Machine Learning Models for Classification

Presentation: V. Vijayarajan, Associate Professor of School of Computing Science and Engineering,

VIT University

Courtesy for Presentation (Edited): Yaser Abu Mostafa, Professor of Electrical Engineering and Computer Science, Caltech

Machine Learning Models for Classification

The Learning Problem

Today's Agenda:

- Examples of Machine Learning
- Components of Learning
- A Simple Model
- Types of Learning

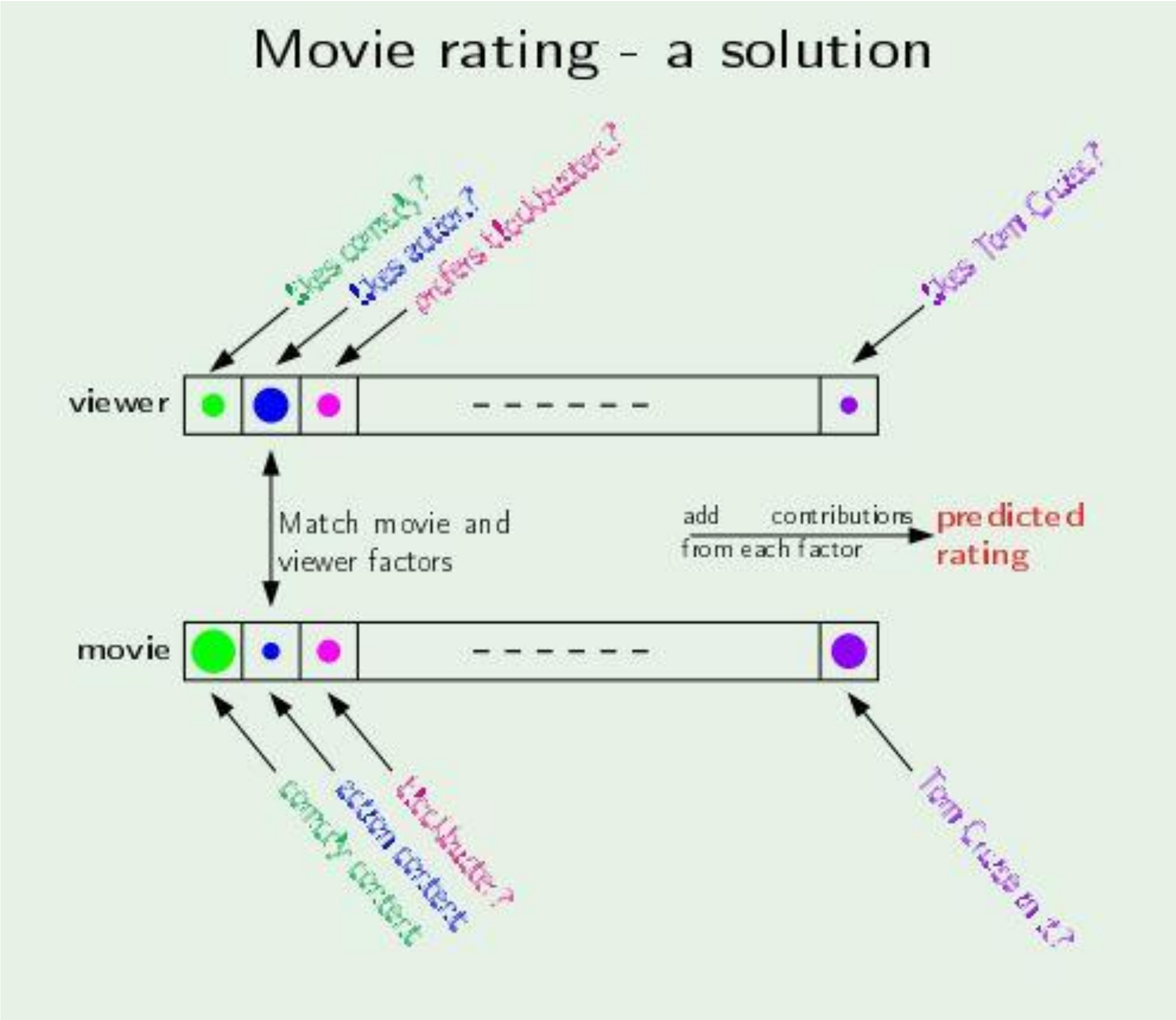
Machine Learning Models for Classification

