What to choose from? So many Machine Learning techniques

Kashif Riaz

Vice President Architecture & Engineering

Ameriprise Financial

Introduction

Something about me

Brief overview of AI

Importance of Machine Learning in various sectors

Today's agenda: Understanding the variety of Machine learning techniques in layman's terms and how to choose the right one

Well maybe a bit technical – Promise no math!

Something about me

- Passionate about human languages
- Expertise in intelligent processing for
  - non-Latin language e.g., Arabic, Chinese. Turkish, Hindi
  - resource scarce languages e.g., Persian, Urdu
  - Morphological Rich Languages (MRL) e.g., Arabic, Russian, German, Hindi, Urdu
- Enterprise Architecture of large-scale systems
- AI and Data Strategy
- Doctorate in Computer Science, specializing in Information Retrieval, NLP, Machine Learning, and computational linguistics
- Patents for question-answering systems, and AI-based reading technology for the learning-disabled, ESL, and geriatric population

Introduction to AI

- The term was coined by Marvin Minsky at Dartmouth College in 1954. He was thinking that there could be a machine that could pass the **Turing Test**
- What is a Turing test?
  - a test requiring that a human being be unable to distinguish the machine from another human being by using the replies to questions put to both.
  - Did ChatGPT pass the Turing test?

- **Eliza**: The results were not conclusive. Machine translation examples between English and Russian were poor.
  - Causes: lack of data, the complexity of languages
  - The machines were slow --> why?
  - AI-Winter

The AI Spring

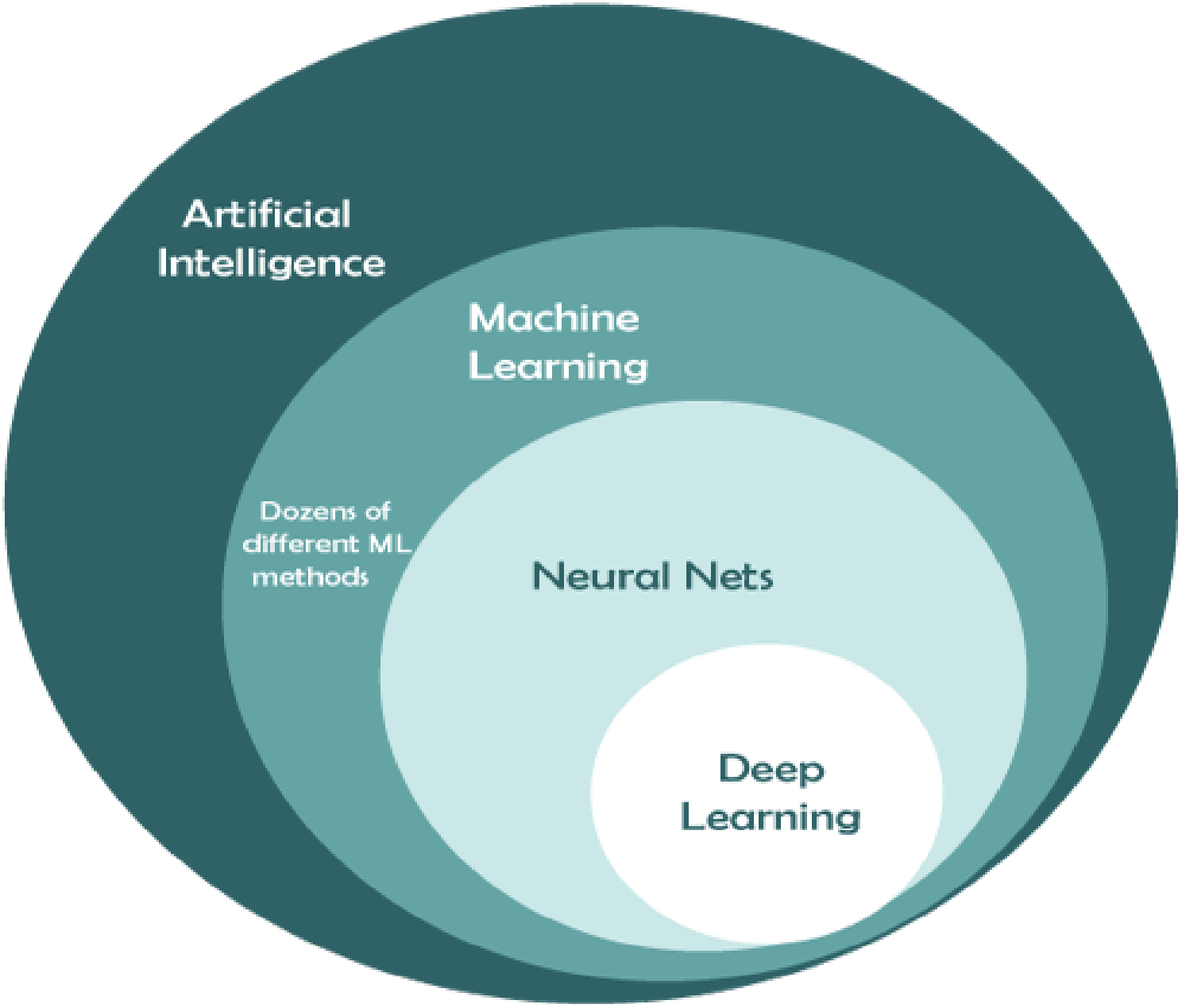
The boost in the field in AI came about as more data became available, more computing power came through, and rule-based methods started to get replaced by statistical and probabilistic-based methods.

