Demystifying AI - Day 1

Topics

- AI in our everyday lives
- Modern AI
- What is AI?
- Data
- Subsets of AI
- Introduction to ML algorithms
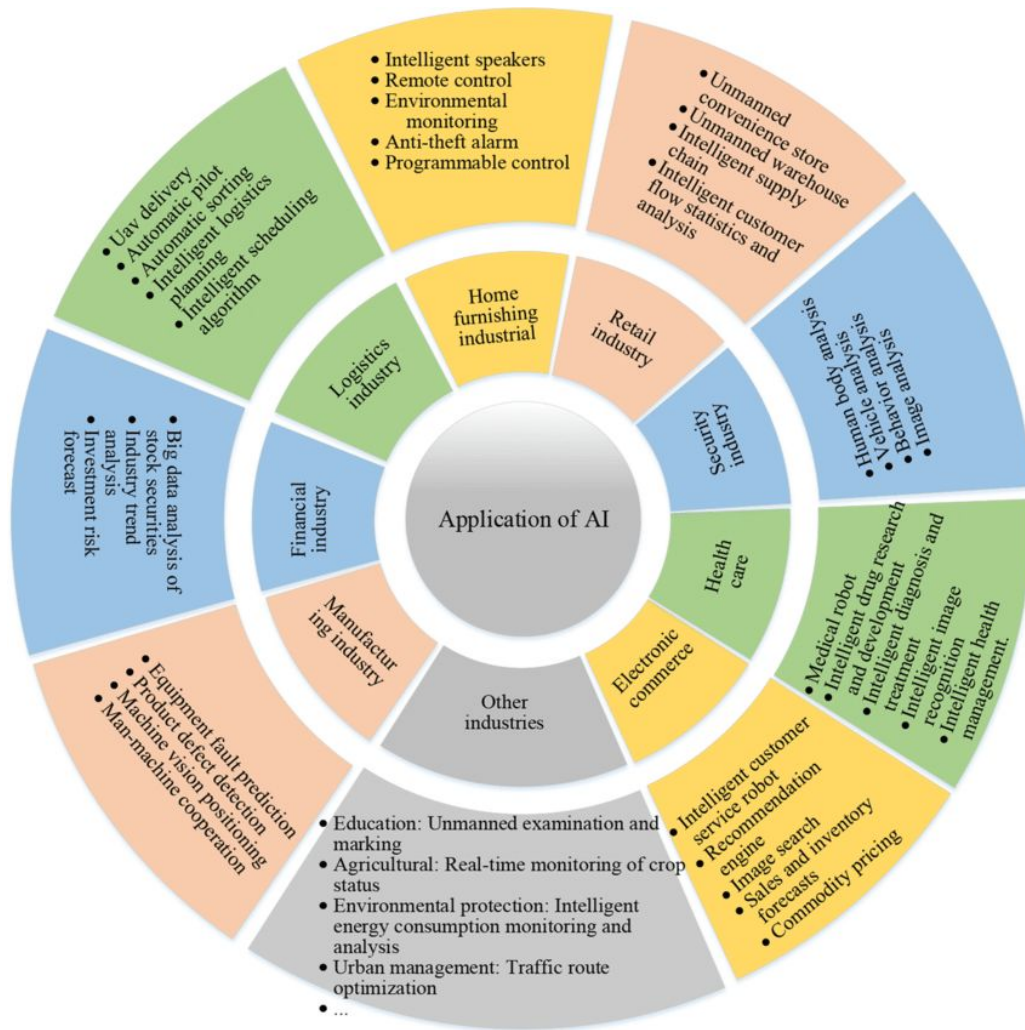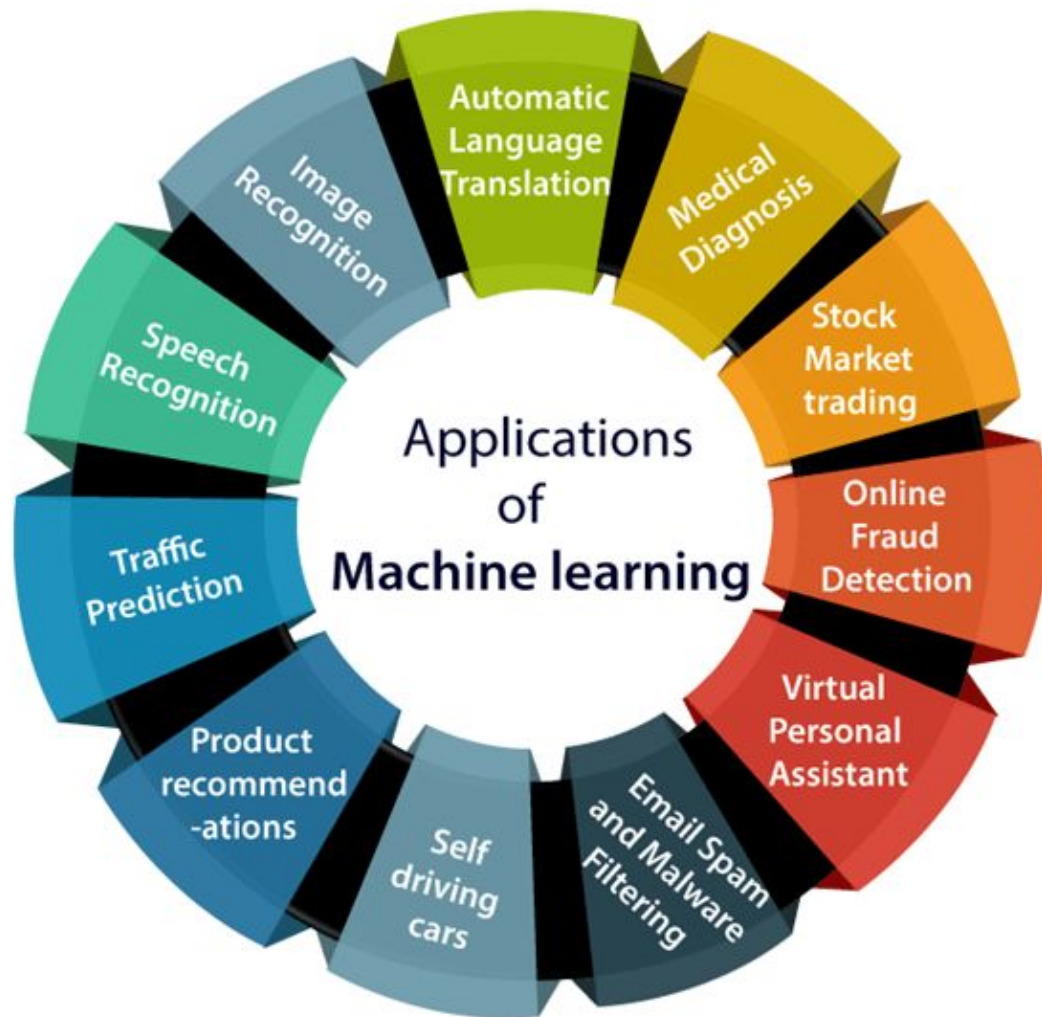- Identifying & Productizing ML Use-Cases
- Conclusion

