Supervised Learning:

Pattern Recognition System

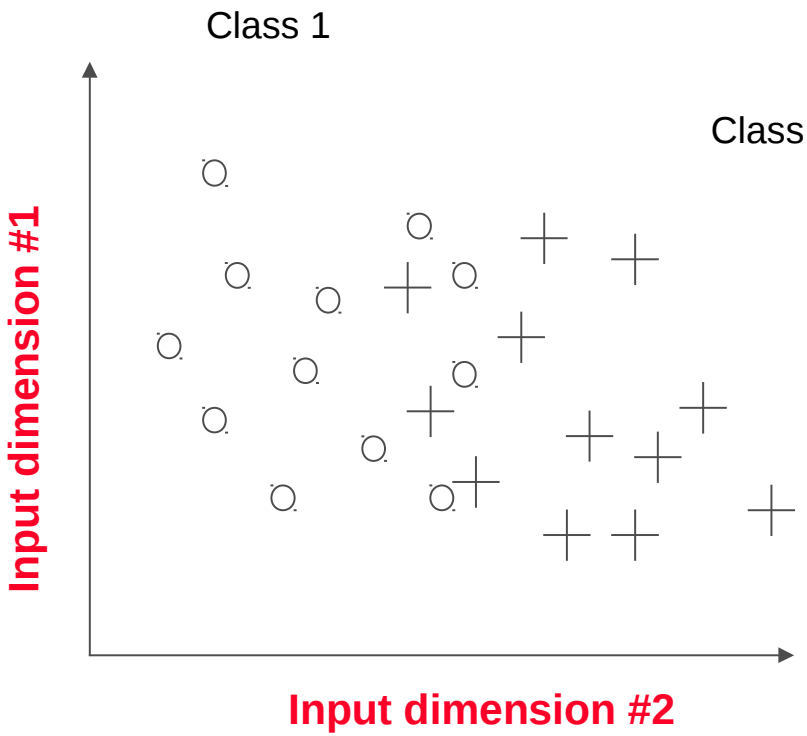
Supervised Learning

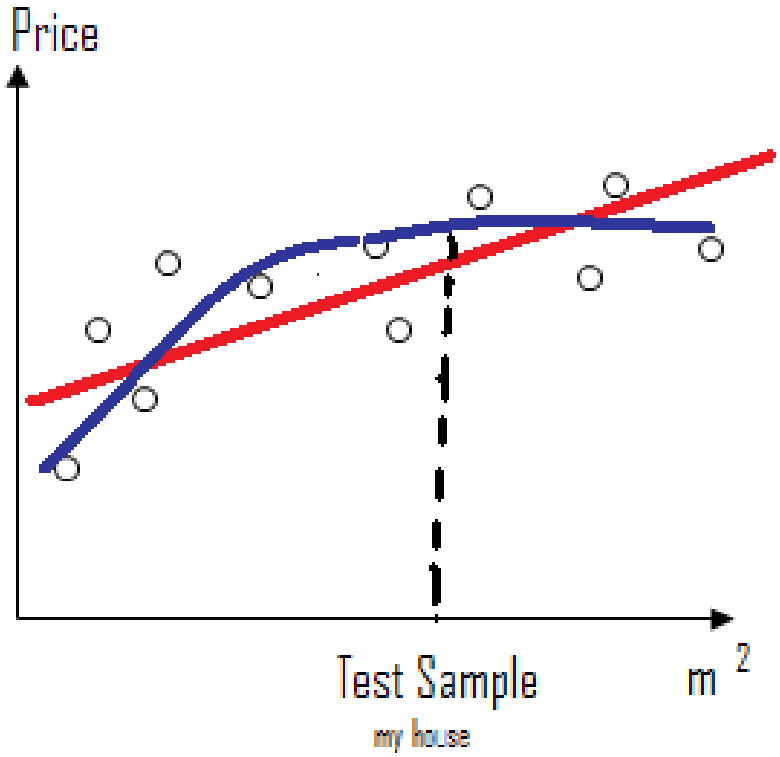
- **In supervised learning, the aim is to learn a mapping from the input to an output whose correct values are provided by a supervisor.**

- **There are many applications of machine learning in *pattern recognition.***
  - ***optical character recognition***
  - ***face recognition***
  - ***medical diagnosis (ECG, EEG signal classification)***
  - ***speech recognition***
  - **time-series prediction**