Today, we are surrounded by AI.

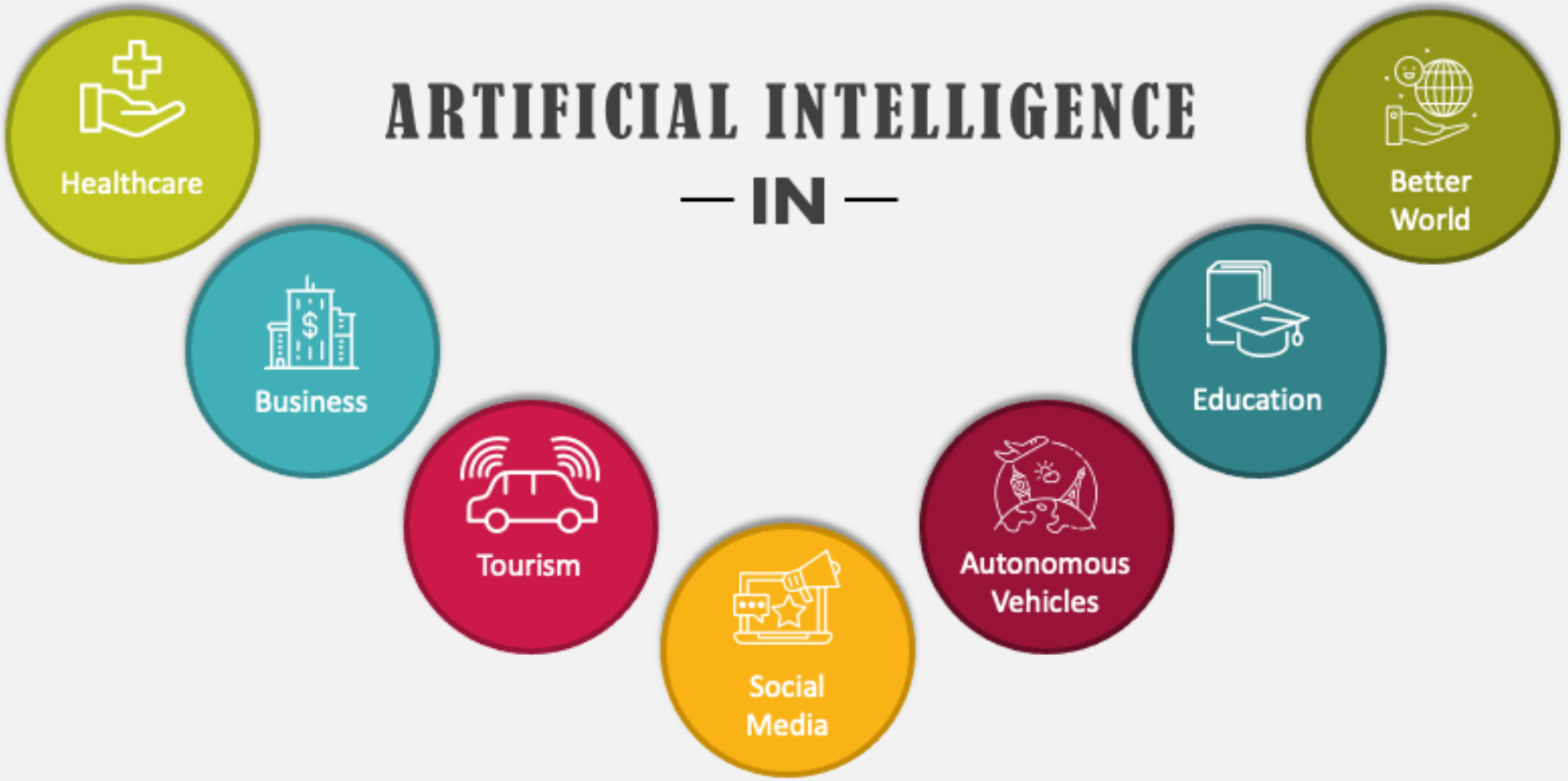
From assistants such as Amazon's Alexa

Predicting what we may like to buy next.

AI is typically defined as the ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning, and problem-solving. For example, computer vision, virtual agents, etc.


Artificial Intelligence
Machine Learning
Dozens of different ML methods
Neural Nets
Deep Learning

APPLICATIONS OF AI

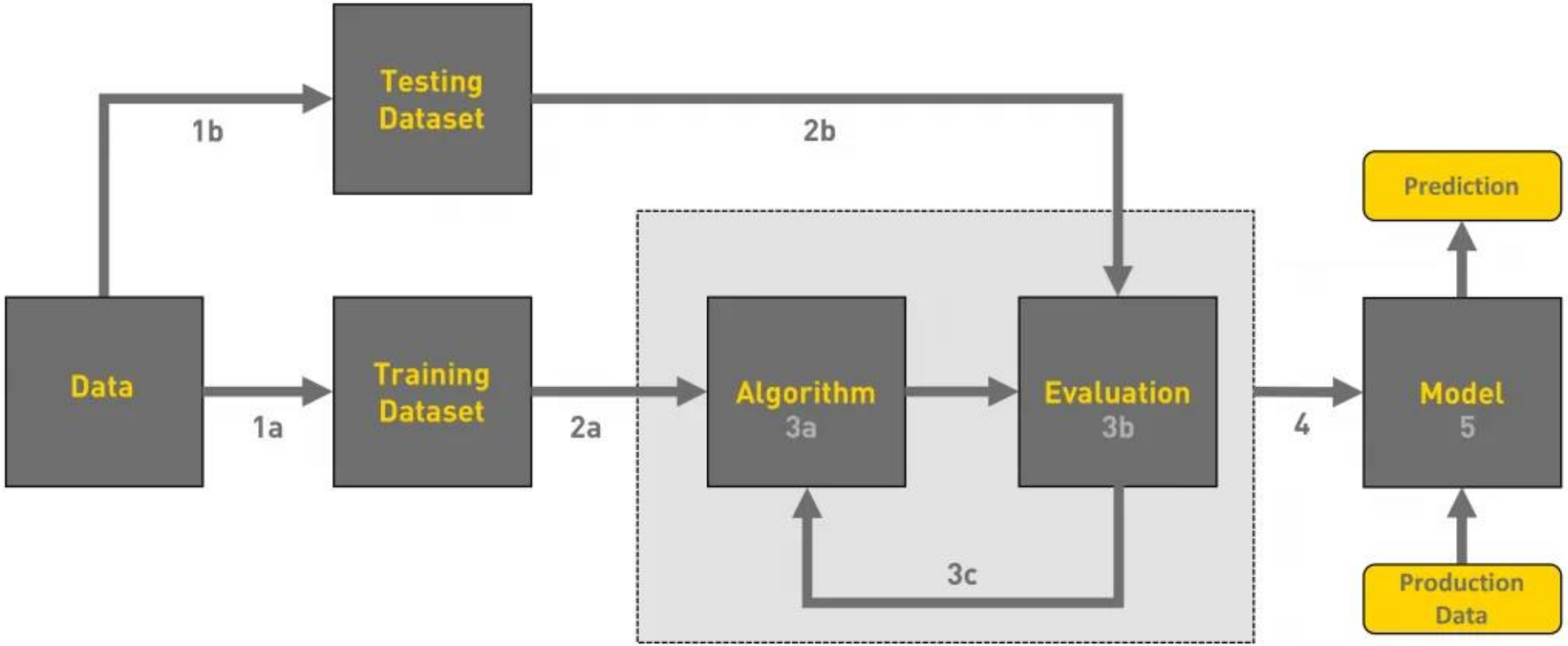


Source : 360DigiTMG.com

Machine Learning – A definition

- Machine learning is a subfield of AI that involves learning patterns from input data. The algorithm makes decisions or predictions, without explicitly being programmed to perform the task.
  - Since it is not programmed, it is a **model**
- A machine learning model is an expression of an algorithm (Nvidia)
- It is a trivial/nontrivial equation that describes a set of input data and minimizes error when unseen data is provided
- Layman's terms machine learning is like teaching computers to learn from experience
- Examples
  - Self-driving cars
  - Virtual agents – conversational bots