The Learning Problem

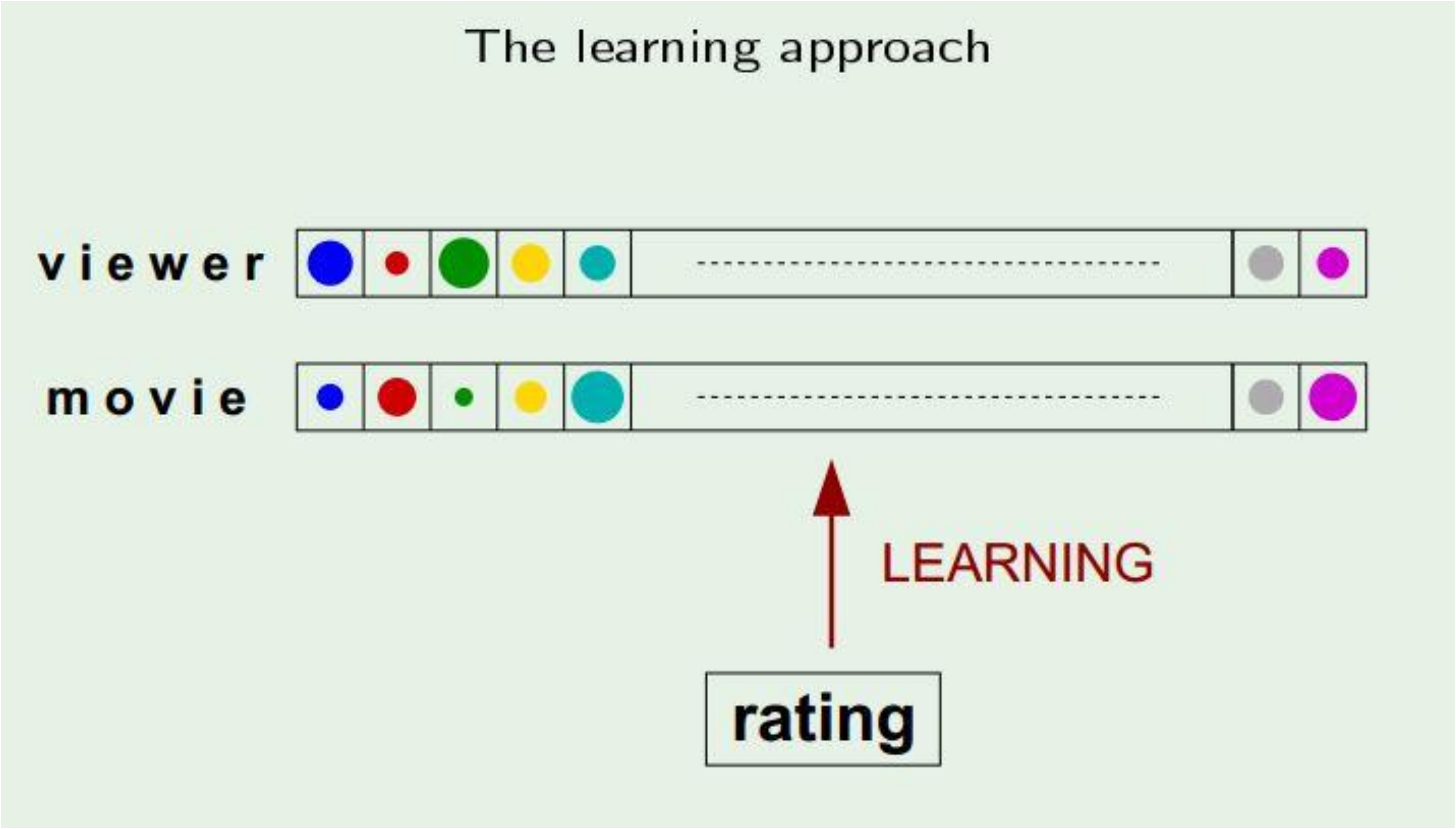
Essene of machine learning:

- Without data no business of taking about Machine Learning
- A pattern must exists in data
- We cannot pin it down mathematically

Machine Learning Models for Classification



Machine Learning Models for Classification



Machine Learning Models for Classification

Components of learning

Metaphor: Credit Approval

Applicant Information:

age	23 Years
gender	male
annual salary	Rs. 30,000
years in residence	1 year
years in job	1 year
current dept	Rs. 15,000
…	…

Approve credit?

