Pattern Recognition
(模式识别)

Instructor

Pengfei FANG (方鹏飞)

Email: fangpengfei@seu.edu.cn

URL: https://fpfcjdsg.github.io/

Southeast University

Spring Semester

About Me

https://fpfcjdsg.github.io/

Pengfei Fang

Ph.D., Associate Professor
Pattern Learning and Mining (PALM) Lab
School of Computer Science and Engineering
Southeast University, Nanjing, China

E-mail:fangpengfei (at) seu.edu.cn
Office: Room 527, Computer Science Building,
       Jiulonghu Campus, Southeast University
Google Scholar

Experience:

- MEng, Australian National University (ANU)

- PhD, Australian National University (ANU)

- Research Fellow, Monash University

Research Interests:

- Computer Vision

- Machine Learning

- Artificial Intelligence

Acknowledgement

This slides are created by Prof. Min-Ling Zhang, and I use his slides for the PR course this year.

Course Information

- Credits
  - 3 credits with 48 course hours and 16 project hours
  - Week 1 – Week 16, Wednesday &Thursday

- Contents
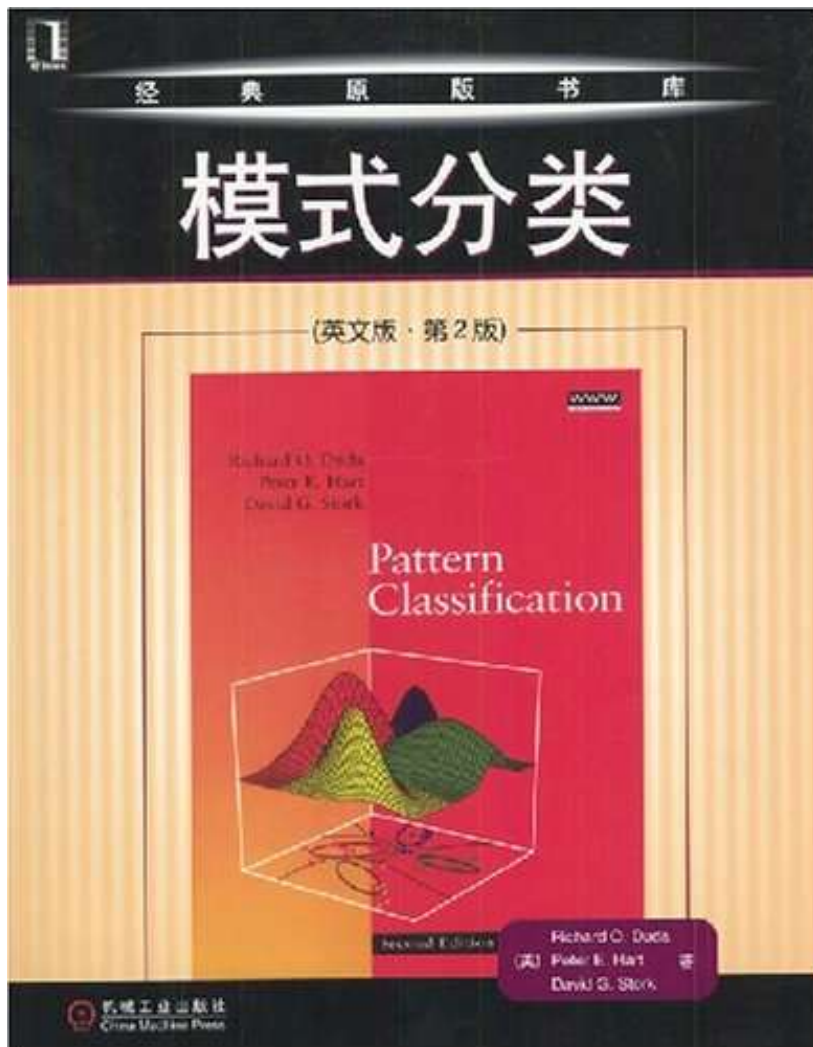 - Chapters 1-6 & 8 & 10

- Evaluation
  - In class(attendance & assignments): 10%
  - Quiz: 10%
  - Project: 30%
  - Final Exam: 50%

Textbook

Richard O. Duda, Peter E. Hart, David G. Stork

Pattern Classification, 2nd edition

John Wiley & Sons, 2001

- -

模式分类

(英文版·第2版)

机械工业出版社, 2004

References

- **Books**
  - 张学工. 模式识别(第三版). 清华大学出版社, 2010.
  - S. Theodoridis, K. Koutroumbas. Pattern Recognition, 4th edition. Elsevier Publishers, 2009.
  - C. Bishop. Pattern Recognition and Machine Learning. Cambridge University Press, 2007.

- **Web Resources**
  - International Association for Pattern Recognition (IAPR)
  - Pattern Recognition Journal (PRJ)
  - List of pattern recognition web sites

Remarks

- Mathematical background
  - **Linear algebra**
  - **Probability theory**
  - Statistics
  - Information theory

Our course isn't a mathematical one → Carefully read and comprehend materials in Appendix "Mathematical Foundation"

Remarks (Cont.)

- **No pain, No gain**

Classroom lectures are important but not enough ➡ Review what have been taught with at least 4~6 hours per week

- **Terminologies and Contents**

It's important to understand the terminologies in both Chinese and English
Try to understand the concepts without first translating them into Chinese (in your mind)

Artificial Intelligence

- 国务院印发了《新一代人工智能发展规划》，提出了六方面的重点任务和一系列保障措施，要求到2020年人工智能总体技术和应用与世界先进水平同步，人工智能核心产业规模超过1500亿元，带动相关产业规模超过1万亿元；到2030年，我国人工智能理论、技术与应用总体达到世界领先水平。

Artificial Intelligence and COVID-19

- **AI + CT Screening**

Artificial Intelligence and COVID-19

- **AI + Targeted Drug Research**

- **AI + Body Temperature and Face Recognition**

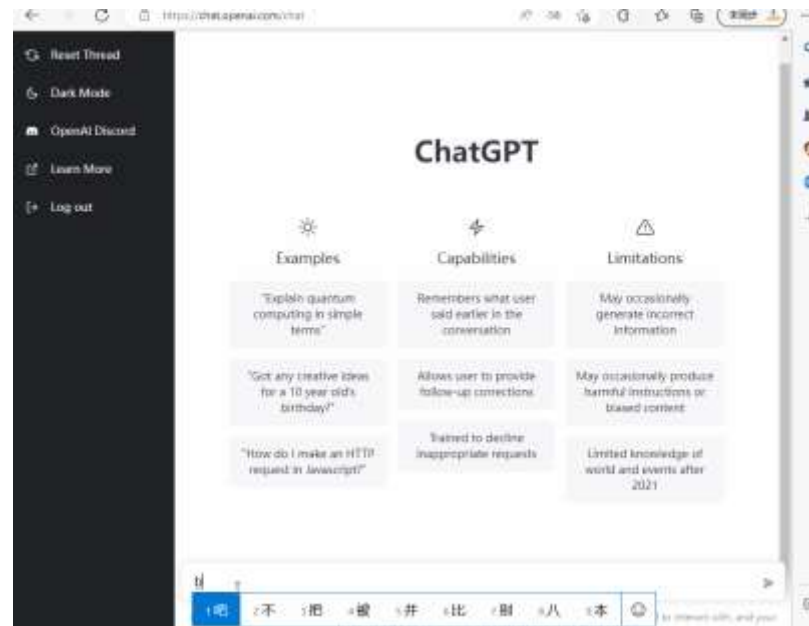# Artificial Intelligence and COVID-19

- AI + Information Gathering

轨迹追踪

信息共享

…

ChatGPT

ChatGPT（全名：Chat Generative Pre-trained Transformer），美国OpenAI研发的聊天机器人程序，于2022年11月30日发布。ChatGPT是人工智能技术驱动的自然语言处理工具，它能够通过学习和理解人类的语言来进行对话，还能根据聊天的上下文进行互动，真正像人类一样来聊天交流，甚至能完成撰写邮件、视频脚本、文案、翻译、代码，写论文等任务。

