

Concepts of Machine Learning

Day 2
GNITS AI Workshop

WHAT DOES IT MEAN FOR SOMETHING TO BE ARTIFICIALLY INTELLIGENT MACHINE?

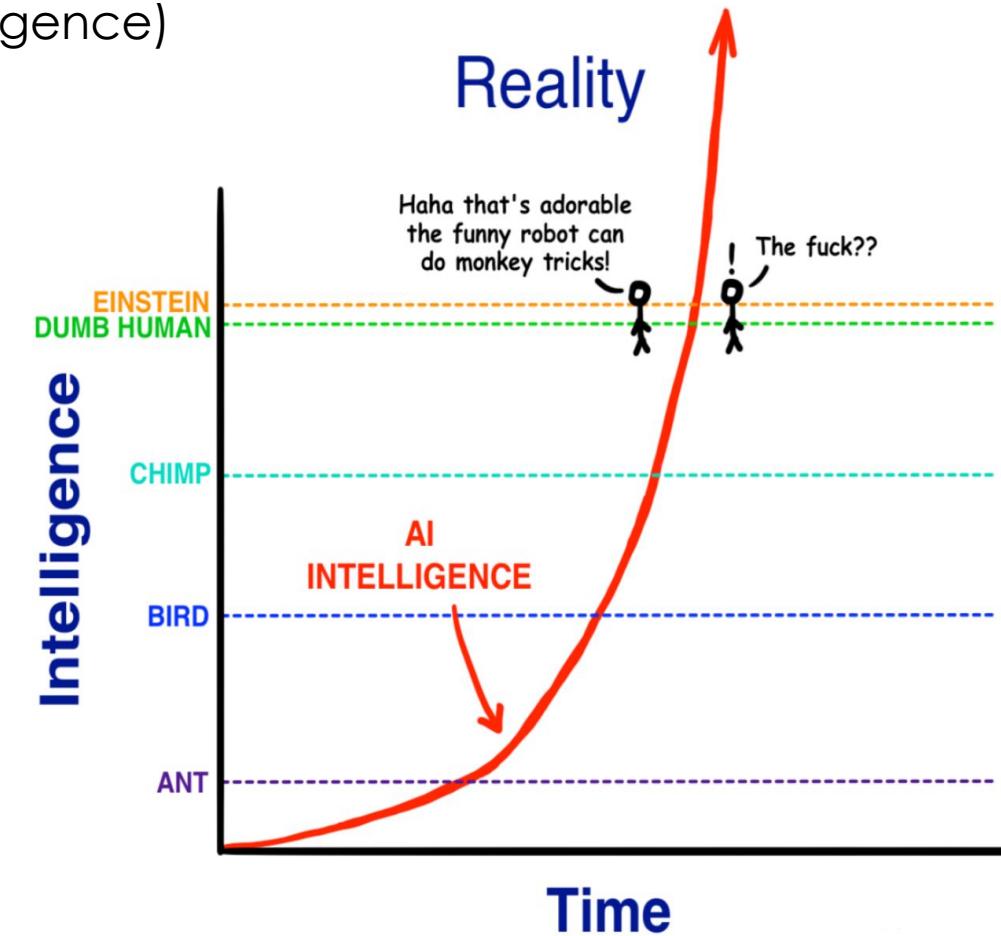
- **Exhibit Intelligence**
- **Perceive their environments**
- **Take actions/ make decision to maximize chance of success at a goal**



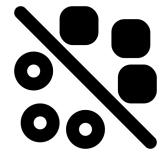
PHASES OF AI



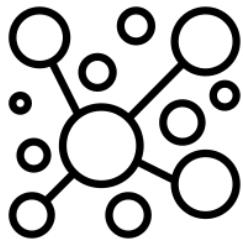
- **Artificial General Intelligence(AGI):** A machine with the ability to apply intelligence to any problem, rather than just a specific problem (human- level intelligence)
- **Artificial Narrow Intelligence(ANI):** Machine Intelligence that equals or exceeds human intelligence or efficiency at a specific task.
- **Artificial Super Intelligence(ASI):** An intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills.



LEARNING APPROACHES



- **Supervised learning:** Learning with a labelled data
Example: email spam detector (classification), Weather Reports (Regression).

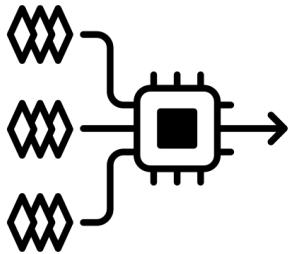


- **Unsupervised learning:** Discovering patterns in unlabeled data
Example: Cluster similar documents based on the content type.



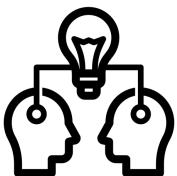
- **Reinforcement Learning:** learning based on feedback or reward
Example: learn to play chess by winning or loosing.

LEARNING APPROACHES



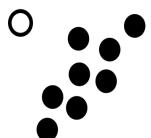
- **Semi - Supervised learning:** involves usage of both labelled and unlabeled data to train a model. This is useful when labeled data is expensive or difficult to obtain.

Example: Fraud Detection, Sentiment Analysis, Image and Video Classification



- **Transfer Learning:** Method where knowledge learnt from one task is applied to another related task.

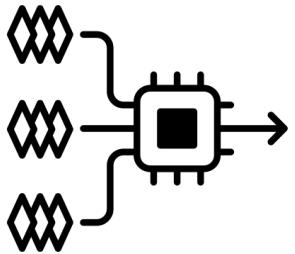
Example: Speech Recognition using BERT, GPT.



- **Anomaly Detection:** involves in identifying data points that are significantly different from the rest of the data.

Example: Network Intrusion Detection, Medical Diagnosis, Fraud detection (credit card).

LEARNING APPROACHES



- **Recommender Systems:** Based on the past behaviours predicting the likelihood that a user will be interested in.

Example: Fraud Detection, Sentiment Analysis, Image and Video Classification



- **Time series forecasting:** Predicting future values of a time dependent variable based on its past behaviour.

Example: Stock market prediction.

Era of Advancements ?

- 1950 – 1980: warm up

Features: Computers are not powerful, no efficient algorithms, no data.

Situations: Chinese people already had a good dream since the inception of computers called “Electronic Brains”.

- 1980 - 2000: Research Driven

Features: Computers are very powerful, efficient algorithms developed, no enough data in most cases

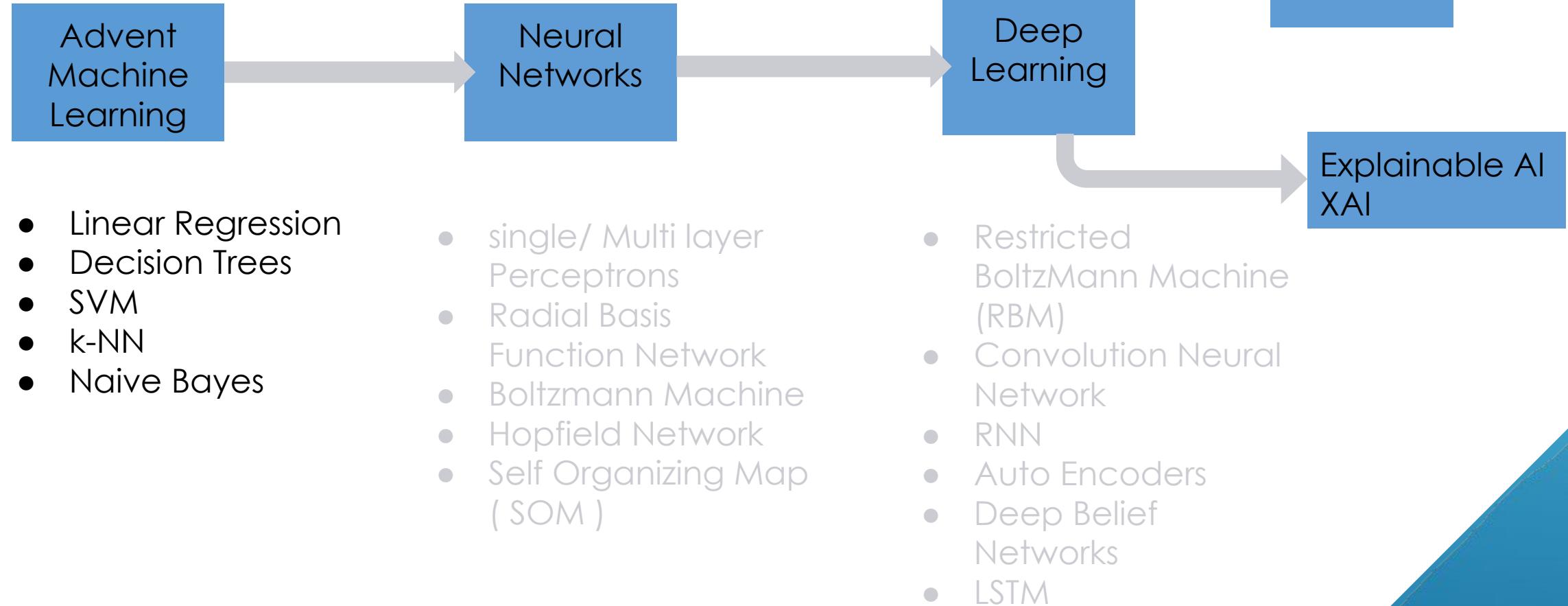
Situations: mostly driven by researchers instead of industries.

- 2000 - present: Data Driven

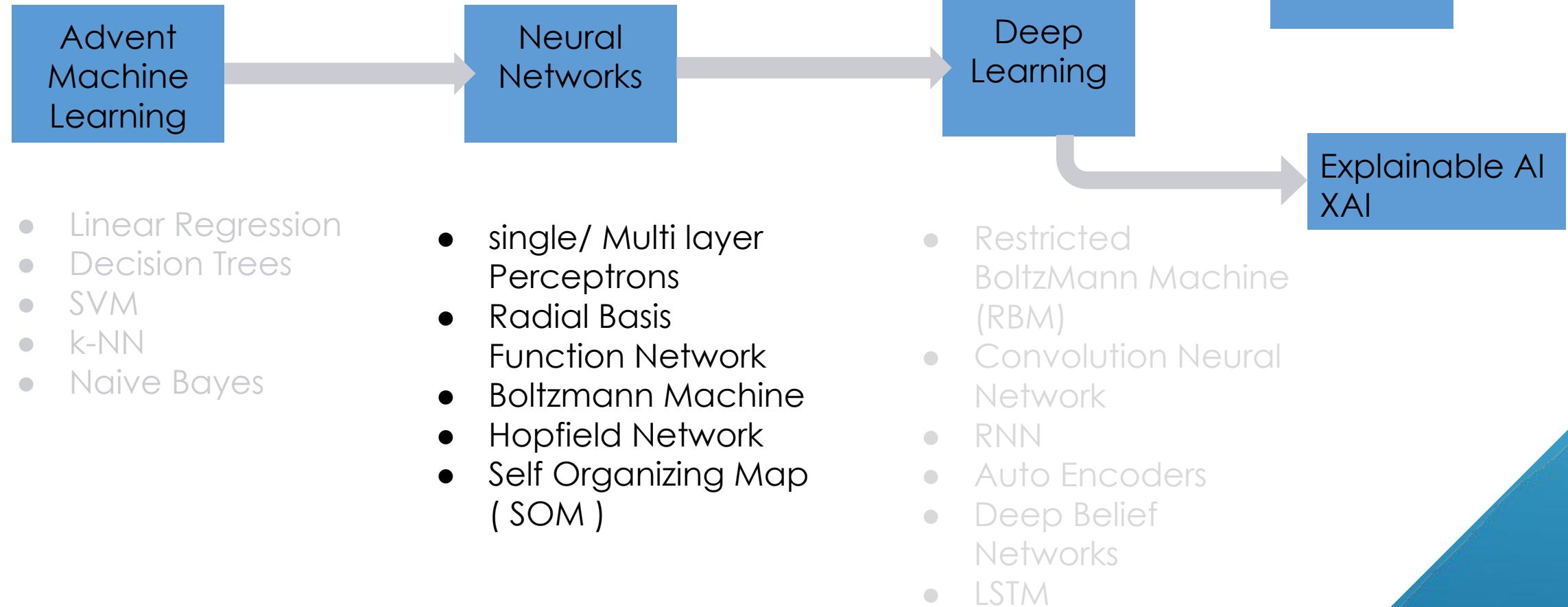
Features: Computers are powerful enough, smart computing sensors/devices everywhere, huge data coming in. Efficient algorithms on the way.

Situations: No matter we like or not, we have to socialize with machines.

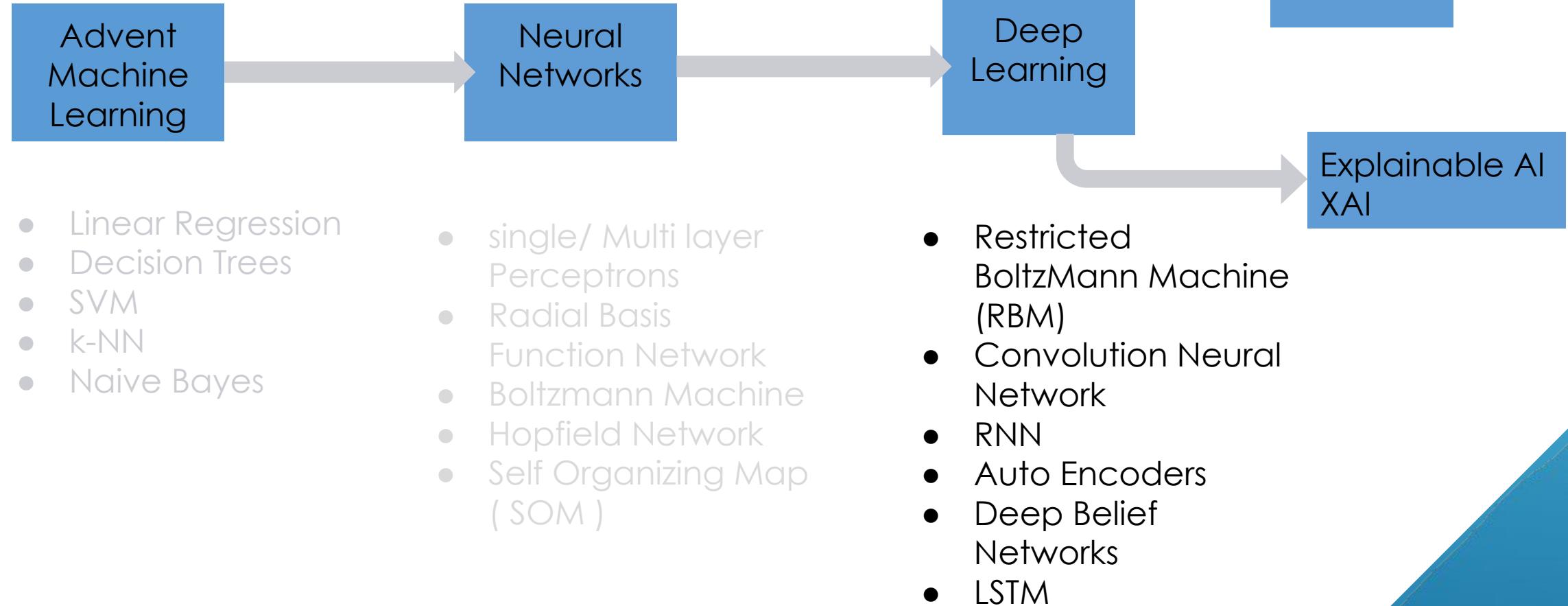
Era of Advancements ?



Era of Advancements ?



Era of Advancements ?



Era of Advancements ?

