Concepts of Machine Learning

Day 2

GNITS AI Workshop

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

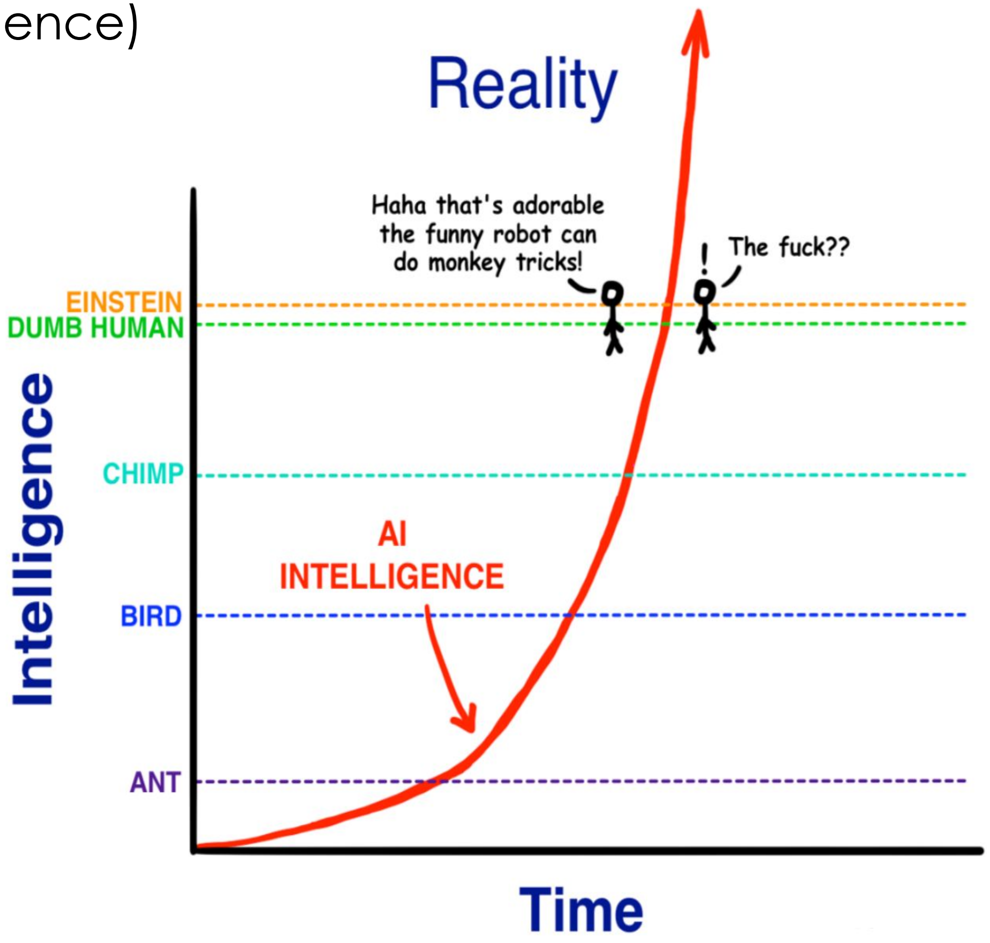
- **Exhibit Intelligence**
- **Perceive their environments**
- **Take actions/ make decision to maximize chance of success at a goal**

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).

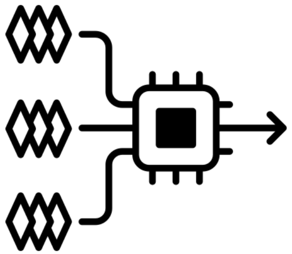
- **Unsupervised learning:** Discovering patterns in unlabeled data

  Example: Cluster similar documents based on the content type.

- **Reinforcement Learning**: learning based on feedback or reward

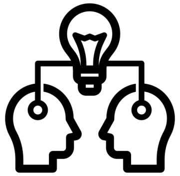
  Example: learn to play chess by winning or loosing.

LEARNING APPROACHES



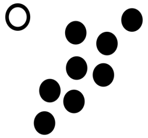
- **Semi - Supervised learning:** involves usage of both labelled and unlabeled data to train a model. This is useful when labeled data is expensive or difficult to obtain.

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



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

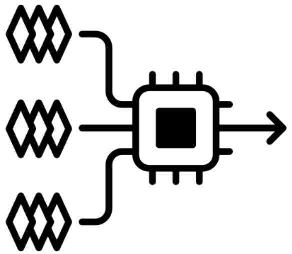
  Example: Speech Recognition using BERT, GPT.



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

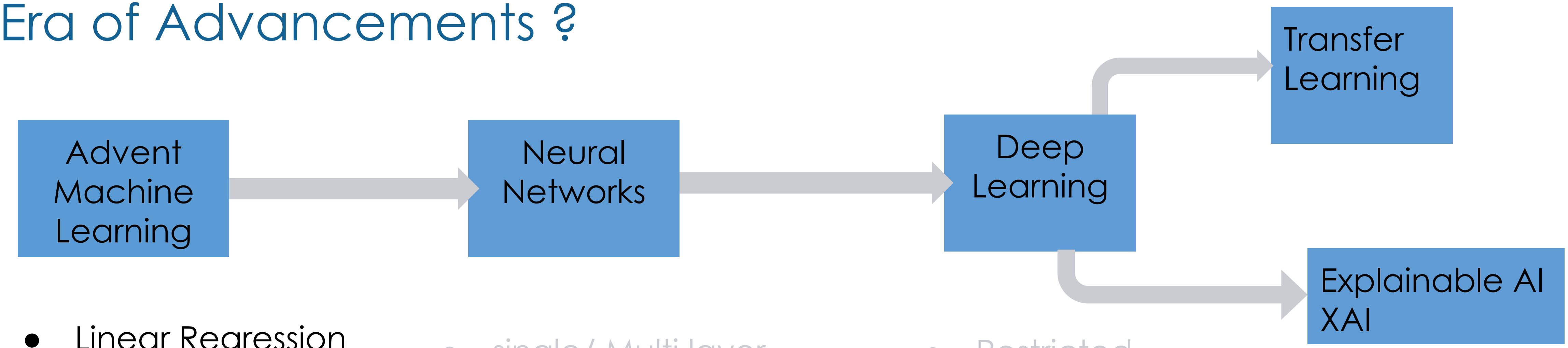
Situations: mostly driven by researchers instead of industries.

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 ?

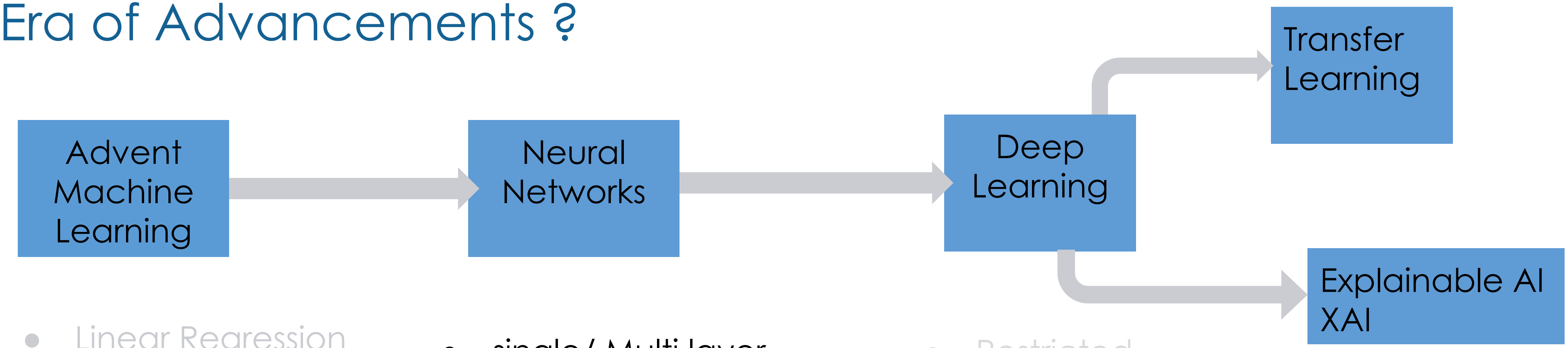


- Linear Regression
- Decision Trees
- SVM
- k-NN
- Naive Bayes

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

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

Era of Advancements ?

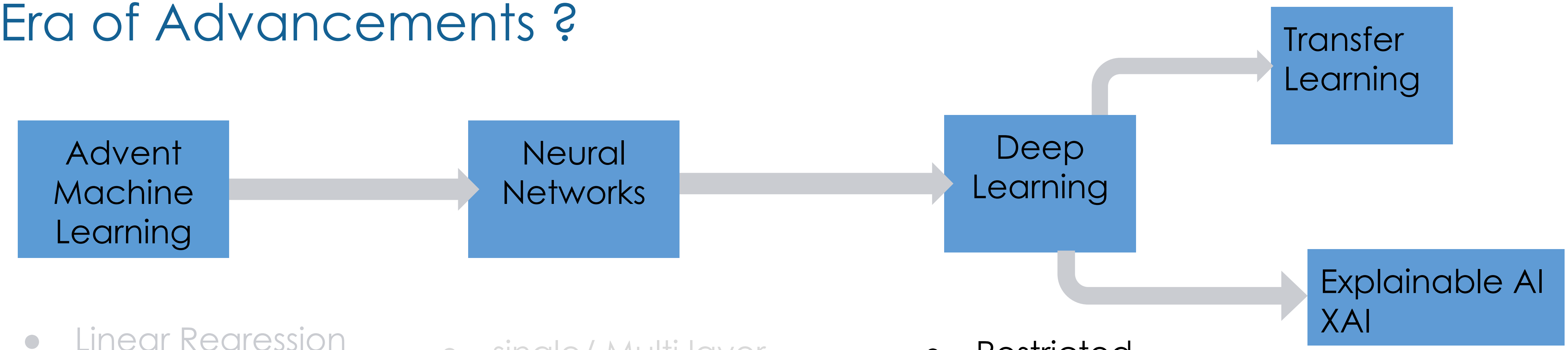


- Linear Regression
- Decision Trees
- SVM
- k-NN
- Naive Bayes

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

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

Era of Advancements ?



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

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

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

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

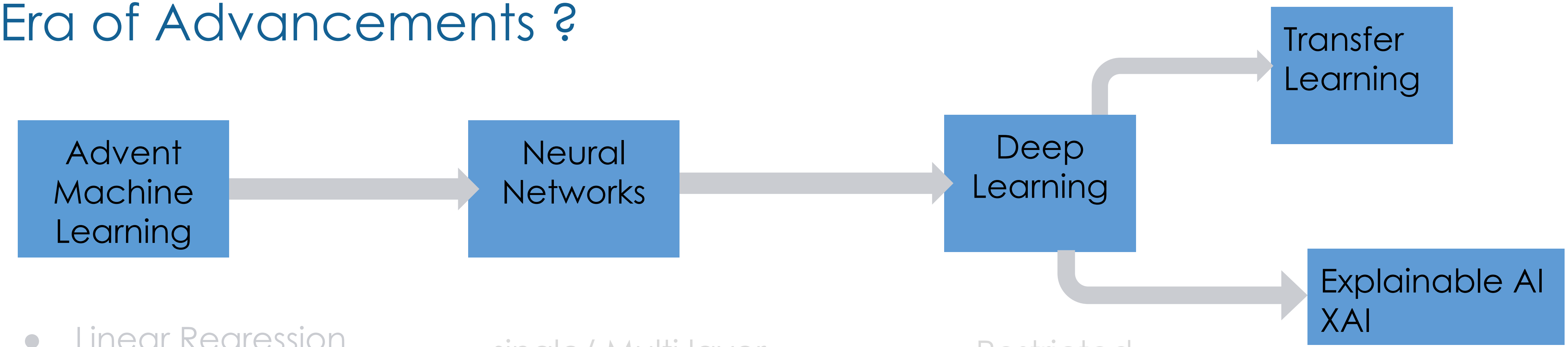
- Restricted BoltzMann Machine (RBM)
- Convolution Neural Network
- RNN
- Auto Encoders
- Deep Belief Networks
- LSTM

Transfer Learning

- Language Models
- GPT
- Transformer Networks

Explainable AI XAI

Era of Advancements ?



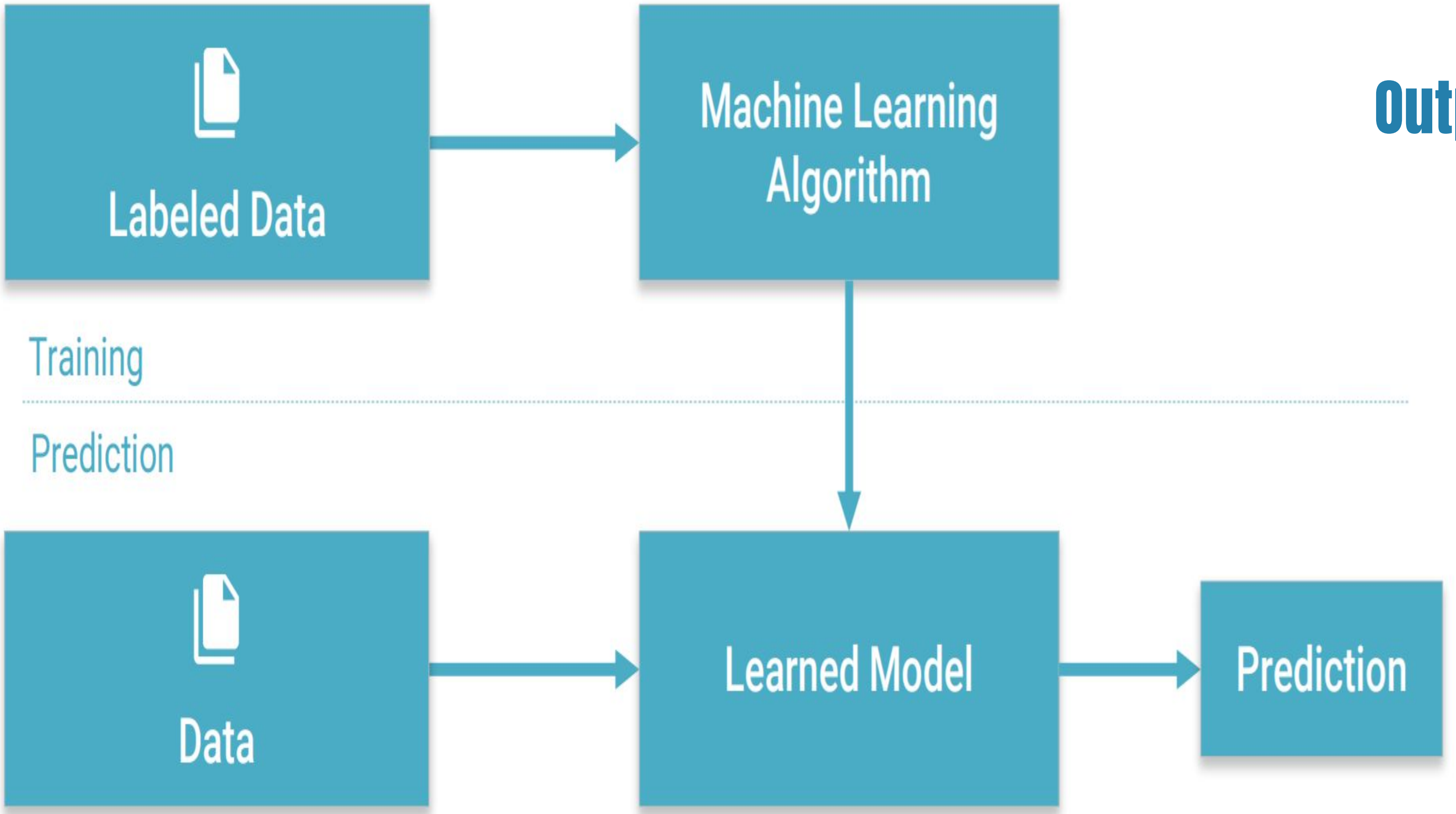
- Linear Regression
- Decision Trees
- SVM
- k-NN
- Naive Bayes

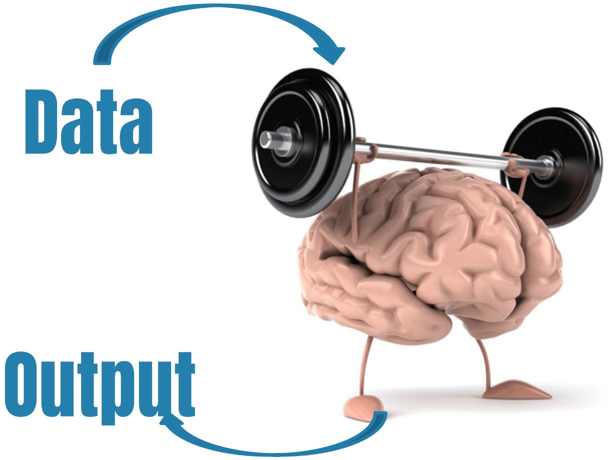
- single/ Multi layer Perceptrons
- Radial Basis Function Network
- Boltzmann Machine
- Hopfield Network
- Self Organizing Map (SOM)

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

- SHAP (Shapley Additive Explanations)
- LIME (Local Interpretable Model-Agnostic Explanations

