Overview: Machine Learning

Slides adapted from lectures by
Nando de Freitas, University of British Columbia

Machine Learning:

What is Machine Learning?

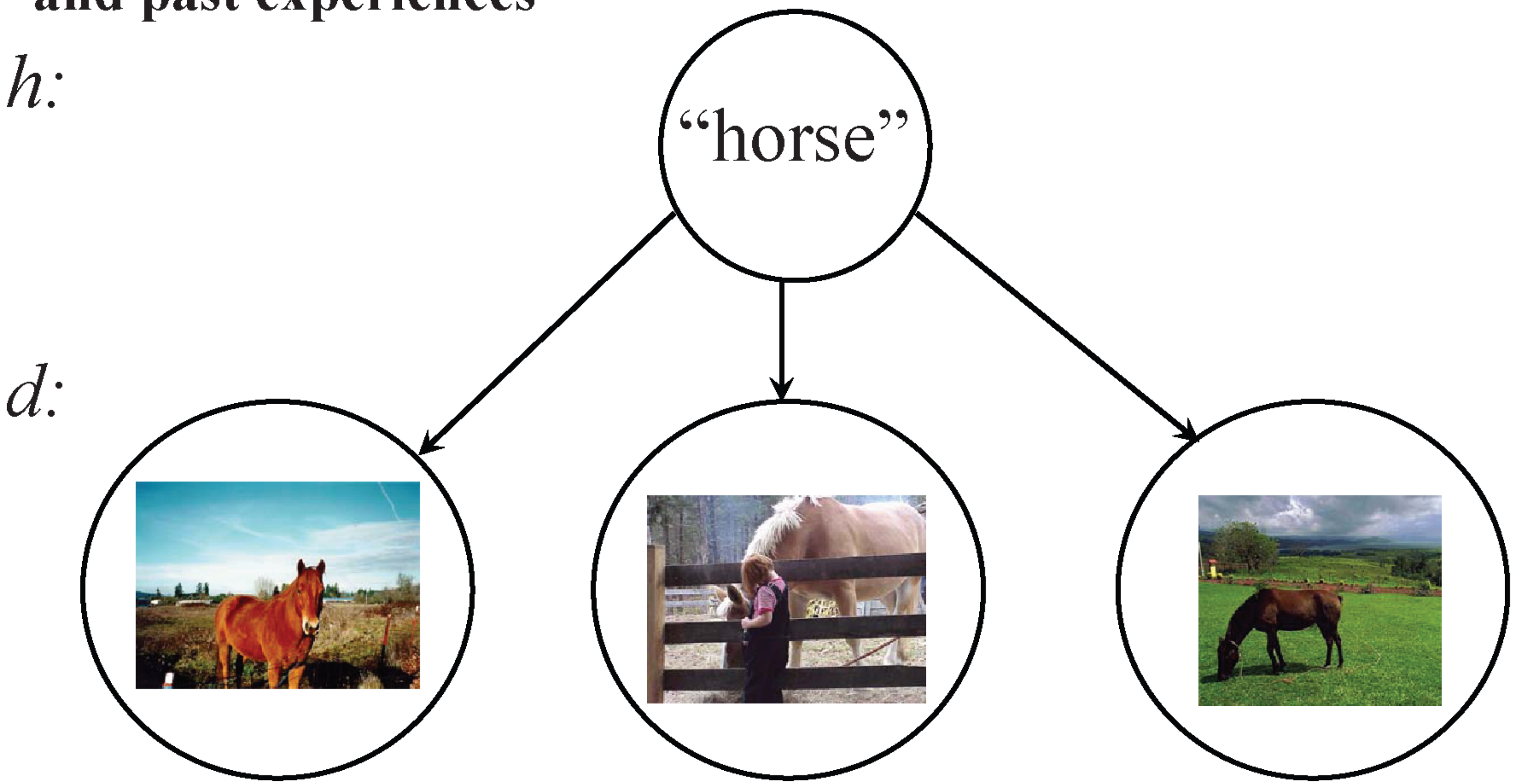
“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” --Tom M. Mitchell

- Closely related to
  - Statistics (fitting models to data and testing them)
  - Data mining / exploratory data analysis (discovering models)
  - Adaptive control theory
  - And of course AI

Machine Learning:

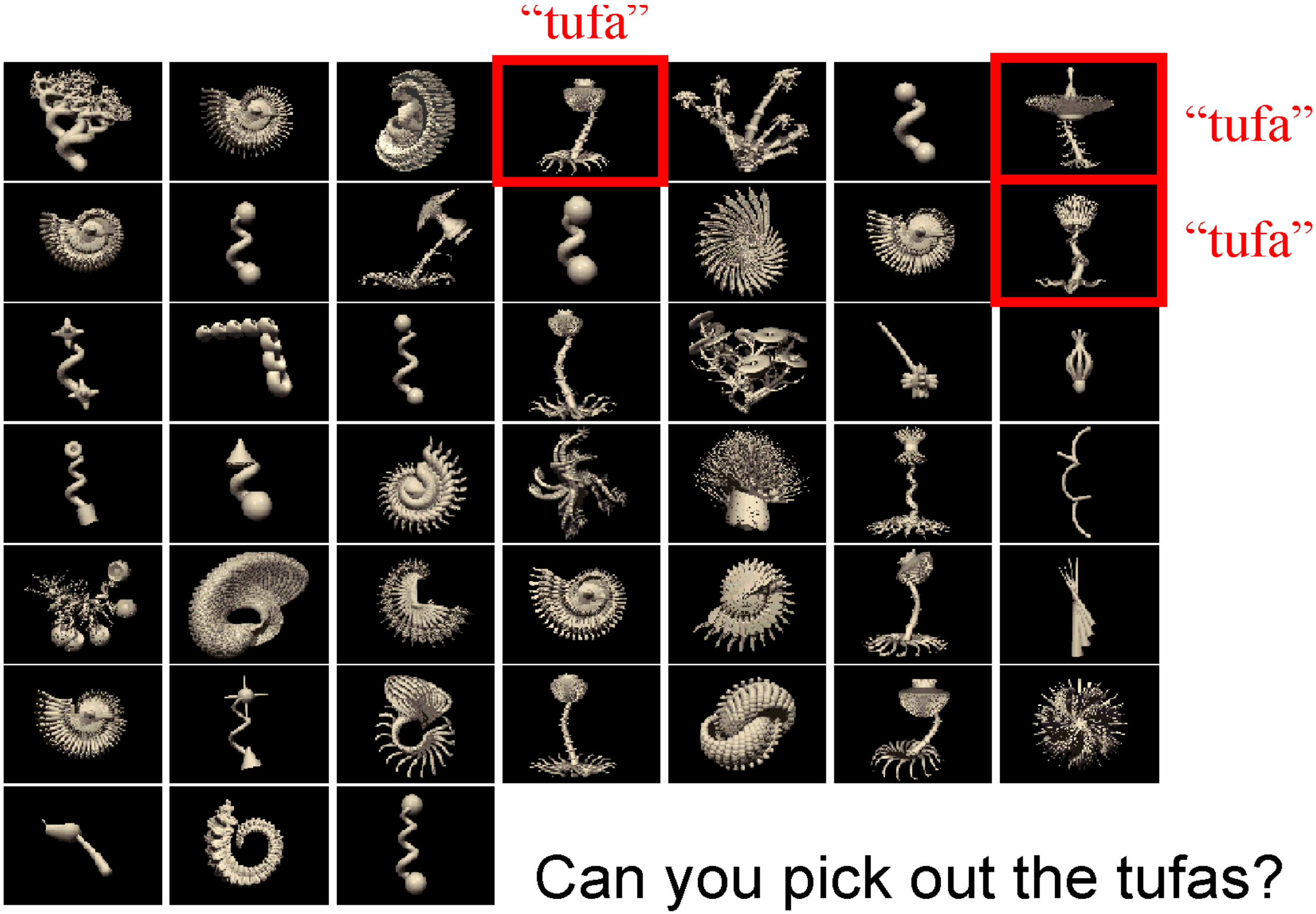
Abstractions from Observation

Learning is the process of automatically constructing abstractions of the real world from a set of observations and past experiences



Machine Learning:

