



Pattern Recognition: An Introduction



Dr. Konstantinos Karampidis



Definition (1/3)

- ❑ Pattern recognition (PR) is a subtopic of machine learning.
- ❑ Is the study of how machines can
 - Observe the environment,
 - Learn to distinguish patterns of interest,
 - Identify regularities, structures or meaningful information in data and assigning observations to categories or concepts.
 - Make sound and reasonable decisions about the categories of the patterns.



Definition (2/3)

- ❑ **Pattern:** a description of an object.
- ❑ **Recognition:** classifying an object to a pattern class.
- ❑ PR techniques are an important component of intelligent systems and are used for
 - Decision making
 - Object & pattern classification
 - Data preprocessing



Definitions from the literature (3/3)

- "The assignment of a physical object or event to one of several prespecified categories" -Duda and Hart
- "A problem of estimating density functions in a high-dimensional space and dividing the space into the regions of categories or classes" - Fukunaga
- "Given some examples of complex signals and the correct decisions for them, make decisions automatically for a stream of future examples" - Ripley
- "The science that concerns the description or classification (recognition) of measurements" -Schalkoff
- "The process of giving names y to observations x ", -Schürmann
- Pattern Recognition is concerned with answering the question "What is this?" -Morse



Why Pattern Recognition?

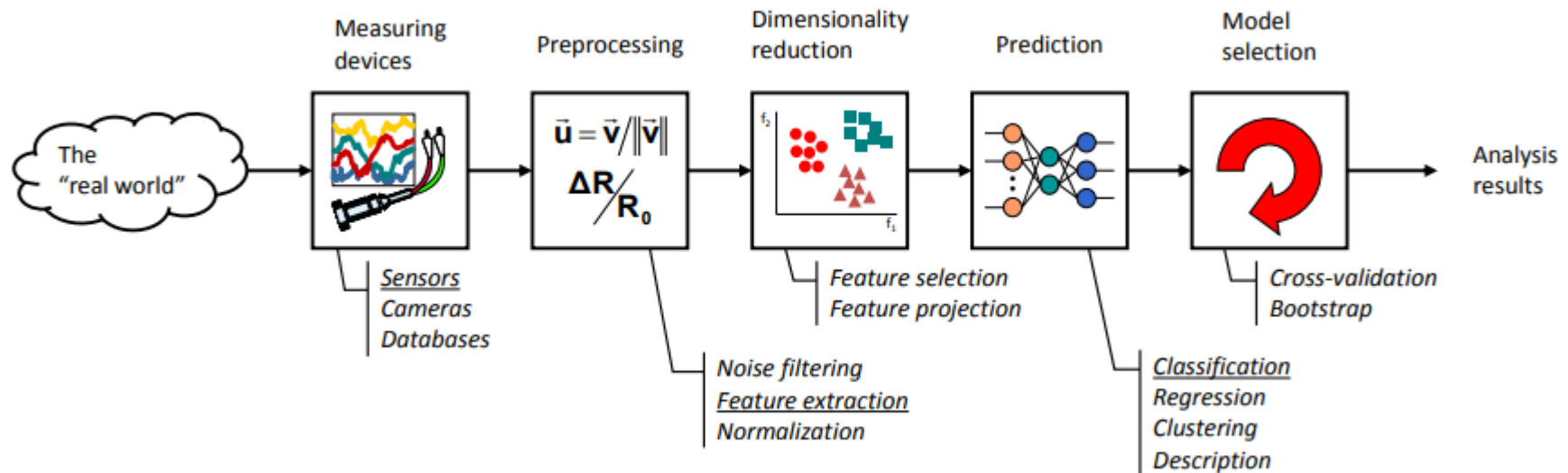
Why do we need automated recognition?

- ❑ Large volumes of data that are difficult to analyze manually.
- ❑ Variability between observations belonging to the same class.
- ❑ Similarities between observations belonging to different classes.
- ❑ Need for consistent, repeatable and efficient decisions.
- ❑ Uncertainty and incomplete information in real-world data.



Pattern Recognition Categories

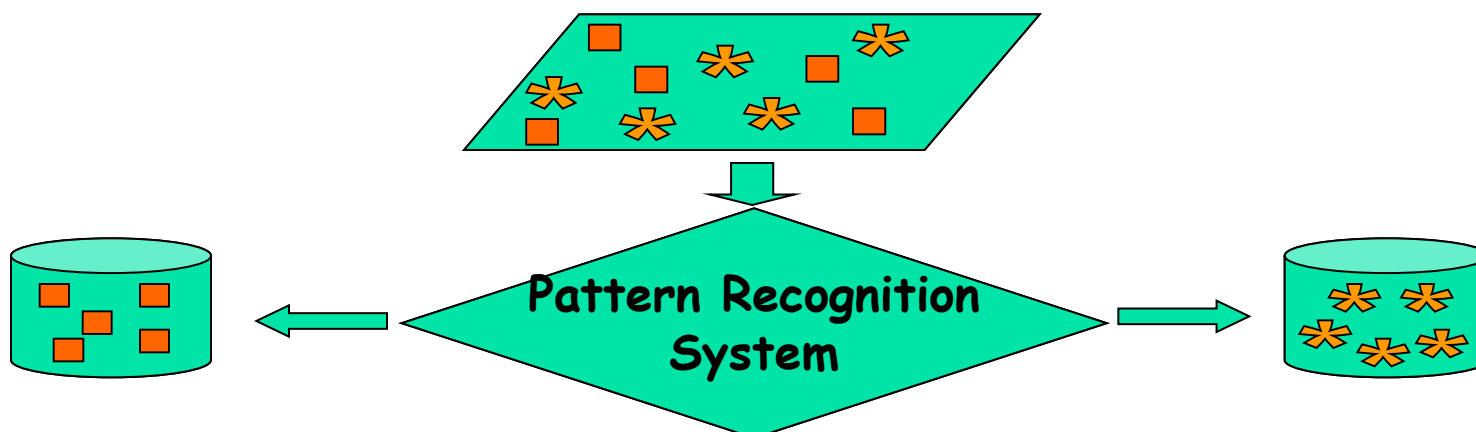
- The act of recognition can be divided in two broad categories:
 - **Concrete** Items. (characters, pictures, objects, sounds)
 - **Spatial** Items: classification of patterns in **space**
 - fingerprints
 - weather maps
 - Pictures
 - **Temporal** Items: classification of patterns in **time**
 - Electrical activity produced by the brain
 - Radar Signatures.
 - Sounds and Music
 - **Abstract** Items (solution of a mathematical problem or a philosophical question)
 - Involves the recognition of a solution to a problem, In other words, recognizing items that do not exist physically.



