IF5181 Pengenalan Pola

Automatic Fish Sorting

Masayu Leylia Khodra

Referensi: Bab 1 dari R. Duda, P. Hart, and D. Stork, Pattern Classification, 2nd ed. John Wiley & Sons, 2001

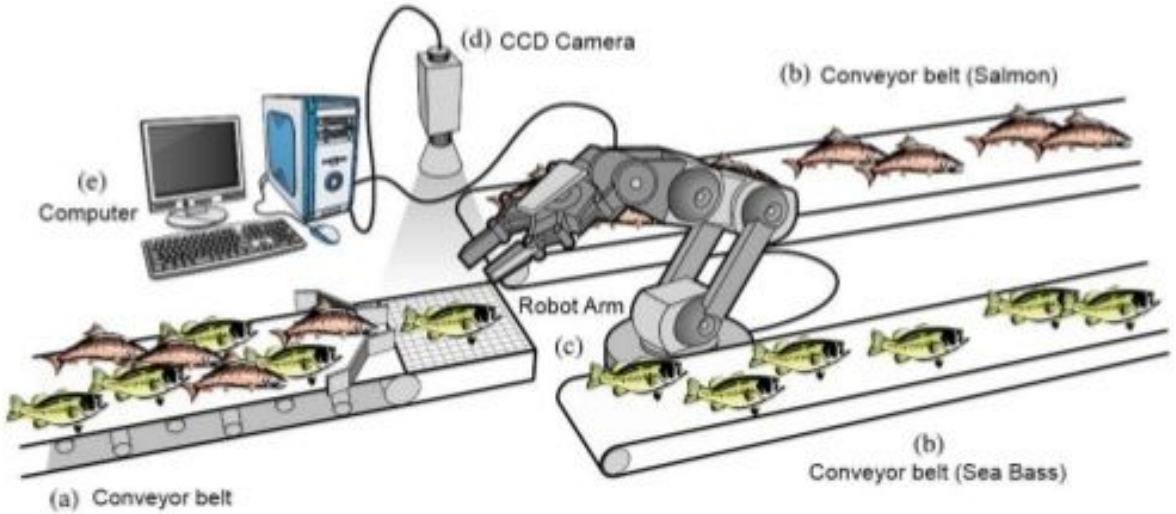
Outline

- Overview
  - Tiga pendekatan pengenalan pola
  - PR - AI - ML - DL
- Kasus: automatic fish sorting pada fish packing plant
  - Deskripsi persoalan
  - Training samples, fish features
  - Tentative vs Optimal Model

Fish Packing Plant

- Fish packing plant wants to automate the process of sorting incoming fish on a conveyor belt according to species.

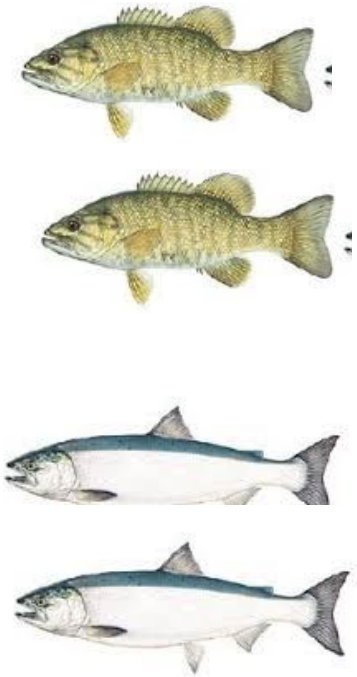
- A: Conveyor belt for fish
- B: Conveyor belt for classified fish
- C : Robot arm for grabbing fish
- D: Machine vision system with CCD camera
- E : Computer that analyze fish image and control the robot arm



