# EC3357:Machine Learning

## Lecture 1: Introduction to Machine Learning

# What is Learning?

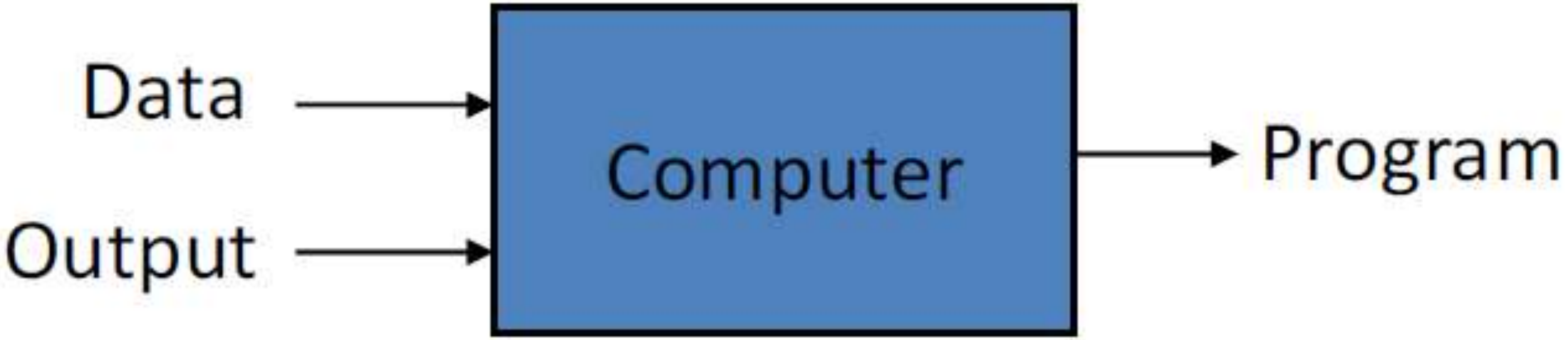
- Herbert Simon: “Learning is any process by which a system improves performance from experience.”

- “A computer 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.” – Tom Mitchell

- From the definition by Tom Mitchell (1998):
  - Machine Learning is the study of algorithms that
    - improve their performance P
    - at some task T
    - with experience E.
  - A well-defined learning task is given by <P, T, E>.

## Traditional Programming


## Machine Learning



# How can we solve a specific problem?

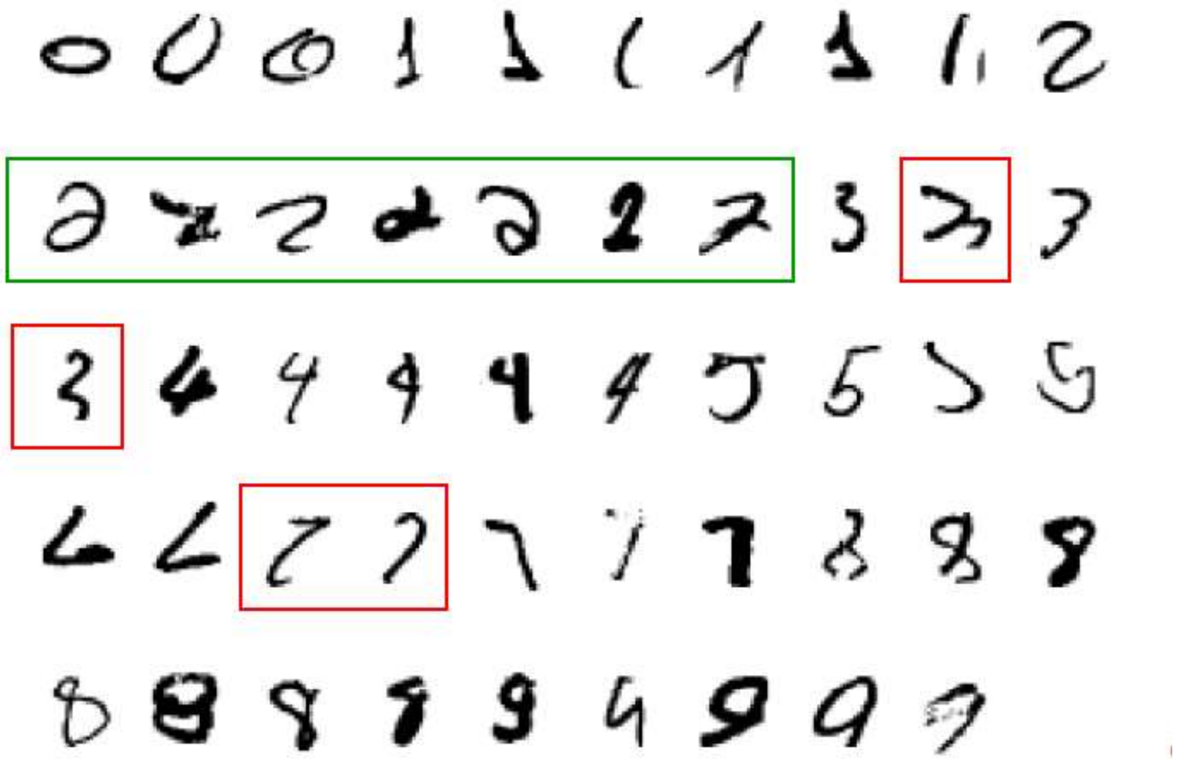
- As computer scientists we write a program that encodes a set of rules that are useful to solve the problem
- In many cases is very difficult to specify those rules, e.g., given a picture determine whether there is a cat in the image.

# Learning

- Learning systems are not directly programmed to solve a problem, instead develop own program based on:
  - Examples of how they should behave
  - From trial-and-error experience trying to solve the problem
- Learning simply means incorporating information from the training(input-output pairs) examples into the system

# A classic example of a task that requires machine learning

- It is very hard to say what makes a 2

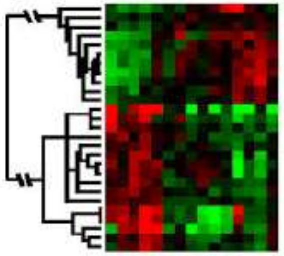


# Why use learning?

- It is very hard to write programs that solve problems like recognizing a handwritten digit
  - What distinguishes a 2 from a 7?
  - How does our brain do it?
- Instead of writing a program by hand, we collect examples that specify the correct output for a given input
- A machine learning algorithm then takes these examples and produces a program that does the job

# When Do We Use Machine Learning

- ML is used when:
  - Human expertise does not exist (navigating on Mars)
  - Humans can’t explain their expertise (speech recognition)
  - Models must be customized (personalized medicine)
  - Models are based on huge amounts of data (genomics)

- Learning isn’t always useful:
  - There is no need to “learn” to calculate payroll

# More examples of tasks that are best solved by using a learning algorithm

- Recognizing patterns:
  - Facial identities or facial expressions
  - Handwritten or spoken words
  - Medical images
- Generating patterns:
  - Generating images or motion sequences
- Recognizing anomalies:
  - Unusual credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
  - Future stock prices or currency exchange rates

# Sample Applications

- Web search
    - Computational biology
    - Finance
    - E-commerce
    - Space exploration
    - Robotics
    - Information extraction
    - Social networks
    - Debugging software
    - [Your favorite area]

# Samuel’s Checkers-Player

- “Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.” - Arthur Samuel (1959)

# Defining the Learning Task

Improve on task T, with respect to performance metric P, based on experience E

- T: Playing checkers
- P: Percentage of games won against an arbitrary opponent
- E: Playing practice games against itself

- T: Recognizing hand-written words
- P: Percentage of words correctly classified
- E: Database of human-labeled images of handwritten words

- T: Driving on four-lane highways using vision sensors
- P: Average distance traveled before a human-judged error
- E: A sequence of images and steering commands recorded while observing a human driver.

