5CS037-Concepts and Technologies of AI

Lecture-06

Learning → Artificial Intelligence.

Introduction to Learning for Artificial Intelligence

Siman Giri

1. What is Learning?

Background and Motivation for Machine Learning.

1.1 Learning: Intuition

- **Example: Identify the spam emails!!!**
- **(Program a machine that learns how to filter spam emails.)**
  - In the early days of **“intelligent”** applications, many systems used **hand-coded rules** of **“if”** and **“else”** decisions to **process data** or **adjust to user input**.
  - A naïve solutions: machine can simply **make a array of all the words**, appearance of whose result in an **email being spam**, when a **new email arrives**, machine can check for those **blacklisted word from array**. If it matches one of them, it can be assigned as **spam** otherwise can be moved to **inbox**.
  - This would be an example of using an expert-designed rule system (“learning by memorization”) to design an **“intelligent”** application.

1.1 Learning: Intuition

- In our example - “learning by memorization” approach might work well but it lacks one important aspects of learning systems
  - – the ability to label unseen email-messages i.e. email messages which may be spam but does not contain any of the word in the black-list(array) will be delivered to our inbox.
- Manually crafting decision rules is feasible for some application, but has following two disadvantages:
  - The logic required to make a decision is specific to a single domain and task. Changing the task even slightly might required to rewrite of the whole system.
  - Designing rules requires a deep understanding of how a decision should be made by a human expert.
  - **{We did not learn from the data we had, instead we memorize a features of data.}**

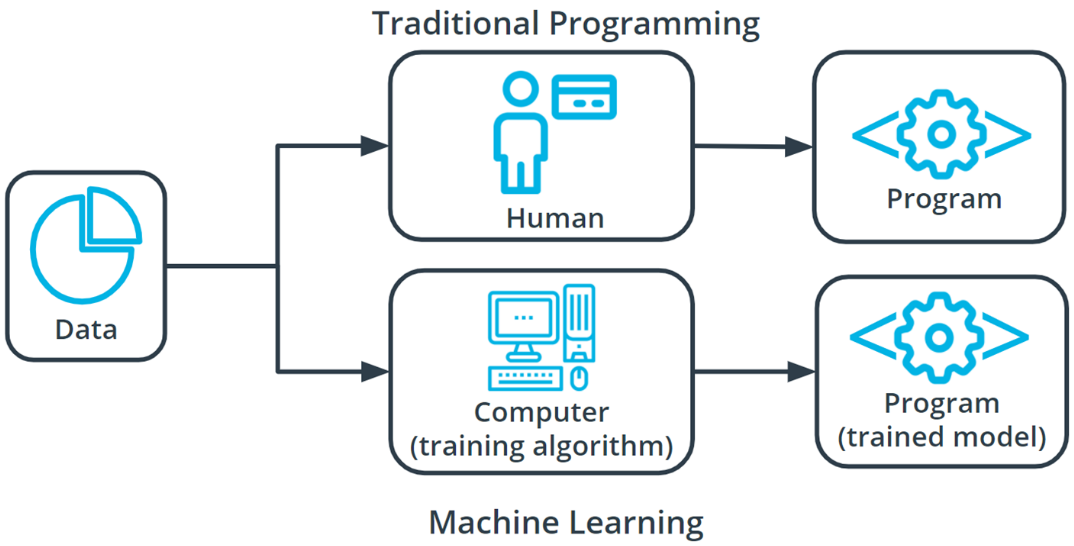
1.2 Learning: When do we need Machine Learning?

- A successful learning system must be able to progress from individual examples to broader generalization – also referred as "inductive reasoning" or "inductive inference".
- Example1: detect cat in an image.

  - way in which pixels (~ which make up an image in a computer) are "perceived" by the computer is very different from how humans perceive a face.
  - This difference in representation makes it basically impossible for a human to come up with a good set of rules to describe what constitutes a cat in a digital image.
  - Using machine learning, however, simply presenting a program with a large collection of images of faces is enough for an algorithm to determine what characteristics are needed to identify a face. {learning from data ~ What does it means learning from data?}

1.3 Learning: Traditional Programming Vs. Machine Learning.

- The main focus of machine learning is making decisions or inferences (predictions) based on data.
- The most successful kinds of machine learning algorithms are those that automate decision-making processes by generalizing from known examples.



