

Pengenalan Pola/ Pattern Recognition

Dasar Pengenalan Pola 1

Imam Cholissodin S.Si., M.Kom.





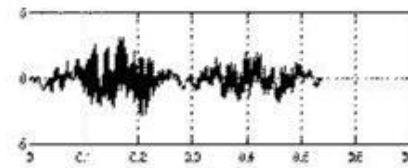
Dasar Pengenalan Pola 1

1. Apa itu pola?
2. Apa itu kelas pola?
3. Apa itu pengenalan pola?
4. Contoh penerapan
5. Perhitungan secara statistik
6. Persepsi manusia dan mesin
7. Proses pengenalan pola
8. Studi kasus



Apa itu Pola (*Pattern*) ?

- Pola adalah Objek, Proses, atau Kejadian yang dapat diberi nama
- Pola adalah himpunan pengukuran yang menggambarkan sebuah objek



John Smith



Apa itu Kelas Pola ?

- Kelas Pola / Kategori merupakan himpunan pola yang memiliki atribut tertentu
- Kumpulan dari beberapa objek yang identik (kemiripan data)



- Selama proses pengenalan objek dimasukkan ke dalam kelas yang ditentukan



Apa itu Pengenalan Pola ?

- Teori, Algoritma, Sistem untuk meletakkan pola-pola ke dalam kategori
- Menemukan hubungan suatu pola terhadap pola-pola sebelumnya
- Belajar membedakan pola yang dianggap penting terhadap latar belakangnya



Persepsi Manusia

- Manusia telah dianugerahi kemampuan untuk menerima rangsangan (indera) dari lingkungan dan memberikan aksi terhadap apa yang diamati,
 - Mengenali wajah
 - Memahami kata yang diucapkan
 - Membaca tulisan tangan
 - Membedakan makanan segar dari baunya
- Tugas kita:
 - Menjadikan mesin (komputer) memiliki kemampuan yang mirip dengan manusia





Contoh Aplikasi

Optical Character Recognition (OCR)

- **Handwritten**: sorting letters by postal code.
- **Printed texts**: reading machines for blind people, digitalization of text documents.

Biometrics

- **Face recognition, verification, retrieval.**
- **Finger prints recognition.**
- **Speech recognition.**

Diagnostic systems

- **Medical diagnosis**: X-Ray, ECG (ElectroCardioGraph) analysis.

Military applications

- **Automated Target Recognition** (ATR).
- **Image segmentation and analysis** (recognition from aerial or satellite photographs).



Aplikasi Pengenalan Pola berdasarkan Domain Permasalahan

Problem Domain	Application	Input Pattern	Pattern Classes
Bioinformatics	Sequence Anaysis	DNA/Protein sequence	Known types of genes/patterns
Data Mining	Seaching for meaningful patterns	Points in multi-dimensional space	Compact and well-seperated cluster
Document classification	Internet search	Text Document	Semantic categories (e.g., business, sports, etc.)
Document image analysis	Reading machine for the blind	Document image	Alphanumeric characters, words
Industrial automation	Printed circuit board inspection	Intensity or range image	Defective / non-defective nature of product
Multimedia database retrieval	Internet search	Video clip	Video genres (e.g., action, dialogue, etc.)



Aplikasi Pengenalan Pola berdasarkan Domain Permasalahan

Problem Domain	Application	Input Pattern	Pattern Classes
Biometric Recognition	Personal Identification	Face, iris, fingerprint	Authorized user for access control
Remote Sensing	Forecasting crop yield	Multispectral image	Land use categories, growth pattern of crop
Speech Recognition	Telephone directory enquiry without operator assistance	Speech waveform	Spoken words



Pendekatan Pengenalan Pola

- **Template Matching** : berdasarkan template
- **Statistical** : berdasarkan model statistik dari pola dan kelas pola yang diberikan
- **Structural (or syntactic)** : kelas pola direpresentasikan oleh struktur formal seperti grammer, string, automata, dll.
- **Neural networks** : mesin klasifikasi yang direpresentasikan oleh model sel neuron dari otak manusia