Remarks

- **What's the use of Pattern Recognition**

  - One of the core components in AI：computer vision、natural language processing、speech recognition，…
  - One of the most hot research fields in computer science
  - Common interests of Facebook、Google、Baidu、…

- **So ...**

  - Find the most cool job
  - Do the most exciting research
  - …

Chapter 1

Introduction

The 3W of Pattern Recognition

- **W**hat is Pattern Recognition (PR)?

  What is **Pattern**?

  What is **Recognition**?

  What is **Pattern Recognition**?

- **W**hy do we need Pattern Recognition?

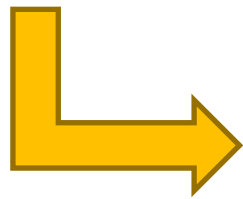 The **necessity and importance** for pattern recognition

- Ho**W** to perform Pattern Recognition?

  The **building blocks** of a pattern recognition system

What is Pattern?

"To understand is to perceive patterns"

– Isaiah Berlin

Patterns are essential for human perception and understanding

"A pattern is the opposite of a chaos(混乱，混沌); it is an entity vaguely defined (定义模糊的实体), that could be given a name."
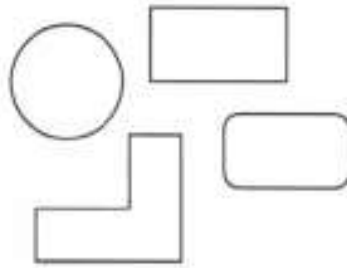
– Satoshi Watanabe

What is Pattern? (Cont.)

Some examples……

Barcode

Geometry

What is Pattern? (Cont.)

"A pattern is the opposite of a chaos; it is an entity vaguely defined, that could be given a name."

- There are various kinds of patterns
  - Visual patterns (视觉模式) such as face, fingerprint, etc.
  - Temporal patterns (时序模式) such as speech, audios, videos, data streams, etc.
  - Logical patterns (逻辑模式) such as characters, strings, images, etc.
  - ......

What is Recognition?

"Recognition *of a pattern as a member of a category we already know, or we are familiar with*"

"识别(Recognition)是将模式鉴定(Identification)为我们已知或者熟悉的类别(Category)的成员"

Two types of recognition

Classification (分类) ➡ Categories are known and the task is to assign a proper class label for each pattern

Clustering (聚类) ➡ Categories are unknown and the task is to learn categories and group the patterns accordingly
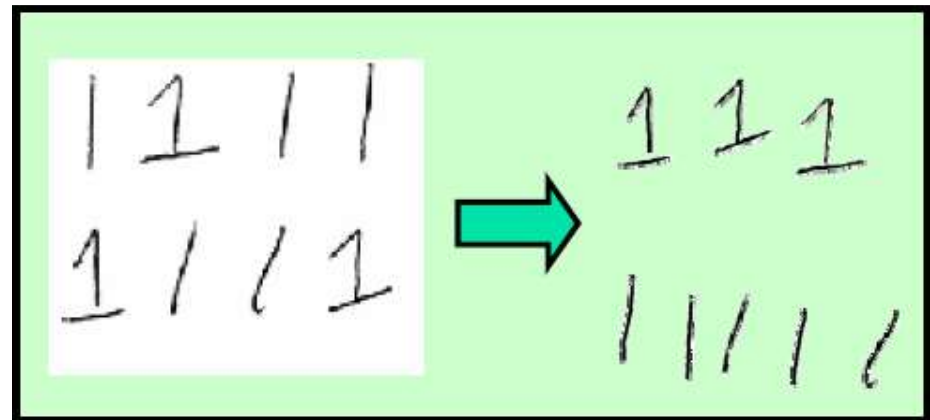
Classification vs. Clustering

Classification: An example

We already know the categories of characters, and then classify the handwritten ones into category "A" and category "B"
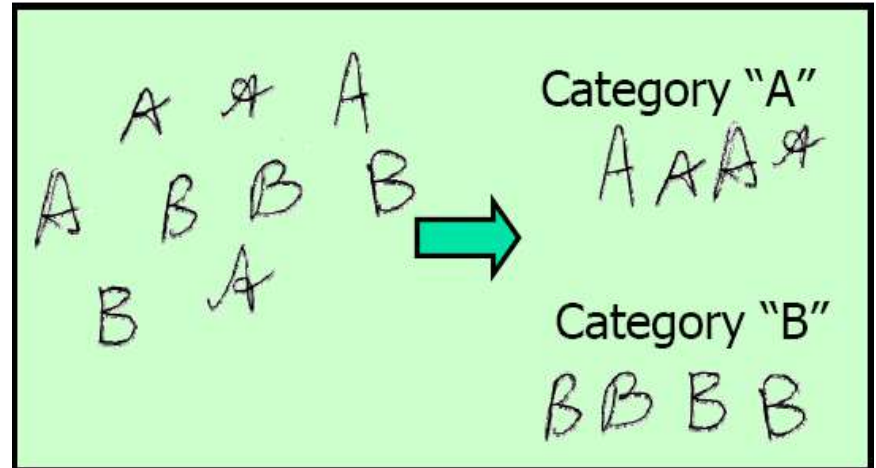
Clustering: An example

We do not know the categories of symbols, and then learn the categories and group the symbols accordingly

What is Pattern Recognition?

Pattern recognition is the procedure of processing and analyzing diverse information (numerical, literal, logical) characterizing the objects or phenomenon, so as to provide descriptions, identifications, classifications and interpretations for them.

对表征事物或现象的各种形式的（数值的，文字的和逻辑关系的）信息进行处理和分析，从而对事物或现象进行描述、辨认、分类和解释的过程。

What is Pattern Recognition? (Cont.)

A "Perceive + Process + Prediction" View

It is the study of how machines can

- ✓ **Perceive**: Observe the environment (i.e. interact with the real-world)

- ✓ **Process**: Learn to distinguish patterns of interest from their background

- ✓ **Prediction**: make sound and reasonable decisions about the categories of the patterns

Why need pattern recognition?

"The real power of human thinking is based on recognizing patterns. The better computers get at pattern recognition, the more humanlike they will become."
  – *Ray Kurzweil @ New York Times, 2003*

"The problem of searching for patterns in data is a fundamental one and has a long and successful history."
  – *Christopher M. Bishop*

Pattern recognition is needed in designing almost all automated and intelligent systems!
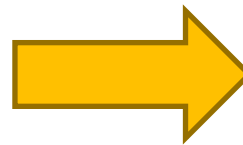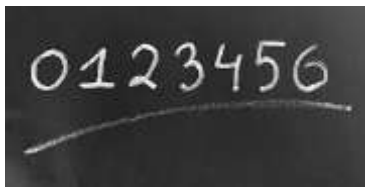
Applications of Pattern Recognition

1) Character Recognition

[字符识别]

Input:
images with characters

Output:
the identified character strings
(0123456)

Useful in scenarios such as automatic license plate recognition (ALPR), optical character recognition (OCR), etc.