2. Introduction to Machine Learning.

Definition

2.1 Definition: Machine Learning.

- Some popular definition from legends of the field:
  - “Learning is any process by which a system improves performance from experience”.
    -- Herbert Simon
  - Definition by Tom Mitchel(1998):
    - Machine Learning is the study of algorithms that:
    - Improve their performance P
    - At some task T
    - With experience E
    - A well defined learning task is given by <P, T, E>.
  - “Field of study that gives computers the ability to learn without being explicitly programmed.”
    - - Arthur Samuel ,1959 (an AI pioneer at IBM).

2.1 Definition: Machine Learning.

- Machine learning is a sub-domain of artificial intelligence (AI) that utilizes **Statistics, Pattern recognition, knowledge discovery and data mining** to **automatically learn and improve with experiences** without **being explicitly programmed**.
- As machine-learning (ML) methods have improved in their capability and scope, ML has become the best way, measured in terms of speed, human engineering time, and robustness, to make many applications.
- Great examples are face detection and speech recognition and many kinds of language-processing tasks. Almost any application that involves understanding data or signals that come from the real world can be best addressed using machine learning.

2.2 Machine Learning: Examples.

- **Handwriting recognition learning problem**
  - Task T: Recognizing and classifying handwritten words within images
  - Performance P: Percent of words correctly classified
  - Training experience E: A dataset of handwritten words with given classifications
- **A robot driving learning problem**
  - Task T: Driving on highways using vision sensors
  - Performance measure P: Average distance traveled before an error
  - Training experience E: A sequence of images and steering commands recorded while observing a human driver
- **A chess learning problem**
  - Task T: Playing chess
  - Performance measure P: Percent of games won against opponents
  - Training experience E: Playing practice games against itself

2.3 Machine Learning: Premises.

- There exists some **pattern/behavior** of interest:

  (Some Task to be solved)

- The **pattern/behavior** is difficult to **describe:**

  (Encoding a rule to understand a behavior is difficult)

- There is **data**

  (past experiences are in abundant)

- Use data to **"learn"** the pattern

2.4 Machine Learning: Human in the loop.

- One crucial aspect of machine learning approaches to solving problems is that human and often undervalued engineering plays an important role.
  - A human still has to frame the problem,
    - acquire and organize data, design a space of possible solutions,
    - select a learning algorithm and its parameters,
    - apply the algorithm to the data,
    - validate the resulting solution to decide whether it’s good enough to use, etc.
- These steps are of great importance.

2.5 Machine Learning: Cautions!!

- Machine learning is a very general and useful framework, but it is not **“magic”** and will not always work.
  - In order to better understand when it will and when it will not work, it is useful to **formalize** the **learning problem** more.
- **Some challenges of Machine Learning:**
  - Why do we think that previously seen data will help us predict the future?
  - **estimation:**
    - When we have data that are noisy reflections of some underlying quantity of interest, we have to aggregate the data and make estimates or predictions about the quantity.
    - How do we deal with the fact that, for example, the same treatment may end up with different results on different trials?
    - How can we predict how well an estimate may compare to future results?
  - **generalization:**
    - How can we predict results of a situation or experiment that we have never encountered before in our data set?

3.Components of Machine Learning.

Framing a Learning Problem.

3.1 Types of Learning.

- There are many different problem classes in machine learning.
- They vary according to **what kind of data** is provided and **what kind of conclusions** are to be drawn from it.
- Some-popular kind are:
  - **Supervised Learning**
  - Unsupervised Learning
  - Reinforcement Learning
- In this course, we will focus on **classification and regression** (two examples of supervised learning) and will touch on unsupervised learning if time permits.

3.2 Datasets

- Data is the driving force of Machine Learning.
- Data comes in various format for examples:
  - Images of cat.
  - Housing prices.
  - Weather information.
- Datasets are made up of individual examples that contain features and a label.
  - Examples that contain both features and a label are called **labeled datasets**.
  - Examples that contain only features are called **unlabeled datasets**.