A basic pattern classification system contains:

- A sensor
- A preprocessing mechanism
- A feature extraction mechanism (manual or automated)
- A classification algorithm
- A set of examples (training set) already classified or described

PR Applications (1/2)



1. Typical Pattern Classification Model

Application	Inputs	Outputs
Optical Character Recognition Speech Recognition Weather Forecast Medical Diagnosis Financial Applications	Character Image Audio Signal Satellite Images Symptoms Financial Data	Character Word Weather Prediction Disease Financial Forecast

2. Pattern Recognition Applications



PR applications (2/2)

- Image processing
- Computer vision
- Speech recognition
- Multimodal interfaces
- Automated target recognition
- Optical character recognition
- Seismic analysis
- Fingerprint identification
- Radar signal classification/analysis
- Face recognition
- Character recognition
- Handwriting analysis
- Industrial inspection
- Financial forecast
- Medical diagnosis
- ECG signal analysis



More Applications (1/3)

□ Speech Recognition:

Converts spoken words into machine readable input.

Microphone interface module makes ideal accessories for Human Computer Interaction

□ Optical Character Recognition - OCR

Translation of images of handwritten, typewritten or printed text

➤ HandWritten Character Recognition

- off line from a piece of paper by optical scanning (OCR).
- on line sensing the movements of a pen tip

□ Machine Vision:

Mass surveillance systems incorporating recognition techniques on data extracted from images.

Example: Automatic number plate recognition on vehicles.



More Applications (2/3)

- ❑ **Medical Diagnosis:** Evaluation in diagnostic hypothesis. Ability to cope with uncertainties and errors in medical information.
 - **Automatic analysis** of medical image, X-ray images, tomography, ultrasound scans etc.
 - **Clustering** of electroencephalograms, cardiograms, scan-detection for genetic irregularities in chromosomes.
- ❑ **Geographical Integration Systems:** Automated analysis of satellite imagery, location of crop diseases, detection of ancient settlements, land use, atmospheric conditions, fossil mineral detection.



More Applications (3/3)

- ❑ **Industrial Applications:** Quality inspection and control, inspection in electronics industry
- ❑ **Economic and Monetary:** detection of irregular transactions through credit card, clustering of loan requests, stock market prediction
- ❑ **Data mining:** search engines, content based image and sound retrieval from large databases



Pattern Recognition as a Decision-Making Problem

The recognition process can be mathematically represented as:

$$f(\mathbf{x}) \rightarrow \omega_i$$

Where:

$\mathbf{x}$: Feature vector representing the input pattern.

f : Recognition function or classifier.

ω_i : Assigned class.

