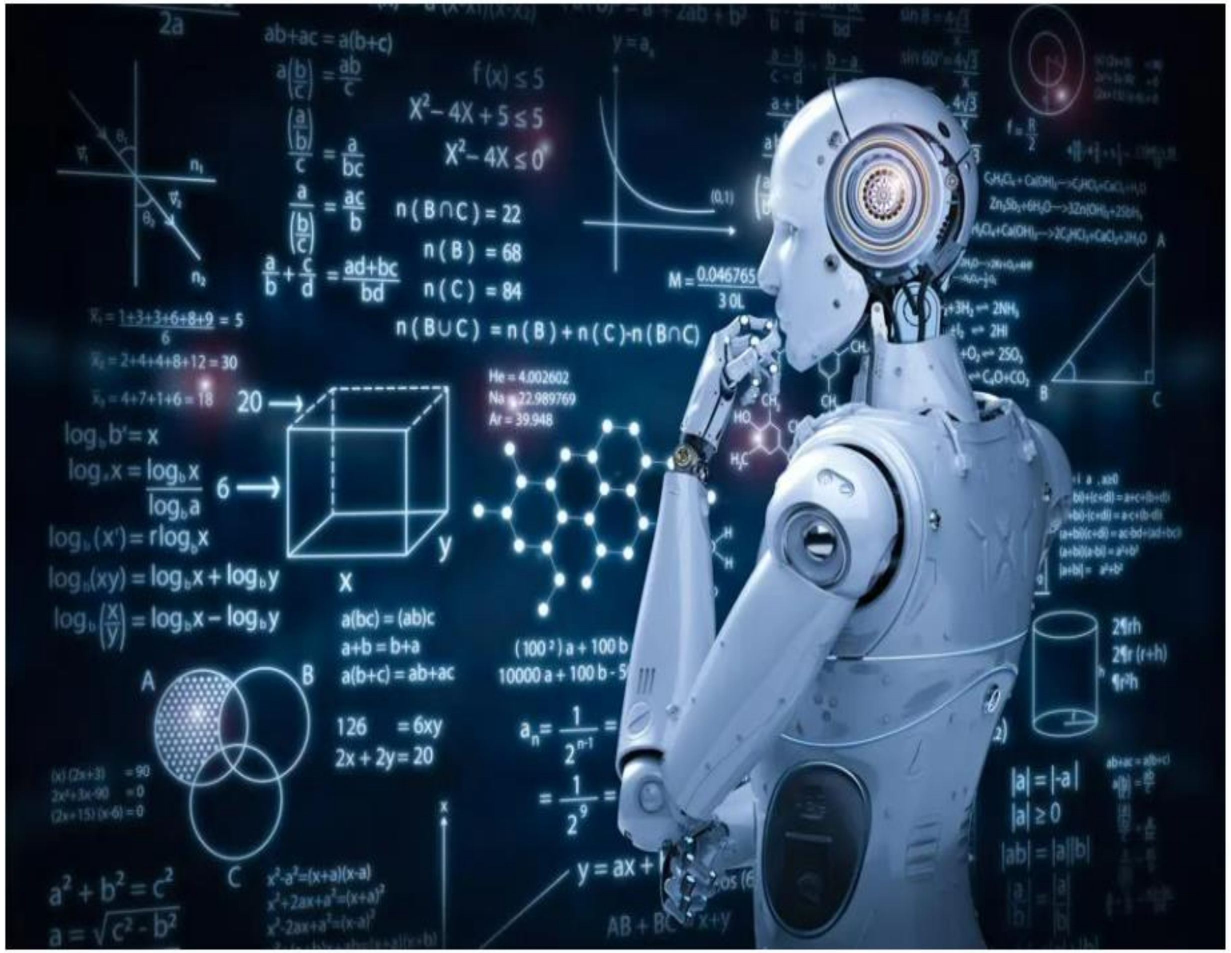


Machine Learning

Course Code:

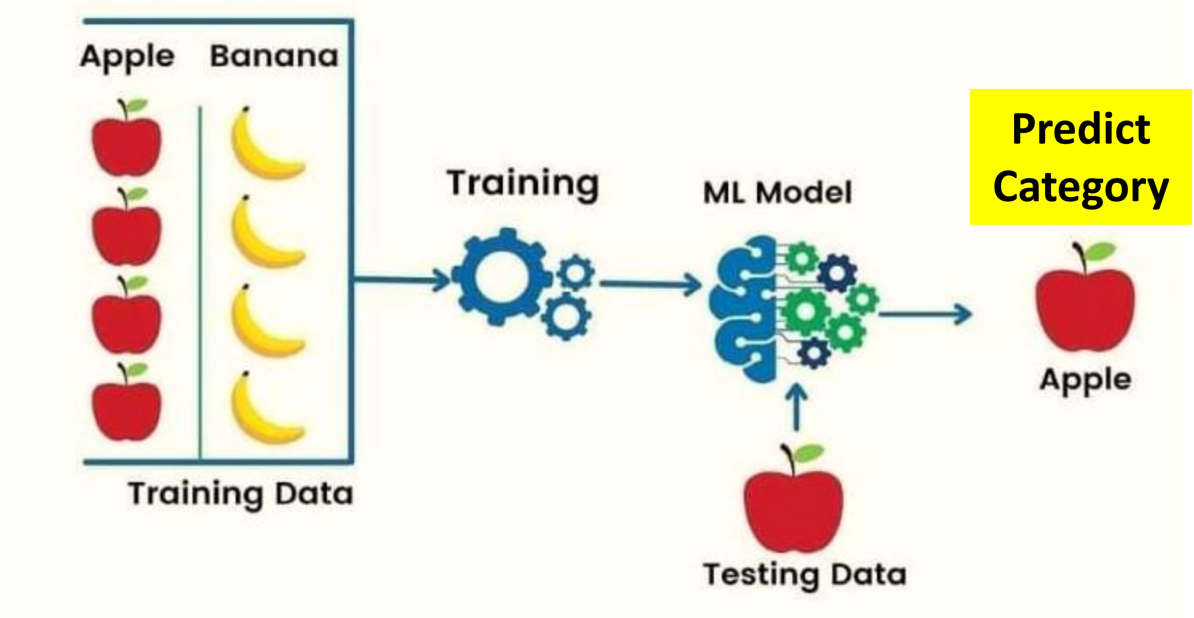
Supervised Learning

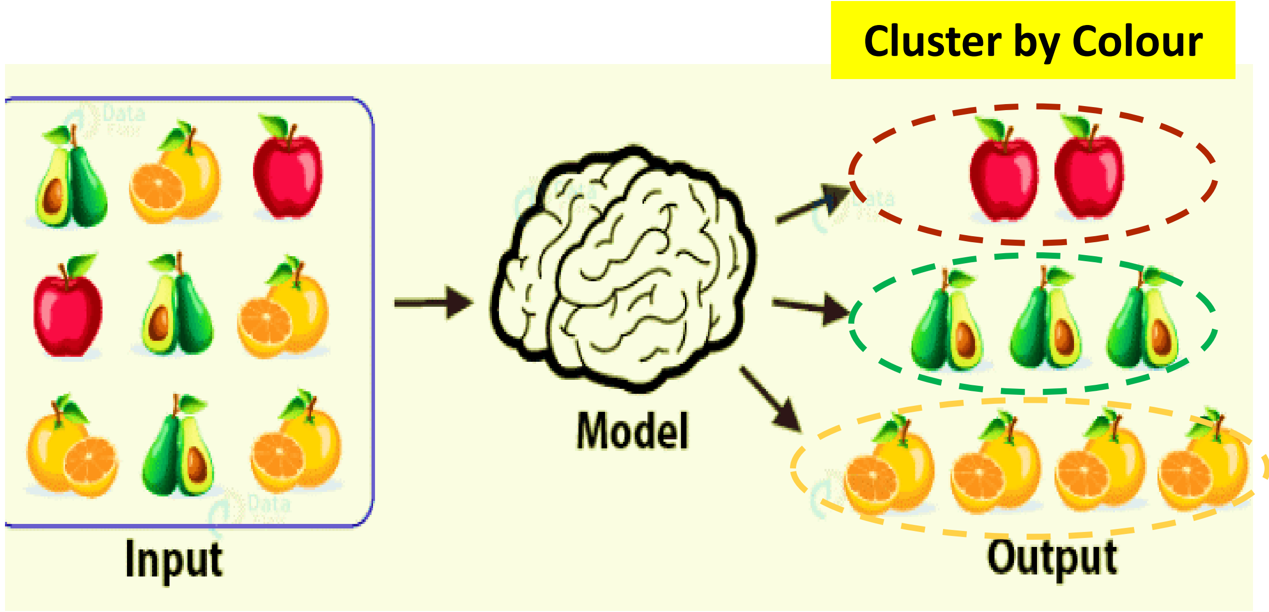
Supervised Learning

- Training Data is Labelled
- Task driven
- Classification/Regression

Unsupervised Learning

- Training Data is Not Labelled
- Data driven
- Clustering, Dimensionality Reduction

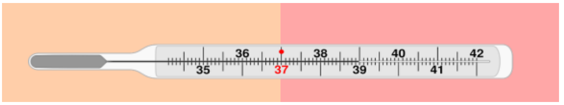




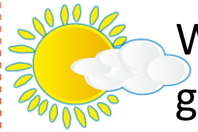
Example of Classification vs. Regression

Parameters	Classification	Regression
Output type	Discrete	Continuous
Trying to find	A boundary	Best Fit Line
Evaluation	Accuracy	Sum of squared errors
Examples	Will it be cold or Hot tomorrow?	What is the temperature going to be tomorrow?

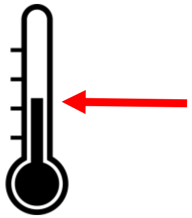
Will it be cold or Hot tomorrow?



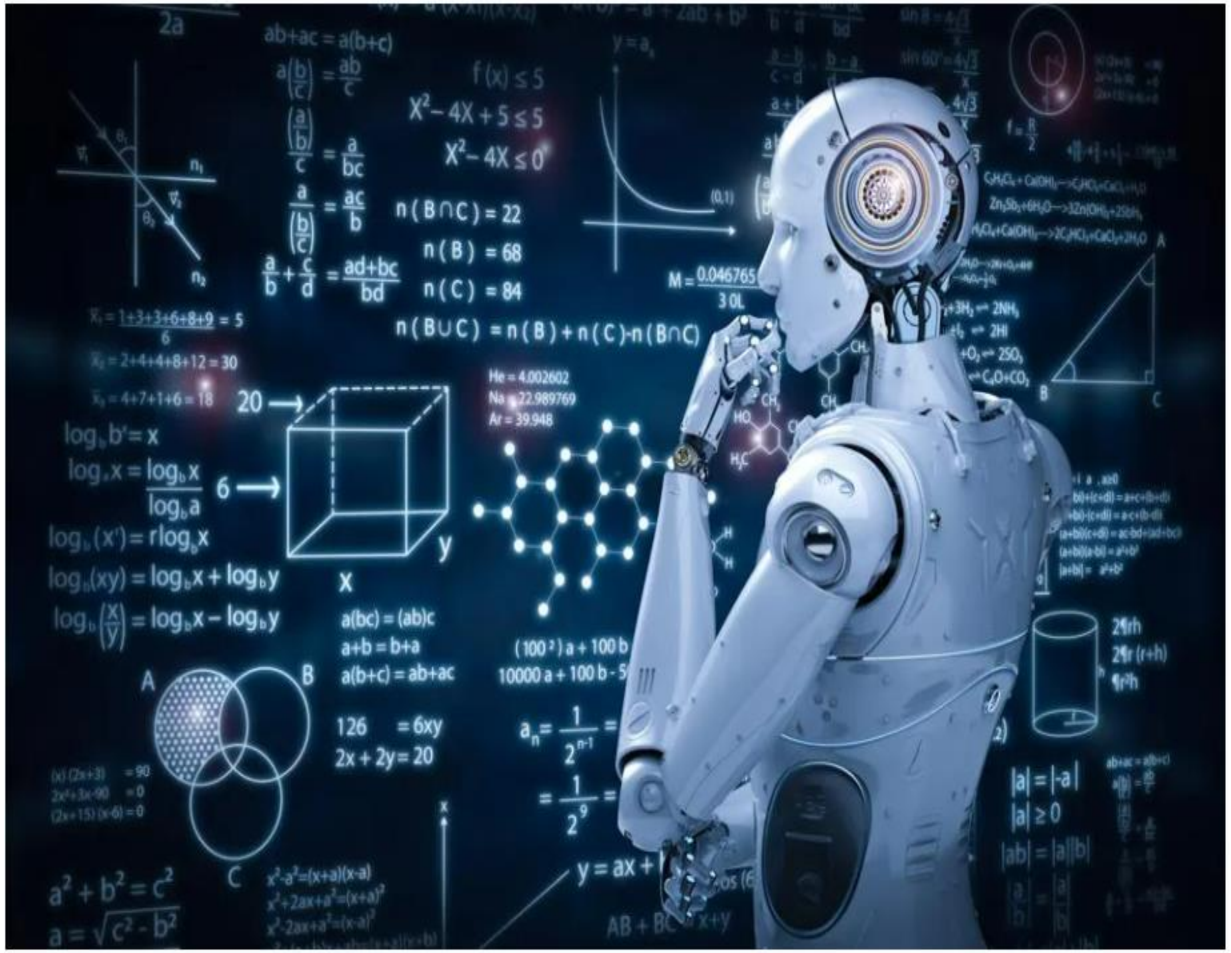
Cold Hot



What is the temperature going to be tomorrow?



42^{o}C
(Prediction)



Machine Learning

Course Code:

Unit 2: Supervised Learning: Regression

Lecture 1: Types of Regression models

Regression

- Estimates the relationship between variables

Regression

- Estimates the relationship between variables
- Predict the value of one variable (dependent variable) based on the values of other variables (independent variables).

Regression

- Estimates the relationship between variables
- Predict the value of one variable (dependent variable) based on the values of other variables (independent variables).
- It predicts a continuous/real values such as salary, age, temperature, price, etc.

Regression

- Estimates the relationship between variables
- Predict the value of one variable (dependent variable) based on the values of other variables (independent variables).
- It predicts a continuous/real values such as salary, age, temperature, price, etc.
- Regression analysis is widely used for prediction and forecasting

Real-time example of Regression Analysis

Based on the scores in the mid-term exams, predict the score of a student in the end-term exam.

Mid-term score	End-term score
30	87
42	92
21	58

Source: freepik, presentermedia

Real-time example of Regression Analysis

Based on the scores in the mid-term exams, predict the score of a student in the end-term exam.

Mid-term score	End-term score
30	87
42	92
21	58

Based on the size of house, predict its selling price.

Size in sq. ft	Price in Lakh
2400	90
1200	50
1800	75

Source: freepik, presentermedia

Real-time example of Regression Analysis

Based on the working hour of an employee, predict the expectancy of their life.

Working hour	Life expectancy
8	70
12	65
15	60

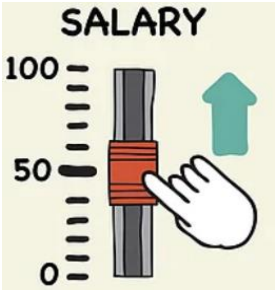
Source: freepik, istockphoto, cleanpng, clipartbest, canstockPhoto

Real-time example of Regression Analysis

Based on the working hour of an employee, predict the expectancy of their life.

Working hour	Life expectancy
8	70
12	65
15	60

Predict the salary of a person based on the three factors (education, working hour, age).



Source: freepik, istockphoto, cleanpng, clipartbest, canstockPhoto

Regression Terminologies

- **Independent Variable**

Regression Terminologies

- **Independent Variable**
  - Independent variables are regarded as inputs to a system and may take on different values freely.

Regression Terminologies

- **Independent Variable**
  - Independent variables are regarded as inputs to a system and may take on different values freely.
  - Independent variable is also called as predictor or explanatory variable.

Regression Terminologies

- **Independent Variable**
  - Independent variables are regarded as inputs to a system and may take on different values freely.
  - Independent variable is also called as predictor or explanatory variable.
- **Dependent Variable**

Regression Terminologies

- **Independent Variable**
  - Independent variables are regarded as inputs to a system and may take on different values freely.
  - Independent variable is also called as predictor or explanatory variable.
- **Dependent Variable**
  - Dependent variables are those values that changes as a consequence of changes in other values in the system.

Regression Terminologies

❖ Independent Variable

- Independent variables are regarded as inputs to a system and may take on different values freely.
- Independent variable is also called as predictor or explanatory variable.

❖ Dependent Variable

- Dependent variables are those values that changes as a consequence of changes in other values in the system.
- Dependent variable is also known as response variable.

Regression Terminologies

❖ Independent Variable

- Independent variables are regarded as inputs to a system and may take on different values freely.
- Independent variable is also called as predictor or explanatory variable.

❖ Dependent Variable

- Dependent variables are those values that changes as a consequence of changes in other values in the system.
- Dependent variable is also known as response variable.

$$Y = a + b * X$$

Regression Terminologies

❖ Independent Variable

- Independent variables are regarded as inputs to a system and may take on different values freely.
- Independent variable is also called as predictor or explanatory variable.

❖ Dependent Variable

- Dependent variables are those values that changes as a consequence of changes in other values in the system.
- Dependent variable is also known as response variable.

$$Y = a + b * X$$

Dependent variable (DV) → Y

Independent variable (IV) → X

Regression Terminologies

- **Outliers**

Regression Terminologies

❖ Outliers

- These are the data points which contains either very high values or very low values in comparison to other data points.

Regression Terminologies

- **Outliers**
  - These are the data points which contains either very high values or very low values in comparison to other data points.
  - An outlier may hamper the result, so it should be avoided.

Regression Terminologies

- **Outliers**
  - These are the data points which contains either very high values or very low values in comparison to other data points.
  - An outlier may hamper the result, so it should be avoided.
- **Overfitting and underfitting**

Regression Terminologies

- **Outliers**
  - These are the data points which contains either very high values or very low values in comparison to other data points.
  - An outlier may hamper the result, so it should be avoided.
- **Overfitting and underfitting**
  - If our algorithm works well with the training dataset but not work well in the test data, then such problem is called overfitting.

Regression Terminologies

- **Outliers**
  - These are the data points which contains either very high values or very low values in comparison to other data points.
  - An outlier may hamper the result, so it should be avoided.
- **Overfitting and underfitting**
  - If our algorithm works well with the training dataset but not work well in the test data, then such problem is called overfitting.
  - If our algorithm does not perform well even in training data, then it is known as underfitting.

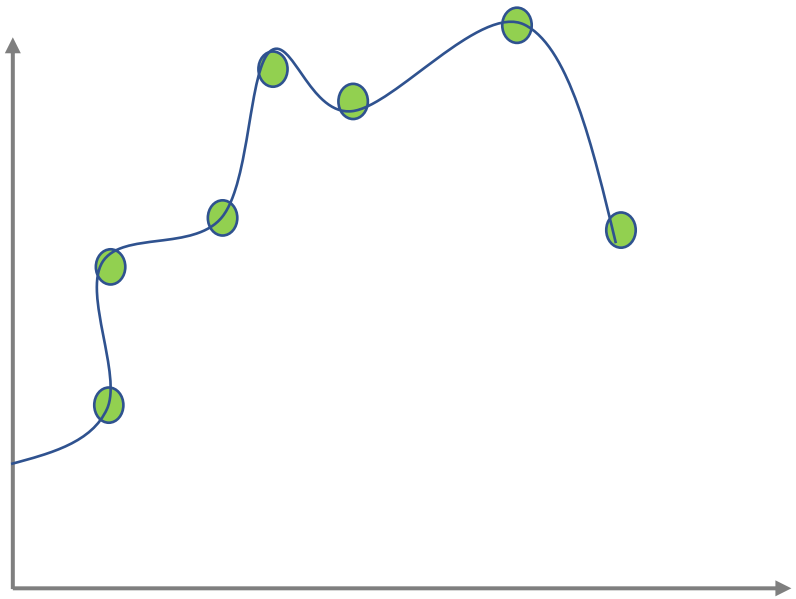
Regression Terminologies

- **Outliers**
  - These are the data points which contains either very high values or very low values in comparison to other data points.
  - An outlier may hamper the result, so it should be avoided.
- **Overfitting and underfitting**
  - If our algorithm works well with the training dataset but not work well in the test data, then such problem is called overfitting.
  - If our algorithm does not perform well even in training data, then it is known as underfitting.

Later, we will discuss this points in detail.

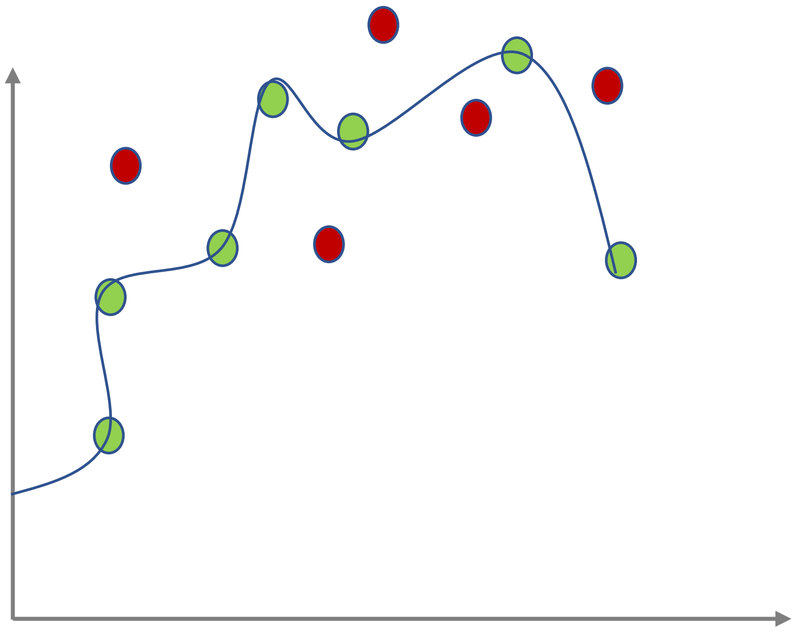
Model Complexity

Overfitting



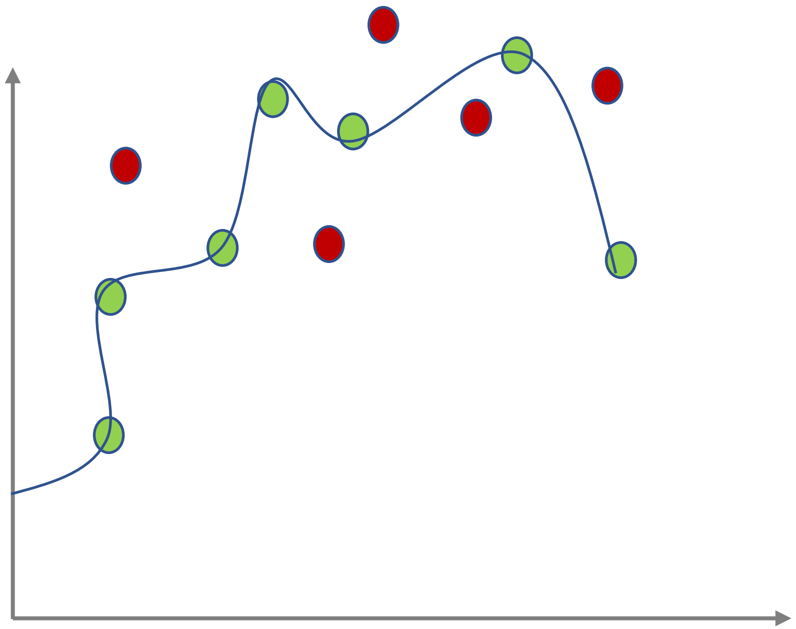
- A scenario where the machine learning model tries to learn from the details along with the noise in the data and tries to fit each data point on the curve.

Overfitting



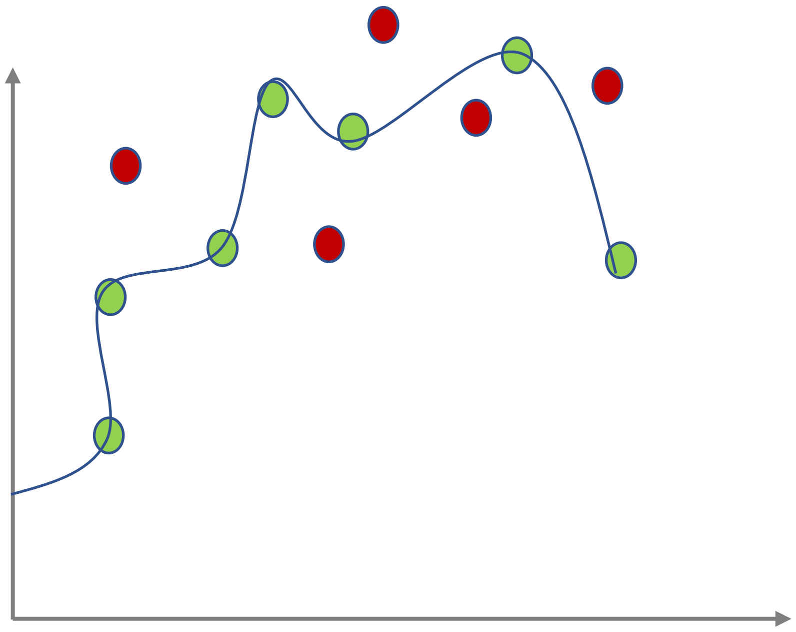
- A scenario where the machine learning model tries to learn from the details along with the noise in the data and tries to fit each data point on the curve.
- If the model has very less flexibility, it fails to predict new data points, and thus the model rejects every new data point during prediction.