Automated Fish Classification System

Training Samples

Preprocessing:
- image enhancement
- resize
- segmentation



Fish features

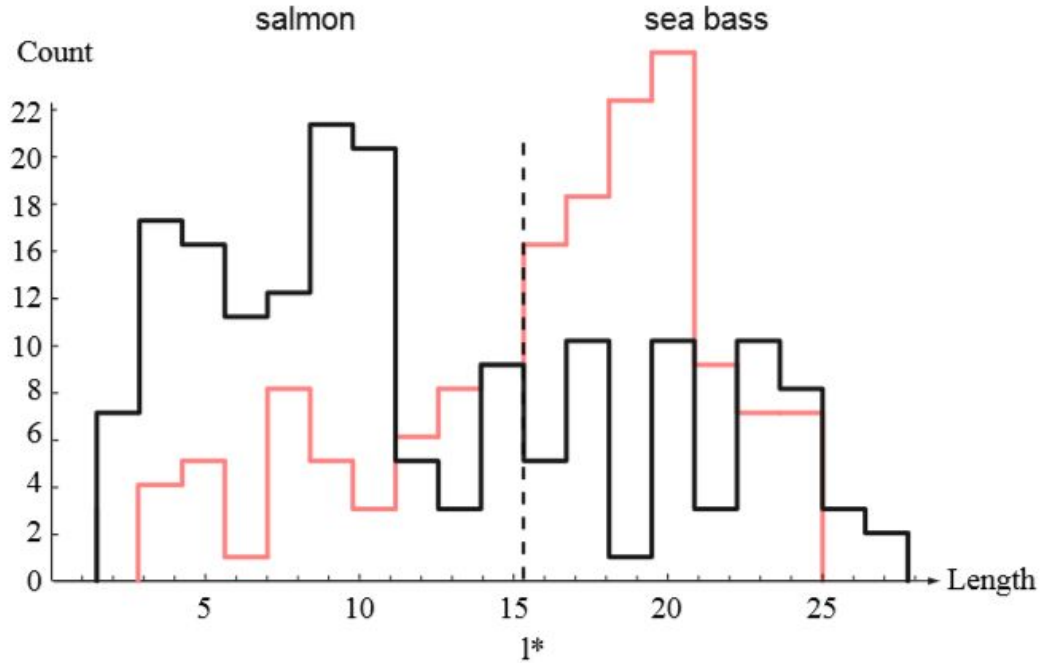
- Piksel
- Descriptor based on : color, shape, textures

Find physical differences between the two types of fish:

- Length
- Lightness
- Width
- Number and shape of fins
- Position of the mouth,
- etc…