Goals

- Understand and notice AI in our daily lives
- Distinguish between the different types of AI tasks
- Understand data in the context of AI
- Conceptualize what algorithms are used on different types of data
- Be able to recognize good ML use-cases

AI in our everyday lives

Application of AI

- Home furnishing industrial
  - Intelligent speakers
  - Remote control
  - Environmental monitoring
  - Anti-theft alarm
  - Programmable control

- Retail industry
  - Unmanned convenience store
  - Unmanned warehouse
  - Intelligent supply chain
  - Intelligent customer flow statistics and analysis

- Logistics industry
  - Uav delivery
  - Automatic pilot
  - Automatic sorting
  - Intelligent logistics planning
  - Intelligent scheduling algorithm

- Security industry
  - Human body analysis
  - Vehicle analysis
  - Behavior analysis
  - Image analysis

- Financial industry
  - Big data analysis of stock securities
  - Industry trend analysis
  - Investment risk forecast

- Health care
  - Medical robot
  - Intelligent drug research and development
  - Intelligent diagnosis and treatment
  - Intelligent image recognition
  - Intelligent health management.

- Manufacturing industry
  - Equipment fault prediction
  - Product defect detection
  - Machine vision positioning
  - Man-machine cooperation

- Electronic commerce
  - Intelligent customer service robot
  - Recommendation engine
  - Image search
  - Sales and inventory forecasts
  - Commodity pricing

- Other industries
  - Education: Unmanned examination and marking
  - Agricultural: Real-time monitoring of crop status
  - Environmental protection: Intelligent energy consumption monitoring and analysis
  - Urban management: Traffic route optimization
  - ...

Applications of Machine learning

- Automatic Language Translation
- Medical Diagnosis
- Stock Market trading
- Online Fraud Detection
- Virtual Personal Assistant
- Email Spam and Malware Filtering
- Self driving cars
- Product recommend-ations
- Traffic Prediction
- Speech Recognition
- Image Recognition