Features | Label

date	lat	long	temp	humidity	cloud_coverage	wind_direction	atmp_pressure	rainfall
2021-09-09	49.71N	82.16W	74	20	3	N	18.6	.01
2021-09-09	32.71N	117.16W	82	42	6	SW	29.94	.23

Example

Labeled Dataset

Features

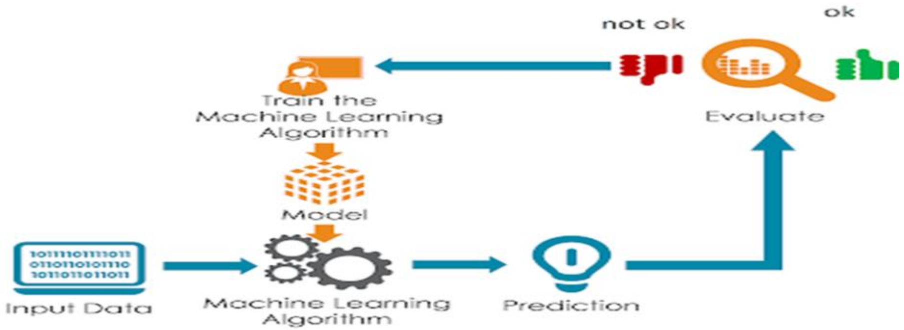
date	lat	long	temp	humidity	cloud_coverage	wind_direction	atmp_pressure
2021-09-09	49.71N	82.16W	74	20	3	N	18.6
2021-09-09	32.71N	117.16W	82	42	6	SW	29.94

Example

Unlabeled Dataset

3.3 Elements of Machine Learning.

- There may be hundreds of machine learning algorithms, all of those algorithms must have following three attributes:
  - A Decision Process (Representation/Model):
    - Machine learning algorithms(Models) are used to make inference or estimate of an output based on input data – labeled or unlabeled.
  - An Error Function (Evaluation):
    - A performance metric used to evaluate the estimate of a model.
    - Metrics depends on types of learning (supervised or unsupervised) and types of task (Classification or Regression)
  - An model Optimization Process:
    - An automated algorithm or process used to update parameters of machine learning models until threshold or accepted evaluation metric has been achieved.



3.4 A Decision Process: Function Approximation.

- Machine learning is concerned with using the right features to build the right models that achieve the right tasks.
- For a given problem, what kind of function better approximates the relationship between input (feature) with output (target/label)
- The function broadly can be classified as:
  - **Numerical Function:**
    - **Linear Regression.**
    - **Support Vector Machines.**
  - **Symbolic Function (Logical or Rule Based):**
    - **Decision Tree and Random Forest**
  - **Instance Based:**
    - **Nearest Neighbor**
  - **Probabilistic Models:**
    - **Naïve Bayes**

