



# Machine Learning for Data Mining

Data Mining and Text Mining (UIC 583 @ Politecnico di Milano)

- What is Machine Learning?
- What are the paradigms?
- Unsupervised Learning
- Supervised Learning
- Reinforcement Learning

What is  
Machine Learning?

- ❑ “The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience.” Tom Mitchell (1997)
- ❑ A program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .
- ❑ A well-defined learning task is defined by  $P$ ,  $T$ , and  $E$ .

- ❑ Task T: playing checkers
  
- ❑ Artificial Intelligence
  - ▶ Design and implement a computer-based system that exhibit intelligent action
  
- ❑ Machine Learning
  - ▶ Write a program that can learn how to play
  - ▶ It can learn from examples of previous games, by playing against another opponent, by playing against itself

- A handwriting recognition learning problem
  - ▶ Task T: recognizing and classifying handwritten words within images
  - ▶ Performance P: percent of words correctly classified
  - ▶ Training experience E: a database of handwritten words with given classification
  
- A robot driving learning problem
  - ▶ Task T: driving on public four-lane highways using vision
  - ▶ Performance P: average distance traveled before an error
  - ▶ Training experience E: a sequence of images and steering commands recorded while observing a human driver

# Unsupervised Learning



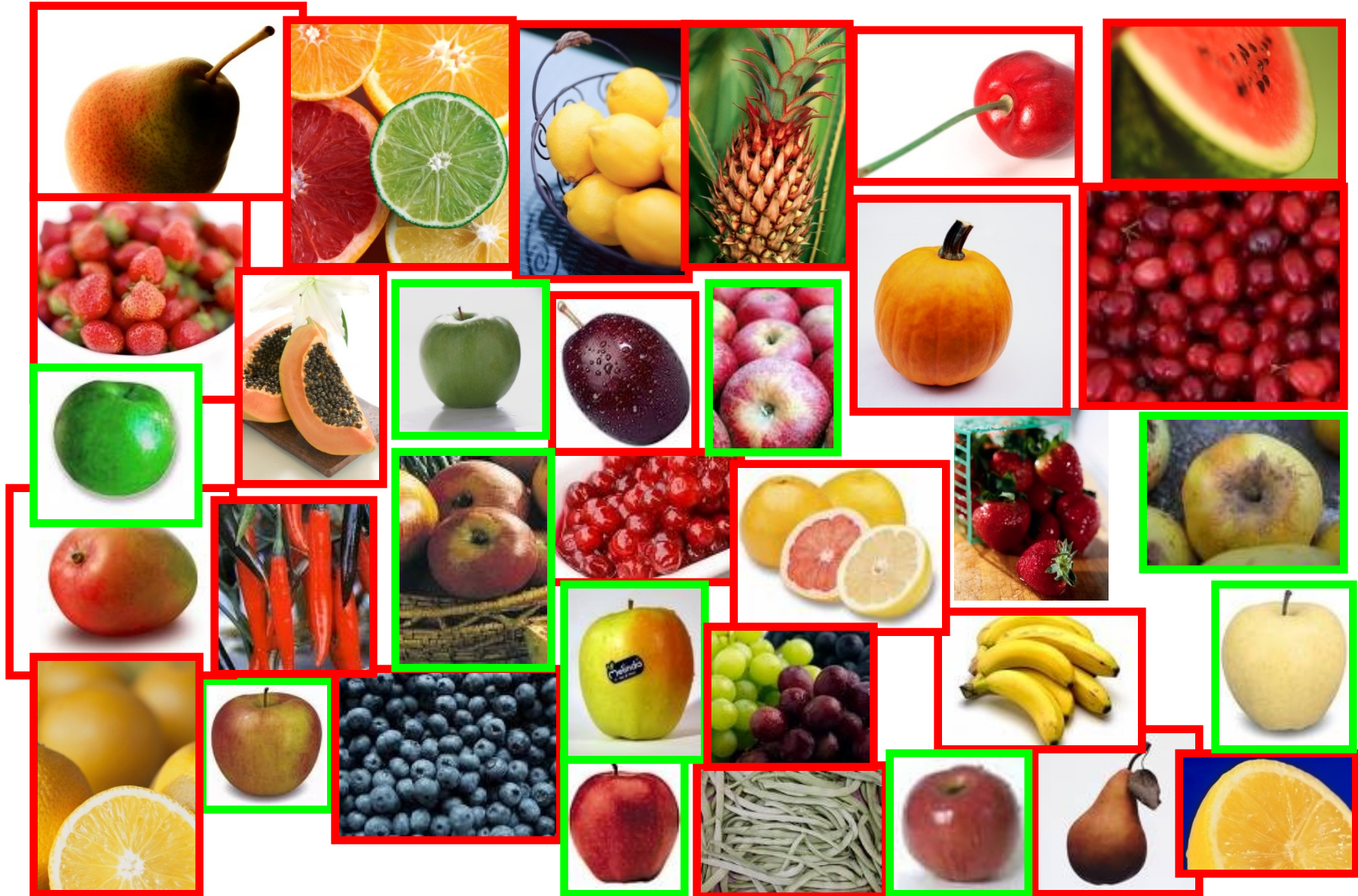
- ❑ Task T: finding interesting groups into data, learning “what normally happens”
- ❑ Performance P: how good, how interesting the groups are
- ❑ Training experience E: raw data
- ❑ Example applications
  - ▶ Customer segmentation in CRM
  - ▶ Color quantization for image compression,
  - ▶ Bioinformatics

