CSC3022H:

Machine Learning: Introduction

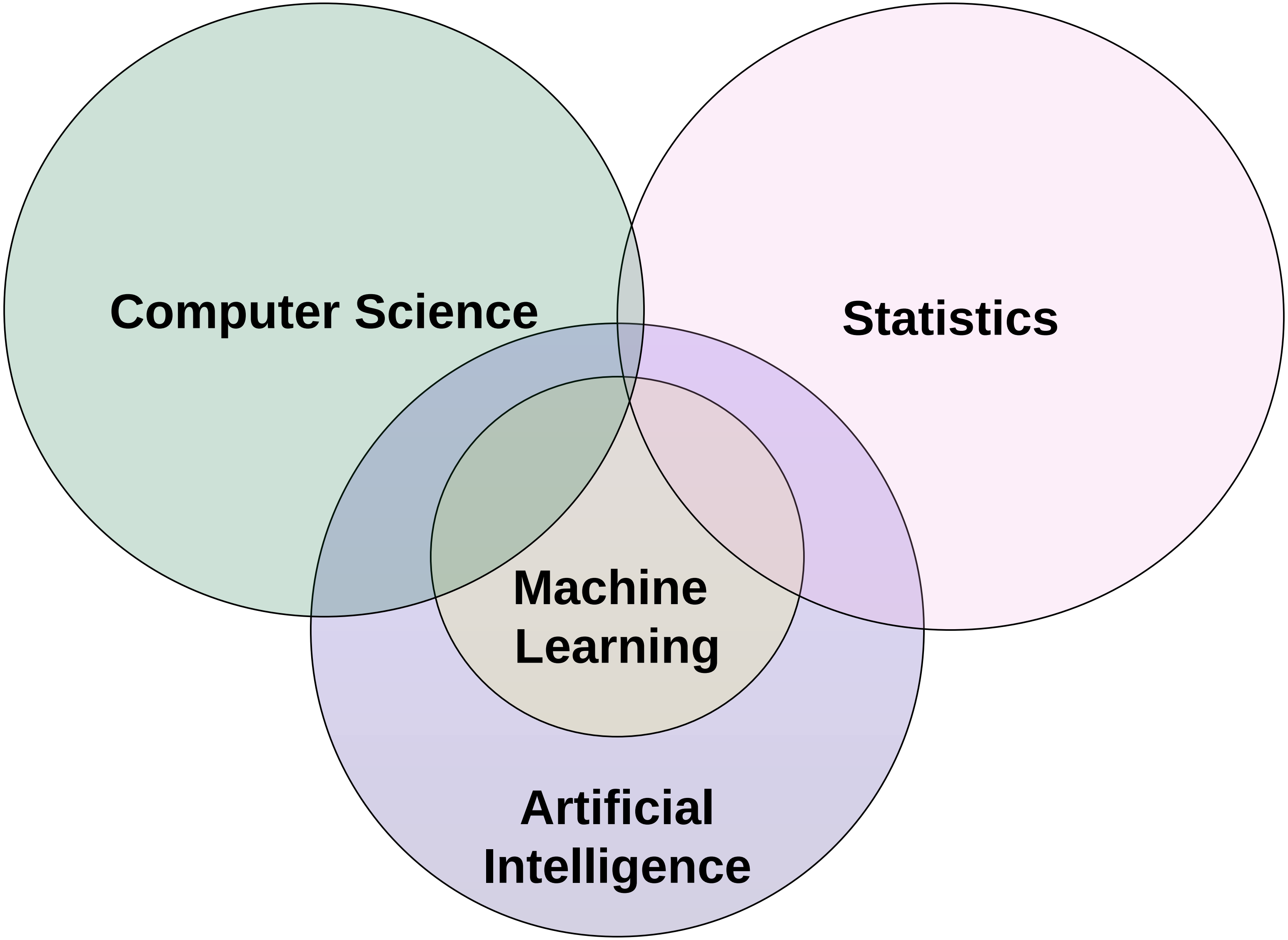
Geoff Nitschke

Department of Computer Science
University of Cape Town, South Africa

Course Syllabus

- **Supervised Learning:**
  - **ANNs: Back propagation**.
  - Generative Learning algorithms: Monte-Carlo.
  - *Concept Learning.*
- **Unsupervised Learning:**
  - Clustering: Hierarchical clustering, *K-means*, EC, NE.
  - PCA, ICA, SOM, ART.
- **Reinforcement Learning:**
  - **Q-learning. Policy and Value function approximation.**

Where Does ML Fit In?



Approaches to AI

- **GOFAI:** Good Old Fashioned AI
(McCarthy, 1955; Haugeland, 1985).

- **New (Biologically Inspired) AI** (Brooks, 1989).

