Introduction

- Computational Linguistics and Natural Language Processing
- Analytics for Industries
- Pattern Recognition and Anomaly Detection

Computational Linguistics and Natural Language Processing

Computational Linguistics

- **Linguistics** (study of language)
- **Computer Science** (coding and algorithms)

What is Computational Linguistics

- Computational Linguistics is the scientific study of language using computers.
- It tries to understand how language works and builds computer models of that understanding.
- **Example:**
- Creating grammar rules that a computer can use
- Understanding how humans form sentences
- Studying language meaning and structure with the help of algorithms

What is Natural Language Processing

- **Natural Language Processing** is a part of **Artificial Intelligence (AI)**.
- It focuses on **making computers understand, interpret, and generate human language** (like English, Hindi, etc.).
- **Real-life Examples:**
- Google Translate
- Chat bots (like Siri, Alexa)
- Spell checkers
- Voice assistants
- Sentiment analysis (positive/negative review detection)

Natural language Processing (NLP)

→ How Human Communicate with each other

→ Computer should replicate the same thing

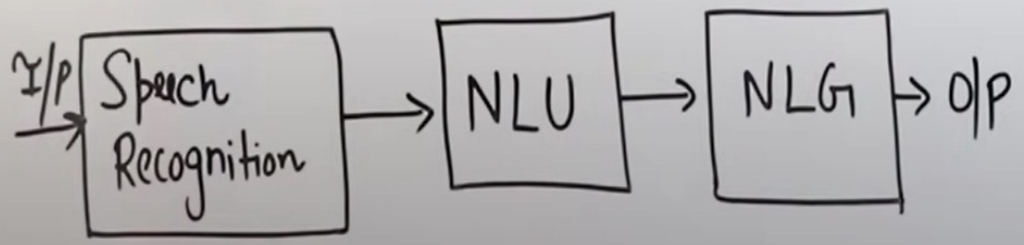
→ Applications of NLP

* Speech Recognition
* Sentimental Analysis
* Machine Translations
* Chat bots etc.

NLU ⇒ What do the users say? their intent? Meaning?

Challenges:
- Lexical Ambiguity
- Syntactic Ambiguity
- Semantic "
- Pragmatic "

→ The tank was full of water.
→ Old men and women were taken to safe place.
→ The Car hit the pole while it was moving.
→ The police are Coming.



