Machine learning

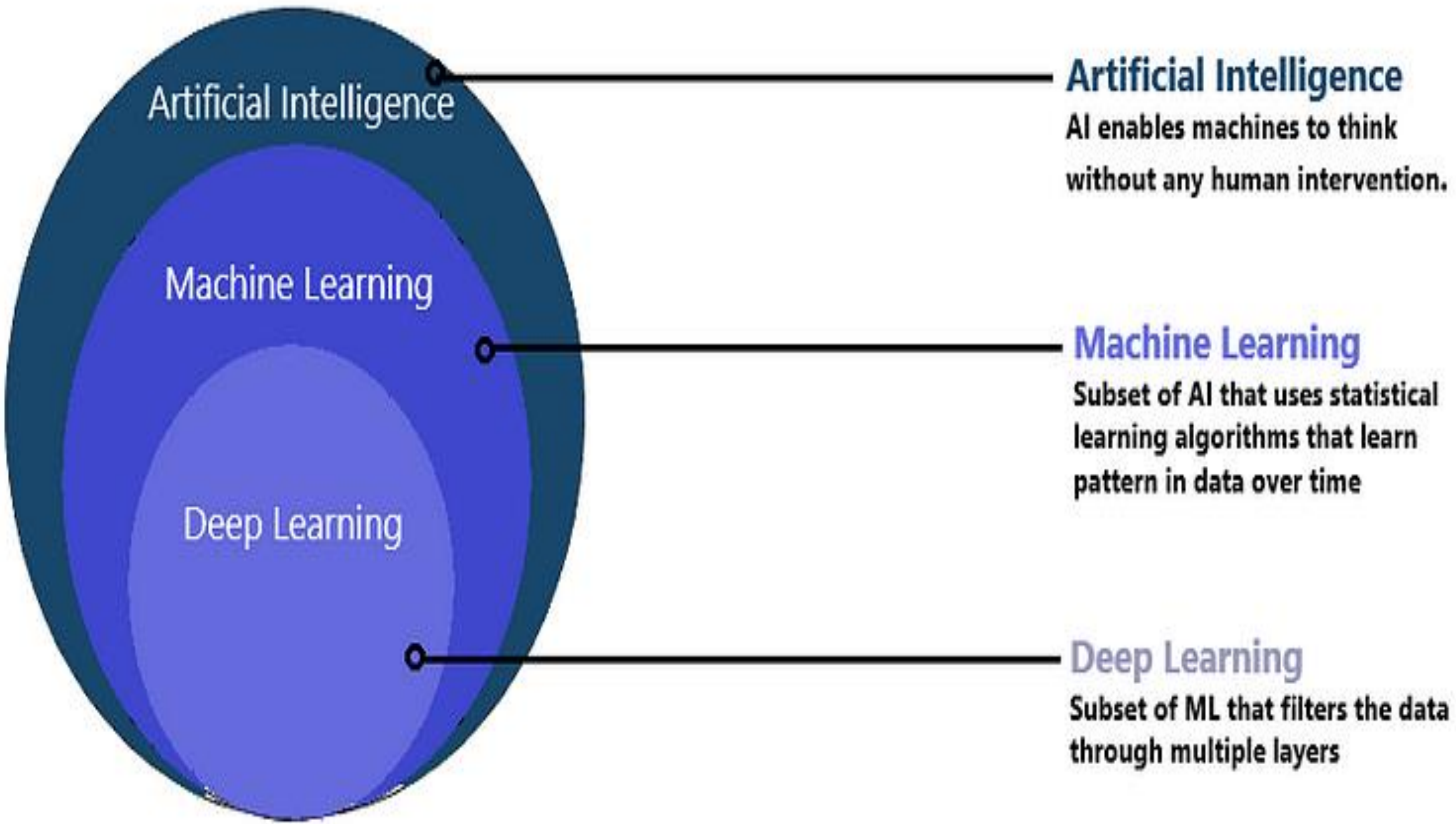
OUTLOOK

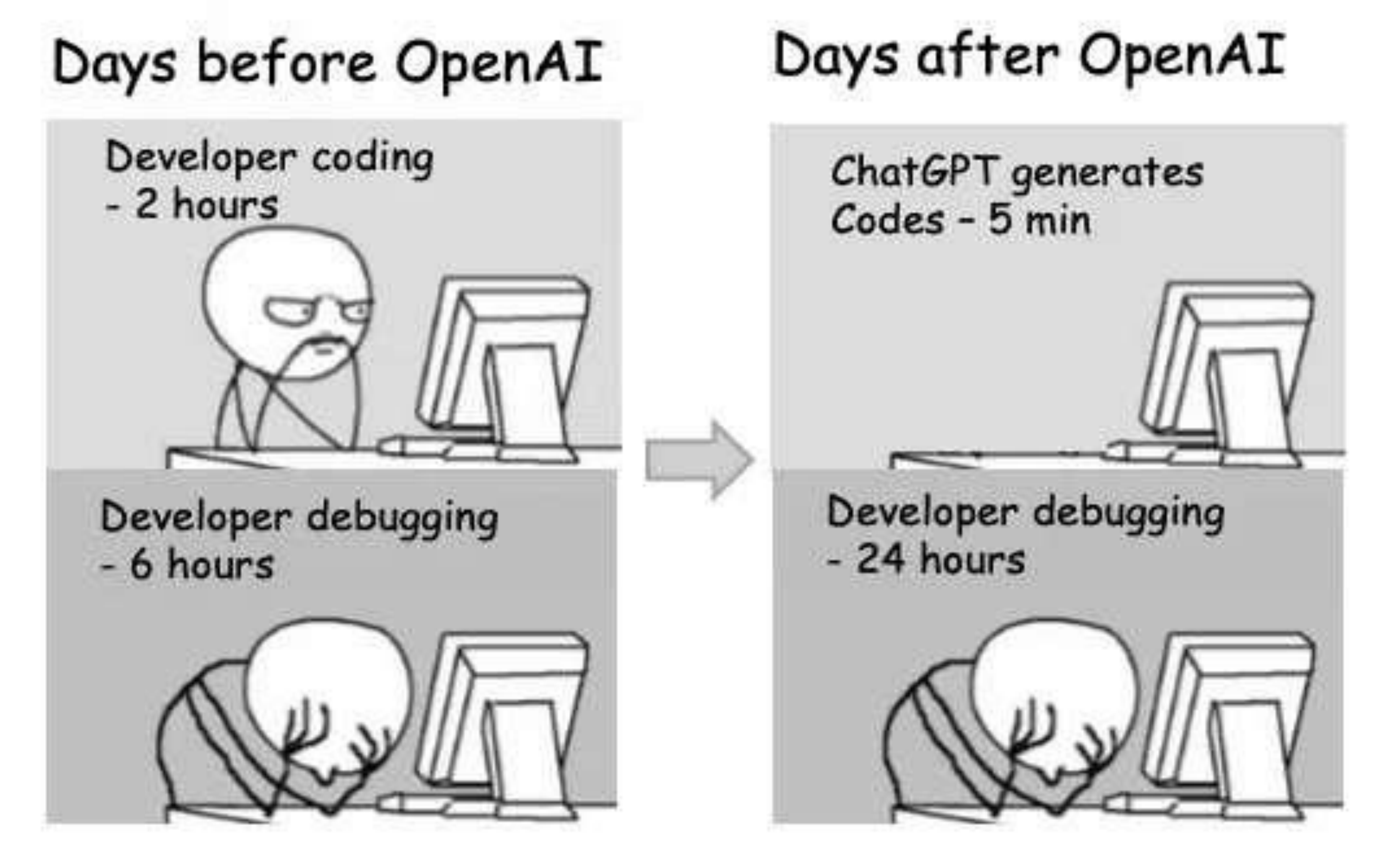
- Artificial Intelligence
- Machine Learning
- Deep Learning

Artificial Intelligence

Robo Revolution: The Age of Automation

What is machine learning ?


Artificial Intelligence
Machine Learning
Deep Learning
Artificial Intelligence
AI enables machines to think
without any human intervention.
Machine Learning
Subset of AI that uses statistical
learning algorithms that learn
pattern in data over time
Deep Learning
Subset of ML that filters the data
through multiple layers


Days before OpenAI
Developer coding - 2 hours
Developer debugging - 6 hours
Days after OpenAI
ChatGPT generates Codes - 5 min
Developer debugging - 24 hours

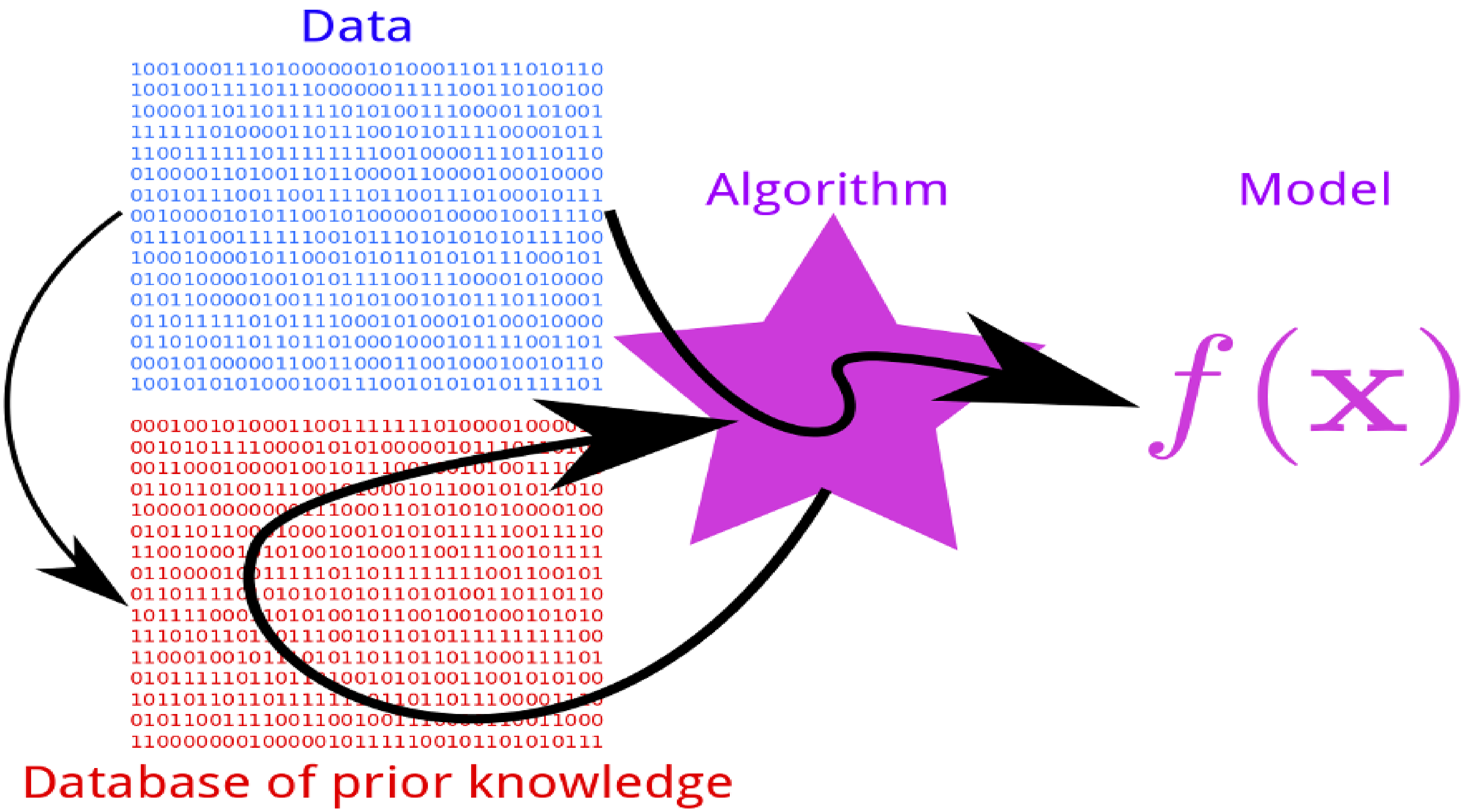
What we need?

Data : For training and improving models.

Algorithms : For processing and analyzing data.

Machine learning models : Such as neural networks.

Computational infrastructure : For processing and storing data.


Data
Algorithm
Model
$f(\mathbf{x})$
Database of prior knowledge

Library

What is python Library? A library is a collection of code that makes everyday tasks more efficient.

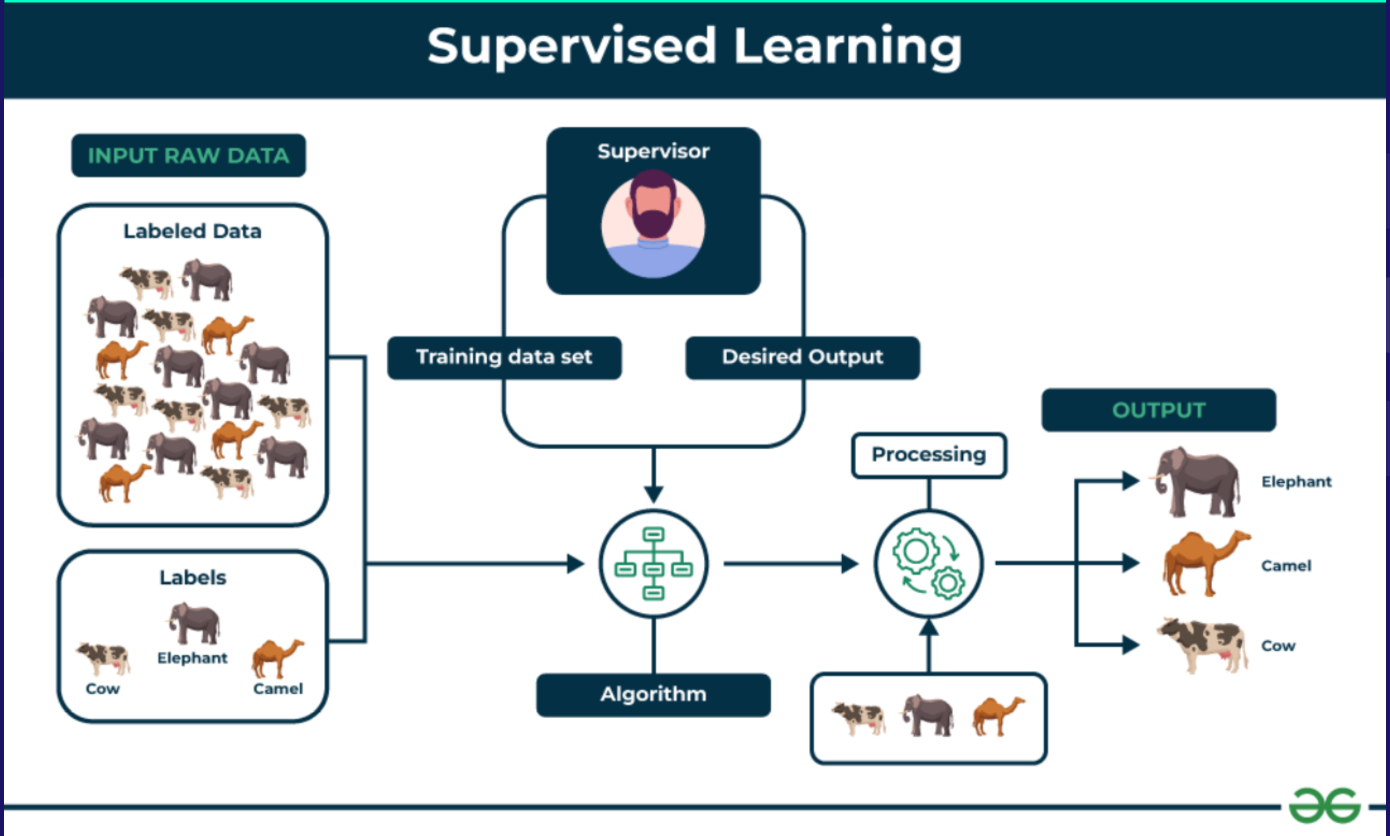
Two important definitions

1-Model :A “model” in machine learning is the output of a machine learning algorithm run on data. A model represents what was learned by a machine learning algorithm.

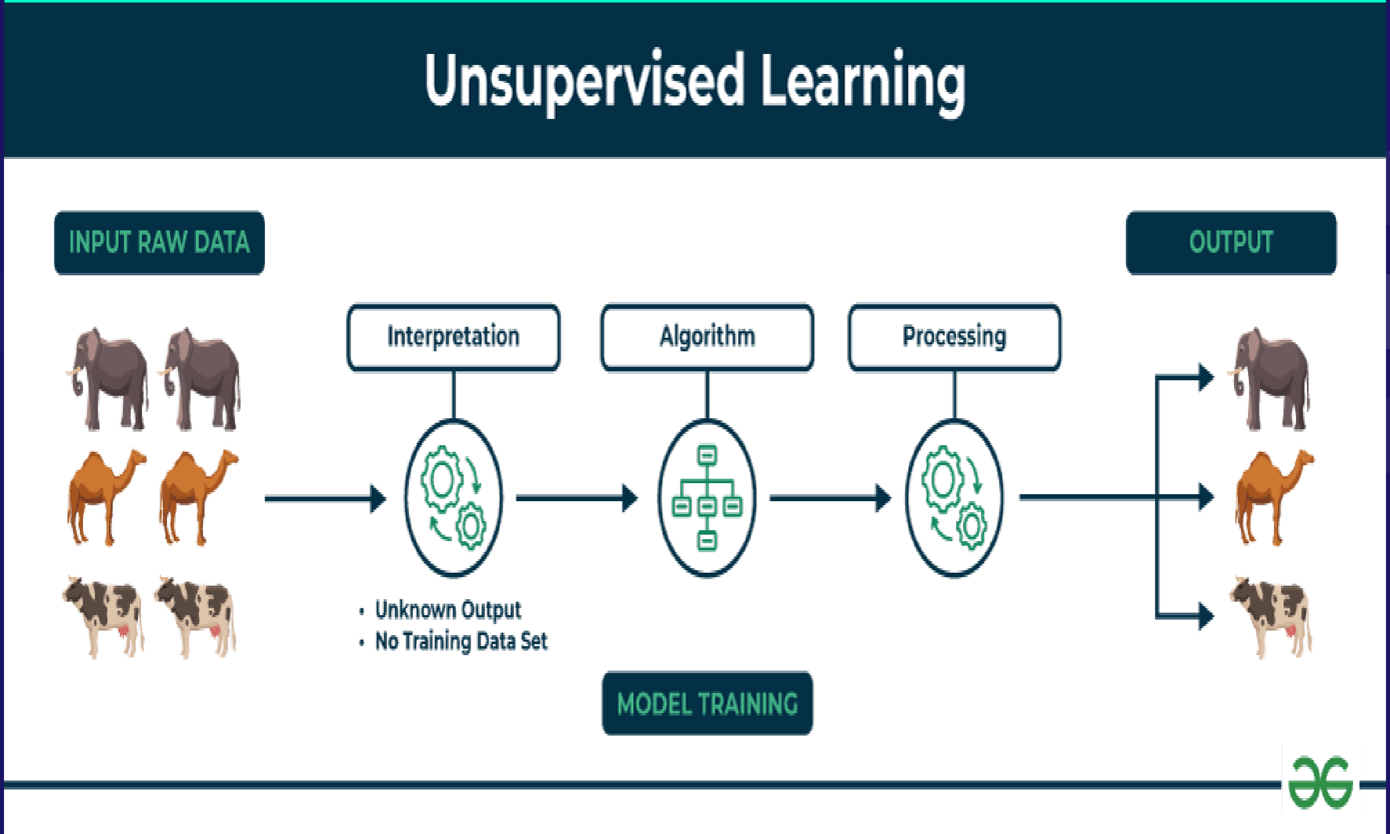
2-algorithm :An “algorithm” in machine learning is a procedure that is run on data to create a machine learning “model.” Machine learning algorithms perform “pattern recognition.” Algorithms “learn” from data, or are “fit” on a dataset.
There are many machine learning algorithms. For example, we have algorithms for classification, such as k-nearest neighbors. We have algorithms for regression, such as linear regression, and we have algorithms for clustering, such as k-means.

