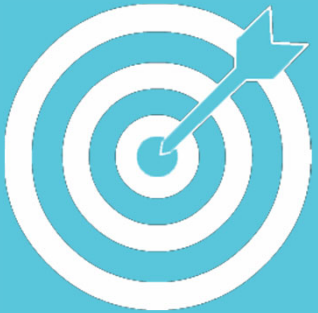# Classical Machine Learning (CML) Techniques

An Introduction

Learning Objectives

- Discuss supervised learning with examples

- Explain unsupervised learning with examples

- Define reinforcement learning

- Define some important models and techniques in ML

Supervised Learning: Case Study

Ever wondered how Amazon makes recommendations?

Supervised Learning: Case Study

Amazon uses **supervised learning** algorithms to predict what items the user may like based on the purchase history of similar classes of users.

New Input

Algorithm Trained
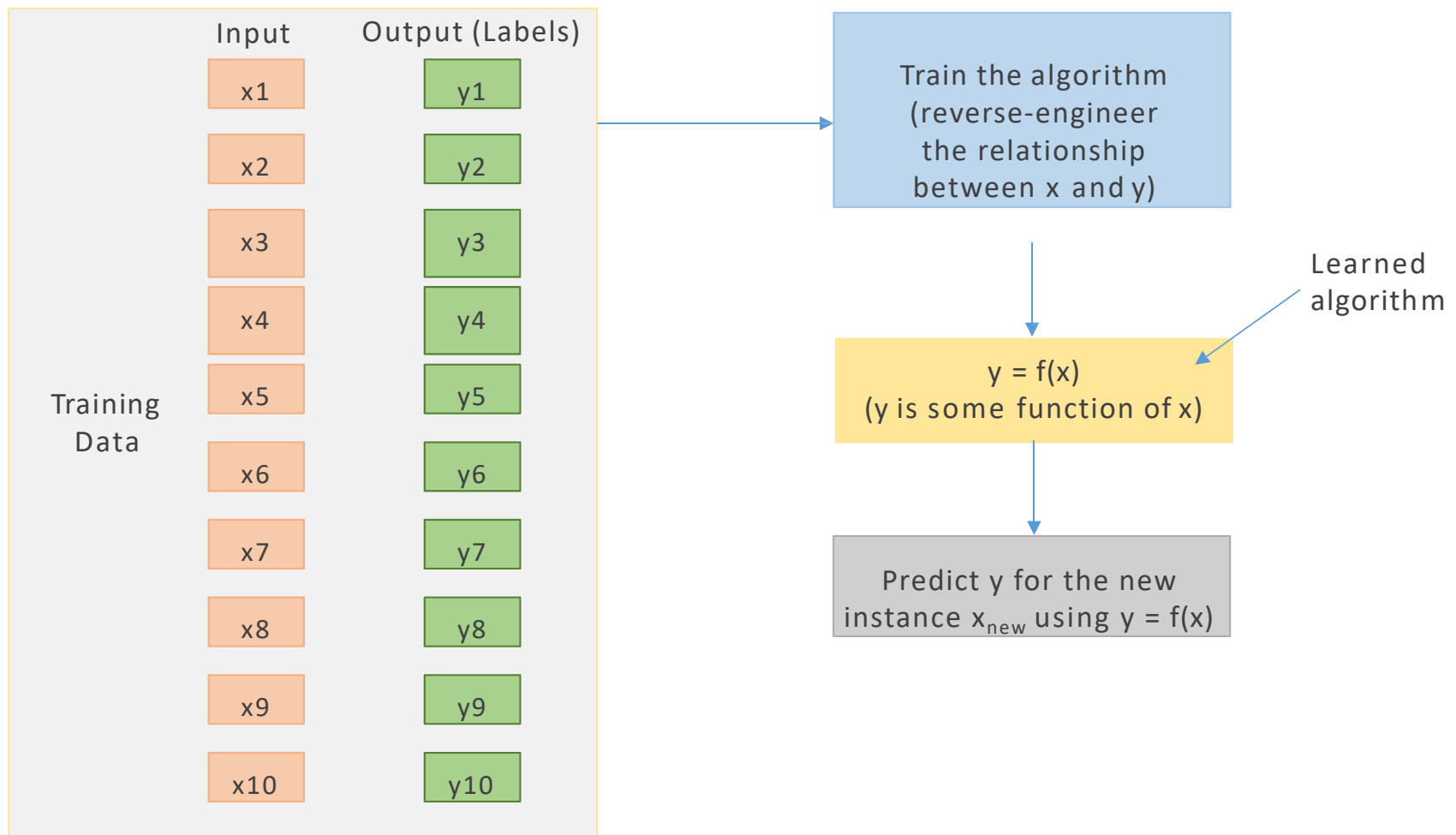on Historical Data

Predicted
Output
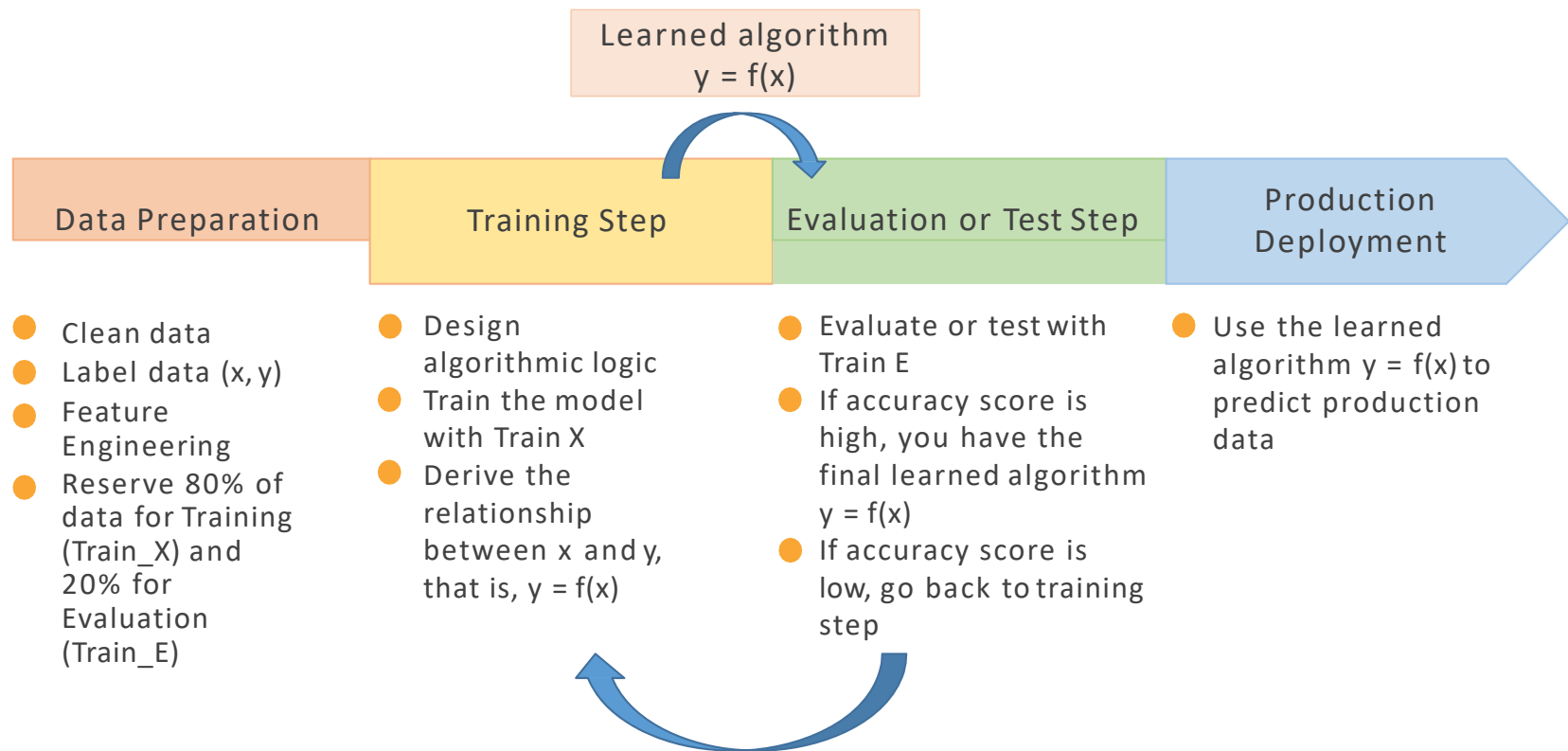
What Is Supervised Learning?

" Supervised Learning is a type of Machine Learning used to learn models
from labeled training data. It allows us to predict output for future or
unseen data. "

Understanding the Supervised Learning Algorithm

Supervised Learning Flow

Learned algorithm
y = f(x)

Data Preparation	Training Step	Evaluation or Test Step	Production Deployment

- Clean data
- Label data (x, y)
- Feature Engineering
- Reserve 80% of data for Training (Train_X) and 20% for Evaluation (Train_E)

- Design algorithmic logic
- Train the model with Train X
- Derive the relationship between x and y, that is, y = f(x)

- Evaluate or test with Train E
- If accuracy score is high, you have the final learned algorithm y = f(x)
- If accuracy score is low, go back to training step

- Use the learned algorithm y = f(x) to predict production data

