BCSE209L- Machine Learning
Module 2

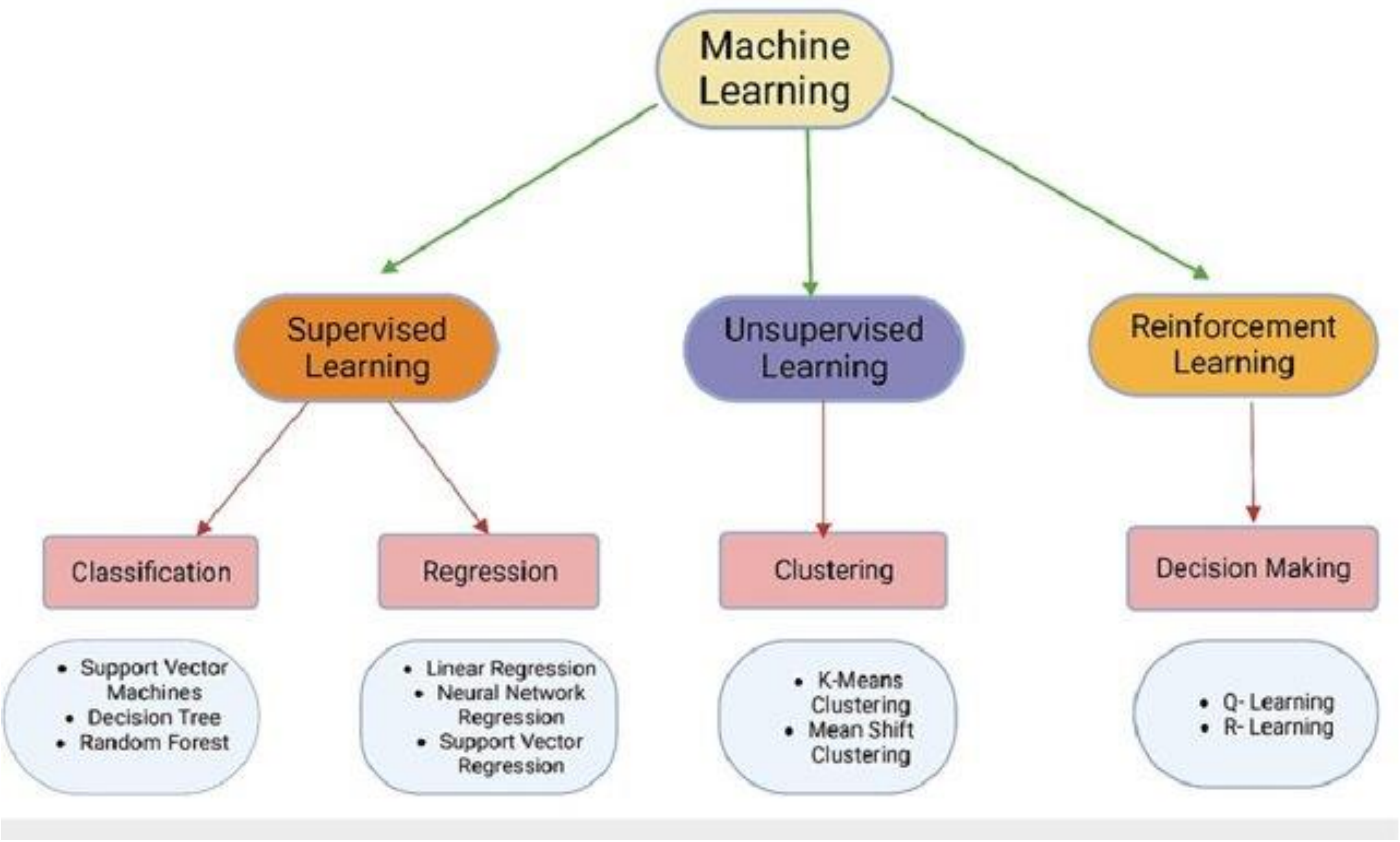
Dr. G. Praveen Kumar
SMEC- Spl in AI& ML

Context :

Linear and Non-Linear examples – Multi–Class & Multi-Label classification –

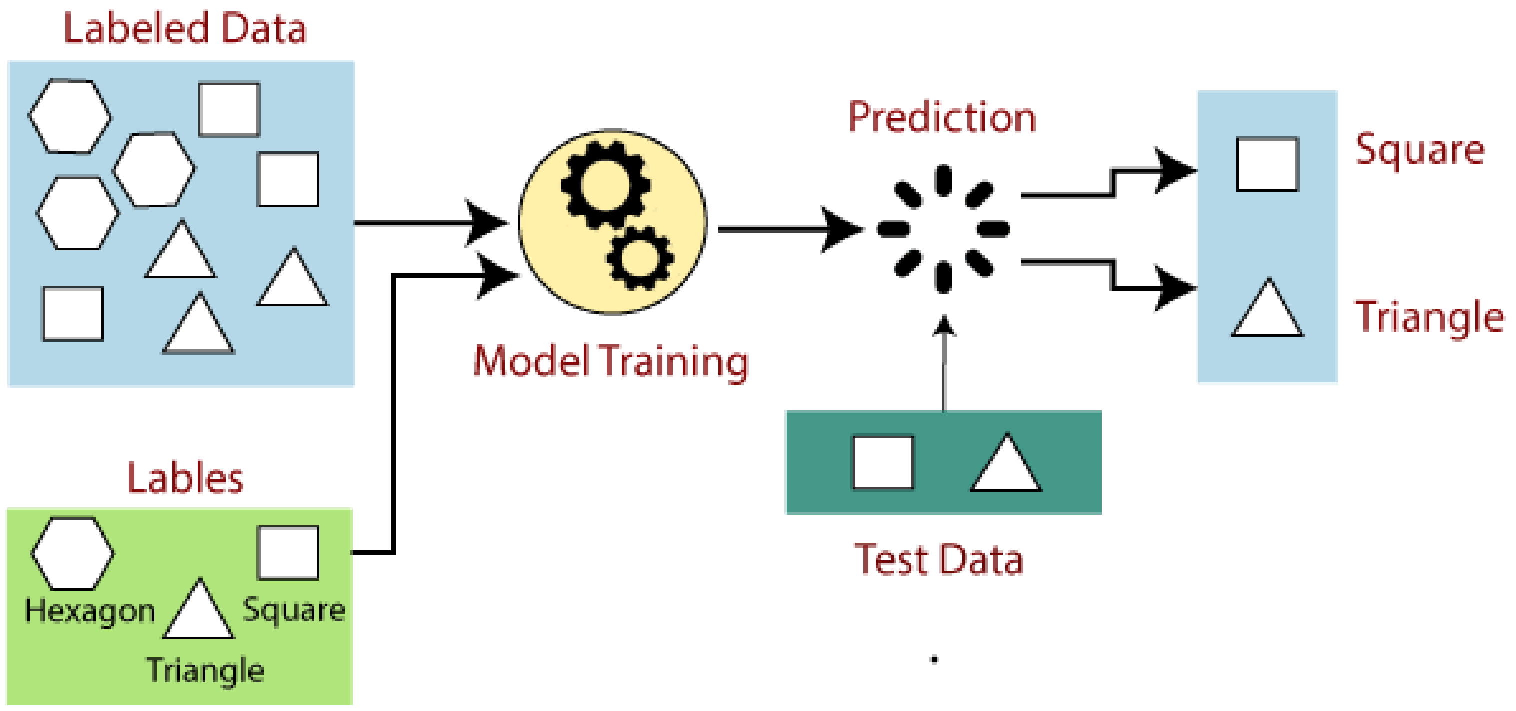
Linear Regression – Multiple Linear Regression – Naïve Bayes Classifier – Decision Trees – ID3 – CART – Error bounds.

Introduction to Machine Learning

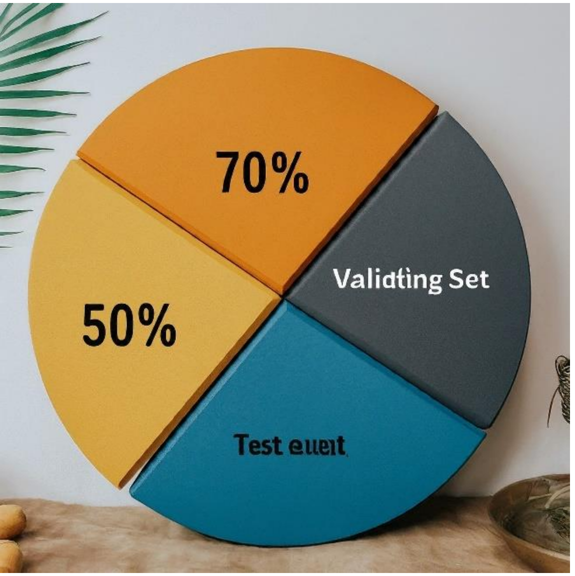


Introduction to Machine Learning

- Supervised learning is a category of machine learning that uses labeled datasets to train algorithms to predict outcomes and recognize patterns.
- Unlike unsupervised learning, supervised learning algorithms are given labeled training to learn the relationship between the input and the outputs.



Introduction to Supervised Learning



1. Gather labeled data (cat/not cat) images to train your program.
2. Label each image (cat or not cat) to help the program learn the difference.
3. Split data into training and test sets (train the program, test its learning).
4. Choose a learning method (like decision trees) to help find patterns in data.
5. Train the program on labeled data (like showing it cat/not cat pictures).
6. Test the program on unseen data (see if it can identify cats in new pictures).

Linear Examples

1. Linear Equation in One Variable:

   $y = 3x + 5$

2. Linear Equation in Two Variables:

   ax + by + c = 0

Eg: Coffee: Find the amount of water to add to coffee grounds to get a desired final brew volume and strength.

