Fundamentals of Machine Learning

Basic concepts and historical context that led to the success of deep learning

Roberto Souza
Assistant Professor
Electrical and Computer Engineering
Schulich School of Engineering

January 2022

UNIVERSITY OF CALGARY

Outline

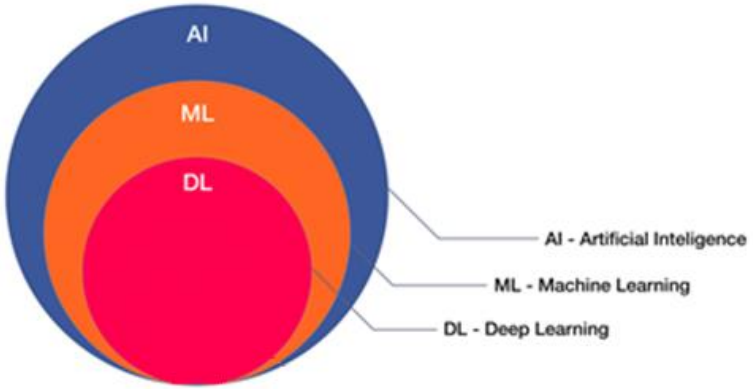
- Learning Goals
- Artificial intelligence (AI), machine learning (ML) and deep learning (DL)
  - Definitions
  - Historical context
- Fundamental ML concepts
- Summary

Learning Goals

- Explain the difference between AI, ML, and DL
- Explain the historical context that led to the success of DL
- Introduce basic ML concepts

Artificial Intelligence (AI) Machine Learning (ML) and Deep Learning (DL)

- AI: the broad discipline of creating intelligent machines
- ML: refers to systems that can learn from experience
- DL: refers to systems that learn from experience on large data sets
- Neural Networks (NN): models of human neural networks that are designed to help computers learn



What is Machine Learning?

- **Algorithms to parse data, learn from it, and make determinations or predictions about something in the world**
- Build models by training with data
- Three aspects:
  - **Data** -> engineer or learn features? how to set the experiment?
  - **Model**-> which model is best? Many times arbitrary
  - **Cost function minimization** -> set model parameters
- Concerns: interpretability, explainability (*i.e.*, black boxes), generalizability

Traditional ML

- Feature engineering
- “Simpler models” -> less parameters to be learned

$$X = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1M} \\ x_{21} & x_{22} & \dots & x_{2M} \\ x_{31} & x_{32} & \dots & x_{3M} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{NM} \end{bmatrix}}_{\text{N samples with M features}}$$

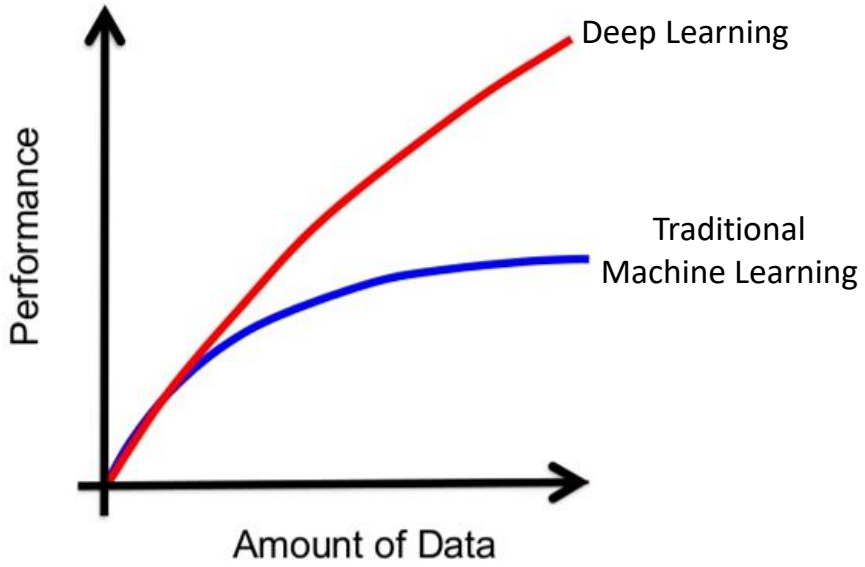
$$Y = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_N \end{bmatrix}}_{\text{Labels}}$$

6

Deep Learning (DL)

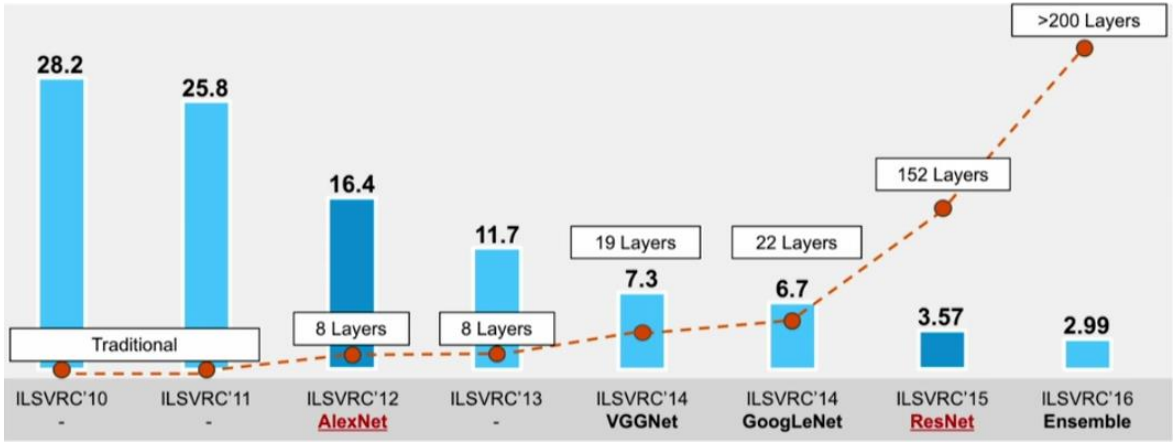
- DL is a data-driven modeling approach, which “learns the features”
  - But which features?