GOFAI: Central Hypothesis

- Knowledge can be represented by symbols and intelligence is reducible to symbol manipulation (Allen and Simon, 1963).
- AI is achieved by manipulation of symbols.
- Dominated AI paradigm until the late 20th century.
- **Philosophical Roots:**
  - Gottfried Leibniz (1646 – 1716): Attempted to create a logical calculus of all human ideas.
  - David Hume (1711 – 1776): Perception is reducible to "atomic impressions“.
  - Immanuel Kant (1724 – 1804): Experience is controlled by formal rules.

GOFAI and Symbols

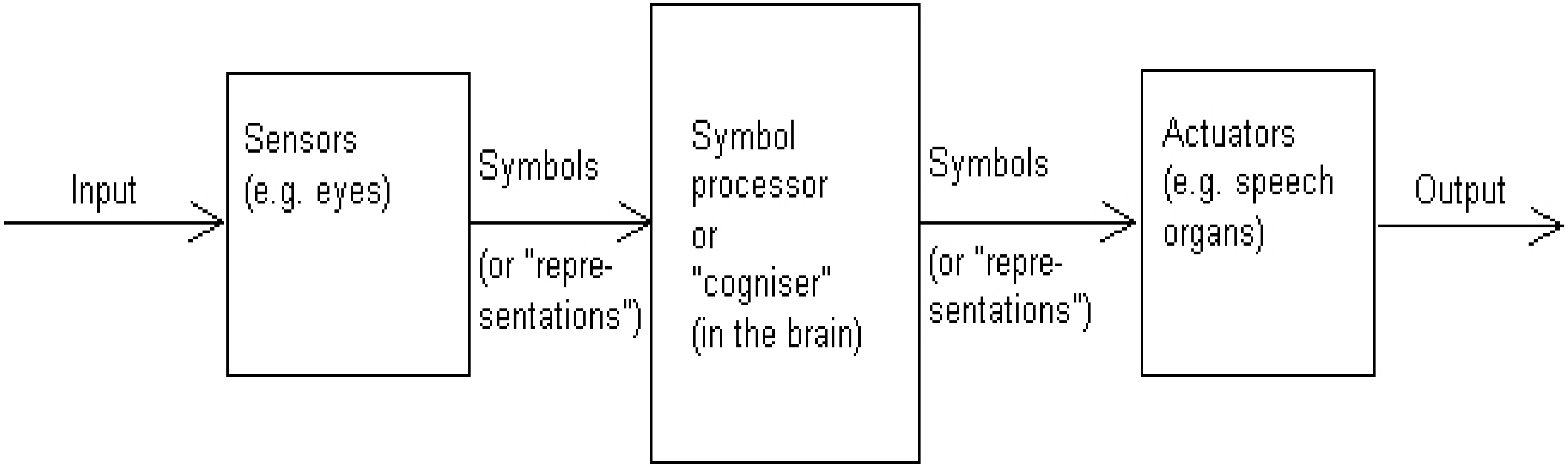
- **Formal Logic:**
  - **Symbols:** AND, OR, NOT, A, B…
  - **Expressions:** TRUE or FALSE statements.
  - **Process:** Rules of logical deduction.

- **Chess:**
  - **Symbols:** The pieces.
  - **Expressions:** All possible board configurations.
  - **Processes:** The legal chess moves.

GOFAI and Symbols

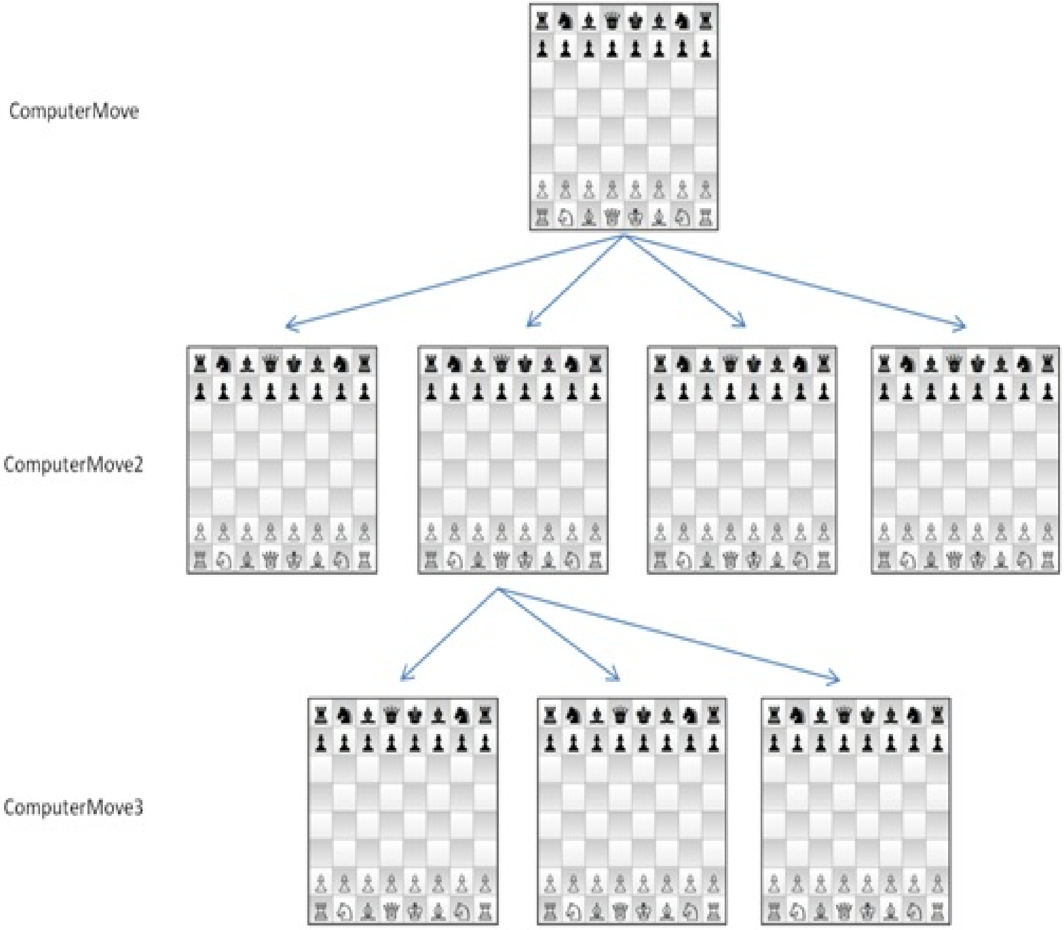
- **Human Thought:**
  - **Symbols:** Encoded in our brains.
  - **Expressions:** Our thoughts.
  - **Processes:** The mental operations of thinking.
- **AI "Thought":**
  - **Symbols:** Data structures.
  - **Expressions:** Sets of data structures.
  - **Processes:** Programs that manipulate the data structures.

