

STOR566: Introduction to Deep Learning

Lecture 2: Overview of Machine Learning

Yao Li
UNC Chapel Hill

Aug 18, 2022

Outline

- Overview of machine learning
- Colab tutorial

Machine Learning: Overview

Human Learning

Observation

 → 0

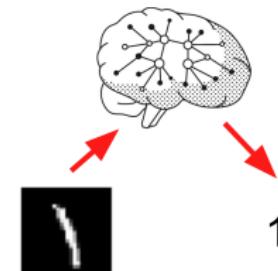
 → 0

 → 1

Learning



Decision rule



Machine Learning

Training Data

 → 0

 → 0

 → 1

Machine Learning



Decision rule

Machine Learning

Training Data

$$\begin{matrix} \text{0} \\ x_1 \end{matrix} \longrightarrow 0$$

$$\begin{matrix} \text{0} \\ x_2 \end{matrix} \longrightarrow 0$$

$$\begin{matrix} \text{l} \\ x_3 \end{matrix} \longrightarrow 1$$

Machine Learning

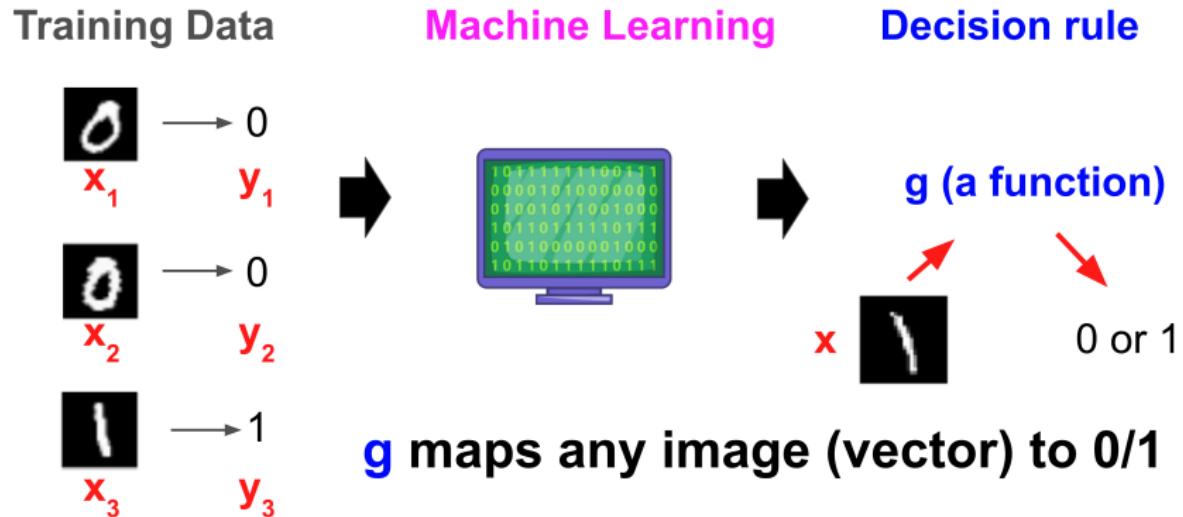


Decision rule

x_1 : vector of pixel values [0, 24, 128, ...]