- Complex models with (b)millions of parameters that need to be tuned



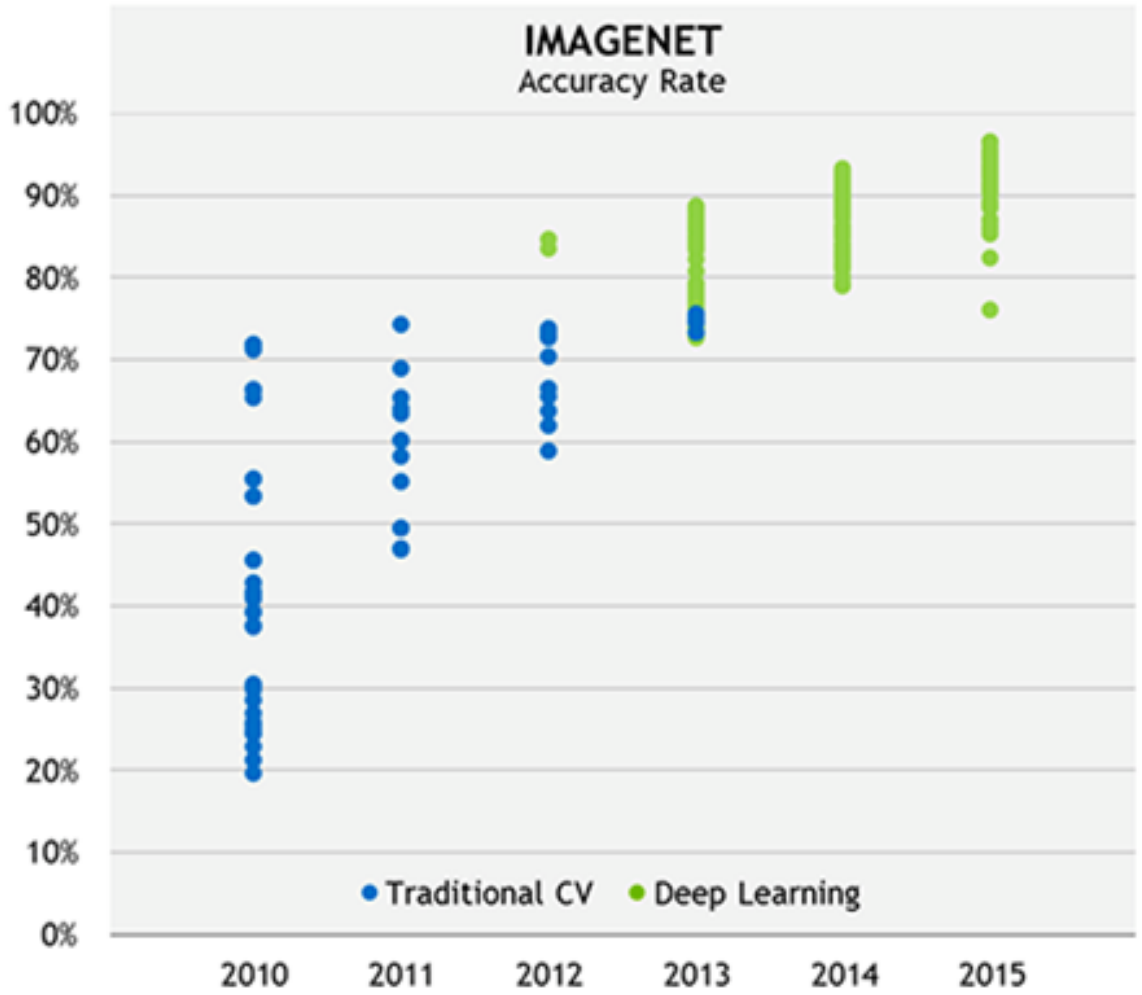
ImageNet Challenge

- ImageNet is a large scale object classification challenge
- >14,000,000 annotated images
- >20,000 classes



- In 2012 teams started using graphics processing units (GPUs)

ImageNet Challenge



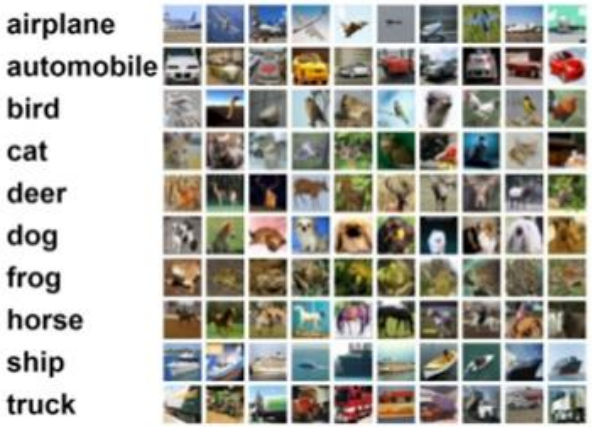
Source: https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-artificial-intelligence-gpus/

9

GPU/TPU Computing

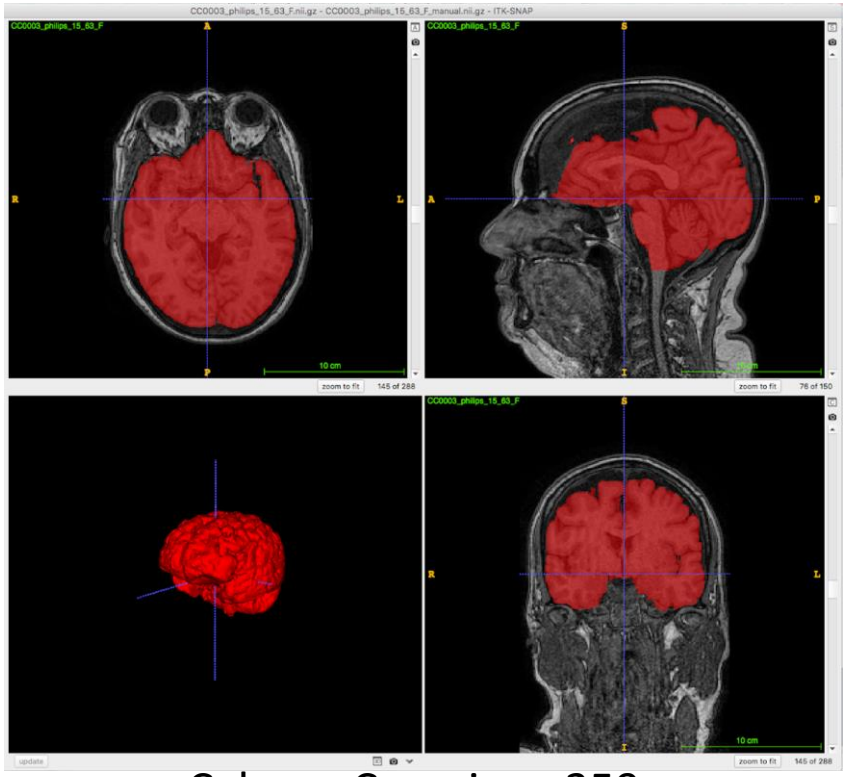
- Hardware and software improvements
- GPUs with more cores and more memory
- Optimized parallel computing platforms