- T: Categorize email messages as spam or legitimate.
- P: Percentage of email messages correctly classified.
- E: Database of emails, some with human-given labels

# Autonomous Driving



# Flying Robots

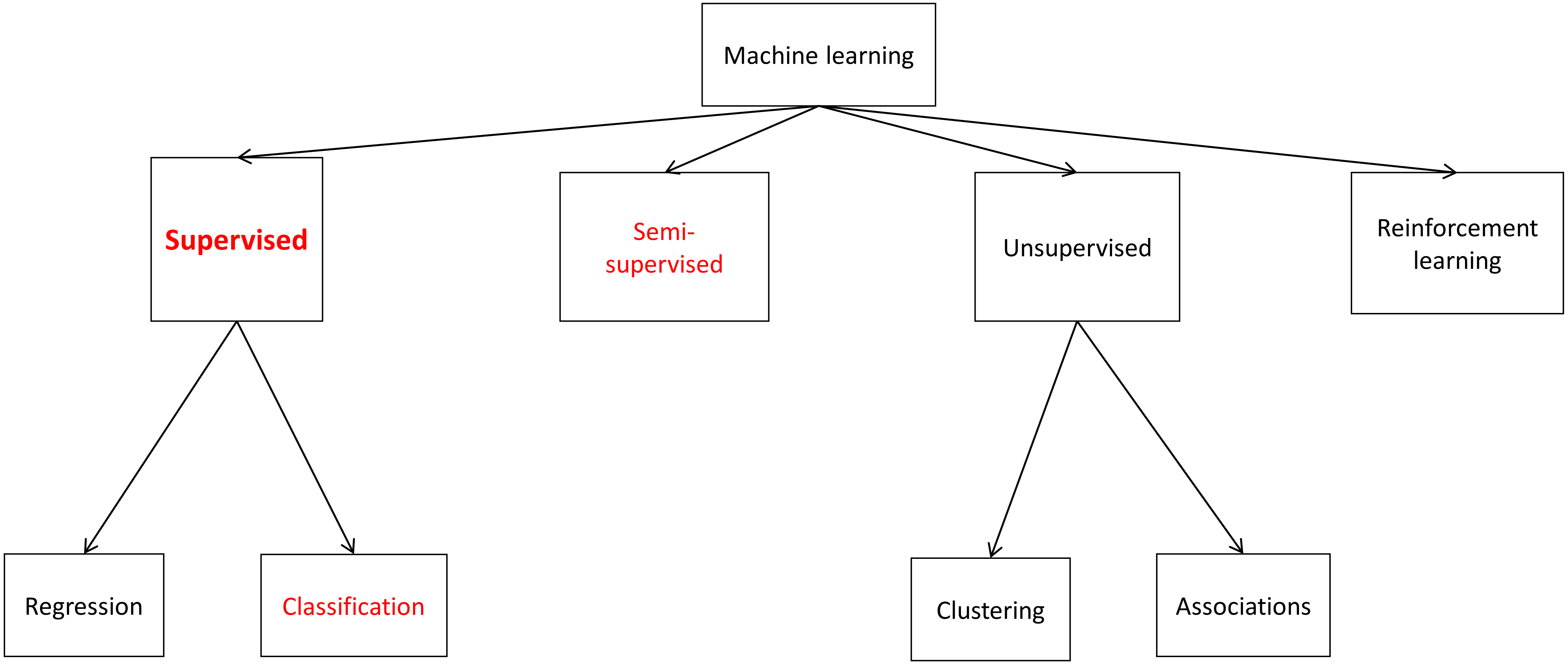
Figure: Video: https://www.youtube.com/watch?v=YQIMGV5vtd4

# Types of Learning

- Supervised (inductive) learning
  – Given: training data + desired outputs (labels)
- Unsupervised learning
  – Given: training data (without desired outputs)
- Semi-supervised learning
  – Given: training data + a few desired outputs
- Reinforcement learning

– Rewards from sequence of actions. An agent interacting with the world makes observations, takes actions, and is rewarded or punished; it should learn to choose actions in such a way as to obtain a lot of reward.

# ML taxonomy

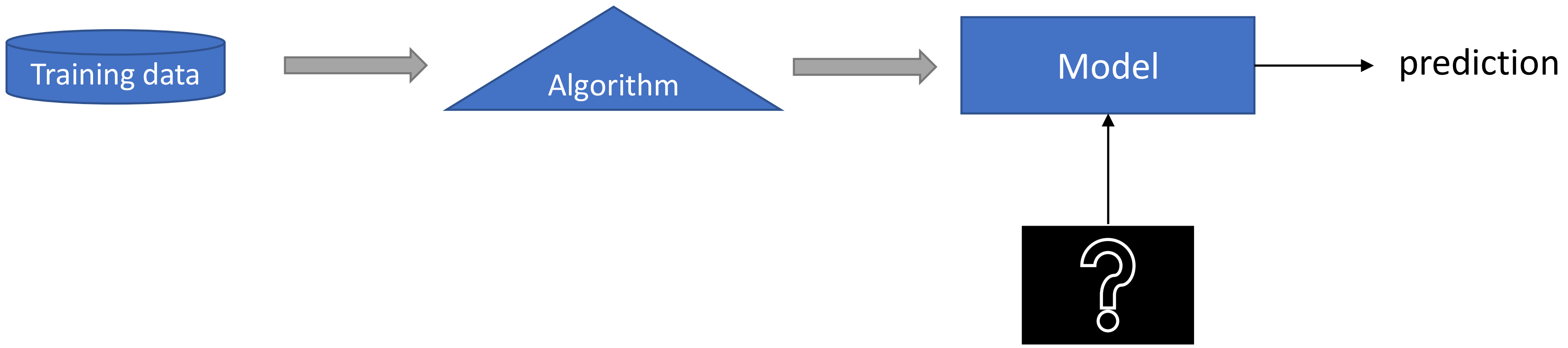

Machine learning
Supervised
Semi-supervised
Unsupervised
Reinforcement learning
Regression
Classification
Clustering
Associations