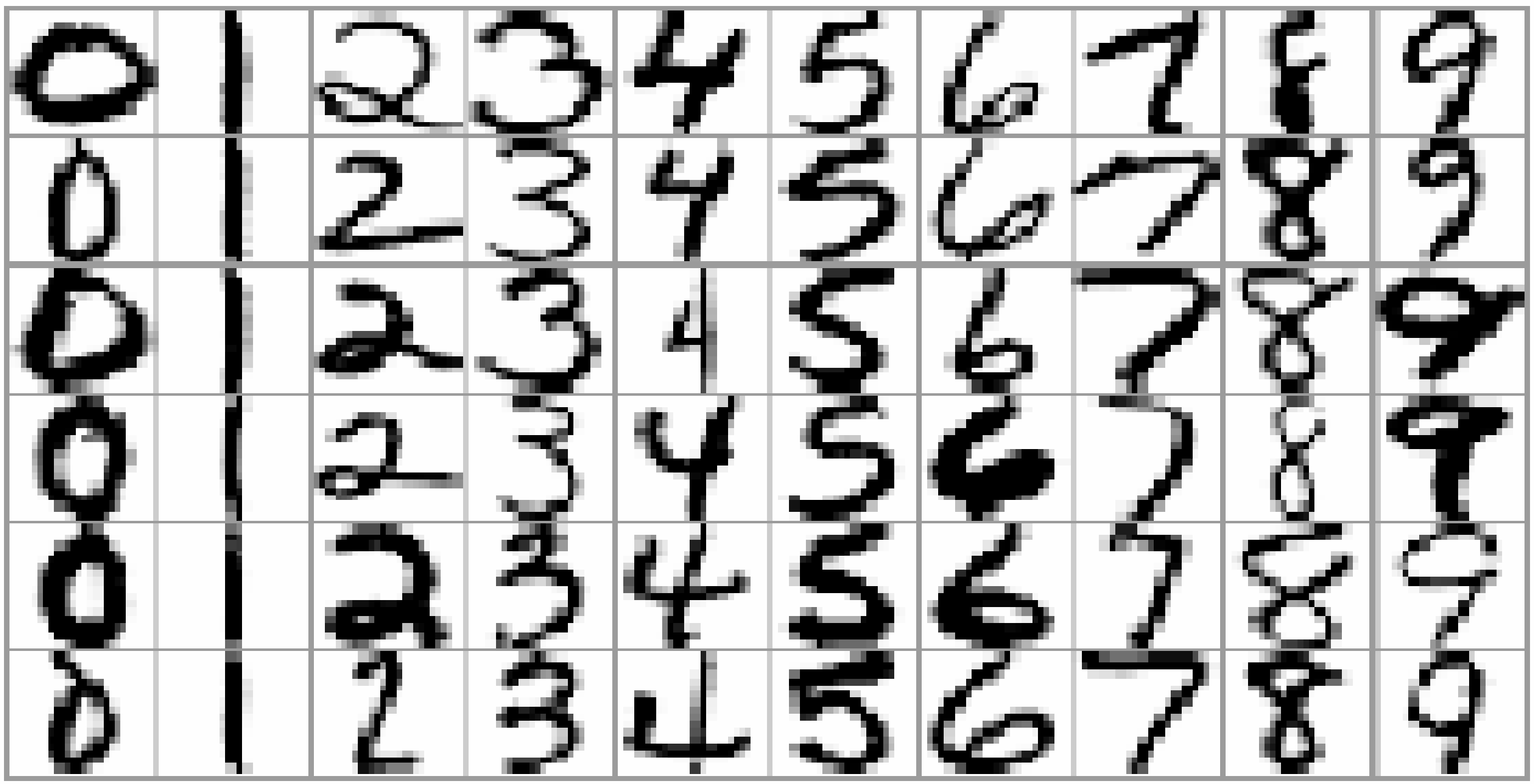
Learning Concepts and Words



Can you pick out the tufas?

Machine Learning:

Recognizing Noisy Input



Machine Learning:

Classic Recognition Problem

Training examples of a person

Test images

AT&T Laboratories, Cambridge UK

http://www.uk.research.att.com/facedatabase.html

Machine Learning:

Why Learn?

- Special Approach to Programming
  - To optimize a performance using example data or past experience.
- Not always needed
  - There is no need to “learn” to calculate payroll
- But used when
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech/image recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

Forms of Learning

Learning technique to use depends on:

- Which component of the agent is to be improved.
- What prior knowledge the agent has.
- What representation is used for the data and the component.
- What feedback is available to learn from.

Components to Improve / Learn

- Condition-Action rules
- Inference of world properties from percepts
- World model / results of actions
- Utility world states
- Utility / cost of actions
- Goals (classes of goal states)
- ...

Possible Representations

- Propositional / First-Order Logic sentences
- Bayesian networks
- Arithmetic functions, e.g., $eval(s) = \sum_i w_i * f_i(s)$
- Neural networks
- ...

Types of Learning by Feedback

- Supervised Learning
  - learning a function from input/output pairs
- Reinforcement Learning
  - learning from reward for a sequence of actions
  - Problem: Which of the actions is responsible for low/high reward?
- Unsupervised Learning
  - learning patterns in the input without feedback
  -

Supervised Learning

Supervised Learning in General

Given a training set of example input-output pairs

$$(x_{1,} y_1), \ldots, (x_N, y_N)$$

generated by an unknown function f such that

$$(\forall i)\, y_i = f(x_i)$$

discover a function h (hypothesis) that approximates f: $(\forall x)\, h(x) \approx f(x)$

Accuracy of h is measured using a test set of examples. $\textit{test set} \cap \textit{training set} = \emptyset$

What are good hypotheses?

Consistent hypothesis is trivial:

$$\left(\forall i : 1 \leq i \leq N\right) h\left(x_i\right) = f\left(x_i\right)$$

Goal: Generalization $\left(\forall x\right) h\left(x\right) \approx f\left(x\right)$

Objective: Minimize error on the training set and achieve good generalization.

Accuracy of h is measured using a test set of examples: $\textit{test set} \cap \textit{training set} = \emptyset$

e.g., $$error\left(h\right) = \sum_i \left|h\left(x_i\right) - y_i\right|$$

Two common learning problems

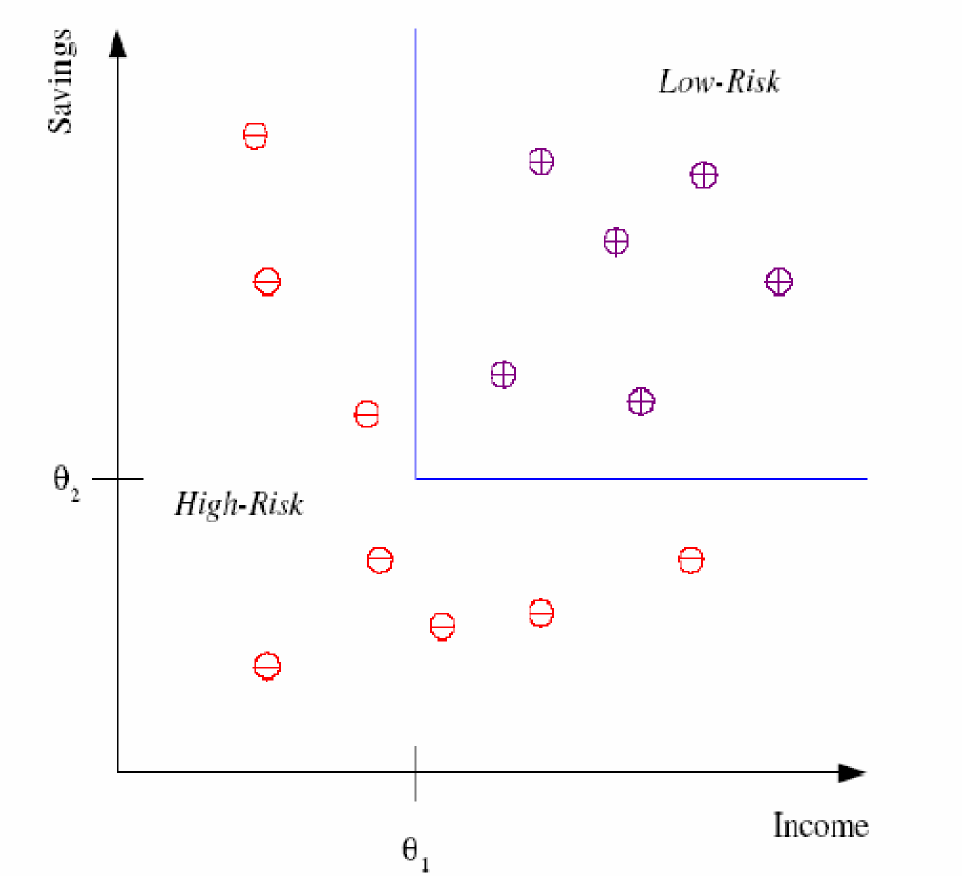
- Classification
  - finite set of output values y
  - e.g., sunny/cloudy/rainy, red/blue/green, ...
- Regression
  - output is a number

Supervised Learning:
Classification

- Example: Credit scoring

  Differentiating between <u>low-risk</u> and <u>high-risk</u> customers from their *income* and *savings*

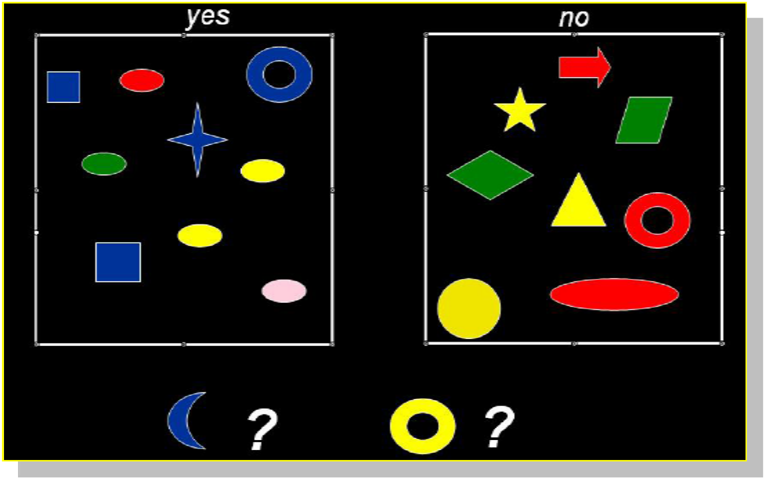
- Input data is two dimensional, output is binary