Recommendation Systems

Examples

- Amazon
- YouTube
- Netflix
- Spotify
- LinkedIn
- Pandora

Benefits

- Increased conversion
- Reduced churn
- Overall increased customer satisfaction

Recommended for you, Thomas

Literature & Fiction
62 ITEMS

Exercise & Fitness Equipment
8 ITEMS

Health, Fitness & Dieting Books
37 ITEMS

Tableware
12 ITEMS

Prime Video – Unlimited Streaming for Prime Members
12 ITEMS

Coffee, Tea & Espresso
98 ITEMS

Biographies & Memoirs
17 ITEMS

Engineering Books
7 ITEMS

Computer Vision & Natural Language

- Self-driving cars
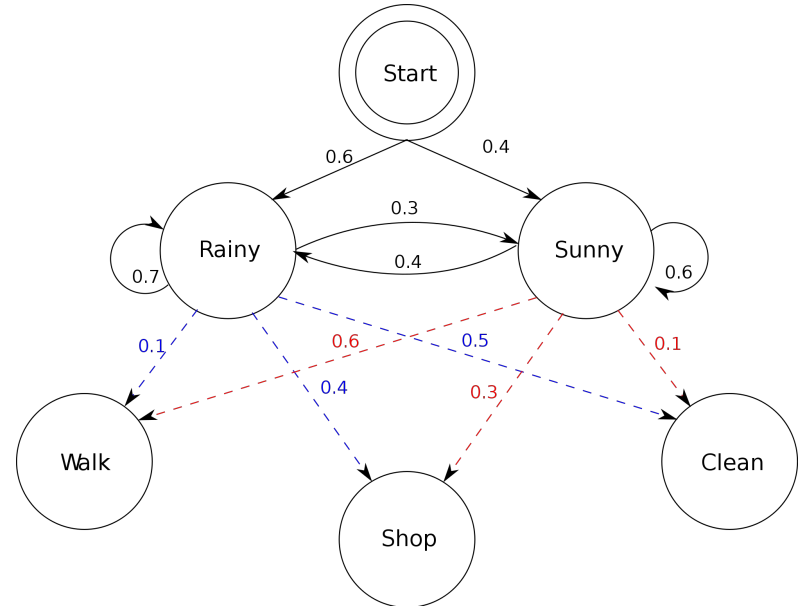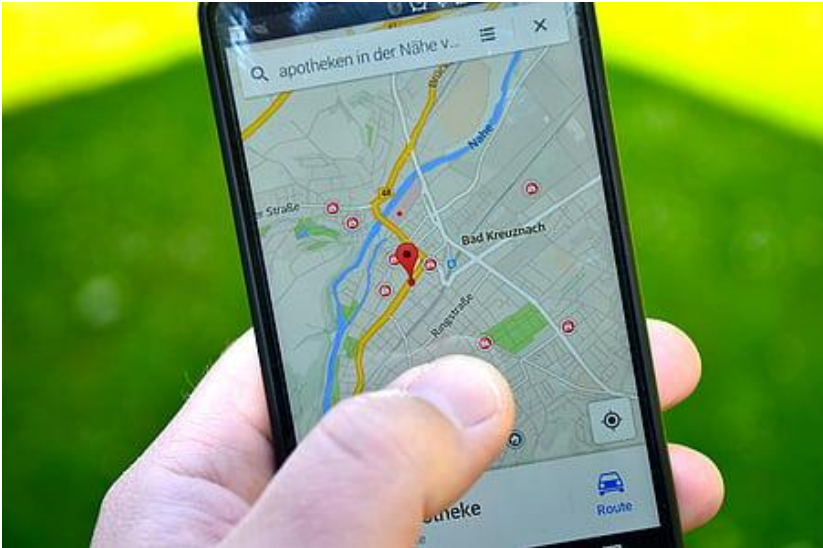- Social media filters (instagram/snapchat)
- Siri
- Alexa

Inventory Optimization

- Grocery Shopping
- Online Shopping
- Supply-chain management

GPS Navigation

- How do our phones always know the fastest way to get from point A to B? AI, of course!
- Modern GPS navigation systems such as Waze and Google Maps are utilizing GNNs (graph-neural-networks).

Fraud Detection

- Fraud detection in financial transactions, loan approvals, insurance rates & more
- When your bank calls and says 'is this really you', yes, that is AI!
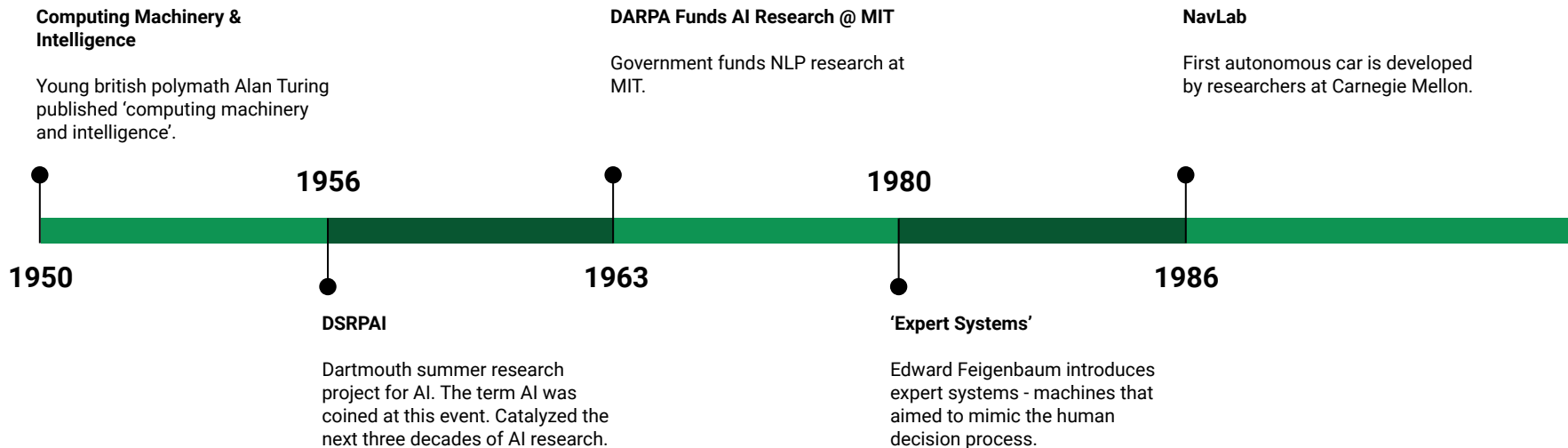- When emails get sent to your spam folder, that is AI!

