Chapter 1

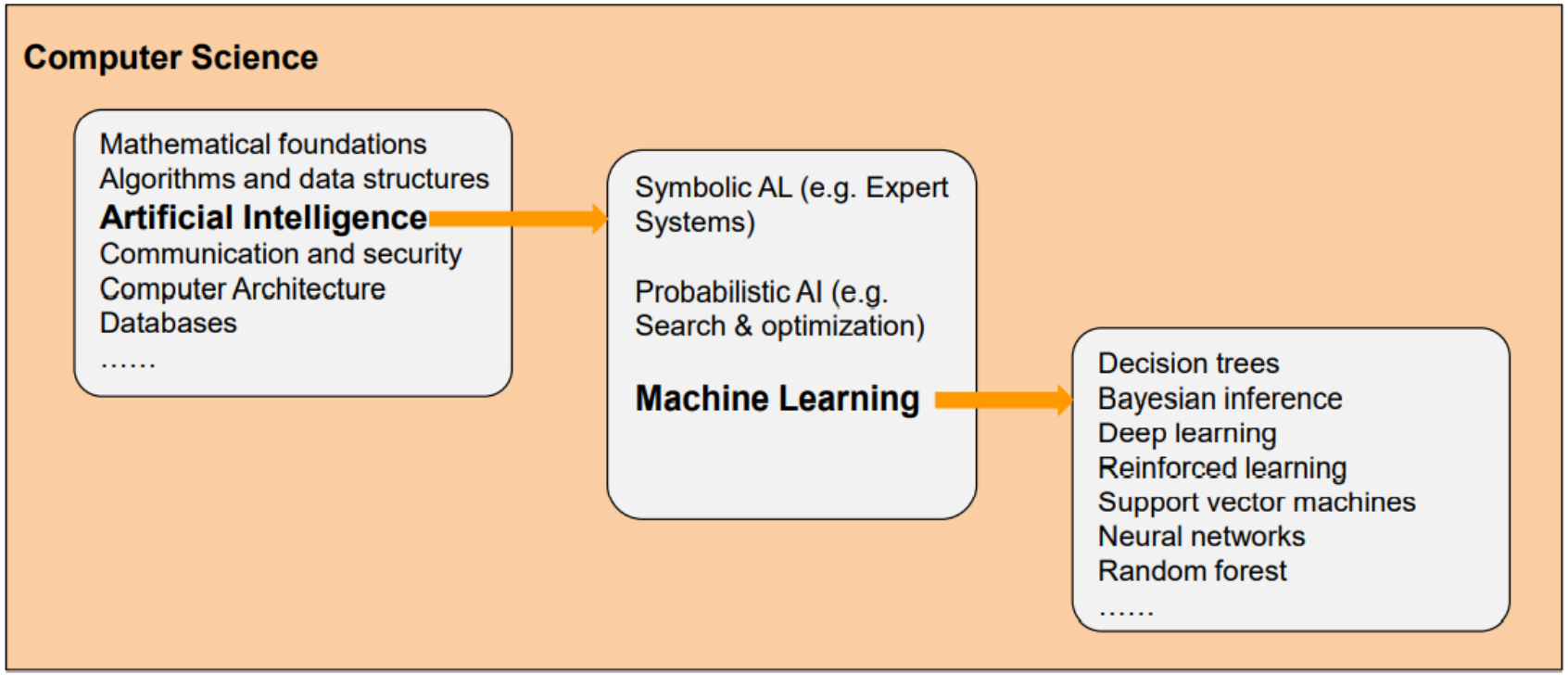
Introduction to Machine Learning

Contents: Introduction to Machine Learning

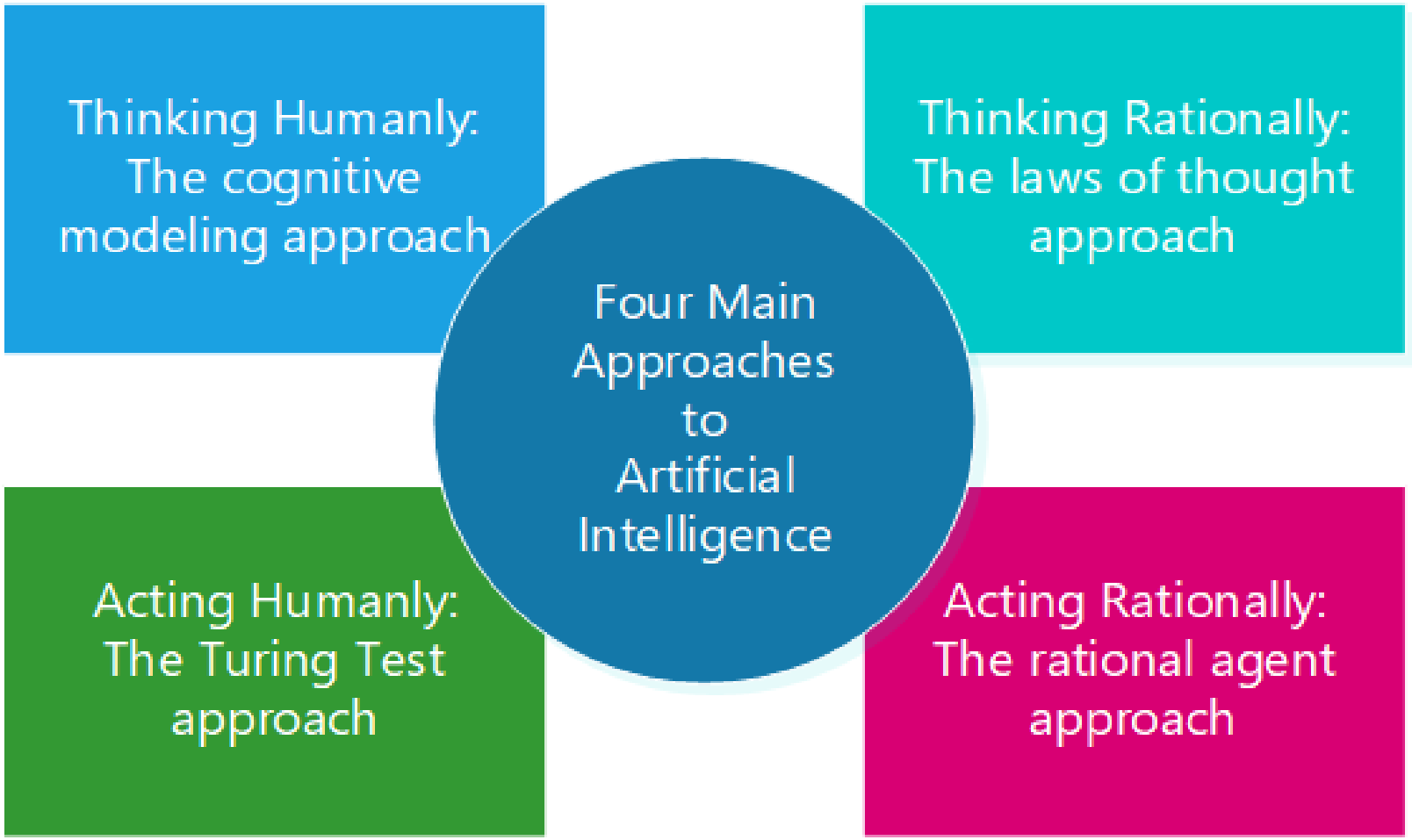
- Artificial Intelligence and Machine Learning
- What is Machine Learning
- Machine Learning problems
- Types of Machine Learning

Artificial Intelligence and Machine Learning

- Artificial Intelligence (AI) is a branch or Computer Science that uses algorithms and techniques to mimic human intelligence
- Machine Learning (ML) is one of several AI techniques for sophisticated cognitive task



The four approached of AI



- Machine Learning is a particularly interesting technique because it represents a paradigm shift within AI

Traditional AI techniques

- **Static** – hard-coded set of steps and scenarios
- **Rule Based** – expert knowledge
- **No generalization** – handling special cases is difficult