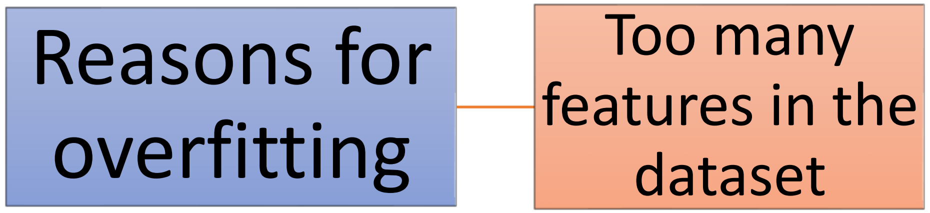
Overfitting



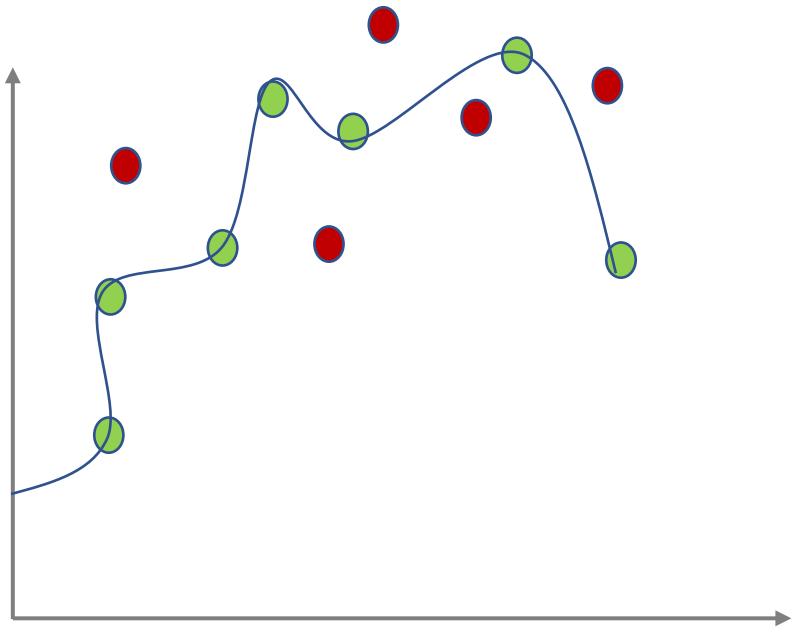
Reasons for overfitting

Overfitting



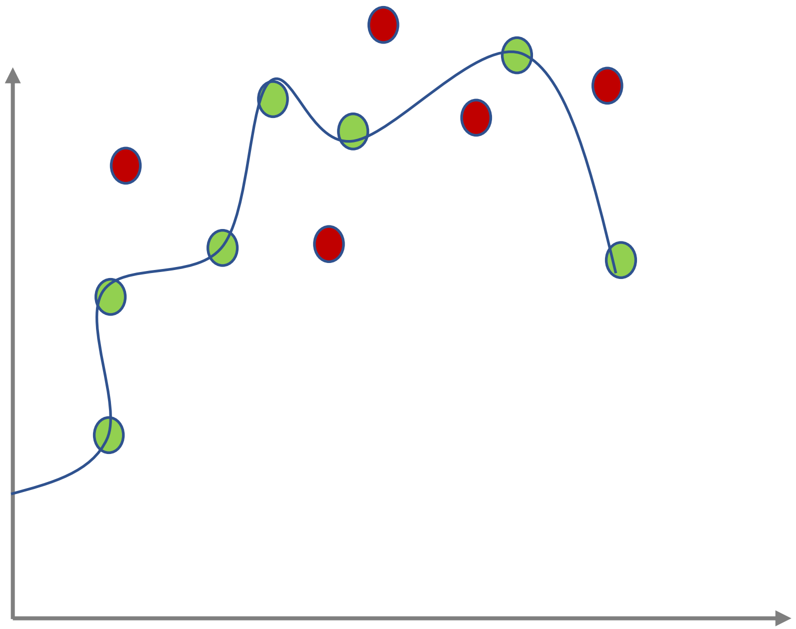


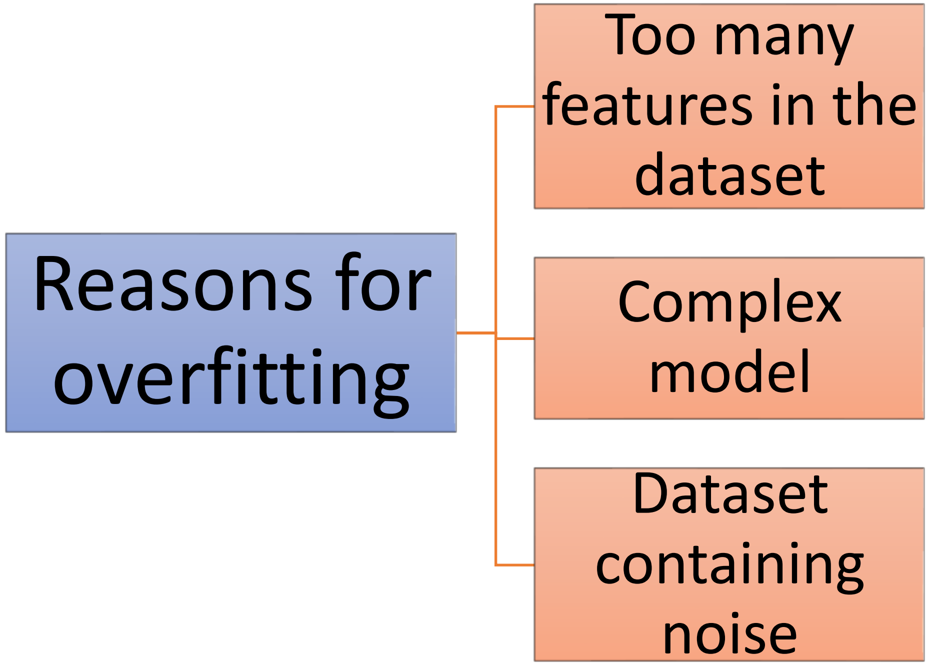
Overfitting



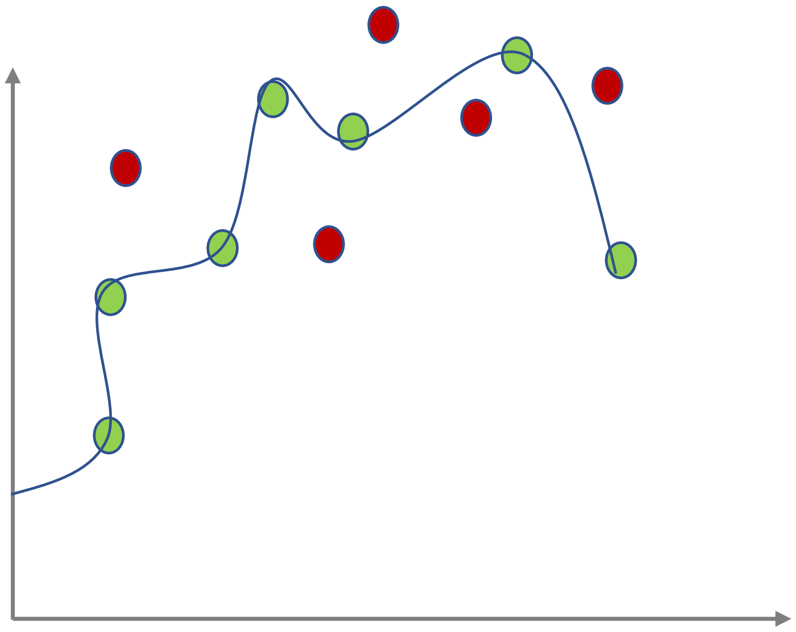
Reasons for overfitting
- Too many features in the dataset
- Complex model

Overfitting





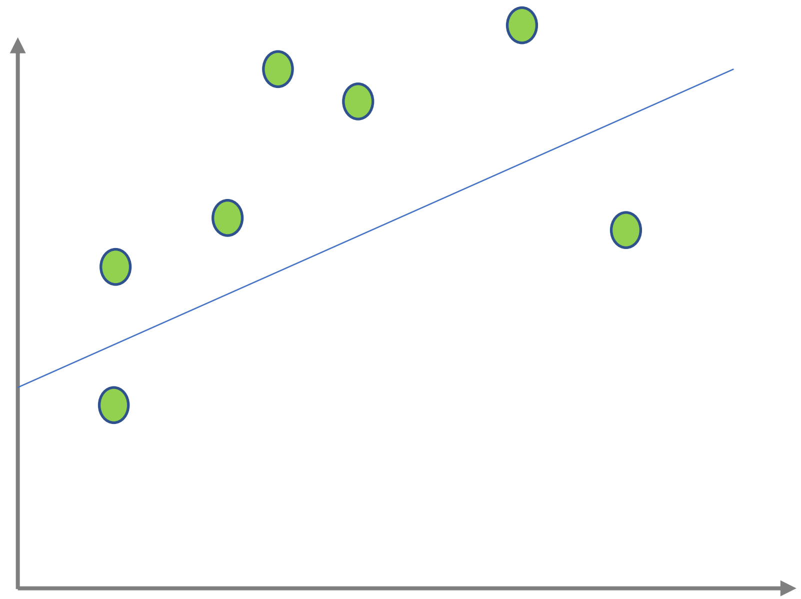
Overfitting



Reasons for overfitting

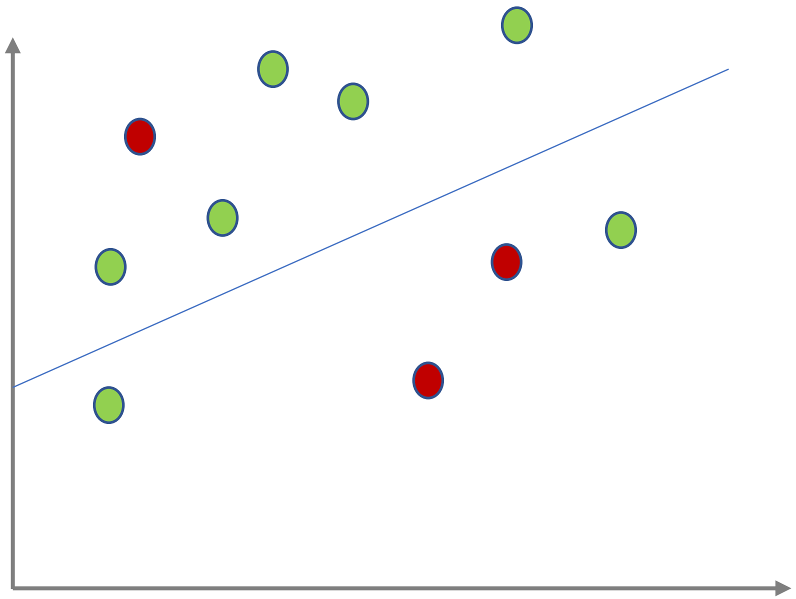
- Too many features in the dataset
- Complex model
- Dataset containing noise
- Size of dataset is very less

Underfitting



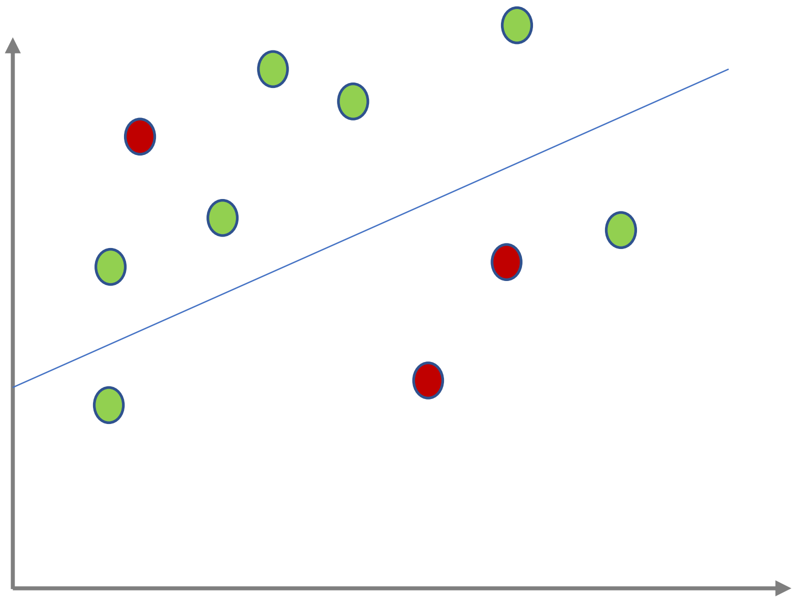
- A scenario where the machine learning model cannot learn the details of the training data and cannot predict new datapoint.

Underfitting



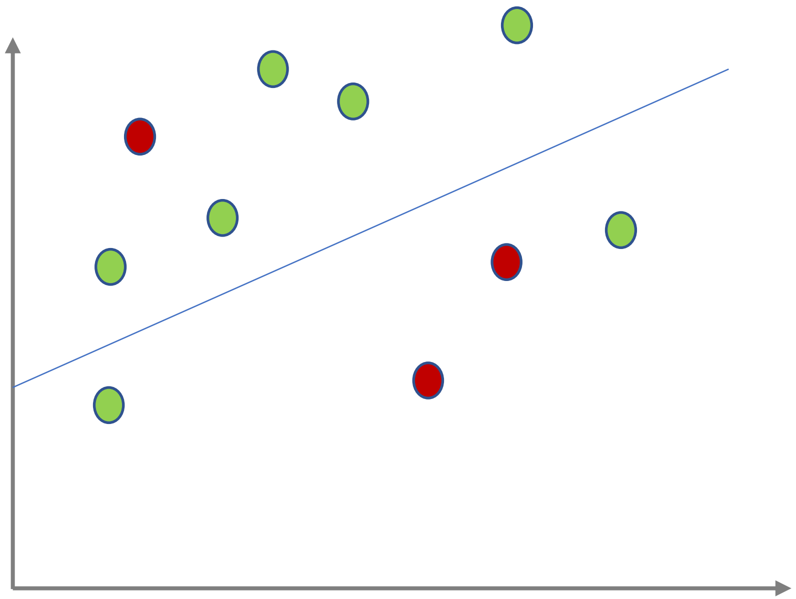
- A scenario where the machine learning model cannot learn the details of the training data and cannot predict new datapoint.
- As the model does not fully learn the patterns, it accepts every new datapoints during the prediction.

Underfitting



Reasons for underfitting

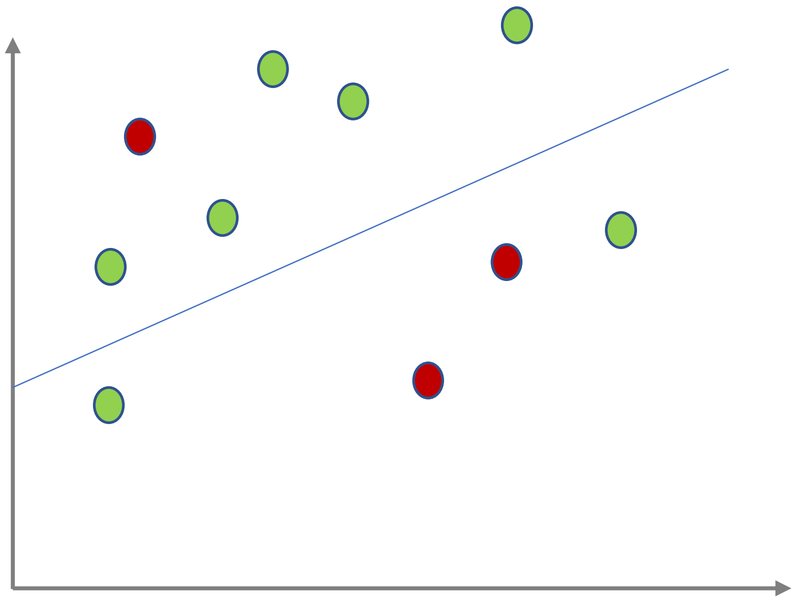
Underfitting



Reasons for underfitting

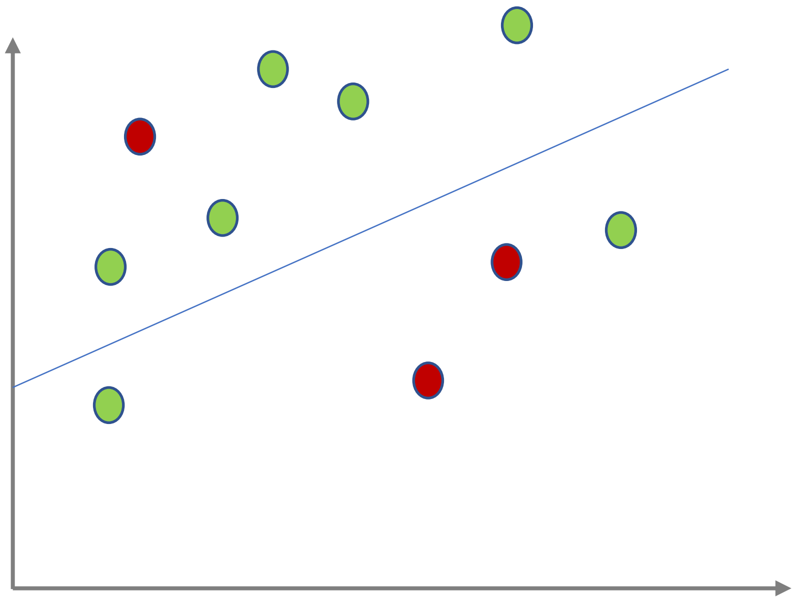
Very simple model

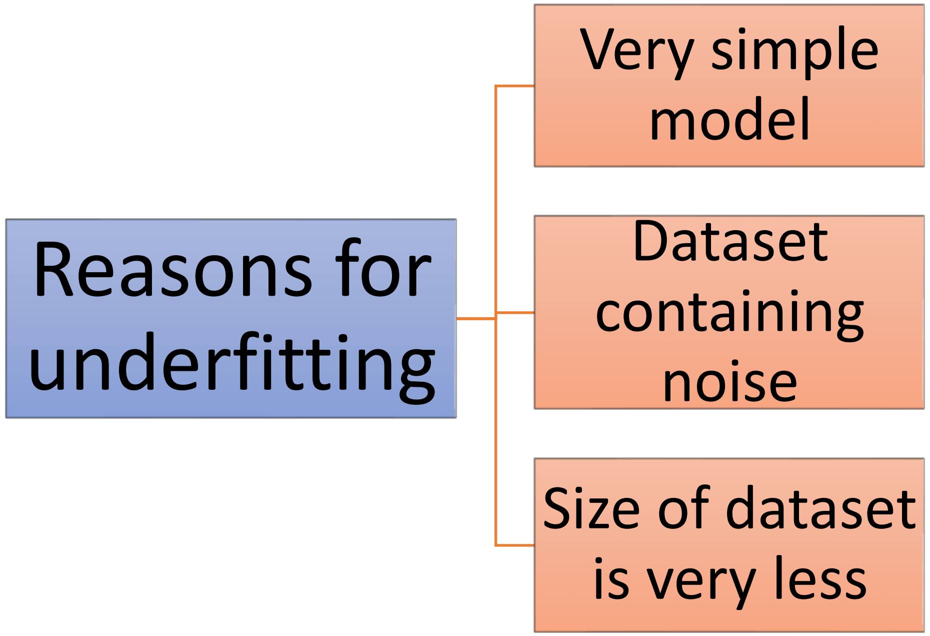
Underfitting



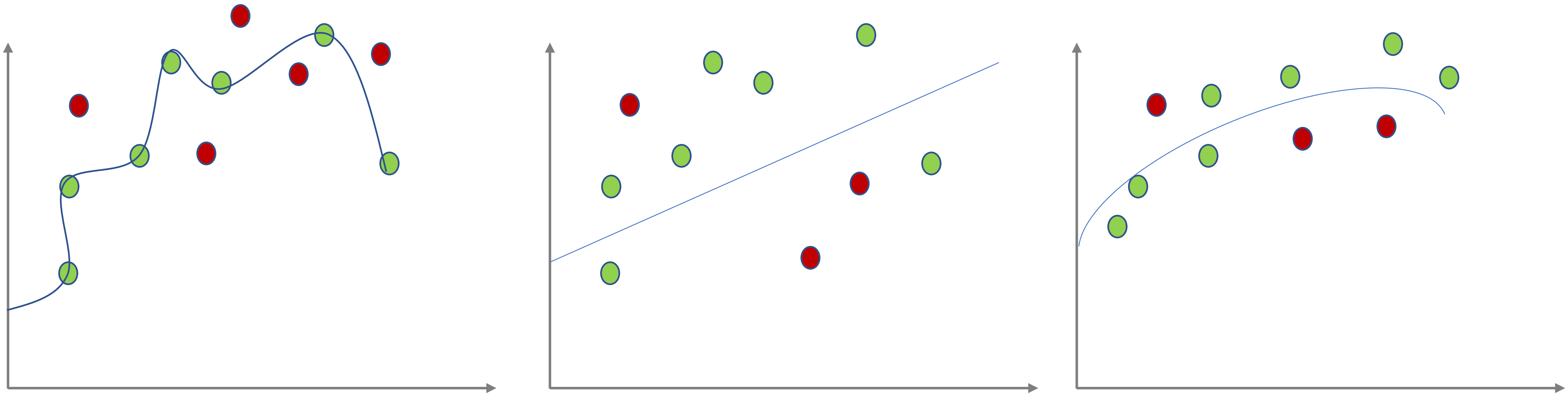
Reasons for underfitting
- Very simple model
- Dataset containing noise

Underfitting





Comparison



Overfit

Underfit

Best fit

Let’s understand these concepts using a real-world example

Suppose there are three students in a classroom.

Let’s understand these concepts using a real-world example

Suppose there are three students in a classroom.

A

Image Source: OpenClipart-Vectors from Pixabay

Let’s understand these concepts using a real-world example

Suppose there are three students in a classroom.

A

B

Image Source: OpenClipart-Vectors from Pixabay, Clker-Free-Vector-Images from Pixabay

Let’s understand these concepts using a real-world example

Suppose there are three students in a classroom.

A

B

C

Image Source: OpenClipart-Vectors from Pixabay, Clker-Free-Vector-Images from Pixabay, OpenClipart-Vectors from Pixabay

Let's understand these concepts using a real-world example

Suppose there are three students in a classroom.

A

Guessing

B

C

Let’s understand these concepts using a real-world example

Suppose there are three students in a classroom.

A

Guessing

B

Cramming

C

Let’s understand these concepts using a real-world example

Suppose there are three students in a classroom.