# Supervised Learning

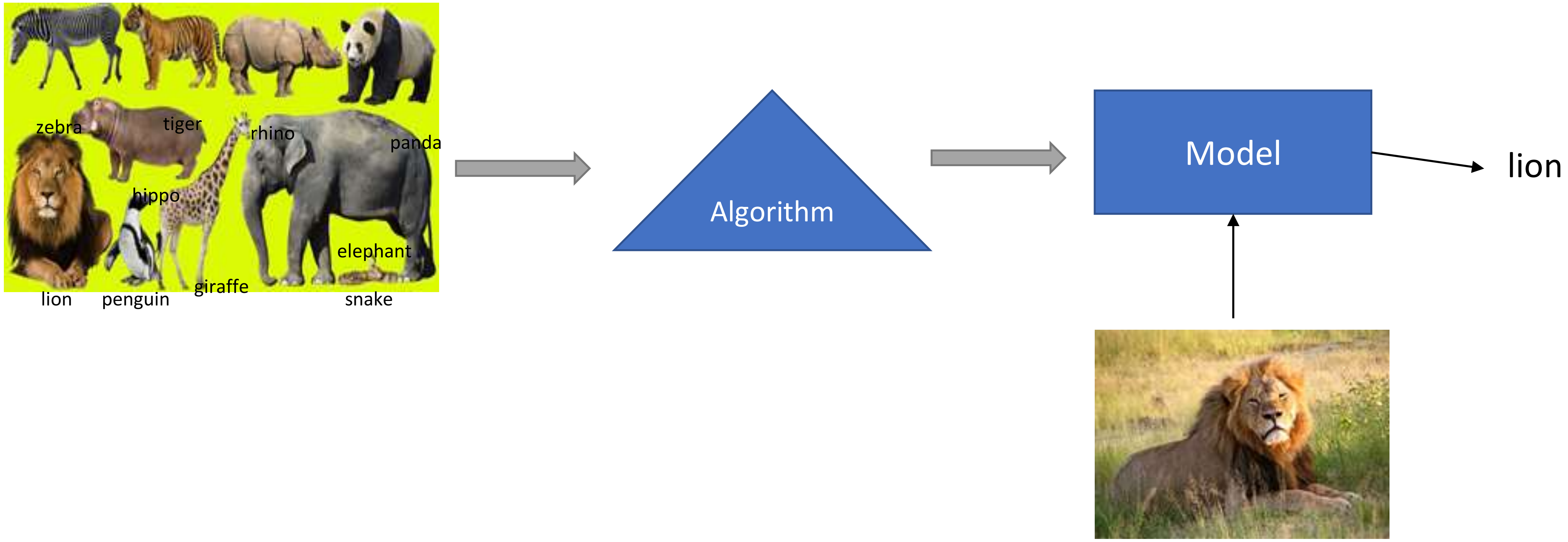
- Learning a discrete function: **Classification**
  - Boolean classification:
    - Each example is classified as true(positive) or false(negative).
    - predict categorical values, i.e., labels
- Learning a continuous function: **Regression**
  - predict numerical values

# Supervised learning

- In supervised learning, the algorithms are presented with a set of classified instances from which they learn a way of classifying unseen instances. When the attribute to be predicted is numeric rather than nominal it is called regression.



# Classification



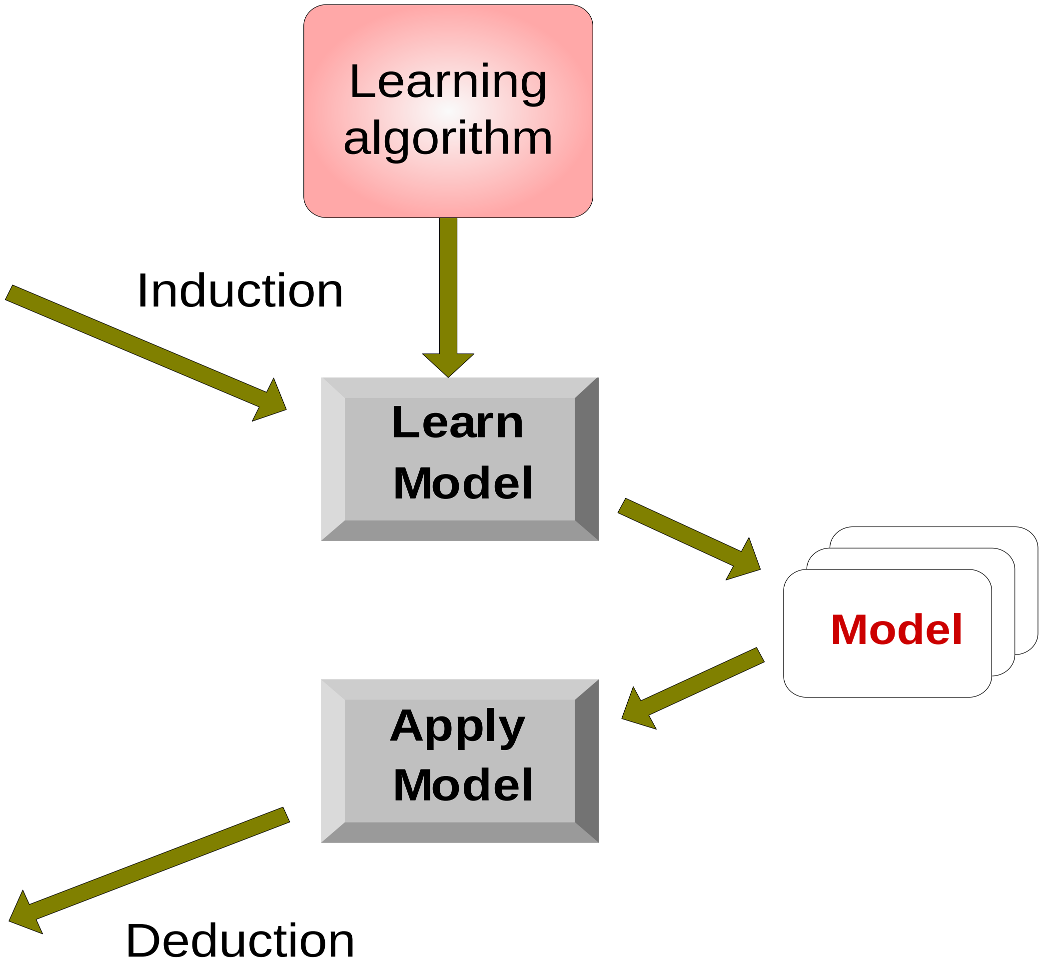
# Classification—A Two-Step Process

- Model construction: describing a set of predetermined classes
  - Each tuple/sample is assumed to belong to a predefined class, as determined by the class label
  - The set of tuples used for model construction is training set
  - The model is represented as **classification rules**, **decision trees**, or **mathematical formulae**
- Model usage: for classifying future or unknown objects
  - Estimate accuracy of the model
    - The known label of test sample is compared with the classified result from the model
    - **Test set is independent of training set**, otherwise over-fitting will occur
  - If the accuracy is acceptable, use the model to classify data tuples whose class labels are not known