Machine Learning Process

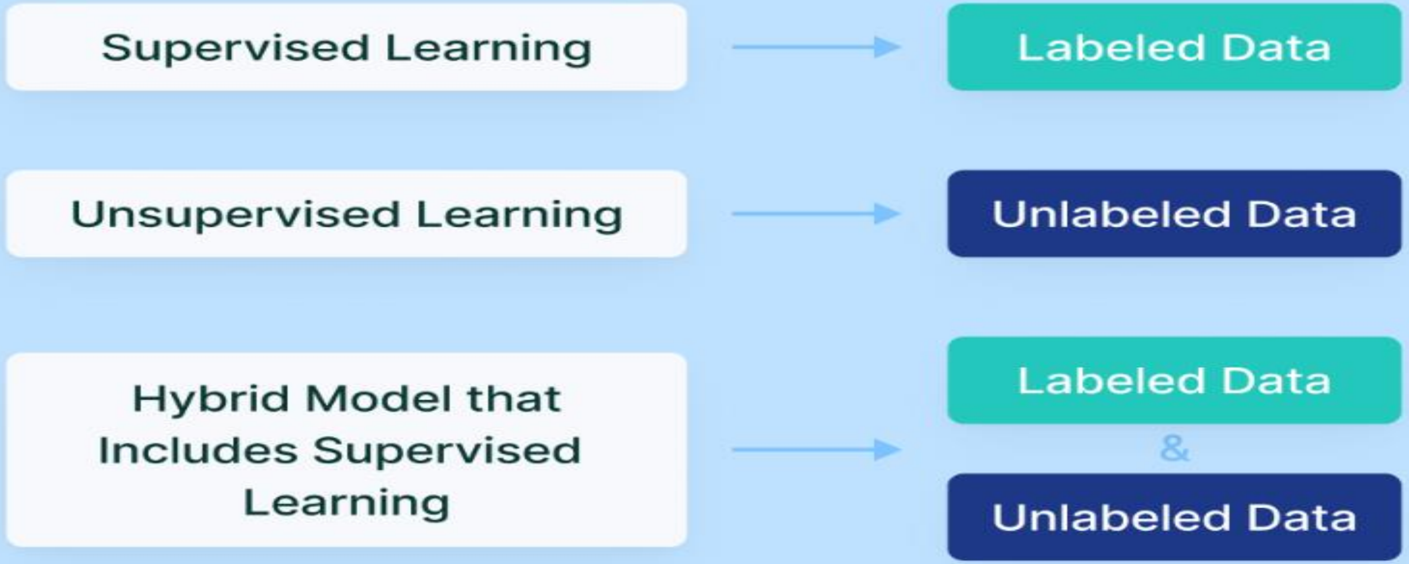
Supervised vs. unsupervised learning



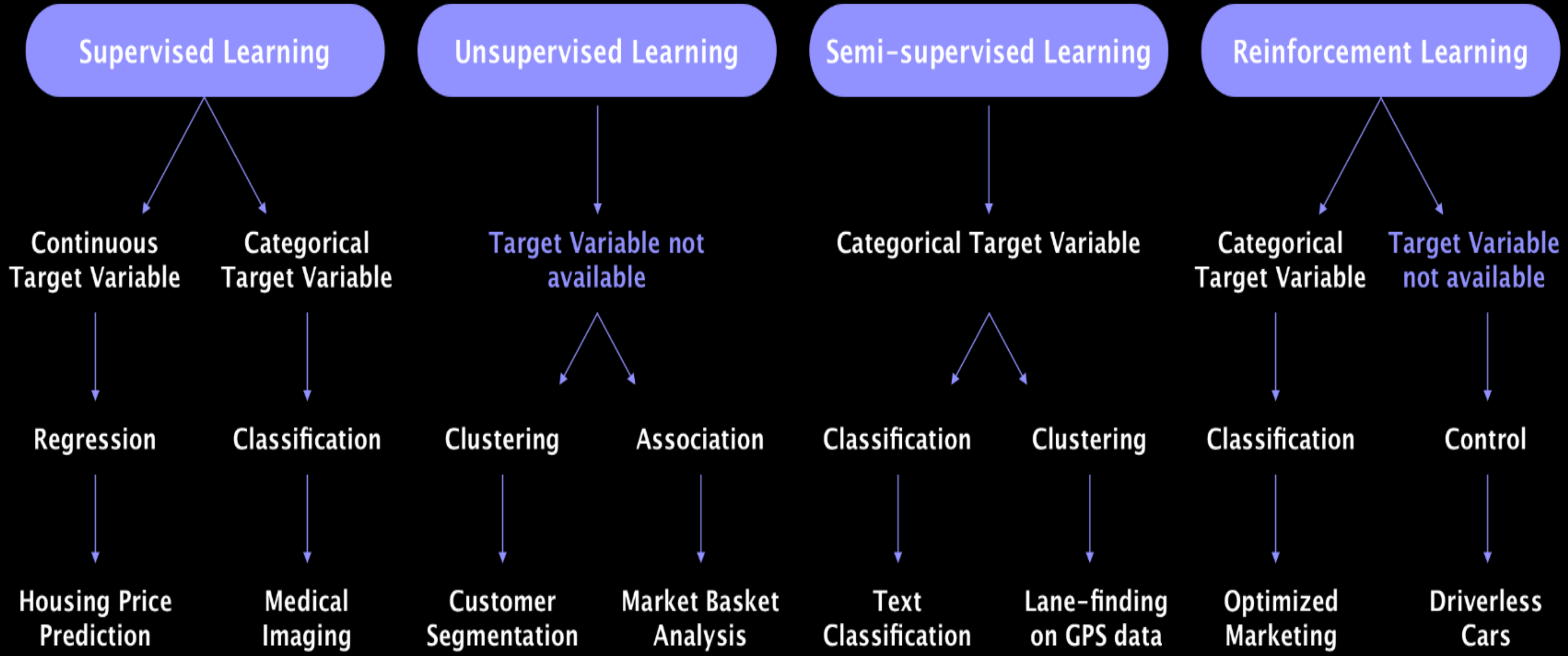
Supervised vs. unsupervised learning



Data in Supervised vs. Unsupervised Learning



Machine Learning Types



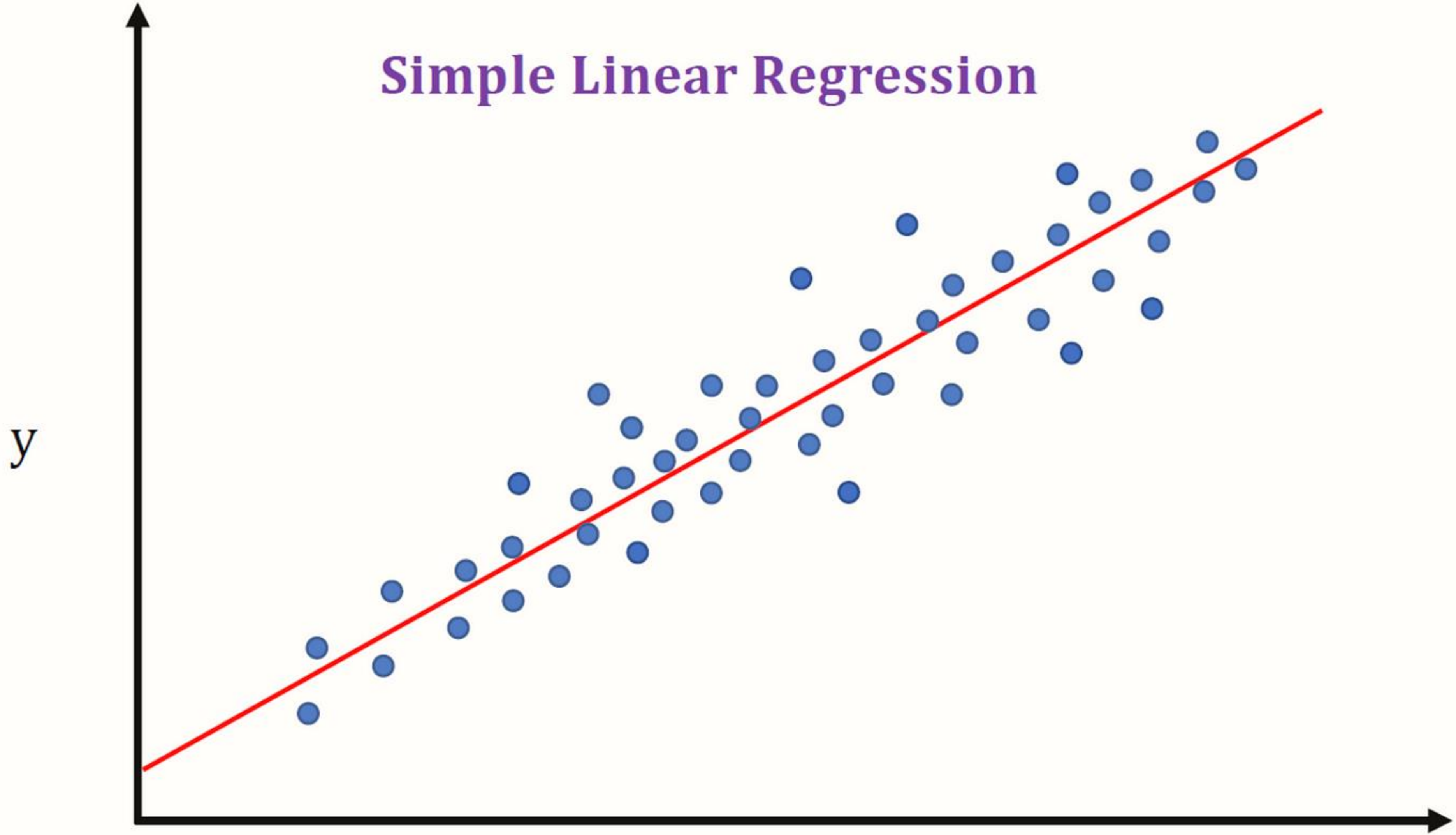
Regression intro

- Continuous value
- Independent / dependent variables
- Y = Continuous
- $Y = ax+b$
- Intercept / coefficient