3.5 Framework of a learning Process.

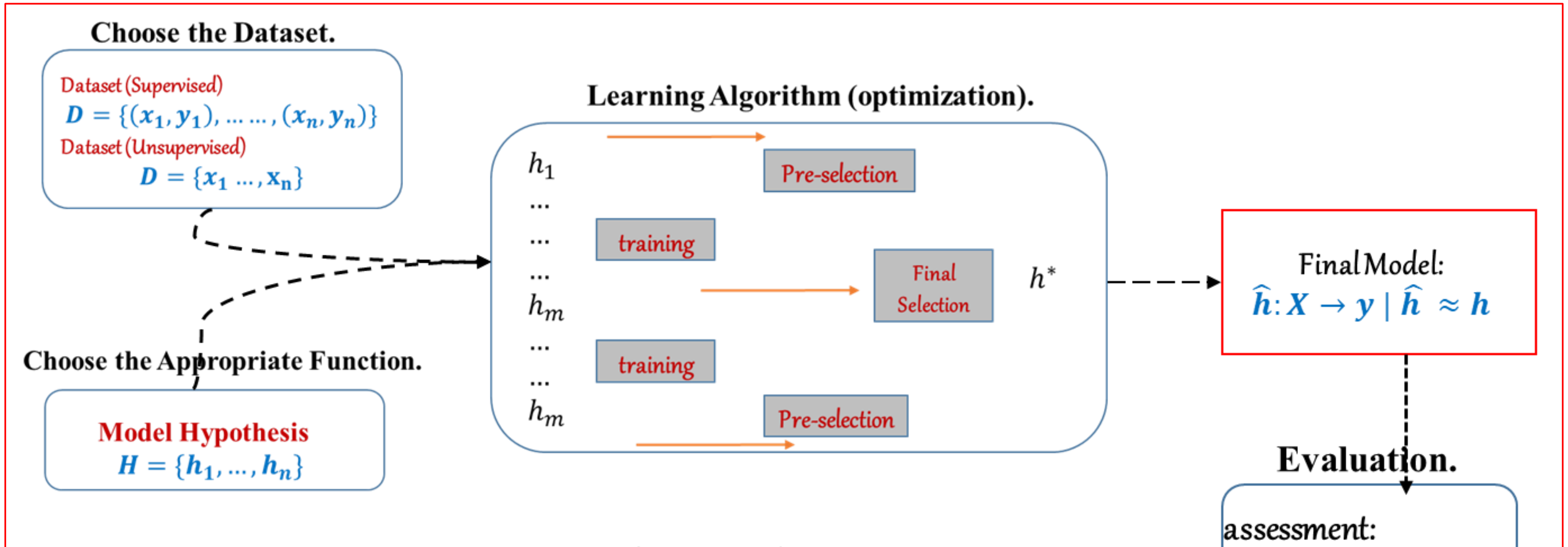


Fig: Schematic of Learning Process

4. Supervised Machine Learning.

4.1 Supervised Learning: Data

- **Data:**
  - For Supervised Learning Setup, **training data** comes in pairs of inputs **(x, y):** where $\boldsymbol{X} \in \boldsymbol{R}^{\boldsymbol{d}}$ is the input instance and $\boldsymbol{Y}$ its label, which can be written as:
    - $\boldsymbol{D} = \{(\boldsymbol{x_1}, \boldsymbol{y_1}) \dots (\boldsymbol{x_n}, \boldsymbol{y_n})\} \subseteq \boldsymbol{R}^{\boldsymbol{d}} * \boldsymbol{C}$
    - Where:
      - $\boldsymbol{R}^{\boldsymbol{d}}$: **d-dimensional feature space.**
      - $\boldsymbol{x_i}$: **input vector of the** $\boldsymbol{i^{th}}$ **sample.**
      - $\boldsymbol{y_i}$: **label of the** $\boldsymbol{i^{th}}$ **sample.**
      - $\boldsymbol{C}$: **label space.**

Terminology Alert!!!

- **Variables:**
    - **Target or output variables** also referred as **dependent variables.**
    - **Predictor, Feature or input variables** also referred as **independent variables**
- **Notations:**
    - **Feature Variables**: **x or X**.
    - **Actual Target Variables**: **y or Y**.
    - **Predicted Target Variables**:$\hat{y}$ **or** $\hat{Y}$.

4.1 Supervised Learning: Data

- **<u>Data: Feature Vector Space</u>**.
  - We call $\boldsymbol{x_i^d}$ a feature vectors, each one of it's dimensions is a features describing the $\boldsymbol{i^{th}}$ sample.
  - Examples: **House Price Prediction**.

Street	OverallCond	YearBuilt	YearRemo	MasVnrArea	TotalBsmt	Heating	CentralAir	BsmtFullB	FullBath	HalfBath	Bedroom/	SaleCondition	Price
Pave	6	1961	1961	0	882	GasA	Y	0	1	0	2	Normal	11622
Pave	6	1958	1958	108	1329	GasA	Y	0	1	1	3	Normal	14267
Pave	5	1997	1998	0	928	GasA	Y	0	2	1	3	Normal	13830
Pave	6	1998	1998	20	926	GasA	Y	0	2	1	3	Normal	9978
Pave	5	1992	1992	0	1280	GasA	Y	0	2	0	2	Normal	5005
Pave	5	1993	1994	0	763	GasA	Y	0	2	1	3	Normal	10000
Pave	7	1992	2007	0	1168	GasA	Y	1	2	0	3	Normal	7980
Pave	5	1998	1998	0	789	GasA	Y	0	2	1	3	Normal	8402
Pave	5	1990	1990	0	1300	GasA	Y	1	1	1	2	Normal	10176
Pave	5	1970	1970	0	882	GasA	Y	1	1	0	2	Normal	8400
Pave	5	1999	1999	0	1405	GasA	Y	1	2	0	2	Normal	5858
Pave	5	1971	1971	504	483	GasA	Y	0	1	1	2	Normal	1680
Pave	5	1971	1971	492	525	GasA	Y	0	1	1	3	Normal	1680
Pave	6	1975	1975	0	855	GasA	Y	0	2	1	3	Normal	2280
Pave	6	1975	1975	0	836	GasA	Y	0	1	0	2	Normal	2280
Pave	5	2009	2010	162	1590	GasA	Y	0	2	1	3	Partial	12858
Pave	5	2009	2010	256	1544	GasA	Y	0	2	0	3	Partial	12883
Pave	5	2005	2005	615	1698	GasA	Y	0	2	0	3	Normal	11520
Pave	5	2005	2006	240	1822	GasA	Y	0	2	0	3	Normal	14122
Pave	5	2003	2004	1095	2846	GasA	Y	1	2	1	3	Normal	14300
Pave	5	2002	2002	232	1671	GasA	Y	1	2	1	3	Normal	13650
Pave	5	2006	2006	178	1370	GasA	Y	0	2	0	2	Normal	7132
Pave	5	2005	2005	0	1324	GasA	Y	0	2	0	3	Normal	18494
Pave	5	2006	2006	14	1145	GasA	Y	0	2	0	2	Normal	3203
Pave	5	2004	2004	0	384	GasA	Y	1	2	1	3	Normal	13300

4.1 Supervised Learning: Tasks

- **<u>Data: label Space</u>**.
  - There can be multiple scenario for the label space c.

