Muhammad Gunzelas 0311-7711699

# NOTES ML



## Machine Learning

Machine learning is the process of training computers, using math and statistical processes to find and recognize patterns in data.

After patterns are found, ML generates and updates training model to make increasingly accurate predictions and inferences about future outcomes based on historical and new data.

For Example:-

ML could help determine the likelihood of a customer purchasing a particular product based on previous purchases by the user or the product's past sales history.

Building ML applications is an iterative process that involves a sequence of steps.

To build an ML application

follow these general steps:

1- Formulate a Problem:-
Frame the core ML problem(s) in term of what is observed and what answer you want the model to predict.

2- Prepare your Data :-
Collect, clean and prepare data to make it suitable for consumption by ML model training algorithms. Visualize and analyze the data to run sanity checks to validate the quality of the data and to understood the data.

3- Train ~~your~~ the Model :- Imp
To train a highly predictive model, the raw data (input variables) and answer (target) can't always

be used effectivly. Preferably, Construct more predictive input representation or features from the raw variables. Feed the resulting features to learning algorithm to to build model. Set aside, or holdout, a subset of the input data to test the model.

4- Test the Model :- Imp

Evaluate the quality of the models on data that was held out from the model building.

5- Deploy your Model :-

Use the model to generate predictions of the target answer for new data instances.

Q: What are the Key terms in machine learning?

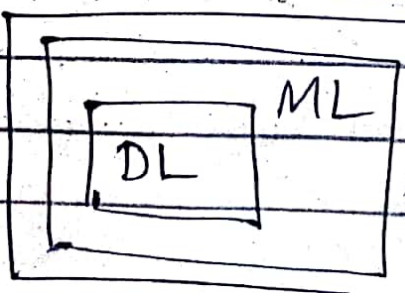
- **Model**
  The output of an ML algorithm trained on a data set, used for data prediction.
- **Training**
  The act of creating a model from past data.
- **Testing**
  Measuring the performance of a model on test data.
- **Deployment**
  Integrating a model into a production pipeline.

Q: What is the difference between ML and AI?

Artificial intelligence ingests data, such as human knowledge, and imitates natural intelligence.

Machine learning is a subset of AI, where data and algorithms continuously improve the training model to help achieve higher-quality output predictions.

Deep learning is a subset of machine learning. It is an approach to realizing ML that relies on a layered architecture mimicking the human brain to identify data patterns and train model.




DL is subset of ML. ML is technique for realizing AI.

Q What is the difference between Machine learning and classical programming?

Machine learning involves teaching a computer to recognize patterns by example, ~~wit~~ rather than programming it with specific rules.

These patterns can be found in the data. In other words ML is about creating Algorithms (or a set of rules) that learn from complex fn. (patterns) from data and make predictions on it (a form of "narrow AI")

ML learns from data and can be reused for unseen future, or new data without rewriting code.