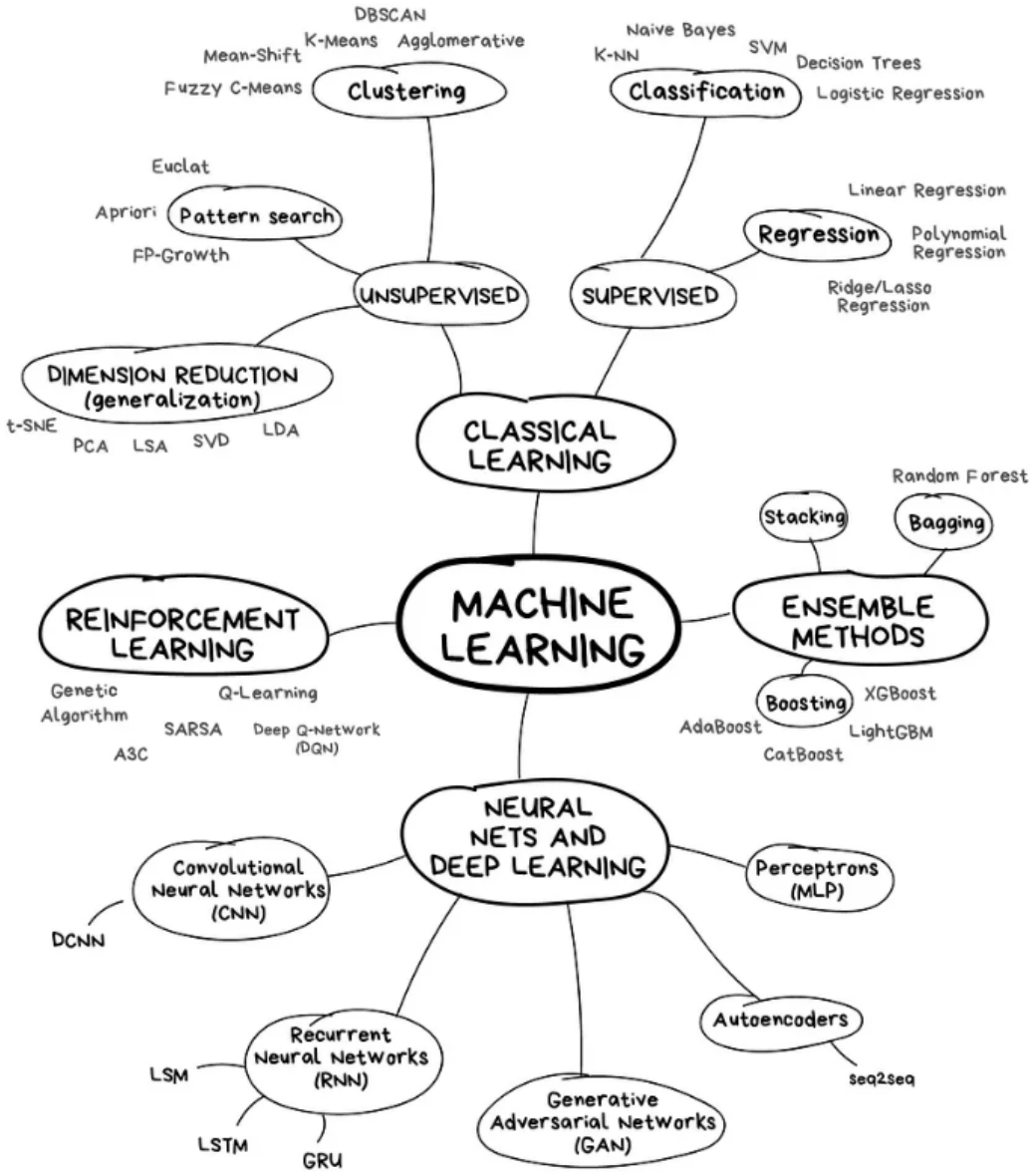
Machine Learning – Process



Why so many techniques

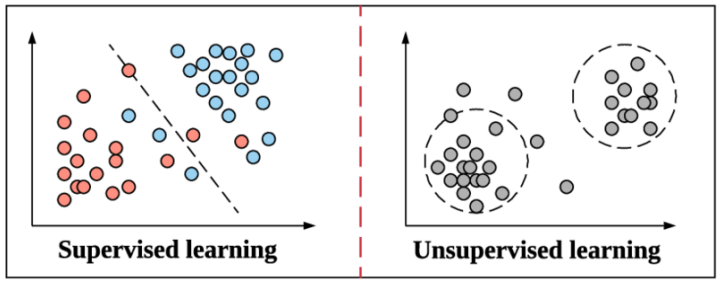
- **It is all about the data**
  - Numerical data, categorical data, textual data, image data
  - Sparse data
  - Data quality
  - Complicated features
- **Diverse applications**
  - Mission-critical applications
  - Different tools for different tasks in the toolbox, we use different techniques for different problems
- **Compliance, explainability**
- **Cost of training**
  - GPU
  - Data accumulation
  - Annotations

The world of Machine Learning

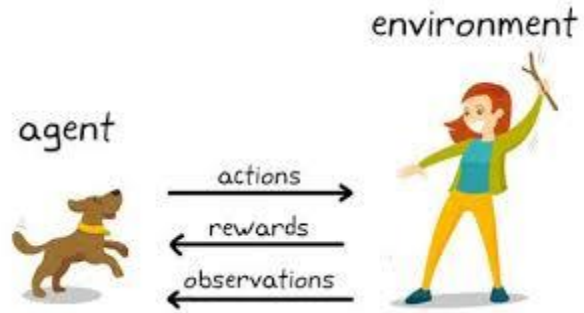


What we will cover today

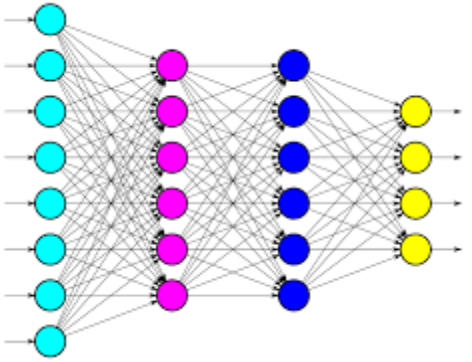
- **Classical Machine Learning**
  - Supervised Learning
  - Unsupervised Learning
  - Semi-supervised Learning



- **Reinforcement Learning**



- **Neural Networks**
  - Deep Learning



Supervised Learning

What it is

An algorithm that uses training data and feedback from humans to learn the mapping of inputs to a given output.

For example, housing prices increase, when interest rates decrease

When to use it

When you have a mechanism to label the input data for training.

We know which behavior we want to predict on unseen data

How it works

A human labels the data for training to create a model.

If the fish swims upstream and is pink in color, then it is salmon

Once the training is complete, try it on unseen data and fix errors if present

Algorithms

- Linear Regression
- Logistic Regression
- Discriminant Analysis
- Decision Trees