3. Linear Function Representing Direct Proportionality:

   $y = kx$

   (where k is a constant)

4. Linear Cost Function:

   $C = 50 + 20x$

   (where C is the total cost, 50 is a fixed cost, and $20x$ is the variable cost dependi number of units x)

5. Linear Temperature Conversion:

   $F = \frac{9}{5}C + 32$

   (where F is the temperature in Fahrenheit and C is the temperature in Celsius)

6. Linear Depreciation:

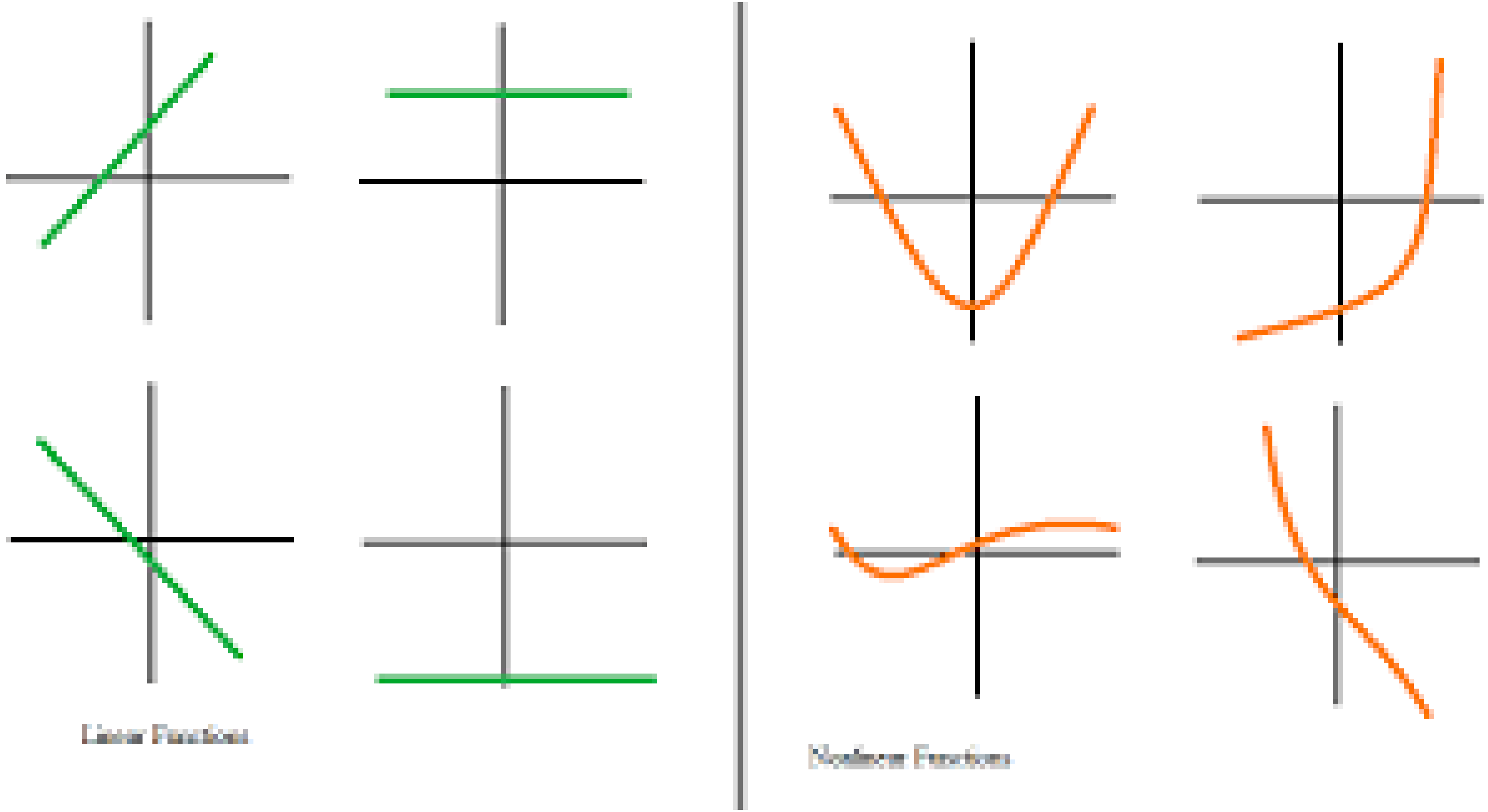
   $V = P - Dt$

   (where V is the value of an asset, P is the initial value, D is the depreciation per time period, and t is time)

Non-Linear Examples

1. Quadratic Function:
   $y = x^2 + 2x + 1$
2. Cubic Function:
   $y = x^3 - 3x^2 + 2x$
3. Exponential Function:
   $y = e^x$
4. Logarithmic Function:
   $y = \log(x)$
5. Trigonometric Function (Sine):
   $y = \sin(x)$
6. Inverse Function:
   $y = \frac{1}{x}$
7. Power Function:
   $y = x^n$
   (where n is a non-linear exponent, e.g., $y = x^{3/2}$)
8. Non-Linear Growth (Logistic Function):
   $y = \frac{L}{1+e^{-k(x-x_0)}}$
   (where L is the curve's maximum value, k is the steepness of the curve, and x_0 is the x-value of the sigmoid's midpoint)
9. Polynomial Function (Degree greater than 1):
   $y = 4x^5 - 3x^3 + 2x^2 - x + 6$
10. Rational Function:
    $y = \frac{x^2+3}{x-2}$

Introduction to Machine Learning



Introduction to Machine Learning

Linear Machine Learning Algorithms

1. **Linear Regression**: Predicting a continuous value (like house price) based on a linear relationship with other features (like square footage).
2. **Logistic Regression**: Classifying data points into two categories (like spam/not spam email) based on a linear decision boundary.
3. **Linear Discriminant Analysis (LDA)**: Used for classification problems with two or more classes.
4. **Principal Component Analysis (PCA):** Reducing the dimensionality of data by finding linear combinations that capture most of the variance.
5. **K-Means Clustering:** Grouping data points into a specific number of clusters (k) based on their linear distance from the cluster centroid. (Note: While K-Means itself uses linear distances, it can be applied for non-linear data distributions)
6. **Support Vector Machine (SVM) for Linear Classification:** Separating data points belonging to different classes (like cats vs dogs in images) with a hyperplane in a linear fashion.

A linear classifier achieves this by making a classification decision based on the value of a linear combination of the characteristics.

A classification algorithm (Classifier) that makes its classification based on a linear predictor function combining a set of weights with the feature vector

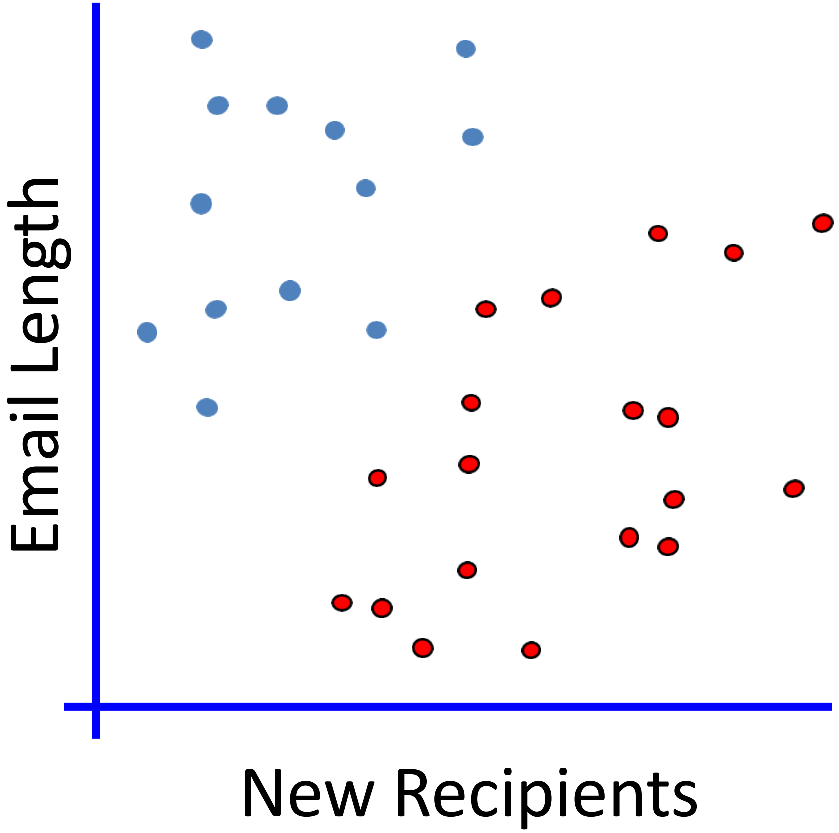
$$y = f(\vec{w} \cdot \vec{x}) = f\left(\sum_j w_j x_j\right),$$

Decision boundaries is flat

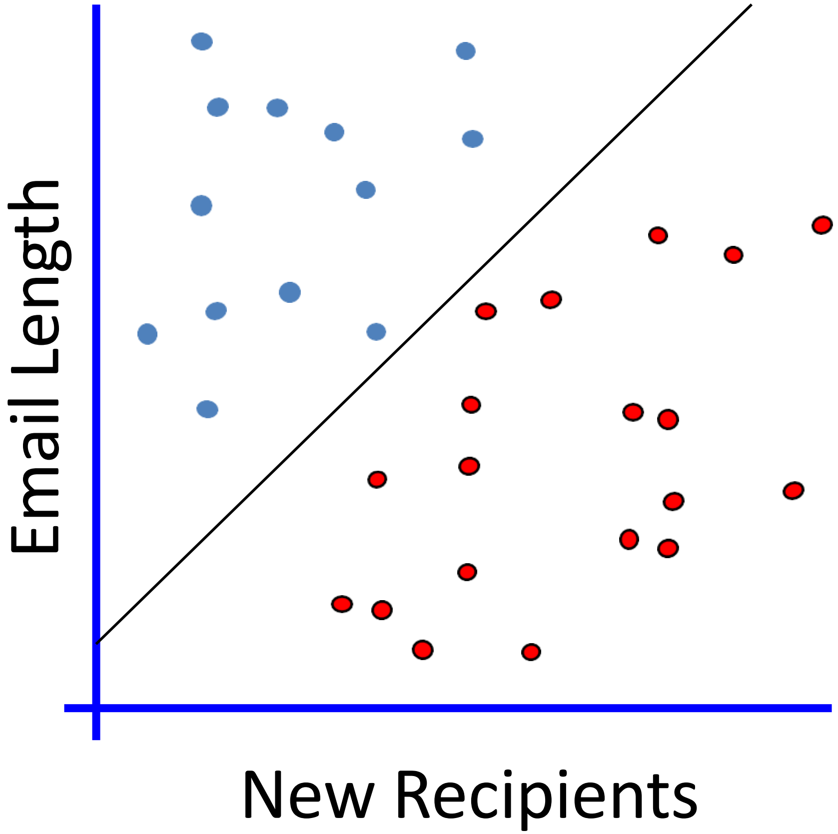
- Line, plane, ….