Machine Learning Models for Classification Contd..

The Learning Problem

Essene of machine learning: (Confirmation)

- A pattern exists
- We cannot pin it down mathematically
- We have data on it

Components of Learning

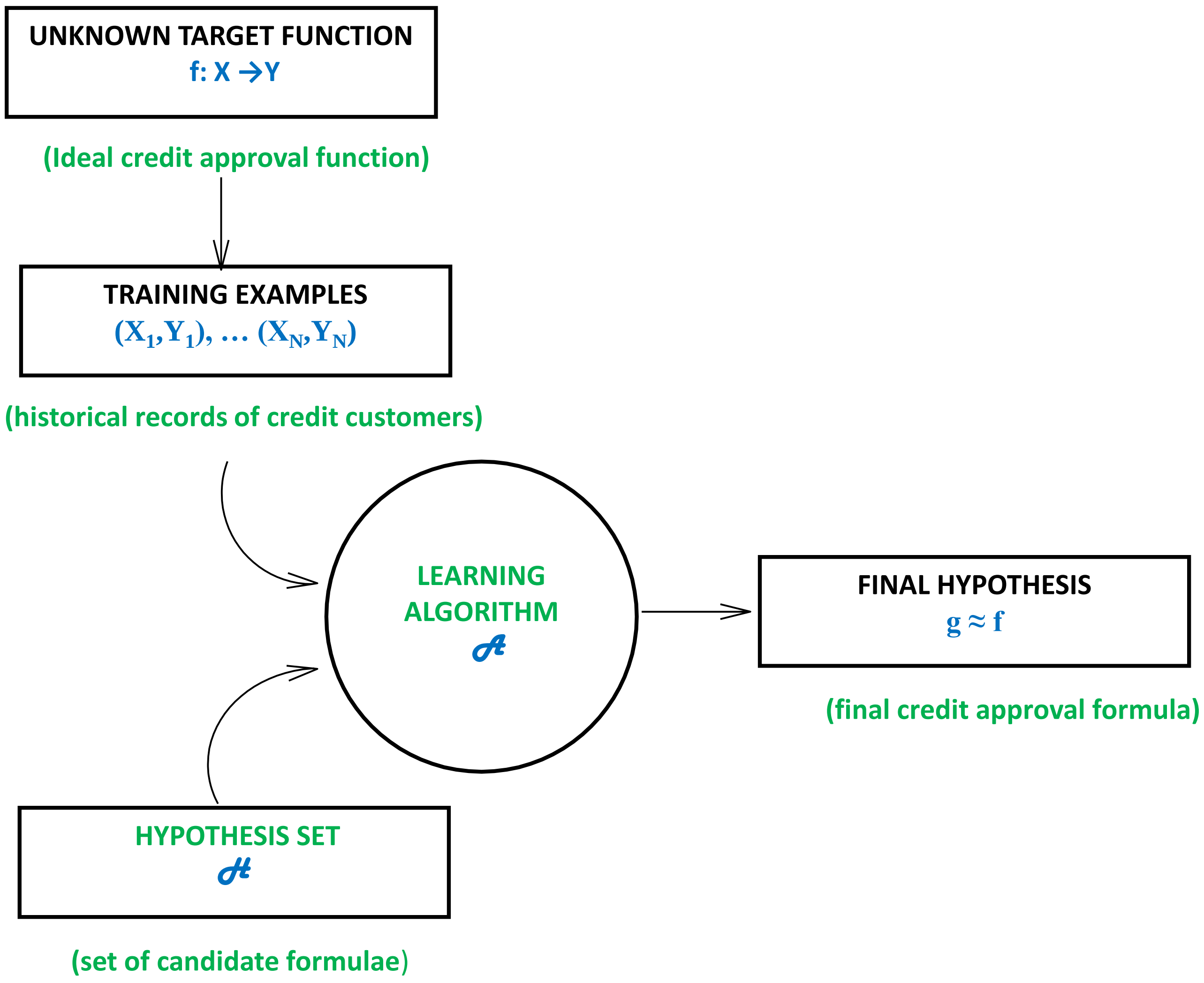
Formalization:

- Input: x (customer application)
- Output: y (good/bad customer)
- Target function: f: $X \rightarrow Y$ (ideal credit approval formula)
- Data: (X_1,Y_1), (X_2,Y_2), …. (X_N,Y_N) (historical records)

  ↓ ↓ ↓

- Hypothesis: g: $X \rightarrow Y$ (formula to be used)

Machine Learning Models for Classification Contd..



Machine Learning Models for Classification

Solution Components:

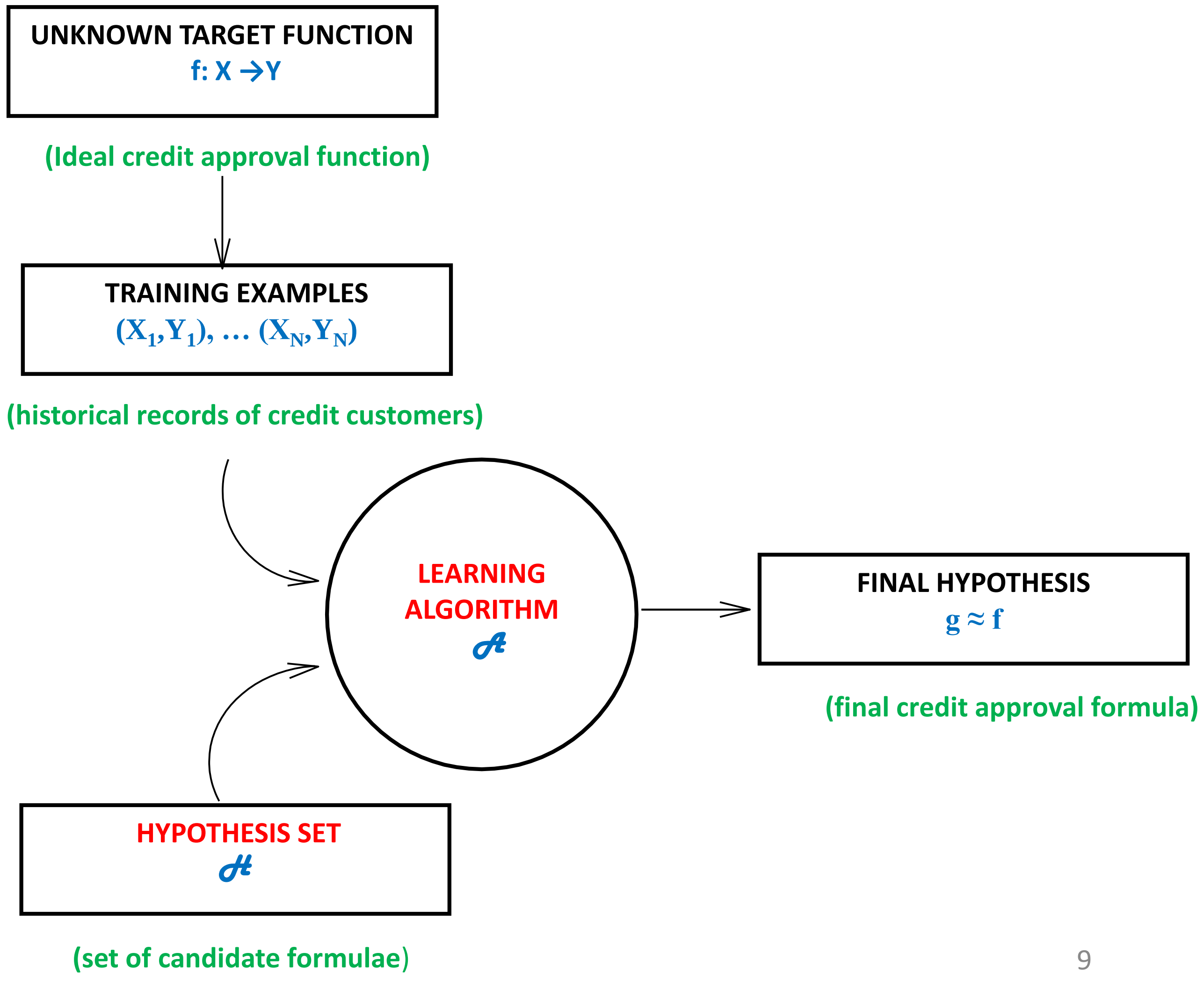
The 2 solution components of the learning problem