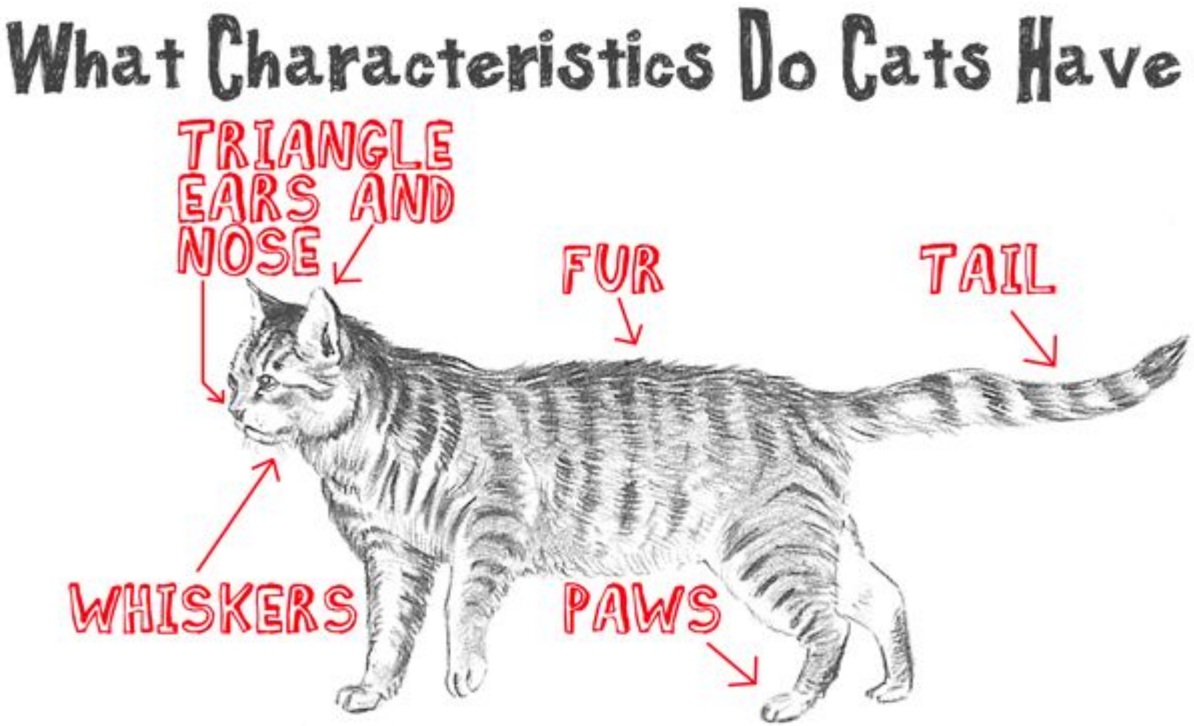
MACHINE LEARNING BASICS

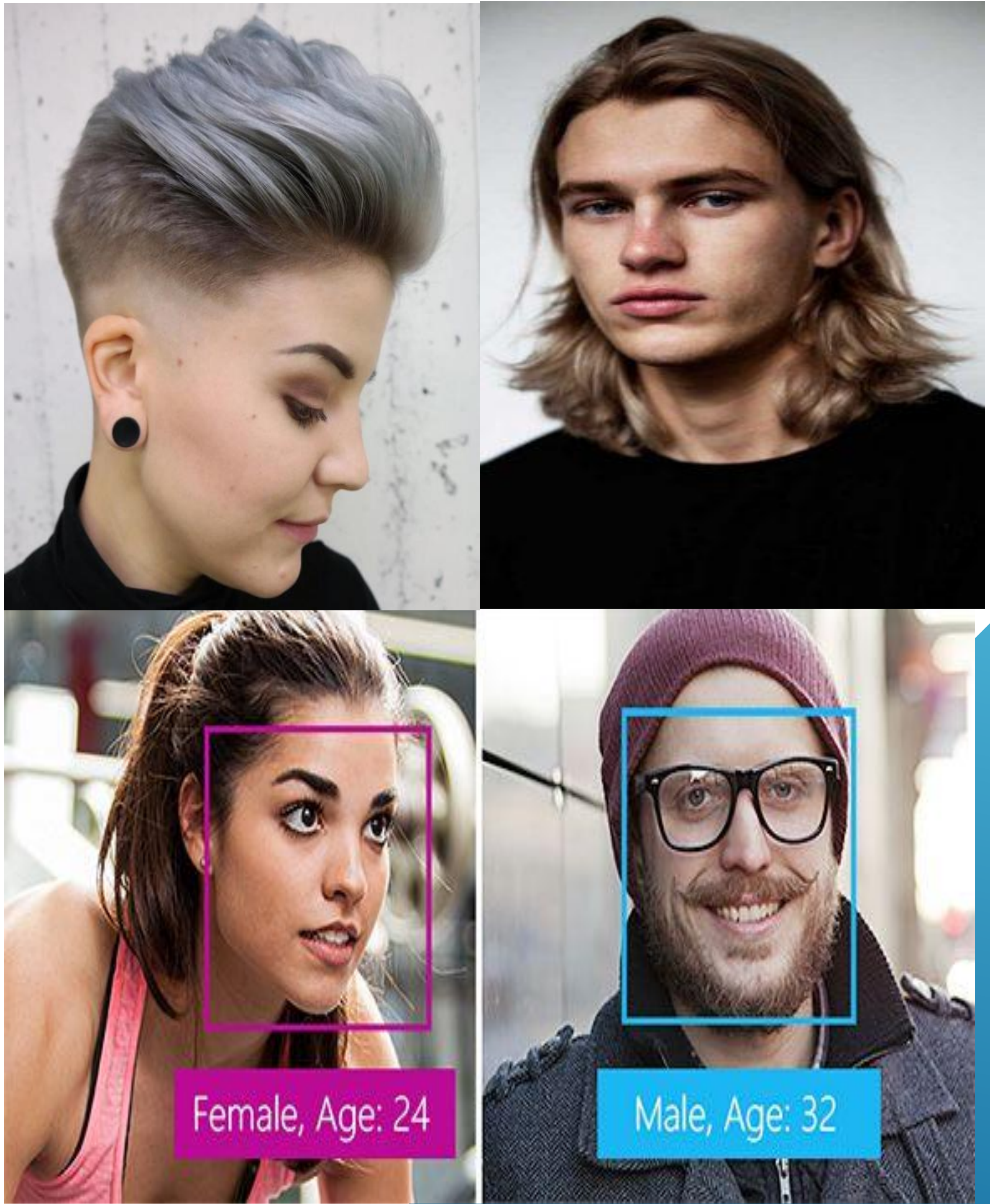




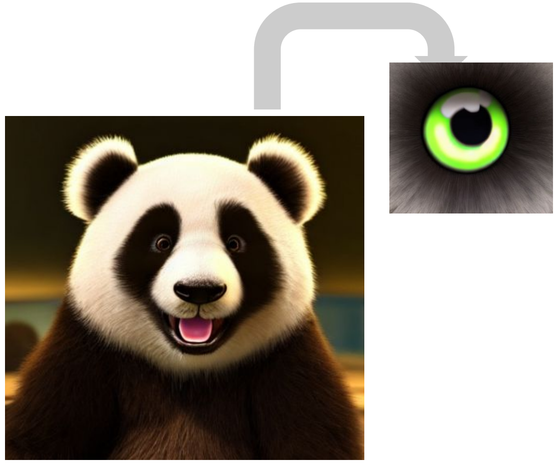
FEATURE SELECTION & FEATURE EXTRACTION

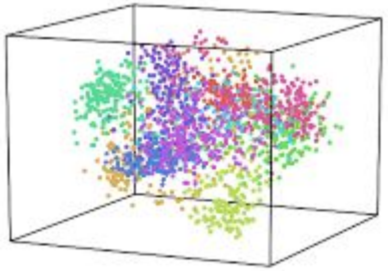
A simple way to think about supervised learning.

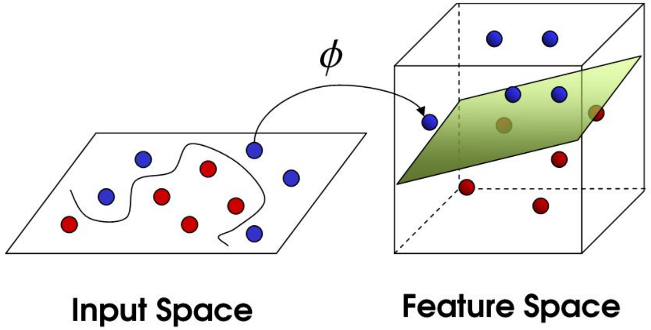




MACHINE LEARNING BASICS

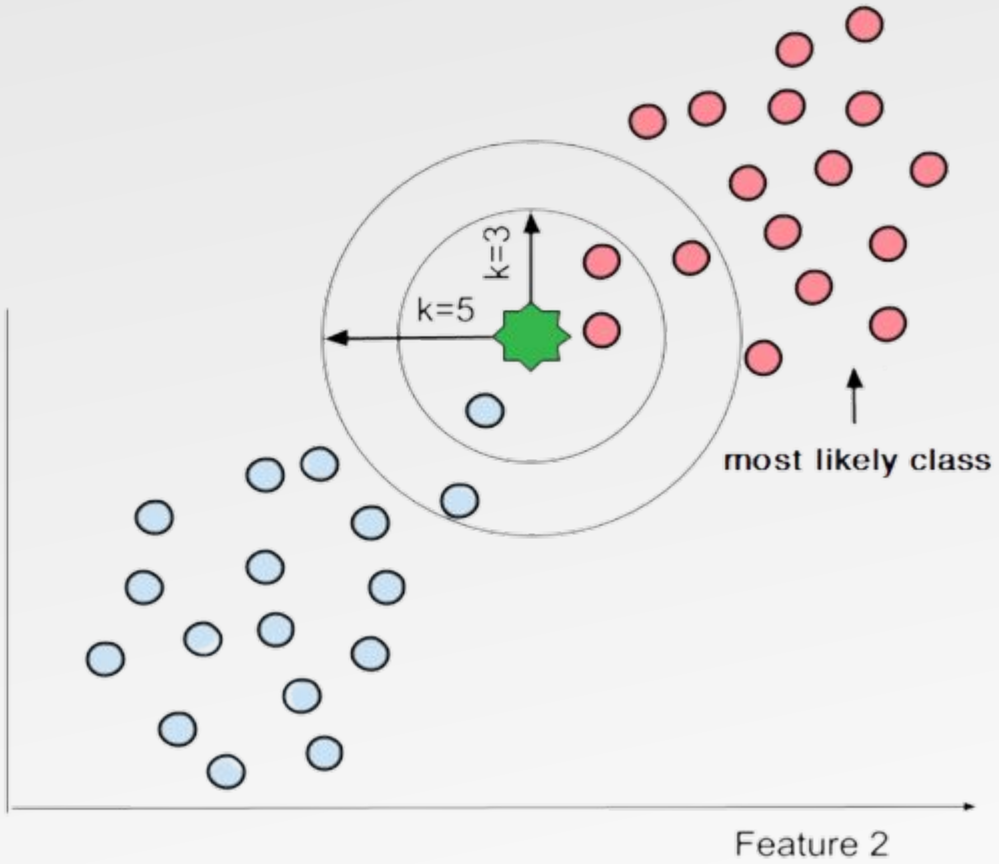






Features Selection → Feature Extraction → Vector Space → ML Algorithms

K NEAREST NEIGHBORS (*k*-NN)



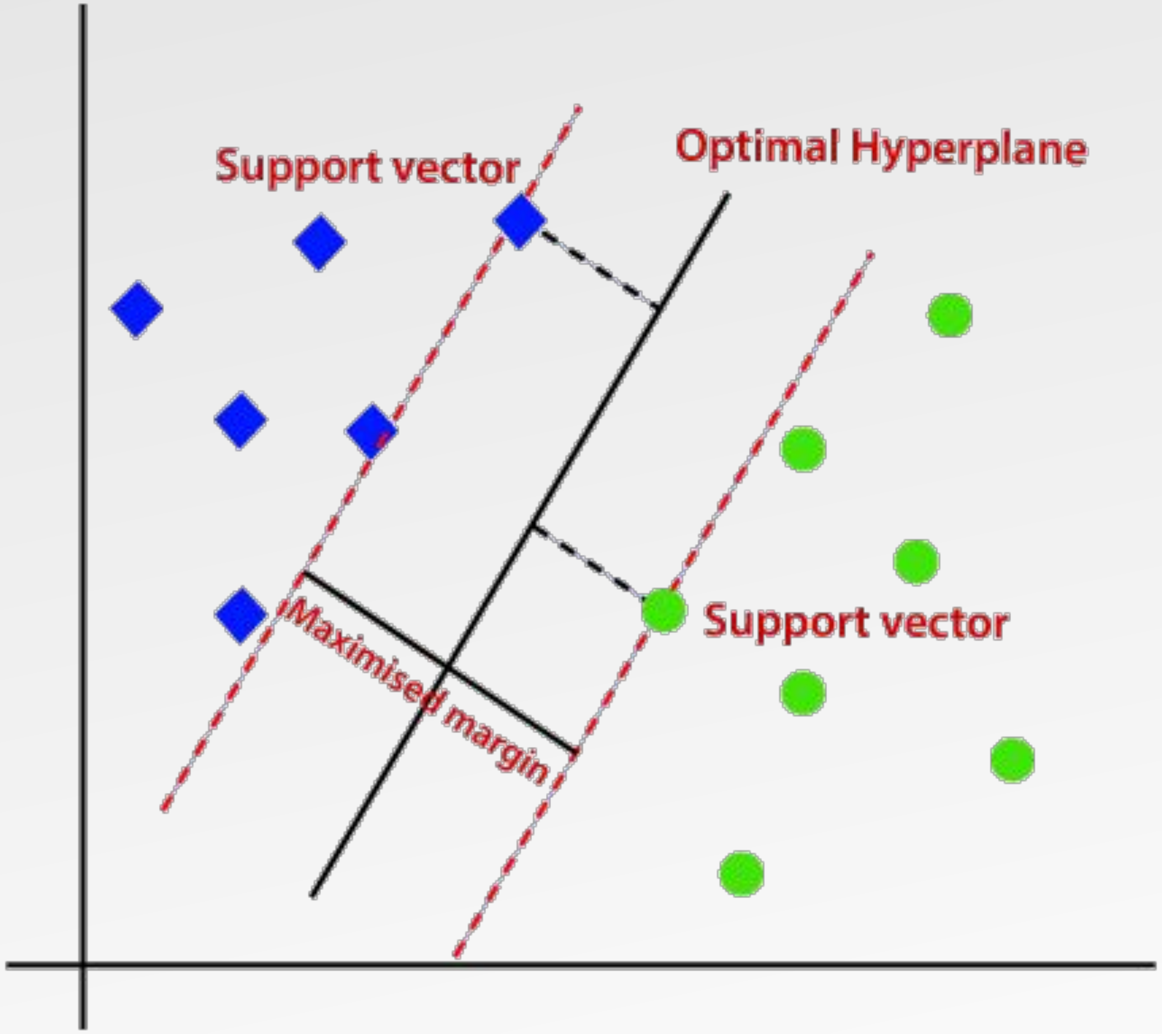
The simplest algorithm

No training is required

Dependent on value of *k*

k value

Support Vector Machines (SVM)



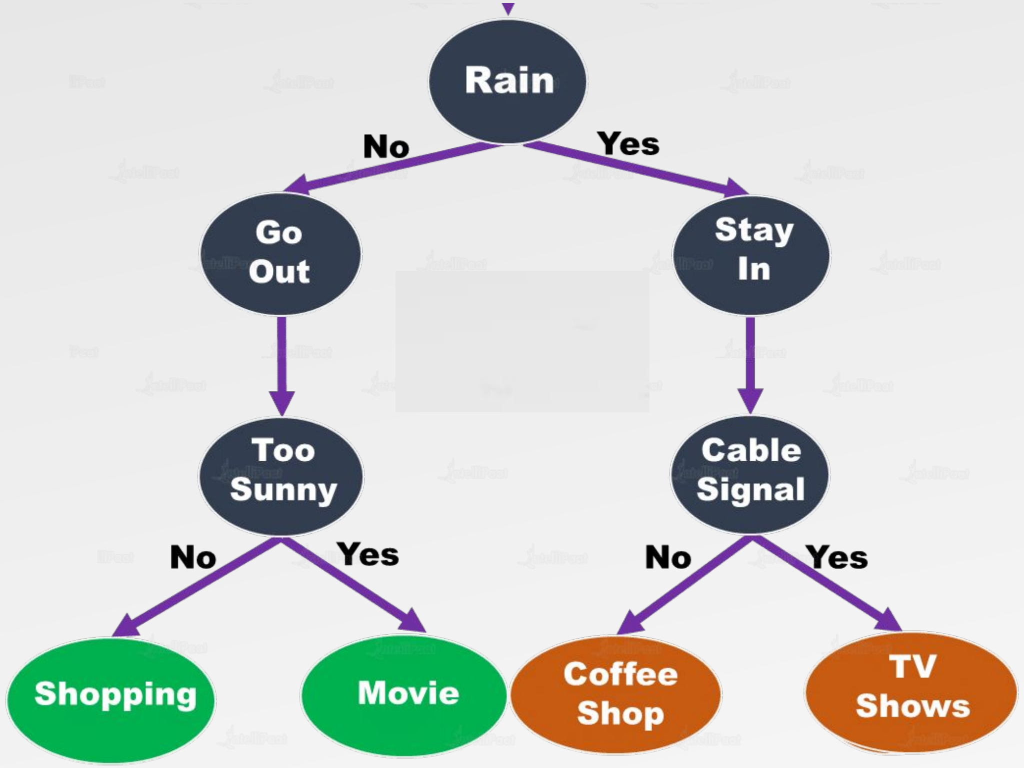
Creates a decision boundary called Hyperplane.

Adversely affected by complexity of data.

C
- regularization parameter.