---

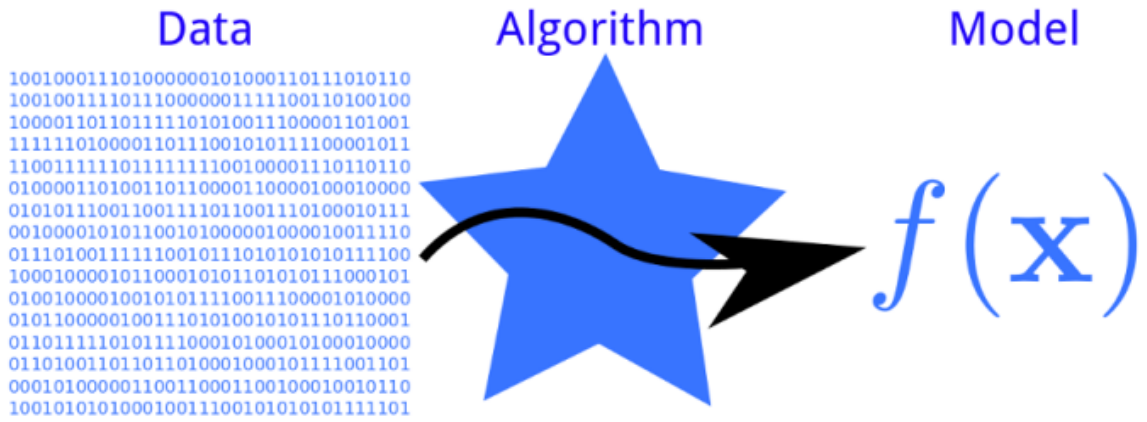
Machine Learning

- **Dynamic** – evolves with data, finds new patterns
- **Data driven** – discovers knowledge
- **Generalization** – adapts to new situations and special cases

Example: how other AI techniques vs. machine learning try to excel at playing the game of chess

Symbolic AI	Mathematical/Statistical AI	Machine Learning Approach
"Let us sit down with the world's best chess player, Mr. X, and put his knowledge into a computer program"	"Let us simulate all the different possible moves and the associated outcomes at each single step and go with the most likely to win"	**"Let us show millions of examples or real life and simulated games (won and lost) to the program, and let it learn from experience"**

What is Machine Learning?



"Machine Learning is the science of getting computers to act without being explicitly programmed." – Andrew Ng (Coursera)

"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 task in T, as measured by P, improves with experience E." – Tom M. Mitchell (1997)

Cont….

- Machine Learning (ML) is one of several AI techniques for sophisticated cognitive tasks
- Machine Learning is a particularly interesting technique because it represents a paradigm shift within AI
- When to used machine learning? ML is particularly good at solving 2 types of problems where other AI techniques fail
  - Tasks programmers can’t describe
  - Complex multidimensional problems that can’t be solved by numerical reasoning

Machine Learning problems

- Machine learning is particularly good at solving **2 types of problems** where other AI techniques fail

Tasks programmers can't describe

Complex multidimensional problems that can't be solved by numerical reasoning

Hand writing

Weather Forecasting

Health Care Outcomes

Cognitive Reasoning

Network Intrusion

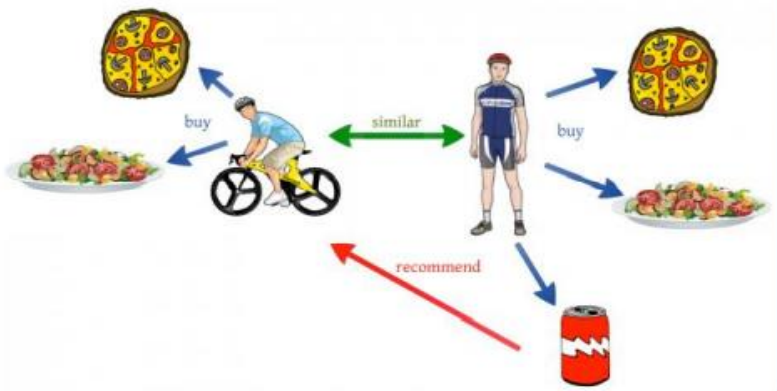
Movie Recommendation

Cont….

- Recommender systems deal with making recommendations based upon previously collected data and leveraging ML techniques.