A

Guessing

B

Cramming

C

Problem solving

Let’s understand these concepts using a real-world example

Training	Class test
Testing	Exam

A

Guessing

B

Cramming

C

Problem solving

Let's understand these concepts using a real-world example

Training	Class test
Testing	Exam

A

Guessing

B

Cramming

C

Problem solving

Class test

Let’s understand these concepts using a real-world example

Training	Class test
Testing	Exam

A

52%

Guessing

B

Cramming

C

Problem solving

Class test

Let’s understand these concepts using a real-world example

Training	Class test
Testing	Exam

A

52%

Guessing

B

97%

Cramming

C

Problem solving

Class test

Let’s understand these concepts using a real-world example

Training	Class test
Testing	Exam

A

52%

Guessing

B

97%

Cramming

C

94%

Problem solving

Class test

Let's understand these concepts using a real-world example

Training	Class work
Testing	Exam

A

52%

48%

Guessing

B

97%

65%

Cramming

C

94%

88%

Problem solving

Class test

Final exam

Types of Regression

Types of Regression

Linear Regression

Types of Regression

Linear Regression

Multiple Linear Regression

Types of Regression

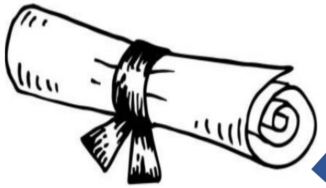
Linear Regression

Multiple Linear Regression

Logistic Regression

Types of Regression

Linear Regression



Multiple Linear Regression

Logistic Regression

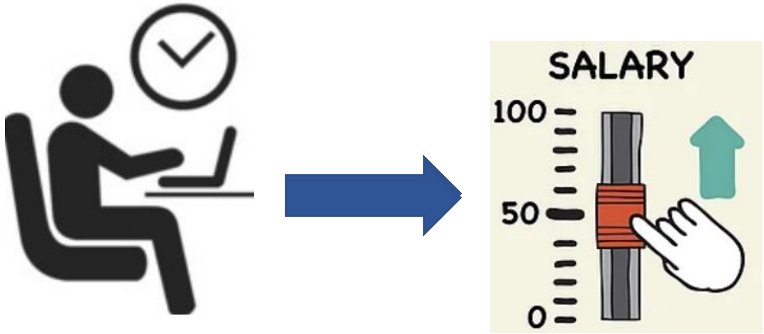
Source: freepik

Types of Regression

Linear Regression

Multiple Linear Regression

Logistic Regression



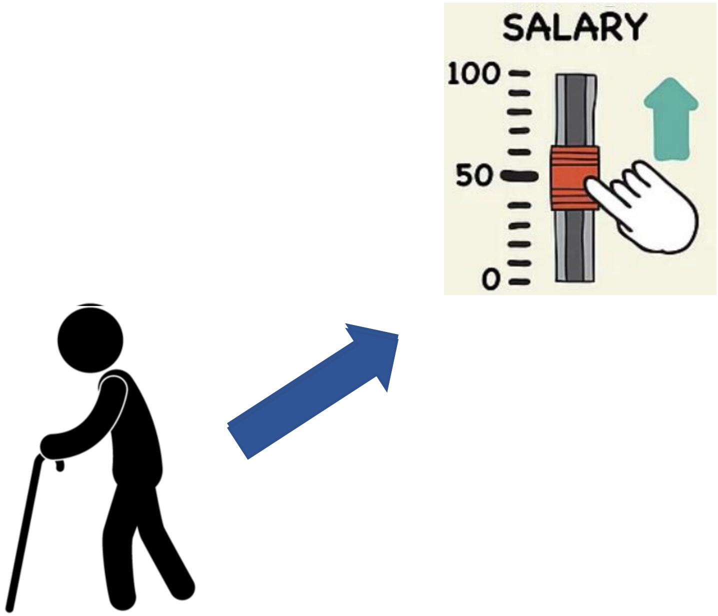
Source: freepik

Types of Regression

Linear Regression

Multiple Linear Regression

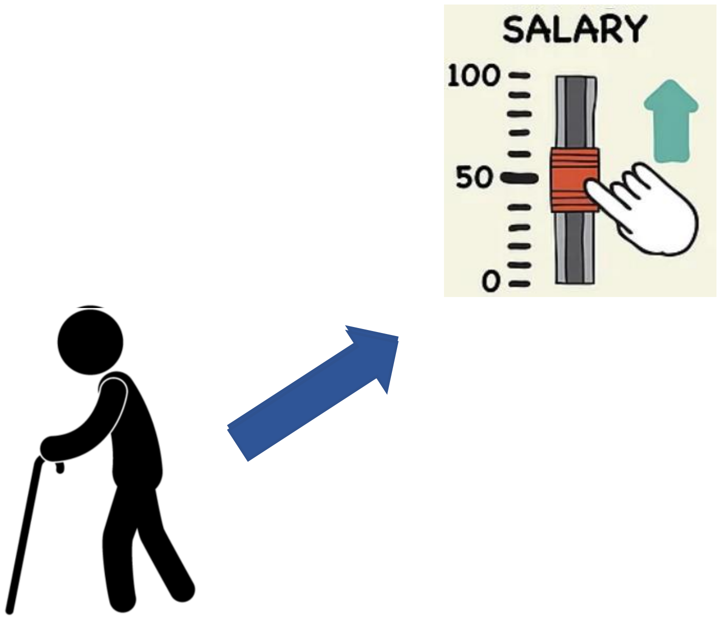
Logistic Regression



Source: freepik

Types of Regression

Linear Regression



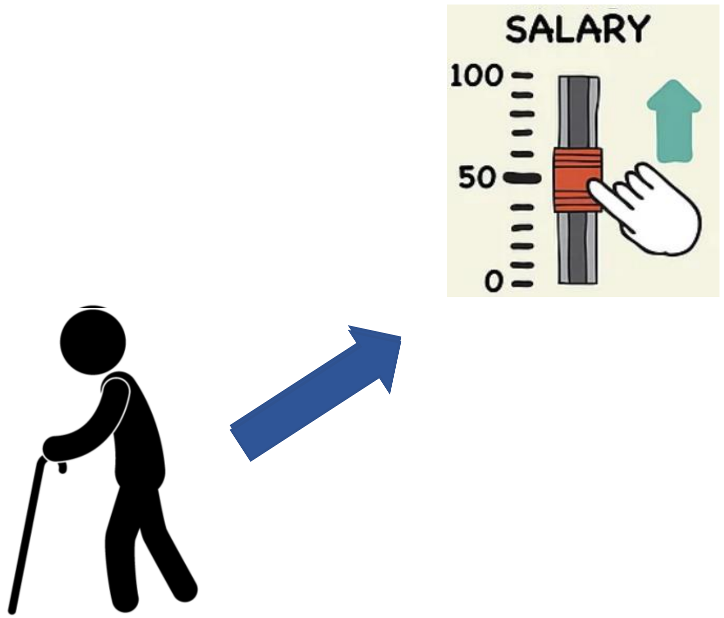
Multiple Linear Regression

Logistic Regression

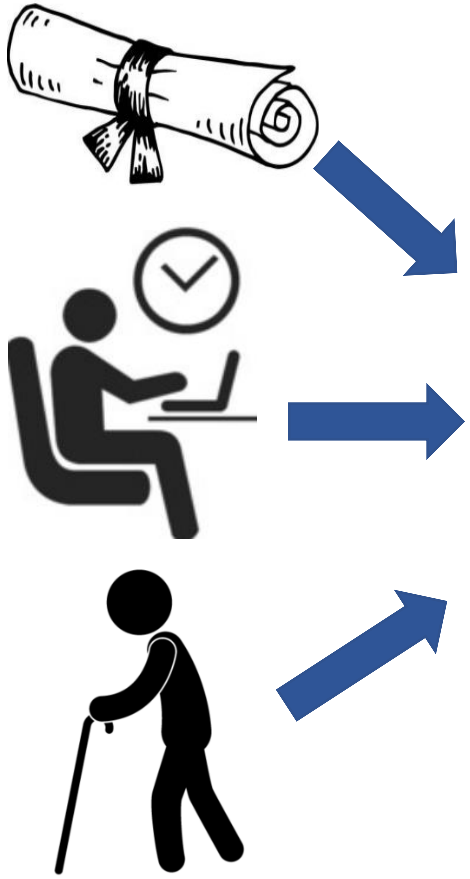
Source: freepik

Types of Regression

Linear Regression



Multiple Linear Regression

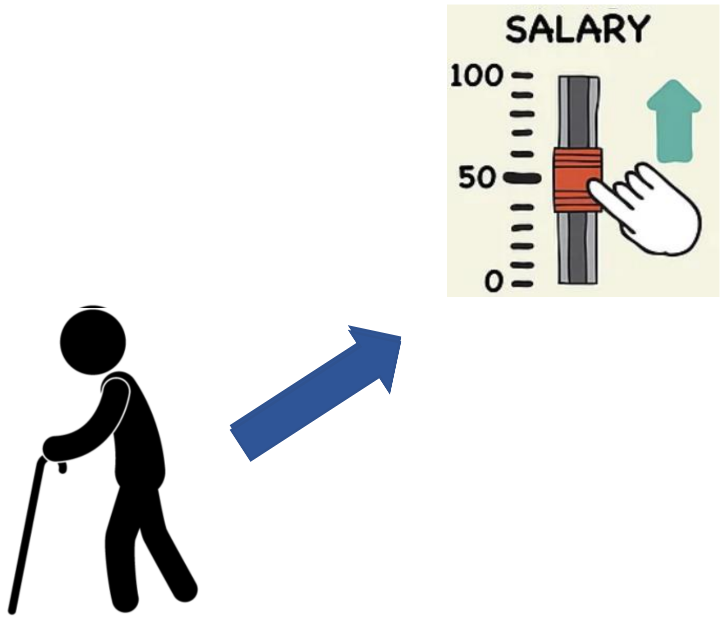


Logistic Regression

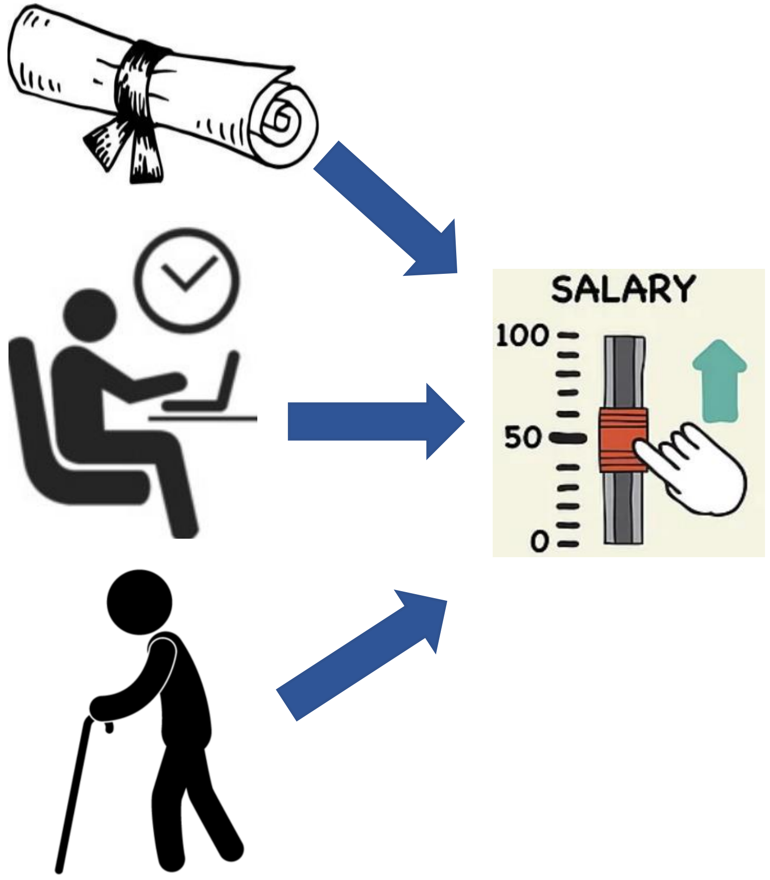
Source: freepik

Types of Regression

Linear Regression



Multiple Linear Regression



Logistic Regression

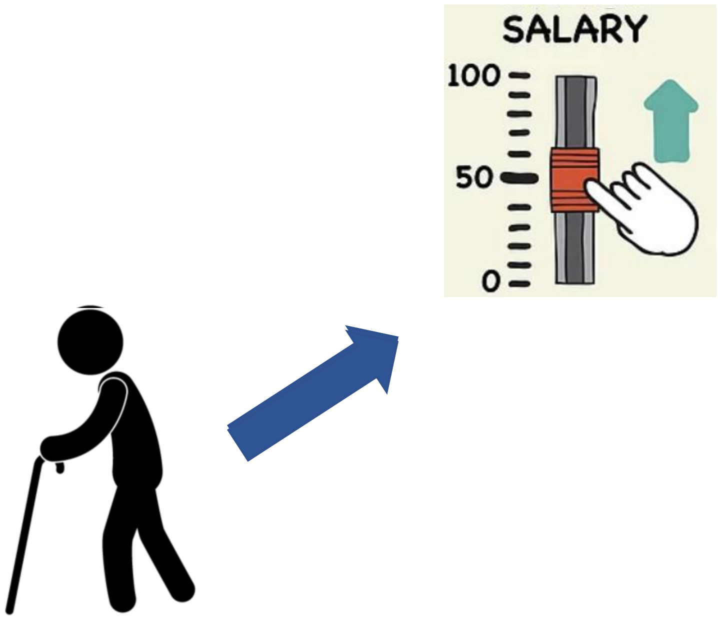
Source: freepik

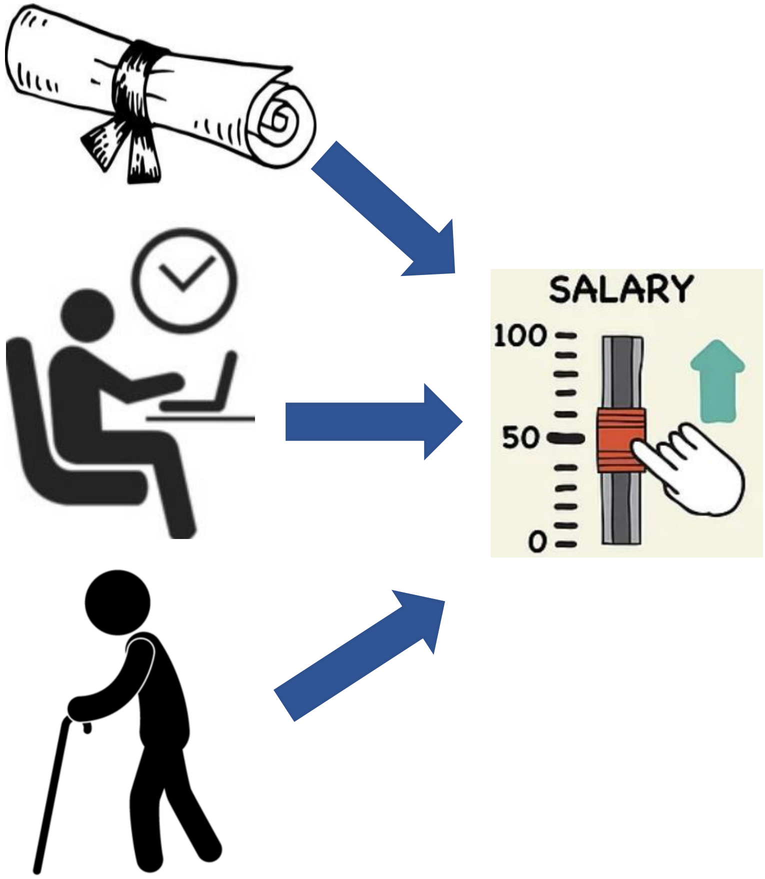
Types of Regression

Linear Regression

Multiple Linear Regression

Logistic Regression

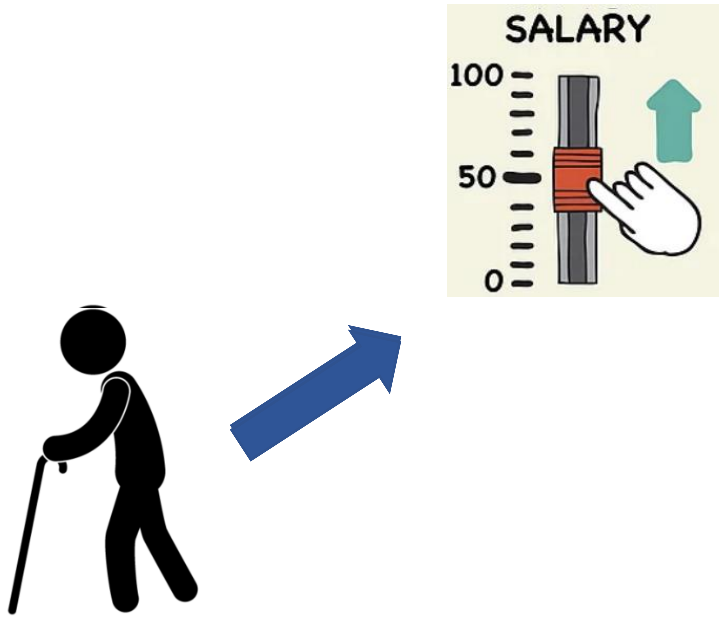




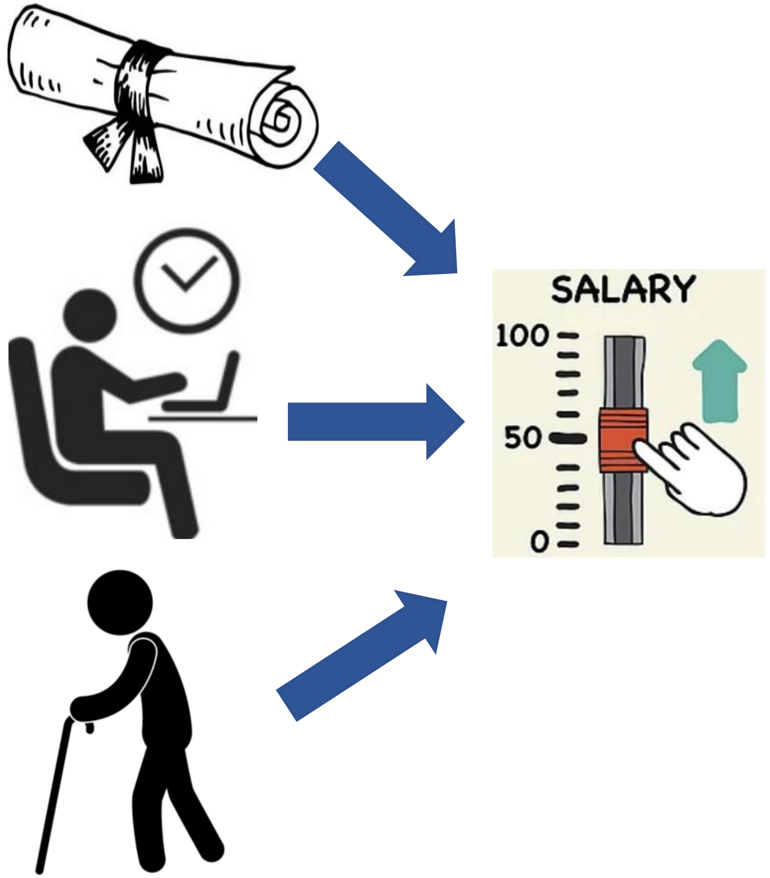
Source: freepik

Types of Regression

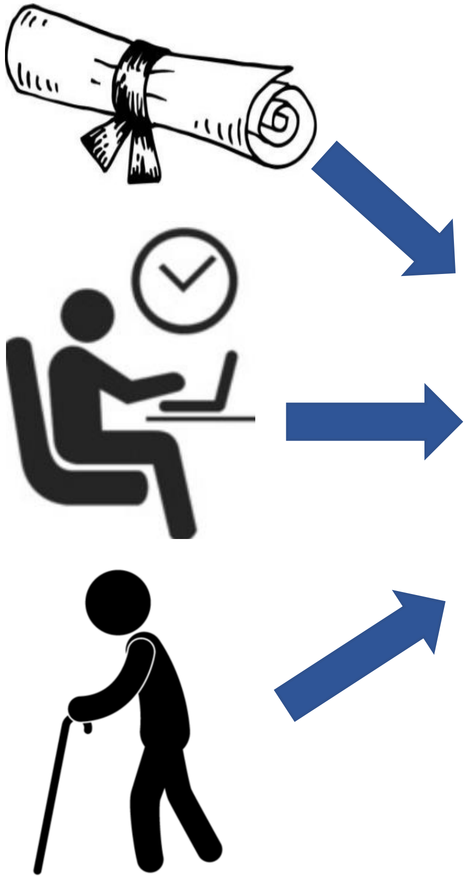
Linear Regression



Multiple Linear Regression



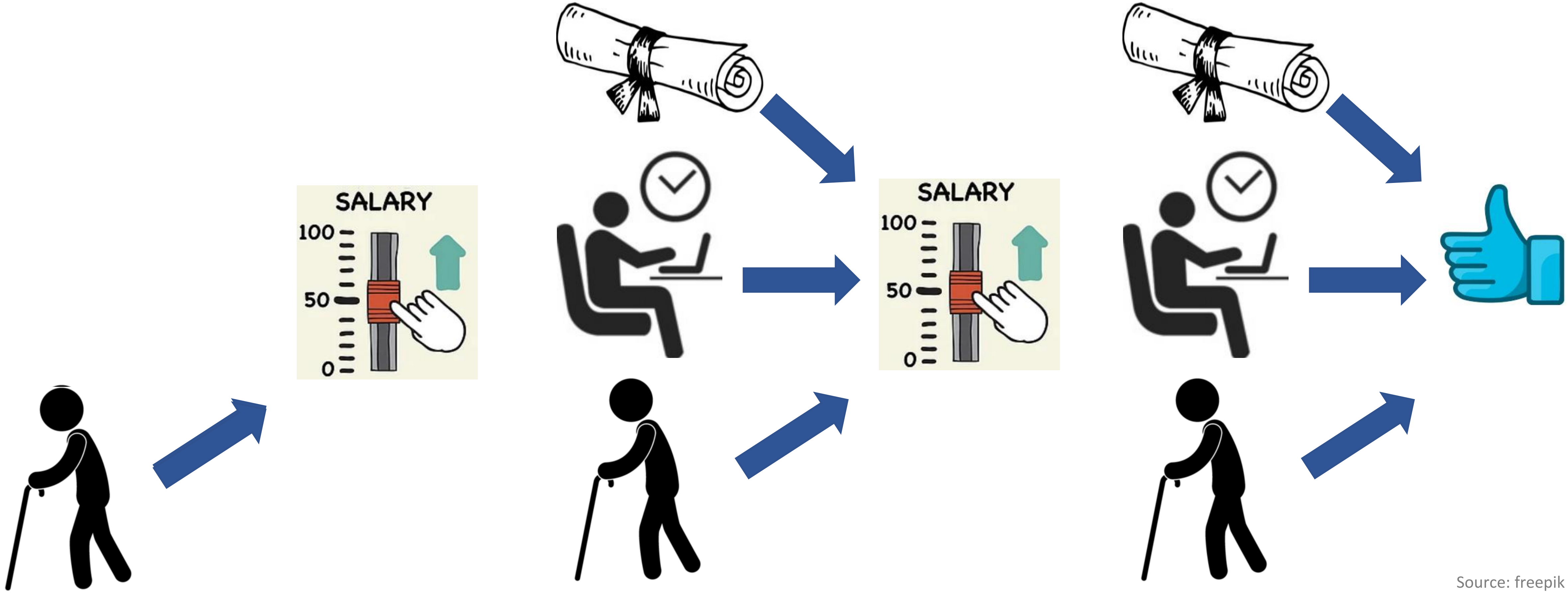
Logistic Regression



Source: freepik

Types of Regression

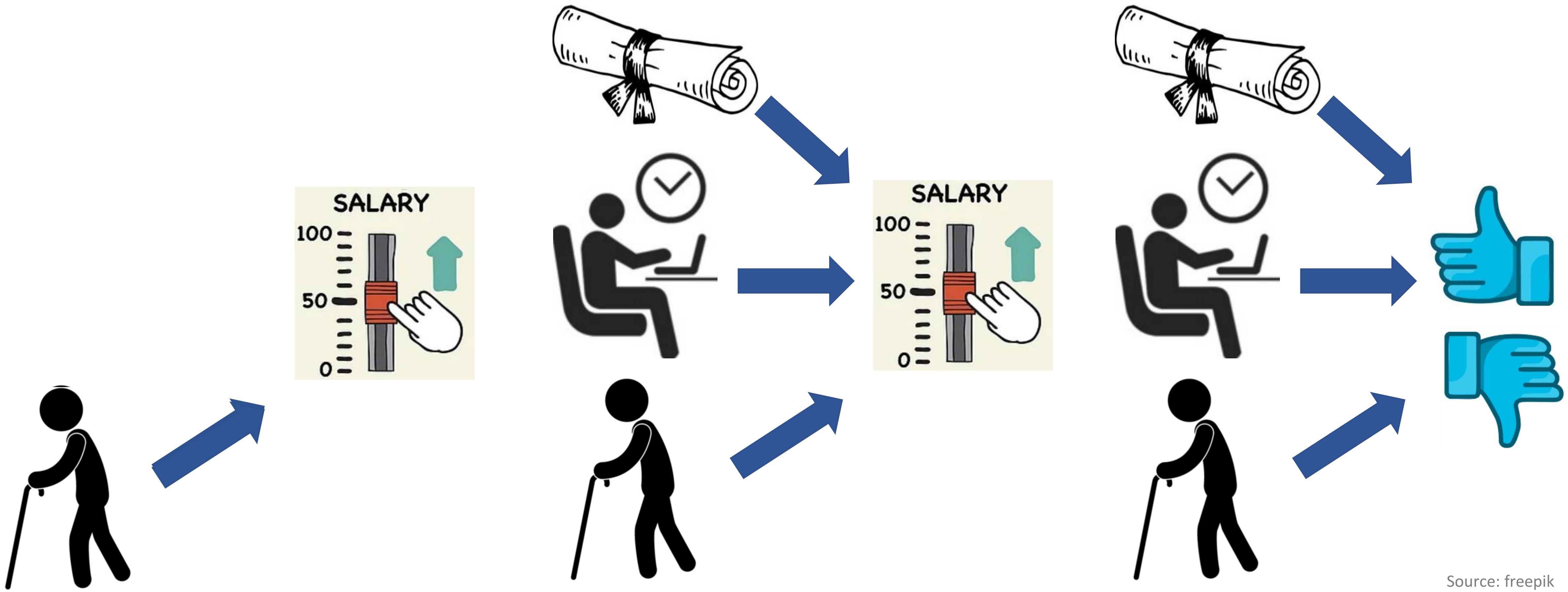
Linear Regression | **Multiple Linear** Regression | **Logistic** Regression



Source: freepik

Types of Regression

Linear Regression | **Multiple Linear** Regression | **Logistic** Regression



Source: freepik

Logistic Regression vs Classification

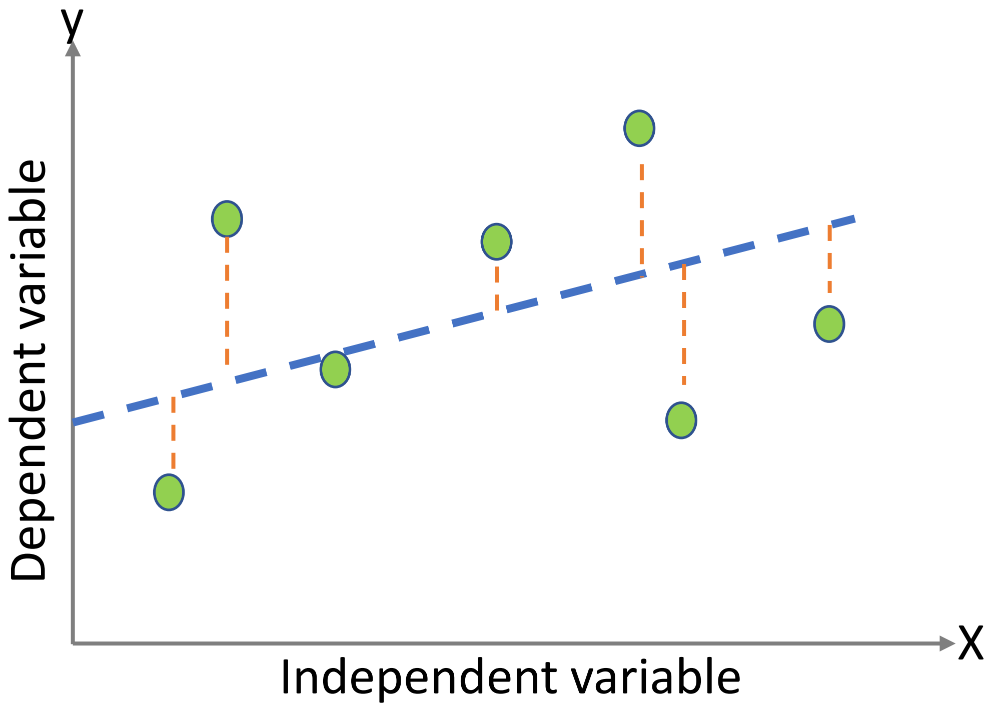
Feature	Logistic regression	Classification
Type of model	Probabilistic	Non-probabilistic
Number of outcomes	Binary	Any number
Interpretability	More interpretable	Less interpretable
Task	Predicting the probability of an outcome	Predicting the category of an outcome

Linear Regression

Linear Regression is a method to predict dependent variable (Y) based on values of independent variables (x). It can be used for the cases where we want to predict some continuous quantity.

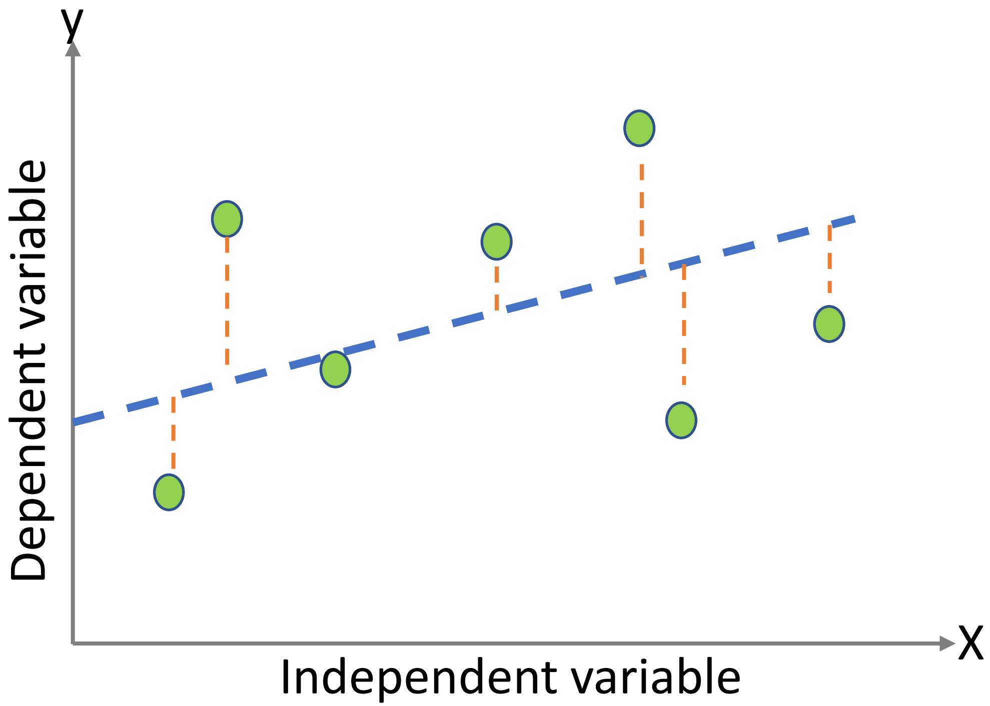
Linear Regression

Linear Regression is a method to predict dependent variable (Y) based on values of independent variables (x). It can be used for the cases where we want to predict some continuous quantity.



Linear Regression

Linear Regression is a method to predict dependent variable (Y) based on values of independent variables (x). It can be used for the cases where we want to predict some continuous quantity.



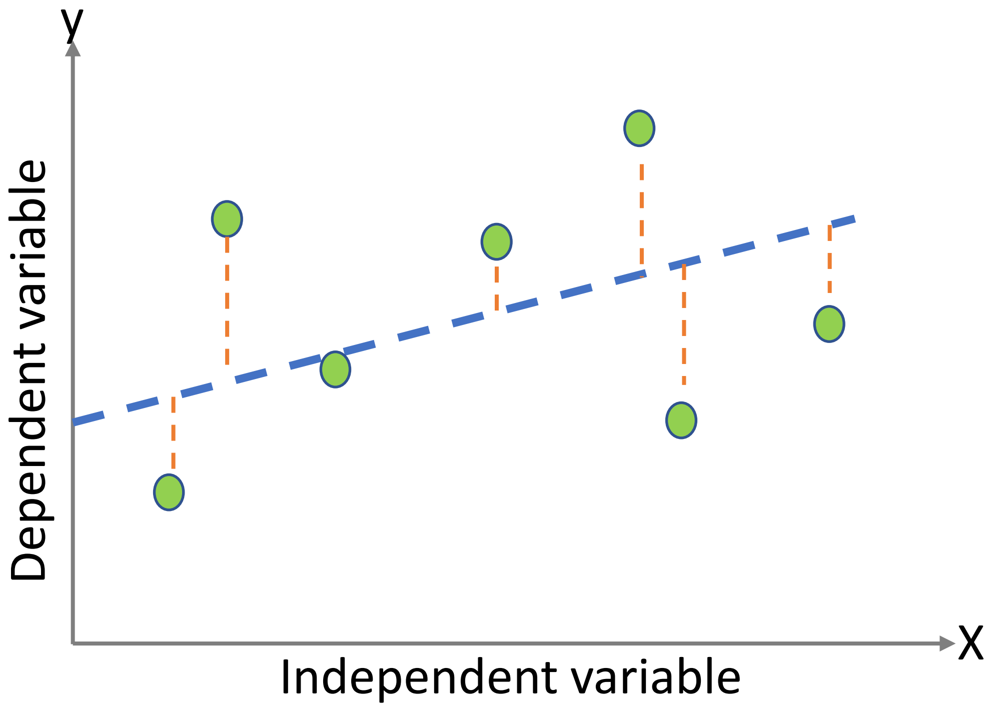
The following equation is used to represent a linear regression model:

$$Y = b_0 + b_1 x + e$$

Linear Regression

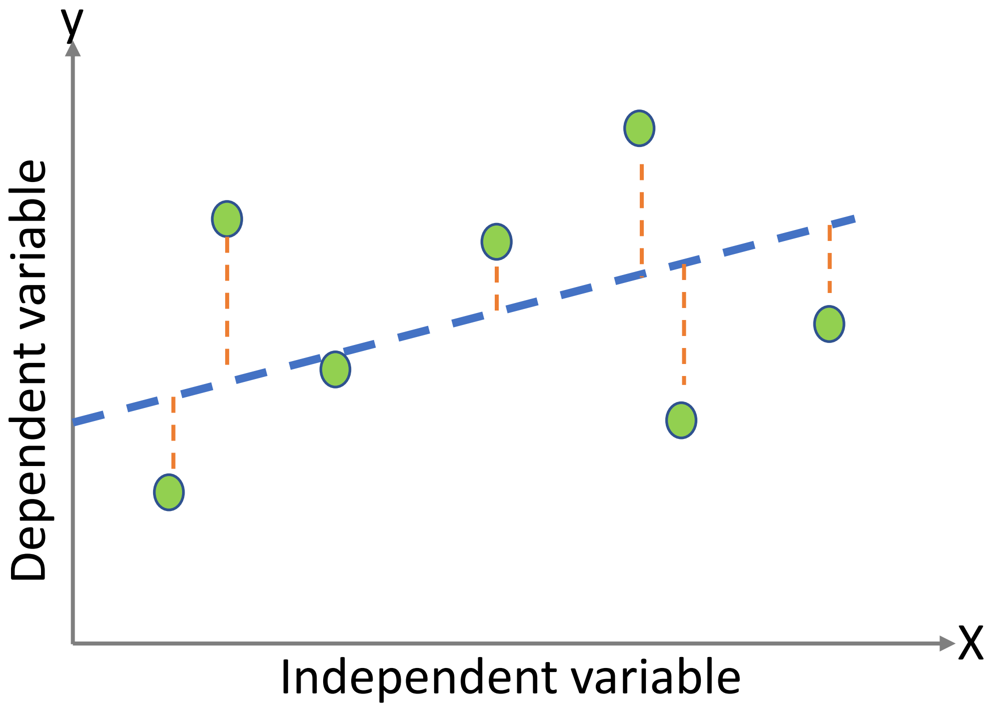
It depicts the relationship between one dependent and one or more independent variable. An example and its components are explained below:

$$Y = b_0 + b_1 x + e$$



Linear Regression

It depicts the relationship between one dependent and one or more independent variable. An example and its components are explained below:

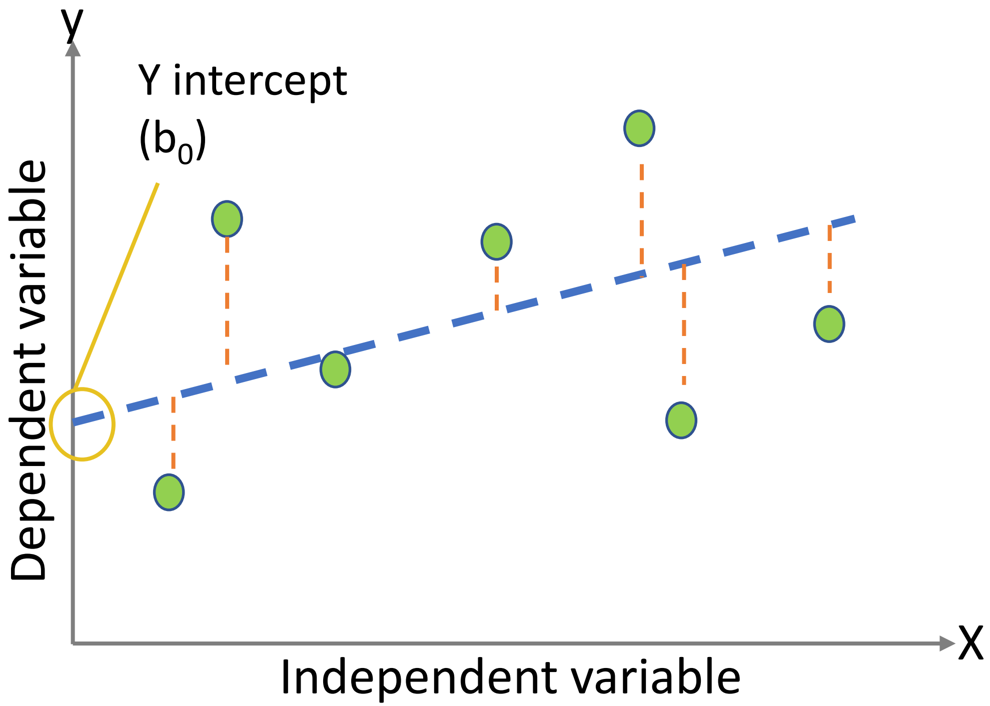


$$Y = b_0 + b_1 x + e$$

Dependent variable

Linear Regression

It depicts the relationship between one dependent and one or more independent variable. An example and its components are explained below:



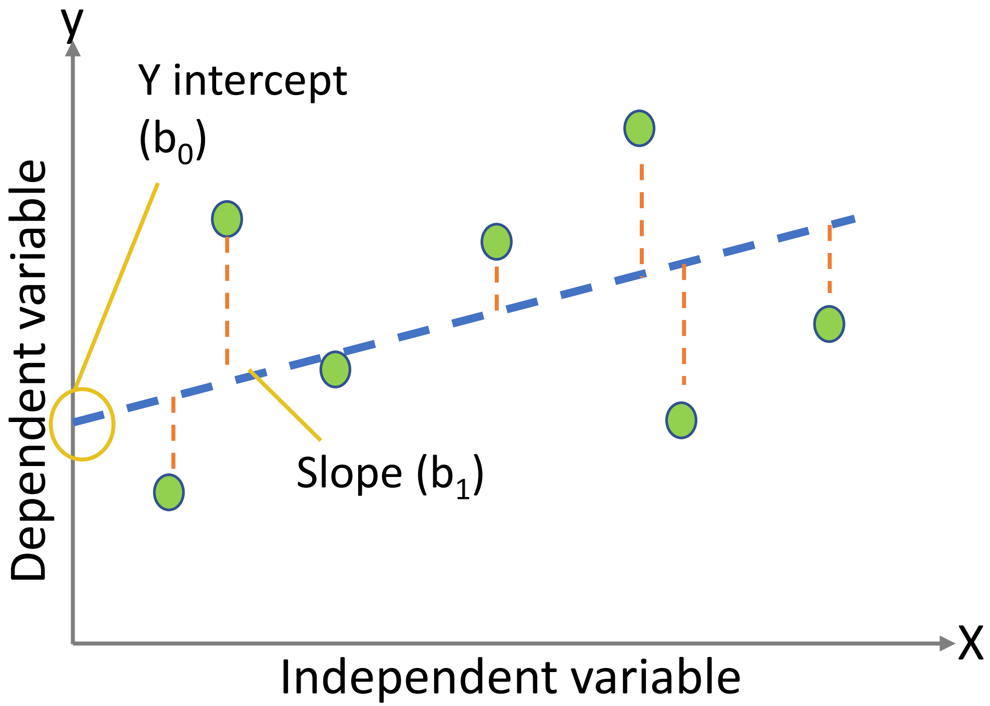
$$Y = b_0 + b_1 x + e$$

Y intercept

Dependent variable

Linear Regression

It depicts the relationship between one dependent and one or more independent variable. An example and its components are explained below:



$$Y = b_0 + b_1 x + e$$

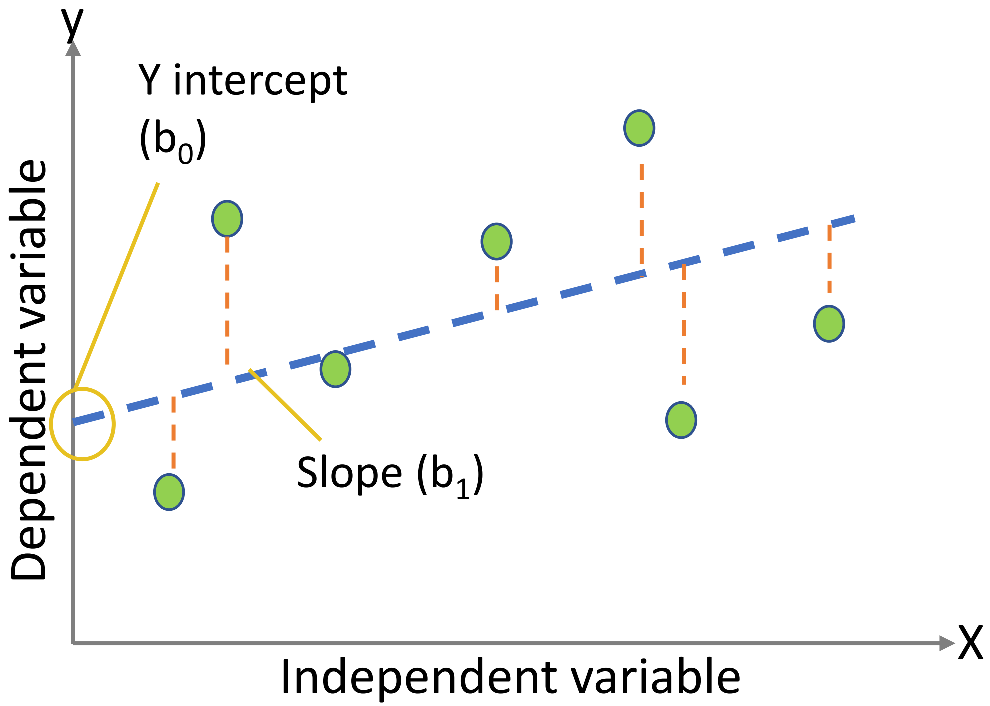
Slope

Y intercept

Dependent variable

Linear Regression

It depicts the relationship between one dependent and one or more independent variable. An example and its components are explained below:



$$Y = b_0 + b_1 x + e$$

Independent variable

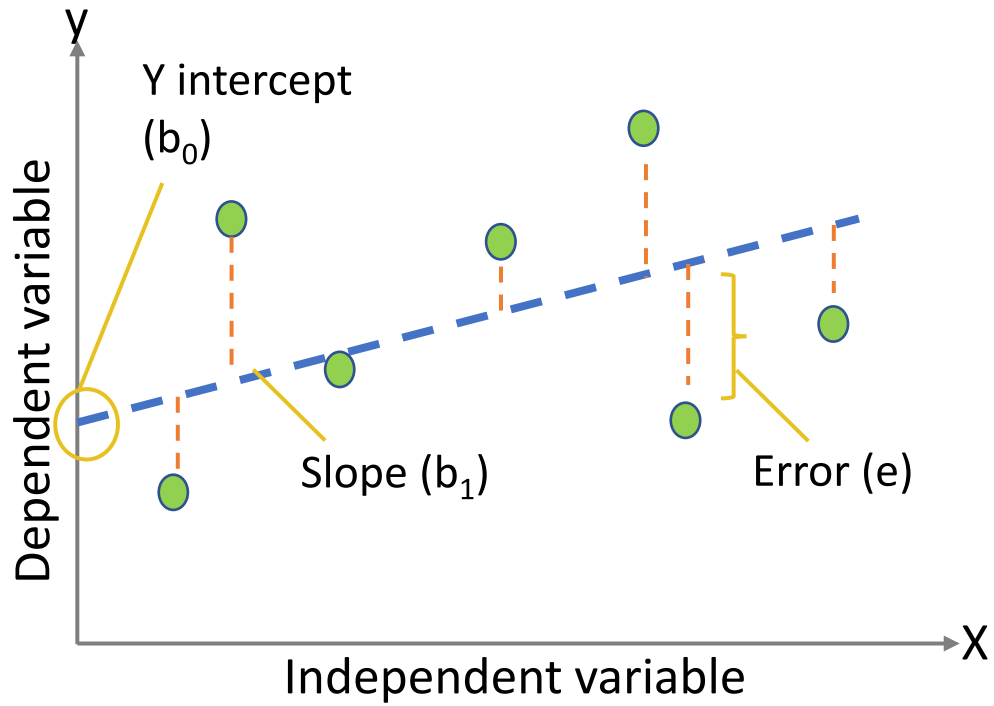
Slope

Y intercept

Dependent variable

Linear Regression

It depicts the relationship between one dependent and one or more independent variable. An example and its components are explained below:



$$Y = b_0 + b_1 x + e$$

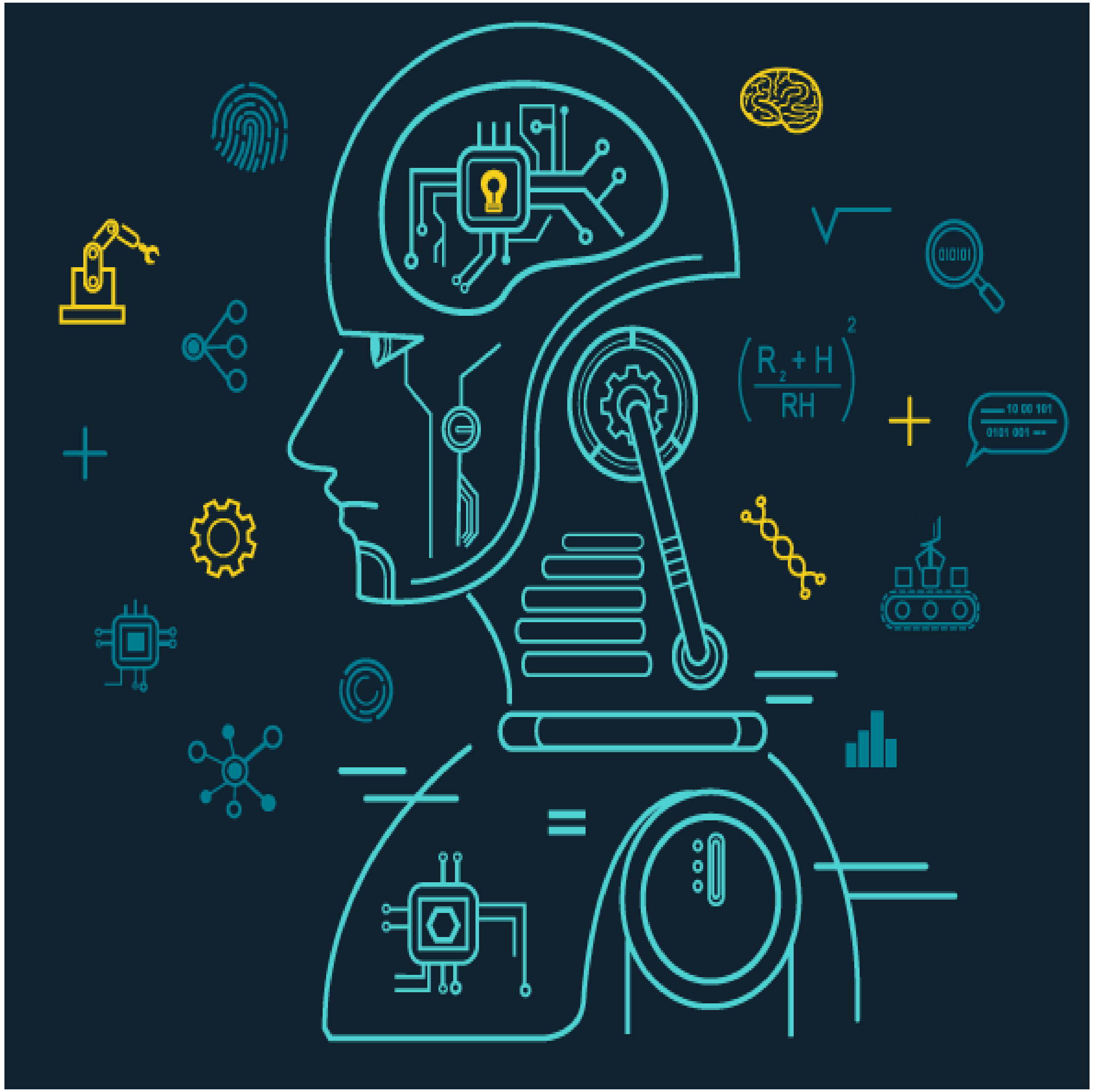
Error

Independent variable

Slope

Y intercept

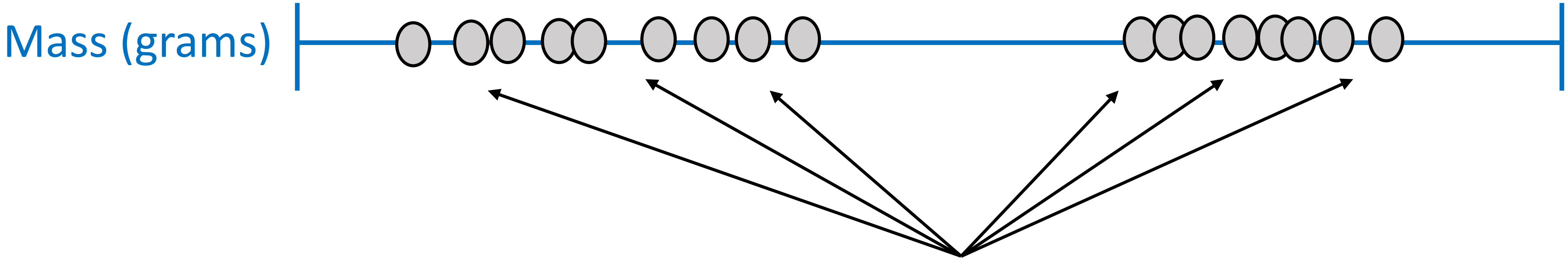
Dependent variable



Machine Learning

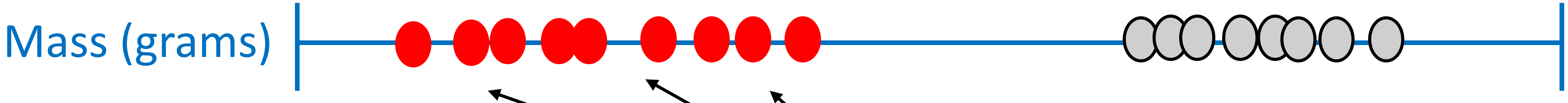
Support Vector Machines

Intuition Behind the Support Vector Machines



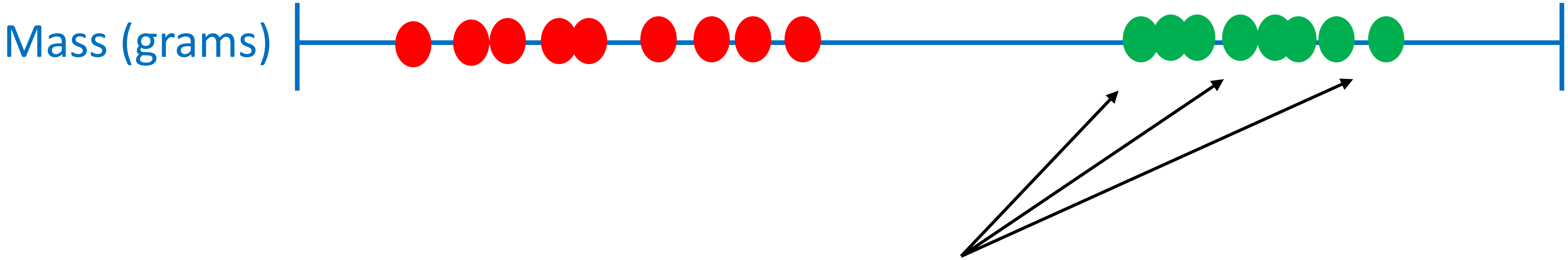
Let's start by imaging we measured the **mass** of a bunch of **mice…..**

Intuition Behind the Support Vector Machines



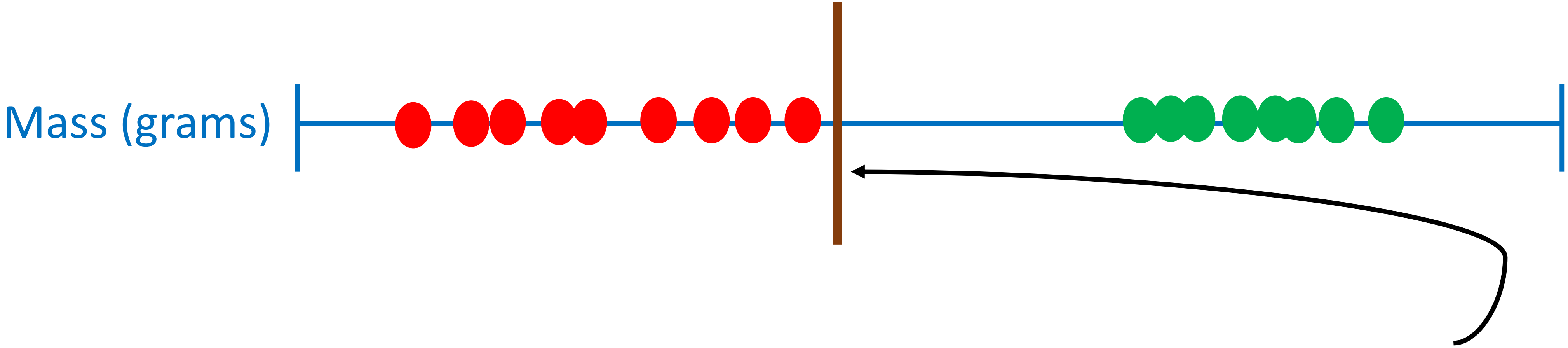
Red dot's represents mice are not ***obese…***

Intuition Behind the Support Vector Machines



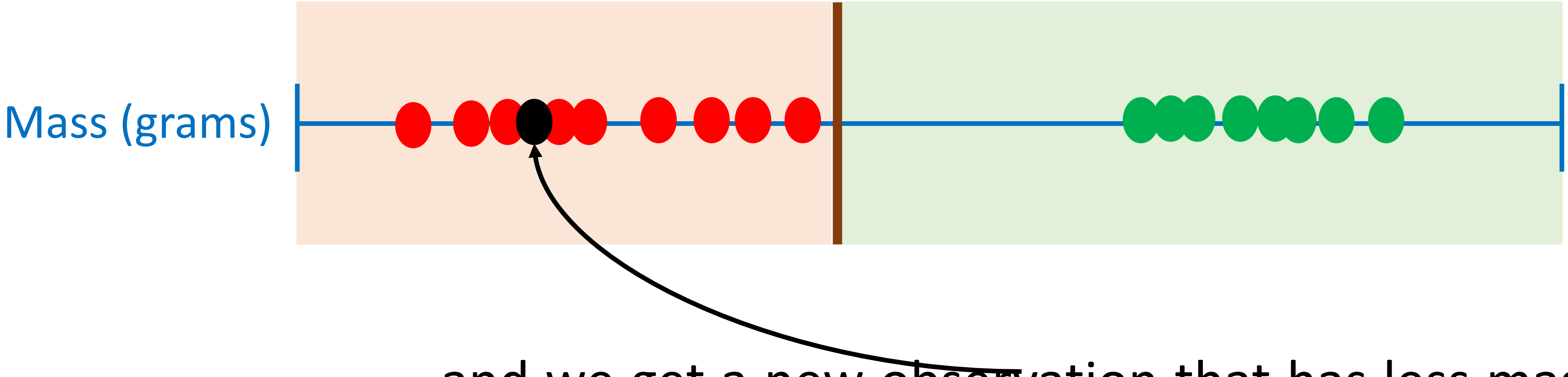
Green dot’s represents mice are ***obese…***

Intuition Behind the Support Vector Machines



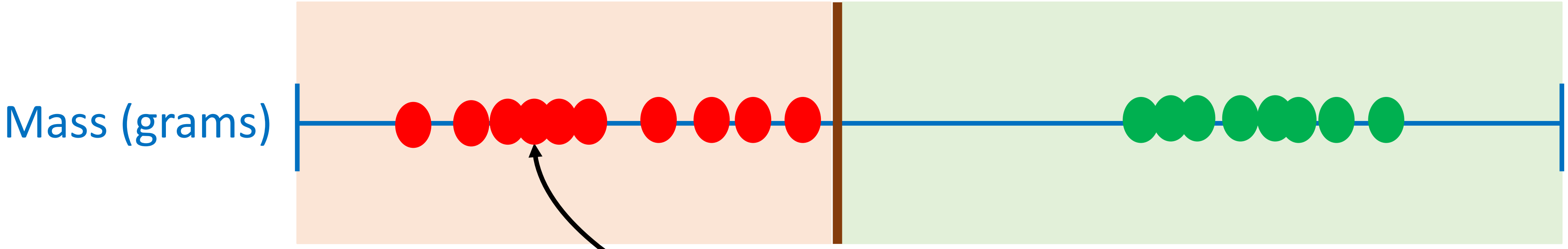
Based on these observations we pick a line (threshold) separating two classes.

Intuition Behind the Support Vector Machines



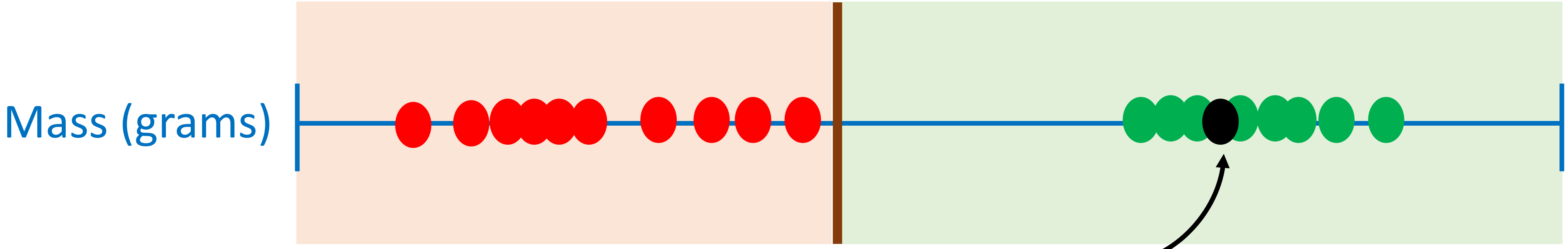
..and we get a new observation that has less mass than the threshold.

Intuition Behind the Support Vector Machines



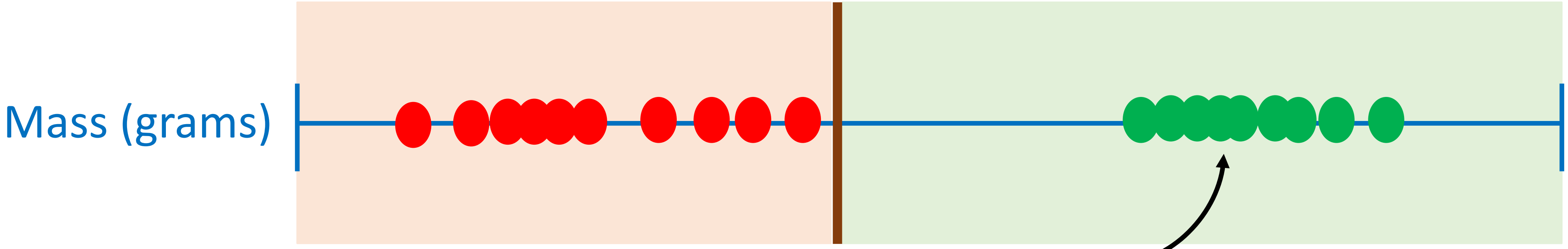
..and we get a new observation that has less mass than the threshold. We can classify it as not obese.

Intuition Behind the Support Vector Machines



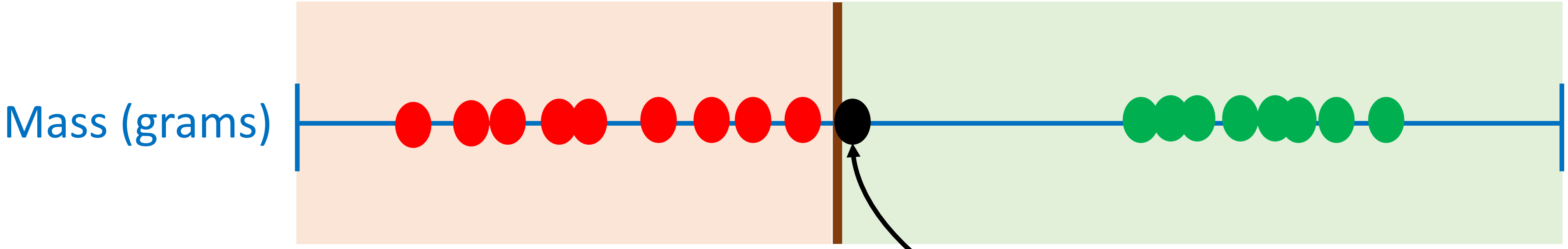
..and we get a new observation that has more mass than the threshold.

Intuition Behind the Support Vector Machines



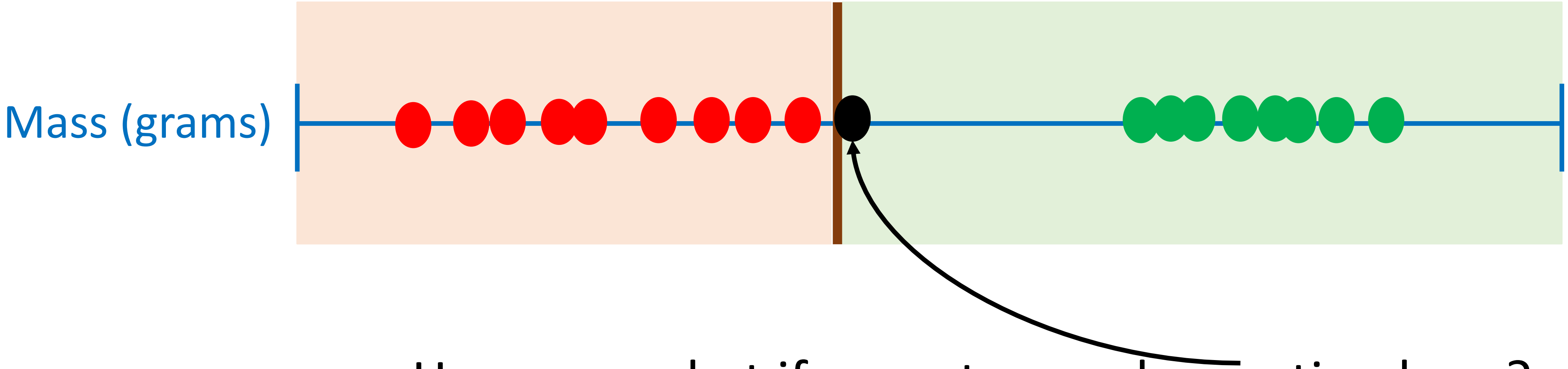
..and we get a new observation that has more mass than the threshold. We can classify it as a obese.

Intuition Behind the Support Vector Machines



However, what if we get new observation here?

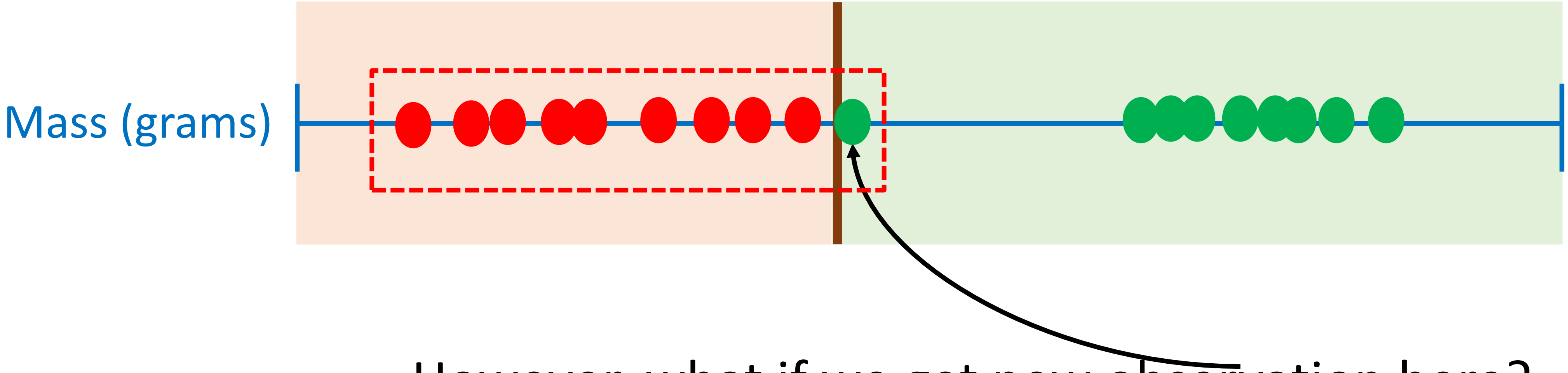
Intuition Behind the Support Vector Machines



However, what if we get new observation here?

Because this observation has more mass than the threshold, we classify it as **obese**.

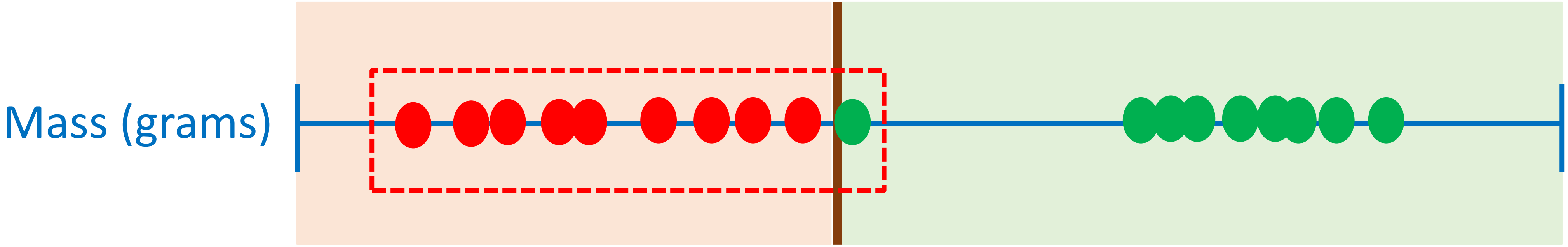
Intuition Behind the Support Vector Machines



However, what if we get new observation here?
Because this observation has more mass than the threshold, we classify it as **obese**.

