

Tutorial 07 for Fundamentals of Artificial Intelligence

Hualie Jiang

School of Science and Engineering
The Chinese University of HongKong, Shenzhen

March 20, 2019



香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Problem 1

1. Answer the following questions:

- (a) What is “Pattern”? According to its definition in L8: P4-7, could you list some patterns in your daily life?
- (b) “Pattern Recognition (PR)” is defined in L8: P8. Could you give some PR samples and tell which features you use?
- (c) PR is an important component in both ML(Machine Learning) and AI. Please explain why? (L8: P9-10)



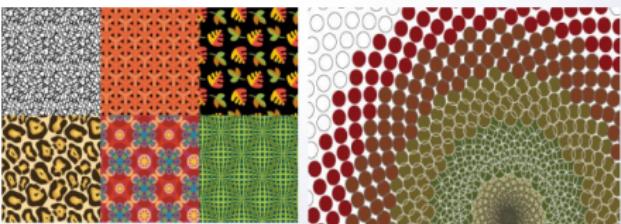
香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

What is Pattern

- A “pattern” is **the form of representation of an objectively existed event or object**. For instance, voice, image and character are patterns.
- There are many kinds of patterns -- visual patterns, temporal patterns, logical patterns. More broadly, *any natural and social phenomenon* may be considered as “Patterns”.
- “Pattern” is **a set of measurements or observations, represented in vector or matrix notation (in mathematics)**.

→ “ We need to seek classification, recognition, or description of a pattern that is **INVARIANT** to some (known) changes or deviation in the pattern from the ‘ideal’ case.



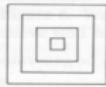
香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

What is Pattern



"Nominal" pattern



Dilation (magnification)

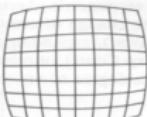
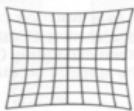
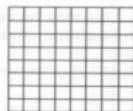


Displacement (translation)

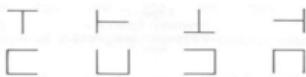


Rotation

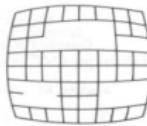
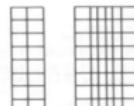
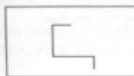
(a) Image pattern



(b) Grid pattern



(c) Character pattern



(d) More extreme pattern

Pattern Distortion

- In many situations a set of patterns from the same class may exhibit wide variations from a single exemplar of the class.
- Invariant features: rotated, scaled, and translated invariant.



香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

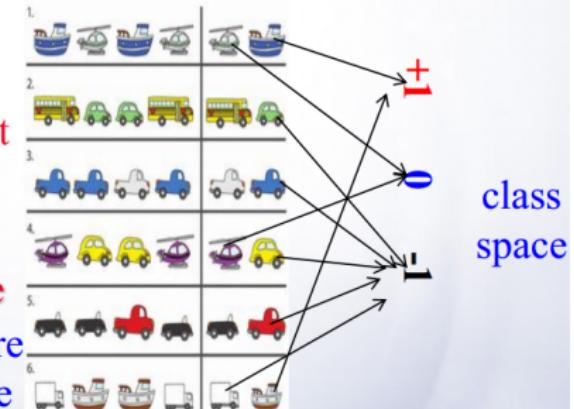
Chinese Character Pattern

苞 胞 包 襄 剥 薄 宦 保 堡 饱 宝 抱 报 暴 猊 鲍 爆 林
碑 悲 驯 执 壅 背 贝 钮 信 狐 备 意 焙 被 奔 苞 本 皋
崩 翊 角 𠵼 蹤 逼 驸 比 部 笔 狼 碧 蘭 蔽 华 峯 峰
帀 庶 痘 闭 敝 幦 必 辟 壁 臂 避 陞 鞍 边 编 蔽 扉 便
叟 卜 辨 辨 辨 遍 林 彪 膻 表 燮 焰 别 瘑 溪 斓 澜
宾 摈 兵 冰 柄 丙 乘 饼 烛 瞭 痘 并 破 菩 墁 援 鮮 波
博 勃 搏 锅 菲 伯 帛 舶 脖 脖 沽 泊 驳 帕 卜 布 补 帛
不 布 步 薄 部 怖 掇 猜 裁 材 才 财 眇 践 采 彩 菜 菜