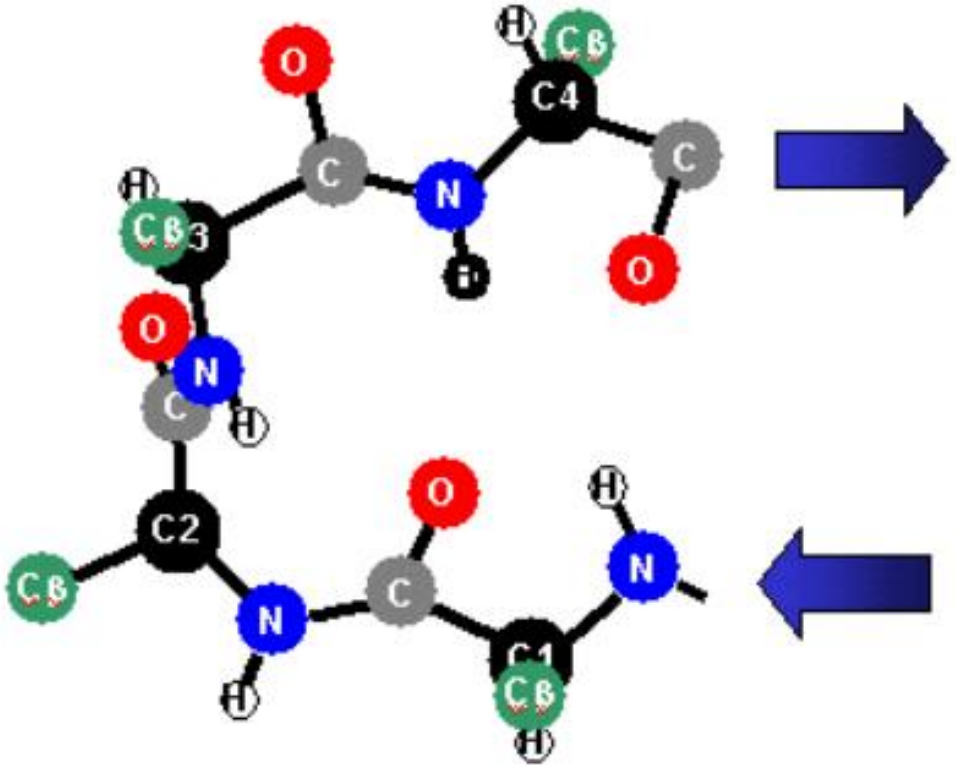
Content Based (Features)
Modified Linear Regression

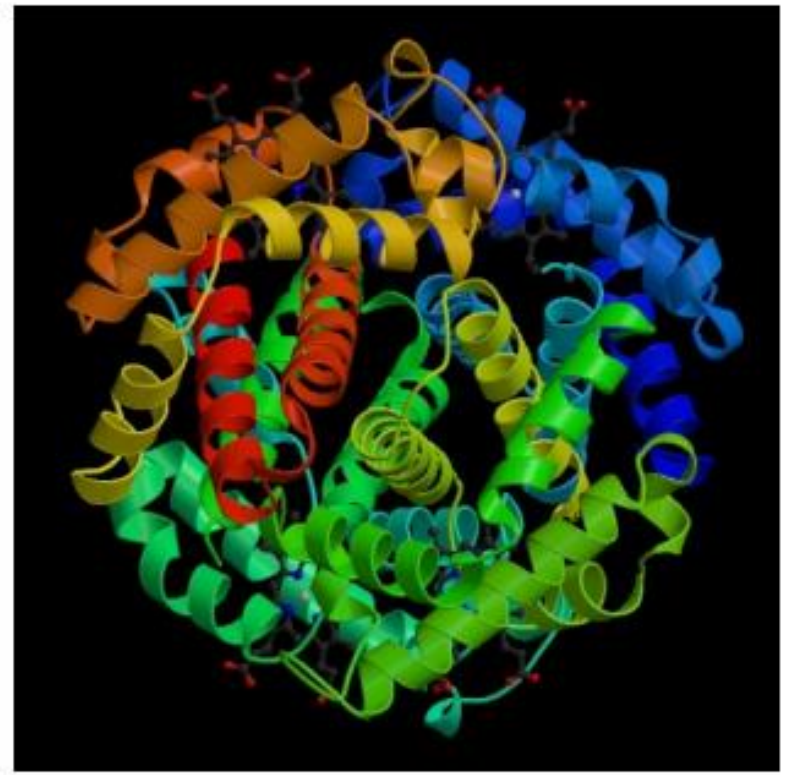
Non-content Based (No Features)
Collaborative Filtering
Matrix Factorization



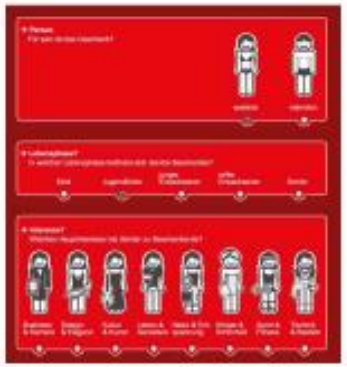
- ML processes are being used to train robots and improve accuracy. It provides robots with adequate computer vision and motion control to better understand the environment and act accordingly.