Supervised Learning

- Training Set
- Input Variable / “Features”
- Output Variable / “Target”





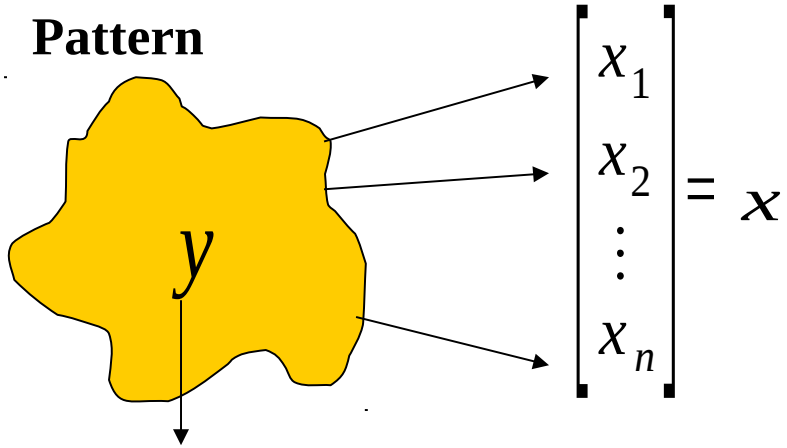
Classification

- **Target: Discrete**

Regression

Target: Continues

Basic concepts



Feature vector $x \equiv X$

- A vector of observations (measurements).
- x s a point in feature space X .

Hidden state $y \in Y$

- Cannot be directly measured.
- Patterns with equal hidden state belong to the same class.

Task

- To design a classifer (decision rule) $q: X \rightarrow Y$ which decides about a hidden state based on an onbservation.

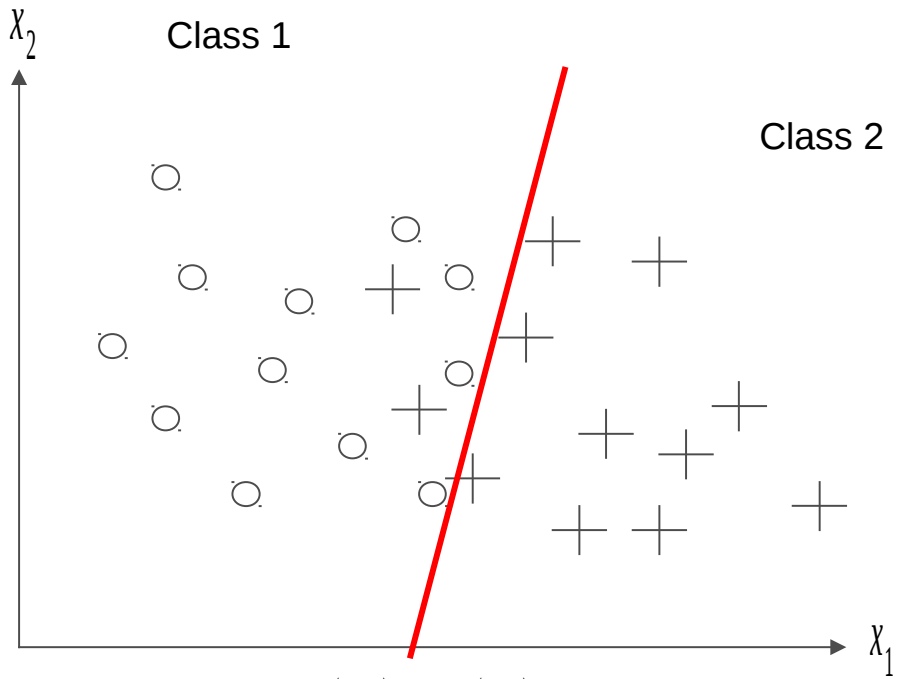
$$y = q(x) = a_0 + a_1 x_1 \ldots + a_n x_n$$

