Introduction to Machine Learning

Machine Learning

By Megha Singhal | 2020

Machine Learning Process

Step 1

Gathering data from various sources

Step 2

Cleaning data to have homogeneity

Step 3

Model Building.Selecting right ml algo

Step 4

Gaining insights from model results.

Step 5

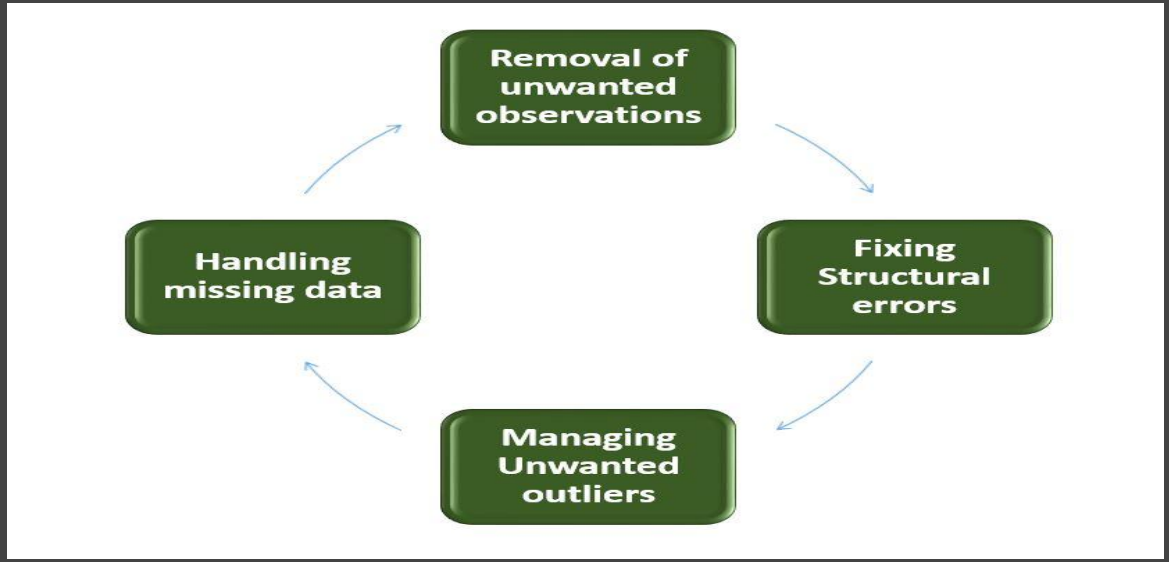
Data Visualization: result into visual graphs

Machine Learning Process

1. **Gathering Data -** Quality and quantity of data that you gather will directly determine how good your predictive model can be. Some models require continuous live fed data.
2. **Data Preparation** – Clean your data to remove invalid, inconsistent values. Split the data into training and test data sets.
3. **Choice of the ML algorithm** – Depending on factors such as nature of data (labelled or unlabelled), type of data (numerical, Audio-visual, categorical), measure of accuracy, cost of human intervention/correction.
4. **Continuous Learning of the model** – Incrementally improve the model's performance, by adjusting output parameters or rewards in each iteration. Evaluate model accuracy.
5. **Data visualization to make predictions / spot trends –** Predict expected results by running the model. Present the output in meaningful human readable forms (Tables, graphs, images, etc).