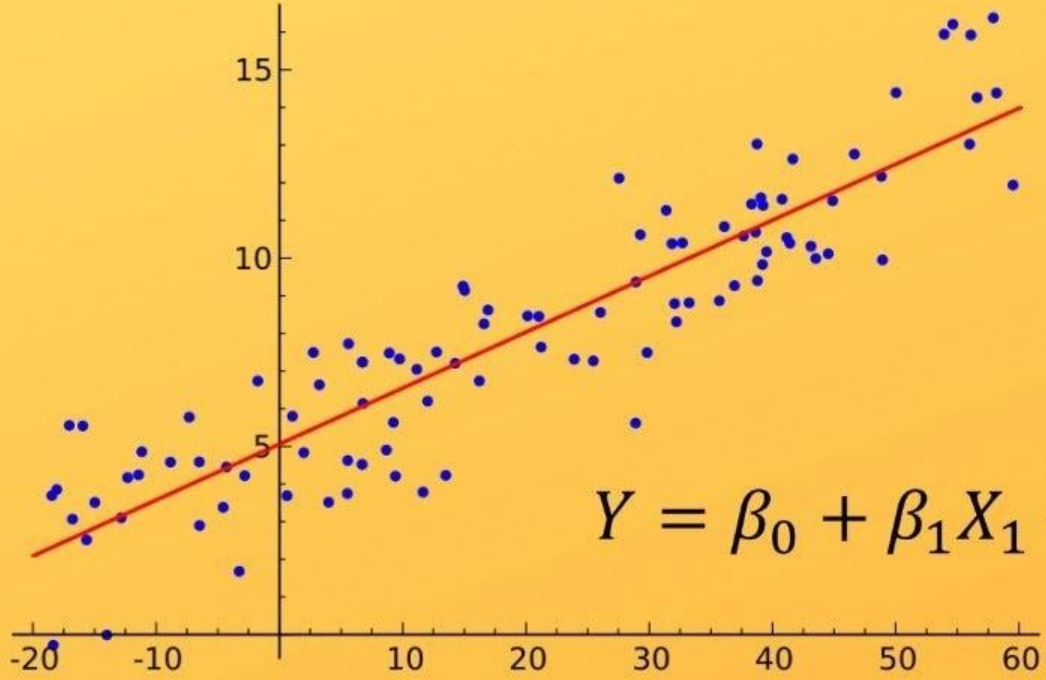
Regression types

- Simple : linear / non-linear
$Y=ax+b$

- Multiple : linear / non-linear
$Y=ax1+bx2+c$

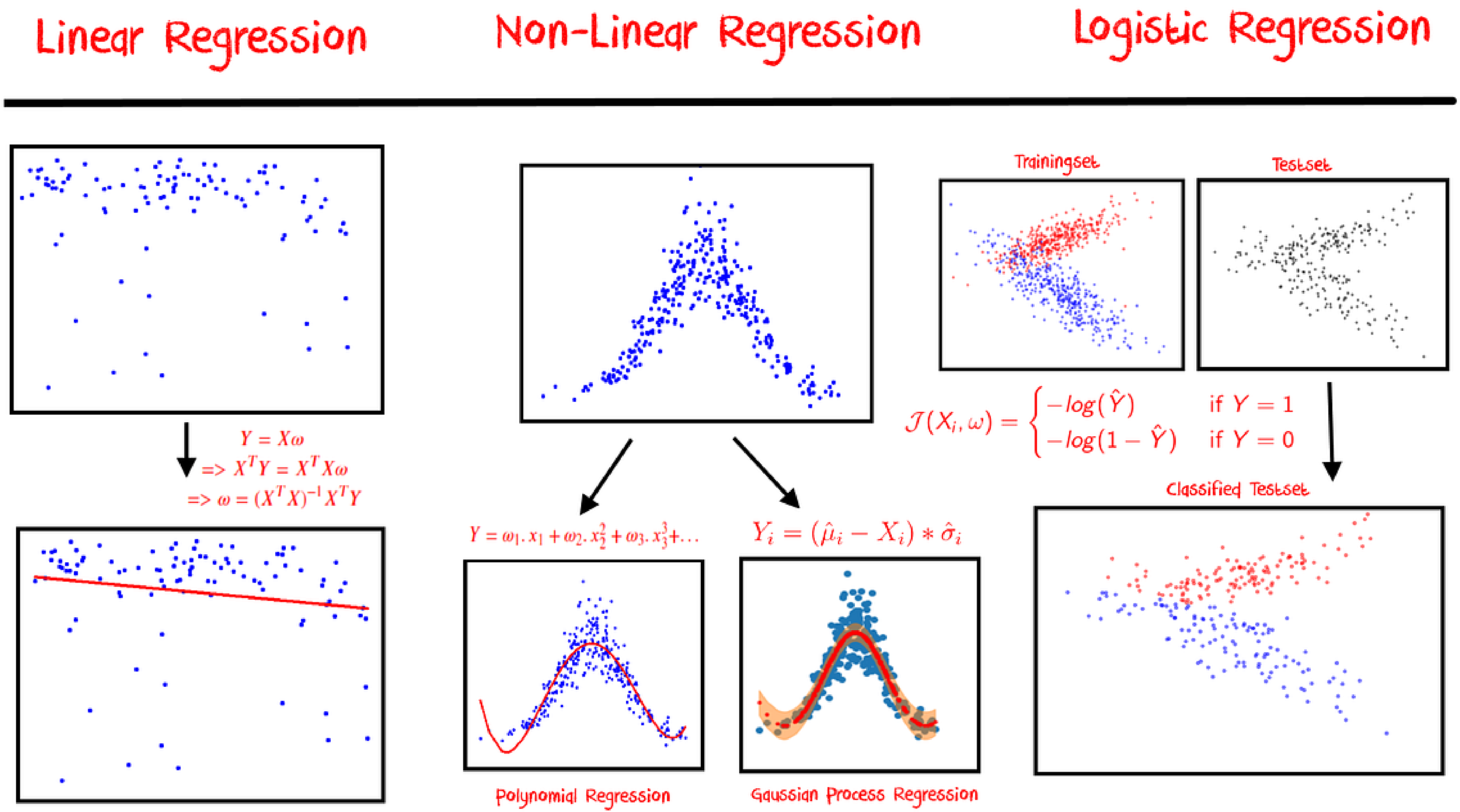

Simple Linear Regression
y

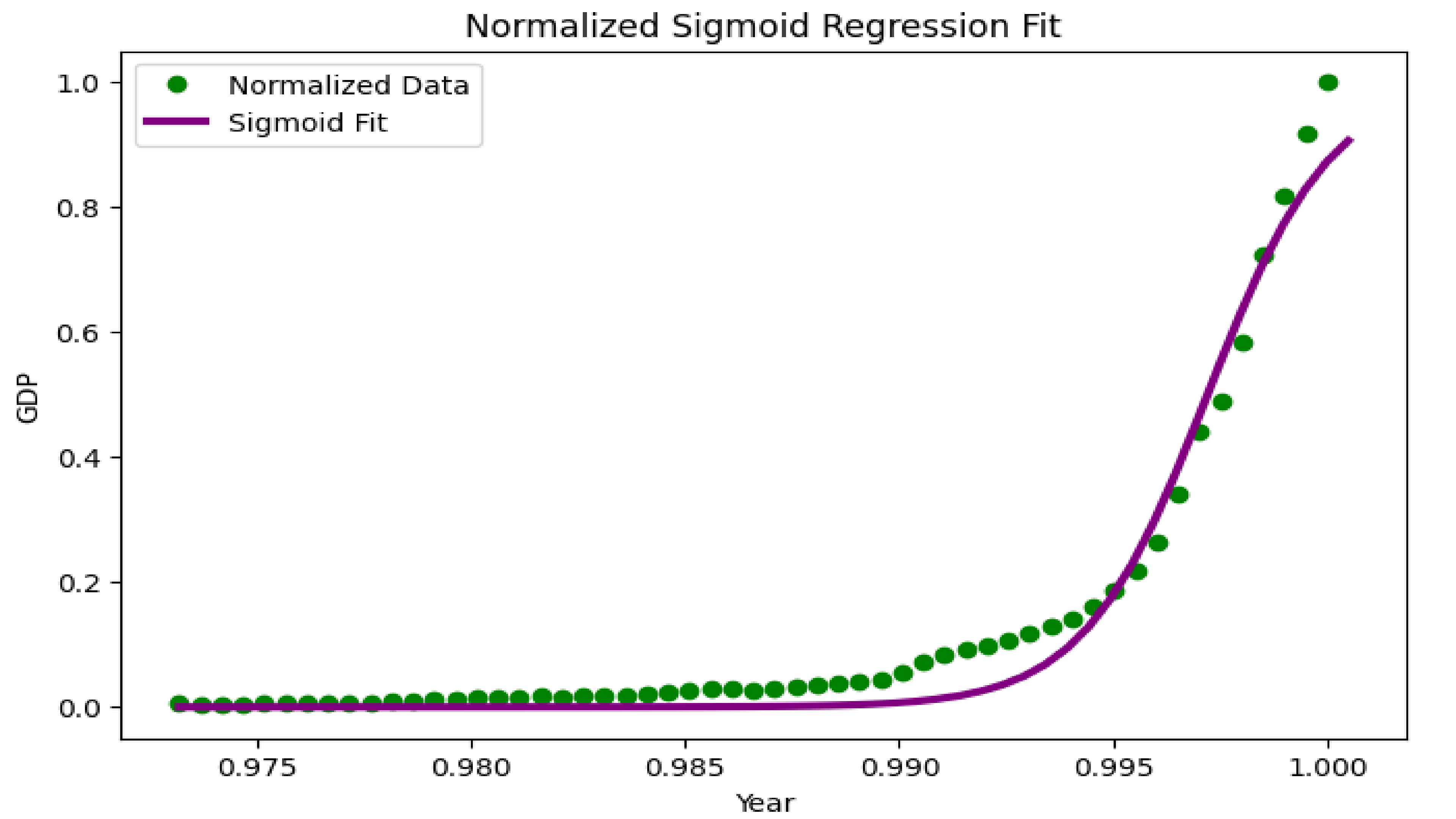
Multiple Linear Regression



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{\boxed{p}} X_{\boxed{p}} + \varepsilon$$

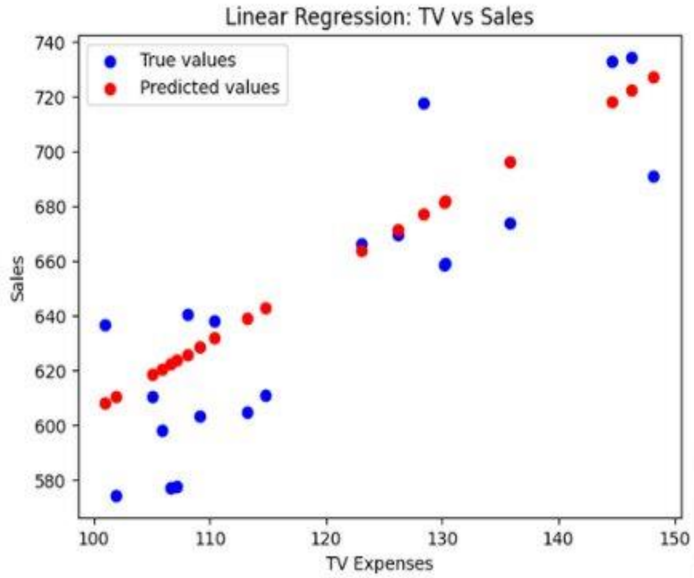
number of predictors

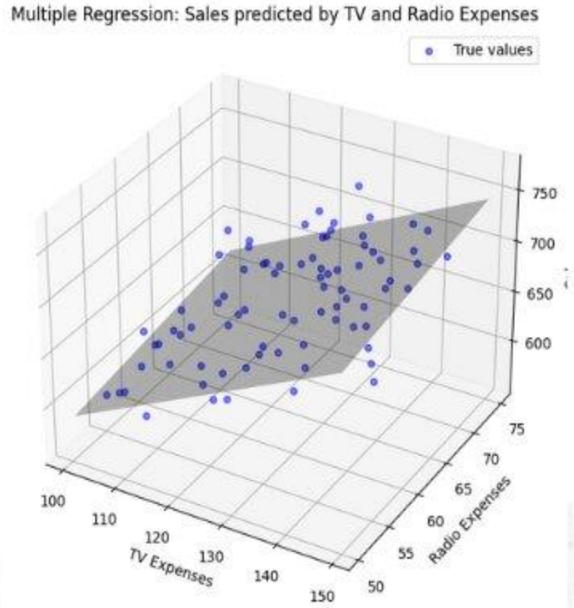

Linear Regression
Non-Linear Regression
Logistic Regression
$Y = X\omega$
$\Rightarrow X^T Y = X^T X\omega$
$\Rightarrow \omega = (X^T X)^{-1} X^T Y$
$Y = \omega_1 . x_1 + \omega_2 . x_2^2 + \omega_3 . x_3^3 + \ldots$
$Y_i = (\hat{\mu}_i - X_i) * \hat{\sigma}_i$
Polynomial Regression
Gaussian Process Regression
Trainingset
Testset
$\mathcal{J}(X_i, \omega) = \begin{cases} -log(\hat{Y}) & \text{if } Y = 1 \\ -log(1 - \hat{Y}) & \text{if } Y = 0 \end{cases}$
Classified Testset



Normalized Sigmoid Regression Fit
Normalized Data
Sigmoid Fit
GDP
Year
0.0
0.2
0.4
0.6
0.8
1.0
0.975
0.980
0.985
0.990
0.995
1.000

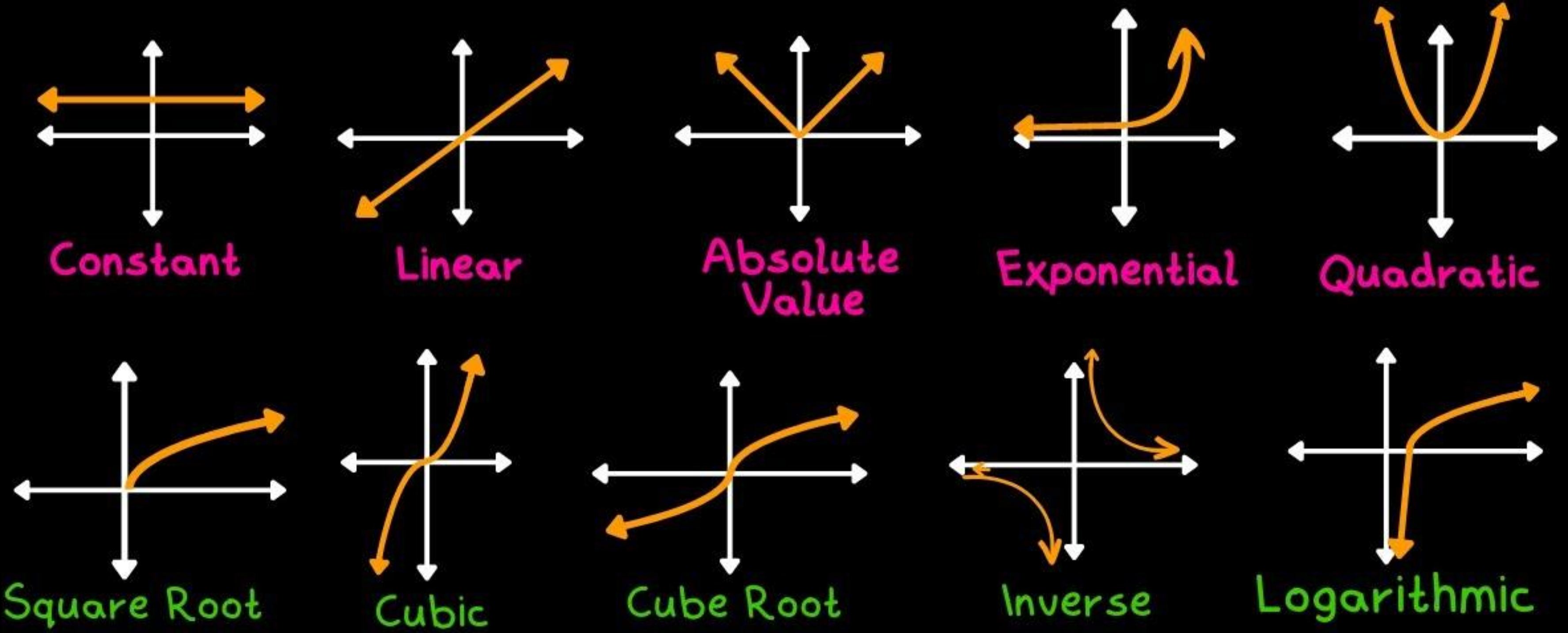
LINEAR REGRESSION VS MULTIPLE REGRESSION





Parent Functions Review

Domain & Range



Model evaluation

For what?

- Train and test on same data
- Train/test split

Model Evaluation

Errors

- mean absolute error (MAE)
- mean squared error (MSE)
- root mean squared error (RMSE); interpretable in the same units as the response vector or y units
- Relative absolute error, also known as residual sum of square (RAE)
- Relative squared error (RSE)
- R2; Popular metric for the accuracy of your model. represents how close the data values are to the fitted regression line. The higher the better

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

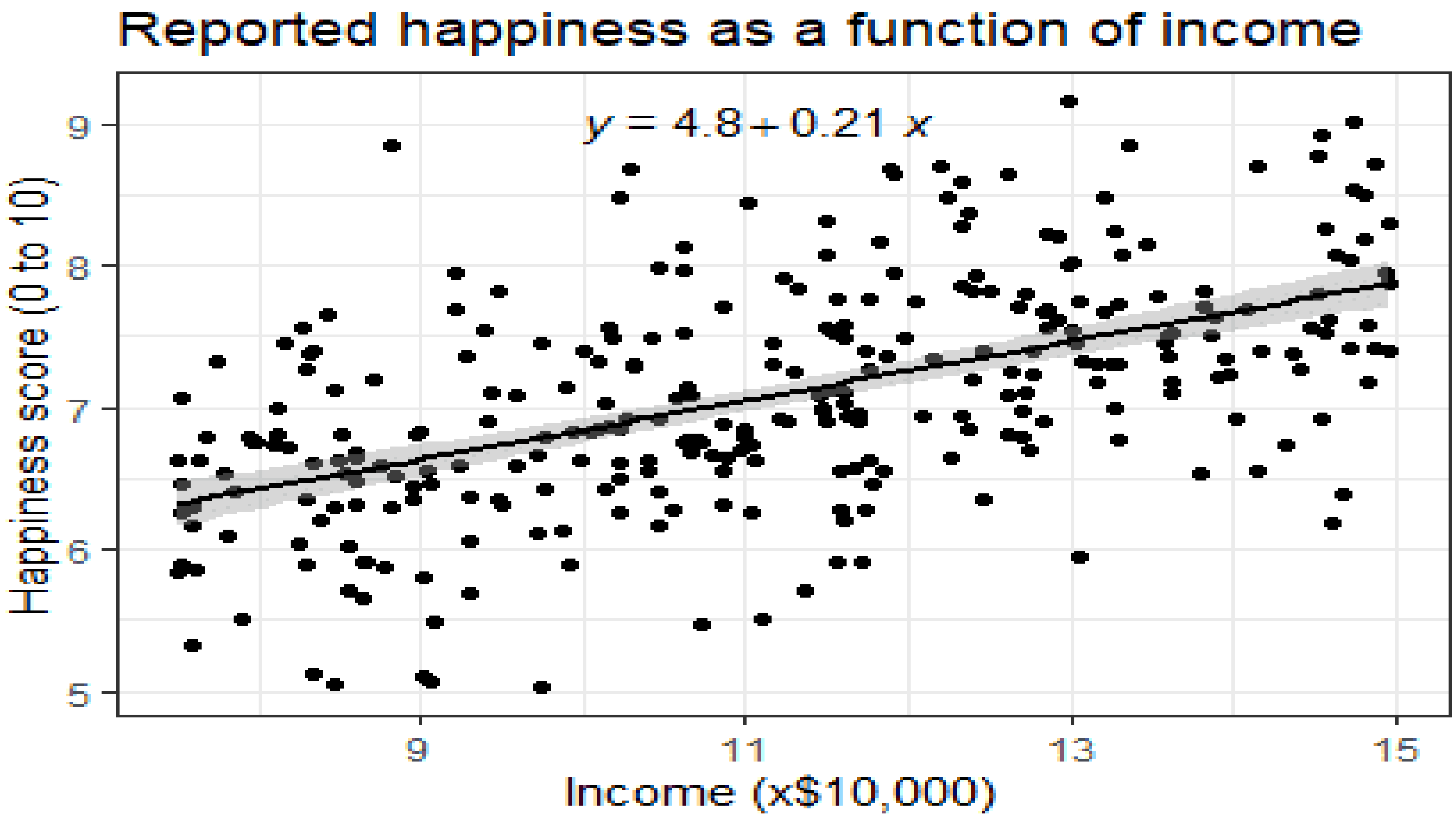
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

$$RAE = \frac{\sum_{j=1}^{n}|y_j - \hat{y}_j|}{\sum_{j=1}^{n}|y_j - \bar{y}|}$$

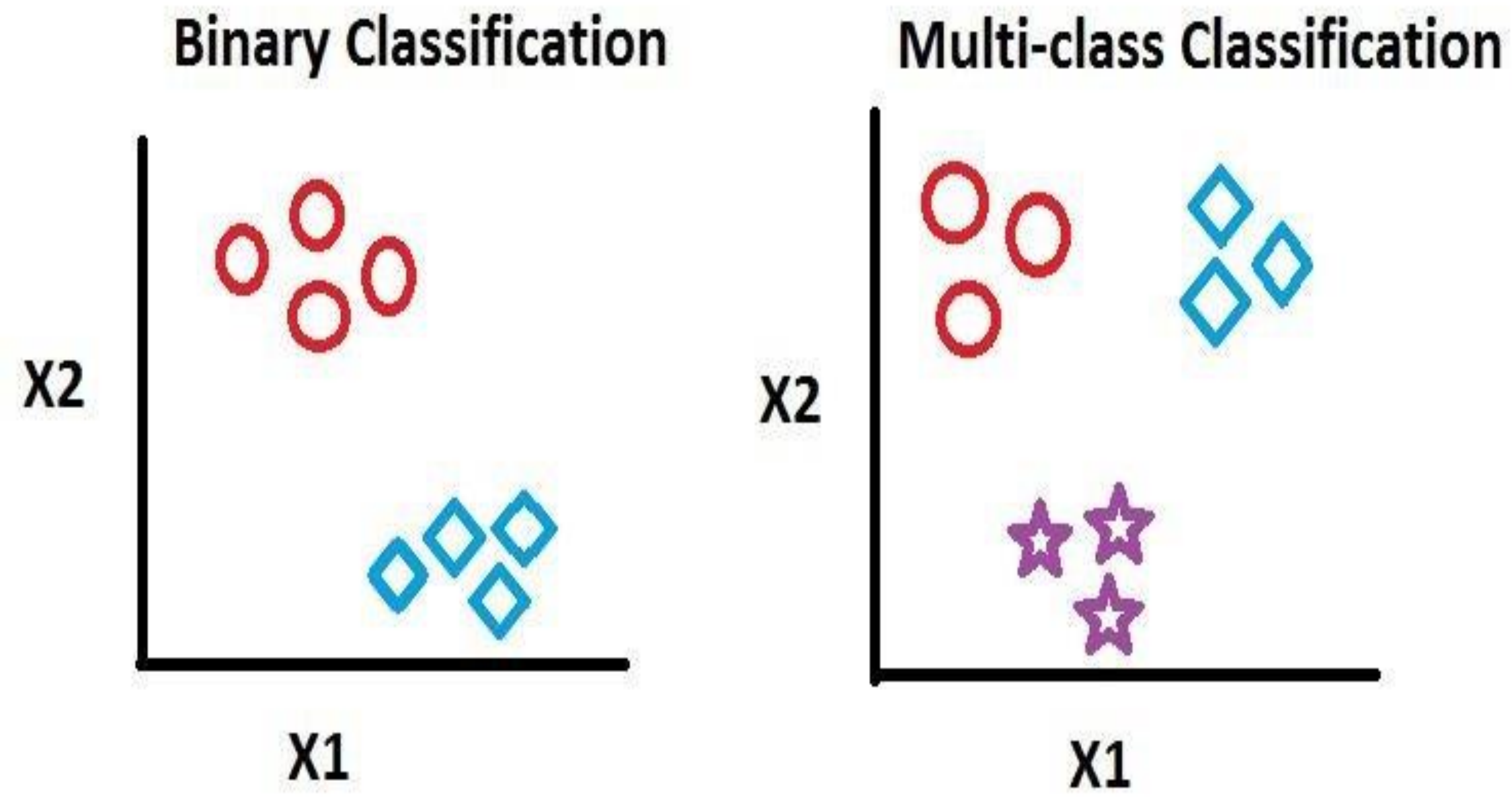
$$RSE = \frac{\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2}$$