Kernel -
kernel type
degree
gamma/coefficient

Decision Trees



Root Node

Interior Nodes

Leaf Nodes

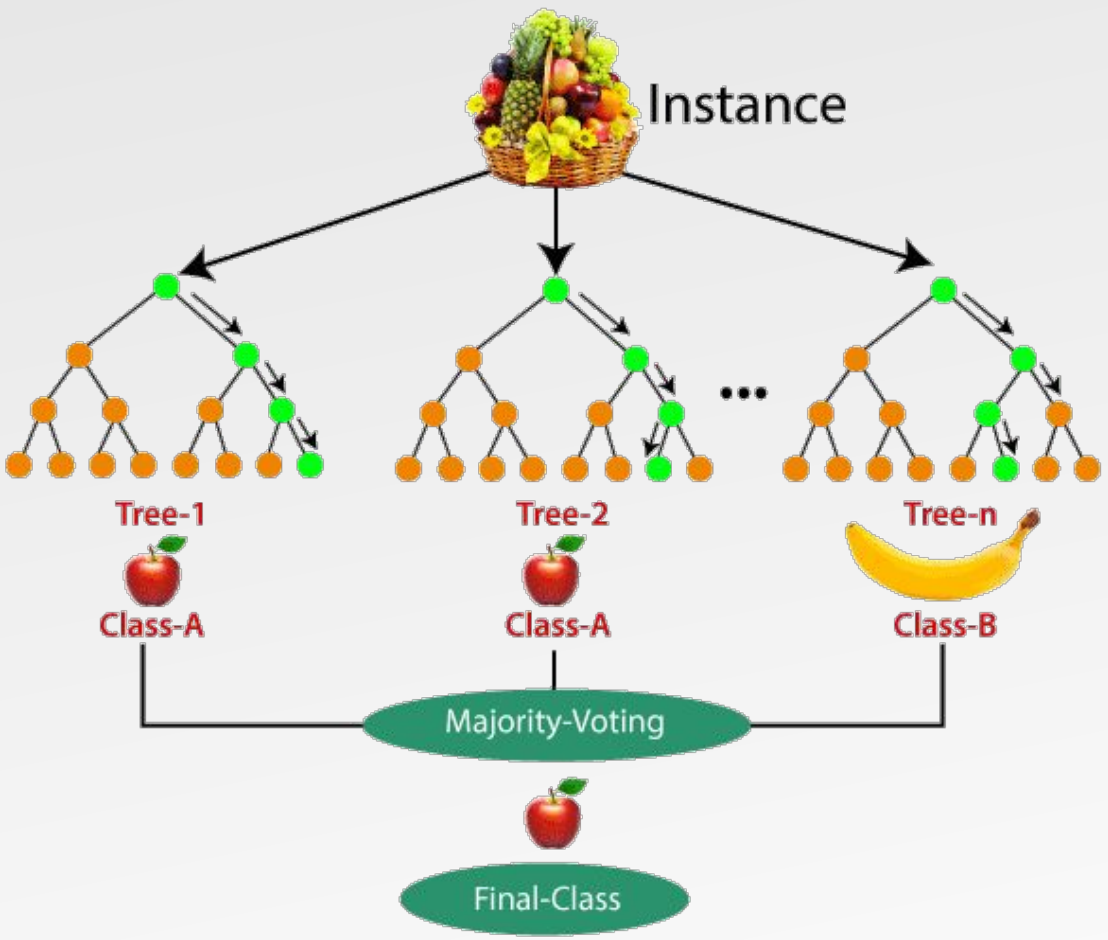
Simple for data exploration

Small Change in data adversely affect the output.

Lot of ambiguity.

Hypothesis Space

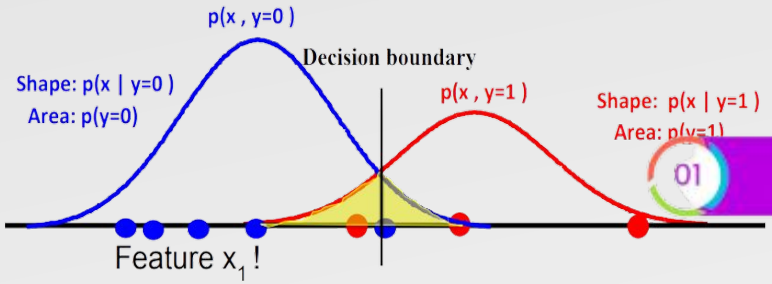
Random Forest

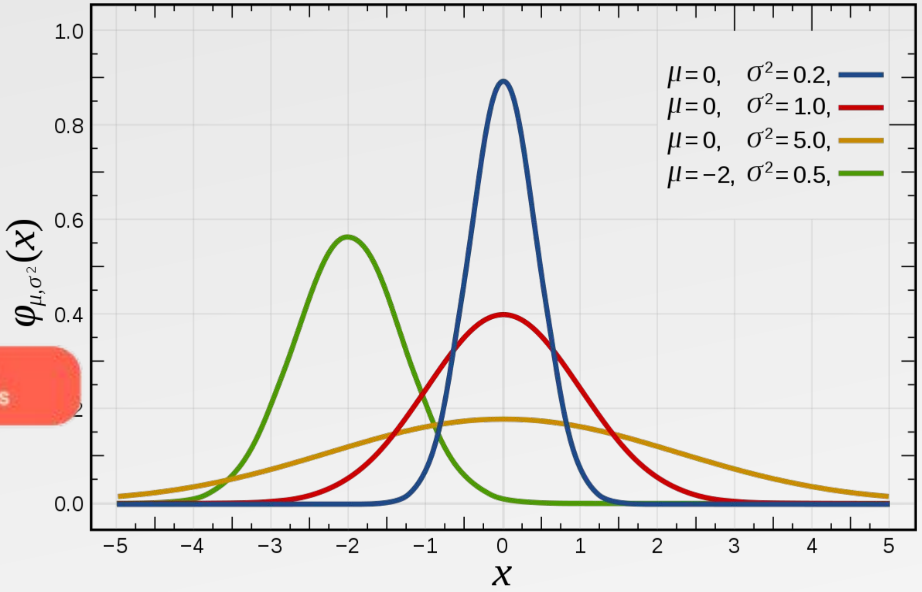


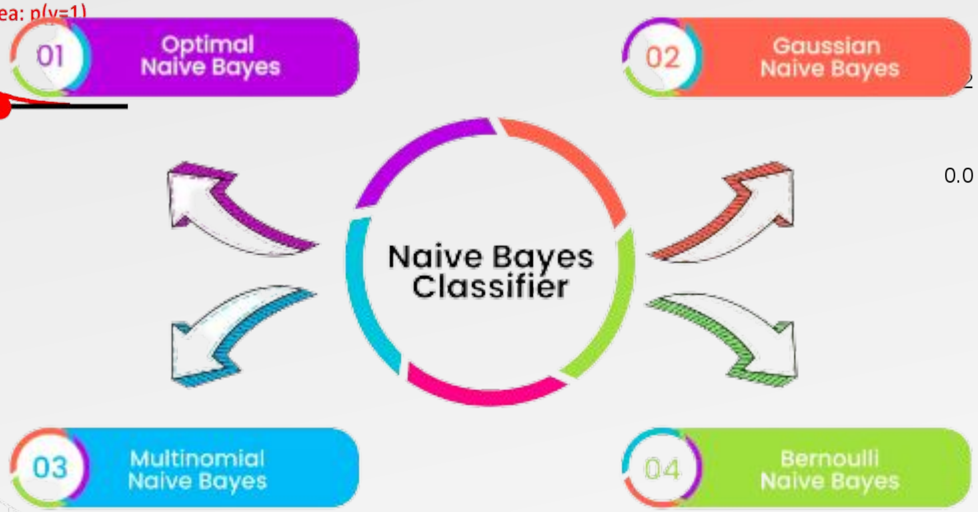
Multi Level Decision Trees
Robust to Outliers
can handle missing values

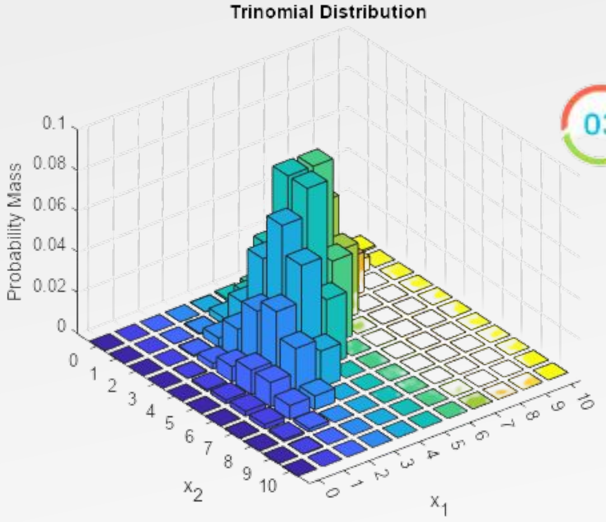
Training Period.
Complexity.

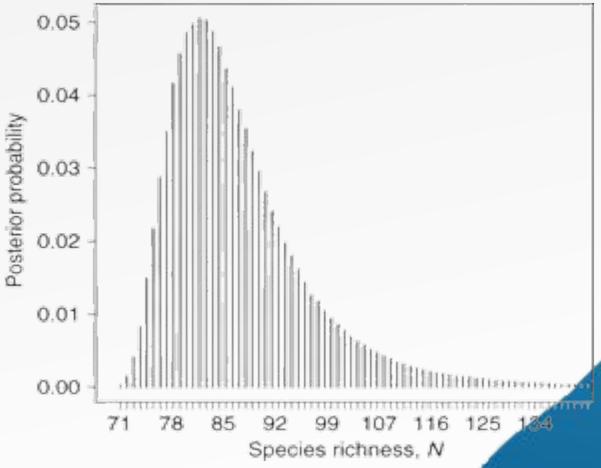
Naive Bayes


p(x , y=0)
Decision boundary
Shape: p(x | y=0)
Area: p(y=0)
p(x , y=1)
Shape: p(x | y=1)
Area: p(y=1)
Feature x_1 !

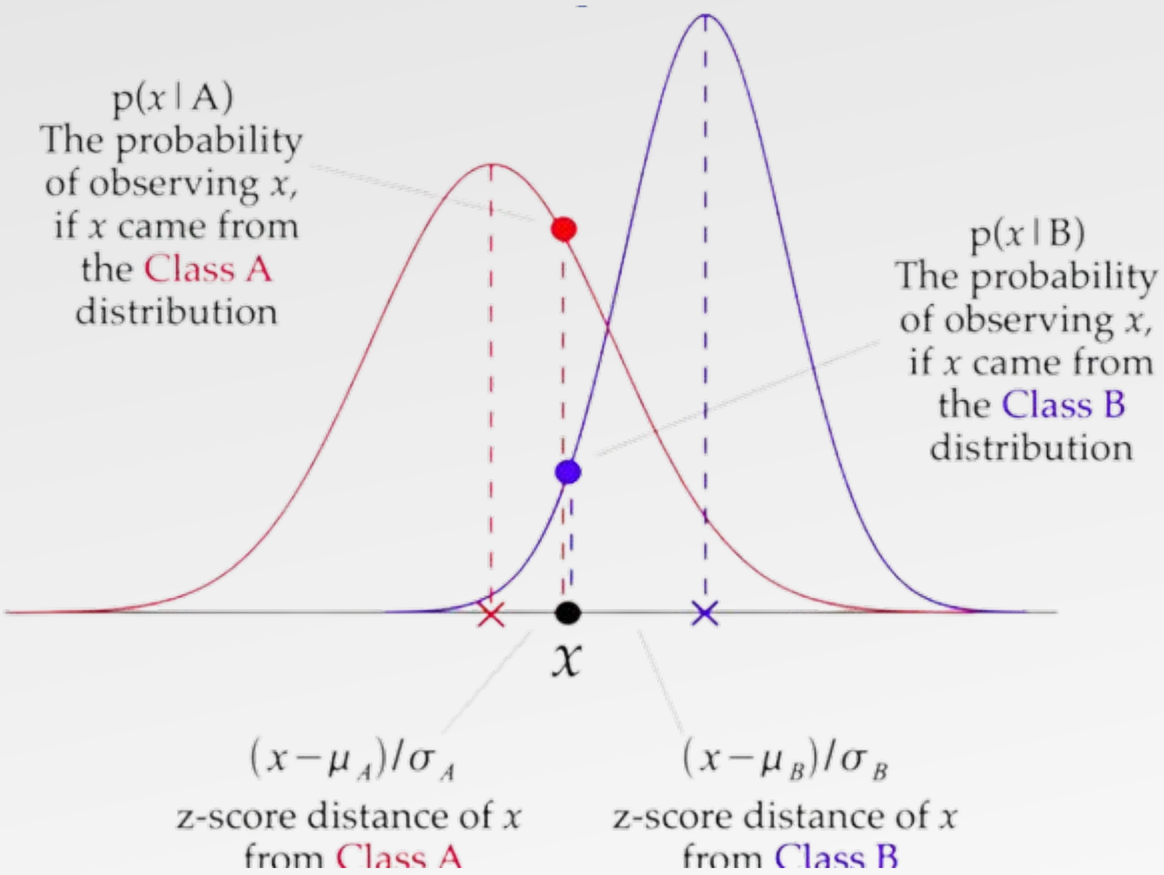

$\mu=0, \quad \sigma^2=0.2,$
$\mu=0, \quad \sigma^2=1.0,$
$\mu=0, \quad \sigma^2=5.0,$
$\mu=-2, \quad \sigma^2=0.5,$
$\varphi_{\mu,\sigma^2}(x)$
x
1.0
0.8
0.6
0.4
0.0
−5 −4 −3 −2 −1 0 1 2 3 4 5


Naive Bayes Classifier
01 Optimal Naive Bayes
02 Gaussian Naive Bayes
03 Multinomial Naive Bayes
04 Bernoulli Naive Bayes


Trinomial Distribution
Probability Mass
0.1
0.08
0.06
0.04
0.02
0
x_2
x_1
0 1 2 3 4 5 6 7 8 9 10


Posterior probability
0.05
0.04
0.03
0.02
0.01
0.00
71 78 85 92 99 107 116 125 134
Species richness, N