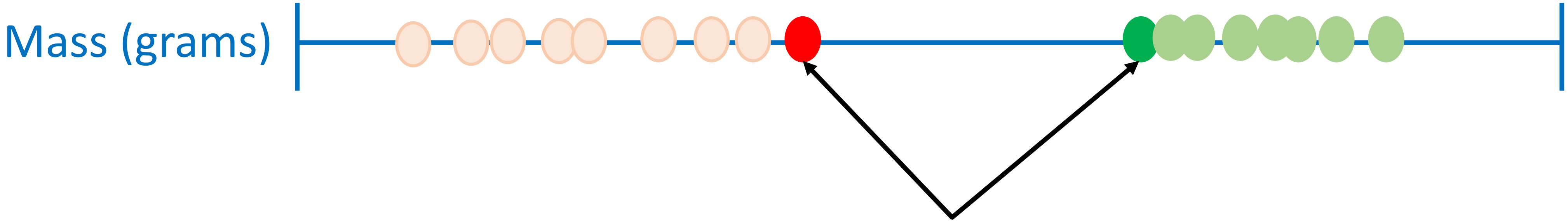
But it doesn't make sense.

Intuition Behind the Support Vector Machines



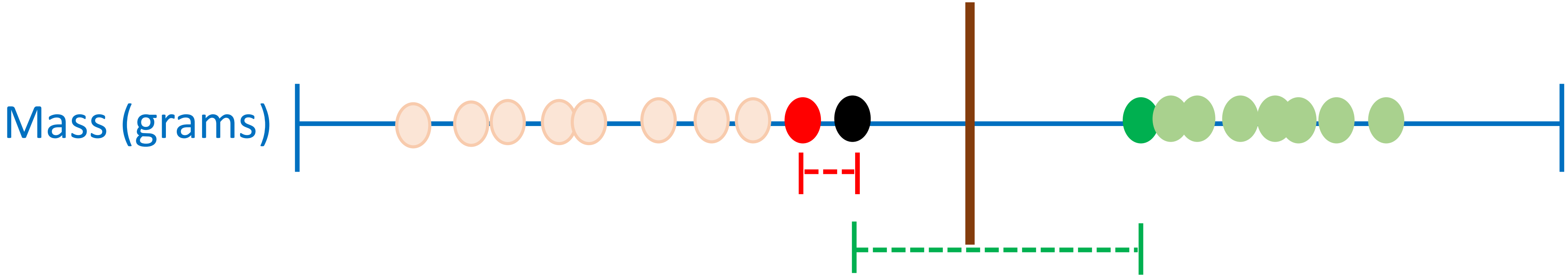
So this threshold is pretty lame. Can we do better?

Support Vector Machines



Focus on the observations on the edges of the each category.

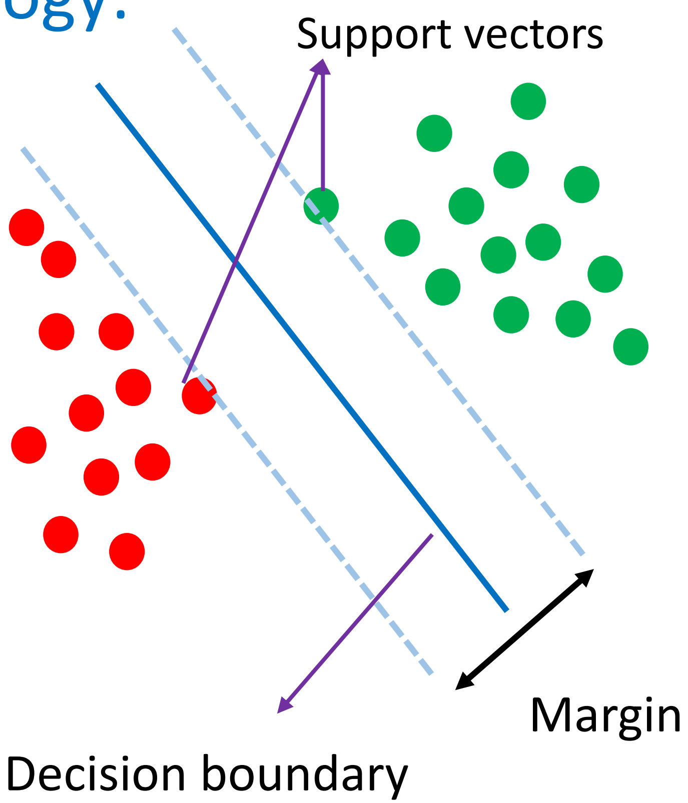
Support Vector Machines



Use mid point between them as a threshold.

Support Vector Machines

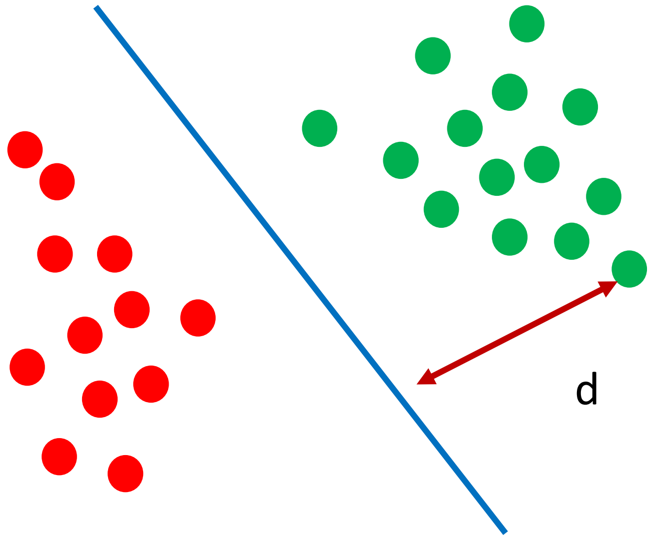
Terminology:



Margin (M): The distance of the closest data point to the decision boundary.

Support Vector Machines

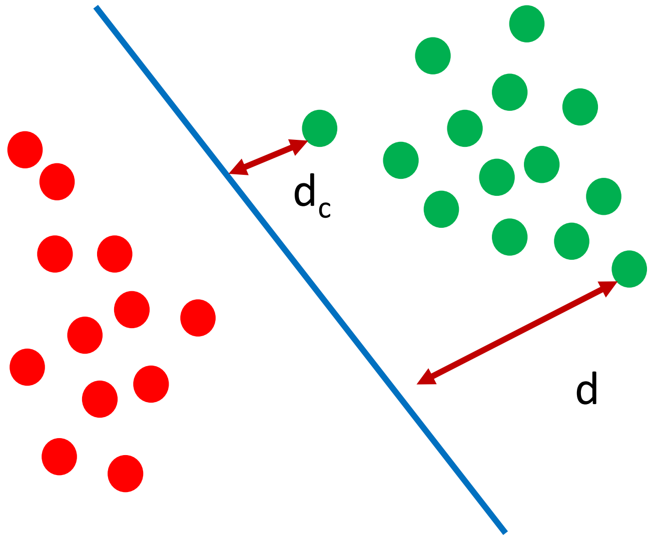
Terminology:



Margin (M): The distance of the closest data point to the decision boundary.

Support Vector Machines

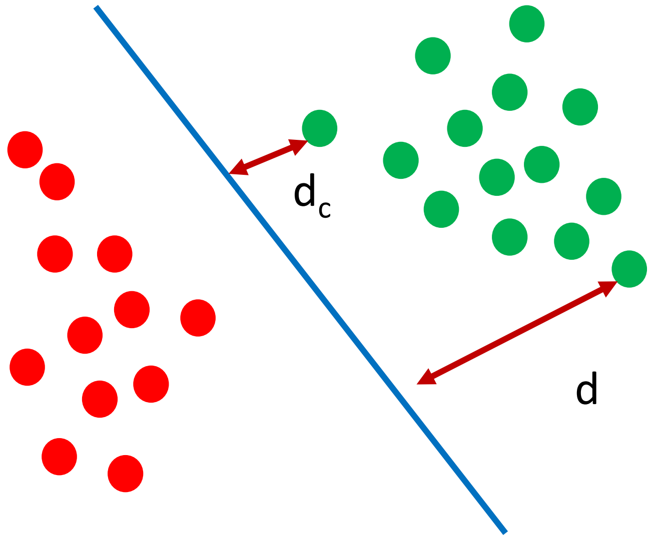
Terminology:



Margin (M): The distance of the closest data point to the decision boundary.

Support Vector Machines

Terminology:



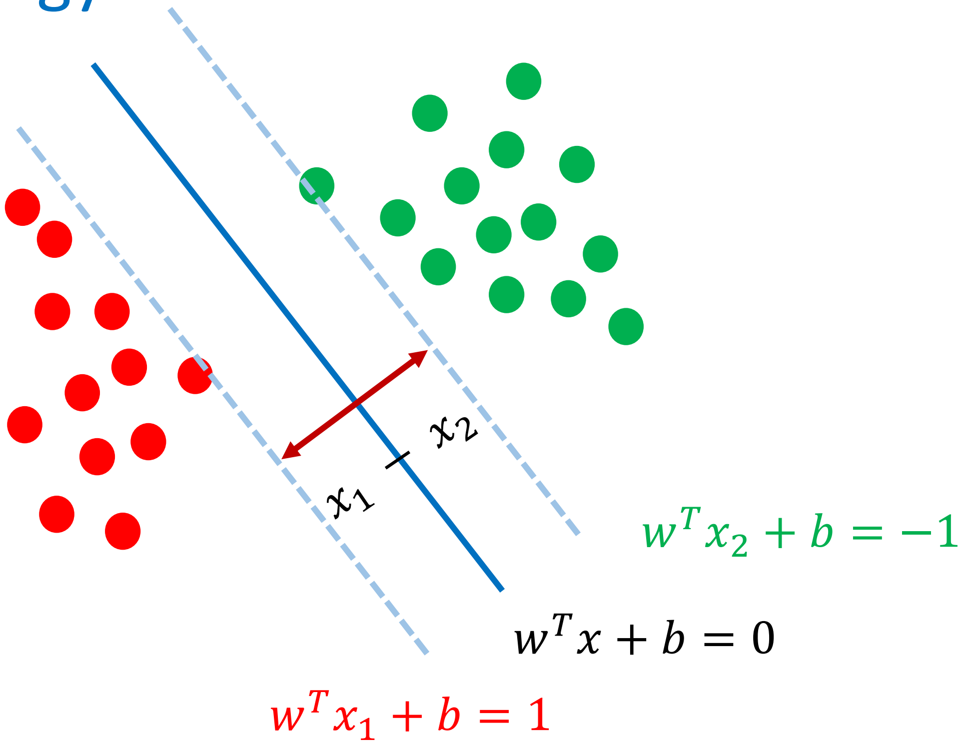
Margin (M): The distance of the closest data point to the decision boundary.

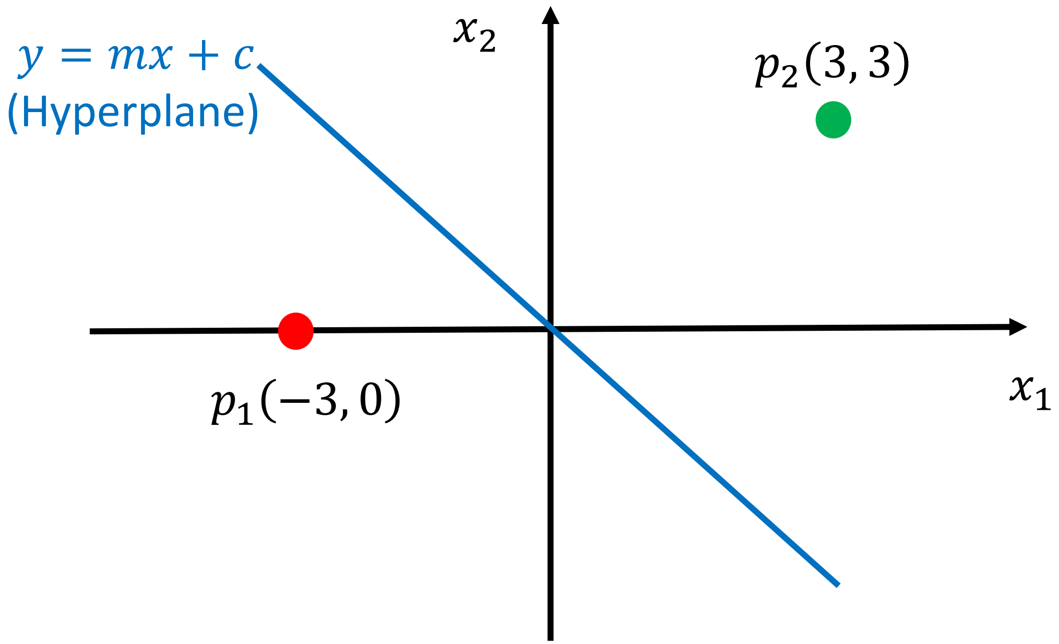
Find out the decision boundary that **maximizes** the margin (M).

This is called Maximal Margin Classifier (Hard Margin).

Support Vector Machines

Terminology:



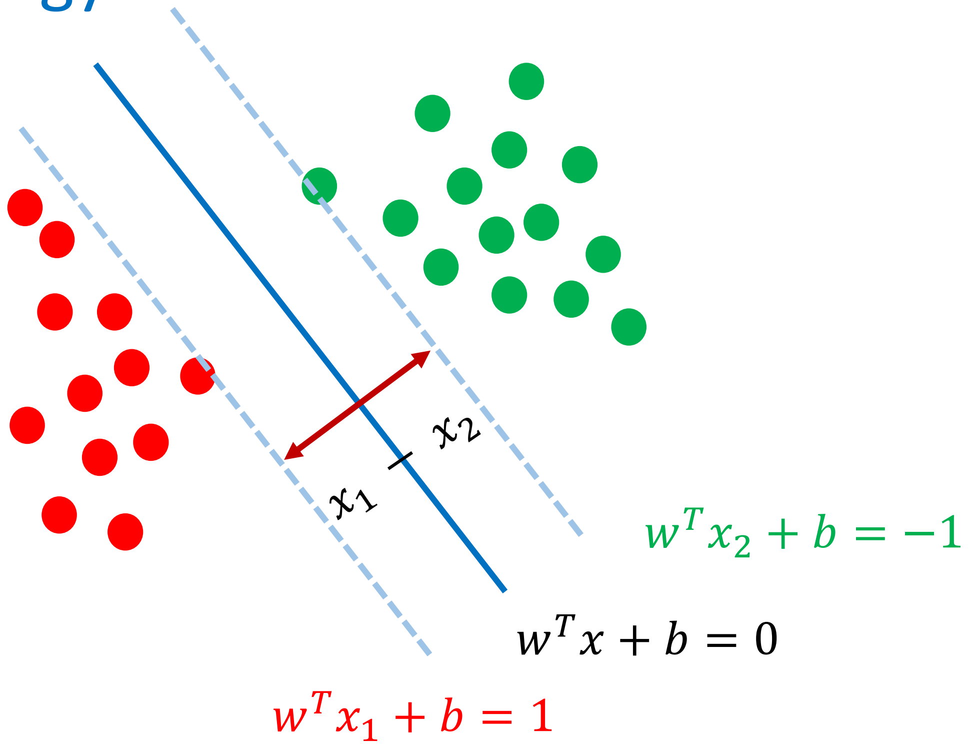


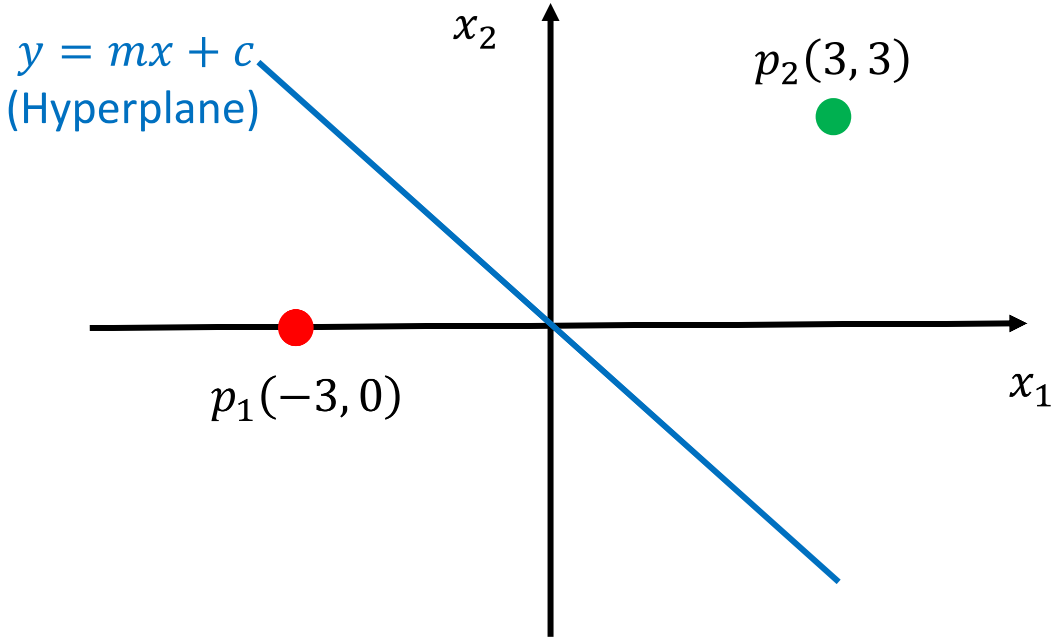
Slope of line m = -1

c = 0

Support Vector Machines

Terminology:





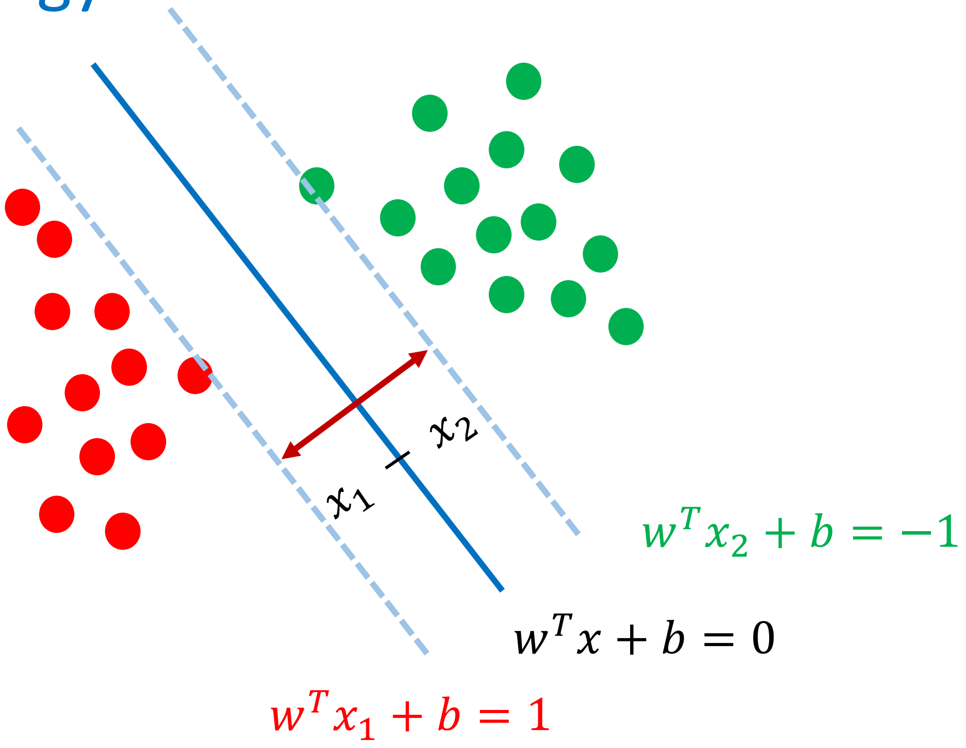
Slope of line m = -1

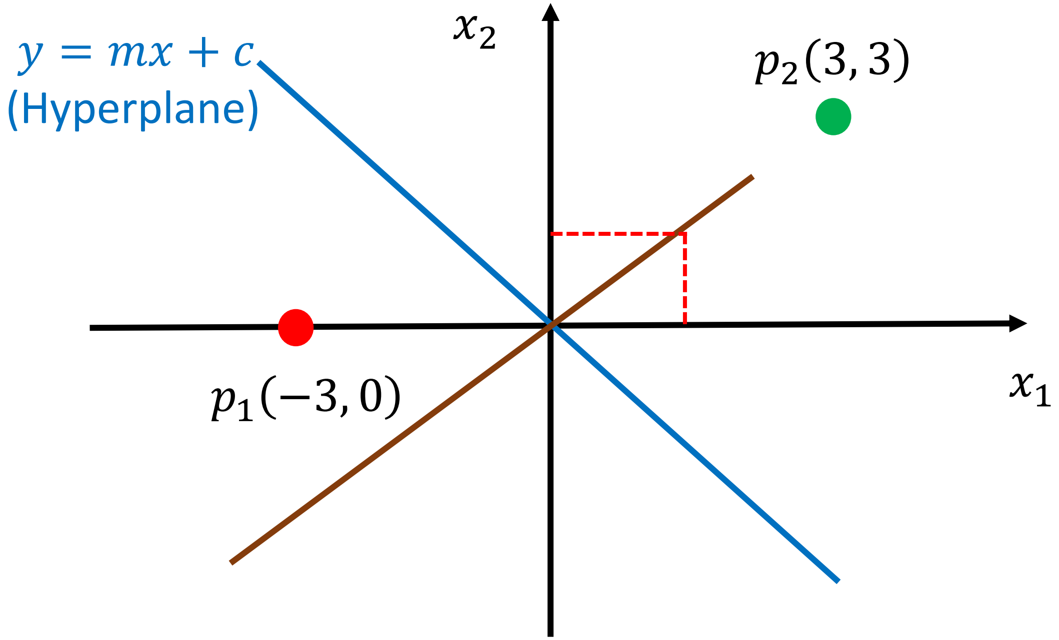
Intercept c = 0

$W \rightarrow parameters\ of\ the\ line\ (m, c) = (-1{,}0)$

Support Vector Machines

Terminology:



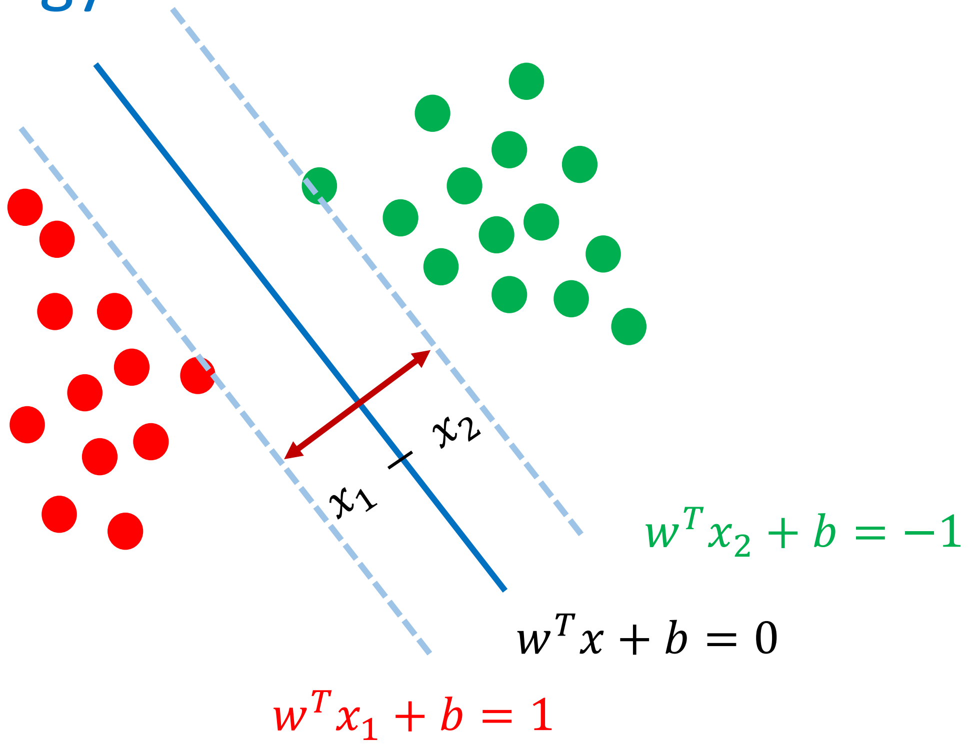


Slope of line m = 1

c = 0

Support Vector Machines

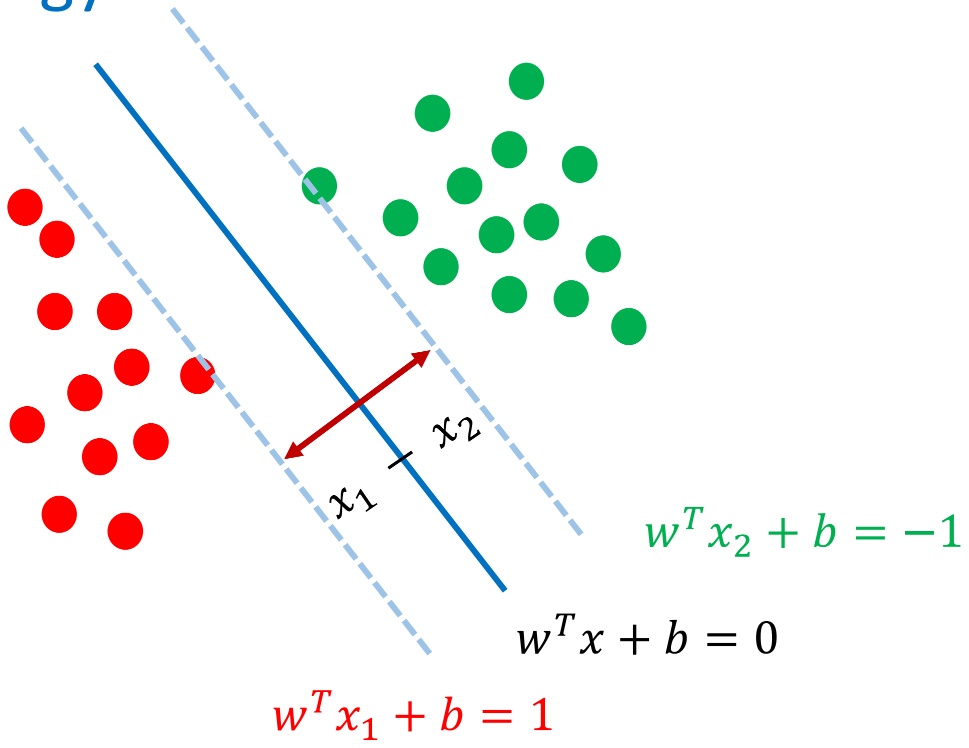
Terminology:



$$
\begin{array}{rl}
 & w^Tx_1 + b = 1 \\
(-) & w^Tx_2 + b = -1 \\
\hline
 & w^T(x_1 - x_2) = 2
\end{array}
$$

Support Vector Machines

Terminology:



$$
\begin{array}{rl}
 & w^T x_1 + b = 1 \\
(-) & w^T x_2 + b = -1 \\
\hline
 & w^T(x_1 - x_2) = 2
\end{array}
$$

Divide by $\|w\|$

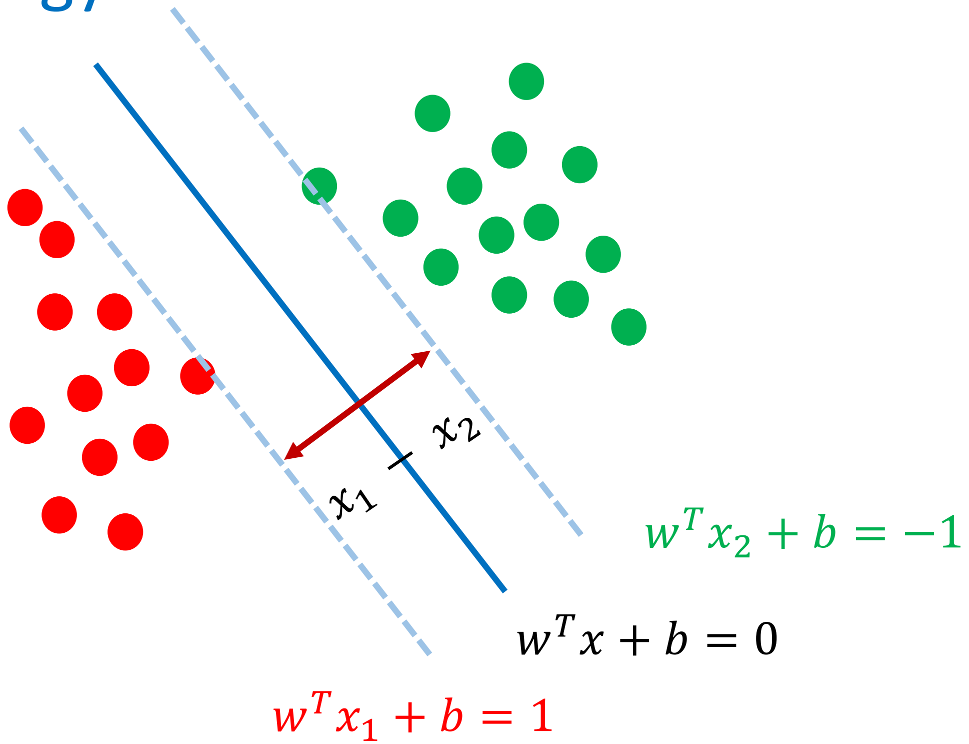
Magnitude of the vector

$$\frac{w^T}{\|w\|}(x_1 - x_2) = \frac{2}{\|w\|}$$

$$(x_1 - x_2) = \frac{2}{\|w\|}$$

Support Vector Machines

Terminology:

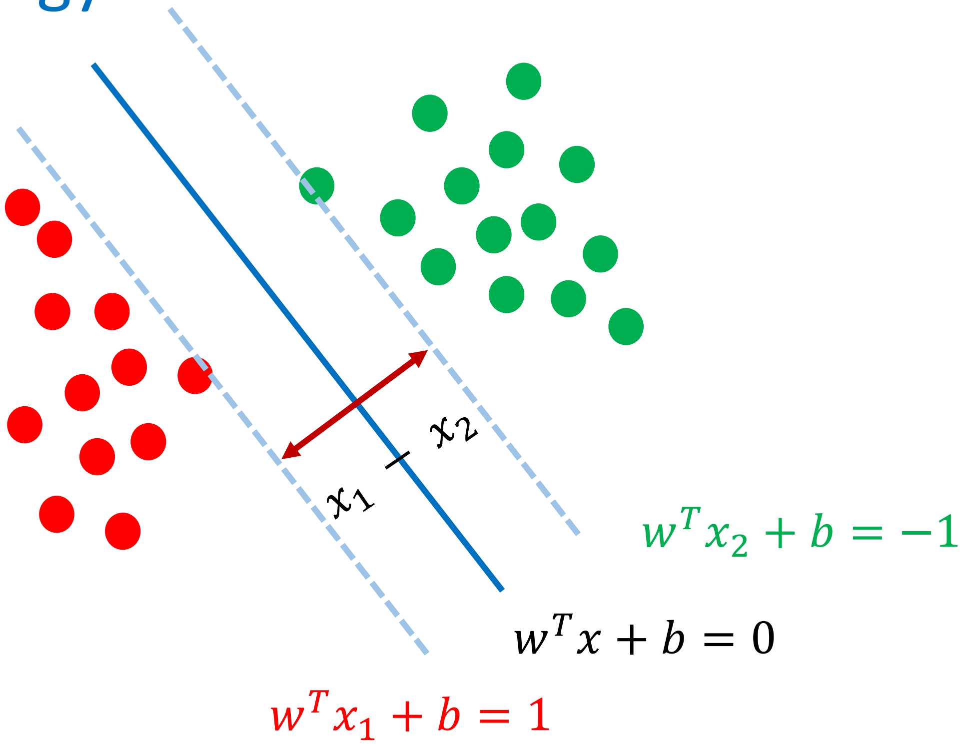


$$max\left(\frac{2}{\|w\|}\right)$$

$$y_i = \begin{cases} -1 & w^T x_1 + b \leq 1 \\ 1 & w^T x_1 + b \geq 1 \end{cases}$$

Support Vector Machines

Terminology:



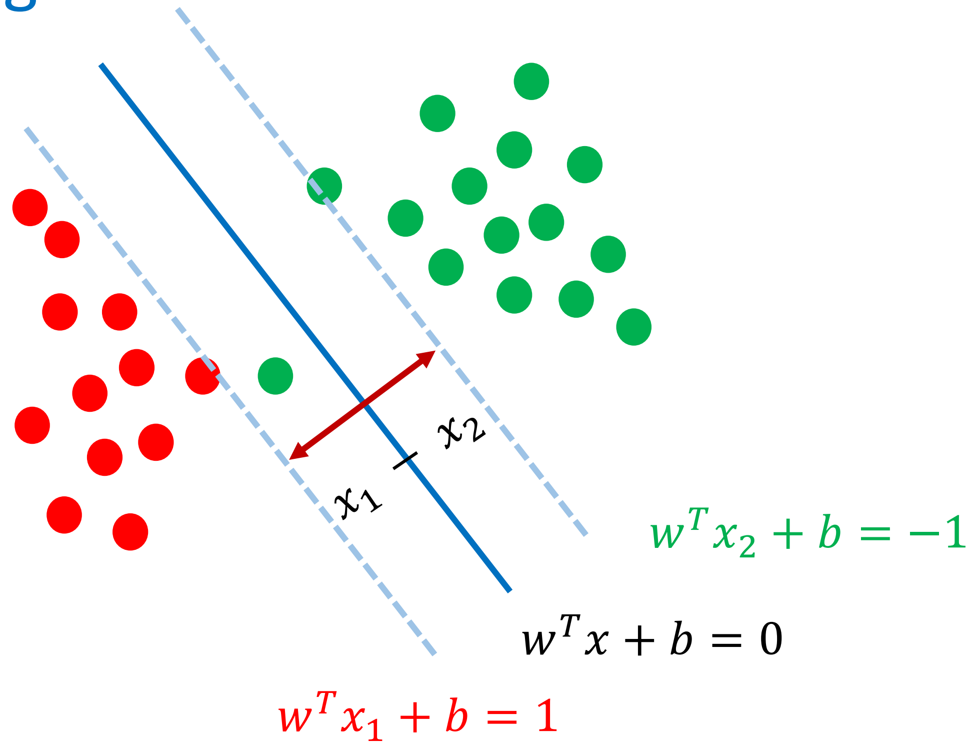
$$max\left(\frac{2}{\|w\|}\right)$$

$$y_i = \begin{cases} -1 & w^T x_1 + b \le 1 \\ 1 & w^T x_1 + b \ge 1 \end{cases}$$

$$min\left(\frac{\|w\|}{2}\right)$$

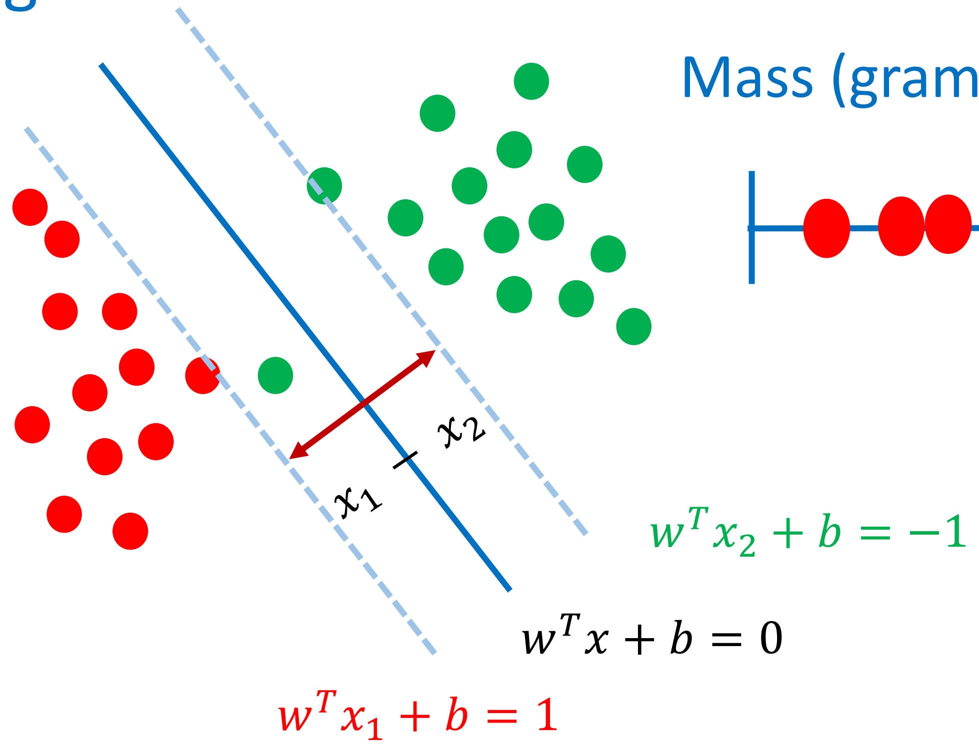
Support Vector Machines

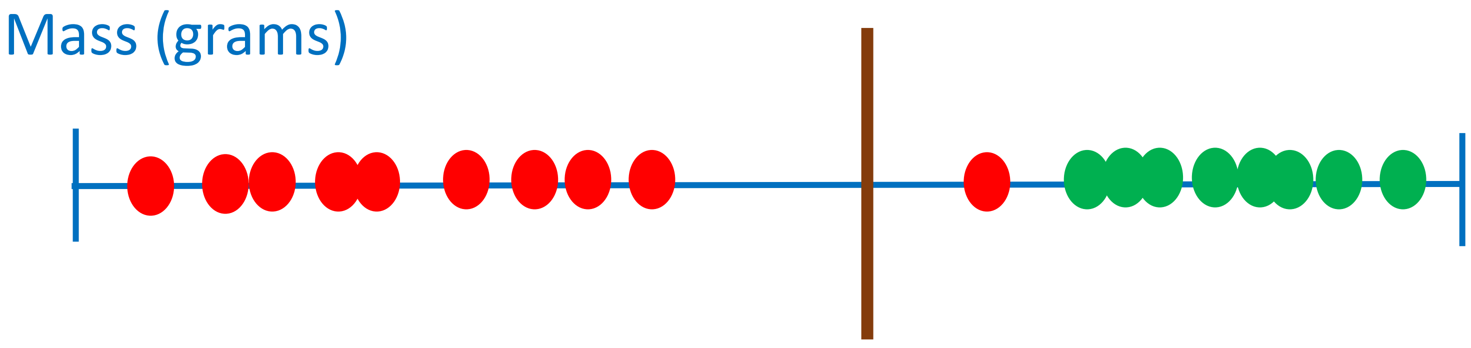
Soft Margin



Support Vector Machines

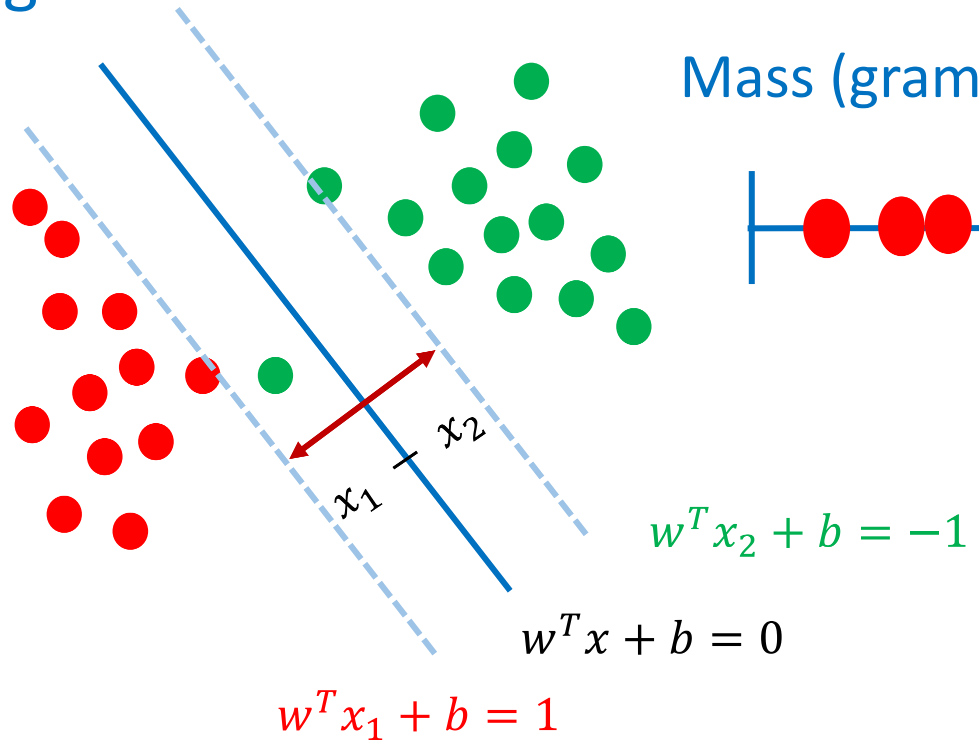
Soft Margin

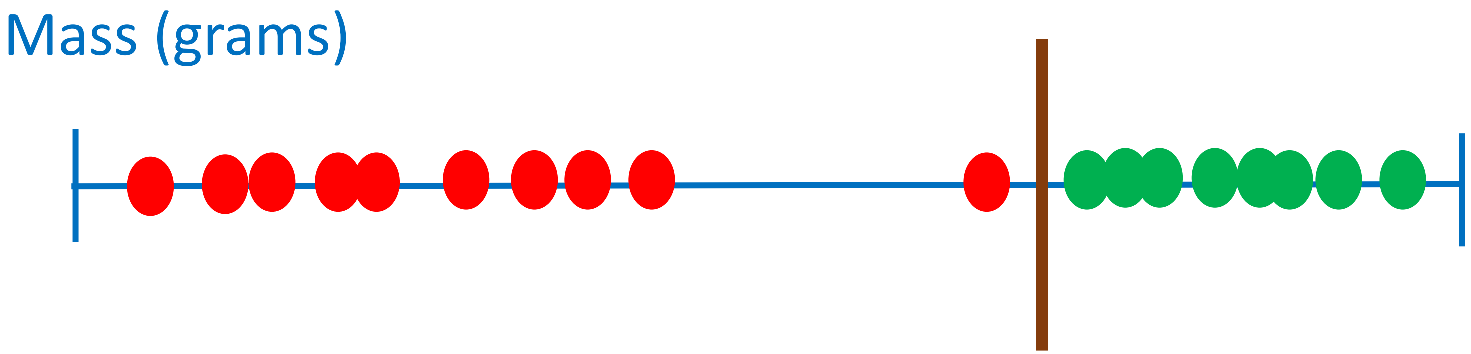




Support Vector Machines

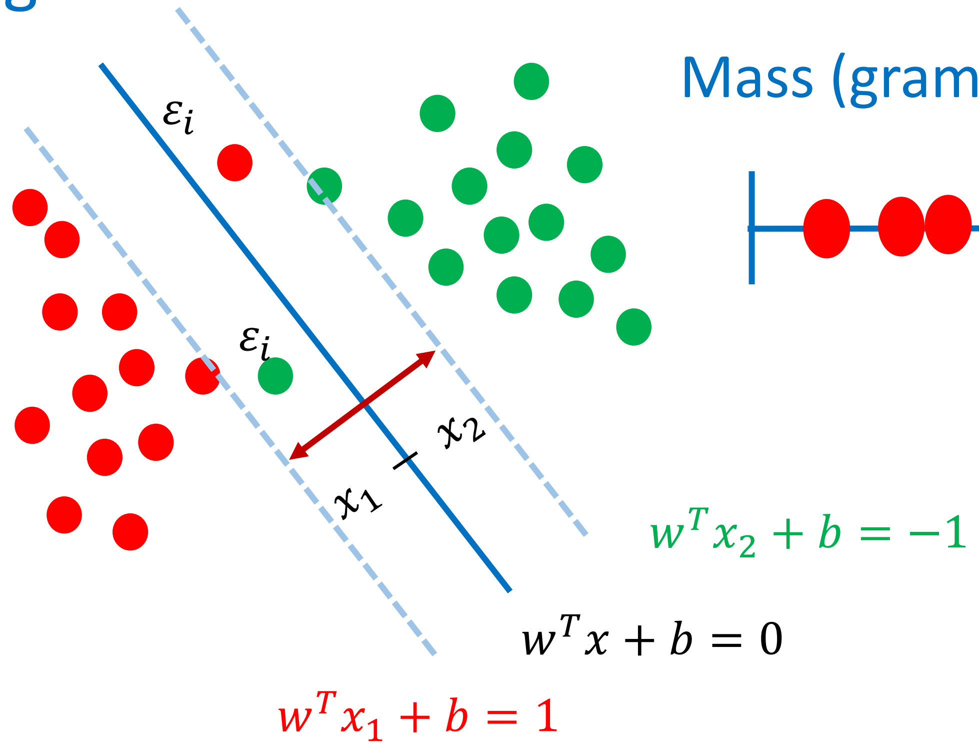
Soft Margin

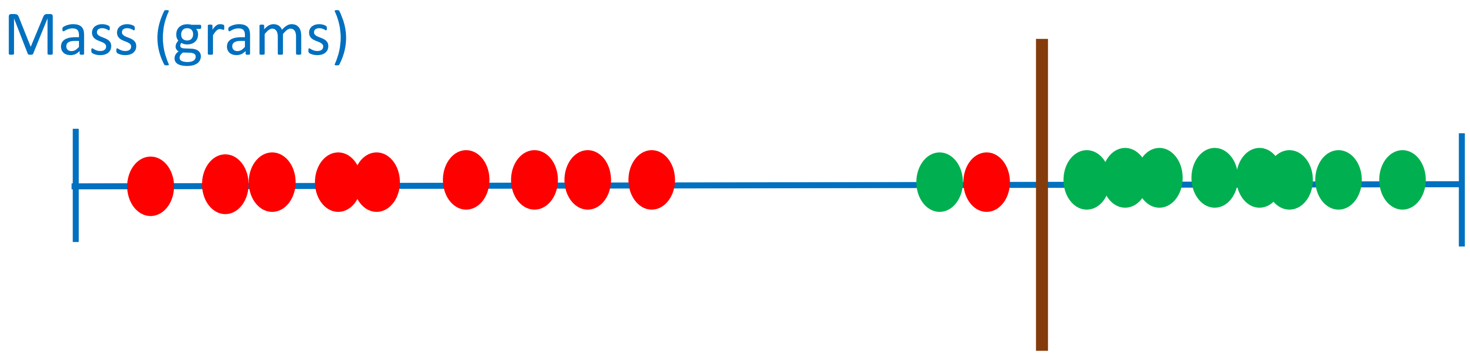




Support Vector Machines

Soft Margin

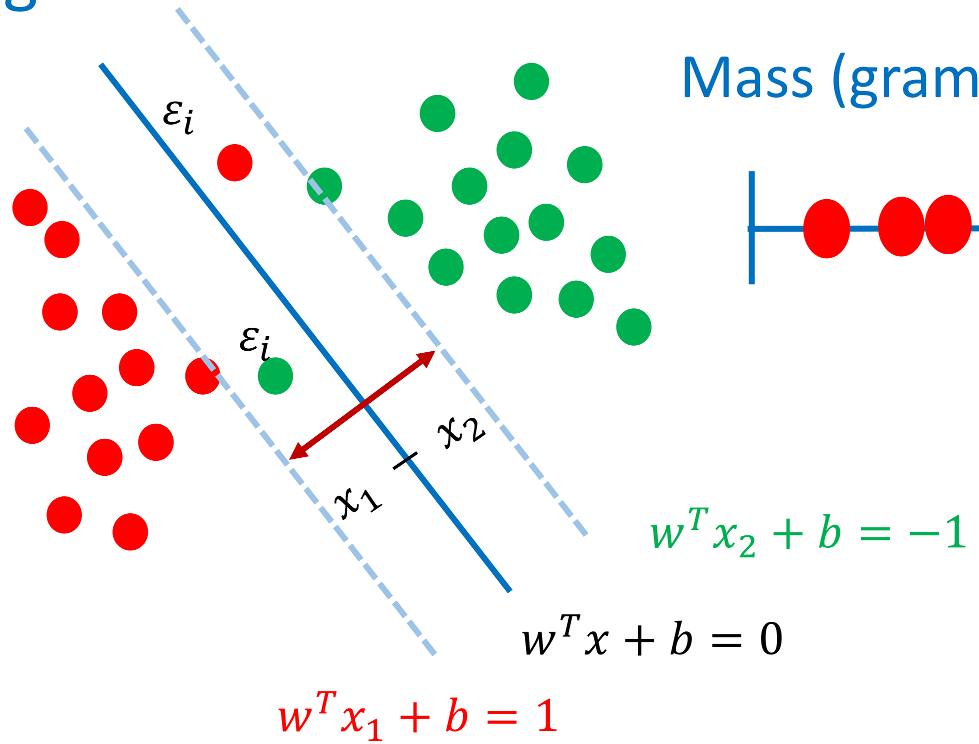


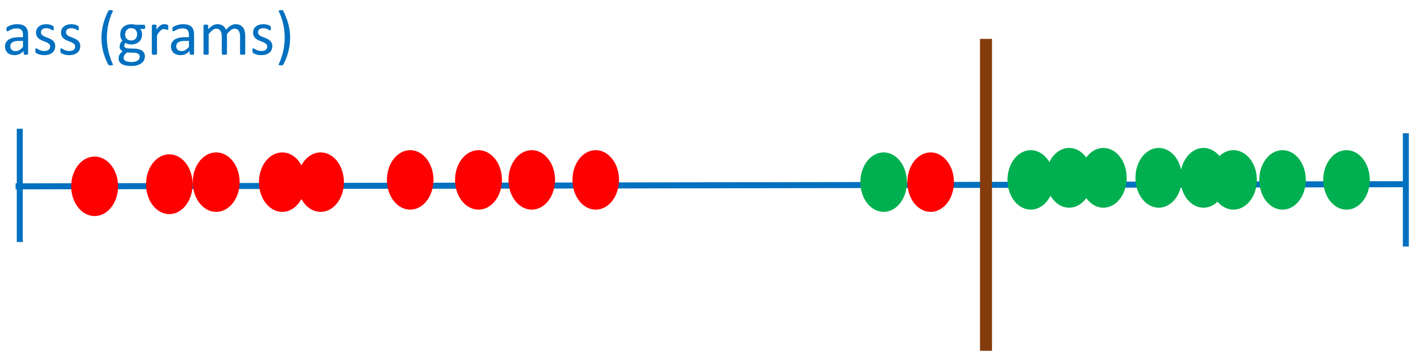


Here, we allow misclassifications.

Support Vector Machines

Soft Margin





Here, we allow misclassifications.

$$min\left(\frac{\|w\|}{2}\right) + c \, * \sum \varepsilon_i$$

Hinge Loss

Loss Function: (Hinge loss)

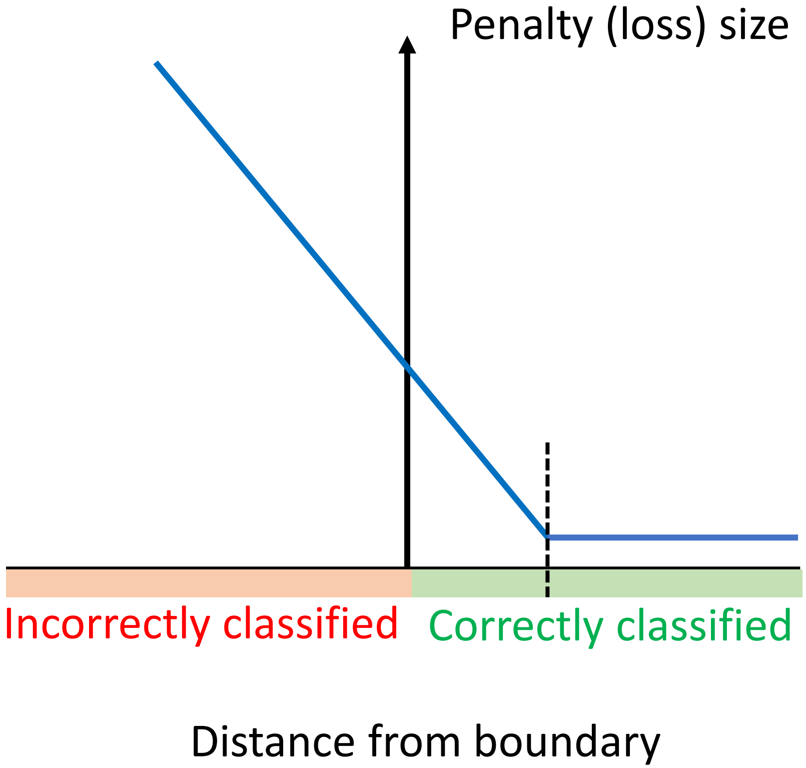
- Hinge loss is one of the types of loss function, mainly used for maximum-margin classification models.

- Hinge loss incorporates a margin or distance from the classification boundary into the calculation loss. Even if new observations are classified correctly, they can incur a penalty if the margin from the decision boundary is not enough.

Hinge Loss

Loss Function: (Hinge loss)

$$L = max\left(0, 1 - y_i(w^T x_I + b)\right)$$

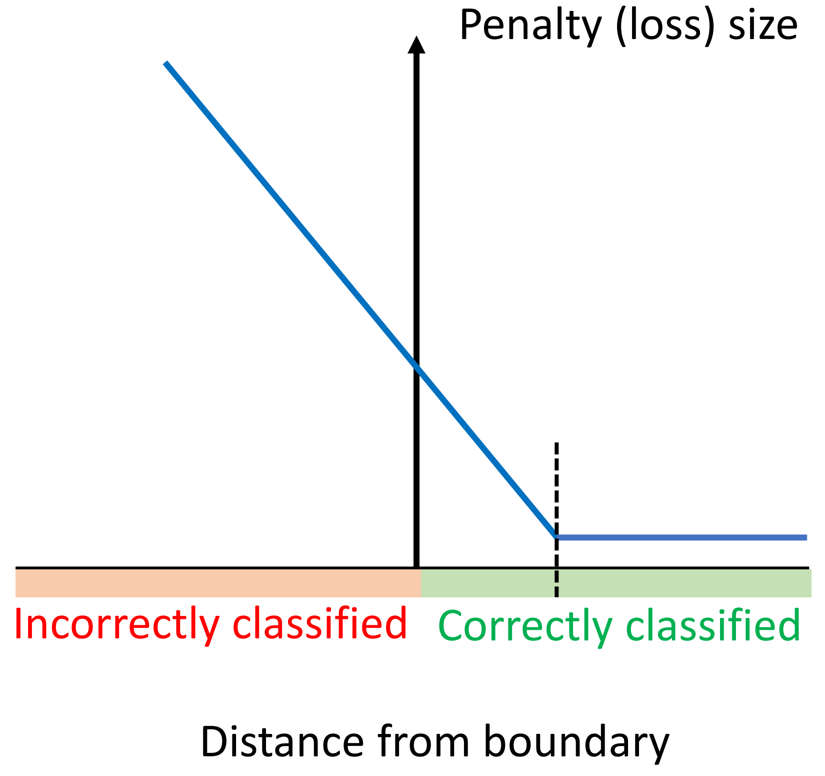


Hinge Loss

Loss Function: (Hinge loss)

$$L = max\left(0, 1 - y_i(w^T x_I + b)\right)$$

0 – for correct classification
1 – for incorrect classification

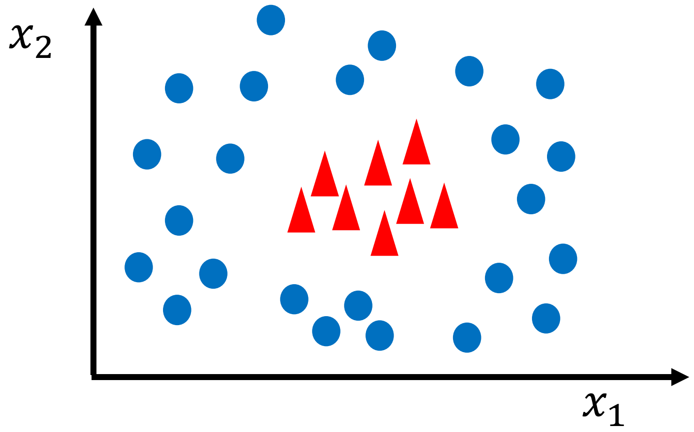


Kernel Trick

- SVM Kernel:
  - Kernel function generally transforms the training set of data so that a non-linear decision surface can be transformed to a linear equation in a high number of dimension spaces.
  - It returns the inner product between two points in a standard feature dimension.

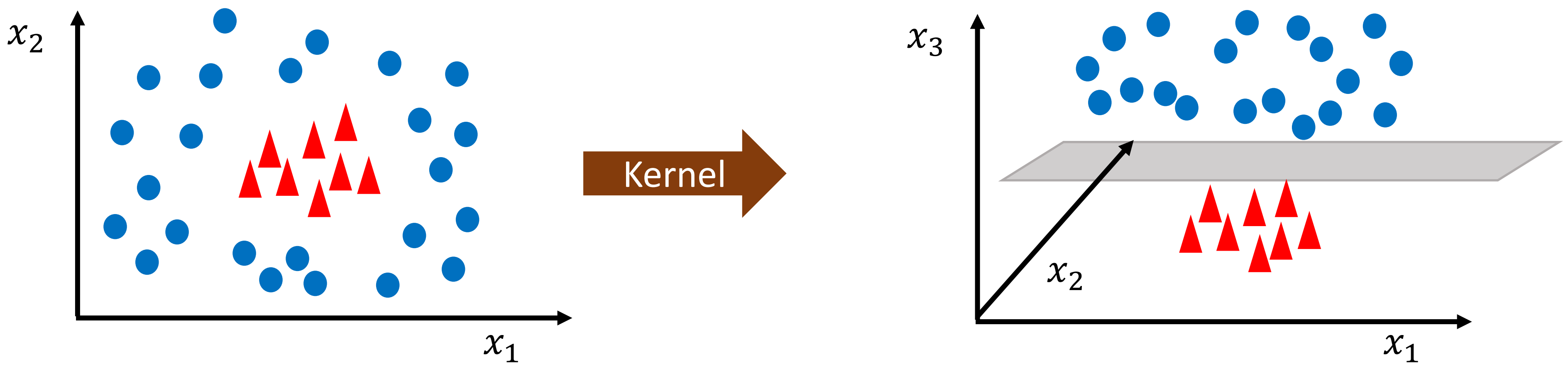
Kernel Trick

- SVM Kernel:
  - Kernel function generally transforms the training set of data so that a non-linear decision surface can be transformed to a linear equation in a high number of dimension spaces.
  - It returns the inner product between two points in a standard feature dimension.



Kernel Trick

- SVM Kernel:
  - Kernel function generally transforms the training set of data so that a non-linear decision surface can be transformed to a linear equation in a high number of dimension spaces.
  - It returns the inner product between two points in a standard feature dimension.



