

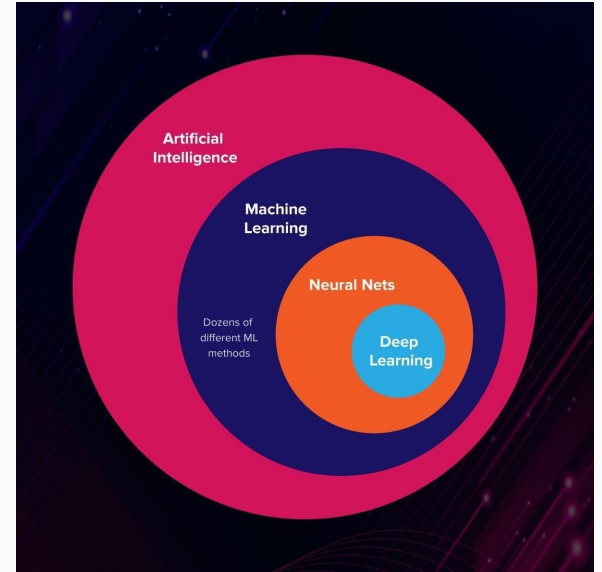
What Is Machine Learning?

CHAPTER 3

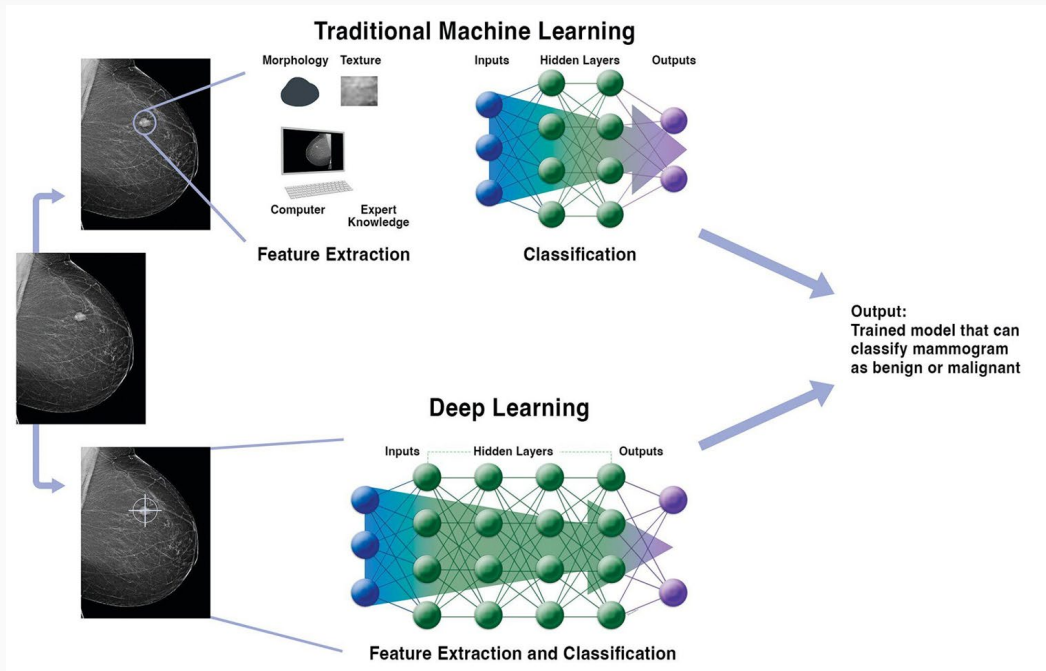


What's the difference between artificial intelligence and machine learning?

- Definition of AI vs ML



Machine Learning in Mammography



What is agent?

- An agent is anything that can be seen as perceiving its environment through sensors and acting on them through effectors.
- An agent can be understood as a mapping between percept sequences and actions.

Autonomy

can learn from experiences over time.

Interface

An interface agent is one that provides an interface to a complex system.

Performance

Performance is the measure used to evaluate the behavior of the agent in the environment. It answers the question, Does the agent do what it's supposed to do in the environment?

Goals

Goals refer to what the agent is trying to achieve.

Utility

Utility refers to the agent's own, internal performance assessment—that is, the agent's own measure of performance at any given state. This may differ from the performance of the agent.

This allows rational decisions in cases where there are several paths to the same goal, allowing an agent to distinguish the paths better than others

Knowledge

Knowledge is acquired by an agent through its sensors or knowledge about the environment. Knowledge can be used to decide how to act.

Environment

- Accessibility
- Determinism
- Episodes
- Type of environment
- Flow of data to environment

Training Data

Training data is the data that will be used by the learning algorithm to learn possible hypotheses.

Target Function

This is the mapping function f from x to $f(x)$

Hypothesis

This is an approximation of f .

Learner

The learner is the learning algorithm or process that creates the classifier.

Validation

Validation includes methods used within machine learning development that provide a method of evaluation of model performance.

Feature

A feature is a data attribute and its value.

Feature Selection

this is the process of choosing the features required to explain the outputs of a statistical model while excluding irrelevant features.