Discriminant:
IF *income*> θ1 AND *savings*> θ2 THEN <u>low-risk</u>
ELSE <u>high-risk</u>

Supervised Learning:

Classification



p features (attributes)

n cases

Training Set:

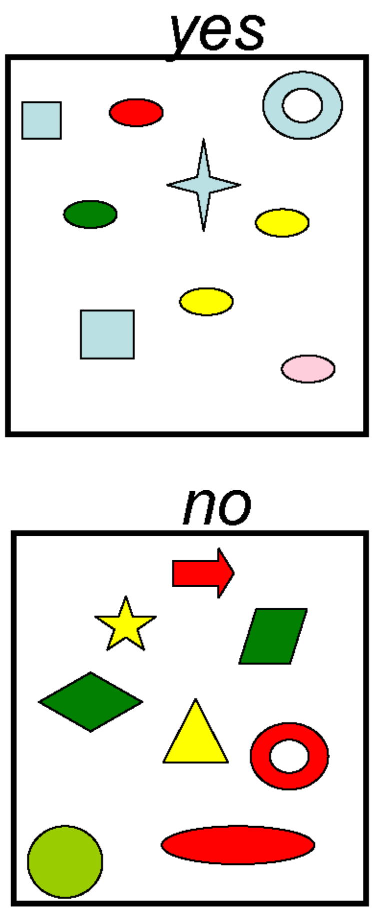
Color	Shape	Size	Label
Blue	Square	Small	Yes
Red	Ellipse	Small	Yes
Red	Ellipse	Large	No

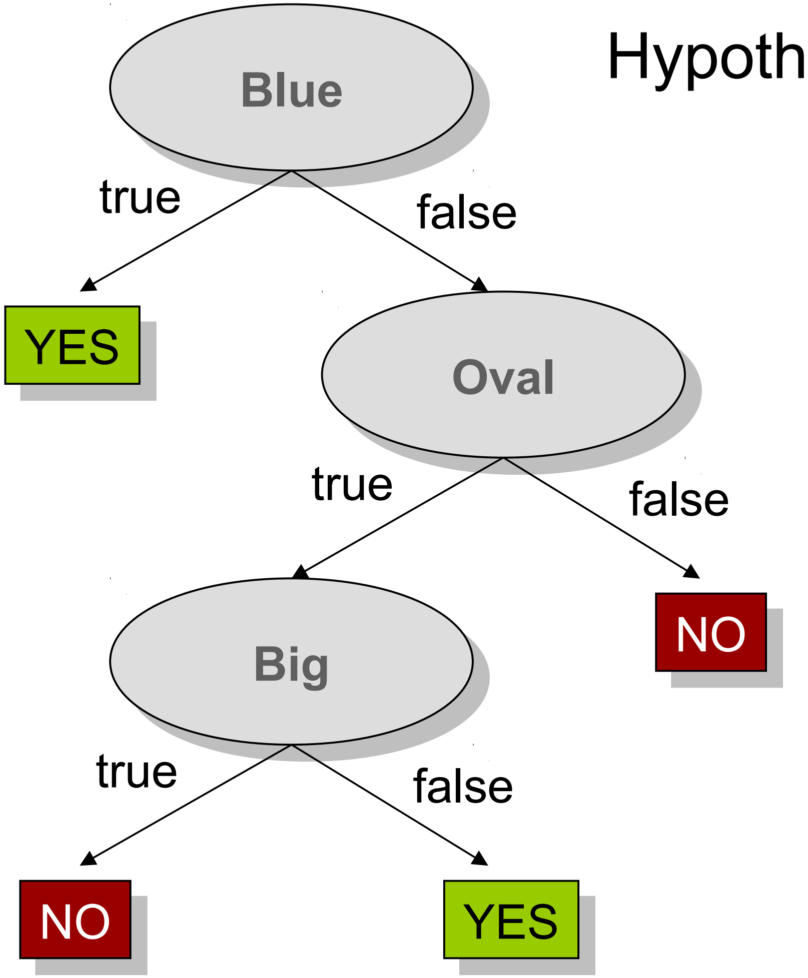
Test Set:

Blue	Crescent	Small	?
Yellow	Ring	Small	?

Supervised Learning:

Classification - Decision Tree



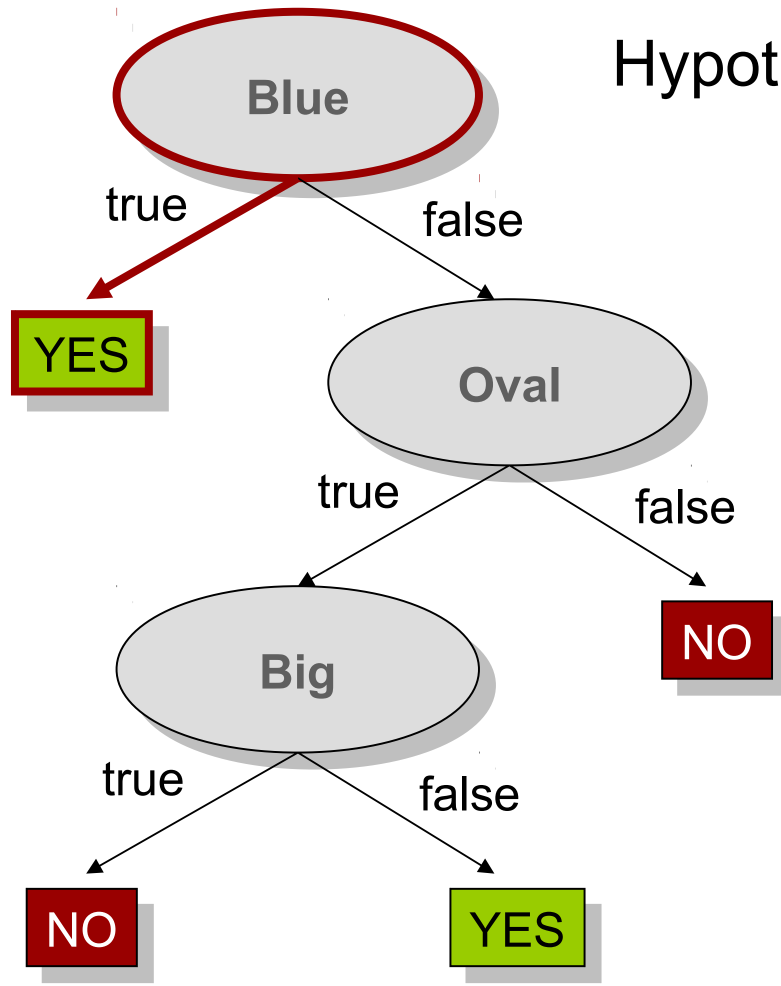


Hypothesis

Supervised Learning:

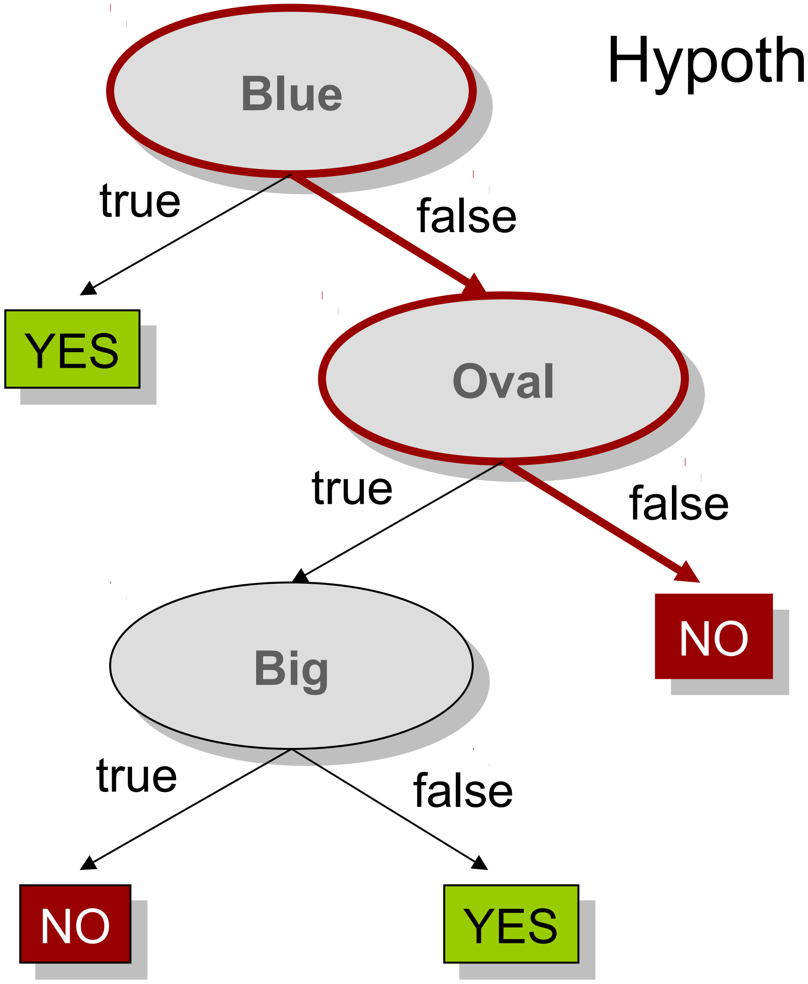
Classification - Decision Tree

Hypothesis



Supervised Learning:

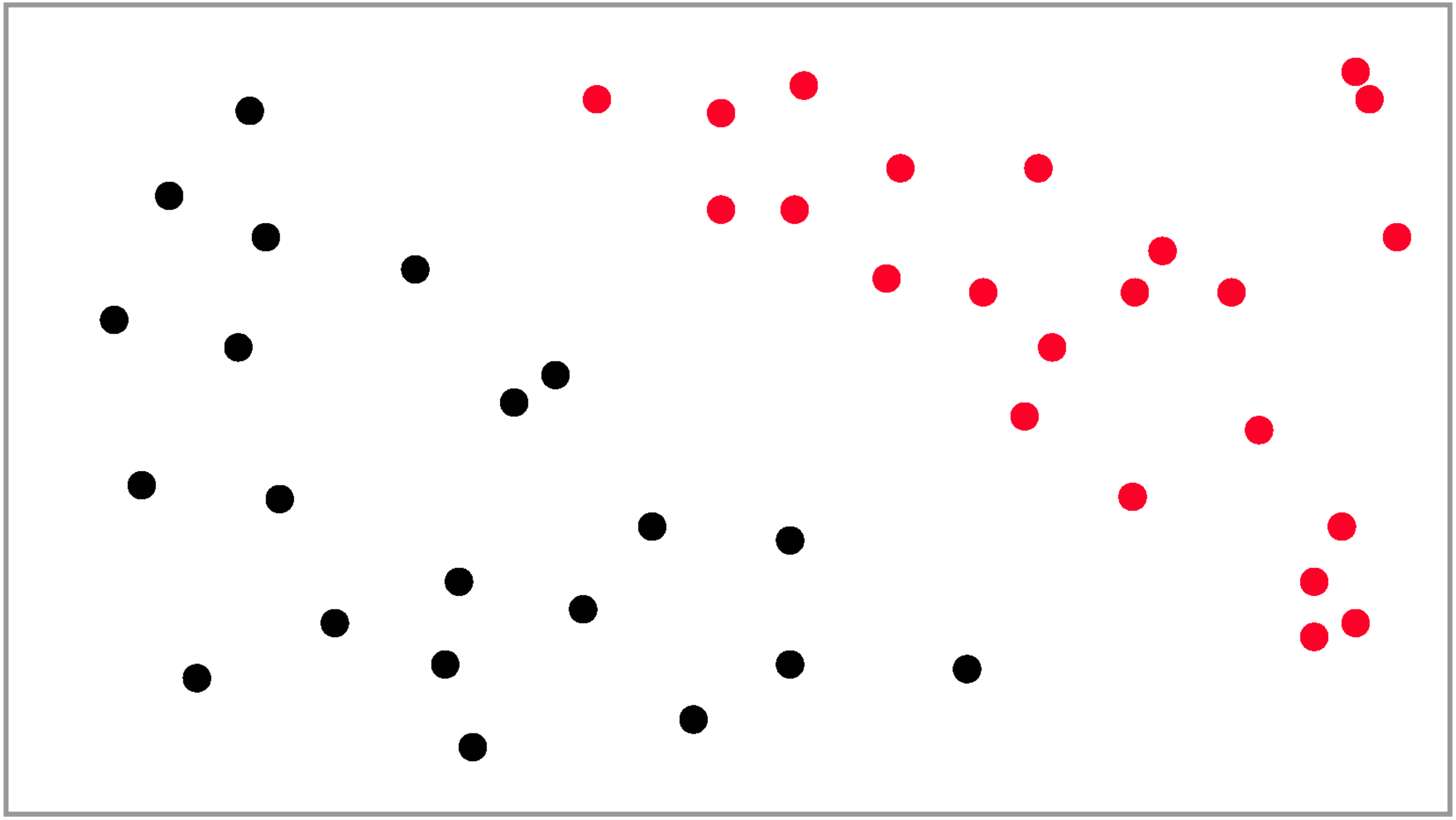
Classification - Decision Tree



Hypothesis

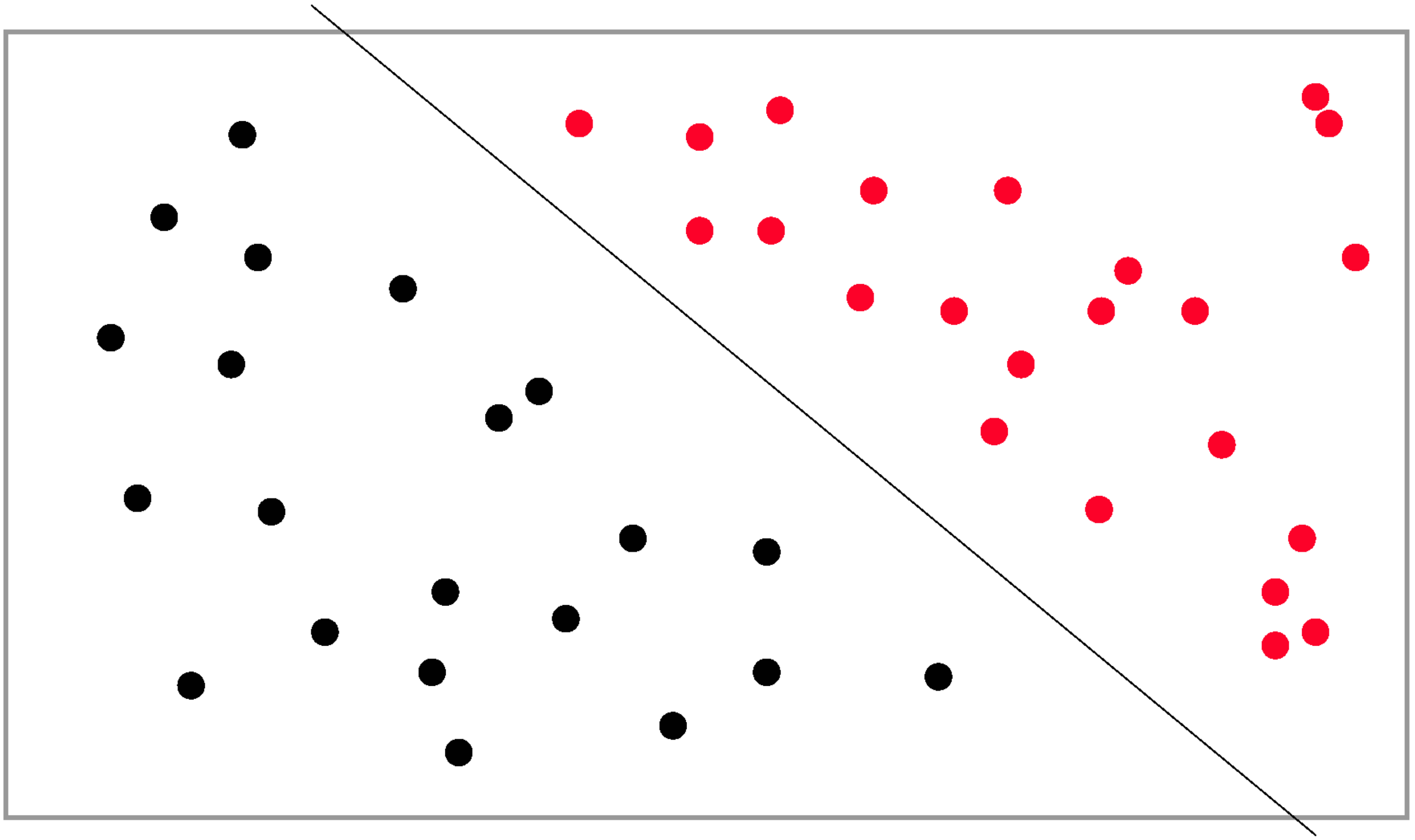
Supervised Learning:

What is the right Hypothesis?