Types of SVM Kernels

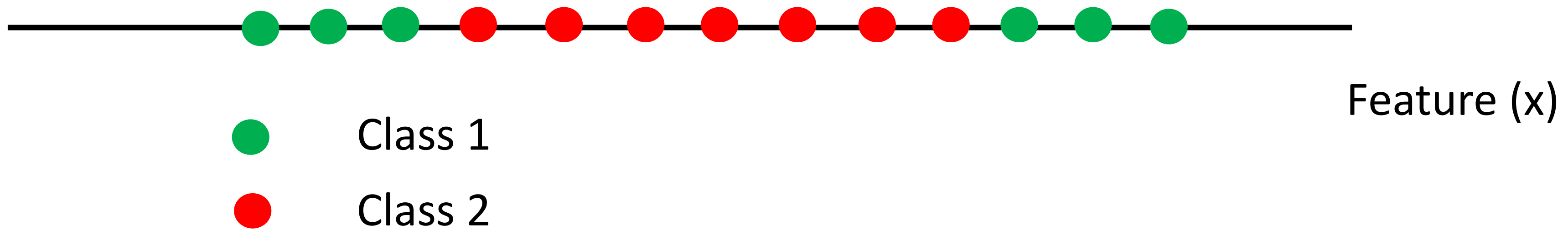
- Linear Kernel
- Polynomial Kernel
- Radial Basis Function
- Sigmoid

SVM Kernel Example

Feature (x)	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6

SVM Kernel Example

Feature (x)	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6

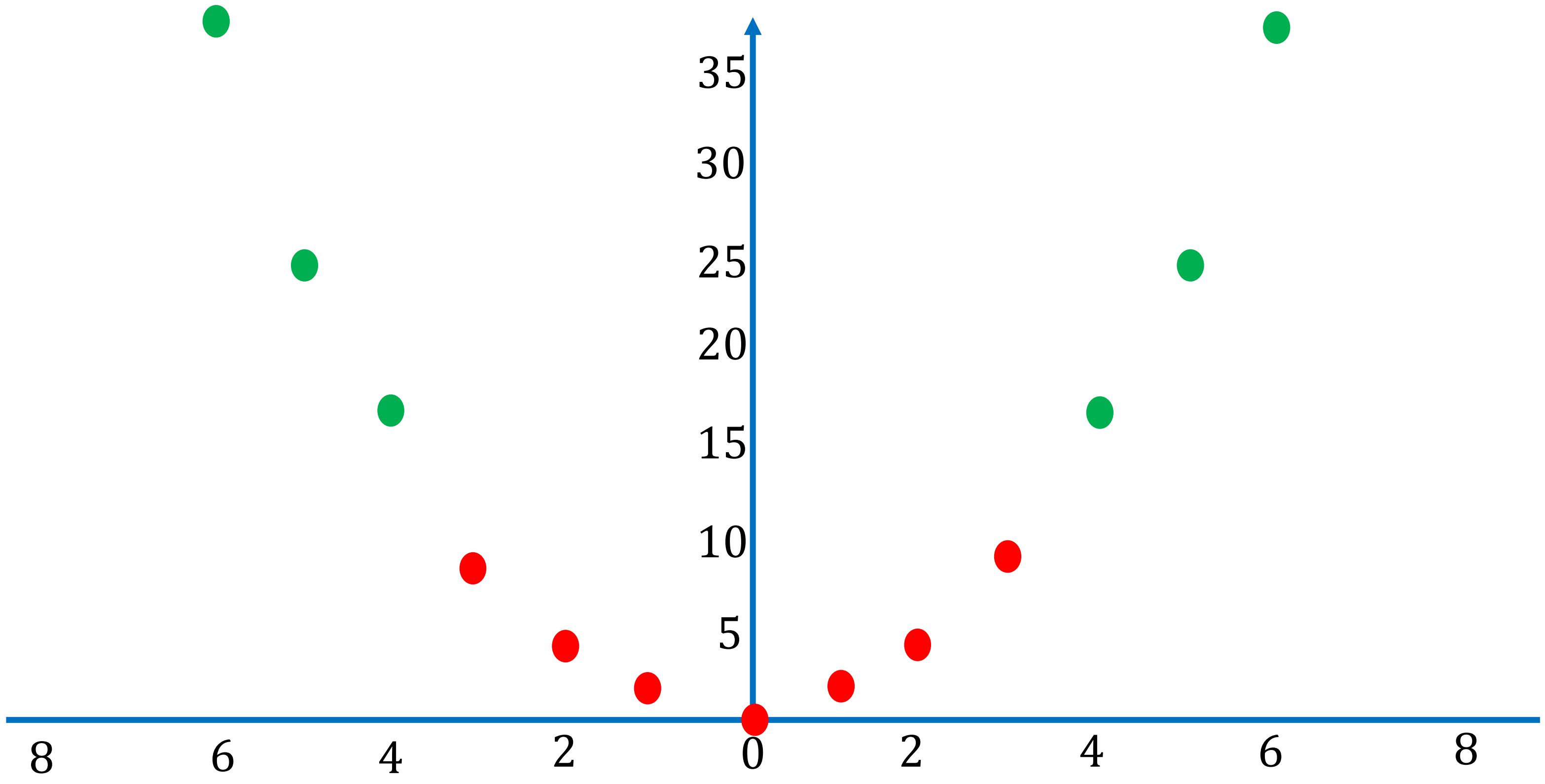


SVM Kernel Example

Feature (x)	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
x^2	36	25	16	9	4	1	0	1	4	9	16	25	36

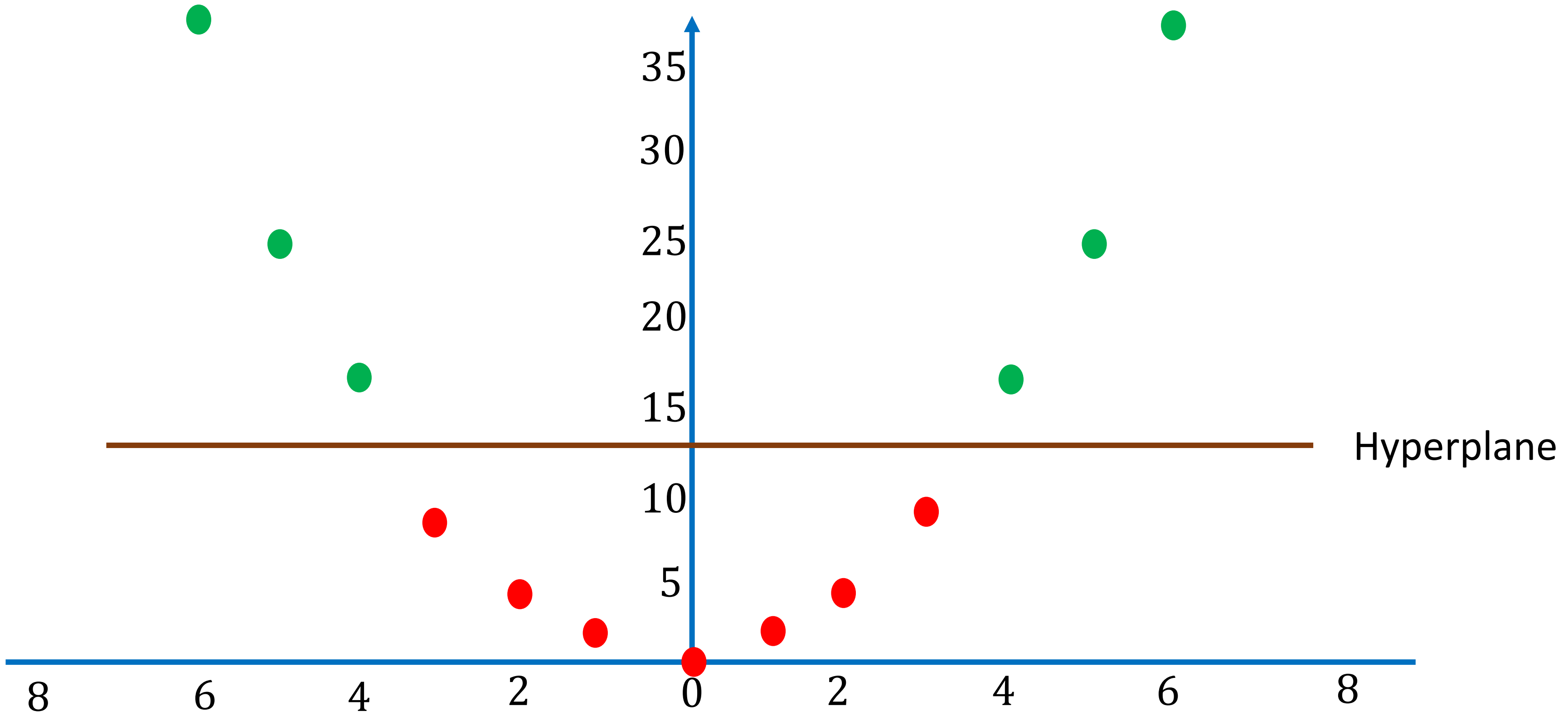
SVM Kernel Example

Feature (x)	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
x^2	36	25	16	9	4	1	0	1	4	9	16	25	36



SVM Kernel Example

Feature (x)	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
x^2	36	25	16	9	4	1	0	1	4	9	16	25	36



Support Vector Machines

Linear Kernel:

$$K(x_1, x_2) = {x_1}^T x_2$$

Polynomial Kernel:

$$K(x_1, x_2) = ({x_1}^T x_2 + r)^d$$

Radial Basis Function:

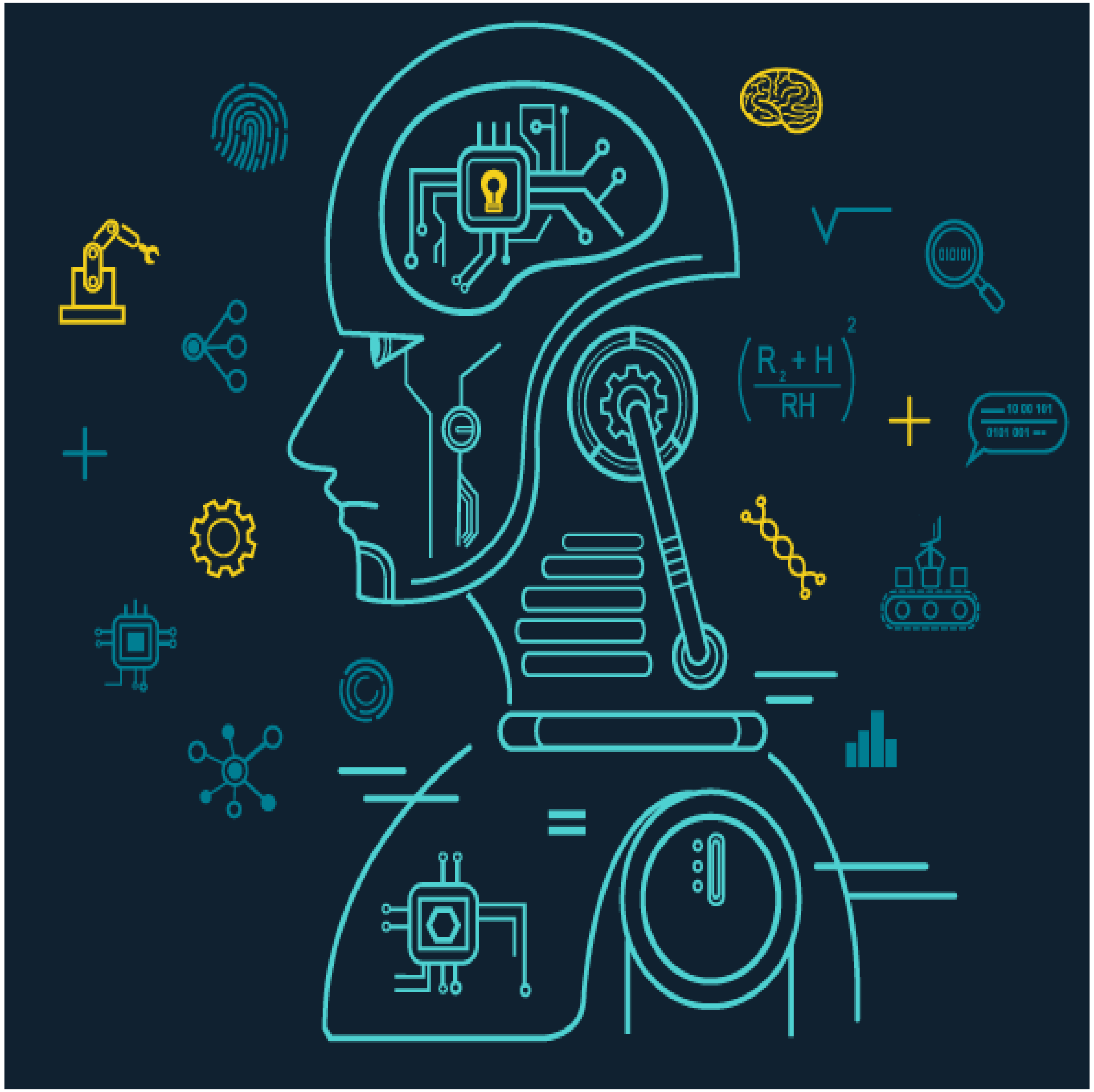
$$K(x_1, x_2) = e^{-\gamma \|x_1 - x_2\|^2}$$

Sigmoid Kernel:

$$K(x_1, x_2) = tanh(\gamma {x_1}^T x_2 + r)$$

References

1. https://link.springer.com/book/10.1007/978-0-387-77242-4
2. https://ineuron.ai/one-neuron/Tech Neuron?campaign=affiliate&coupon_code=SID5.
3. http://faculty.marshall.usc.edu/gareth-james/ISL/.



Machine Learning

K-Nearest Neighbor

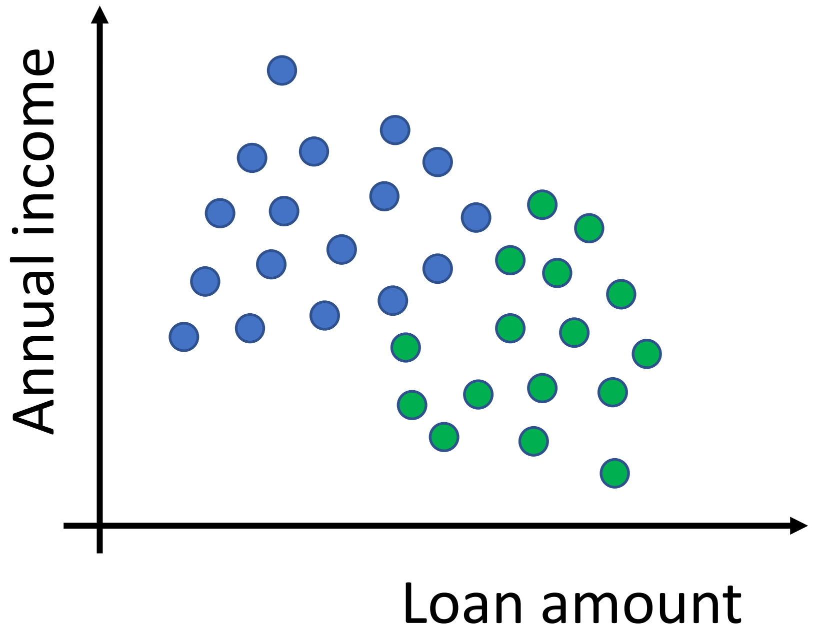
K-Nearest Neighbor

- By now, we all know machine learning models makes predictions by learning from the past data available.
- Supervised learning model.
- Used for both classification and regression problem.
- Can be used for non-linear data.
- Based on K-Neighbor.

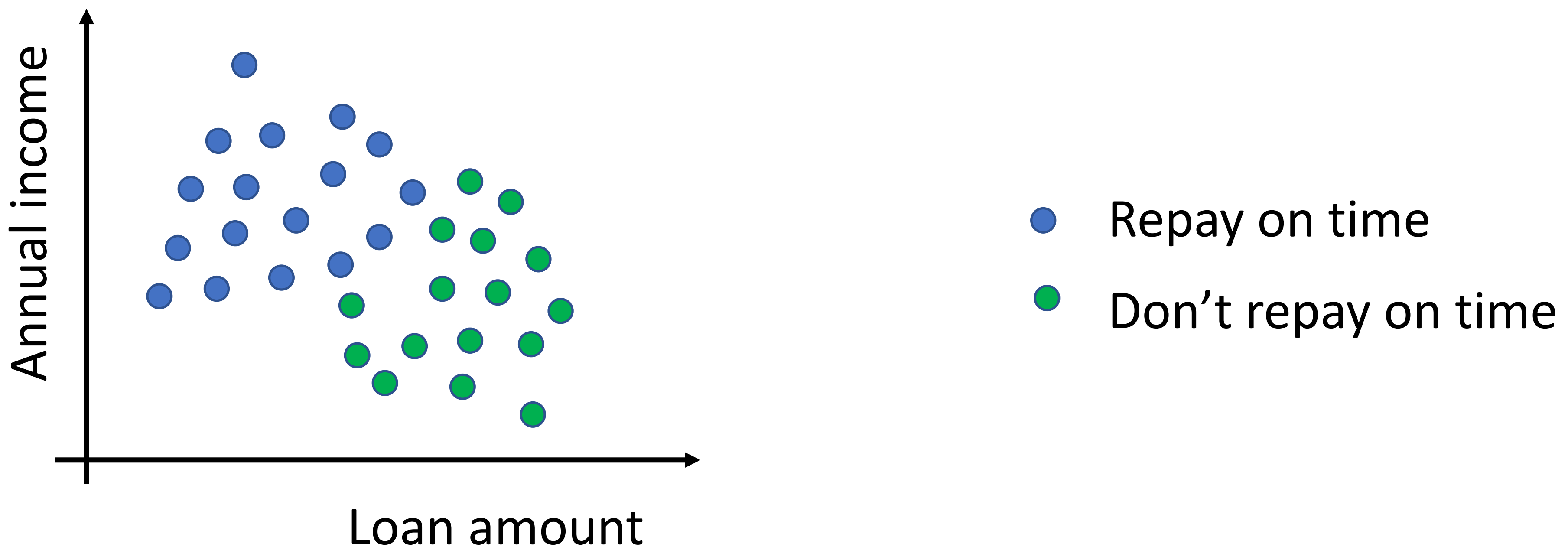
K-Nearest Neighbor

- KNN is a non-parametric classification method.
- The algorithm can be continuously updated with new training data.

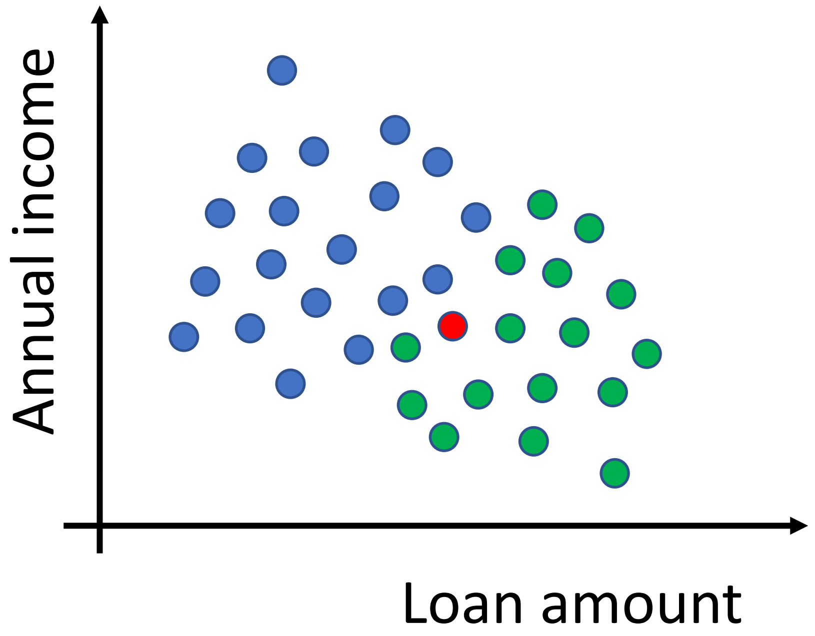
K-Nearest Neighbor with Example



K-Nearest Neighbor with Example

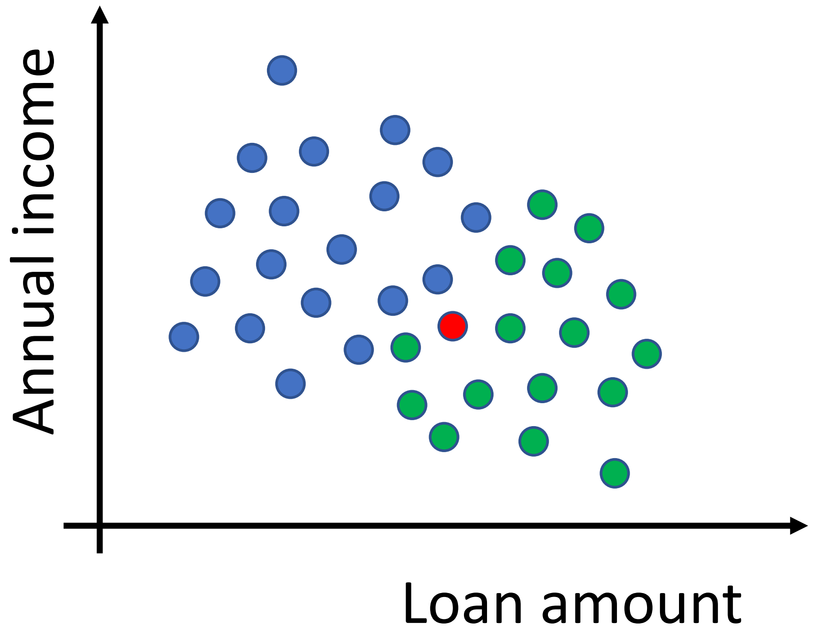


K-Nearest Neighbor with Example



- Repay on time
- Don’t repay on time

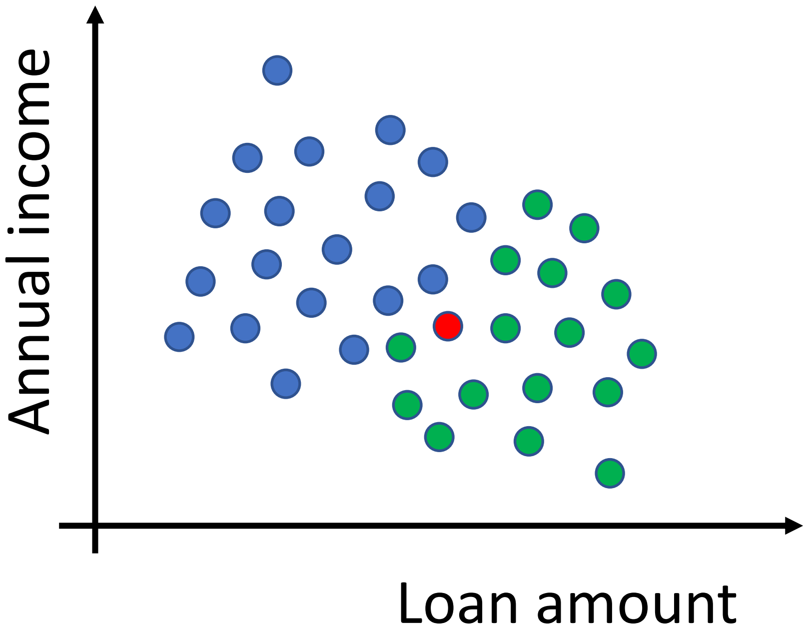
K-Nearest Neighbor with Example



K = 5

- Repay on time
- Don’t repay on time

K-Nearest Neighbor with Example

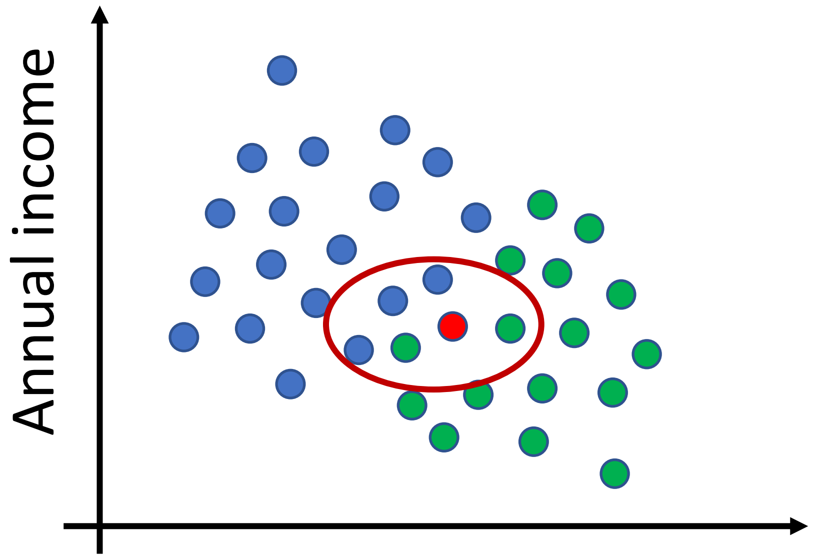


K = 5

To measure the distance between the data points:

- Euclidean distance.
- Manhattan distance.

K-Nearest Neighbor with Example



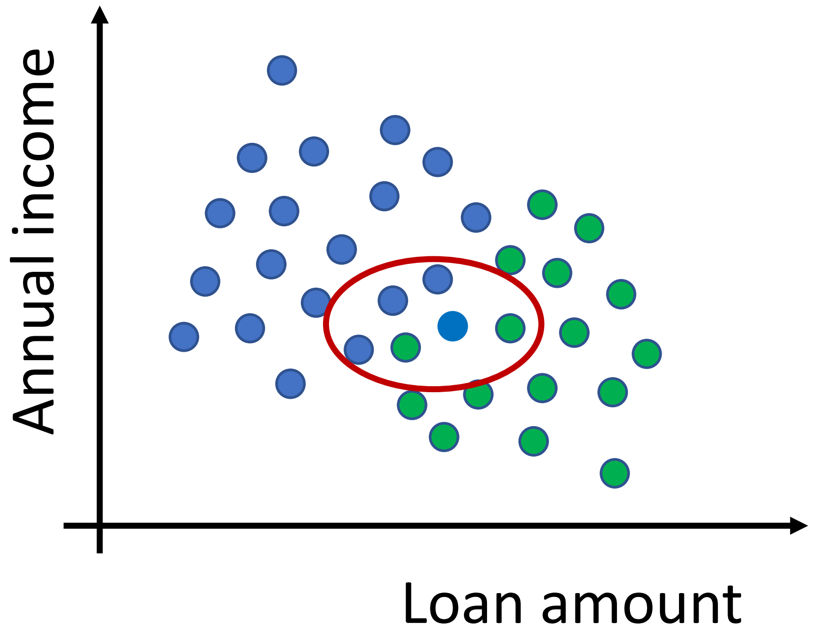
K = 5

- Repay on time (blue)
- Don't repay on time (green)

To measure the distance between the data points:

- Euclidean distance.
- Manhattan distance.

K-Nearest Neighbor with Example



K = 5

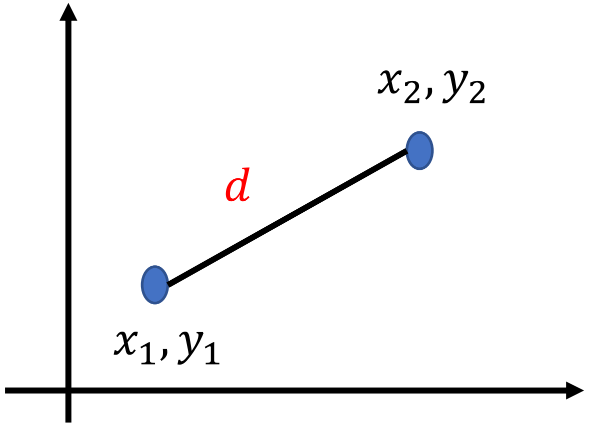
- Repay on time
- Don’t repay on time

To measure the distance between the data points:

- Euclidean distance.
- Manhattan distance.

