**Module 3: Supervised and Unsupervised Learning Part 2**

In this module, the student will be able to:

1. differentiate supervised from unsupervised learning by explaining their goals, input data requirements, and applications.
2. appreciate the importance of machine learning in real-world decision-making, recognizing its role in industries such as healthcare, finance, and retail.
3. apply a clustering algorithm (e.g., K-Means) to group unlabeled data and interpret the results meaningfully.

Lesson summary:

The module discusses the concepts of supervised and unsupervised learning in machine learning, highlighting their differences, applications, and advantages. Supervised learning uses labeled data to train models for prediction and classification tasks such as email spam detection, medical diagnosis, house price prediction, and sales forecasting. In contrast, unsupervised learning works with unlabeled data to discover hidden patterns, commonly applied in clustering, association, and dimensionality reduction, with examples like customer segmentation, image segmentation, market basket analysis, recommendation systems, and fraud detection. A comparison table further outlines distinctions in their goals, input data, human supervision, complexity, algorithms, and accuracy, helping learners understand when and how to apply each approach in real-world scenarios.

Lesson proper:

Basic Concepts associated with Learning

These machine learning algorithms are used across many industries to identify patterns, make predictions, and more. Explore the differences between supervised and unsupervised learning to understand better what they are and how you might use them.

Supervised learning and unsupervised learning are two common types of machine learning models. You can use machine learning in descriptive, predictive, and prescriptive analyses to answer questions, predict events, and guide decisions. Discover the uses of each approach, their pros and cons, and how to decide which is right for your purposes.

**Supervised and unsupervised learning uses**

[Machine learning](https://www.coursera.org/articles/what-is-machine-learning), a subset of [artificial intelligence](https://www.coursera.org/articles/what-is-artificial-intelligence) (AI), uses algorithms to parse data, gather information, and output predictions or decisions without being specifically programmed to do so. Various disciplines use supervised and unsupervised learning algorithms in machine learning processes, each with its own strengths and best-case uses.

By understanding how the unique features of each learning algorithm can benefit different functions, you can make informed decisions about how to use these tools to answer questions and guide decision-making.

**Supervised learning**

You use supervised machine learning algorithms when you have defined, known output data. This learning method requires labeled input and output data to train the model, which can then make predictions by learning from the provided data set. For instance, supervised learning can perform applications like email spam filtering and object recognition.

**1. Classification (KNN, Random Forests, Decision Trees, etc.)**

* **Email Spam Detection**: An email service classifies incoming emails as "spam" or "not spam" based on their content and sender behavior.
* **Medical Diagnosis**: A system predicts whether a tumor is "benign" or "malignant" based on patient data like age, test results, and medical history.

**2. Regression (Linear, Polynomial, Logistic, etc.)**

* **House Price Prediction**: A real estate company forecasts the selling price of houses by analyzing factors like location, number of rooms, and square footage.
* **Sales Forecasting**: A retail store predicts future monthly sales based on past sales data, seasonal trends, and promotional campaigns.

**Unsupervised learning**

You might choose unsupervised machine learning, on the other hand, when the target output is unknown and the data is unlabeled. This type of learning discovers hidden patterns in data. It is commonly used for clustering data points in different groups (such as populations), which can help with tasks like market segmentation. Other applications include anomaly detection, such as detecting faulty equipment or security concerns.

**1. Clustering (K-Means, etc.)**

* **Customer Segmentation**: A retail company groups its customers based on purchasing behavior to create targeted marketing strategies (e.g., budget buyers, luxury buyers).
* **Image Segmentation**: In medical imaging, clustering is used to separate different tissues (like bones, organs, tumors) in an MRI or CT scan.

**2. Association (Apriori, etc.)**

* **Market Basket Analysis**: A supermarket discovers that customers who buy bread often also buy butter, so they place those items closer together to increase sales.
* **Streaming Services**: Netflix finds that users who watch a certain type of movie (say, action films) also tend to watch superhero series, so it recommends related titles.

**3. Dimensionality Reduction (PCA, t-SNE, etc.)**

* **Facial Recognition**: Reducing the number of features from an image dataset (like pixel values) to keep only the most important ones, making recognition faster and more accurate.
* **Financial Fraud Detection**: Simplifying massive transaction datasets with many variables (location, time, amount, device used, etc.) into fewer key dimensions to detect unusual patterns more easily.

Differences between Supervised and Unsupervised Learning

The table below shows some key differences between supervised and unsupervised machine learning −

|  |  |  |
| --- | --- | --- |
| **Basis** | **Supervised Learning** | **Unsupervised Learning** |
| Definition | Supervised learning algorithms train data, where every input has a corresponding output. | Unsupervised learning algorithms find patterns in data that has no predefined labels. |
| Goal | The goal of supervised learning is to predict or classify based on input features. | The goal of unsupervised learning is to discover hidden patterns, structures and relationships. |
| Input Data | Labeled: Input data with corresponding output labels. | Unlabeled: Input data is raw and unlabeled. |
| Human Supervision | Supervised learning algorithms needs human supervision to train the model. | Unsupervised learning algorithms does not any kind of supervision to train the model.. |
| Tasks | Regression, Classification | Clustering, Association and Dimensionality Reduction |
| Complexity | supervised machine learning methods are computationally simple. | Unsupervised machine learning methods are computationally complex. |
| Algorithms | Linear regression, K-Nearest Neighbors, Decision Trees, Naive Bayes, SVM | K- Means clustering, DBSCAN, Autoencoders |
| Accuracy | Supervised machine learning methods are highly accurate. | Unsupervised machine learning methods are less accurate. |
| Applications | Image classification, Sentiment Analysis, Recommendation systems | Customer Segmentation, Anomaly Detection, Recommendation Engines, NLP |

References:

* [Supervised vs. Unsupervised Learning: Pros, Cons, and When to Choose | Coursera](https://www.coursera.org/articles/supervised-vs-unsupervised-learning?utm_medium=sem&utm_source=gg&utm_campaign=b2c_apac_x_multi_ftcof_career-academy_cx_dr_bau_gg_pmax_gc_s2_all_m_hyb_24-08_x&campaignid=21573875733&adgroupid=&device=c&keyword=&matchtype=&network=x&devicemodel=&creativeid=&assetgroupid=6511393614&targetid=&extensionid=&placement=&gad_source=1&gad_campaignid=21584159401&gbraid=0AAAAADdKX6Z__oeMcew1YOWhNB44t6chd&gclid=Cj0KCQjwzt_FBhCEARIsAJGFWVk8JW5euYMmyb4QXsYbCAc-RgbTJ9ImDbOL_FMsg6J_7dIKp5ZWMZgaAncOEALw_wcB)
* [Supervised vs. Unsupervised Learning](https://www.tutorialspoint.com/machine_learning/machine_learning_supervised_vs_unsupervised.htm)