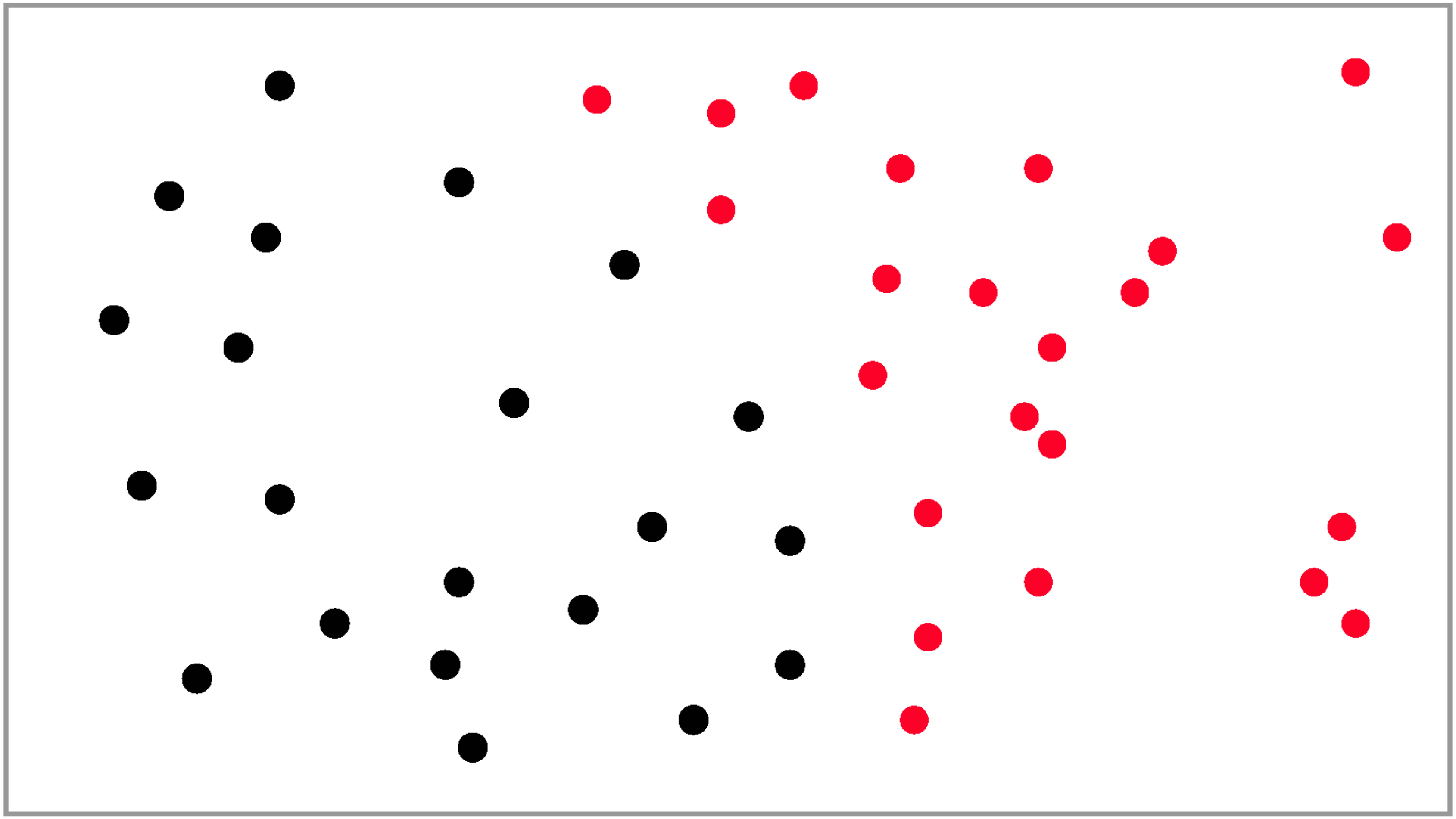
Supervised Learning:

Hypothesis – Linear Separation



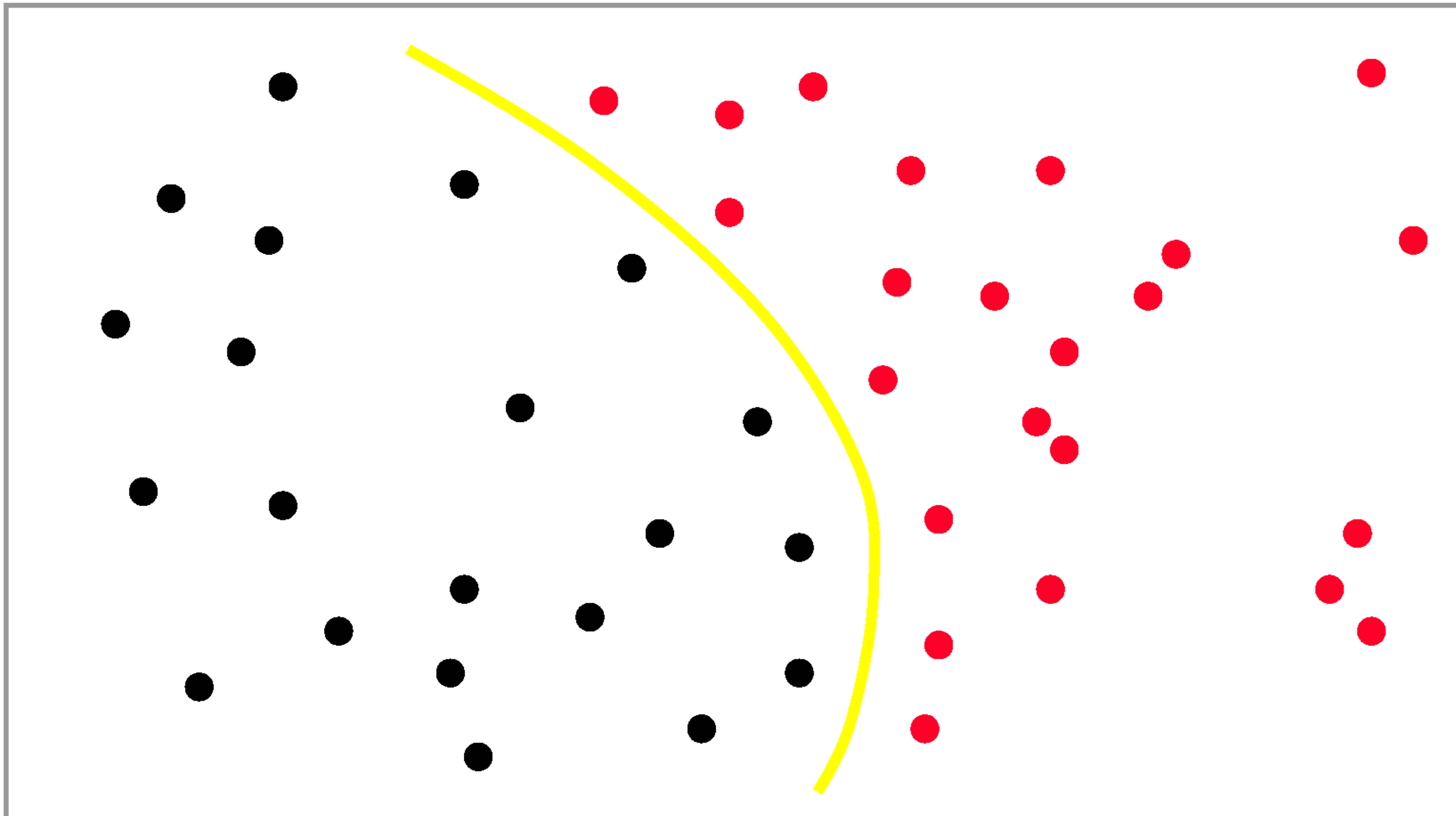
Supervised Learning:

Hypothesis – Linear Separation?



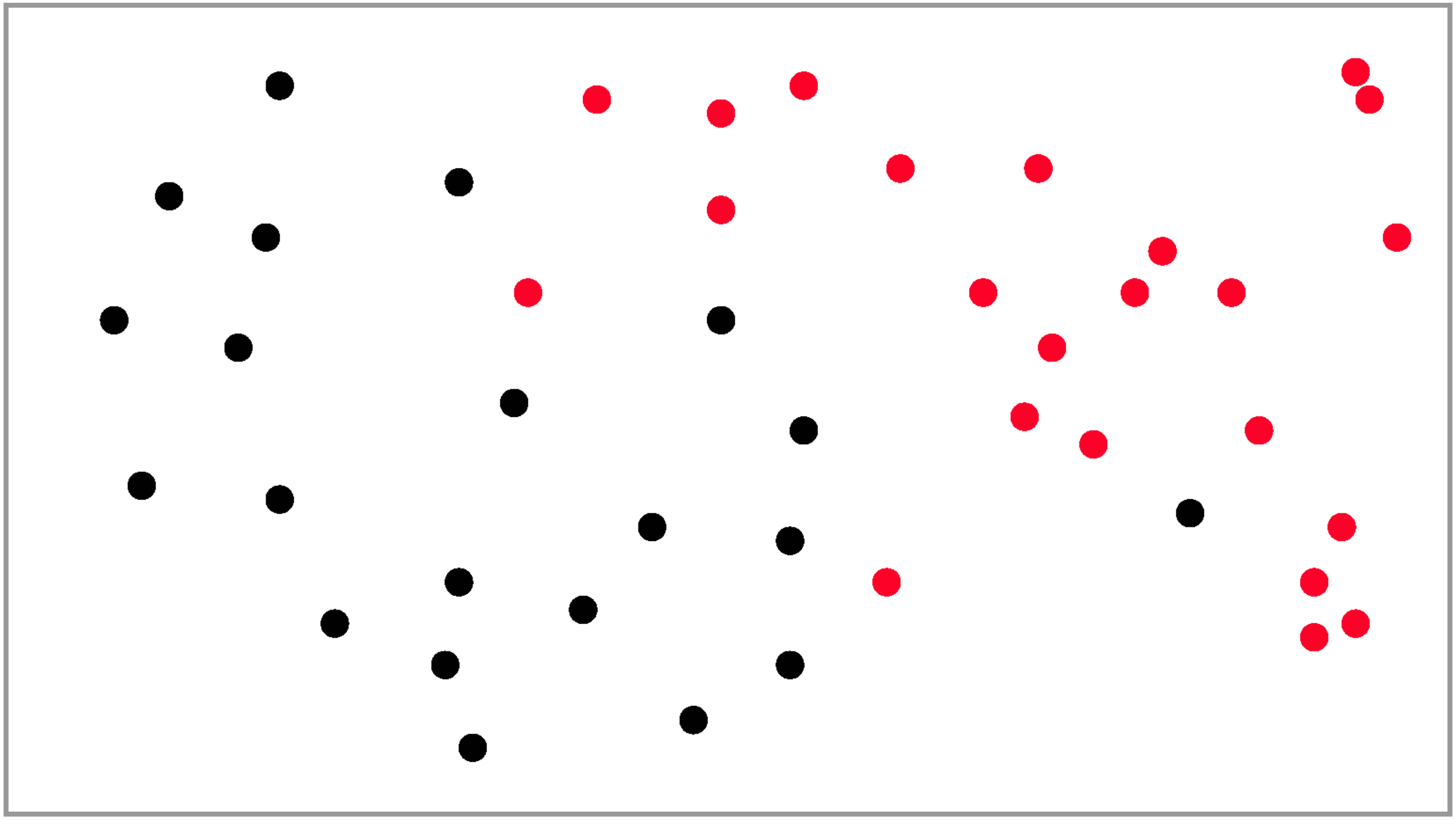
Supervised Learning:

Hypothesis – Quadratic Separation



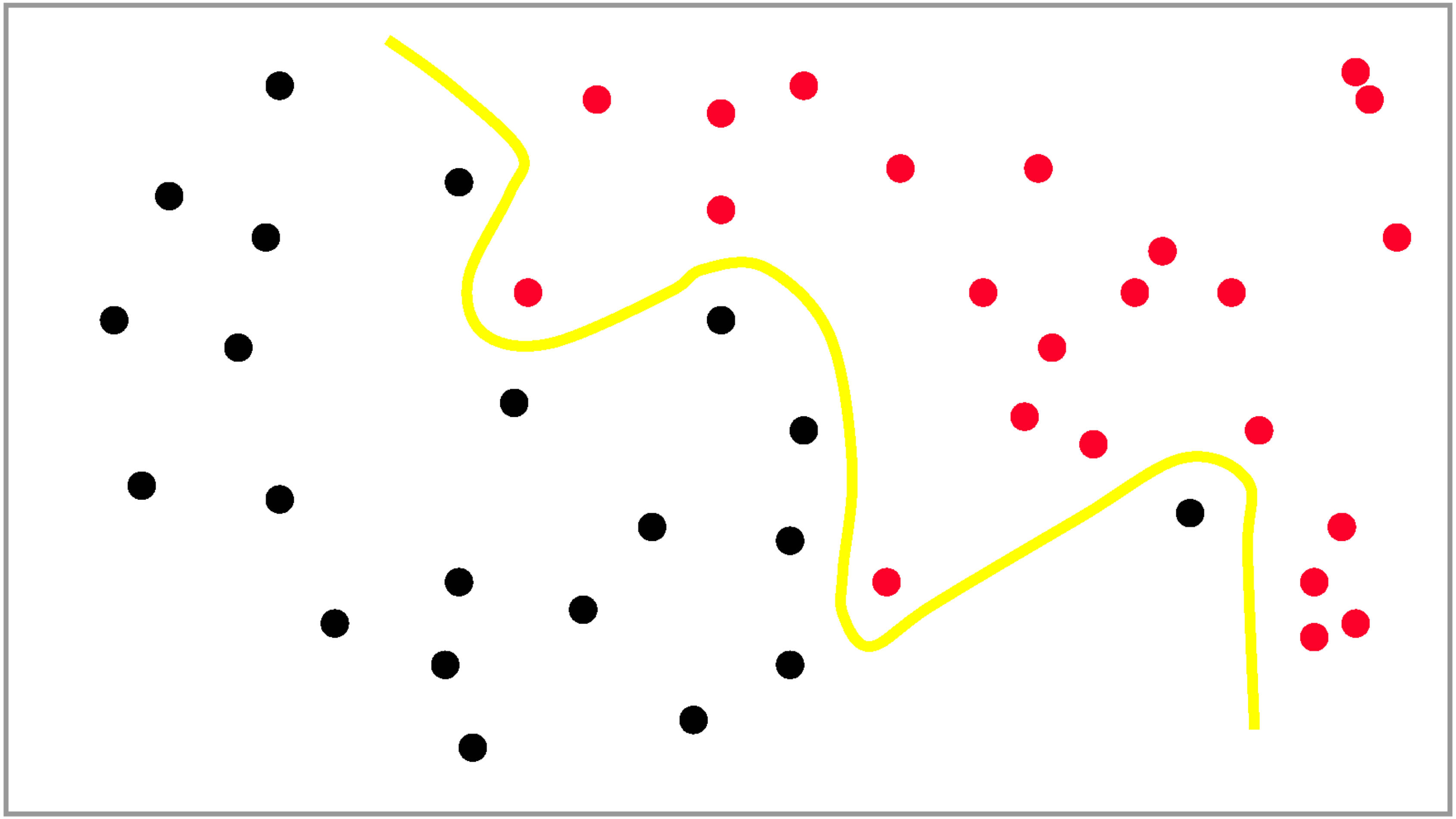
Supervised Learning:

Problem – Noisy/Mislabeled Data



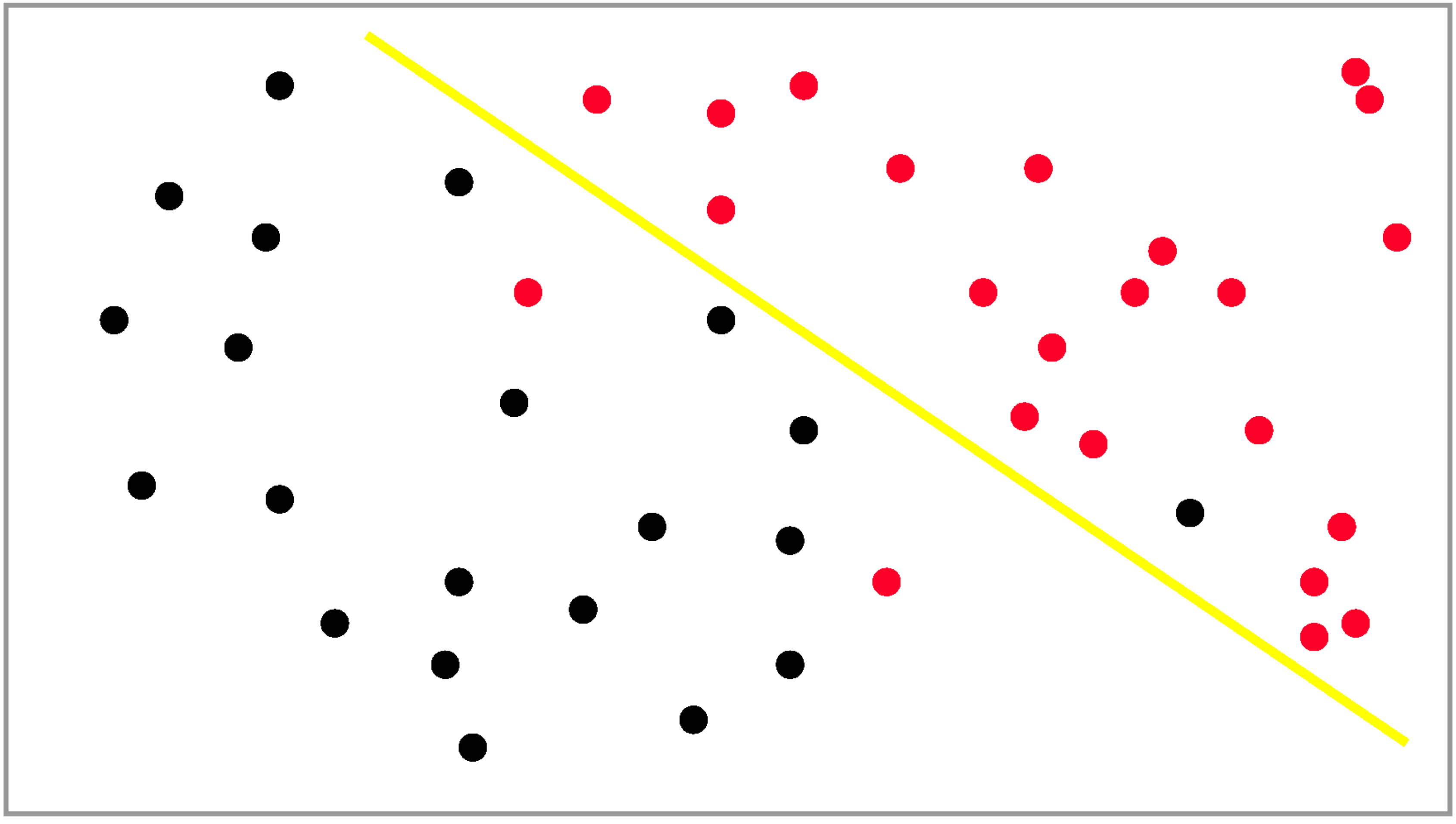
Supervised Learning:

Avoid: Overfitting



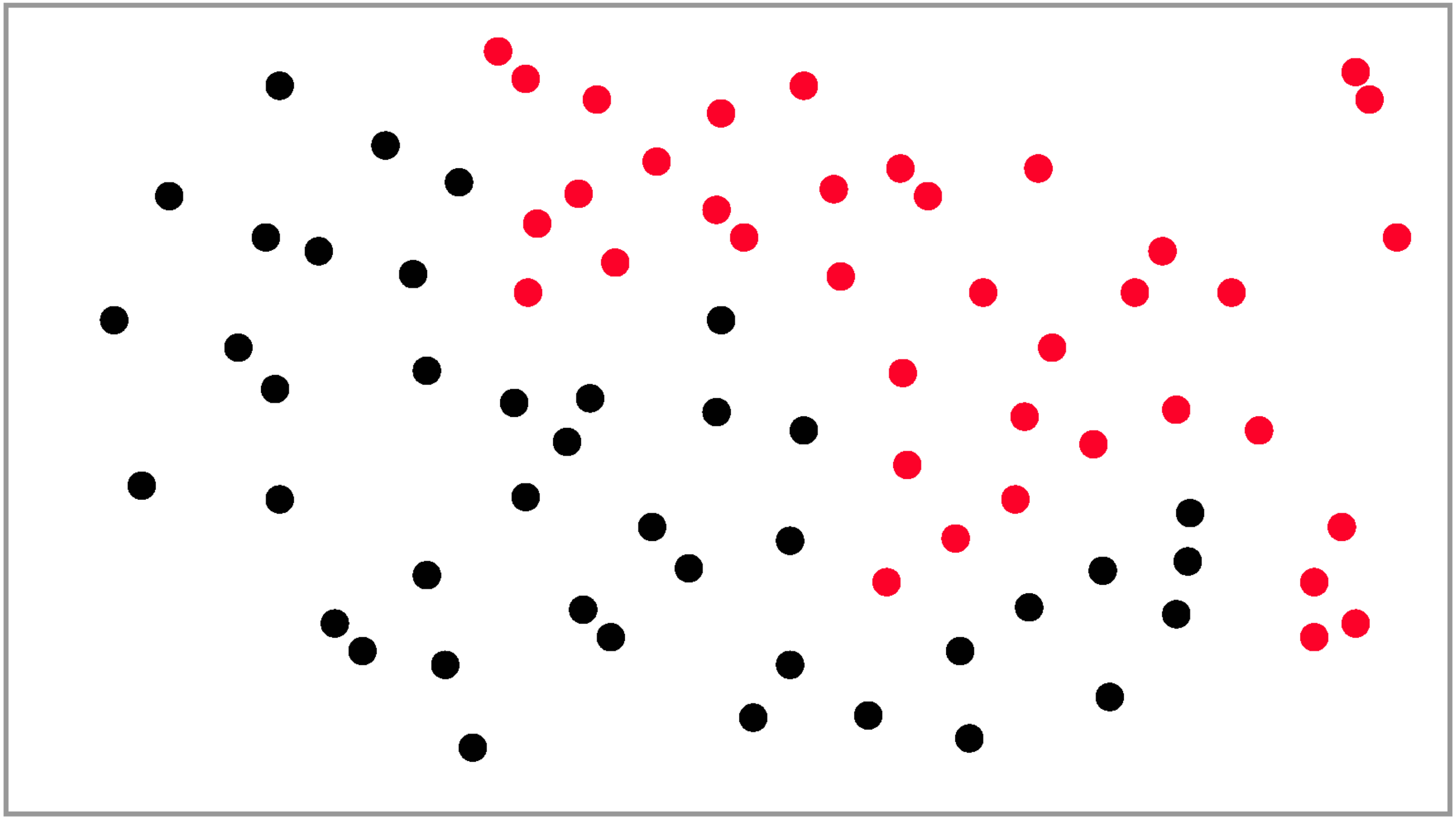
Supervised Learning:

Is the hypothesis underfitting?



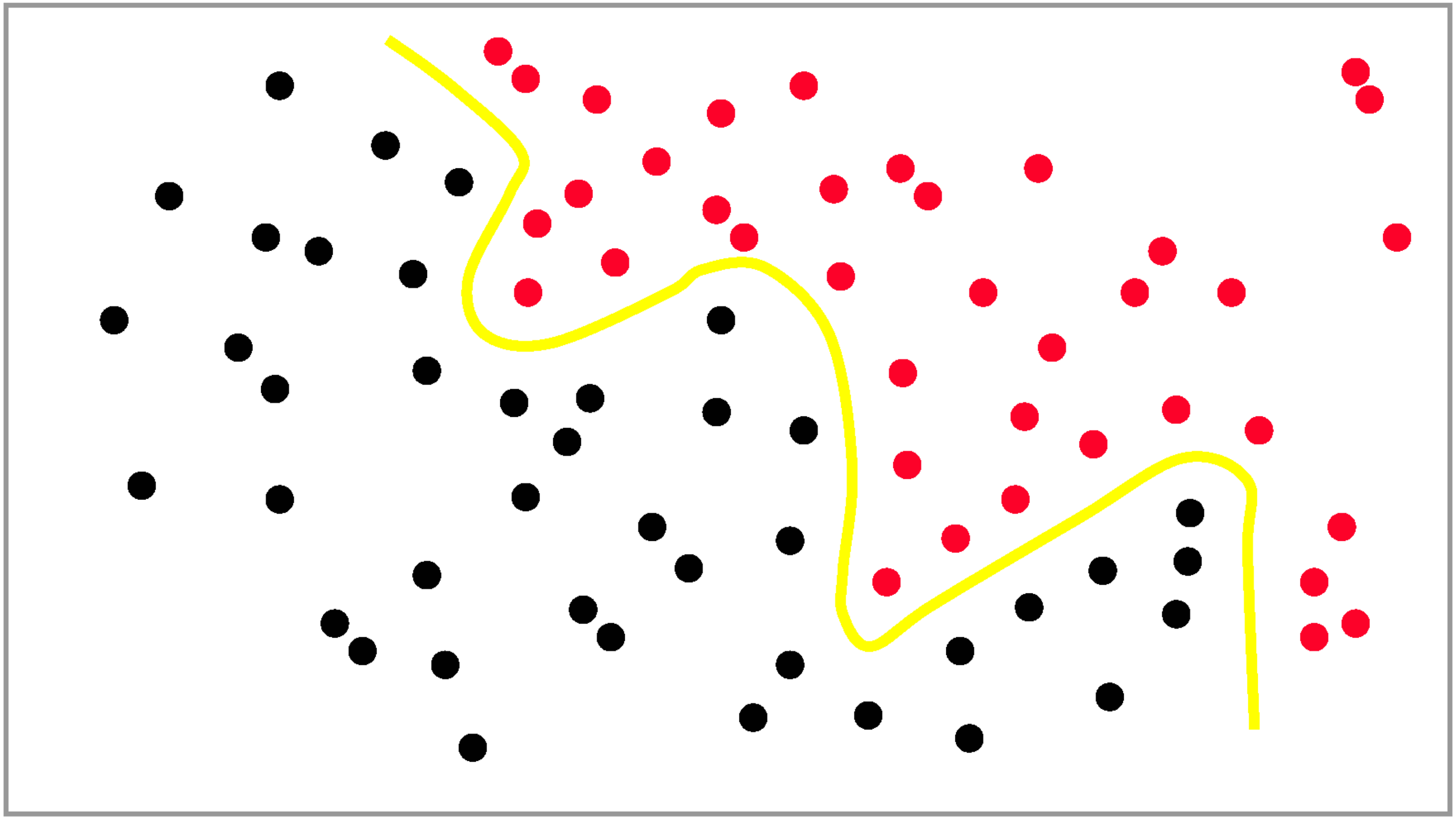
Supervised Learning:

More data $\rightarrow$ ...



Supervised Learning:

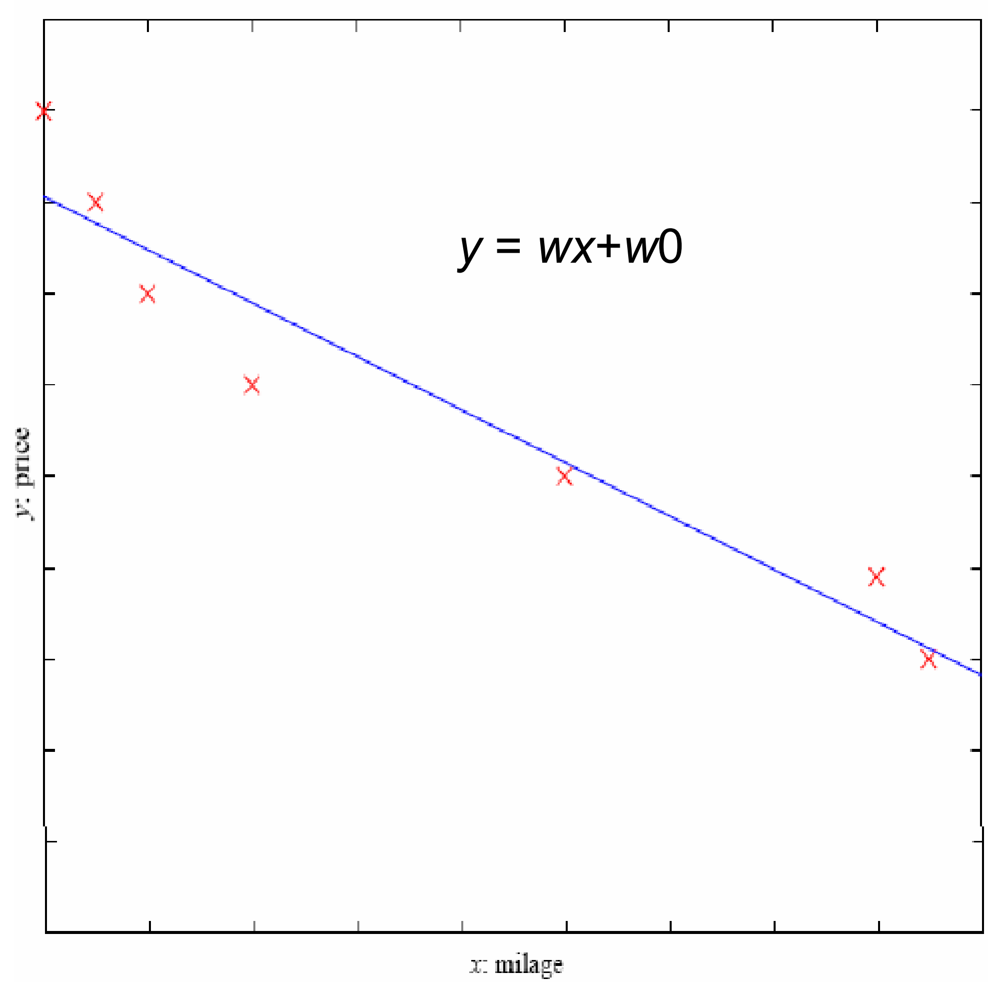
... → more complex hypothesis



Supervised Learning: Linear Regression

- Example:
  Price of a used car
  x : car attribute
  y : price

- $y = g(x \mid \theta)$
  model:
  $g(\,)$
  parameters:
  $\theta = (w, w_0)$

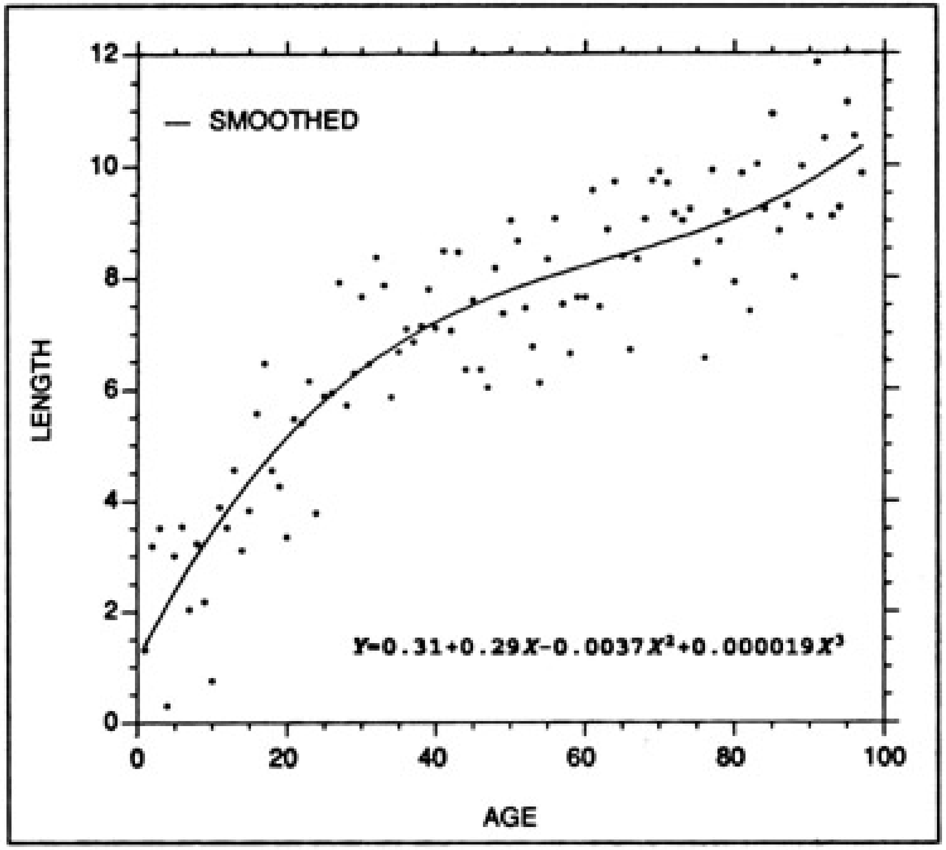


Supervised Learning:

Polynomial Regression

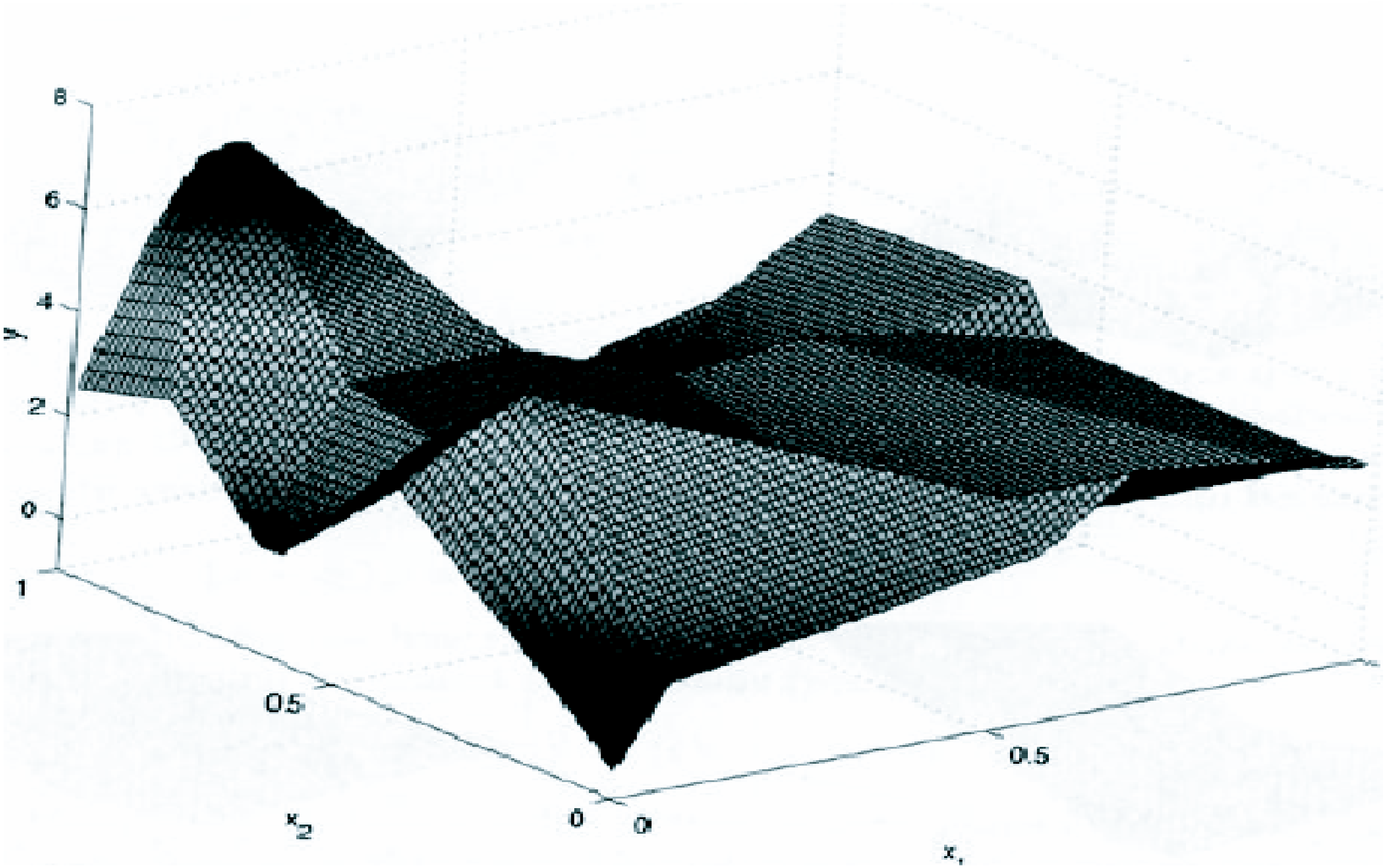
- Example:
  Growth of a species
  x : age
  y : length

- $y = g\ (x \mid \theta)$
  model:
  $g\ (\ )$
  parameters:
  $\theta = (w_3, w_2, w_1, w_0)$



Supervised Learning:

Piecewise Linear 2D Regression



Supervised Learning:

Some Regression Applications

- Cost estimation
  - Energy consumption
- Control
  - Angle of steering wheel for robot car
  - Kinematics of a robot arm

Supervised Learning:

Range of Methods

- Methods differ in terms of
  - The form of hypothesis space
  - The way to find best hypothesis given data
- There are many successful approaches
  - Decision trees
  - Support vector machines
  - Neural networks
  - Case-based reasoning
  - ...

Supervised Learning:

General 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)

Reinforcement Learning

Reinforcement Learning:

Overview

- Characteristics
  - Learning a Policy: A sequence of outputs
  - No supervised output, but a delayed reward
  - Credit assignment problem:
    - Which action led me to winning the game?
- Examples
  - Elevator scheduling
  - Backgammon and Chess
  - Robot control

Unsupervised Learning

Unsupervised Learning:

Overview

- General characteristics
  - Learning “what normally happens”
  - No output available
- Examples
  - Clustering
  - Dimensionality reduction
  - Abnormality detection
  - Latent/hidden variable estimation

Unsupervised Learning:

Example: Image Clustering