May involve non-linear operations

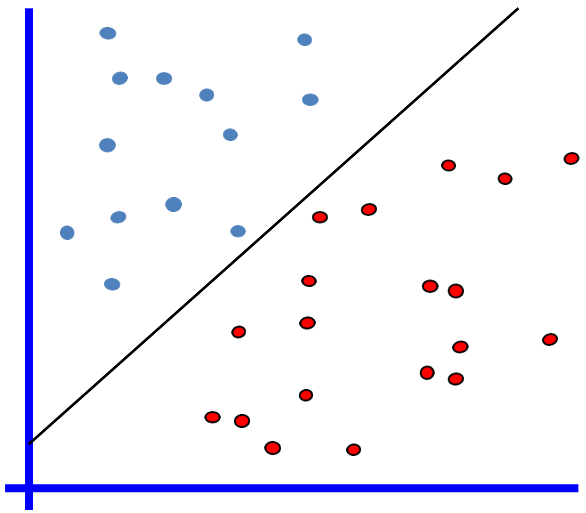
Linear Binary Examples

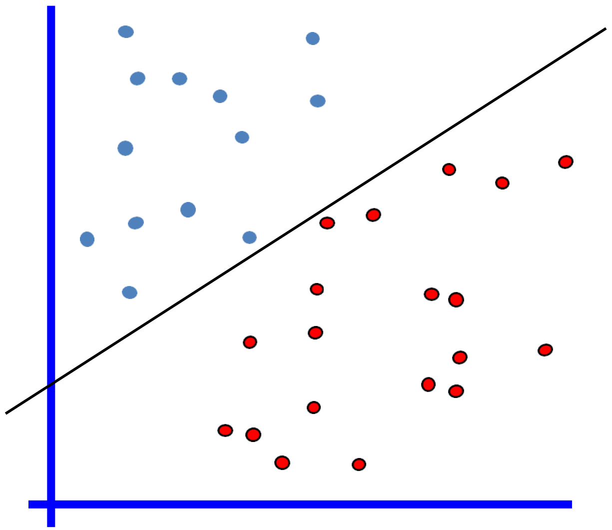


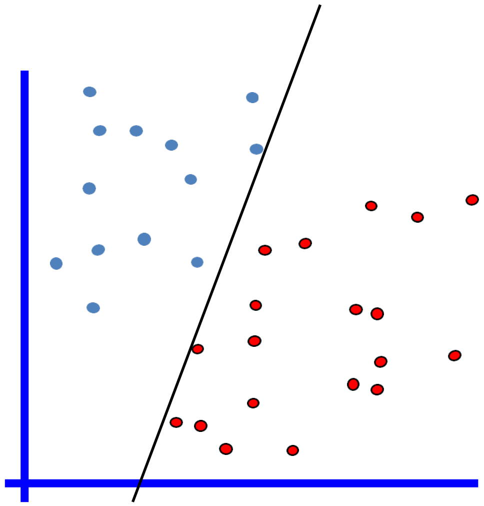
How would you classify this data?



Linear Binary Examples





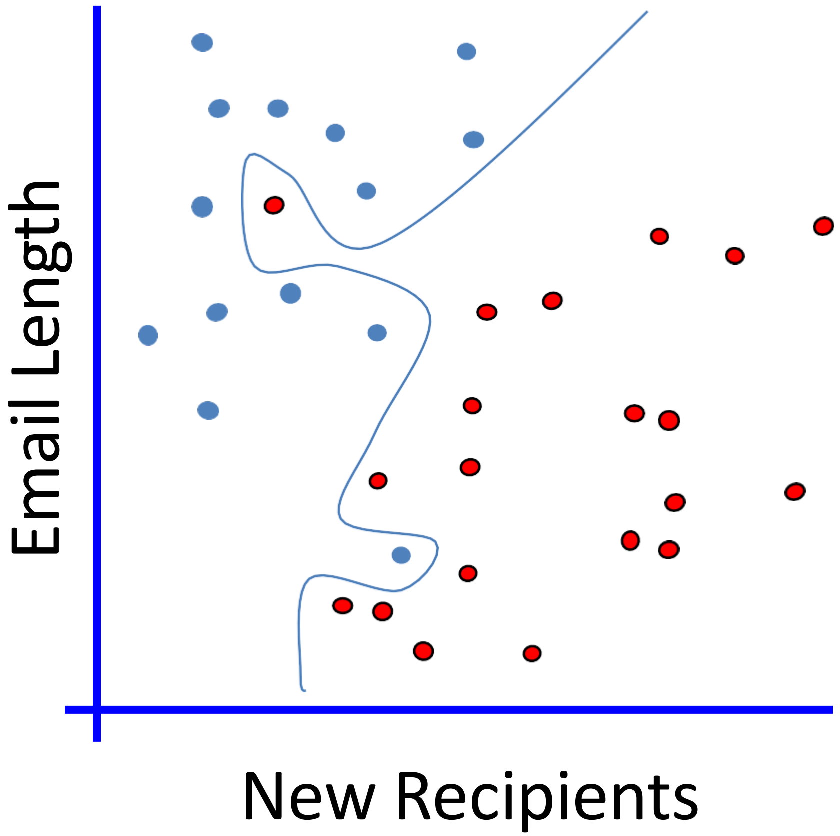


Non-Linear Examples

Non-Linear Machine Learning Algorithms

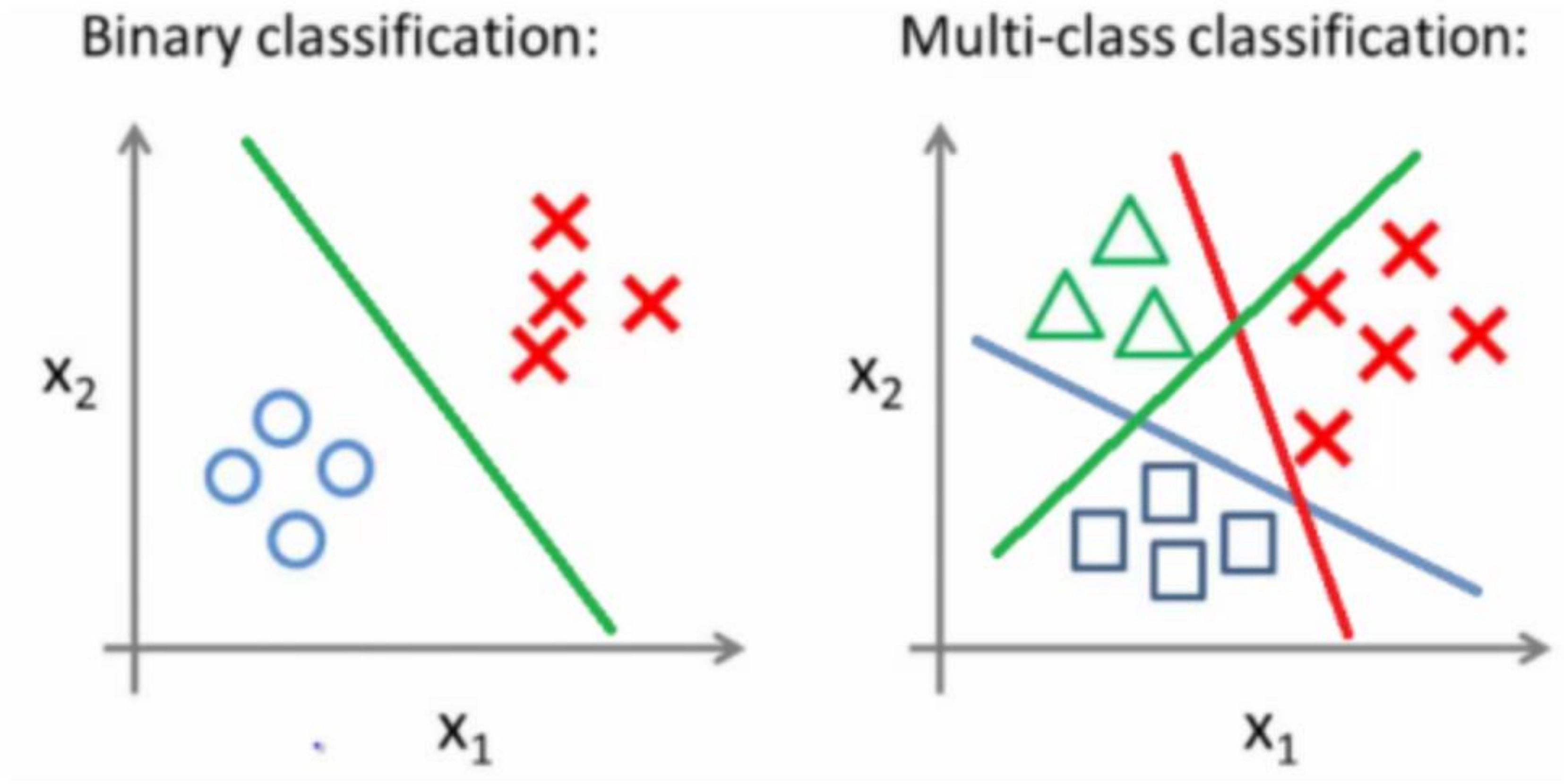
1. **Decision Trees**: A model that makes decisions based on the feature values.
2. **Random Forest**: An ensemble of decision trees to improve prediction accuracy.
3. **Support Vector Machines (SVM) with Non-Linear Kernels (RBF, Polynomial)**: Used for complex classification tasks.
4. **Neural Networks**: Used for both regression and classification tasks with non-linear relationships.
5. **K-Nearest Neighbors (KNN)**: A non-parametric method used for classification and regression.
6. **Gradient Boosting Machines (GBM)**: An ensemble technique that builds models sequentially.
7. **XGBoost**: An optimized version of gradient boosting that performs well in competitions.
8. **Multilayer Perceptron (MLP)**: A type of neural network with multiple layers.
9. **Artificial Neural Networks (ANNs):** Learning complex, non-linear relationships between features and outputs through interconnected layers of artificial neurons. (This includes popular architectures like Convolutional Neural Networks (CNNs) for image recognition and Recurrent Neural Networks (RNNs) for sequence data)

Non-Linear Binary Examples



Binary Vs Multi-class

Multi-class classification refers to a type of classification problem where an instance can belong to one of three or more classes. Each instance is assigned exactly one label from the set of possible classes.