Large Datasets



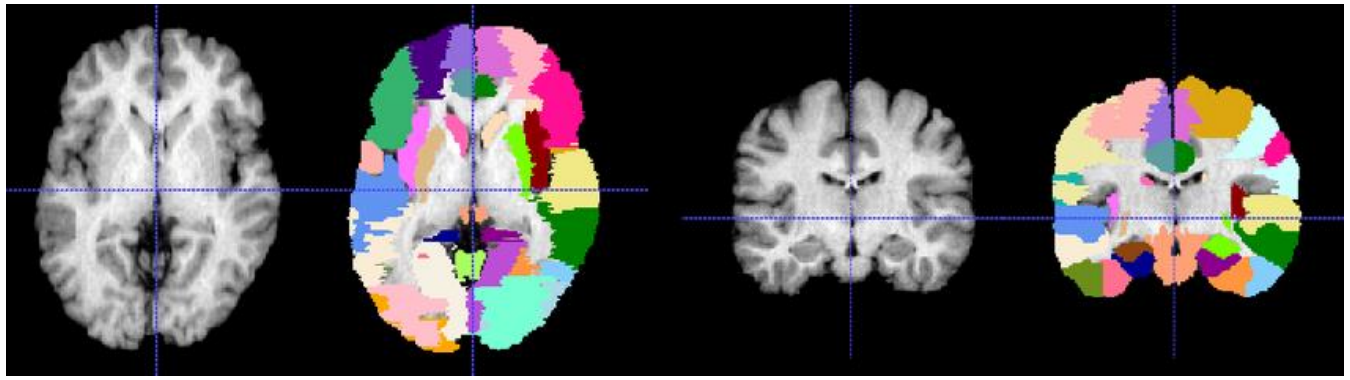
CIFAR-10

MNIST



Calgary-Campinas-359

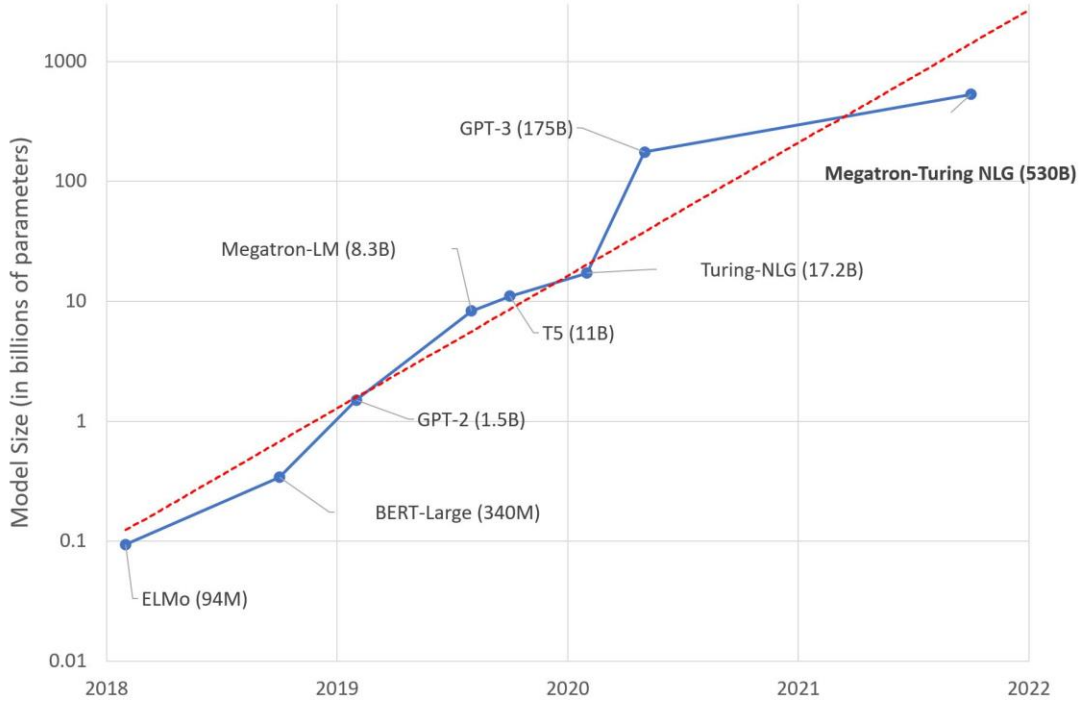
https://sites.google.com/view/calgary-campinas-dataset/home