Applications of PR (Cont.)

2) Fingerprint Recognition

[指纹识别]

Input:
fingerprints of some person

Output:

the person's identity

Useful in scenarios such as computerized access control, criminal pursuit, etc.
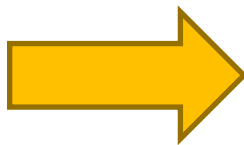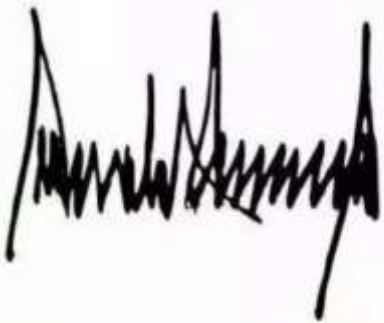
Applications of PR (Cont.)

3) Signature Identification

[签名验证]

Input:

signature of some person
(sequence of dots)

Output:

the signatory's identity

Useful in scenarios such as digital signature verification, credit card anti-fraud, etc.

Applications of PR (Cont.)

4) Speech Recognition
[语音识别]

Input:
acoustic signal
(e.g. sound waves)
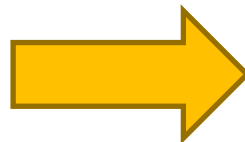
Output:

contents of the speech
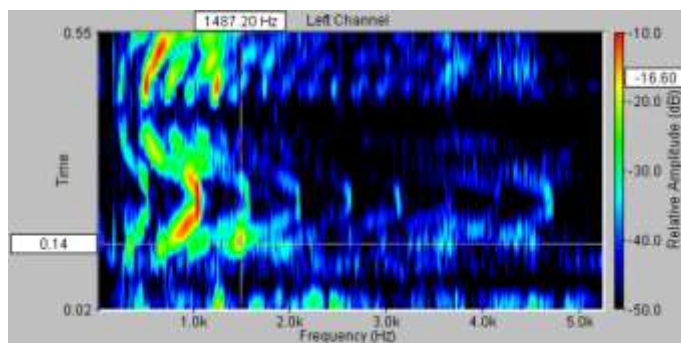
"Imagine all the people.
Living life in peace"

Useful in scenarios such as speech-to-text (STT), voice command & control, etc.

Applications of PR (Cont.)

5) Face Detection
[人脸检测]

Useful in scenarios such as digital camera capturing, video surveillance, etc.

Input:
images with several people

Output:
locations of the peoples' faces in the image
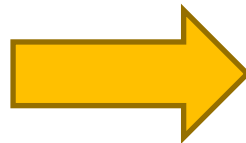
Applications of PR (Cont.)

6) Text Categorization

[文档分类]

Input:

document, web pages, etc.

Output:

category of the text, such as political, economic, military, sports, etc.

Useful in scenarios such as information retrieval, document organization, etc.

Applications of PR - More

Problem	Input	Output
Detection and diagnosis of disease	Electrocardiogram (ECG) waveforms, Electroencephalogram (EEG) waveforms	Types of cardiac(心脏疾病) conditions, classes of brain conditions
Natural resource identification	Multispectral images	Terrain forms(地形形态), vegetation cover
Aerial reconnaissance (空中侦察)	Visual, infrared, radar images	Tanks, airfields
Identification and counting of cells	Slides of blood samples, micro-sections of tissues	Type of cells
Inspection (PC boards, IC masks, textiles)	Scanned image (visible, infrared)	Acceptable/unacceptable
Manufacturing	3-D images (structured light, laser, stereo)	Identify objects, pose, assembly
Web search	Key words specified by a user	Text relevant to the user
……	……	……

How to recognize patterns?

For humans, pattern recognition is natural & easy

recognize a face   understand spoken words

read handwritten characters   identify items by feel

decide whether an apple is ripe by its smell   ……

For computers, pattern recognition is never easy

All in all, pattern recognition is important, useful, attractive, but rather challenging

Challenges ➔ Opportunities

Basic Concepts

Model (模型)

Descriptions which are typically mathematical in form
[以数学形式表达的性质]

e.g. image ➜ *matrix; sound waves* ➜ *frequency vector*

Sample (样本)

Representatives of the patterns we want to classify
[分类的基本对象，模式的实例]

e.g. fingerprint of a suspect; Electrocardiogram of a patient

Training Set (训练集)

A set of samples used to train classifiers
[用于训练分类器的样本集合]

Basic Concepts (Cont.)

Test Set (测试集)

A set of samples to be classified, usually being mutually exclusive to training set

[用于测试分类器的样本集合,通常与训练集无交集]

"Training set" vs. "Test set" ⇔ "Homeworks" vs. "Exams"

Feature (特征)

Attributes which characterize properties of the samples
[用于刻画样本性质的属性]

e.g. to characterize a person, we may use features such as height, weight, age, salary, occupation, etc.

Basic Concepts (Cont.)

Feature Vector (特征向量)

Vector formed by a group of features, usually in column form
[由一组特征组成的向量，通常表示为列向量]

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

vector component
(in *italic*)

feature vector
(in **boldface**)

dimensionality
(number of features)

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

$$\mathbf{x} = (x_1, x_2, \ldots, x_d)^t$$

transpose
operator
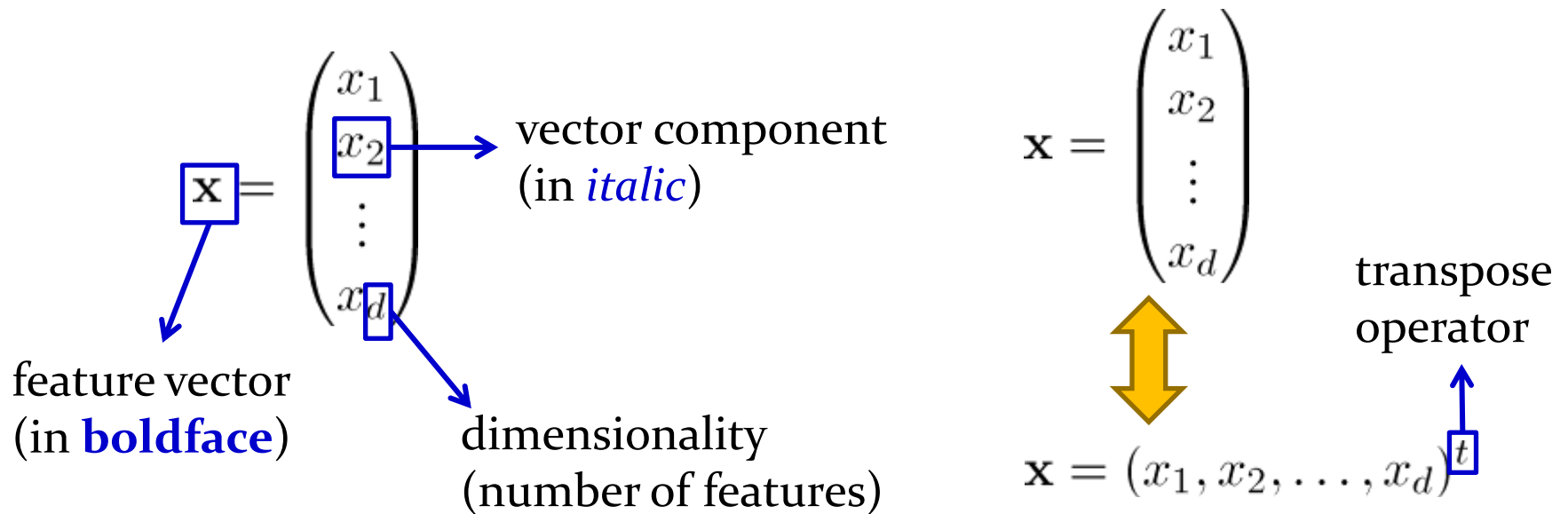
Basic Concepts (Cont.)

