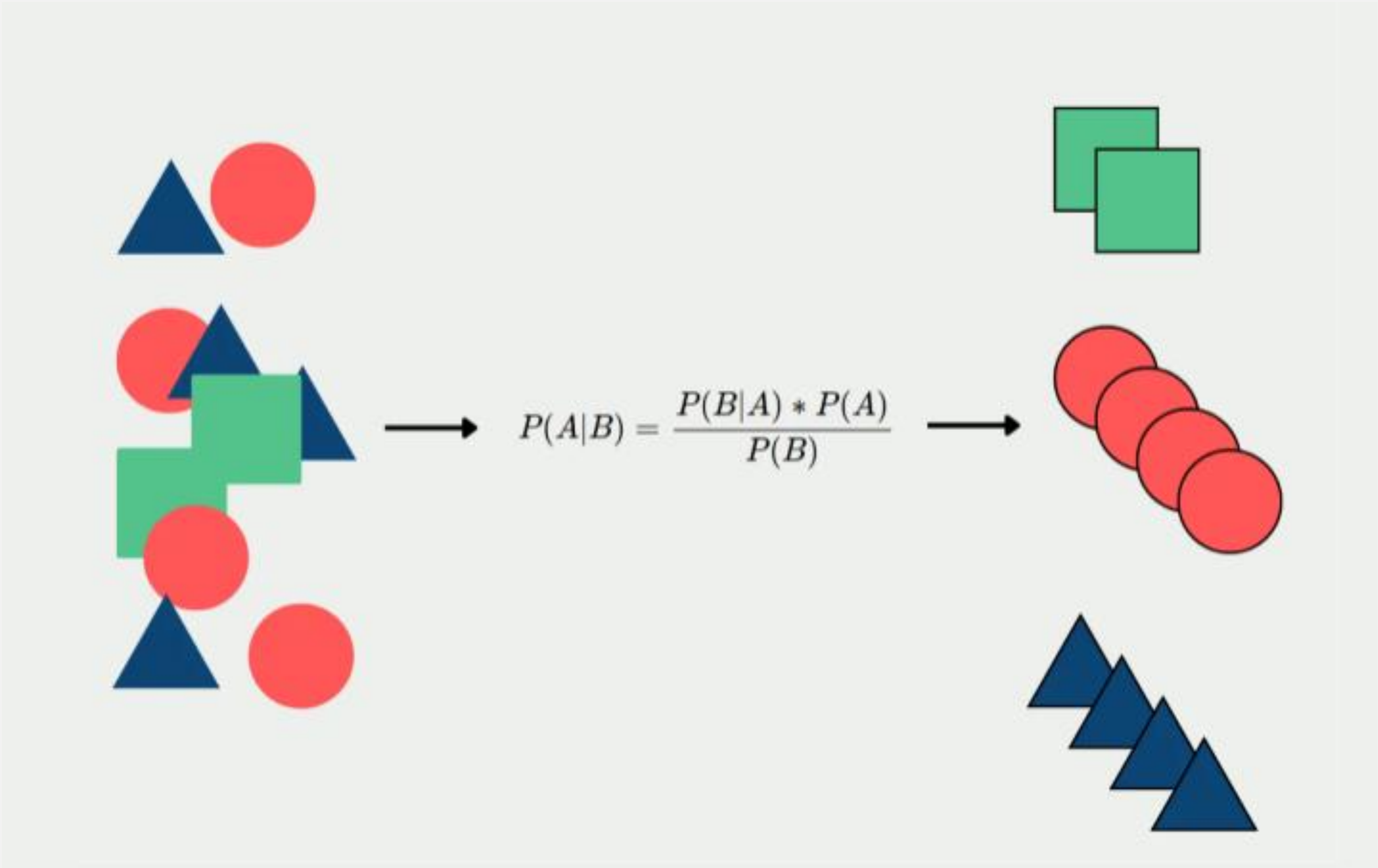
Baye’s Theorem

- Posterior = Likelihood × Prior / Evidence
- Prior: initial probability of class
- Likelihood: probability of evidence given class
- Evidence: probability of observing feature
- Posterior: updated probability after evidence
- Class with highest posterior selected