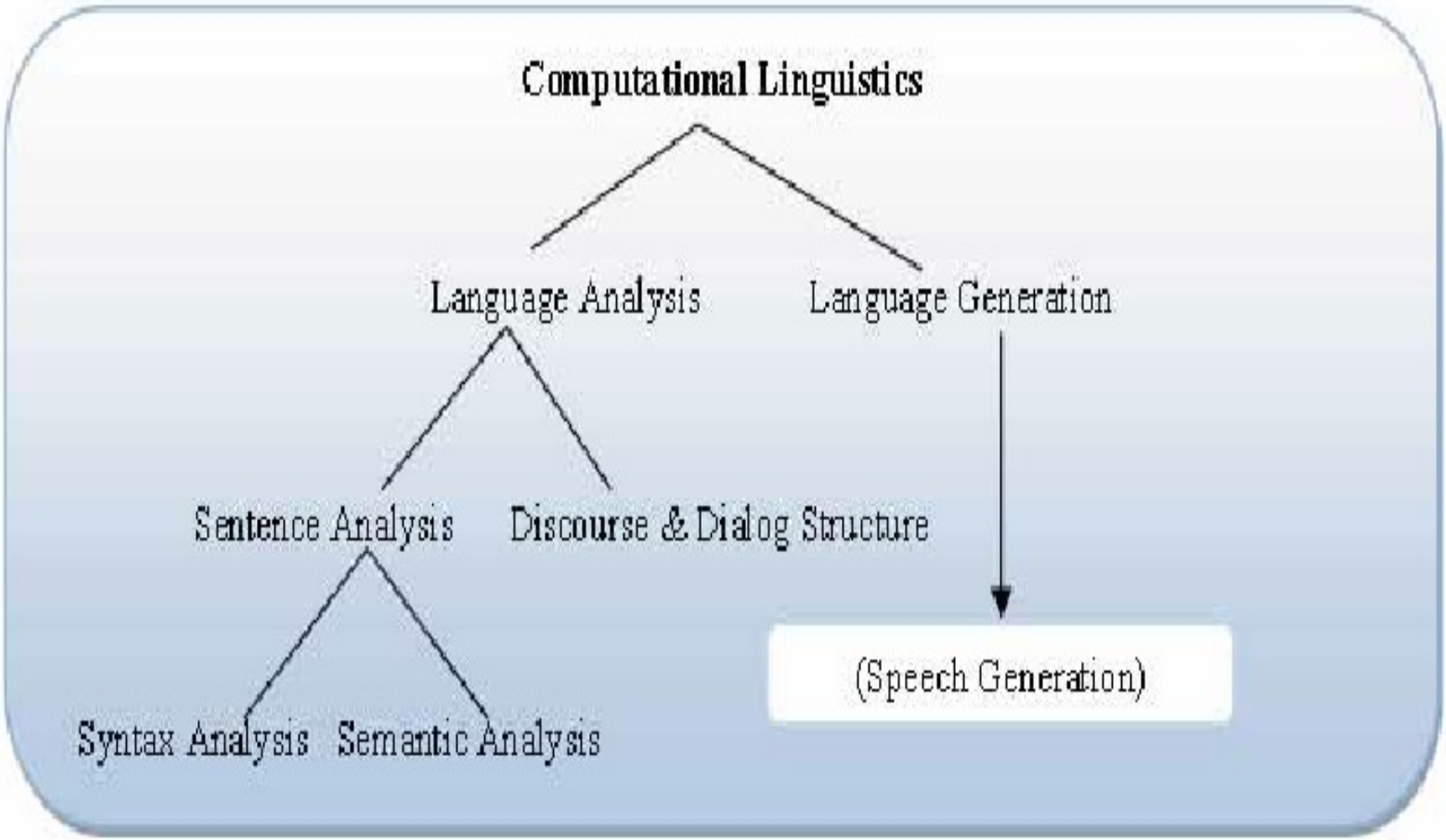
NLG → What should we say to User?
→ it should be Intelligent and Conversational.
→ Deal with structured data.
→ Text / Sentence Planning