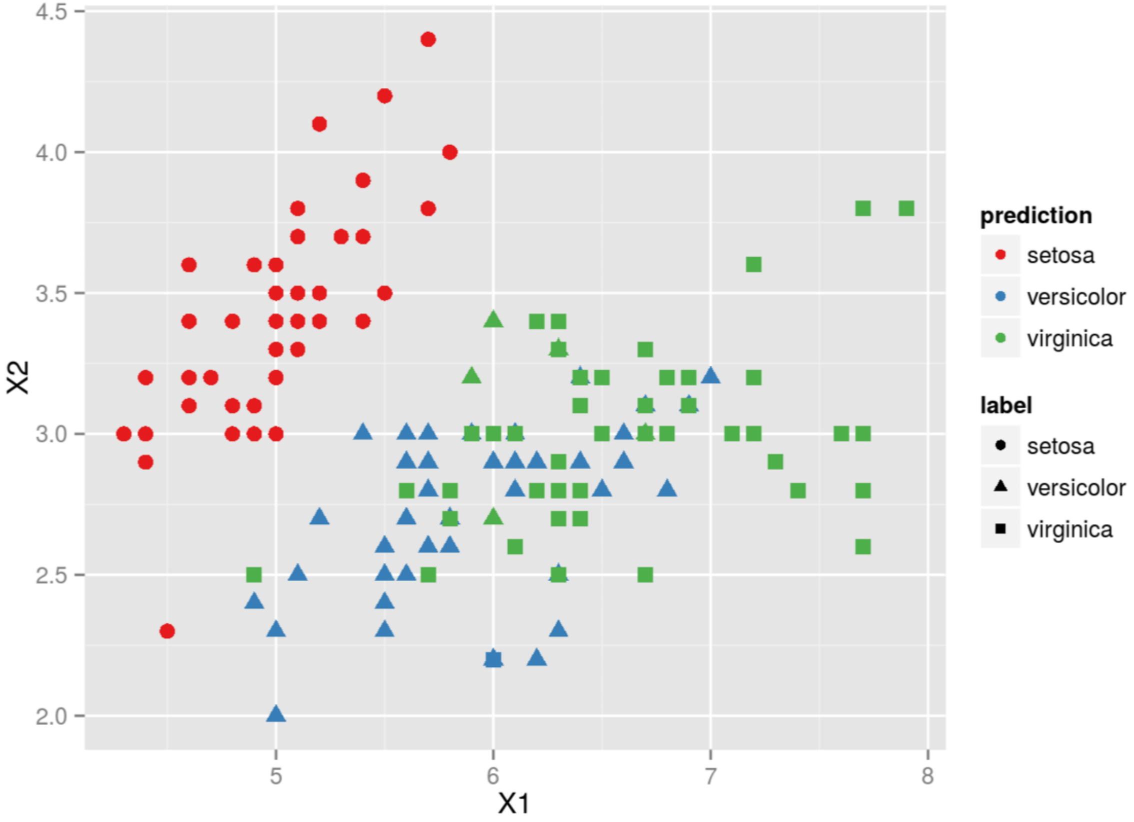


Multi-Class Classification

Output $\in \{1,2,3, \ldots K\}$

In some cases, output space can be very large (i.e., K is very large)

Each input belongs to exactly one class
(c.f. in multilabel, input belongs to many classes)



SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
6.8	3.2	5.9	2.3	Iris-virginica
6.9	3.1	5.1	2.3	Iris-virginica
4.9	3.0	1.4	0.2	Iris-setosa
5.6	3.0	4.5	1.5	Iris-versicolor
4.8	3.1	1.6	0.2	Iris-setosa
5.8	2.8	5.1	2.4	Iris-virginica
7.2	3.6	6.1	2.5	Iris-virginica
5.1	3.5	1.4	0.3	Iris-setosa
4.7	3.2	1.6	0.2	Iris-setosa
6.6	3.0	4.4	1.4	Iris-versicolor

Multi-Class Classification

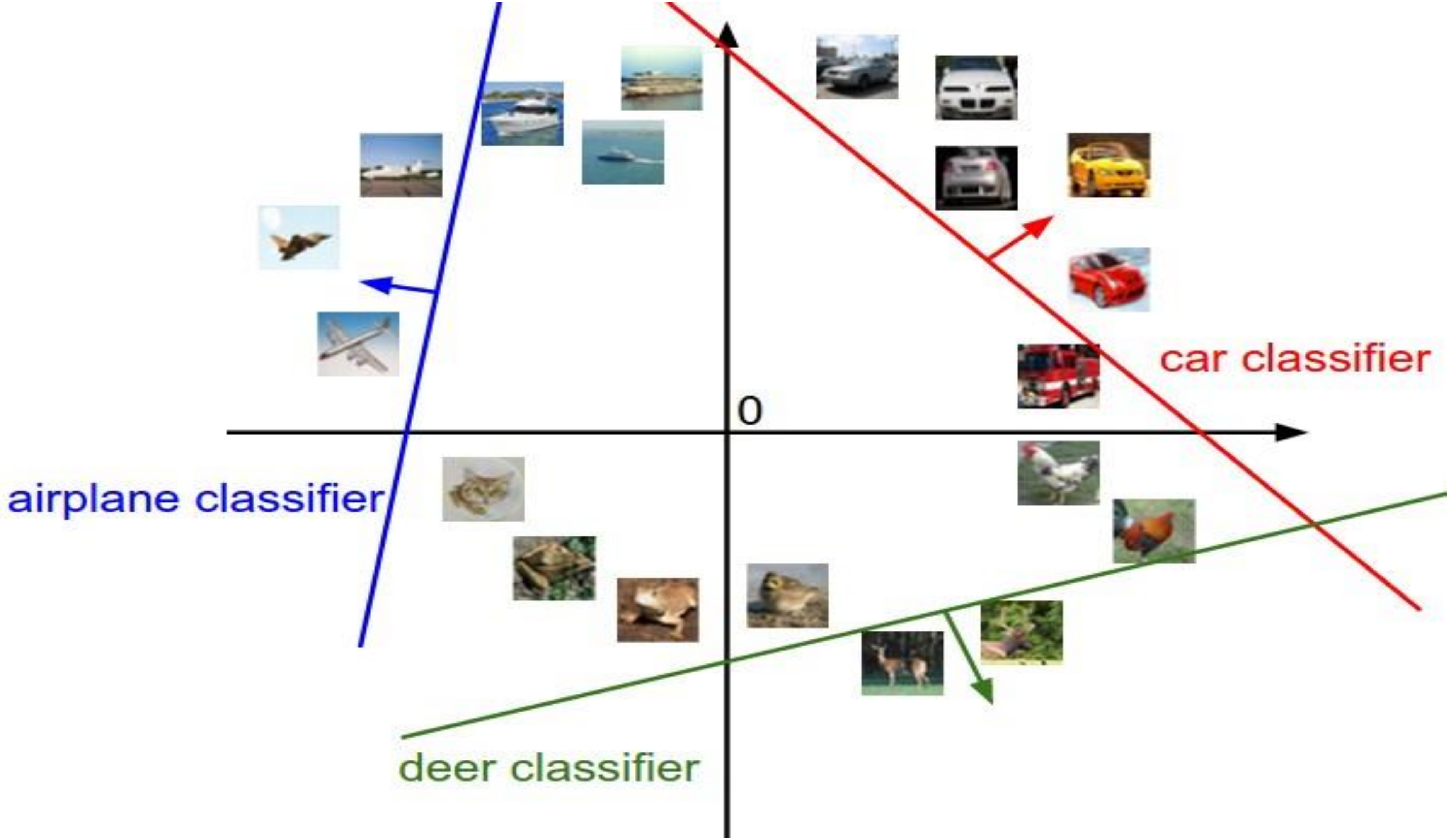
Multi-class classification is simply classifying objects into any one of multiple categories. Such as classifying just into either a dog or cat from the dataset.

1. When there are more than two categories in which the images can be classified, and

2. An image does not belong to more than one class

If both of the above conditions are satisfied, it is referred to as a multi-class image classification problem

Multi-Class Classification



Multi-label classification

When we can classify an image into more than one class (as in the image beside), it is known as a multi-label image classification problem.

Multi-label classification is a type of **classification** in which an object can be categorized into more than one class.

For example, In the image dataset, we will **classify** a picture as the **image** of a dog or cat and also **classify** the same **image** based on the breed of the dog or cat

These are all labels of the given images. **Each image here belongs to more than one class and hence it is a multi-label image classification problem.**

Multi-label classification

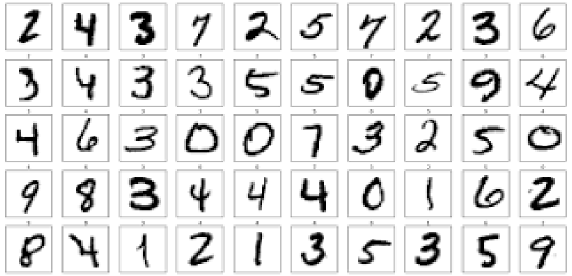
Binary Classification

•Only two class instances are present in the dataset.

•It requires only one classifier model.

•Confusion Matrix is easy to derive and understand.

•Example:- Check email is spam or not, predicting gender based on height and weight.



Multi-class Classification

•Multiple class labels are present in the dataset.

•The number of classifier models depends on the classification technique we are applying to.