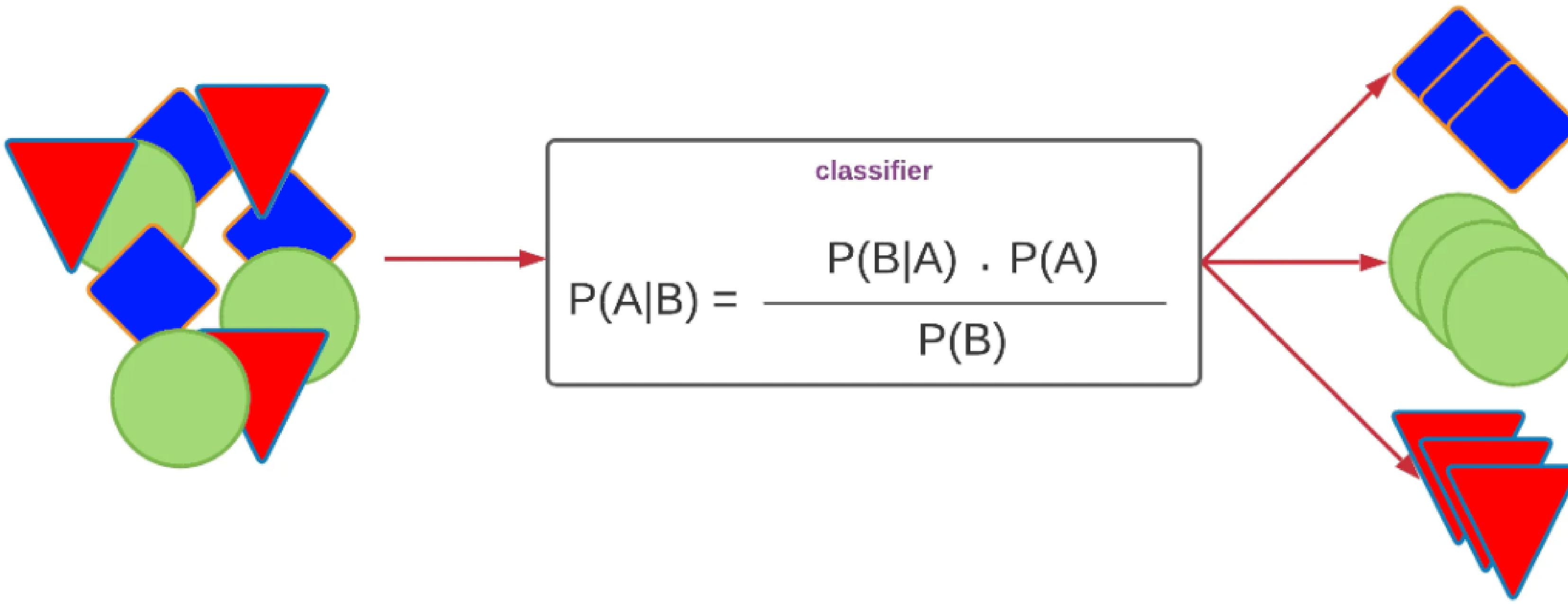
Naive Bayes Working Steps

- Calculate prior probability for each class
- Compute likelihood for features per class
- Multiply probabilities using independence assumption
- Normalize values using Baye’s formula
- Select class with highest posterior probability
- Repeat for every new unseen data

Simple Representation of the Naive Bayes Classification



Formula



Bayes Theorem

Conditional prob: given condition + event & we have to find prob.

Bayes → given event : choice is asked for prob.

Eg

Bag 1	Bag 2
3 R, 4 B	5 R, 6 B

One ball drawn and it is Red. Find the prob. if it is drawn from bag 2.

$$\text{Prob} = \frac{\text{fav. outcome}}{\text{possible outcome}}$$

Prob. of Bag 1 choose $= \frac{1}{2}$ — $P(E_1)$

Bag 2 choose $= \frac{1}{2}$ — $P(E_2)$

Red $= \frac{3}{7}$ — $P(A|E_1)$

Red $= \frac{5}{11}$ — $P(A|E_2)$

Prob of red ball from bag 2 $= \frac{fo}{po}$

$$= \frac{\frac{1}{2} \times \frac{5}{11}}{\frac{1}{2} \times \frac{3}{7} + \frac{1}{2} \times \frac{5}{11}}$$

posterior probability ~~P(A/E₁)~~

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

likelihood

prior prob

~~R / Bag~~ marginal ~~likelihood~~

Naive Bayes Example

- **Email classification:** spam vs ham (non-spam)
- Words act as independent features here
- Count word frequencies across training dataset
- Estimate likelihoods per class using frequencies
- Calculate posterior probabilities per email category
- Assign email to higher posterior probability

Discretization in Naive Bayes

- Naïve Bayes uses frequency tables
- Each feature must be categorical valued
- Numeric features lack inherent categories
- Algorithm fails directly with numeric data
- **Solution:** discretize numbers into bins
- Discretization also known as binning