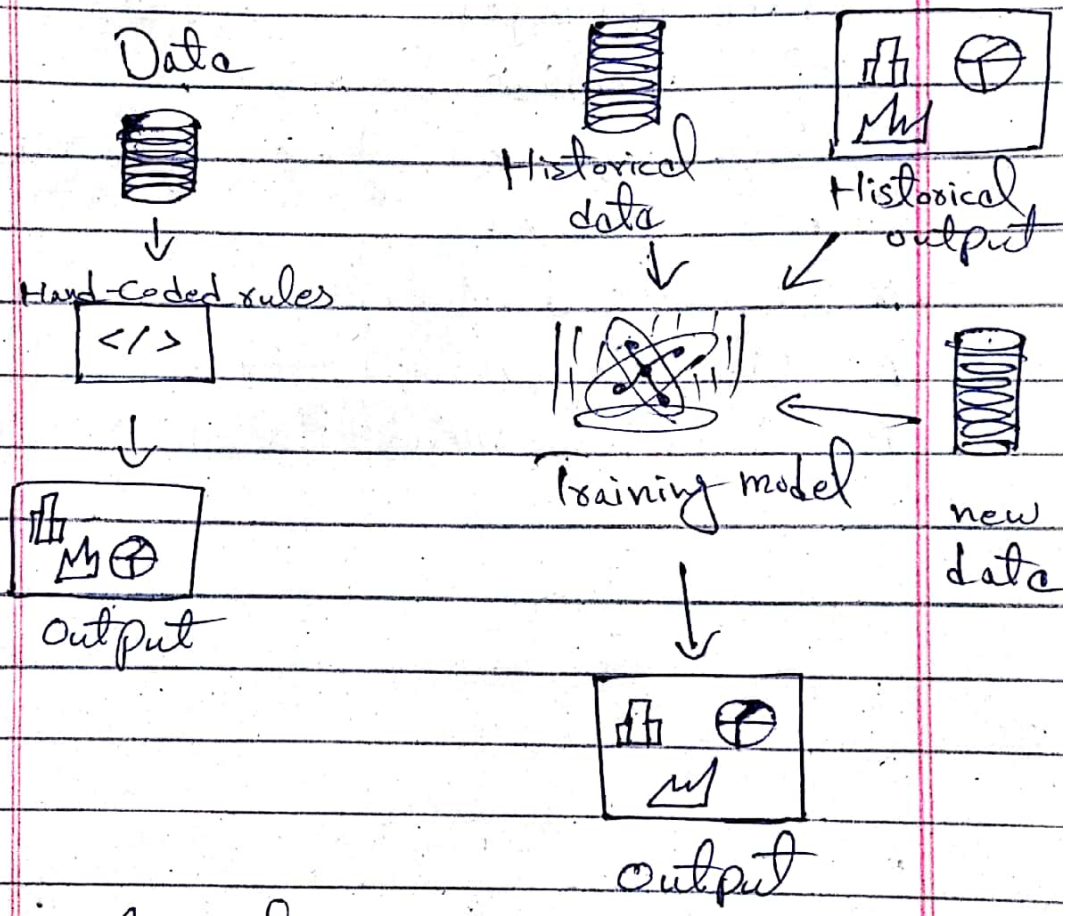
Put another way, with ML, you start with a

ML use for future predictions



problem, identify data associated with that problem, use an algorithm to then model that problem, and generate output.

Classical programming Machine learning




* Classical programming uses data and static, hand-coded rules to generate result. ML uses historical data or output, combined with

data (new), to generates new ~~data~~ rules in a training model.

Q How does Machine learning work?

Three major categories of machine learning, depending on the specific use case.

* Supervised Learing
* Un- Supervised Learning
* Reinforcement Learning

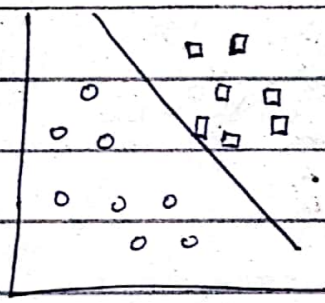
* Supervised Learning

Supervised Learning is a method where a model learns from a data set containing input values and output values (paired) that you would like to predict.

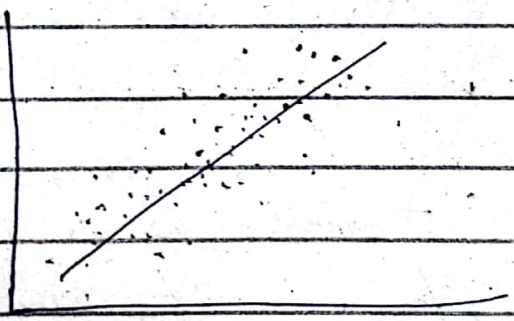
(OR) labeled data use in Supervised learning.

Examples

Classification - Classifying Documents



Regression - Forecasting demand for a product



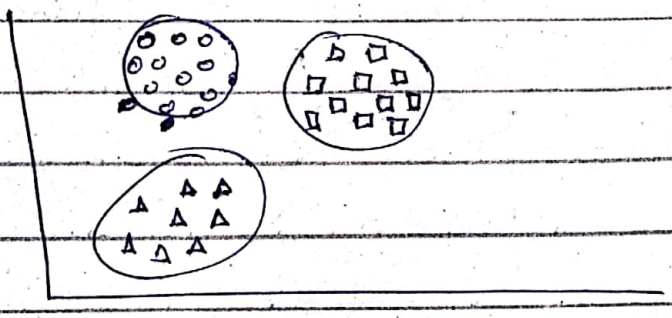
## ✱ Un-Supervised learning

Un-supervised learning is a method in which the training model learns from data without any guidance. The objective pattern and structure recognition

(or) no-label data is un-supervised learning.