The algorithm can be improved by more training data, capacity, or algo redesign.

Testing the Algorithm

1

Once the algorithm is trained, test it with test data (a set of data instances that do not appear in the training set).

2

3

4

Testing the Algorithm

Once the algorithm is trained, test it with test data (a set of data instances that do not appear in the training set)

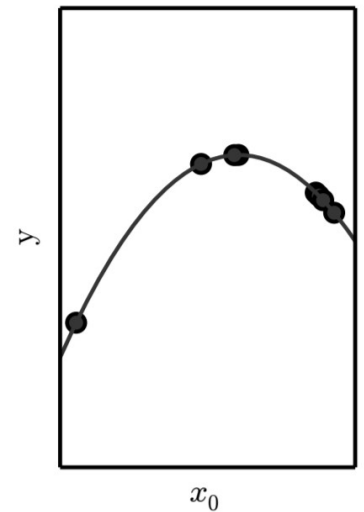A well-trained algorithm can predict well for new test data.

Appropriate capacity

y

x_0

Testing the Algorithm

1	Once the algorithm is trained, test it with test data (a set of data instances that do not appear in the training set)
2	A well trained algorithm can predict well for new test data.
3	If the learning is poor, you have an **underfitted** situation. The algorithm will not work well on test data. Retraining may be needed to find a better fit.
4	

Underfitting

y

x_0

Testing the Algorithm

1   Once the algorithm is trained, test it with test data (a set of data instances that do not appear in the training set)
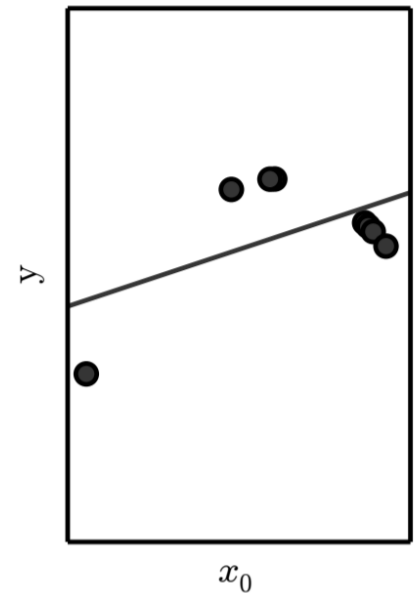
**2**   A well trained algorithm can predict well for new test data.

3   If the learning is poor, we have an **underfitted** situation. The algorithm will not work well on test data. Retraining may be needed to find a better fit.

4   If learning on training data is not proper, it may lead to **overfitting** – a situation where the algorithm is not able to handle new testing data that it has not seen before. The technique to keep data generic is called regularization.
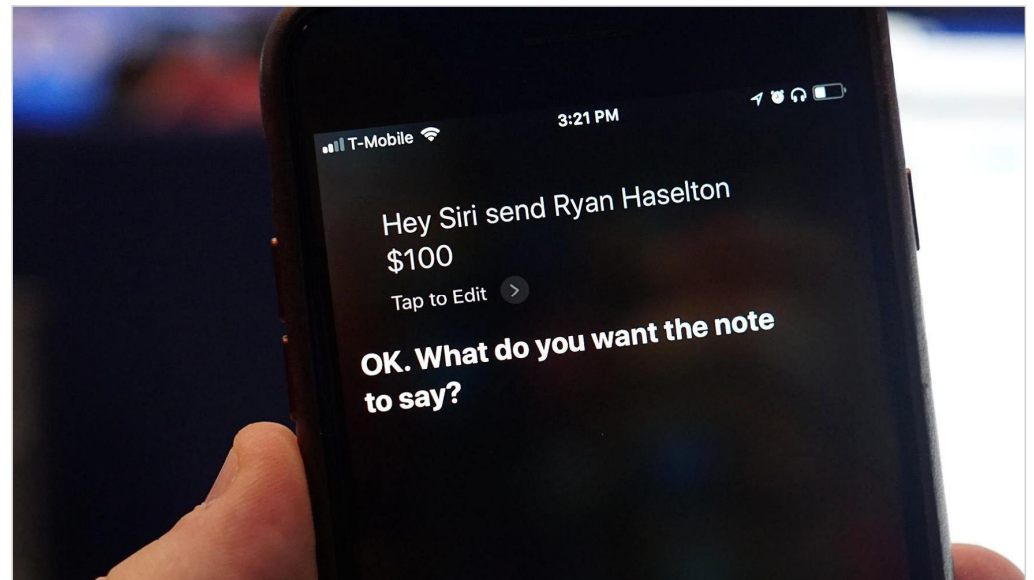
Overfitting