•One vs. All:- **N-class instances** then **N binary classifier models**

•One vs. One:- **N-class instances** then **N* (N-1)/2 binary classifier models**

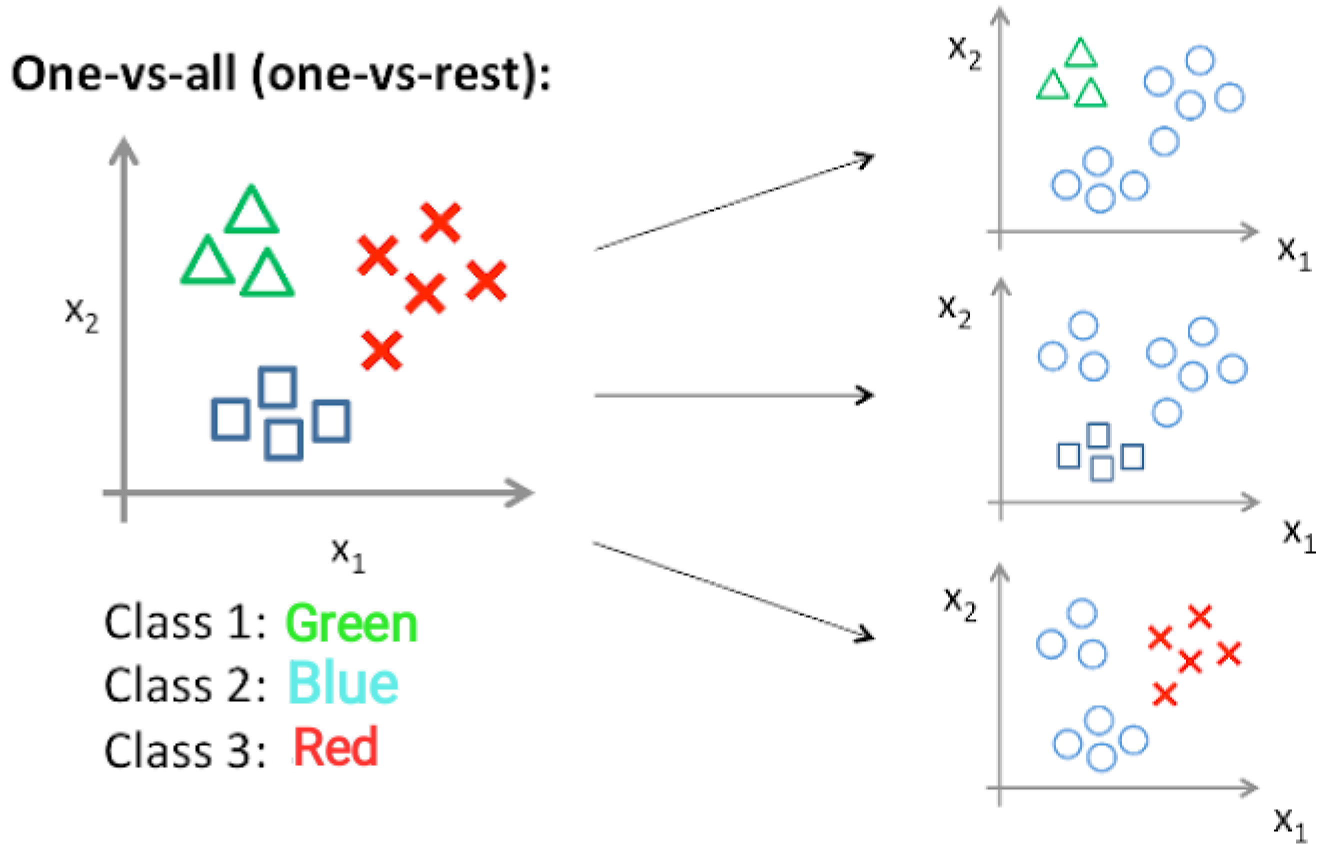
•The Confusion matrix is easy to derive but complex to understand.

•Example:- Check whether the fruit is apple, banana, or orange.

Multi-label classification

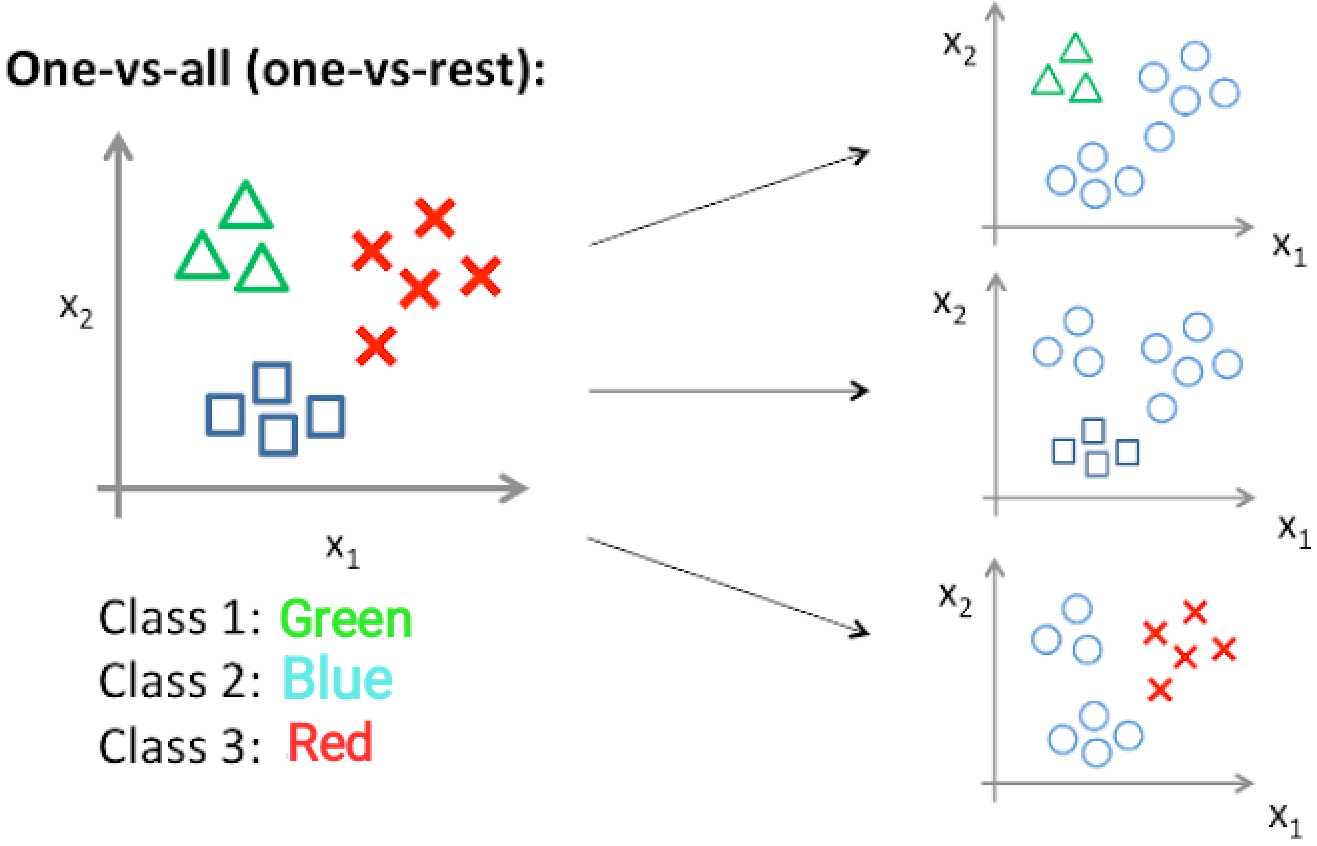
One-Vs-Rest (OVR) Classification Model for Multi-Class Classification

One-Vs-One (OVO) Classification Model for Multi-Class Classification



Multi-label classification

One-Vs-Rest (OVR) Classification Model for Multi-Class Classification



The number of class labels present in the dataset and the number of generated binary classifiers must be the same.

- **Classifier 1:- [Green] vs [Red, Blue]**
- **Classifier 2:- [Blue] vs [Green, Red]**
- **Classifier 3:- [Red] vs [Blue, Green]**

Multi-label classification

Primary Dataset

Features			Classes
x1	x2	x3	G
x4	x5	x6	B
x7	x8	x9	R
x10	x11	x12	G
x13	x14	x15	B
x16	x17	x18	R

Class 1 :- Green

Class 2 :- Blue

Class 3 :- Red

Training Dataset 1
Class :- Green

Features			Green
x1	x2	x3	**+1**
x4	x5	x6	-1
x7	x8	x9	-1
x10	x11	x12	**+1**
x13	x14	x15	-1
x16	x17	x18	-1