Example

Clustering - Customer Segmentation



Association - Discovering regularities among products

Customers who purchase item A will usually purchase item B.

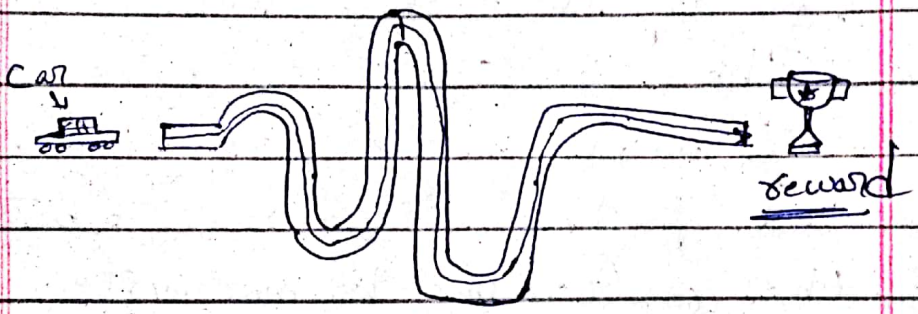
Customers who purchase items X and Y usually purchase item Z.

* Reinforcement learning

Reinforcement learning is a method in which the training model learns from its environment by being rewarded for correct moves and punished for incorrect moves.

Example.

Autonomous Driving




Q what kind of solutions can ML provide?

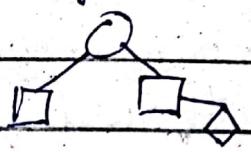
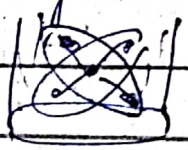
ML can provide predictive solutions (regression and classification), prioritization (ranking and scores), and behavior patterns (recommending and clustering).

Q What are Simple and Complex models?

Simple and Complex models differ when balancing a model's accuracy (number of correctly predicted data points) and a model's explainability (how much of the ML System can be explained in "human term".

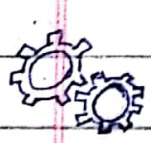
The output of a Simple ML model may be explainable and produce faster results, but the results may be inaccurate.

The output of a Complex ML model may be accurate, but the results may be difficult to communicate.

| Simple model | Complex model |
| --- | --- |
|  |  |
| ↓ Accuracy ↑ Explainable | ↑ Accuracy ↓ Explainable |

Q. What are the reasons to use machine learning?

Ans:- An example of a business problem where the use of ML would be appropriate is generating personalized recommendations. In this case, the solution to the problem requires complex logic, and we would want to provide personalized recommendations at scale with quick turnaround times.



- Requires Complex logic
  Since developing personalized recommendations requires complex logic, ML is an appropriate tool to consider.

- Requires Scalability
  Serving million of requests for personalized recommendations every second is a challenge

- Requires Personalization
  Delivering personalized recommendations at scale and being responsive at the same time is difficult to achieve with classical programming techniques.

- Requires Responsiveness
  The ability to deliver personalized ~~w~~ recommendation with ⩵ in few seconds even while handling millions requests per second is expected.

Q What is an example business case for machine learning?

Ans:- Consider Financial institution that needs to determine which category of products and offering is most interesting to a customer.

The problem might not

be effectively solved using simple hand coded rules since the outcomes might depend on many factors and overlapping rules. ML could solve this problem

Q What are the reasons to NOT use machine learning?

Ans:- Business reasons to avoid ML depend on whether traditional methods and rules are viable options, if there are few or no requirements to adapt to new data, if business goals include 100% outcomes accuracy, or if models must be explained or transparated.

- Can be solved with traditional algorithms.
- Do not require adapting to new data.

- Requires 100% accuracy
- Requires full interpretability.

Q What types of data are available today?

Ans:- ML use training data optimized for learning and generalization. Models can ingest several types of data.

- Document
- Audio
- Images
- Video
- Weather reports
- Website interaction
- Social media connections
- Industrial monitoring

Q Is my data ready for machine learning solution?

Ans:- Data readiness depends on the quality, quantity, diversity, and complexity of the data collected. After discovering and collecting all relevant data, the data should be cleansed, validate, transformed and store.

Q Is my data high Quality?

Ans:- Data used in an Machine learning project should be relevant to produce valuable results. Data should be timely so that training data is as close to the actual data as possible. Data should be representative of the data across all data sources. Data selection should be unbiased.

