Pattern Recognition: An introduction

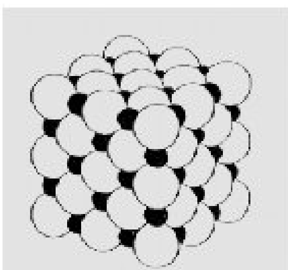
Jun Hu

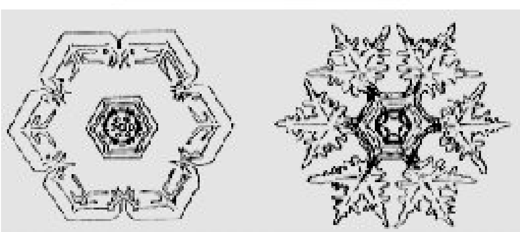
October, 2007

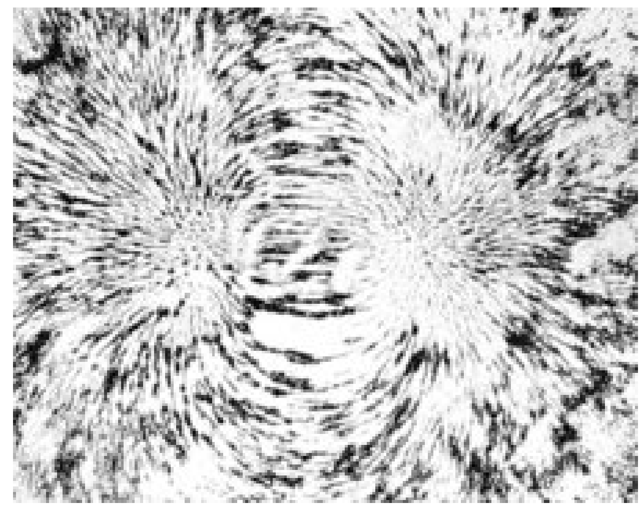
Designed Intelligence
Department of Industrial Design
Eindhoven University of Technology

What are patterns?

- Patterns in physics and chemistry

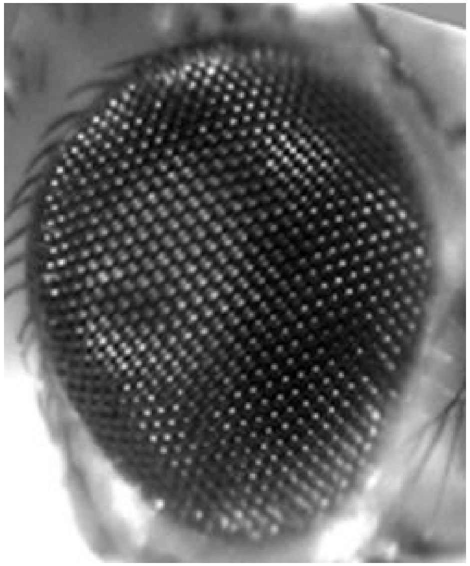






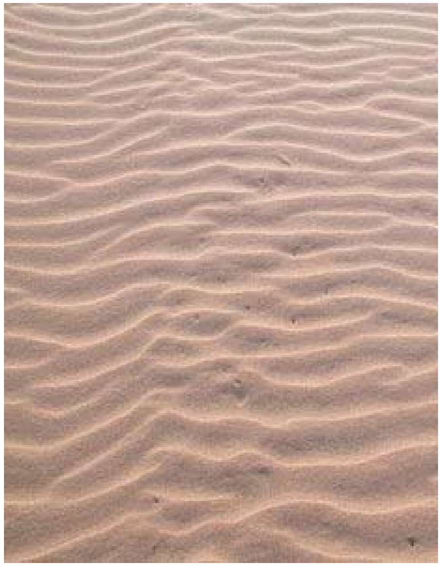
What are patterns?

- Patterns in biology



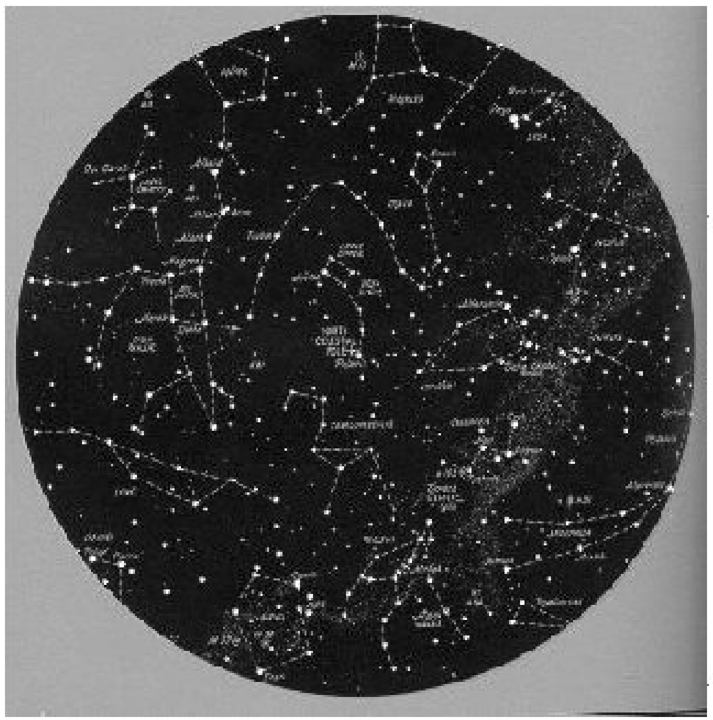
What are patterns?

