Industry Application

1. How will AI affect industry?

“…AI is the new electricity” – Andrew Ng

* Add $15.7 trillion to global economy by 2030 (PWC)
* Labor productivity improvements and product enhancements (PWC)

1. Sectors with great potential

* Retail/E-commerce: pricing and promotion and customer service management
  + Such as password reset activities
  + Better search result and newsletter
* Consumer goods: supply-chain management, demand forecasting
  + Better routing of goods and services
* Finance: marketing and sales, assessing and managing risk
  + How do I know when someone will likely to get married so that I can share them different insurance?

Affecting Industry

1. Sense and Decide: Blue River Technology

* Technology to put on tractors to distinguish weed vs plant to help them spray out the herbicide on the weeds which helps them reduce the usage of herbicide by 90%

Machine Learning Concept

1. What is machine learning?

* It’s a branch of Artificial Intelligence; computer algorithms that improve automatically through experience (Tom Mitchell, Professor and Former Chair of Machine Learning Department of Carnegie Mellon University)

1. Machine Learning Techniques

* Supervised Learning – uses pre-labeled data to train the model to predict new outcomes for information that hasn’t been exposed to before.
  + Classification
  + Regression
* Unsupervised Learning – uses non-labeled data and self organizes to predict patterns or outcomes such as clustering or associations
  + Clustering
  + Association
* Reinforcement Learning – is giving feedback to an algorithm when it does something right or wrong based on a discrete outcome, this can either be in:
  + Real-time
  + Offline

Supervised Learning

1. Supervised Learning is what most of machine learning and production today is actually using

* It’s mapping inputs to a particular output



* Example can be the image can be the input data and label is the output (for Classification)

1. Types of Supervised Learning

* Classification: Categorical outcome, categorizing unstructured data into particular categories or classes.
* Regression: the output variable in a regression is numerical or continuous whereas for classification it is discrete

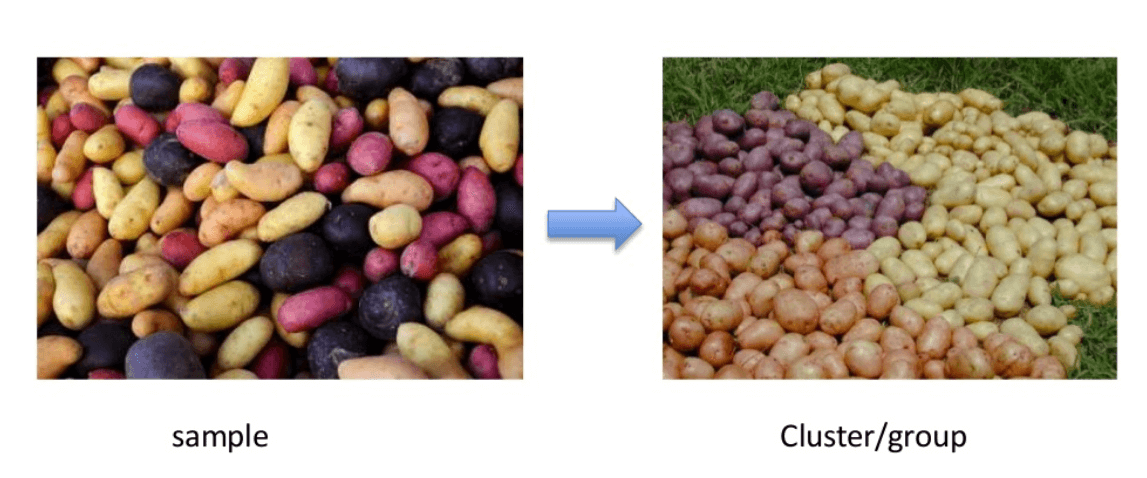
1. Applications

* Image classification
* Optical character recognition
* Face recognition
* Sentiment analysis
* Natural language processing
* Machine translation
* Audio transcription
* Event detection

Unsupervised Learning

1. Unsupervised learning finds pattern in data.
2. Clustering

* It mainly deals with finding a structure or pattern in a collection of uncategorized data. Clustering algorithms will process your data and find natural clusters(groups) if they exist in the data. You can also modify how many clusters your algorithms should identify. It allows you to adjust the granularity of these groups.