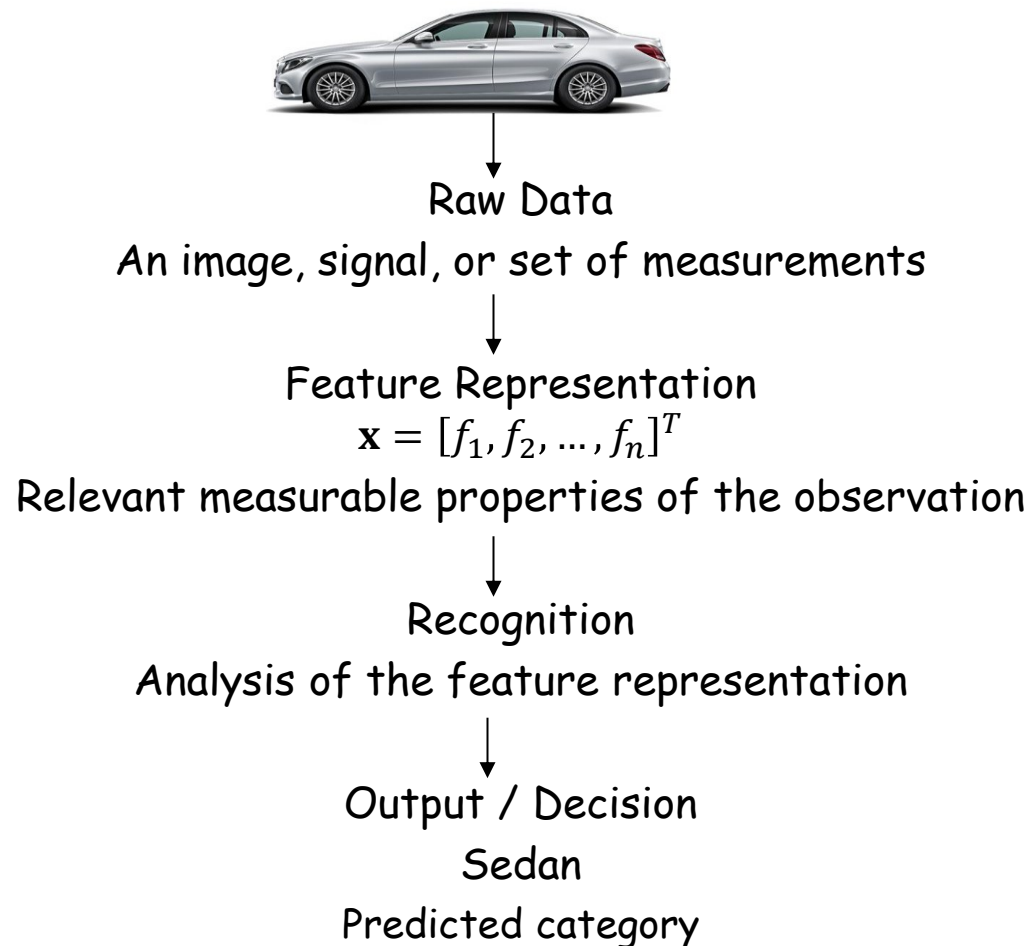
Example: Vehicle Recognition

Given a side-view image of a vehicle, a recognition system may assign it to one of the following classes:

- ❑ Sedan
- ❑ SUV
- ❑ Hatchback

From Raw Data to Decision

A Pattern Recognition system transforms raw observations into a representation that can be used to make decisions.





Historical Reference (1/2)

■ Fundamental elements of Pattern Recognition:

➤ Plato and Aristoteles:

Among the Pioneers to draw the discriminating between

- **Essential attribute** (shared among the members of a category)
- **Non essential attribute** (different members)
- **Pattern Recognition:** Procedures to detect essential attributes in a category of objects.

➤ **Aristoteles:** Constructed a clustering system to arrange animals. The system was based in the blood colour.

- **Red Colour** -> Vertebrate
- **All Other Colours** -> Invertebrate.
- Further clustering involved subcategories derived from the two main categories.



Historical Reference (2/2)

- ❑ **Theophrastos** made a relative clustering system for plants
Categorization
- ❑ **Carolus Linnaeus** constructed more systemic taxologies about animals, plants, stratum and diseases, bringing into play, state of the art knowledge.
- ❑ **Hertzprung, Russell:** Taxonomy about stars
Two Variables:
 - Brightness
 - Temperature.
- ❑ First systemic effort for mathematical formulation, **Fisher**, 1936.

Ivan Petrovich Pavlov (1/2)

- ❑ Ivan Petrovich Pavlov (1849-1936) was a scientist whose study of the digestive system led him to study reflexes as well
- ❑ Famous example of Pavlov's dog





Ivan Petrovich Pavlov (2/2)

The Experiment and Discovery

- ❑ **The setup:** Ivan Pavlov studied digestion in dogs in the late 1890s and early 1900s.
- ❑ **The observation:** He noticed that dogs began to salivate not only when food was placed in their mouths, but also at the sight of food, lab coats, or the technicians who fed them.
- ❑ **The conditioning:** Pavlov paired a neutral stimulus—such as the ringing of a bell or a metronome—with the presentation of food.
- ❑ **The result:** After repeated pairings, the dogs learned to associate the sound with food and would salivate upon hearing the bell alone, even when no food was given.



Pavlovian Generalization (1/2)

- Further studies were done in the style of Pavlov's dog, and as long as stimulus S was given, the reaction R would be the same
- Then, if a stimulus similar to S , S' was given instead, R would be the same
- This shows a different type of pattern recognition: the similarity between S and S' was recognized and generalized so that the same output, R , was given



Pavlovian Generalization (2/2)

Meaning and Legacy

- ❑ **Classical conditioning:** This foundational concept shows how a biological response can become tied to a previously unrelated environmental signal.
- ❑ **Behavioral science:** It helped lay the groundwork for modern behavioral psychology, influencing areas like habit formation, learning theories, and phobia treatments.
- ❑ **Everyday usage:** The phrase "Pavlov's dog" is also used to describe a person who reacts automatically or instinctively to a cue without thinking



From Observation to Recognition

- ❑ A pattern is first **observed** through some form of data.
- ❑ The observation must be represented in a way that allows analysis.
- ❑ Relevant characteristics can be identified and used for recognition.
- ❑ Recognition requires a way to compare the observed pattern with other patterns.



Different Ways to Recognize Patterns

- ❑ Patterns can be described in **different ways**.
- ❑ One approach is to represent patterns using **measurable features**.
- ❑ Another approach is to describe patterns through their **structure and relationships**.
- ❑ The choice of representation determines how recognition can be performed.