Modern AI

Timeline

Computing Machinery & Intelligence

Young british polymath Alan Turing published 'computing machinery and intelligence'.

1956

1950

DARPA Funds AI Research @ MIT

Government funds NLP research at MIT.

1980

1963

NavLab

First autonomous car is developed by researchers at Carnegie Mellon.

1986

DSRPAI

Dartmouth summer research project for AI. The term AI was coined at this event. Catalyzed the next three decades of AI research.

'Expert Systems'

Edward Feigenbaum introduces expert systems - machines that aimed to mimic the human decision process.

Source: The History of Artificial Intelligence - Science in the News (harvard.edu)

The Only Constant is Change

- In the field of AI/ML - you are a lifelong learner
- LLMs (GPT3, Alexa), Diffusion (DALL-E, Stability AI)
- Prompt Engineering

Text-to-Image with Stable Diffusion
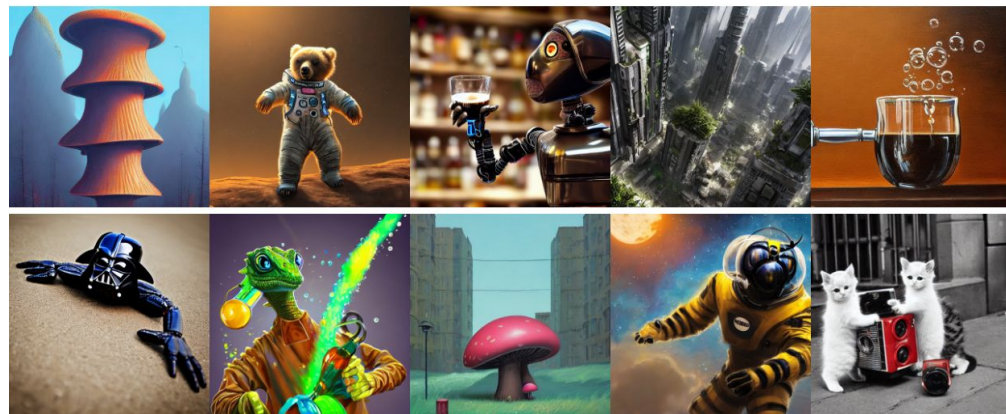
Prompt

Summarize this for a second-grade student:

Jupiter is the fifth planet from the Sun and the largest in the Solar System. It is a gas giant with a mass one-thousandth that of the Sun, but two-and-a-half times that of all the other planets in the Solar System combined. Jupiter is one of the brightest objects visible to the naked eye in the night sky, and has been known to ancient civilizations since before recorded history. It is named after the Roman god Jupiter. [19] When viewed from Earth, Jupiter can be bright enough for its reflected light to cast visible shadows,[20] and is on average the third-brightest natural object in the night sky after the Moon and Venus.

Sample response

Jupiter is a planet that is bigger than all the other planets in our solar system and is very bright when you see it in the night sky. It is named after the Roman god Jupiter. When viewed from Earth, it is usually one of the three brightest objects in the sky.

What is AI?

AI = COMPUTERS + MATH + DATA

- Statistics (probabilities)
- Linear Algebra (vectors, matrices)
- Calculus (optimization)

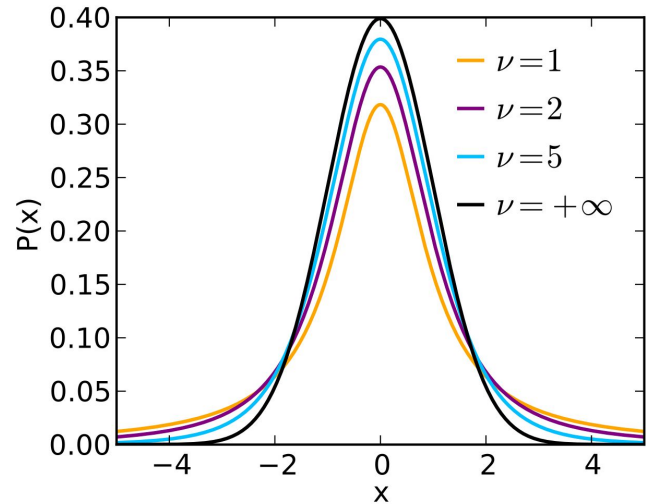Probability Density Function

$$A = [x0, x1, x2, \ldots, xn]$$

$$B = [y0, y1, y2, \ldots, yn]$$

$$A \odot B = \sum_{i=0}^{n} xi \cdot yi$$

Dot Product

$$H(X) = H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Shannon Entropy Formula

How Does It Work?