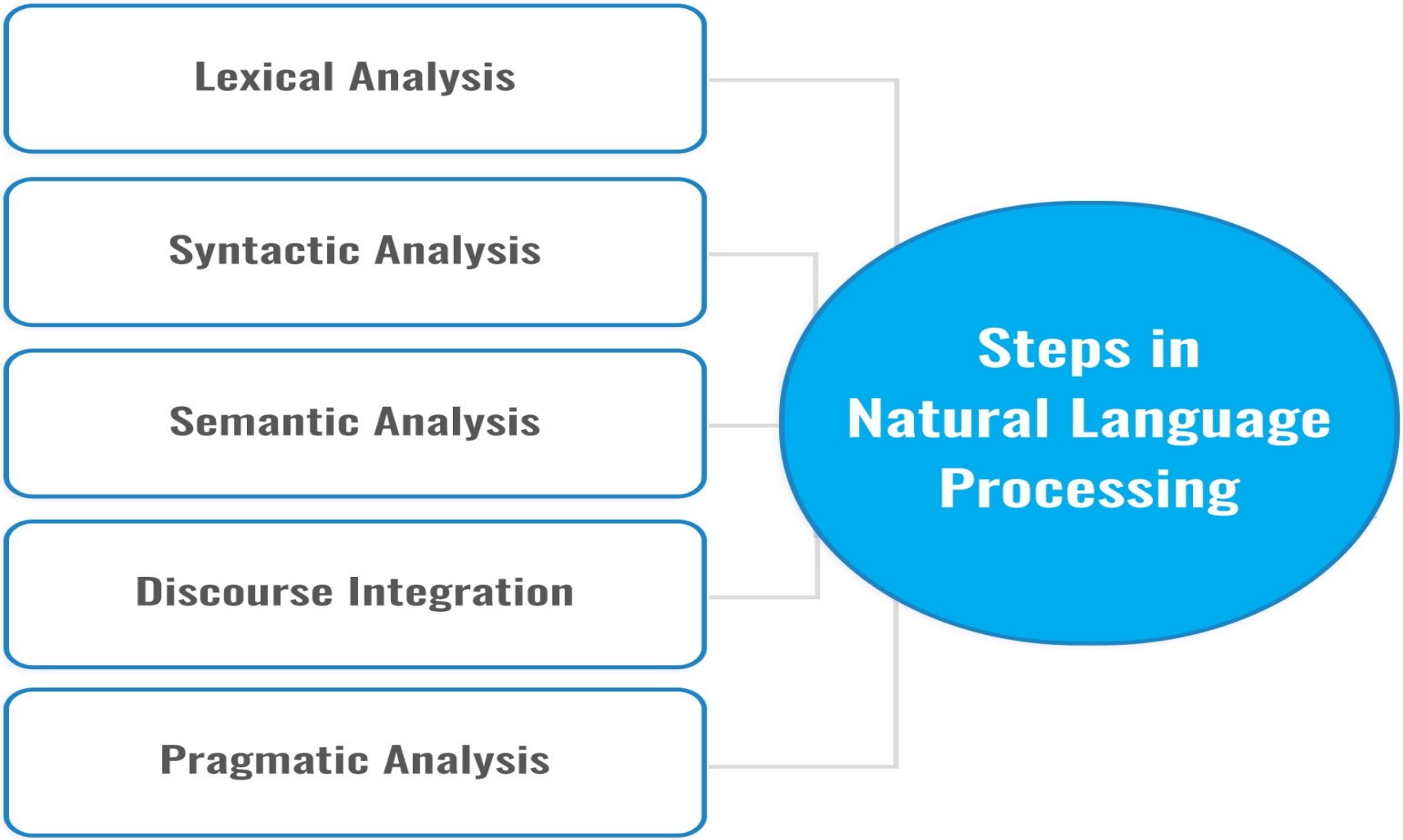
Difference Between CL and NLP

Feature	Computational Linguistics	Natural Language Processing
Purpose	Study of language using computers	Use of language in real-world applications
Approach	Rule-based, theory-driven	Data-driven, machine learning focused
Focus	Understanding how language works	Making machines *do* language tasks

Summary

- **CL** helps us **understand language** using computer tools.
- **NLP** helps us **use language** in computer applications.
- Both fields are essential to build smart tools like translation systems, chatbots, and voice assistants.


Computational Linguistics
Language Analysis
Language Generation
Sentence Analysis
Discourse & Dialog Structure
(Speech Generation)
Syntax Analysis
Semantic Analysis


Lexical Analysis
Syntactic Analysis
Semantic Analysis
Discourse Integration
Pragmatic Analysis
Steps in Natural Language Processing

Analytics for Industries

What is Analytics?

- Analytics is the process of using data to discover meaningful patterns, draw conclusions, and support decision-making.
- It involves various techniques including statistics, machine learning, and data visualization.

Why Analytics in Industries?

- In the modern world, industries generate vast amounts of data from operations, customers, sensors, and transactions.

Analytics helps convert this raw data into actionable insights, which:

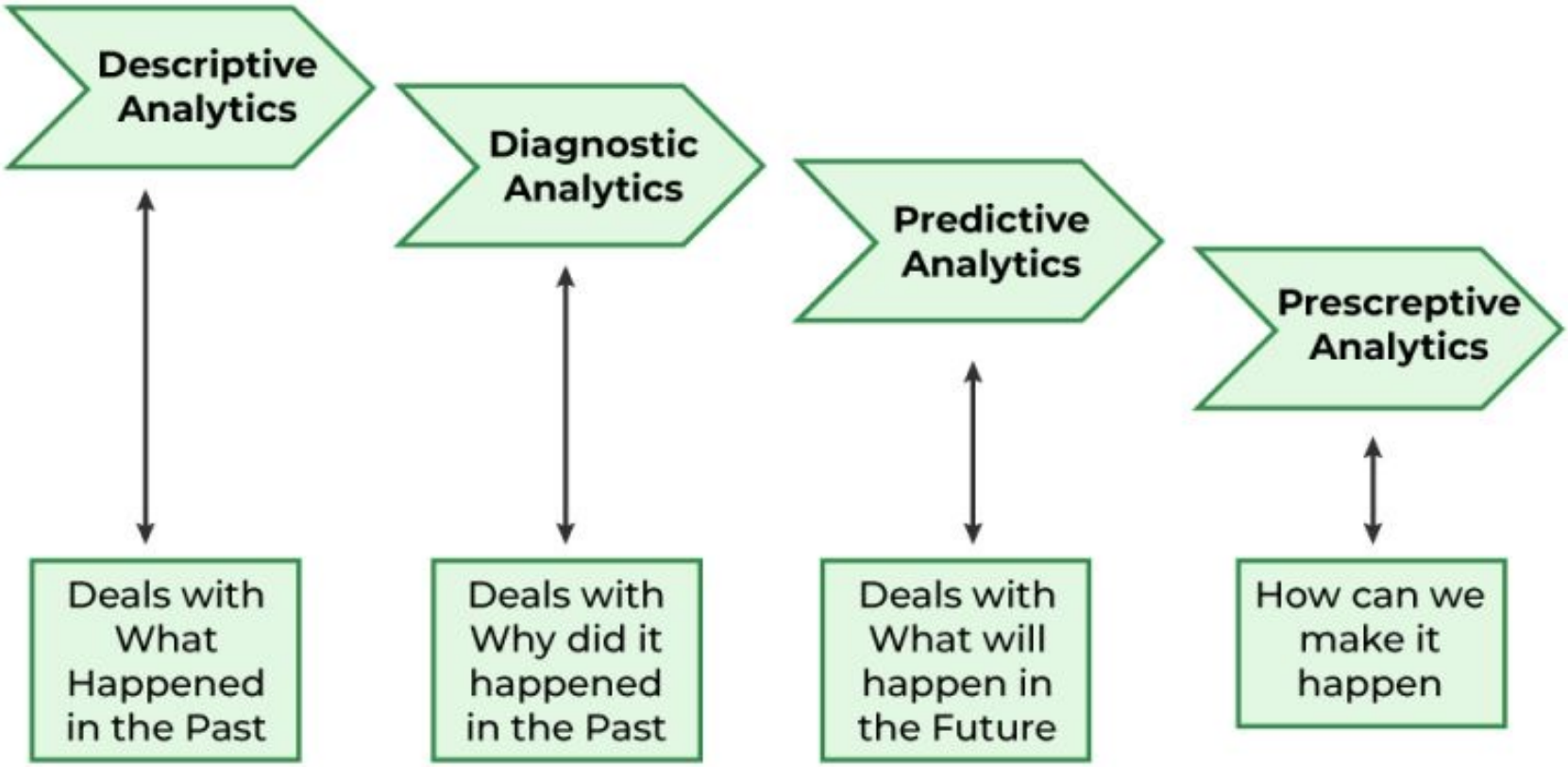
- Improve operational efficiency
- Predict market trends
- Enhance Customer Satisfaction
- Reduce Cost

Types of Analytics

- **Descriptive Analytics** – What happened? (e.g., sales reports)
- **Diagnostic Analytics** – Why did it happen? (e.g., root cause analysis)
- **Predictive Analytics** – What is likely to happen? (e.g., demand forecasting)
- **Prescriptive Analytics** – What should we do? (e.g., inventory optimization)

Key Tools & Technologies

- Programming: Python, R, SQL
- Tools: Power BI, Tableau, Excel, SAS
- Technologies: Big Data (Hadoop, Spark), Cloud platforms (AWS, Azure, GCP)
- AI/ML: Machine Learning, Deep Learning, Natural Language Processing


Descriptive Analytics
Diagnostic Analytics
Predictive Analytics
Prescreptive Analytics
Deals with What Happened in the Past
Deals with Why did it happened in the Past
Deals with What will happen in the Future
How can we make it happen

Pattern Recognition and Anomaly Detection

What is pattern recognition?