Two fundamental viewpoints:

- ❑ **Statistical** → *What can we measure about the pattern?*
- ❑ **Syntactic / Structural** → *How is the pattern organized?*

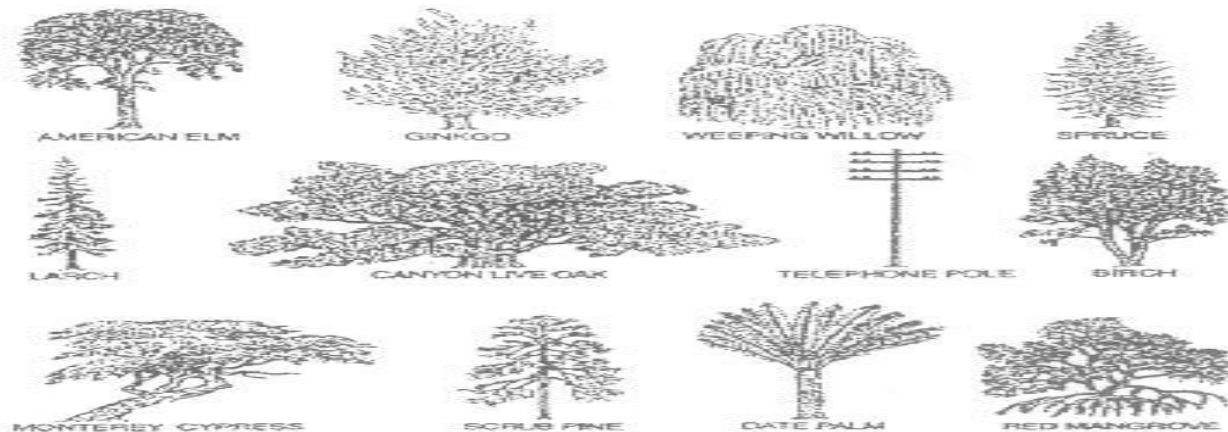


PR Methodologies

- ❑ Basically two methodologies
 - **Statistical Pattern Recognition:** clustering based in statistical analysis of objects and features
 - Extraction of intrinsic characteristics
 - Feature vector formation
 - Mathematical - statistical methods, linear algebra, probability theory.
 - **Syntactic Pattern Recognition:** pattern structures which can take into account more complex interrelationships between features than simple numerical
 - Sophisticated hierarchical descriptions
 - Decision trees, logical and grammatical rules
 - **Final Result:** series of rules describing a clustering process or grammar describing the object.

Syntactic PR

- ❑ **Syntactic Methodologies:** complex and sensitive to noise, slight variations, missed or incomplete information
 - Can be used as alternative in cases statistical methodologies are not suitable or applicable.
 - In cases that pattern description related to a problem is obscure, doubtful, or not fully specified.
 - Logical Rules to cluster trees;





Syntactic vs Statistical PR

❑ Statistical Pattern Recognition:

- Strong mathematical foundation. Number of elements and order of the elements of an object feature vector is always fixed.

❑ Syntactic Pattern Recognition: Based mostly in logical and/or intuition rules

- The number and order of the elements corresponding to a feature vector varies between the population of patterns



Focus of This Course

- ❑ The course focuses primarily on **statistical pattern recognition**.
- ❑ Patterns are represented using **feature vectors**.
- ❑ Recognition is formulated as a **decision or classification problem**.
- ❑ Statistical methods provide the foundation for most of the techniques studied in this course.
- ❑ Syntactic approaches are introduced as an alternative perspective on pattern representation and recognition.



Fields of Science related to PR

Pattern Recognition is an interdisciplinary field that draws concepts and techniques from:

- ❑ Statistics & Probability — uncertainty, distributions, inference, decision-making
- ❑ Machine Learning
- ❑ Artificial Neural Networks
- ❑ Computer Vision
- ❑ Speech recognition
- ❑ Cognitive Science & Neuroscience — human perception and recognition
- ❑ Mathematics — vectors, matrices, optimization, geometry
- ❑ Signal & Image Processing — representation and transformation of data

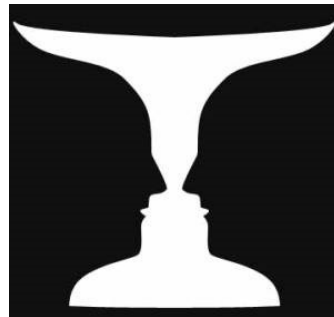


What Is a Pattern? (1/5)

- ❑ Watanabe describes a pattern as the opposite of chaos because it represents the extraction of order and minimum entropy from otherwise random or unorganized data
- ❑ An entity
- ❑ Anything that could be given a name or a specific description
- ❑ Any image that we recognize is a pattern
- ❑ Two or more patterns can exist within an image or thing
- ❑ Humans can only actively see one pattern at a time

What Is a Pattern? (2/5)

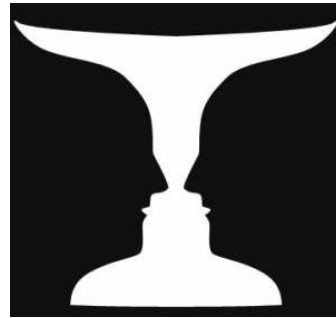
Examples of this are visual illusions



What do you see in the image?

This image is a famous optical illusion known as Rubin's vase developed by Danish psychologist Edgar Rubin.

What Is a Pattern? (3/5)



Because your brain cannot process a single outline as two separate boundaries simultaneously, it alternates your awareness between two distinct patterns:

- The White Vase/Chalice: If you perceive the white shape as the figure, your brain treats the black area as empty background space.
- The Two Black Faces: If you perceive the two side-profile faces looking at each other as the figure, your brain treats the white center as the background.



What Is a Pattern? (4/5)

- ❑ A **pattern** is an identifiable regularity, structure, or characteristic in observed data.
- ❑ A pattern may correspond to an **object, event, signal, behavior, or class**.
- ❑ Patterns can be recognized because they share **common characteristics**.
- ❑ The characteristics may be numerical, structural, spatial, temporal, or symbolic.
- ❑ A pattern does not have to be identical to another pattern to belong to the same class.