- **<u>Optimization</u>**
  - Minimizing or Maximizing a function
- Give algorithm input data, and labeled output data
- Randomly initialize parameters
- Iteratively adjust parameters until the mathematical "distance" between algorithm output and labeled output is small.

$$y = mx + b$$

Simple Regression

$$MSE = \frac{1}{n} \Sigma \left(y - \widehat{y} \right)^2$$

The square of the difference between actual and predicted

Measure of "incorrectness"

Trainable Parameters

- In the previous slide, it was mentioned we iteratively adjust our parameters to tighten the gap between the predicted output and the desired output.
- **<u>Example:</u>**
  - X = [1.2, 1.7, 2.9, 4]
  - Y = [0]

$$\hat{y} = X \bullet W + b$$

$$\hat{y} = [1.2, 1.7, 2.9, 4] \bullet [-0.1, -0.1, -0.1, -0.1] + [0, 0, 0, 0]$$

$$\hat{y} = -0.98$$

Now we adjust our 'W' and 'b' parameters to minimize the MSE, and run the formula again. This is called training.

$$MSE = 0.9603999999999999 = (0 - (-0.98))^2/1$$

Data

What is data in AI?

- Data is the most vital piece of any machine learning algorithm
- Everything you do everyday, generates data that can be utilized by a ML algorithm
- **Fun Facts:**
  - 2.5 quintillion bytes of data created everyday
  - Humans on avg generate 1.7MB of data per second

Source: How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read (forbes.com)

Our current love affair with social media certainly fuels data creation. According to Domo's Data Never Sleeps 5.0 report, these are numbers generated **every minute** of the day:

- Snapchat users share 527,760 photos
- More than 120 professionals join LinkedIn
- Users watch 4,146,600 YouTube videos
- 456,000 tweets are sent on Twitter
- Instagram users post 46,740 photos

Different Types of Data

- Structured Data & Unstructured data
- **Structured Data**
    - Tabular data (any data in table format)
    - Think excel
- **Unstructured Data**
    - Documents (PDFs/invoices/receipts)
    - Text data (articles, books, reviews, comments, tweets, etc)
    - Image data (images/videos)
    - Audio data (mp3/wav)

Data Sources

- **YOU** are the data source
- Here is an example:
  - You are planning a vacation - you google different hotels, routes, and activities
  - You go on the trip, download a book on audible, order ubers, go shopping, maybe eat some nice meals, post pictures, tweet about it, etc
- Every step above generated tons of data that will be used to sell you things, show you ads, and essentially aim to "program" you.
- Companies of all kinds across all industries will purchase this data, and then develop ML algorithms to increase business capacity or reduce costs

Data Storage

- Generally, we are dealing with large volumes of data, too large to maintain on a local machine
- Companies have different systems
    - CRM (salesforce, SAP)
    - SQL (GraphQL, PostgreSQL)
    - Cloud data-store (Azure, AWS, GCP)
- When building an ML model, we generally need to access multiple types of data, from multiple systems. The more data the better the algorithm.
- **Accessing the data**
    - API
    - Queries
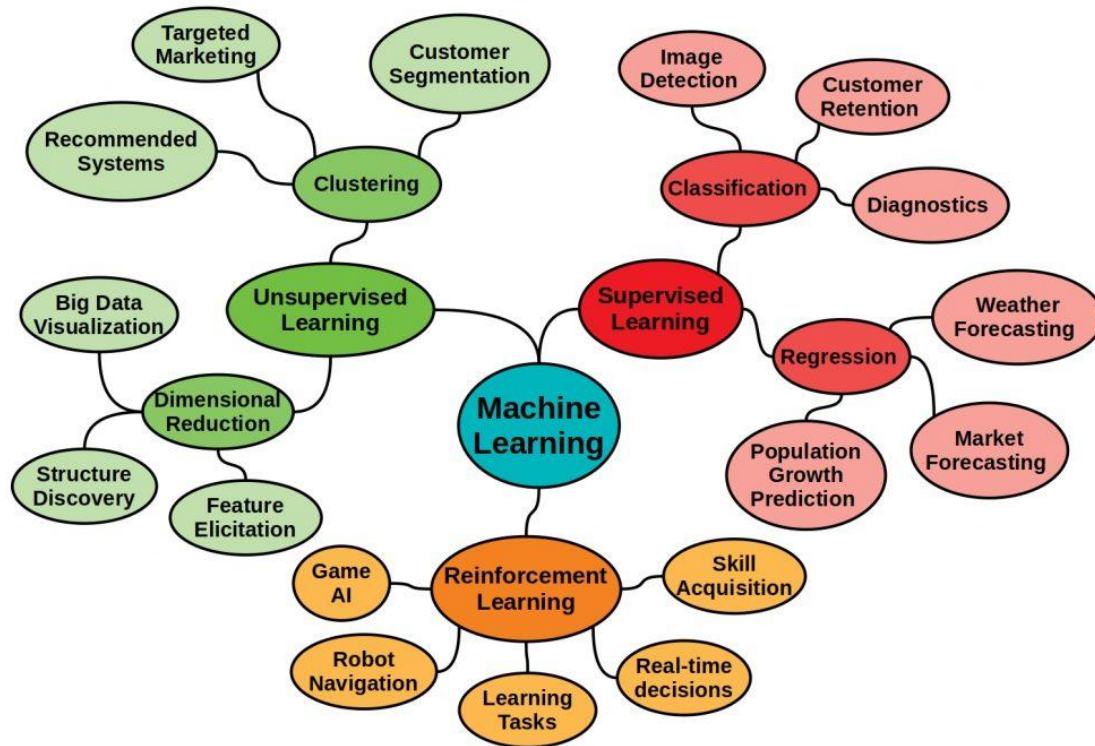   - & sometimes it's as simple as hitting 'export'! Thank you SalesForce!