- Naïve Bayes
- Support Vector Machine
- Random Forest
- AdaBoost
- Gradient-boosting trees
- Simple Neural network

Supervised Learning models and Business use cases

Algorithm	Description	sample business use case
Linear Regression	A method to understand the correlation	Optimize price points
Logistic Regression	Extension of linear regression that's used for classification tasks. Output is binary	Predict if the skin lesion is benign by looking color, shape, size etc.
Linear / Quadratic Discriminant	Upgrades a logistic regression to deal with non-linear problems. The correlation is not linear with outputs	Predict client churn Predict a sales likelihood of closing
Decision Tree	Classification that splits the data into branches	Decision tree for interview screening
Support Vector Machines	Draws an optimal division between classes	Predict how likely Ad will be clicked
Naïve Bayes	Determines probability based on previous events	Spam filters: Misspelling in the email
Random Forest	Generates multiple decision trees and takes the majority vote of them to predict the output	Predict call volumes in call centers.
AdaBoost	Generates multiple models and picks the best one	Fraud detection on credit card
Gradient Descent	Creates multiple decision trees, each fixing the error of the previous one. Output is a prediction from all trees	Forecast product demand and inventory levels
Simple Neural Network	Simple NN, of an input layer, a hidden layer, and an output layer to classify	Prediction if a patient will join the health care program.