Binary Classification	$c = \{0\ or\ 1\}$	E.g.: An email is either spam or not a spam.
Multi Class Classification	$c = \{1,2,..k\}(k \geq 2)$	E.g.: Traffic sign Classification.
Regression	$c = \mathbb{R}$	E.g.: Height of the person.

4.2 Supervised Learning: Regression.

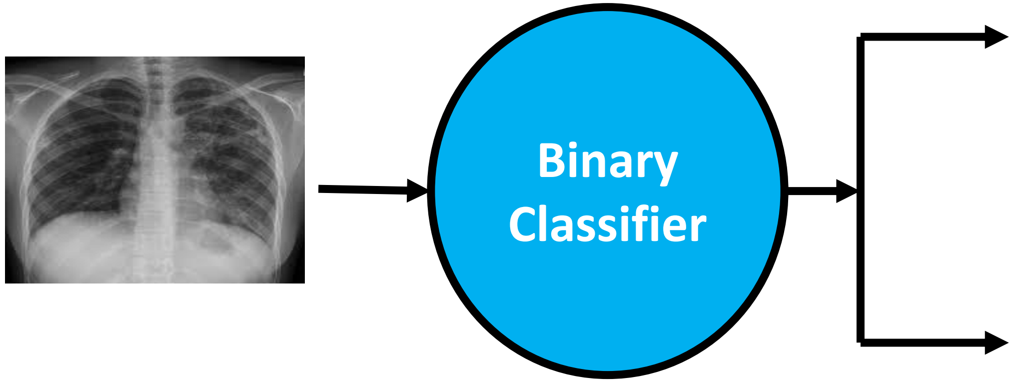
- House Price Prediction:

Street	OverallCond	YearBuilt	YearRemo	MasVnrArea	TotalBsmt	Heating	CentralAir	BsmtFullB	FullBath	HalfBath	BedroomA	SaleCondition	Price
Pave	6	1961	1961	0	882	GasA	Y	0	1	0	2	Normal	11622
Pave	6	1958	1958	108	1329	GasA	Y	0	1	1	3	Normal	14267
Pave	5	1997	1998	0	928	GasA	Y	0	2	1	3	Normal	13830
Pave	6	1998	1998	20	926	GasA	Y	0	2	1	3	Normal	9978
Pave	5	1992	1992	0	1280	GasA	Y	0	2	0	2	Normal	5005
Pave	5	1993	1994	0	763	GasA	Y	0	2	1	3	Normal	10000
Pave	7	1992	2007	0	1168	GasA	Y	1	2	0	3	Normal	7980
Pave	5	1998	1998	0	789	GasA	Y	0	2	1	3	Normal	8402
Pave	5	1990	1990	0	1300	GasA	Y	1	1	1	2	Normal	10176
Pave	5	1970	1970	0	882	GasA	Y	1	1	0	2	Normal	8400
Pave	5	1999	1999	0	1405	GasA	Y	1	2	0	2	Normal	5858
Pave	5	1971	1971	504	483	GasA	Y	0	1	1	2	Normal	1680
Pave	5	1971	1971	492	525	GasA	Y	0	1	1	3	Normal	1680
Pave	6	1975	1975	0	855	GasA	Y	0	2	1	3	Normal	2280
Pave	6	1975	1975	0	836	GasA	Y	0	1	0	2	Normal	2280
Pave	5	2009	2010	162	1590	GasA	Y	0	2	1	3	Partial	12858
Pave	5	2009	2010	256	1544	GasA	Y	0	2	0	3	Partial	12883
Pave	5	2005	2005	615	1698	GasA	Y	0	2	0	3	Normal	11520
Pave	5	2005	2006	240	1822	GasA	Y	0	2	0	3	Normal	14122
Pave	5	2003	2004	1095	2846	GasA	Y	1	2	1	3	Normal	14300
Pave	5	2002	2002	232	1671	GasA	Y	1	2	1	3	Normal	13650
Pave	5	2006	2006	178	1370	GasA	Y	0	2	0	2	Normal	7132
Pave	5	2005	2005	0	1324	GasA	Y	0	2	0	3	Normal	18494
Pave	5	2006	2006	14	1145	GasA	Y	0	2	0	2	Normal	3203
Pave	5	2004	2004	0	384	GasA	Y	1	2	1	3	Normal	13300