Machine Learning Basics

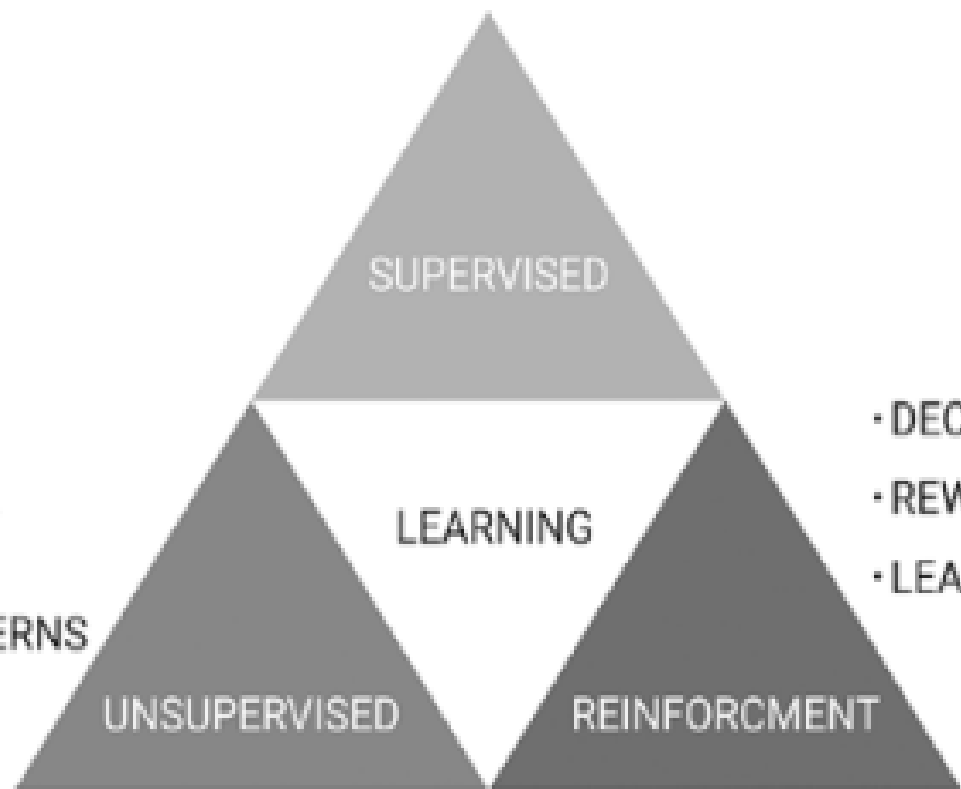
Supervised learning

Unsupervised learning

Semi-supervised learning

Reinforcement learning

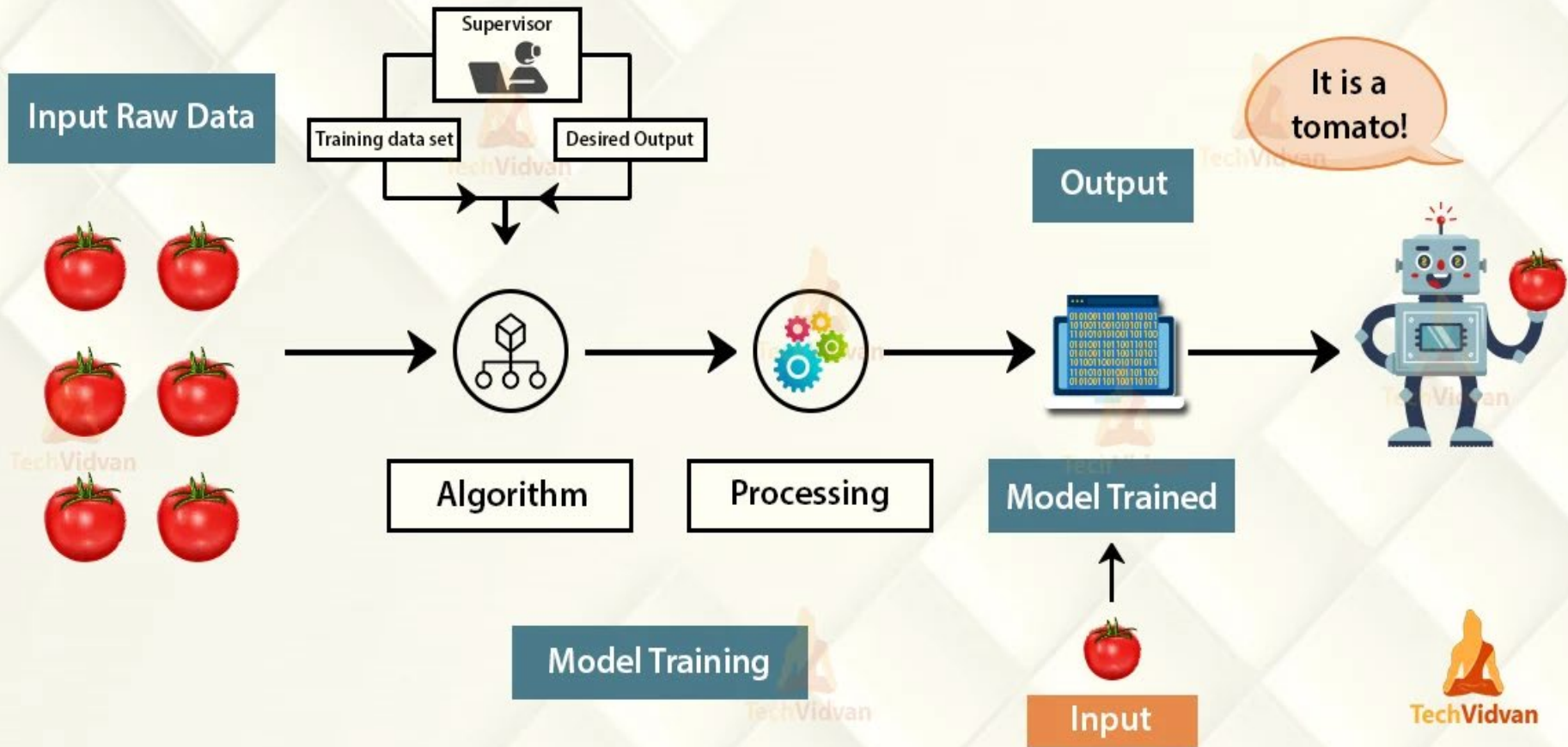
- LABELED DATA
- DIRECT FEEDBACK
- PREDICT OUTCOME/FUTURE