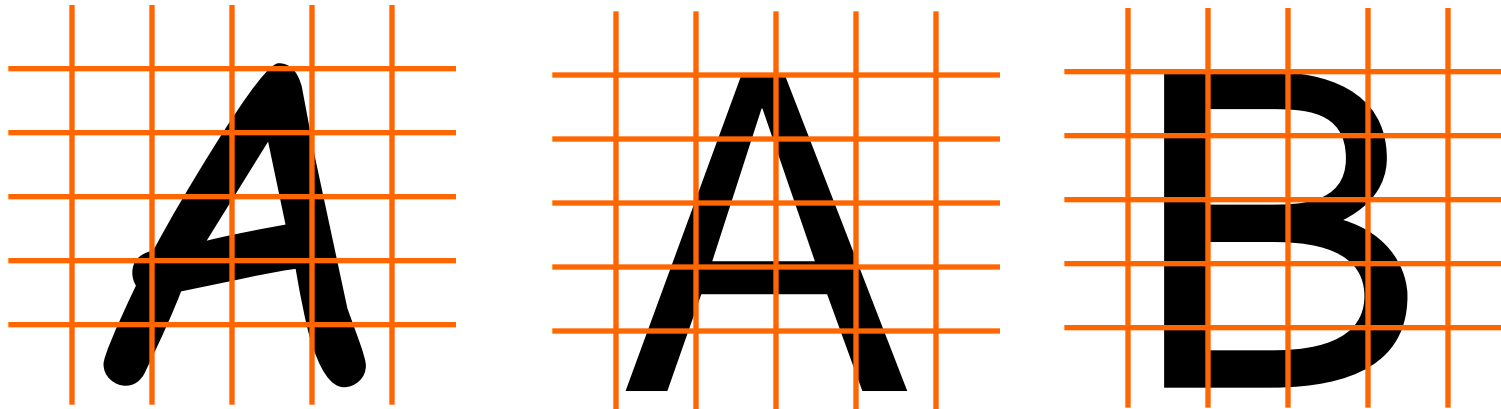


Model Pengenalan Pola

Approach	Representation	Recognition Function	Typical Criterion
Template matching	Samples, pixels, curves	Correlation, distance measure	Classification error
Statistical	Features	Discriminant function	Classification error
Syntactic or structural	Primitives	Rules, grammar	Acceptance error
Neural networks	Samples, pixels, features	Network function	Mean square error



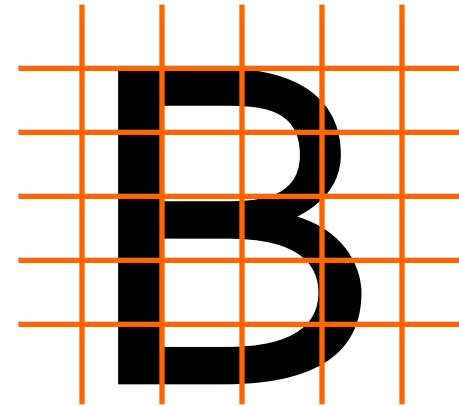
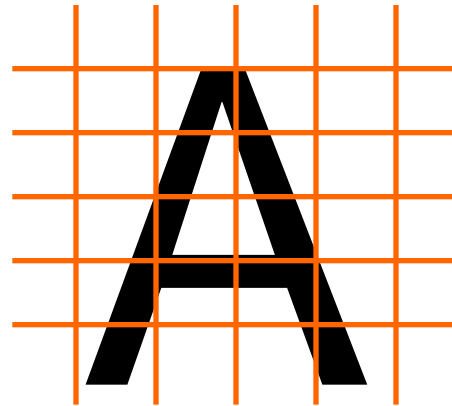
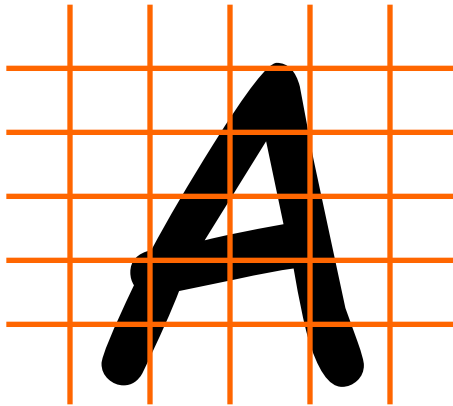
Pendekatan Statistik