Hypothesis (q) _ Decision Surface

$$y = q(x) = a_0 + a_1 x_1 \ldots + a_n x_n$$

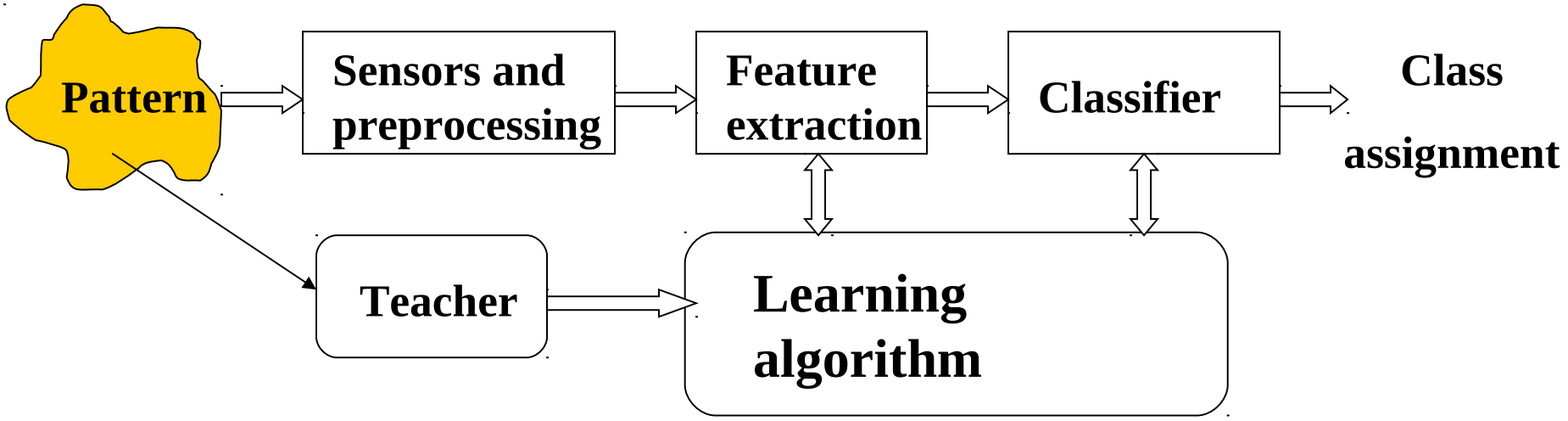
$$y = q(x) = \sum_{i=0}^{i=n} a_i x_i$$

N: number of features



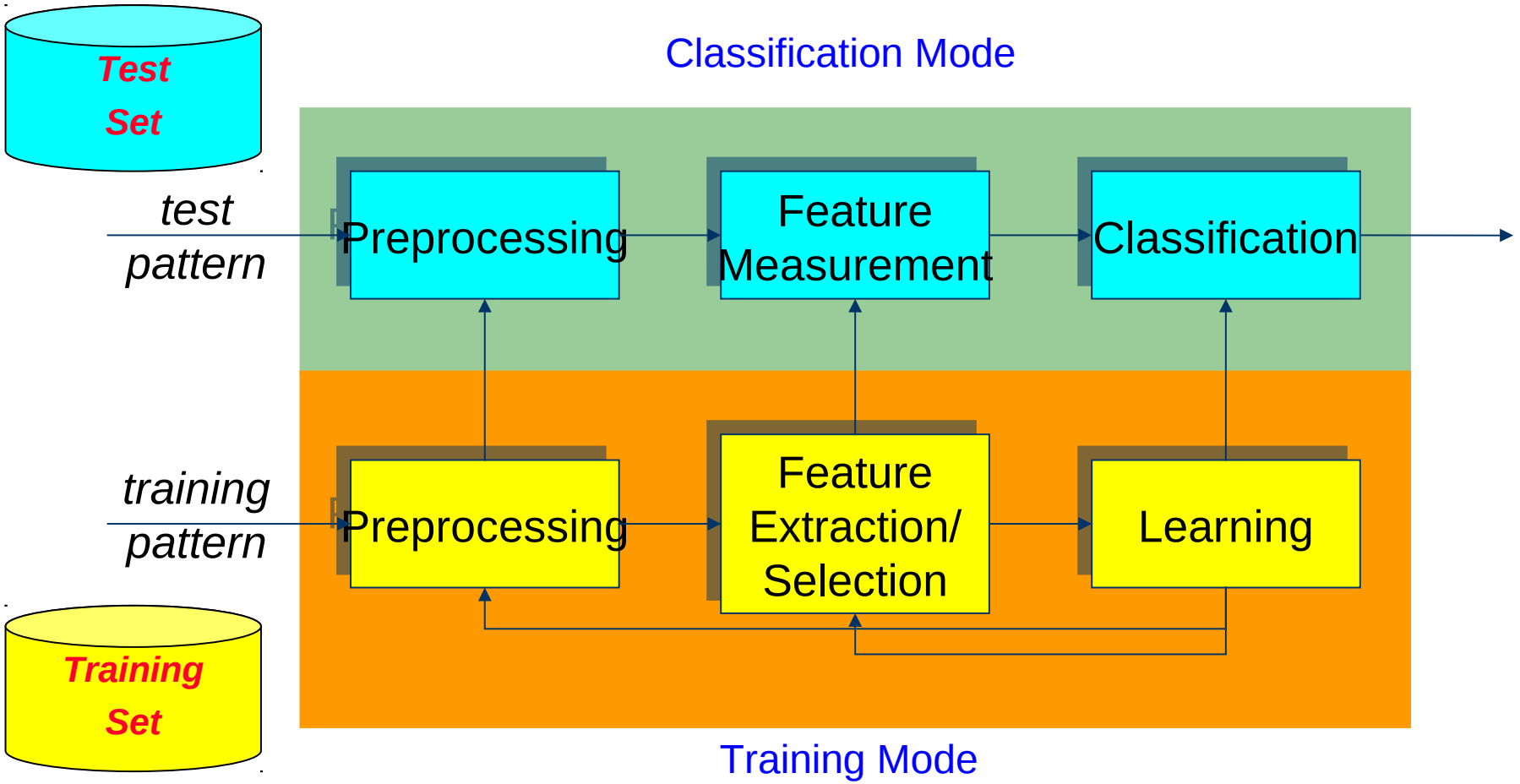
$$y = h(x) = q(x) = a_0 + a_1 x_1 + a_2 x_2$$