What Is a Pattern? (5/5)

Examples:

- ❑ A handwritten digit → visual pattern
- ❑ A speech signal → temporal pattern
- ❑ A fingerprint → biometric pattern
- ❑ An ECG signal → physiological pattern
- ❑ A customer's transaction history → behavioral pattern



Features & Patterns (1/3)

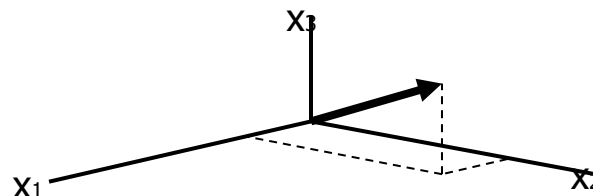
□ Feature

Feature is any distinctive aspect, quality or characteristic. Features may be symbolic (i.e., color) or numeric (i.e., height).

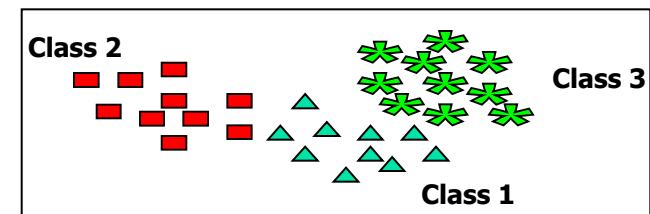
- The combination of n features is represented as a n -dimensional column vector called a **feature vector**
- The n -dimensional space defined by the feature vector is called the **feature space**
- Objects are represented as points in feature space. This representation is called a **scatter plot**

$$X = [x_0, x_1, \dots, x_n]$$

1. Feature Vector



2. Feature Space (3D)



3. ScatterPlot (2D)



Features & Patterns (2/3)

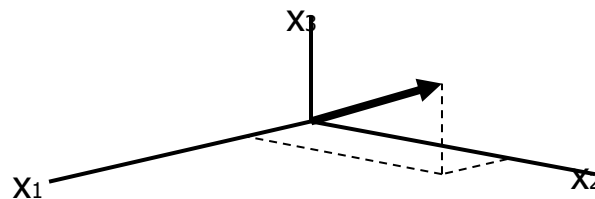
- Different patterns may share some features but differ in others.
- The choice of features strongly affects the recognition process.
- Features should capture information that is **relevant for distinguishing patterns**.

Example — Vehicles:

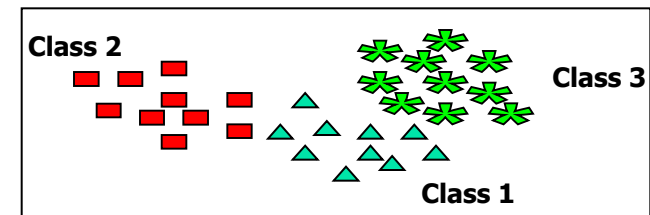
- Length
- Height
- Number of wheels
- Aspect ratio
- Color

$$X = [x_0, x_1, \dots, x_n]$$

1. Feature Vector



2. Feature Space (3D)

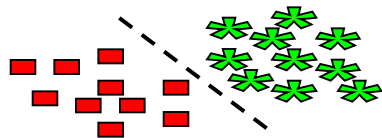


3. ScatterPlot (2D)

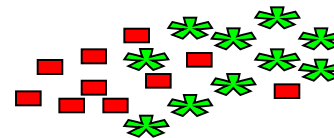
Features & Patterns (3/3)

❑ What makes a “good” feature vector?

- The quality of a feature vector is related to its ability to discriminate examples from different classes
- Examples from the same class should have similar feature values
- Examples from different classes have different feature values



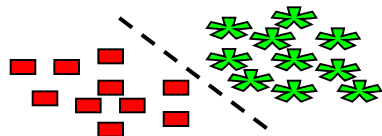
1. "Good" Features



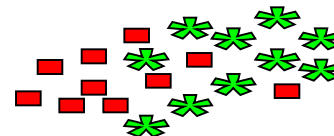
2. "Bad" Features

Classification in Feature Space (1/2)

- Once patterns are represented by features, they can be viewed as points in a **feature space**.
- Each pattern corresponds to a point in the feature space.
- Different classes may form different **regions or groups**.
- Classification aims to assign each new pattern to the appropriate class.
- The geometry of the feature space can help us understand the classification problem.



1. "Good" Features



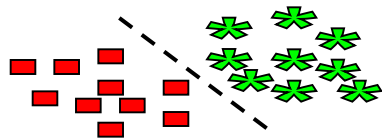
2. "Bad" Features

Classification in Feature Space (2/2)

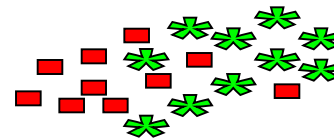
Example: Vehicle Classification

- ❑ x_1 : vehicle length
- ❑ x_2 : vehicle height
- ❑ Each vehicle $\rightarrow$ one point in the 2D feature space
- ❑ Different vehicle types $\rightarrow$ different regions

Feature representation transforms a recognition problem into a problem that can be studied geometrically.



1. "Good" Features



2. "Bad" Features



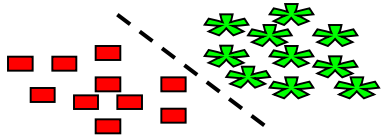
Decision Boundaries (1/2)

A **decision boundary** separates regions of the feature space associated with different classes.