**Membandingkan
Grid per Grid**



Pendekatan Statistik



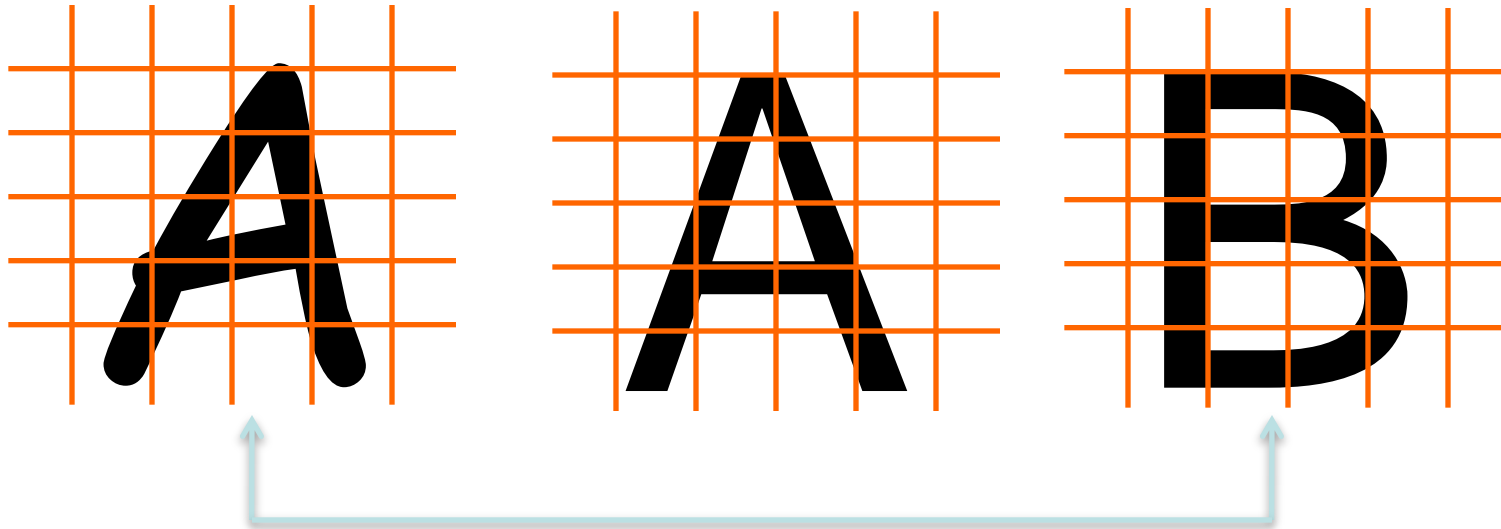
0	0	1	0
0	0	1	0
0	1	1	1
1	0	0	1
1	0	0	1

0	1	1	0
0	1	1	0
0	1	1	0
1	0	0	1
1	0	0	1

Jumlah grid
yang tidak
sesuai = 3



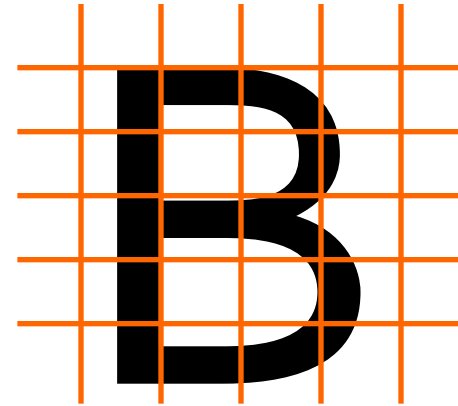
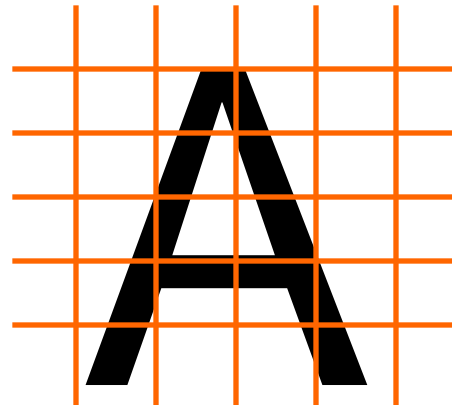
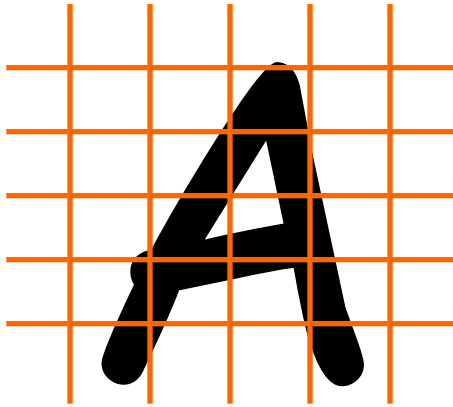
Pendekatan Statistik



**Membandingkan
Grid per Grid**



Pendekatan Statistik



0	0	1	0
0	0	1	0
0	1	1	1
1	0	0	1
1	0	0	1

Jumlah grid
yang tidak
sesuai = 10

1	1	1	0
0	1	0	1
0	1	1	1
0	1	0	1
1	1	1	0



Permasalahan

- Waktu yang dibutuhkan untuk mengenali pola (sesuai jumlah pola yang disimpan)

A-Z

a-z

0-9



Solution
***Artificial
Intelligence***



Persepsi Manusia dan Mesin

- Kita sering dipengaruhi oleh pengetahuan tentang bagaimana pola dimodelkan dan dikenali secara alami ketika kita membangun algoritma pengenalan pola
- Penelitian tentang persepsi mesin juga membantu kita mendapatkan pemahaman lebih dalam dan apresiasi untuk sistem pengenalan pola secara alami
- Sampai saat ini, kita telah mengaplikasikan beberapa teknik yang murni secara numerik dan tidak ada korespondensinya dengan sistem alamiah

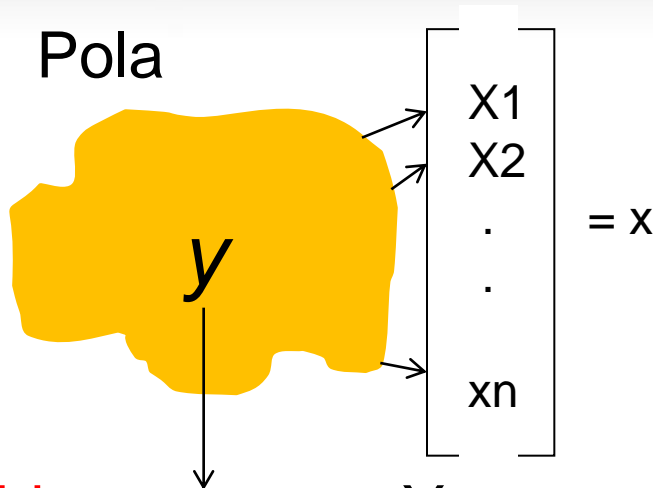


Pengenalan Pola

- Dua Tahap
 - Learning
 - Detection
- Waktu Learning lebih besar
- Sulit untuk belajar, tetapi sekali terpelajar sistem akan menjadi “**natural**”
- Dapat menggunakan metode AI :
 - Neural Network
 - Machine Learning