Components of PR system



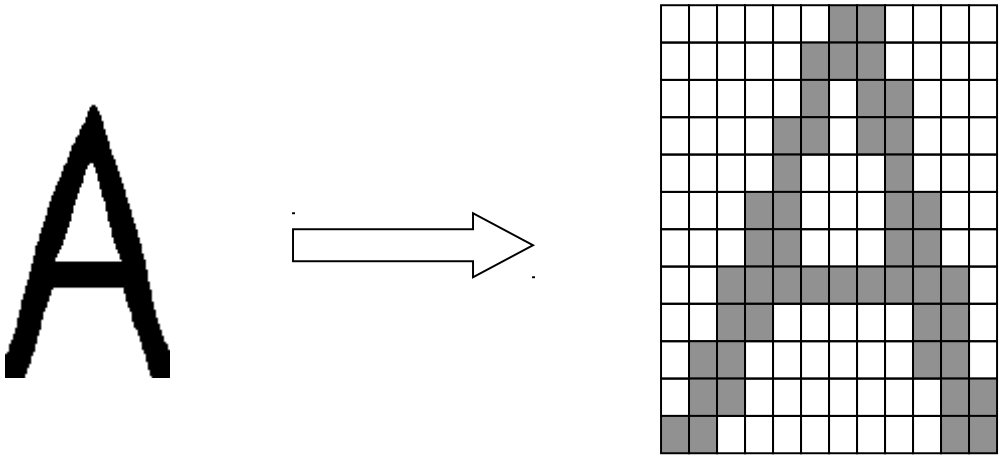
- Sensors and preprocessing.
- A feature extraction aims to create discriminative features good for classification.
- A classifier.
- A teacher provides information about hidden state – supervised learning.
- A learning algorithm sets PR from training examples.