4.2 Supervised Learning: Classification.

Input:

Output:



Benign **0**

Malignant **1**

- Tasks in Supervised Learning-Classification Task:
  - **Binary Classification.**
  - **Multi-Class Classification.**

Binary Response

$$Y = \begin{cases} \text{Non} - \text{smoker} \\ \text{Smoker} \end{cases}$$

$$Y = \begin{cases} \text{Survives} \\ \text{Dies} \end{cases}$$

Ordinal Response

$$Y = \begin{cases} \text{Agree} \\ \text{Neutral} \\ \text{Disagree} \end{cases}$$

4.3 Supervised Learning: Problem Setup.

- **Data → Learning**.
- Our assumptions:
  - We belief that **datapoints** $(\boldsymbol{x_i}, \boldsymbol{y_i})$ are drawn from some (unknown) **distribution** $\boldsymbol{P(X, Y)}$.
  - We would like to learn a function "$\boldsymbol{h}$“ such that for a new pair $(\boldsymbol{x}, \boldsymbol{y})$, we have $\boldsymbol{h(x) = \hat{y}}$ where $\boldsymbol{\hat{y} \approx y}$.
- Learning:
  - **Step-I:** .From a hypothesis class **“H”** i.e. "$\boldsymbol{h \in H}$“**;** we need to select the machine learning algorithm that we think is appropriate for the particular problem.
  - **Step-II:** Find the best function with in the class "$h \in H$“
    - How can we find best function?
    - For this we need be able to evaluate what it means for one function to be better than another. (Aka Loss function.)

4.4 Supervised Learning: Evaluation Metric.

- **Loss function($\mathbb{L}$):**
- Aka Error Function/Cost Function.
- A loss function evaluates a hypothesis $h \in H$; on our training data and tells us how bad it is.
  - Higher the loss, the worse it is
  - A loss of zero means it makes **perfect predictions**.
- It is common practice to normalize the loss by the total number of training samples,"n", so that the output can interpreted as the average loss per sample.

4.4 Supervised Learning: Evaluation Metric.

- **Loss function($\mathbb{L}$): Examples.**
- **Zero-One Loss:**
  - It counts how many mistakes and hypothesis function "h" makes on the training set, given by:
    - $\mathbb{L}_{0\backslash 1}(\boldsymbol{h}) = \frac{1}{n}(\sum \boldsymbol{\delta_h}(\boldsymbol{x_i}) \neq \boldsymbol{y_i}; \{\boldsymbol{\delta_h}(\boldsymbol{x_i}) \neq \boldsymbol{y_i} = \begin{cases} \mathbf{1}, \boldsymbol{if}\ \boldsymbol{h}(\boldsymbol{x_i}) \neq \boldsymbol{y\_i} \\ \mathbf{0}, \boldsymbol{other\ wise}. \end{cases}$
  - For every single example: if it is miss-predicted then loss is 1 and 0 otherwise.
  - The normalized zero-one loss returns the fraction of misclassified training samples.
  - Can be used to evaluate classifier but are seldom used because it is rarely is useful to guide **optimization procedures** because the function is non-differentiable and non-continuous.

4.4 Supervised Learning: Evaluation Metric.

- **Loss function($\mathbb{L}$): Examples.**
  - **Squared Loss:**
    - Typically used in regression settings. It iterates over all training sample and computes the loss as:
      - $\mathbb{L}_{sq}(h) = \frac{1}{n}(\sum_i^n (h(x_i) - y_i)^2$.
  - **Absolute Loss:**
    - Expressed as:
      - $\mathbb{L}_{abs}(h) = \frac{1}{n}(\sum_i^n |h(x_i) - y_i|$.

4.5 What is (Supervised) Machine Learning!!!

- It is an attempt to find the function ***"h"*** that minimizes the selected loss such that:
  - $\boldsymbol{h = argmin_{h \in H} \mathbb{L}(h)}$
- A big part of machine learning focuses on the question, how to do this minimization efficiently?
- If you find a function ***h(.)*** with low loss on your data ***D***, how do you know whether it will still get examples right that are not in ***D***?
  - Generalization!!!