Konsep Dasar



Feature vector $x \in X$

- Vector dari hasil pengamatan (pengukuran).
- x adalah sebuah titik dalam ruang vektor X

Hidden state $y \in Y$

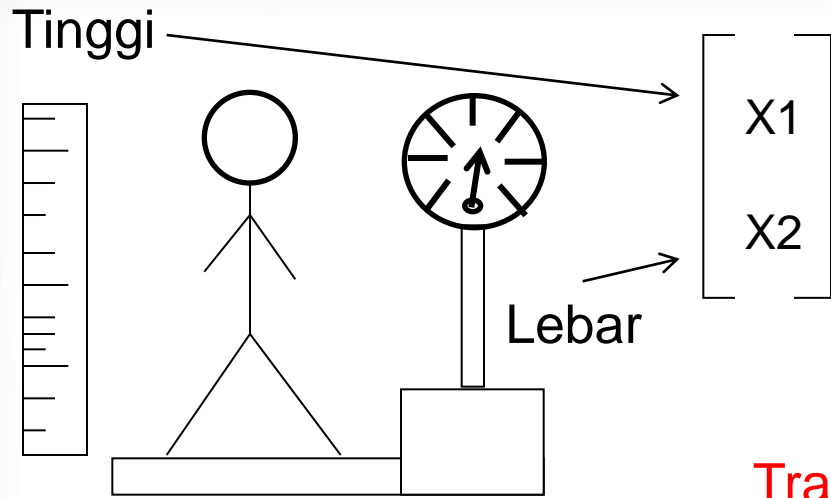
- Tidak dapat diukur secara langsung.
- Pola dengan hidden state yang sama terdapat dalam kelas yang sama

Task

- Mendesain classifier (decision rule) $q : X \rightarrow Y$ yang menentukan hidden state berdasarkan pengamatan



Example



Task: jockey-hooper recognition

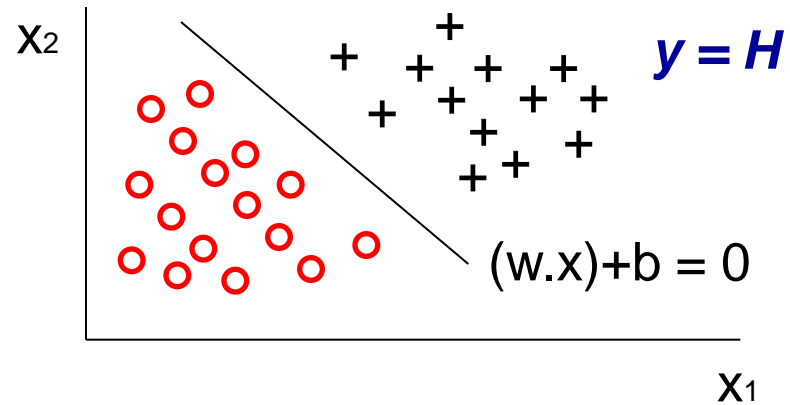
Himpunan hidden state $Y = \{H, J\}$

Ruang vektor $X = \mathbb{R}^2$

Training examples : $\{(x_1, y_1), \dots, (x_i, y_i)\}$

Linier classifier :

$$q(x) = \begin{cases} H & \text{if } (w \cdot x) + b \geq 0 \\ J & \text{if } (w \cdot x) + b < 0 \end{cases}$$





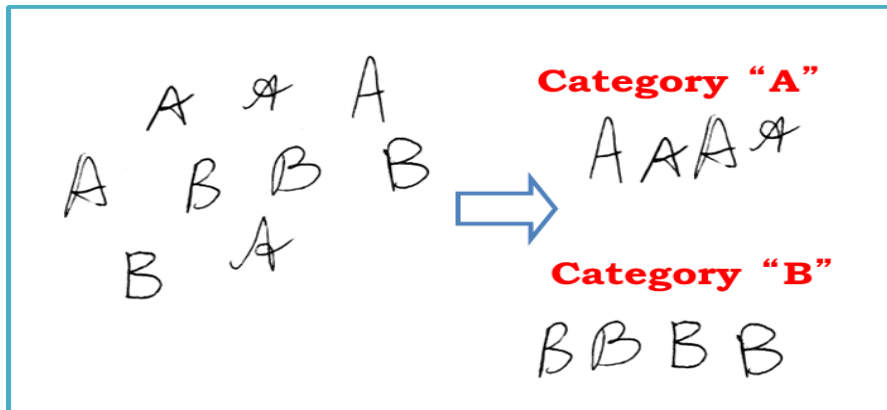
Learning

- Bagaimana mesin dapat belajar aturan (rule) dari data.
 - **Supervised learning :**
User menyediakan label kategori atau value/ bobot untuk masing-masing pola dalam data training.
 - **Unsupervised learning :**
Sistem membentuk cluster atau pengelompokan secara alami dari pola yang dimasukan.