* Different Types of Clustering:
  + Exclusive (partitioning) - data are grouped in such a way that one data can belong to one cluster only.
    - Example: K-means
  + Agglomerative - every data is a cluster. The iterative unions between the two nearest clusters reduce the number of clusters.
    - Example: Hierarchical clustering
  + Overlapping - fuzzy sets is used to cluster data. Each point may belong to two or more clusters with separate degrees of membership. Here, data will be associated with an appropriate membership value.
    - Example: Fuzzy C-Means
  + Probabilistic - uses probability distribution to create the clusters
    - Example: Following keywords
      * "man's shoe."
      * "women's shoe."
      * "women's glove."
      * "man's glove."
    - can be clustered into two categories "shoe" and "glove" or "man" and "women."

1. Association

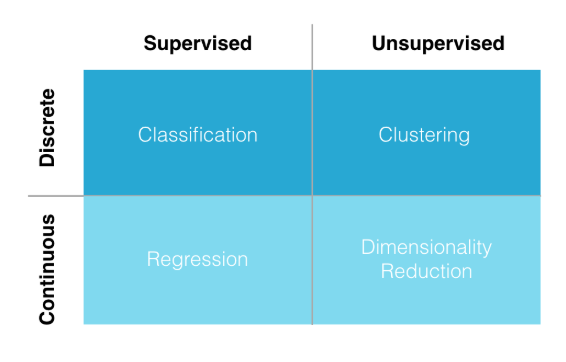
* Association rules allow you to establish associations amongst data objects inside large databases. This unsupervised technique is about discovering interesting relationships between variables in large databases. For example, people that buy a new home most likely to buy new furniture.

1. Applications

* E-commerce website: i.e. find the common customer’s journey to help retain customers
* Clustering automatically split the dataset into groups base on their similarities
* Anomaly detection can discover unusual data points in your dataset. It is useful for finding fraudulent transactions
* Association mining identifies sets of items which often occur together in your dataset
* Latent variable models are widely used for data preprocessing. Like reducing the number of features in a dataset or decomposing the dataset into multiple components

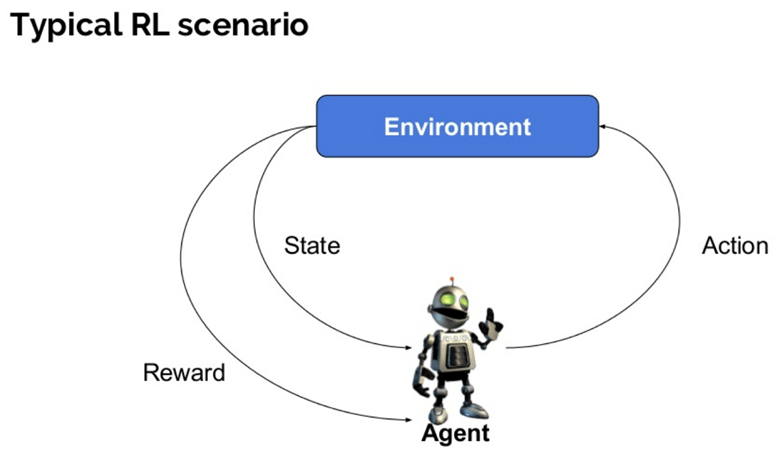
Supervised vs Unsupervised Learning

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Supervised machine learning technique** | **Unsupervised machine learning technique** |
| Process | In a supervised learning model, input and output variables will be given. | In unsupervised learning model, only input data will be given |
| Input Data | Algorithms are trained using labeled data. | Algorithms are used against data which is not labeled |
| Algorithms Used | Support vector machine, Neural network, Linear and logistics regression, random forest, and Classification trees. | Unsupervised algorithms can be divided into different categories: like Cluster algorithms, K-means, Hierarchical clustering, etc. |
| Computational Complexity | Supervised learning is a simpler method. | Unsupervised learning is computationally complex |
| Use of Data | Supervised learning model uses training data to learn a link between the input and the outputs. | Unsupervised learning does not use output data. |
| Accuracy of Results | Highly accurate and trustworthy method. | Less accurate and trustworthy method. |
| Real Time Learning | Learning method takes place offline. | Learning method takes place in real time. |
| Number of Classes | Number of classes is known. | Number of classes is not known. |
| Main Drawback | Classifying big data can be a real challenge in Supervised Learning. | You cannot get precise information regarding data sorting, and the output as data used in unsupervised learning is labeled and not known. |