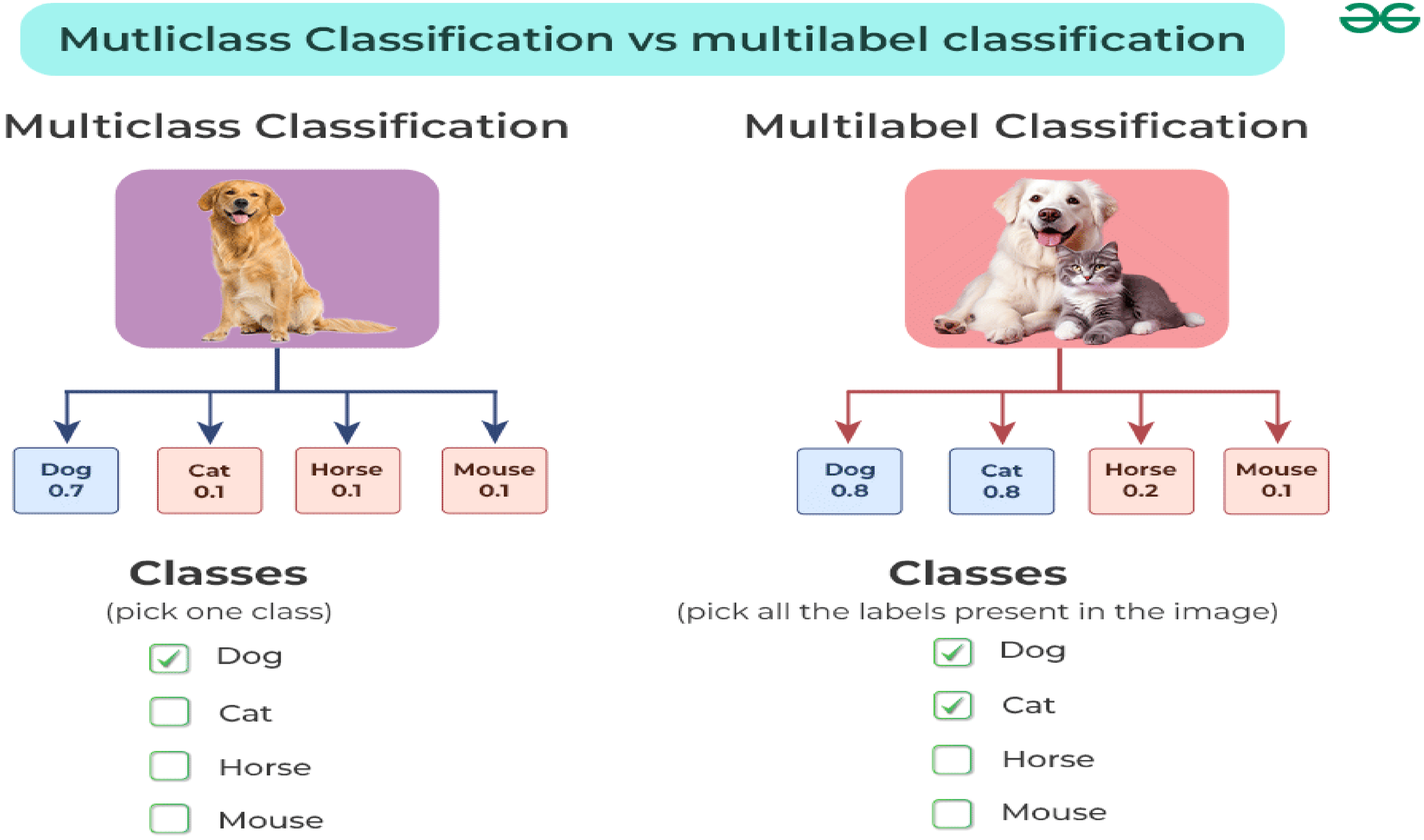
$$\boxed{R^2 = 1 - RSE}$$

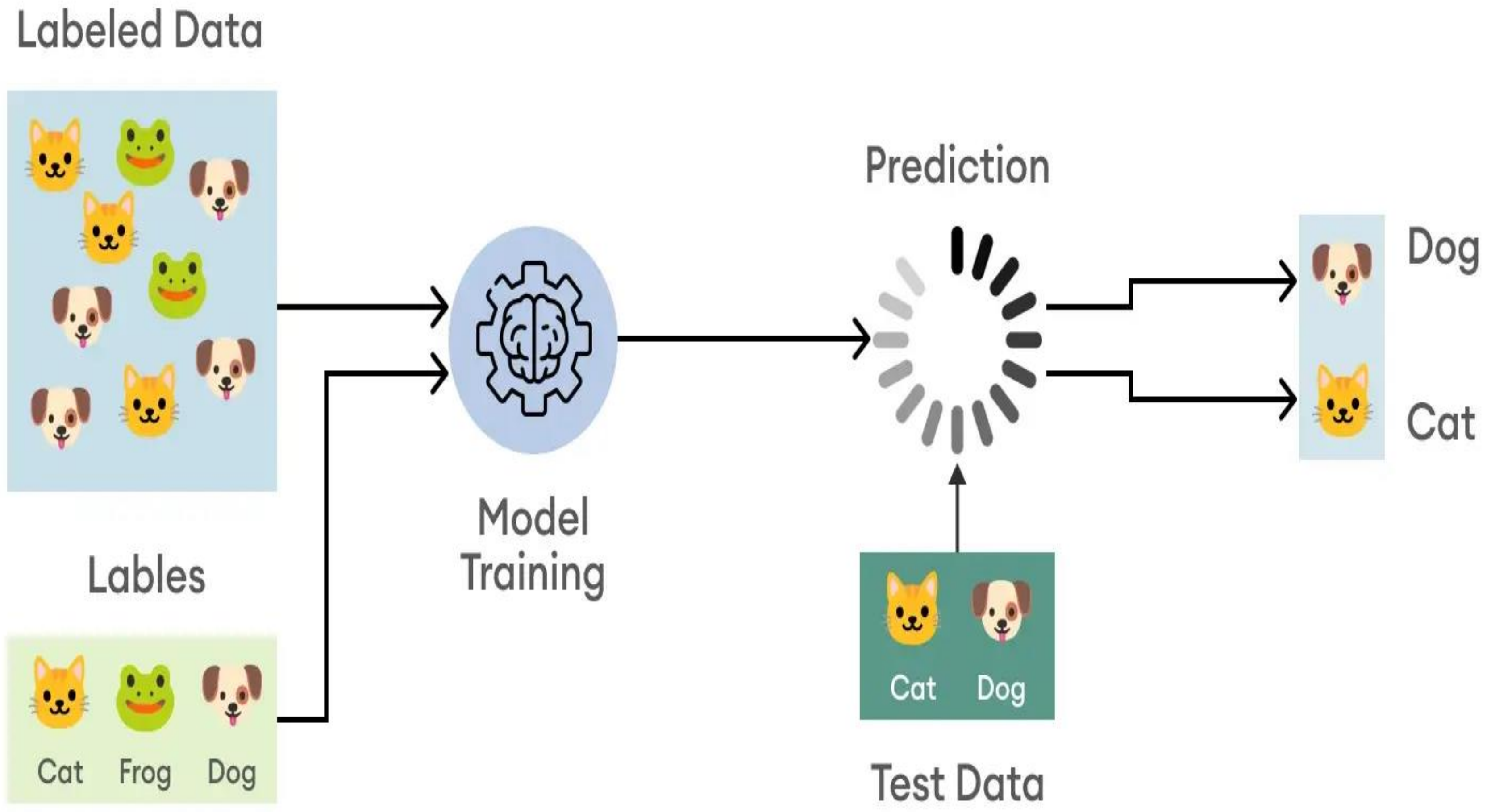

Reported happiness as a function of income
y = 4.8 + 0.21 x
Happiness score (0 to 10)
9
8
7
6
5
9
11
13
15
Income (x$10,000)

Classification intro

- Supervised
- Categorizing
- Multi-class/binary
- KNN / Decision Trees / Logistic Regression / SVM


Binary Classification
X2
X1
Multi-class Classification
X2
X1

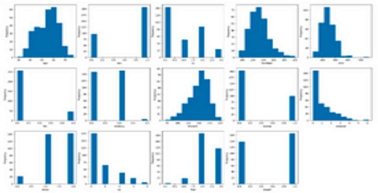

Mutliclass Classification vs multilabel classification
Multiclass Classification
Dog
0.7
Cat
0.1
Horse
0.1
Mouse
0.1
Classes
(pick one class)
Dog
Cat
Horse
Mouse
Multilabel Classification
Dog
0.8
Cat
0.8
Horse
0.2
Mouse
0.1
Classes
(pick all the labels present in the image)
Dog
Cat
Horse
Mouse

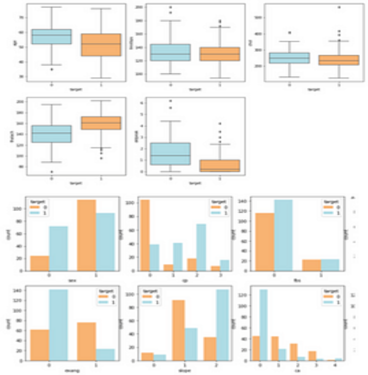

Labeled Data
Lables
Cat
Frog
Dog
Model
Training
Prediction
Dog
Cat
Cat
Dog
Test Data

Machine Learning Algorithms - Classification

Exploratory Data Analysis (EDA)

1) **Histogram:** df.plot(kind = 'hist')
2) **Box Plot:** sns.boxplot()
3) **Grouped Bar Chart:** sns.countplot()

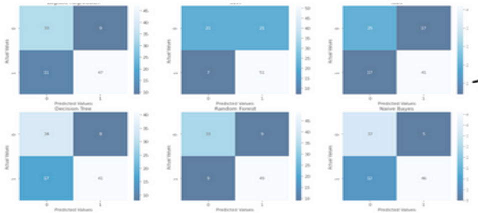




Model Evaluation

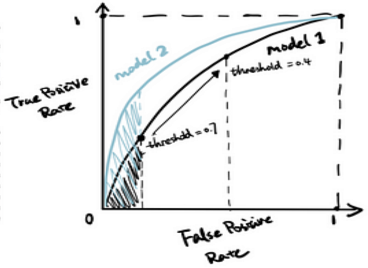
Confusion Matrix

confusion_matrix(y_test, y_pred)

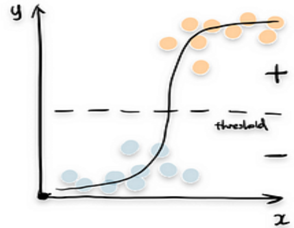


ROC & AUC

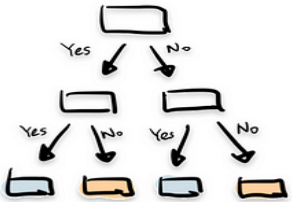
metrics.auc(fpr, tpr)



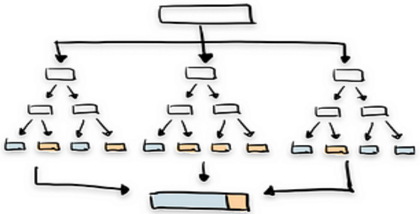
Logistic Regression



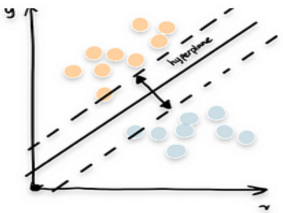
Decision Tree



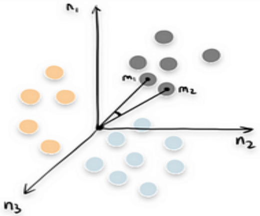
Random Forest



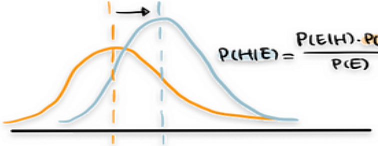
Support Vector Machine



K Nearest Neighbour



Naive Bayes



$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

Classification

Intro

x

y

m

Age	Sex	BP	Cholesterol	Na	K	Drug
23	F	HIGH	HIGH	0.793	0.031	drugY
47	M	LOW	HIGH	0.739	0.056	drugC
47	M	LOW	HIGH	0.697	0.069	drugC
28	F	NORMAL	HIGH	0.564	0.072	drugX
61	F	LOW	HIGH	0.559	0.031	drugY
22	F	NORMAL	HIGH	0.677	0.079	drugX
49	F	NORMAL	HIGH	0.79	0.049	drugY
41	M	LOW	HIGH	0.767	0.069	drugC
60	M	NORMAL	HIGH	0.777	0.051	drugY
43	M	LOW	NORMAL	0.526	0.027	drugY

Categorical Variable

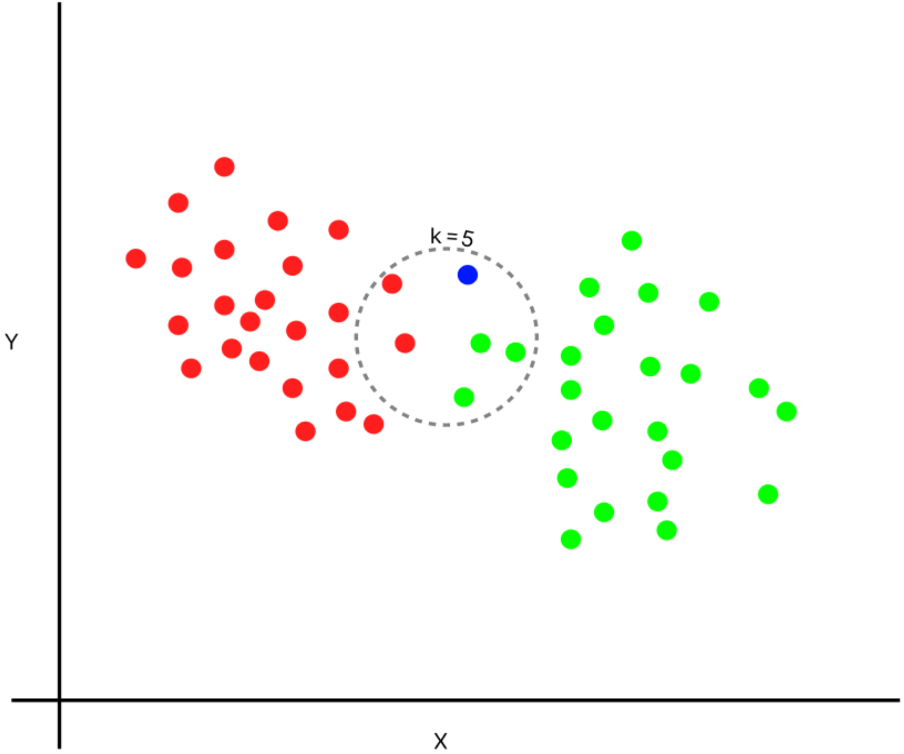
Age	Sex	BP	Cholesterol	Na	K	Drug
36	F	LOW	HIGH	0.697	0.069	

