Machine Learning

CPSC 383: Explorations in Artificial Intelligence and Machine Learning
Fall 2025

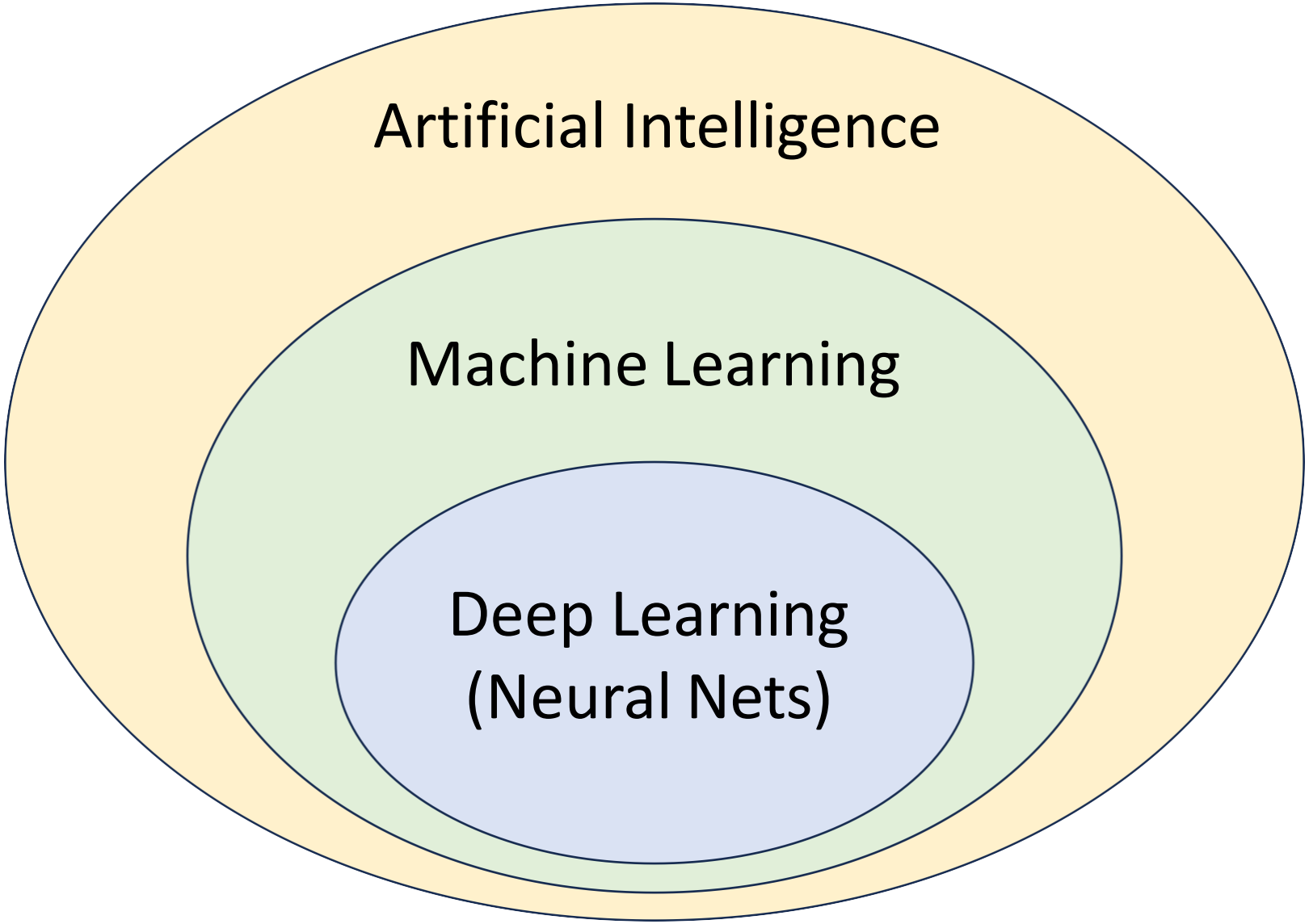
Jonathan Hudson, Ph.D
Associate Professor (Teaching)
Department of Computer Science
University of Calgary

August 27, 2025

UNIVERSITY OF CALGARY

Machine Learning

- **Machine learning** is about learning patterns from data in order to make useful predictions.
  - We do this all the time in other areas! But with different goals.
  - These days, the field is dominated by neural networks, but we'll be starting with other models to understand the basic principles.



Induction

Machine learning is built on the premise of **induction**, i.e. that data we have seen in the past can be used to predict future data.

If you measured the temperature of the room every day for the past 10 years, can you use this to make a good guess at the temperature of the room tomorrow?

Human element

Even the best ML approaches require a “smart human” to set them up.

"A neural network is the second best way to solve any problem. The best way is to actually understand the problem.“