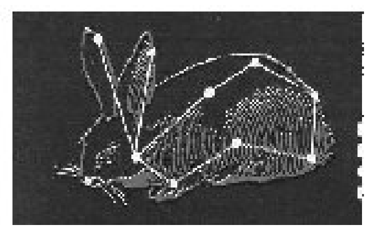
- Patterns in nature



What are patterns?

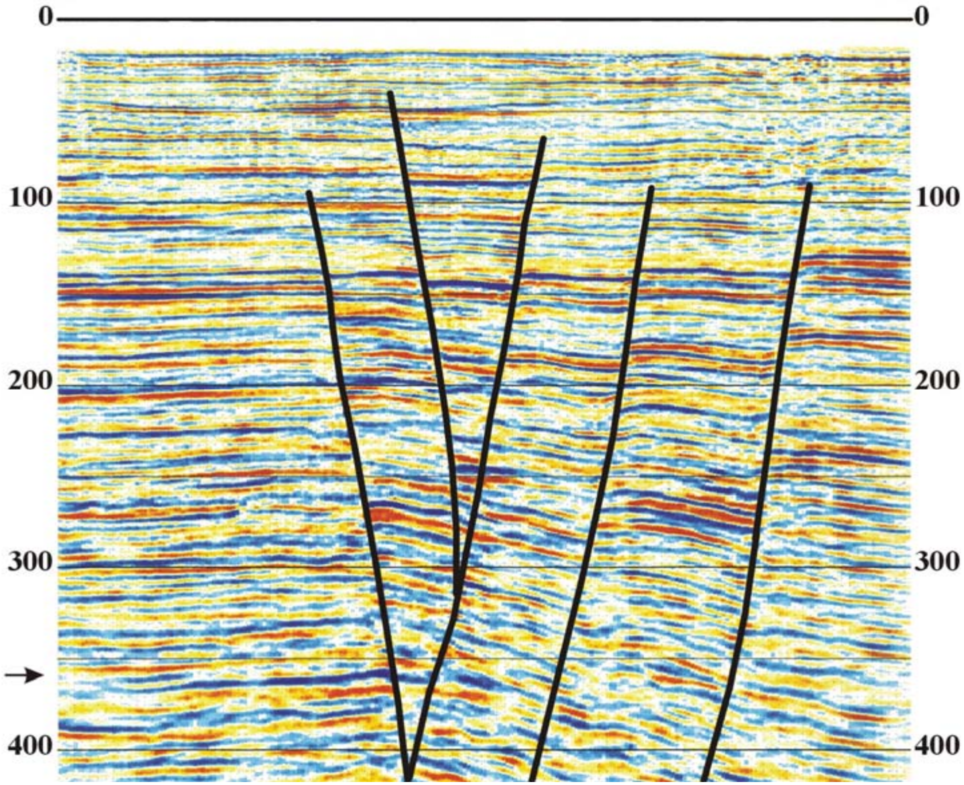
- Patterns in Astronomy





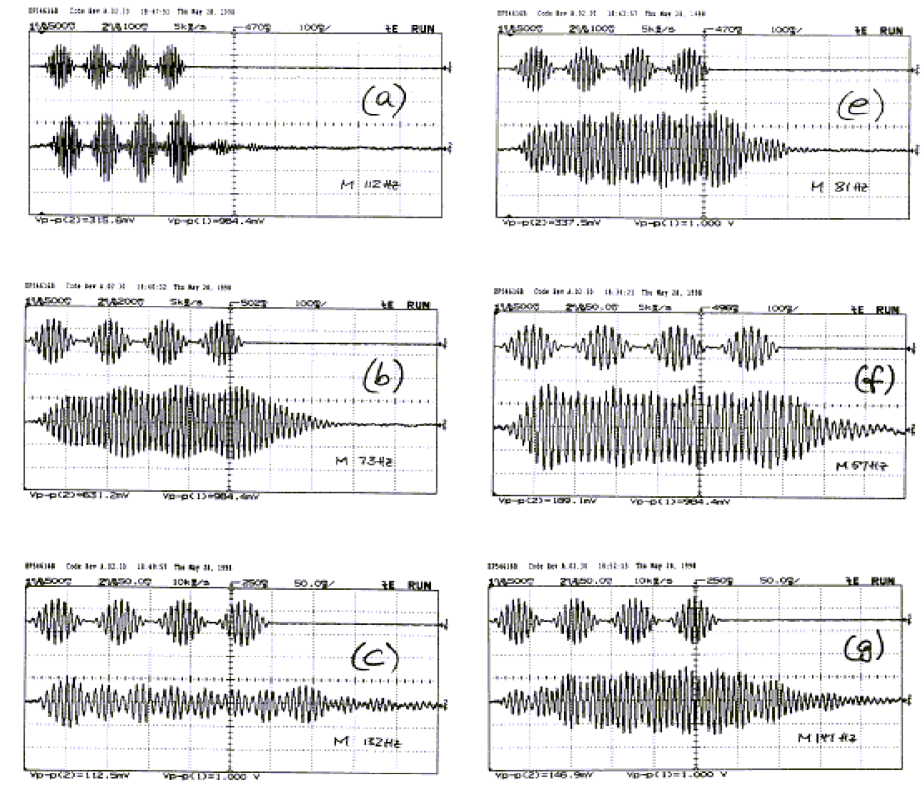
What are patterns?