# Illustrating Classification Task

| Tid | Attrib1 | Attrib2 | Attrib3 | Class |
|---|---|---|---|---|
| 1 | Yes | Large | 125K | No |
| 2 | No | Medium | 100K | No |
| 3 | No | Small | 70K | No |
| 4 | Yes | Medium | 120K | No |
| 5 | No | Large | 95K | Yes |
| 6 | No | Medium | 60K | No |
| 7 | Yes | Large | 220K | No |
| 8 | No | Small | 85K | Yes |
| 9 | No | Medium | 75K | No |
| 10 | No | Small | 90K | Yes |

Training Set




| Tid | Attrib1 | Attrib2 | Attrib3 | Class |
|---|---|---|---|---|
| 11 | No | Small | 55K | ? |
| 12 | Yes | Medium | 80K | ? |
| 13 | Yes | Large | 110K | ? |
| 14 | No | Small | 95K | ? |
| 15 | No | Large | 67K | ? |

Test Set

# Issues: Data Preparation

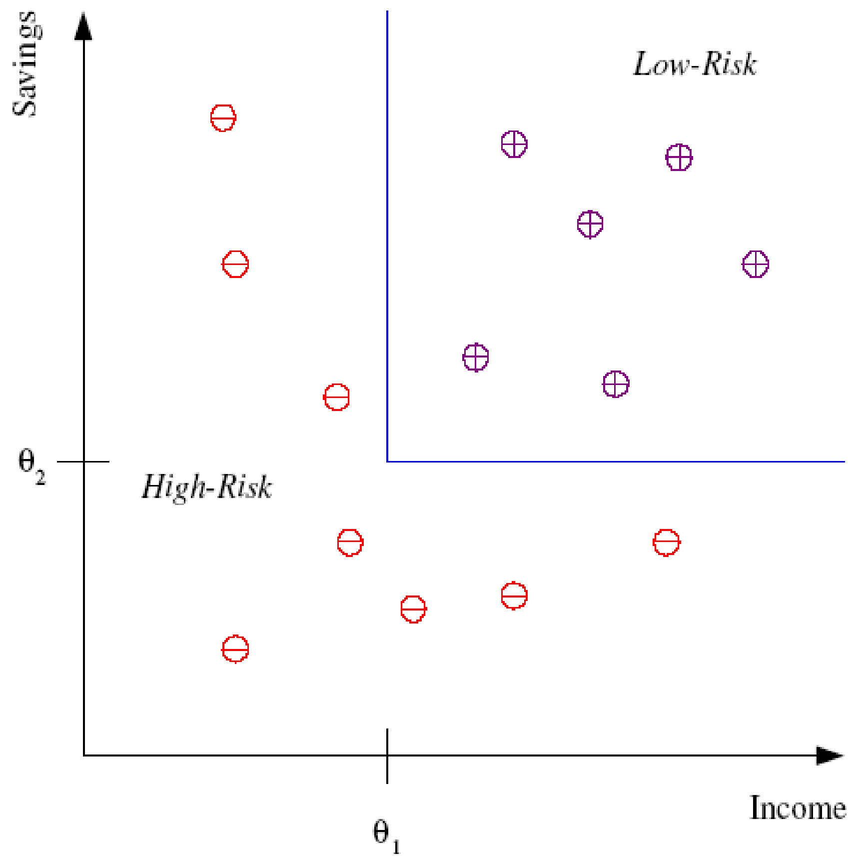
- Data cleaning
  - Preprocess data in order to reduce noise and handle missing values
- Relevance analysis (feature selection)
  - Remove the irrelevant or redundant attributes
- Data transformation
  - Generalize data to (higher concepts, discretization)
  - Normalize attribute values

# Classification Techniques

- Decision Tree based Methods
- Rule-based Methods
- Naïve Bayes and Bayesian Belief Networks
- Neural Networks
- Support Vector Machines
- and more...

# Classification

- Example: Credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*




Discriminant: IF *income* > $\theta_1$ AND *savings* > $\theta_2$
THEN low-risk ELSE high-risk

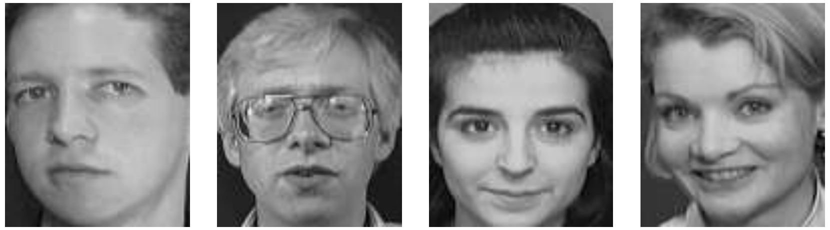
# Classification: Applications

- Aka Pattern recognition
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Medical diagnosis: From symptoms to illnesses
- ...

# Face Recognition

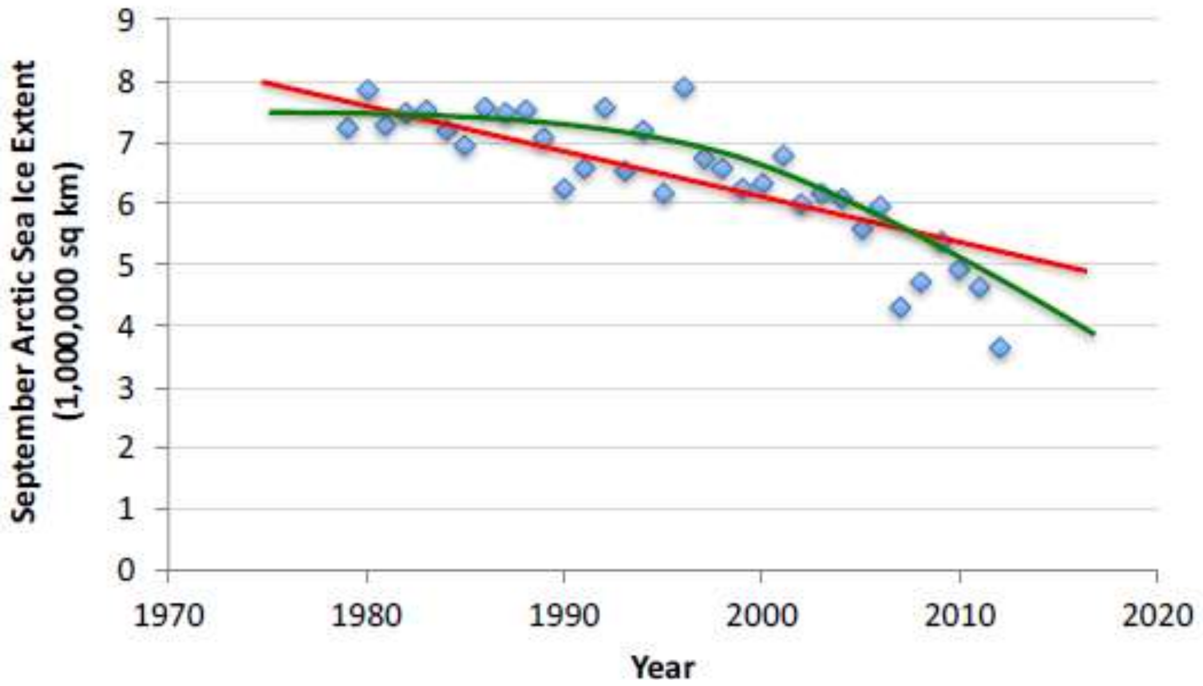
Training examples of a person

Test images



# Supervised Learning: Regression

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict $y$ given $x$
  - $y$ is real-valued == regression



