Details of Artificial Intelligence

Dr. Tassadaq Hussain

Assistant Professor Riphah International University

Collaborations:

Microsoft Research and Barcelona Supercomputing Center

Barcelona, Spain

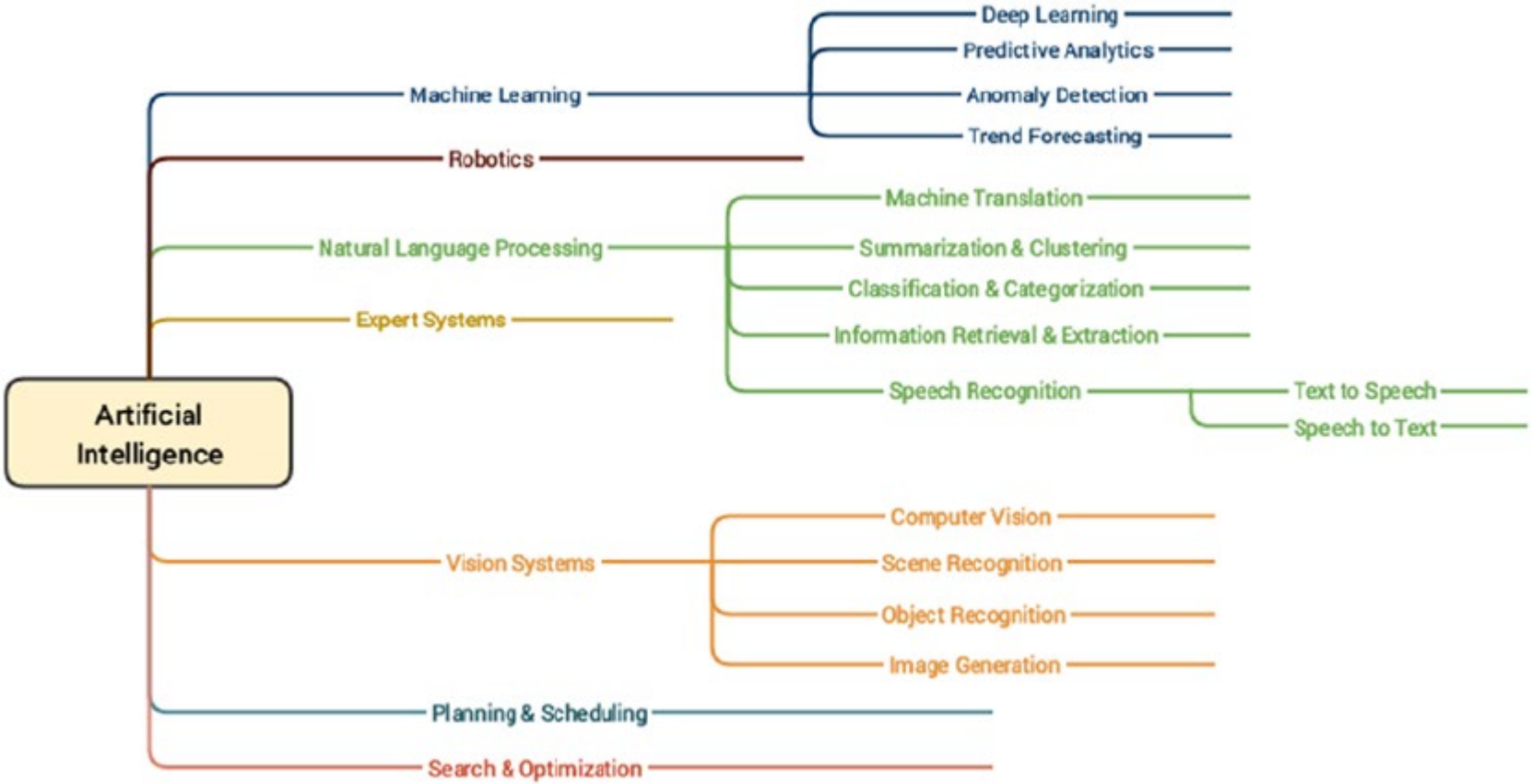
UCERD Pvt Ltd Islamabad

UCERD
Gathering
Intellectuals
www.ucerd.com

Contents

- **Machine Learning Understanding**
- ML Basic Model
- Machine Learning Methods
- Deep Learning
- Data Understanding

UCERD


Artificial Intelligence
Machine Learning
Deep Learning
Predictive Analytics
Anomaly Detection
Trend Forecasting
Robotics
Natural Language Processing
Machine Translation
Summarization & Clustering
Classification & Categorization
Information Retrieval & Extraction
Speech Recognition
Text to Speech
Speech to Text
Expert Systems
Vision Systems
Computer Vision
Scene Recognition
Object Recognition
Image Generation
Planning & Scheduling
Search & Optimization

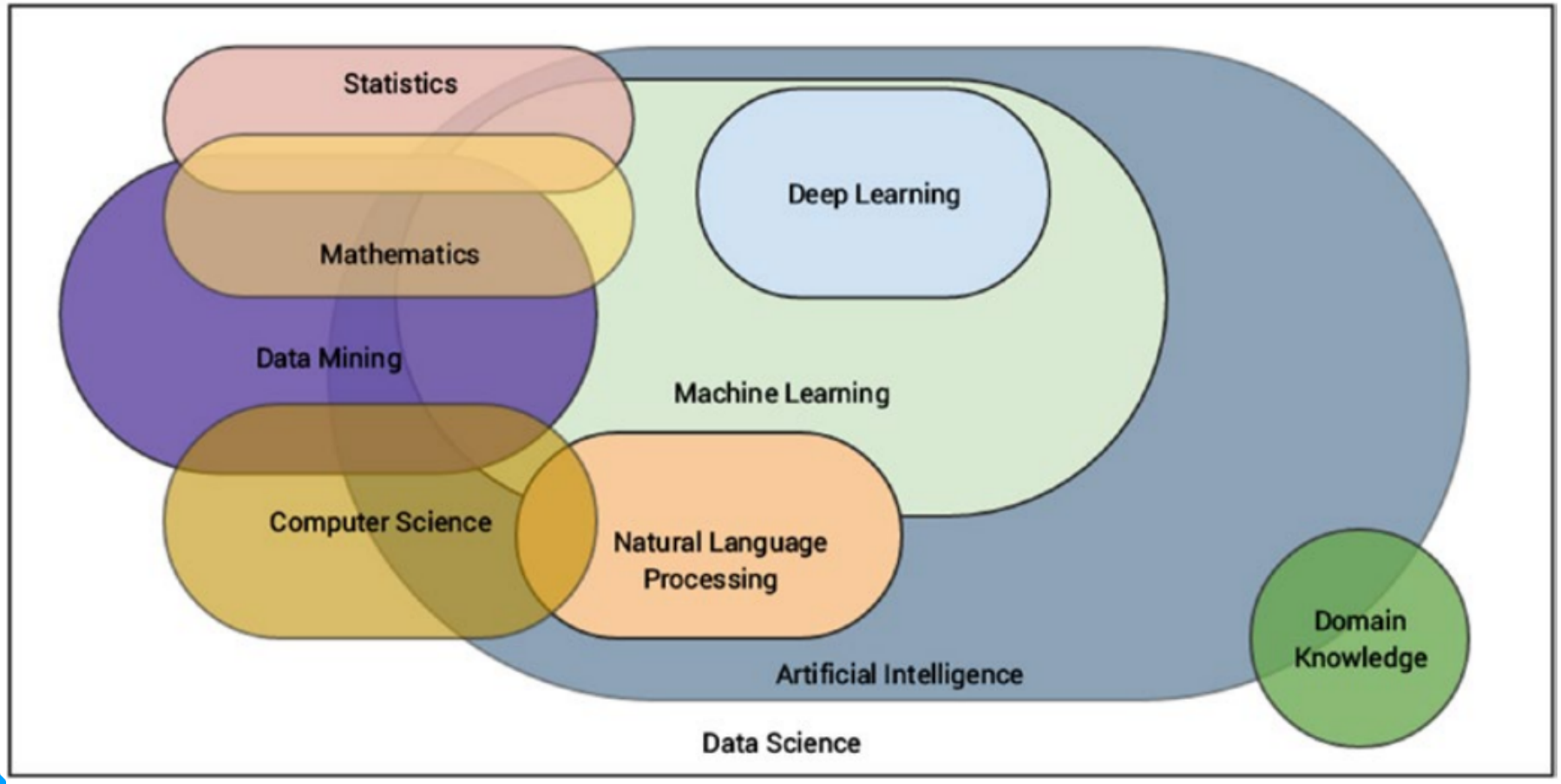
UCERD

Why Machine Learning

- Lack of sufficient human expertise in a domain (e.g., simulating navigations in unknown territories or even spatial planets).
- Scenarios and behavior can keep changing over time (e.g., availability of infrastructure in an organization, network connectivity, and so on).
- Humans have sufficient expertise in the domain but it is extremely difficult to formally explain or translate this expertise into computational tasks (e.g., speech recognition, translation, scene recognition, cognitive tasks, and so on).
- Addressing domain specific problems at scale with huge volumes of data with too many complex conditions and constraints.

UCERD

ML: A Multi-disciplinary Field



UCERD

- Artificial intelligence
- Natural language processing
- Data mining
- Mathematics
- Statistics
- Computer science
- Deep Learning
- Data Science

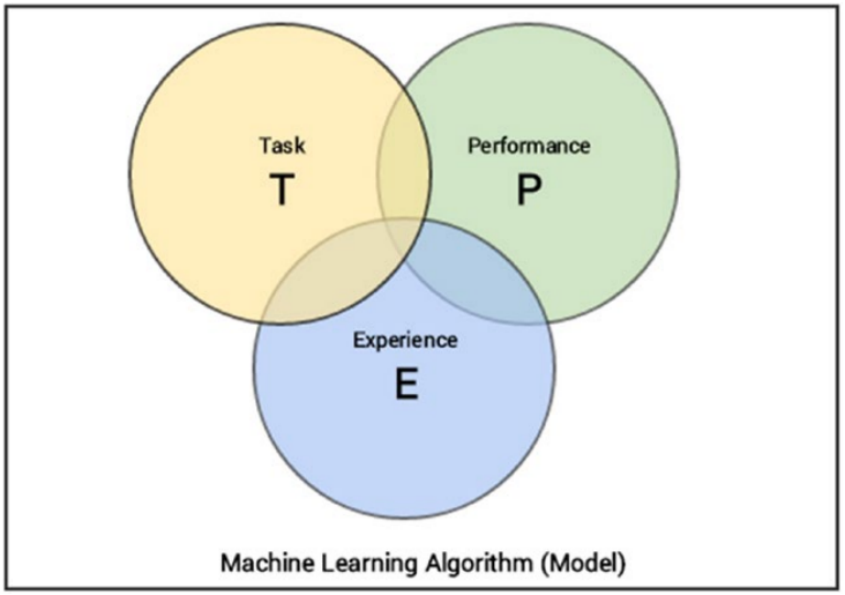
UCERD

Contents

- Machine Learning Understanding
- **ML Basic Model**
- Machine Learning Methods
- Deep Learning
- Data Understanding

UCERD

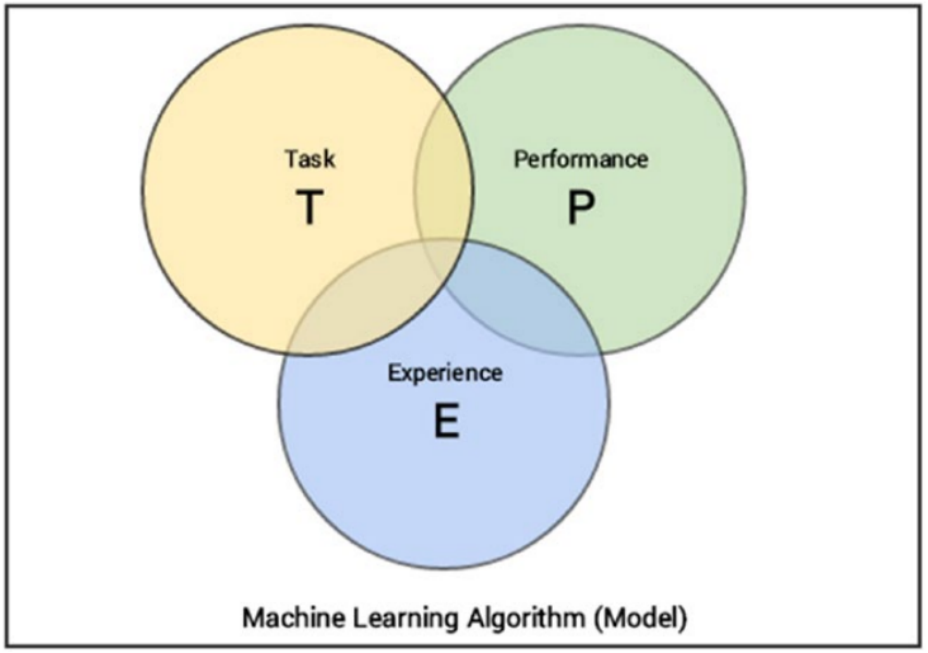
- “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.”



UCERD

A Machine Learning Model

- Classification or categorization
- Regression
- Anomaly detection
- Structured annotation
- Translation
- Clustering or grouping
- Transcriptions



UCERD

Machine Learning Tasks

- Classification or categorization: This typically encompasses the list of problems or tasks where the machine has to take in data points or samples and assign a specific class or category to each sample. A simple example would be classifying animal images into dogs, cats, and zebras.
- Regression: These types of tasks usually involve performing a prediction such that a real numerical value is the output instead of a class or category for an input data point. The best way to understand a regression task would be to take the case of a real-world problem of predicting housing prices considering the plot area, number of floors, bathrooms, bedrooms, and kitchen as input attributes for each data point.
- Anomaly detection: These tasks involve the machine going over event logs, transaction logs, and other data points such that it can find anomalous or unusual patterns or events that are different from the normal behavior. Examples for this include trying to find denial of service attacks from logs, indications of fraud, and so on.

- Structured annotation: This usually involves performing some analysis on input data points and adding structured metadata as annotations to the original data that depict extra information and relationships among the data elements. Simple examples would be annotating text with their parts of speech, named entities, grammar, and sentiment. Annotations can also be done for images like assigning specific categories to image pixels, annotate specific areas of images based on their type, location, and so on.
- Translation: Automated machine translation tasks are typically of the nature such that if you have input data samples belonging to a specific language, you translate it into output having another desired language. Natural language based translation is definitely a huge area dealing with a lot of text data.
- Clustering or grouping: Clusters or groups are usually formed from input data samples by making the machine learn or observe inherent latent patterns, relationships and similarities among the input data points themselves. Usually there is a lack of pre-labeled or pre-annotated data for these tasks hence they form a part of unsupervised Machine Learning (which we will discuss later on). Examples would be grouping similar products, events and entities.
- Transcriptions: These tasks usually entail various representations of data that are usually continuous and unstructured and converting them into more structured and discrete data elements. Examples include speech to text, optical character recognition, images to text, and so on.

Machine Learning Methods

Machine Learning has multiple algorithms, techniques, and methodologies that can be used to build models to solve real-world problems using data. This section tries to classify these Machine Learning methods under some broad categories to give some sense to the overall landscape of Machine Learning methods that are ultimately used to perform specific Machine Learning tasks we discussed in a previous section. Typically the same Machine Learning methods can be classified in multiple ways under multiple umbrellas. Following are some of the major broad areas of Machine Learning methods.

- Supervised learning
- Unsupervised learning
- Semi-supervised learning
- Reinforcement learning

UCERD

Supervised Learning

- Supervised learning methods or algorithms include learning algorithms that take in data samples (known as training data) and associated outputs (known as labels or responses) with each data sample during the model training process. The main objective is to learn a mapping or association between input data samples x and their corresponding outputs y based on multiple training data instances. This learned knowledge can then be used in the future to predict an output y′ for any new input data sample x′ which was previously unknown or unseen during the model training process. These methods are termed as supervised because the model learns on data samples where the desired output responses/labels are already known beforehand in the training phase.

UCERD

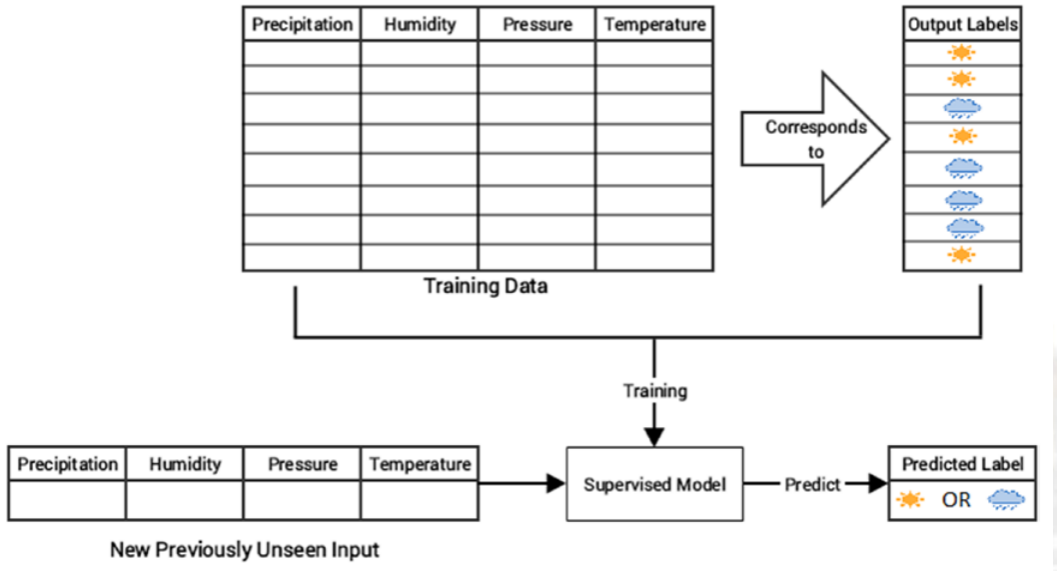
Supervised Learning

- Supervised learning methods or algorithms include learning algorithms that take in data samples (known as training data) and associated outputs (known as labels or responses) with each data sample during the model training process. The main objective is to learn a mapping or association between input data samples x and their corresponding outputs y based on multiple training data instances. This learned knowledge can then be used in the future to predict an output y′ for any new input data sample x′ which was previously unknown or unseen during the model training process. These methods are termed as supervised because the model learns on data samples where the desired output responses/labels are already known beforehand in the training phase.

- Classification
- Regression

UCERD

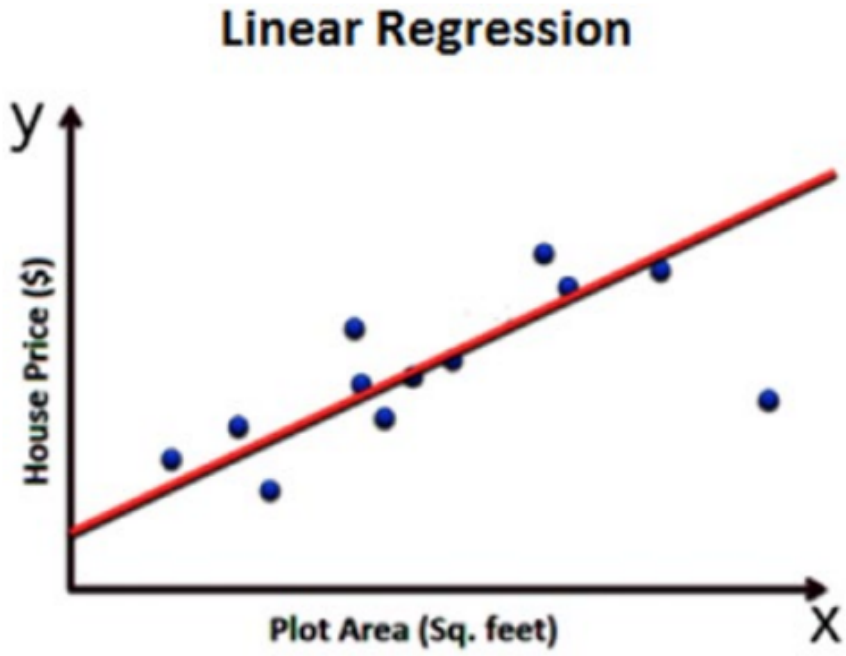
- The classification based tasks are a sub-field under supervised Machine Learning, where the key objective is to predict output labels or responses that are categorical in nature for input data based on what the model has learned in the training phase. Output labels here are also known as classes or class labels are these are categorical in nature meaning they are unordered and discrete values. Thus, each output response belongs to a specific discrete class or category.

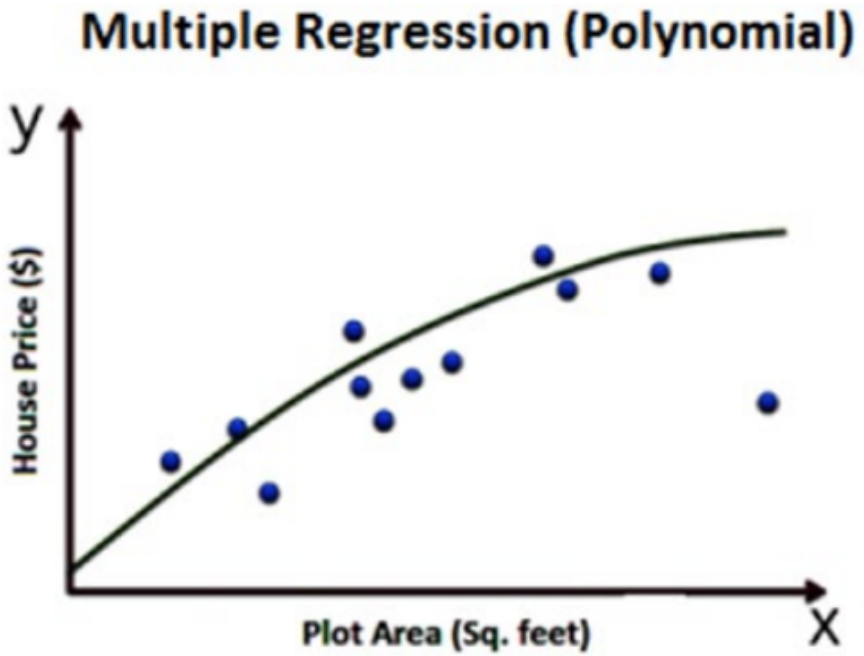


UCERD

ML: Regression

- Machine Learning tasks where the main objective is value estimation can be termed as regression tasks. Regression based methods are trained on input data samples having output responses that are continuous numeric values unlike classification, where we have discrete categories or classes.





Unsupervised Learning

- Unsupervised learning methods usually require some training data where the outcomes which we are trying to predict are already available in the form of discrete labels or continuous values. However, often we do not have the liberty or advantage of having pre-labeled training data and we still want to extract useful insights or patterns from our data. In this scenario, unsupervised learning methods are extremely powerful.
- These methods are called unsupervised because the model or algorithm tries to learn inherent latent structures, patterns and relationships from given data without any help or supervision like providing annotations in the form of labeled outputs or outcomes.

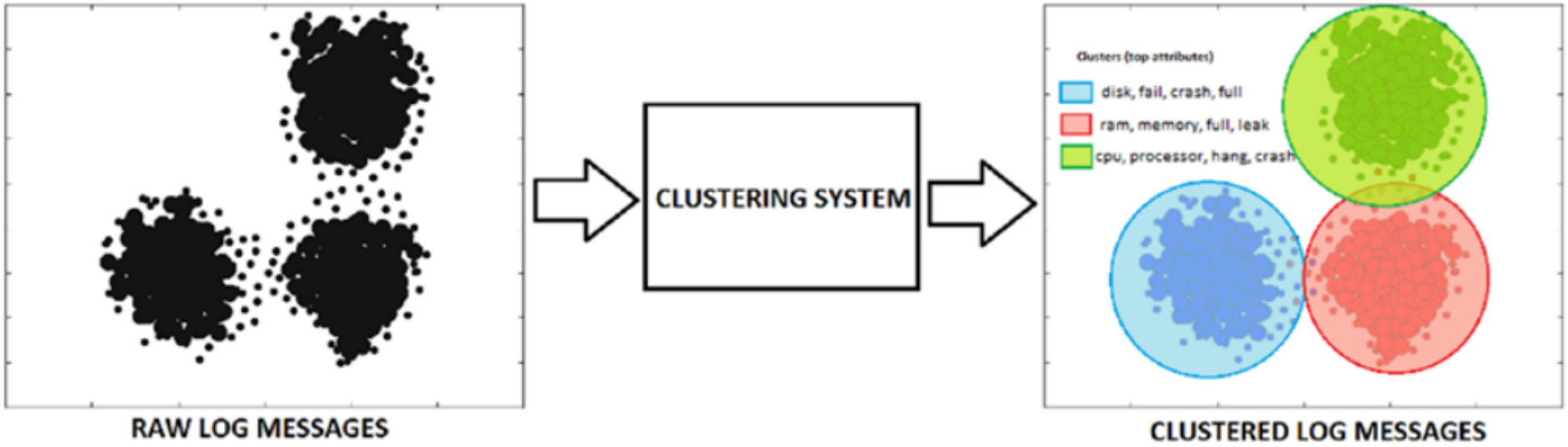
UCERD

- Clustering
- Dimensionality reduction
- Anomaly detection
- Association rule-mining

UCERD

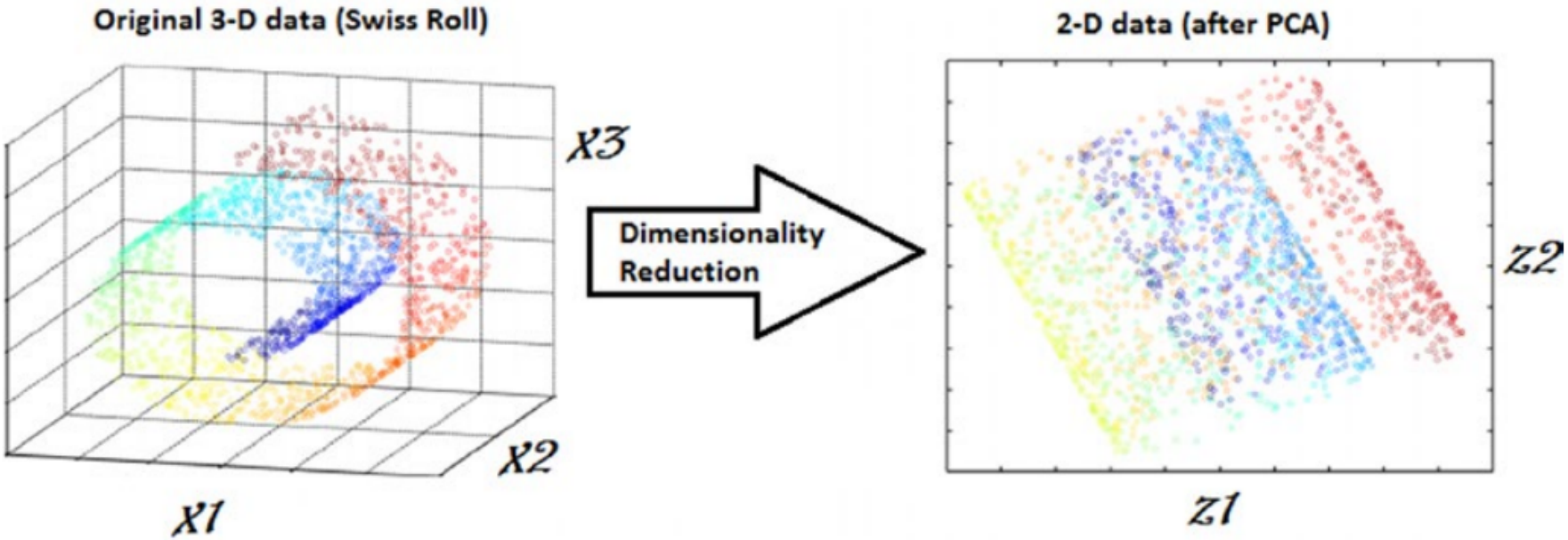
Clustering

- Clustering methods are Machine Learning methods that try to find patterns of similarity and relationships among data samples in our dataset and then cluster these samples into various groups, such that each group or cluster of data samples has some similarity, based on the inherent attributes or features. These methods are completely unsupervised because they try to cluster data by looking at the data features without any prior training, supervision, or knowledge about data attributes, associations, and relationships.



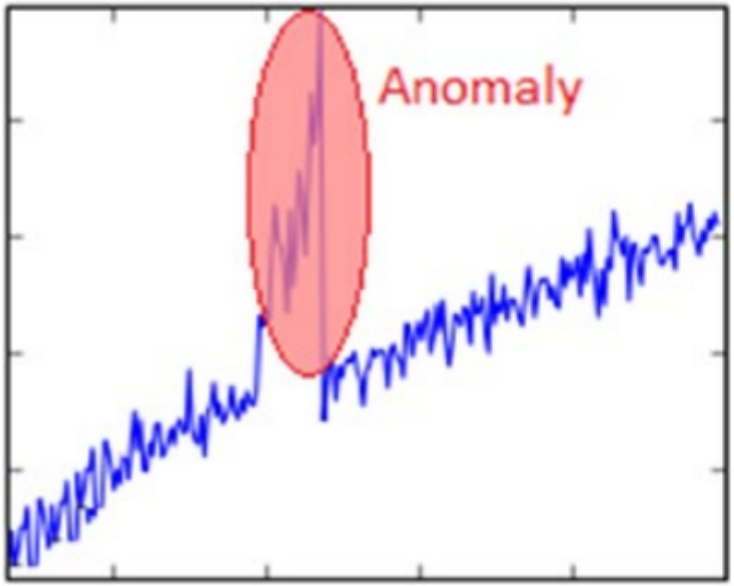
•Dimensionality Reduction

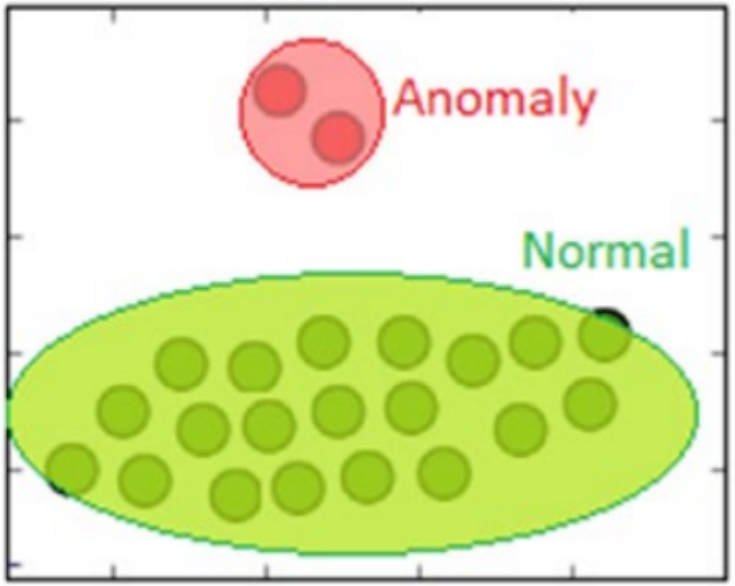
- Dimension reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables.
- Once we start extracting attributes or features from raw data samples, sometimes our feature space gets bloated up with a humongous number of features. This poses multiple challenges including analyzing and visualizing data with thousands or millions of features, which makes the feature space extremely complex posing problems with regard to training models, memory, and space constraints. In fact this is referred to as the “curse of dimensionality”. Unsupervised methods can also be used in these scenarios, where we reduce the number of features or attributes for each data sample.
- These methods reduce the number of feature variables by extracting or selecting a set of principal or representative features. There are multiple popular algorithms available for dimensionality reduction like Principal Component Analysis (PCA), nearest neighbors, and discriminant analysis. Figure shows the output of a typical feature reduction process applied to a Swiss Roll 3D structure having three dimensions to obtain a two-dimensional feature space for each data sample using PCA.



Anomaly Detection

The process of anomaly detection is also termed as outlier detection, where we are interested in finding out occurrences of rare events or observations that typically do not occur normally based on historical data samples. Sometimes anomalies occur infrequently and are thus rare events, and in other instances, anomalies might not be rare but might occur in very short bursts over time, thus have specific patterns.

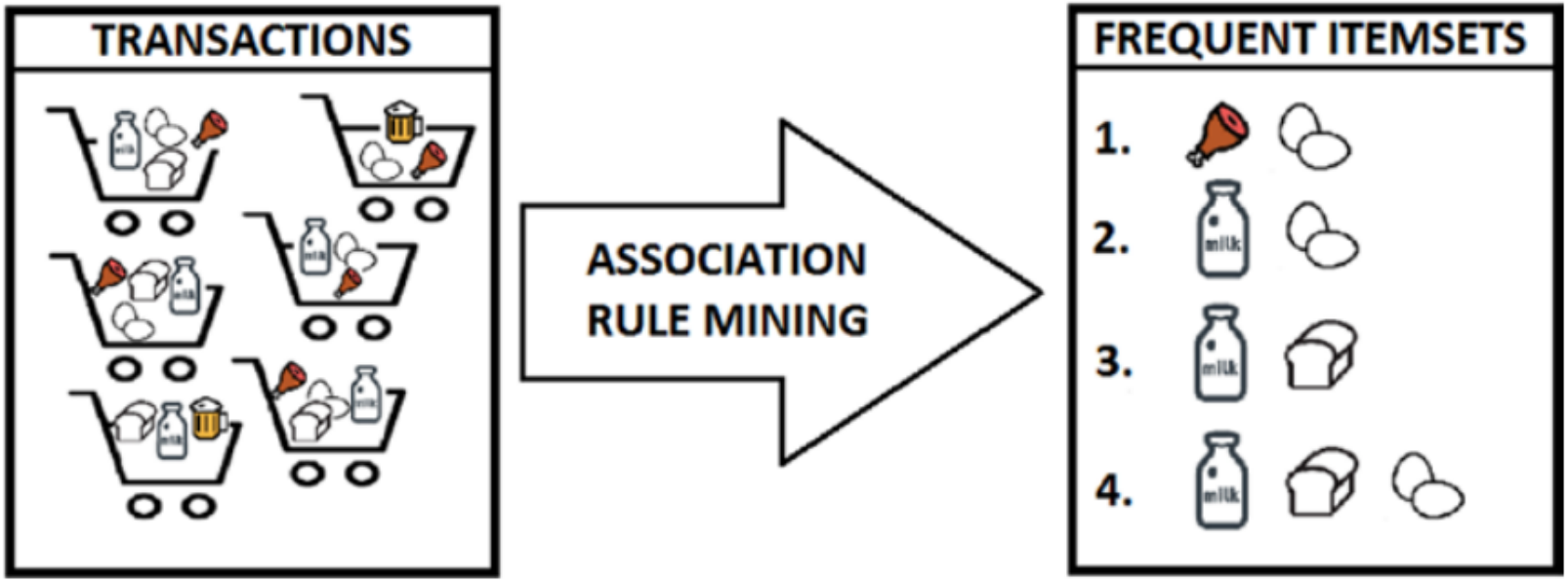




UCERD

Association Rule-Mining

Typically association rule-mining is a data mining method use to examine and analyze large transactional datasets to find patterns and rules of interest. These patterns represent interesting relationships and associations, among various items across transactions. Association rule-mining is also often termed as market basket analysis, which is used to analyze customer shopping patterns.



UCERD

Reinforcement Learning

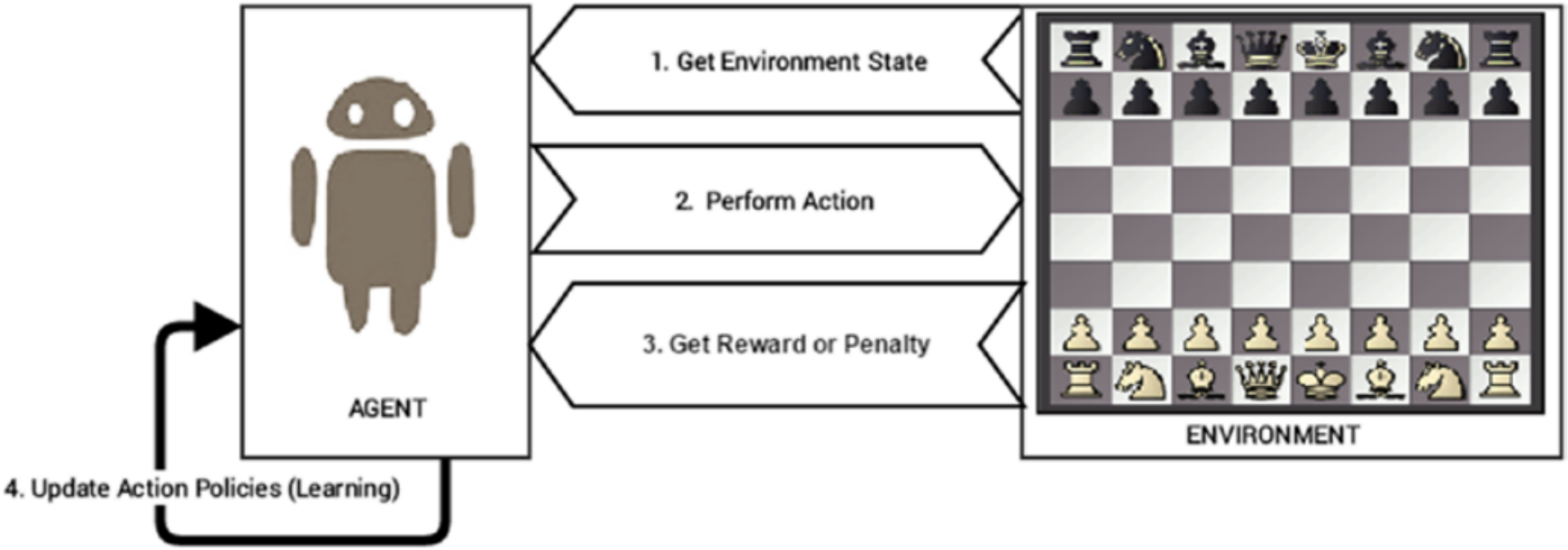
The reinforcement learning methods are a bit different from conventional supervised or unsupervised methods. In this context, we have an agent that we want to train over a period of time to interact with a specific environment and improve its performance over a period of time with regard to the type of actions it performs on the environment. Typically the agent starts with a set of strategies or policies for interacting with the environment. On observing the environment, it takes a particular action based on a rule or policy and by observing the current state of the environment. Based on the action, the agent gets a reward, which could be beneficial or detrimental in the form of a penalty. It updates its current policies and strategies if needed and this iterative process continues till it learns enough about its environment to get the desired rewards.

-

UCERD

- 1. Prepare agent with set of initial policies and strategy
- 2. Observe environment and current state
- 3. Select optimal policy and perform action
- 4. Get corresponding reward (or penalty)
- 5. Update policies if needed
- 6. Repeat Steps 2 - 5 iteratively until agent learns the most optimal policies
-

UCERD


1. Get Environment State
2. Perform Action
3. Get Reward or Penalty
AGENT
ENVIRONMENT
4. Update Action Policies (Learning)

UCERD

Contents

- Machine Learning Understanding
- ML Basic Model
- **Machine Learning Methods**
- Deep Learning
- Data Understanding

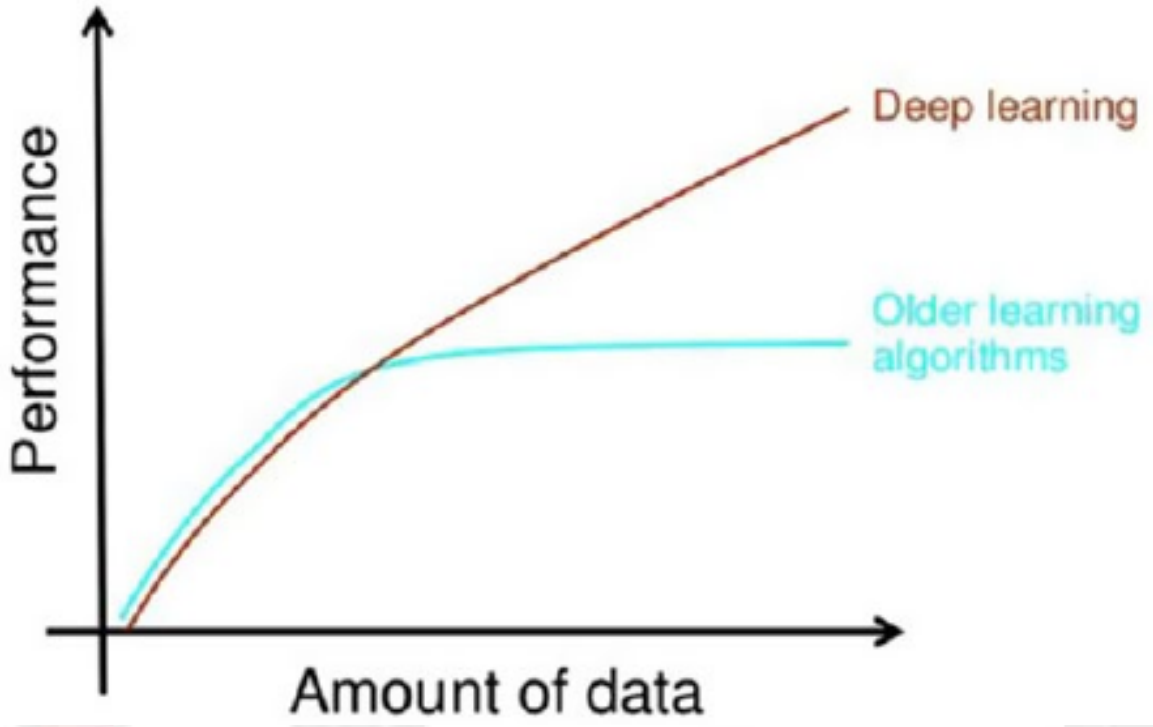
UCERD

Deep Learning

- The field of Deep Learning is a sub-field of Machine Learning that has recently come into much prominence.
- Its main objective is to get Machine Learning research closer to its true goal of “making machines intelligent”.
- Deep Learning is often termed as a rebranded fancy term for neural networks.
- Deep Learning based algorithms involves the use of concepts from representation learning where various representations of the data are learned in different layers that also aid in automated feature extraction from the data.

UCERD

Why deep learning



UCERD

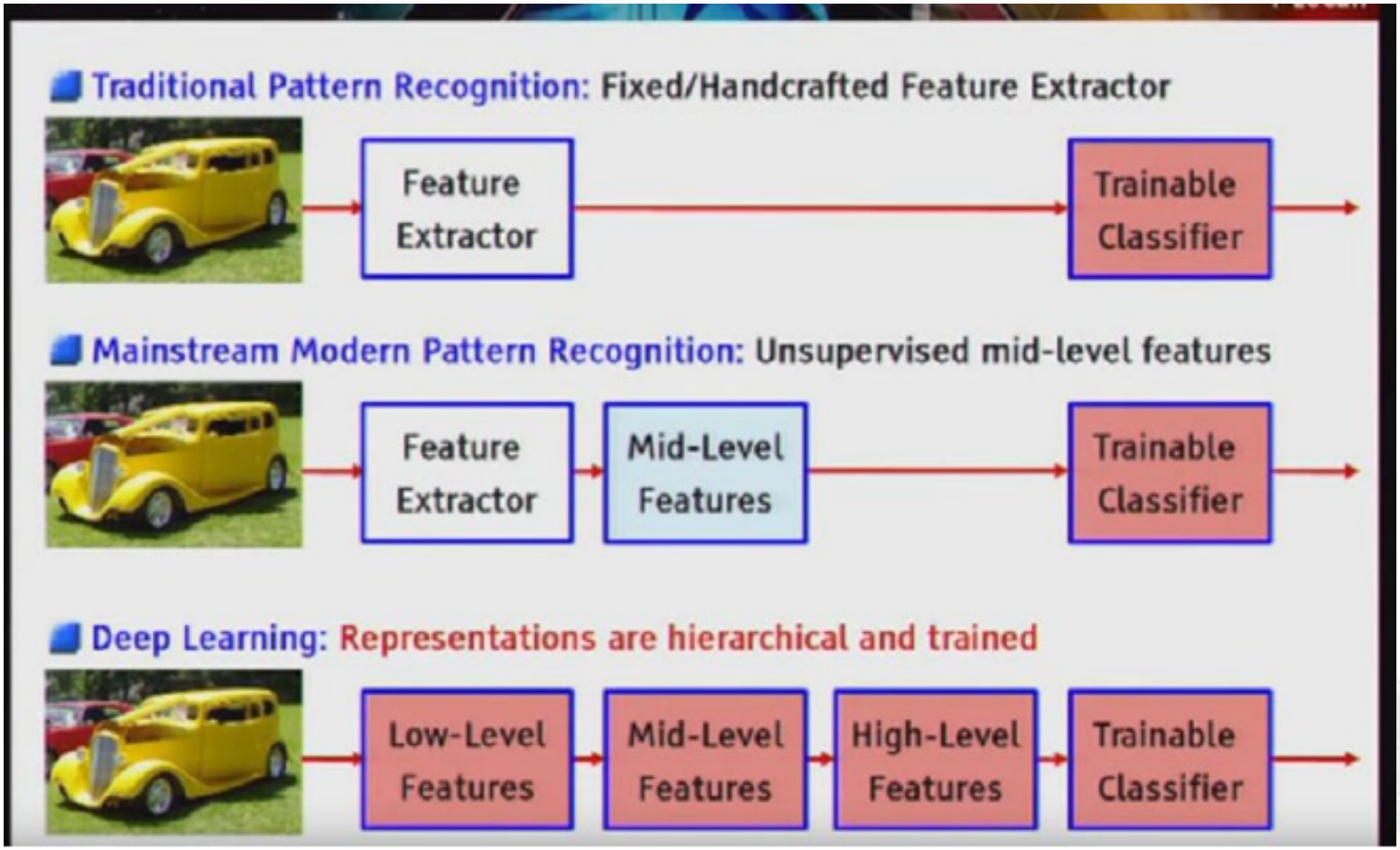
- Deep Learning algorithms are based on distributed representational learning and they start performing better with more data over time.
- Deep Learning could be said to be a rebranding of neural networks, but there is a lot into it compared to traditional neural networks.
- Better software frameworks like tensorflow, theano, caffe, mxnet, and keras, coupled with superior hardware have made it possible to build extremely complex, multi-layered Deep Learning models with huge sizes.
- Deep Learning has multiple advantages related to automated feature extraction as well as performing supervised learning operations, which have helped data scientists and engineers solve increasingly complex problems over time.

UCERD

Salient features of most Deep Learning algorithms

- Hierarchical layered representation of concepts. These concepts are also called features in Machine Learning terminology (data attributes).
- Distributed representational learning of the data happens through a multi-layered architecture (unsupervised learning).
- More complex and high-level features and concepts are derived from simpler, low-level features.
- A “deep” neural network usually is considered to have at least more than one hidden layer besides the input and output layers. Usually it consists of a minimum of three to four hidden layers.
- Deep architectures have a multi-layered architecture where each layer consists of multiple non-linear processing units. Each layer’s input is the previous layer in the architecture. The first layer is usually the input and the last layer is the output.
- Can perform automated feature extraction, classification, anomaly detection, and many other Machine Learning tasks.

UCERD


Traditional Pattern Recognition: Fixed/Handcrafted Feature Extractor
Feature Extractor
Trainable Classifier
Mainstream Modern Pattern Recognition: Unsupervised mid-level features
Feature Extractor
Mid-Level Features
Trainable Classifier
Deep Learning: Representations are hierarchical and trained
Low-Level Features
Mid-Level Features
High-Level Features
Trainable Classifier

UCERD

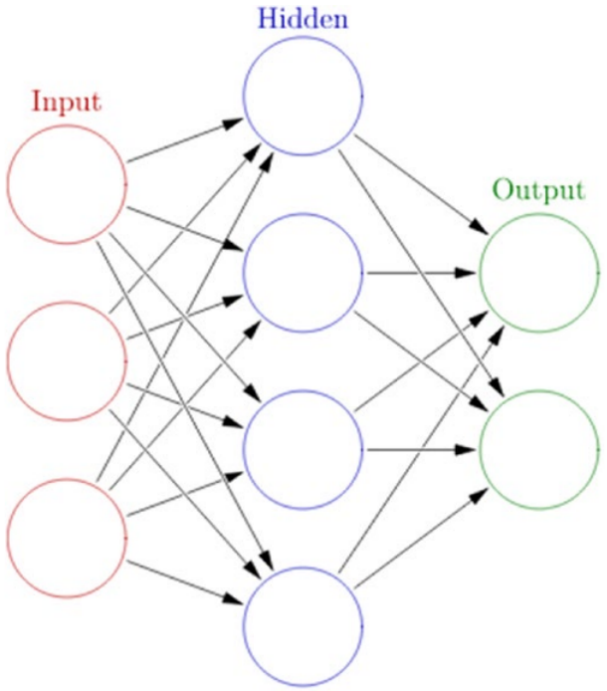
Important Concepts of DL Algorithm

- Artificial Neural Networks
  - Backpropagation
- Multilayer Perceptrons
- Convolutional Neural Networks
- Recurrent Neural Networks
- Long Short-Term Memory Networks
- Autoencoders

UCERD

Artificial Neural Network (ANN)

- An ANN is a computational model and architecture that simulates biological neurons and the way they function in our brain. Typically, an ANN has layers of interconnected nodes.
- The nodes and their inter-connections are analogous to the network of neurons in our brain. A typical ANN has an input layer, an output layer, and at least one hidden layer between the input and output with inter-connections.



UCERD

Backpropagation

- The backpropagation algorithm is a popular technique to train ANNs and it led to a resurgence in the popularity of neural networks in the 1980s. The algorithm typically has two main stages—propagation and weight updates. They are described briefly as follows.

  1. Propagation

     a. The input data sample vectors are propagated forward through the neural network to generate the output values from the output layer.

     b. Compare the generated output vector with the actual/desired output vector for that input data vector.

     c. Compute difference in error at the output units.

     d. Backpropagate error values to generate deltas at each node/neuron.

  2. Weight Update

     a. Compute weight gradients by multiplying the output delta (error) and input activation.

     b. Use learning rate to determine percentage of the gradient to be subtracted from original weight and update the weight of the nodes.

UCERD

Multilayer Perceptrons

A multilayer perceptron, also known as MLP, is a fully connected, feed-forward artificial neural network with at least three layers (input, output, and at least one hidden layer) where each layer is fully connected to the adjacent layer.

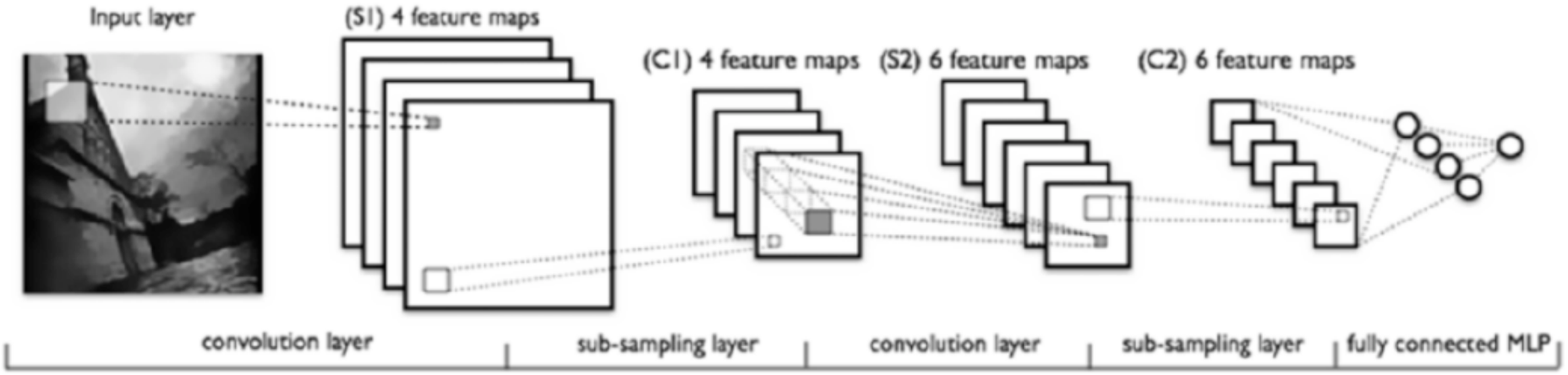
- Each neuron usually is a non-linear functional processing unit. Backpropagation is typically used to train MLPs and even deep neural nets are MLPs when they have multiple hidden layers. Typically used for supervised Machine Learning tasks like classification.

UCERD

Convolutional Neural Networks

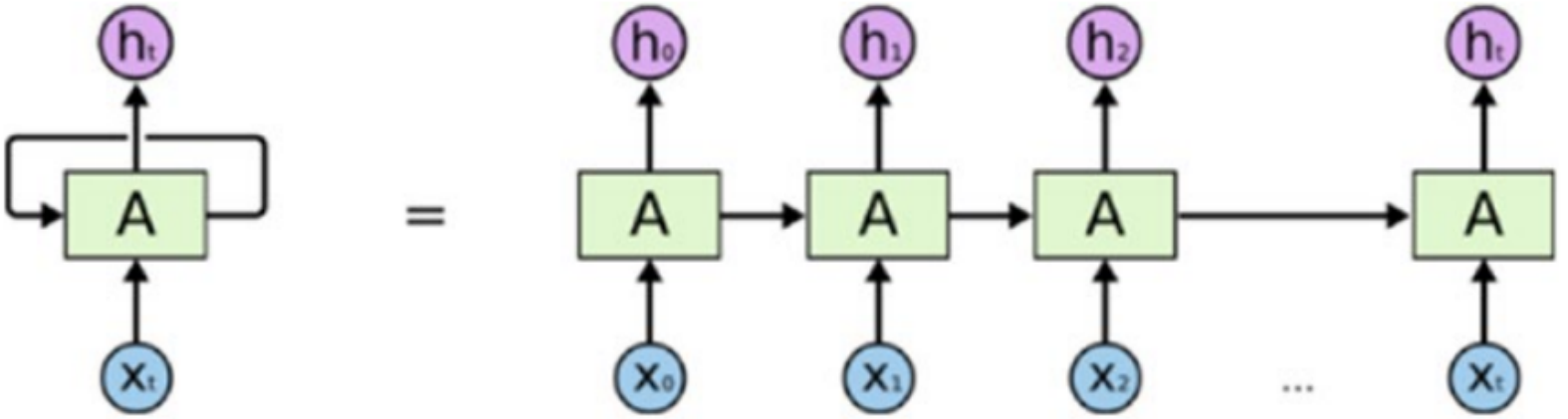
A convolutional neural network, also known as convnet or CNN, is a variant of the artificial neural network, which specializes in emulating functionality and behavior of our visual cortex. CNNs typically consist of the following three components.

- Multiple Convolutional layers
- Pooling layers
- Fully connected MLPs



Recurrent Neural Networks

- A recurrent neural network, also known as RNN, is a special type of an artificial neural network that allows persisting information based on past knowledge by using a special type of looped architecture. They are used a lot in areas related to data with sequences like predicting the next word of a sentence. These looped networks are called recurrent because they perform the sam operations and computation for each and every element in a sequence of input data. RNNs have memory that helps in capturing information from past sequences.



UCERD

Long Short-Term Memory Networks

- RNNs are good in working on sequence based data but as the sequences start increasing, they start losing historical context over time in the sequence and hence outputs are not always what is desired. This is where
- Long Short-Term Memory Networks, popularly known as LSTMs, come into the picture! Introduced by Hochreiter & Schmidhuber in 1997, LSTMs can remember information from really long sequence based data and prevent issues like the vanishing gradient problem, which typically occurs in ANNs trained with backpropagation. LSTMs usually consist of three or four gates, including input, output, and a special forget gate.

UCERD

Autoencoders

- An autoencoder is a specialized Artificial Neural Network that is primarily used for performing unsupervised Machine Learning tasks. Its main objective is to learn data representations, approximations, and encodings.
- Autoencoders can be used for building generative models, performing dimensionality reduction, and detecting anomalies.

UCERD

Contents

- Machine Learning Understanding
- ML Basic Model
- Machine Learning Methods
- Deep Learning
- **Data Understanding**

UCERD

Contents

- Machine Learning Understanding
- ML Basic Model
- Machine Learning Methods
- Deep Learning
- **Data Understanding**

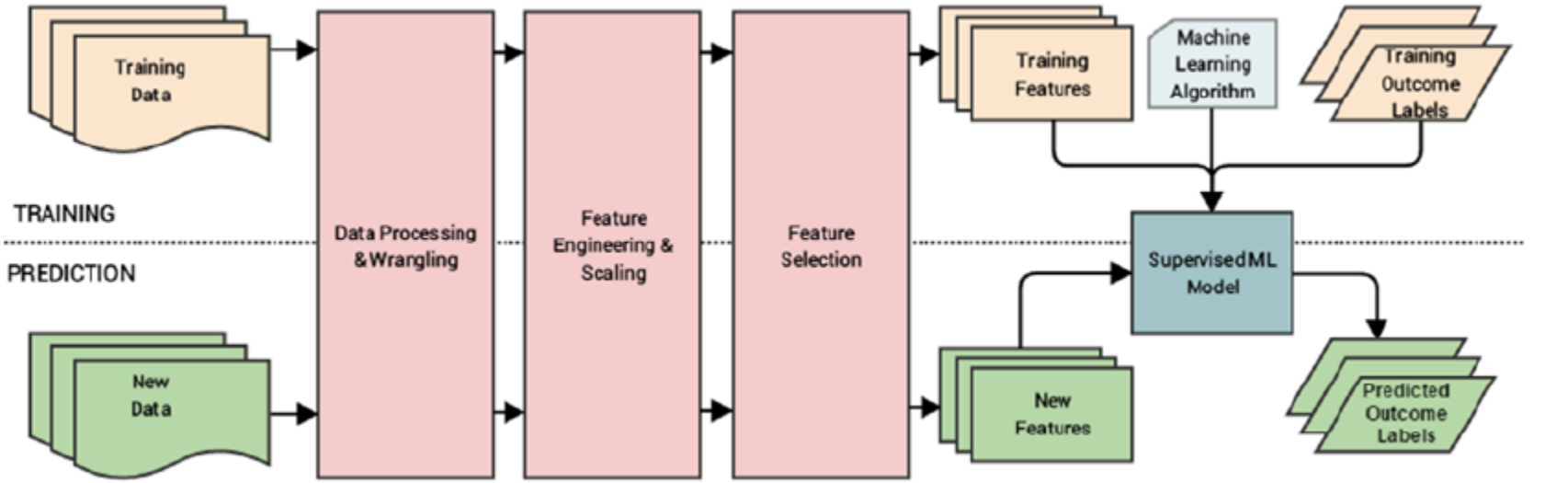
UCERD

Data

- Data Collection
- Data Description
- Exploratory Data Analysis
- Data Quality Analysis
- Data Preparation
- Data Integration
- Data Wrangling

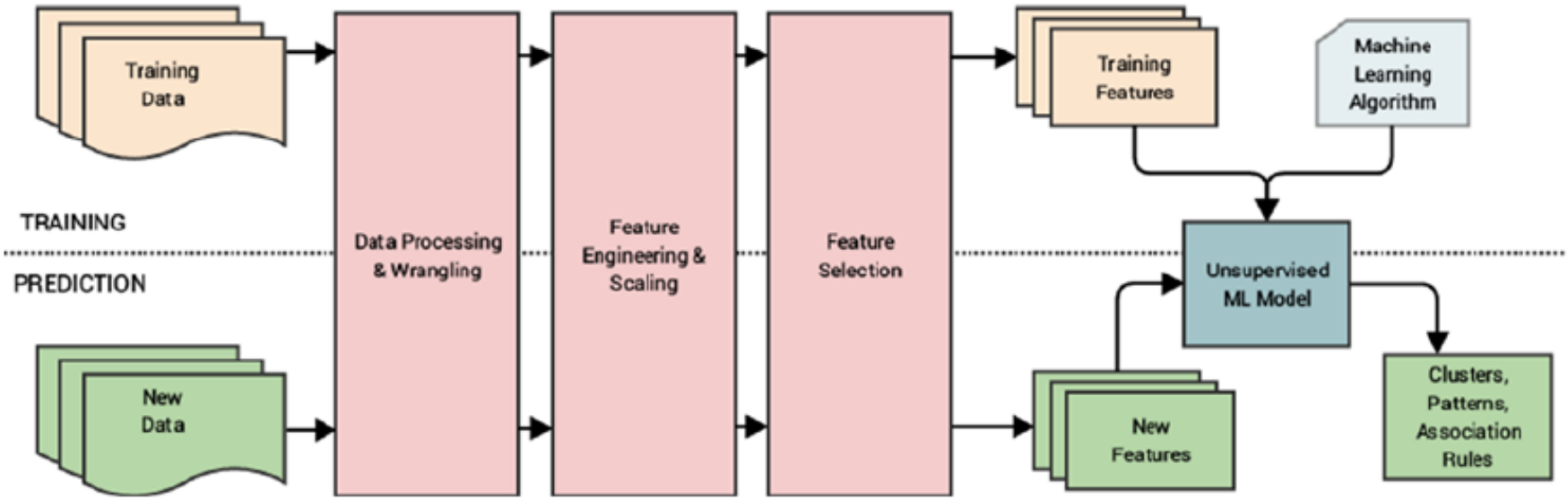
UCERD

Supervised Machine Learning Pipeline



UCERD

Unsupervised Machine Learning Pipeline



UCERD

Datasets: Datatypes

- Scalar
- Vector
- Matrix
- Tensor
  - You can think of a tensor as a generic array. Tensors are basically arrays with a variable number of axes. An element in a three-dimensional tensor T can be denoted by Tx,y,z where x, y, z denote the three axes for specifying element T.

UCERD

Databases Format

- A CSV data file is one of the most widely available formats of data.
- Java Script Object Notation (JSON) is one of the most widely used data interchange formats across the digital realm. JSON is a lightweight alternative to legacy formats like XML.
- XML
- SQL => sqlalchemy and pyodbc

UCERD

Data Description

- Numeric
- Text
- Categorical

UCERD

Data Wrangling / Munging

- Understand Data
- Filtering Data
  - Clean
  - Typecasting,Duplicate Handling
- Transforming
- Normalize
- Visualization

	Date	Price	Product ID	Quantity Purchased	Serial No	User ID	User Type
0	NaN	3021.06	417	13	1000	5958	NaN
1	NaN	1822.62	731	1	1001	5351	c
2	2016-07-01	542.36	829	2	1002	5799	a
3	2016-01-20	2323.30	905	0	1003	5480	d
4	2016-01-19	243.43	158	37	1004	5790	a
5	2016-01-16	274.26	754	33	1005	5820	a
6	NaN	5836.68	341	18	1006	5468	c
7	2016-01-19	NaN	819	34	1007	5486	b
8	2016-01-23	1171.88	929	12	1008	5143	a
9	2016-07-01	668.80	718	31	1009	5510	d

UCERD

Statistics

- The field of statistics can be defined as a specialized branch of mathematics that consists of frameworks and methodologies to collect, organize, analyze, interpret, and present data. Generally this falls more undeapplied mathematics and borrows concepts from linear algebra, distributions, probability theory, and inferential methodologies. There are two major areas under statistics that are mentioned as follows.
- Descriptive statistics
- Inferential statistics

UCERD

Challenges in Machine Learning

- Data quality issues lead to problems, especially with regard to data processing and feature extraction.
- Data acquisition, extraction, and retrieval is an extremely tedious and time consuming process.
- Lack of good quality and sufficient training data in many scenarios.
- Formulating business problems clearly with well-defined goals and objectives.
- Feature extraction and engineering, especially hand-crafting features, is one of the most difficult yet important tasks in Machine Learning. Deep Learning seems to have gained some advantage in this area recently.
- Overfitting or underfitting models can lead to the model learning poor representations and relationships from the training data leading to detrimental performance.
- The curse of dimensionality: too many features can be a real hindrance. Complex models can be difficult to deploy in the real world.

Real-World Applications of Machine Learning

- Product recommendations in online shopping platforms
- Sentiment and emotion analysis
- Anomaly detection
- Fraud detection and prevention
- Content recommendation (news, music, movies, and so on)
- Weather forecasting
- Stock market forecasting
- Market basket analysis
- Customer segmentation
- Object and scene recognition in images and video
- Speech recognition
- Churn analytics
- Click through predictions
- Failure/defect detection and prevention
- E-mail spam filtering

UCERD