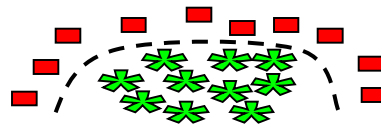
- ❑ A boundary defines where the classification decision changes from one class to another.
- ❑ It can be **linear** or **nonlinear**.
- ❑ A good boundary should separate patterns belonging to different classes.
- ❑ The position and shape of the boundary depend on the characteristics of the data and the classification method.

Decision Boundaries (2/2)

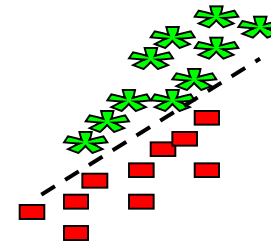
- More complex models result in more complex boundaries



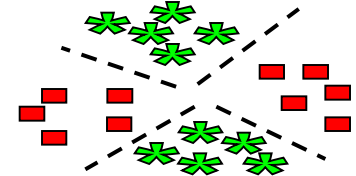
1. Linear separability



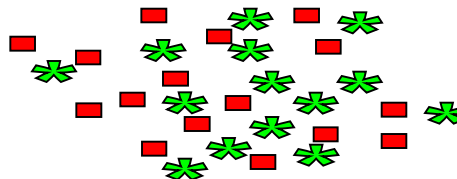
2. Non-linear separability



3. Correlated features



4. Multi-modal



What can be done if data cannot be separated with a hyperplane?



What Is a Classifier? (1/5)

A **classifier** is a method or model that assigns an observed pattern to one of a set of predefined classes.

- ❑ Input: a pattern represented by a **feature vector**
- ❑ Output: an assigned **class label**
- ❑ The classifier uses a **decision rule** to make the assignment.
- ❑ The decision rule determines how the feature space is divided into different **decision regions**.

Conceptual flow: Pattern→Feature Vector→Classifier→Class Label



Classifiers (2/5)

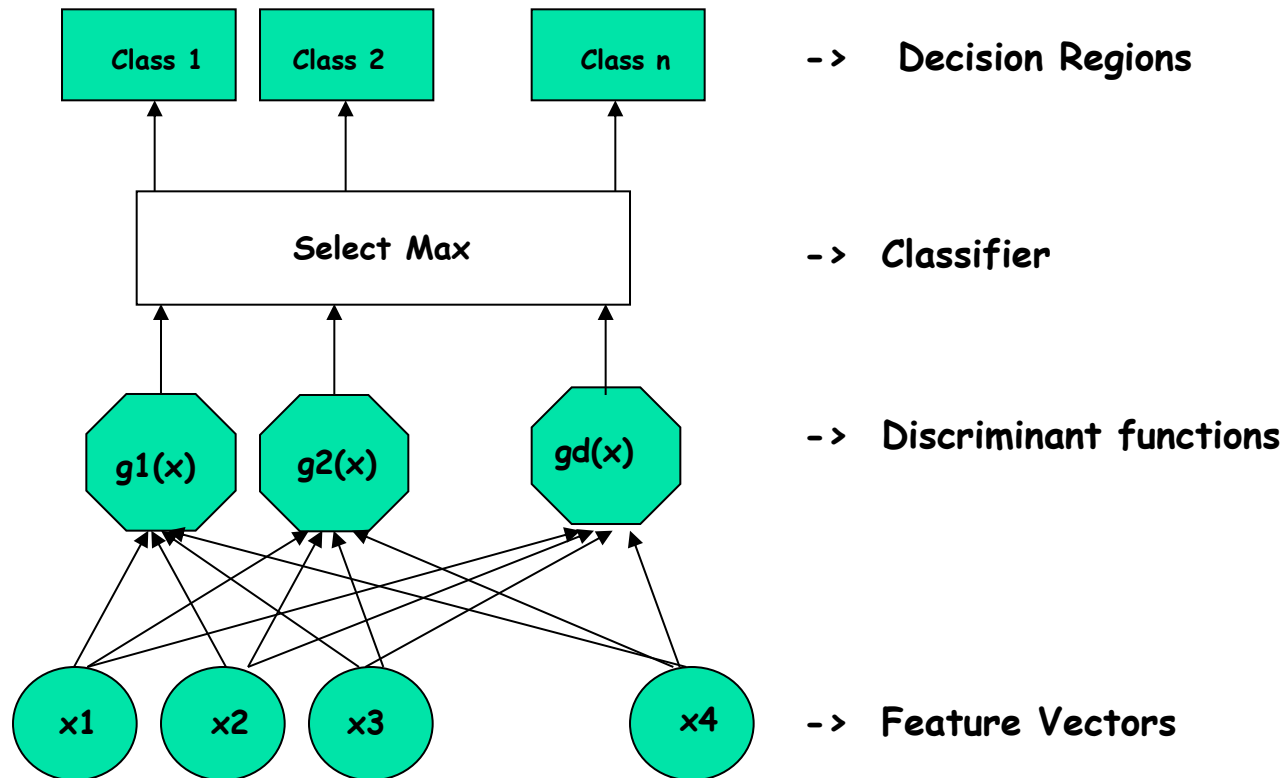
- The classification of feature vector $\mathbf{x}$ consists of determining which decision region it belongs to, and assign $\mathbf{x}$ to this class

Example:

- $\mathbf{x} \rightarrow \text{Classifier} \rightarrow \text{Car} / \text{Truck} / \text{Motorcycle}$
- A classifier can be represented as a set of discriminant functions

The classifier assigns a feature vector $\mathbf{x}$ to class w if $g_j(\mathbf{x}) > g_i(\mathbf{x}) \forall j \neq i$

Classifiers (3/5)





Types of Classifiers (4/5)

Classifiers can differ according to the way they make decisions.