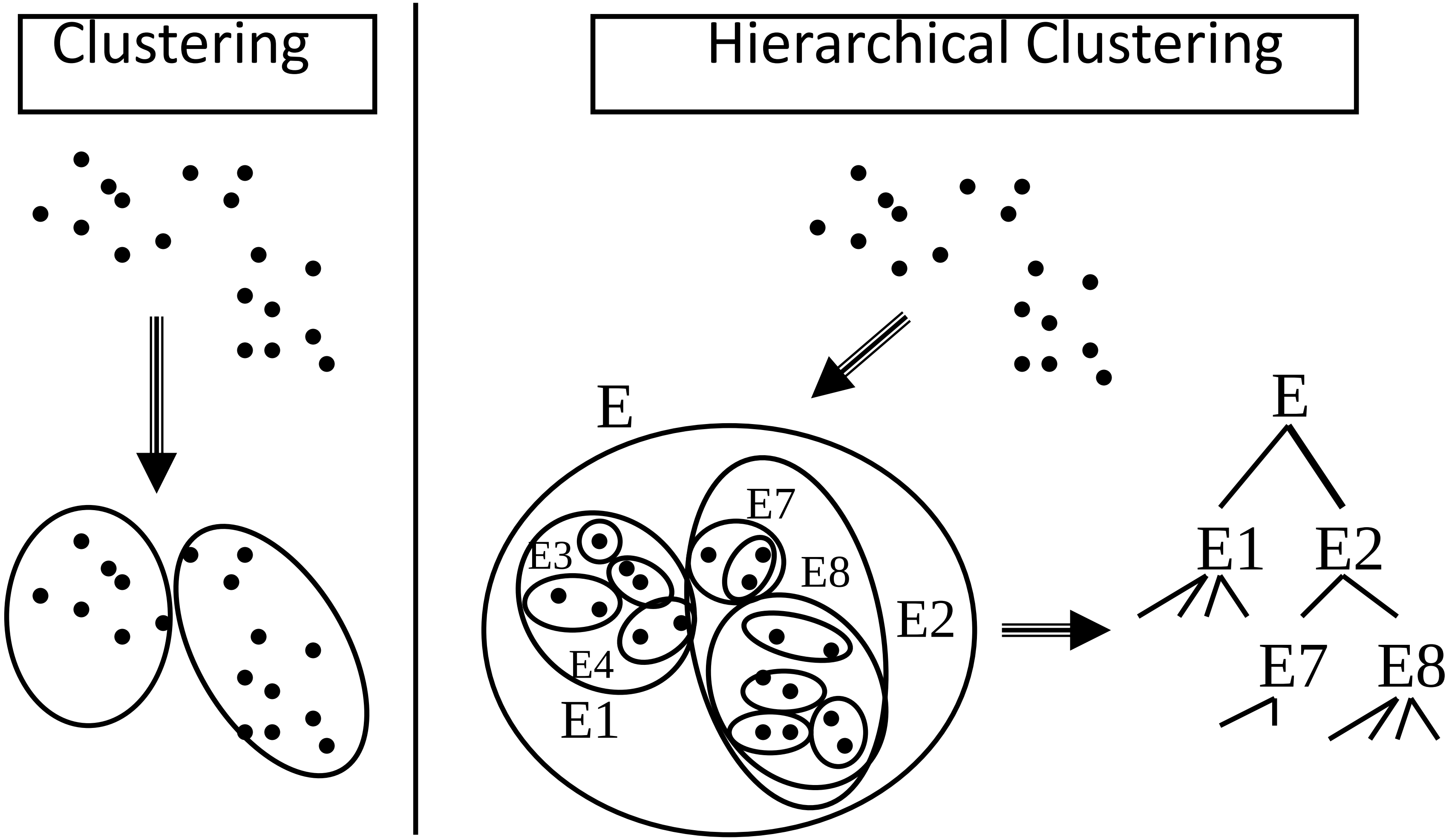
# Supervised Learning: Uses

- Prediction of future cases: Use the rule to predict the output for future inputs
- Knowledge extraction: The rule is easy to understand
- Compression: The rule is simpler than the data it explains
- Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

# Unsupervised Learning: Overview

- So far, in all the learning techniques we considered, a training example consisted of a set of attributes (or features) and either a class (in the case of classification) or a real number (in the case of regression) attached to it.
- Unsupervised Learning takes as training examples the set of attributes/features alone.
- The purpose of unsupervised learning is to attempt to find natural partitions in the training set.
- Two general strategies for Unsupervised learning include: ***Clustering*** and ***Hierarchical Clustering***.