x_0
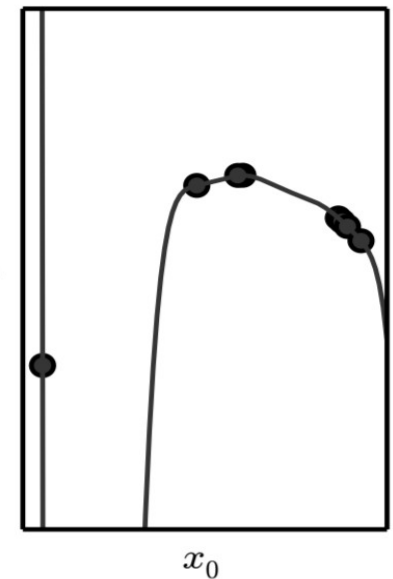
Examples of Supervised Learning

Example 1: Voice Assistants

Voice assistants like Apple Siri, Amazon Alexa, Microsoft Cortana, and Google Assistant are trained to understand human speech and intent.

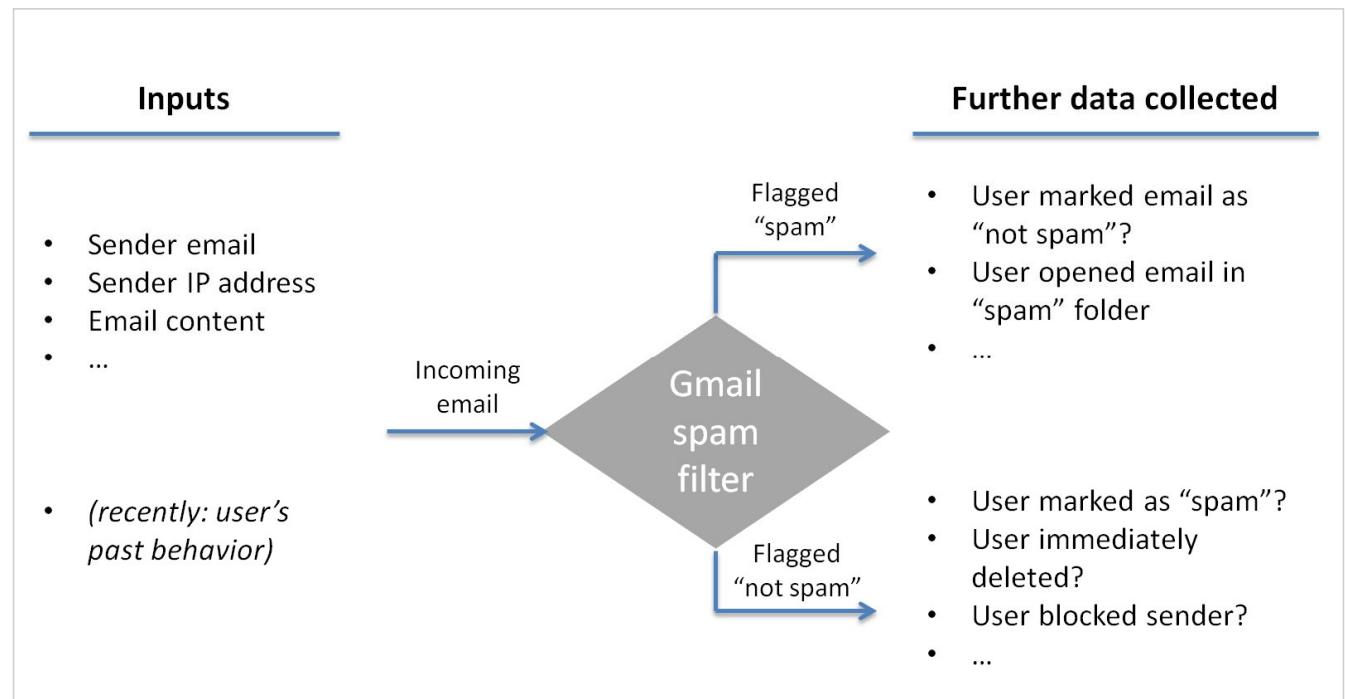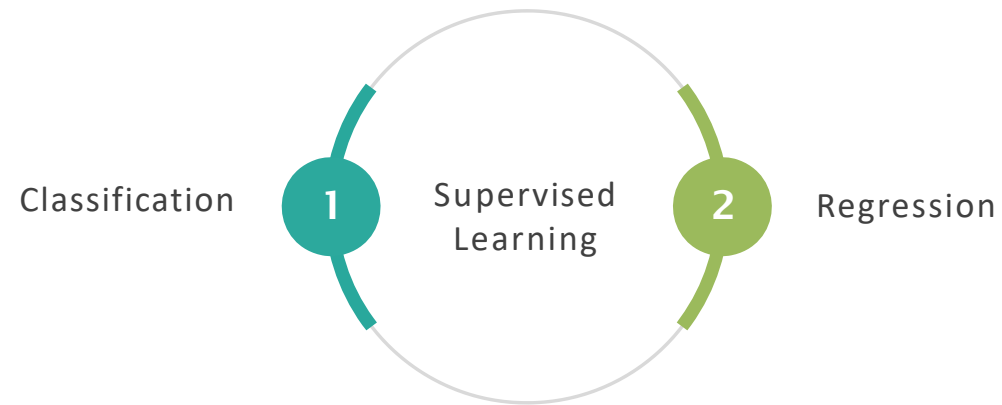Based on human interactions, these chatbots take appropriate action.

https://www.cnbc.com/2017/08/11/how-to-send-money-using-siri-on-an-iphone.html

Examples of Supervised Learning

Example 2: Gmail Filters

Gmail filters a new email into Inbox (normal) or Junk folder (Spam) based on past information about what you consider spam.