- The Hypothesis set

  $\mathcal{H} = \{h\}$ $\quad\quad g \in \mathcal{H}$

- The Learning Algorithm

Together, they are referred to as the learning model



Machine Learning Models for Classification Contd..

A simple hypothesis set – the 'perceptron':

For input $x=(x_{1,}\ x_{2,}\ \ldots\ x_d)$ attribute of a customer

$$\text{Approve credit if } \sum_{i=1}^{d} w_i x_i > \text{threshold}$$

$$\text{Deny credit if } \sum_{i=1}^{d} w_i x_i < \text{threshold}$$

This linear formula $h \in \mathcal{H}$ can be written as

$$h(x) = \text{sign}\left(\sum_{i=1}^{d} w_i x_i - \text{threshold} \right)$$

Machine Learning Models for Classification Contd..

A simple hypothesis set – the ‘perceptron’:

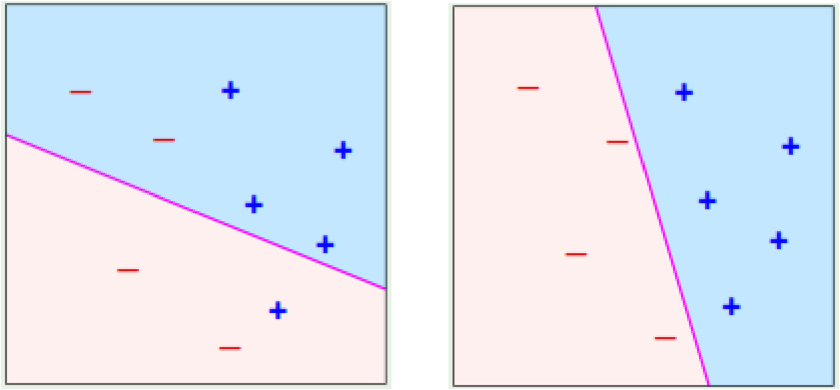
$$h(x) = \text{sign}\left(\left(\sum_{i=1}^{d} w_i x_i\right) + w_0\right)$$

Introduce an artificial co-ordinate $x_0=1$:

$$h(x) = \text{sign}\left(\sum_{i=0}^{d} w_i x_i\right)$$

In vector form, the perceptron implements

$$h(x) = \text{sign}(w^T x)$$



‘linearly separable’ data

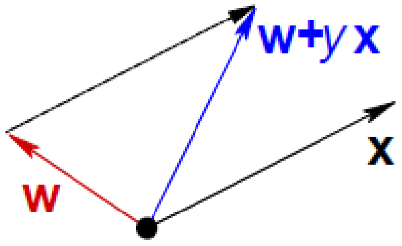
Machine Learning Models for Classification Contd..

A simple learning algorithm PLA:

The perceptron implements

$$h(x) = \text{sign}\ (w^T x)$$

$\boxed{y = +1}$



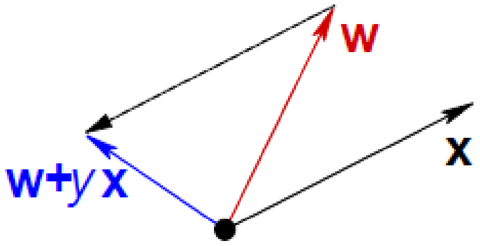
Given the training set:

$$(x_1,y_1), (x_2,y_2), \ldots. (x_N,y_N)$$

pick a misclassified point:

$$\text{sign}\ (w^T x_n) \neq y_n$$

$\boxed{y = -1}$



and update the weight vector

$$\boxed{w \leftarrow w + y_n x_n}$$

Machine Learning Models for Classification Contd..