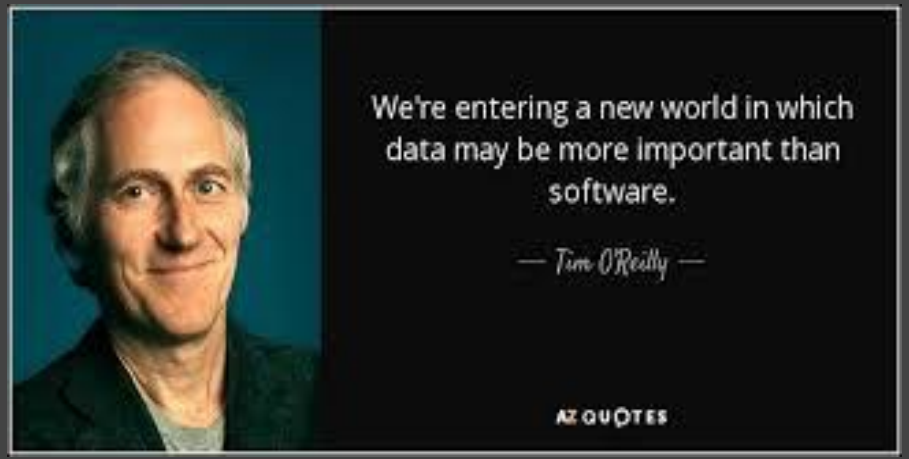
Data Cleaning

- Data cleansing or data cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database.
- Refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.



Why is it important?

- It is commonly said that data scientists spend 80% of their time cleaning and manipulating data, and only 20% of their time actually analysing it !!
- Administratively, incorrect/inconsistent data can lead to false conclusions and misdirected investments.
- In the business world, incorrect data can be costly. Many companies use customer databases that record data like contact information, addresses and preferences.



Megha Singhal

Types of Machine Learning

SUPERVISED LEARNING

- Makes machine learn explicitly.
- Data with clearly defined output is given.
- Direct feedback is given.
- Resolves classification and regression problems.

UNSUPERVISED LEARNING

- Machine learning understands data(patterns)
- Evaluation is qualitative or indirect.
- Does not predict or find anything specific.

REINFORCEMENT LEARNING

- An approach to AI
- Reward based Learning
- Machine learn from positive and negatives
- The goal is to maximize rewards.

Supervised Learning

Supervised Learning is a process of inferring a function from labelled training data. A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples.

Supervised learning problems can be further grouped into:

Regression: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Classification: A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.

A vast majority of practical machine learning uses supervised learning.

Applications - It’s the basis for features in everyday apps like search functionality in Google Photos or Voice Search, wherein ML can differentiate locations, people, and time of day without any written information.

Steps in Supervised Learning

Input: Available **Output:** Available

Predictive , Task driven model. “Labelled" data.

Train and Validate the model on training data (show required output to the model - hence the name Supervised)