4.6 What after learning '*h*'?

- **Prediction:**
  - **Learned model(hypothesis) h(.)** is used to predict the label Y for data without label, the predicted label is represented as $\hat{Y}$.
- **Inference:**
  - Understanding the association between Y and X.
    - Which predictors are associated with the response?
    - What is the relationship between the response and predictor?
    - Can the relationship between Y and each predictor be adequately summarized using a linear equation?

4.7 Examples of Supervised Learning Task

- **Identifying the zip code from handwritten digits on an envelope:**
  - Here the input is a scan of the handwriting, and the desired output is the actual digits in the zip code. To create a dataset for building a machine learning model, you need to collect many envelopes. Then you can read the zip codes yourself and store the digits as your desired outcomes.
- **Determining whether a tumor is benign based on a medical image:**
  - Here the input is the image, and the output is whether the tumor is benign. To create a dataset for building a model, you need a database of medical images. You also need an expert opinion, so a doctor needs to look at all of the images and decide which tumors are benign and which are not. It might even be necessary to do additional diagnosis beyond the content of the image to determine whether the tumor in the image is cancerous or not.
- **Detecting fraudulent activity in credit card transactions:**
  - Here the input is a record of the credit card transaction, and the output is whether it is likely to be fraudulent or not. Assuming that you are the entity distributing the credit cards, collecting a dataset means storing all transactions and recording if a user reports any transaction as fraudulent.

5. Unsupervised Learning

5.1 Unsupervised Learning: Introduction.

- Unsupervised learning focuses on a set of **statistical tool/ methods** intended for setting in which we have only a **set of features** i.e.:
  - $D = \{(x_1) \ldots (x_n)\} \subseteq R^d$
  - Here:
    - R^d: d-dimensional feature space.
    - x_i: input vector of the i^{th} sample measured on **n observations**.
- In unsupervised learning we are **not interested in prediction**, because we do not have an associated response variable **Y**.
- Then, What can be the **goal/objective** of unsupervised learning?

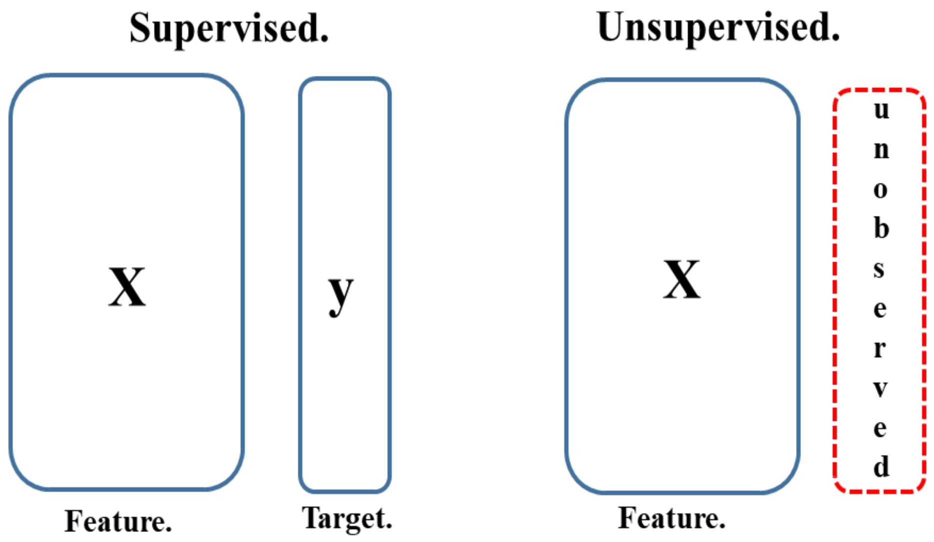


Fig: Supervised Vs. Unsupervised Learning.

5.2 Unsupervised Learning: Need.

- What can be the **goal/objective** of unsupervised learning?
  - Goal in unsupervised learning is to discover interesting things about the measurements/features. Thus tasks in unsupervised learnings are defined purely from exploratory perspective.
- Unsupervised learning explores the relationship between the feature variables and tries to answer the questions such as:
  - Is there an informative way to visualize the data?
  - Can we discover subgroups among the variables or among the observations?