GOFAI → Thinking Machine?



GOFAI Approaches

- **Top-Down Approach:** Hierarchical symbolic based algorithm.



- **GOFAI algorithms?**

Some GOFAI Algorithms

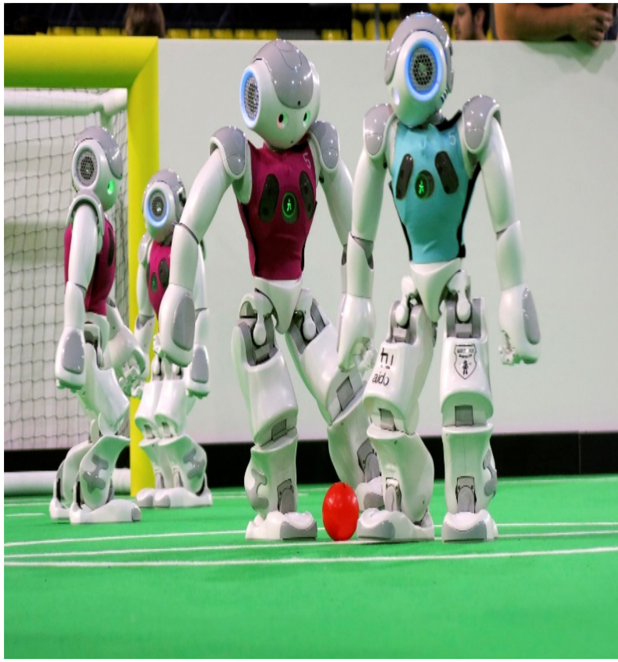
- Finite State Machines (FSMs).
- *MINIMAX, Alpha-Beta Pruning*.
- *Monte-Carlo Search*.
- Rule (Knowledge) Based Systems / Bayesian inference (Expert Systems).
- ***Concept Learning.***

New (Biologically Inspired) AI

- **Bottom-Up (Synthetic) Approach:** Individual components interact (self-organise) to produce global (system-level) behaviour.

- **New AI methods:** How should simple components interact to produce “intelligent” behaviour?





Some New AI Methods

- Evolutionary Algorithms (EAs).
- ***Artificial Neural Networks (ANNs)***
- ***Reinforcement Learning (RL)***
- Particle Swarm Optimisation (PSO).

“Weak” AI?

- We can build machines that act as if they were intelligent.
- Most AI research is in this area (and successful for many applications).
- Constrained problem sets / domains.
- Specific techniques to “simulate” intelligent decisions/actions.
- Does not try to solve the problem of general intelligence.
- All AI applications today are “weak” AI.