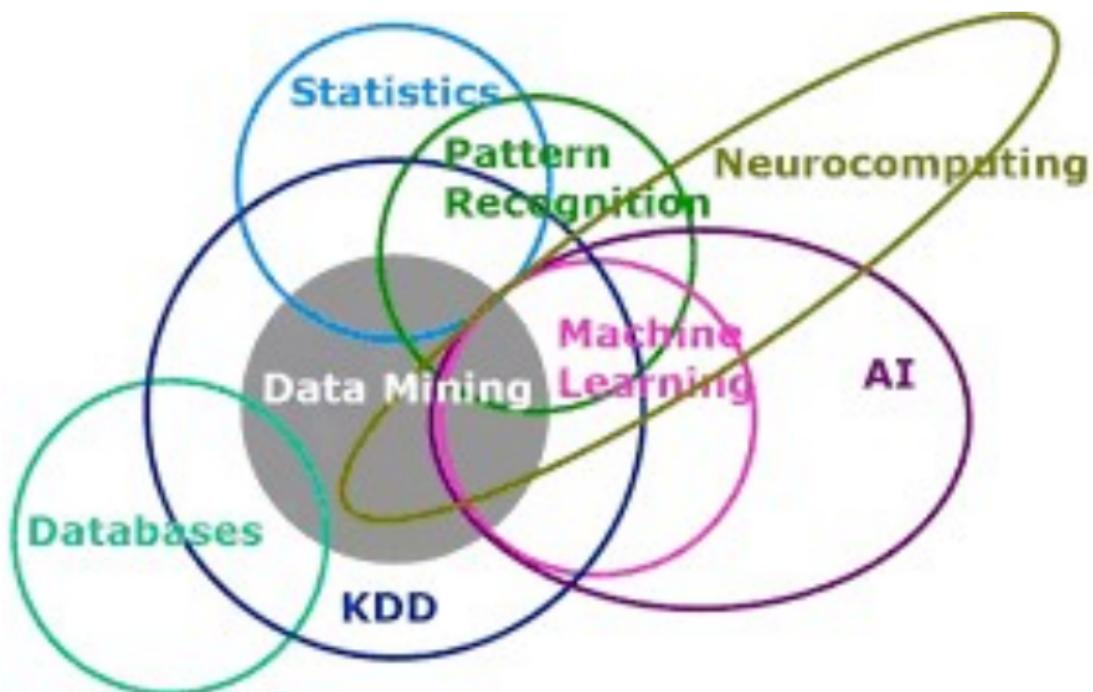


Pattern Recognition

- A basic intelligent ability of human being or animal; Using a broad enough interpretation, we can find PR in every intelligent activity. For instance, you are able to recognize the number of classroom, which is the ability of *number recognition*; on the class you have to be able to understand what the teacher says and writes on the blackboard, this is the ability of *speech and character recognition*.
- From the system viewpoint, PR is an important component of intelligent systems.
- From the theoretical concept, PR is a mapping from **feature space** to **class space**.