- A **pattern** is an object, process or event that can be given a name.
- A **pattern class** (or category) is a set of patterns sharing common attributes and usually originating from the same source.
- In **recognition** or **classification** process given objects are assigned to prescribed classes.
- A **classifier** is a machine which performs classification.

Pattern Recognition

System to classify four characters – A, B, C and D

In all four characters, we have some lines and curves -

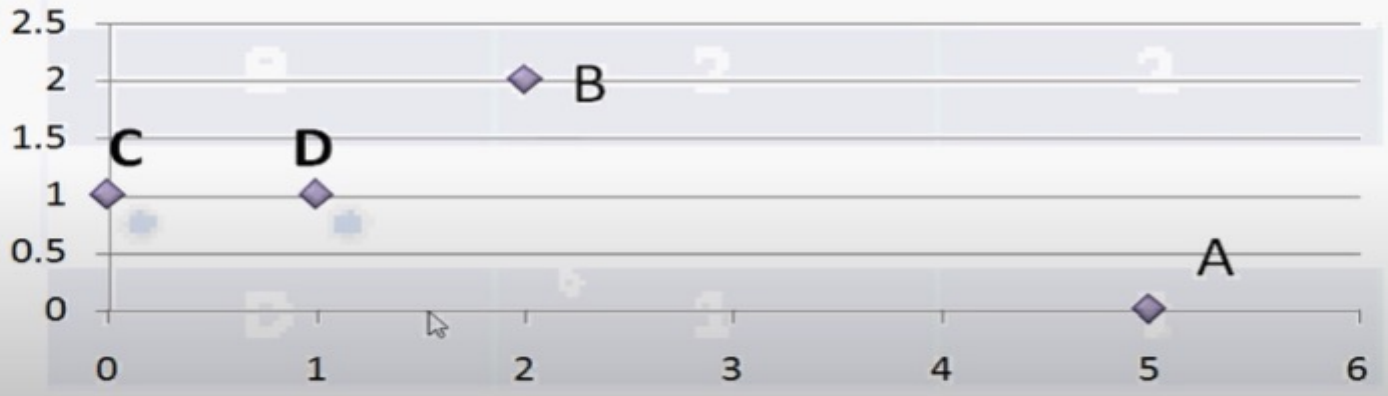
A B C D

Pattern Recognition

Character	No. of Lines	No. of Curves
A	5	0
B	2	2
C	0	1
D	1	1

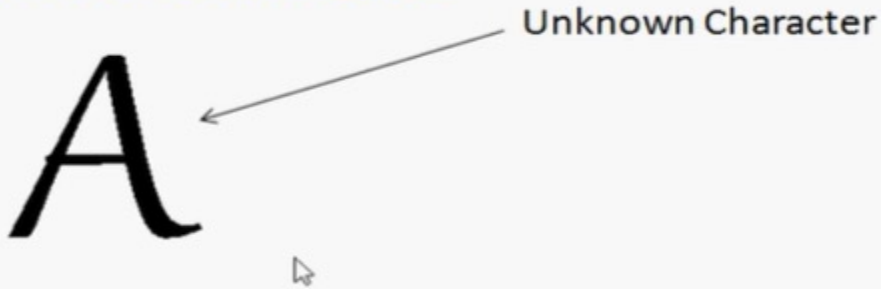
Pattern Recognition

With these two features, all the four characters can be represented in a two dimensional space-