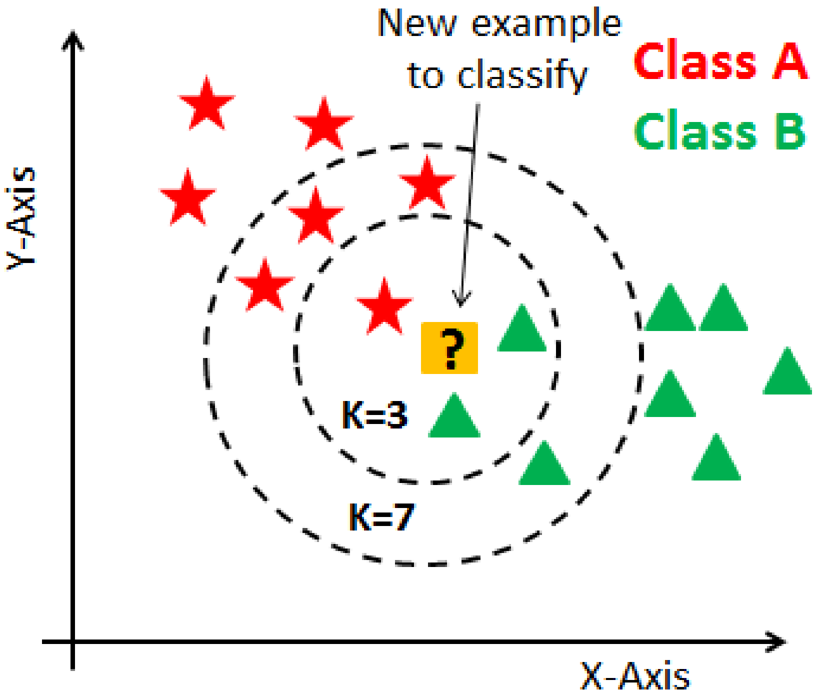
M

KNN
k-nearest neighbors

- Reservoir
- Overfitting/underfitting

KNN
k-nearest neighbors





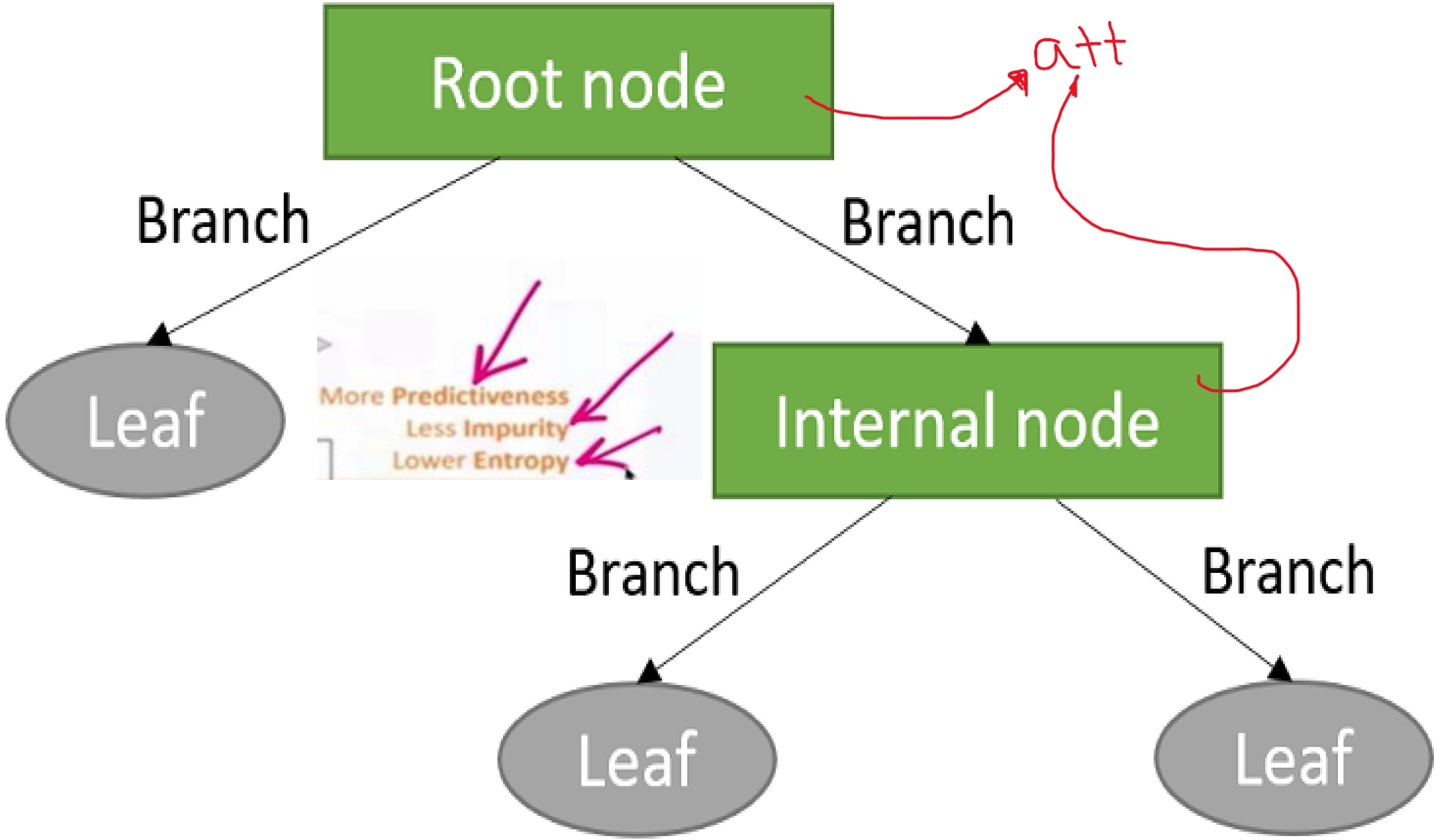
Decision trees intro

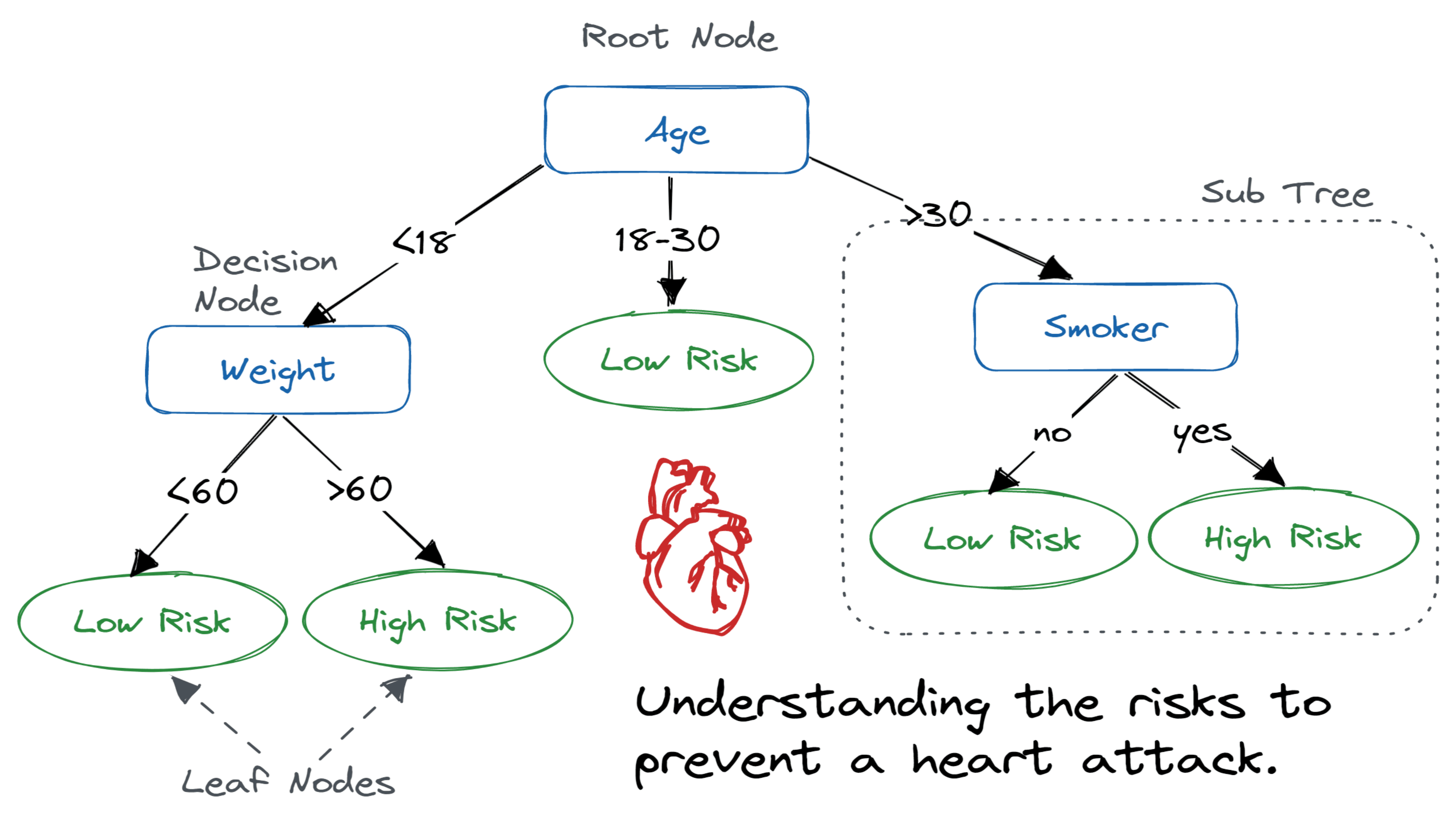
- Internal node / branch / leaf
- Choose attribute from dataset
- Split data based on value of the best attribute

$$Gain(T,X) = Entropy(T) - Entropy(T,X)$$

$$\mathbf{G}(\text{PlayGolf, Outlook}) = \mathbf{E}(\text{PlayGolf}) - \mathbf{E}(\text{PlayGolf, Outlook})$$

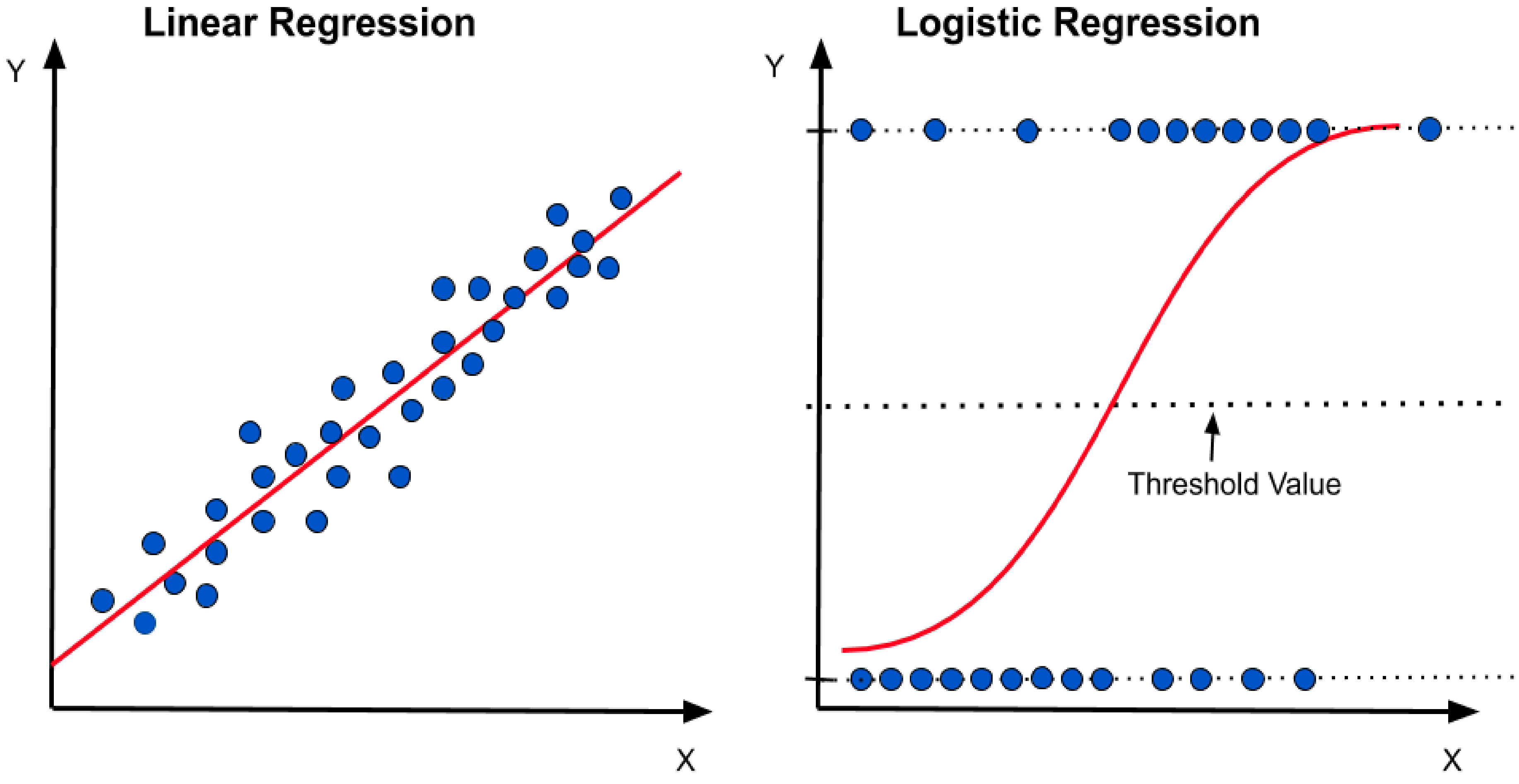
$$= 0.940 - 0.693 = 0.247$$


Root node
att
Branch
Branch
Leaf
More Predictiveness
Less Impurity
Lower Entropy
Internal node
Branch
Branch
Leaf
Leaf

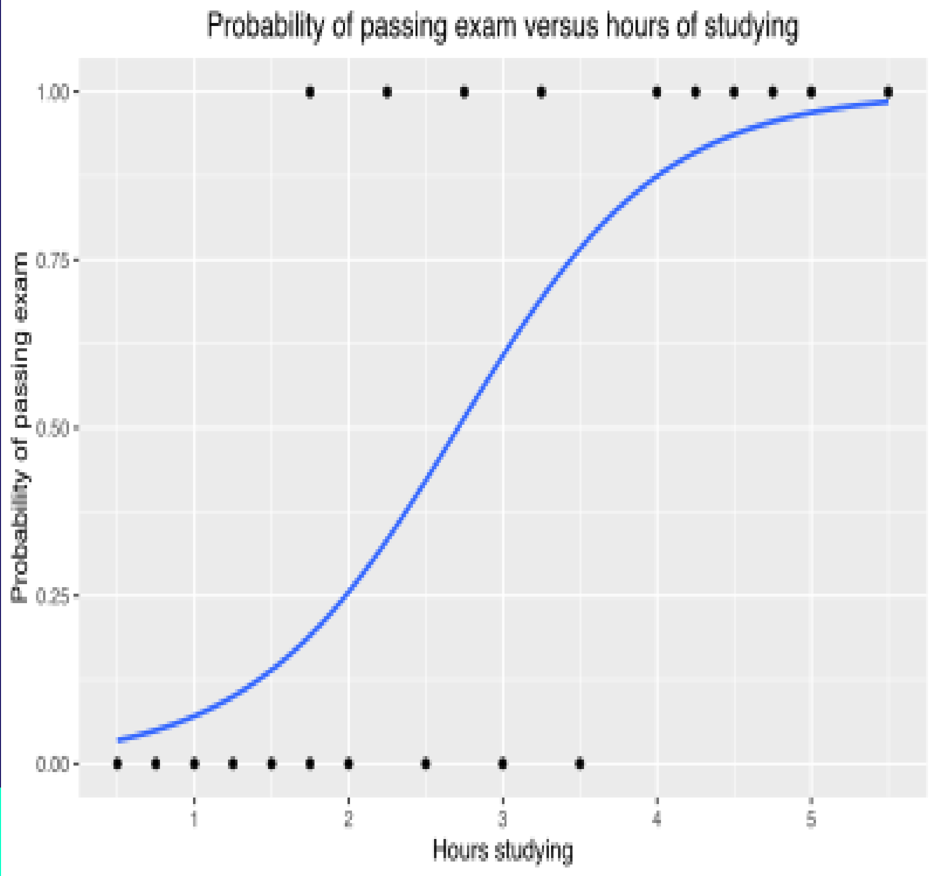

Root Node
Age
<18
18-30
>30
Sub Tree
Decision
Node
Weight
Low Risk
Smoker
<60
>60
no
yes
Low Risk
High Risk
Low Risk
High Risk
Leaf Nodes
Understanding the risks to prevent a heart attack.

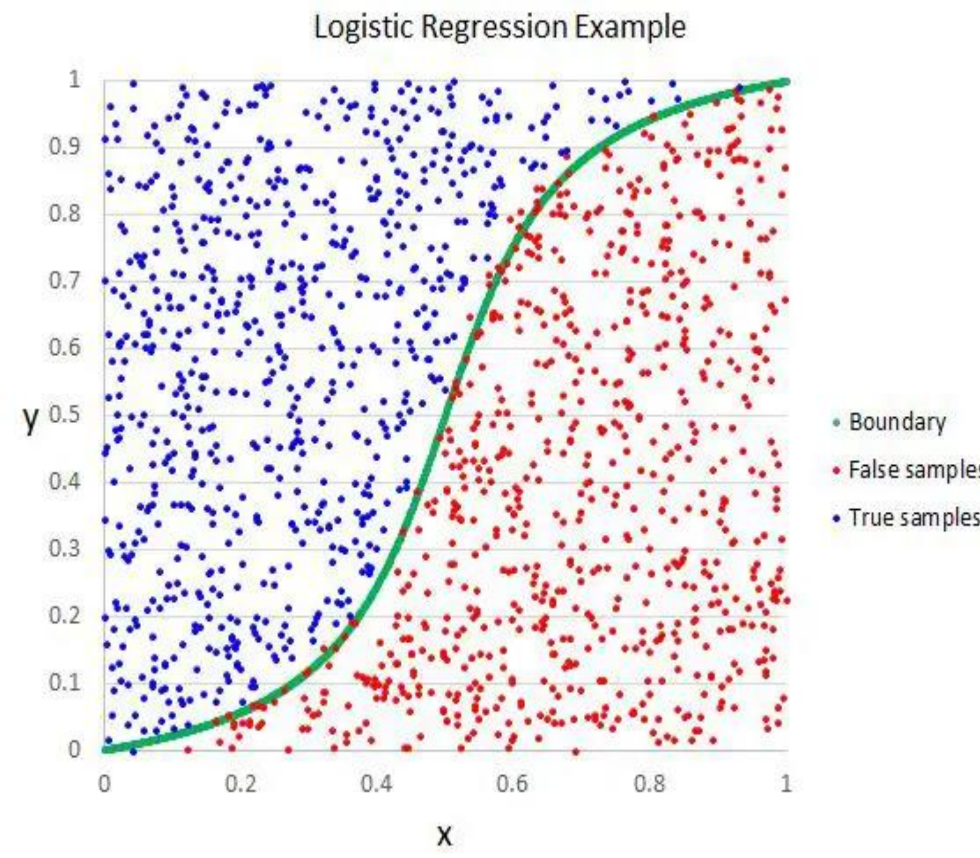
Logistic regression

- supervised
- Close to regression but here, y is a categorical or binary
- All y should be continues