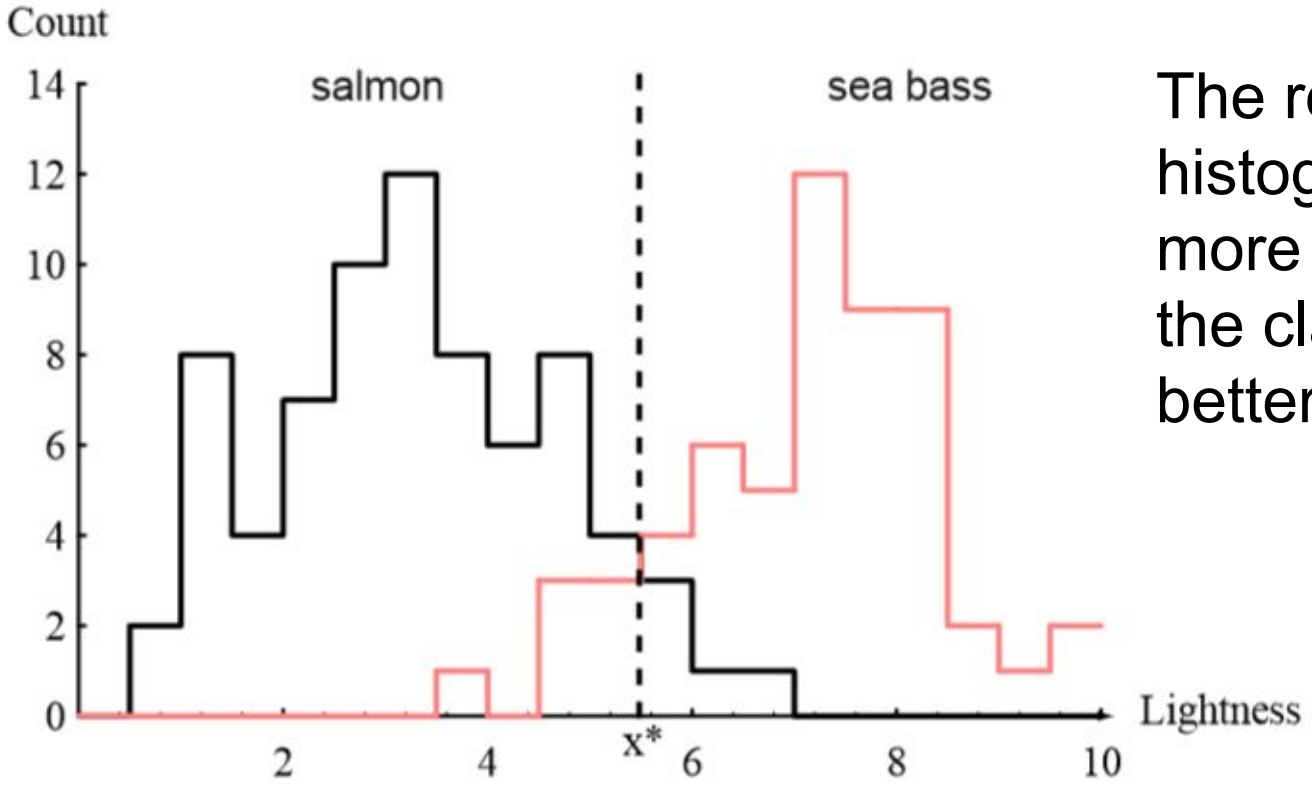
Tentative Model: Length Feature

Suppose somebody at the fish plant tells us that a sea bass is generally longer than a salmon.



- From histograms, sea bass are longer than salmon, on average, but it is clear that this single criterion is quite poor;
- No matter how we choose $l*$, we cannot reliably separate sea bass from salmon by length alone.
- The value $l*$ marked will lead to the smallest number of errors, on average.

Tentative Model: Lightness Feature

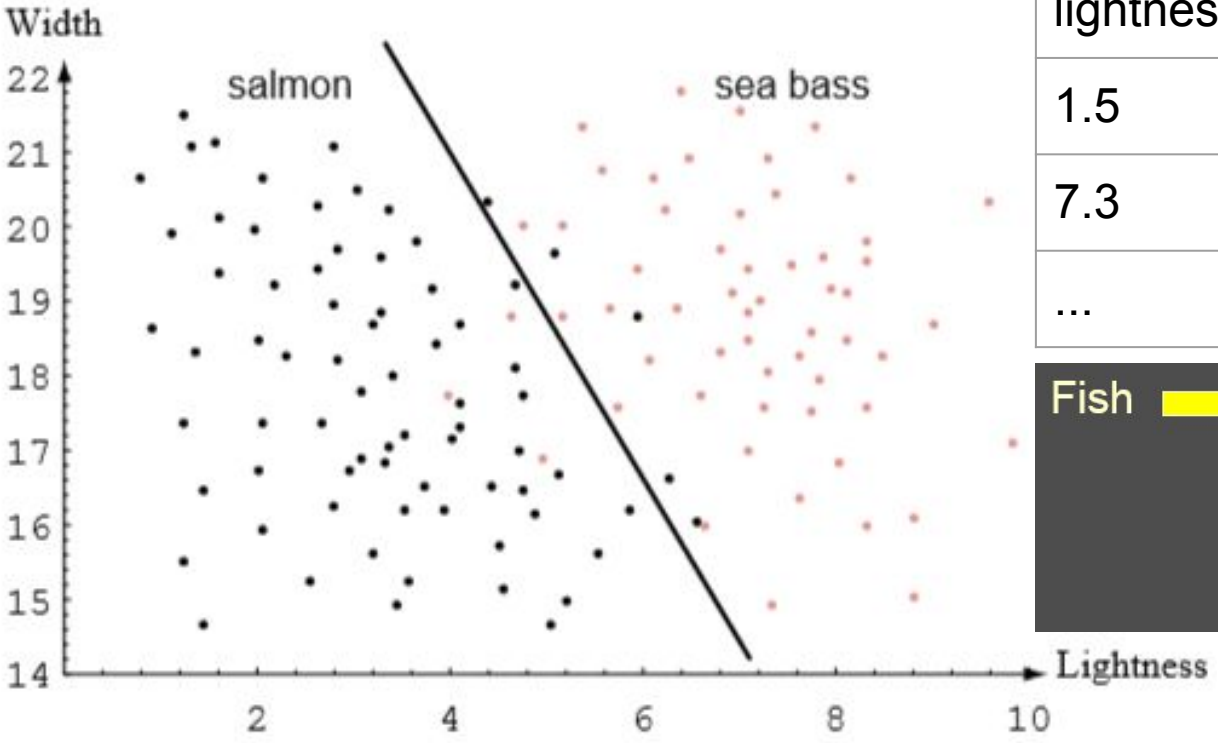


The resulting histograms are much more satisfactory — the classes are much better separated.

