# Introduction to Artificial Intelligence and Machine Learning

What is AI?

* Making computers think like humans = Cognitive Science
* Making computers act like humans = Turing test, NLP
* Making computers think rationally = Logic
* Making computers act rationally = Autonomous

Agent = Acts in environment

* Acts intelligently if:
  + Acts appropriately in its situation
  + Is flexible to changes
  + Learns
  + Makes the right choices depending on its perspective
* Computational Agent: Agent that’s decisions can be explained in terms of computation
* Rational Agent: Achieve best outcome (Goal of AI)

AI { ML (Logical mapping from data to output) { DL (Deeper representations)

Machine Learning = Learns from experience

* Give data and output and Machine Learning makes it understand how it got there
  + 22 = 4 instead given 2 and 4 and returns 2

## Supervised Learning (Input and Output)

**Maps an Input to a Particular Output**

**Presented with Data and their Correctly labelled output**

**Learns from experience to predict new unseen data**

Categories:

* Classification = Predicts a category learned from labelled data
* Regression = Predicts a continuous valued output learned from labelled data

Process:

* Input Labelled Data
* Train Model using algorithms
* Give Model Unseen Data
* Output

Algorithms: Training by making and correcting predictions until a certain accuracy is achieved

* Logistic Regression
* Back Propagation Neural Network

## Unsupervised Learning (Input Only)

**Gives only input data and no output data**

**Models the underlying structure of the data and to find patterns**

Categories:

* Clustering = Finding inherent groups in data
* Association = Finding rules linking data

Process:

* Finds structures in input data
* Extracts rules, reduce redundancy or organise data by similarity

Algorithms:

* K-Means
* Apriori

## Semi Supervised

**Some but not all data is labelled**

**Combines supervised and unsupervised**

# Features in Machine Learning and Feature Extraction

## Features and Labels

**Feature = Individual Characteristic that is Measurable of something observable**

**Label = Tag attached to feature**

## Feature Extraction

**Maps input features to new output features**

**Transforming raw data into numerical features that can be processed whilst preserving information in original dataset**

**Finds interesting parts of the image as a compact feature vector**

**Transforms picture data to numeric data**

Manual or Automated

**Feature Vector = Representing Features in a Mathematically Analysable Way**

Detecting interesting parts of an image as a compact feature vector

Good Features are:

* Repeatable
* Descriptive
* Compact

Process:

* Feature Detection = Used to find features
  + Methods = Edge Detection
* Feature extraction = Gets the features

Text:

* Foundational Assumption = Knowing a word by understanding its neighbours
* Vector Representation = Find similarity between words

Methods:

* + 1 Hot Vector
  + Bag of Words
    - Document = Unit of text to be analysed
    - **Set of words in document and their frequency**
  + Term Frequency - Inverse Document Frequency (TF IDF)
    - TFt,d = Frequency Count of Each Word *t* in Document *d* / Number of Words in Document *d*
    - IDFt  = log2()
    - TF-IDF = TF x IDF
  + PPMI = Positive Pointwise Mutual Information
    - **Shows how dependent words are on each other**
    - PMI = Log2
    - PPMI = Positive PPMI = max(Log2, 0)
  + Word Embedding (Word2Vec, Glove)
  + Distance Measure (Cosine, Euclidean)

# Computer Vision and Natural Language Processing

Convolution = Taking two Functions in and producing a third function

A picture containing text, font, white, handwriting

Description automatically generated

(Output Image)(x, y) = Σ (Input Image)(x - i, y - j) \* (Kernel)(i, j)

where the summation (Σ) is performed over i and j indices for all elements in the kernel matrix.

* (Output Image)(x, y): Pixel value at coordinates (x, y) in the output image, which is the result of the convolution operation.
* (Input Image)(x - i, y - j): Pixel value of the input image at the shifted coordinates (x - i, y - j). The shift is determined by the current i and j values during the summation.
* (Kernel)(i, j): Weight value at the i-th row and j-th column in the kernel matrix.
* Σ: Sum of the element-wise product of the input image and kernel values.

Properties of Convolution:

* Commutativity: k \* u = u \* k
* Associativity: 𝑘 ∗ ℎ ∗ 𝑢 = 𝑘 ∗ ℎ ∗ 𝑢
* Distributivity: 𝑢 ∗ 𝑘 + ℎ = 𝑢 ∗ 𝑘 + 𝑢 ∗ ℎ
* Associativity with scalar multiplication: 𝑎 (𝑘 ∗ 𝑢) = (𝑎𝑘) ∗ 𝑢 where 𝑎 ∈ ℂ (complex numbers)
* Integration:

**Semantic Analysis**

* Grammatical POS Tagging: Tagging each word with its part-of-speech
* Rule Based POS Tagging: Apply most likely tag for each word then retagging

# Clustering and Classification

Clustering = Finding Patterns and Groups amongst the data

* Hierarchical = Build a hierarchy of clusters
  + Agglomerative (Bottom up) = Each observation starts in own cluster and pairs of clusters are merged as you move up
  + Divisive (top down) = All observations start in one cluster and then split recursively as they move down the hierarchy

Process:

1. Initialise each item
2. Merge closest pairs

* + Dendrogram used to split clusters into different levels

Features

* + Deterministic
  + Inefficient
  + Multilevel Representation
  + Flexible
* K Means Clustering = Optimisation Problem – Minimizing variance

Process:

1. Choose K Amount of Clusters and assign to random points to form the initial centroids
2. Create K Clusters by assigning each point to the closest centroid
3. Compute new centroids by averaging points in each cluster
4. Repeat steps 2 and 3 until centroids don’t change

Features:

* + Efficient
  + Non-deterministic
  + Results dependent on how many clusters

**If we have chosen K = 2 in a dataset with dog, cat and horse then we will have too few clusters and get strange results**

Choosing a good K Value:

* + Prior knowledge of data space
  + Use elbow method = Try different values and look for change
  + Run hierarchical clustering on data

Choosing good clusters:

* + Use a random number seed and then define a minimum distance between clusters

Classification = Categorising the data into groups or classes

* Aim = find *f* such that f(x) = y ∈ {1, … , k}
  + x = feature vector (List of features)
  + y = class label

* Algorithms :
  + K-Nearest Neighbour = Group to K nearest neighbours
  + SVM Binary
  + SVM Multiclass
  + Decision Tree
  + Naïve Bayes
  + Discriminant Analysis
  + Ensembles

Accuracy = Correct Labels / All Elements

* Non-categorical Accuracy = || L1 – L2 ||2
* Confusion Matrices = TP, TN, FP, FN Accuracy = (TP + TN) / (TP + TN + FP + FN)

Other measures:

* + Sensitivity
  + Specificity
  + Precision
  + Negative Predictive Value
  + F1 Score

# 5. Artificial Neural Networks

# 6. Genetic Algorithms

# 7. Naïve Bayesian Classifier

# 8. Decision Tree Classifier

# 9. Introduction to Deep Neural Networks

# 10. Introduction to Convolutional Neural Networks