Pattern Recognition

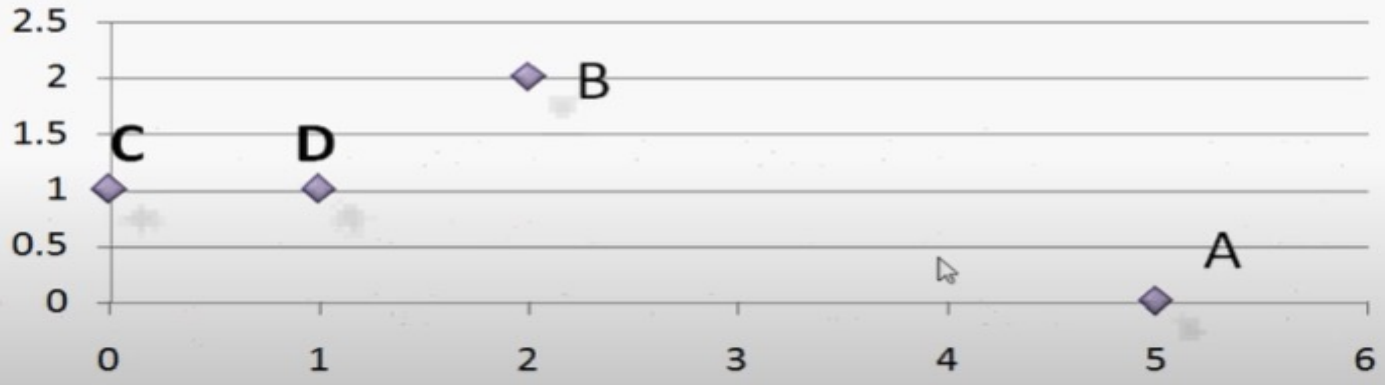
Suppose anyone draws a character -

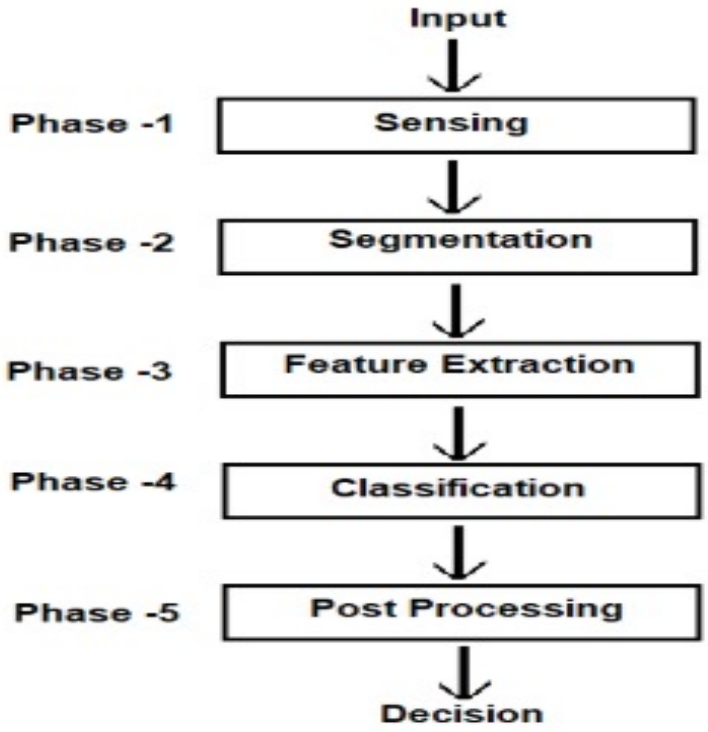


System first extracts the features from handwritten characters- *Number of lines (4)* and *number of curves (1)* The pattern for above character is (4,1).

Pattern Recognition

With these two features, all the four characters can be represented in a two dimensional space-