- Pattern in seismic data



What are patterns?

- Patterns in sound



What are patterns?

- Patterns in faces

What are patterns?

- A **pattern** is an object, process or event that can be given a name.
- A pattern **class** (or category) is a set of patterns sharing common features.

What is pattern recognition?

- Our ability
  - to recognize a face,
  - to understand spoken words,
  - to read handwritten characters
- … all these abilities belong to the complex processes of
- pattern recognition
  - *The act of taking in raw data and taking an action based on the “category” of the pattern.*

Goal of Pattern recognition

- Recognize Patterns. Make decisions about patterns.
- Examples:
  - Visual – is this person happy or sad?
  - Speech – did the speaker say “Yes” or “No”?
  - Physics – is this an atom or a molecule?

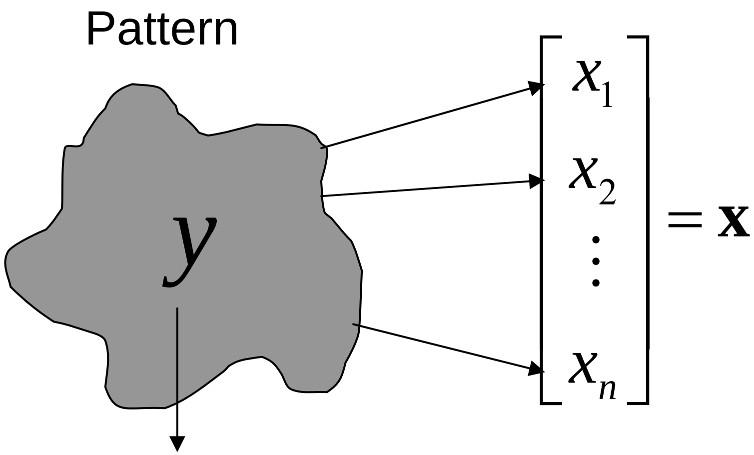
Applications of Pattern Recognition

- OCR
  - Handwritten digit/letter recognition
  - Printed texts: reading machines for blind people,
- Biometrics:
  - Face recognition
  - Speech recognition
  - Finger prints recognition.
  - Interpreting DNA sequences
- Medical diagnosis
  - X-Ray, EEG, EKG
- Military applications
  - Automated Target Recognition (ATR)
  - recognition from aerial or satelite photographs
- Spam detection
- Smell recognition (e-nose, sensor networks)

Approaches

- **Statistical PR**
  - based on underlying statistical model of patterns and pattern classes.
- **Structural (syntactical) PR**
  - pattern classes represented by means of formal structures as grammars, automata, strings, etc.
- **Connectionist's approach**
  - classifier is represented as a network of cells modeling neurons of the human brain.
  - (Neural networks)

Basic concepts



<u>Feature vector</u> $\mathbf{x} \in X$

- A vector of observations (measurements).

- $\mathbf{x}$ is a point in feature space X.

<u>Hidden state</u> $y \in Y$

- Cannot be directly measured.

- Patterns with equal hidden state belong to the same class.

<u>Task</u>

- To design a classifer (decision rule) $\mathrm{q} : X \rightarrow Y$

which decides about a hidden state based on an onbservation.