Bioinformatics: predict the 3d structure of a molecule based on its sequence.



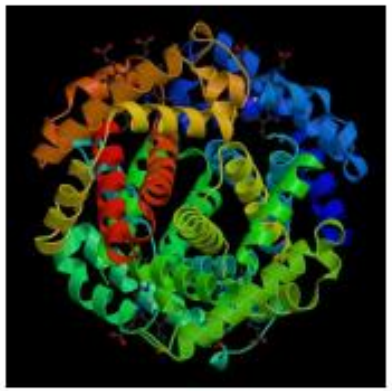


Information Systems



Robotics

Bioinformatics



Many Further Applications!

MACHINE LEARNING

OPTIMIZATION

NUMERICS

One area of research, many names (and aspects)

machine learning
: historically, stresses learning logical or rule-based models (vs. probabilistic models).

data mining, **big data**
: stresses the aspect of large datasets and complicated tasks.

knowledge discovery in databases (KDD)
: stresses the embedding of machine learning tasks in applications, i.e., preprocessing & deployment.

data analysis historically, stresses multivariate regression and unsupervised tasks.

pattern recognition
: name preferred by engineers, stresses cognitive applications such as image and speech analysis.

data science, **applied statistics**
: stresses underlying statistical models, testing and methodical rigor.

predictive analytics, **business analytics**, **data analytics**
: stresses business applications.

Types of Machine Learning

- Machine Learning is generally categorized into three types: Supervised Learning, Unsupervised Learning, Reinforcement learning
- **Supervised Learning:**
  - In supervised learning the machine experiences the examples along with the labels or targets for each example. The labels in the data help the algorithm to correlate the features.
  - Two of the most common supervised machine learning tasks are **classification** and **regression**.

Cont....

- In **classification** problems the machine must learn to predict **discrete/categorical** values. That is, the machine must predict the most probable category, class, or label for new examples. Applications of classification include predicting whether a stock's price will rise or fall.
- In **regression** problems the machine must predict the value of a **continuous** response variable. Examples of regression problems include predicting the sales for a new product.

- **Unsupervised Learning:**
  - When we have unclassified and unlabeled data, the system attempts to uncover patterns from the data . There is no label or target given for the examples. One common task is to group similar examples together called **clustering**.

Cont....

- **Reinforcement Learning:**
  - Reinforcement learning refers to goal-oriented algorithms, which learn how to attain a complex objective (goal) or maximize along a particular dimension over many steps.
  - This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance.
  - Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal. For example, maximize the points won in a game over many moves.
  - A class of learning problems where the correct / optimal action never is shown, but only positive or negative feedback for an action actually taken is given. It is usually applied in computer game developments.

Supervised Learning:

Predicting values. **Known** targets.
User inputs correct answers to learn from. Machine uses the information to guess new answers.

REGRESSION:
Estimate continuous values
(Real-valued output)

CLASSIFICATION:
Identify a unique class
(Discrete values, Boolean, Categories)

Unsupervised Learning:

Search for structure in data. Unknown targets.
User inputs data with undefined answers. Machine finds useful information hidden in data

CLUSTER ANALYSIS:
Group into sets

DENSITY ESTIMATION:
Approximate distribution

DENSITY REDUCTION:
Select relevant variables

Supervised Learning:

Regression

- Linear Regression
- Ordinary Least Squares Regression
- LOESS (Local Regression)
- Neural Networks

Classification

- Decision Trees
- K-Nearest Neighbors
- Support Vector Machine
- Logistic Regression
- Naïve Bayes
- Random Forests

Unsupervised Learning:

Cluster Analysis

- K-Means Clustering
- Hierarchical Clustering

Dimension Reduction

- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)