Inputs

- Sender email
- Sender IP address
- Email content
- …

- *(recently: user's past behavior)*

Incoming email → Gmail spam filter

Flagged "spam"

Flagged "not spam"

Further data collected

- User marked email as "not spam"?
- User opened email in "spam" folder
- …

- User marked as "spam"?
- User immediately deleted?
- User blocked sender?
- …

Types of Supervised Learning

Classification

1

Supervised Learning

2

Regression

Types of Supervised Learning

CLASSIFICATION

Classification **1** Supervised Learning **2** Regression
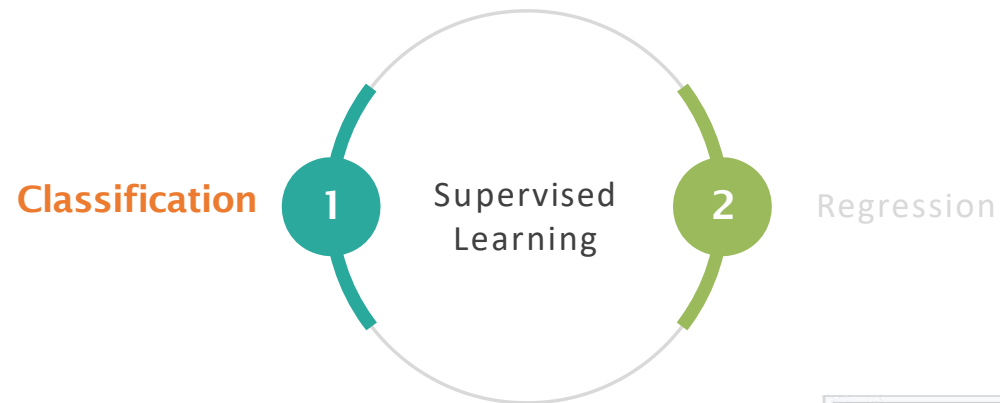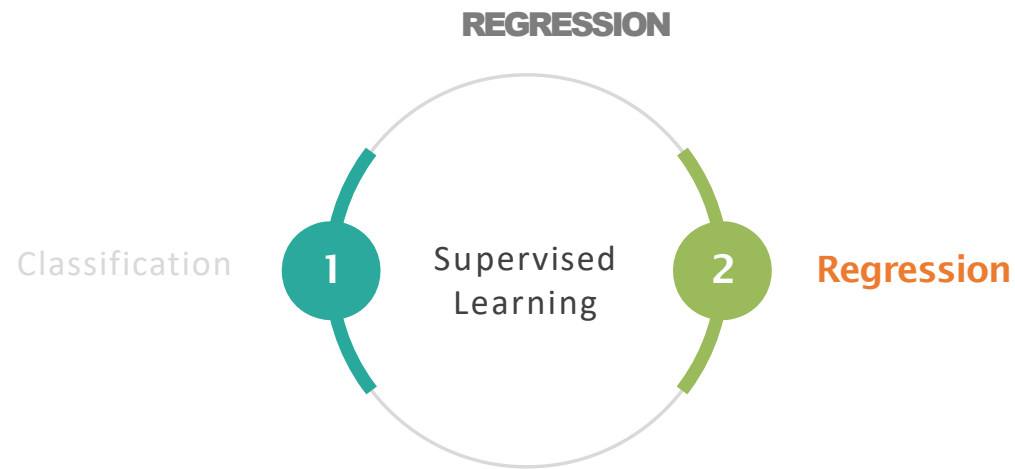
- Answers "What class?"

- Applied when the output has finite and discreet values Example: Social media sentiment analysis has three potential outcomes, positive, negative, or neutral

Types of Supervised Learning

Classification   **1**   Supervised Learning   **2**   **Regression**
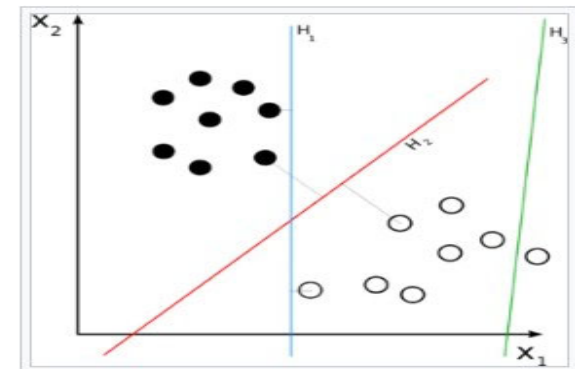
- Answers "How much?"

- Applied when the output is a continuous number

- A simple regression algorithm: $y = wx + b.$ Example: relationship between environmental temperature (y) and humidity levels (x)
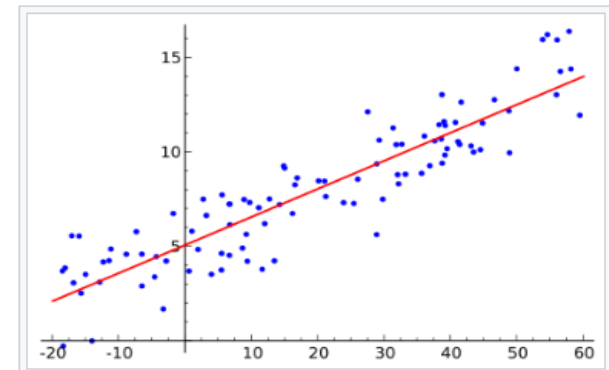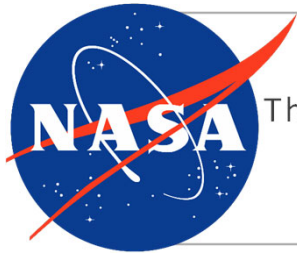
Unsupervised Learning

Unsupervised Learning: Case Study

Ever wondered how NASA discovers a new heavenly body and identifies that it is different from a previously known astronomical object?

It has no knowledge about these new bodies but classifies them into proper categories.

Unsupervised Learning: Case Study

This is possible through **unsupervised learning** to create clusters of heavenly bodies, with each cluster containing objects of a similar nature.
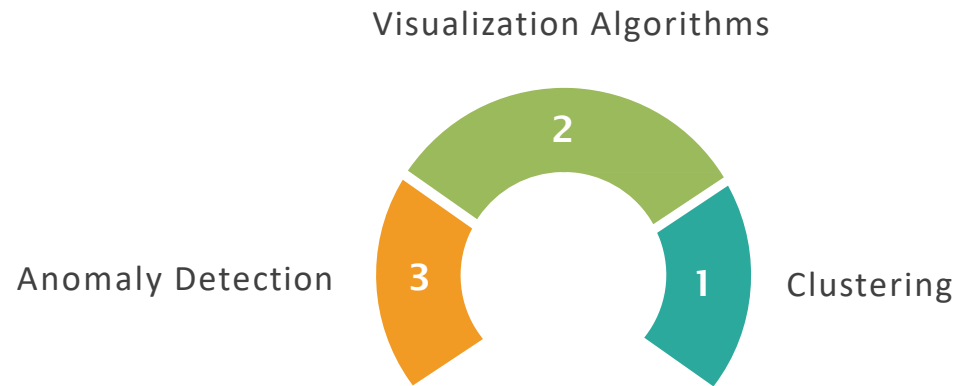
What Is Unsupervised Learning?

> Unsupervised Learning is a subset of Machine Learning used to extract inferences from datasets that consist of input data without labeled responses.
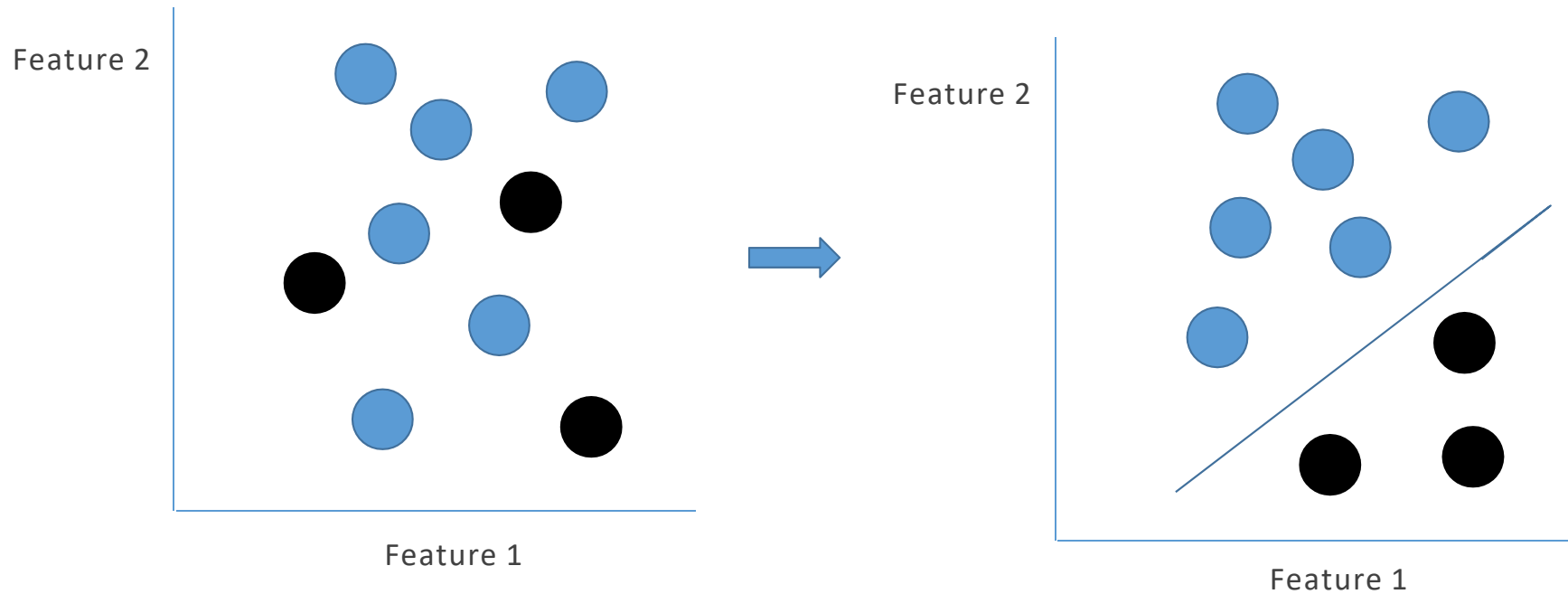
Types of Unsupervised Learning

Visualization Algorithms

Anomaly Detection

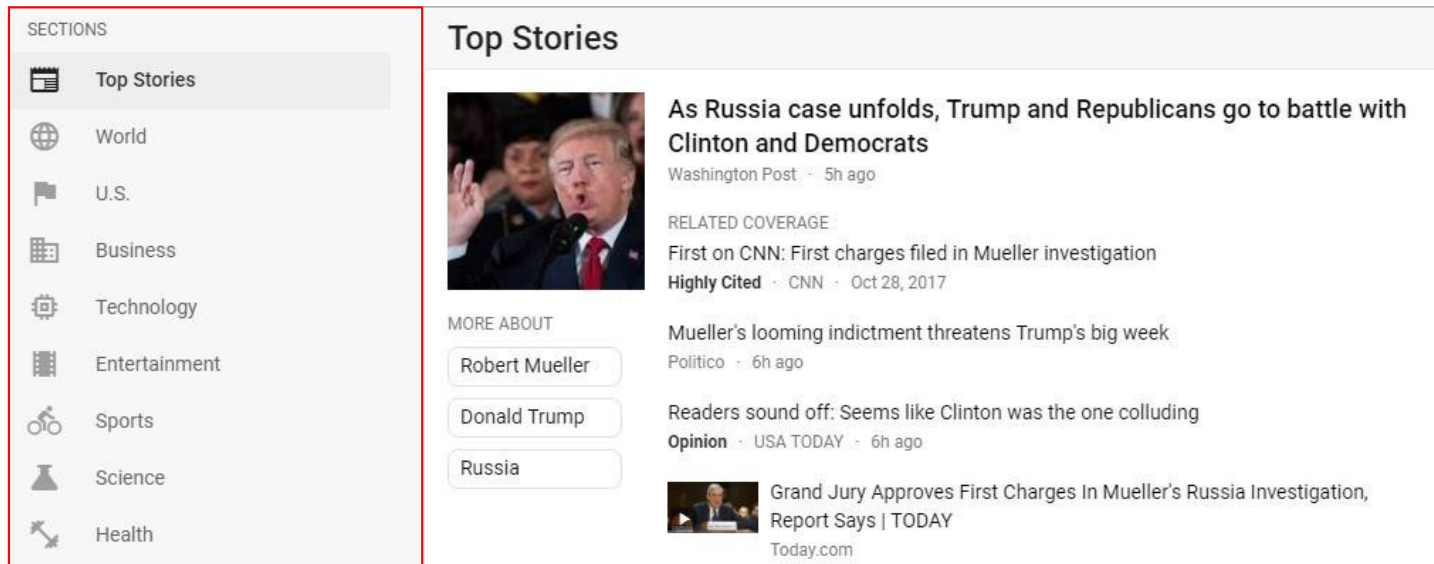Clustering

Types of Unsupervised Learning

CLUSTERING

The most common unsupervised learning method is cluster analysis. It is used to find data clusters so that each cluster has the most closely matched data.

Types of Unsupervised Learning

CLUSTERING

Example: An online news portal segments articles into various categories like Business, Technology, Sports, etc.

Types of Unsupervised Learning

VISUALIZATION ALGORITHMS

Visualization algorithms are unsupervised learning algorithms that accept unlabeled data and display this data in an intuitive 2D or 3D format. The data is separated into somewhat clear clusters to aid

Example:

In the figure, the animals are rather well separated from vehicles. Horses are close to deer but far from birds, and so on.
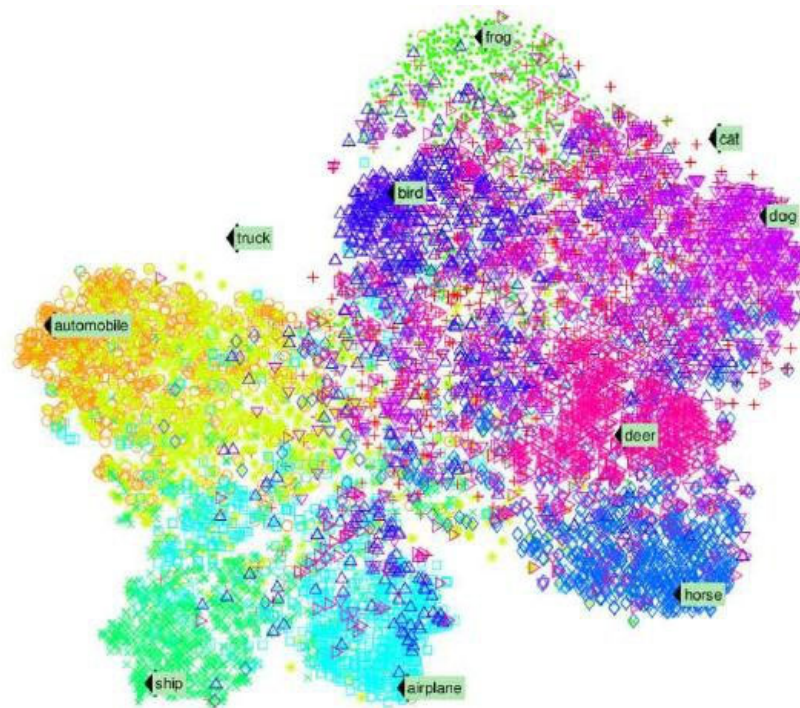
Types of Unsupervised Learning

ANOMALY DETECTION

This algorithm detects anomalies in data without any prior training.

Example:

It can detect suspicious credit card transactions and differentiate a criminal from a set of people.
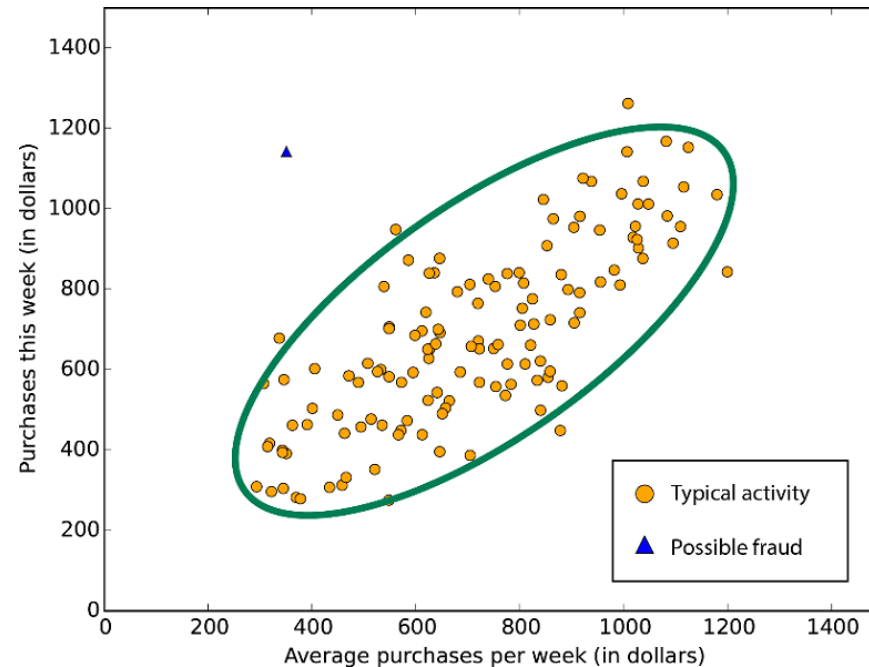
What Is Reinforcement Learning?

> Reinforcement Learning is a type of Machine Learning that allows the learning system to observe the environment and learn the ideal behavior based on trying to maximize some notion of cumulative reward.

Features of Reinforcement Learning

- The learning system (agent) observes the environment, selects and takes certain actions, and gets rewards in return (or penalties in certain cases).

- The agent learns the strategy or policy (choice of actions) that maximizes its rewards over time.
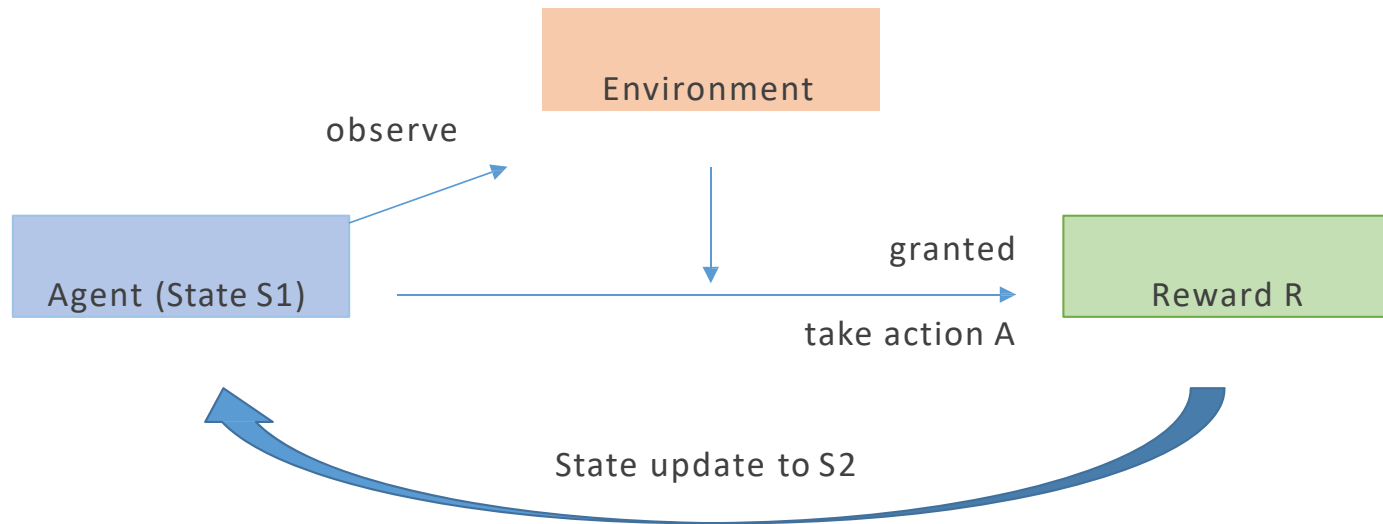
Example of Reinforcement Learning

In a manufacturing unit, a robot uses deep reinforcement learning to identify a device from one box and put it in a container.

The robot learns this by means of a rewards-based learning system, which incentivizes it for the right action.

http://4tsi.com/automation-not-trade-main-cause-of-factory-job-losses/

Some Important Considerations in Machine Learning

Bias and Variance Tradeoff

- Bias refers to error in the machine learning model due to wrong assumptions. A high-bias model will **underfit** the training data.

- Variance refers to problems caused due to overfitting. This is a result of over-sensitivity of the model to small variations in the training data. A model with many degrees of freedom (such as a high-degree polynomial model) is likely to have high variance and thus **overfit** the training data.

Bias and Variance Dependencies

Low variance

High bias

(simple model, underfitting)

High variance

Low bias

(complex model, overfitting)
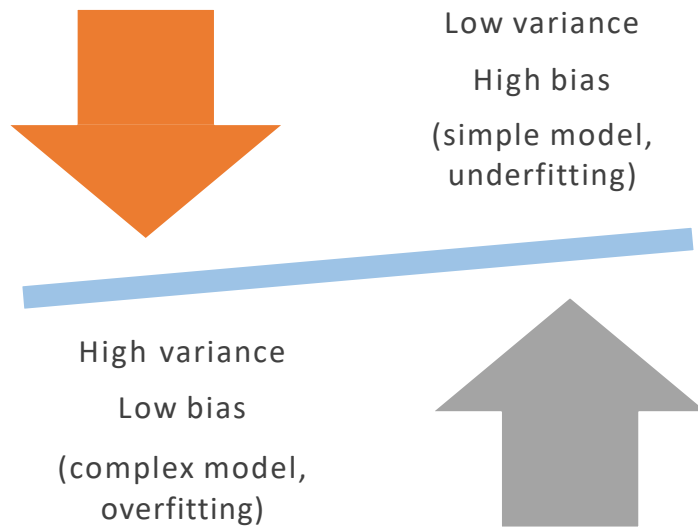
- Increasing a model's complexity will reduce its bias and increase its variance.

- Conversely, reducing a model's complexity will increase its bias and reduce its variance. This is why it is called a tradeoff.

Linearly Separable or Not

- The convergence of the learning algorithms (like perceptron) is only guaranteed if the two classes are linearly separable and the learning rate is sufficiently small.

- If the two classes can't be separated by a linear decision boundary, you can set a maximum number of passes over the training dataset (epochs) and/or a threshold for the number of tolerated misclassifications. The perceptron would never stop updating the weights otherwise.
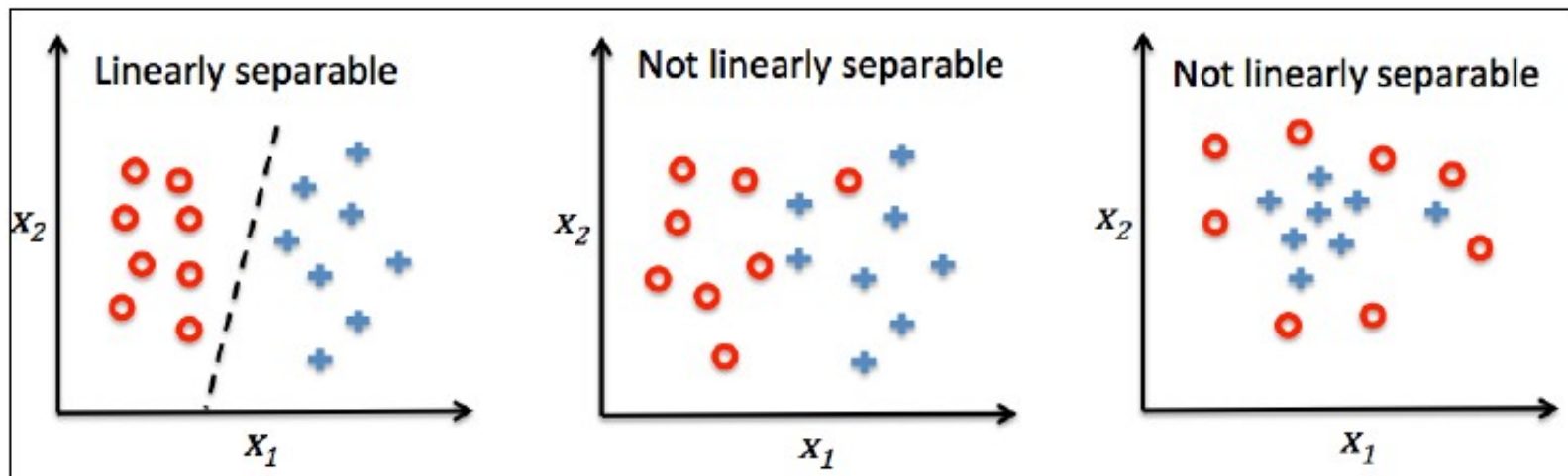
Image Credit : "Python Machine Learning" by Sebastian Raschka

Other Machine Learning Techniques

Probabilistic Models	Decision Trees	Clustering	Associated Rules	Deep Learning	Support Vector Machines
• Model the probability distribution of a data set and use it to predict future outcomes	• Arrive at a hierarchical decisioning tree structure	• Classify data based on closest data points appearing in the same cluster	• A method to discover what items tend to occur together in a sample space	• Based on Artificial Neural Network models • Contains inter-connected neurons organized in layers. This network learns to predict outcomes	• Method to classify data by discovering hyperplanes (separating layers) that segregate types of data

Summary

- Supervised Learning is a type of Machine Learning used to learn models from labeled training data. It allows us to predict output for future or unseen data. Two major types of Supervised Learning are Regression and Classification.

- The ML process (for supervised learning) entails data pre-processing, training the model, and testing the trained model and production deployment.

- If the training is poor, it may lead to underfitting (model does not satisfy the test data). If the training is too intensive, it may lead to overfitting (the model is not able to handle new unseen test data).

- Unsupervised Learning is a subset of Machine Learning used to extract inferences from datasets that consist of input data without labelled responses. Some examples of Unsupervised Learning include Clustering and Visualization algorithms.

In Class Activity

Hands On Activity

https://www.analyticsvidhya.com/blog/2022/01/a-guide-to-understand-machine-learning-pipeline-with-case-study/

Hands On Activity

https://www.analyticsvidhya.com/blog/2020/01/build-your-first-machine-learning-pipeline-using-scikit-learn/

Hands On Activity

https://www.kaggle.com/code/pouryaayria/a-complete-ml-pipeline-tutorial-acu-86

Upcoming Deliverables

- Quiz #1 (Due Jan 27)
- Syllabus Feedback Assignment (Due Jan 20)
- Student Introduction Discussion (Due Jan 20)
- Beginning of Semester – Self Assessment (Due Jan 20)
- Assignment 1 (Due Feb 10)

Next Class Topic

- Math Basics
    - Read Text Book – Appendix A – Math Review
    - Read Ref. Book 1 – pp. 18-25; 30-38.

Appendix

ML Pipeline Phases

- **Business Understanding and Problem Definition**

  Business **objectives and requirements understanding, machine learning problem definition**

- **Data Understanding and Ingestion**

  Initial data collection and familiarization, Data quality problems identification

- **Data Preparation and Segregation**

  Table, record and attribute selection, Data transformation and cleaning. Segregate prepared data to training and testing data

- **Modeling (Model Training)**

  Modeling techniques selection and application, Parameters calibration

- **Model Evaluation**

  Evaluate model in addressing **Business objectives and solving the problem**

- **Deployment and Performance Monitoring**

  Result model deployment, Repeatable ML process implementation and monitor the results

What Is Semi-supervised Learning?

"

It is a hybrid approach (combination of Supervised and Unsupervised Learning) with some labeled and some non-labeled data.

"

Example of Semi-Supervised Learning

Google Photos automatically detects the same person in multiple photos from a vacation trip (clustering – unsupervised).

One has to just name the person once (supervised), and the name tag gets attached to that person in all the photos.
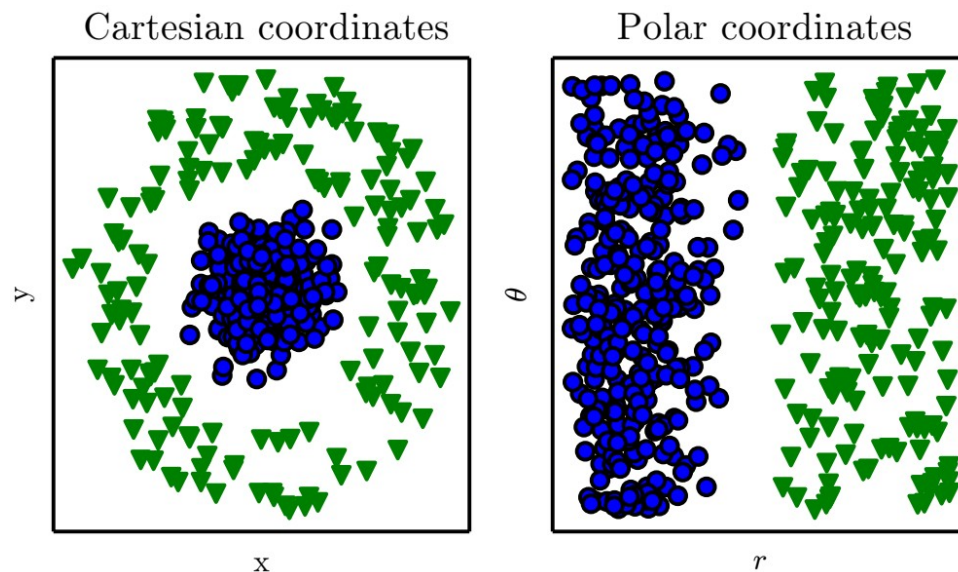

https://play.google.com/store/apps/details?id=com.google.android.apps.photos
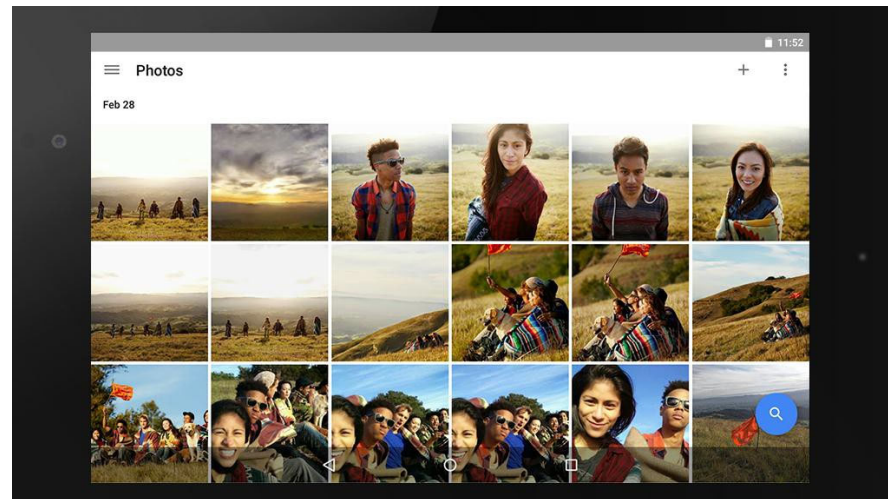
What Is Representation Learning?

> In Machine Learning, Representation refers to the way the data is presented. This often make a huge difference in understanding.

Example of Representation Learning

The figure shows sample data in Cartesian coordinates and polar coordinates.

In this particular case, categorization becomes easier when data is presented in a different coordinate system. Hence, representation matters.

Examples of Supervised Learning

Example 3: Weather Apps

The predictions made by weather apps at a given time are based on some prior knowledge and analysis of how the weather has been over a period of time for a particular place.