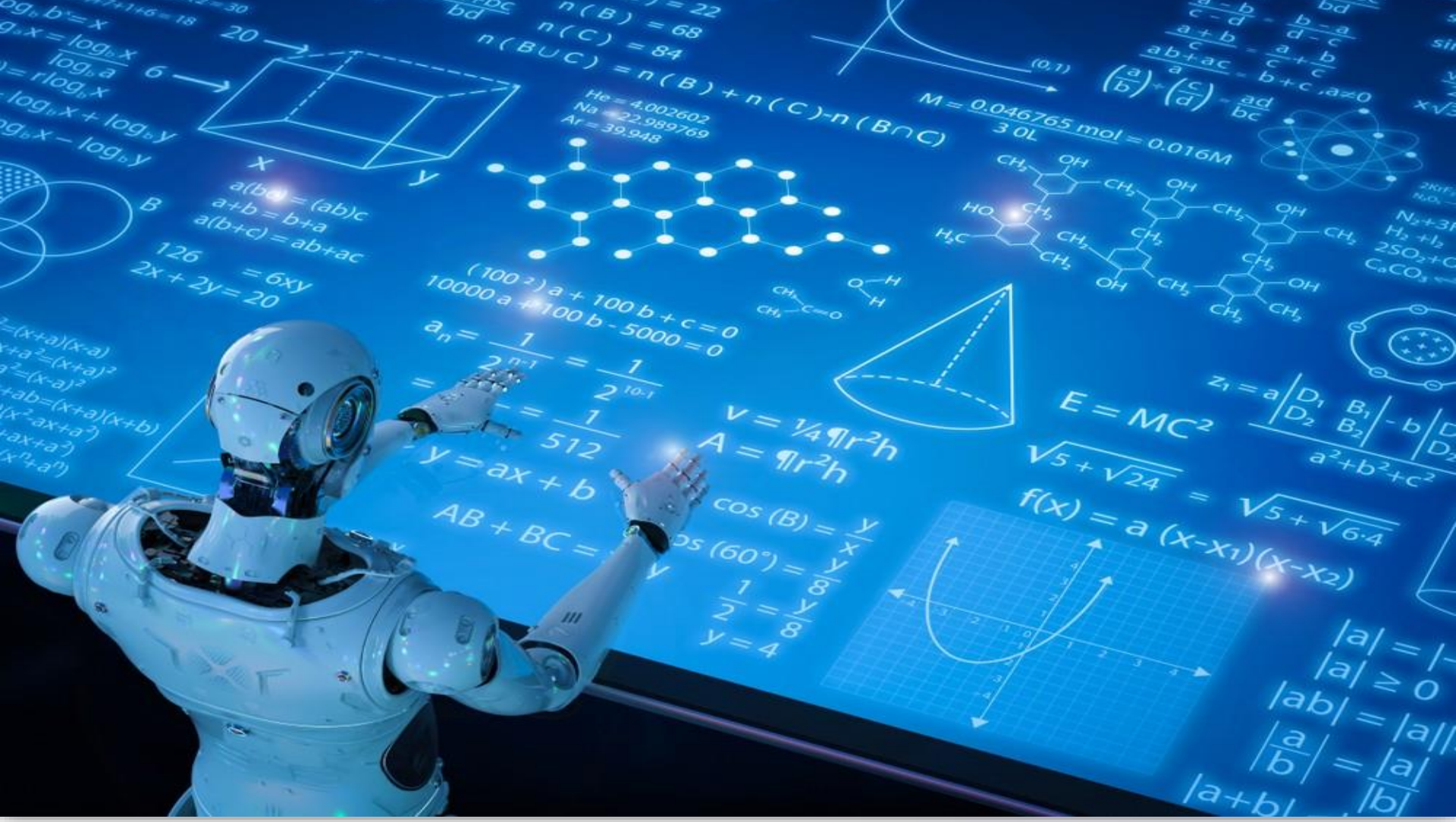


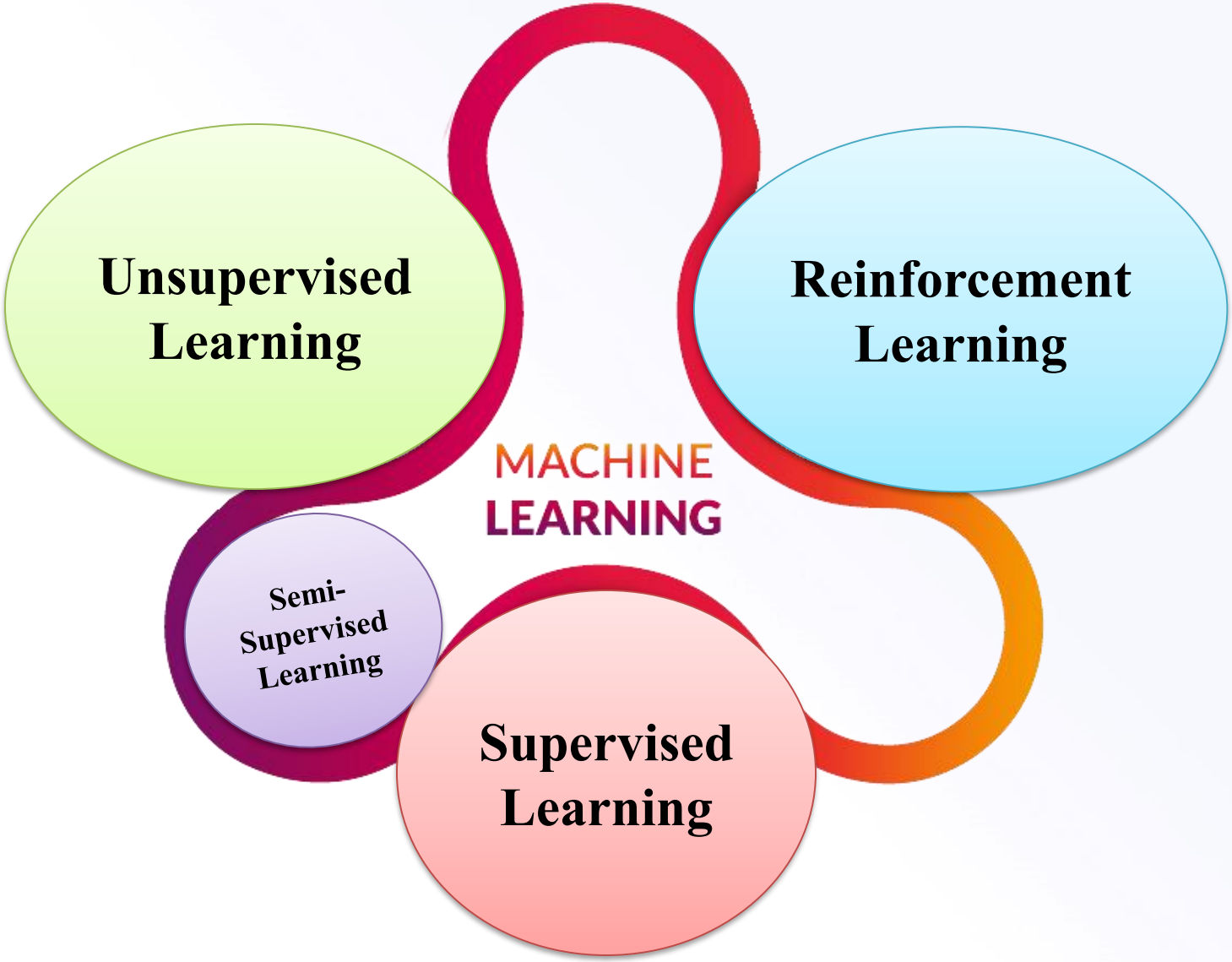
Machine Learning

Md. Jalil Piran, PhD
Asst. Professor
Computer Science and Engineering
Sejong University
Fall, 2020

Outline

- Overview of Machine Learning (ML)
- The History of ML
- Tools
- Types of ML

Types of Learning



Machine learning and our focus

- Like human learning from past experiences.
- A computer does not have “experiences”.
- A computer system learns from **data**, which represent some “past experiences” of an application domain.
- Our focus: learn **a target function** that can be used to predict the values of a discrete class attribute, e.g., approve or not-approved, and high-risk or low risk.
- The task is commonly called: **Supervised learning**, **classification**, or **inductive learning.**

The data and the goal

- **Data**: A set of data records (also called examples, instances or cases) described by
  - k attributes: $A_1, A_2, \ldots A_k$.
  - a class: Each example is labelled with a pre-defined class.

- **Goal**: To learn a classification model from the data that can be used to predict the classes of new (future, or test) cases/instances.

Use-case

Emergency Reception

- An emergency room in a hospital measures 17 variables (e.g., blood pressure, age, etc.) of newly admitted patients.
- **A decision is needed**: whether to put a new patient in an intensive-care unit.
- Due to the high cost of ICU, those patients who may survive less than a month are given higher priority.
- **Problem**: to predict high-risk patients and discriminate them from low-risk patients.

Use-case

Credit Card

- A credit card company receives thousands of applications for new cards. Each application contains information about an applicant,
  - age
  - Marital status
  - annual salary
  - outstanding debts
  - credit rating
  - etc.