Siri


KASPAROV
DEEP BLUE


nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
NATURE.COM/NATURE


THINK
DENKE
$300,000
$1,000,000
$200,000
KEN
WATSON
BRAD

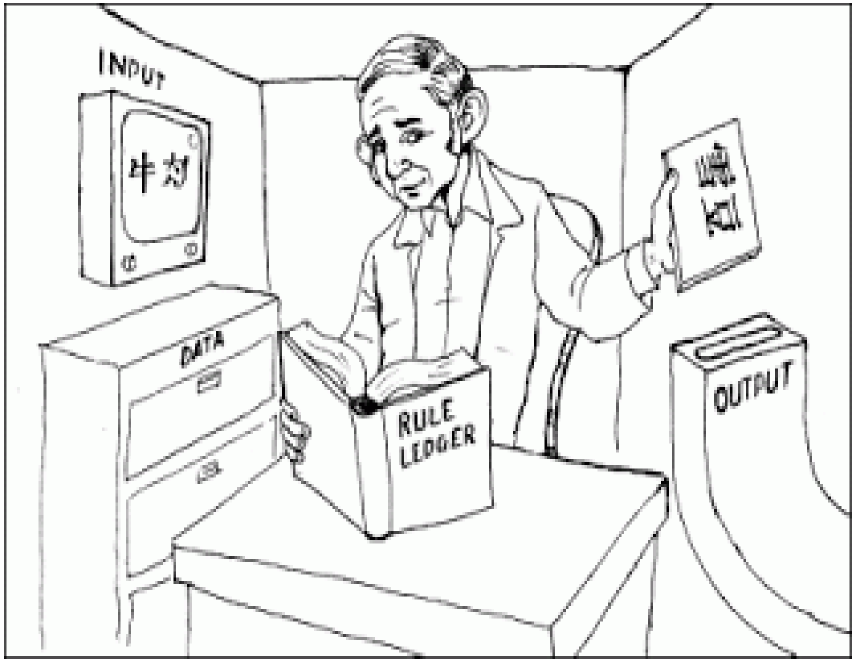

Powered by Diesel
03
Red Bull
MDV
MOHR DAVIDOW VENTURES
android

“Strong” AI?

- The goal is to build machines that are actually thinking “like people” (as opposed to just simulating thinking)
- The Chinese room.
- Examples … ?

Chinese Room Argument (Searle, 1980)

- Person in room speaks English but not Mandarin.
- Receives notes in Mandarin.
- Has English ***rule-book*** for how to write new Chinese characters given input Chinese characters – returns notes.



- Person = CPU, Rule-book = AI Program, Notes = Data
- From outside observer's point of view, the room appears to speak perfect Mandarin!

The Learning Problem

- *Machine Learning* (ML) is programming computers to optimise a performance criterion using example data or past experience.

- Learning is used when:

  - Human expertise does not exist (e.g.: Robots navigating on Mars).

  - Humans are unable to explain their expertise (e.g.: Speech and facial recognition).

  - Solution changes in time (e.g.: Routing on a computer network).

  - Solution needs to be adapted to particular cases (e.g.: Biometrics).

Types of Data

- **Discrete:** One of a finite number of values (e.g.: Address).
- **Continuous:** Within a range (e.g.: Salary).
- **Ordinal:** Ranking for numerical value (e.g.: Age).
- **Relational** (e.g.: Employee records).
- **Independent identical distributed vectors** (e.g. Employee X record).
- **Time series dependent vectors** (e.g. Financial indicators for time t, related to $t - 1$).
- **Images** (Matrices).
- **Variable-size non-vector data** (e.g. Strings, trees, graphs, text).
- **Objects** (e.g. Within a relational schema).

What is Machine Learning (ML)?

- Data is cheap and abundant – Knowledge is expensive and scarce.

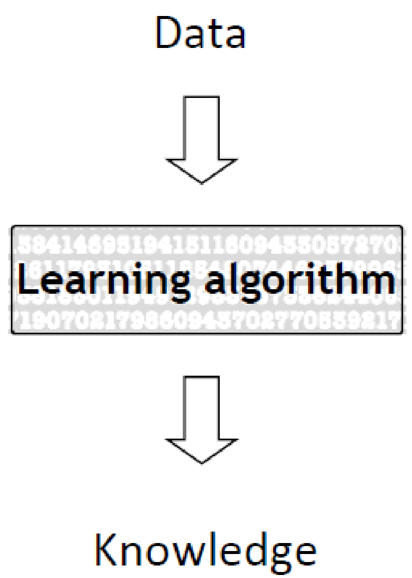
- Build model that is ***a good and useful approximation*** to the data:
  - Learn general models from data of particular examples.

- ML: Design and Analysis of algorithms that improve their performance at some task with experience (Mitchell, 1997).

What is ML ?