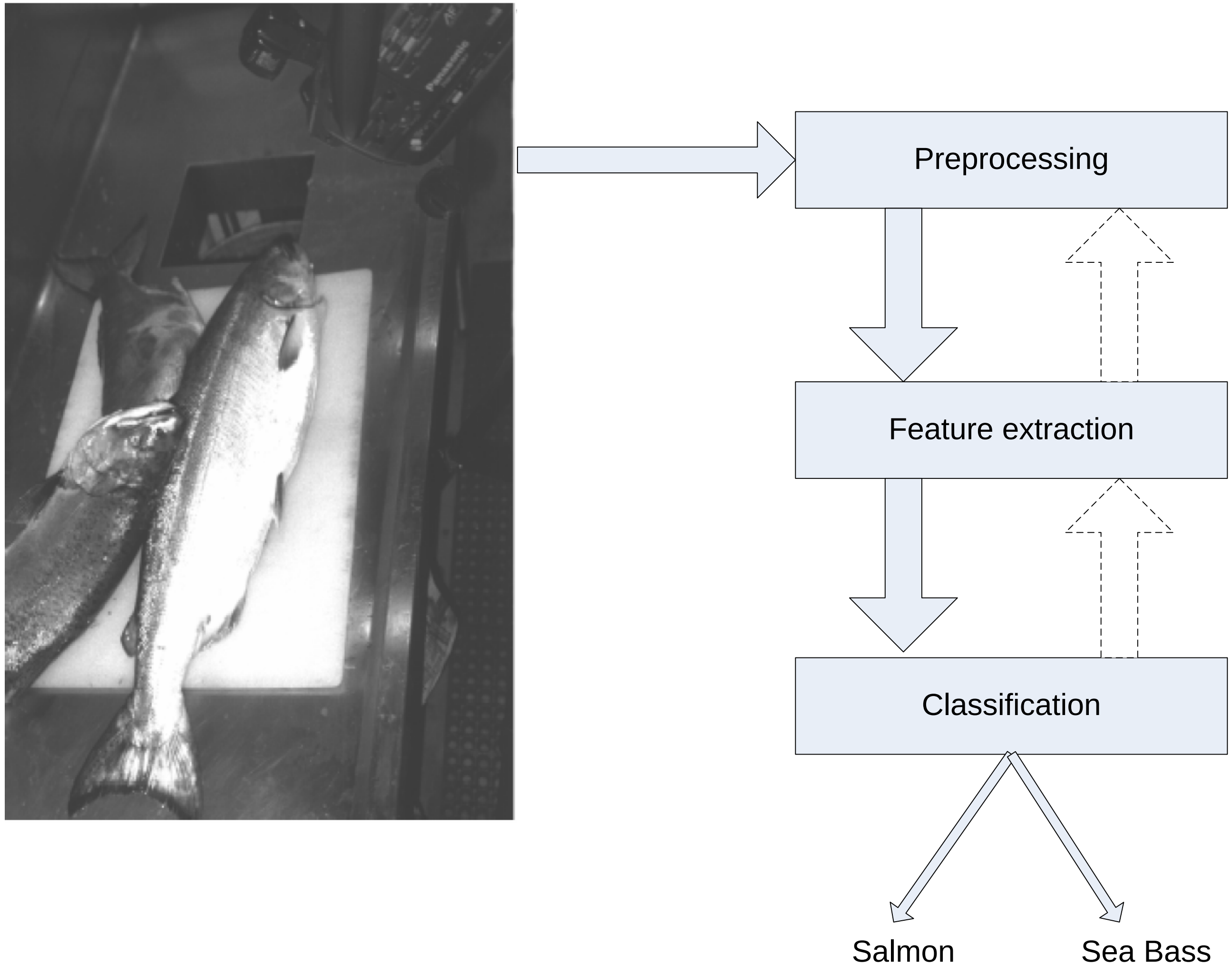
An example: Statistical PR

- Fish recognition using optical sensing
- Salmon vs. Sea bass

Analysis

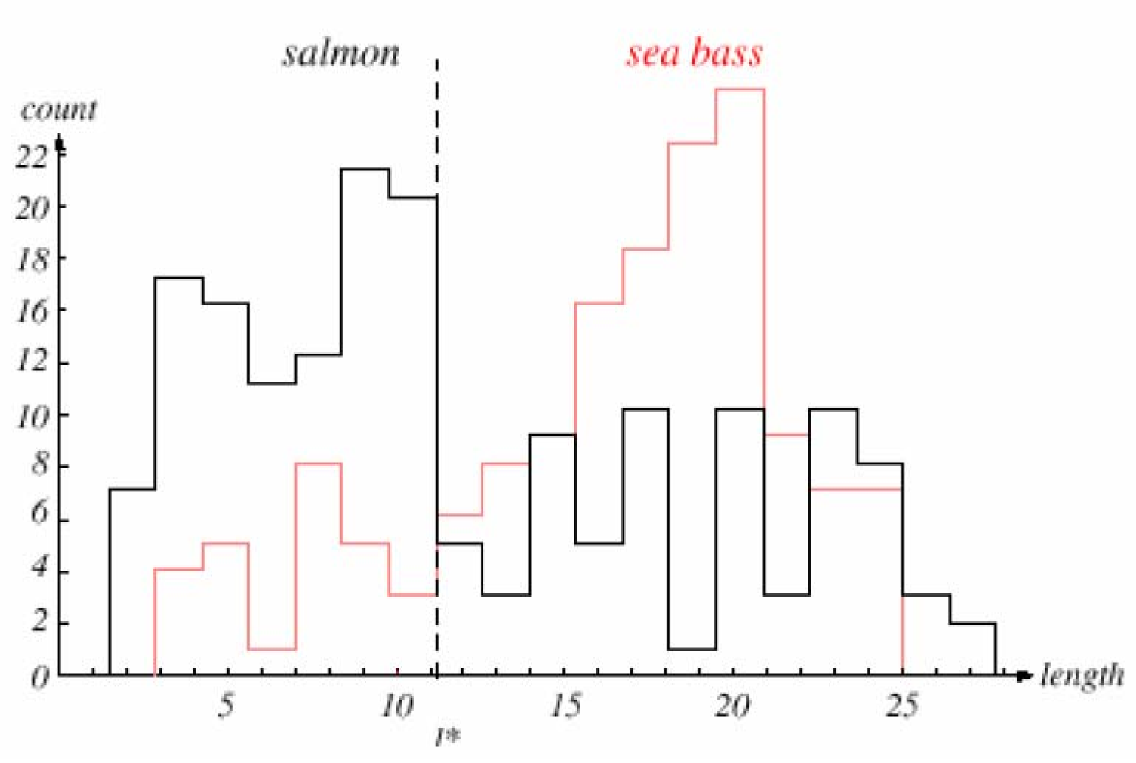
- First take some sample images to extract features that could help to distinguish them
  - Length
  - Lightness
  - Width
  - Number and shape of fins
  - Position of the mouth
  - etc…