Linear Regression
Y
X
Logistic Regression
Y
Threshold Value
X

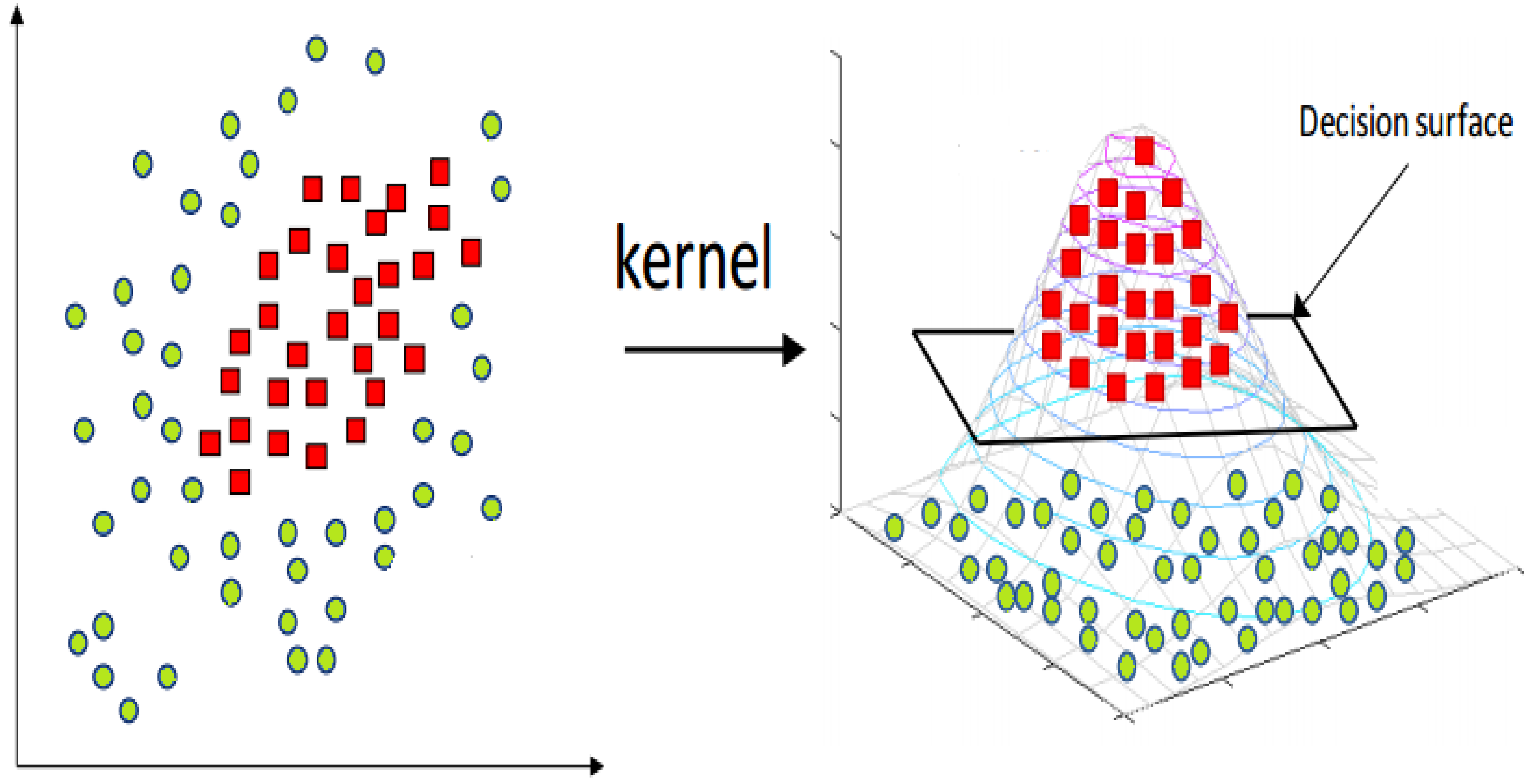
Logistic regression



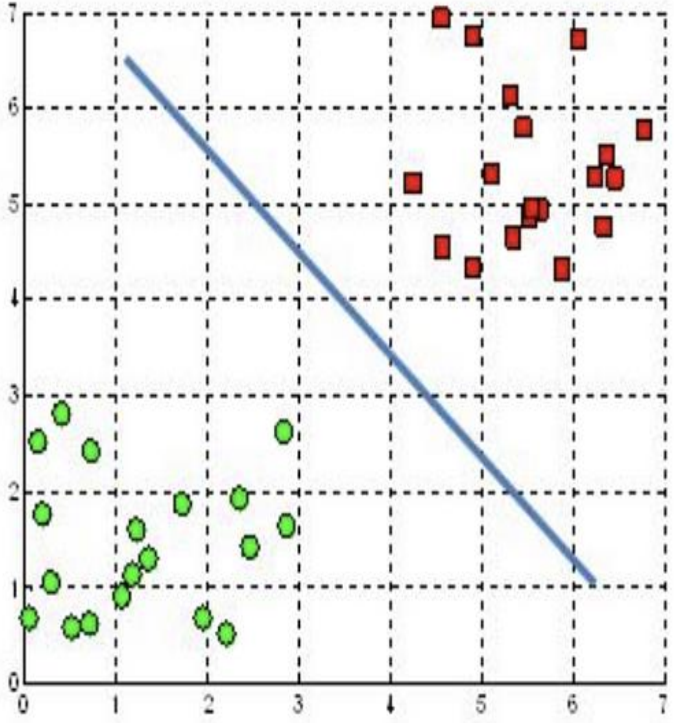


SVM
support vector machines

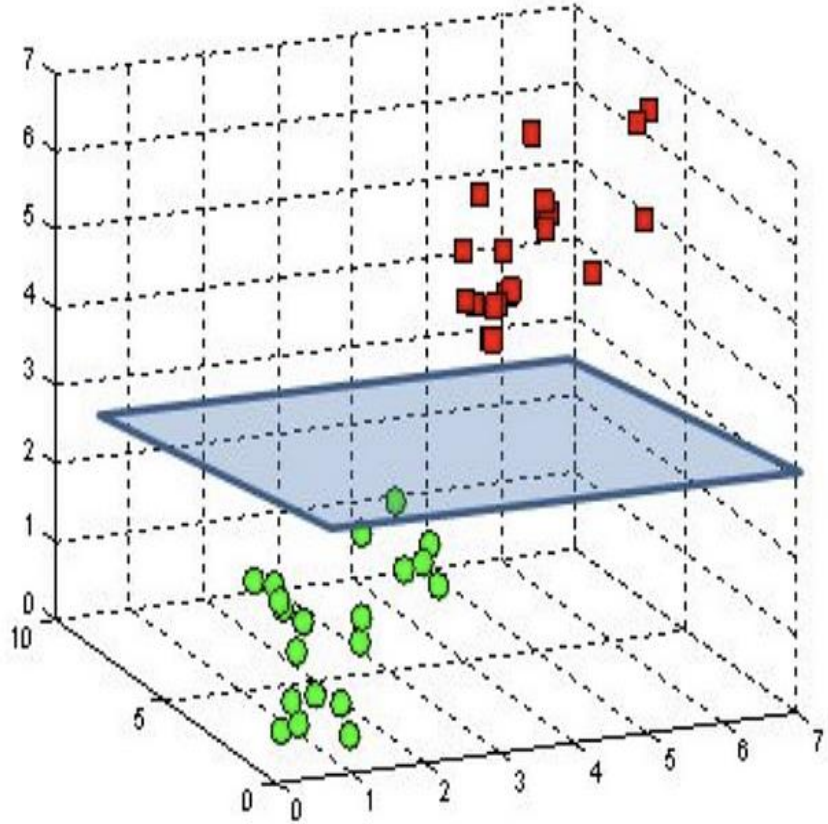
- Supervised
- Based on separator
- High-dimensional/hyper plane

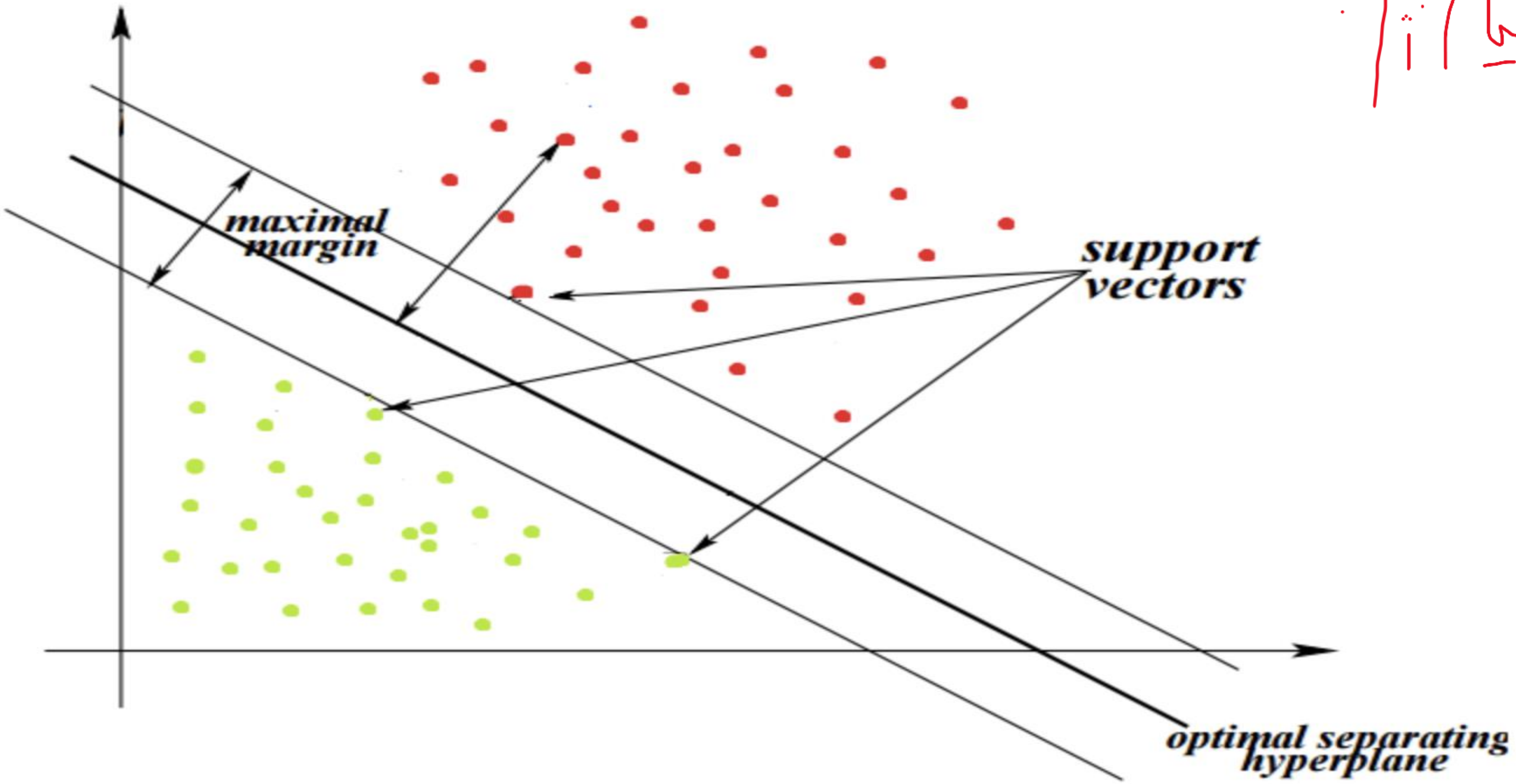

kernel
Decision surface

A hyperplane in $\mathbb{R}^2$ is a line



A hyperplane in $\mathbb{R}^3$ is a plane





source:**stackexchange.com**

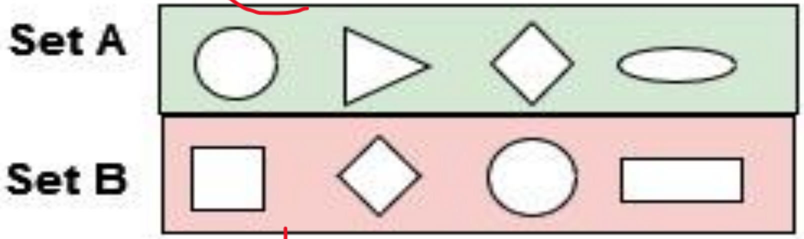
Other definitions

- Libraries
- Normalization
- Imbalanced Dataset

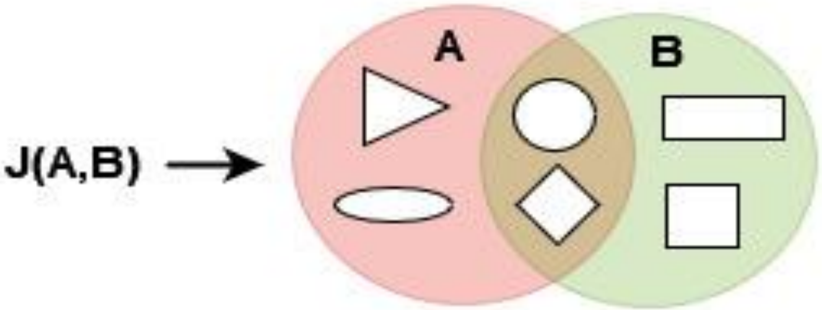
Evaluation

- Y(t)-Y(h) = evaluation
- Jaccard index
- F1_score/confusion matrix
- Log loss/Logarithmic Loss/Cross-Entropy Loss

Jaccard Similarity between two sets



Sets

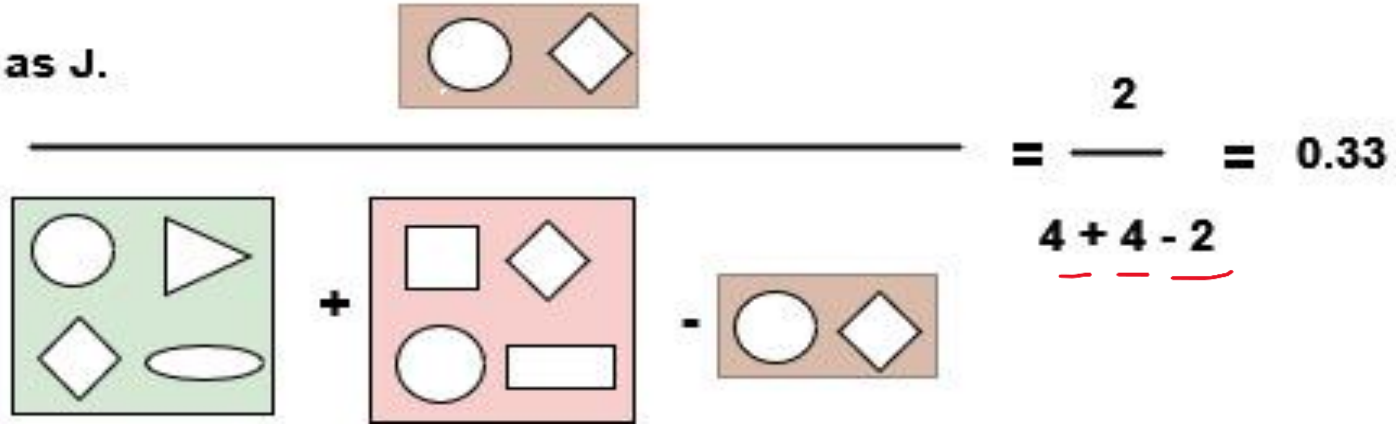


Venn Diagram

Jaccard Similarity/ Jaccard Index/ Jaccard Coefficient

J(A, B) = |A ∩ B| / |A ∪ B| = |A ∩ B| / |A| + |B| - |A ∩ B|

It is represented as J.



$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

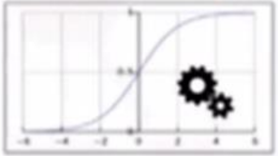
$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

		Predicted	
		+ Positive 1	Negative − 0
Actual	Positive + 1	**True positive**	**False negative**
	Negative − 0	**False positive**	**True negative**

Classification

KNN Evaluation / LogLoss

Performance of a classifier where the predicted output is a probability value between 0 and 1.



Test set

	tenure	age	address	income	ed	employ	equip	callcard	wireless	churn
0	11.0	33.0	7.0	136.0	5.0	5.0	0.0	1.0	1.0	1
1	33.0	33.0	12.0	33.0	2.0	0.0	0.0	0.0	0.0	1
2	23.0	30.0	9.0	30.0	1.0	2.0	0.0	0.0	0.0	0
3	38.0	35.0	5.0	76.0	2.0	10.0	1.0	1.0	1.0	0
4	7.0	35.0	14.0	80.0	2.0	15.0	0.0	1.0	0.0	0

Test

Actual Labels y

Predicted churn	LogLoss
0.91	0.11
0.13	2.04
0.04	0.04
0.23	0.26
0.43	0.56

1

0 0 0

0

$LogLoss = 0.60$

$\hat{y}$ Predicted Probability

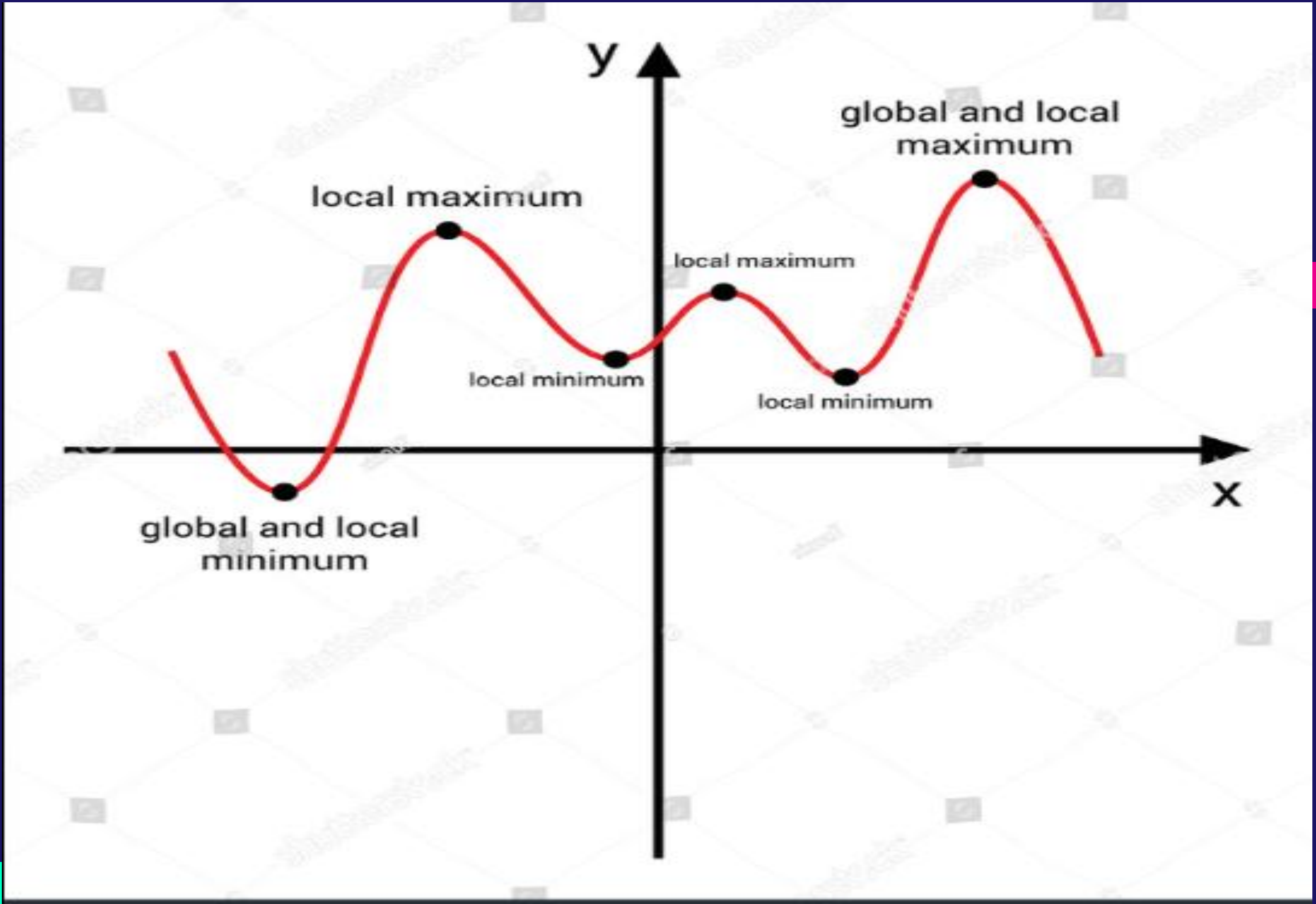
$$LogLoss = -\frac{1}{n}\sum(y \times \log(\hat{y}) + (1-y) \times \log(1-\hat{y}))$$