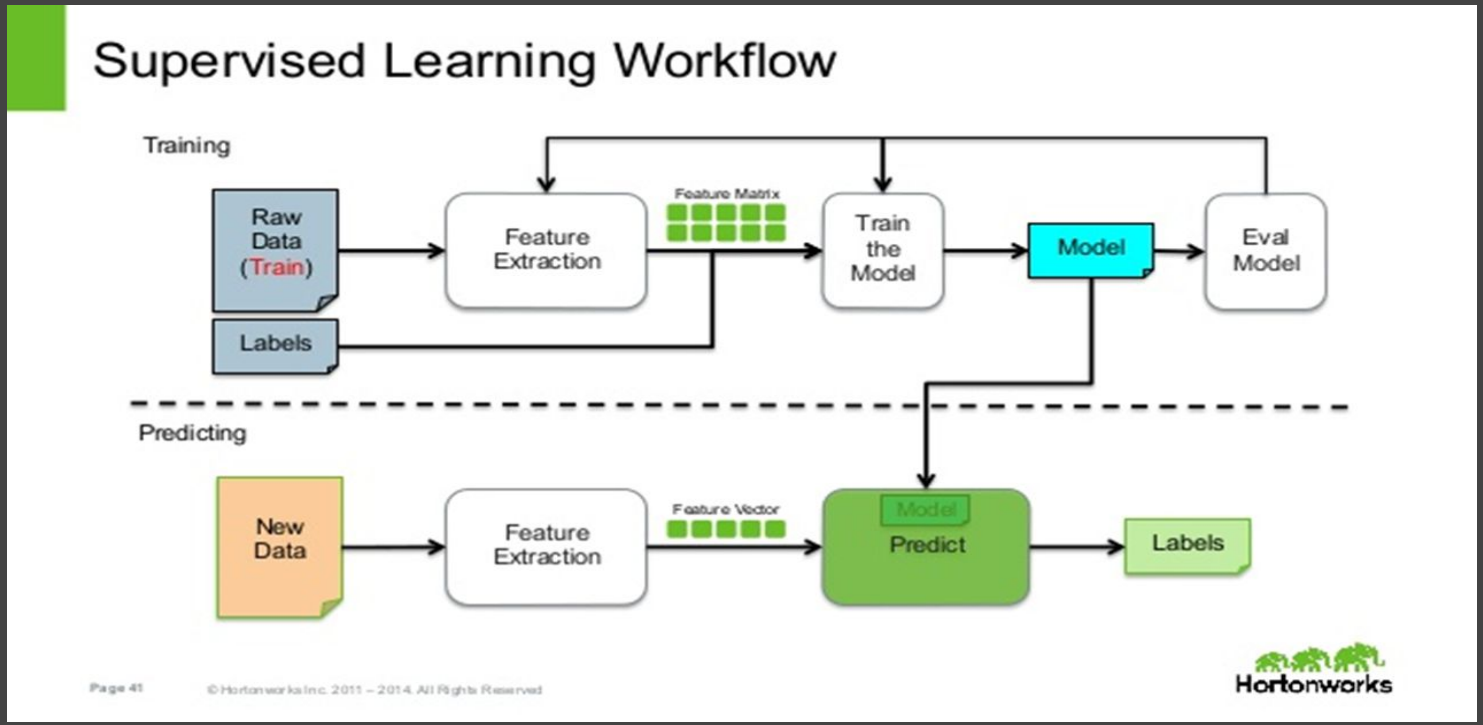
Test on a chunk of data the model has never seen (and this time don’t show the expected outputs to the model)

If the model was able to identify the output (almost matching the actual output - you hide from the model) THEN you’re ready to deploy your model

EXAMPLE - “I know how to classify this data, I just need you (the model) to do this task.”

USAGE - To classify labels or to produce real numbers

Steps in Supervised Learning



Megha Singhal

Unsupervised Learning

Unsupervised Learning is an ML technique to find patterns in data, in an exploratory manner.

The data is not labelled, which means only the input variables(X) are given with no corresponding output variables.

Algorithms are left to themselves to discover interesting patterns in the given data set.

Since data is unlabelled, there is no easy way to evaluate the accuracy of the algorithm —one feature that distinguishes unsupervised learning from supervised learning and reinforcement learning.

Unsupervised learning problems can be further grouped into:

Clustering – Grouping of similar data into groups or clusters.

Dimensionality Reduction – Compression of the data to reduce its complexity without altering its structure

Steps in Unsupervised Learning

Input: Available **Output:** Not known.

Descriptive , Data driven model. “Unlabelled" data

Train the model on training data (just explore, no idea which variables in the data are output targets - hence the name Unsupervised)