Process


Preprocessing
Feature extraction
Classification
Salmon
Sea Bass

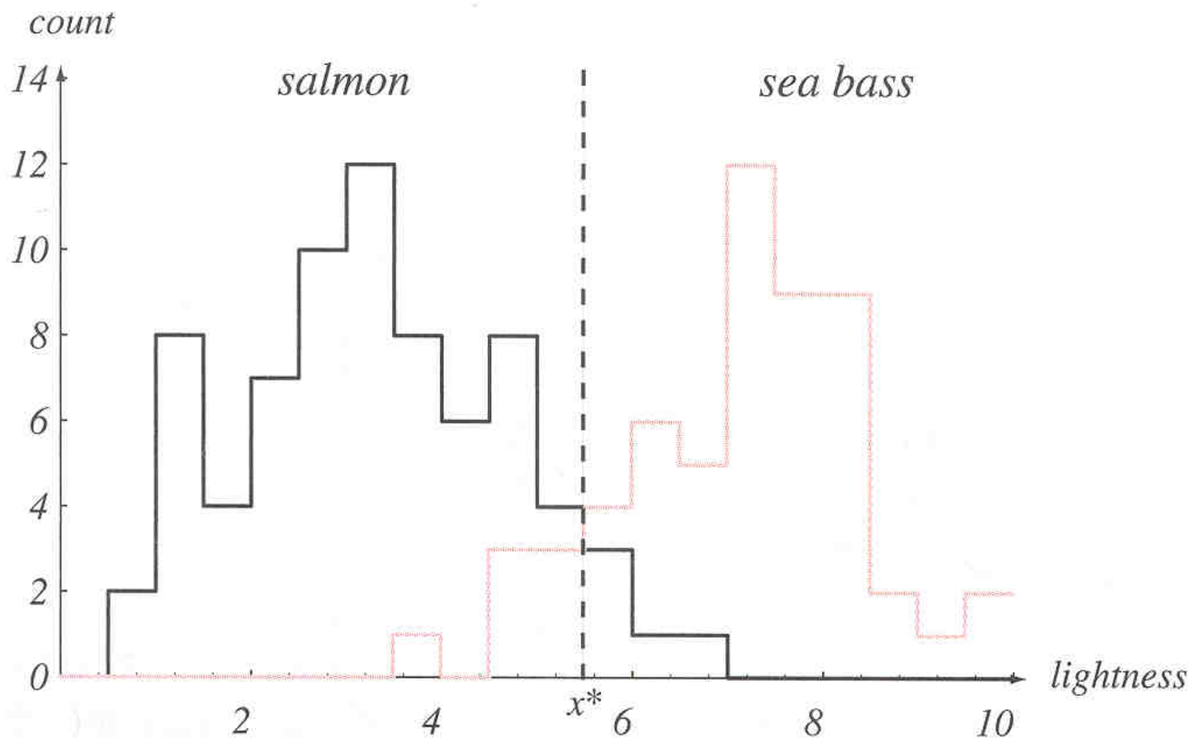
Analysis

- Length?



Analysis

- Lightness?



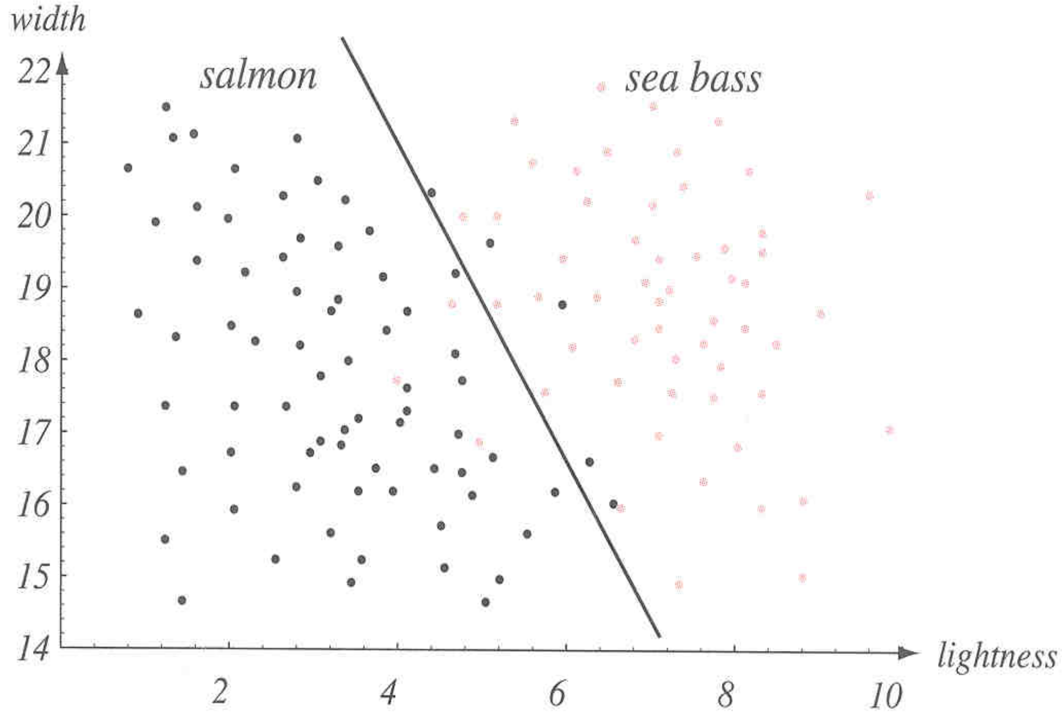
Analysis

- More than one feature?
  - Adopt the lightness and add the width
  - Sea bass is typically wider than salmon.

$$x = (\underbrace{x_1}_{\textit{Lightness}}, \underbrace{x_2}_{\textit{width}})$$