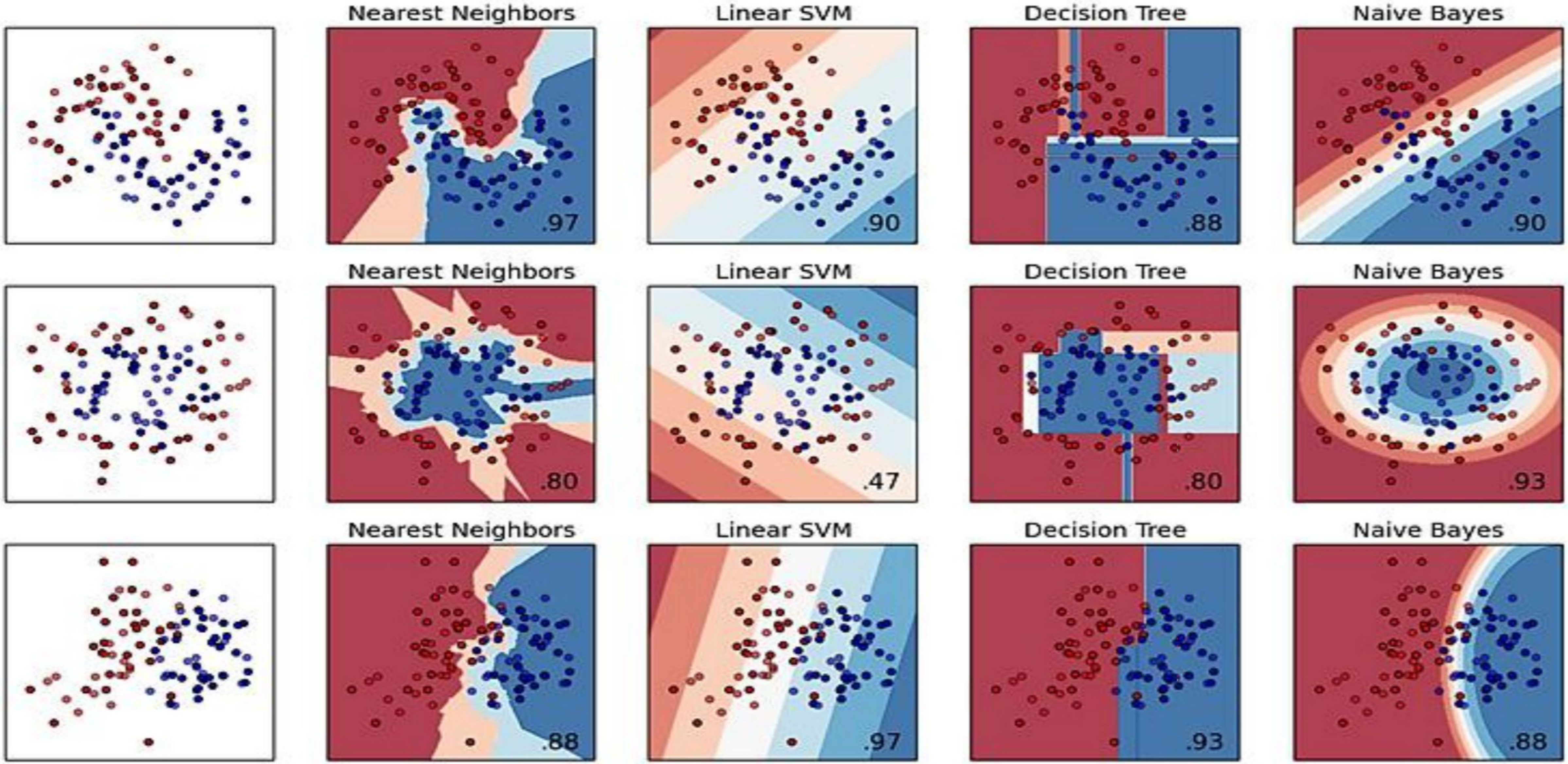
Naive Bayes



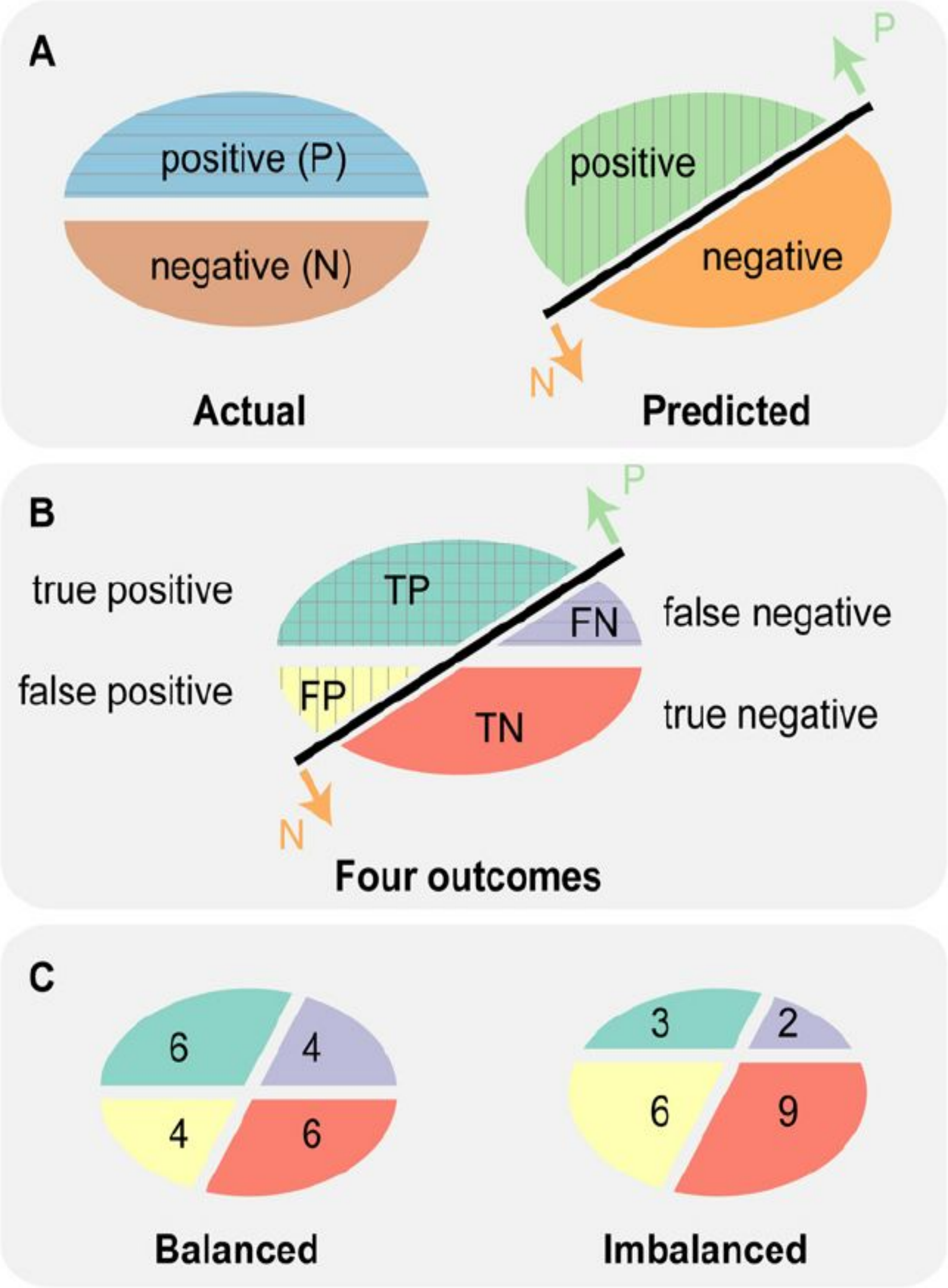
p(x | A)
The probability of observing x, if x came from the Class A distribution
p(x | B)
The probability of observing x, if x came from the Class B distribution
x
$(x-\mu_A)/\sigma_A$
z-score distance of x from Class A
$(x-\mu_B)/\sigma_B$
z-score distance of x from Class B

Different type of algorithms

MACHINE LEARNING CLASSIFICATION TASK



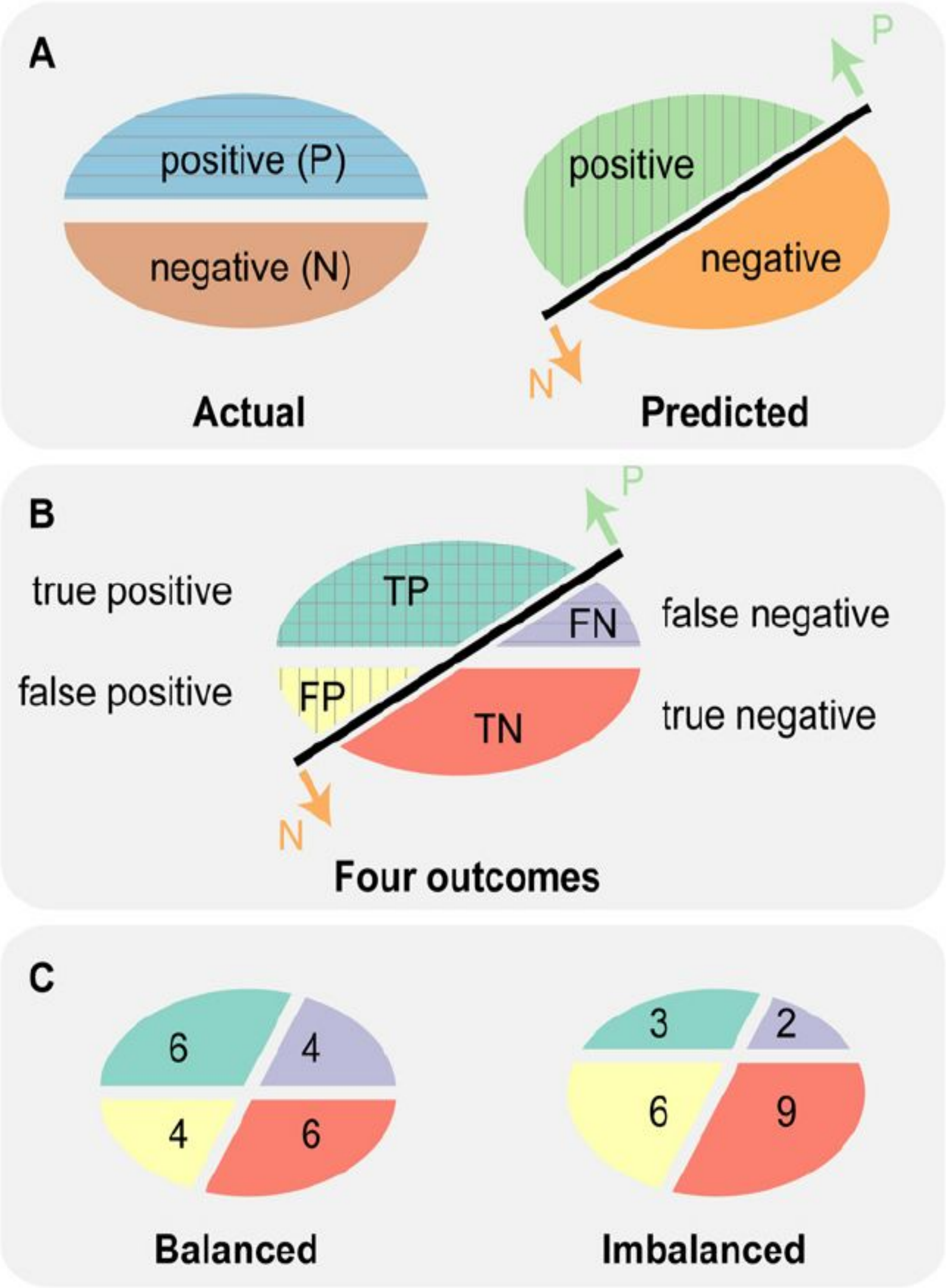
Confusion Matrix



Measure	Formula
ACC	(TP + TN) / (TP + TN + FN + FP)
ERR	(FP + FN) / (TP + TN + FN + FP)
SN, TPR, REC	TP / (TP + FN)
SP	TN / (TN + FP)
FPR	FP / (TN + FP)
PREC, PPV	TP / (TP + FP)
MCC	$(TP * TN — FP * FN) / ((TP + FP)(TP + FN)(TN + FP)(TN + FN))^{1/2}$
$F_{0.5}$	1.5 * PREC * REC / (0.25 * PREC + REC)
F_1	2 * PREC * REC / (PREC + REC)
F_2	5 * PREC * REC / (4 * PREC + REC)

ACC: accuracy; ERR: error rate; SN: sensitivity; TPR: true positive rate; REC: recall; SP: specificity; FPR: false positive rate; PREC: precision; PPV: positive predictive value; MCC: Matthews correlation coefficient; F: F score; TP: true positives; TN: true negatives; FP: false positives; FN: false negatives

The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets



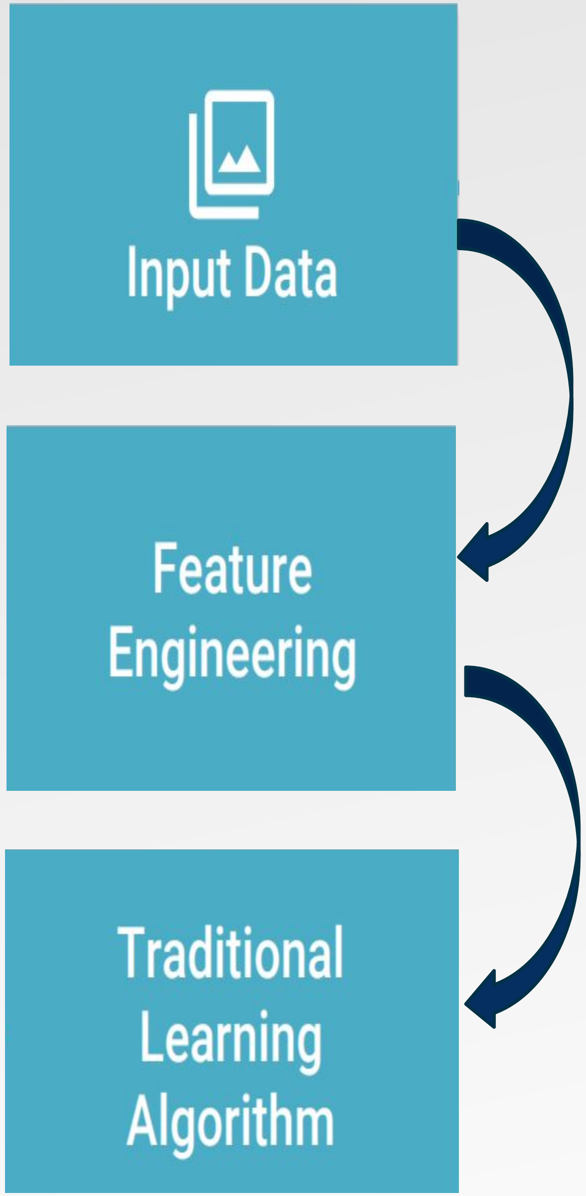
Confusion Matrix

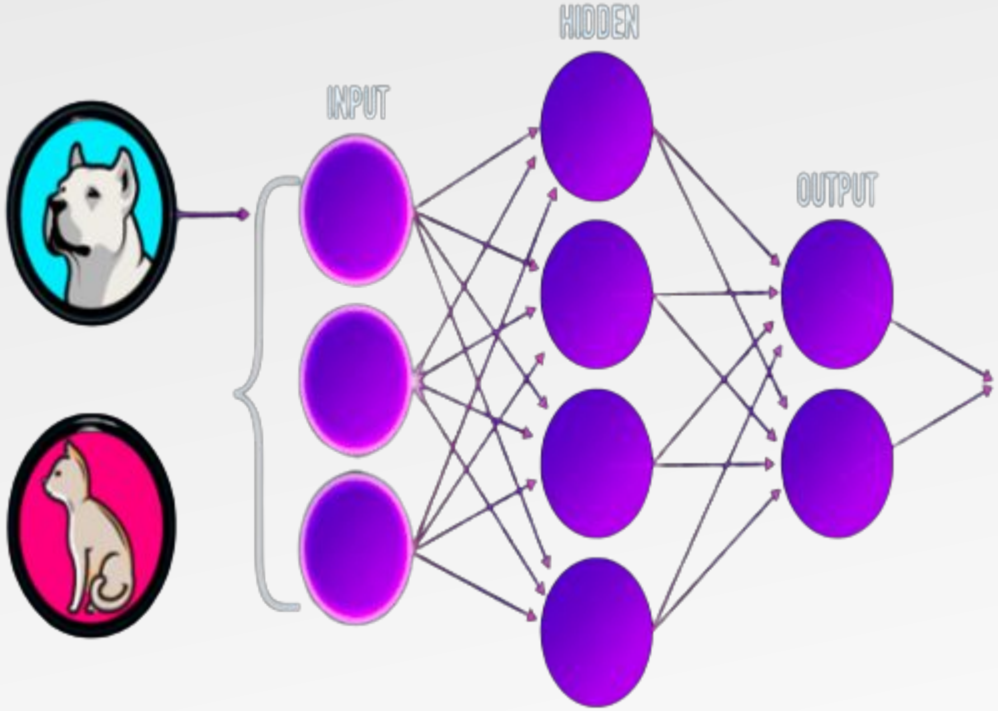
- Precision and recall are two evaluation metrics used to measure the performance of a classifier in binary and multiclass classification problems.
- Precision measures the accuracy of positive predictions, while recall measures the completeness of positive predictions.
- High precision and high recall are desirable, but there may be a trade-off between the two metrics in some cases.
- Precision and recall should be used together with other evaluation metrics, such as accuracy and F1-score, to get a comprehensive understanding of the performance of a classifier.

The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets

Questions Time

Neural Network

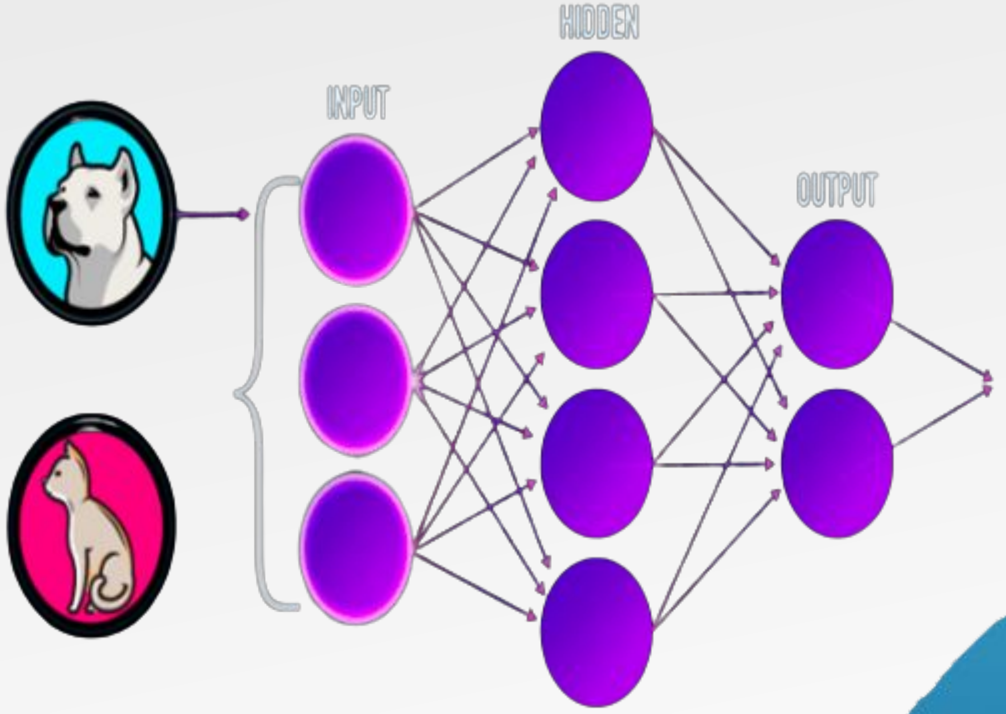




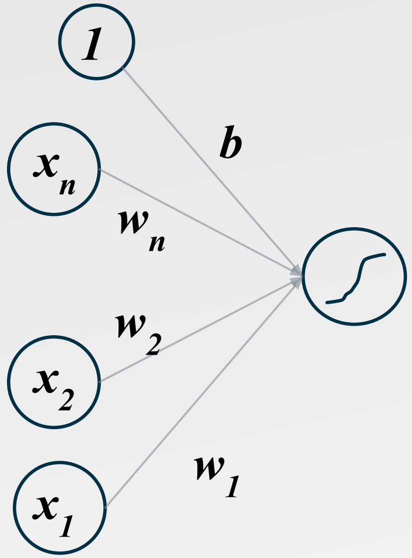
CONNECTING THE DOTS … !