Reinforcement Learning

1. It’s also another category for unsupervised learning, which concerns how machine learning agents ought to take actions in an environment so as to maximize a specific outcome
2. a machine learning method that is concerned with how software agents should take actions in an environment. Reinforcement Learning is a part of the deep learning method that helps you to maximize some portion of the cumulative reward. This neural network learning method helps you to learn how to attain a complex objective or maximize a specific dimension over many steps.



1. Here are some important terms used in Reinforcement AI:

* *Agent*: It is an assumed entity which performs actions in an environment to gain some reward.
* *Environment (e)*: A scenario that an agent has to face.
* *Reward (R)*: An immediate return given to an agent when he or she performs specific action or task.
* *State (s)*: State refers to the current situation returned by the environment.
* *Policy (π)*: It is a strategy which applies by the agent to decide the next action based on the current state.
* *Value (V)*: It is expected long-term return with discount, as compared to the short-term reward.
* *Value Function*: It specifies the value of a state that is the total amount of reward. It is an agent which should be expected beginning from that state.
* *Model of the environment*: This mimics the behavior of the environment. It helps you to make inferences to be made and also determine how the environment will behave.
* *Model based methods*: It is a method for solving reinforcement learning problems which use model-based methods.
* *Q value or action value (Q)*: Q value is quite similar to value. The only difference between the two is that it takes an additional parameter as a current action.

1. Reinforcement Learning Algorithms

* Value-Based - In a value-based Reinforcement Learning method, you should try to maximize a value function V(s). In this method, the agent is expecting a long-term return of the current states under policy π.
* Policy-based - In a policy-based RL method, you try to come up with such a policy that the action performed in every state helps you to gain maximum reward in the future.
  + Two types of policy-based methods are:
    - Deterministic: For any state, the same action is produced by the policy π.
    - Stochastic: Every action has a certain probability, which is determined by the following equation. Stochastic Policy : n{a\s) = P\A, = a\S, =S]
* Model-Based - In this Reinforcement Learning method, you need to create a virtual model for each environment. The agent learns to perform in that specific environment.

1. Types of Reinforcement Learning

* Positive - It is defined as an event, that occurs because of specific behavior. It increases the strength and the frequency of the behavior and impacts positively on the action taken by the agent. This type of Reinforcement helps you to maximize performance and sustain change for a more extended period. However, too much Reinforcement may lead to over-optimization of state, which can affect the results.
* Negative - defined as strengthening of behavior that occurs because of a negative condition which should have stopped or avoided. It helps you to define the minimum stand of performance. However, the drawback of this method is that it provides enough to meet up the minimum behavior.

1. One chat bot from Microsoft called “Tay” had gone awry because this type of learning can be quite powerful and might learn wrong things when not there’s no human in the loop
2. Human in the loop (HITL) refers to having a human-moderator or data annotator that can help with quality control of a product
3. Applications

* Robotics for industrial automation.
* Business strategy planning
* Machine learning and data processing
* It helps you to create training systems that provide custom instruction and materials according to the requirement of students.
* Aircraft control and robot motion control

Quiz:

Which of the following statements are incorrect:

1. Reinforcement Learning is also an unsupervised learning
2. Having human in the loop is not necessary especially when the data are getting bigger because it will cost more and AI helps to be more cost-effective
3. Supervised learning is a complex method compared to unsupervised learning
4. Facial recognition is part of unsupervised learning
5. Regression has continuous output variable that makes it a part of unsupervised learning
6. All of the above

Answer: B, C, D, E

References:

<https://www.guru99.com/supervised-machine-learning.html>

<https://www.guru99.com/unsupervised-machine-learning.html>

<https://www.guru99.com/reinforcement-learning-tutorial.html>