Pattern Recognition



香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

Problem 1

1. Answer the following questions:

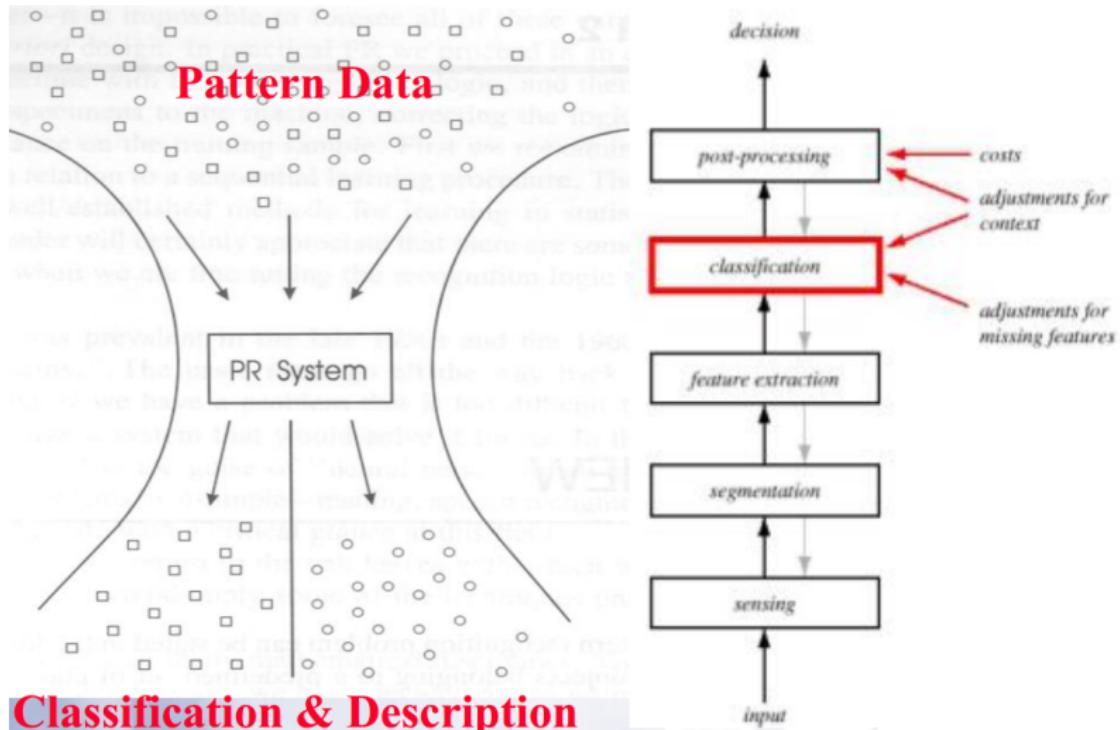
- (d) In P11-13, a PR system is given. Please understand each function in the system.
- (e) Understand the definitions related to PR: Classification, Recognition, Description, Pattern Class and Preprocessing (L8: P14-15). Why classification and recognition require the different classes, c and $c+1$, respectively.
- (f) There are the main two parts, feature extraction/selection and classifier design, in a PR model. Please understand each basic contents.
(L8: P19-28)