Iterations of PLA:

- One iteration of the PLA:

$$w \leftarrow w + y_n x_n$$

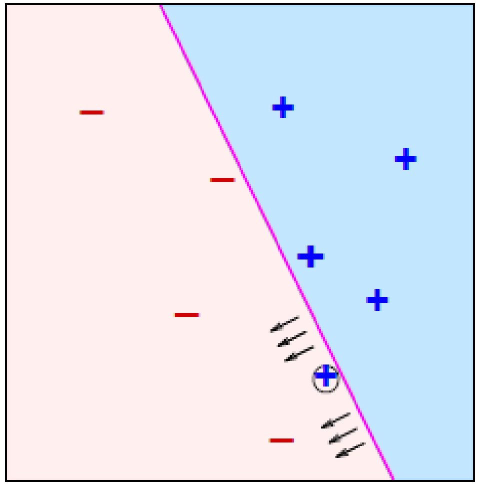
where (x, y) is a misclassified training point

- At iteration t = 1, 2, 3, …. pick a misclassified point from

$(x_1,y_1), (x_2,y_2), \ldots. (x_N,y_N)$

and run a PLA iteration on it.

- That's it!



Types of Learning

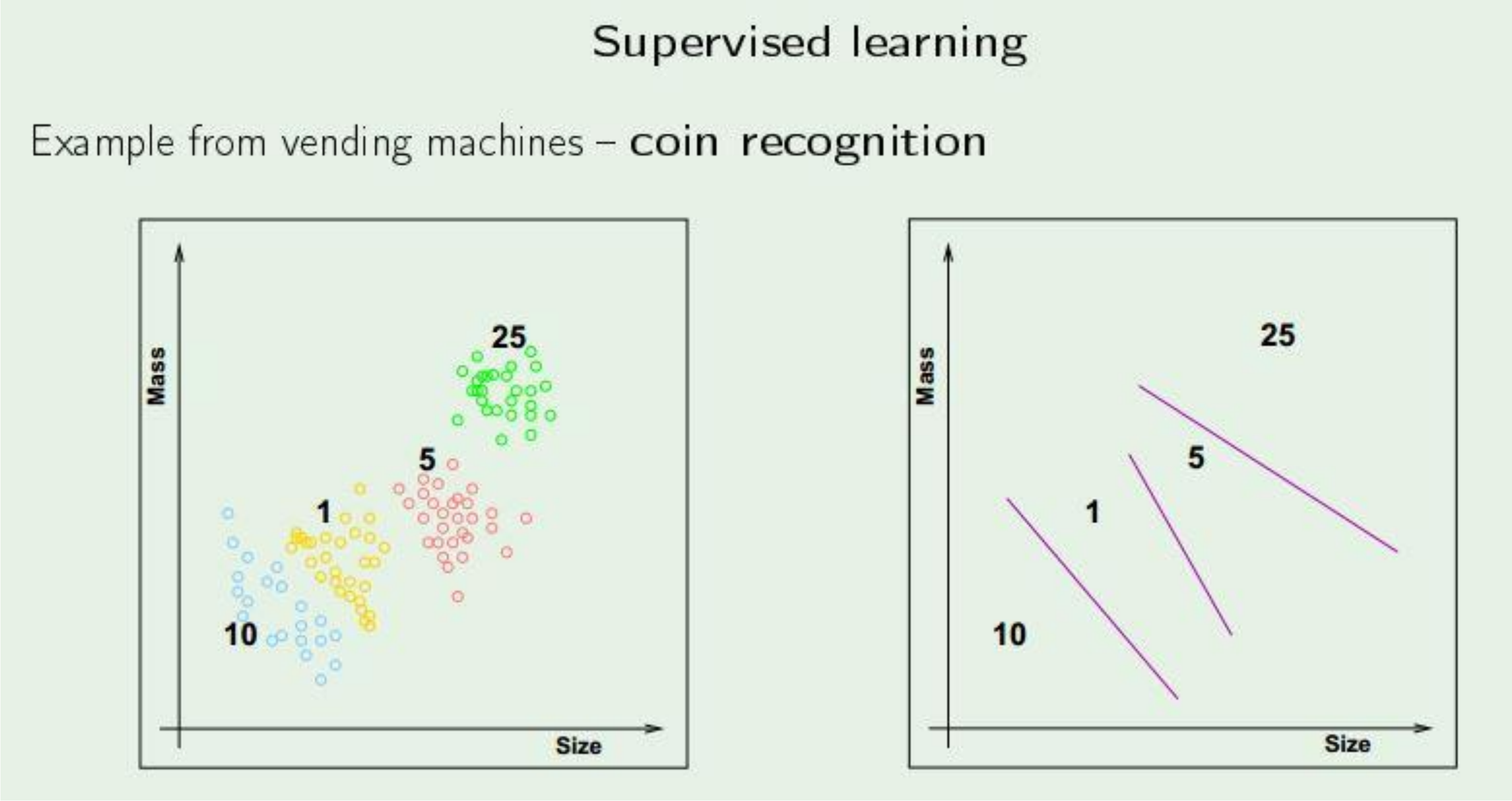
Basic premise of learning

"using a set of observations to uncover an underlying process"

broad premise $\Longrightarrow$ many variations

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

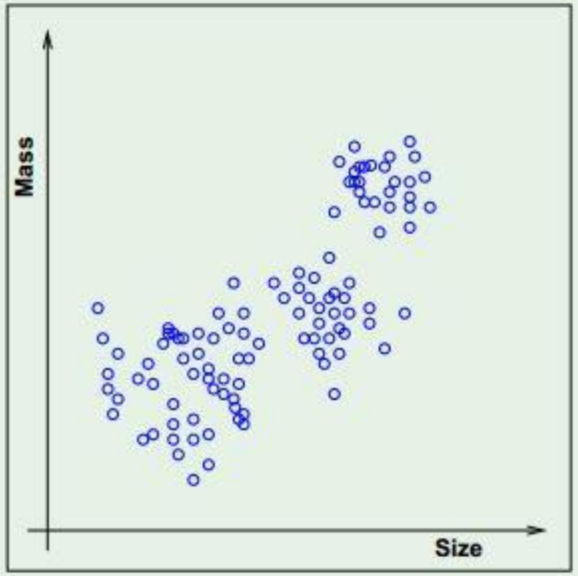
Types of Learning



Types of Learning

Unsupervised learning

Instead of (input,correct output), we get (input, ?)



Types of Learning

Reinforcement learning

Instead of (input, correct output),

we get (input, *some* output, grade for this output)

Other Machine Learning Models for Classification

- Multilayer Perceptron Neural Networks (MLP)
- Radial Basis Function Networks (RBFN)
- Support Vector Machines (SVMs)
- Single Decision Tree (SDT)
- Decision Tree Forests (DTF)
- Deep Learning
- Transfer Learning

Biologically Inspired Learning Algorithms for Classification

- Mathematical optimization is the selection of the best solution from an available set of alternatives.

- Optimization functions are generally
  a) Convex – function having unique minimum and hence converge faster
  b) Non Convex - function having many local minima and hence stuck in local minima without converging to global minima

- The following statements are true about the convex minimization problem:
  a) if a local minimum exists, then it is a global minimum.
  b) the set of all (global) minima is convex.
  c) for each strictly convex function, if the function has a minimum, then the minimum is unique.

- Global Optimization is a NP complete problem and heuristic approaches like Genetic Algorithms (GA), Particle Swarm Optimization and Simulated Annealing have been used to provide near optimum solutions for non-convex optimization problems.

- Some of the existing biologically inspired optimization algorithms are:
  a) Genetic Algorithm (GA)
  b) Particle Swarm Optimization (PSO)
  c) Bee Colony Optimization (BCO)

My Inspirational Research Quote for ever is: "*It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong*"

- ***Richard P. Feynman, Nobel Laureate, Quantum Physics***