Subsets of AI

Overview
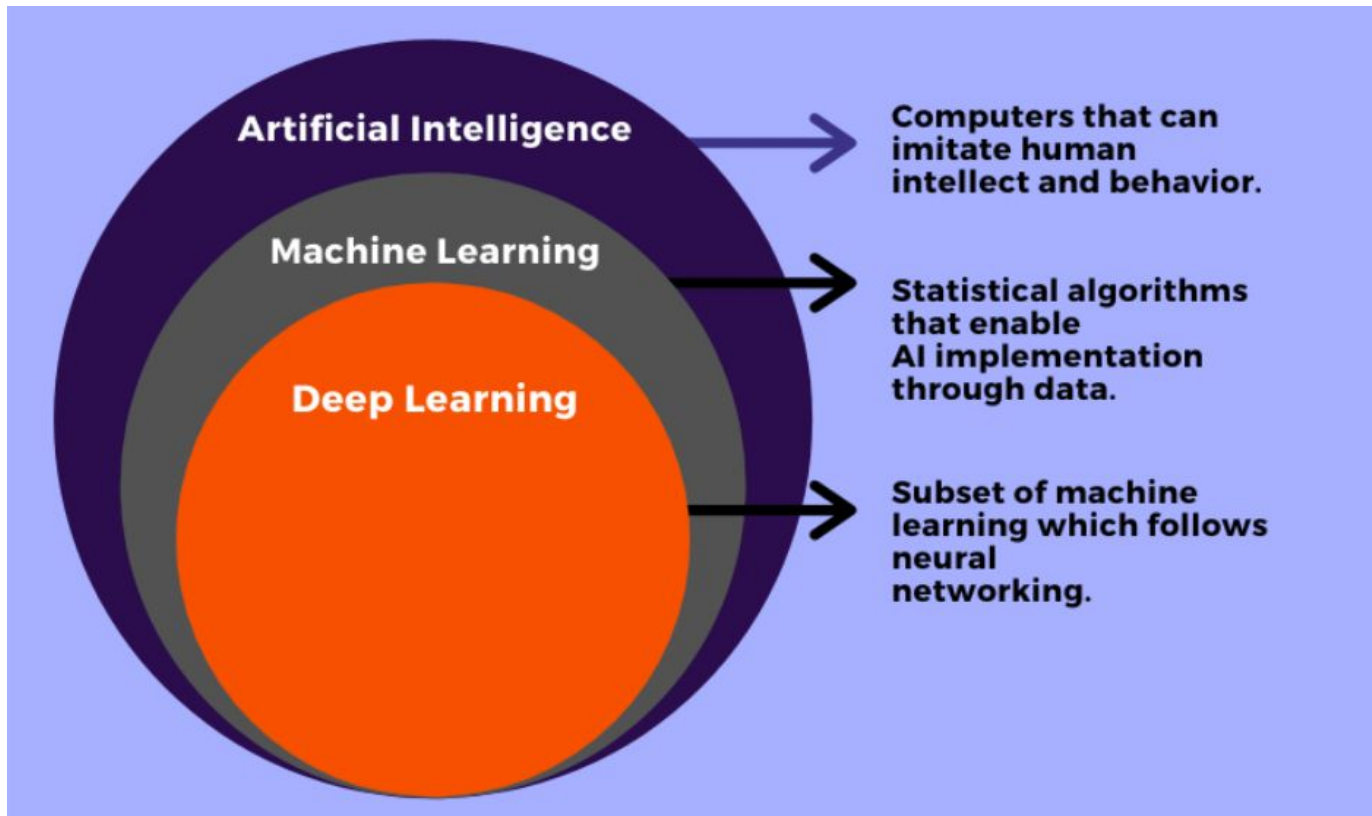
Supervised vs Unsupervised vs RL

Supervised Learning

- Supervised learning means we have **labeled** data
- **Classification**
  - Seeks to classify inputs
  - Is this an airplane?
  - Will the Yankees win or lose?
  - Will it rain tomorrow?
- **Regression**
  - Seeks to predict a continuous value
  - Close price of $AAPL tomorrow
 - Quarterly revenue
  - Age

Unsupervised Learning

- Unsupervised learning means the dataset is not labeled
- **Clustering**
  - Discover hidden patterns and data-groupings without the need for human intervention
  - Clustering is a data mining technique which groups unlabeled data based on their similarities or differences.
  - Its ability to discover similarities and differences in information make it the ideal solution for exploratory data analysis, cross-selling strategies, customer segmentation, and image recognition.
- We are essentially creating "classes", using the raw data.