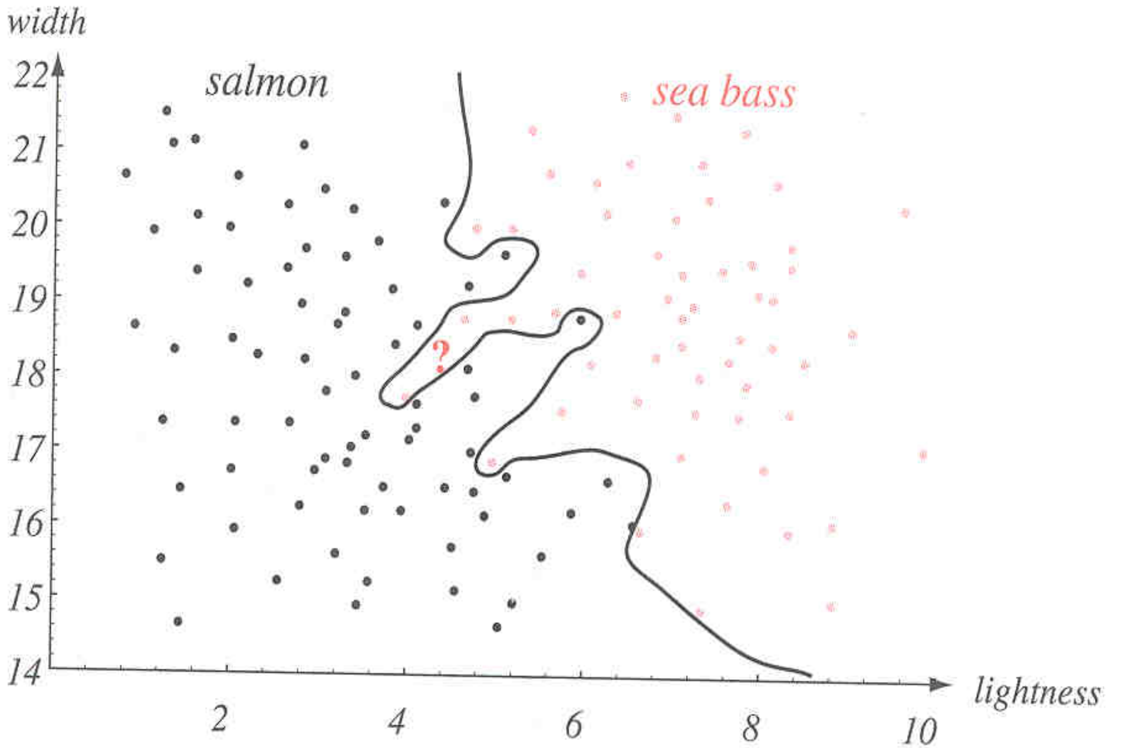
Solution 1

- Underfitting
  - Shall we add another feature?



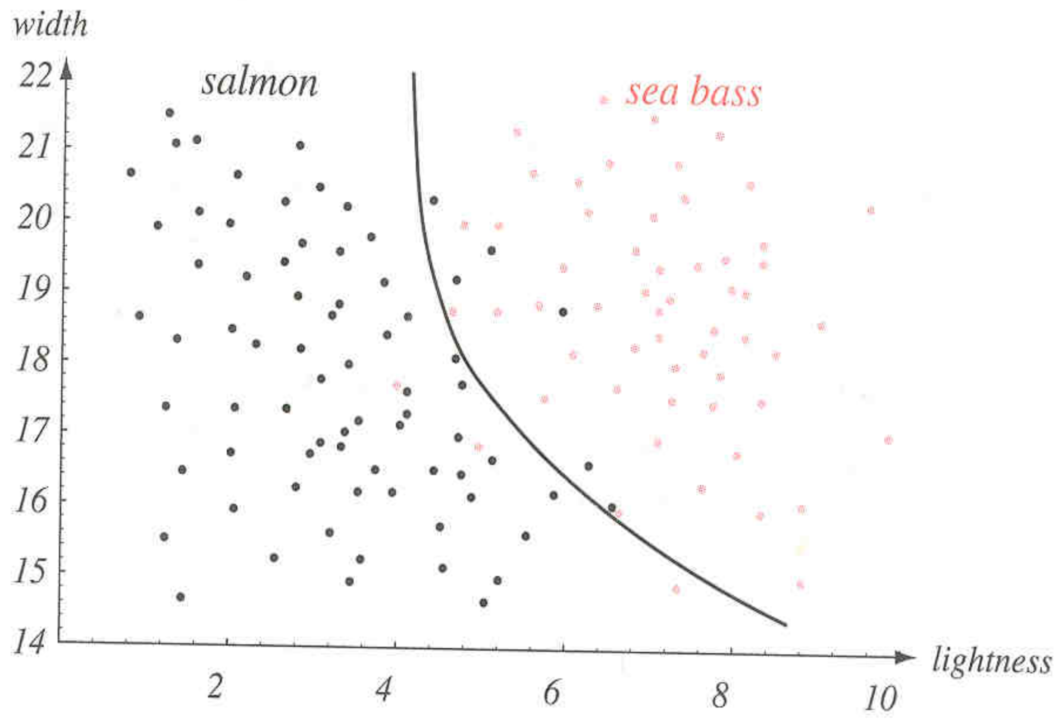
Solution 2

- Better. But what about generalization?