# Clustering and Hierarchical Clustering

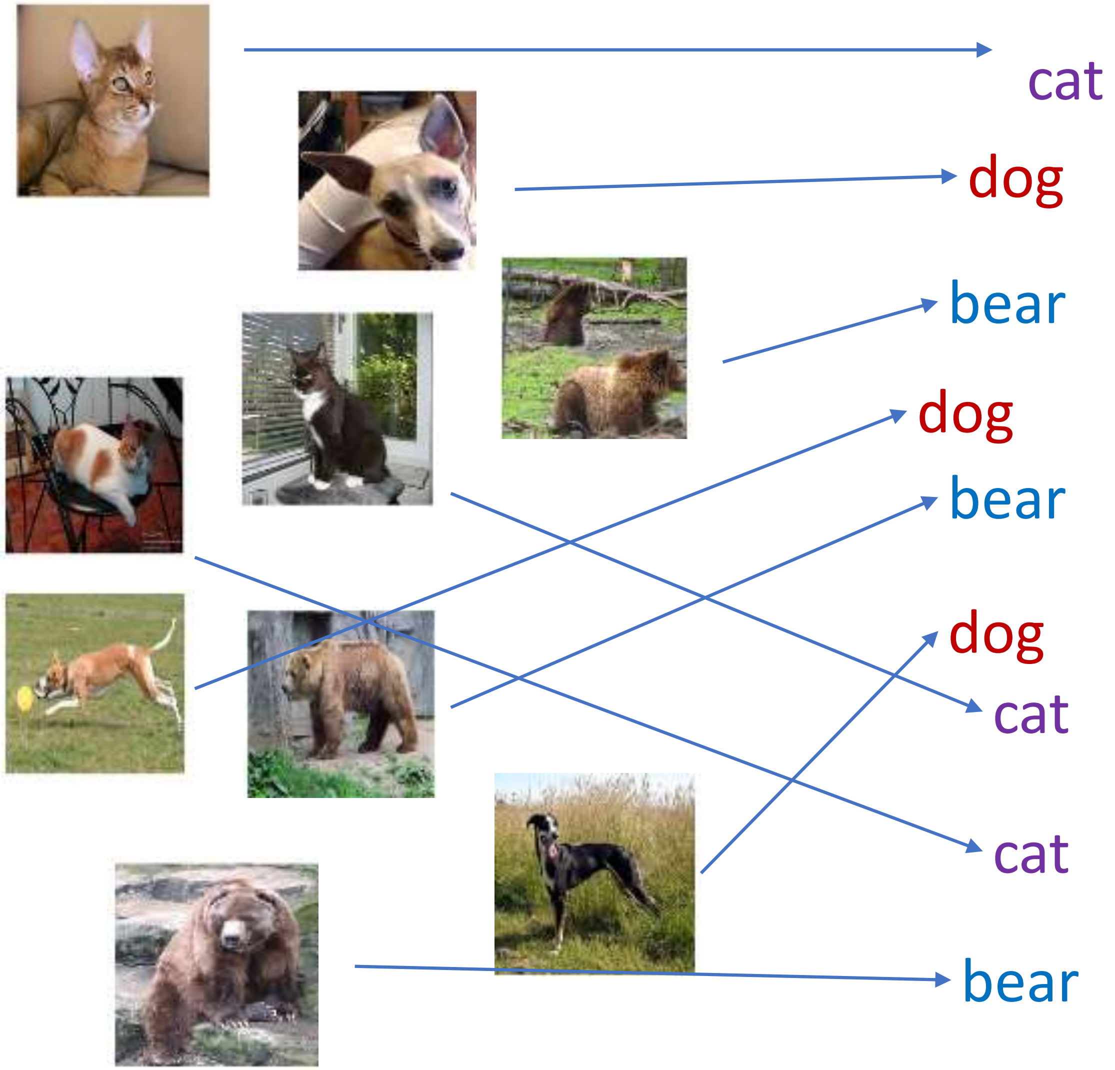


# Other Unsupervised Methods:

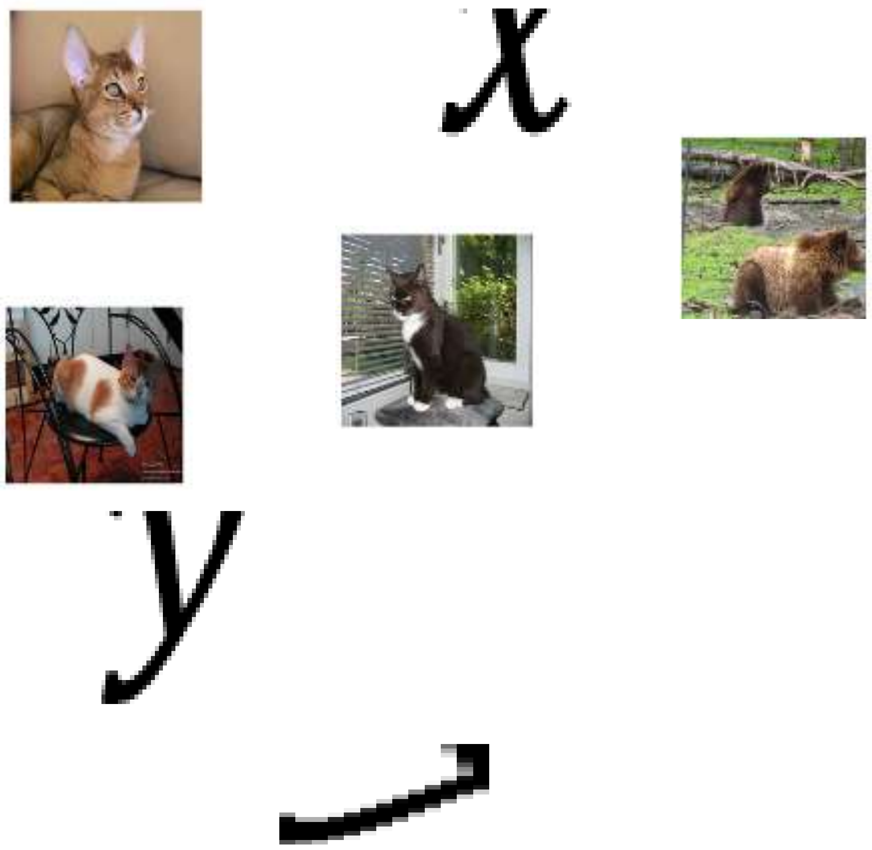
- There are a lot of other Unsupervised Learning Methods.
- **Examples:**
  - k-means
  - The EM Algorithm
  - Competitive Learning
  - Kohonen’s Neural Networks: Self-Organizing Maps
  - Principal Component Analysis, Autoassociation