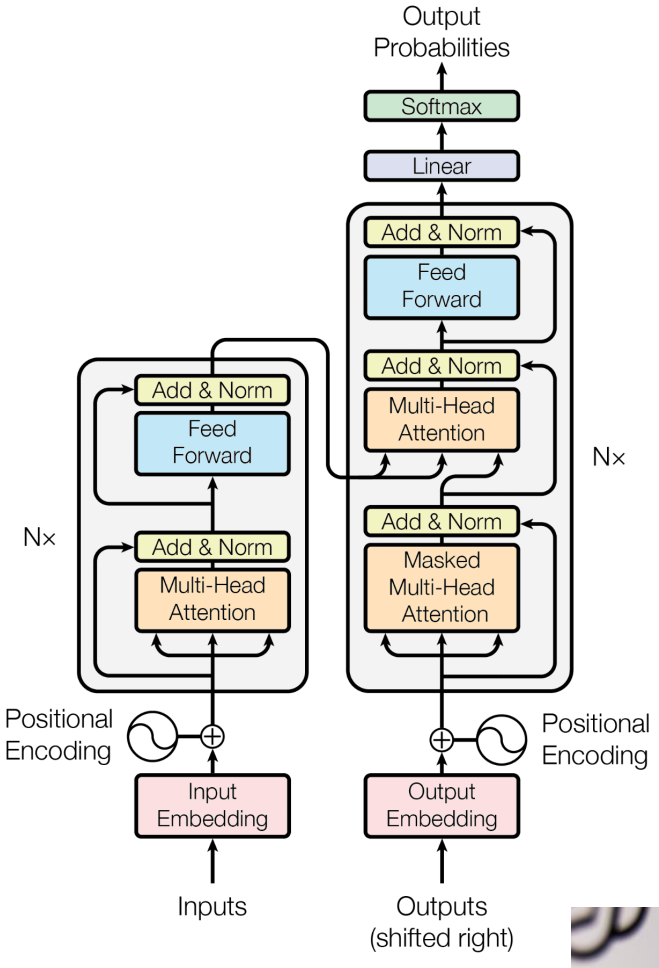


LPBA40

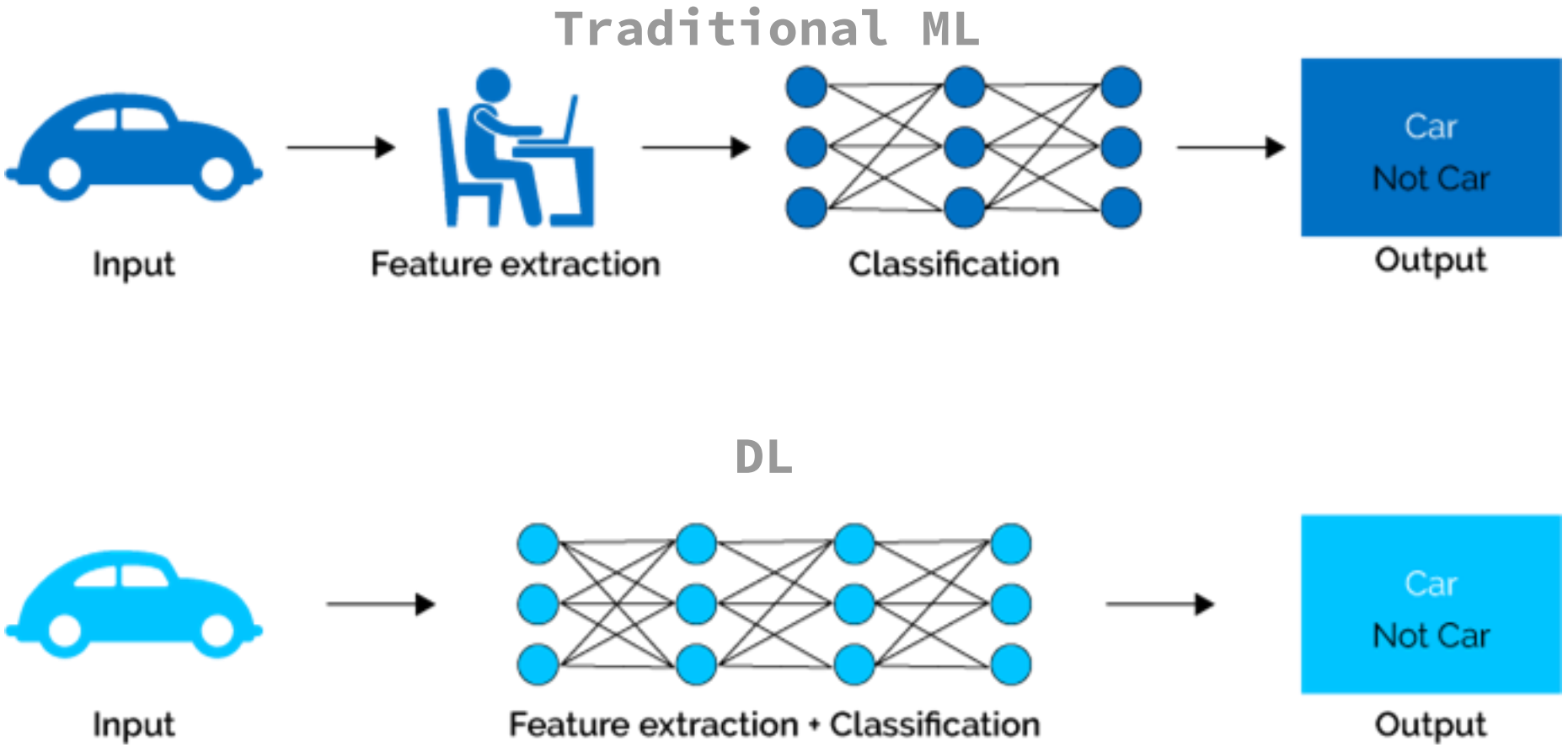
11

Large Language Models

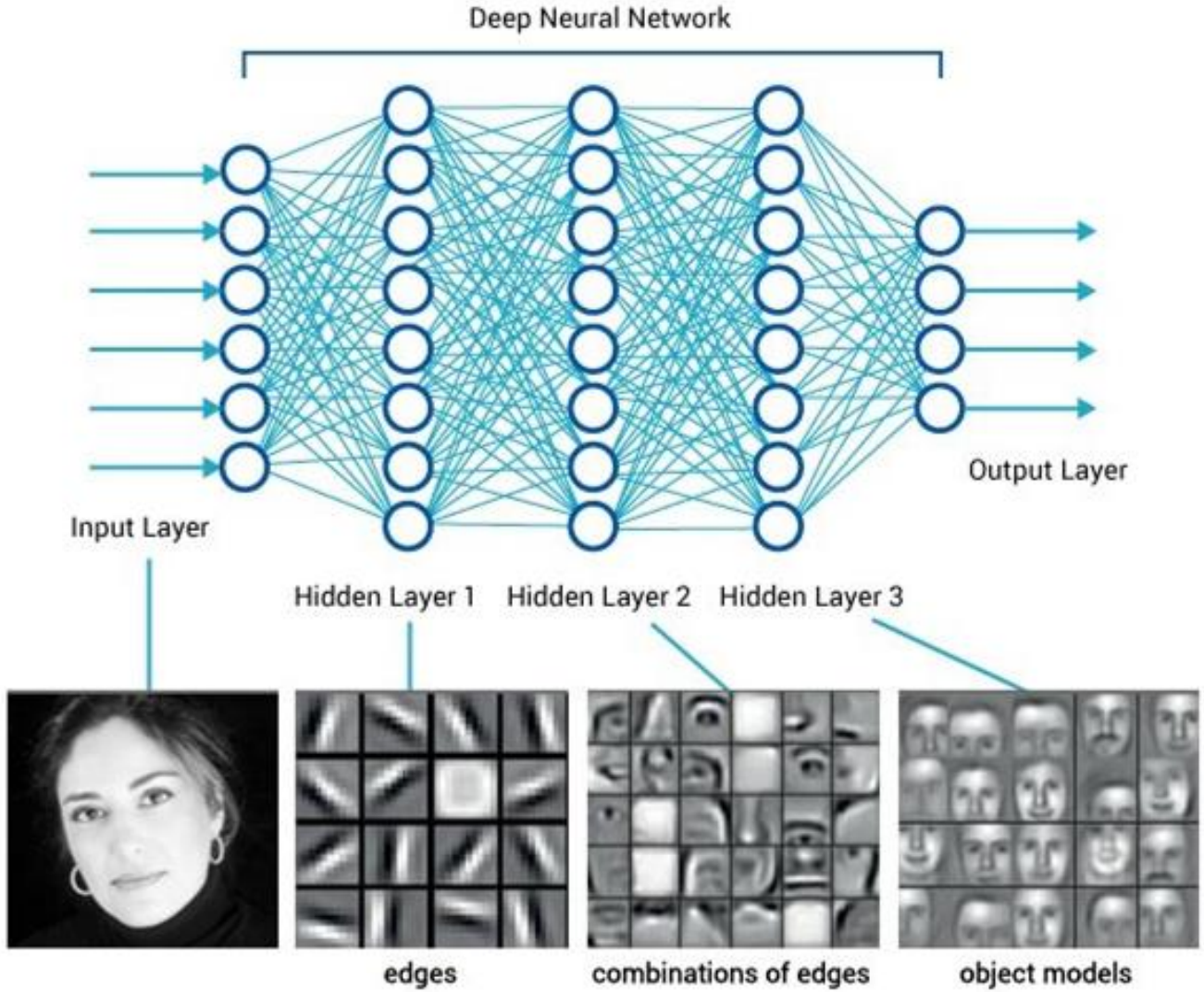




Traditional ML versus DL



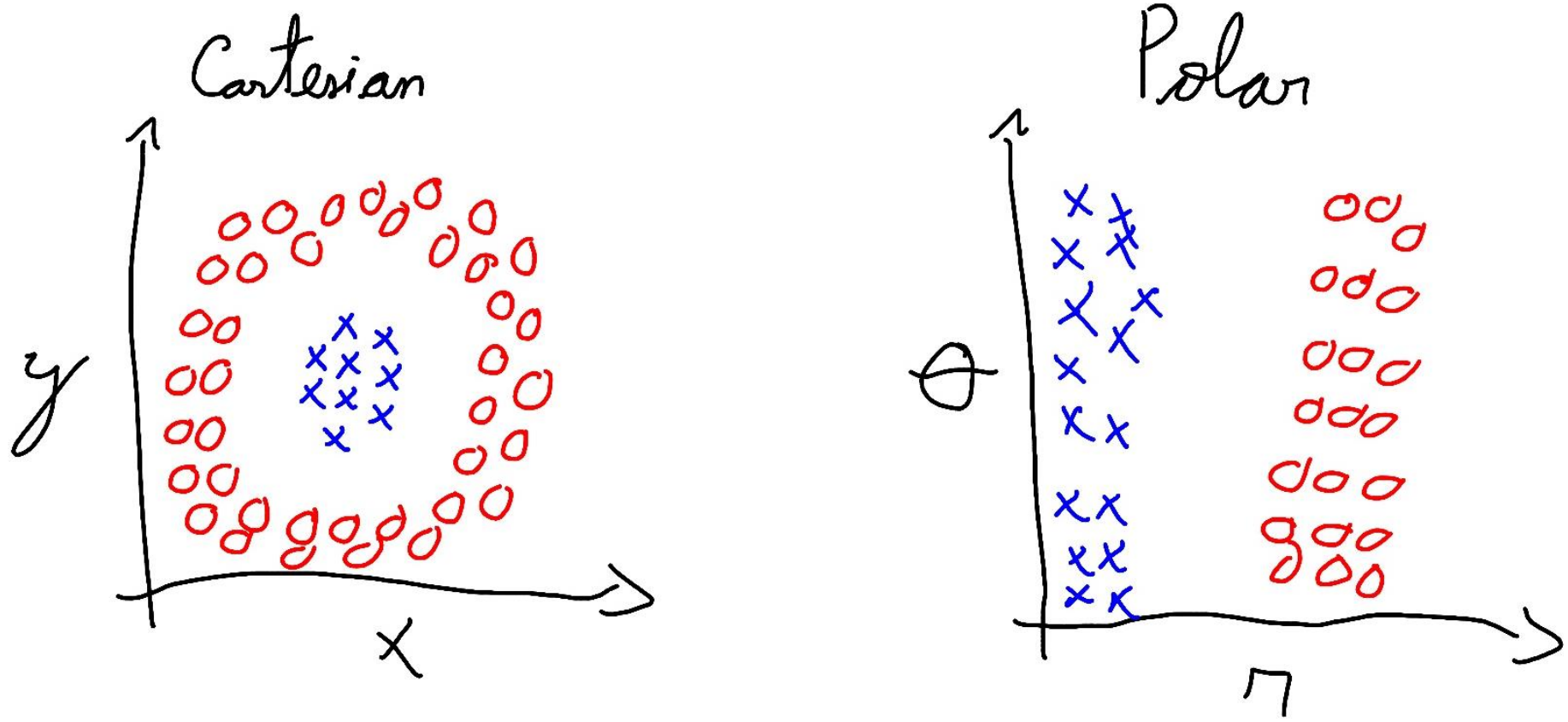
DL Hierarchy of Concepts



It is all about data representation….

- Roman numbers arithmetic:
  - CCCXXVII + CXXIII = ? CDL
- Arabic numbers arithmetic:
  - 327 + 123 = ? 450

It is all about data representation….



Scientific Community is Paying Attention…

Review Article | Published: 27 May 2015

Deep learning

Yann LeCun ✉, Yoshua Bengio & Geoffrey Hinton

Nature **521**, 436–444 (28 May 2015) | Download Citation ↓

Letter | Published: 25 January 2017

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva ✉, Brett Kuprel ✉, Roberto A. Novoa ✉, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun ✉