Explaining:-

- Relevant

  if you are developing a machine learning solution for a forecasting model, but you don't have data directly related to that goal, the data will not help you power the model.

- Fresh

  if you are constructing a predictive model for future call-center capacity, but the only data available

is from a few years ago, the data will not help you build a reliable prediction model.

- Representative
  if you are building a model to forcast sales, does the data contain all available products.
- UnBiased
  A predictive industrial model is trained using sensor data collected using one type of machine, but sensor data from other types of machine is ignored.

Q What does a machine learning lifecycle look like?

Ans:- The Machine learning lifecycle consist of nine stages. ~~Etc~~

- Problem definition
  ML requires a clear idea of the problem to be ~~sol~~ solved

and the business value gained by solving the problem.

- Data exploration

Identifying the data needed for your ML model and evaluating the various means available for collecting that data to train your model. Confirm data availability, both quantity and quality, and understand your data before preparing it for downstream consumption.

- Data preparation

Data preparation start with small statistically valid sample and iteratively improve with different data preparation while maintaining data integrity. Optimized training data is essential for model learning and generalization.

- Model Exploration
  Evaluate several key factors regarding the data being used to train the model. The amount of data, complexity and diversity, quality, freshness and bias may affect the quality of the model in development.
- Model training
  Model training requires the selection of a machine learning algorithm that is appropriate for your problem and then train the ML model. As part of training provide algorithm with representative training data to learn from and set model parameters to optimize the training process.
- Model testing
  Model testing uses different variants of your model and compare how

each variant performs relative to each other. if new version delivers performance better to the previously existing version you replace the older model.

- Evaluation

After the model has been trained, evaluate to determine if its performance and accuracy will enable to achieve your business goals. Evaluate your model using a portion of the dataset that is never used for model training or validation, but it is used to evaluate errors in the final model.

- Production Deployement

Production model takes a trained model, ingest data, and permits the resulting predictions available to user or other systems.

ML model typically takes week or even month.

- Model updates
After a model is deployed to production environment, compare certain key statistics of incoming data to those of the test data on a continuous basis. Drift b/w these statistics could call for retraining the model on newer data to maintain model quality.

"Timeline"

| | |
|---|---|
| Problem Defination | 1 week |
| Data Exploration | 1 week |
| Data preparation | 2 week |
| Model, training, testing | 3 week |
| Evaluation | 1 week |
| Deployement | 4 week. |

Q How do I take my machine learning solution Production?

A While Production deployement of an ML model is one of the ~~first~~ last stages. ML Production code differs in many ways from research code. The purpose of research code is to promote exploration and validate models using iterative processes which might lack ~~formality~~ formal Quality, stability or scaling requirements. However, Production code must meet objective and fixed requirements, facilitate Collaboration through ~~Ves~~ Version control, maintain Code deployment history and meet code reliability standards.

| Research Code | Production code |
| --- | --- |
| Exploratory | → Fixed requirement |
| Iterative | → version controlled |
| Few test or error handling | → production level reliability |

Q How do I move my organization into machine learning?

Ans:- An organization can progress from having few or no ML progress to advanced production ML expertise after completing multiple projects and deployements.

☆ Key strategy

- Robust ML strategy
- Data M'L Strategy
- A Culture of learning and Collaboration

Q How do I evaluate my data strategy?

Ans:- An effective data strategy begins with evaluating the organization's current data and determining the types of solutions needed to improve data Quality. Examples of how to evaluate data Quality ratings.

## ✡ Acceptable

- Data is raw, unlabeled and requires work before its can be used for ML.

Solution:-
Label data using tools.

- Better

Data is labeled, is accessible to select teams in the organization.