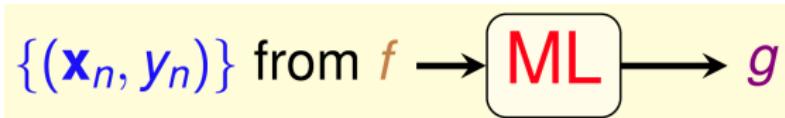
y_1 : 0 or 1

Machine Learning

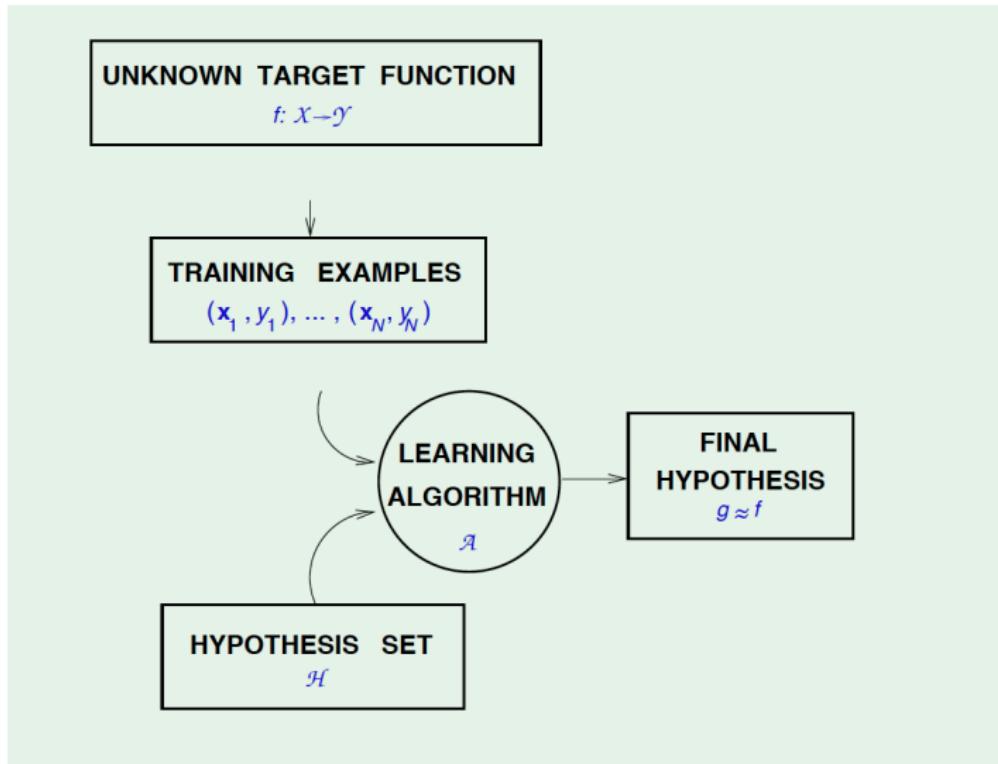


Formalize the Learning Problem

- Input: $\mathbf{x} \in \mathcal{X}$ (an image)
- Output: $y \in \mathcal{Y}$ (class)
- Target function to be learned:
 $f : \mathcal{X} \rightarrow \mathcal{Y}$ (ideal image classification function)
- Data:
 $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$
- Hypothesis (model)
 $g : \mathcal{X} \rightarrow \mathcal{Y}$ (**learned** formula to be used)



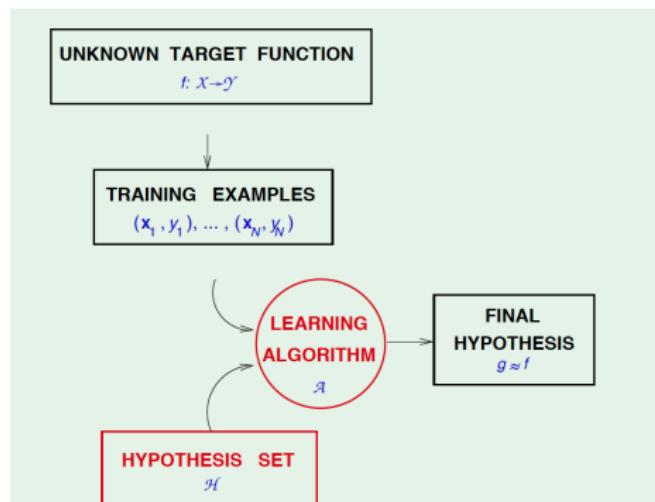
Basic Setup of Learning Problem



(Figure from “Learning from Data”)

Learning Model

- A learning model has two components:
 - The **hypothesis set** \mathcal{H} :
Set of candidate hypothesis (functions)
 - The **learning algorithm**:
To pick a hypothesis (function) from the \mathcal{H}
Usually **optimization algorithm** (choose the best function to minimize the **training error**)



Binary classification

- Data:
 - Features for each training example: $\{\mathbf{x}_n\}_{n=1}^N$, each $\mathbf{x}_n \in \mathbb{R}^d$
 - Labels for each training example: $y_n \in \{+1, -1\}$
- Goal: learn a function $f : \mathbb{R}^d \rightarrow \{+1, -1\}$
- Examples:
 - Credit approve/disapprove
 - Email spam/not-spam
 - patient sick/not sick
 - ...