Nature **542**, 115–118 (02 February 2017) | Download Citation ↓

Letter | Published: 21 March 2018

Image reconstruction by domain-transform manifold learning

Bo Zhu, Jeremiah Z. Liu, Stephen F. Cauley, Bruce R. Rosen & Matthew S. Rosen ✉

Nature **555**, 487–492 (22 March 2018) | Download Citation ↓

Supervised x Unsupervised Learning

- **Supervised**: the data present associated outputs (labels/classes)
- **Unsupervised**: no labels are given to the learning algorithm
  - The goal is to discover groups in the data (clustering) or to determine the distribution of data within the input space (density estimation)

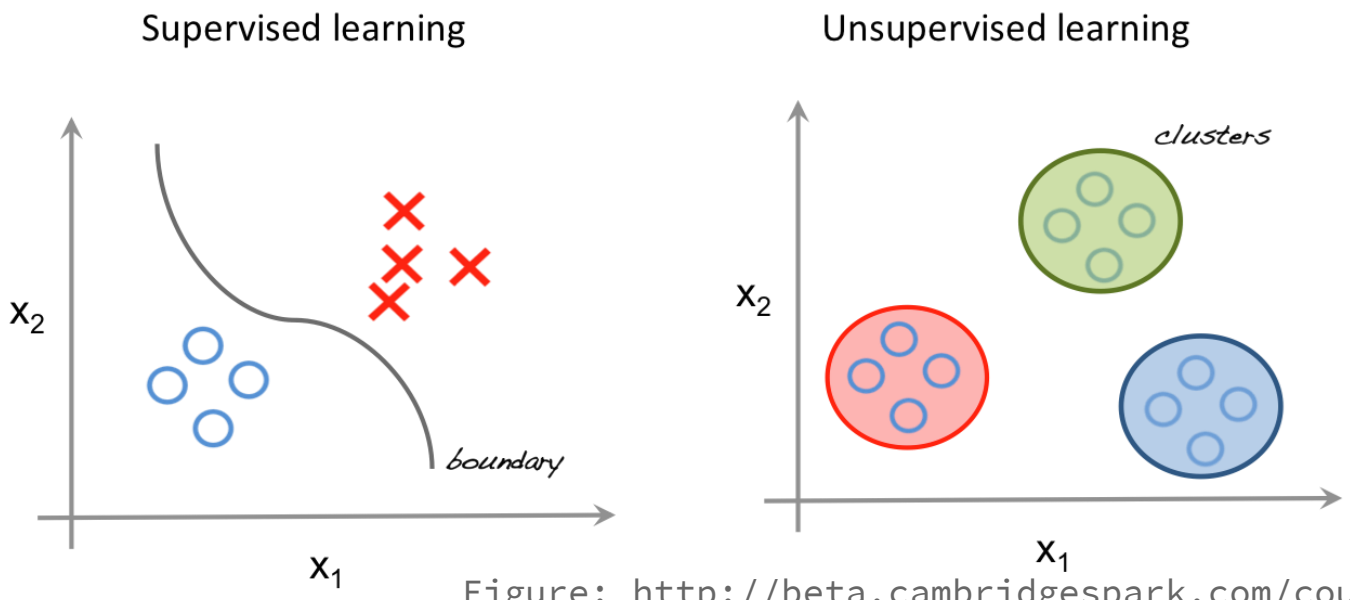
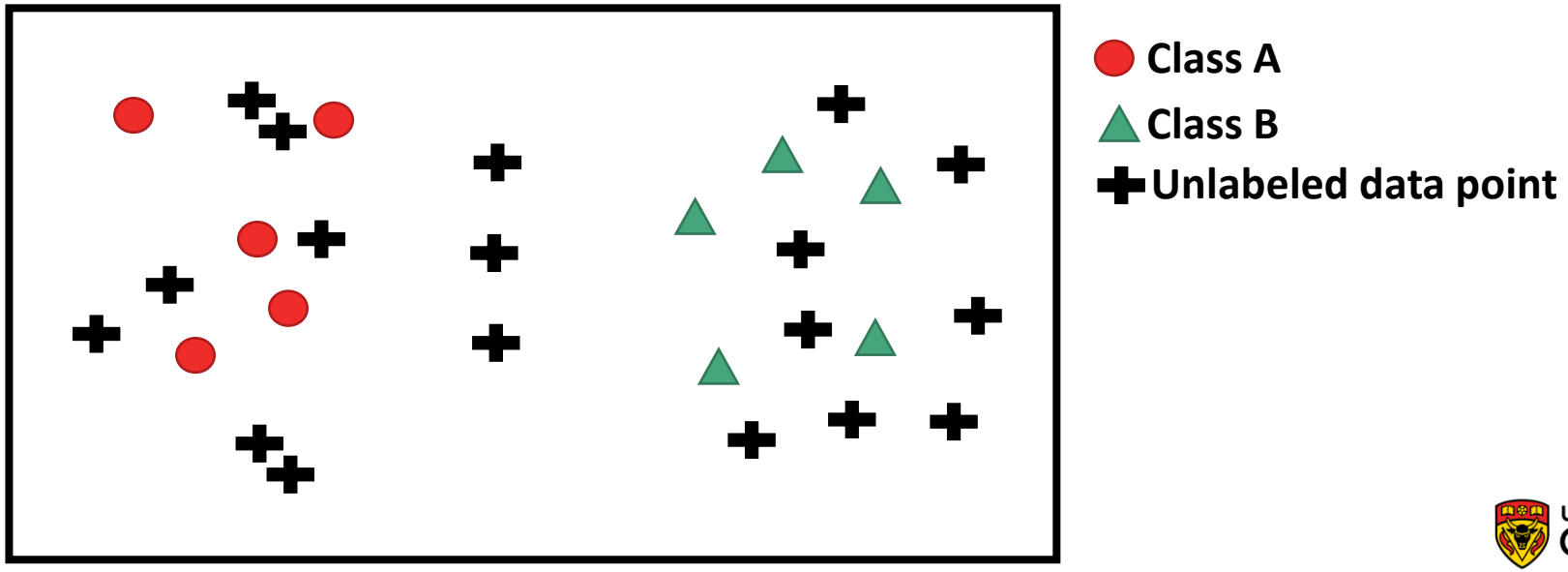