5.3 Unsupervised Learning: ExampleTask.

- **Clustering**:
  - The method of dividing the objects into clusters which are similar between them and are dissimilar to the objects belonging to another cluster
    - K means Clustering,
- **Association**:
  - Rule based machine learning algorithms, that discovers the probability of the co- occurrence of items in a collection i.e. finding relationships between variables in a given dataset.
    - Apriori Algorithm, FP-Growth Algorithm
- **Dimensionality reduction**:
  - *Principal component analysis*

6. Brief History of Machine Learning.

6.1 Some Major Moments in History of AI -1.

- **Pre-2000s**
    - **1950-Alan Turing**
        - Published a paper on "Computing Machinery and Intelligence".
        - Argued Why can not Machine Thinks?
        - Coined a Term "**Turing Test**".
    - **1955/56-Dartmouth College and First AI Conference**
        - John McCarthy and his team hosted the first **AI conference** and coined a term "**Artificial Intelligence**".
        - Allen Newell, Cliff Shaw, and Herbert Simon's published a paper titled "**Logic Theorist**".
    - **1957-Frank Rosenblatt**
        - He built and proposed a first working **Artificial Neural Network-Perceptron**.
    - **1966-72-Shakey the Robot**
        - Artificial Intelligence Center of Stanford Research Institute, developed a general purpose mobile robot named Shakey
        - It was able to do the following:
            - Shakey could perform tasks that required **planning, route-finding, and the rearranging of simple objects**.
    - **1974-1980s: AI winter**
    - **1997-Deep Blue.**
        - Supercomputer "Deep Blue" developed by IBM defeated the world champion chess player- **Garry Kasparov.**

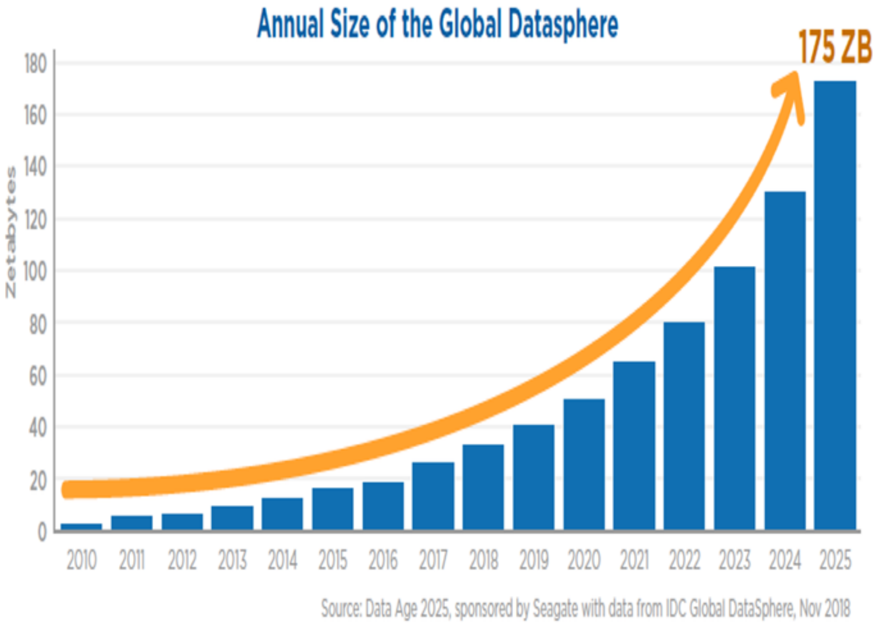
6.2 Some Major Moments in History of AI -2.

- **Post-2000s**
  - **2002-Roomba Vacuum.**
    - The first commercially successful robotic vacuum cleaner was created: “Roomba Vacuum” by **iRobot.**
  - **2011-Neural Net and Cats**
    - Google Engineer **Jeff Dean** and Professor **Andrew Ng** created a **Neural Nets** that trained itself to recognize **“cats”** from **YouTube** videos.
  - **2012-Geoffrey Hinton and Deep Neural Net**
    - He built a Deep Neural Network called **ALEXNET**, which won the **ImageNet** Competition.
  - **2016-AlphaGO beats GO Champion**
    - **Deepmind** built a machine learning system that beat **GO** Champion.
  - **2018-Google Assistant/BERT**
    - **In Google I/O events of 2018 Google Assistant made a booking on saloon.**
  - **2022-chatGPT(GPT4)**
    - **You know it.**

6.3 Emergence of AI-I

- Advancement of Computer storage devices
- Data Everywhere



Thank You any Question!!!