Pattern Recognition System



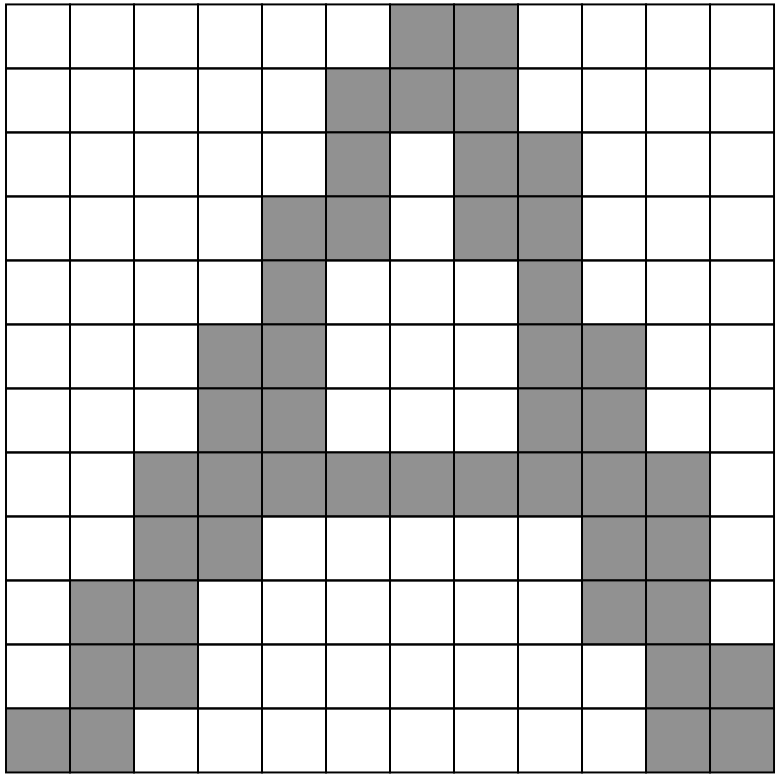
Template Matching

Image is converted into 12x12 bitmap.



Template Matching

Bitmap is represented by 12x12-matrix or by 144-vector with 0 and 1 coordinates.



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

Template Matching

Training samples – templates with corresponding class:

$$t_1 = \{(0,0,0,0,1,1,\ldots,0), 'A'\}$$
$$t_2 = \{(0,0,0,0,0,1,\ldots,0), 'A'\}$$
$$\ldots\ldots\ldots$$
$$t_k = \{(0,0,1,1,1,1,\ldots,0), 'B'\}$$
$$\ldots\ldots\ldots$$

Template of the image to be recognized:

$$T = \{(0,0,0,0,1,1,\ldots,0), 'A'\}$$

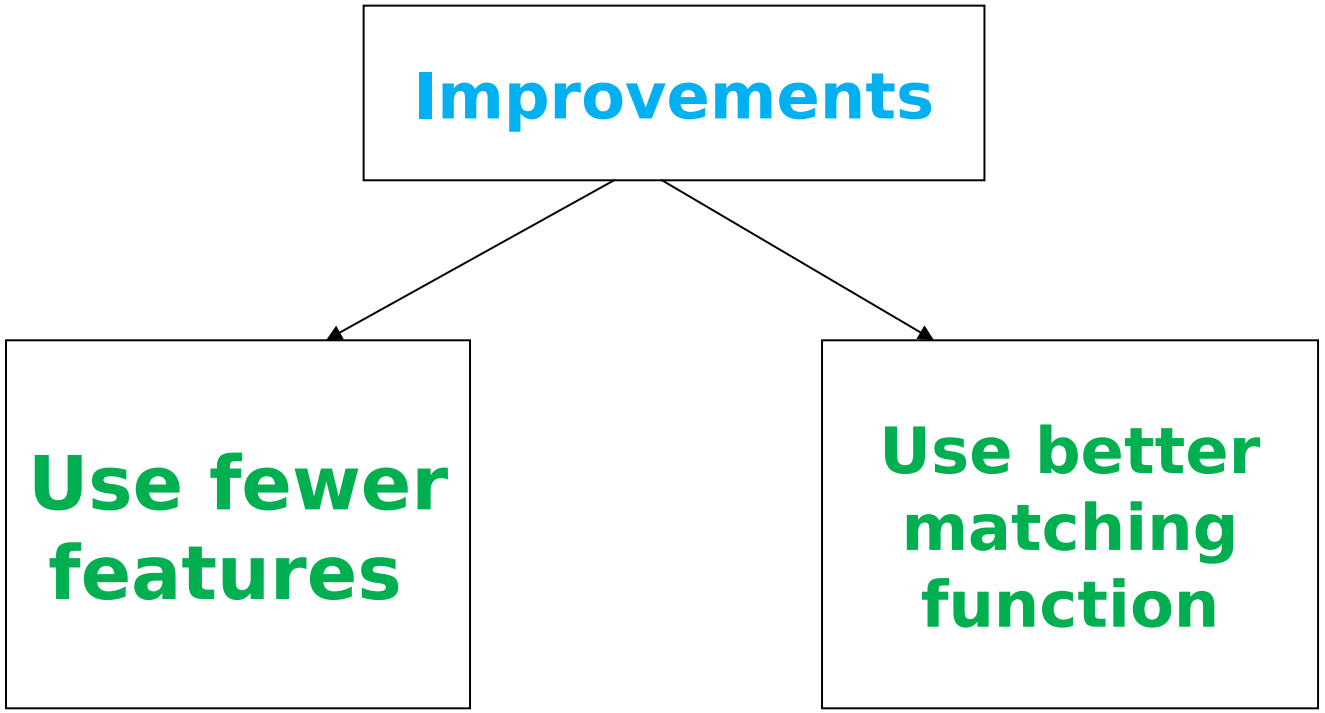
Algorithm:

1. Find t_i, so that $t_i = T$.
2. Assign image to the same class as t_i.

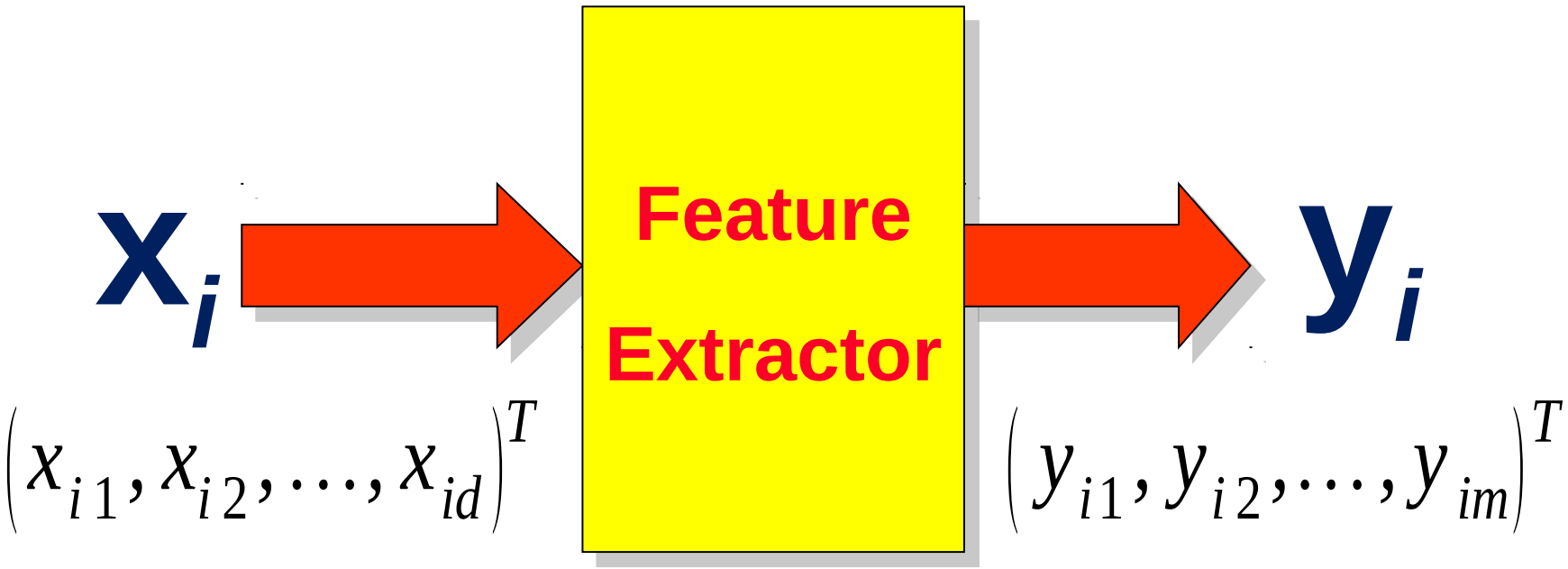
Template Matching

Number of templates to store: 2^{144}

If fewer templates are stored, some images might not be recognized.