Advent
Machine
Learning

- Linear Regression
- Decision Trees
- SVM
- k-NN
- Naive Bayes

Neural
Networks

- single/ Multi layer Perceptrons
- Radial Basis Function Network
- Boltzmann Machine
- Hopfield Network
- Self Organizing Map (SOM)

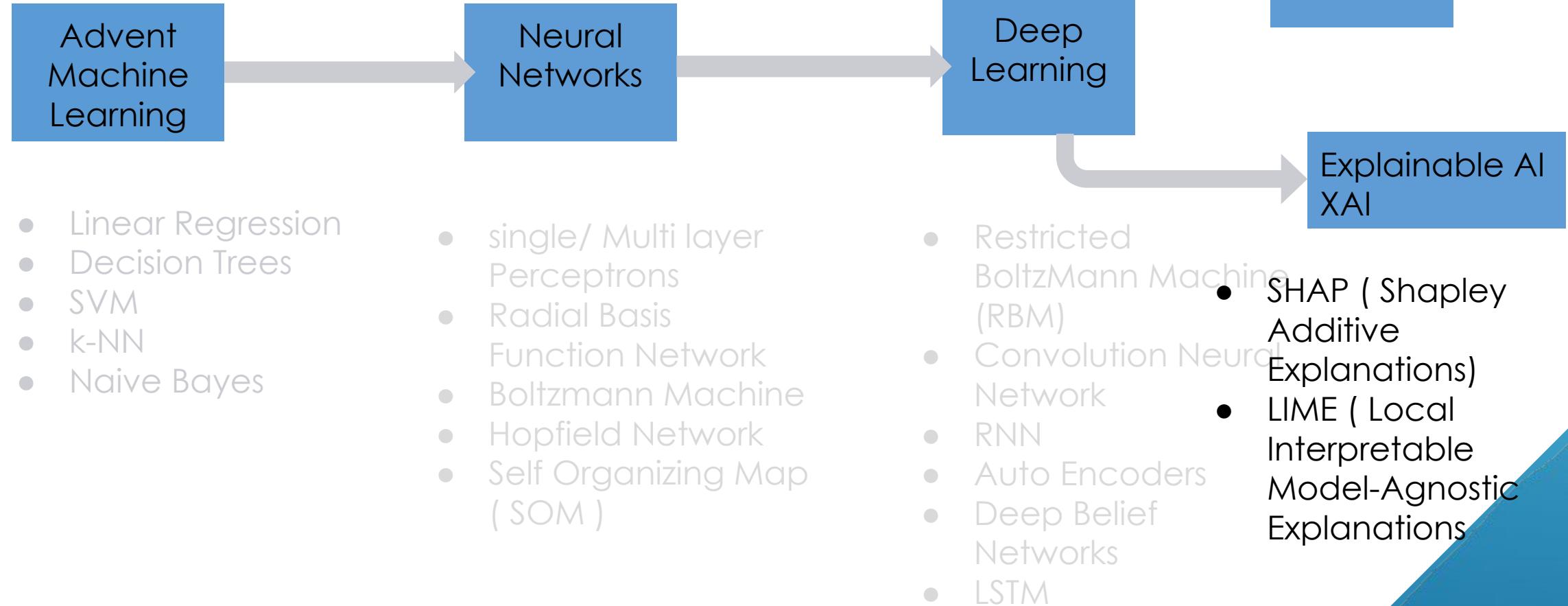
Deep
Learning

- Restricted Boltzmann Machine (RBM)
- Convolution Neural Network
- RNN
- Auto Encoders
- Deep Belief Networks
- LSTM

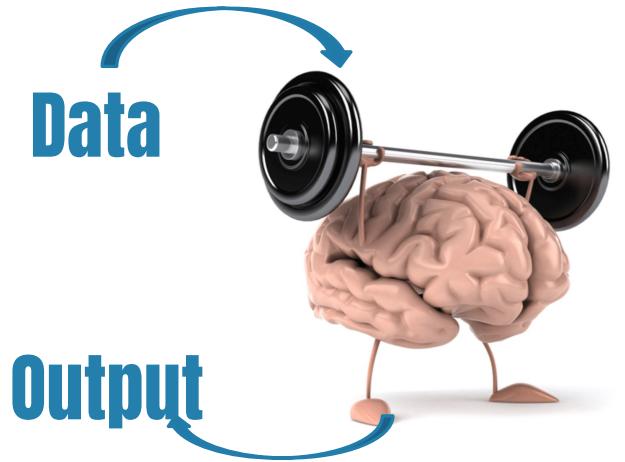
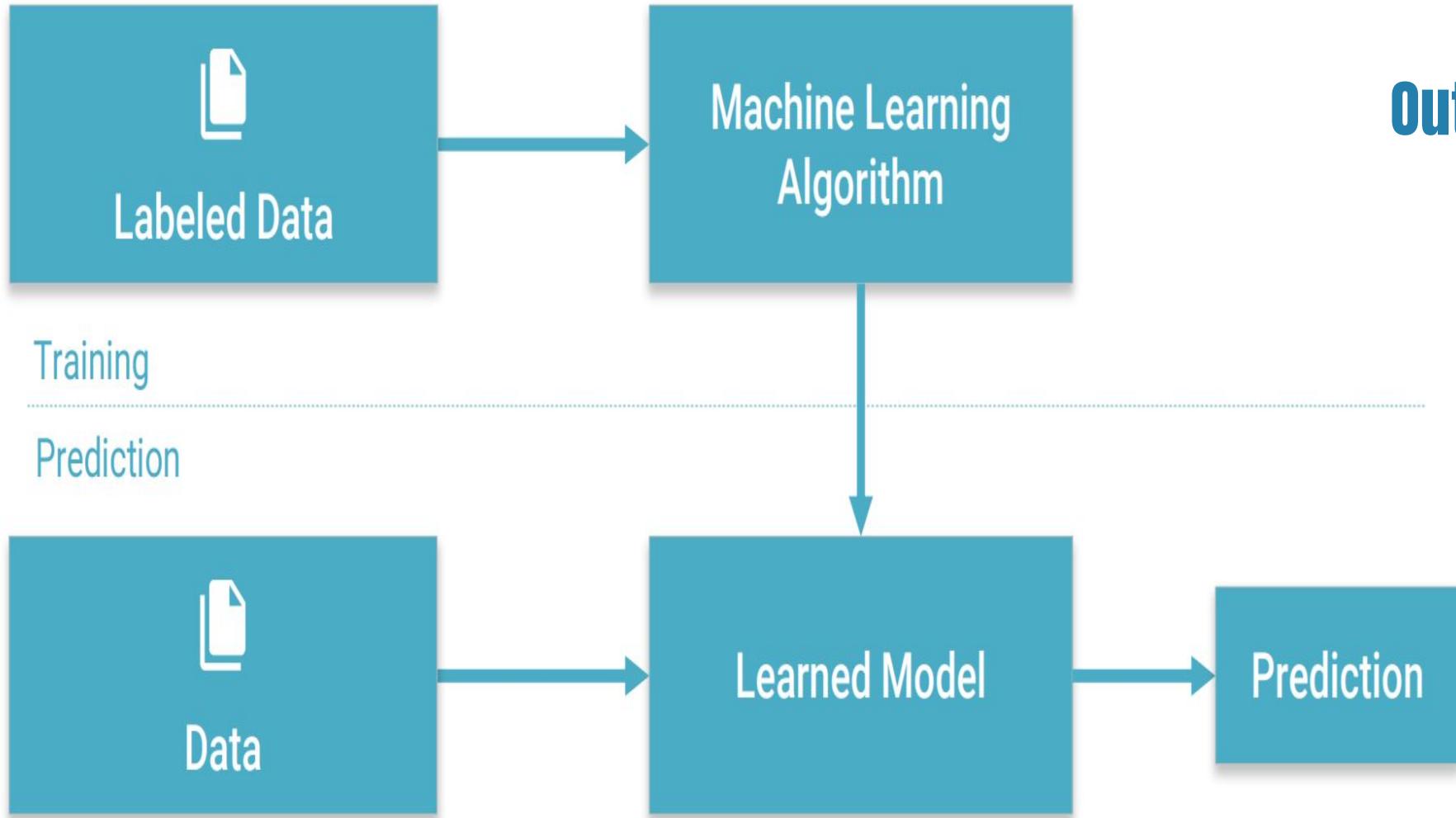
Transfer
Learning

- Language Models
- GPT
- Transformer Networks

Era of Advancements ?



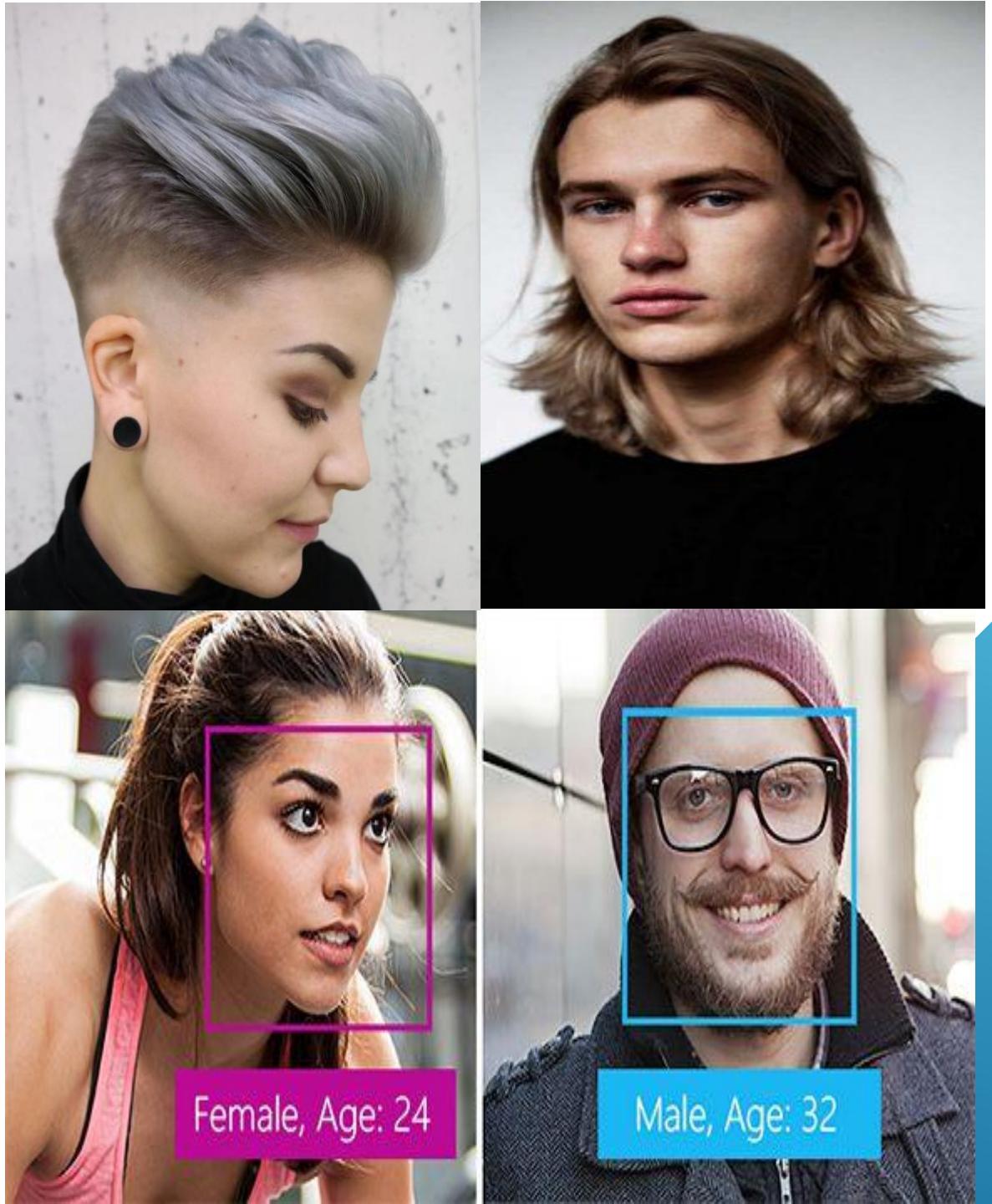
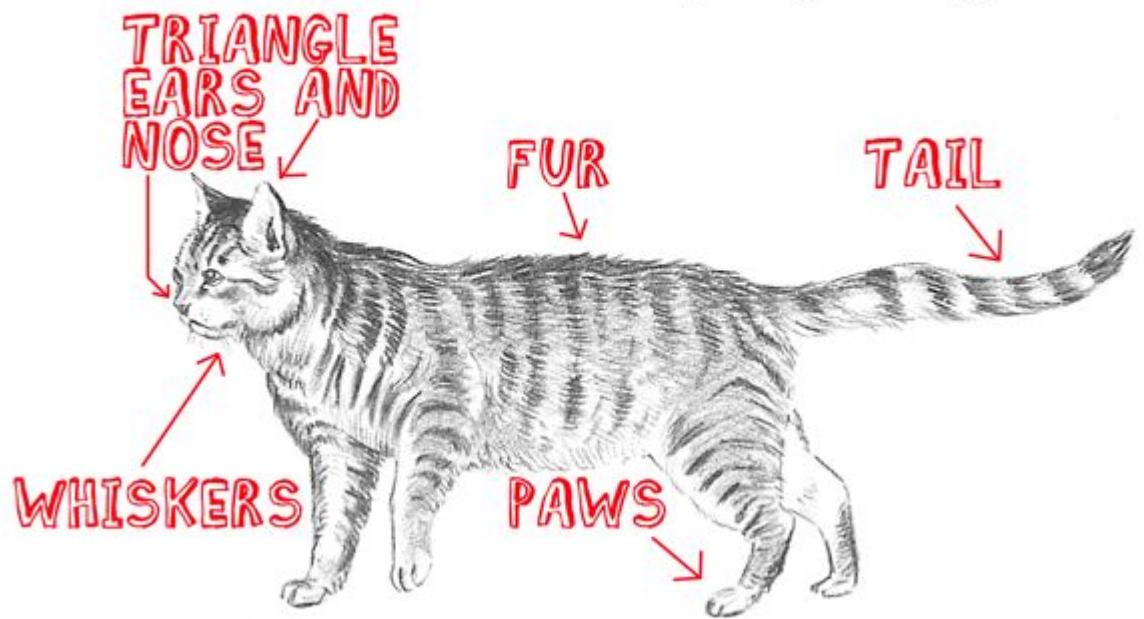
MACHINE LEARNING BASICS



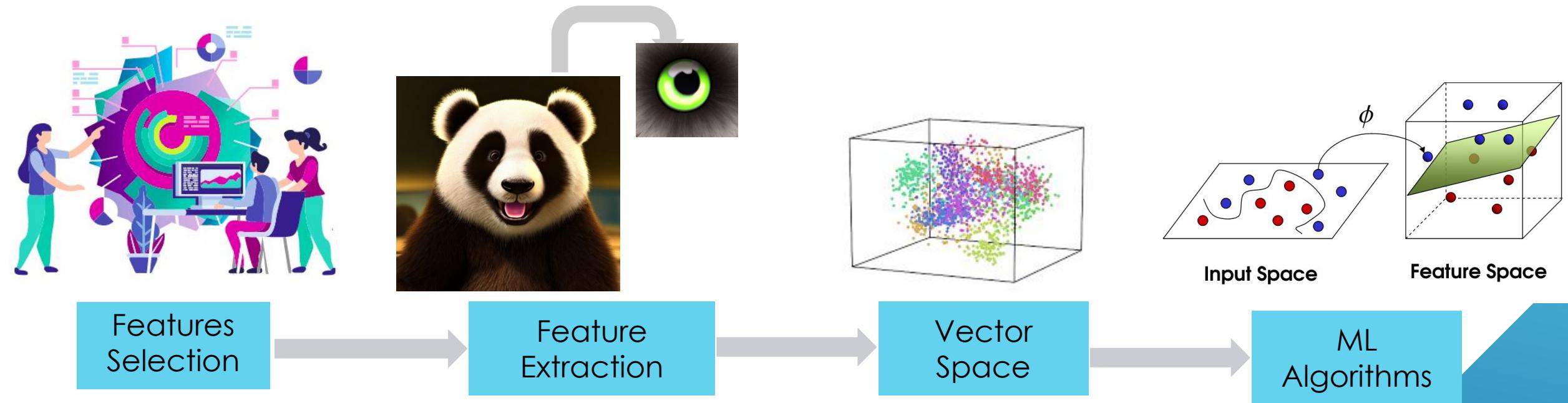
FEATURE SELECTION & FEATURE EXTRACTION

A simple way to think about supervised learning.