Types of model (hypothesis)

- Linear hypothesis space:

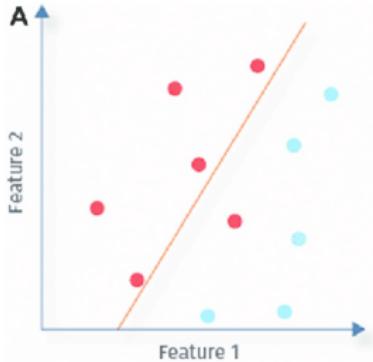
$$h(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^d w_i x_i - \text{threshold}\right)$$

- Feed forward (fully connected) network:

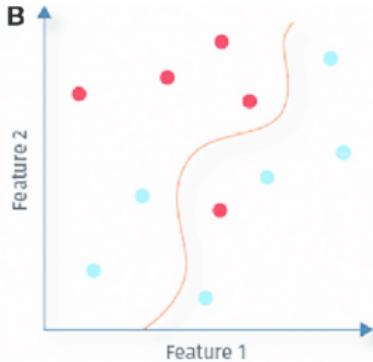
$$h(\mathbf{x}) = \text{sign}(W_L \cdots \sigma(W_2 \sigma(W_1 \mathbf{x} + b_1) + b_2) + b_L)$$

- Tree-based models
- ...

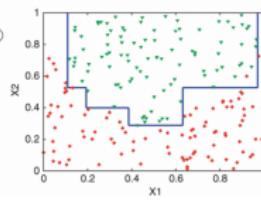
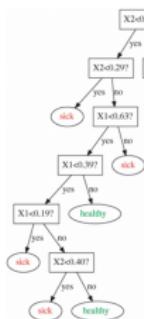
Types of model



Linear classification



Nonlinear classification



Tree-based classification

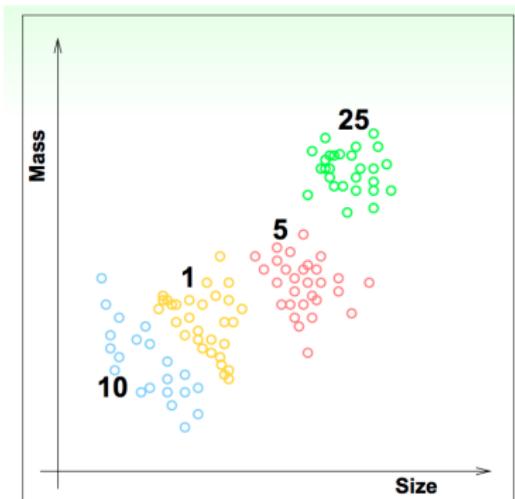
Other types of output space - Regression

- Regression: $y_n \in \mathbb{R}$ (output is a real number)
- Example:
 - Stock price prediction
 - Movie rating prediction
 - ...

Other types of output space - Multi-class prediction

Multi-class classification:

- $y_n \in \{1, \dots, C\}$ (C -way classification)
- Example: Coin recognition
 - Classify coins by two features (size, mass) ($x_n \in \mathbb{R}^2$)
 - $y_n \in \mathcal{Y} = \{1c, 5c, 10c, 25c\}$
($\mathcal{Y} = \{1, 2, 3, 4\}$)
- Other examples: hand-written digits, ...



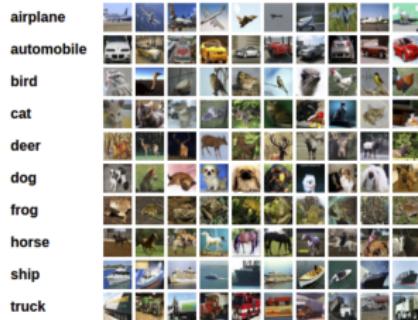
Other types of output space - Multi-class prediction

Multi-class classification:

- More examples: hand-written digit recognition, object classification, ...

0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4 4 4 4 4
5 5 5 5 5 5 5 5 5 5 5 5
6 6 6 6 6 6 6 6 6 6 6 6
7 7 7 7 7 7 7 7 7 7 7 7
8 8 8 8 8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9 9 9 9

MNIST



CIFAR

Other types of output space - Multi-label prediction

- Multi-class problem: Each sample only has **one label**
- Multi-label problem: Each sample can have **multiple labels**

Other types of output space - Multi-label prediction

- Multi-class problem: Each sample only has **one label**
- Multi-label problem: Each sample can have **multiple labels**
- Example:
 - Document categorization (news/sports/economy/· · ·)
 - Document/image tagging
 - · · ·

Other types of output space - Multi-label prediction

- **Extreme classification** (large output space problems):
 - Millions of billions of labels (but usually each sample only has few labels)
 - Recommendation systems: Predict a subset of preferred items for each user
 - Document retrieval or search: Predict a subset of related articles for a query
- Other examples:



1. A red stop sign sitting on the side of a road.
2. A stop sign on the corner of a street.
3. A red stop sign sitting on the side of a street.