# Supervised Learning

# What is an apple?



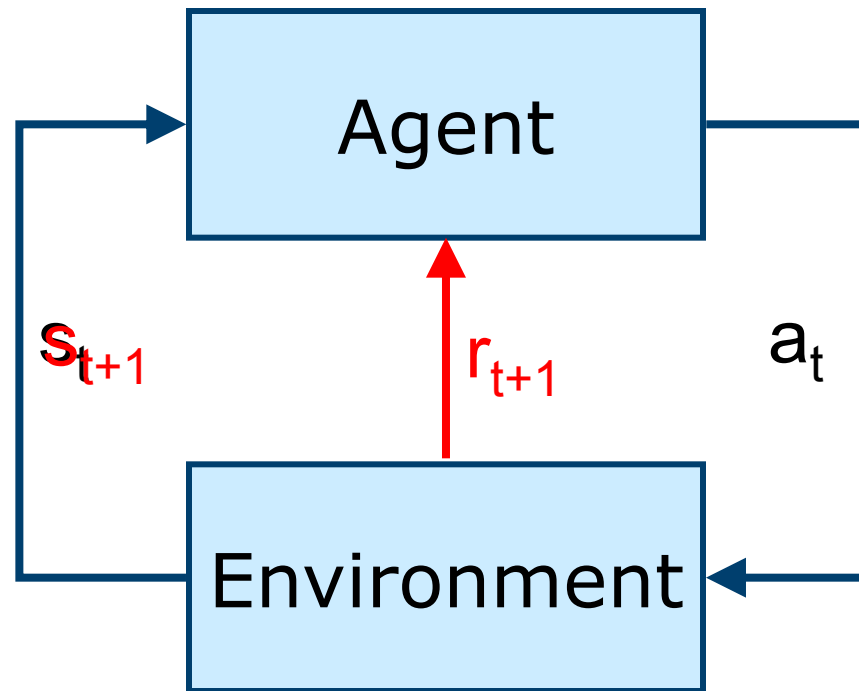
# What is an apple?





- ❑ Training experience  $E$ : examples labeled by a supervisor
- ❑ Task  $T$ : to extract a description of a concept from the data.  
Use the description to predict the output for future examples
- ❑ Performance  $P$ : how accurate the description is
- ❑ Example applications
  - ▶ Credit approval
  - ▶ Target marketing
  - ▶ Medical diagnosis
  - ▶ Fraud detection

# Reinforcement Learning



- ❑ The agent learn through trial-and-error interactions
- ❑ The goal is to maximize the amount of reward received from the environment
- ❑ Compute a value function  $Q(s_t, a_t)$  mapping state-action pairs into expected future payoffs

- ❑ Training experience  $E$ : online interactions with the environment
- ❑ Task  $T$ : collect as much reward as possible
- ❑ Performance  $P$ : the amount of reward
- ❑ Example applications
  - ▶ Robot learning
  - ▶ Games
  - ▶ Multiagent learning

# Data Mining & Machine Learning

## □ Applications

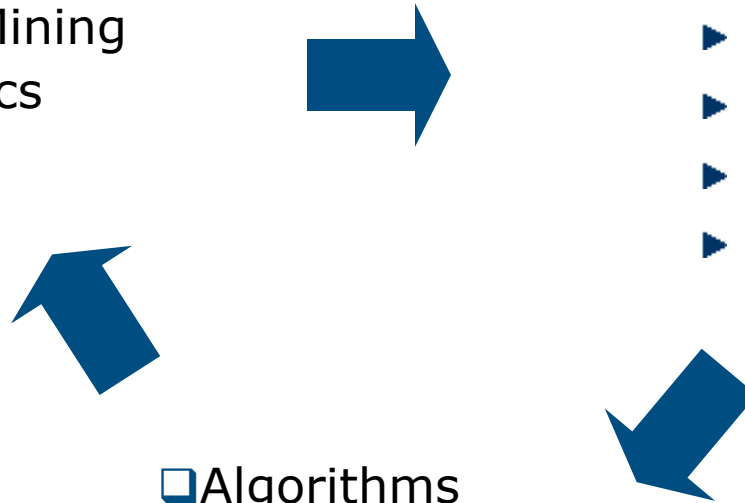
- ▶ Agents
- ▶ Data Mining
- ▶ Robotics
- ▶ ...

## □ Paradigms

- ▶ Unsupervised Learning
- ▶ Supervised Learning
- ▶ Reinforcement Learning
- ▶ ...

## □ Algorithms

- ▶ Clustering
- ▶ Association Rules
- ▶ Decision trees
- ▶ ...



- ❑ Machine learning algorithms acquire structural descriptions from examples
  
- ❑ Structural descriptions represent patterns explicitly
  - ▶ They can be used to predict outcomes in new situations
  - ▶ They can be used to understand and explain how predictions are derived
  
- ❑ Unsupervised learning
  - ▶ Clustering
  - ▶ Association rules
  
- ❑ Supervised learning
  - ▶ Decision trees
  - ▶ Decision rules
  - ▶ Bayesian classifiers