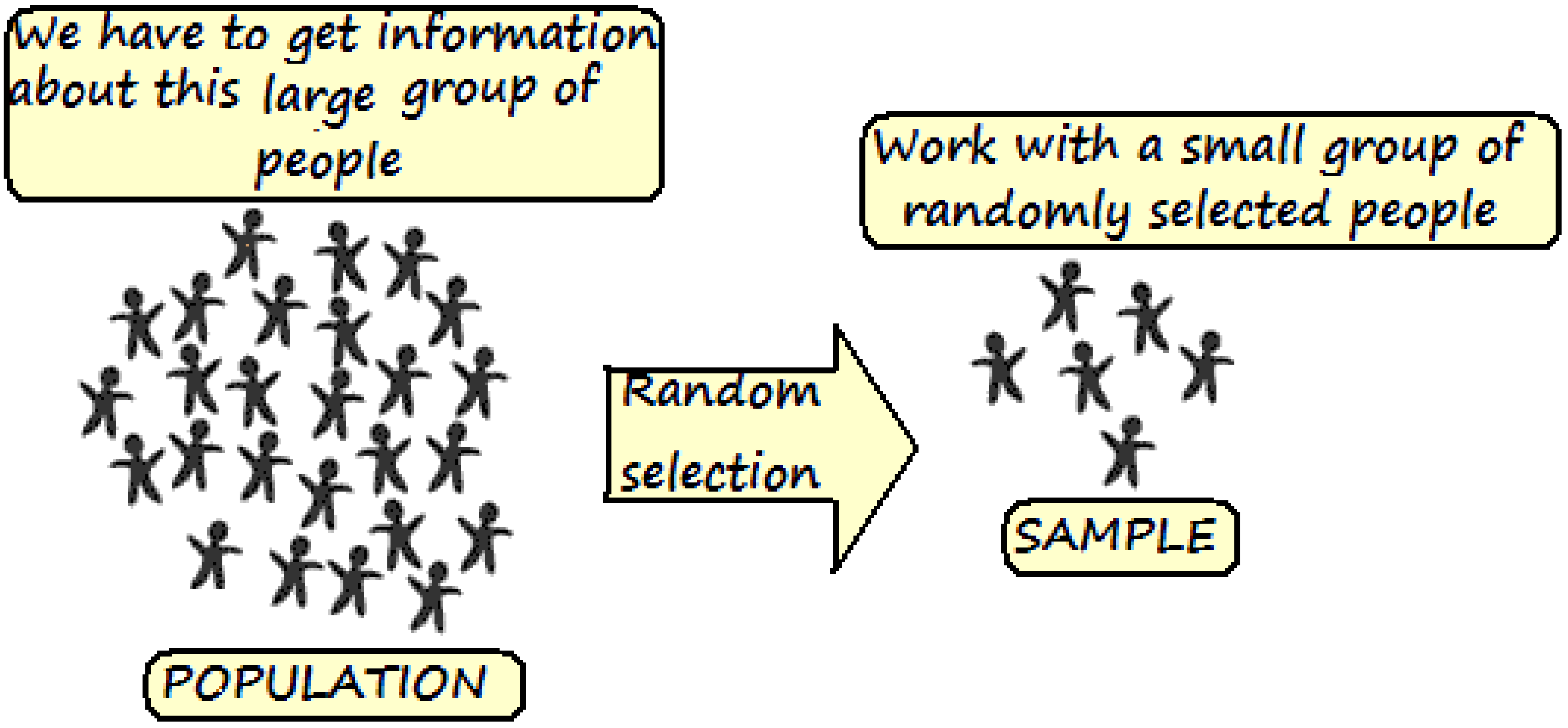
- Optimise performance criteria using example data (past experience).

- **Statistics:** Inference from a sample.



- **Main Goal of Learning:** *Prediction*
  - Obtain a model of some *training data*, through a *learning* process.
  - Use that model to *predict* something about data not seen before.
  - Learn the same distribution as training data using *test data*.

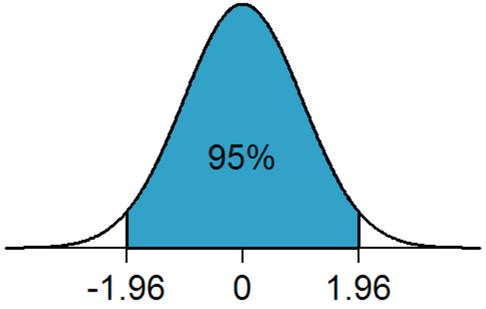
Inferential Statistics



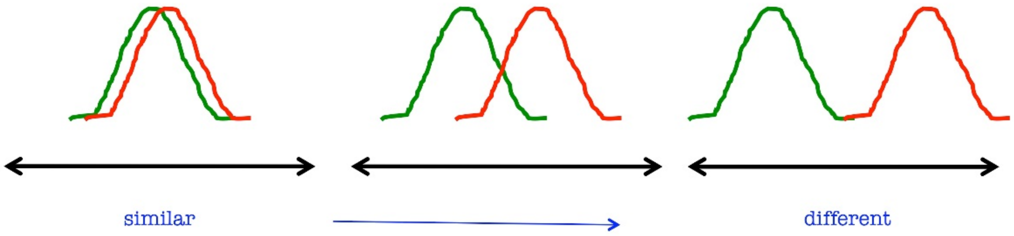
What is ML?

- *ML* ≈ Inferential + Multivariate + Computational statistics:
  - ***Inferential statistics* ≈** Inference from data sample.
  - ***Multivariate statistics* ≈** Prediction of values of a function assumed to describe a multivariate dataset.
  - ***Computational statistics* ≈** Computational methods for statistical problems.