Figure: http://beta.cambridgespark.com/courses/jpm/01-module.html

18

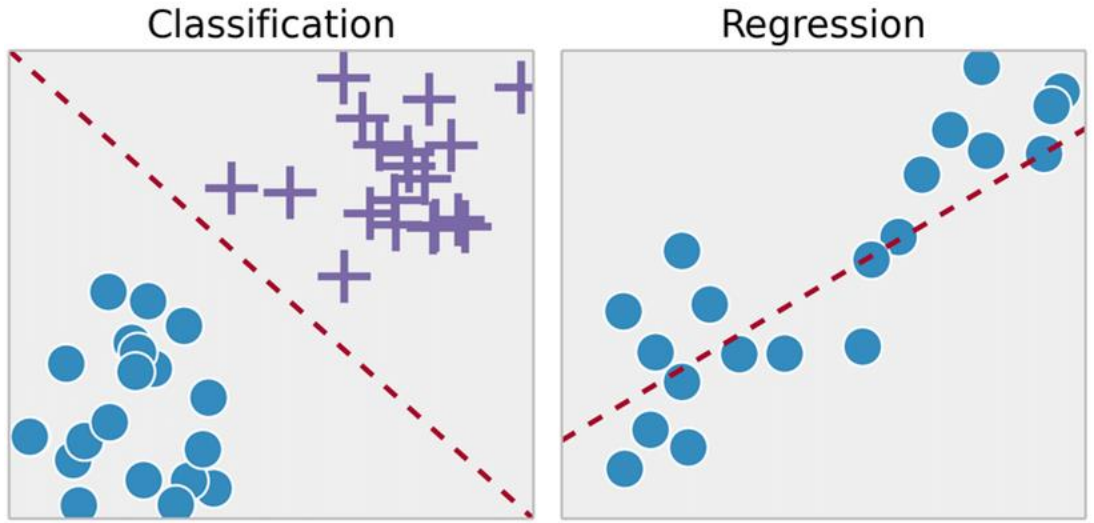
Semi-Supervised Learning

- **Semi-supervised learning** combines a small amount of labeled data with a large amount of unlabeled data during training.
  - Falls in between supervised and unsupervised learning
  - It is a case of weak supervision



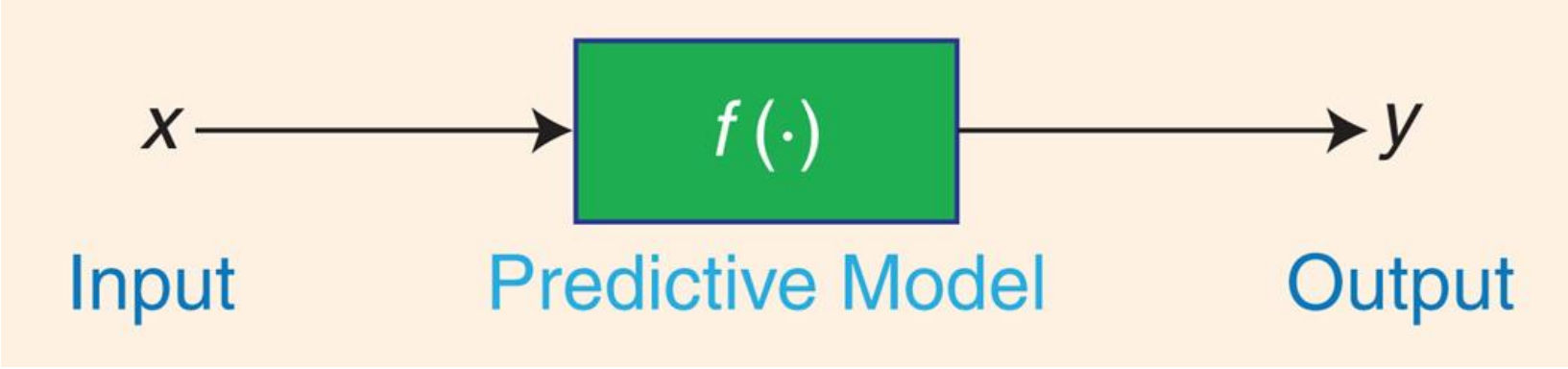
Classification × Regression

- **Classification** refers to decision among a discrete and typically small set of choices (*e.g.*, identifying a tumor as malignant or benign)
- **Regression** refers to estimating a continuous output variable (*e.g.*, diagnostic assessment of disease severity)



Supervised Classification

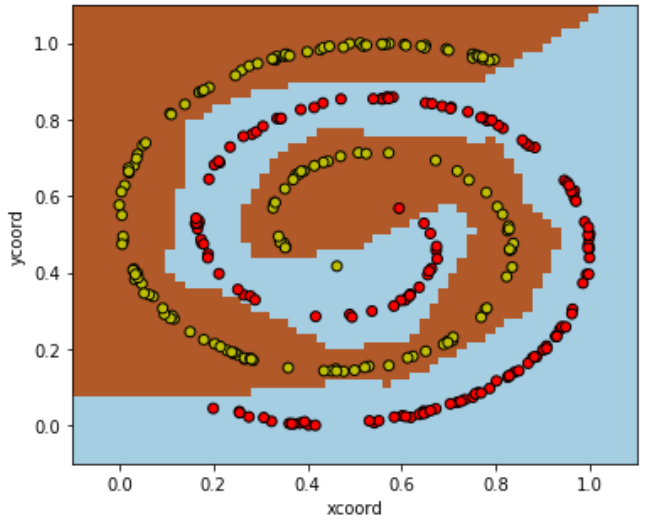
- Predictive model represents the assumed relationship between input variables in **x** and output variable **y**-> y=f(x)
  - the output of the predictive model can be a vector
- **x** is composed of *M* variables (called features), so that $\mathbf{x}_i \in \mathbb{R}^M$
- **y** can be a vector (e.g., in multi-class classifiers)