What Characteristics Do Cats Have



MACHINE LEARNING BASICS

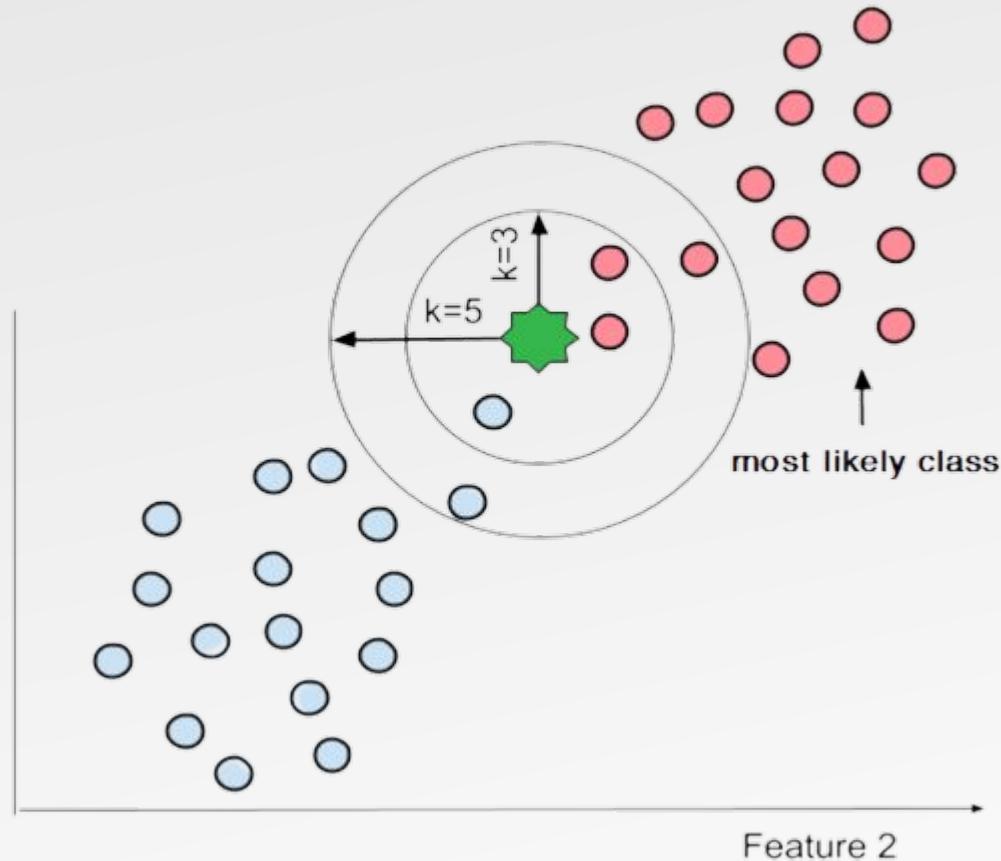


K NEAREST NEIGHBORS (k-NN)

The simplest algorithm

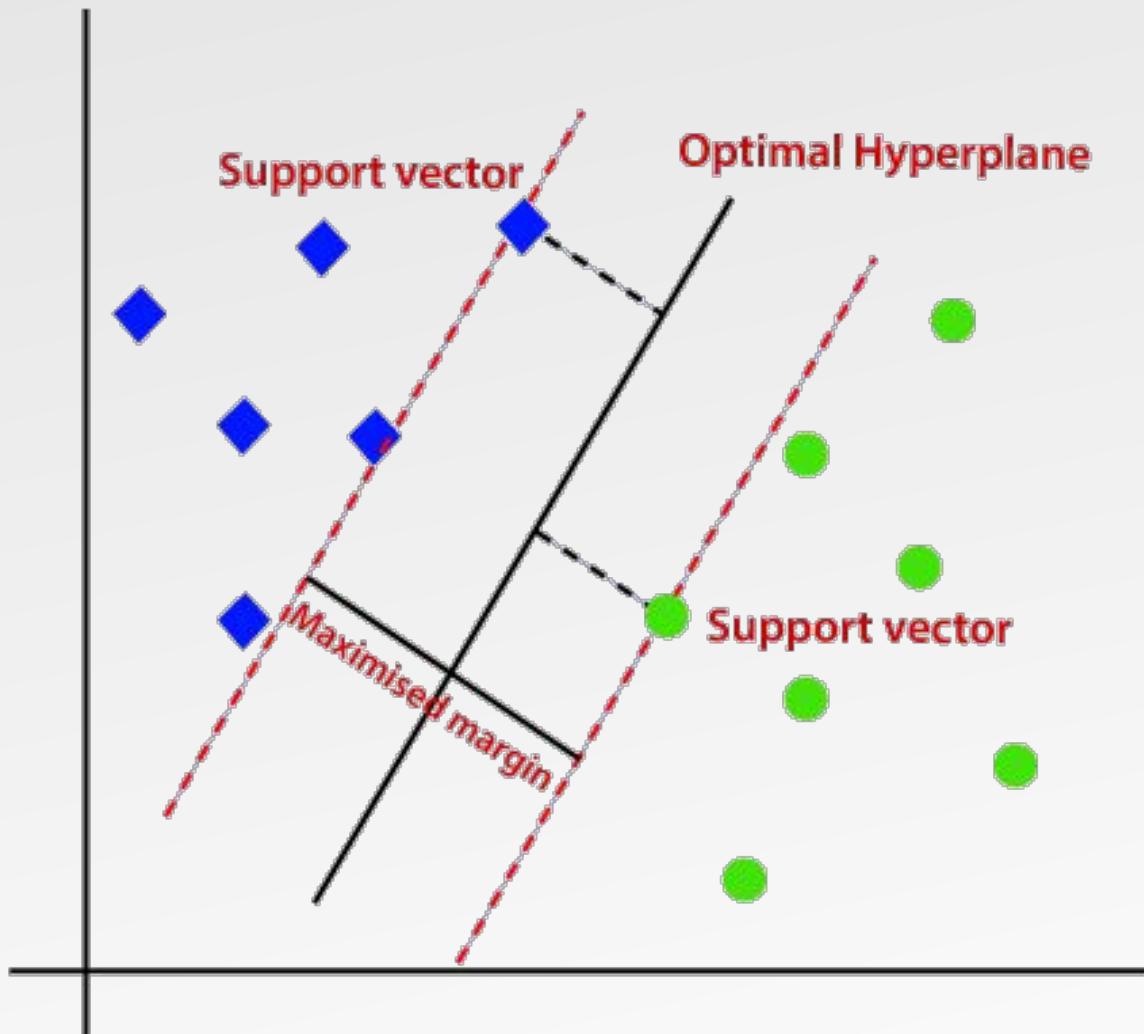
No training is required

Dependent on value of k



k value

Support Vector Machines (SVM)



Creates a decision boundary called Hyperplane.

Adversely affected by complexity of data.

C

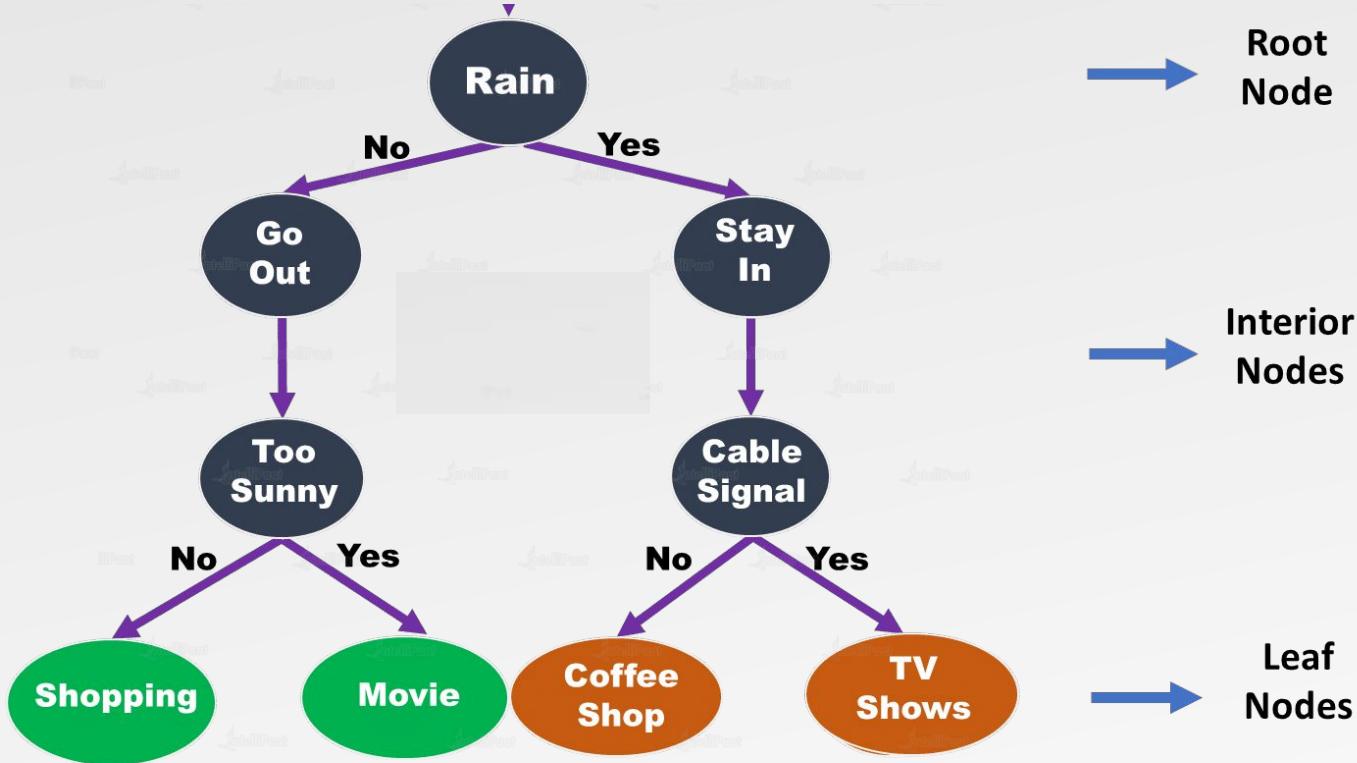
- regularization parameter.

Kernel -

kernel type
degree

gamma/coefficient

Decision Trees



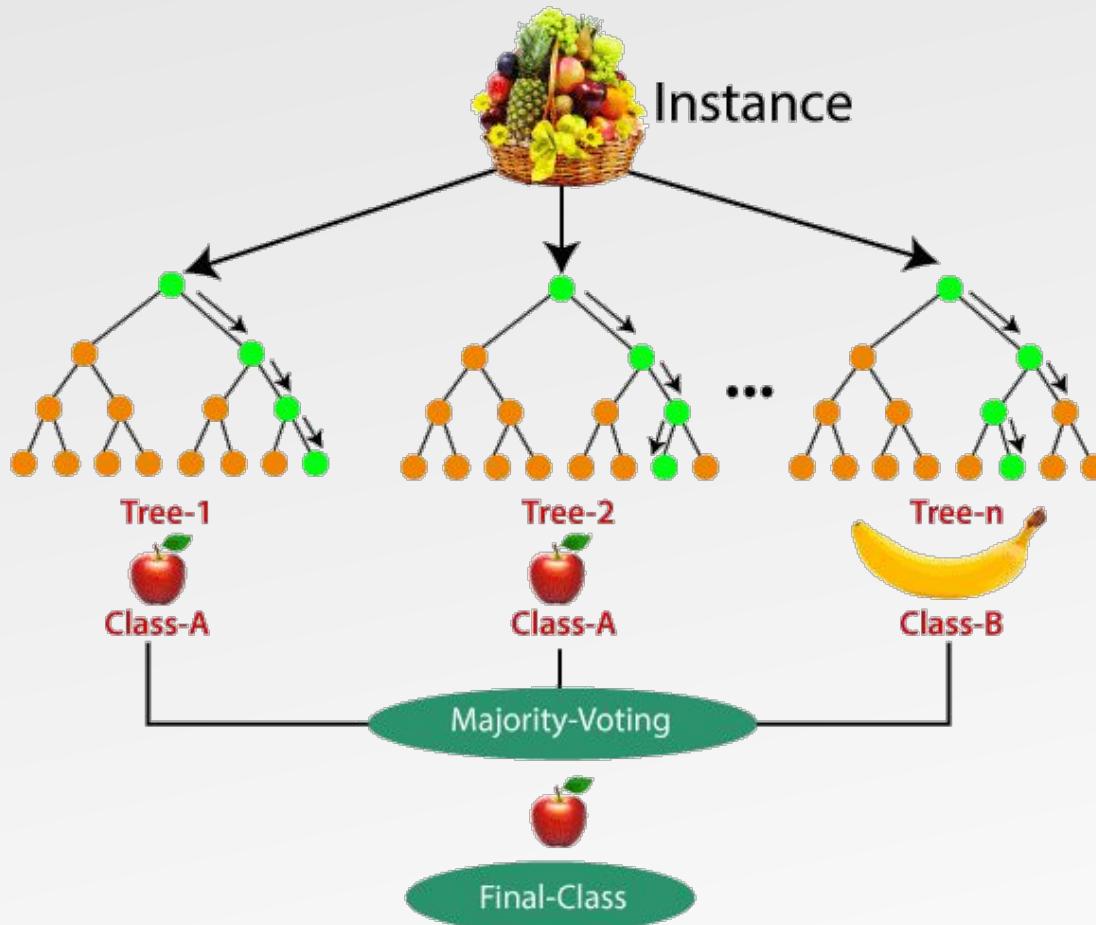
Simple for data exploration

Small Change in data adversely affect the output.

Lot of ambiguity.

Hypothesis Space

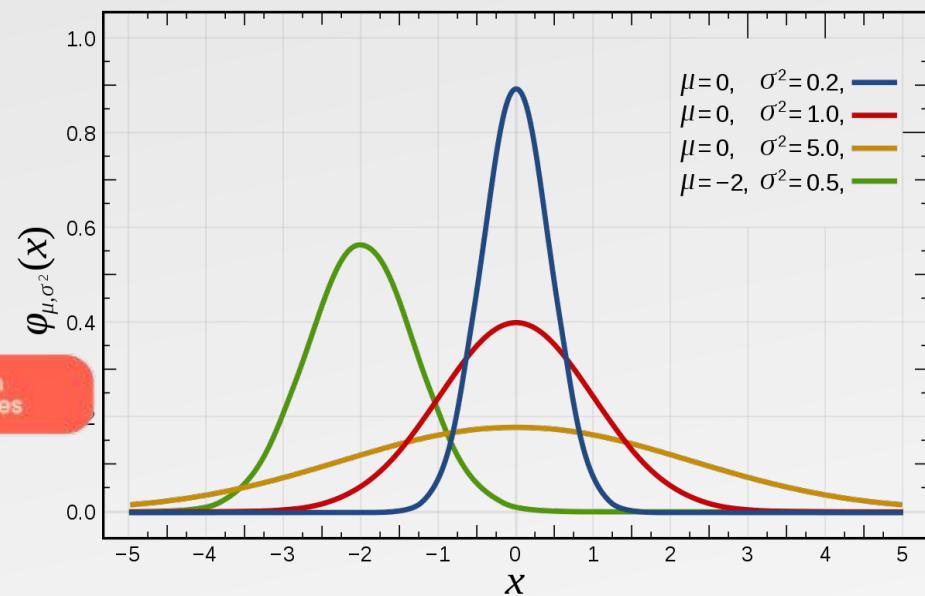
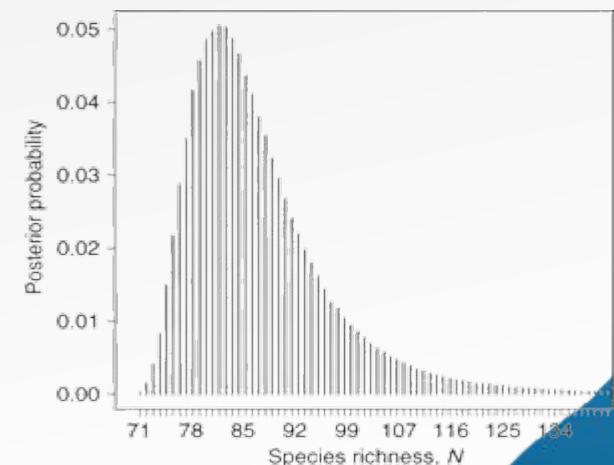
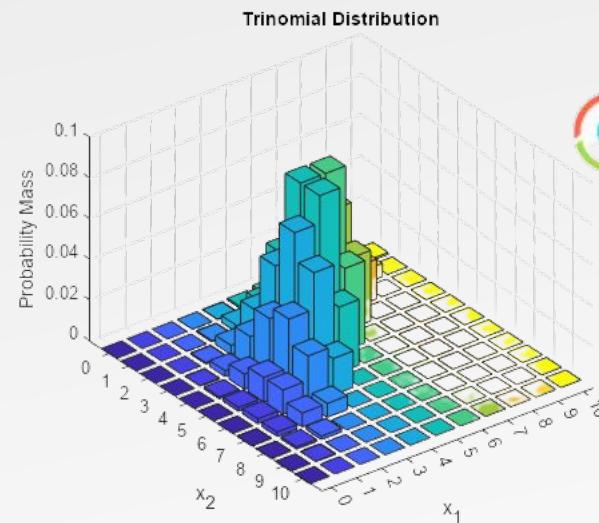
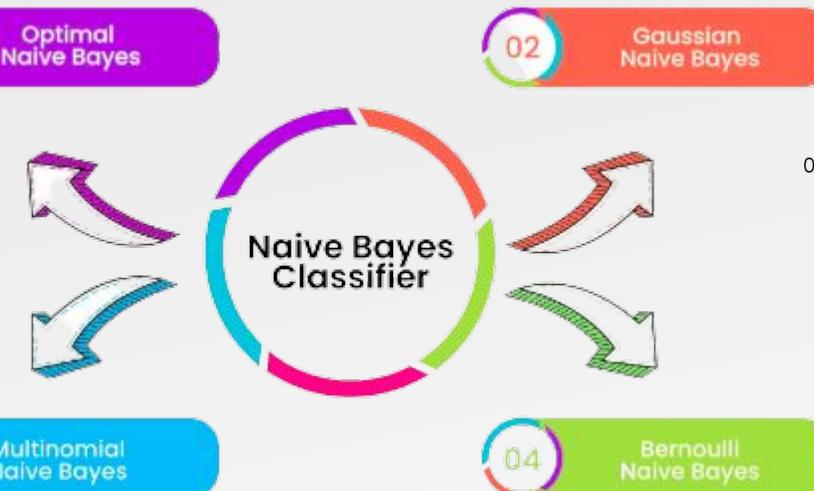
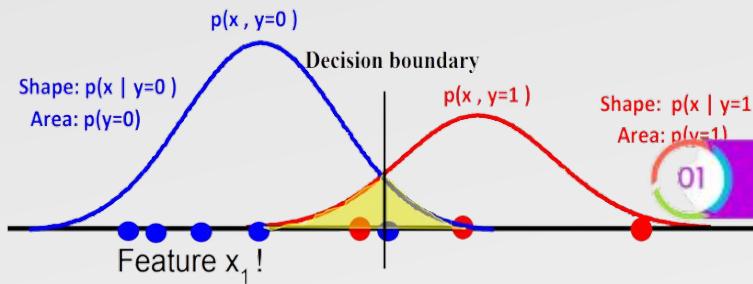
Random Forest