Feature Space (特征空间)

Space containing all the possible feature vectors
(由所有可能的特征向量组成的数据空间)

e.g. the d-dimensional Euclidean space $\mathbf{R}^d$

Scatter Plot (散布图)

Each sample is plotted as a
point in the feature space
(将每个样本表示为特征空间
中的一个点)
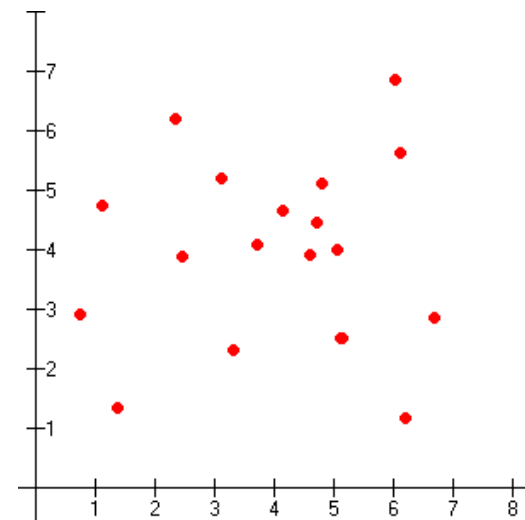
a 2D scatter plot

Basic Concepts (Cont.)

Decision Boundary (决策边界)

Boundaries in feature space which separate different categories (特征空间中区分各个类别的边界)
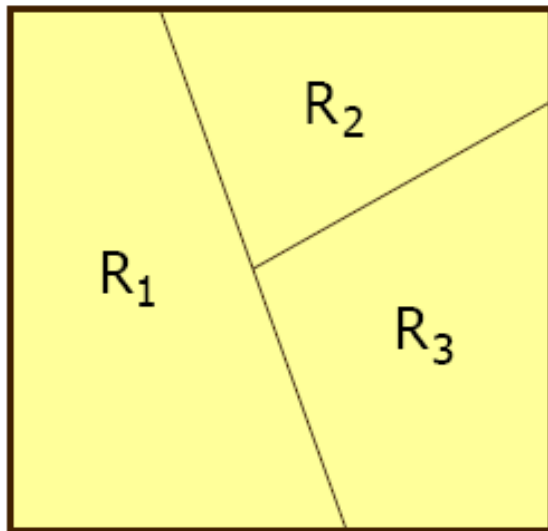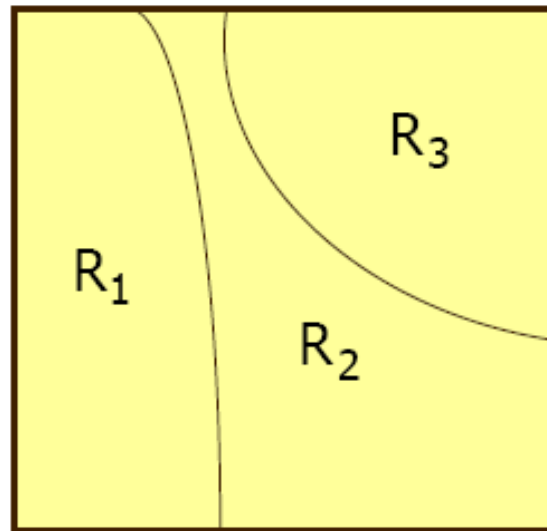
linear boundary   quadratic boundary   complex boundary
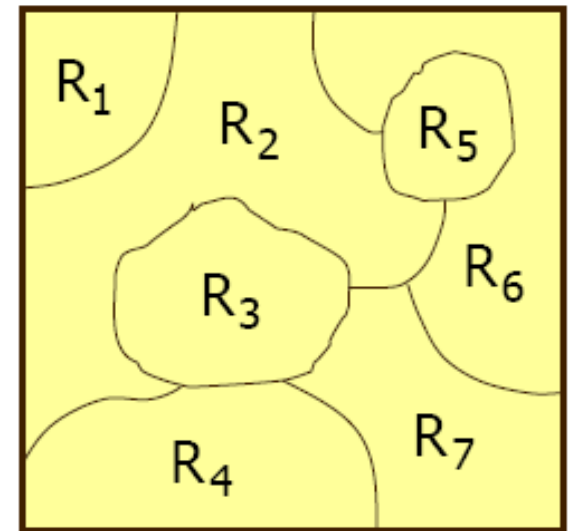
How to do pattern recognition?

An Example

The task: Automate the process of separating animals according to species
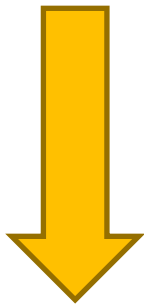
Separate *goat* from *sheep*
[山羊 vs. 绵羊]

Preprocessing

Feature extraction

Classification

goat — sheep
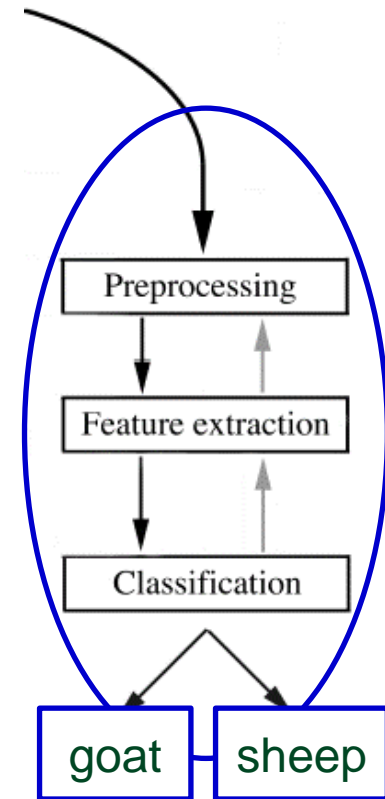
Three basic steps

Example: "goat" vs. "sheep" (Cont.)

Step I: Preprocessing (预处理)

Goal: Preprocess the image captured by the camera, such that subsequent operations could be simplified without losing relevant information

Routine image processing

☐ Adjust the level of illumination

☐ Denoising

☐ Enhance the level of contrast

segmentation

☐ Isolate different objects from one another

☐ Isolate objects from the background

......

Example: "goat" vs. "sheep" (Cont.)

Step II: Feature Extraction (特征抽取)

Goal: Extract features (with good distinguishing ability) from the preprocessed image to be used for subsequent classification

Goat is usually larger than sheep ➡ "size" could be a good candidate for features

Sheep wool is usually more curly than goat ➡ "curliness" could be another good candidate for features

......

Example: "Goat" vs. "Sheep" (Cont.)