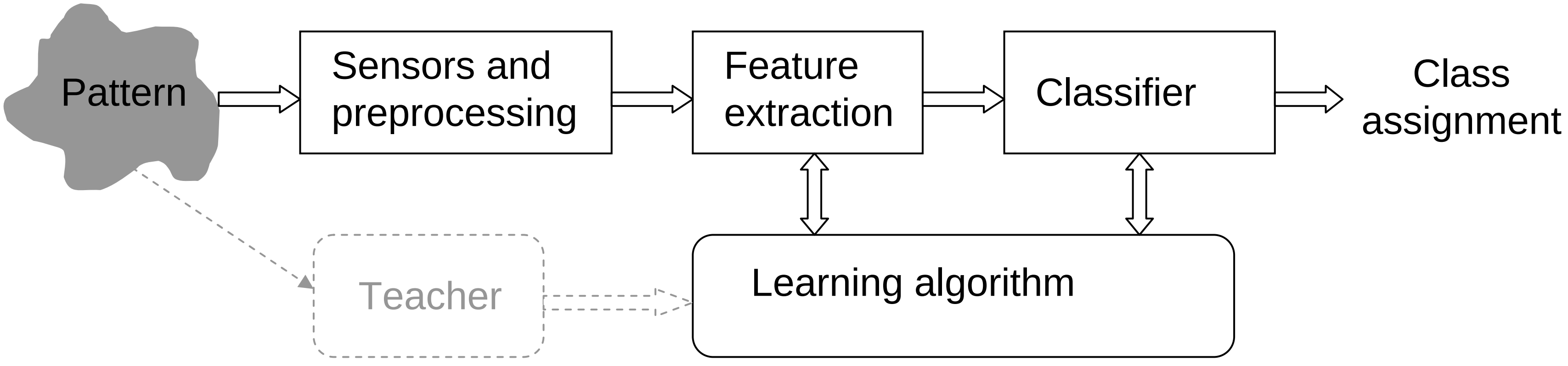


Solution 3

- Good fit. Hope it will perform well on novel patterns.



Components of PR system



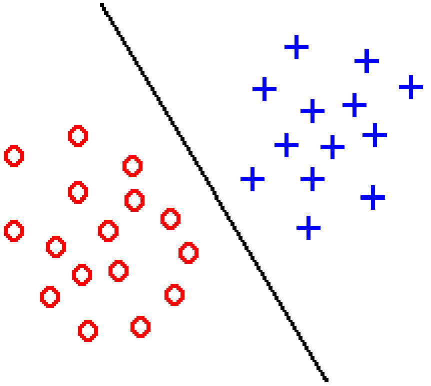
- **Sensors and preprocessing**.
- **A feature extraction** to create discriminative features good for classification.
- **A classifier**.
- **A teacher** provides information about hidden state -- supervised learning.
- **A learning algorithm** sets PR from training examples.

Feature extraction

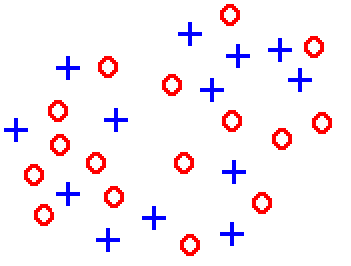
<u>Task:</u> to extract features which are good for classification.

Good features:

- Objects from the same class have similar feature values.
- Objects from different classes have different values.



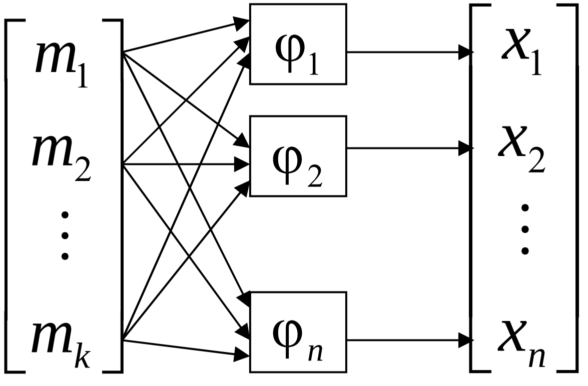
“Good” features



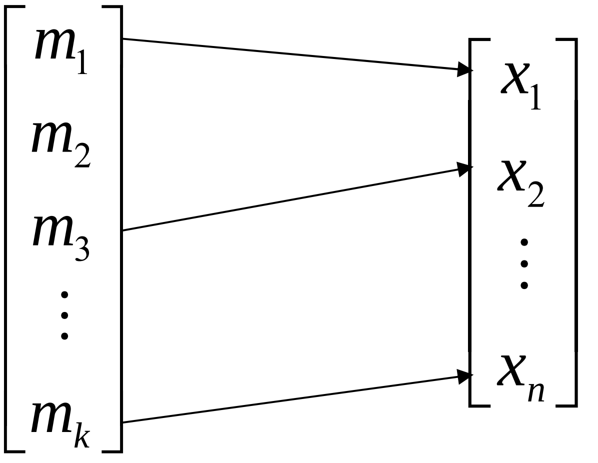
“Bad” features

Feature extraction methods

Feature extraction



Feature selection



Problem can be expressed as optimization of parameters of featrure extractor $\varphi(\theta)$.