- NO LABELS
- NO FEEDBACK
- SEARCH FOR HIDDEN PATTERNS

- DECISION PROCESS
- REWARD SYSTEM
- LEARN SERIES OF ACTIONS

Supervised Learning in ML



Supervised learning (Classification)

support vector machines

naïve Bayes

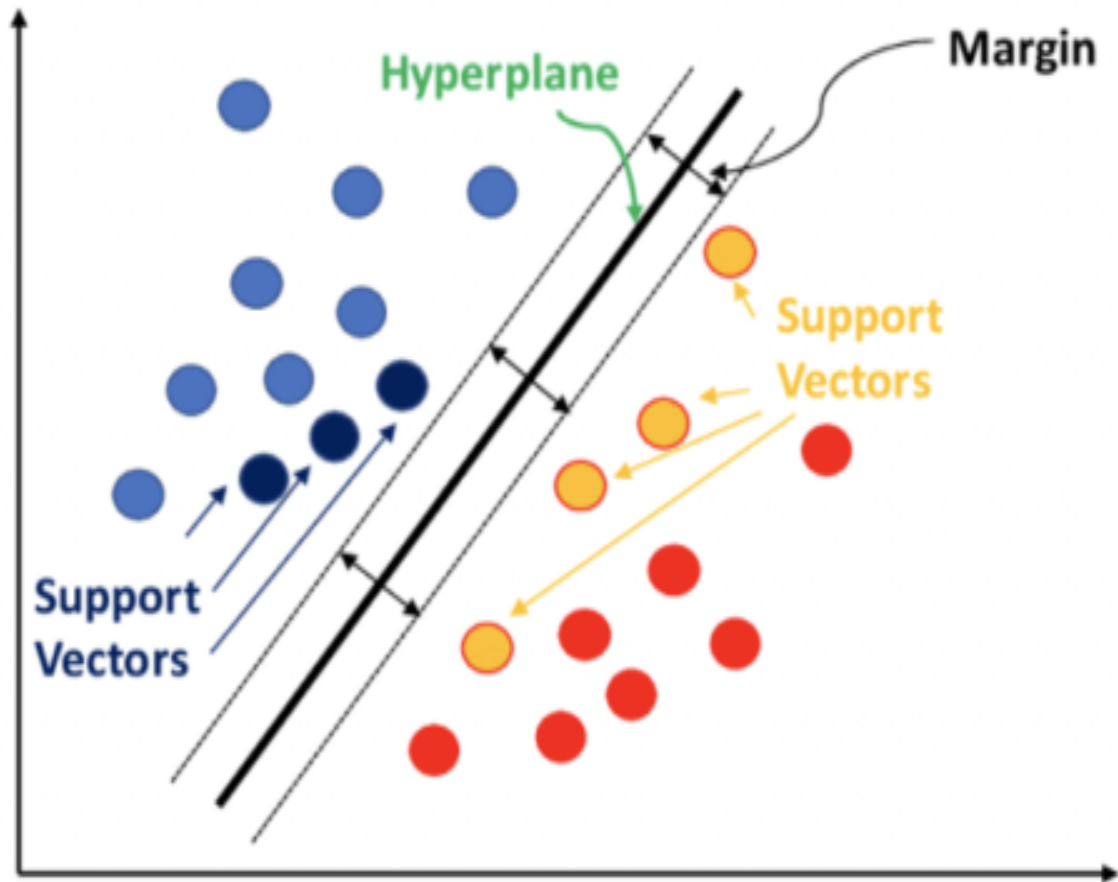
Gaussian Bayes

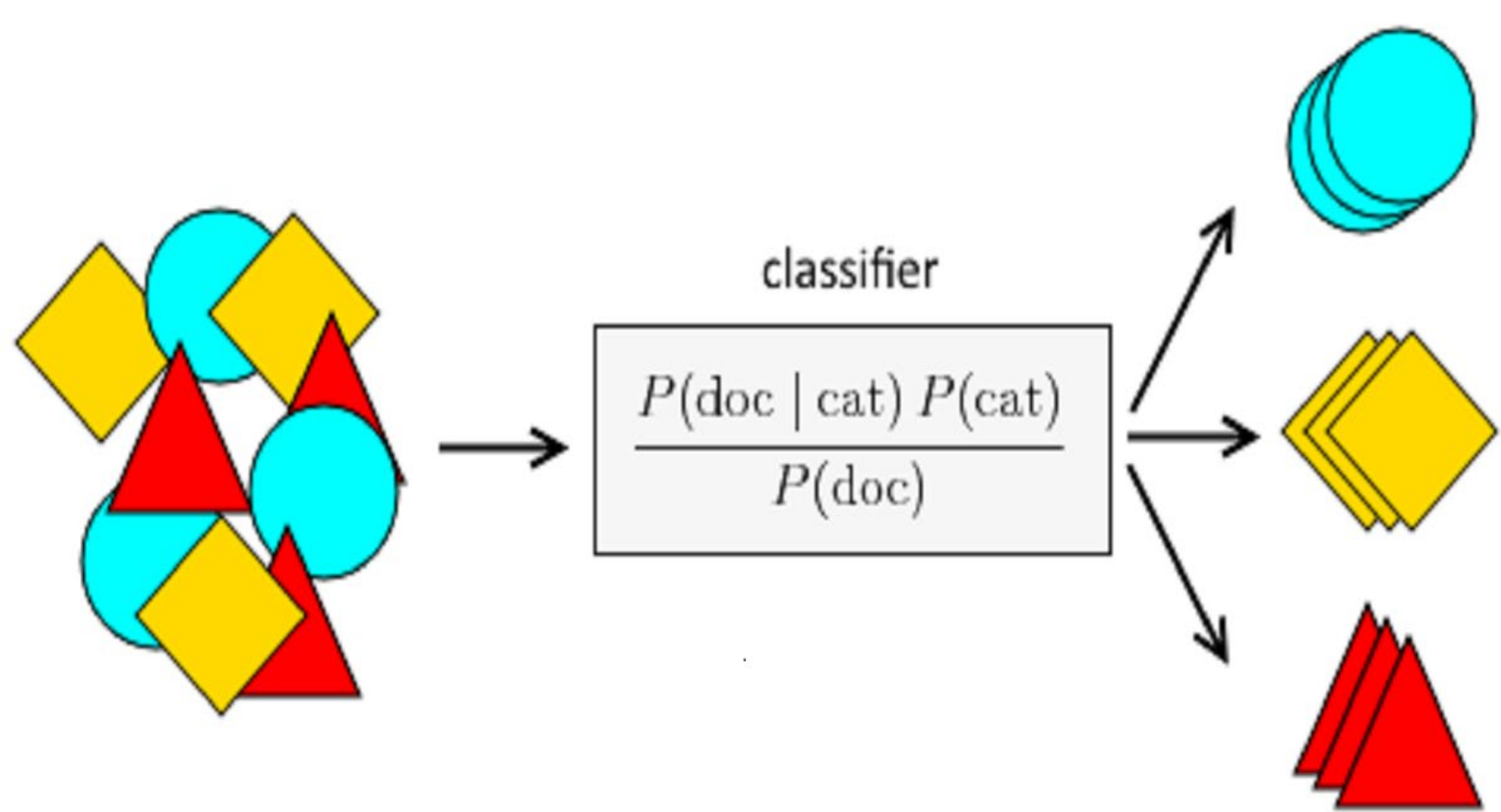
k-nearest neighbors (KNN)