Step III: Classification (分类)

Goal: To distinguish different types of objects (in this case, *Goat* vs. *Sheep*) based on the extracted features

histogram for *size*
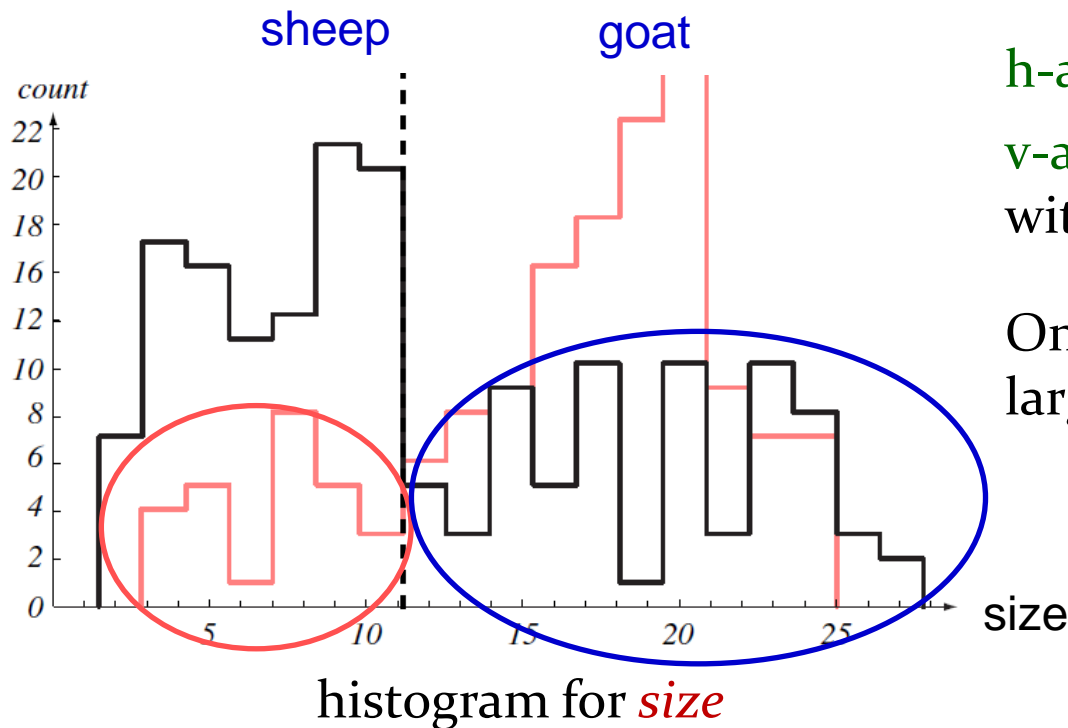
h-axis: size of the animals

v-axis: number of animals with a certain size

On average, goat is somewhat larger than sheep

Too much overlaps ➔ poor separation with the length feature

Example: "Goat" vs. "Sheep" (Cont.)

histogram for *curliness*
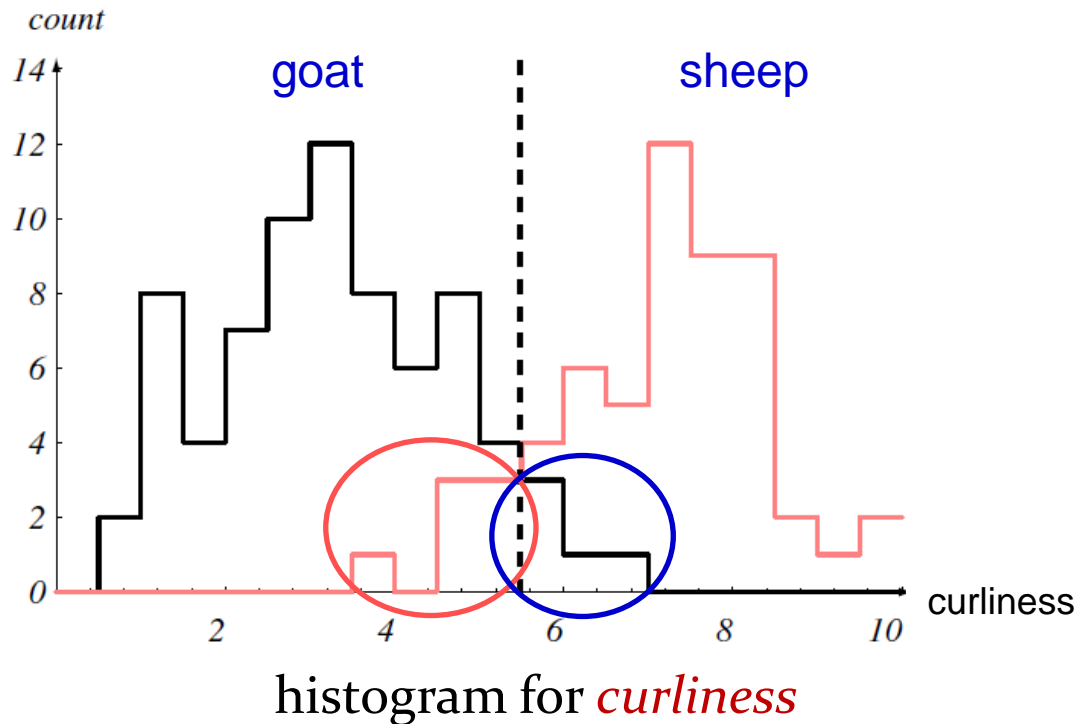
h-axis: curliness of wool

v-axis: number of animals with a certain curliness

On average, sheep wool is much curlier than goat

Less overlaps ➔ better separation with the lightness feature, but still a bit unsatisfactory

What if no other single feature yields better performance?
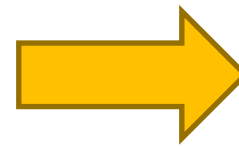
Use more features at the same time!

Example: "Goat" vs. "Sheep" (Cont.)

Using two features
simultaneously

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

curliness

sheep   goat

black dots: sheep samples

red dots: goat samples

Linear decision boundary:

$$f(x_1, x_2) = a \cdot x_1 + b \cdot x_2 + c$$

$$f(x_1, x_2) > 0 \implies \text{sheep}$$

$$f(x_1, x_2) \leq 0 \implies \text{goat}$$

size
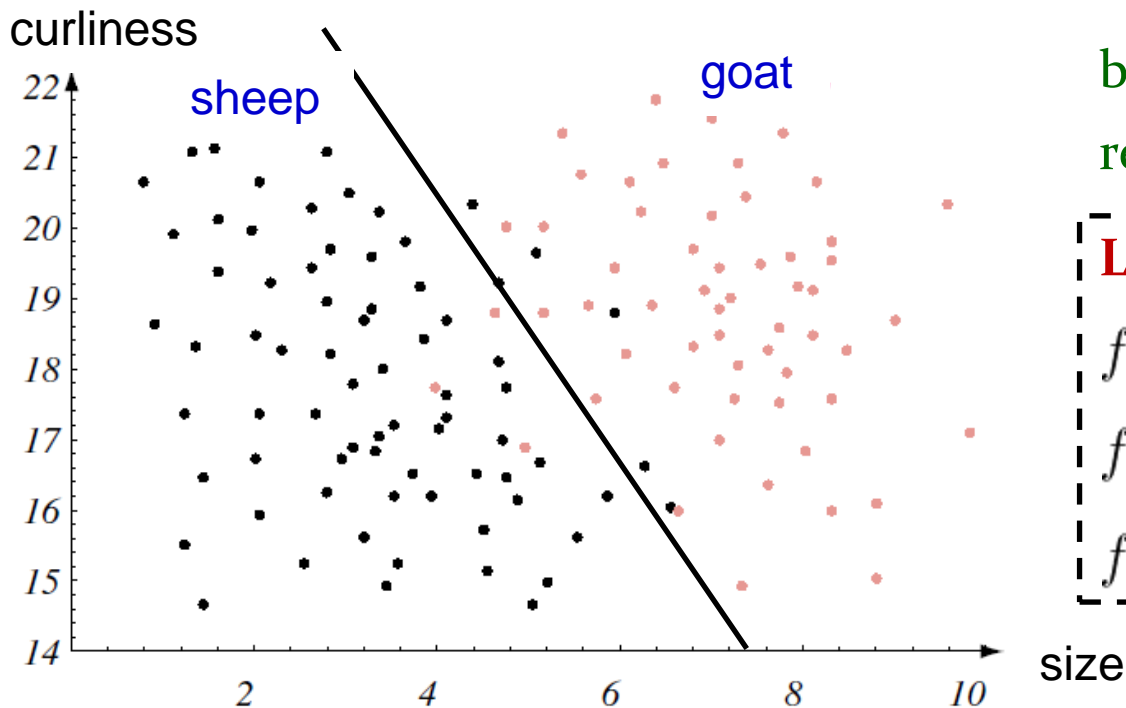
scatter plot for the feature vectors

Much better than
single feature

Example: "Goat" vs. "Sheep" (Cont.)

Linear decision boundary:

⬇
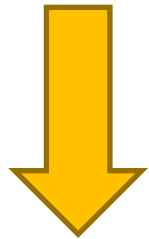
Complex decision boundary

curliness

sheep   goat

size
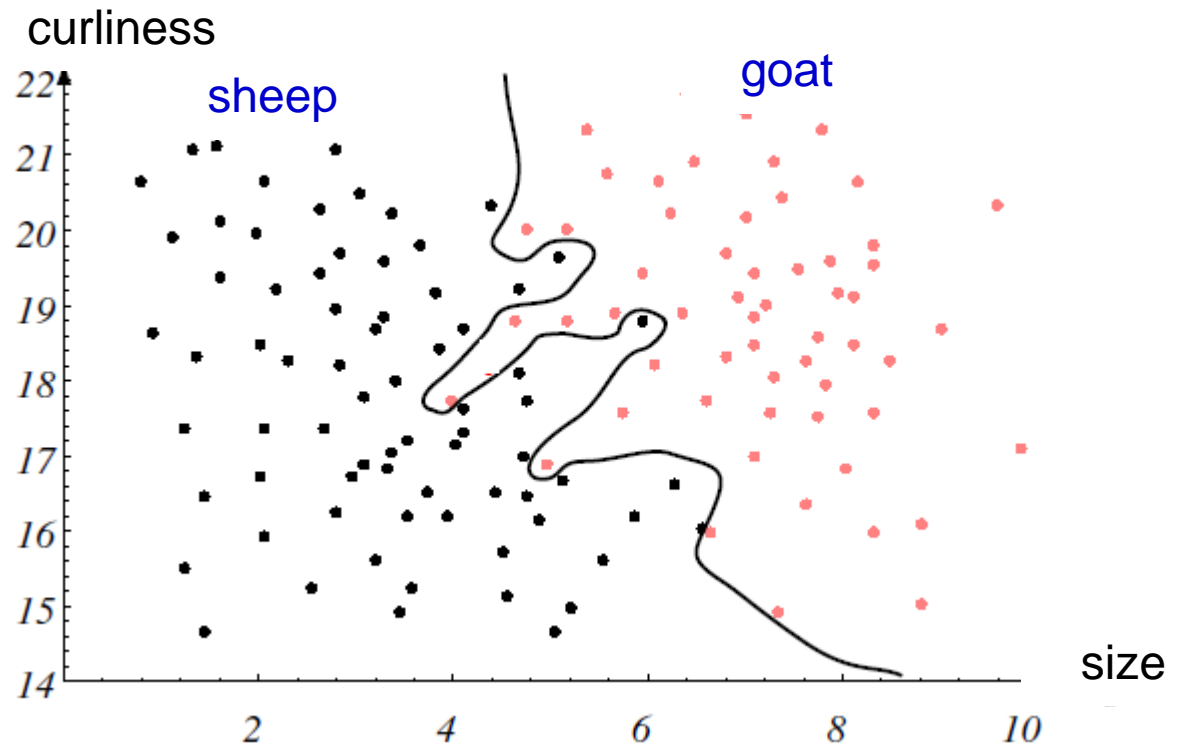
All the training samples (i.e. known patterns) have been separated perfectly

Linear vs non-linear, which one is better?

Example: "Sheep" vs. "Goat" (Cont.)

Generalization

[泛化能力/推广能力] **The ultimate goal!**

The central aim of designing a classifier is to make correct decisions when presented with *novel* (*unseen/test*) patterns, not on training patterns whose labels are already known

Performance on the training set

Simplicity of the classifier

Tradeoff

Related Fields to PR

Pattern Recognition: **Pattern ➔ Category**

- Hypothesis Testing (假设检验)
    - Null hypothesis ➔ Rejection or Not
        - *E.g.: To determine whether a drug is effective; Null hypothesis: it has not effect*
- Image Processing (图像处理)
    - Image ➔ Image
- Regression (回归分析)
    - Pattern ➔ Real Value
- Interpolation (插值)
    - Pattern (unexplored input range) ➔ Interpolated Value
- Density Estimation (概率密度估计)
    - Patterns ➔ Probability density function (pdf) for different categories

Often employed as preliminary steps in pattern recognition

Pattern Recognition System

In addition to the usual "bottom-up" flow of data, some systems also employ feedback from higher levels back down to lower levels (gray arrows)

decision

post-processing — A post processor decide on the appropriate action based on the classification

classification — A classifier uses extracted features to assign the sensed object to a category

feature extraction — A feature extractor measures object properties that are useful for classification

segmentation — A segmentor isolates sensed objects from the background or from other object

sensing — A sensor converts physical inputs (e.g. images, sounds) into digital signal data

input

Design Cycle of PR System

The design of a PR system usually entails a number of different activities, such as *data collection, feature choice, model choice, classifier training, classifier evaluation.*

- ☐ Data collection accounts for a large part of the cost of developing a PR system

- ☐ Feature choice and model choice are highly domain-dependent, where *prior knowledge* (先验知识) plays very important role

  e.g.: the orientation of the tails are the most reliable feature to distinguish sheep vs goat!