Multi-label classification

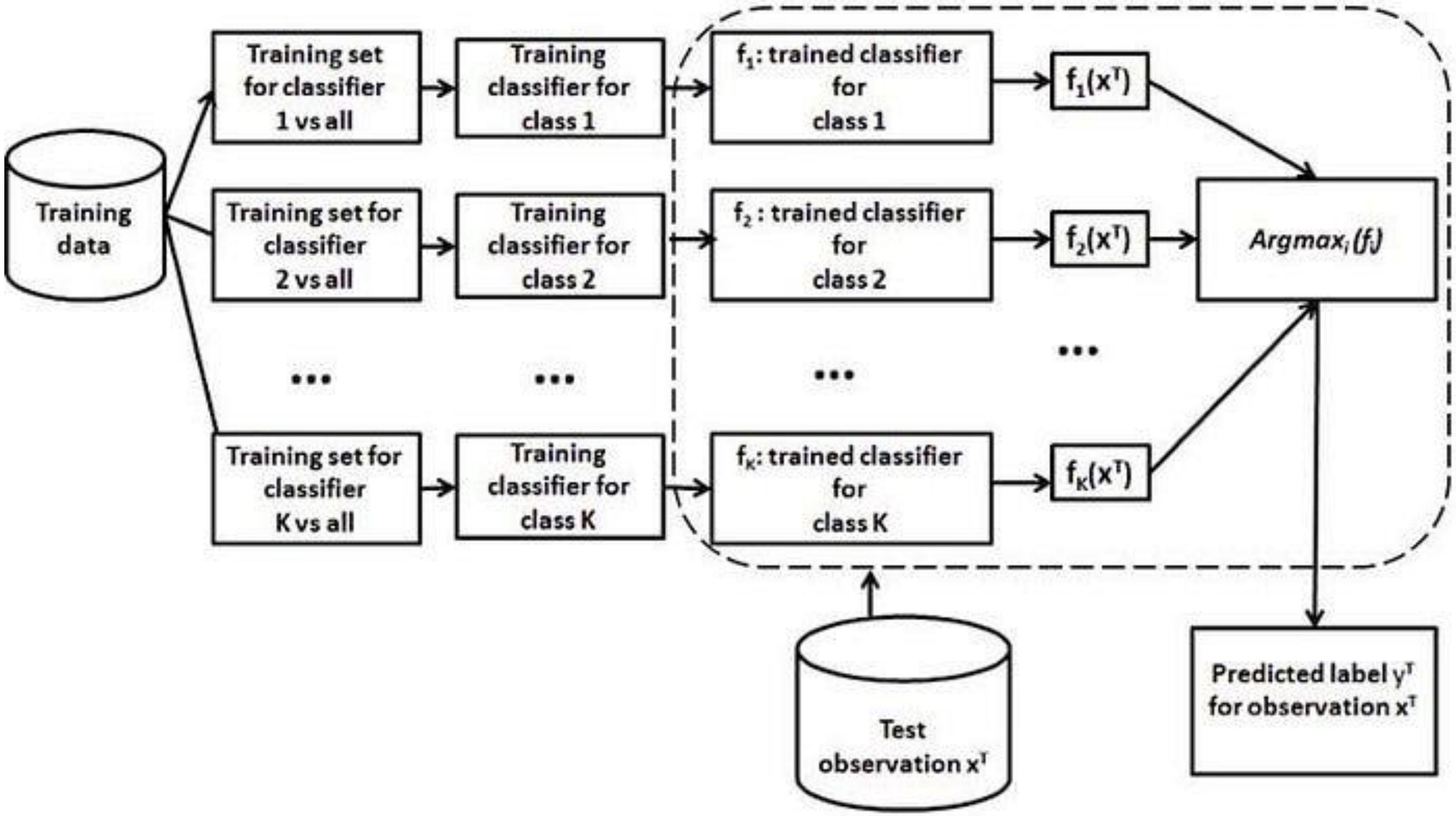
Training Dataset 2
Class :- Blue

Features			Blue
x1	x2	x3	-1
x4	x5	x6	+1
x7	x8	x9	-1
x10	x11	x12	-1
x13	x14	x15	+1
x16	x17	x18	-1

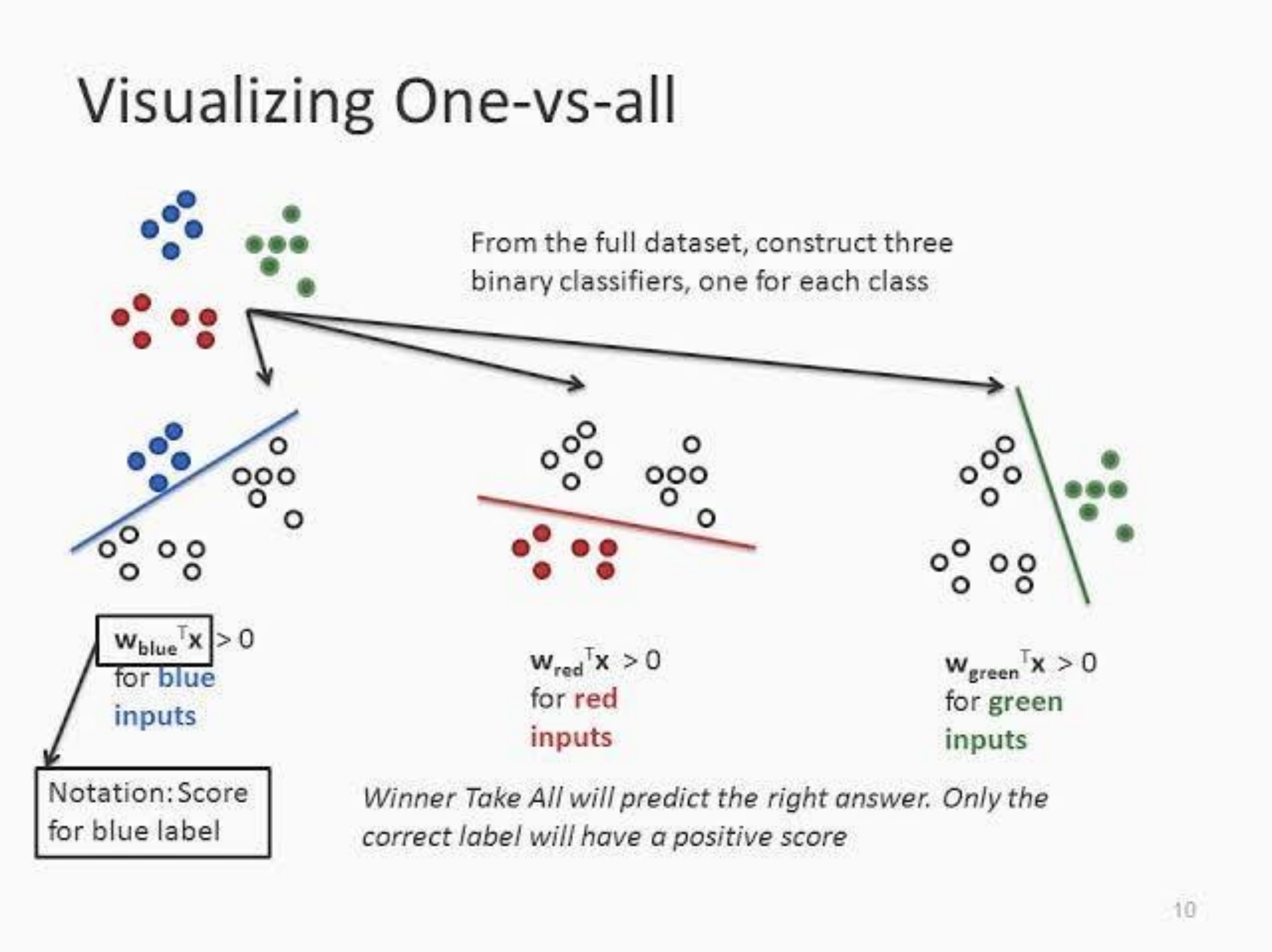
Training Dataset 3
Class :- Red

Features			Red
x1	x2	x3	-1
x4	x5	x6	-1
x7	x8	x9	+1
x10	x11	x12	-1
x13	x14	x15	-1
x16	x17	x18	+1