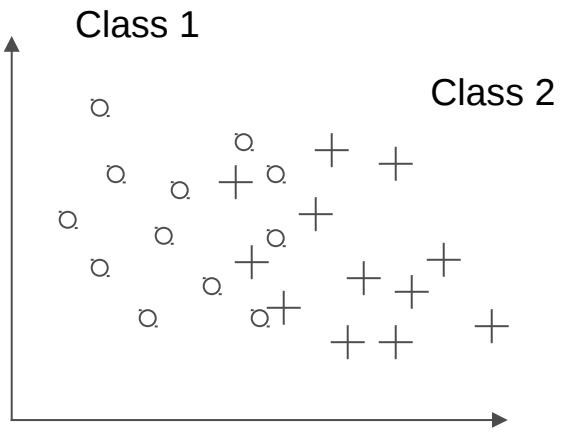
Feature Extractor



$m \leq d$, usually

Feature

- **Features** are numerically expressed properties of the signal.
- The set of features used for pattern recognition is called **feature vector.**
- The number of used features is the **dimensionality** of the feature vector.
- n-dimensional feature vectors can be represented as points in n-dimensional **feature space.**



Some Important Methods

- Principal Component Analysis (PCA)
  - or Karhunen-Loeve Expansion
- Independent Component Analysis (ICA)
- Factor Analysis
- Discriminate Analysis

Linear Approaches

- Kernel PCA
- Multidimensional Scaling (MDS)