start

collect data

choose features

choose model

train classifier

evaluate classifier

end
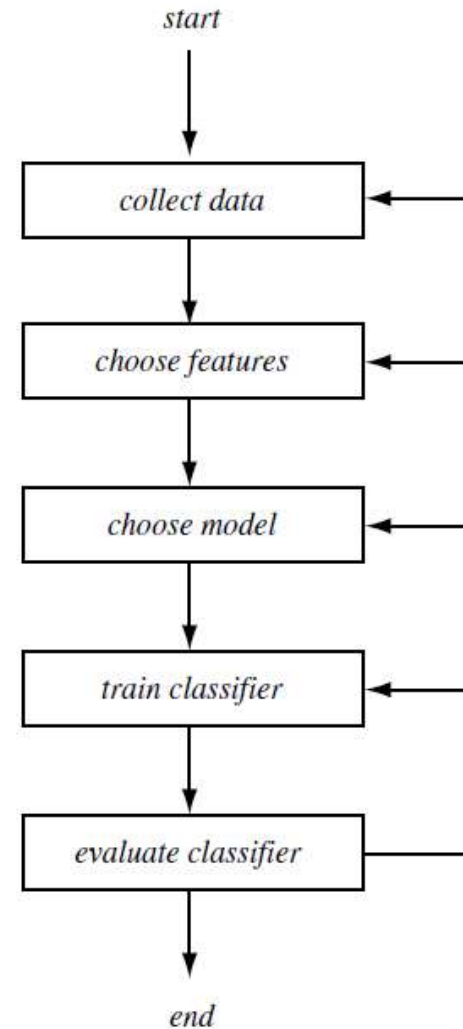
Important Issues in Pattern Recognition

- Noise
  (噪声)

- Segmentation
  (分割)

- Data Collection
  (数据采集)

- Domain Knowledge
  (领域知识)

- Feature Extraction
  (特征抽取)

- Pattern Representation
  (模式表示)

- Missing Features
  (特征缺失)

- Model Selection
  (模型选择)

- Overfitting
  (过配)

- Context
  (上下文)

- Classifier Ensemble
  (分类器集成)

- Costs and Risks
  (代价与风险)

- Computational Complexity
  (计算复杂度)

- ......

Noise

- **General definition**
  - Any property of the sensed pattern which is not due to the true underlying model but instead to intrinsic randomness of the world or the sensors

- **Various types of noise exist**
  - shadows, conveyor belt might shake, etc.

- **Noise can reduce the reliability of the feature values measured**

- **Knowledge of the noise process can help improve performance**

Segmentation

- Individual patterns have to be segmented for subsequent pattern recognition operations

- One of the deepest as well as hardest problems in pattern recognition
  - How can we segment the images without having categorized them firstly?
  - On the other hand, how can we categorize the images without having segmented them firstly?

- How do we "group" together the proper number of elements
  - **BEATS** ➔ BE, BEAT, EAT, AT, EATS?

Data Collection

- A small set of "typical" examples ➜ Preliminary study of system feasibility

- Much more data ➜ Assure good performance in the fielded system

- How do we know that we have collected:
  - Adequately large set of examples for training and testing the system?
  - Representative set of examples for training and testing the system?

- The efforts of data collection could be rather demanding

Domain Knowledge

- There is not sufficient data for training ➔ Incorporate domain knowledge (a.k.a. prior knowledge)

- **Type I:** Incorporate domain knowledge on the patterns themselves – Difficult!
  - ❑ To recognize all types of chairs
  - ❑ Astounding variety in *number of legs*, *material*, *shape*, and *so on* ➔ What is the commonness for chairs which could be regarded as domain knowledge?

- **Type II:** Incorporate domain knowledge on the pattern generation procedure
  - ❑ Optical character recognition ➔ Assume handwritten characters are written as a sequence of strokes
  - ❑ First try to recover stroke representations ➔ deduce the character from the identified strokes

Feature Extraction

r=0.791
P<0.0001
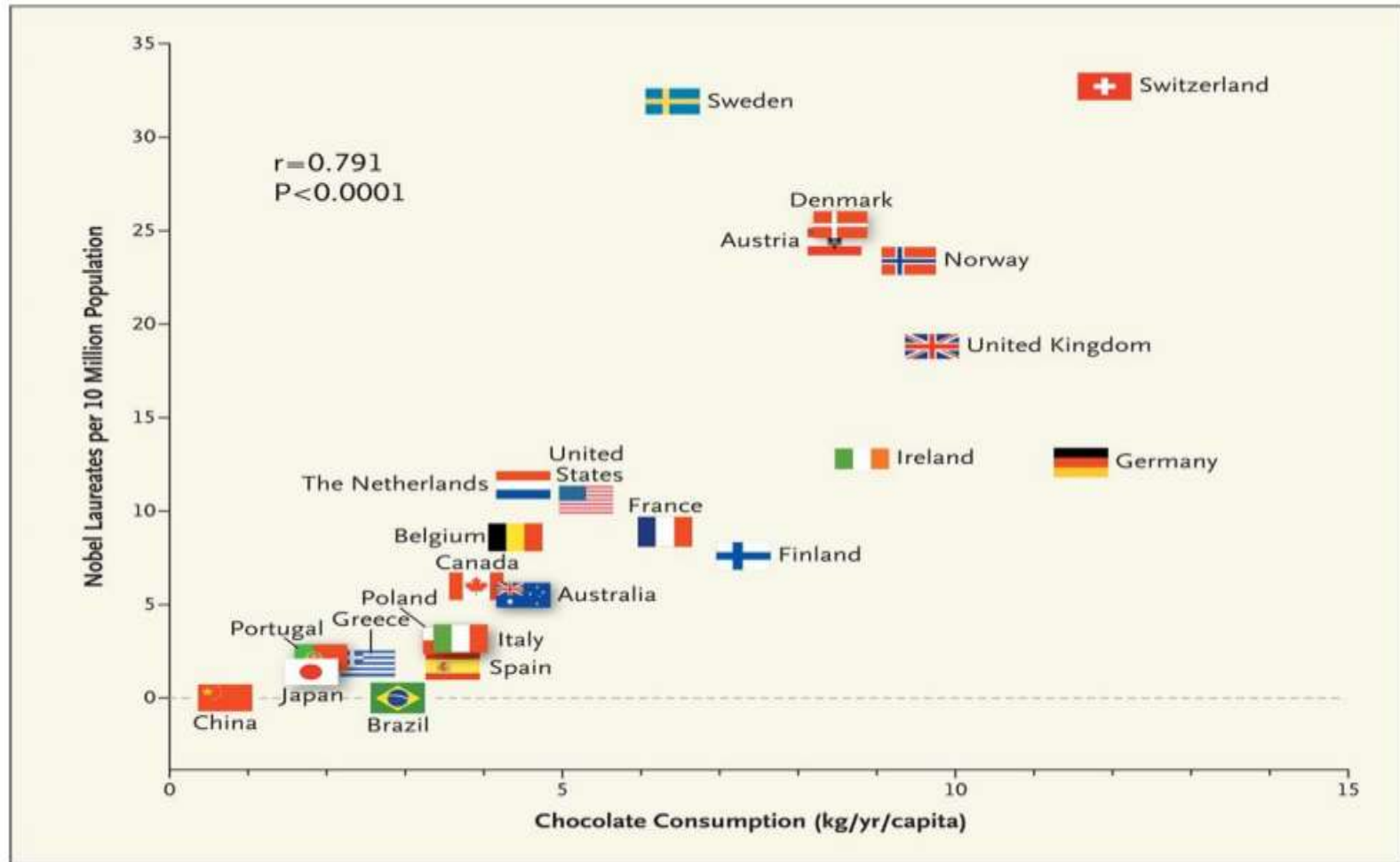
F. H. Messerli: *Chocolate Consumption, Cognitive Function, and Nobel Laureates*, N Engl J Med 2012

Pattern Representation

- Various ways for pattern representation
  - **Statistical:** *feature vector* (the most popular)
  - **Template Matching:** *prototype templates*
  - **Syntactic:** *rules or grammars*

- Desired Properties
  - Patterns from the same classes should have similar representations
  - Patterns from different classes should have dissimilar representations
  - Pattern representations should be invariant to transformations such as *translations, rotations, resizes, reflections, non-rigid deformations*
  - Intra-class variation should be small
  - Inter-class variation should be large
  - ......