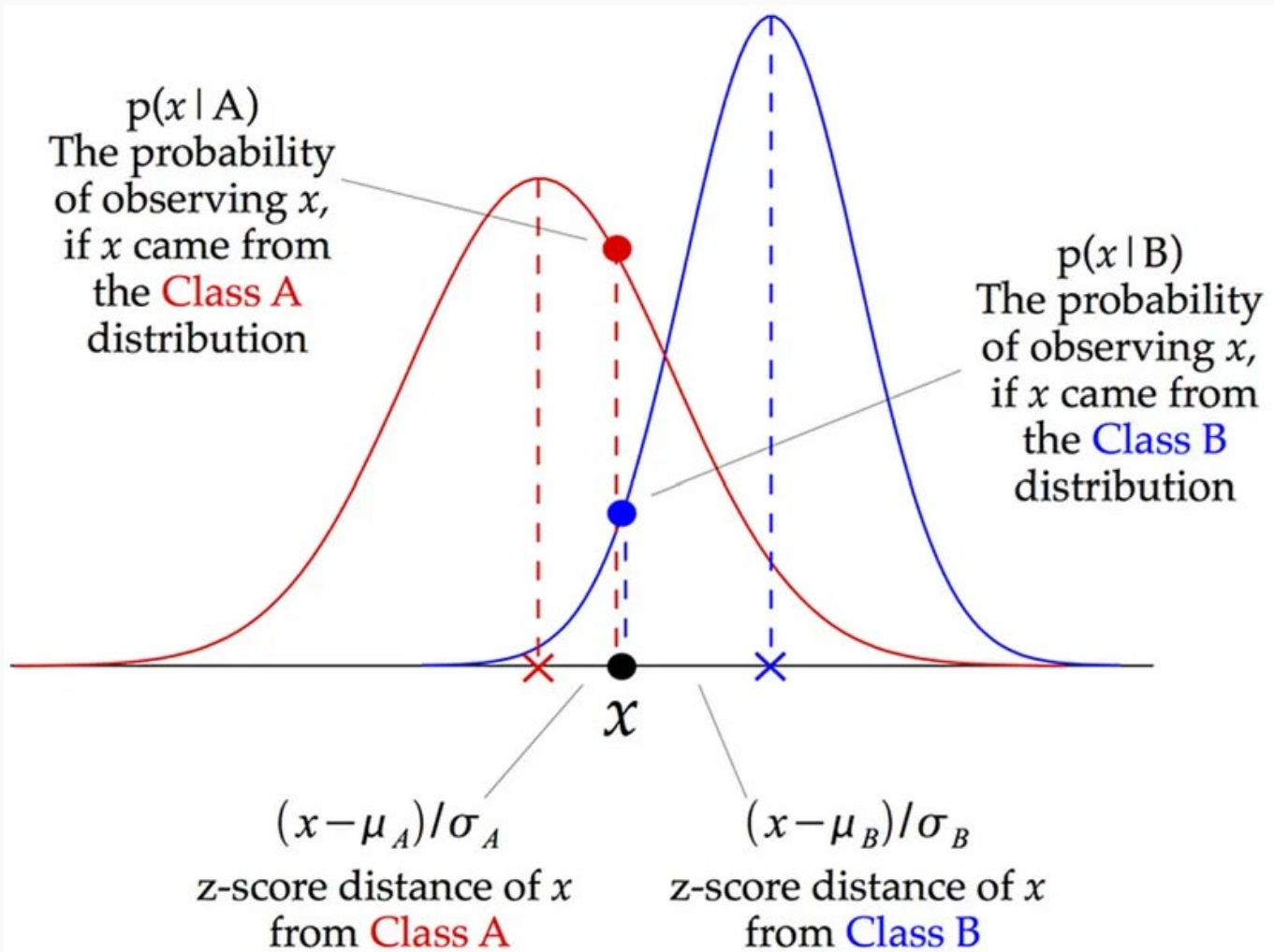
logistic regression

WHAT IS A

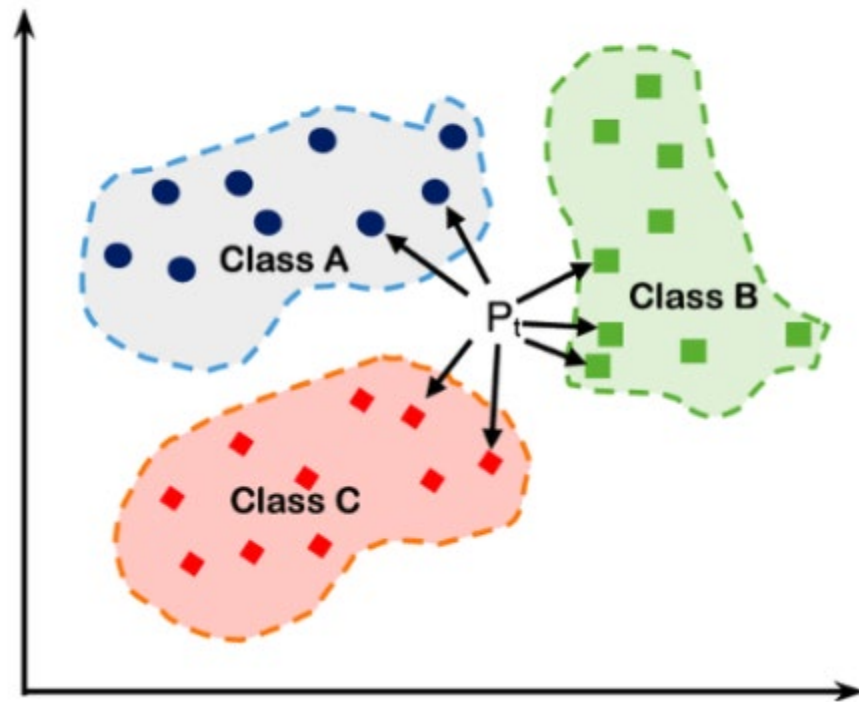
SUPPORT VECTOR MACHINE?