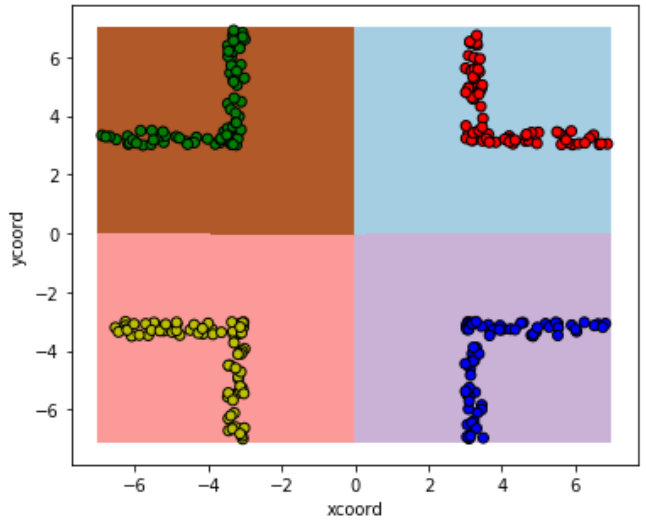
21

Binary x Multi-class x Multi-label Classification

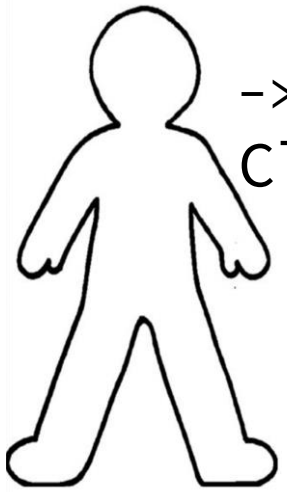
- **Binary:** 2 possible classes (labels).
- **Multi-class:** C (C>2) possible classes.
- **Multi-label:** A sample can belong to more than one class.