香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

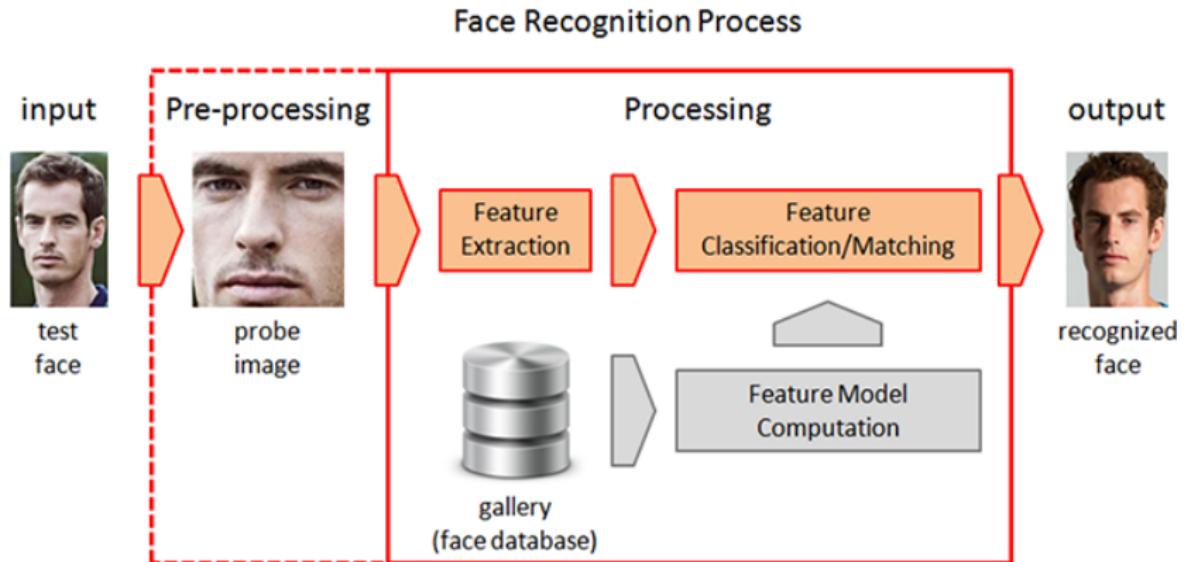
Pattern Recognition System



What is PR System?

- The design of a PR system essentially involves the following three aspects :
 - 1) data acquisition and preprocessing;
 - 2) data representation;
 - 3) decision making.
- The problem domain dictates the choice of sensor, preprocessing technique, representation scheme, and the decision making model.
- It is generally agreed that a **well-defined** and **sufficiently constrained recognition problem** (small intraclass variations and large interclass variations) will lead to a compact pattern representation and a simple decision making strategy.

Face Recognition System

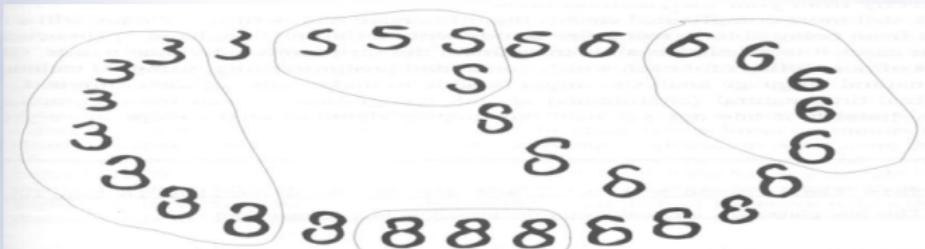


香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

Definitions

- **Classification** - Assigns input data into one of c pre-specified classes based on extraction of significant features and the processing or analysis of these features. It is common to resort to probabilistic or grammatical models in classification. A **Classifier** partitions feature space into class-labeled decision regions.
- **Recognition** is the ability to classify. Often we formulate PR with a $c + 1$ st class, corresponding to the ‘unclassifiable’ or ‘can’t decide’ class.



Definitions

- Pattern class - A set of patterns (hopefully sharing some common attributes) originated from the same source. The key in many PR applications is to identify suitable attributes (e.g., features) and to form a good measure of similarity and an associating matching process.
- Preprocessing is the filtering or transforming of the raw input data to get discriminative feature and minimize noise.
- Description is an alternative to classification where a structural description of the input pattern is desired. It is common to resort to linguistic or structural models in description.

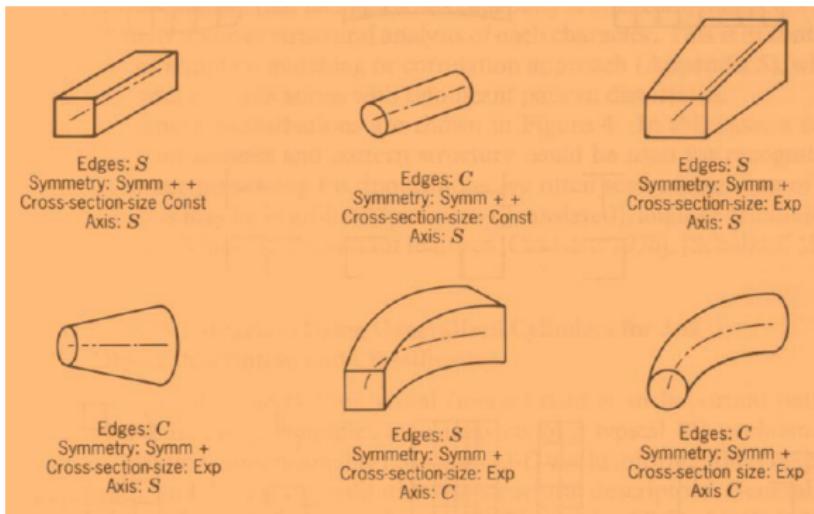


香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

Problem 2

2. A 3-D example for pattern extraction is defined in L8: P16-18. There are four kinds of features are extracted. Please select these features according to some guidelines (L8: P27-28) to classify the given six objects. (One selection: Edge + Symmetry + Axis)