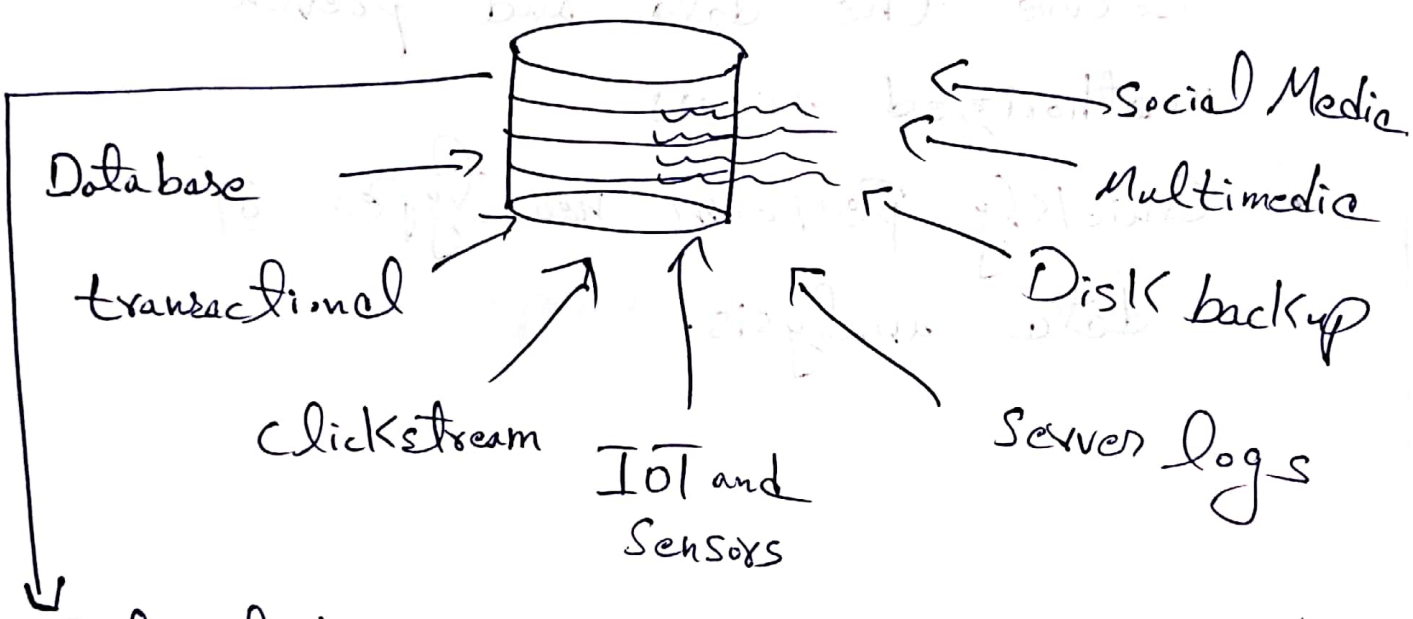
Solution
Create a strategy to unify all the data sources into one source of truth repository

- Best

Data is labeled, lives in source of truth repository, is accessible to all the teams with in the organization

Q How can I improve data strategy?

Ans:- Organization can develop strategic plans to collect and use data, even at the POC stage, to successfully change industries with ML.




Data lake are an increasingly Popular core component of the most efficient ML models. Data lake capture and store all above activity in a single core. Data lake can also grow to any scale and can also demonstrate data access and analysis.

## Benefits of data lake :-

- Catalog, search, and find the relevant data in the central repository.
- Store any type of data, at any scale, and at a low cost.
- Secure the data and prevent unauthorized access
- Quickly perform new types of data analysis.

Q :- What is a Data Scientist?

Ans :- Data Scientist work in both business and technical worlds. with deep data analysis to achieve specific outcomes. In the Machine Learning (ML), data Scientist design and build models from data, Create and work on algorithms, and train models to predict and achieve business goals.

Q :- What skills should a data Scientist have?

Ans :- A typicall data Scientist skillset consists of multiple domains. ↓

- Machine Learning

  ML is a process of training computers, using math and statistical processes, to find and recognize patterns in data. After patterns are found, ML generates and updates training model to make increasingly accurate predictions and if inferences about future outcomes based on historical and new data.

- **Computer Science**

  Computer Science is the study of algorithmic processes, computers and computational machines. Computer Science spans a range of topics including software, software systems, artificial intelligence, computer systems and networks, security database systems, human interaction, vision and graphics, numerical analysis, programming languages, and software engineering.

- **Domain Knowledge**

  Domain Knowledge is knowledge of a specific, specialized discipline or field of expertise which can be a business, industry, or data collection mechanism.

- **Big Data (7) V's**

  Big Data is a large, diverse set of information which uses tools to manage data volume, velocity, veracity, variety, value, viability, and vagueness.

- **Math / Statistics**

  Math and Statistics are collection, analysis, presentations, and interpretation of data. Also the application of probability theory which include linear algebra and differential Equations.

- **Technology Skills**

  Technology skills are abilities or expertise to perform specific, technology tasks.