Machine Learning Problems

Machine learning problems can usually be categorized into

- Supervised learning: every x_n comes with y_n (label)
(semi-supervised learning)
- Unsupervised learning: only x_n , no y_n
- Transfer learning

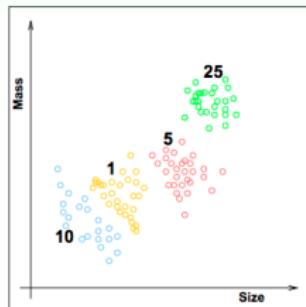
Unsupervised Learning (no y_n)

- Example: clustering

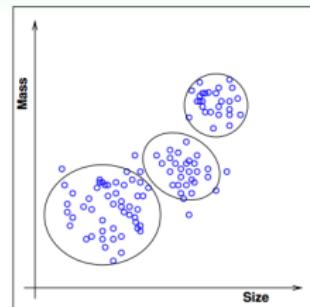
Given examples x_1, \dots, x_N , classify them into K classes

- Other unsupervised learning:

- Outlier detection: $\{x_n\} \Rightarrow \text{unusual}(x)$
- Dimensional reduction
- ...

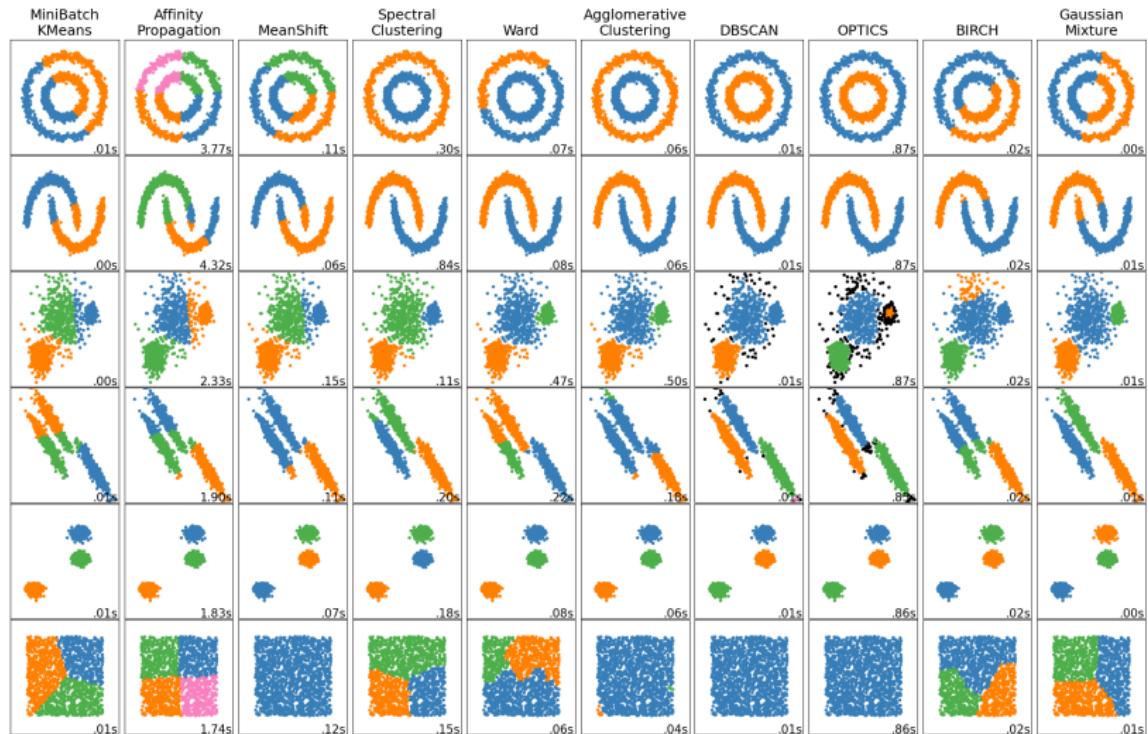


supervised multiclass classification



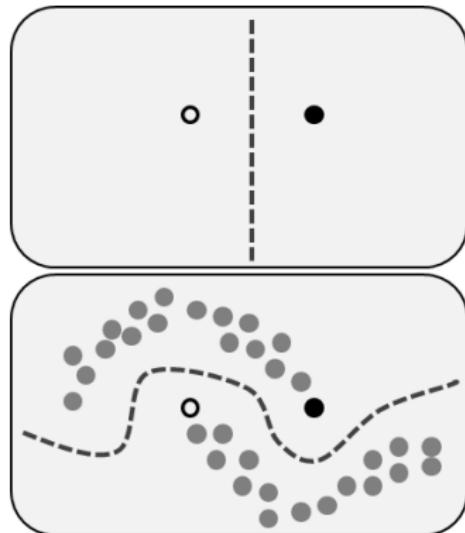
unsupervised multiclass classification
↔ ‘clustering’

Clustering



Semi-supervised learning

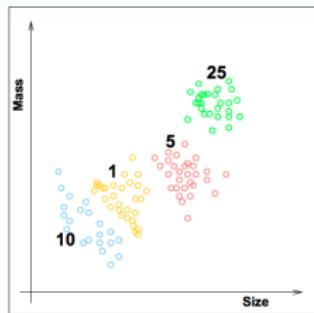
- Only some (few) x_n has y_n
- Labeled data is much more expensive than unlabeled data



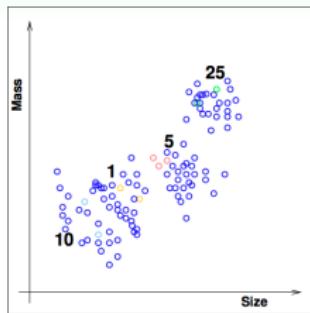
(From Wikipedia)

Semi-supervised learning

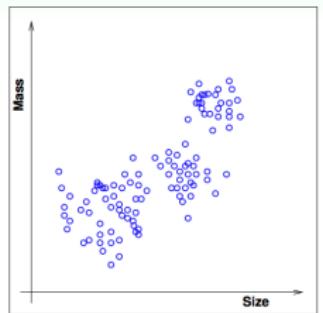
- Only some (few) x_n has y_n
- Labeled data is much more expensive than unlabeled data



supervised



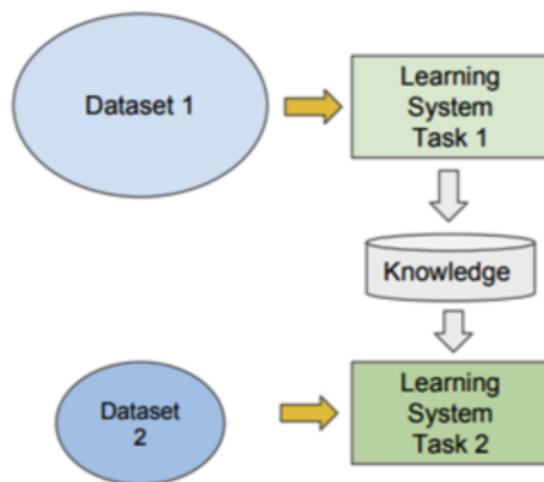
semi-supervised



unsupervised (clustering)

Transfer learning

- Source dataset D_{source} and target dataset D_{target}
- How to leverage the information of D_{source} to improve the performance of target task?



Self-supervised learning

- The pretraining can be done with **unlabeled** data (easy to collect gigantic unlabeled data)
 - Example: We can get almost unlimited unlabeled text from Internet
- Define the training task based on unlabeled data
 - Example: predict a word in a sentence
- Transfer the model to end task

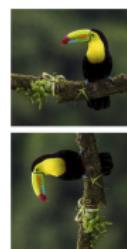
Original sentence:
In Autumn the **leaves** fall from the trees.

Masked sentence:
In Autumn the [] fall from the trees.

leaves
apples
raindrops
branches

Predicted words by the model

Masked language modeling



Are those the same images?



Are those the same images?

Contrastive learning

Conclusions

- Basic concept of learning:
 - Set up a hypothesis space (model class/potential functions)
 - Define an error measurement (define the quality of each function based on data)
 - Develop an algorithm to choose a good hypothesis based on the error measurement (optimization)
- Binary classification, multiclass, multilabel, etc.
- Different learning scenarios

Questions?