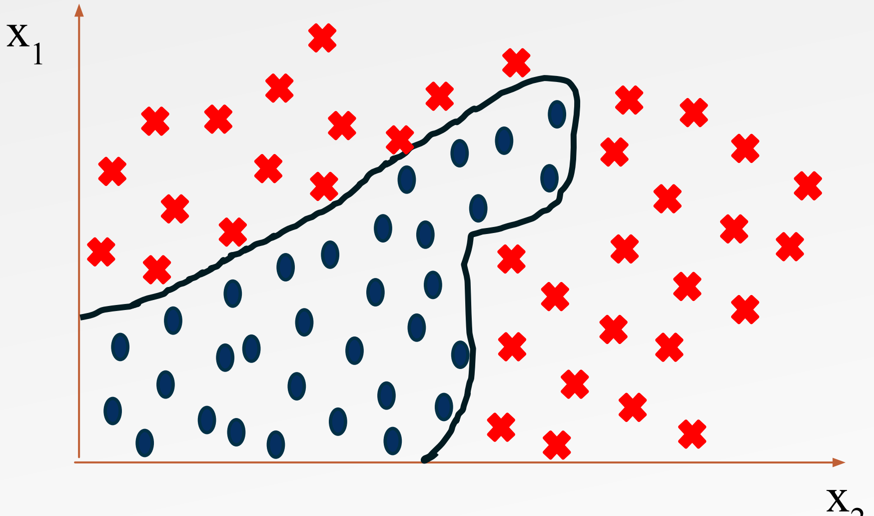
Build a Neural Network

- Artificial Neurons
  - Activation Functions
  - Synaptic Weights
  - Bias
- Layers
- Data Normalization
- Gradient Descent, Learning Rate and Loss Function
- Epoch, Cross Validation



CONNECTING THE DOTS … !





$$y = mx + c$$

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

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

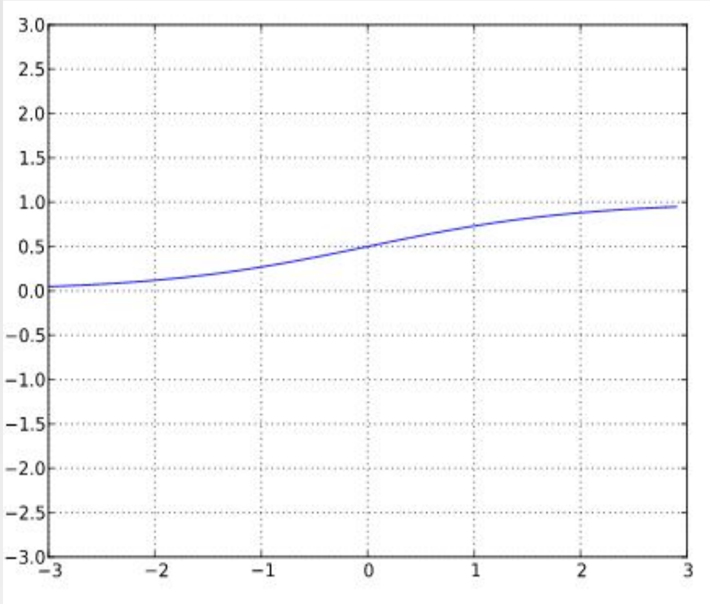
$$h(x) = g(a(x)) = g\left(b + \sum_i w_i x_i\right)$$

w – connected weights

b – neuron bias

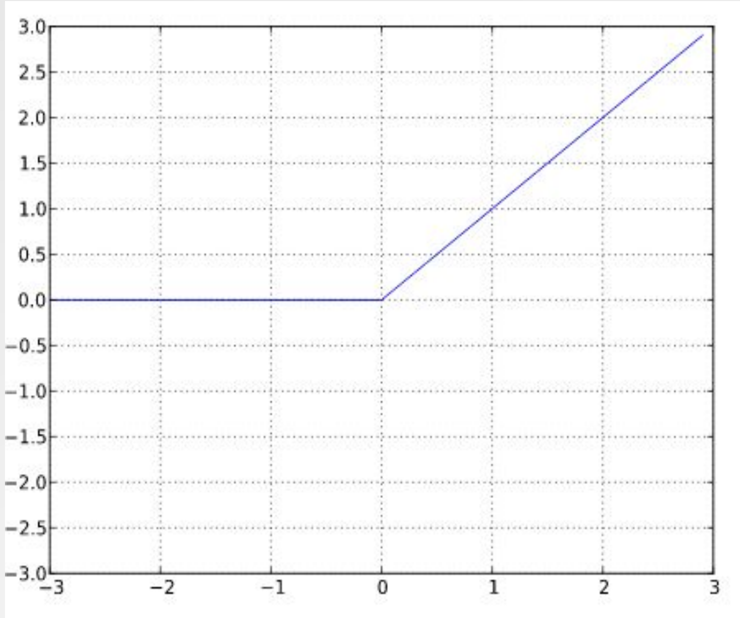
g(.) – activation function

CONNECTING THE DOTS … !



Sigmoid Activation

Squashes the neuron's output between 0 and 1

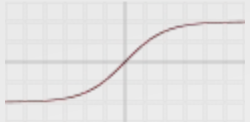


ReLU Activation

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

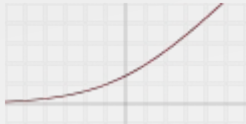
Other popular activation functions

✔ Hyperbolic tangent (TanH)



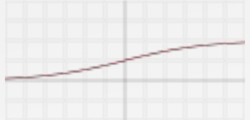
$$f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$$

✔ SoftPlus



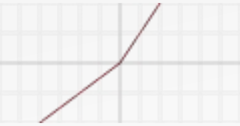
$$f(x) = \ln(1+e^x)$$

✔ Logistic



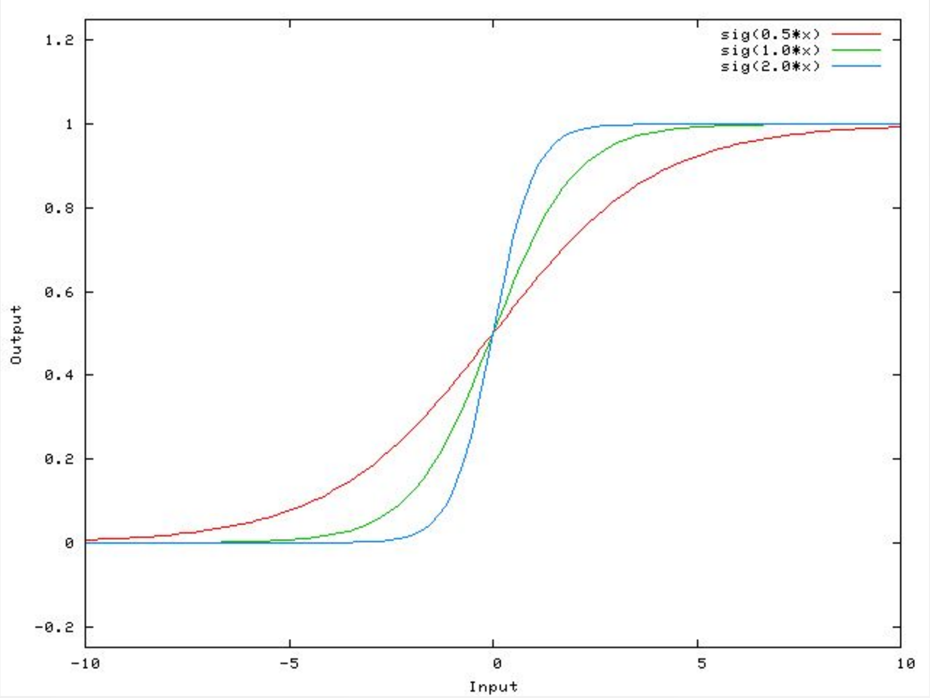
$$f(x) = \frac{1}{1+e^{-x}}$$

✔ Leaky ReLU



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

CONNECTING THE DOTS … !



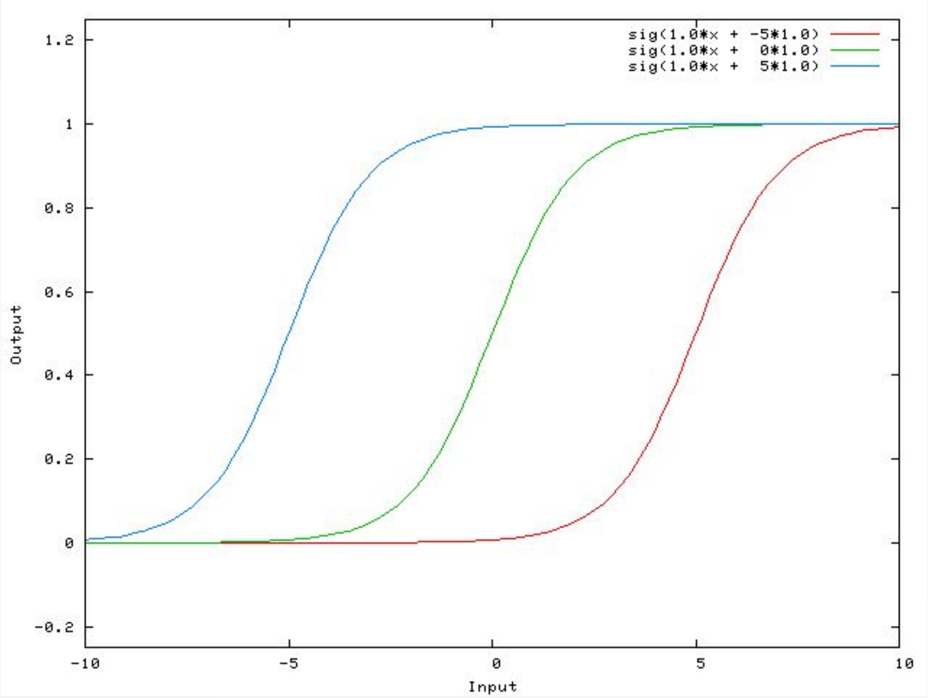


Figure: effect of weights and bias on Sigmoid activation

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

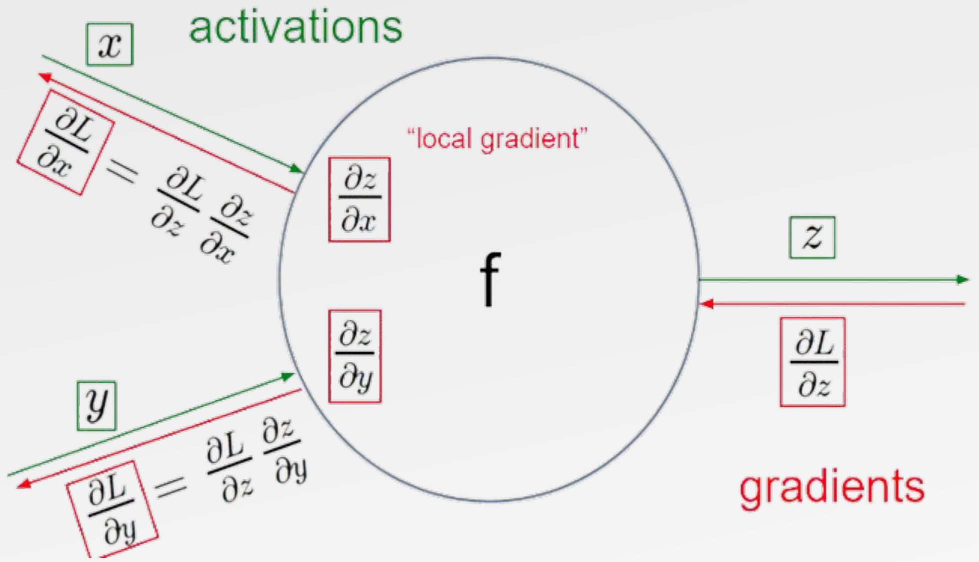
Epoch, Cross Validation

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

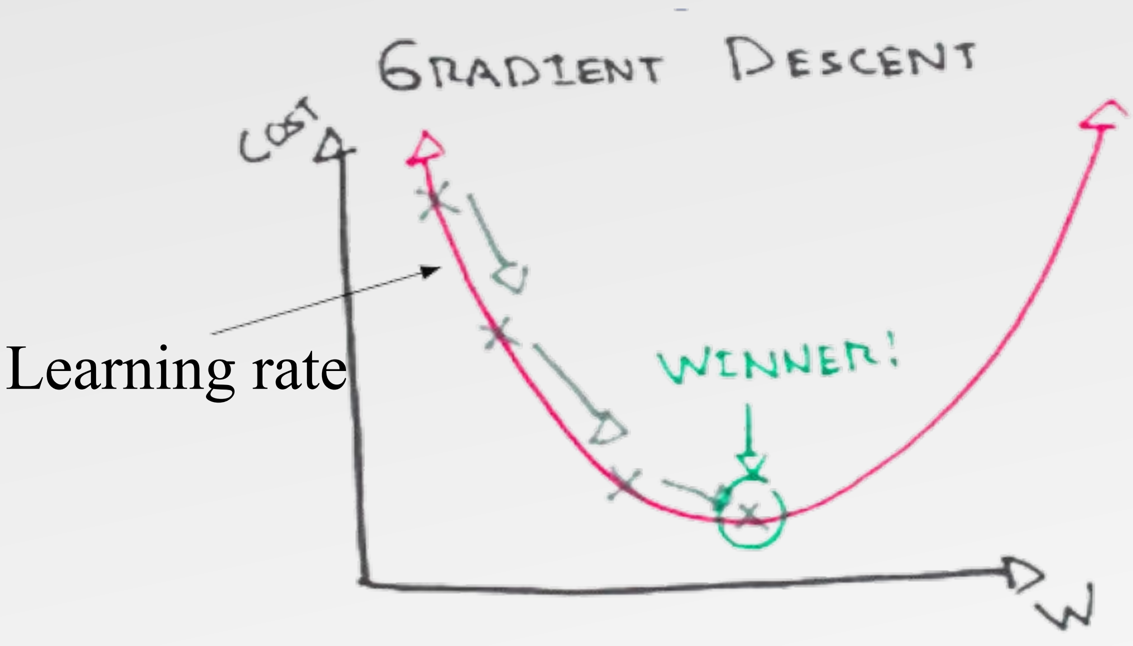
- Activation Functions
- Synaptic Weights
- Bias

Layers

Data Normalization

Gradient Descent, Learning Rate and Loss Function

Epoch, Cross Validation

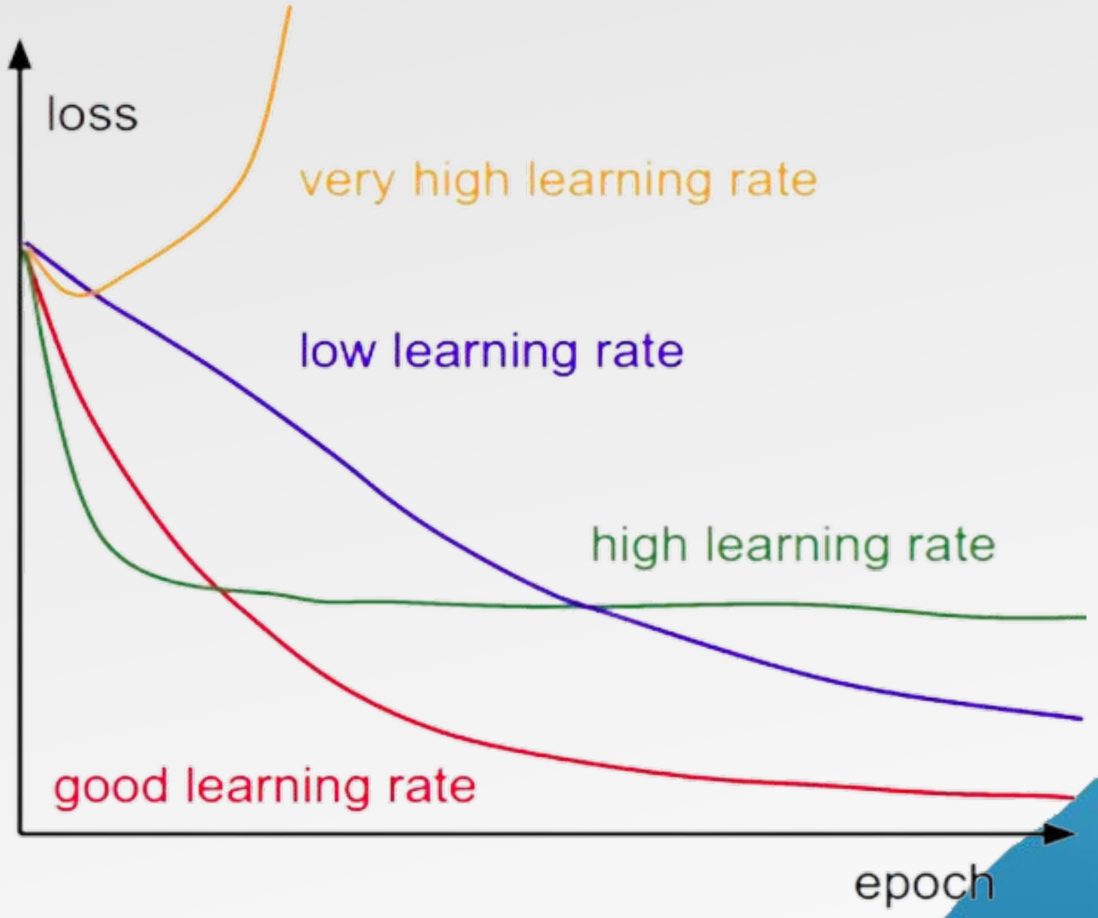


A Loss function tells us “how good” our model is at making predictions for a given set of parameters