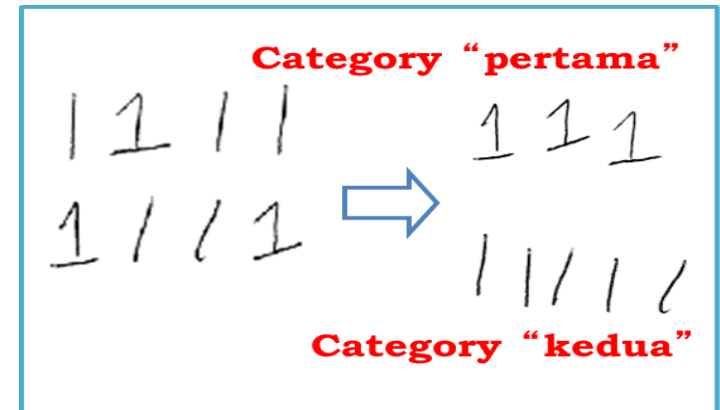


Classification Vs Clustering

- Classification (Kategori/ Label kelasnya diketahui).
- Clustering (Kategori/ Label kelompoknya tidak diketahui dan biasanya dinamai dengan kelas pertama, kedua dst).



Classification
(Supervised Classification)

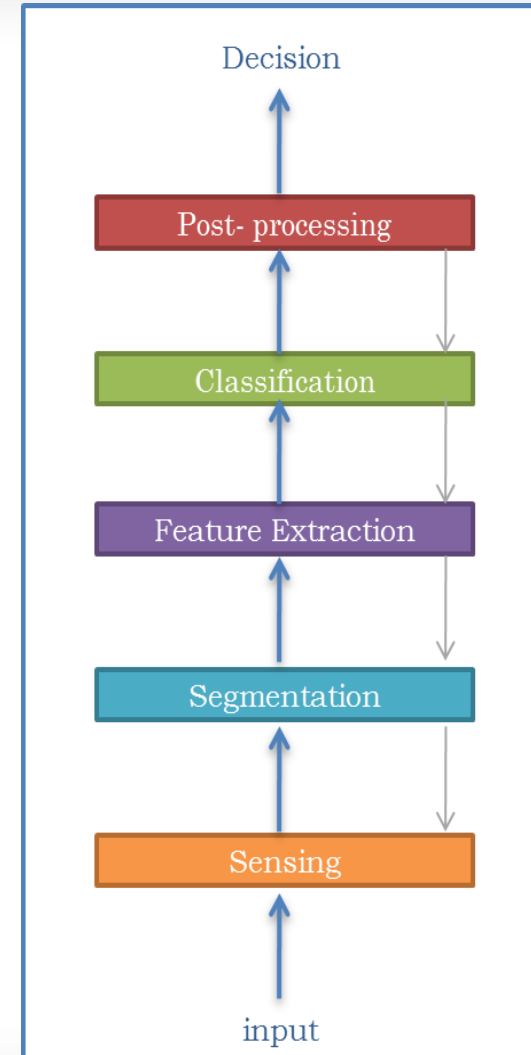


Clustering
(Unsupervised Classification)



Pattern Recognition Process

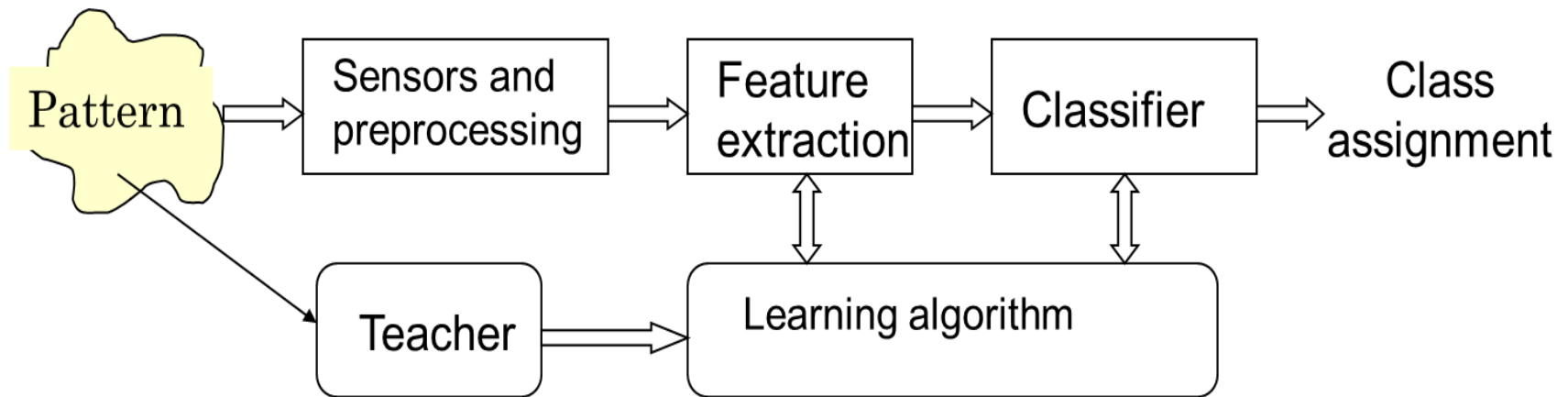
- **Input & Sensing :**
 - Proses pengambilan data (Acquisition) & memasukkan data.
 - Pemrosesan data sesuai dengan karakteristik/ fakta yang ada & pengukuran untuk nilai dari variable data.
- **Pre-processing :**
 - Menghilangkan noise pada data.
 - Melakukan pemisahan pola-pola yang menarik (pattern of interest) dari data.
- **Segmentation :**
 - Proses pemisahan objek satu dengan objek yang lain dalam suatu gambar.
- **Feature extraction :**
 - Menemukan representasi baru/ perwakilan/ ciri khas data dari segi fitur.
- **Classification :**
 - Menggunakan fitur dan model/ algoritma pembelajaran untuk menetapkan pola pada suatu kelas tertentu.
- **Post-processing :**
 - Melakukan evaluasi tingkat keberhasilan dalam keputusan.