What is ML ?



- Main types of inference problems:
  - Point estimation.
  - Confidence sets.
  - Hypothesis testing.



- ML is mostly about point estimation:
  - A statistic (best guess) derived from sample data.
- Data mining: “Knowledge extraction” !

Types of Learning

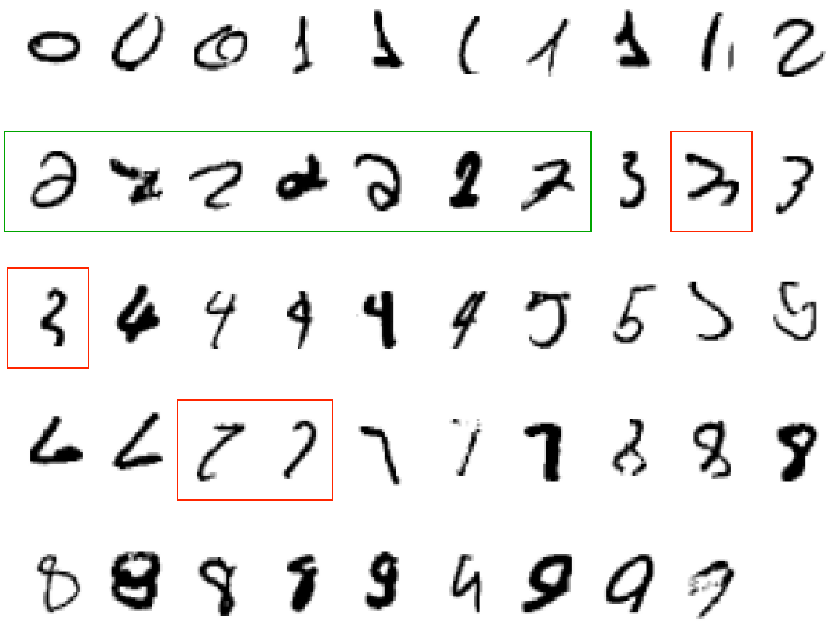
- **Supervised Learning**:
  - Classification / Regression.
- **Unsupervised Learning:**
  - Clustering / Dimensionality reduction.
- **Reinforcement Learning:**
  - Value and policy iteration / Q Learning.

Supervised Learning

- Predicting a target variable for which we get to see examples.
  - **Classification:** Predict a discrete target variable.
  - **Regression:** Predict a continuous target variable.
- **Prediction of future cases:**
  - Find a rule that predicts output for future inputs.
- **Knowledge extraction:**
  - Finding a rule that is easy to understand.
- **Compression:**
  - Finding rule that is simpler than the data it explains.

Supervised Learning

- Useful when humans cannot define a decision rule, but can perform the *classification* task – for example:



Supervised Learning: Classification

- Given: Set of labeled examples (***training data***), each described by a set of attributes, and labeled with a class:
  - Find a model for the class attribute as a function of the values of other attributes.

- Goal: Classify previously unseen examples (***test data***) accurately.

Feature Space $\mathcal{X}$

Label Space $\mathcal{Y}$

"Sports"
"News"
"Science"
...

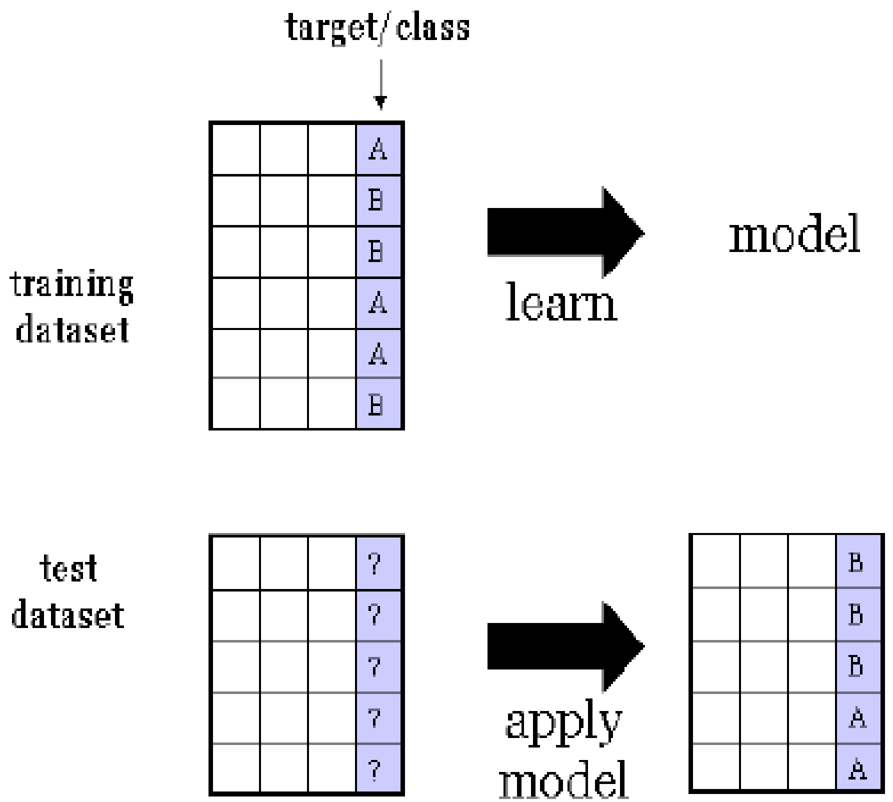
Discrete Labels
Classification

Supervised Learning: Classification

- Facial Recognition: Predicting a discrete target variable:

Training examples

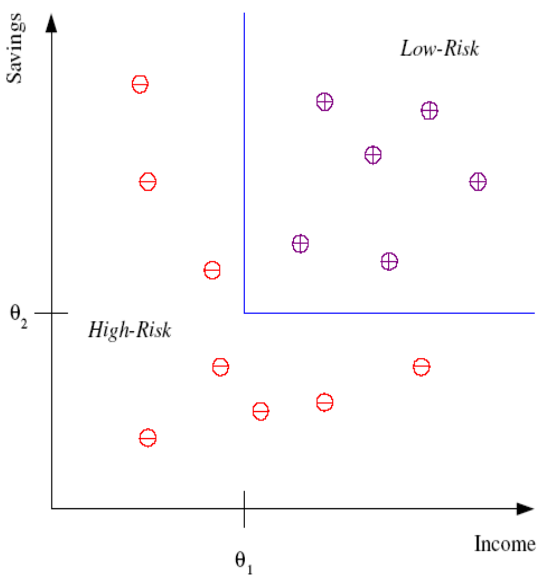
Test images



Classification Examples

- **Fraud Detection:**
  - Predict fraudulent cases in credit card transactions.
  - *Training data:* Previous transactions of given account holder.
  - *Attributes:* Time of purchase, product type, cost, location,…
  - *Class:* Label transactions as fraudulent or fair.

- **Credit Scoring:**
  - Differentiate between *low-risk* and *high-risk* credit applications.
  - *Training data: Incoming and savings*
  - *Attributes:* Income, expenses, debts,…
  - *Class*: Label application as good or bad.

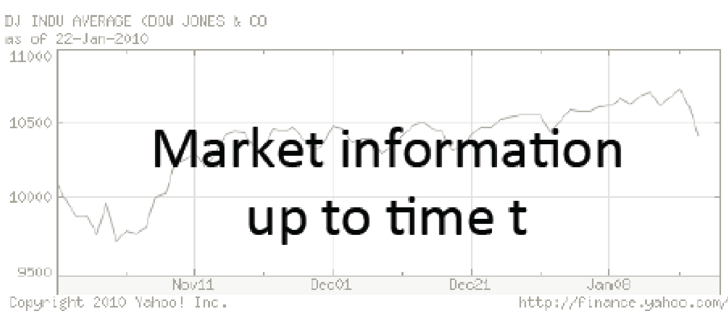


Supervised Learning: Regression

- Predict a continuous valued variable based on attribute values.
  - e.g.: Stock price prediction
  - Predict the value based on a combination of the last *k* values (linear regression):

Feature Space $\mathcal{X}$

Label Space $\mathcal{Y}$



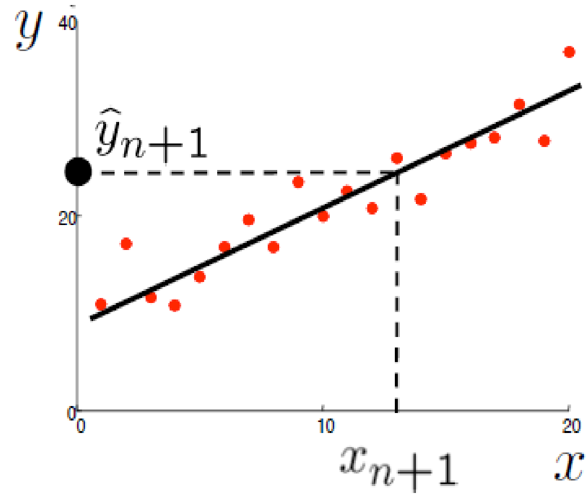
Share Price
"$ 24.50"