Unsupervised Learning

What it is

An algorithm that explores input data when there is not output class is given.

Discovers groupings and identify patterns

For example, customer demographic data or topics in news feeds.

When to use it

When you don't have labeled data

When you don't know the output classes.

You want to discover the patterns. Think of the night sky

The cost of labeling the data is too high

How it works

The algorithm receives the unlabeled data containing e.g. thousands of news articles.

It infers as structure from the data. The algorithm identifies groups of data **(clusters)** of topics

Algorithms

- K-means clustering
- Gaussian mixture model
- Hierarchical clustering
- Recommender systems – Netflix
- Agglomerative Clustering
- Principal Component Analysis
- Latent Semantic Analysis

Unsupervised Learning models and Business use cases

Algorithm	Description	sample business use case
K-means Clustering	Categorizes data into **K** number of groups. Each group contains data with similar characteristics automatically determined by the algorithm	Segment customers into groups based on age, and income for marketing campaign.
Gaussian Mixture model	The input data is assumed to be generated from several Gaussian distributions	Segment employees based on likelihood of attrition
Hierarchical Clustering	Unlike k-means, it does not require the number of clusters to be specified in advance. It creates a tree of clusters called a dendrogram	Cluster loyalty card customers into progressively more micro-segmented groups
DB Scan	Identifies the cluster based on high high-density data points	Social network analysis, Movie watching preferences, anomaly detection
Principal component analysis	It is a dimensional reduction technique that transforms data into a new coordinate system	Clustering of the night sky. SETI program

Reinforcement Learning

What it is

- An algorithm that learns the task by maximizing the rewards it receives from its actions.
- Maximizing points, it receives on return on investments
- Search engine algorithm as it improves by getting feedback from users

When to use it

- When you don't have lots of training data.
- The end state is not clearly defined.
- The only way to learn about the environment is through interacting with it.

How it works

- The algorithm (**agent**) takes an **action** on the **environment**
- Trading a financial stock and learning that you have increased the asset value
- The algorithm optimizes over time by taking the best series of actions to maximize rewards.

Algorithms

- Q-Learning
- Deep Q-networks (DQN)
- Policy Gradients
- Actor Critic Methods
- Proximal Policy Optimization (PPO)
- Temporal Difference Learning

Reinforcement Learning Business use cases

- Optimize trading strategy for an options-trading portfolio
- Rob advisors
- Balance the load of electricity grids in varying demand cycles
- Self-driving cars
- Optimize pricing in real-time for online auction of a product with limited supply
- Uber or Lyft pricing

Deep Learning

What it is

An algorithm based on how the human brain works.

It is based on neural network design.

It can process large amounts of input data.

Learns like a child through examples

When to use it

When the problem is increasingly complex

The output or results require high accuracy

How it works

The algorithm mimics a human brain where the information is stored in multiple layers.

Each layer gives more context and is connected through nodes

The connection between the nodes has weights that are learned by examples.

Algorithms

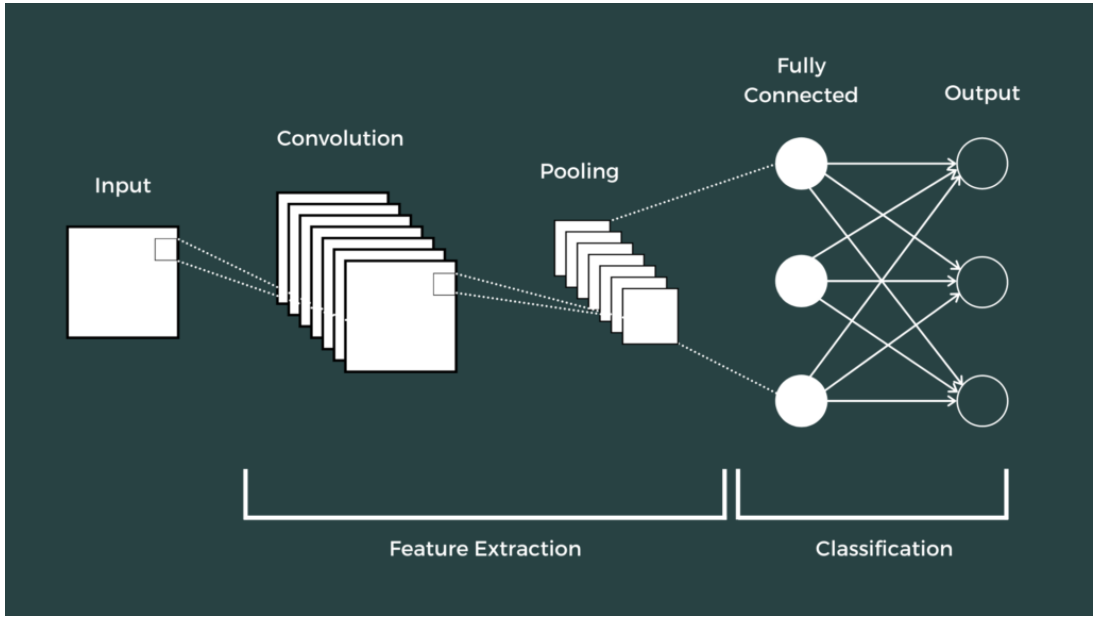
- Convolution Neural Network
- Recurrent Neural Network