Pattern Recognition System

- Sistem adalah keseluruhan bagian dari pengenalan pola mulai dari Input, Proses dan Output.



- Teacher/ Human sebagai salah satu sumber basis pengetahuan (knowledge base)



Case Study

- Klasifikasi Ikan Salmon dan Sea Bass.



Salmon



Sea-bass

- Permasalahan : Bagaimana memisahkan ikan yang masuk secara otomatis pada permukaan yang berjalan (mesin) sesuai dengan spesiesnya.
- Asumsikan bahwa kita hanya memiliki dua jenis ikan :
 - Sea bass.
 - Salmon.



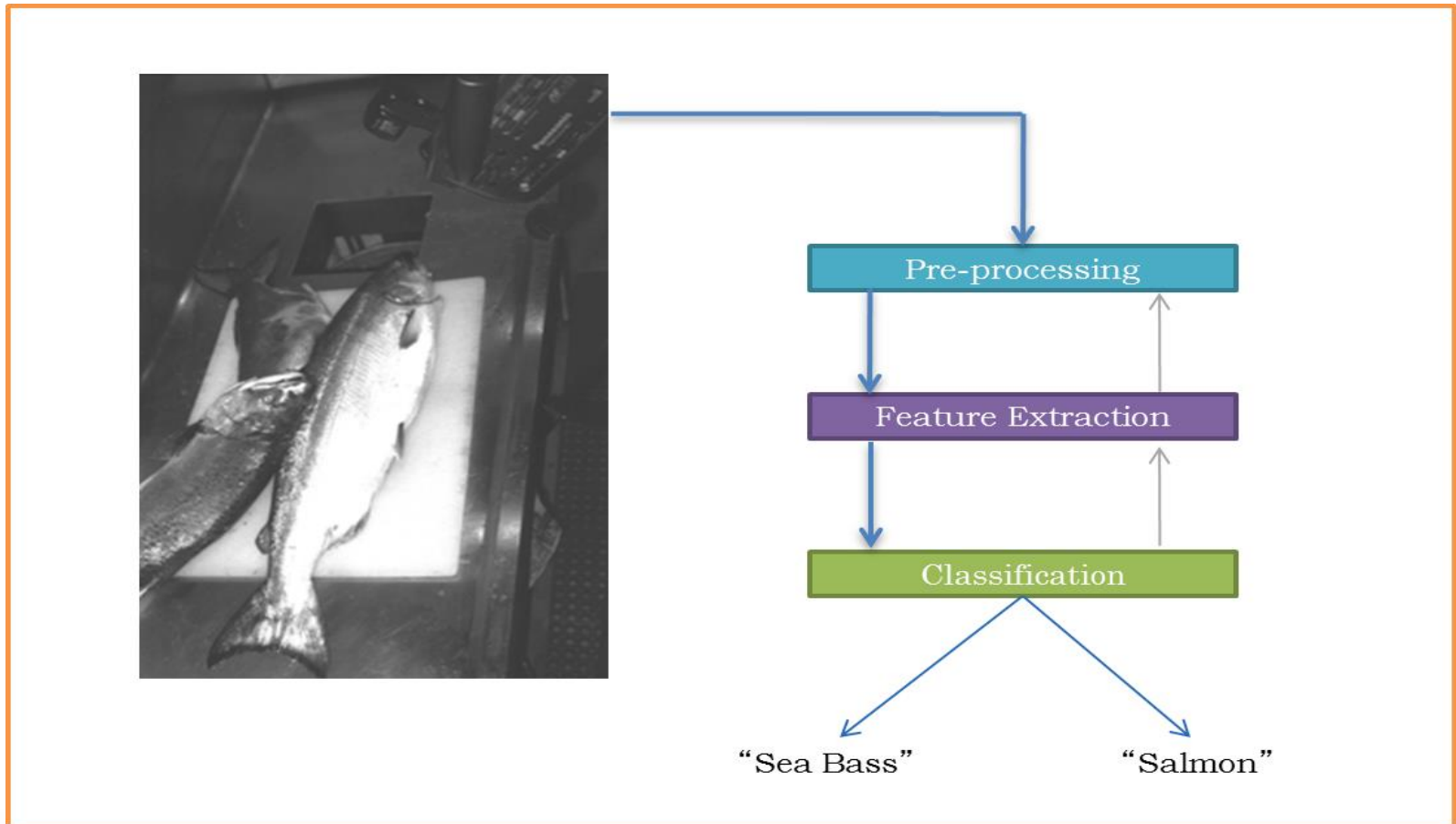
Case Study (Cont.)