Missing Features

- In practical problems, values for certain features may be missing
  - Occlusion between fishes ➔ fish width can't be measured

- How could we train classifiers with missing features?
  - Naïve method could be used, but may not be optimal
    - Assuming the value of missing features is zero
    - Assigning the average value of patterns already seen for the missing feature
  - Sophisticated method might be better, but requires extra efforts in terms of storage and time
    - Fill in the missing values with regression techniques

Model Selection

- Each pattern recognition method employs certain *model hypothesis*

- Every pattern recognition problem has its own *underlying true model*

- Fundamental questions on model selection

  - How do we know whether the hypothesized model is (relatively) consistent with the underlying true model?

  - How are we to know to reject a class of models and try another one?

  - Can we automate the process of model selection, instead of trial and error which is random and tedious?

Overfitting

- We can get perfect classification performance on the training data by choosing complex models
  - Complex models are tuned to the particular training samples, rather than the characteristics of the true model
- Models overly complex than necessary lead to overfitting
  - Good performance on the training data, but poor performance on novel data
- How can we find principled ways to obtain best complexity?
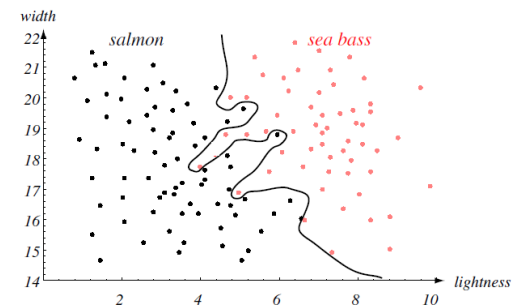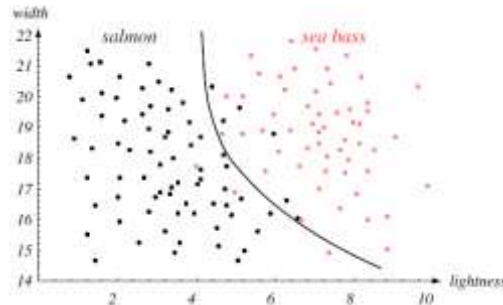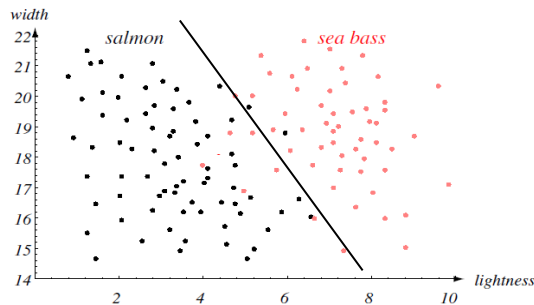
Context

- **Context:** Input-dependent information, other than from the pattern itself

  - context of language, context of videos, etc.

- The same pattern within different context might have different meanings

  - Use the context of a conversation to infer the meaning of the speaker

- Context is very helpful!

How   m ch   info mat on
a e   y u   mi sing

Classifier Ensemble

- Classifier ensemble aims to improve generalization performance by employing a number of classifiers for the same task

  - To improve the performance of speech recognizer: combine the results of *acoustic recognition* and *lip reading*

  - a.k.a. *Multi-classifier System*, *Mixture of Experts*, *Classifier Fusion*, etc.

  - Diverse ensemble techniques: *Bagging*, *Boosting*, *Random subspace*, etc. [ref. pp.475]

- How to combine different classifiers?

  - **Majority voting:** vote for the category where most classifiers agree

  - **Weighted voting:** weight each vote by classifier's confidence

  - **Stacking:** learn the rule of combination (more complicated)

Costs and Risks

- **Cost is the loss after making incorrect decisions**
  - **Equal cost:** In OCR, the cost of mistaking "6" as "9" might be equal to that of mistaking "9" as "6"
  - **Unequal cost:** In Covid-19 diagnosis, the cost of mistaking "positive (阳性)" as "negative (阴性)" would be much higher than that of mistaking "negative" as "positive"

- **Risk is total expected cost which we want to optimize**
  - Error rate (percentages of test patterns being wrongly classified)
  - Precision, Recall, Area under the ROC curve (AUC), etc.

- **Questions on costs and risks**
  - How do we incorporate knowledge of costs, e.g. unequal cost?
  - Can we estimate the *lowest* possible risk of any classifier?
  - ......

Computational Complexity

- How does an algorithm scale with
    - The number of features (dimensionality)
    - The number of training patterns
    - The number of possible categories

- Brute force (蛮力) approaches might lead to perfect classification, but with impractical time and storage requirements
    - In OCR, label all possible 20x20 binary pixel images with a category ➔ use simple table lookup (查表) to classify incoming patterns
    - Labeling each of the $2^{20 \times 20}$ ($\approx 10^{120}$) patterns is prohibitive

- How can we find a good tradeoff between computational ease and classifier performance?

Summary

- ## What is Pattern Recognition?
  - ### Pattern
    - The opposite of chaos
    - Various kinds: visual patterns, temporal patterns, logical patterns, etc.
  - ### Recognition
    - Identification of a pattern as a member of a category
    - Classification: categories known ➔ assign proper class label for each pattern
    - Clustering: categories unknown ➔ learn categories and group patterns
  - ### Pattern Recognition
    - Perceive: observe the environment (i.e. interact with the real-world)
    - Process: learn to distinguish patterns of interest
    - Prediction: make sound and reasonable decisions about the categories

Summary (Cont.)

- ## Why Pattern Recognition?

  - ❑ Pattern recognition is needed in designing almost all automated and intelligent systems

  - ❑ Applications of pattern recognition are ubiquitous

    - Character recognition (images ➜ characters)

    - Speech recognition (speech ➜ text)

    - Fingerprint recognition (fingerprints ➜ person's identity)

    - Signature identification (signature ➜ signatory's identity)

    - Face detection (images ➜ face locations)

    - Text categorization (documents ➜ semantic categories)

    - ……

Summary (Cont.)

- ## How Pattern Recognition?
  - ❑ Basic concepts
    - model, sample, training set, test set, feature, feature vector, feature space, scatter plot, decision boundary
    - An illustrative example: "sheep" vs. "goat"
    - **Generalization**: Make correct decisions given novel patterns
  - ❑ Related fields
    - hypothesis testing, image processing, associative memory, regression, interpolation, density estimation
  - ❑ Components of Pattern Recognition System
    - sensing ➔ segmentation ➔ feature extraction ➔ classification ➔ post-processing ➔ ......

Summary (Cont.)

- ## How Pattern Recognition?

  - ### Design Cycle of Pattern Recognition System

    - collect data ➔ choose features ➔ choose model ➔ train classifier ➔ evaluate classifier ➔ ……

  - ### Important Issues

    - ☐ Noise
    - ☐ Segmentation
    - ☐ Data Collection
    - ☐ Domain Knowledge
    - ☐ Feature Extraction
    - ☐ Pattern Representation
    - ☐ Missing Features

    - ☐ Model Selection
    - ☐ Overfitting
    - ☐ Context
    - ☐ Classifier Ensemble
    - ☐ Costs and Risks
    - ☐ Computational Complexity
    - ☐ ……