Reinforcement Learning

- AREA
  - Agent, Reward, Environment, Action
- We are training a model to respond to a changing environment.
- Taking suitable action to maximize reward in a particular situation.
- Differs from supervised learning because we do not have an answer key (labels)
- The agent learns from its experience in the environment
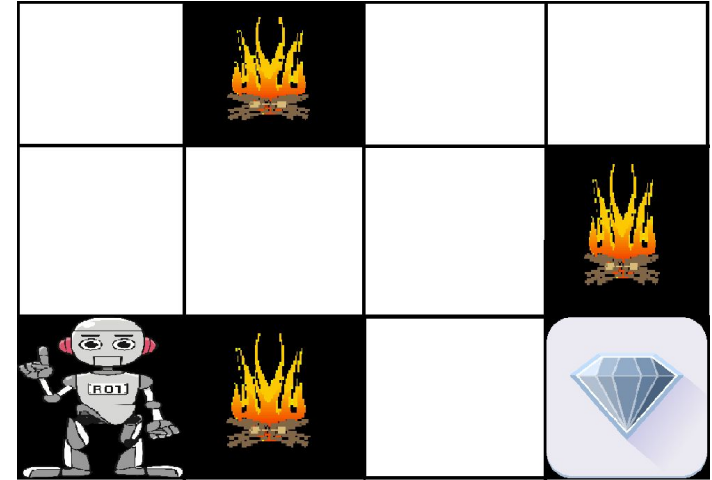
The above image shows the robot, diamond, and fire. The goal of the robot is to get the reward that is the diamond and avoid the hurdles that have fire. The robot learns by trying all possible paths and then choosing the path which gives the reward with the least hurdles. Each right step will give the robot a reward and each wrong step will subtract the reward of the robot. The total reward will be calculated when it reaches the final reward that is the diamond.
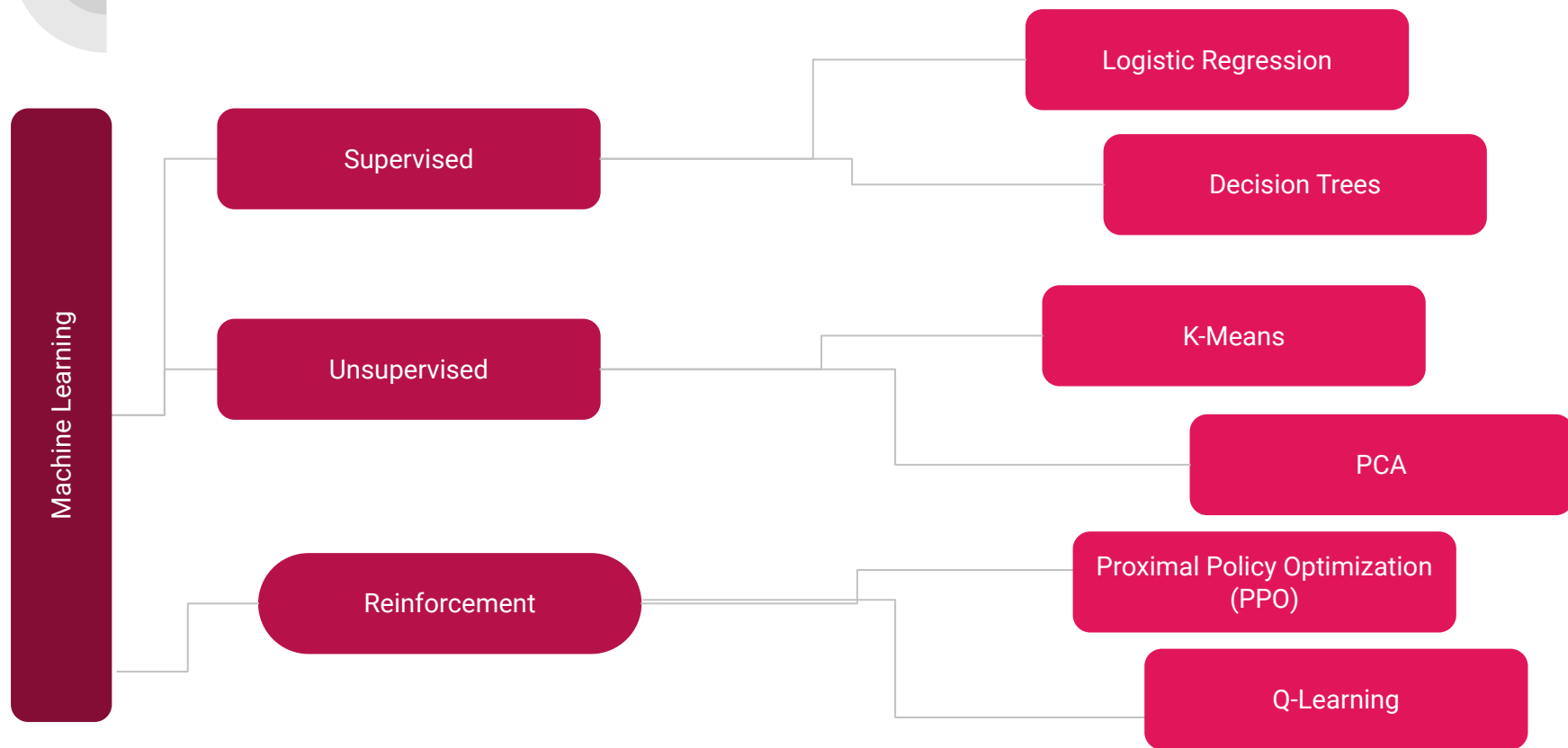
Introduction – ML Algorithms