- Apa yang bisa menyebabkan masalah selama proses Sensing ?
 - Kondisi pencahayaan.
 - Posisi ikan di permukaan yang berjalan (mesin).
 - Noise dari kamera dan hal-hal lainnya.
- Apa saja langkah-langkah dalam Proses?
 - Mengambil gambar.
 - Memisahkan setiap gambar ikan.
 - melakukan pengukuran.
 - membuat keputusan.



Case Study (Cont.)

- Diagram proses klasifikasi :





Case Study (Cont.)

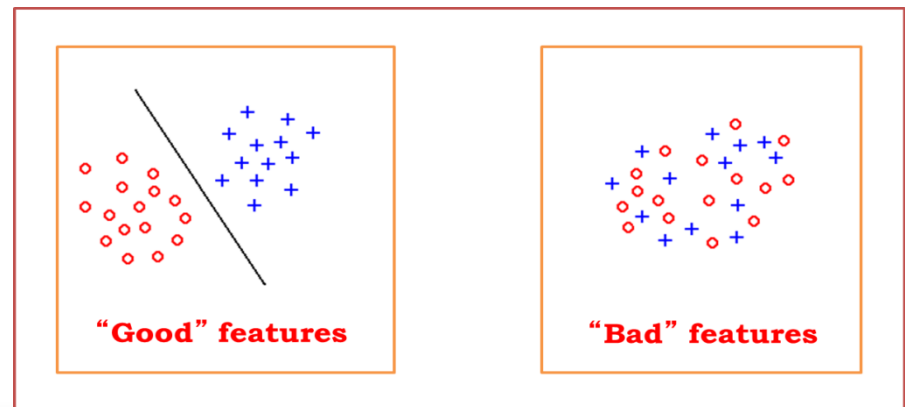
- Pre-Processing :
 - Peningkatan kualitas gambar.
 - Memisahkan titik data yang merupakan representasi dari kedua ikan yang hampir saling bersinggungan atau saling overlap.
 - Menemukan garis boundary/batas pemisah kedua ikan.
- How to separate Sea Bass from Salmon?
 - Menggunakan fitur yang memungkinkan untuk bisa memisahkan : (Panjang ikan, Tingkat kecerahan, Lebar ikan, Jumlah dan Bentuk sirip, Posisi mulut dan fitur lainnya jika masih ada).
 - Asumsi beberapa nelayan mengatakan bahwa "Sea Bass" umumnya lebih panjang daripada "Salmon".
 - Meskipun rata-rata "Sea Bass" lebih panjang dari "Salmon", namun ada banyak contoh di mana ikan yang dilakukan pengamatan menggunakan fitur ini, malah tidak berlaku.



Case Study (Cont.)

- How to separate Sea Bass from Salmon?
 - Untuk meningkatkan hasil recognition/ pengenalan, kita mungkin harus menggunakan lebih dari satu fitur pada suatu waktu.
 - Fitur tunggal kemungkinan besar tidak akan dapat menghasilkan kinerja terbaik.
 - Kombinasi fitur kemungkinan besar akan menghasilkan kinerja yang lebih baik.
- Feature Extraction :