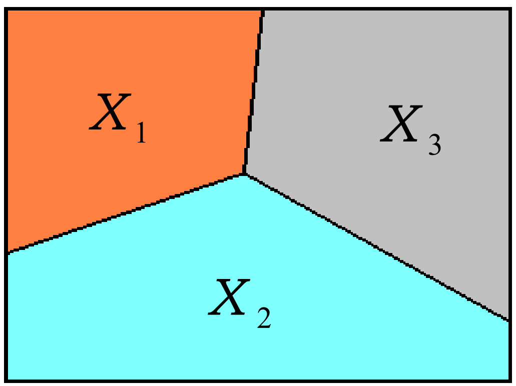
Supervised methods: objective function is a criterion of separability (discriminability) of labeled examples, e.g., linear discriminat analysis (LDA).

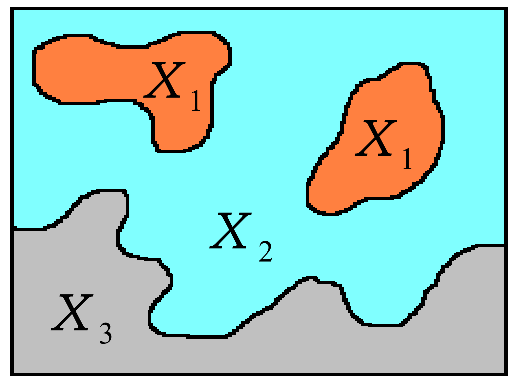
Unsupervised methods: lower dimesional representation which preserves important characteristics of input data is sought for, e.g., principal component analysis (PCA).

Classifier

A classifier partitions feature space *X* into **class-labeled regions** such that

$$X = X_1 \cup X_2 \cup \ldots \cup X_{|Y|} \quad \text{and} \quad X_1 \cap X_2 \cap \ldots \cap X_{|Y|} = \{0\}$$





The classification consists of determining to which region a feature vector **x** belongs to.

Borders between **decision boundaries** are called decision regions.

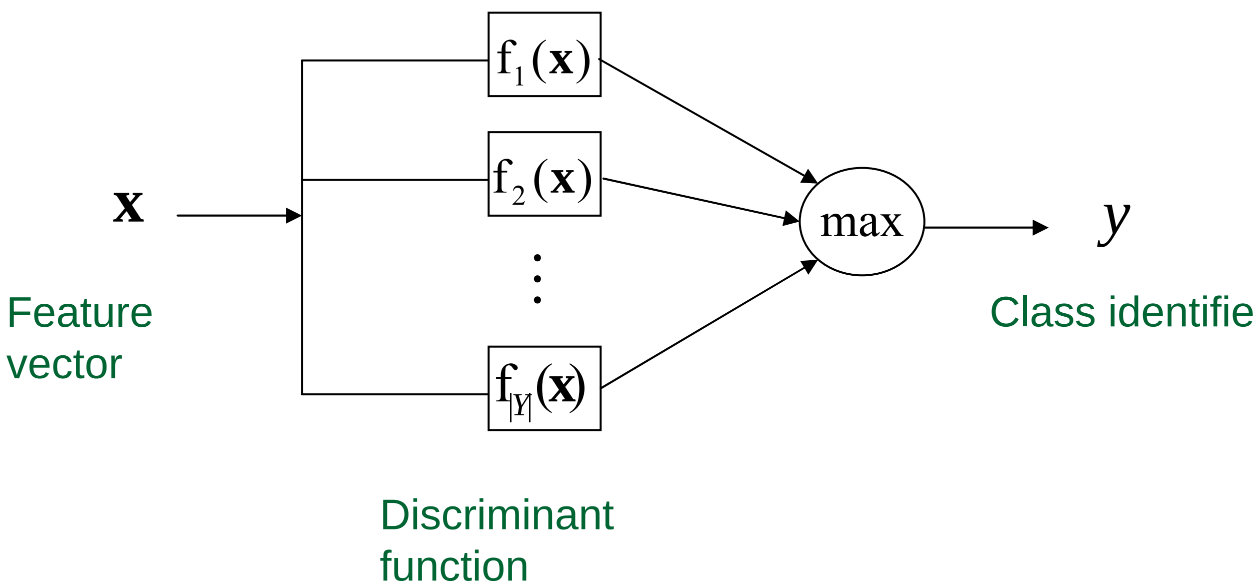
Representation of classifier

A classifier is typically represented as a set of discriminant functions

$$\mathrm{f}_i(\mathbf{x}) : X \to \Re, i = 1, \ldots, |Y|$$

The classifier assigns a feature vector **x** to the *i*-the class if

$$\mathrm{f}_i(\mathbf{x}) > \mathrm{f}_j(\mathbf{x}) \quad \forall j \neq i$$



Bayesian decision making

• The Bayesian decision making is a fundamental statistical approach which allows to design the optimal classifier if complete statistical model is known.

Definition:	Obsevations	X	A loss function	$W : Y \times D \to R$
	Hidden states	Y	A decision rule	$\mathrm{q} : X \to D$
	Decisions	D	A joint probability	$\mathrm{p}(\mathbf{x}, y)$

Task: to design decision rule q which minimizes Bayesian risk

$$\mathrm{R}(\mathrm{q}) = \sum_{y \in Y} \sum_{x \in X} \mathrm{p}(\mathbf{x}, y)\, \mathrm{W}(\mathrm{q}(\mathbf{x}), y)$$

Wrap-up

- Statistical and Syntactical approaches
- Neural networks: non-linear statistical
- Every step is important: