① ①

"A computer program is said to learn from _____ with respect to some _____ and some _____ measure

② ③

- Machine learning can be defined by three key aspects:

② ▶ [redacted]

① ▶ [redacted]

③ ▶ [redacted]

_____ is what you want the machine to achieve. It's the primary goal of the learning process.

_____ corresponds to the training data or the input that the machine learns from. This data provides examples or instances for the machine to develop a model or understanding.

_____ is the measure of how well the machine is doing its task after being trained on the experience. It's essentially a measure of its accuracy, efficiency, and overall capability to perform its task.

_____ program an algorithm to automatically learn from data, or from experience

- Why might you want to use a learning algorithm?
  1. hard to code up a solution by hand (e.g. vision, speech)
  2. system needs to adapt to a changing environment (e.g. spam detection)
  3. want the system to perform better than the human programmers

What is the difference between traditional programming and machine learning

| Aspect | Traditional Programming | Machine Learning |
|---|---|---|
| **Data** | Predefined and static datasets. | Dynamic, real-time, and historical. |
| **Flexibility** | Fixed and rigid. | Adaptive to changes. |
| **Methodology** | Hardcoded rules. | Data-driven predictions. |
| **Adaptability** | Limited; struggles with unforeseen cases. | Highly adaptive to new scenarios. |
| **Efficiency** | Resource-intensive for complex tasks. | Optimized for real-time problems. |
| **Goal** | Solve predefined problems. | Learn and improve over time. |

1. _____ Have labeled examples of the correct behavior.
2. _____ No labeled examples – instead, looking for “interesting” patterns in the data.
3. _____Learns through trial and error. The machine’s algorithms are rewarded when it performs a correct action and penalized when it doesn’t.

## Machine Learning Types and Examples

1. Supervised Learning

- Classification
Example: Image Classification, Identity Fraud Detection, Customer Retention, Diagnostics.
- Regression
Example: Advertising Popularity Prediction, Weather Forecasting, Market Forecasting, Estimating Life Expectancy, Populatio
Growth Prediction.

2. Unsupervised Learning

- Clustering
Example: Customer Segmentation, Targeted Marketing, Recommender Systems.
- Dimensionality Reduction
Example: Big Data Visualization, Meaningful Compression, Structure Discovery, Feature Elicitation.

3. Reinforcement Learning

- Learning Tasks
Example: Real-Time Decisions, Game AI, Robot Navigation, Skill Acquisition.
This categorization highlights tasks and examples of each machine learning type for better understanding.

is an area of ML that focuses on unifying ML with artificial intelligence.

## Machine learning concept and terminology

- Data: forms the main source of learning in machine learning.
+ It can be structured (e.g., tables) or unstructured (e.g., images, text).
+ Features (columns) and instances (rows) are fundamental components of datasets.

- Task:  is the problem that the ML algorithm is designed to solve.
+ It defines the goal, and the performance is measured based on how well the task is achieved.

- Algorithms: are the methods or processes applied to data to create models.
+ Different algorithms are suited for different tasks and datasets.

Models: are the outputs of algorithms applied to datasets.
+ They represent the learned patterns or relationships in the data.
+ Models can be logical, geometric, or probabilistic.

[redacted] a subset of attributes used to describe a property of data. [redacted]

- For example [redacted]

A data dimension consists of three attributes: day, month, year

## Data Types:

- Attributes or features can have different data types, such as:
  - Categorical (for example: young, old),
  - Ordinal (for example: 0, 1),
  - Numeric (for example: 1.3, 2.1, 3.2, and so on).

- [redacted] Refers to the data examples that are used to learn or build a classifier. [redacted]
- [redacted] Refers to the data examples that are verified against the built classifier and can help tune the accuracy of the output. [redacted]
- [redacted] Refers to the data examples that help assess the overall performance of the classifier.

What is the difference between testing dataset and validation dataset?

| Aspect | Validation Dataset | Testing Dataset |
|---|---|---|
| **Purpose** | Tunes the model parameters. | Evaluates the final model's accuracy. |
| **When Used** | During development. | After finalizing the model. |
| **Modifications** | Model can be adjusted based on results. | No adjustments are made. |
| **Outcome** | Optimizes performance. | Confirms generalization capability. |

_____ is usually the raw form of the data.

It consists of samples of natural or human-created artifacts. For example, video streams, audio, photos, and tweets.

_____ are mapped or defined by humans and are significantly more expensive to obtain than the unlabeled raw data.

## Machine learning phases

| Phase | Description | Output |
| --- | --- | --- |
| **Phase 1 - Training** | Uses training data to pair inputs with expected outputs to build the model. | A trained learning model. |
| **Phase 2 - Validation and Testing** | Measures the model's performance using validation/testing datasets, estimating metrics like accuracy and precision. | A refined and validated model. |
| **Phase 3 - Application** | Applies the model to real-world data to derive results and predictions. | Real-world outputs and predictions. |

## Machine learning process

1. ML Problem Framing
2. Data Collection
3. Data Integration
4. Data Preparation and Cleaning
5. Data Visualization and Analysis
6. Feature Engineering (or Feature Augmentation if required)
7. Model Training and Parameter Tuning

8. Model Evaluation
- Condition - Are Business Goals Met?
* Yes: Proceed to Model Deployment.
* No: Return to Feature Augmentation, Data Augmentation

9. Model Deployment
10. Monitoring and Debugging

# Examples of ML problems

| Example | Description |
| --- | --- |
| **Image Recognition** | Machine learning algorithms analyze images to identify objects, people, or patterns. Applications include healthcare (diagnosing diseases), retail (categorizing products), and security (detecting threats). |
| **Spam Detection** | ML algorithms classify emails as spam or not spam based on rules derived from features of the email data. |
| **Digit Recognition** | Interprets handwritten numbers (e.g., zip codes) and groups them for faster processing. |
| **Fraud Detection** | Identifies anomalies in transactional data to detect fraud in industries like finance, healthcare, and e-commerce. |
| **Predictive Maintenance** | Analyzes sensor and maintenance data to predict equipment failures, reducing downtime and optimizing performance. |
| **Speech Recognition** | Automated systems interpret and execute user requests by mapping them to specific tasks (e.g., call center automation). |
| **Face Detection** | Tags and categorizes individuals across digital photographs; widely used in social media and photo management applications. |

# Real world ML applications

| Application | Description |
| --- | --- |
| **Speech Recognition** | Natural language understanding in narrowly bounded domains. |
| **Image Recognition** | Identifies objects, people, or patterns in images. |
| **Driving Assistance** | Enables vehicles to drive safely along curving roads. |
| **Real-Time Translation** | Translates spoken Chinese into spoken English in real time. |