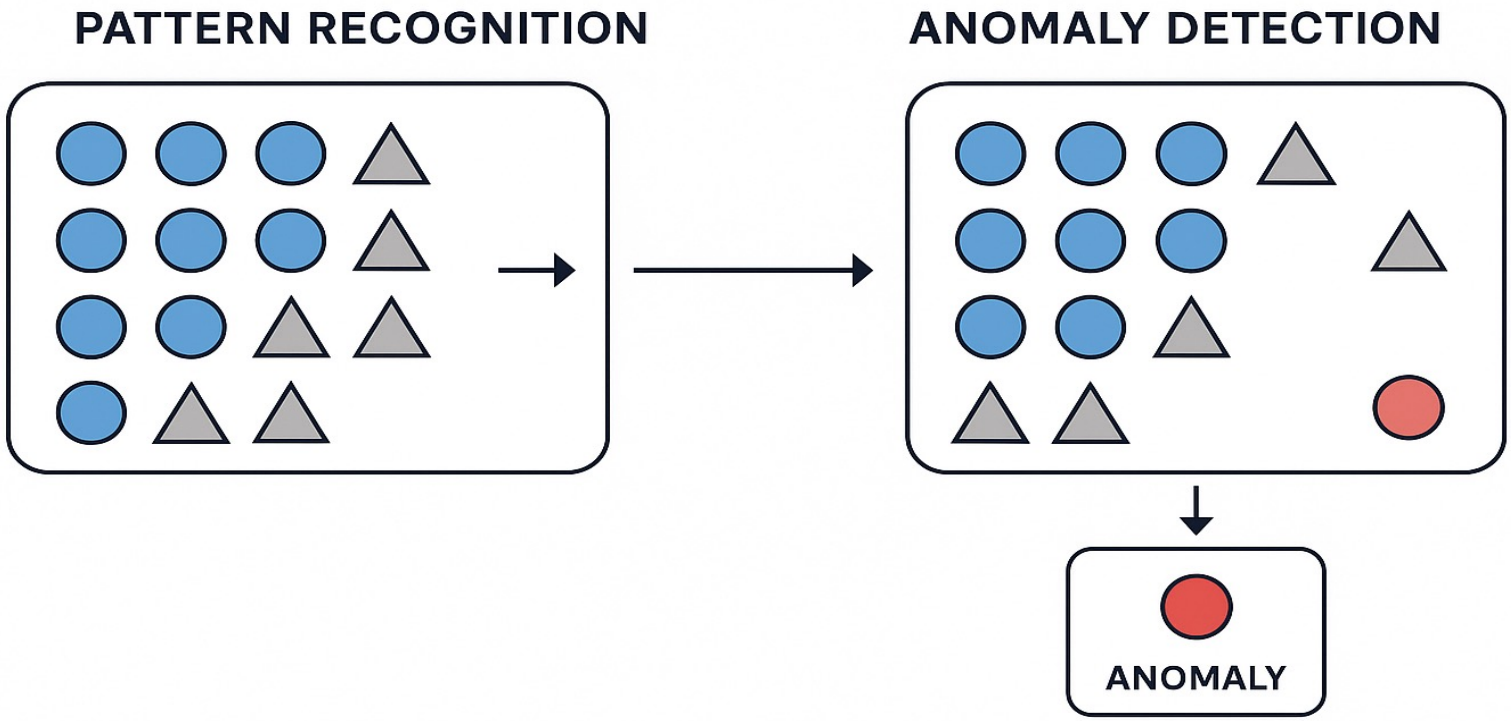


Phases in Pattern Recognition

Anomaly Detection

- **Anomaly Detection**, also called **Outlier Detection**, is a technique used to identify rare or unusual patterns in data that do not conform to expected behavior.
- These “anomalies” or “outliers” could indicate:
- Fraud (in banking/finance)
- Faults (in machines or sensors)
- Errors (in data entry)
- Health issues (in medical diagnosis)

PATTERN RECOGNITION AND ANOMALY DETECTION



Example Use Cases

- **Banking**: Detecting fraudulent transactions.
- **Healthcare**: Identifying abnormal heart rates or test results.
- **Cybersecurity**: Spotting unauthorized access or attacks.
- **Manufacturing**: Noticing defective products on the line.

Types of Anomalies

- **Point Anomalies** – Single data points far from others.
- **Contextual Anomalies** – Unusual in a specific context (e.g., high temperature in winter).
- **Collective Anomalies** – A group of data points behaving unusually together.

Common Techniques

- Statistical methods (mean, standard deviation)
- Machine learning (Isolation Forest, SVM)
- Neural networks (Autoencoders)
- Clustering (K-means, DBSCAN)