Convolutional Neural Network (CNN)

What it is	When to use it	How it works
A multi-layered neural network with a special architecture to extract complex features at each layer to determine the output	Unstructured data set • Images recognition • Anomaly detection • Facial recognition • NLP • Video Analysis	A CNN takes images of cats as an input. It processes all the pixels in the image. This image is received through the input layer The hidden or inner layers of the model identify unique features like eyes, a tail, whiskers The CNN can now classify a different image of a cat with a different color. If it identifies the unique features that it learned. It identified it as a cat

Convolution Neural Network Business use cases

- Diagnose health disease from a medical scan
- Detect company logo in social media
- Sentiment analysis of customers by reading faces after trying a product
- Recognizing detective products in a assembly line



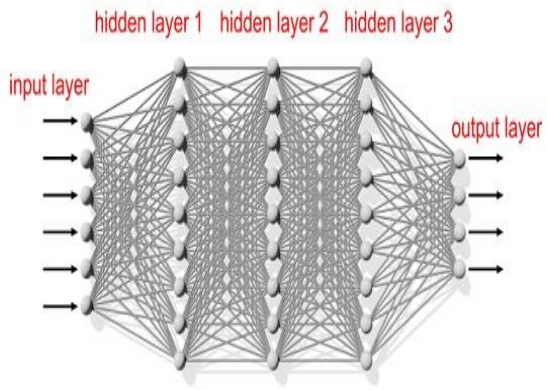
Recurrent Neural Network (CNN)

What it is

- A multilayered neural network that stores its learnings and information in the context nodes.
- It can learn data sequences and the output is another sequence or a value