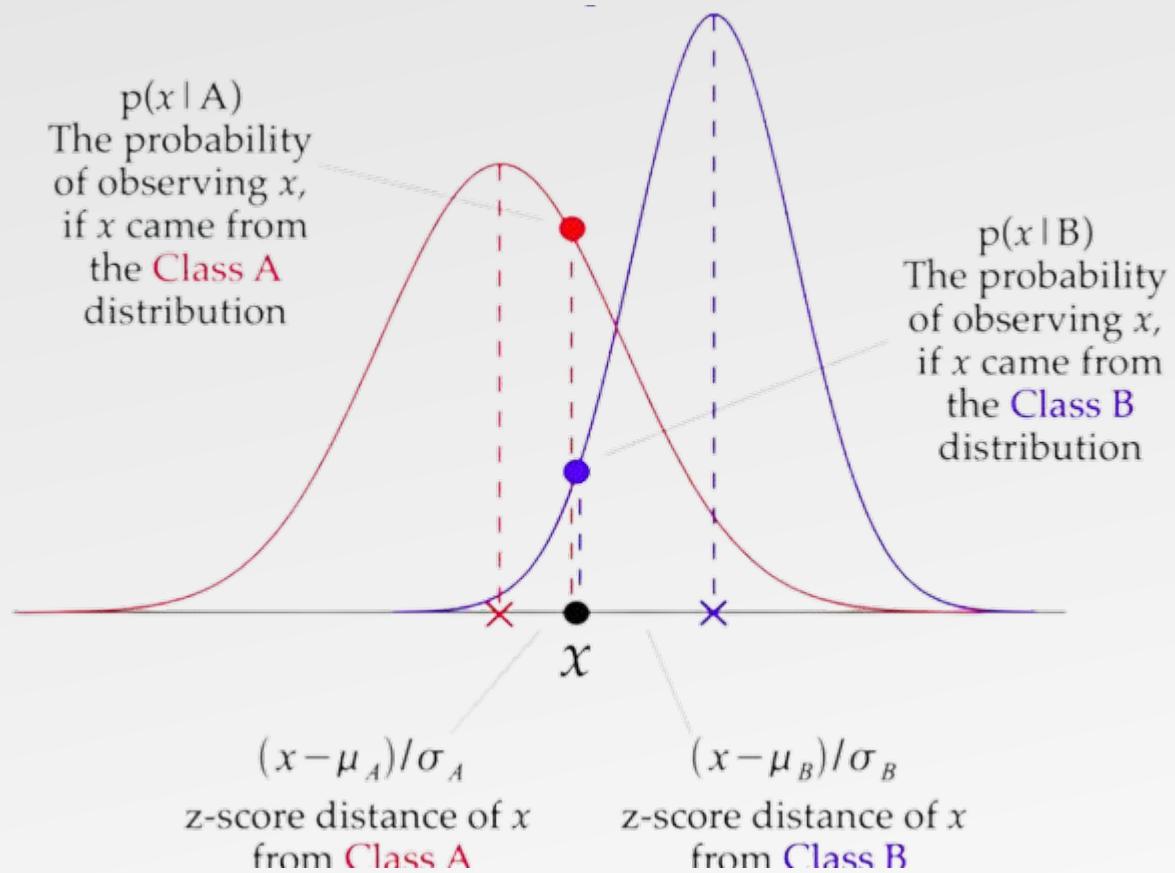
**Multi Level Decision Trees
Robust to Outliers
can handle missing values**

**Training Period.
Complexity.**

Naive Bayes

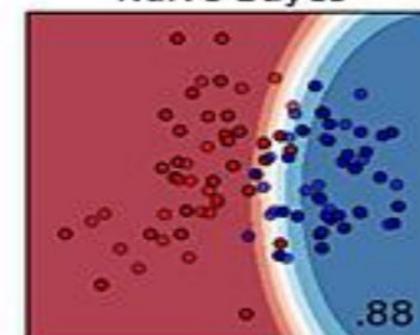
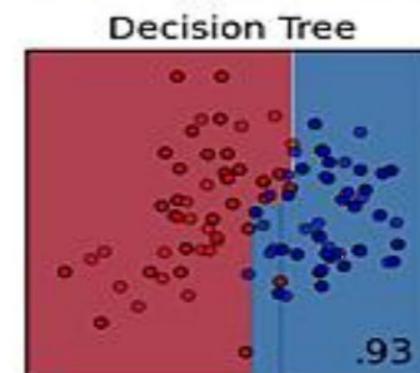
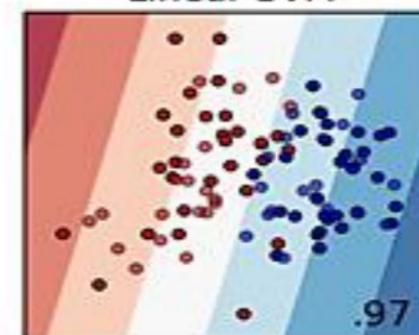
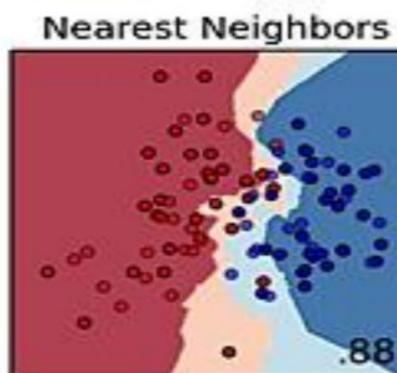
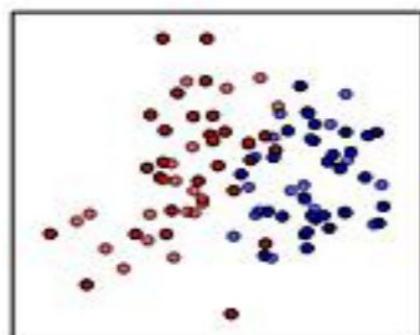
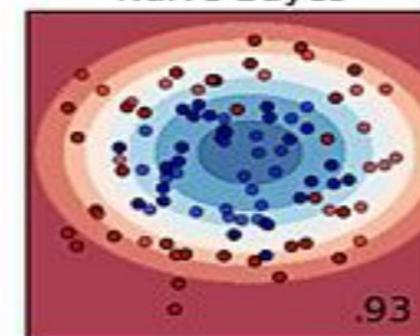
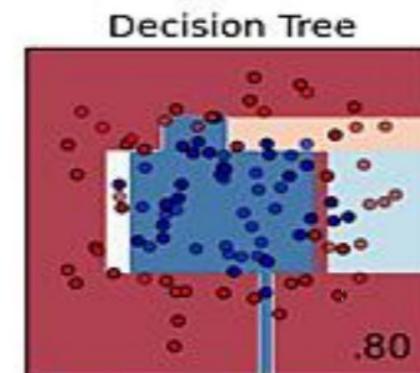
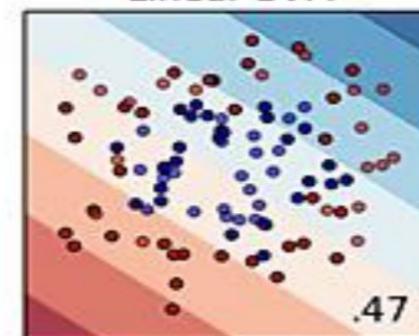
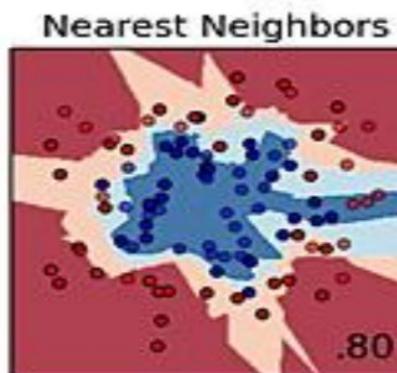
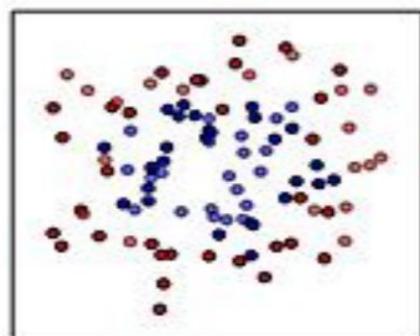
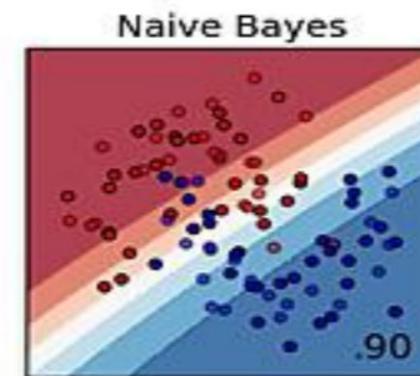
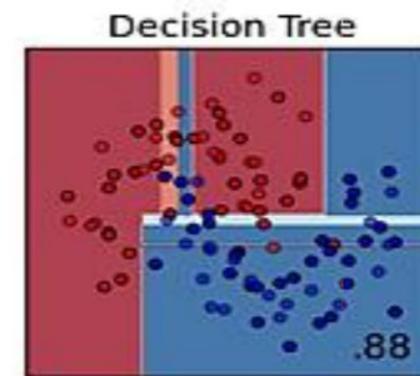
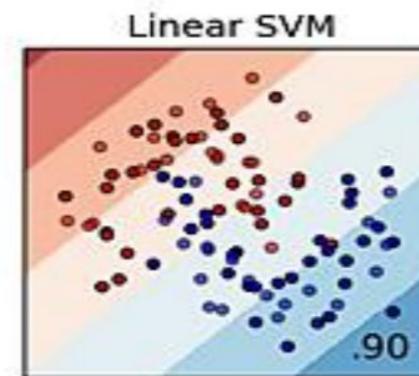
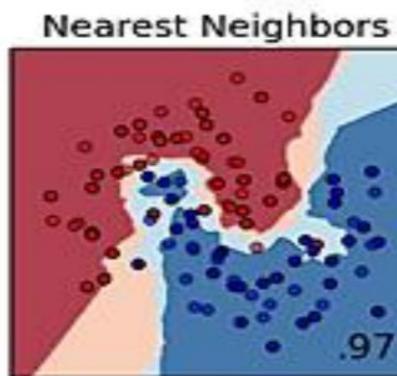
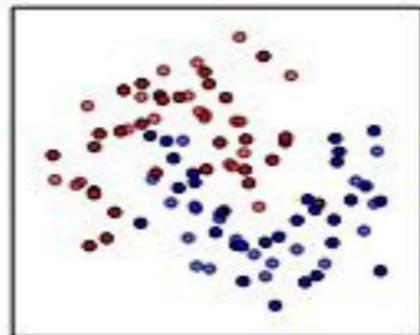


Naive Bayes



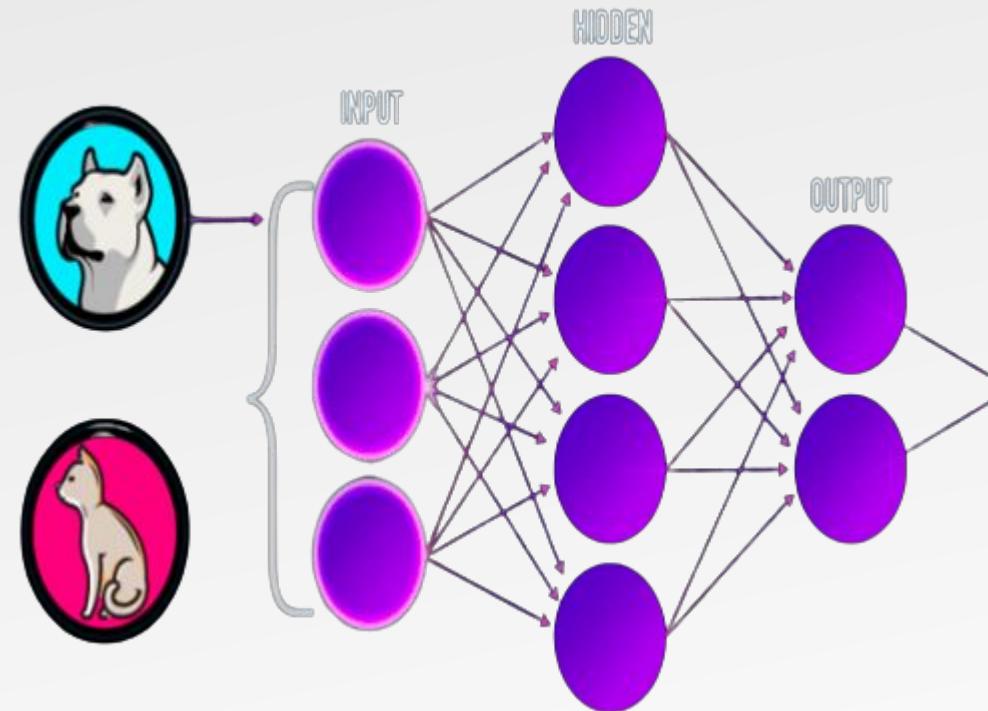
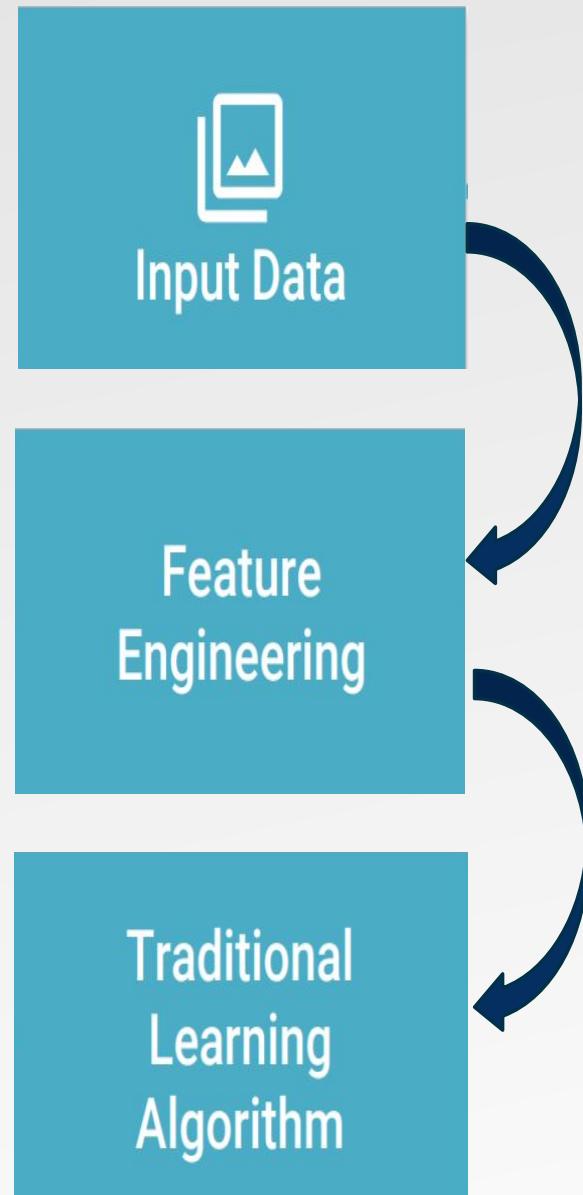
Different type of algorithms

MACHINE LEARNING CLASSIFICATION TASK



Questions Time

Neural Network



CONNECTING THE DOTS ... !