Binary



Multi-class

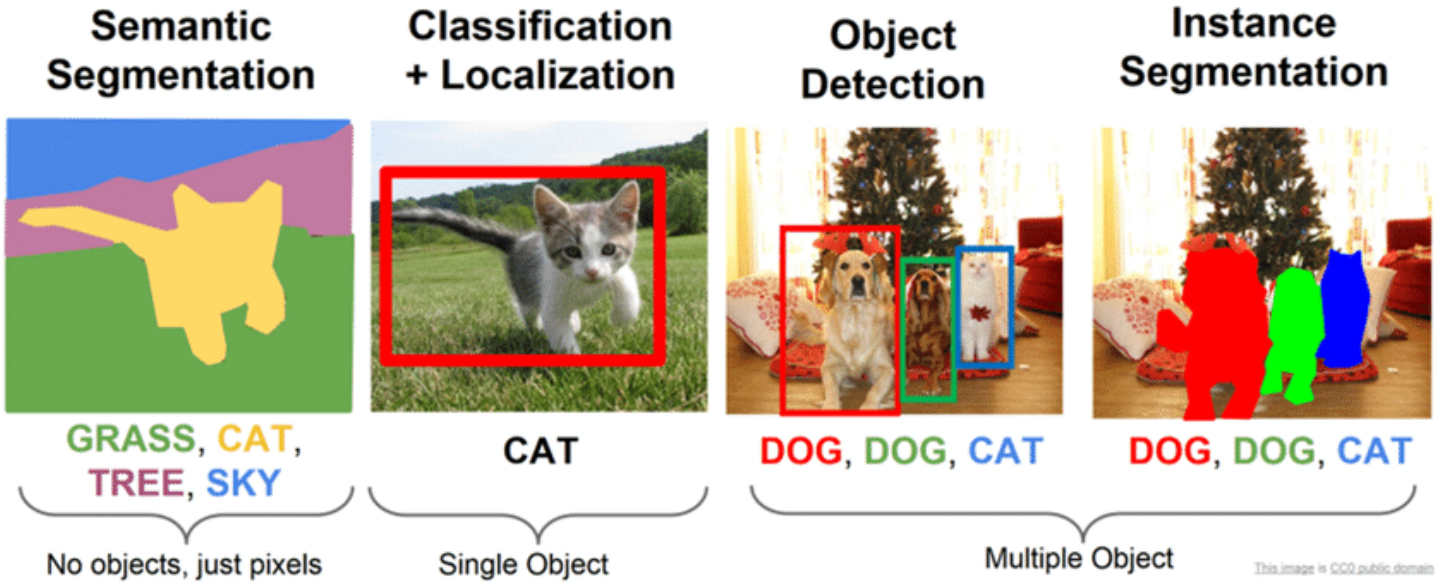


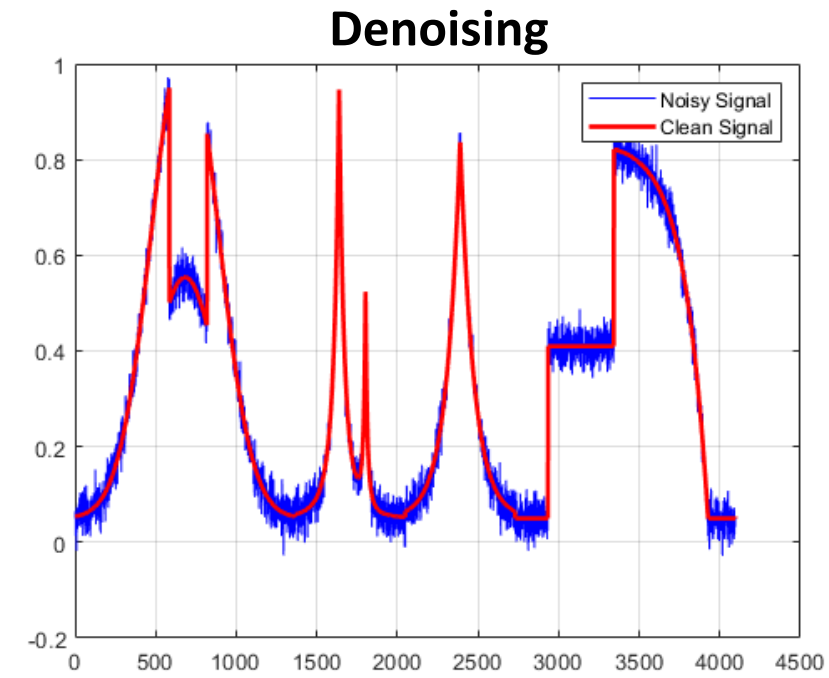
-> Class A and Class B

Multi-label

22

Types of Problems

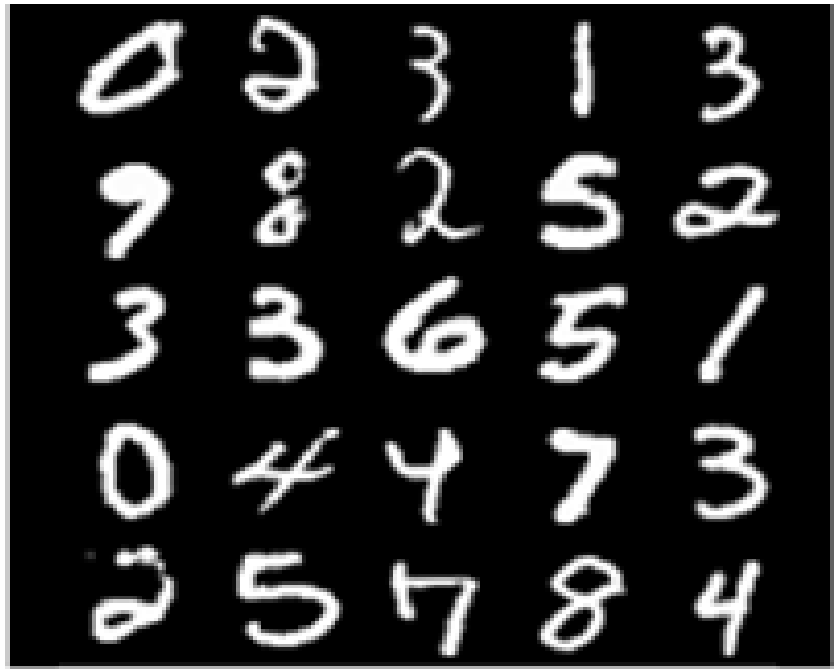




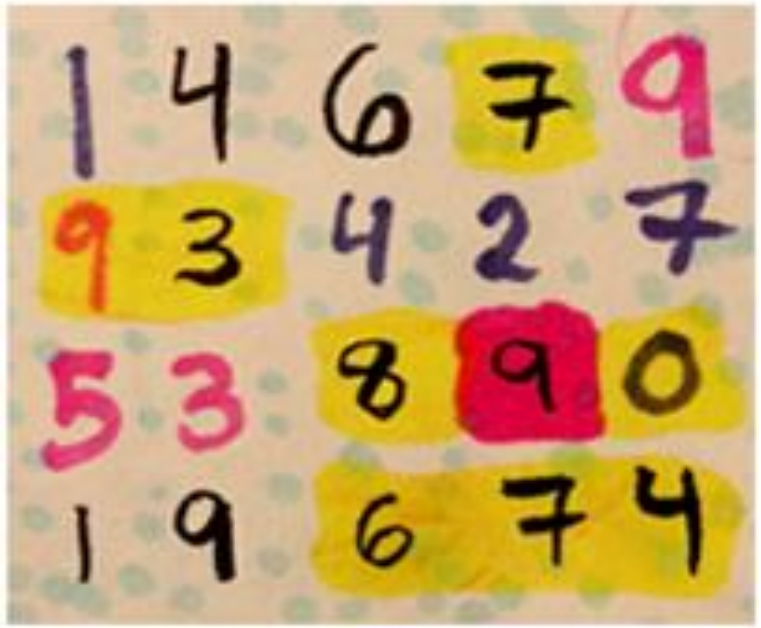
UNIVERSITY OF CALGARY

Domain Shift

- Domain shift occurs when the source data distribution is different (but related) to the target data distribution



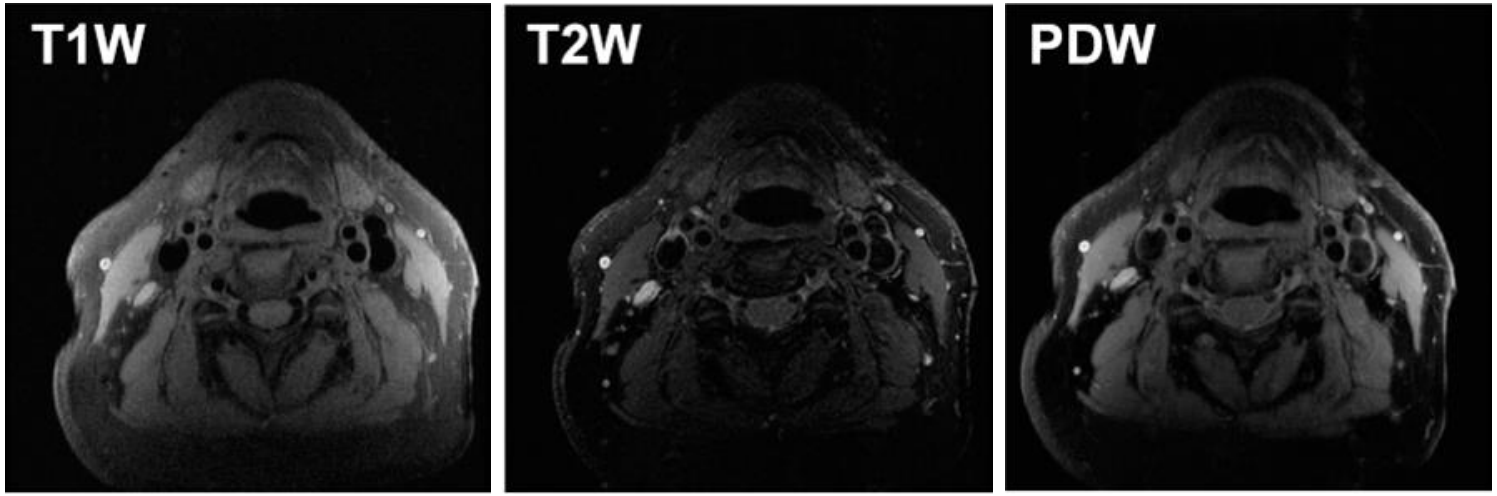
Source domain



Target domain

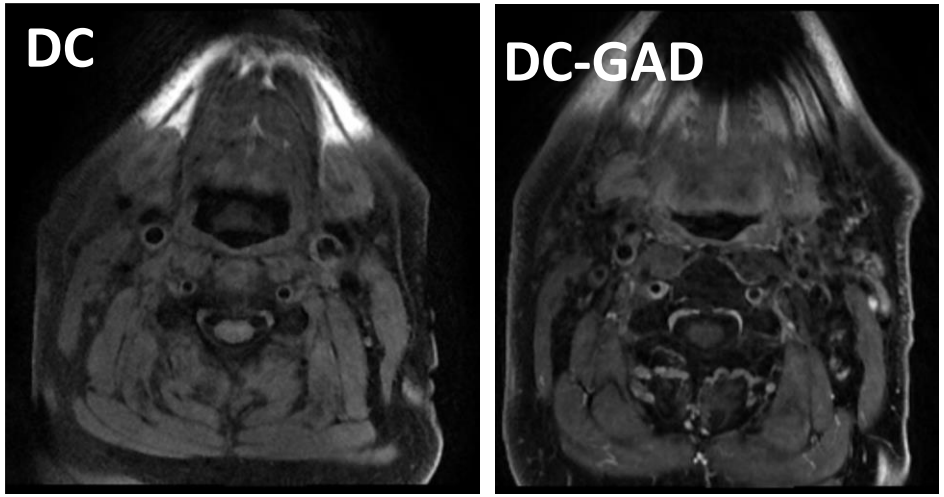
Domain Shift

AIM-HIGH Study



- The carotid arteries were manually annotated at the time of the study

CARDIS Study



- Leverage AIM-HIGH annotated data to create a segmentation model for the data being collected at CARDIS study

25

Summary

- DL < ML < AI
- The success of DL methods came with the development in hardware (GPUs/TPUs), software and availability of data (ImageNet)
- DL models can learn the features from the data
- DL models performance scales better with the amount of data available

Thanks!