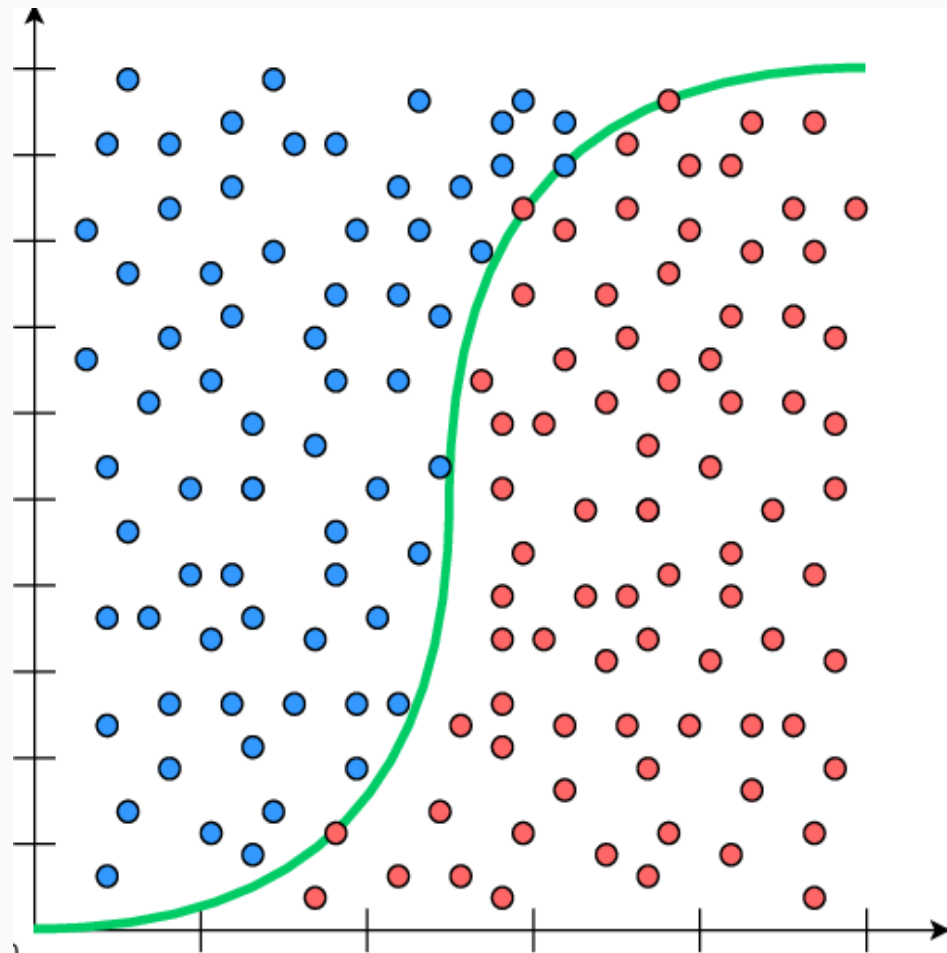






K Nearest Neighbors





Supervised learning (Regression)

Linear regression

polynomial regression

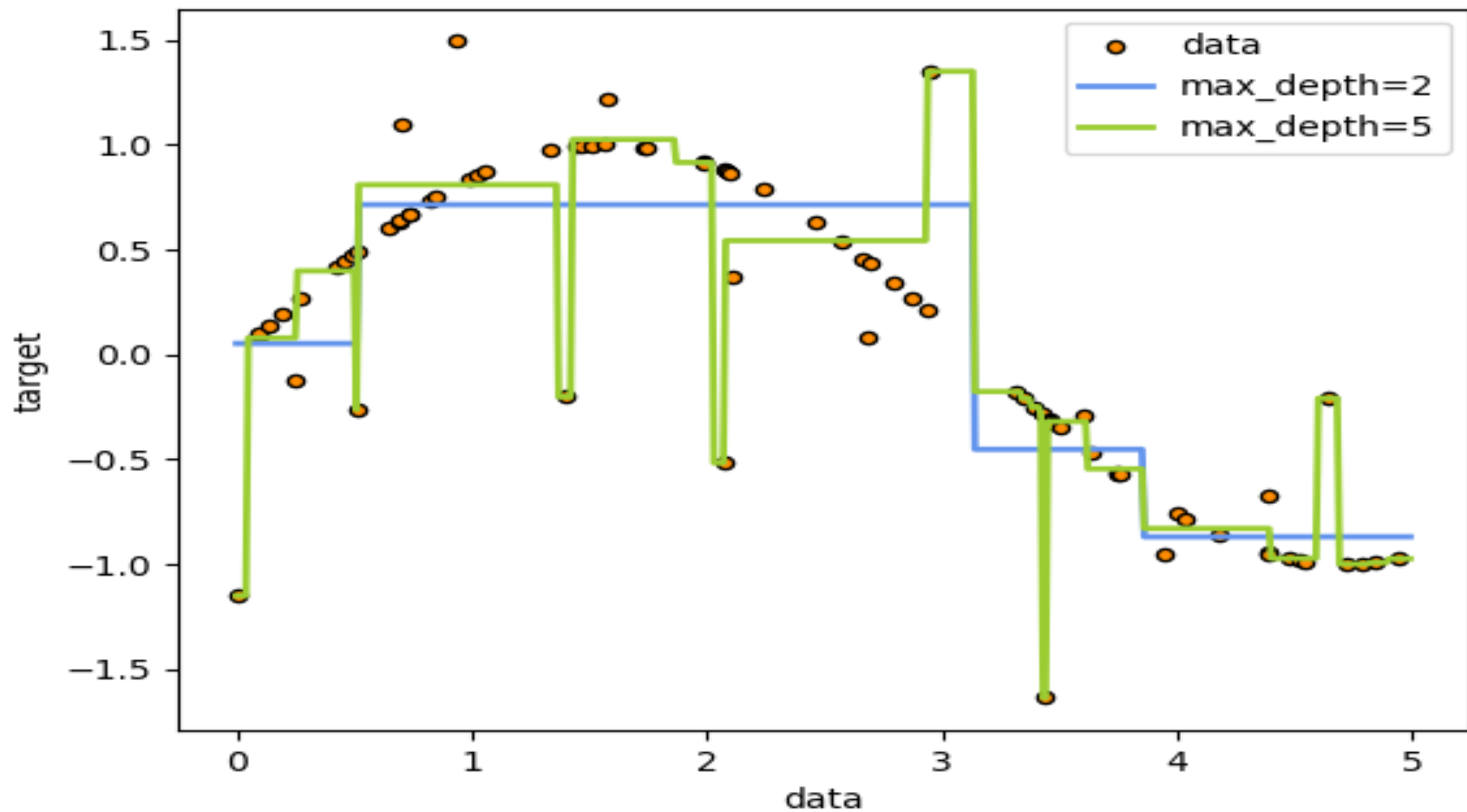
support vector machine (SVR)