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

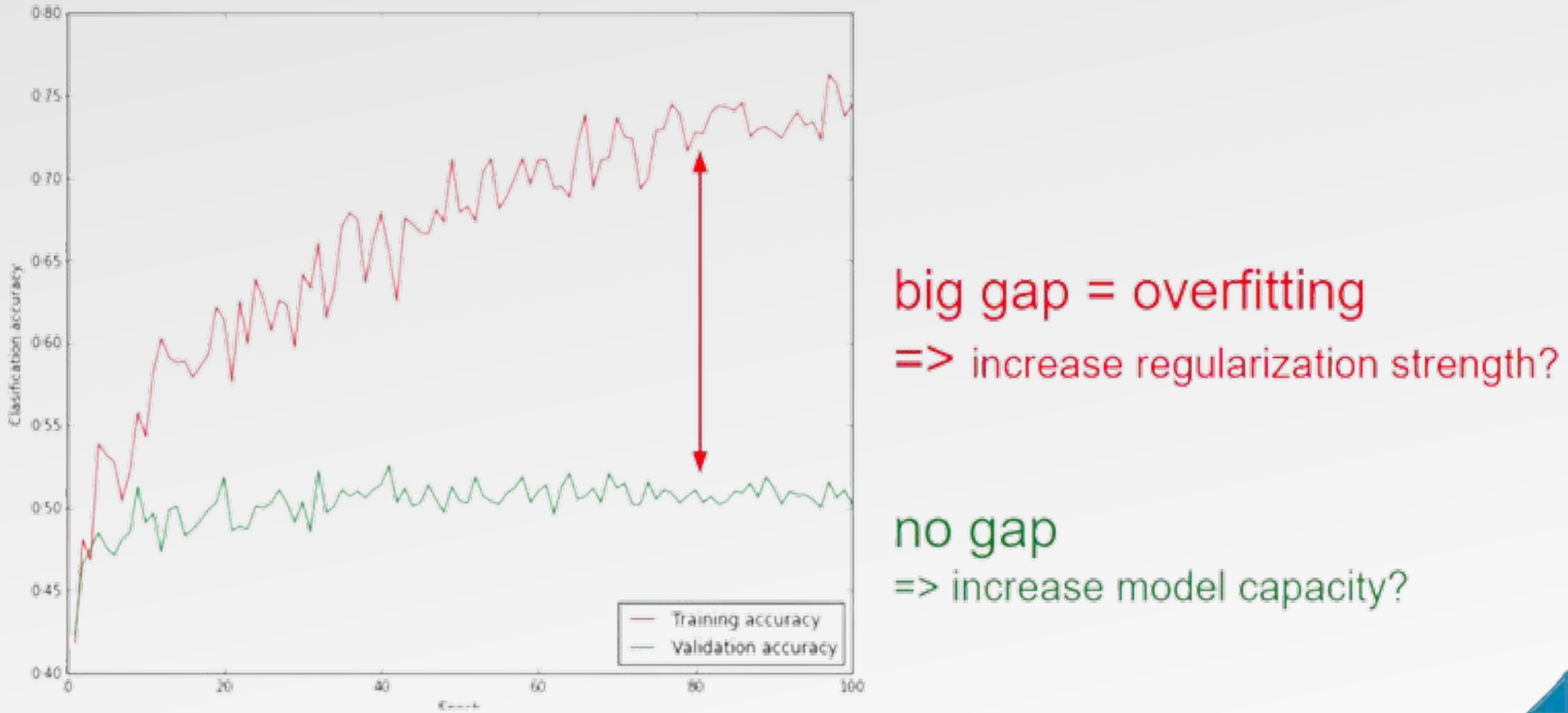
Gradient Descent, Learning Rate and Loss Function

Epoch, Cross Validation

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

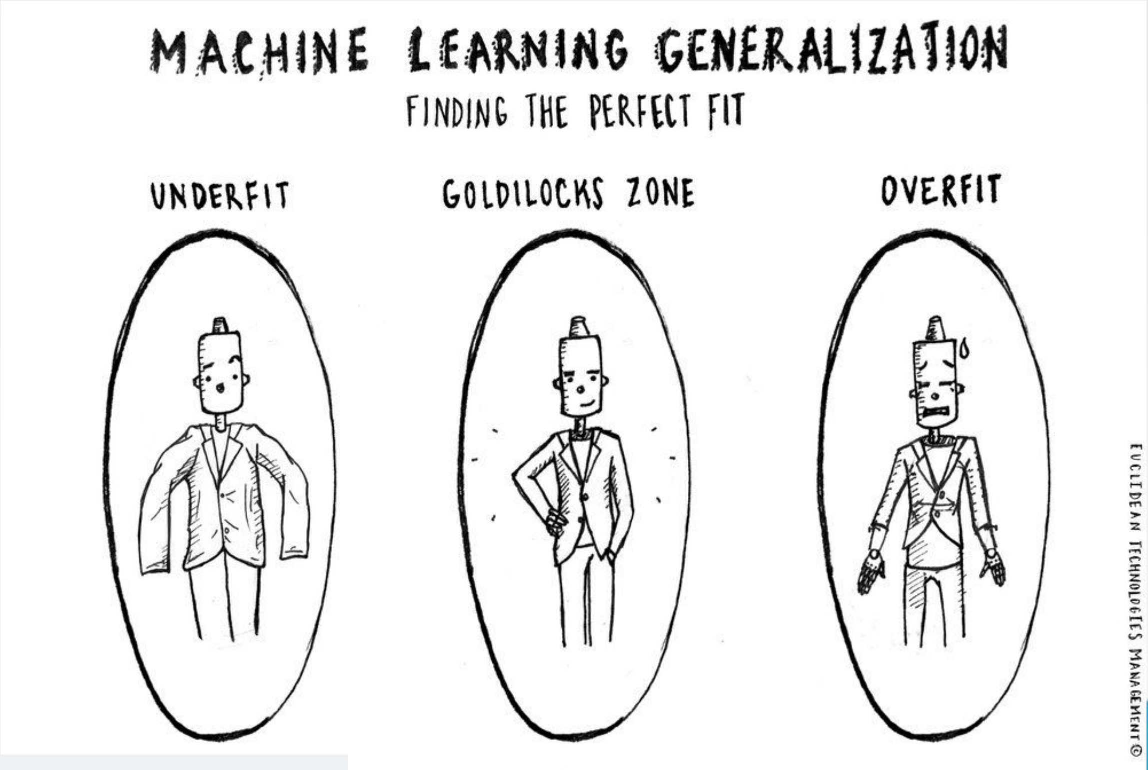
Cross validate the accuracy in order to get the accurate results.

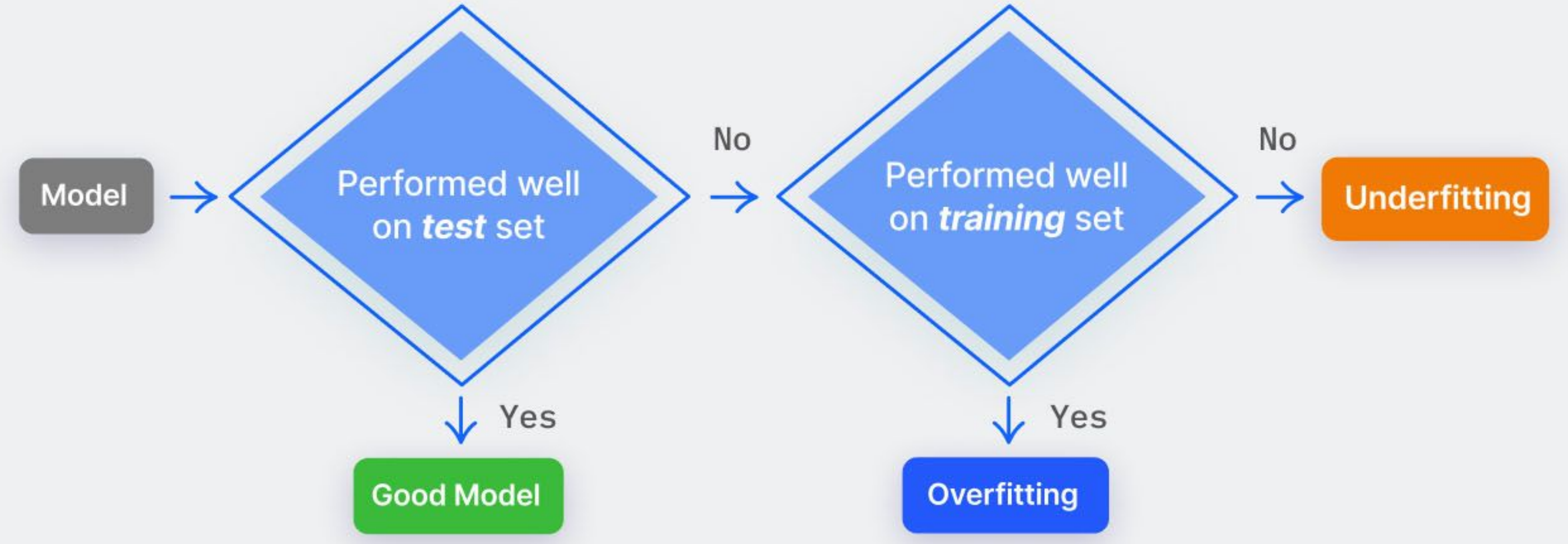
CONNECTING THE DOTS … !



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

CONNECTING THE DOTS … !





CONNECTING THE DOTS … !

Underfitting happens when:

1. Unclean training data containing noise or outliers can be a reason for the model not being able to derive patterns from the dataset.
2. The model has a high bias due to the inability to capture the relationship between the input examples and the target values.
3. The model is assumed to be too simple. For example, training a linear model in complex scenarios.

Overfitting happens when:

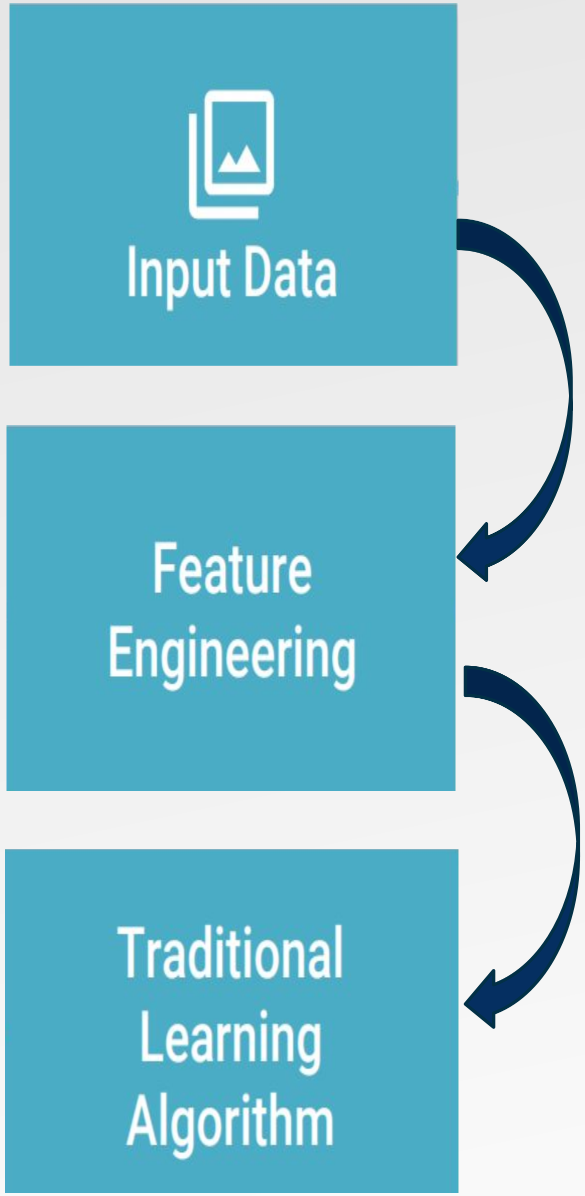
1. The data used for training is not cleaned and contains garbage values. The model captures the noise in the training data and fails to generalize the model's learning.
2. The model has a high variance.
3. The training data size is not enough, and the model trains on the limited training data for several epochs.
4. The architecture of the model has several neural layers stacked together. Deep neural networks are complex and require a significant amount of time to train, and often lead to overfitting the training set.

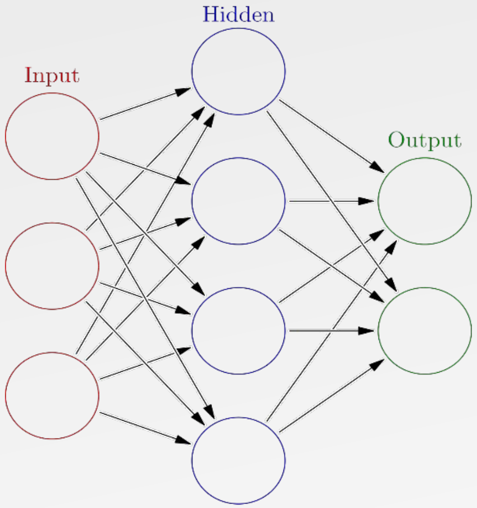
CONNECTING THE DOTS … !

Avoid Overfitting

- Train with more data
- Data augmentation
- Addition of noise to the input data
- Feature selection
- Cross-validation
- Simplify data
- Regularization
- Ensembling

CONNECTING THE DOTS … !

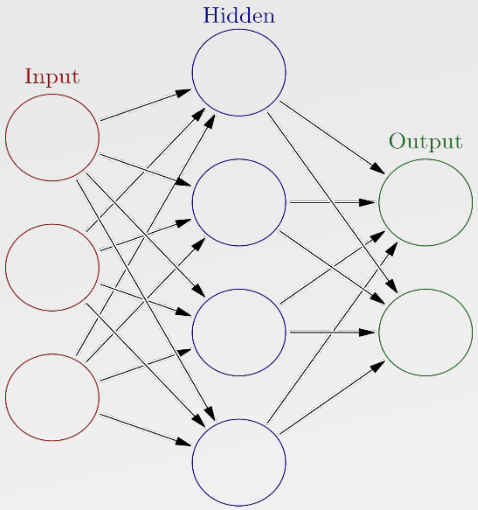




[3 2]

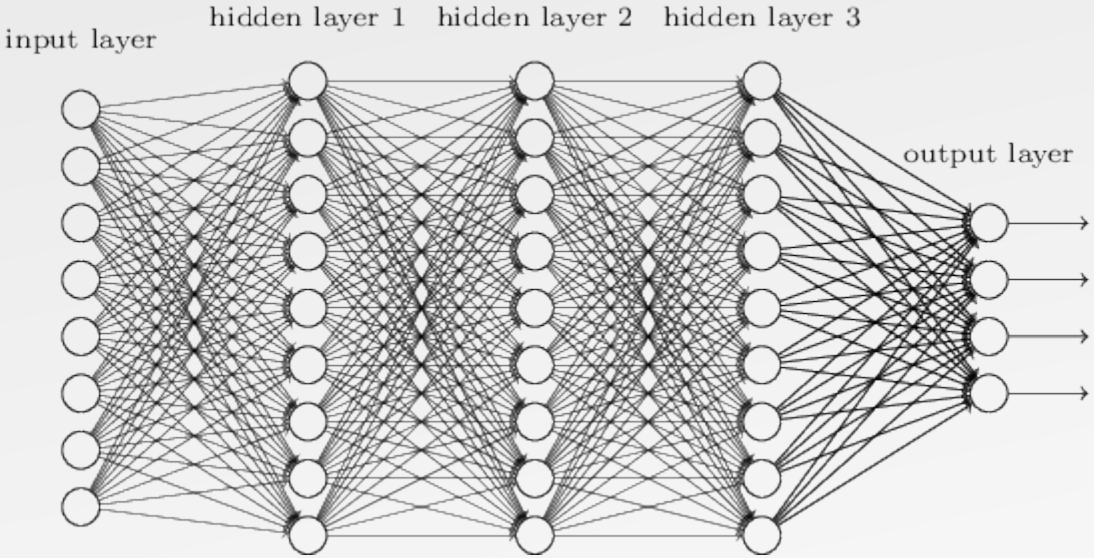
Artificial Neural Network

CONNECTING THE DOTS … !



[3 2]

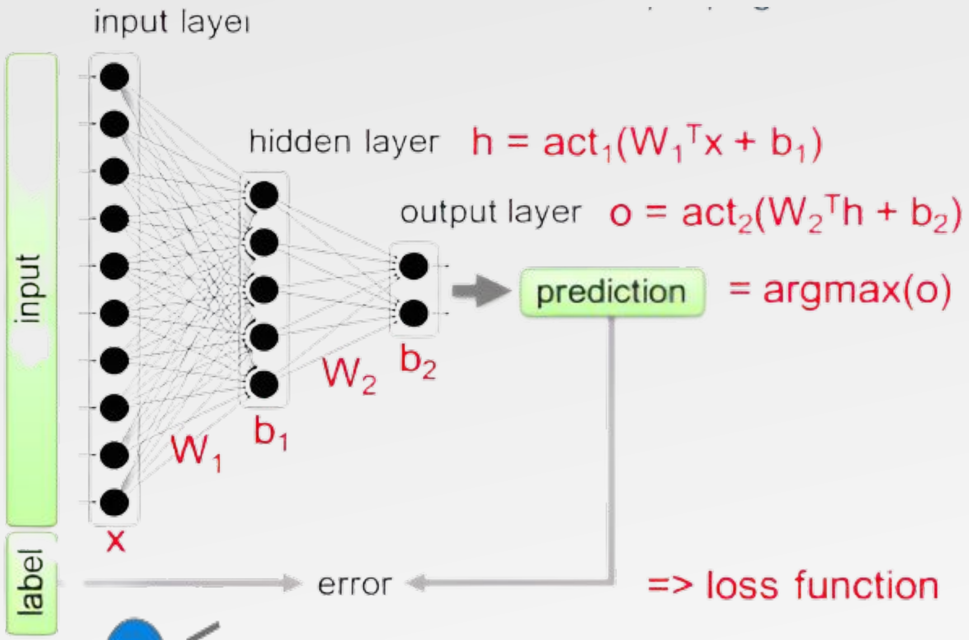
Artificial Neural Network

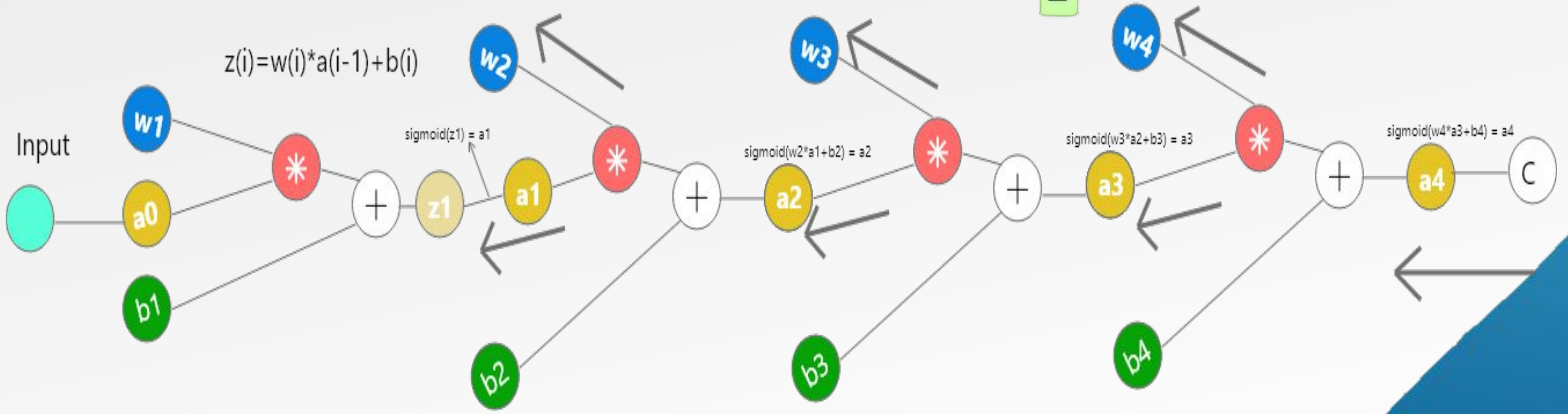


[9 9]   [9 9]   [9 4]