Build a Neural Network

Artificial Neurons

Activation Functions

Synaptic Weights

Bias

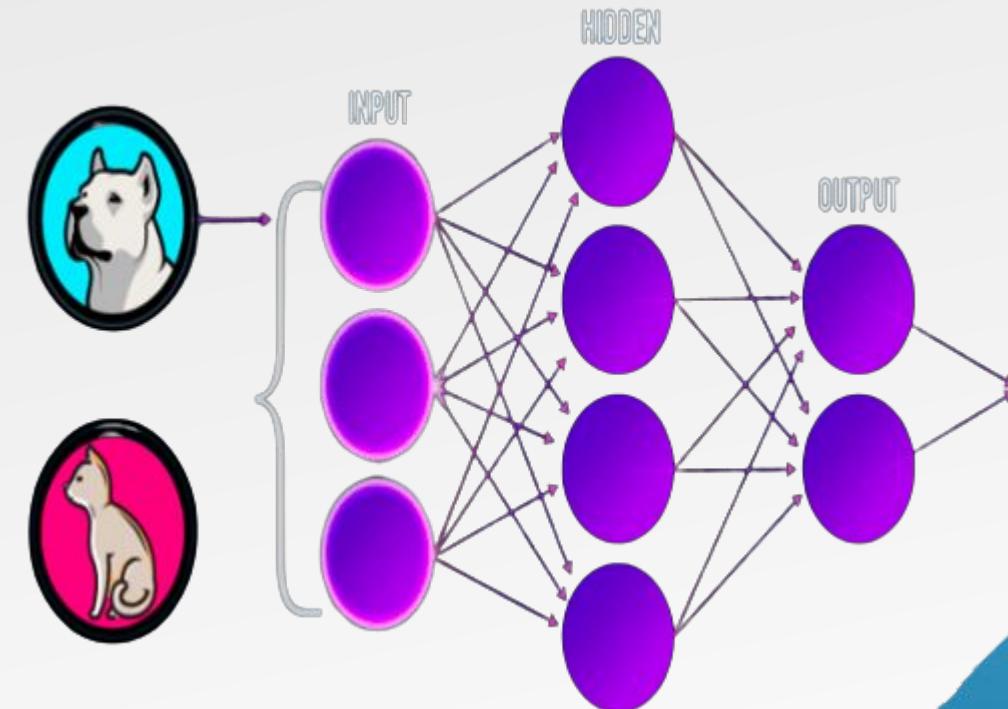
Layers

Data Normalization

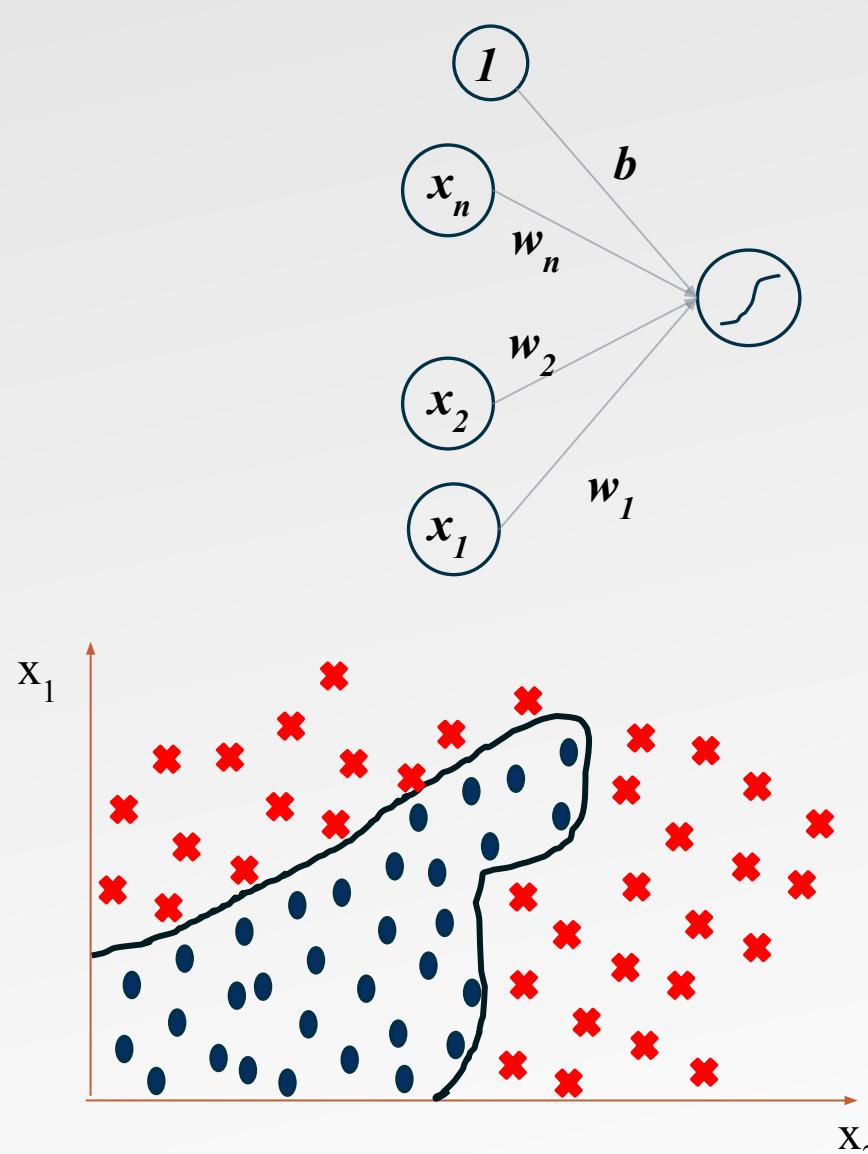
Gradient Descent, Learning

Rate and Loss Function

Epoch, Cross Validation



CONNECTING THE DOTS ... !



Artificial Neuron

- ✓ Connects weights, bias, activation function
 - Neuron pre-activation (or input activation)

$$a(x) = b + \sum_i w_i x_i = b + w^T x$$

- ✓ Neuron (output) activation

$$h(x) = g(a(x)) = g(b + \sum_i w_i x_i)$$

w – connected weights

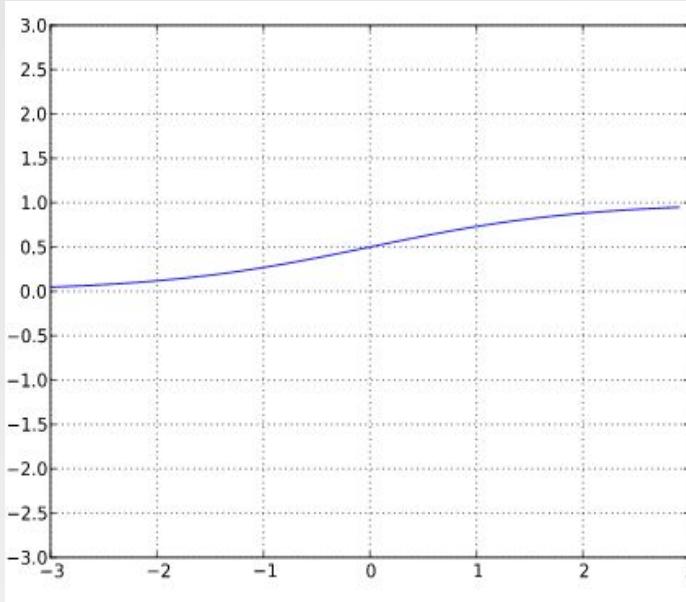
b – neuron bias

g(.) – activation function

$$y = mx + c$$

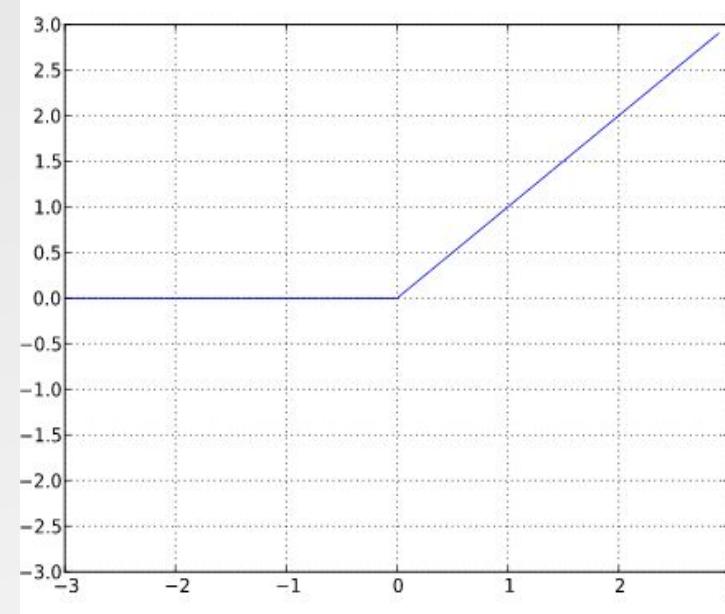
$$y = \sum_{n=1}^N m_n x_n + c_n$$

CONNECTING THE DOTS ... !



Sigmoid Activation

Squashes the neuron's output **between 0 and 1**

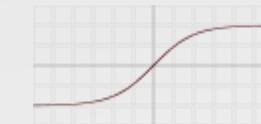


ReLU Activation

Squashes the neuron's output **0 for -ve and linear for positive**

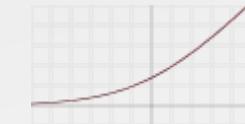
Other popular activation functions

- ✓ Hyperbolic tangent (TanH)



$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$$

- ✓ SoftPlus



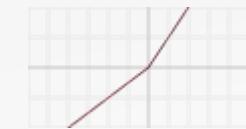
$$f(x) = \ln(1 + e^x)$$

- ✓ Logistic



$$f(x) = \frac{1}{1 + e^{-x}}$$

- ✓ Leaky ReLU



$$f(x) = \begin{cases} 0.01x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

CONNECTING THE DOTS ... !

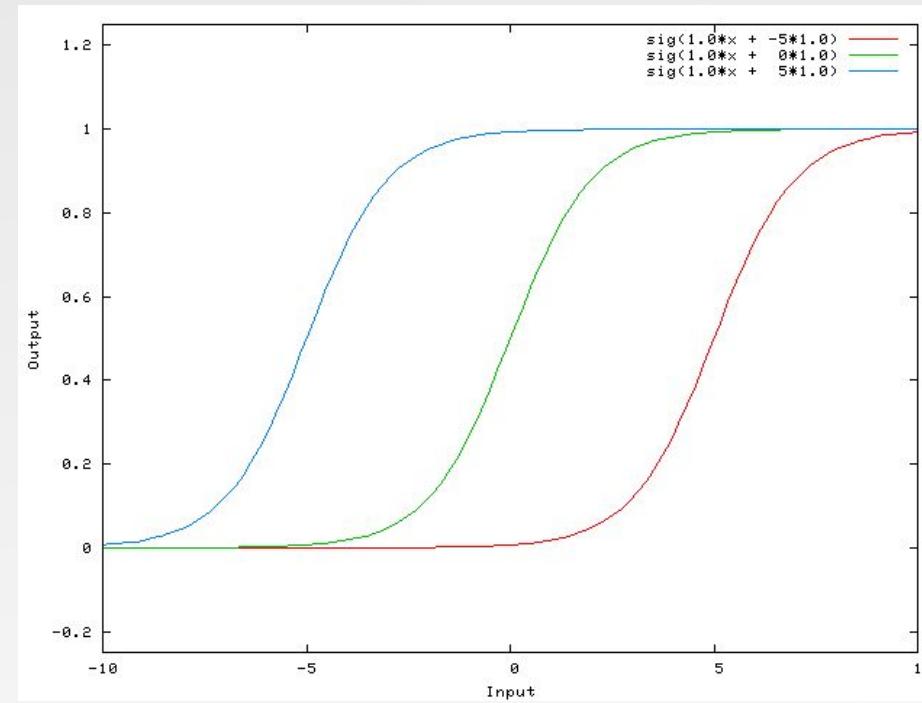
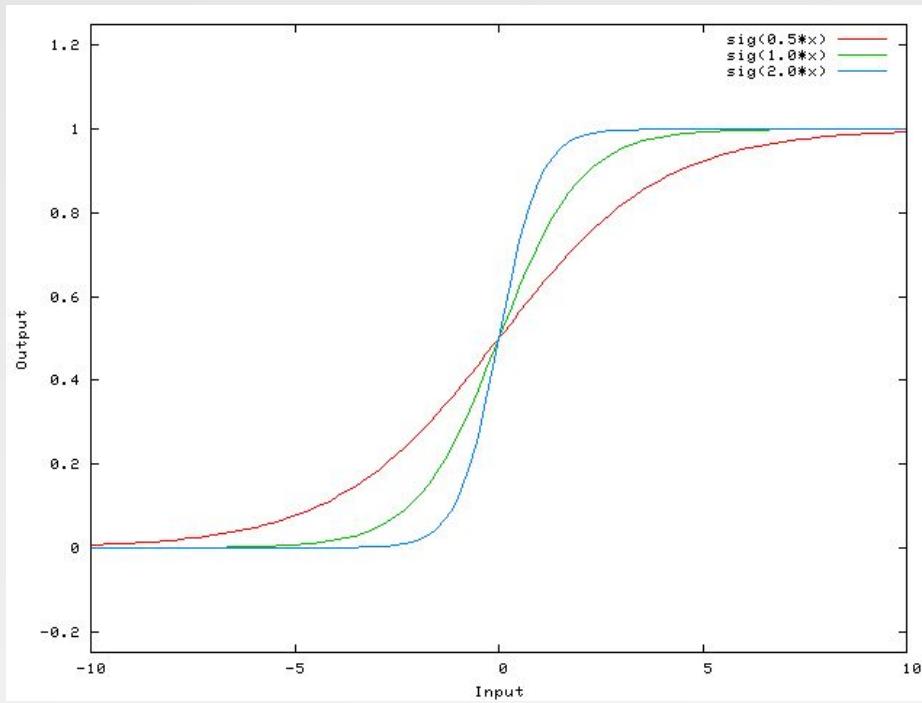


Figure: effect of weights and bias on Sigmoid activation

CONNECTING THE DOTS ... !

Build a Neural Network

Artificial Neurons

Activation Functions

Synaptic Weights

Bias

Layers

Data Normalization

Gradient Descent, Learning

Rate and Loss Function

Epoch, Cross Validation



CONNECTING THE DOTS ... !

Build a Neural Network

Artificial Neurons

Activation Functions

Synaptic Weights

Bias

Layers

Data Normalization

Gradient Descent, Learning

Rate and Loss Function

Epoch, Cross Validation

Normalizing the data is done so that all the input variables have the same treatment in the model and the coefficients of a model are not scaled with respect to the units of the inputs.

CONNECTING THE DOTS ... !

Build a Neural Network

Artificial Neurons

Activation Functions

Synaptic Weights

Bias

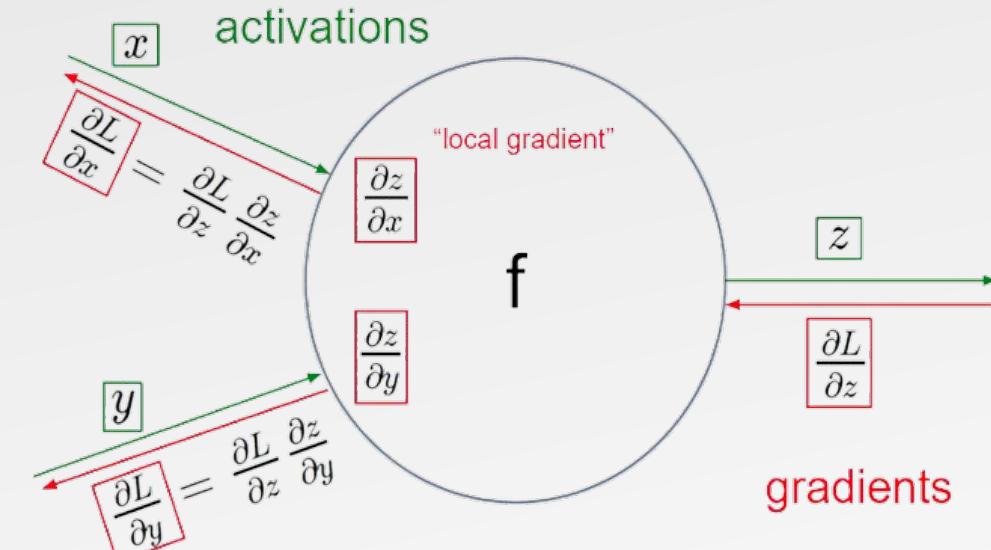
Layers

Data Normalization

Gradient Descent, Learning

Rate and Loss Function

Epoch, Cross Validation



CONNECTING THE DOTS ... !

Build a Neural Network

Artificial Neurons

Activation Functions

Synaptic Weights

Bias

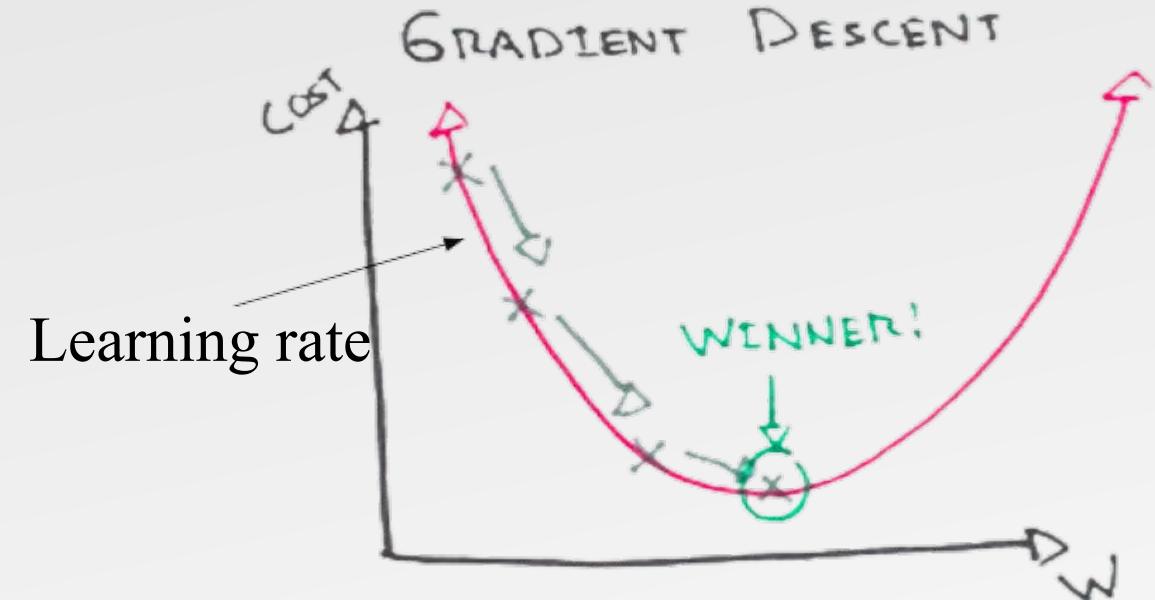
Layers

Data Normalization

Gradient Descent, Learning

Rate and Loss Function

Epoch, Cross Validation



A Loss function tells us “how good” our model is at making predictions for a given set of parameters

CONNECTING THE DOTS ... !