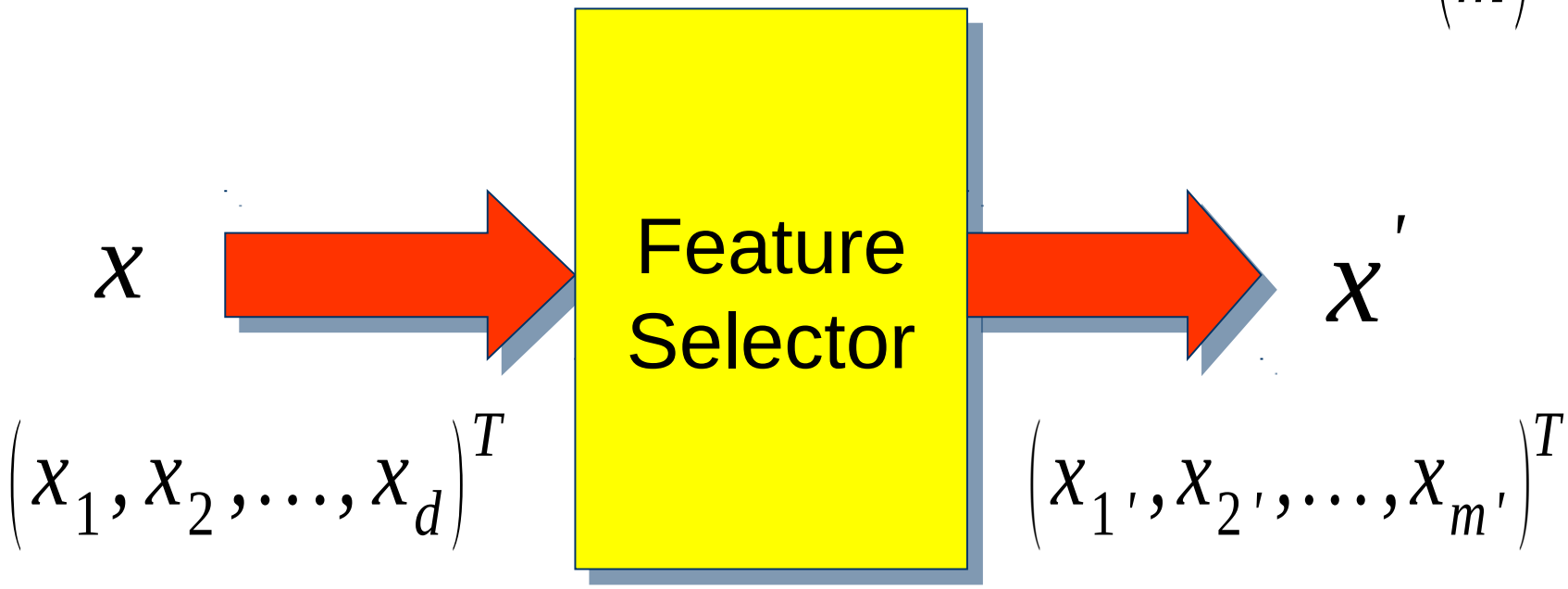
Nonlinear Approaches

- Feed-Forward Neural Networks
- Self-Organizing Map

Neural Networks

Feature Selector •Feature Fusio

possible Selections $\binom{d}{m}$

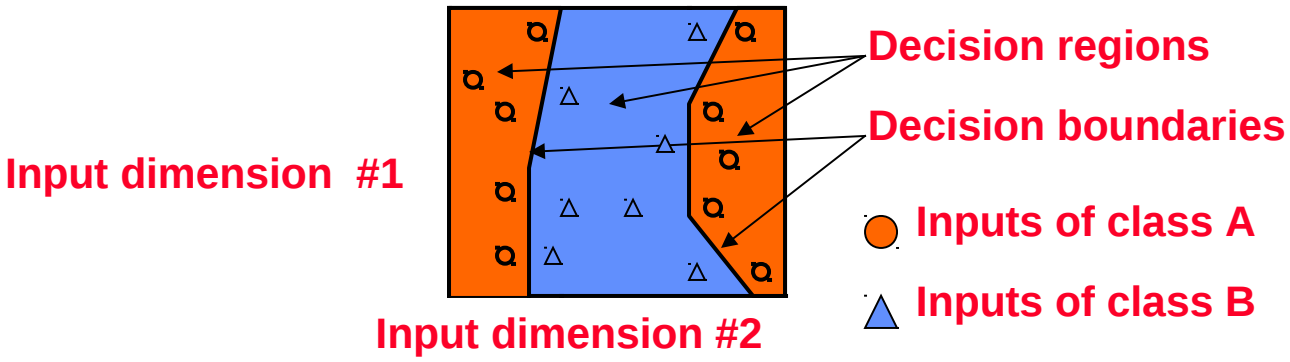


$m \leq d$, usually

Classification: Terminology

A *classifier* can be viewed as a function of block. A classifier assigns one class to each point of the input space. The input space is thus partitioned into disjoint subsets, called *decision regions*, each associated with a class.

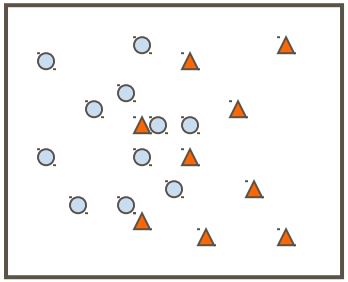
Classification: Terminology (cont.)



- **The way a classifier classifies inputs is defined by its decision regions.**

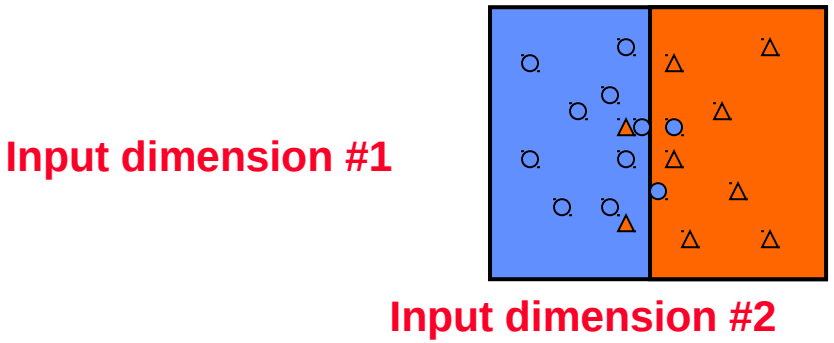
- **The borderlines between decision regions are called *decision-region boundaries* or simply *decision boundaries*.**

Classification: Terminology (cont.)



In practice, input vectors of different classes are rarely so neatly distinguishable. Samples of different classes may have same input vectors. Due to such a *uncertainty*, areas of input space can be clouded by a mixture of samples of different classes.

Classification: Terminology (cont.)



- **The *optimal classifier* is the one expected to produce the least number of misclassifications(Error Reduction).**

- **Such misclassifications are due to uncertainty in the problem rather than a deficiency in the decision regions.**

Classification: Terminology (cont.)

- **A designed classifier is said to *generalize well* if the classifier achieves similar classification accuracy to both training samples and real world samples**

Training Error Reduction # Generalization