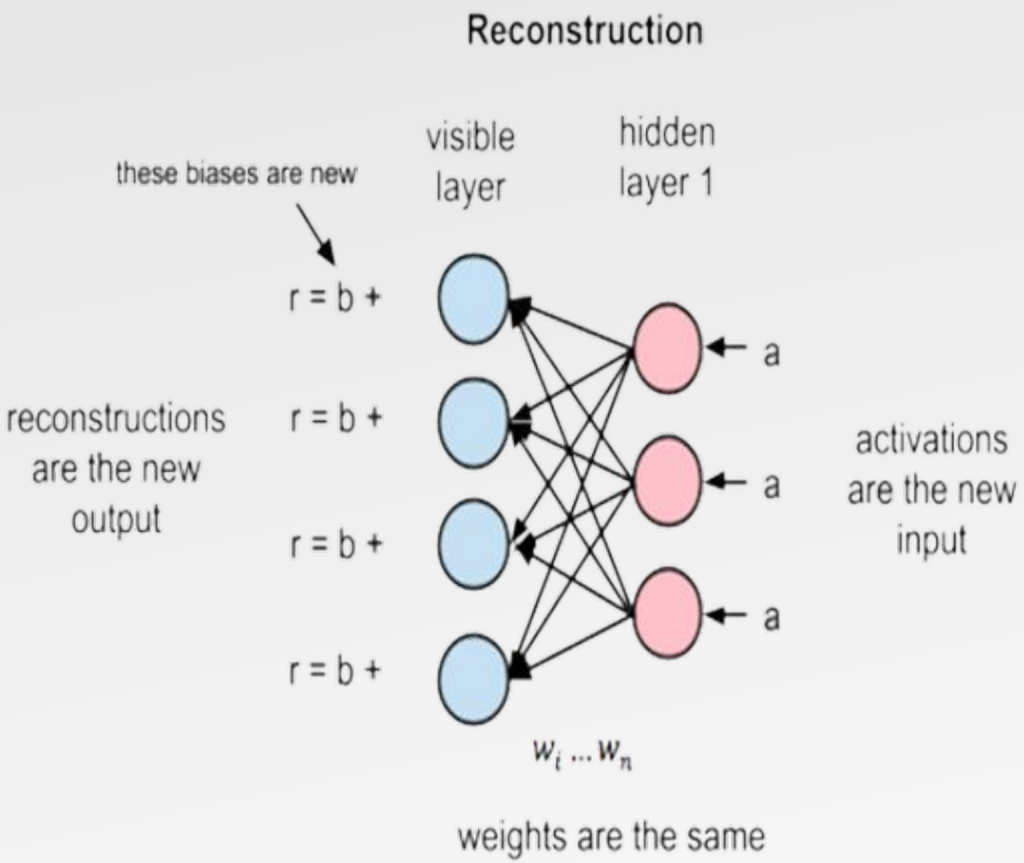
Deep Neural Network

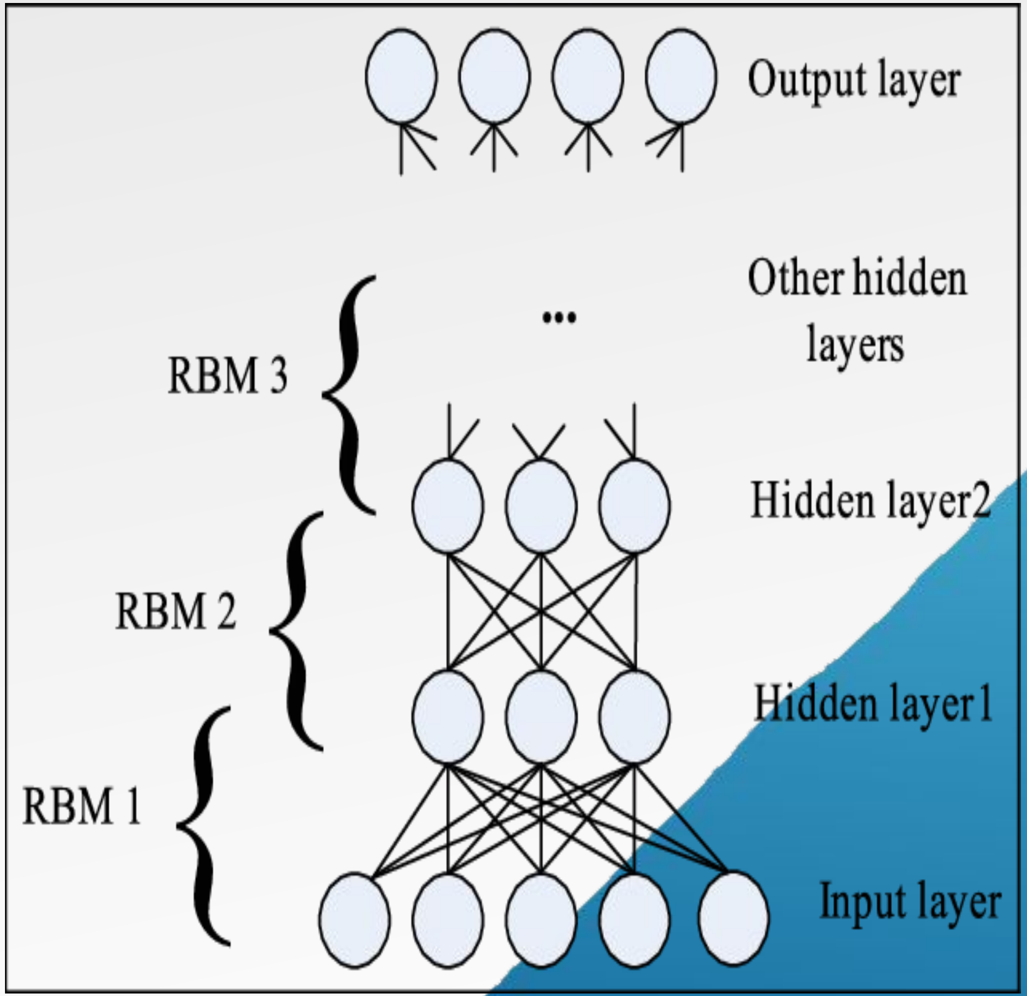
WHAT MADE VANISHING GRADIENT PROBLEM VANISH ?



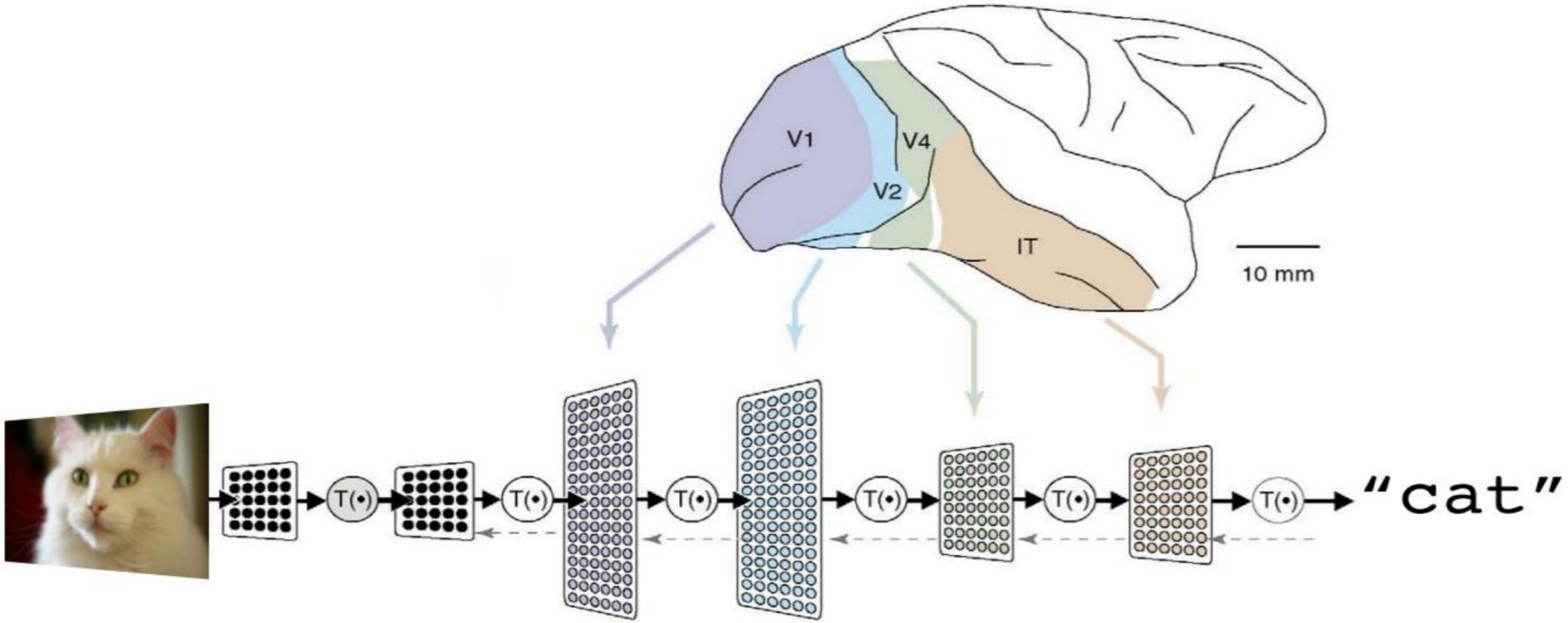


WHAT MADE VANISHING GRADIENT PROBLEM VANISH ?



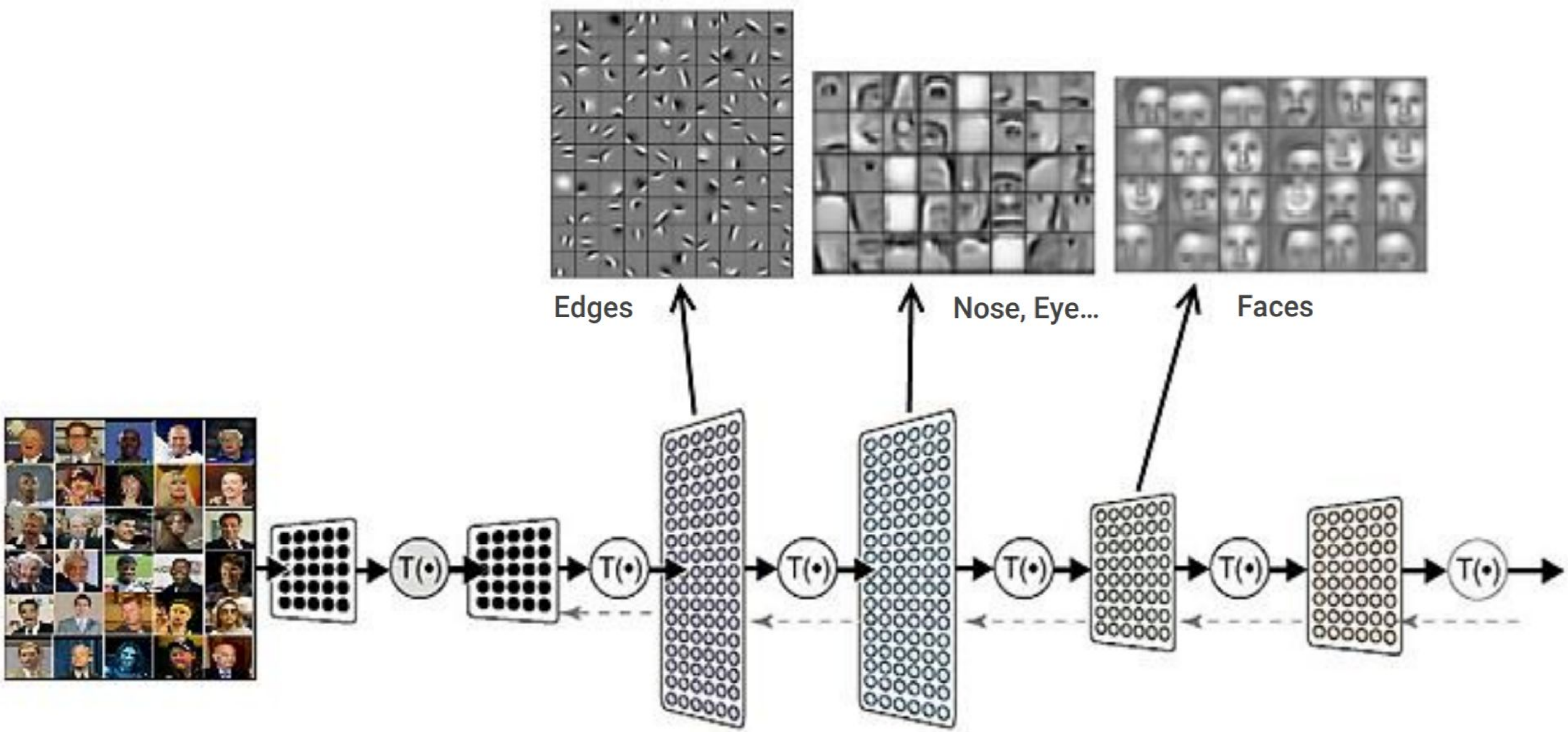


DEEP LEARNING ARCHITECTURE

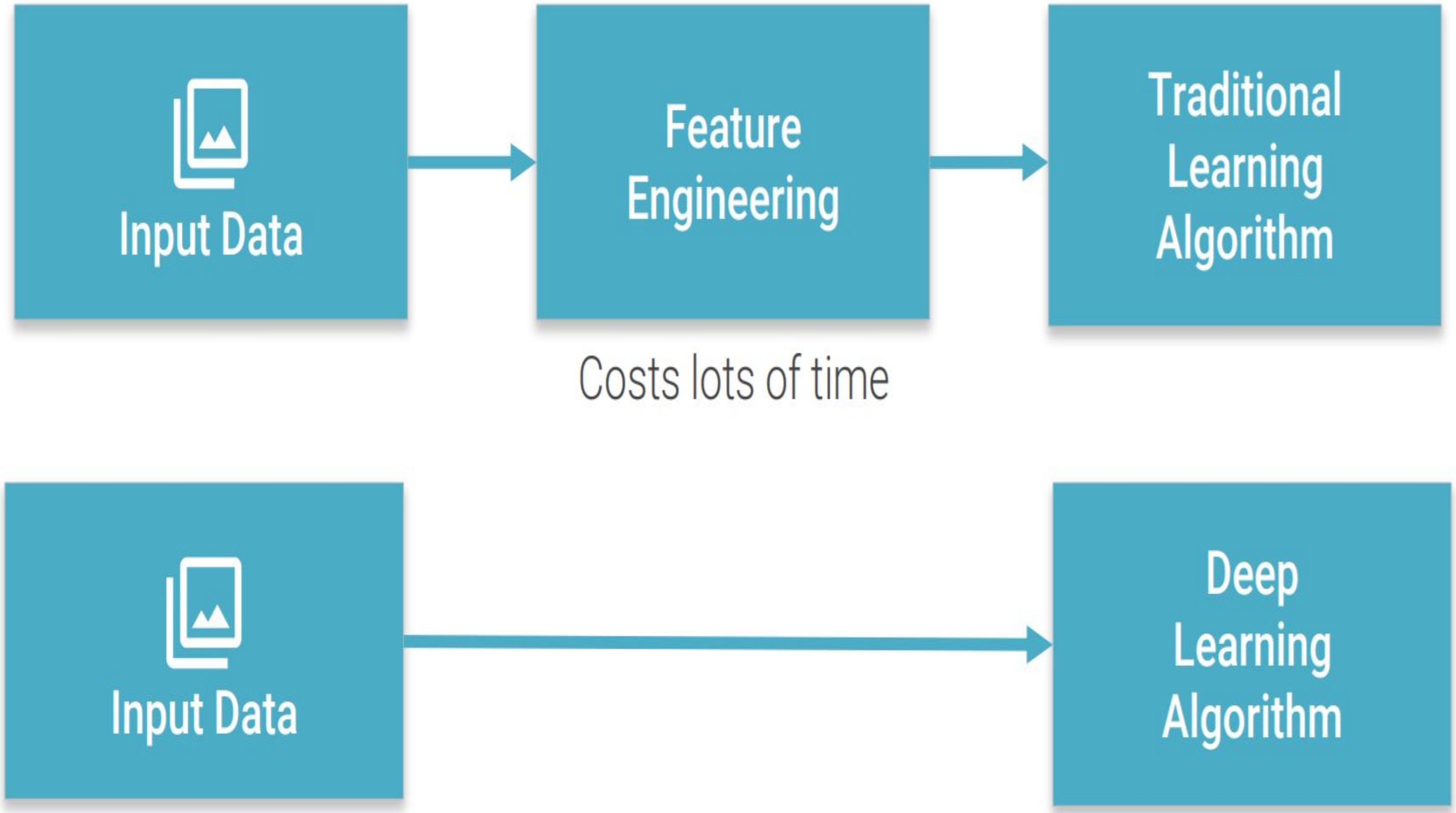


- A deep neural network consists of a hierarchy of layers, whereby each layer transforms the input data into more abstract representations (e.g. edge -> nose -> face). The output layer combines those features to make predictions.