When to use it

- Time Series data or text or audio sequences



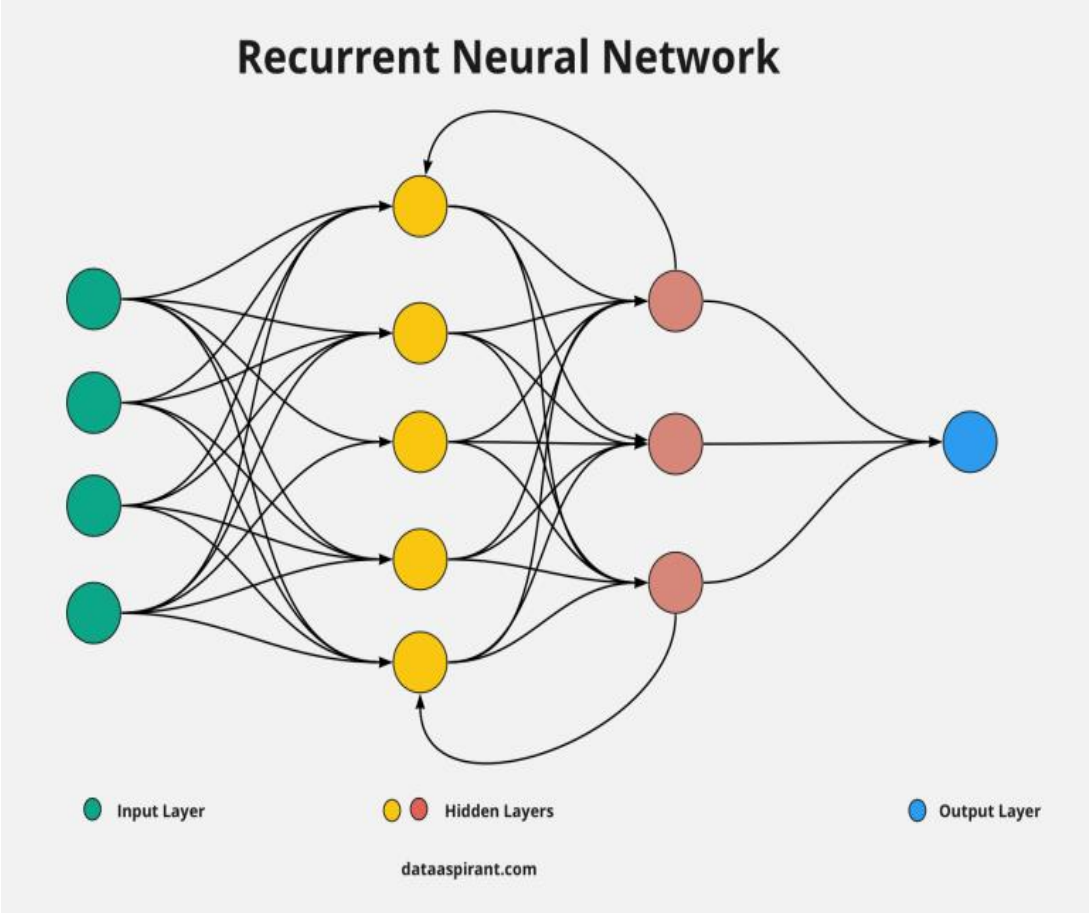
deep neural network

How it works

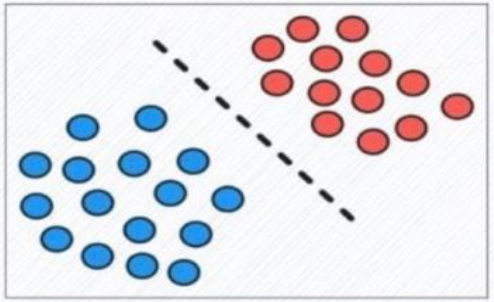
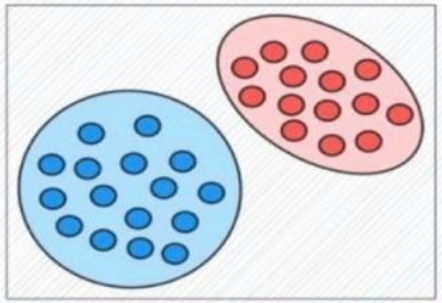
- When we are trying to predict the next word in the sentence. "Are you available …….?"
- The RNN receives the first letter "Are" and changes it into a vector. This vector is stored as the first input of the sequence. The process repeats for 2nd (you) and the 3rd sequence (available).
- After seeing "available". The nodes assign probability from the complete English vocabulary of what could be the next possible word to complete the sequence
- For example, tomorrow will have a higher probability than yesterday

Recurrent Neural Network Business use cases

- Generate analysis report for securities trade
- Provide language translation
- Summarize earning report
- Generate captions for image
- Power chatbots to answer more nuanced questions



Machine Learning – Statistical Distinction

Discriminative Models	Generative Models
	
Learns the decision boundary between classes	Learns the input distribution
Maximizes the conditional probability: P(Y|X)	Maximizes the joint probability: P(X, Y)
Directly estimates P(Y|X)	Estimates P(X|Y) to find P(Y|X) using Bayes' rule
Cannot generate new data	Can generate new data
Specifically meant for classification tasks	Typically, they are NOT used to solve classification tasks
Discriminative models don't possess generative properties	Generative models possess discriminative properties
Logistic Regression, Random Forests, SVMs	Hidden Markov Models, Naive Bayes, Gaussian Mixture Models
Neural Networks, Decision Tree, kNN	Gaussian Discriminant Analysis, LDA, Bayesian Networks

Conclusion

- Always try to understand the problem you are trying to solve
- Understand the error tolerance of the solution
- Do you want/require a “man in the middle” solution?
- Understand the regulatory and compliance requirements
- Do you want a disruptive solution or want to stay with the pack?
- If multiple algorithms can solve a business problem, always try the simplest one first

Questions