Build a Neural Network

Artificial Neurons

Activation Functions

Synaptic Weights

Bias

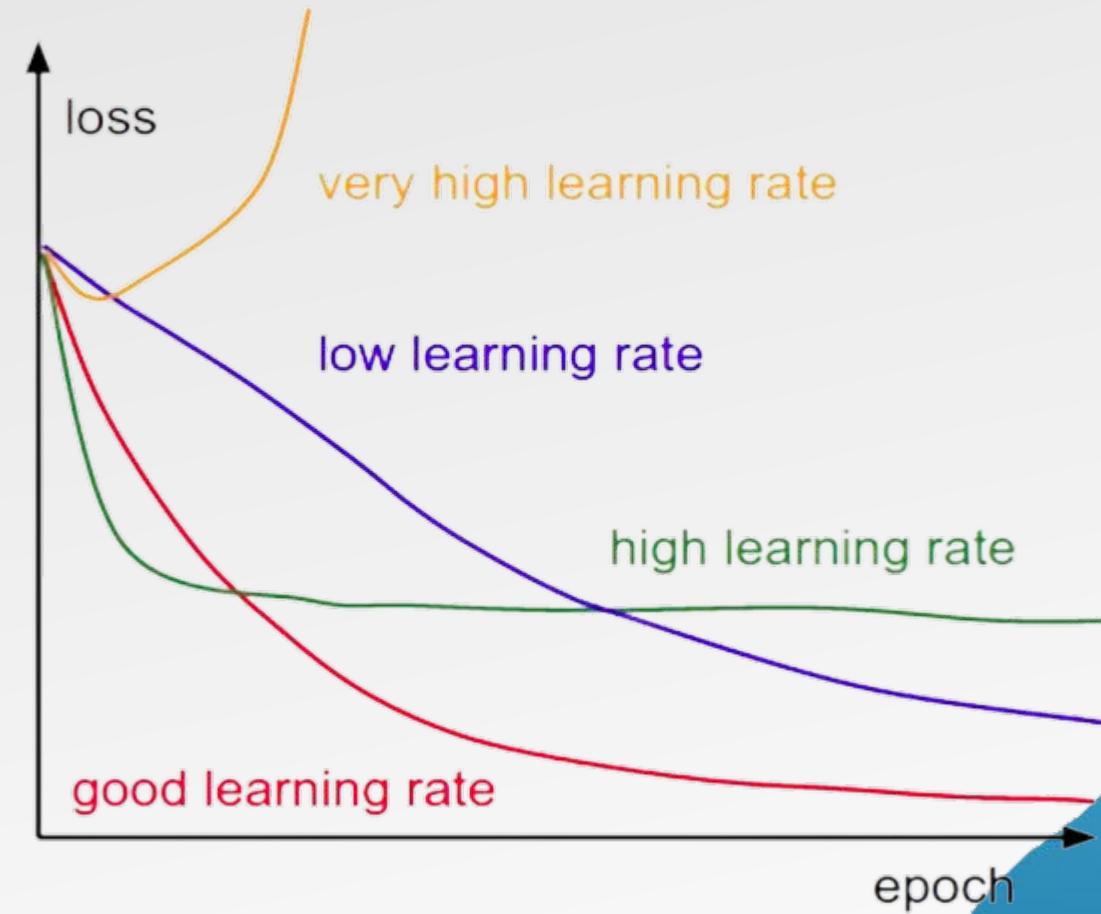
Layers

Data Normalization

Gradient Descent, Learning

Rate and Loss Function

Epoch, Cross Validation



CONNECTING THE DOTS ... !

Build a Neural Network

Artificial Neurons

Activation Functions

Synaptic Weights

Bias

Layers

Data Normalization

Gradient Descent, Learning

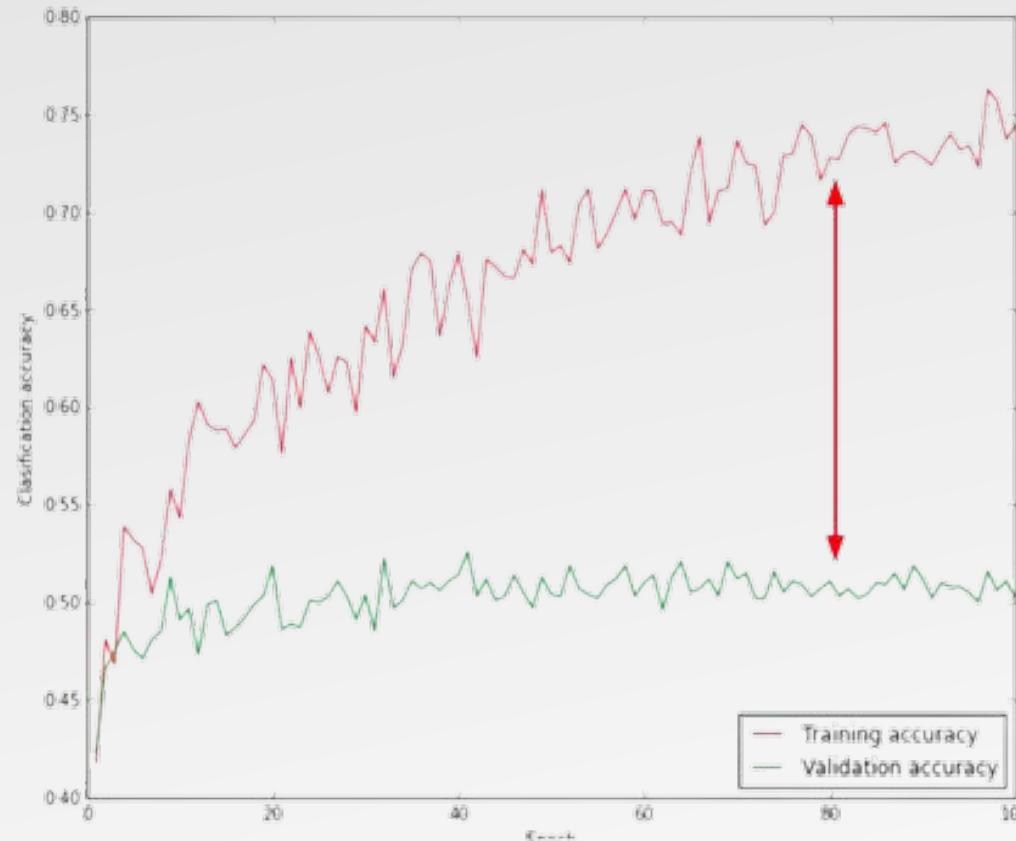
Rate and Loss Function

Epoch, Cross Validation

Epoch is number of iteration taken to get the loss to minimum.

Cross validate the accuracy in order to get the accurate results.

CONNECTING THE DOTS ... !

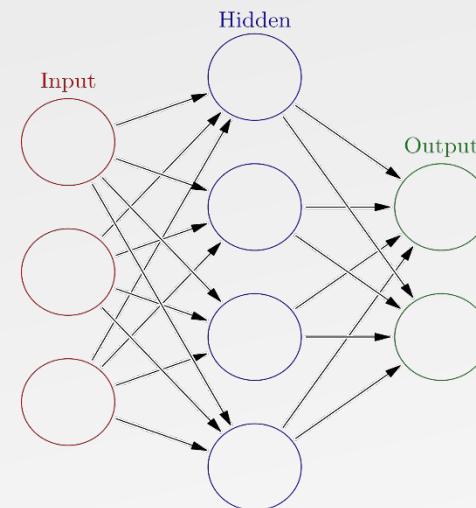
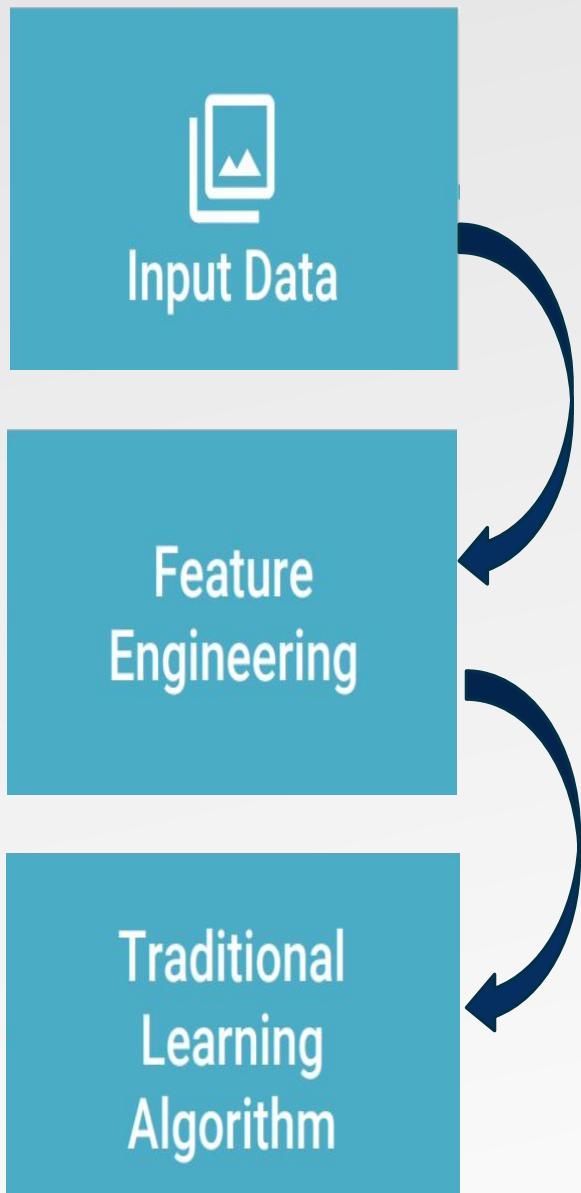


big gap = overfitting
=> increase regularization strength?

no gap
=> increase model capacity?

Given limited amounts of labeled data, training via back-propagation does not work well

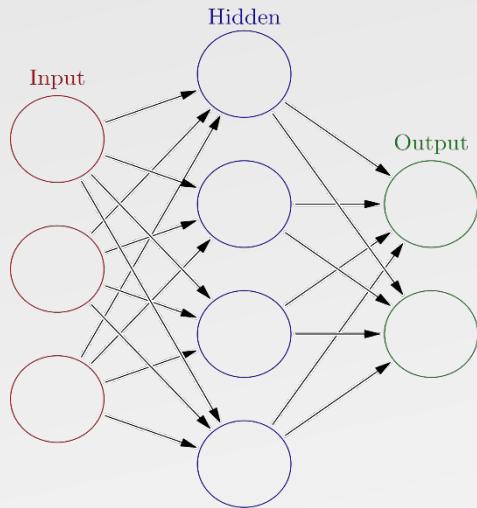
CONNECTING THE DOTS ... !



[3 2]

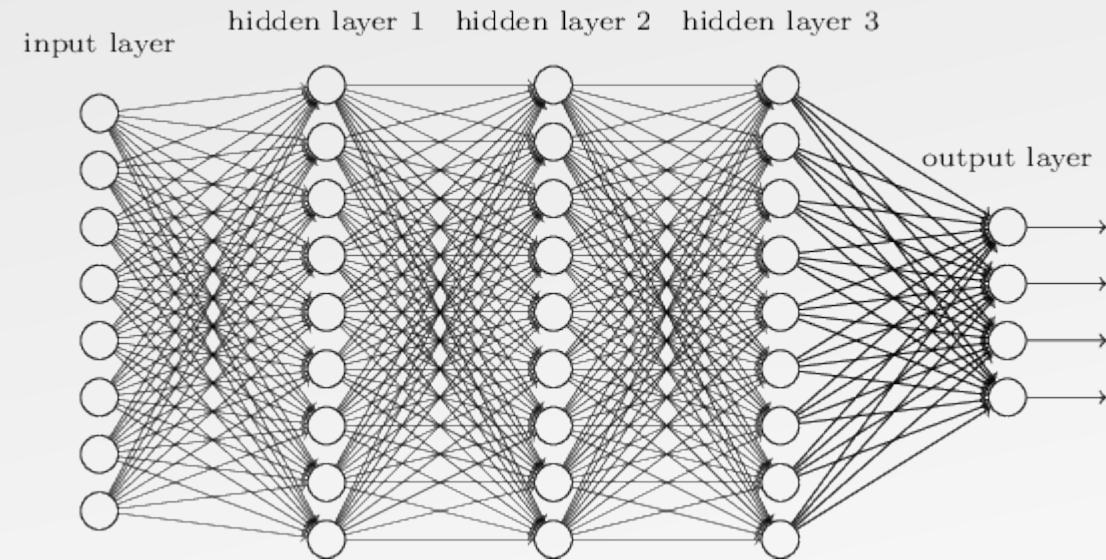
Artificial Neural Network

CONNECTING THE DOTS ... !



[3 2]

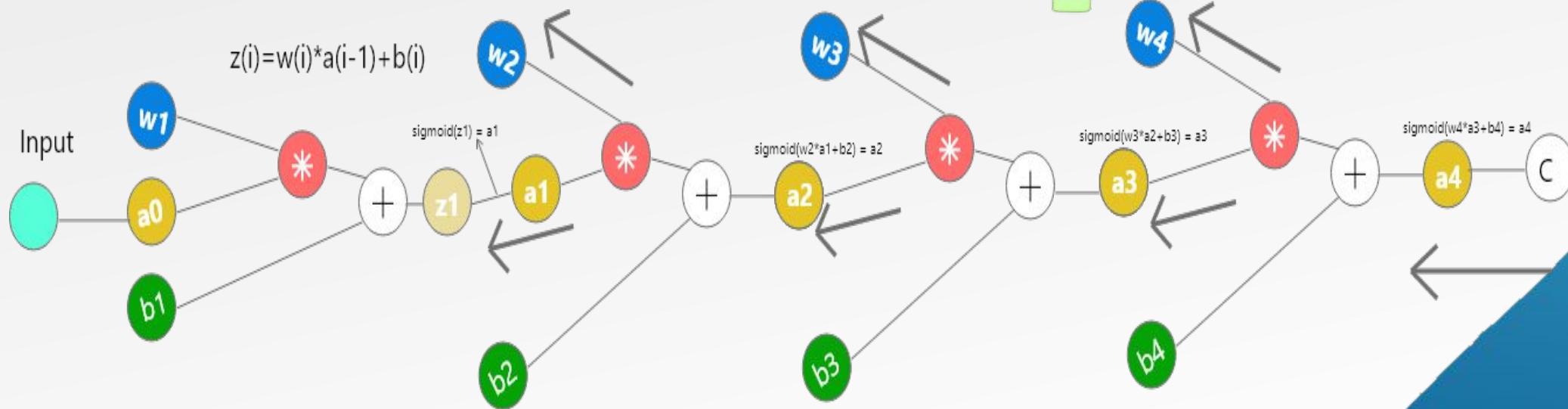
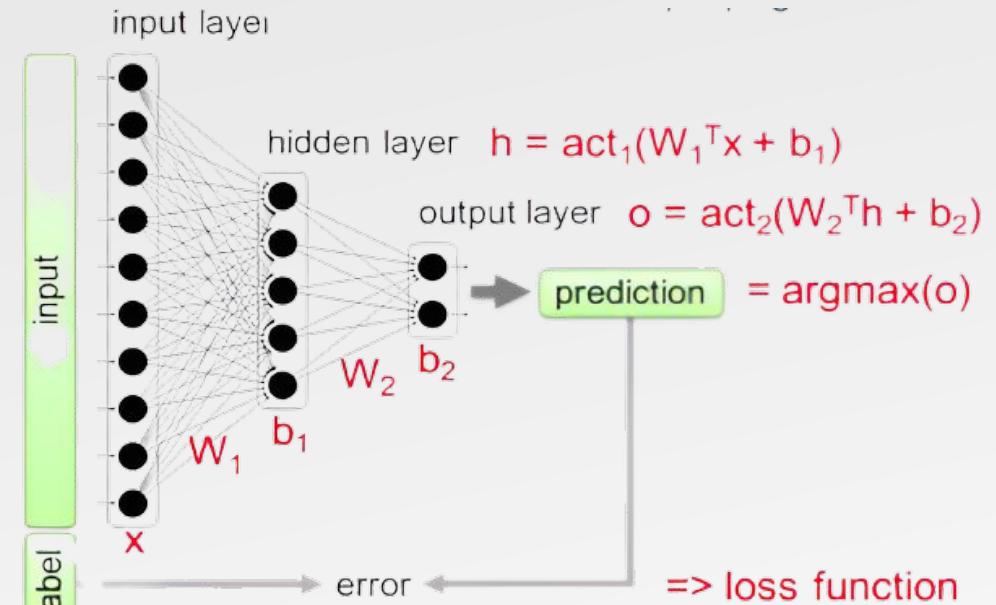
Artificial Neural Network