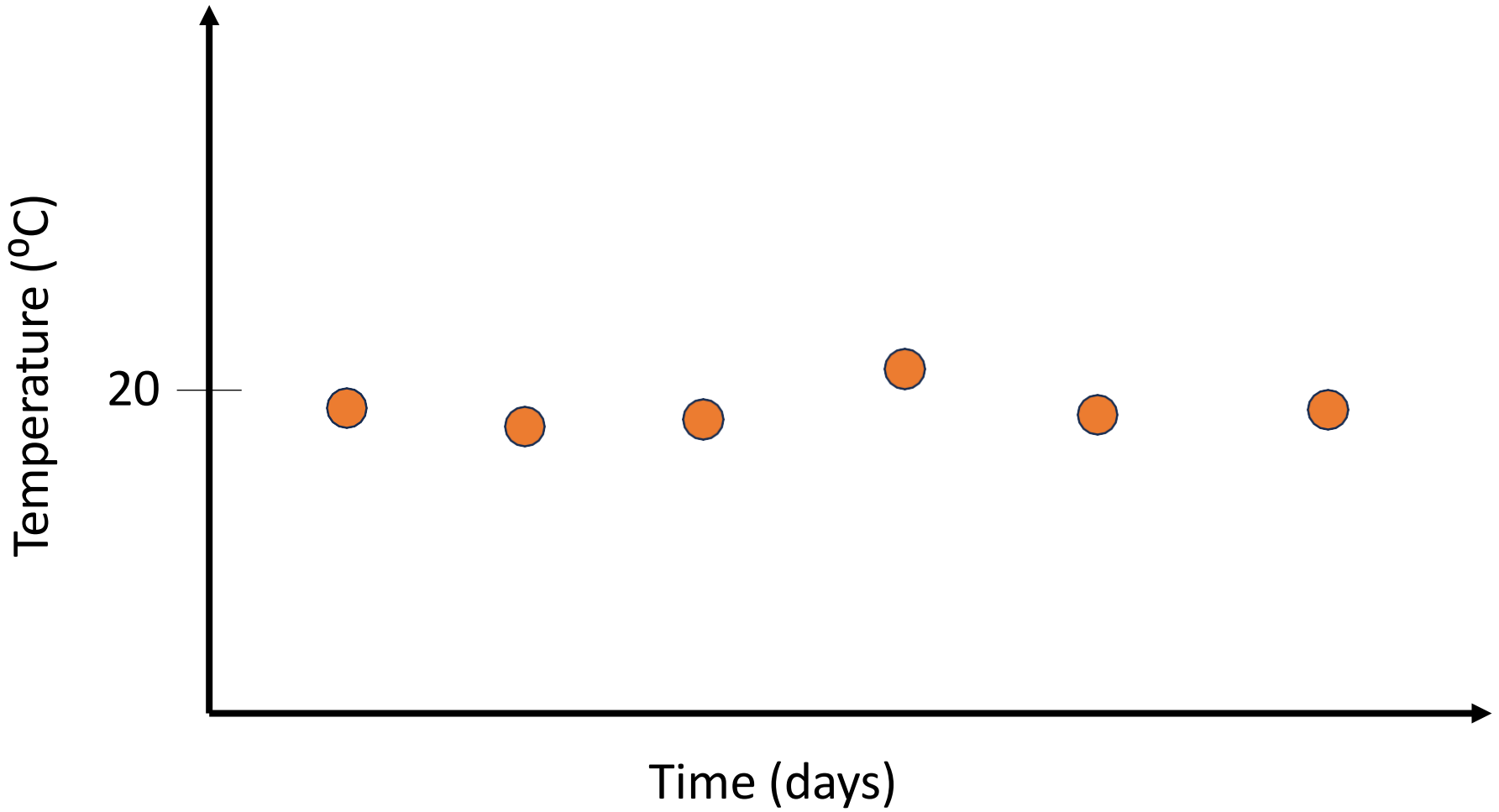
- John Denker, 1994

Human element

A human needs to choose:

- Approach
- Assumptions
- Model (type of solution you would expect to see)
- Data source, encoding
- Training algorithm
- How to evaluate the model’s performance

Example



Categories

There are three main types of machine learning:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

… as well as many others, and often they are used in combination.

Compare Supervised vs Unsupervised

- Supervised Learning requires labeled data!
  - Takes time
  - Takes resources
  - Costs money!
- ImageNet
  - Dr. Fei-Fei Li, using Princeton students
  - But even after heavily optimizing the labeling process—for example, pre-downloading candidate images so they're instantly available for students to review—Li and her graduate student Jia Deng calculated that it would take more than 18 years to select and label millions of images.
  - Used Amazon Mechanical Turk to do labelling -> loss of control of data quality

Compare Supervised vs Unsupervised

- For Unsupervised Learning you just need the data (not the label)
  - Find patterns without prior knowledge
  - More about learning structure of data
- Challenges
  - Evaluation -> if not labelled then what is correct?
    - Have to infer correctness from quality of learned patterns
- Advantage
  - Very good for exploration of data to learn things about it
- Disadvantage
  - Sensitive to quality of data
    - In supervised we can often ‘assume’ a model type based on our knowledge and labelling, and this can help method ignore or adjust for bad (or outlier) data
    - For unsupervised we often don’t know what good model is yet

Review

- **In unsupervised learning**, the system is given a data set and must find some inherent patterns or structure.
  - e.g. Netflix recommendations, identifying high-value customers, detecting unusual bank transactions, topic modelling
- In **supervised learning**, the learning system is given pairs of input/output values as training data
  - e.g. (x, y), where x is an input value and y is the "correct" or "target" output
- **Reinforcement Learning**
  - In reinforcement learning, the goal is to learn a mapping of inputs to outputs that maximizes the earned "reward".

Plan

- **Unsupervised Learning**
  - Clustering
    - K-means
  - Association Rule Learning
    - Apriori algorithm
- **Supervised Learning**
  - Decision Trees
    - Random Forest -> Ensemble version of decision trees
  - Neural networks and deep learning in unit 3
- **Reinforcement Learning**
  - Multi-armed Bandit

Next…unsupervised machine learning

Jonathan Hudson, Ph.D.
jwhudson@ucalgary.ca
https://cspages.ucalgary.ca/~jwhudson/

UNIVERSITY OF CALGARY