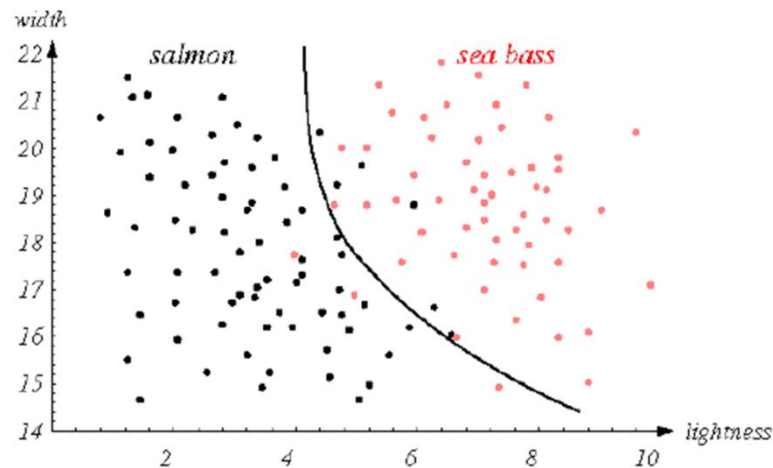
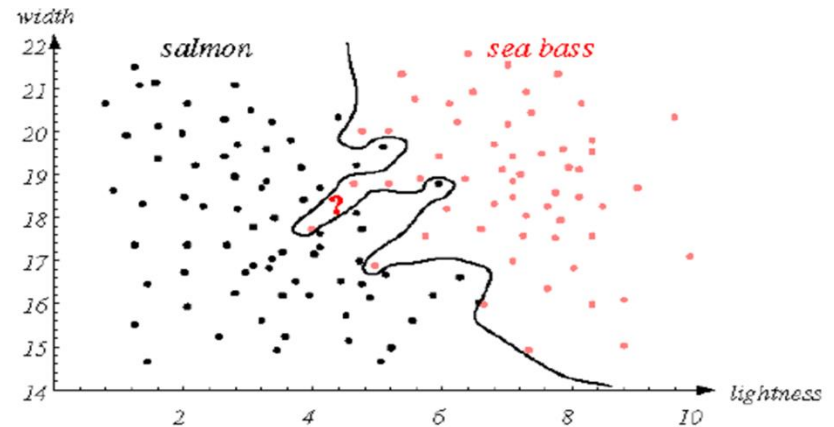
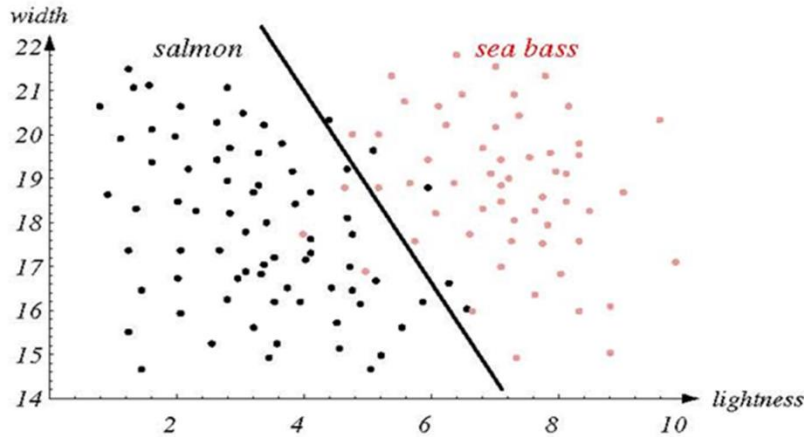
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \begin{array}{l} x_1 : \textit{lightness} \\ x_2 : \textit{width} \end{array}$$





Case Study (Cont.)

- Decision Boundary :



- Model yang lebih kompleks akan menghasilkan boundary/batas pemisah yang lebih kompleks pula.
- Penggunaan fitur yang berbeda akan menghasilkan boundary/batas pemisah keputusan yang berbeda pula.



Case Study (Cont.)

- Decision Boundary :
 - Apakah ada kemungkinan bahwa pelanggan/ pembeli menemukan ikan "Sea Bass" di dalam kumpulan ikan "Salmon"?
 - Kita juga harus mempertimbangkan biaya sesuai dengan tingkat kesalahan yang berbeda dalam setiap keputusan yang kita buat.
 - Sebagai contoh, jika perusahaan melakukan pengepakan ikan dan mengetahui bahwa :
 - Pelanggan yang membeli ikan Salmon akan sangat kecewa jika mereka melihat ikan Sea Bass dalam kaleng mereka.
 - Pelanggan yang membeli ikan Sea Bass tidak akan merasa senang jika mereka sering melihat ikan Salmon yang lebih mahal di kaleng yang mereka beli. Ini sepertinya buka masalah harga, tapi selera ikan dari pelanggan.
 - Bagaimanapun, pengetahuan yang dimiliki pelanggan akan mempengaruhi keputusan untuk membeli ikan kaleng.



Case Study (Cont.)

- Isu-Isu terkait dengan Ekstraksi Fitur :
 - Apakah ada kemungkinan bahwa pelanggan/ pembeli menemukan ikan "Sea Bass" di dalam kumpulan ikan "Salmon"?
 - Fitur yang berkorelasi besar tidak akan meningkatkan kinerja.
 - Kemungkinan akan ada kesulitan untuk mengekstraksi fitur tertentu.
 - Dibutuhkan komputasi yang cukup mahal untuk mengekstraksi banyak fitur.
 - Adanya fitur yang hilang.
 - Domain pengetahuan yang digunakan.



The Design Cycle

- **Collect Data**
 - Mengumpulkan data training dan data testing
- **Choose Features**
 - Disesuaikan dengan domain data.
- **Choose Model**
 - Disesuaikan dengan domain data.
- **Training**
 - Supervised learning.
 - Unsupervised learning.
- **Evaluate**
 - Menghitung kinerja dengan fitur data yang digunakan.

Selesai