Ensembles

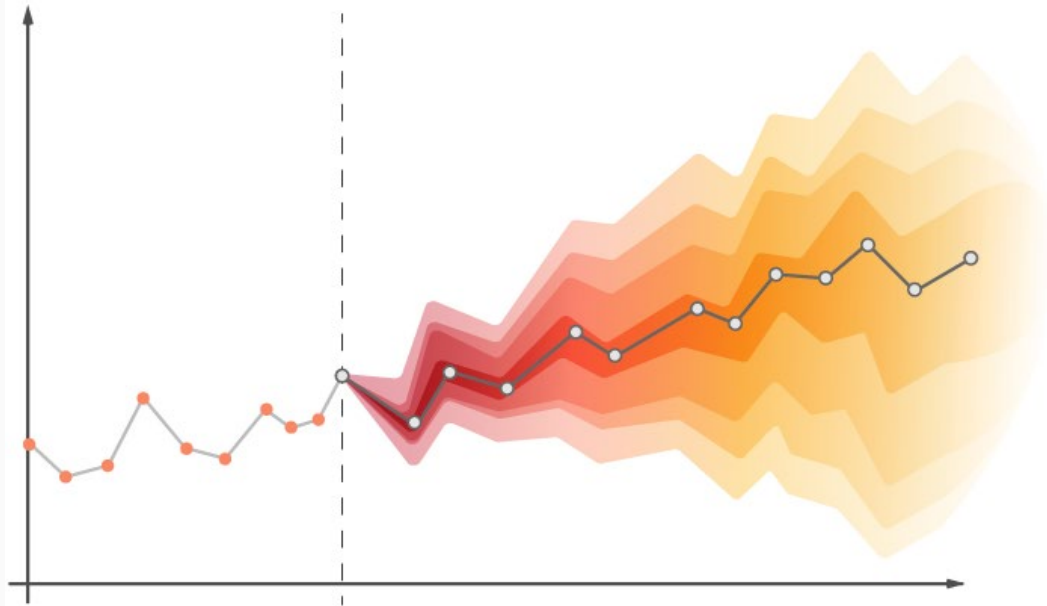
decision trees

neural networks

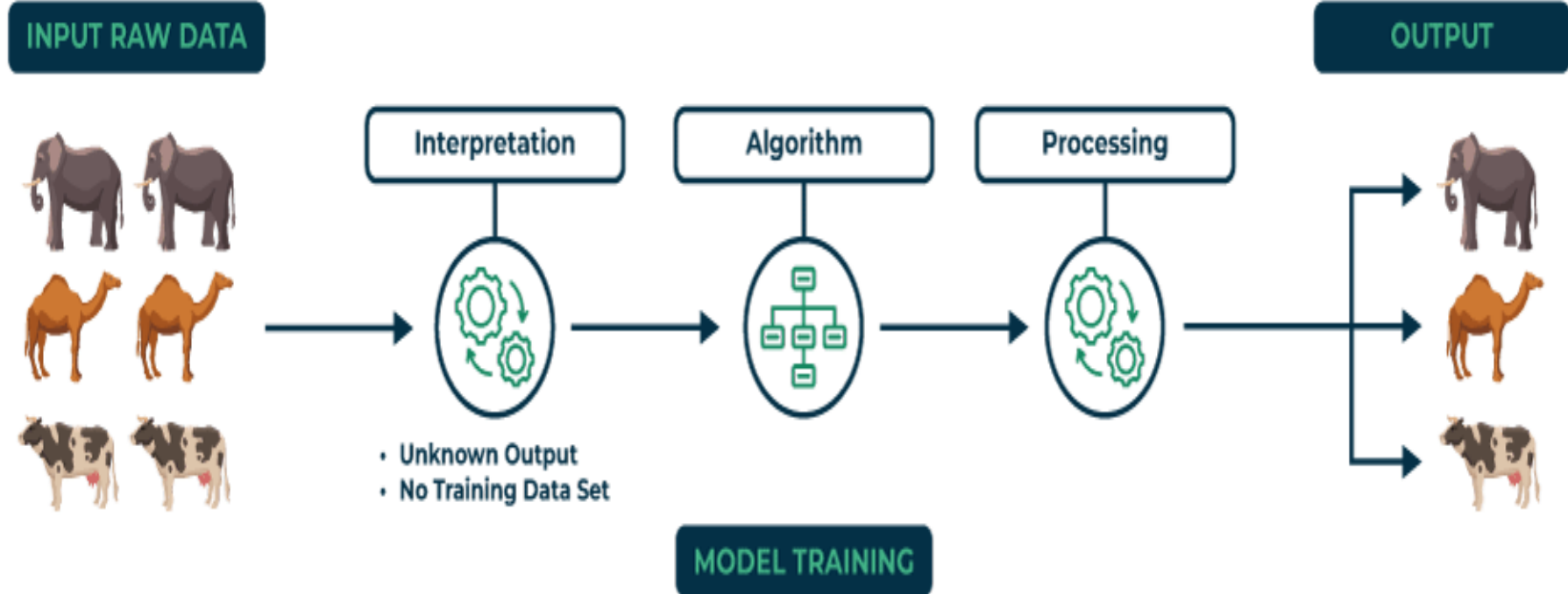
Decision Tree Regression



Supervised learning (forecasting)



Unsupervised Learning



Supervised Learning Algorithms

Regression

- Linear
- Polynomial
- Ridge Regression
- Lasso Regression

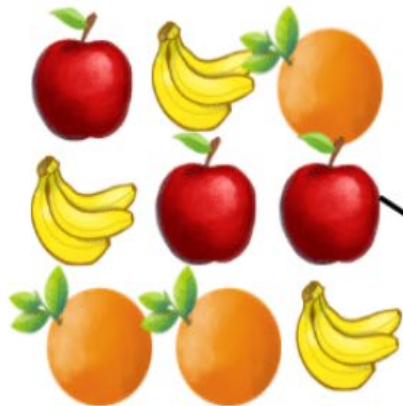
Classification

- Logistic Regression
- Linear Discriminant Analysis
- Naïve Bayes

Regression & Classification

- Decision Trees
- K- Nearest Neighbors (KNN)
- Support Vector Machine (SVM)