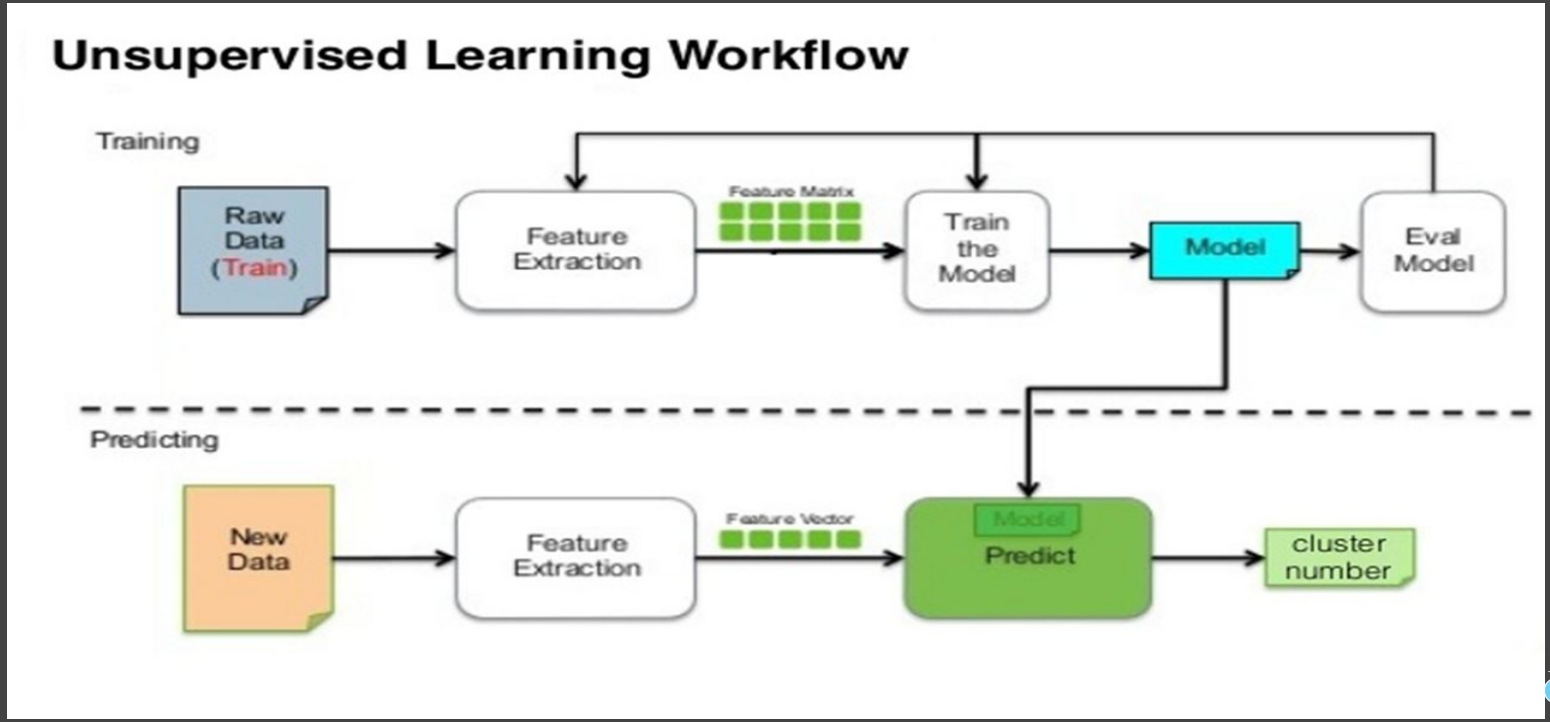
Simplify and group the data (so that it can be categorized into distinct sets)

If the model helps to identify useful real world patterns, your model is successful. Measuring accuracy of prediction is domain specific and subjective.

EXAMPLE - “I am not looking for any specific output, I just need you (the model) to classify this data meaningfully. I can tell whether the grouping is useful or not.”

USAGE - To recognize established patterns in image, voice and other rich data processing

Steps in Unsupervised Learning



Reinforcement Learning

Reinforcement learning algorithm (called the agent) continuously learns from the environment in an iterative fashion.

Aims at using observations gathered from the interaction with the environment to take actions that would maximize the reward or minimize the risk.

In the process, the agent learns from its experiences of the environment until it explores the full range of possible states.

Decision making function is used to make the agent perform an action.

After the action is performed, the agent receives a reward or reinforcement from the environment.

The state-action pair information about the reward is stored.

Some applications of the reinforcement learning algorithms are computer played board games (Chess, Go), robotic hands, and self-driving cars

Steps in Reinforcement Learning

Input: Available - Initial state (point from where the model will start) **Output:** No single output. Input state is fed into the model and observed by the agent.

Train/Validate on input data. Based on the input, the model returns a STATE. Decision making function is used to make the agent perform an action.