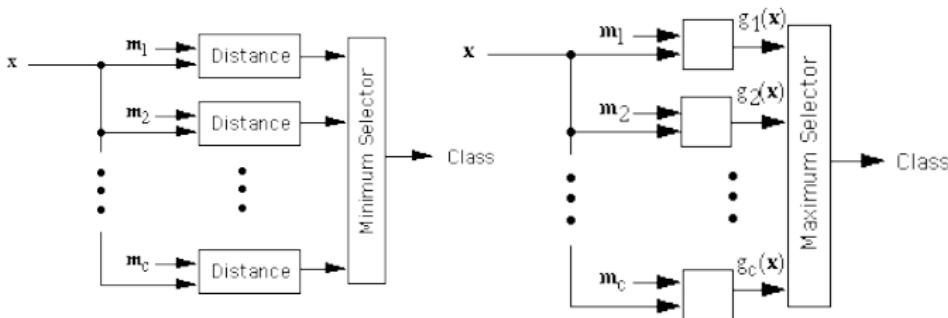
Problem 2

$$[1, 2, 3, 4, 5, 6] \left\{ \begin{array}{l} Edges : S \rightarrow [1, 3, 5] \\ Edges : C \rightarrow [2, 4, 6] \end{array} \right. \left\{ \begin{array}{l} Symmetry : ++ \rightarrow [1] \\ Symmetry : + \rightarrow [3, 5] \\ Symmetry : ++ \rightarrow [2] \\ Symmetry : + \rightarrow [4, 6] \end{array} \right. \left\{ \begin{array}{l} Axis : S \rightarrow [3] \\ Axis : C \rightarrow [5] \\ Axis : S \rightarrow [4] \\ Axis : C \rightarrow [6] \end{array} \right.$$



Problem 3

3. A minimum-error classifier based on $\|x - m_k\|$ is defined in L8: P41. How to get a linear discriminant function, $g(x) = m'x - 0.5\|m\|^2$ in L8: P43? Which difference between the following two classifiers using the approaches mentioned above.



香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

Problem 3

As the solution given in P10:43, using the inner product to express the Euclidean Distance from \mathbf{x} to \mathbf{m}_k , we can write

$$\|\mathbf{x} - \mathbf{m}_k\|^2 = (\mathbf{x} - \mathbf{m}_k)^\top (\mathbf{x} - \mathbf{m}_k) = \mathbf{x}^\top \mathbf{x} - \mathbf{m}_k^\top \mathbf{x} - \mathbf{x}^\top \mathbf{m}_k + \mathbf{m}_k^\top \mathbf{m}_k,$$

Note that $\mathbf{m}_k^\top \mathbf{x} = \mathbf{x}^\top \mathbf{m}_k$,

$$\text{So, } \|\mathbf{x} - \mathbf{m}_k\|^2 = -2(\mathbf{m}_k^\top \mathbf{x} - 0.5\mathbf{m}_k^\top \mathbf{m}_k) + \mathbf{x}^\top \mathbf{x}$$

For each k , $\mathbf{x}^\top \mathbf{x}$ is the same.