Multi-label classification



Multi-label classification

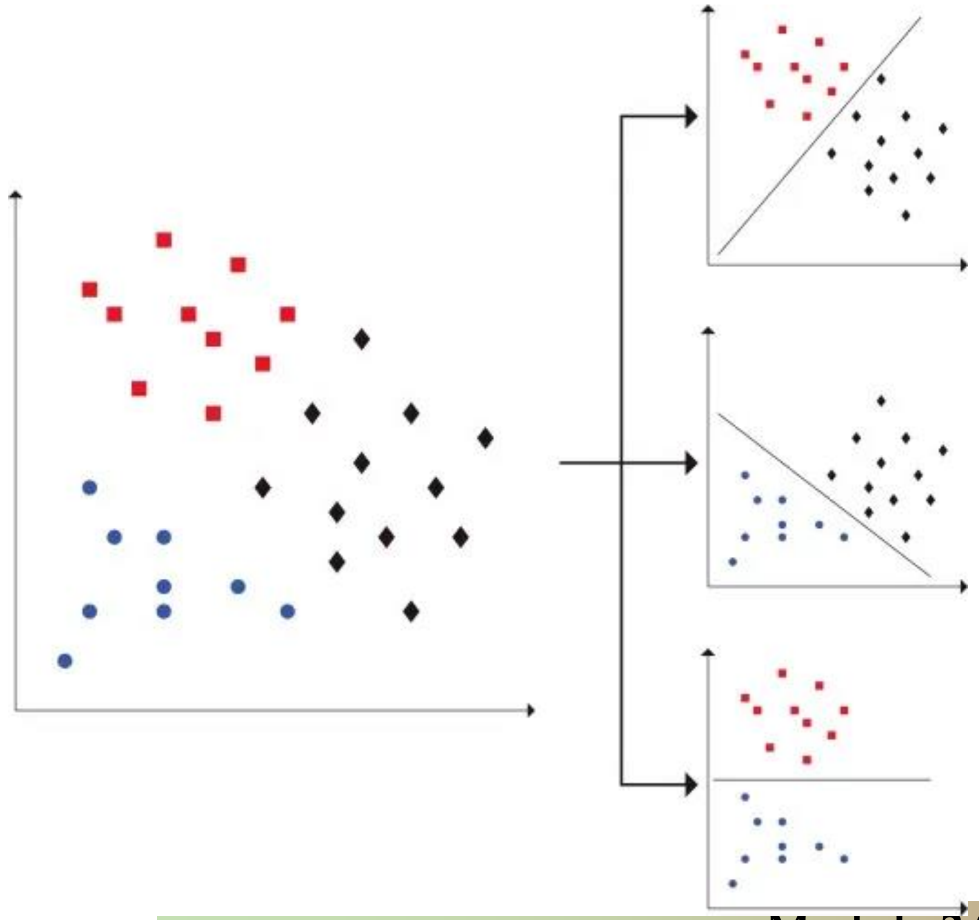


Multi-label classification

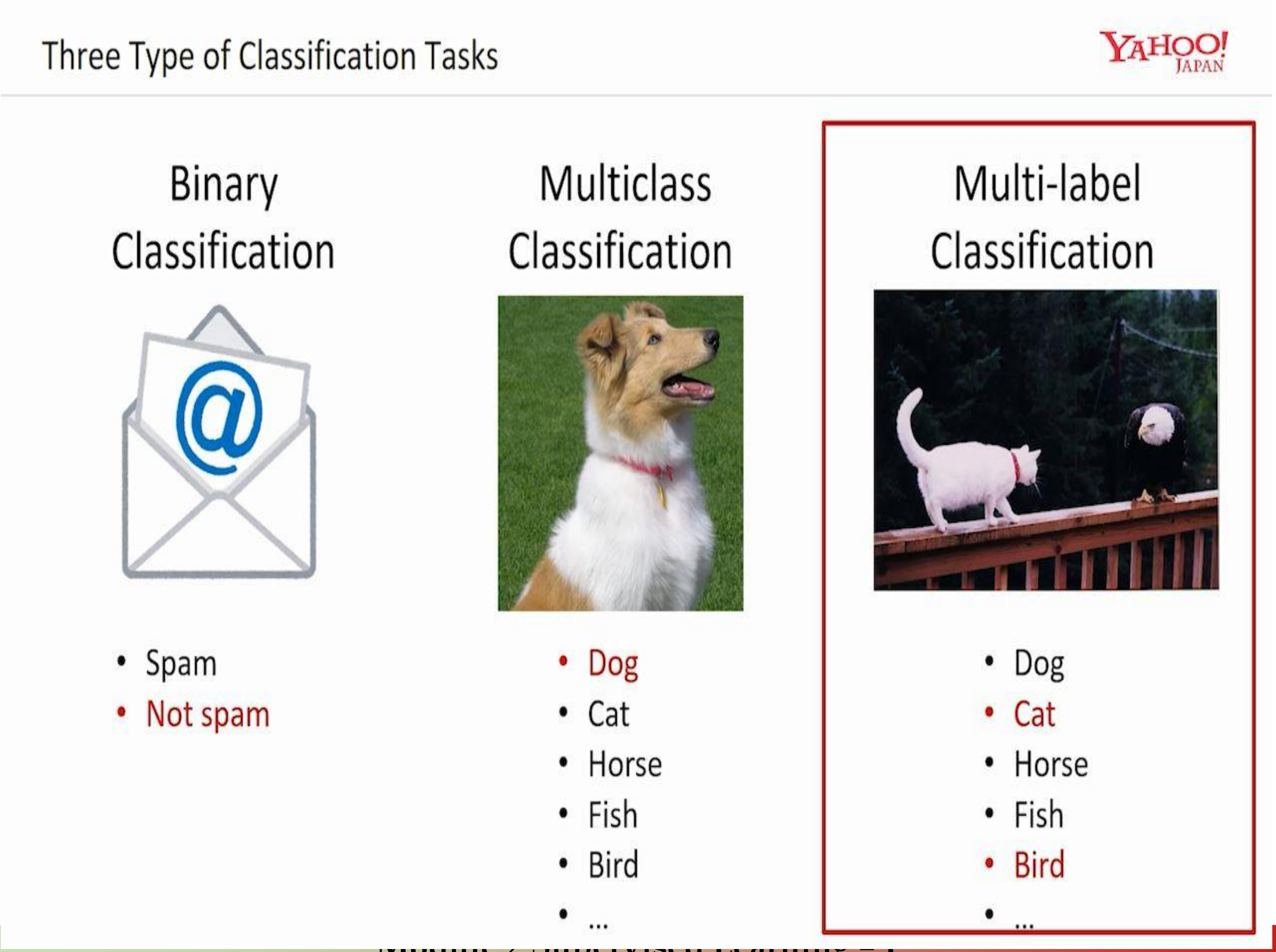
One-Vs-One (OVO) Classification Model for Multi-Class Classification

N class : N* (N-1)/2 binary classifier

- Classifier 1: Green vs. Blue
- Classifier 2: Green vs. Red
- Classifier 3: Blue vs. Red



Multi-Class Classification



Multi class Vs multi label classification

Table 1

x	y
x_1	t_1
x_2	t_2
x_3	t_1
x_4	t_2
x_5	t_1

Binary Classification

Table 2

x	y
x_1	t_2
x_2	t_3
x_3	t_4
x_4	t_1
x_5	t_3

Multi-class Classification

Table 3

x	y
x_1	$[t_2, t_5]$
x_2	$[t_1, t_2, t_3, t_4]$
x_3	$[t_3]$
x_3	$[t_2, t_4]$
x_3	$[t_1, t_3, t_4]$

Multi-label Classification

Multi class Vs multi label classification

Binary classification: Is this a picture of the sea?

$\in \{\texttt{yes}, \texttt{no}\}$

Multi-*class* classification: What is this a picture of?

$\in \{\texttt{sea}, \texttt{sunset}, \texttt{trees}, \texttt{people}, \texttt{mountain}, \texttt{urban}\}$

Multi class Vs multi label classification

Multi-label classification: Which labels are relevant to this picture?

$\subseteq$ {sea, sunset, trees, people, mountain, urban}

	$K = 2$	$K > 2$
$L = 1$	binary	multi-class
$L > 1$	multi-label	multi-output†

† also known as multi-target, multi-dimensional.

Figure: For L target variables (labels), each of K values.

Multiclass Classification:

Objective: In multiclass classification, the goal is to assign instances to one of multiple predefined classes or categories. Each instance belongs to exactly one class.

Example: Imagine you're building a model to classify fruits. You have three classes: "apple," "orange," and "banana." When presented with an image of a fruit, the model predicts a single label (e.g., "apple").

Application: Handwriting recognition, email categorization, and image classification with more than two distinct categories fall under multiclass classification.

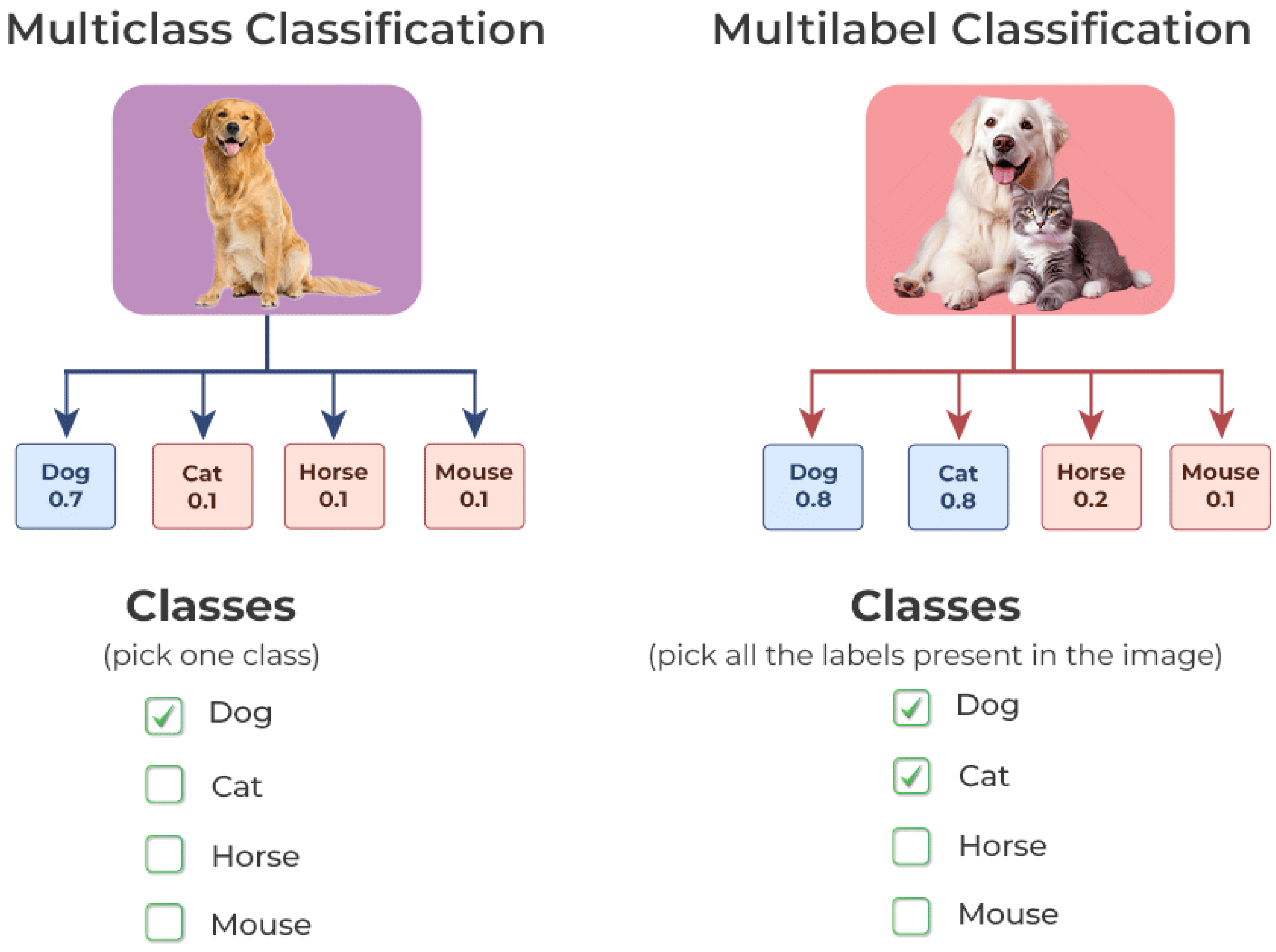
Multilabel Classification:

Objective: In multilabel classification, each instance can be associated with multiple labels simultaneously. This allows for the assignment of multiple binary labels to the same instance.