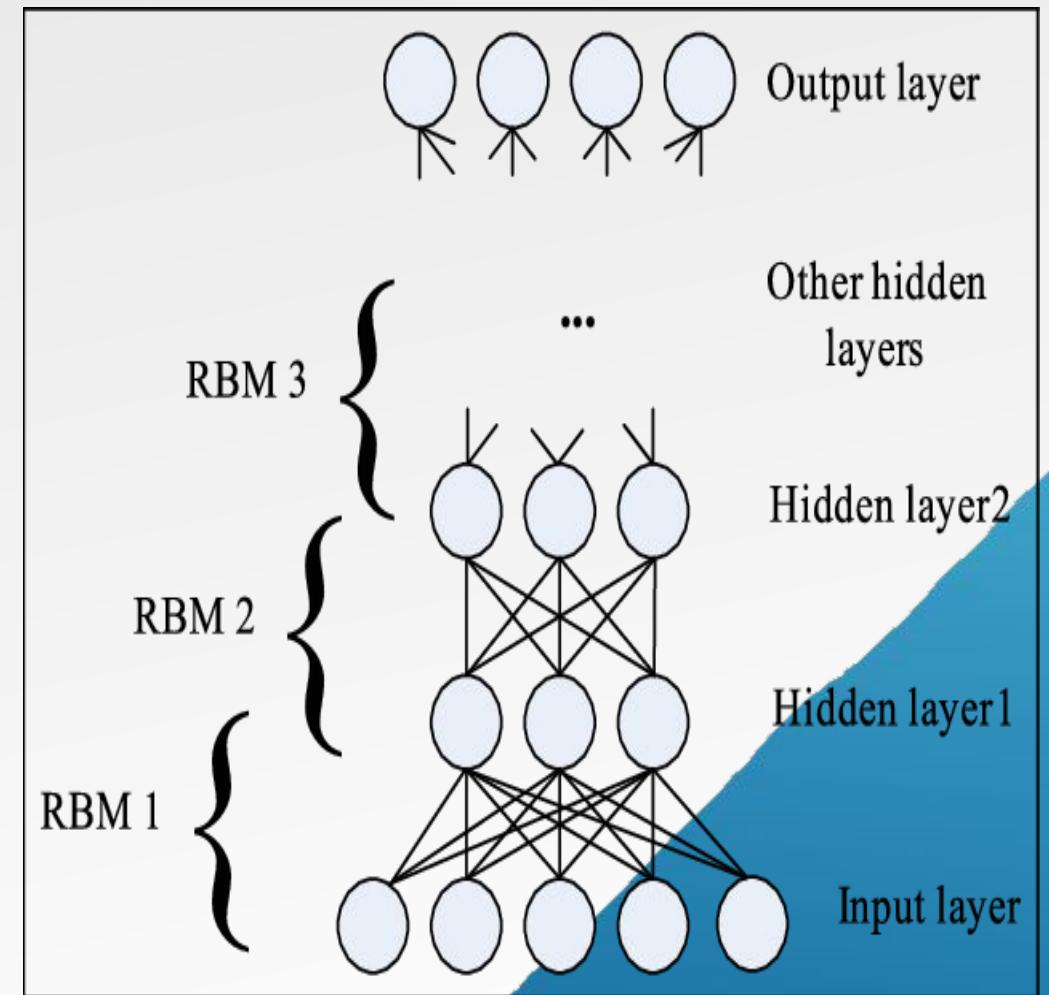
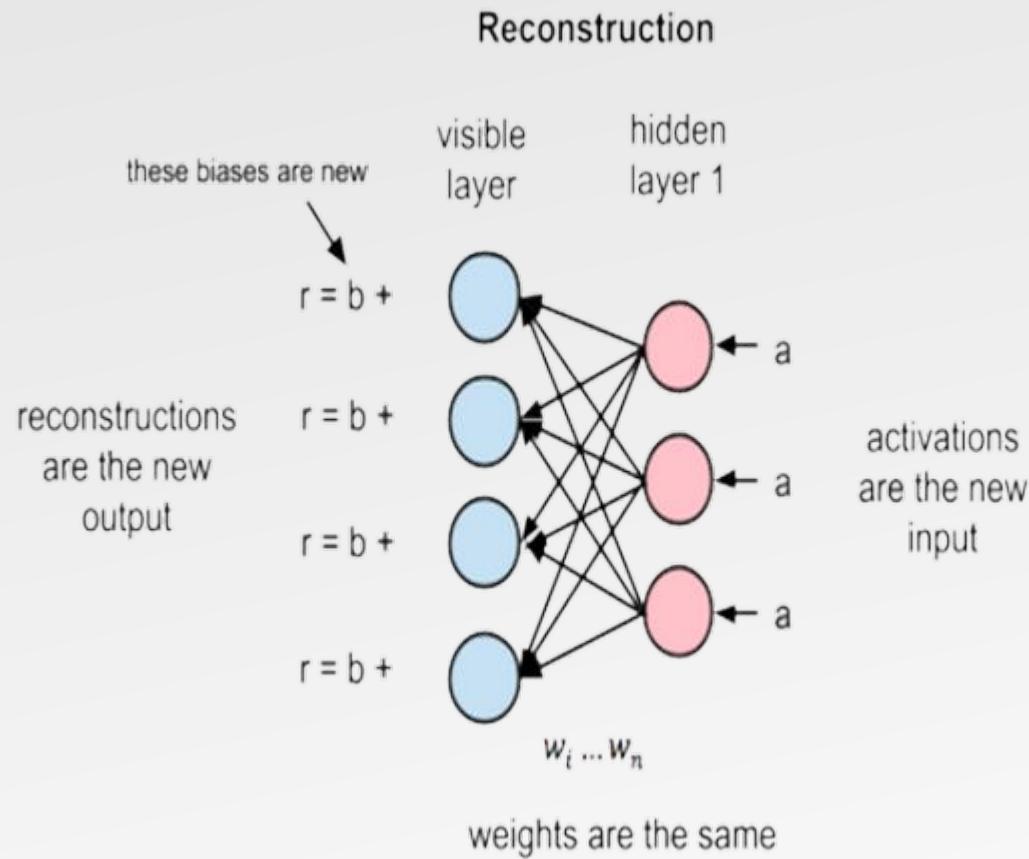
[9 9] [9 9] [9 4]

Deep Neural Network

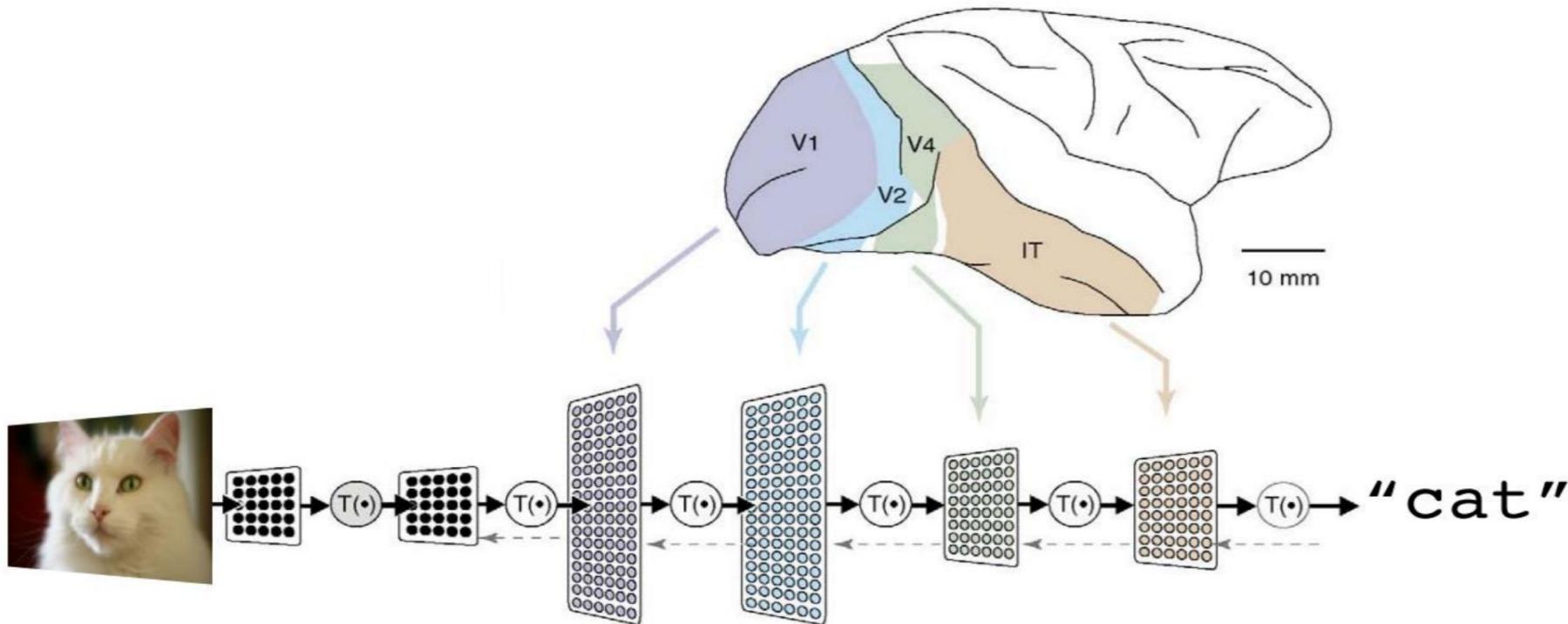
WHAT MADE VANISHING GRADIENT PROBLEM VANISH ?



WHAT MADE VANISHING GRADIENT PROBLEM VANISH ?

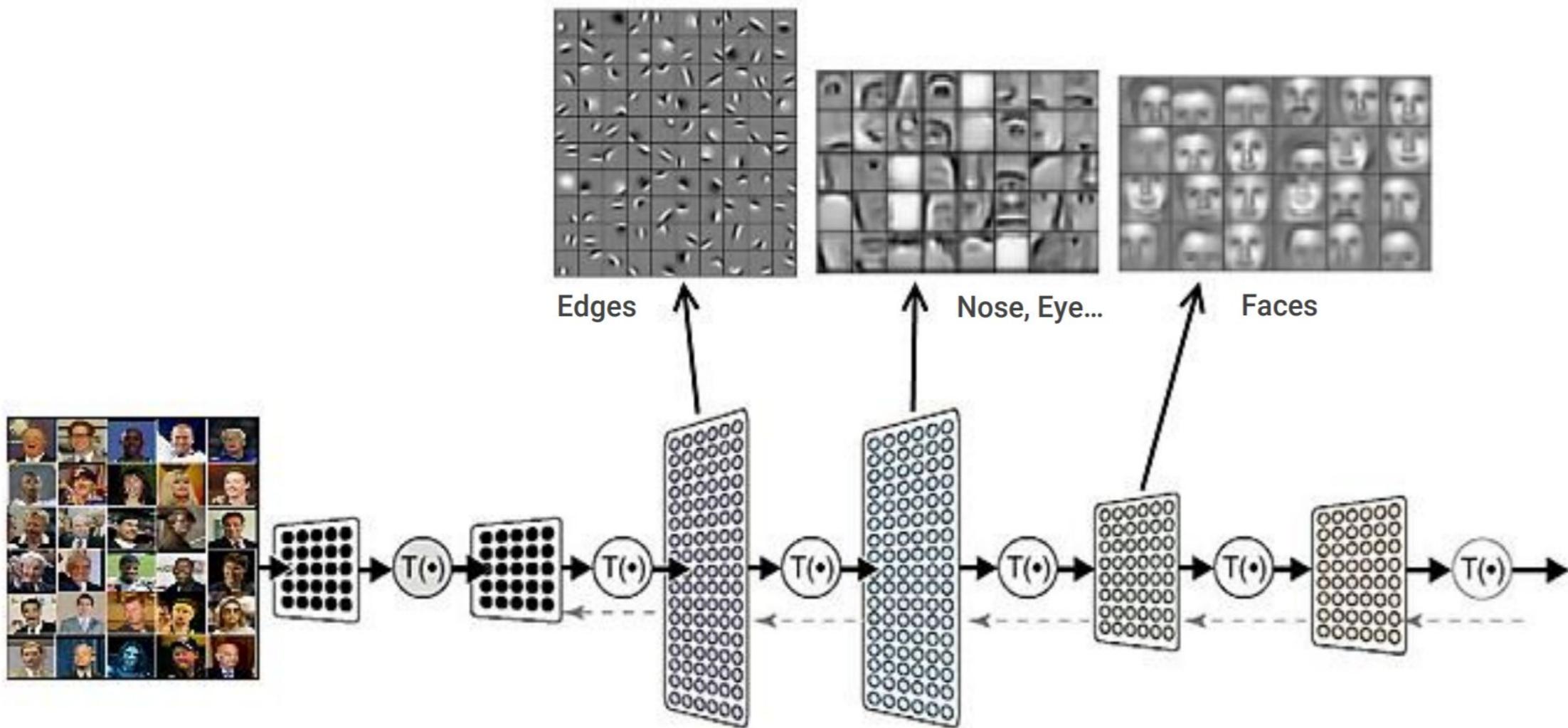


DEEP LEARNING ARCHITECTURE

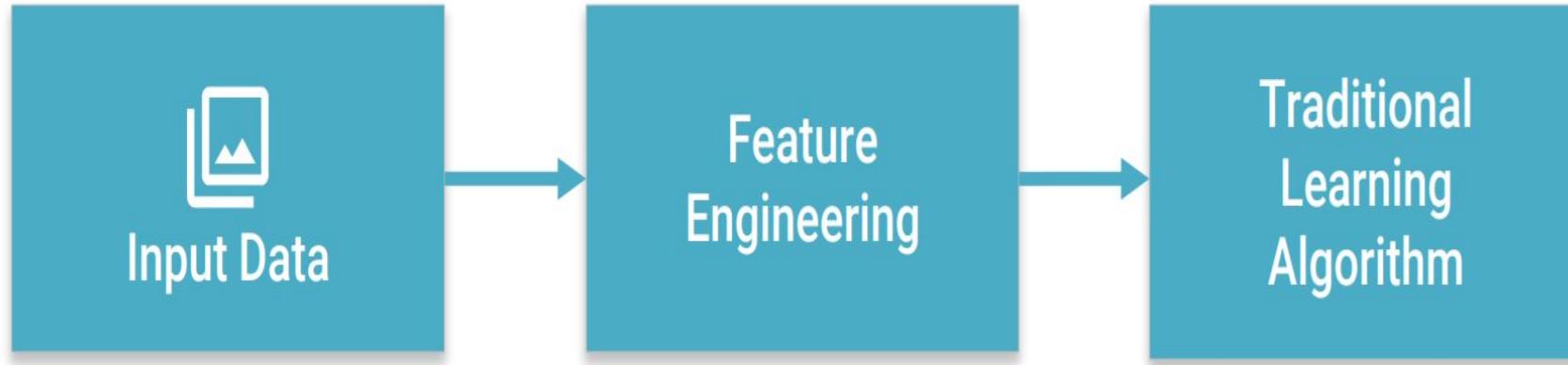


- A deep neural network consists of a hierarchy of layers, whereby each layer transforms the input data into more abstract representations (e.g. edge -> nose -> face). The output layer combines those features to make predictions.

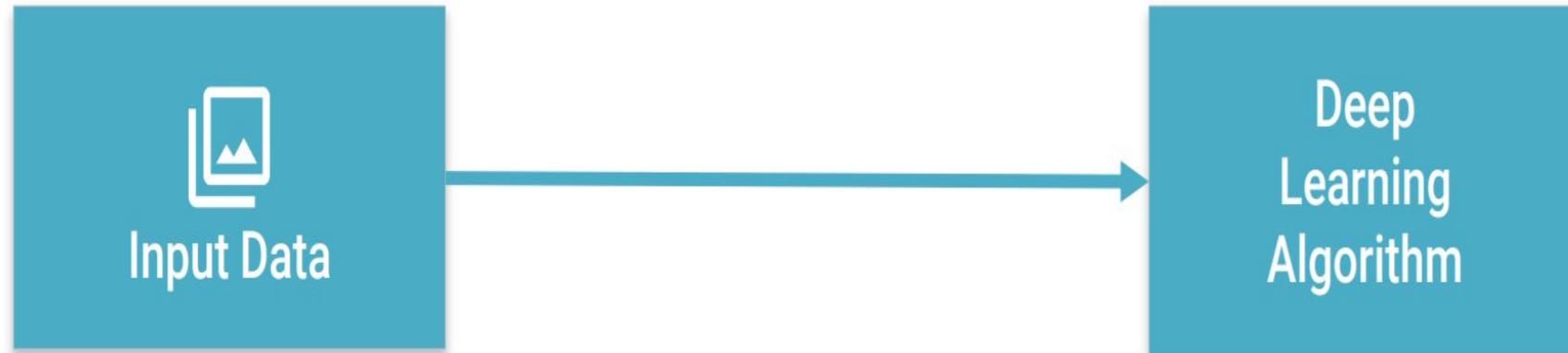
DEEP LEARNING BASICS



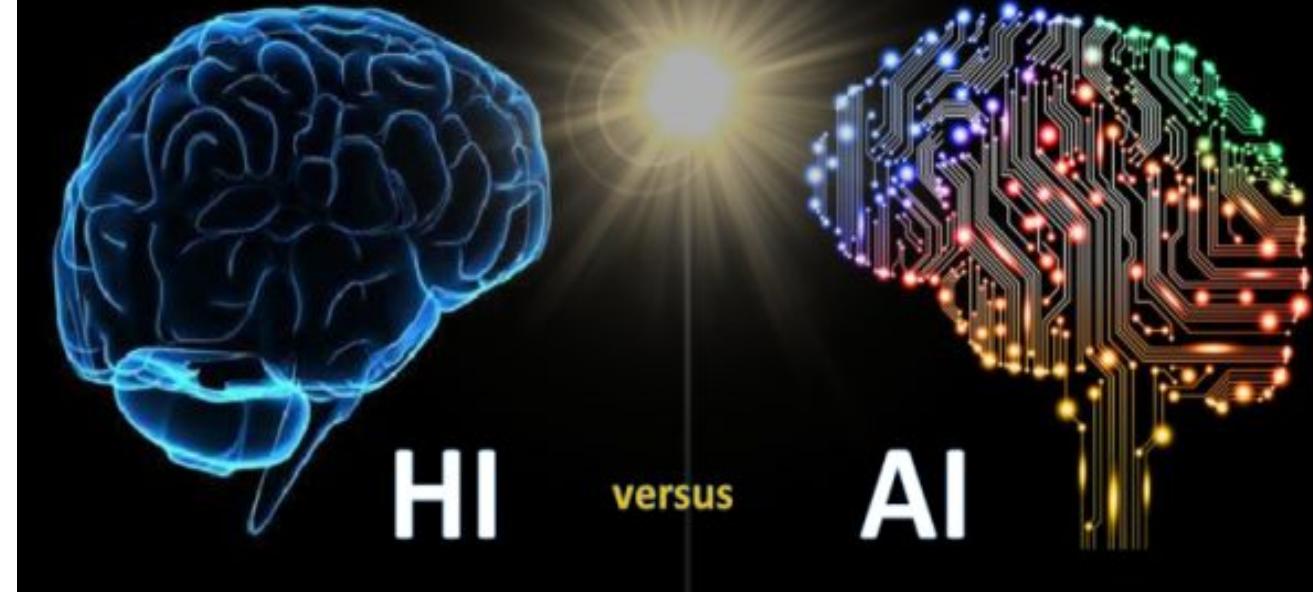
DIFFERENCES BETWEEN TRADITIONAL NETWORKS AND DEEP NETWORKS



Costs lots of time



The Impossible War



COGNITIVE COMPUTERS ARE

- Made with algorithms.
- Knowledgeable only what are taught.
- Control only what we give them control of
- Aware of nuances and can continue to learn more.
- Do very boring work for you.
- Often make better, more consistent decision than humans.
- Be efficient, won't get tired

Human Intelligence is **General**.
Machine Intelligence is **Limited**.