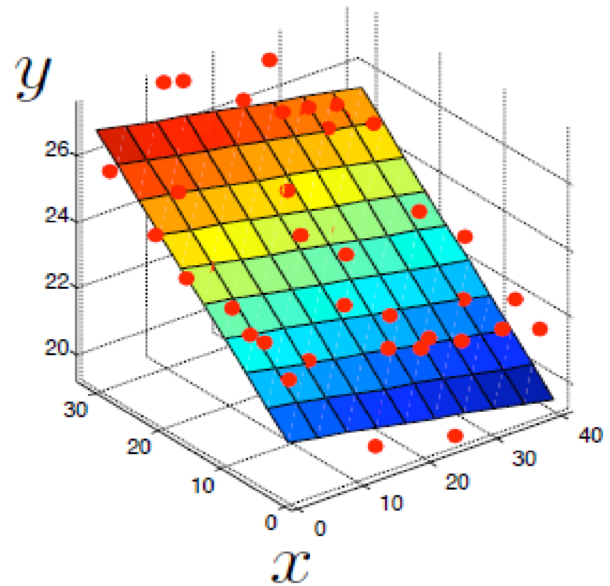
Continuous Labels
Regression

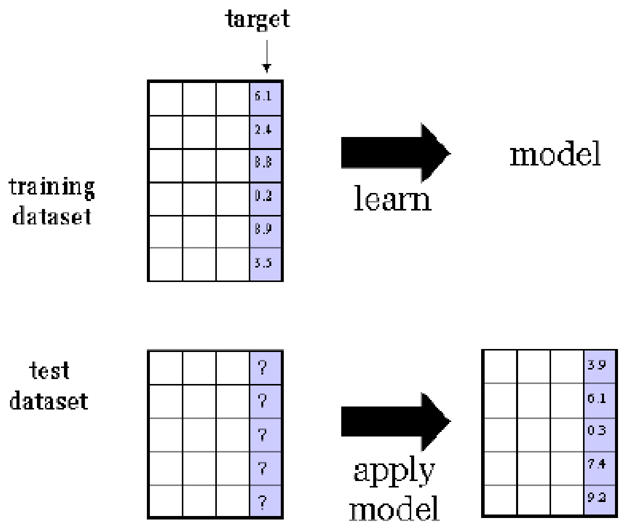
Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

Regression Example

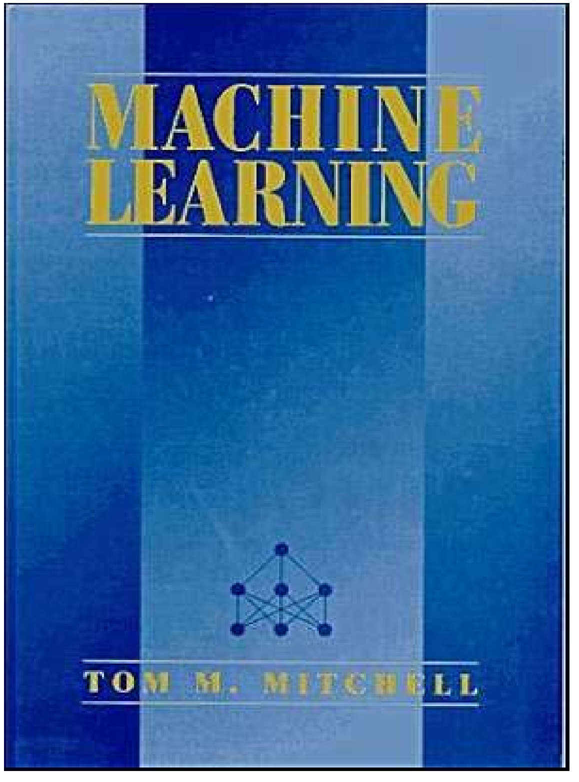
- Predicting a continuous target variable: Price of a car:
  - x : car attributes.
  - y : price.
  - $y = g(x \mid \theta)$.
  - $g(\,)$ model.
  - θ parameters.



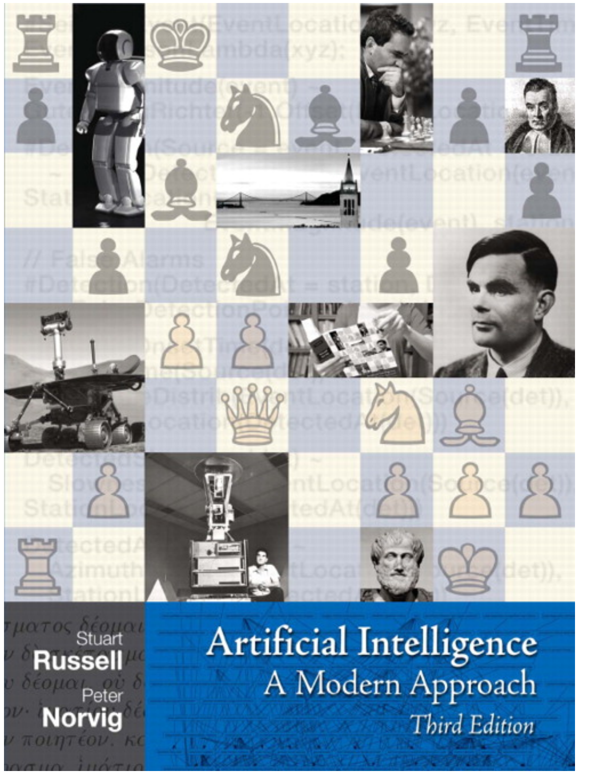




Reading



Chapter 14 – Key Ideas in Machine Learning



Chapter 1 – Introduction