DEEP LEARNING BASICS

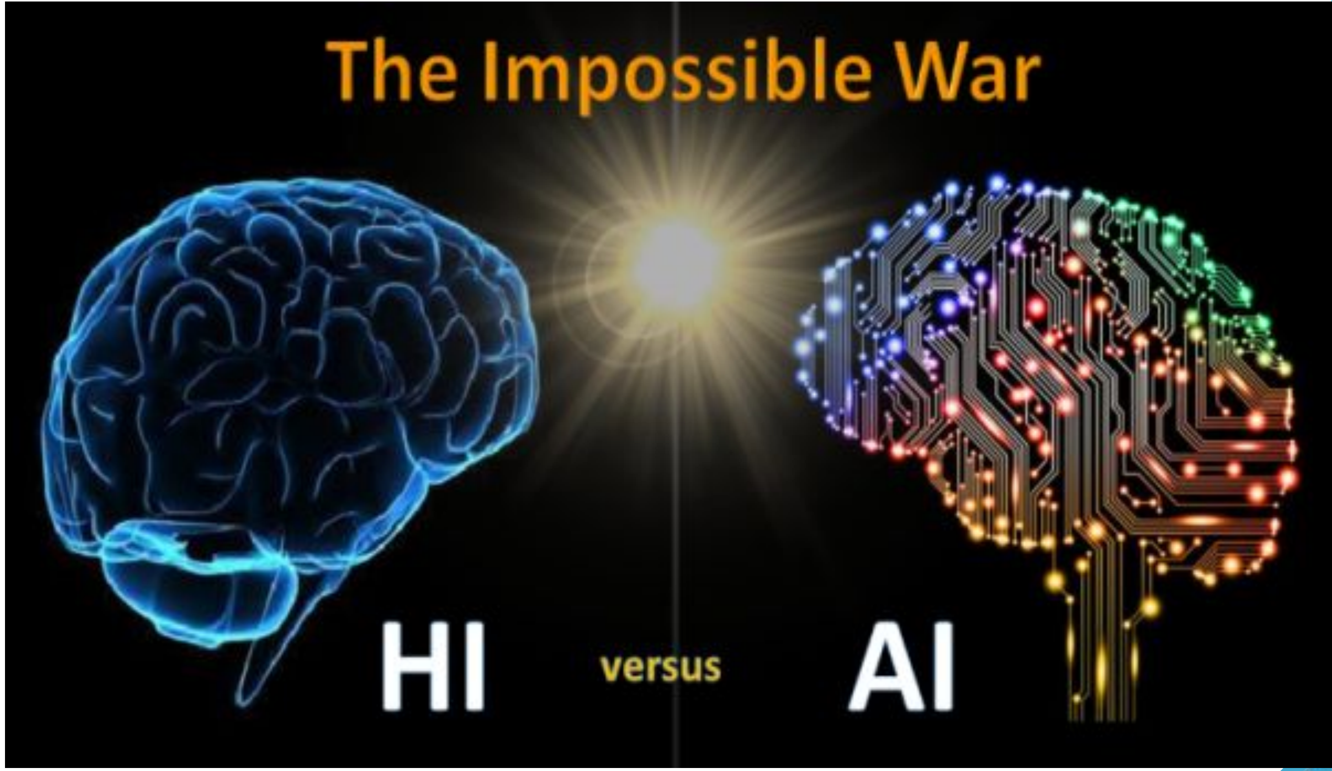


DIFFERENCES BETWEEN TRADITIONAL NETWORKS AND DEEP NETWORKS



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

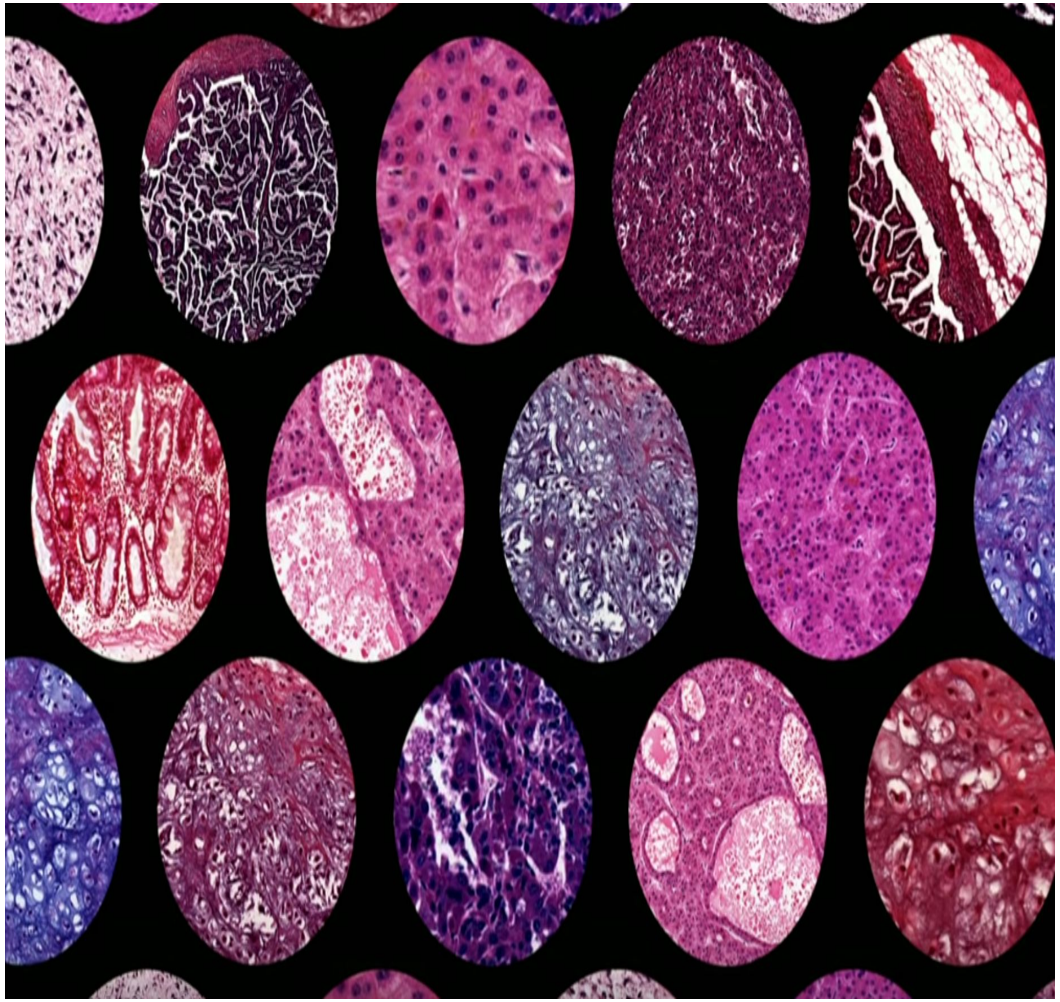


Human Intelligence is **General.**
Machine Intelligence is **Limited.**

CHALLENGES OF AI

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

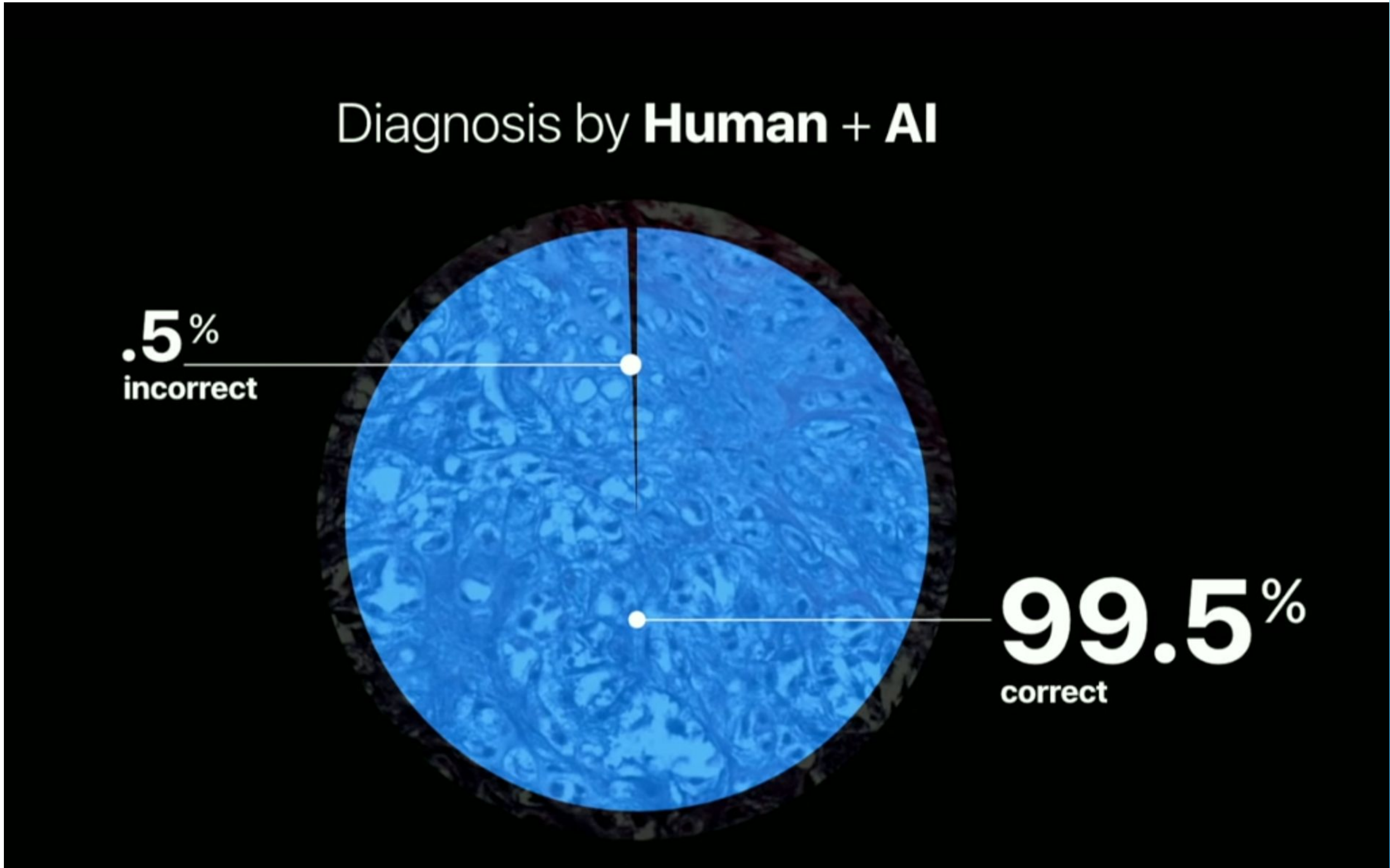
Diagnosing Cancer



HUMANISTIC AI

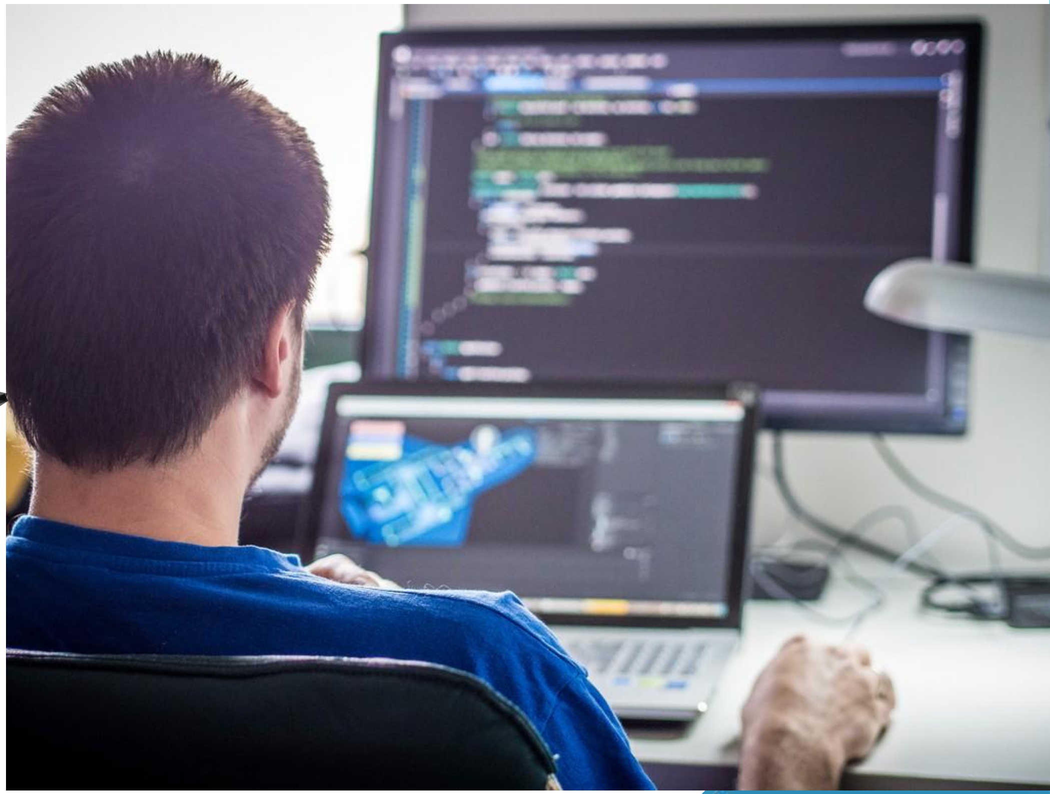
Diagnosing Cancer

- The human-machine team delivers **Super-Human performance**



Design

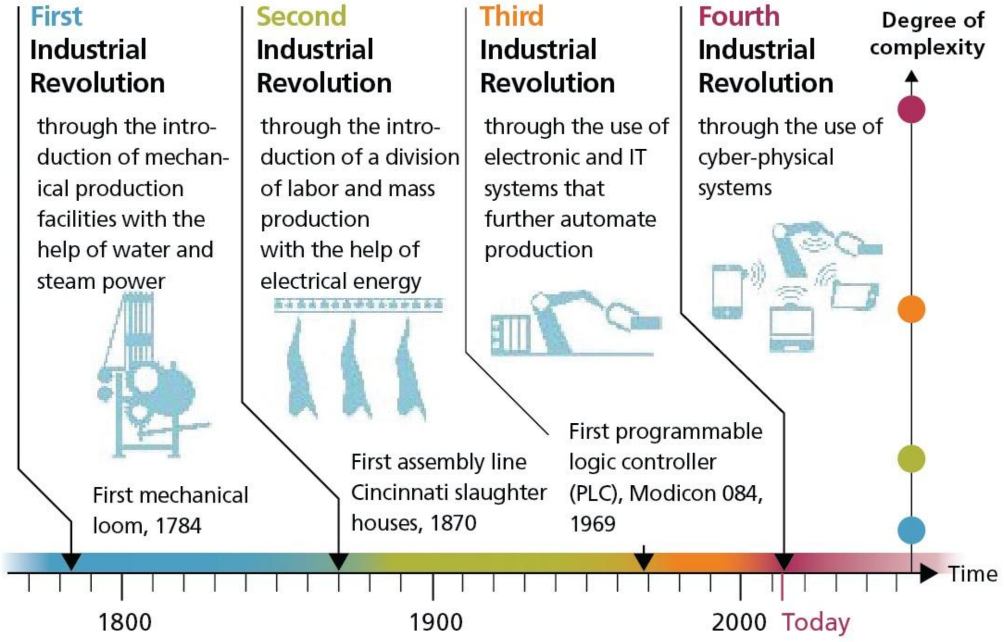
HUMANISTIC AI



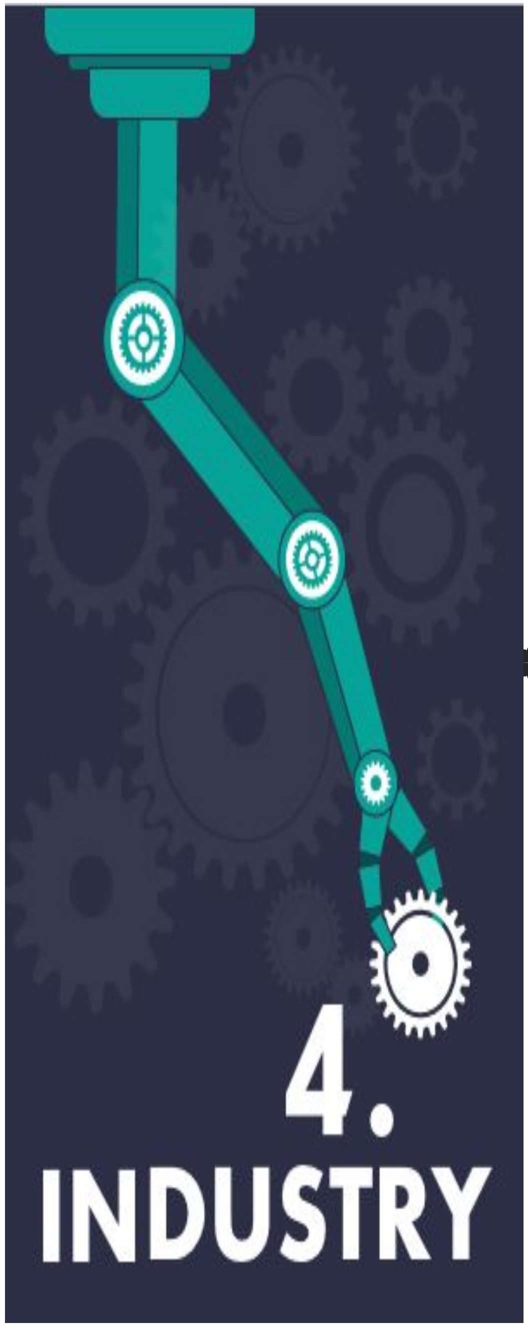
Design

HUMANISTIC AI

ARTIFICIAL SUPER INTELLIGENCE **EXTENDED**



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



- Productivity
- Flexibility
- Speed
- Quantity
- Safety
- Working Conditions
- Environmental Protection
- Innovative Capability