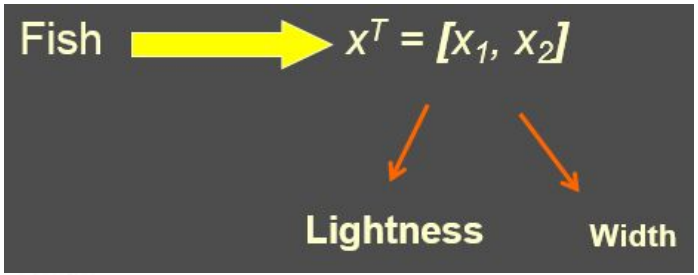
Threshold decision boundary and cost relationship

- Deciding the fish was a sea bass when in fact it was a salmon was just as undesirable as the converse.
  - salmon in “sea bass” cans vs sea bass in “salmon” cans ?
- Move our decision boundary (x*) toward smaller values of lightness in order to minimize the cost.
  - Reduce the number of sea bass that are classified salmon.
  - More salmon makes its way into the cans of sea bass.

Tentative Model: Lightness and Width Feature



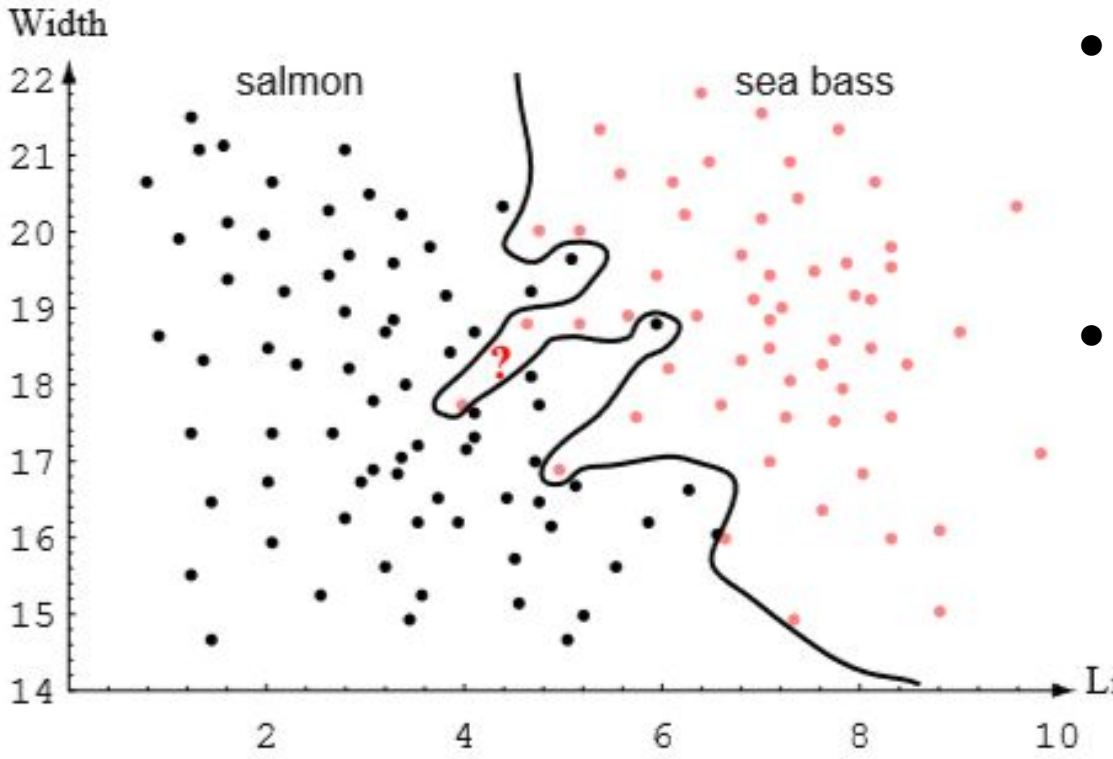
lightness	width	category
1.5	14.6	salmon
7.3	15	Sea bass
...		