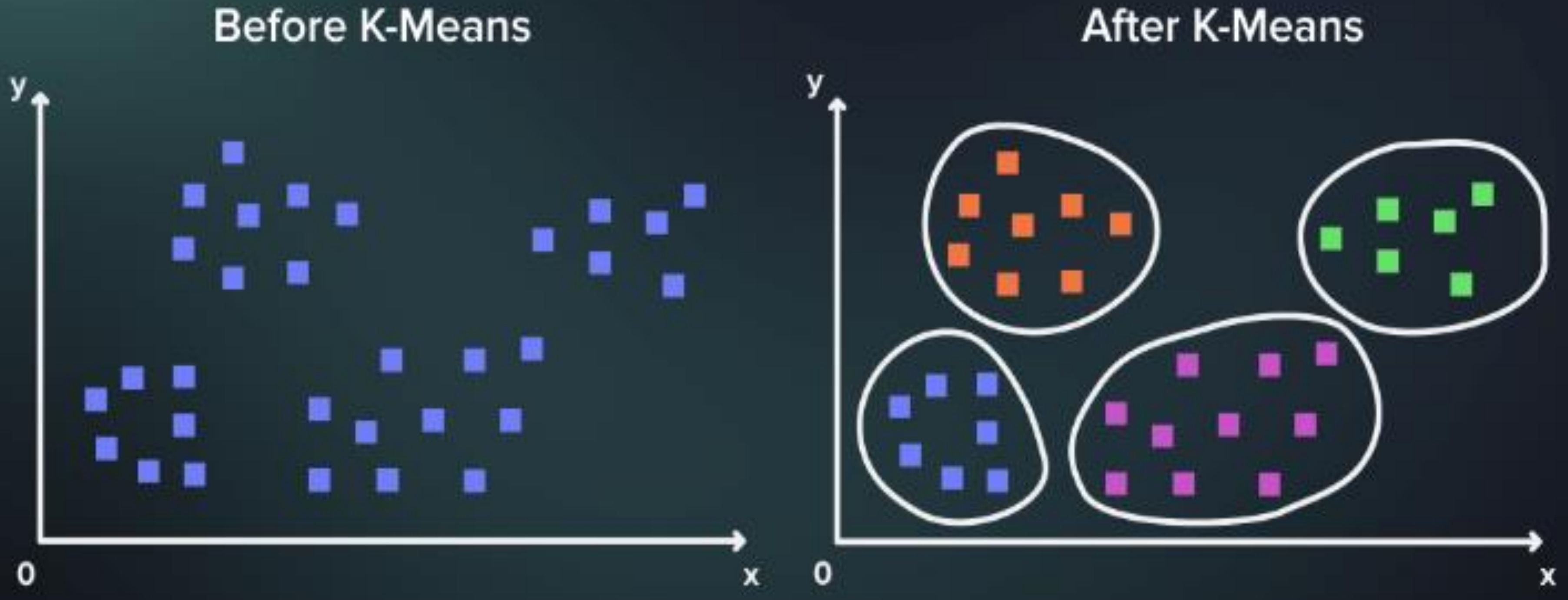
Clustering intro

- Unsupervised
- Partitioning / clustering
- Finding clusters in datasets (data pointents)
- There is no Y !
- prediction is not the goal
- Retail/marketing/banking/insurance/biology

K-means

- Unsupervised
- Divides data into K non-overlapping subset
- There is no Y !
- Understand the similarity and dissimilarity
- Distance(x1,x2) & distance(clu1,clu2)
- K=centroid / elbow method


y
global and local
maximum
local maximum
local maximum
local minimum
local minimum
x
global and local
minimum



Before K-Means
After K-Means
y
0
x
y
0
x

Hierarchical

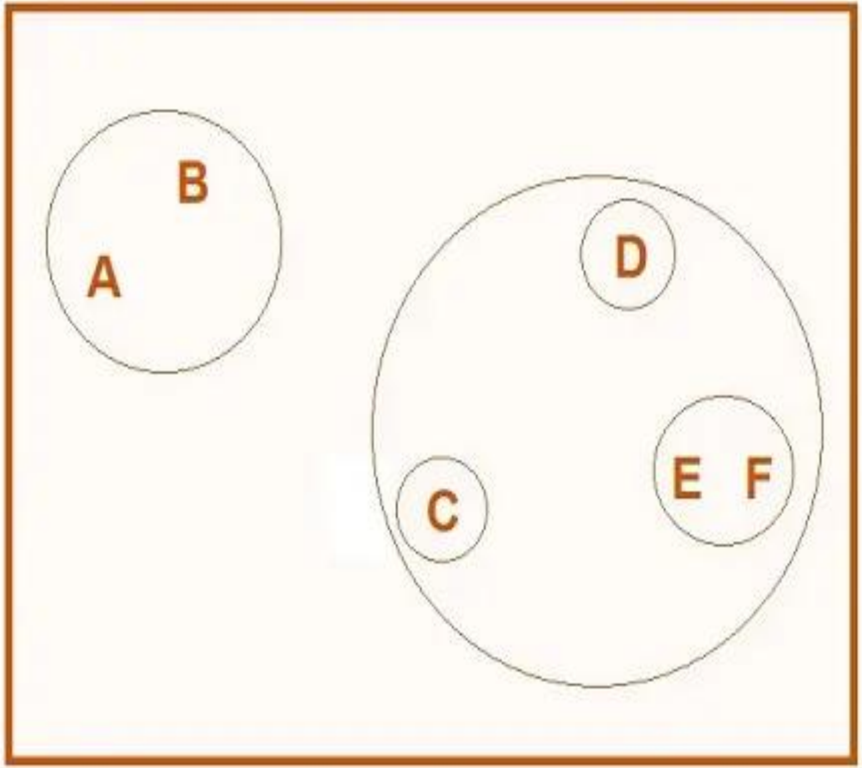
- Unsupervised
- Agglomerative(bottom up) / divisive
- Distance matrix / Similarity matrix

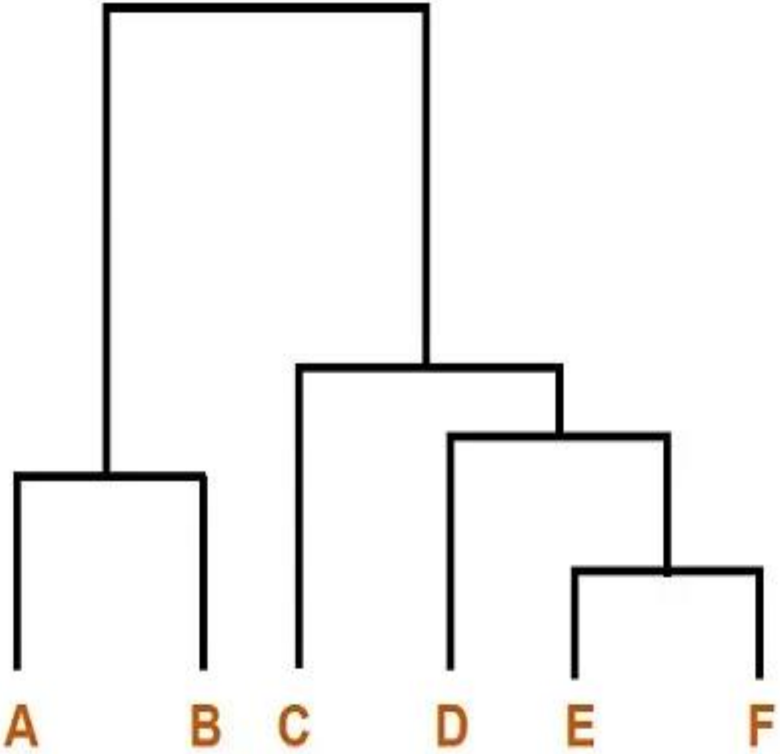
Upper Triangular Matrix

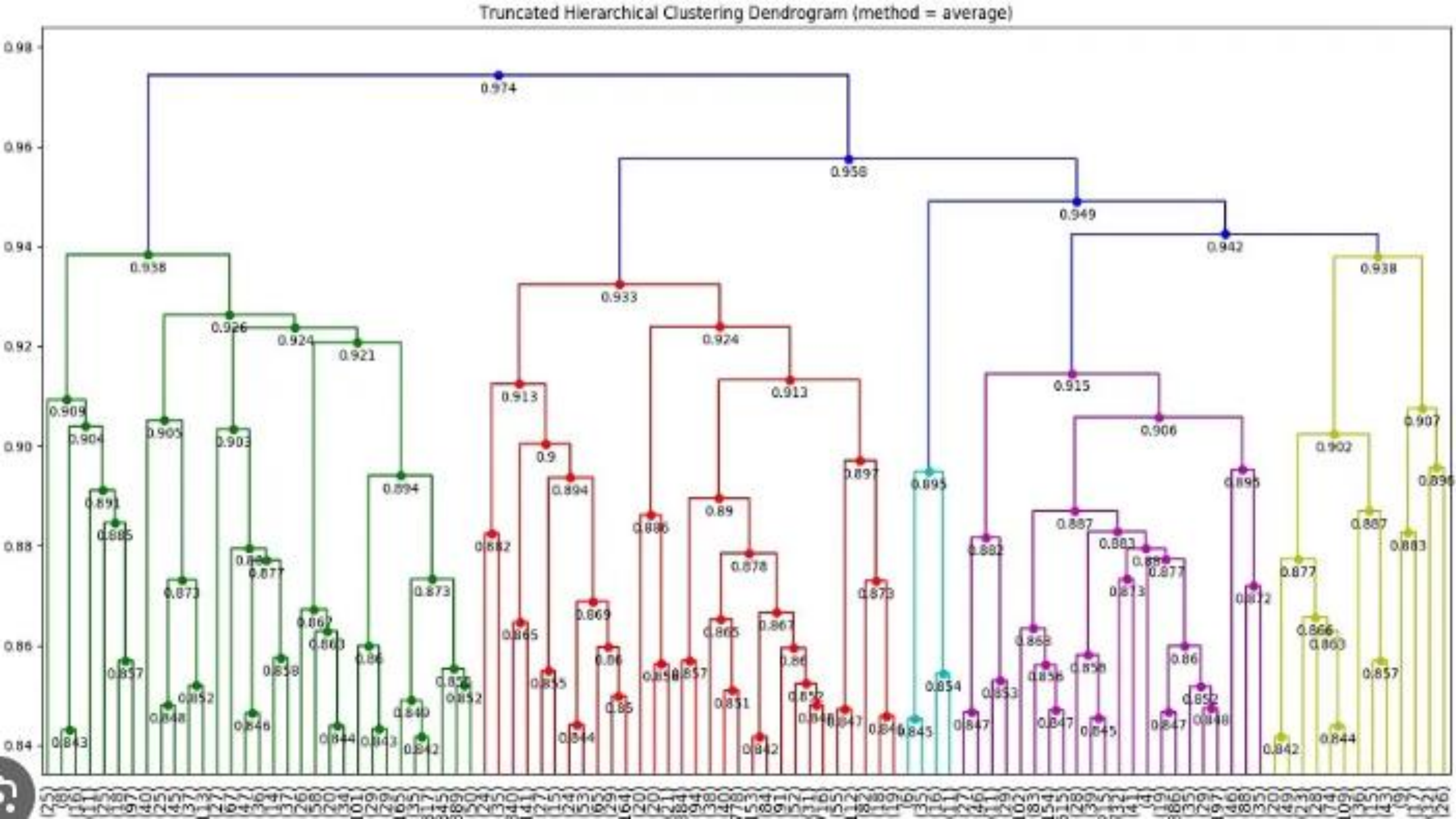
Lower Triangular Matrix

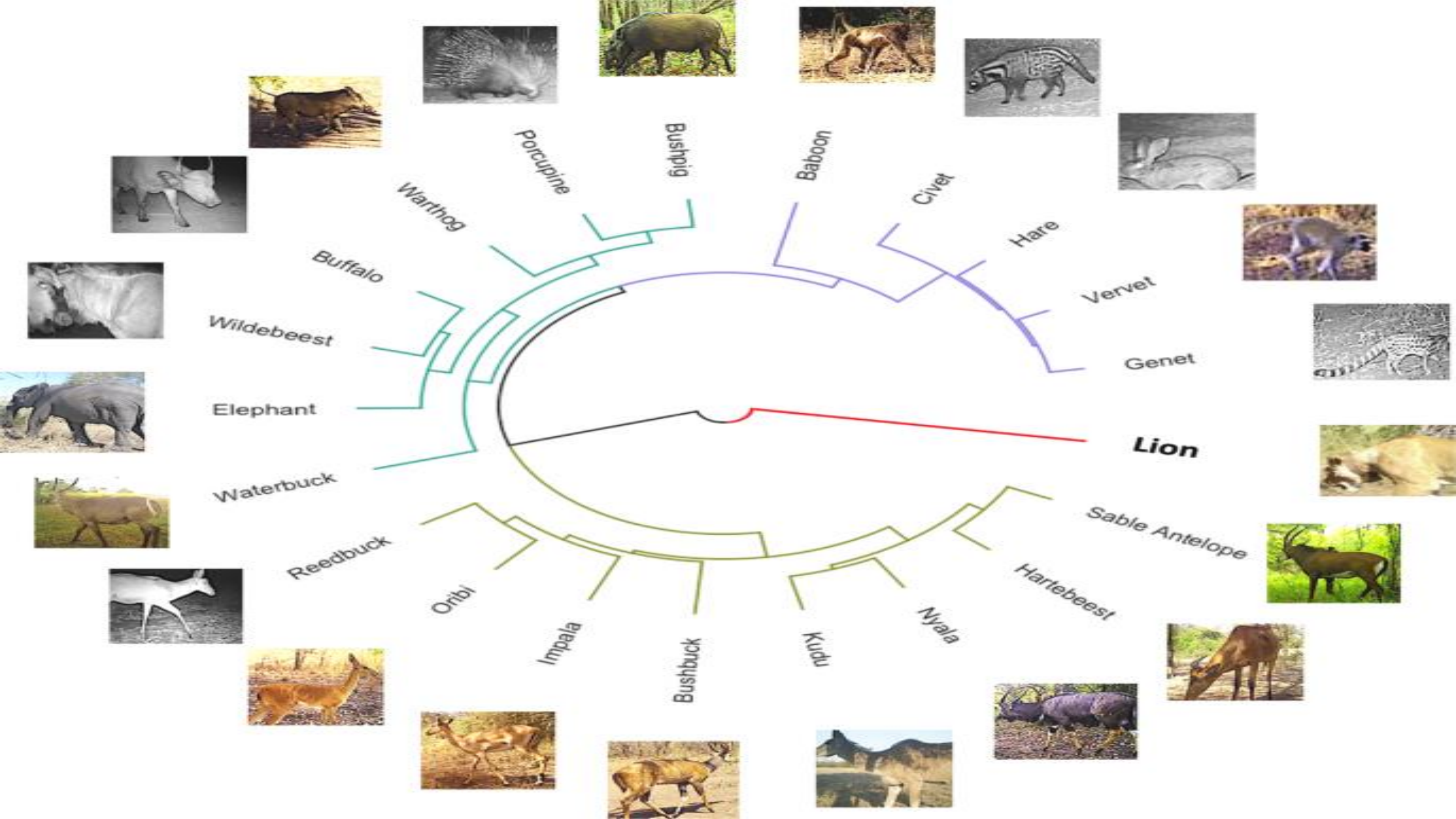
$$U = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}_{4\times 4} \quad L = \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}_{4\times 4}$$

Grouping Similar Clusters → Dendrogram






Truncated Hierarchical Clustering Dendrogram (method = average)
0.974
0.958
0.949
0.942
0.938
0.933
0.926
0.924
0.921
0.938
0.915
0.913
0.913
0.906
0.909
0.905
0.903
0.907
0.902
0.897
0.895
0.895
0.894
0.894
0.9
0.89


Porcupine
Bushpig
Baboon
Civet
Hare
Vervet
Genet
Warthog
Buffalo
Wildebeest
Elephant
Waterbuck
Lion
Sable Antelope
Hartebeest
Nyala
Kudu
Bushbuck
Impala
Oribi
Reedbuck

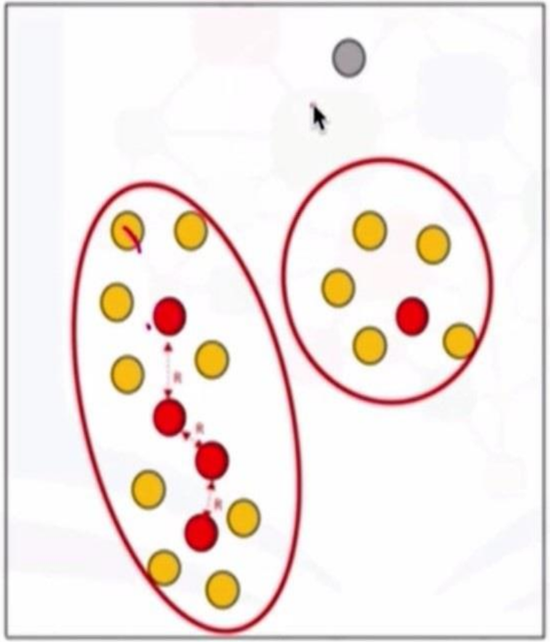
DBSCAN

- Unsupervised
- Density-based spatial clustering of applications with noise
- K-means = assign every data point to a cluster no outlier/always spherical shape clusters
- Good for anomaly detection
- density-based