Input Data



Machine Learning Model



Prediction

It's an Apple

Partial Labels



Orange
Banana

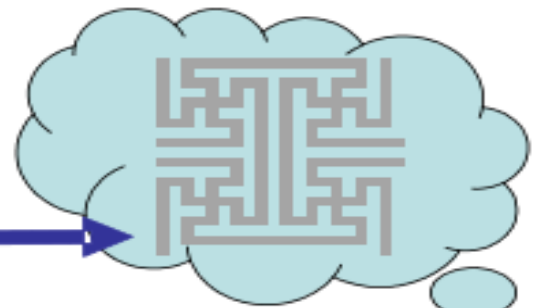


Unlabelled Data

internal state



reward



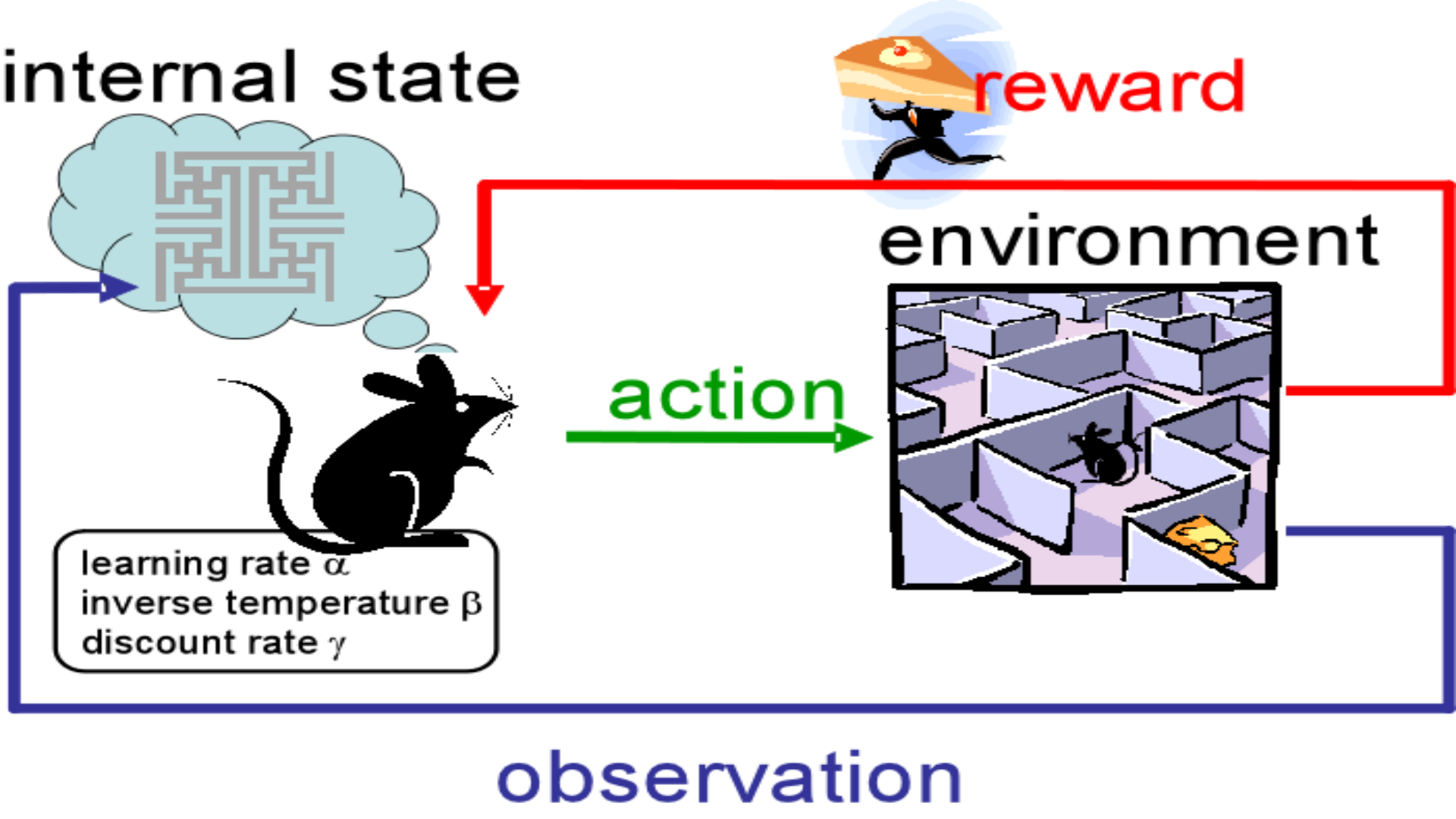
environment

action



learning rate α
inverse temperature β
discount rate γ

observation



Data mining

Pattern predictions based on trends and behaviors

Prediction based on probable outcomes

Analysis of large datasets (particularly unstructured)

Clustering through identification of facts previously unknown

Expanding And Exploring Business

Classification

Clustering

Regression

Time Series

Detection Of Anomaly

Analysis Of Association

**Tasks Of
Data Mining**

	Parametric	Non-parametric
Assumed distribution	Normal	Any
Assumed variance	Homogeneous	Homogenous and Heterogeneous
Typical data	Ratio or Interval	Ordinal or Nominal
Data set relationships	Independent	Any
Usual central measure	Mean	Median
Benefits	Can draw more conclusions	Simplicity; Less affected by outliers