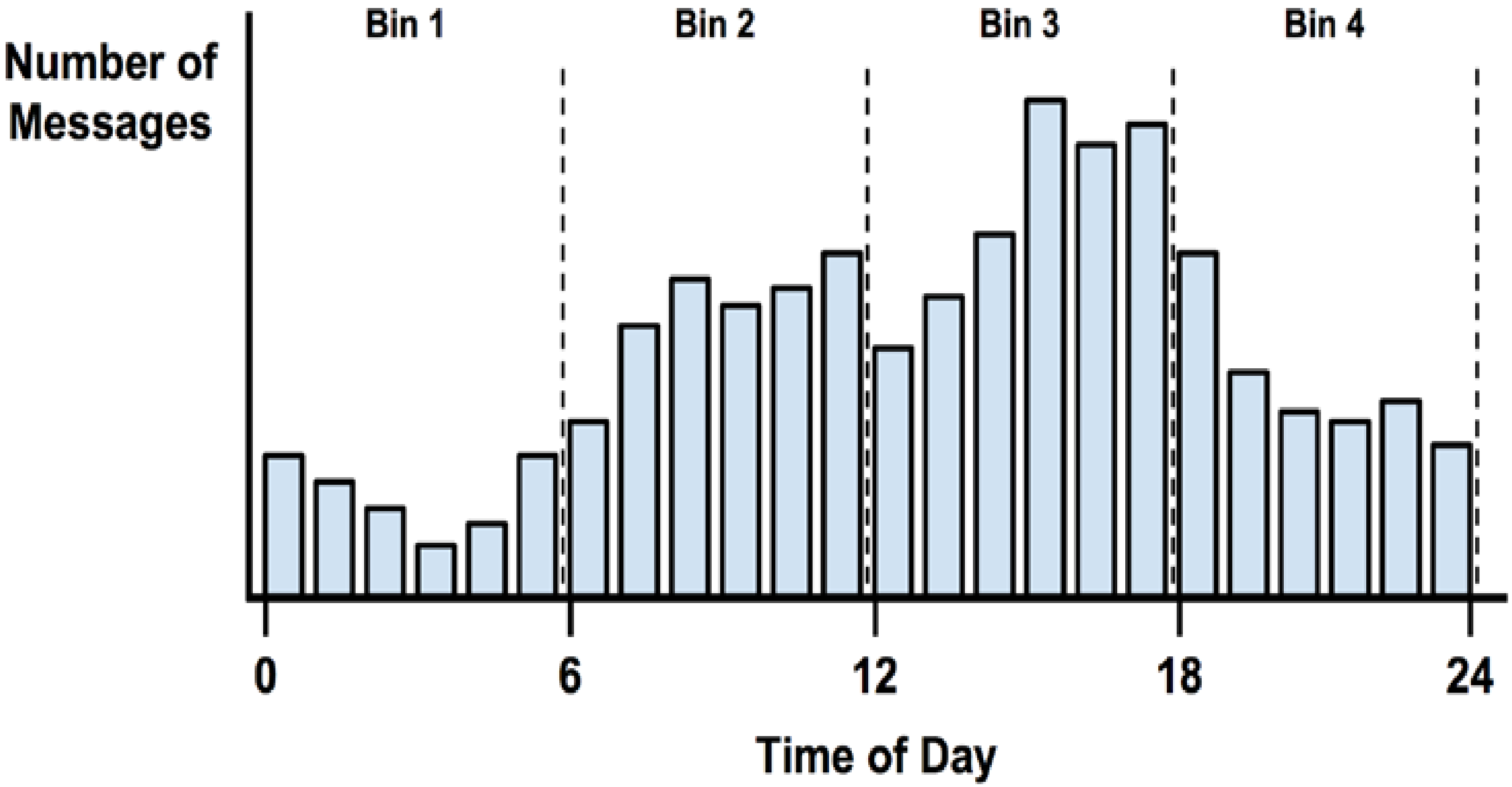
Methods of Discretization

- Discretization means grouping numeric values
- Places values into bins or categories
- Several methods available for binning
- **Most common:** natural cut points distribution
- Explore data for natural boundaries
- Creates categorical ranges from continuous data

Natural cut points:

- boundaries in data where values naturally separate
- E.g. clear gaps or clusters in the data distribution.
- These cuts divide data into meaningful intervals instead of random ranges.
- E.g. If exam scores cluster around 40, 60, and 80, those points can be natural cuts for “Low,” “Medium,” and “High.”

- **Example of Discretization**
- Consider spam dataset with time feature
- Time measured 0 to 24 hours
- Divide into ranges: morning, afternoon, night
- Each bin acts as category input
- Frequency table created for each bin
- Enables Naive Bayes probability calculation


Bin 1
Bin 2
Bin 3
Bin 4
Number of Messages
0
6
12
18
24
Time of Day

Advantages & Disadvantages of Naive Bayes

- Advantages: simple, fast, scalable, interpretable
- Performs well with small training datasets
- Effective with high-dimensional sparse data
- **Disadvantages:** independence assumption rarely true
- May underperform if features correlated

KNN vs Naive Bayes

- KNN is instance-based learning technique
- Naive Bayes is probabilistic learning approach
- KNN slower during prediction phase
- Naive Bayes faster prediction after training
- KNN sensitive to irrelevant noisy features
- Naive Bayes sensitive to independence assumption