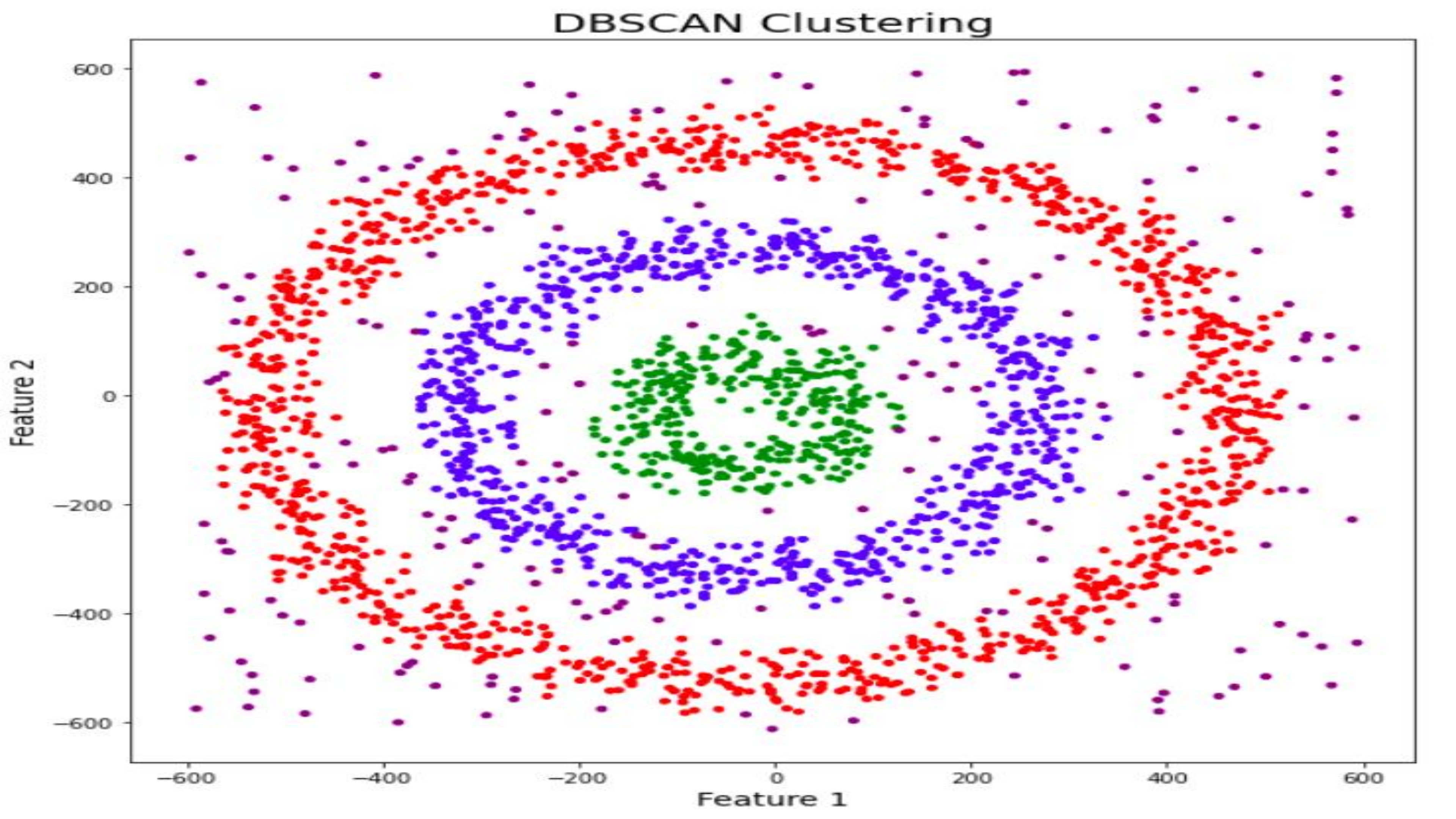
Clustering

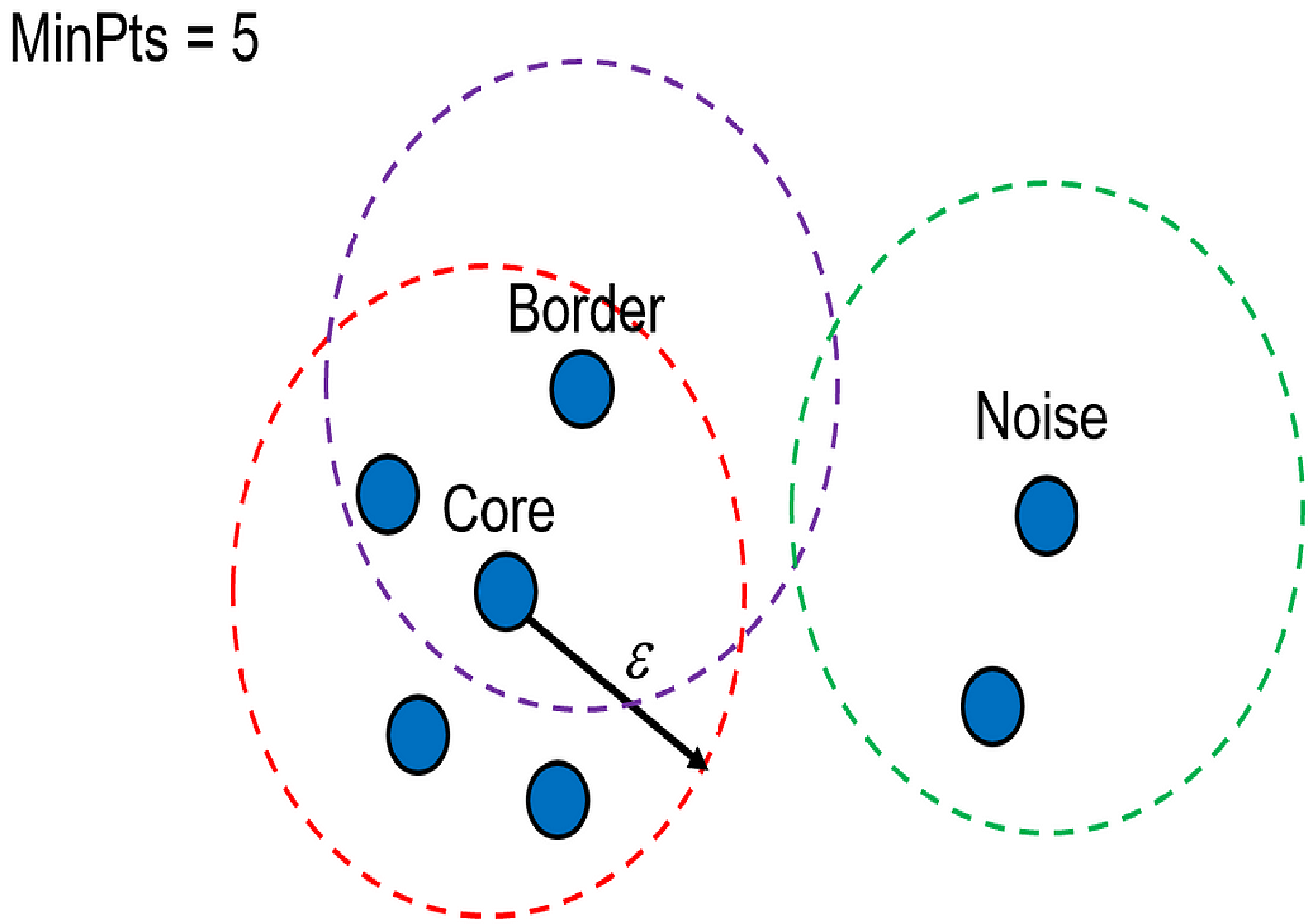
DBSCAN

- point types:
  - core: within our neighborhood of the point there are at least M points.
  - Border:
    - less than M in neighborhood
    - reachable from a core point
  - outlier is not core neigher a border



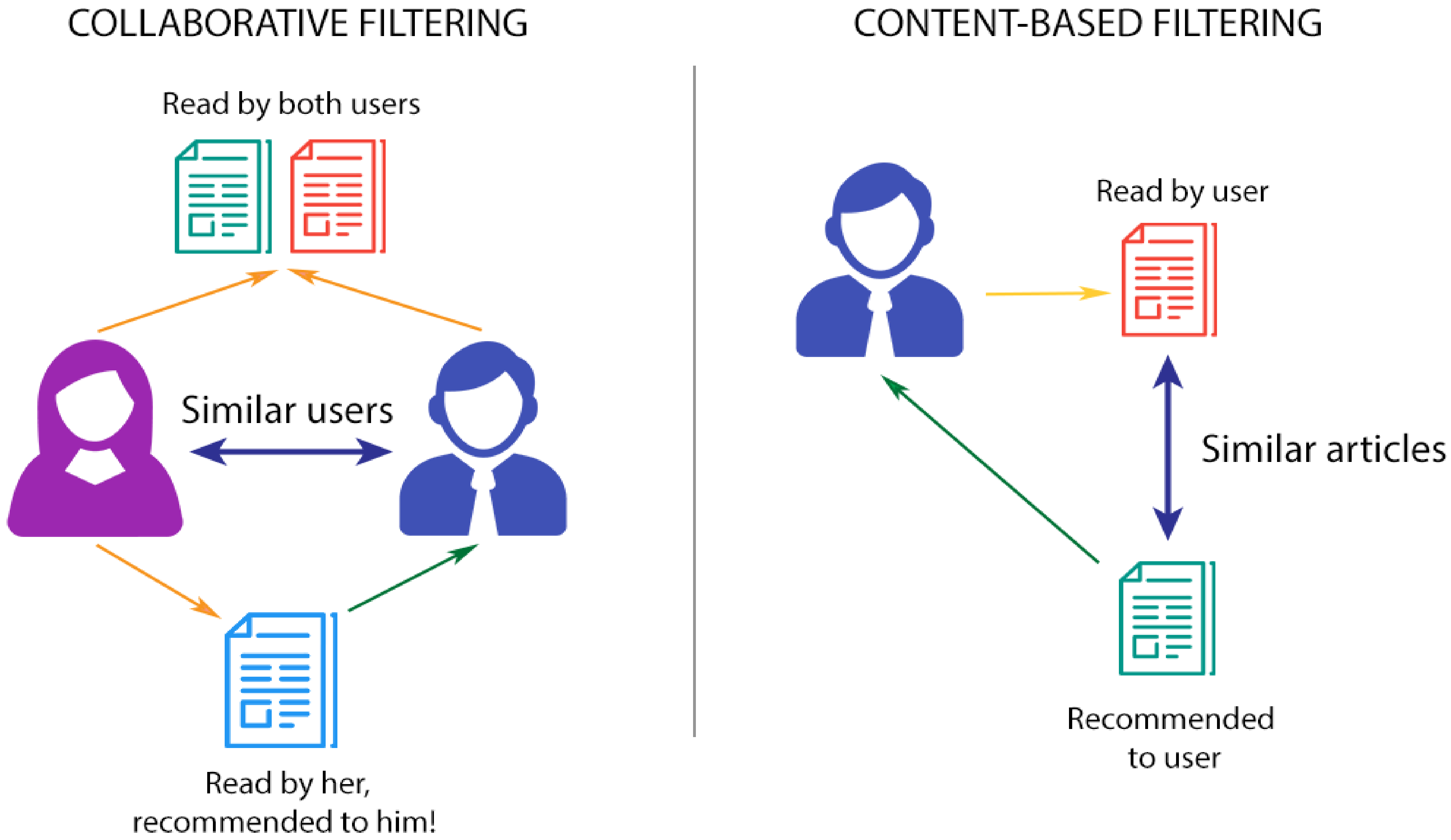
R = 2unit , M = 6

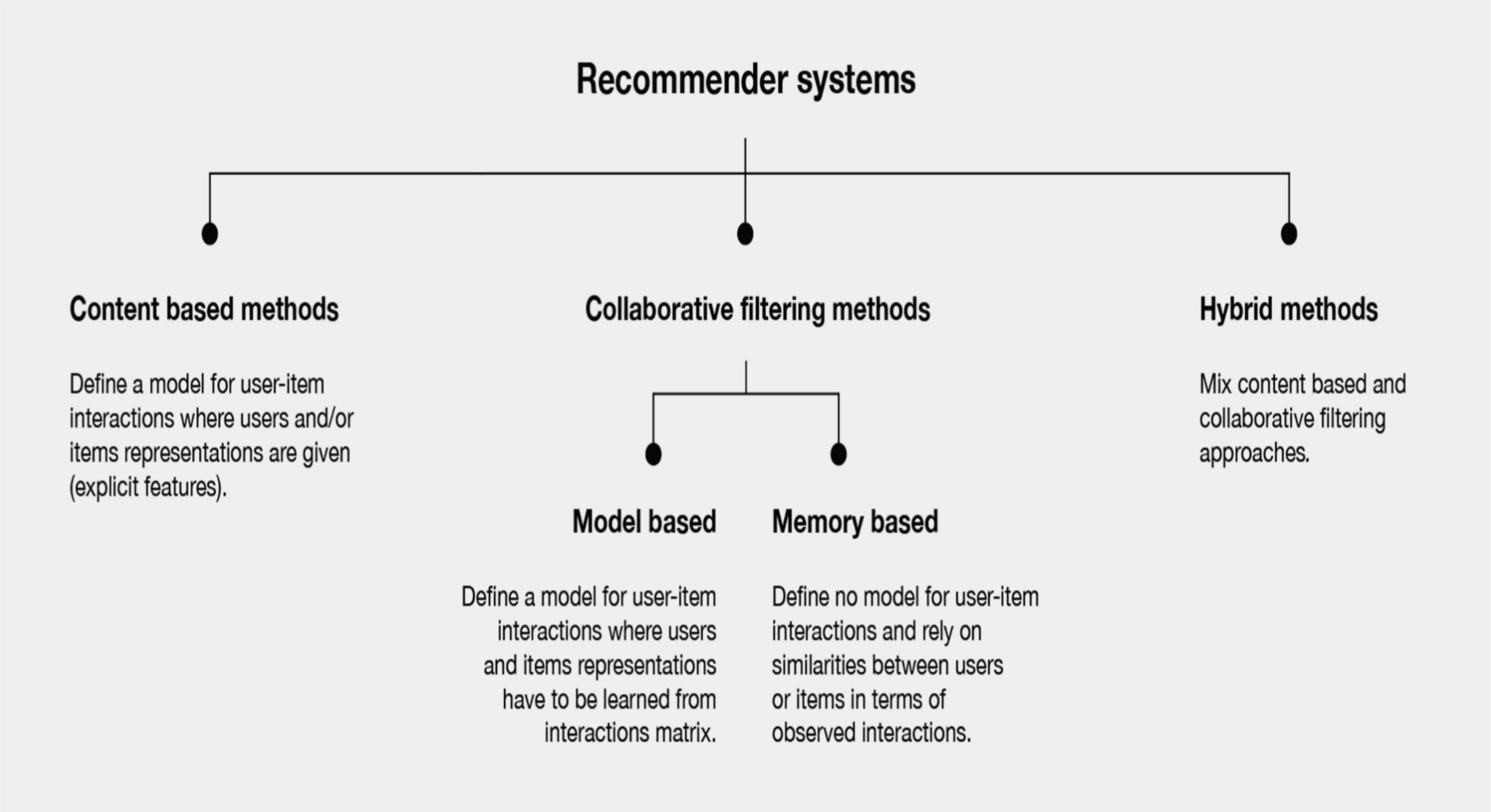

DBSCAN Clustering
Feature 2
Feature 1
600
400
200
0
−200
−400
−600
−600
−400
−200
0
200
400
600


MinPts = 5
Border
Core
ε
Noise

Recommenders intro

- Behavioral patterns
- Recommender system capture the pattern of people's behavior and use it to predict what else they might want or like
- Amazon / Digikala / Spotify / Netflix / Snapfood


COLLABORATIVE FILTERING
Read by both users
Similar users
Read by her,
recommended to him!
CONTENT-BASED FILTERING
Read by user
Similar articles
Recommended
to user


Recommender systems
Content based methods
Define a model for user-item interactions where users and/or items representations are given (explicit features).
Collaborative filtering methods
Model based
Define a model for user-item interactions where users and items representations have to be learned from interactions matrix.
Memory based
Define no model for user-item interactions and rely on similarities between users or items in terms of observed interactions.
Hybrid methods
Mix content based and collaborative filtering approaches.

Recommenders / content based

- Works based user profiles(like , view ,…

Recommenders / collaborative

- Based on the user's similarity or neighborhoods
- Finds similarity between users
- Based on items similarity
- Item based / user based

Collaborative Filtering VS Content-based Filtering

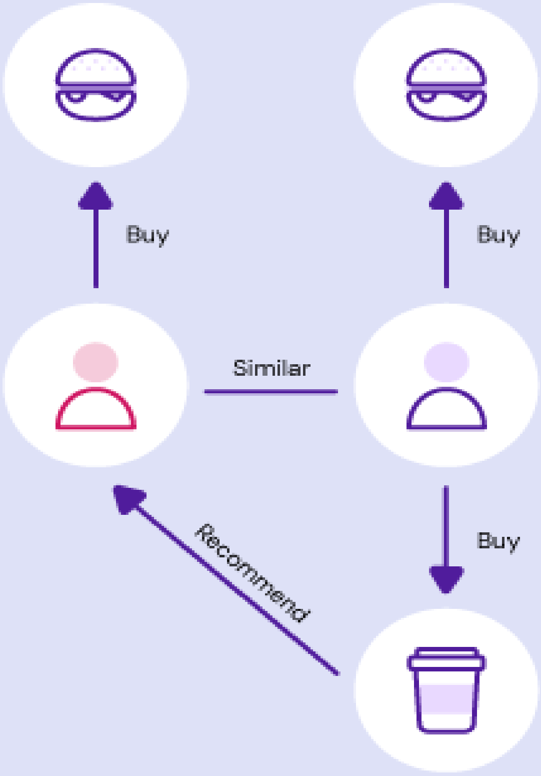
Collaborative Filtering

- Uses past interactions to recommend new items
- Item features are not required

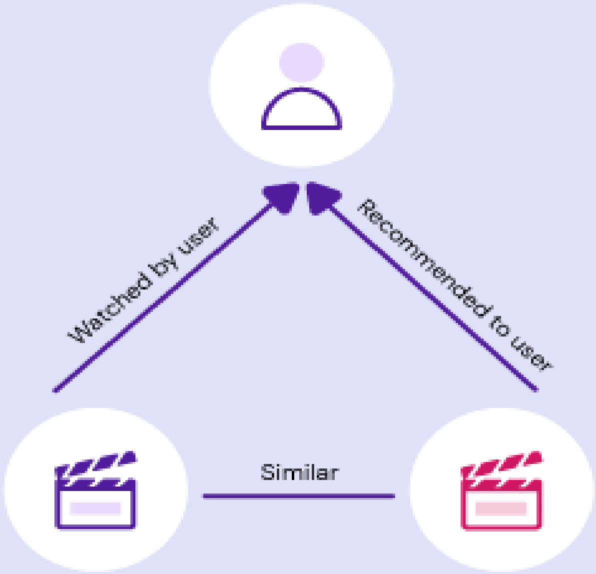
Content-based Filtering

- Uses ML algorithms to predict and recommend new items
- Item features are used to group similar items

Collaborative Filtering



Content-Based Filtering



ALTERNATIVE RESOURCES

PHOTOS:

- Close up of hacker
- Teacher talking with his students online
- Hacking concept
- Young woman enjoying new technologies
- Close up of hacker
- Man using laptop in cafe

RESOURCES

Did you like the resources on this template?
Get them for free at our other websites.

VECTORS:

- Landing page template with programming concept
- Development Icon Pack
- Realistic multimedia player
- Concept of flat computer engineering