# Supervised Learning vs Unsupervised Learning

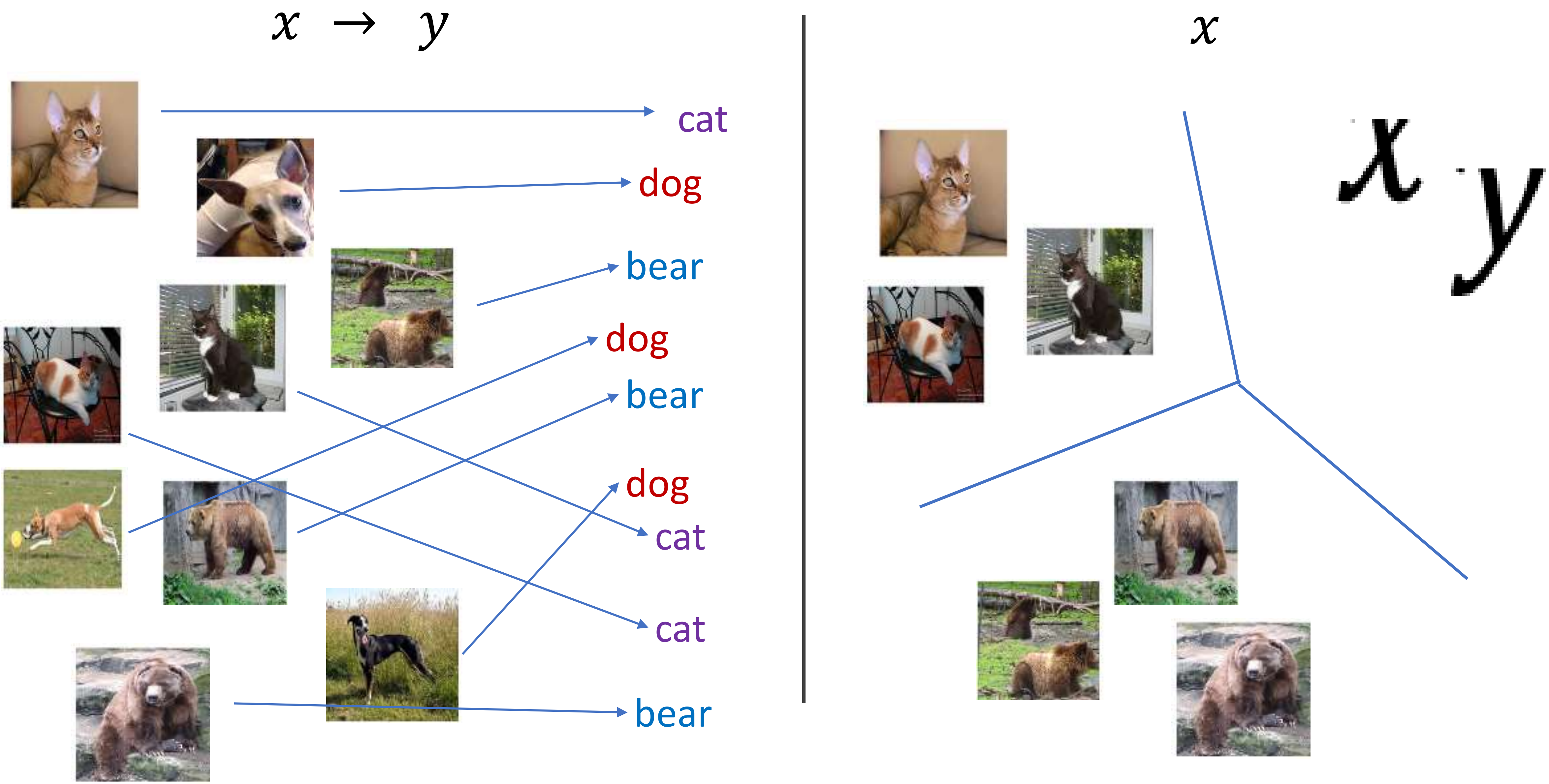
$x \rightarrow y$



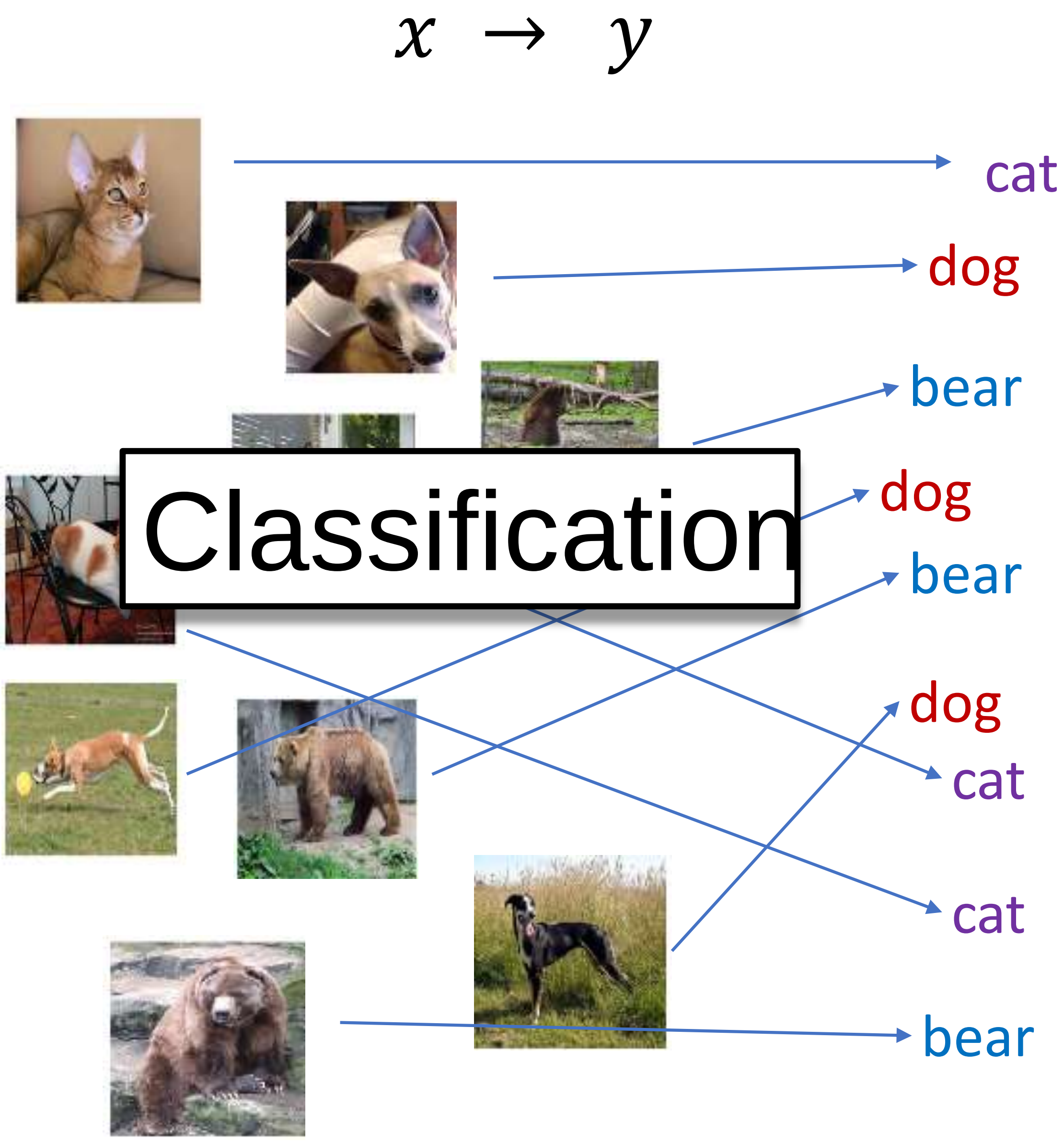

$x$

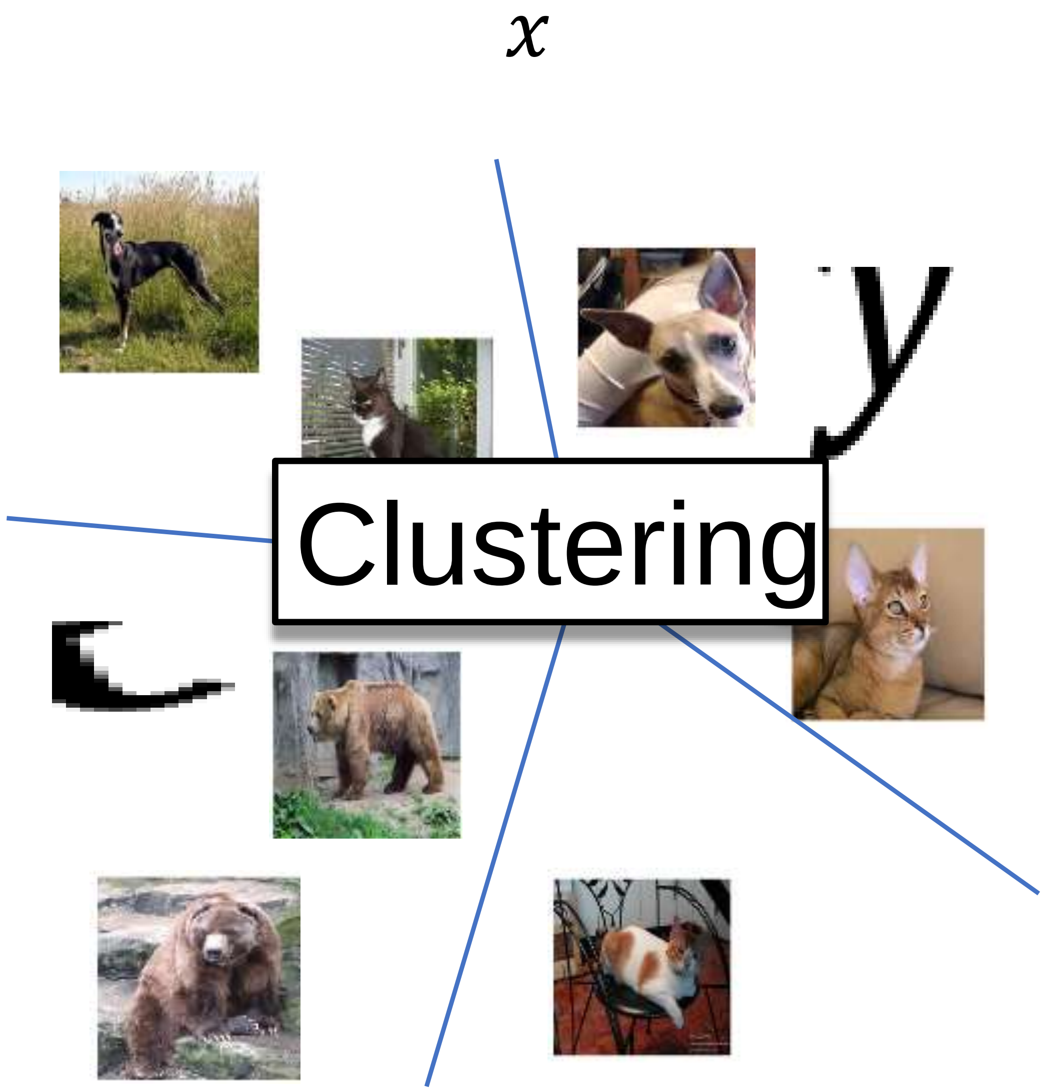


# Supervised Learning vs Unsupervised Learning



# Supervised Learning vs Unsupervised Learning



# Reinforcement Learning

- Given a sequence of states and actions with (delayed) rewards, output a policy
  - Policy is a mapping from statesàactions that tells you what to do in a given state
- Examples:
  - Credit assignment problem
  - Game playing
  - Robot in a maze
  - Balance a pole on your hand

# Reinforcement Learning

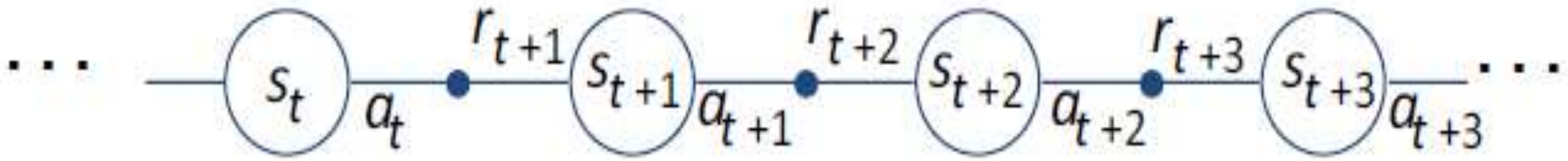
Agent and environment interact at discrete time steps : $t = 0, 1, 2, K$

Agent observes state at step $t$ : $s_t \in S$

produces action at step $t$ : $a_t \in A(s_t)$

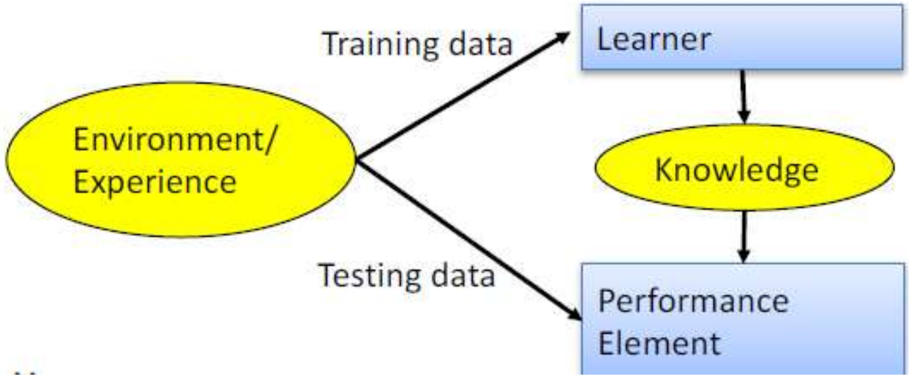
gets resulting reward : $r_{t+1} \in \Re$

and resulting next state : $s_{t+1}$



# Designing a Learning System

- Choose the training experience
- Choose exactly what is to be learned
  - i.e. the ***target function***
- Choose how to represent the target function
- Choose a learning algorithm to infer the target function from the experience



# Case Study

- What grade will I get in this course?
- Data: entry survey and marks from this and previous years
- Process the data
  - Split into training set; and test set
  - Determine representation of input;
  - Determine the representation of the output;
- Choose form of model: linear regression
- Decide how to evaluate the system's performance: objective function
- Set model parameters to optimize performance
- Evaluate on test set: generalization