- ❑ **Distance-based** classifiers
Use similarity or distance between patterns.
- ❑ **Probabilistic / Statistical** classifiers
Use probability models and statistical decision rules.
- ❑ **Linear** classifiers
Use linear decision boundaries in the feature space.
- ❑ **Nonlinear** classifiers
Can represent more complex decision boundaries.



Types of Classifiers (5/5)

Different classifiers make different assumptions about the data and the structure of the classes.

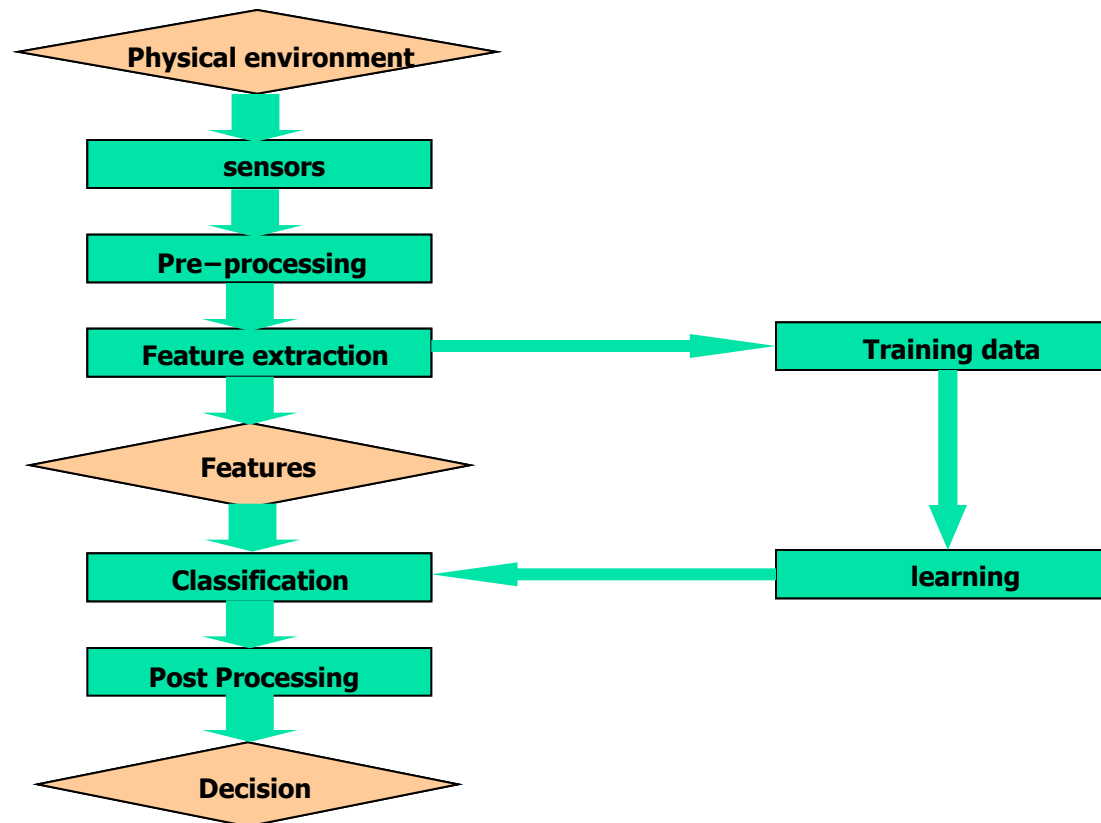
Examples studied later in this course:

□ **k-NN → Bayesian Classification → Logistic Regression → SVM → Decision Trees**

Different methods, same fundamental goal: **assign patterns to the appropriate classes.**



PR Systems



Process Diagram for typical Pattern Recognition System



Components of a PR system

- ❑ Sensorial Data
 - Important Issues
 - Noise
 - Bandwidth
 - Sensitivity
- ❑ Pre-processing
 - Noise Cancelation
 - Signal conditioning
- ❑ Feature extraction
 - build feature vector
- ❑ Learning
 - Build decision regions based on a training set of feature ventors
- ❑ Classification
 - Use the decision regions to map evaluation feature vectors
- ❑ Post Processing
 - Evaluation
 - Optimization



Design Cycle



□ Data Collection

- Collect training and evaluation information
- ➤ **But** difficult to determine appropriate number of samples

□ Feature Selection

- Computational cost (multidimensional vectors)
- Discriminative features depend on prior knowledge
- Translation or rotation invariant features
- Robust features with respect to partial occlusions, distortions or deformations



Design Cycle

❑ Model Selection

- Design criteria and requirements
- Missing or incomplete patterns
- Computational complexity
- Syntactic or structural

❑ Train Classifier

- Supervised training: a teacher dictates the correct cluster
- Unsupervised training: automatic cluster forming
- Reinforcement learning: no a-priori categories, system feedback provides the decision for right or wrong

❑ Evaluate Classifier

- Estimation of the performance with non training data
- Performance prediction with future data
- Problems of overfitting and generalization



Learning in PR (1/3)

- ❑ Any method that incorporates information from training samples in the design of a classifier employs learning.
- ❑ We use learning because all practical or interesting PR problems are so hard that we cannot guess classification decision ahead of time.
- ❑ **Approach:**
 - Assume some general form of model
 - Use training patterns to learn or estimate the unknown parameters.