Rule based on 2 Features

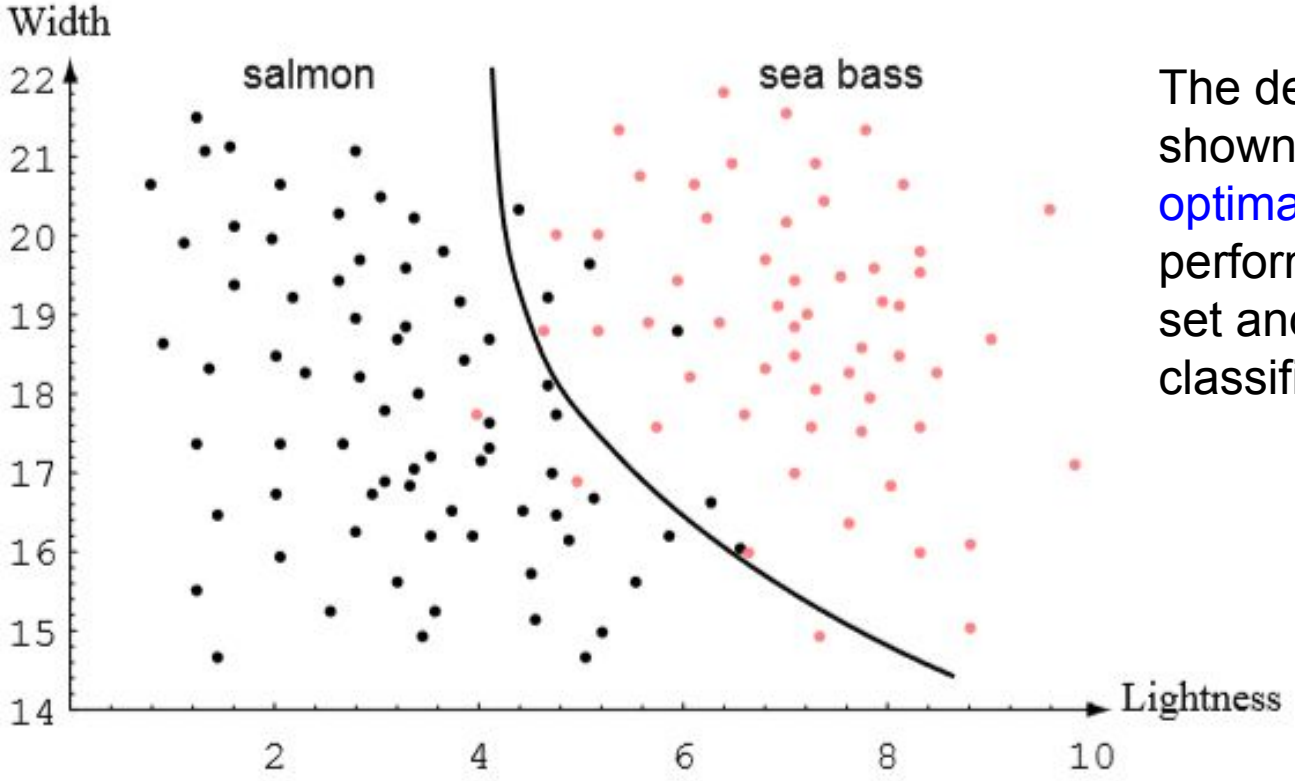
Given decision boundary, the following rule for separating the fish: Classify the fish as sea bass if its feature vector falls above the decision boundary shown, and as salmon otherwise.

Tentative Complex Model: Training Accuracy 100%



- While such a decision may lead to perfect classification of our training samples, it would lead to poor performance on future patterns.
- The novel test point marked ? is evidently most likely a salmon, whereas the complex decision boundary shown leads it to be misclassified as a sea bass.

Optimal Model: Better Generalization



The decision boundary shown might represent the optimal tradeoff between performance on the training set and simplicity of classifier.

Representasi

- Task berbeda membutuhkan set fitur berbeda ⇒ decision boundary berbeda.
- Tantangan sulit: satu general purpose artificial pattern recognition device, yaitu untuk berbagai task yang bervariasi.
- Representasi ideal: pola berkategori sama “dekat” satu sama lain (variasi intra class rendah), dan “jauh” dengan pola berkategori lain (variasi inter class besar).

Representasi (2)

- Semakin sedikit fitur, semakin simple decision boundary, dan semakin mudah proses training-nya.
- Robust features: relatif tidak sensitif terhadap noise atau error lainnya.