# ML in a Nutshell

- Tens of thousands of machine learning algorithms
  - Hundreds new every year
- Every ML algorithm has three components:
  - **Representation**
  - **Optimization**
  - **Evaluation**

# Various Function Representations

- • Numerical functions
    - Linear regression
    - Neural networks
    - Support vector machines
- Symbolic functions
    - Decision trees
    - Rules in propositional logic
    - Rules in first-order predicate logic
- Instance-based functions
    - Nearest-neighbor
    - Case-based
- Probabilistic Graphical Models
    - Naïve Bayes
    - Bayesian networks
    - Hidden-Markov Models (HMMs)
    - Probabilistic Context Free Grammars (PCFGs)
    - Markov networks

# Various Search/Optimization Algorithms

- Gradient descent
  - Perceptron
  - Backpropagation
- Dynamic Programming
  - HMM Learning
  - PCFG Learning
- Divide and Conquer
  - Decision tree induction
  - Rule learning
- Evolutionary Computation
  - Genetic Algorithms (GAs)
  - Genetic Programming (GP)
  - Neuro-evolution

# Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- etc.

# ML in Practice

- Understand domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learn models
- Interpret results
- Consolidate and deploy discovered knowledge

# Lessons Learned about Learning

- Learning can be viewed as using direct or indirect experience to approximate a chosen target function.
- Function approximation can be viewed as a search through a space of hypotheses (representations of functions) for one that best fits a set of training data.
- Different learning methods assume different hypothesis spaces (representation languages) and/or employ different search techniques