Learning in PR (2/3)

□ Supervised Learning

➤ Teacher provides a label or cost for each pattern in a training set.

➤ Objective: Reduce the sum of the costs for these patterns

□ . ➤ Issues: How to make sure that the learning algorithm

- can learn the solution.

- Will be stable to parameter variation.

- Will converge in finite time.

- Scale with # of training patterns & # of input features.

- Favors "simple" solutions



Learning in PR (3/3)

❑ **Unsupervised Learning (Clustering)**

- There is no explicit teacher.
- System forms clusters or "natural grouping" of the input patterns.

❑ **Reinforcement Learning (Learning with a critic)**

- No desired category is given. Instead, the only teaching feedback is that the tentative category is right or wrong.
- Typical way to train a classifier:
 - Present an input
 - Compute its tentative label
 - Use the known target category label to improve the classifier.



Adaptation in Pattern Recognition (1/2)

In many applications, the characteristics of the data may **change over time**.

- ❑ Sensors and acquisition conditions may change.
- ❑ The environment may change.
- ❑ The statistical properties of the data may change.

A fixed recognition system may therefore become less effective.

Adaptation refers to modifying or updating a recognition system in response to such changes.



Adaptation in Pattern Recognition (2/2)

Example:

A face recognition system may encounter changes in:

- ❑ illumination
- ❑ camera characteristics
- ❑ viewpoint
- ❑ appearance over time

An adaptive system can adjust its behavior as the operating conditions change.



Learning and Adaptation in PR

Learning and adaptation address different aspects of Pattern Recognition.

<i>Learning</i>	<i>Adaptation</i>
Derives a model from data	Updates the system when conditions change
Usually associated with training	Can continue during operation
Establishes the initial recognition capability	Maintains or improves performance over time

Data → Learning → Initial Model

Changing Environment → Adaptation → Updated Model

A recognition system may need both learning and adaptation to perform reliably in real-world environments.



The subproblems of PR (1/3)

A Pattern Recognition system can be divided into several interconnected subproblems:

- ☐ Data Acquisition
 - ☐ How should the observations be collected?
- ☐ Preprocessing
 - ☐ How can noise, artifacts, and unwanted variability be reduced?
- ☐ Feature Representation
 - ☐ Which characteristics provide useful information about the patterns?
- ☐ Learning / Model Construction
 - ☐ How can a recognition rule be learned from available data?
- ☐ Classification / Decision
 - ☐ How should patterns be assigned to classes?
- ☐ Evaluation
 - ☐ How can we determine whether the system performs well?



The subproblems of PR (2/3)

❑ Invariants:

- Translation invariant: absolute position on conveyor belt is irrelevant. Orientation invariant, size invariant, etc...

❑ Evidence Pooling:

- Can design several classifiers and combine them.
- How to pool the evidence to achieve the best decision?

❑ Costs and Risks:

- A classifier is used to recommend an action, and each action has an associated cost or risk.
- A classifier might be designed to minimize some total expected cost or risk.



The Subproblems of PR (3/3)

- ❑ How to incorporate knowledge about such risks, and how will they affect the classification decision?
 - Can we estimate the lowest possible risk of any classifier, to see how close ours meet this ideal?
- ❑ **Computational Complexity:**
 - How an algorithm scales as a function of the
 - *feature dimensions?*
 - *what Features?*
 - *what categories?*
- ❑ What is the tradeoff between computational ease & performance?



Summary

- ❑ Pattern recognition techniques find applications in many areas: machine learning, statistics, mathematics, computer science, biology, etc.
- ❑ There are many sub-problems in the design process.
 - ❑ Many of these problems can indeed be solved.
- ❑ More complex learning, searching and optimization algorithms are developed with advances in computer technology.
 - ❑ There remain many fascinating unsolved problems



Summary

Key Takeaways

- ❑ Pattern Recognition is an interdisciplinary field.
- ❑ Two classical perspectives are:
 - Statistical Pattern Recognition
 - Syntactic / Structural Pattern Recognition
- ❑ Recognition systems can learn from data and may adapt to changing conditions.
- ❑ PR system design is an iterative process.
- ❑ The different stages of a PR system are strongly interconnected.
- ❑ The central idea:
Represent patterns appropriately → learn or define a decision rule → recognize new patterns → evaluate and refine the system.



Questions ?



References

□ Journals

- Journal of Pattern Recognition Society.
- IEEE transactions on Neural Networks.
- Pattern Recognition and Machine Learning.

□ Books

- Duda, Hart: Pattern Classification and Scene Analysis. J. Wiley & Sons, New York, 1982. (2nd edition 2000).
- Fukunaga: Introduction to Statistical Pattern Recognition. Academic Press, 1990.
- Bishop: Neural Networks for Pattern Recognition. Clarendon Press, Oxford, 1997.
- Schlesinger, Hlaváč: Ten lectures on statistical and structural pattern recognition. Kluwer Academic Publisher, 2002.
- Sato Watanabe Pattern Recognition: Human and Mechanical, Wiley, 1985
- E. Gose, R. Johnsonbaugh, S. Jost, Pattern recognition and image analysis, Prentice Hall, 1996.
- Sergios Thodoridis, Kostantinos Koutroumbas, Pattern recognition, Academic Press, 1998.