Overview

- Now that we have defined the subsets of ML, we can identify some prevalent algorithms used in each subset.
- **Different tasks require different algorithms**
- We cannot use the same algorithm to classify breeds of dogs and also predict next months revenue.
- Generally, we use the following algorithms
  - Decision Trees/Logistic Regression for tabular data
  - Convolutional-Neural-Net (CNN) for vision tasks
  - Recurrent-Neural-Net (RNN) for language tasks
  - Deep-Q Learning for RL tasks
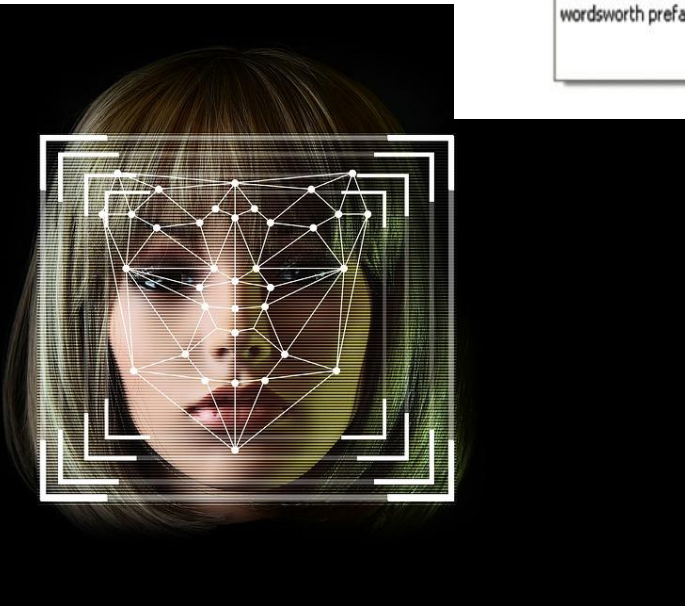
Some Flavors of ML Algorithms

Machine Learning

- Supervised
  - Logistic Regression
  - Decision Trees
- Unsupervised
  - K-Means
  - PCA
- Reinforcement
  - Proximal Policy Optimization (PPO)
  - Q-Learning

Examples

RNN

CNN

RL

Identifying ML Use-Cases

How To Identify ML Use-Cases

- A process where large, consistent volume of data is being ingested
- A tedious/time-consuming/repetitive process
- A decision making process
- Any process involving predictions & forecasting

Productizing ML Use-Cases
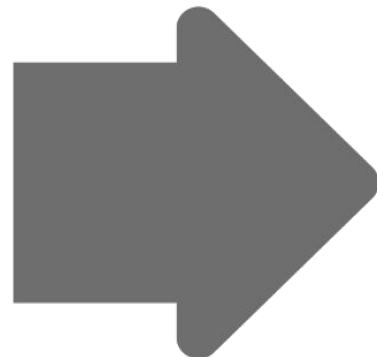
The Machine Learning Process

Step 1
Gathering data from various sources

Step 2
Cleaning data to have homogeneity

Step 3
Model Building- Selecting the right ML algorithm

Step 4
Gaining insights from the model's results

Step 5
Data Visualization- Transforming results into visuals graphs

Types of ML Deployments

- Interactive vs Non-Interactive
- Single Record vs Batch
- Interactive
  - Client facing
  - Client sends data and gets results
- Non-Interactive
  - Not client facing
  - Client only receives results
- Single Record
  - Model is receiving one input at a time and returning one prediction at a time
- Batch
  - Model is receiving large batches of input data and returning batches of predictions

Productizing ML Use-Cases Cont.

- This will be discussed in depth later on in the course.
- The ML process starts with data engineering, and ends with deploying an API.
- **Birds-eye view:**
  - Collect data from data source(s)
  - Develop ML model
  - Build query pipelines for the data
  - Integrate pipeline + model into API
  - Deploy the API

Conclusion

Questions

- I want to forecast my branch's next quarter revenue. What subset of ML is this?
    a. Unsupervised Learning
    b. Supervised Learning
    c. Reinforcement Learning
- I have identified the subset of ML for this project, what type of task is this?
    a. Regression
    b. Clustering
    c. Classification
- I have a dataset on my customer behaviors, what task is best suited to group the customers together and profile them based on behavior?
    a. Classification
    b. Reinforcement Learning
    c. Clustering
- **BONUS:** What algorithm should I use to generate text? Think auto-complete on emails/texts.
    a. CNN (Convolutional)
    b. Decision Tree
    c. RNN (Recurrent)