So the $\mathbf{m}_k^\top \mathbf{x} - 0.5\mathbf{m}_k^\top \mathbf{m}_k$ term can be extracted as the discriminant function

$$g_k(\mathbf{x}) = \mathbf{m}_k^\top \mathbf{x} - 0.5\mathbf{m}_k^\top \mathbf{m}_k = \mathbf{m}_k^\top \mathbf{x} - 0.5\|\mathbf{m}_k\|^2.$$

Therefore, $\|\mathbf{x} - \mathbf{m}_k\|^2 = -2g_k(\mathbf{x}) + \mathbf{x}^\top \mathbf{x}$. To find the template \mathbf{m}_k that minimizes $\|\mathbf{x} - \mathbf{m}_k\|^2$ is equivalent to find the \mathbf{m}_k that maximizes $g_k(\mathbf{x}) = \mathbf{m}_k^\top \mathbf{x} - 0.5\|\mathbf{m}_k\|^2$.

Write \mathbf{m}_k as \mathbf{m} , we get a linear discriminant function: $g(\mathbf{x}) = \mathbf{m}^\top \mathbf{x} - 0.5\|\mathbf{m}\|^2$.

Minimum-distance classifier measures the distance from a feature vector \mathbf{x} to the templates $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_c$ and assigns \mathbf{x} to the class of the nearest template. In contrast, minimum-Euclidean-distance classifier classifies an input feature vector \mathbf{x} by computing c linear discriminate functions $g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_c(\mathbf{x})$ and assigns \mathbf{x} to the class corresponding to the maximum discriminant function.



Problem 4

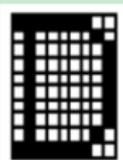
4. In L8: P33 there are two templates ("D" and "O") and ten input samples, each five. Please classify the third input example in the "D" row by using the maximum correlation approach and minimum error approach, respectively. (For Maximum Correlation Approach: "D"-Input: $82/90=91\%$ and "O"-Input: $66/90=73\%$)

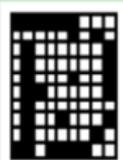
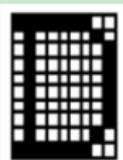


Problem 4

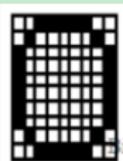
- Maximum correlation approach: Count the number of agreements (black matching black and white matching white). Pick the class that has the maximum number of agreements.

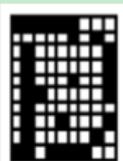
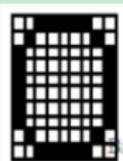
When the third template in the “D” row is input:



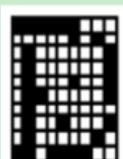
Compare  with  and count the agreements of each row and add them up to calculate the percentage.

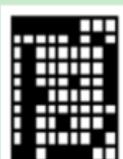
$$(8+7+9+9+8+8+8+8+9+8) / (9 \times 10) = 82 / 90 = 91\%$$



Compare  with  and count the agreements of each row and add them up to calculate the percentage.

$$(6+7+7+7+6+6+6+6+9+6) / (9 \times 10) = 66 / 90 = 73\%$$

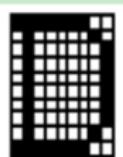


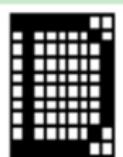
Therefore,  should be classified to the “D” class.

Problem 4

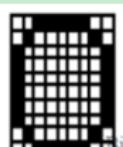
- 2) Minimum error approach: Count the number of disagreements(black where white should be or white where black should be). Pick the class that has the minimum number of disagreements.

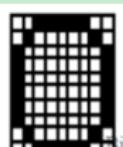
When the third template in the “D” row is input:



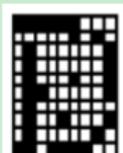
Compare  with  and count the disagreements of each row and add them up to calculate the percentage.

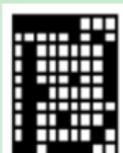
$$(1 + 2 + 0 + 0 + 1 + 1 + 1 + 1 + 0 + 1) / (9 \times 10) = 8 / 90 = 9\%$$



Compare  with  and count the disagreements of each row and add them up to calculate the percentage.

$$(3 + 2 + 2 + 2 + 3 + 3 + 3 + 3 + 0 + 3) / (9 \times 10) = 24 / 90 = 27\%$$



Therefore,  should be classified to the “D” class.