After action is performed, based on its output, the agent receives reward or reinforcement from the environment/user. The state-action pair information about the reward is stored.

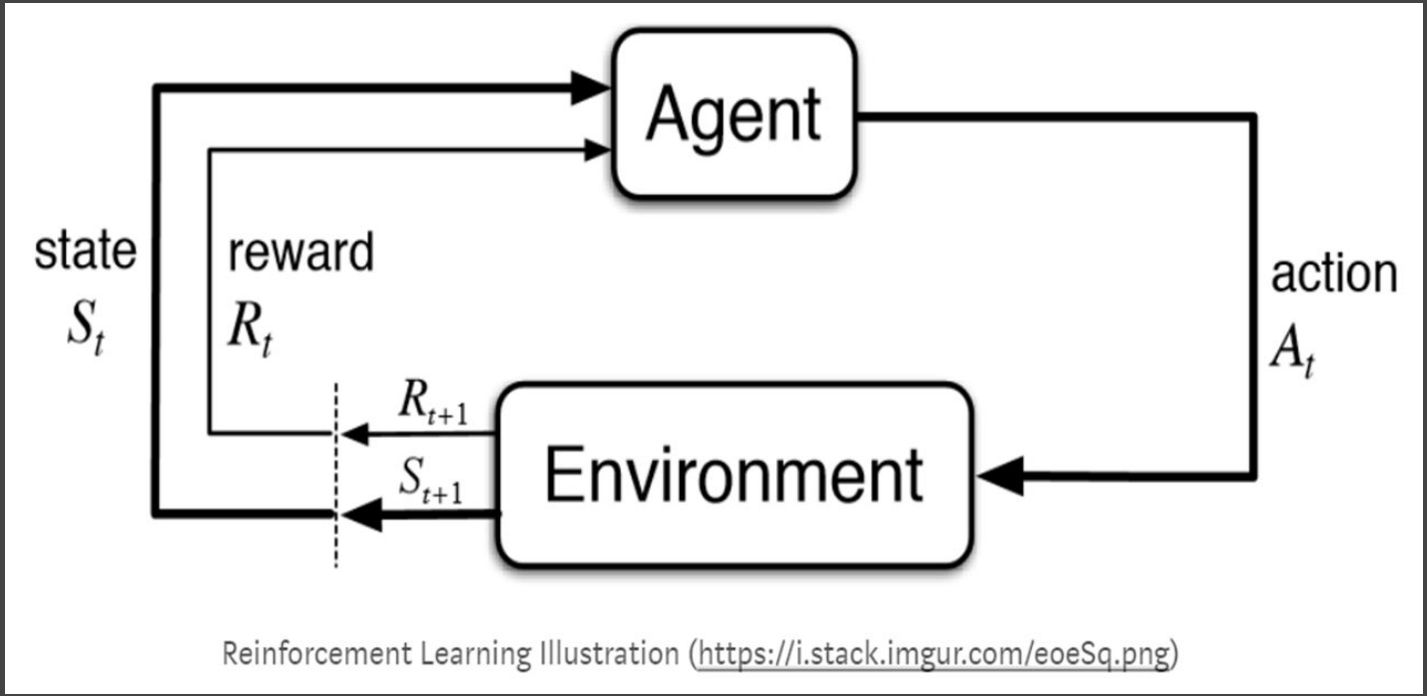
This process continues in iterations and the model continuously keeps on learning from live data. At every step, it presents actions from states.

Agent choosing the right step at each iteration is based on the Markov Decision Process.

EXAMPLE - "I don't know how to act in this environment. Can you find a good policy/behaviour and meanwhile I'll give you feedback."

USAGE - To find an optimal policy which maximizes the reward for the model/agent

Steps in Reinforcement Learning



Reinforcement Learning Illustration (https://i.stack.imgur.com/eoeSq.png)

Next tutorials on algorithms

Like and Subscribe

Thank you and Keep Learning