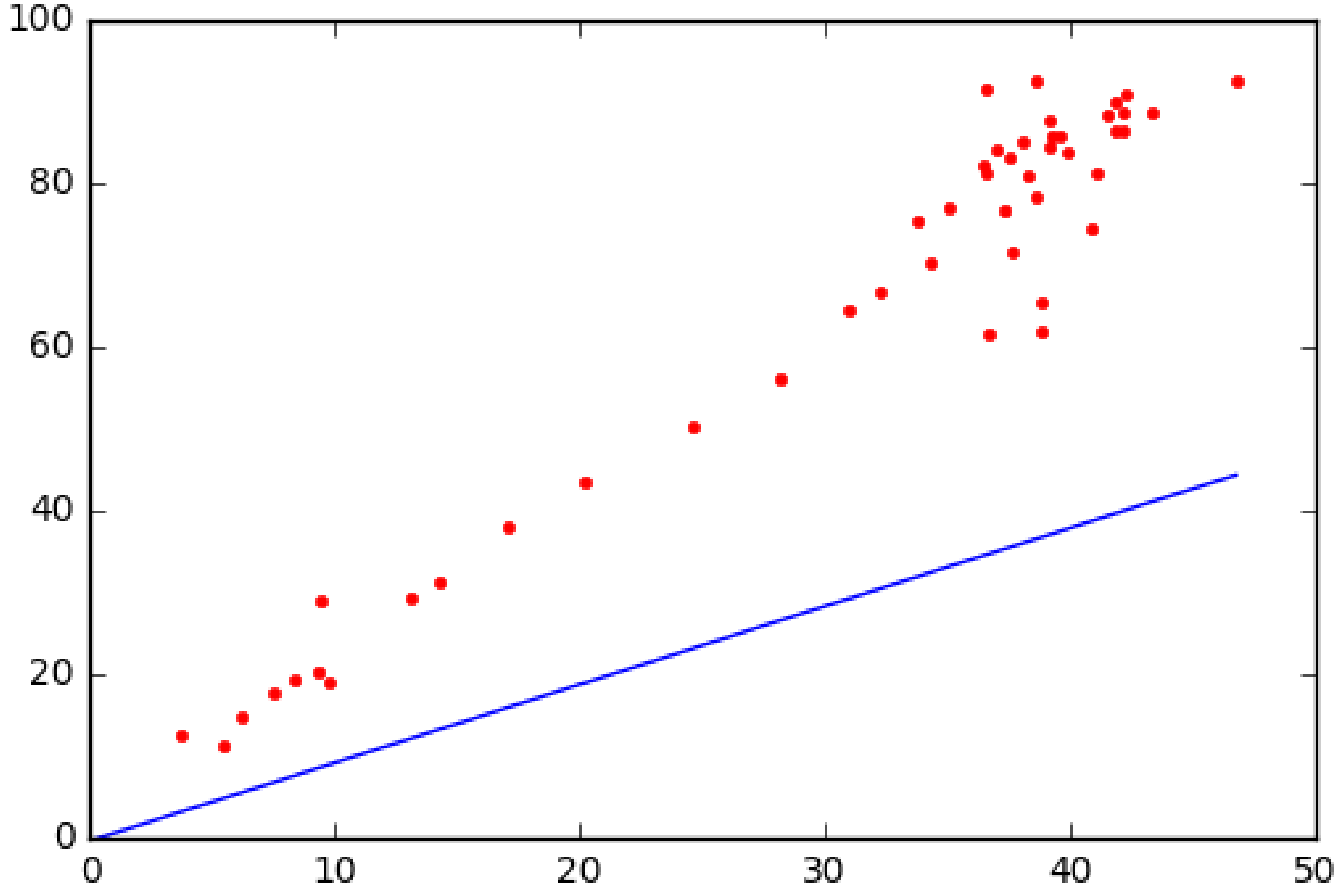
Example: Suppose you're working on a movie recommendation system. Each movie can belong to several genres (e.g., action, comedy, drama). The model predicts multiple labels (genres) for a given movie.

Application: Recommender systems, tagging content (e.g., articles with multiple topics), and sentiment analysis where a text can express mixed emotions

Mutliclass Classification vs multilabel classification



Linear Regression



Linear Regression

Hours Studied	Grade on Exam
2.00	69.00
9.00	98.00
5.00	82.00
5.00	77.00
3.00	71.00
7.00	84.00
1.00	55.00
8.00	94.00
6.00	84.00
2.00	64.00

Using least mean square (LMS) method ,

(i) Find regression line,

(ii) If hour studied=3, Find grade on exam

Linear Regression

Car	Odometer	Price
1	37388	14636
2	44758	14122
3	45833	14016
4	30862	15590
5	31705	15568
6	34010	14718
.	.	.
.	Independent variable x	Dependent variable y
.	.	.

Using least mean square (LMS) method ,

(i) Find regression line,

(ii) If odometer=30000, Find price

Multiple Linear Regression

y	X_1	X_2
140	60	22
155	62	25
159	67	24
179	70	20
192	71	15
200	72	14
212	75	14
215	78	11

Using least mean square (LMS) method ,

(i) Find regression line,

(ii) If odometer=30000, Find price

Where is Naive Bayes Used?

Face Recognition
As a classifier, it is used to identify the faces or its other features, like nose, mouth, eyes, etc.

Weather Prediction
It can be used to predict if the weather will be good or bad.

Medical Diagnosis
Doctors can diagnose patients by using the information that the classifier provides. Healthcare professionals can use Naive Bayes to indicate if a patient is at high risk for certain diseases and conditions, such as heart disease, cancer, and other ailments.

News Classification
With the help of a Naive Bayes classifier, Google News recognizes whether the news is political, world news, and so on.

Naïve Bayes Classifier

Naive Bayes (based on principle of conditional probability)

Bayes’ Theorem finds the probability of an event occurring given the probability of another event that has already occurred.

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior x likelihood}}{\text{evidence}}$$

We are trying to find probability of event A, given the event B is true. Event B is also termed as evidence.

P(A) is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance(here, it is event B).

P(B) is Marginal Probability: Probability of Evidence.

P(A|B) is a posteriori probability of B, i.e. probability of event after evidence is seen.

P(B|A) is Likelihood probability i.e the likelihood that a hypothesis will come true based on the evidence.

Naïve Bayes Classifier

	Outlook	Temperature	Humidity	Windy	Play Golf
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

Naïve Bayes Classifier

Outlook

	Yes	No	P(yes)	P(no)
Sunny	3	2	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Total	9	5	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

today = (Sunny, Hot, Normal, False)

$$P(Yes|today) = \frac{P(SunnyOutlook|Yes)P(HotTemperature|Yes)P(NormalHumidity|Yes)P(NoWind|Yes)P(Yes)}{P(today)}$$

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

$$P(No|today) = \frac{P(SunnyOutlook|No)P(HotTemperature|No)P(NormalHumidity|No)P(NoWind|No)P(No)}{P(today)}$$

$$P(Yes|today) \propto \frac{3}{9} \cdot \frac{2}{9} \cdot \frac{6}{9} \cdot \frac{6}{9} \cdot \frac{9}{14} \approx 0.02116$$

and

$$P(No|today) \propto \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{2}{5} \cdot \frac{5}{14} \approx 0.0068$$

Normalization

$$P(Yes|today) + P(No|today) = 1$$

$$P(Yes|today) = \frac{0.02116}{0.02116+0.0068} \approx 0.0237$$

and

$$P(No|today) = \frac{0.0068}{0.0141+0.0068} \approx 0.33$$

Gaussian Naive Bayes

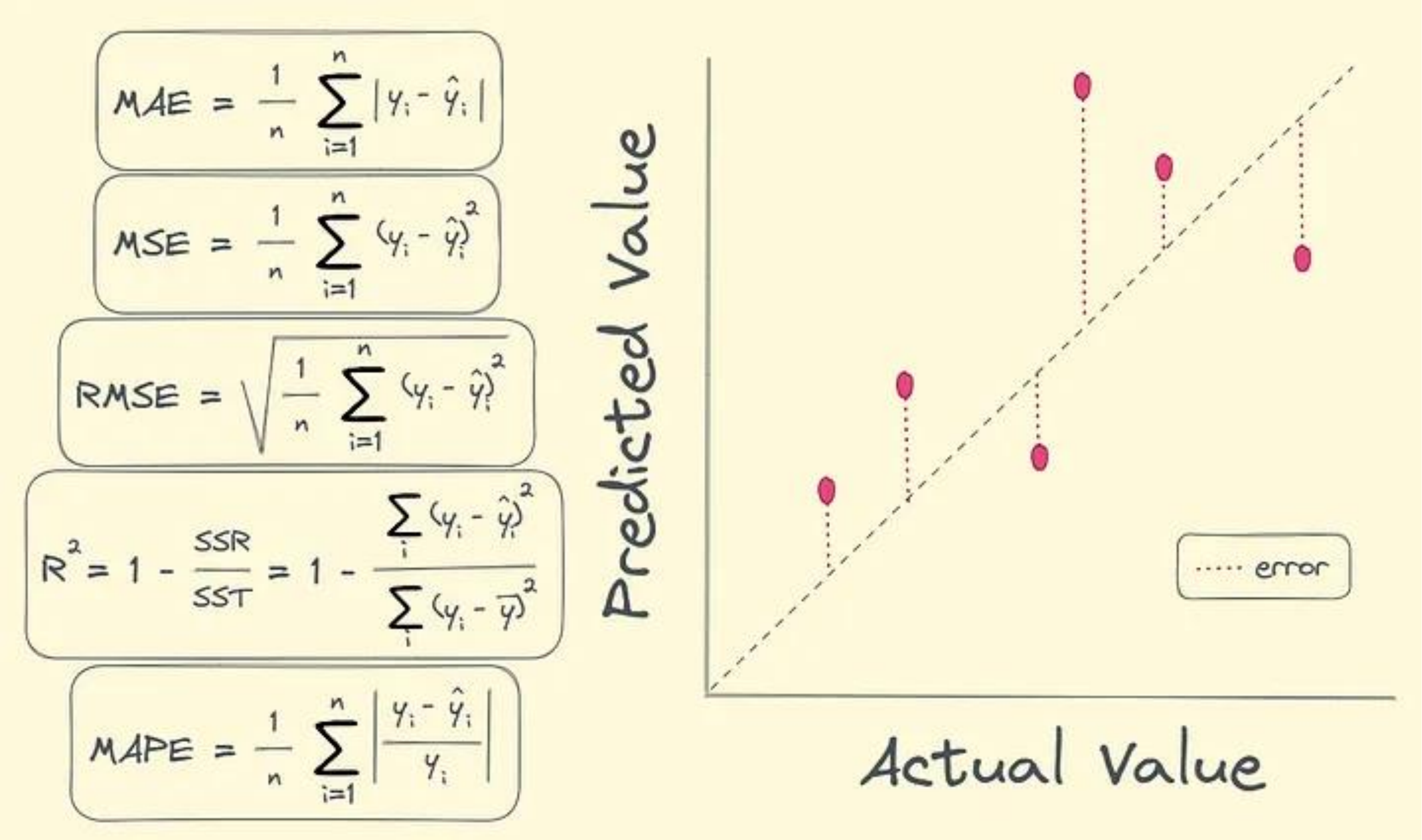
Gaussian Naive Bayes is a type of Naive Bayes method where continuous attributes are considered and the data features follow a Gaussian distribution throughout the dataset.

outlook	temperature	humidity	windy	play
sunny	85	85	false	no
sunny	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

Given the training data in the table below (Tennis data with some numerical attributes), predict the class of the following new example using Naïve Bayes classification:

outlook=overcast, temperature=60, humidity=62, windy=false

Gaussian Naive Bayes



Mean Squared Error (MSE)

Root Mean Squared Error (RMSE)

Mean Absolute Error (MAE)

R-squared (R^2)

Mean Absolute Percentage Error (MAPE)

Decision Trees

ID3

CART

Gaussian Naive Bayes

Gaussian Naive Bayes

Gaussian Naive Bayes

Gaussian Naive Bayes

Gaussian Naive Bayes