- **Problem**: to decide whether an application should approved, or to classify applications into two categories, approved and not approved.

Potential Tasks

(1) **Classification** – assigning a category to each item

(2) **Regression** – predicting a value for each item

(3) **Ranking** – learning to order items.

(4) **Clustering** – partitioning items into homogeneous subsets

(5) **Dimensionality reduction** (manifold learning) – transforming a representation of items into a lower-dimensional one

(6) **Anomaly Detection**

Ingredients of ML

- **Training examples**
- **Feature vectors (patterns)**
- **Labels**
- **Hyperparameters**
  - free parameters not determined by the learning algorithm but specified as input to the learning algorithm
- **Validation sample** – for tuning the hyperparameters Test sample
- **Hypothesis set** – a set of functions mapping to the set of labels

Types of Machine Learning

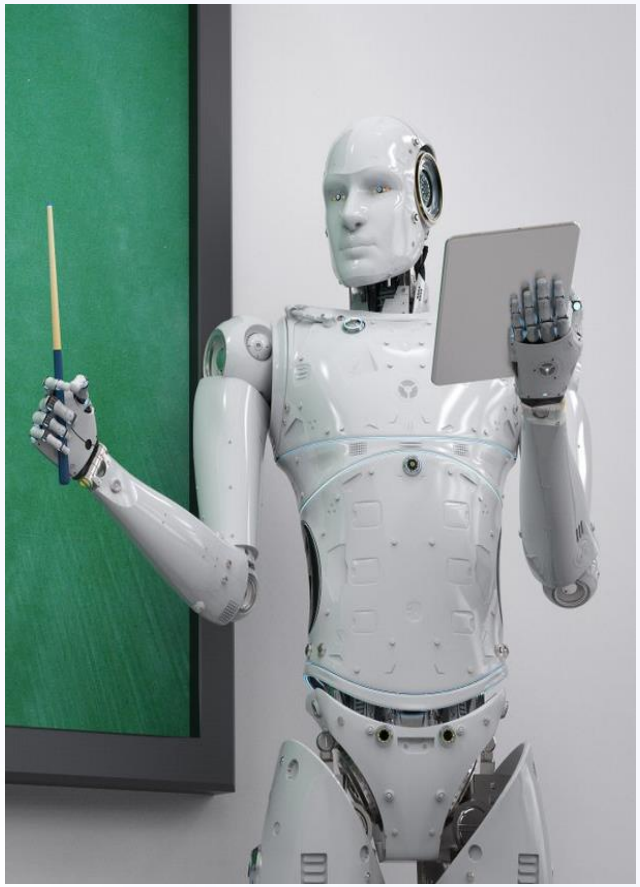
- **Supervised Learning**
  - Learn a mapping from the input to an output using a set of labeled examples
- **Unsupervised Learning**
  - Find the regularities of data using a set of unlabeled examples
- **Semi-supervised Learning**
  - The training sample consisting of both labeled and unlabeled data

Types of Machine Learning

- **Reinforcement Learning**
  - During learning, the correct answers are not provided but hints or delayed rewards
- **On-line Learning**
  - Training and testing phases are intermixed.
- **Active Learning**
  - The learner adaptively or interactively collects training examples

Supervised Learning

- Supervised learning algorithms try to **model** relationships and dependencies between the target prediction output and the input features
- Used to **predict** the output values for new data based on those relationships which it learned from the previous data sets.



Supervised Learning

- The majority of practical machine learning uses supervised learning.

$$Y = f(X)$$

X: input variables

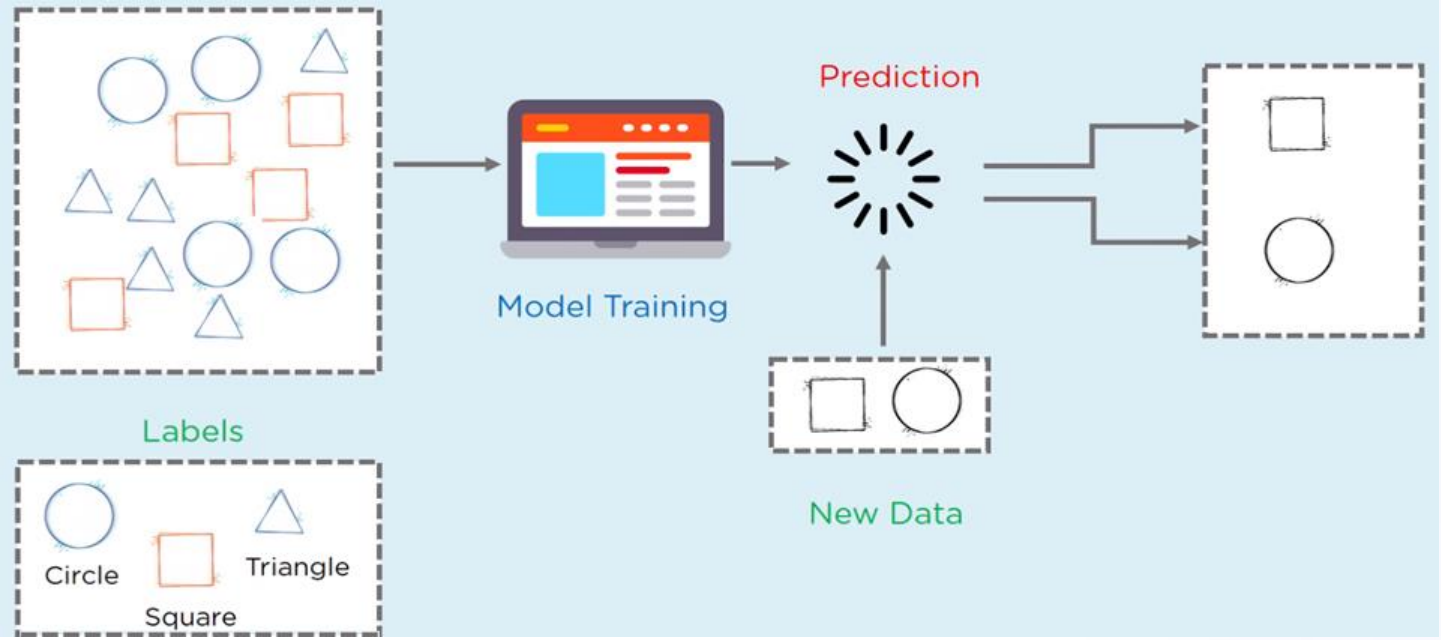
Y: an output variable

$f(.)$: an algorithm to learn the mapping function from the input to the output.

- **Goal**: to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

Supervised Learning

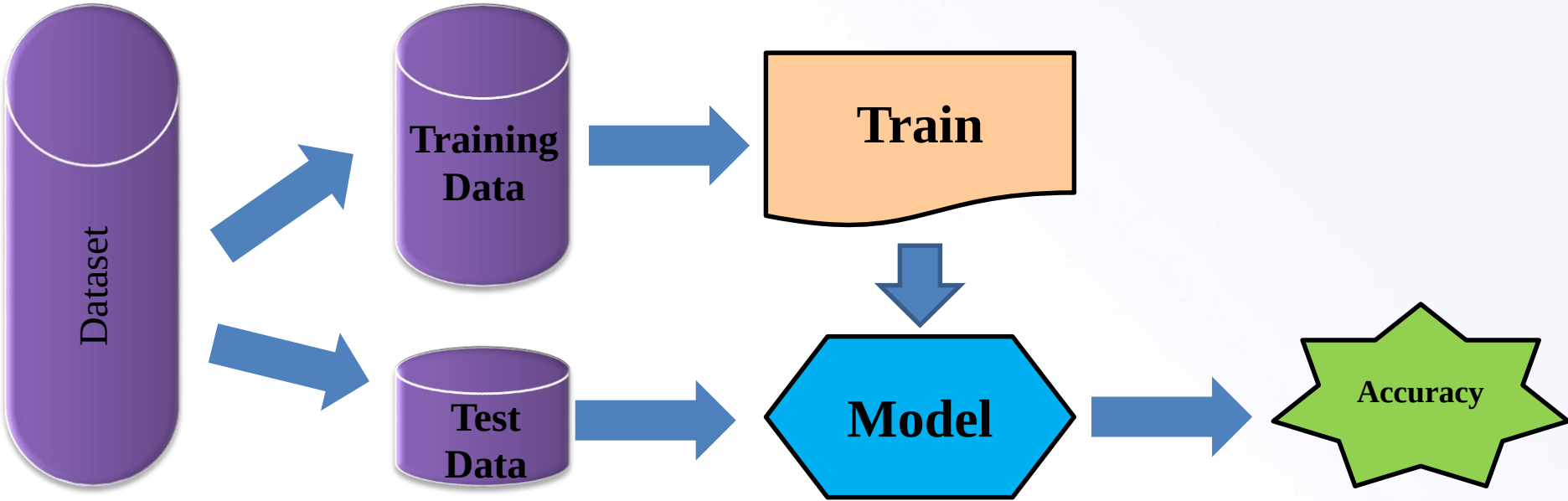
- Supervised Learning enable machines to **classify** / **predict** objects, problems or situations based on labeled data fed to the machine.



Supervised learning process: two steps

- **Learning (training):** Learn a model using the **training data**
- **Testing:** Test the model using **unseen test data** to assess the model accuracy

$$Accuracy = \frac{Number\ of\ correct\ classification}{total\ number\ of\ test\ cases}$$



What is Learning?

Given

a **data set** D,

a **task** T, and

a **performance measure** M,

a computer system is said to **learn** from ***D*** to perform the task ***T*** if after learning the system's performance on ***T*** improves as measured by ***M***.

In other words, the learned model helps the system to perform ***T*** better as compared to no learning.

Supervised Learning

Supervised learning problems

- **Classification**: A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.
- **Regression**: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.
- Some common types of problems built on top of classification and regression include **recommendation** and **time series prediction** respectively.

Supervised Learning

Algorithms

- **Regression**
  - Ordinary Least Squares Regression (OLSR)
  - Linear Regression
  - Logistic Regression
  - Statistical Logistic Regression
  - Stepwise Regression
  - Multivariate Adaptive Regression Splines (MARS)
  - Locally Estimated Scatterplot Smoothing (LOESS)
- **Random forest** for classification and regression problems
- **Support vector machines** for classification problems
- **Artificial neural networks** (ANN)
- **Deep neural networks (DNN)**
- **Linear discriminant analysis**
- **Decision trees**
- **Similarity learning**
- **Bayesian logic**
- **Supervised classifier**
- **Probabilistic Learning**

Supervised Learning

Which algorithm?

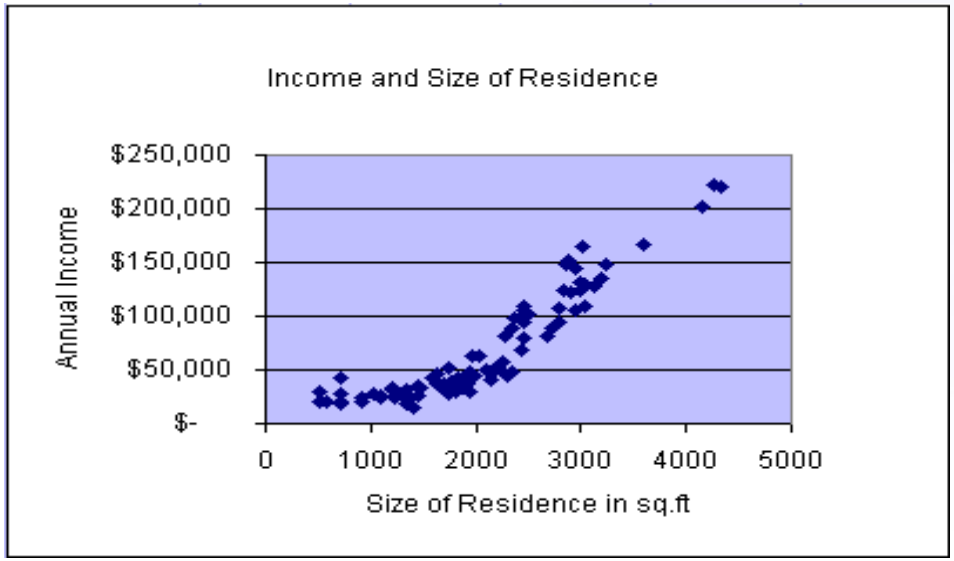
- The first is the bias and variance that exists within the algorithm as there is a fine line between being flexible enough and too flexible.

- Another is the complexity of the model or function that the system is trying to learn. Additionally, the heterogeneity, accuracy, redundancy and linearity of the data should be analyzed before choosing an algorithm.

Supervised Learning

Linear Regression

- Regression maps an input to an output based on example input-output pairs,
- It predicts continuous valued output.
- The Regression analysis is the statistical model which is used to predict the numeric data instead of labels.
- It can also identify the distribution trends based on the available data or historic data.



Supervised Learning

Classification

- Given a collection of records (**training set**)
  - Each record contains a set of **attributes**, one of the attributes is the **class**.
- Find a **model** for class attribute as a function of the values of other attributes.
- Goal: previously unseen records should be assigned a class as accurately as possible.
  - A **test set** is used to determine the accuracy of the model.
  - Usually, the given data set is divided into training and test sets
  - with **training set** used to **build the model** and **test set** used to **validate it**.

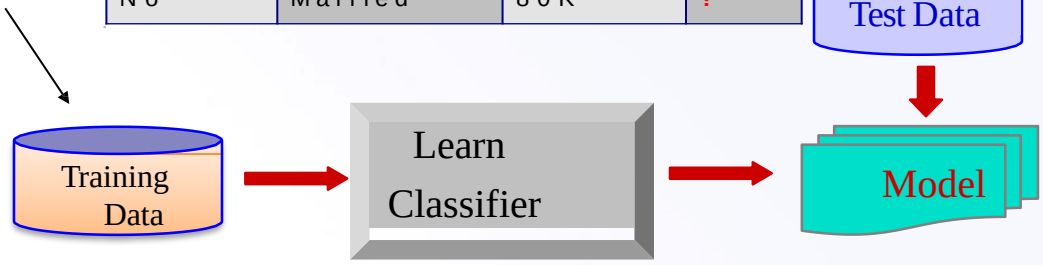
Supervised Learning

Classification

categorical categorical continuous class

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Refund	Marital Status	Taxable Income	Cheat
No	Single	75K	?
Yes	Married	50K	?
No	Married	150K	?
Yes	Divorced	90K	?
No	Single	40K	?
No	Married	80K	?



Supervised Learning

Many classifiers to choose;

- Logistic regression
- Neural networks
- Naïve Bayes
- Bayesian network
- SVM
- Randomized Forests
- Boosted Decision Trees
- K-nearest neighbor
- RBMs
- …

Supervised Learning

Classification:

Application 1

- **Direct Marketing**
  - **Goal**: Reduce cost of **mailing** by **targeting** a set of consumers likely to buy a new cell-phone product.
  - **Approach**:
    - Use the data for a similar product introduced before.
    - We know which customers decided to buy and which decided otherwise. This *{buy, don't buy}* decision forms the *class attribute*.
    - Collect various demographic, lifestyle, and company-interaction related information about all such customers.
      - Type of business, where they stay, how much they earn, etc.
    - Use this information as input attributes to learn a classifier model.

Supervised Learning

Classification:

Application 2

- **Fraud Detection**
  - **Goal**: Predict fraudulent cases in credit card transactions.
  - **Approach**:
    - Use credit card transactions and the information on its account-holder as attributes.
      - When does a customer buy, what does he buy, how often he pays on time, etc.
    - Label past transactions as fraud or fair transactions. This forms the class attribute.
    - Learn a model for the class of the transactions.
    - Use this model to detect fraud by observing credit card transactions on an account.

Supervised Learning

Summary

- Supervised learning is best suited to problems where there is a **set of available reference points** or a **ground truth** with which to train the algorithm.

But those aren't always available!!!

Unsupervised Learning

- UL algorithms try to use techniques on the input data to **mine for rules**, **detect patterns**, and **summarize and group the data points** which help in deriving meaningful insights and describe the data better to the users.

Learn patterns from (unlabeled) data.

Unsupervised Learning

- **Unsupervised learning** is where you only have input data (X) and no corresponding output variables.

- The goal for unsupervised learning is to **model** the underlying structure or distribution in the data in order to learn more about the data.

Unsupervised Learning

Problems

- **Clustering**: where you want to discover the inherent groupings in the data, such *as grouping customers by purchasing behavio*r.

- **Association**: where you want to discover rules that describe large portions of your data, *such as people that buy X also tend to buy Y*.

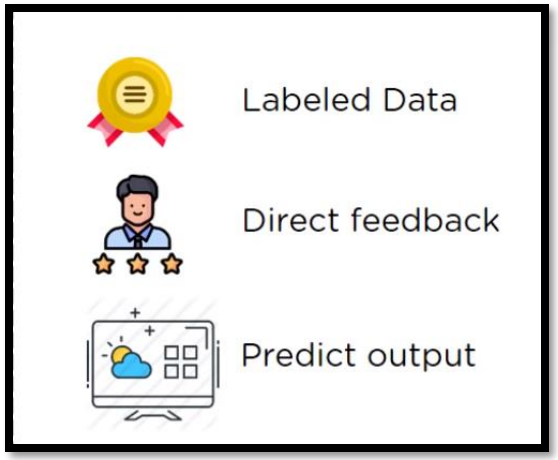
Unsupervised Learning

Approaches

- Clustering (similarity-based)
- Density estimation (e.g., EM algorithm)
- Performance Tasks
- Understanding and visualization
- Anomaly detection
- Information retrieval
- Data compression

Comparison

Supervised vs. Unsupervised



Unsupervised Learning

Algorithms

- K-means clustering,
- Hierarchical clustering,
- Unsupervised soft-clustering,
- Affinity propagation clustering
- Self-organizing map learning
- Autoencoders
- Adversarial autoencoders
- Non-parametric Bayesian Learning
- Generative Deep Neural networks (GDNN)
- Apriori,
- PCA
- Mixture model
- Gaussian mixture model, Expectation Maximization (EM)
- Dirichlet process

Unsupervised Learning

Clustering

- Clustering is the task of partitioning the dataset into groups, called clusters.
- The goal is to split up the data in such a way that points within single cluster are very similar and points in different clusters are different.
- It determines grouping among unlabeled data.
- Given **a set of data points**, each having a **set of attributes**, and a **similarity measure among them**, find clusters such that
  - Data points in one cluster are more similar to one another.
  - Data points in separate clusters are less similar to one another.
- Similarity Measures:
  - **Euclidean Distance** if attributes are continuous.
  - Other Problem-specific Measures.

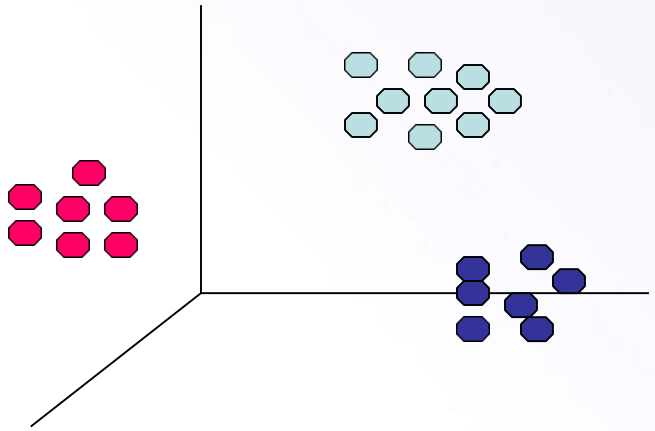
Unsupervised Learning

Clustering

- **Euclidean Distance-**based Clustering in 3-D space.

Intracluster distances are minimized

Intercluster distances are maximized



Unsupervised Learning

Clustering Strategies

- **K-means**
  - Iteratively re-assign points to the nearest cluster center.
- **Agglomerative clustering**
  - Start with each point as its own cluster and iteratively merge the closest clusters.
- **Mean-shift clustering**
  - Estimate modes of probability density function.
- **Spectral clustering**
  - Split the nodes in a graph based on assigned links with similarity weights.

Unsupervised Learning

Clustering:
Application 1

- **Market Segmentation:**
  - **Goal**: subdivide a market into distinct subsets of customers where any subset may conceivably be selected as a market target to be reached with a distinct marketing mix.

  - **Approach**:
    - Collect different attributes of customers based on their geographical and lifestyle related information.
    - Find clusters of similar customers.
    - Measure the clustering quality by observing buying patterns of customers in same cluster vs. those from different clusters.

Unsupervised Learning

Clustering:
Application 2

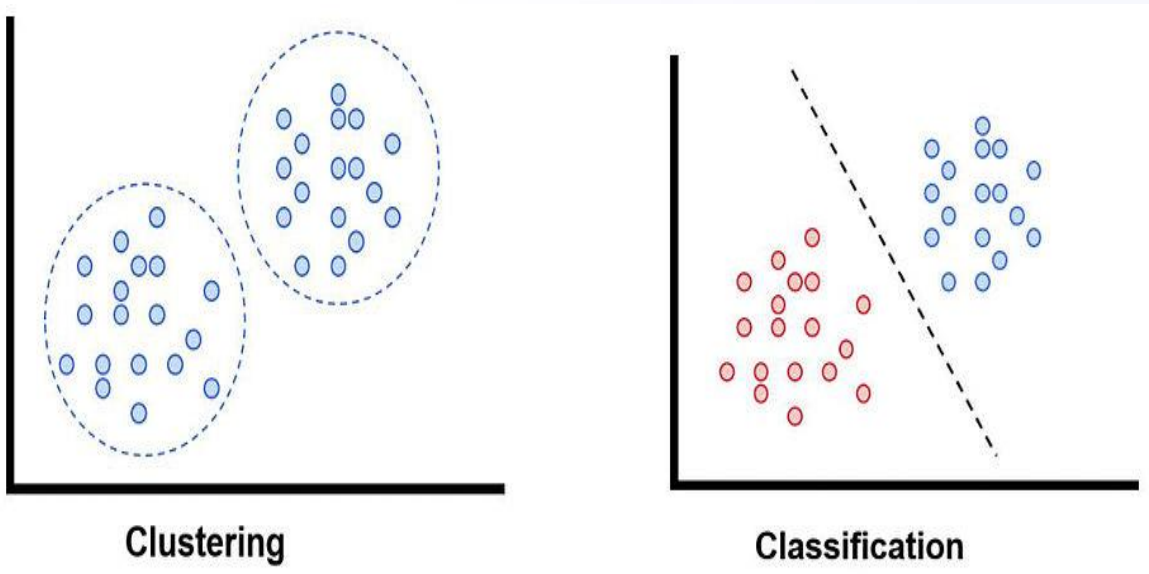
- **Document Clustering**:
  - **Goal**: To find groups of documents that are similar to each other based on the important terms appearing in them.
  - **Approach**: To identify frequently occurring terms in each document. Form a similarity measure based on the frequencies of different terms. Use it to cluster.
  - Gain: Information Retrieval can utilize the clusters to relate a new document or search term to clustered documents.
  - **Gain**: Information retrieval can utilize the clusters to relate a new document or search term to clustered documents.

Comparison

Classification vs. Clustering

- **Classification** is used in supervised learning technique where predefined labels are assigned to instances by properties.

- **Clustering** is used in unsupervised learning where similar instances are grouped, based on their features or properties.

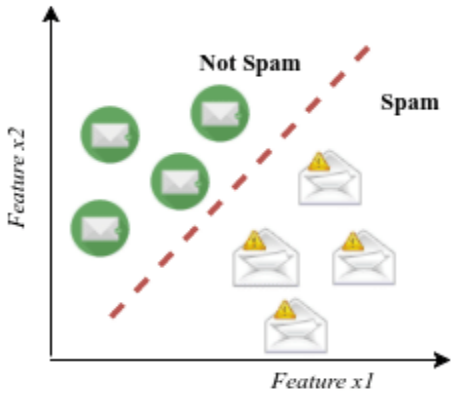


Comparison

Regression	Classification	Clustering
supervised learning	supervised learning	Unsupervised learning
map an input to an output based on example input-output pairs		partitioning the dataset into groups, e.g. clusters
predicts continuous valued output	predicts discrete number of values	Split up the data in such a way that points within single cluster are very similar and points in different clusters are different.
used to predict the numeric data instead of labels.	data is categorized under different labels	determines grouping among unlabeled data.
identify the distribution trends based on the available data or historic data	the labels are predicted for the data.	

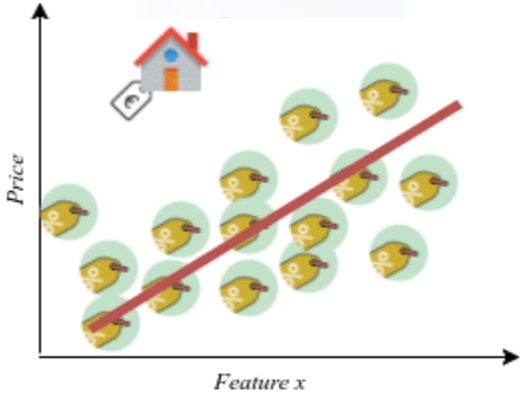
Unsupervised Learning

Classification



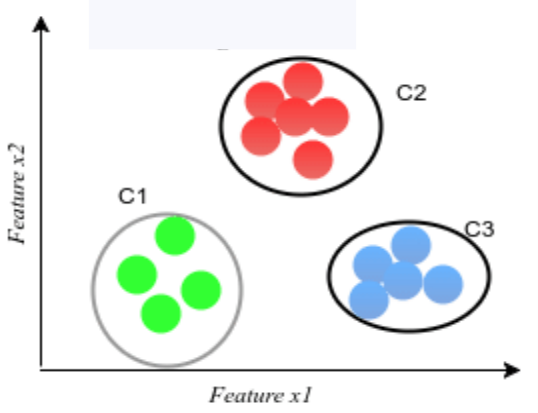
Spam filtering as a classification task

Regression



House price estimation as a regression task

Clustering



customers are grouped into three different categories based on their purchasing behavior.

Unsupervised Learning

Association Rule Discovery

- Given a set of **records** each of which contain some number of items from a given collection;
  - Produce **dependency rules** which will **predict** occurrence of an item based on occurrences of other items.

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Rules Discovered:

{Milk} --> {Coke}

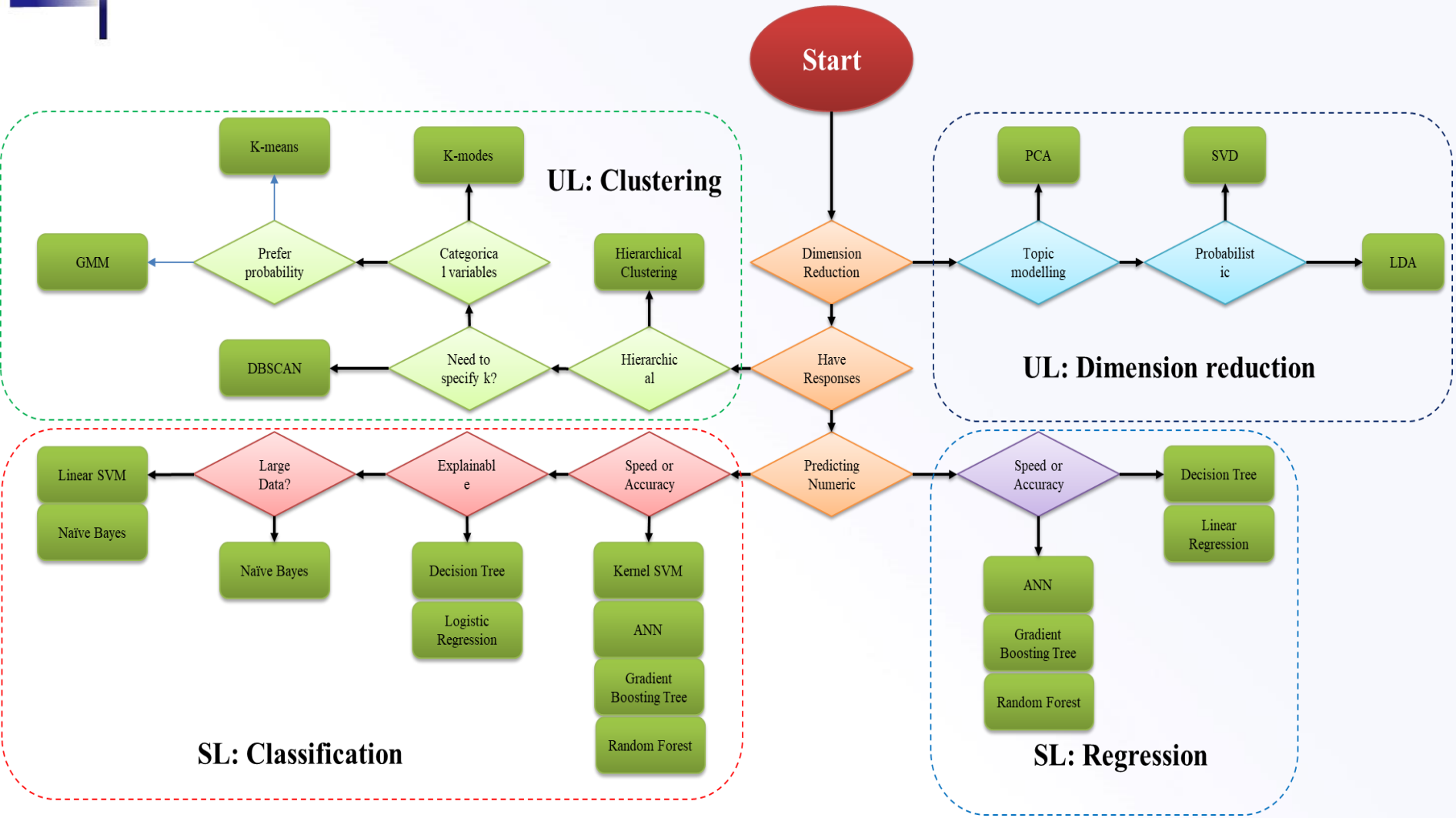
{Diaper, Milk} --> {Beer}

Unsupervised Learning

Association Rule Discovery:
Application 2

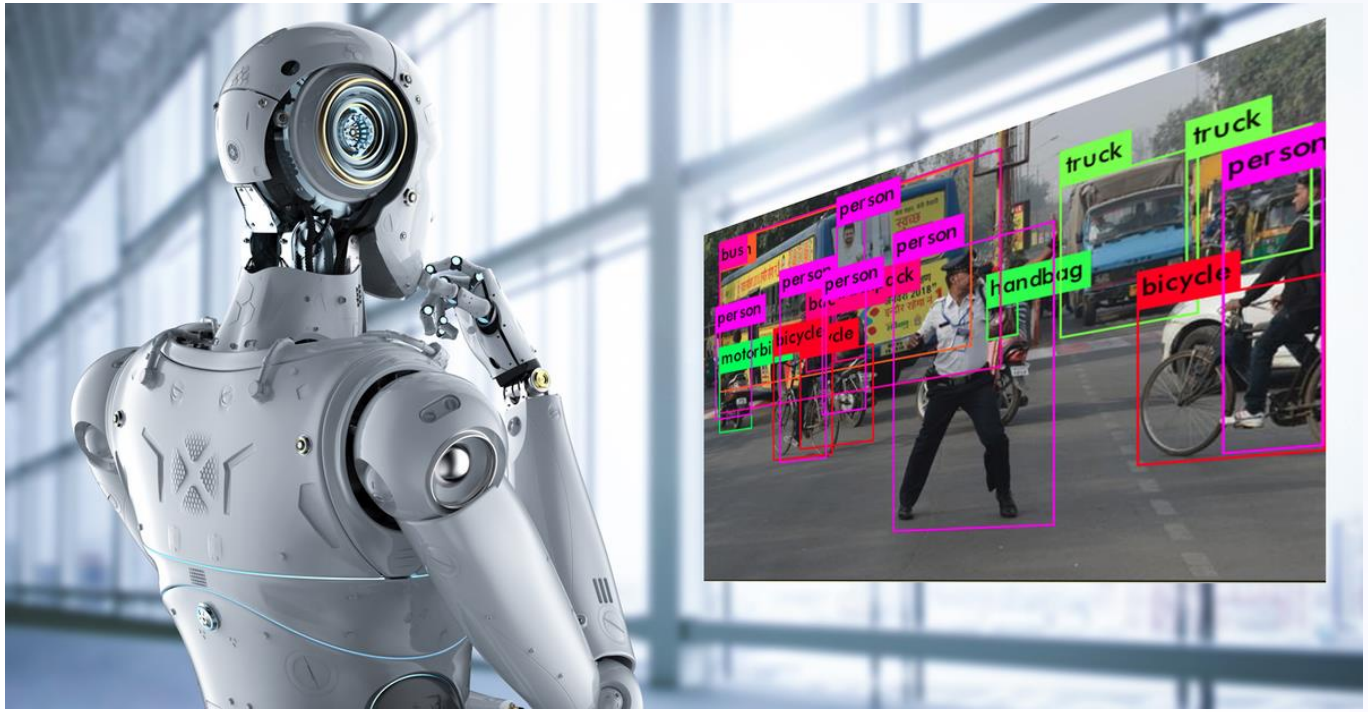
- **Supermarket shelf management**.
  - **Goal**: To identify items that are bought together by sufficiently many customers.
  - **Approach**: Process the point-of-sale data collected with barcode scanners to find dependencies among items.
  - A classic rule --
    - If a customer buys diaper and milk, then he is very likely to buy beer.
    - So, don't be surprised if you find six-packs stacked next to diapers!

SL and USL Cheat-sheet



Semi-supervised Learning

- A training dataset with both **labeled** and **unlabeled** data.
- This method is particularly useful when extracting relevant features from the data is difficult,
- and labeling examples is a time-intensive task for experts.



Semi-supervised Learning

- Problems where you have a large amount of input data (X) and only some of the data is labeled (Y) are called semi-supervised learning problems.

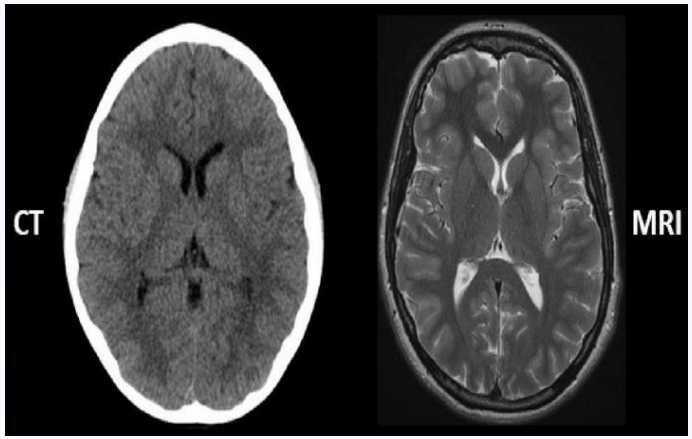
- These problems sit in between both supervised and unsupervised learning.

- A good example is a photo archive where only some of the images are labeled, (e.g. dog, cat, person) and the majority are unlabeled.

Semi-supervised Learning

Applications

- **Medical images**
- e.g. CT scans or MRIs
- A trained radiologist can go through and label a small subset of scans for tumors or diseases.
- It would be too time-intensive and costly to manually label all the scans
- but the deep learning network can still benefit from the small proportion of labeled data and improve its accuracy compared to a fully unsupervised model.



Reinforcement learning

- An agent learns how to behave in an environment by performing actions and seeing the results.

Reinforcement learning

- **Video games** are full of reinforcement cues.
- Complete a level and earn a badge.
- Defeat the bad guy in a certain number of moves and earn a bonus. Step into a trap — game over.
- These cues help players learn how to improve their performance for the next game.
- Without this feedback, they would just take random actions around a game environment in the hopes of advancing to the next level.

Comparison

Supervised Learning

- Labeled dataset
- Establish relationship between input and output
- Generate output for new data points
- Reliable models but expensive and limited

- Classification: Associative classifiers, Decision Trees, Instance Learning, Bayesian Learning, Kernel machines, Neural Networks, Genetic Algorithms, etc.

- Regression: Linear Regression, …

Unsupervised Learning

- Unlabeled dataset
- Decipher structure of the data
- Output attributes are not defined
- Clustering: K-means, DBScan, Hierarchical algorithms, Self Organizing Maps, etc.
- Associations: Apriori, FP-Growth, …

Reinforcement Learning

- Maximizing the rewards from the results
- Aka. credit assessment learning
- Additional decision about rewards
- Explore the tradeoff between exploring and exploiting the data

Reinforcement learning

- Topics:
  - **Policies**: what actions should an agent take in a particular situation
  - **Utility estimation**: how good is a state (→used by policy)
- No supervised output but delayed reward
- Credit assignment problem (what was responsible for the outcome)
- Applications:
  - Game playing
  - Robot in a maze
  - Multiple agents, partial observability, ...

Reinforcement learning

Steps

- In order to produce intelligent programs (also called agents), reinforcement learning goes through the following steps:
  - **Input state** is observed by the agent.
  - **Decision making function** is used to make the agent perform an action.
  - After the action is performed, the agent receives **reward** or reinforcement from the environment.
  - The **state-action pair information** about the reward is stored.

Reinforcement learning

Algorithms

- Q-Learning
- MDP / POMDP
- Temporal Difference (TD)
- Deep Adversarial Networks
- Multi-armed bandit
- Actor-critic
- Deep reinforcement learning (DRL)

ML Algorithms Cheat Sheet

What do you want to do?

Extract information from text

Text Analytics

Derives high-quality information from text

Answers questions like: What info is in this text?

- **Extract N-Gram Features from Text** ← Creates a dictionary of n-grams from a column of free text
- **Feature Hashing** ← Converts text data to integer encoded features using the Vowpal Wabbit library
- **Preprocess Text** ← Performs cleaning operations on text, like removal of stop-words, case normalization
- **Word2Vector** ← Converts words to values for use in NLP tasks, like recommender, named entity recognition, machine translation

Predict between several categories

Multiclass Classification

Answers complex questions with multiple possible answers

Answers questions like: Is this A or B or C or D?

- **Multiclass Logistic Regression** ← Fast training times, linear model
- **Multiclass Neural Network** ← Accuracy, long training times
- **Multiclass Decision Forest** ← Accuracy, fast training times
- **One-vs-All Multiclass** ← Depends on the two-class classifier
- **Multiclass Boosted Decision Tree** ← Non-parametric, fast training times and scalable

Predict between two categories

Two-Class Classification

Answers simple two-choice questions, like yes or no, true or false

Answers questions like: Is this A or B?

- **Two-Class Support Vector Machine** ← Under 100 features, linear model
- **Two-Class Averaged Perceptron** ← Fast training, linear model
- **Two-Class Decision Forest** ← Accurate, fast training
- **Two-Class Logistic Regression** ← Fast training, linear model
- **Two-Class Boosted Decision Tree** ← Accurate, fast training, large memory footprint
- **Two-Class Neural Network** ← Accurate, long training times

Predict values

Regression

Makes forecasts by estimating the relationship between values

Answers questions like: How much or how many?

- **Fast Forest Quantile Regression** ← Predicts a distribution
- **Poisson Regression** ← Predicts event counts
- **Linear Regression** ← Fast training, linear model
- **Bayesian Linear Regression** ← Linear model, small data sets
- **Decision Forest Regression** ← Accurate, fast training times
- **Neural Network Regression** ← Accurate, long training times
- **Boosted Decision Tree Regression** ← Accurate, fast training times, large memory footprint

Generate recommendations

Recommenders

Predicts what someone will be interested in

Answers the question: What will they be interested in?

- **SVD Recommender** ← Collaborative filtering, better performance with lower cost by reducing dimensionality

Discover structure

Clustering

Separates similar data points into intuitive groups

Answers questions like: How is this organized?

- **K-Means** ← Unsupervised learning

Find unusual occurrences

Anomaly Detection

Identifies and predicts rare or unusual data points

Answers the question: Is this weird?

- **One Class SVM** ← Under 100 features, aggressive boundary
- **PCA-Based Anomaly Detection** ← Fast training times

Classify images

Image Classification

Classifies images with popular networks

Answers questions like: What does this image represent?

- **DenseNet** ← High accuracy, better efficiency

Most of the knowledge in the world in the future is going to be extracted by machines and will reside in machines.

Yann LeCun, Director of AI Research, Facebook

Office: #432, Dayang AI Center, **Phone**: 010-8999-8586, **email**: *piran@sejong.ac.kr*