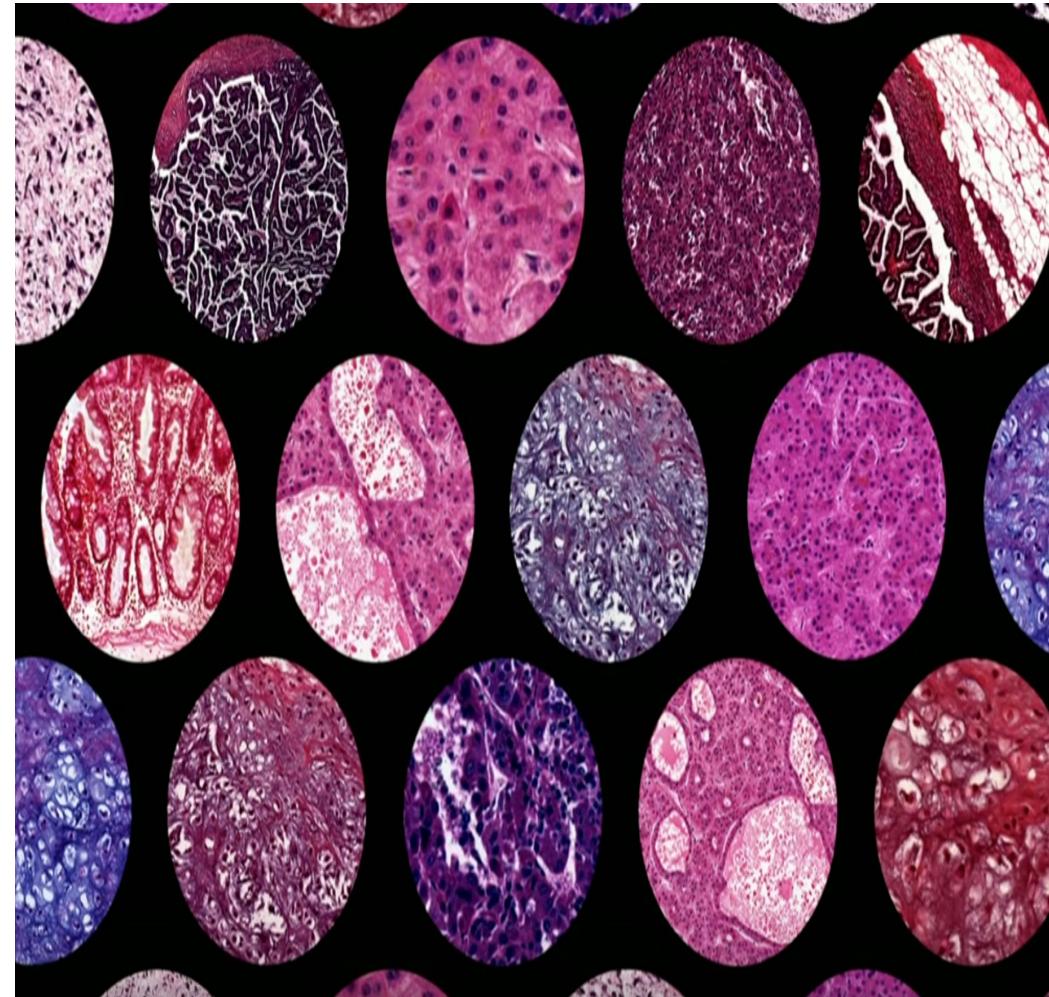
CHALLENGES OF AI

- **Building Trust**
- **AI human interface**
- **Investment**
- **Higher Expectations**



HUMANISTIC AI

Diagnosing Cancer



HUMANISTIC AI

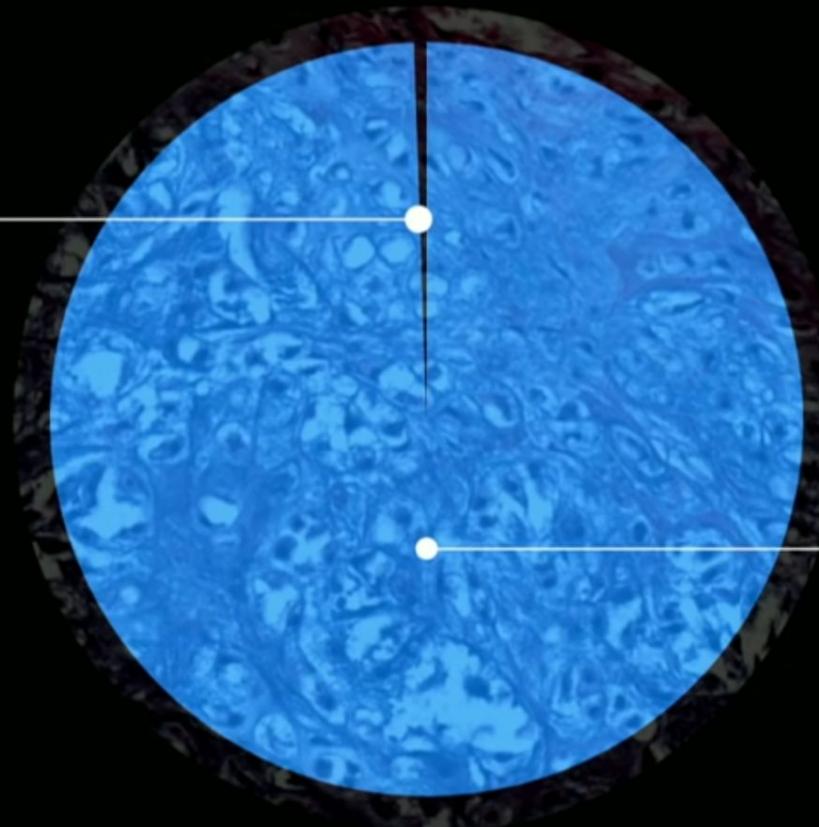
- The human-machine team delivers

Super-Human performance

Diagnosing Cancer

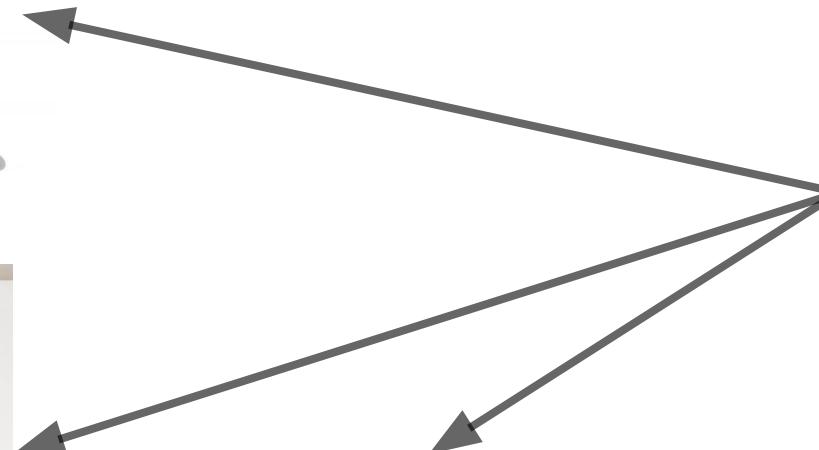
Diagnosis by **Human + AI**

.5%
incorrect



Design

HUMANISTIC AI

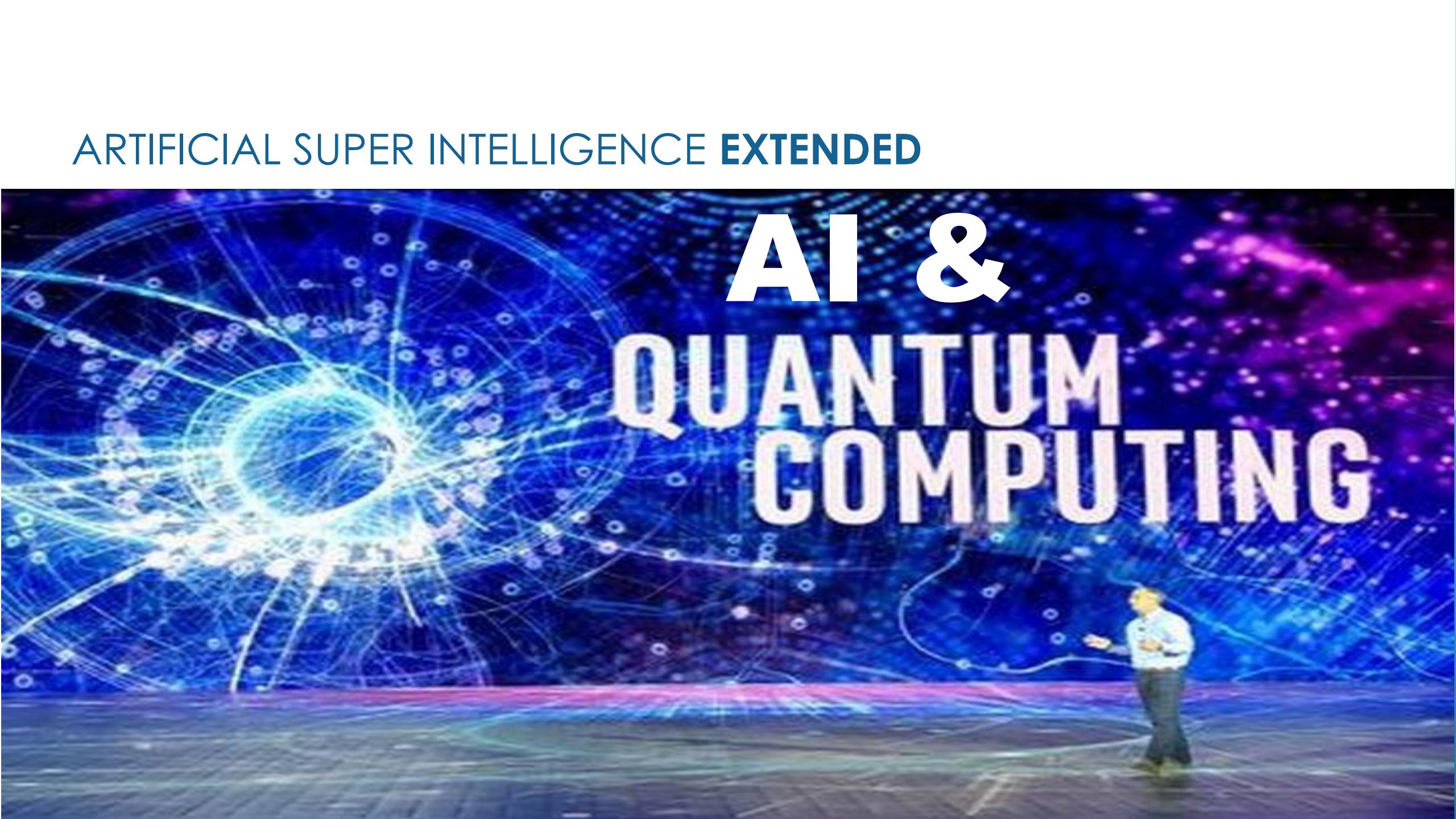


Design

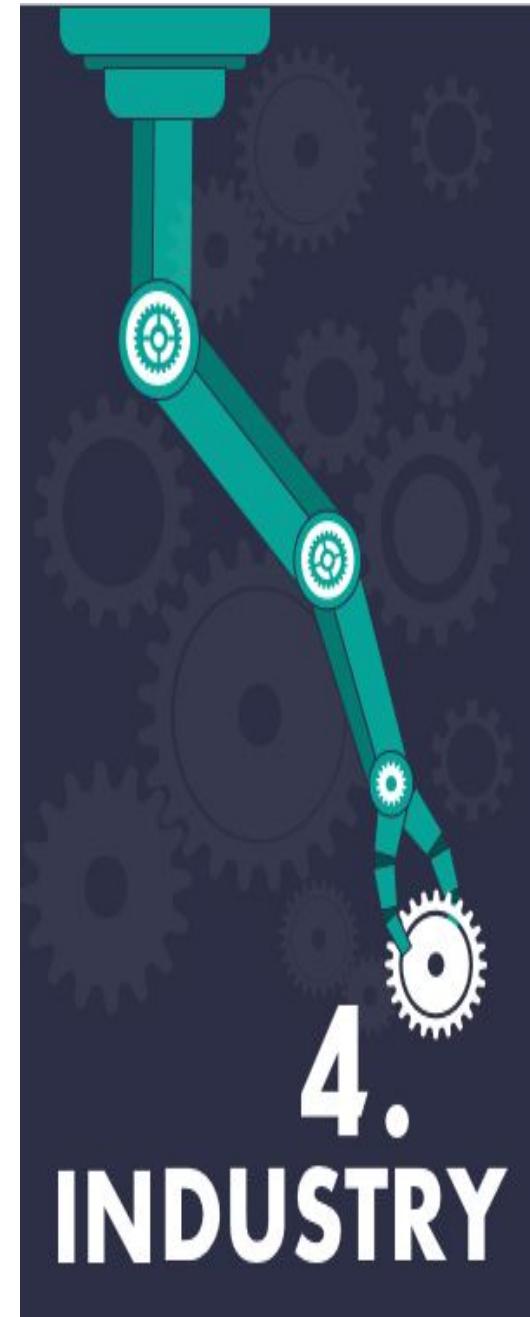
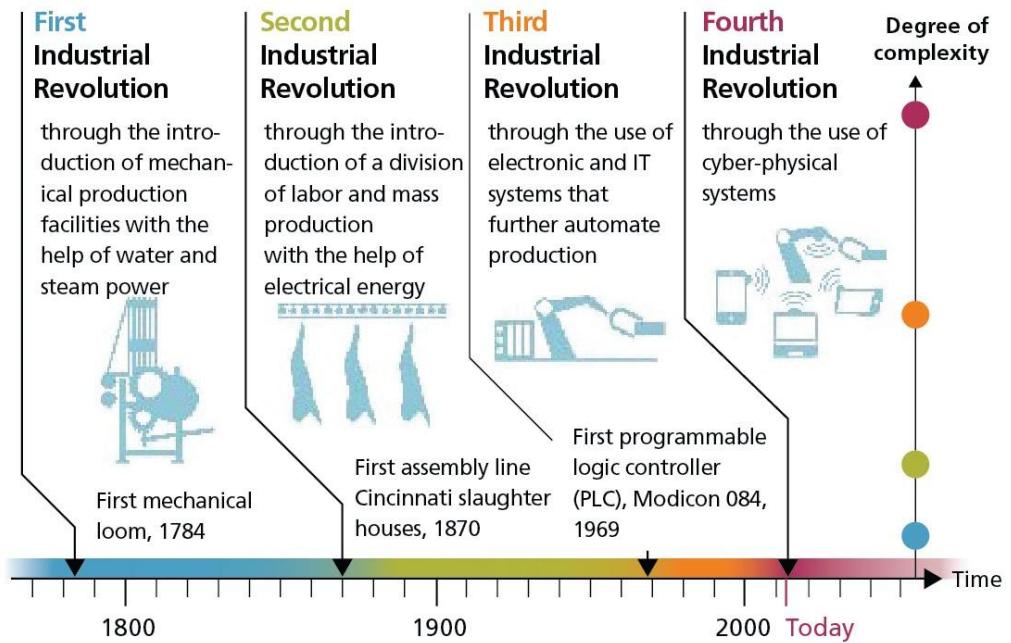
HUMANISTIC AI



ARTIFICIAL SUPER INTELLIGENCE EXTENDED

A man stands on a stage, gesturing towards a massive, glowing digital network visualization that fills the background. The visualization consists of a complex web of blue and white lines forming a globe-like structure, with numerous small circular nodes of varying sizes scattered across it, suggesting a global-scale AI or quantum computing system.

AI & QUANTUM COMPUTING



Ref Image: Visualization of 4 industrial revolution –
www.automationeverywhere.com