K-Nearest Neighbor Distance Measures

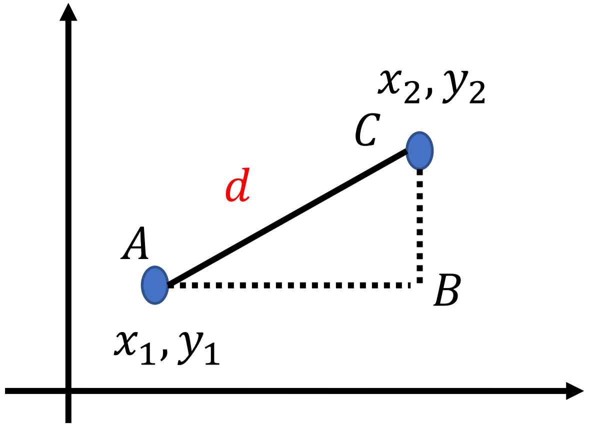
- **Euclidian distance:**



- **Manhattan distance:**

K-Nearest Neighbor Distance Measures

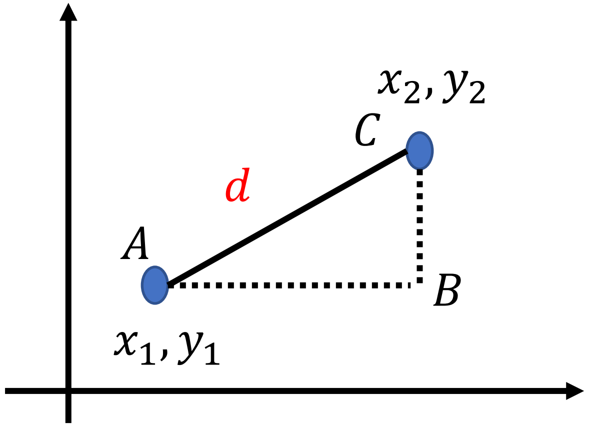
- **Euclidian distance:**



- **Manhattan distance:**

K-Nearest Neighbor Distance Measures

- **Euclidian distance:**



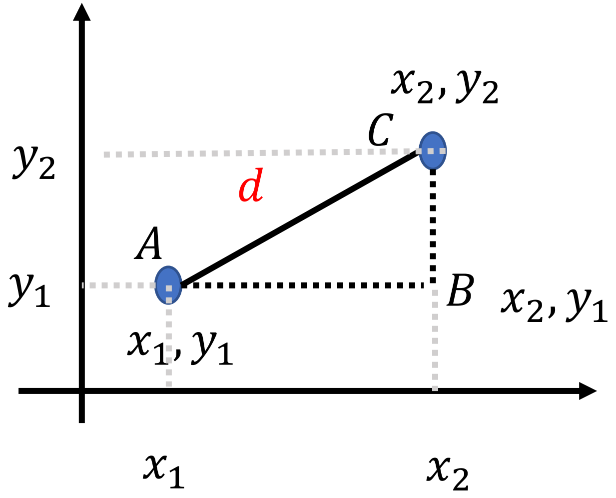
$$AC^2 = AB^2 + BC^2$$

$$AC = \sqrt{AB^2 + BC^2}$$

- **Manhattan distance:**

K-Nearest Neighbor Distance Measures

- **Euclidian distance:**



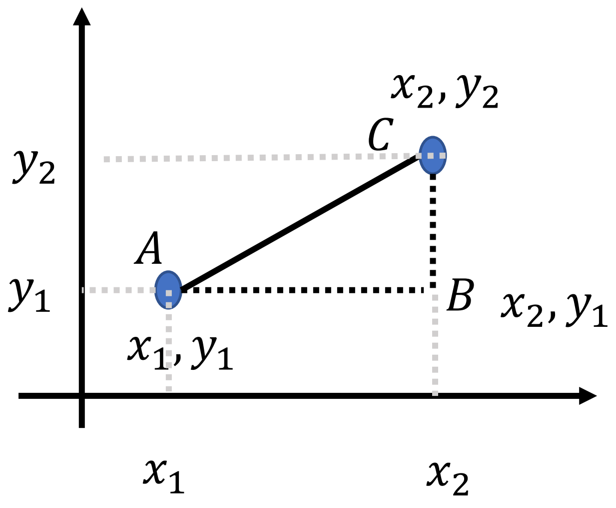
$$AC^2 = AB^2 + BC^2$$

$$AC = \sqrt{AB^2 + BC^2}$$

- **Manhattan distance:**

K-Nearest Neighbor Distance Measures

- **Euclidian distance:**



$$AC^2 = AB^2 + BC^2$$

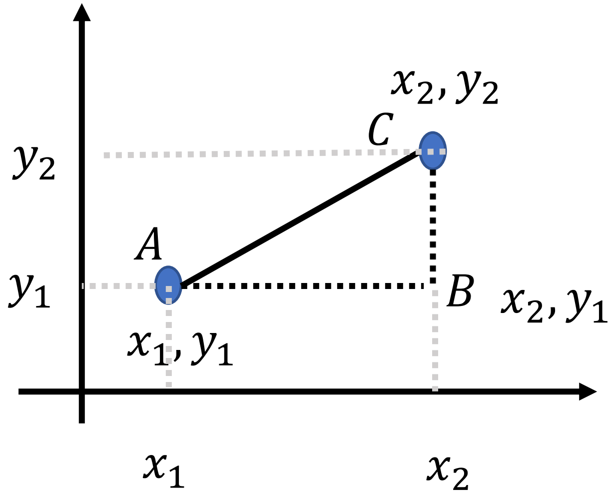
$$AC = \sqrt{AB^2 + BC^2}$$

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

- **Manhattan distance:**

K-Nearest Neighbor Distance Measures

- **Euclidian distance:**

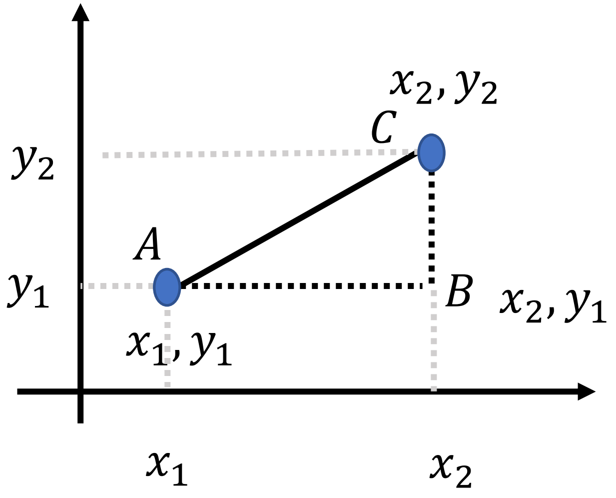


$$AC^2 = AB^2 + BC^2$$

$$AC = \sqrt{AB^2 + BC^2}$$

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

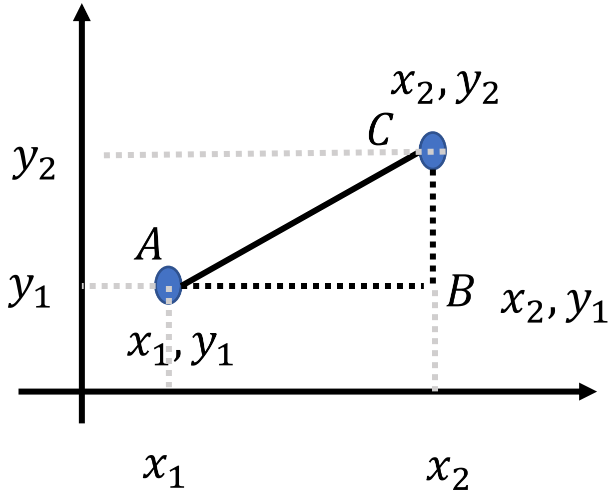
- **Manhattan distance:**



$$D = \sqrt{|x_2 - x_1| + |y_2 - y_1|}$$

K-Nearest Neighbor Distance Measures

- **Euclidian distance:**



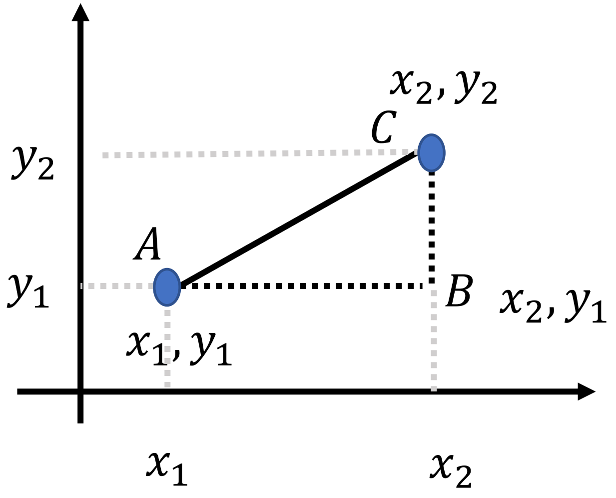
$$AC^2 = AB^2 + BC^2$$

$$AC = \sqrt{AB^2 + BC^2}$$

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

More influenced by outliers.

- **Manhattan distance:**



$$D = \sqrt{|x_2 - x_1| + |y_2 - y_1|}$$

Less influenced by outliers.

Numerical Example with KNN

Sepal Length	Sepal Width	Species
5.3	3.7	Setosa
5.1	3.8	Setosa
7.2	3.0	Virginica
5.4	3.4	Setosa
5.1	3.3	Setosa
5.4	3.9	Setosa
7.4	2.8	Virginica
6.1	2.8	Verscicolor
7.3	2.9	Virginica
6.0	2.7	Verscicolor
5.8	2.8	Virginica
6.3	2.3	Verscicolor
5.1	2.5	Verscicolor
6.3	2.5	Verscicolor
5.5	2.4	Verscicolor

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Numerical Example with KNN

Sepal Length	Sepal Width	Species
5.3	3.7	Setosa
5.1	3.8	Setosa
7.2	3.0	Virginica
5.4	3.4	Setosa
5.1	3.3	Setosa
5.4	3.9	Setosa
7.4	2.8	Virginica
6.1	2.8	Verscicolor
7.3	2.9	Virginica
6.0	2.7	Verscicolor
5.8	2.8	Virginica
6.3	2.3	Verscicolor
5.1	2.5	Verscicolor
6.3	2.5	Verscicolor
5.5	2.4	Verscicolor

Sepal Length	Sepal Width	Species
5.2	3.1	?

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Numerical Example with KNN

Sepal Length	Sepal Width	Species
5.3	3.7	Setosa
5.1	3.8	Setosa
7.2	3.0	Virginica
5.4	3.4	Setosa
5.1	3.3	Setosa
5.4	3.9	Setosa
7.4	2.8	Virginica
6.1	2.8	Verscicolor
7.3	2.9	Virginica
6.0	2.7	Verscicolor
5.8	2.8	Virginica
6.3	2.3	Verscicolor
5.1	2.5	Verscicolor
6.3	2.5	Verscicolor
5.5	2.4	Verscicolor

Sepal Length	Sepal Width	Species
5.2	3.1	?

1. Distance

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Numerical Example with KNN

Sepal Length	Sepal Width	Species
5.3	3.7	Setosa
5.1	3.8	Setosa
7.2	3.0	Virginica
5.4	3.4	Setosa
5.1	3.3	Setosa
5.4	3.9	Setosa
7.4	2.8	Virginica
6.1	2.8	Verscicolor
7.3	2.9	Virginica
6.0	2.7	Verscicolor
5.8	2.8	Virginica
6.3	2.3	Verscicolor
5.1	2.5	Verscicolor
6.3	2.5	Verscicolor
5.5	2.4	Verscicolor

Sepal Length	Sepal Width	Species
5.2	3.1	?

1. Distance

$$\text{Distance (Sepal Length, Sepal Width)} = \sqrt{(5.2 - 5.3)^2 + (3.1 - 3.7)^2}$$

$$\text{Distance (Sepal Length, Sepal Width)} = 0.608$$

Sepal Length	Sepal Width	Species	Distance
5.3	3.7	Setosa	0.608

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Numerical Example with KNN

Sepal Length	Sepal Width	Species	Distance
5.3	3.7	Setosa	0.608
5.1	3.8	Setosa	0.707
7.2	3.0	Virginica	2.002
5.4	3.4	Setosa	0.36
5.1	3.3	Setosa	0.22
5.4	3.9	Setosa	0.82
7.4	2.8	Virginica	2.22
6.1	2.8	Verscicolor	0.94
7.3	2.9	Virginica	2.1
6.0	2.7	Verscicolor	0.89
5.8	2.8	Virginica	0.67
6.3	2.3	Verscicolor	1.36
5.1	2.5	Verscicolor	0.60
6.3	2.5	Verscicolor	1.25
5.5	2.4	Verscicolor	0.75

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Numerical Example with KNN

Sepal Length	Sepal Width	Species	Distance
5.3	3.7	Setosa	0.608
5.1	3.8	Setosa	0.707
7.2	3.0	Virginica	2.002
5.4	3.4	Setosa	0.36
5.1	3.3	Setosa	0.22
5.4	3.9	Setosa	0.82
7.4	2.8	Virginica	2.22
6.1	2.8	Verscicolor	0.94
7.3	2.9	Virginica	2.1
6.0	2.7	Verscicolor	0.89
5.8	2.8	Virginica	0.67
6.3	2.3	Verscicolor	1.36
5.1	2.5	Verscicolor	0.60
6.3	2.5	Verscicolor	1.25
5.5	2.4	Verscicolor	0.75

2. Find Rank

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Numerical Example with KNN

Sepal Length	Sepal Width	Species	Distance	Rank
5.3	3.7	Setosa	0.608	3
5.1	3.8	Setosa	0.707	6
7.2	3.0	Virginica	2.002	13
5.4	3.4	Setosa	0.36	2
5.1	3.3	Setosa	0.22	1
5.4	3.9	Setosa	0.82	8
7.4	2.8	Virginica	2.22	15
6.1	2.8	Verscicolor	0.94	10
7.3	2.9	Virginica	2.1	14
6.0	2.7	Verscicolor	0.89	9
5.8	2.8	Virginica	0.67	5
6.3	2.3	Verscicolor	1.36	12
5.1	2.5	Verscicolor	0.60	4
6.3	2.5	Verscicolor	1.25	11
5.5	2.4	Verscicolor	0.75	7

2. Find Rank

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Numerical Example with KNN

Sepal Length	Sepal Width	Species	Distance	Rank
5.3	3.7	Setosa	0.608	3
5.1	3.8	Setosa	0.707	6
7.2	3.0	Virginica	2.002	13
5.4	3.4	Setosa	0.36	2
5.1	3.3	Setosa	0.22	1
5.4	3.9	Setosa	0.82	8
7.4	2.8	Virginica	2.22	15
6.1	2.8	Verscicolor	0.94	10
7.3	2.9	Virginica	2.1	14
6.0	2.7	Verscicolor	0.89	9
5.8	2.8	Virginica	0.67	5
6.3	2.3	Verscicolor	1.36	12
5.1	2.5	Verscicolor	0.60	4
6.3	2.5	Verscicolor	1.25	11
5.5	2.4	Verscicolor	0.75	7

3. Find the Nearest Neighbor

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Numerical Example with KNN

Sepal Length	Sepal Width	Species	Distance	Rank
5.3	3.7	Setosa	0.608	3
5.1	3.8	Setosa	0.707	6
7.2	3.0	Virginica	2.002	13
5.4	3.4	Setosa	0.36	2
5.1	3.3	Setosa	0.22	1
5.4	3.9	Setosa	0.82	8
7.4	2.8	Virginica	2.22	15
6.1	2.8	Verscicolor	0.94	10
7.3	2.9	Virginica	2.1	14
6.0	2.7	Verscicolor	0.89	9
5.8	2.8	Virginica	0.67	5
6.3	2.3	Verscicolor	1.36	12
5.1	2.5	Verscicolor	0.60	4
6.3	2.5	Verscicolor	1.25	11
5.5	2.4	Verscicolor	0.75	7

3. Find the Nearest Neighbor

If k = 1 – Setosa

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Numerical Example with KNN

Sepal Length	Sepal Width	Species	Distance	Rank
5.3	3.7	Setosa	0.608	3
5.1	3.8	Setosa	0.707	6
7.2	3.0	Virginica	2.002	13
5.4	3.4	Setosa	0.36	2
5.1	3.3	Setosa	0.22	1
5.4	3.9	Setosa	0.82	8
7.4	2.8	Virginica	2.22	15
6.1	2.8	Verscicolor	0.94	10
7.3	2.9	Virginica	2.1	14
6.0	2.7	Verscicolor	0.89	9
5.8	2.8	Virginica	0.67	5
6.3	2.3	Verscicolor	1.36	12
5.1	2.5	Verscicolor	0.60	4
6.3	2.5	Verscicolor	1.25	11
5.5	2.4	Verscicolor	0.75	7

3. Find the Nearest Neighbor

If k = 1 – Setosa

If k = 2 – Setosa

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Numerical Example with KNN

Sepal Length	Sepal Width	Species	Distance	Rank
5.3	3.7	Setosa	0.608	3
5.1	3.8	Setosa	0.707	6
7.2	3.0	Virginica	2.002	13
5.4	3.4	Setosa	0.36	2
5.1	3.3	Setosa	0.22	1
5.4	3.9	Setosa	0.82	8
7.4	2.8	Virginica	2.22	15
6.1	2.8	Verscicolor	0.94	10
7.3	2.9	Virginica	2.1	14
6.0	2.7	Verscicolor	0.89	9
5.8	2.8	Virginica	0.67	5
6.3	2.3	Verscicolor	1.36	12
5.1	2.5	Verscicolor	0.60	4
6.3	2.5	Verscicolor	1.25	11
5.5	2.4	Verscicolor	0.75	7

3. Find the Nearest Neighbor

If k = 1 – Setosa

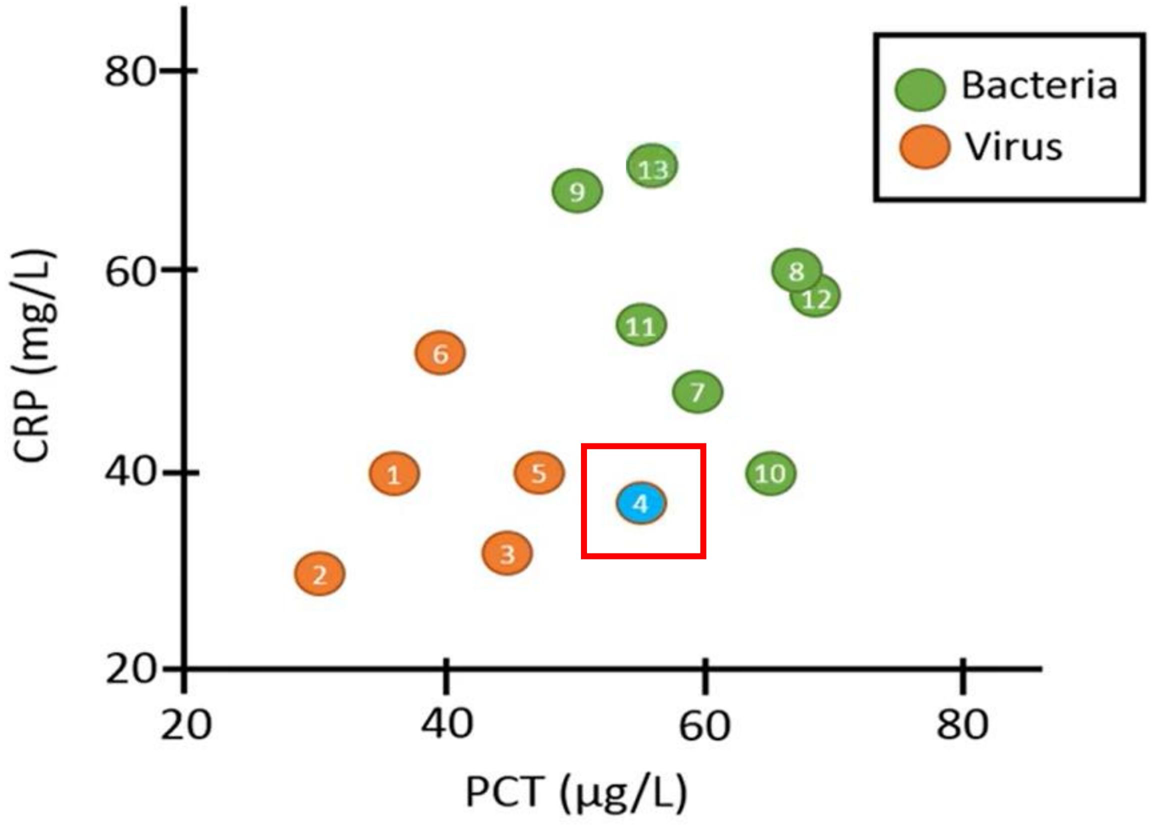
If k = 2 – Setosa

If k = 5 – Setosa

Ref. https://www.youtube.com/watch?v=Vk9lGGODaJA

Choosing optimal K-value

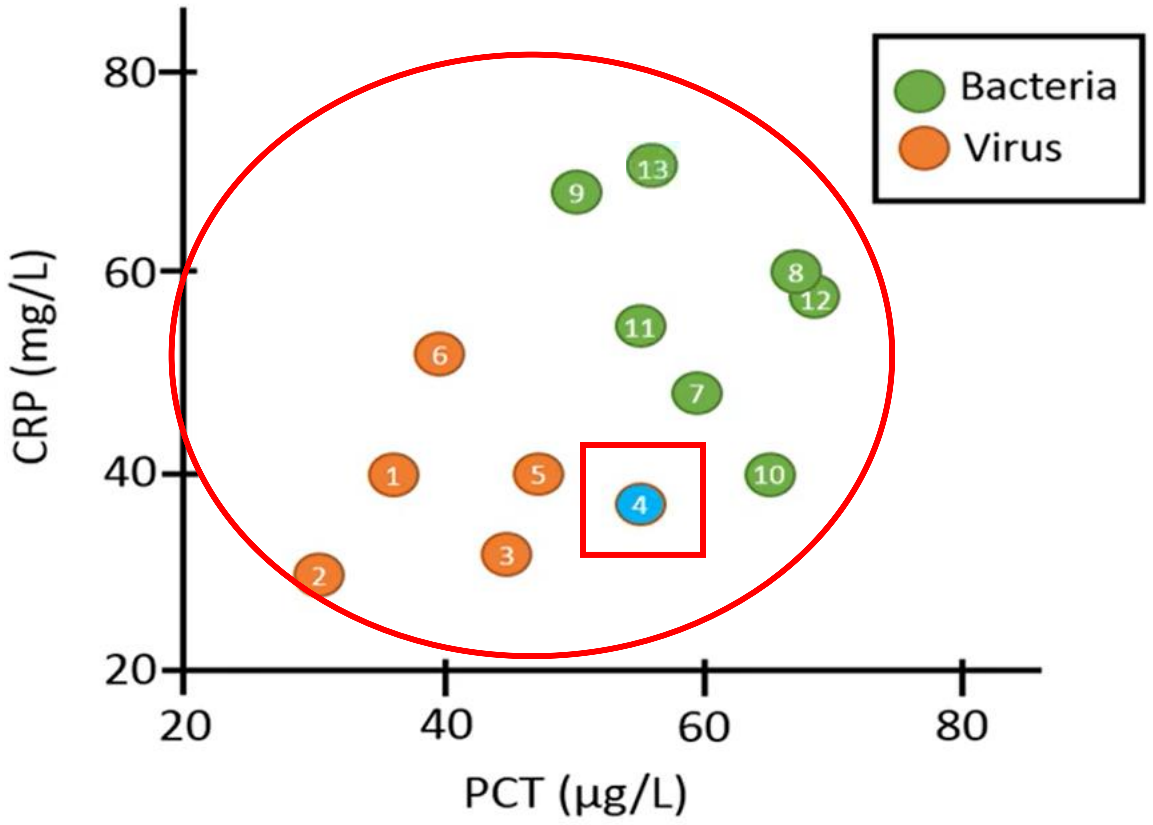
- The value of K should not be too high.



Ref. https://www.youtube.com/watch?v=48RqX4HTtCE

Choosing optimal K-value

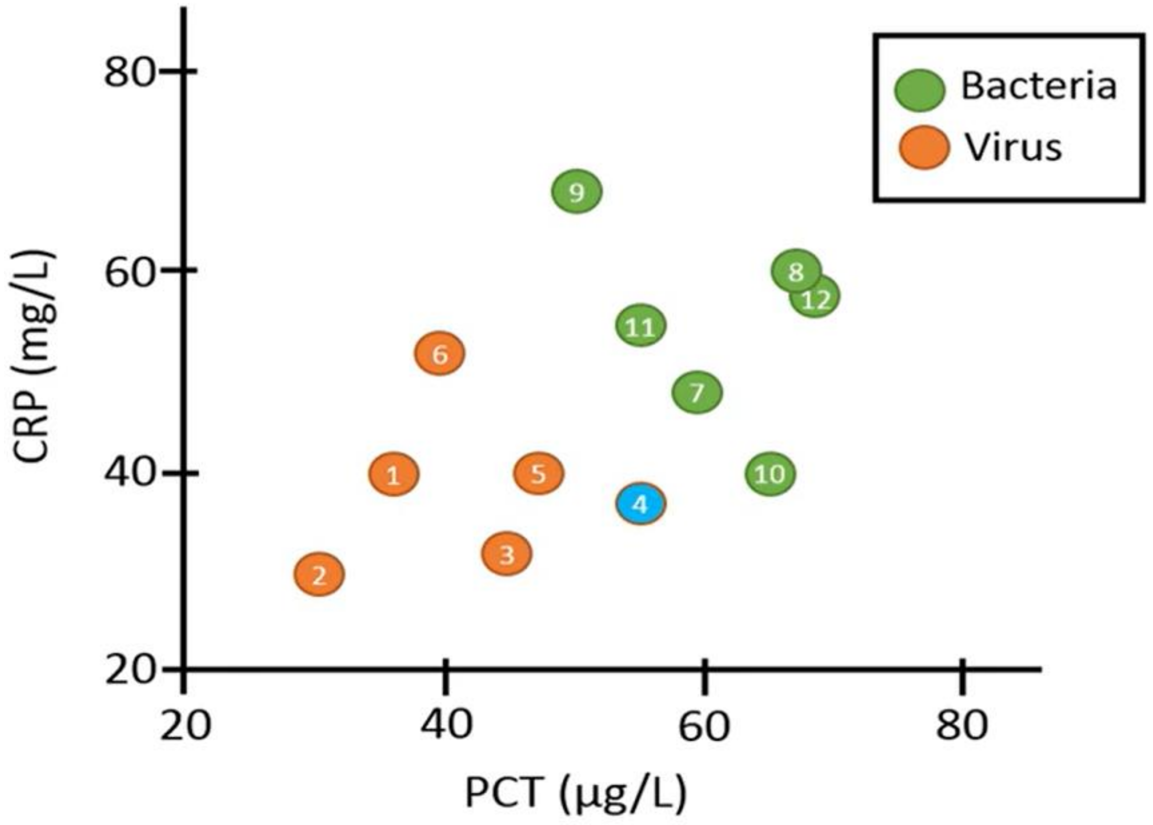
- The value of K should not be too high.



Ref. https://www.youtube.com/watch?v=48RqX4HTtCE

Choosing optimal K-value

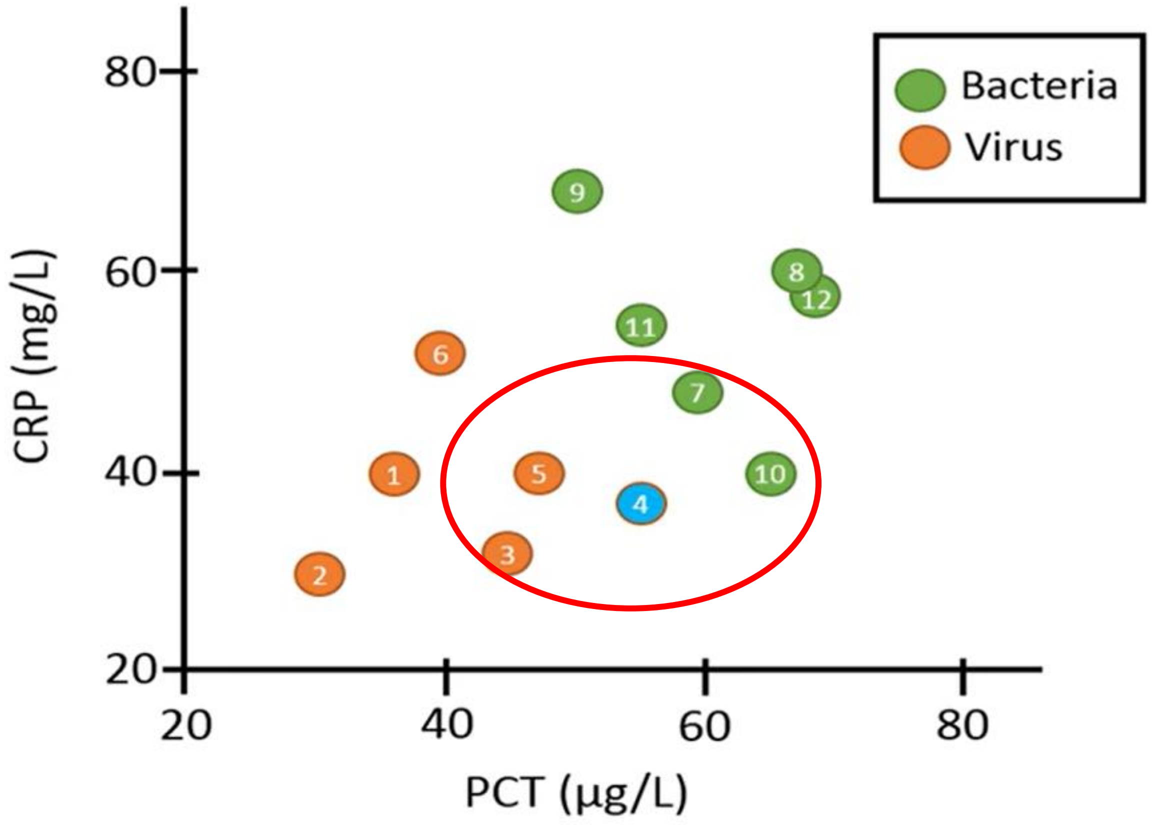
- The value of K should not be too high.
- The value of K should be an odd number.



Ref. https://www.youtube.com/watch?v=48RqX4HTtCE

Choosing optimal K-value

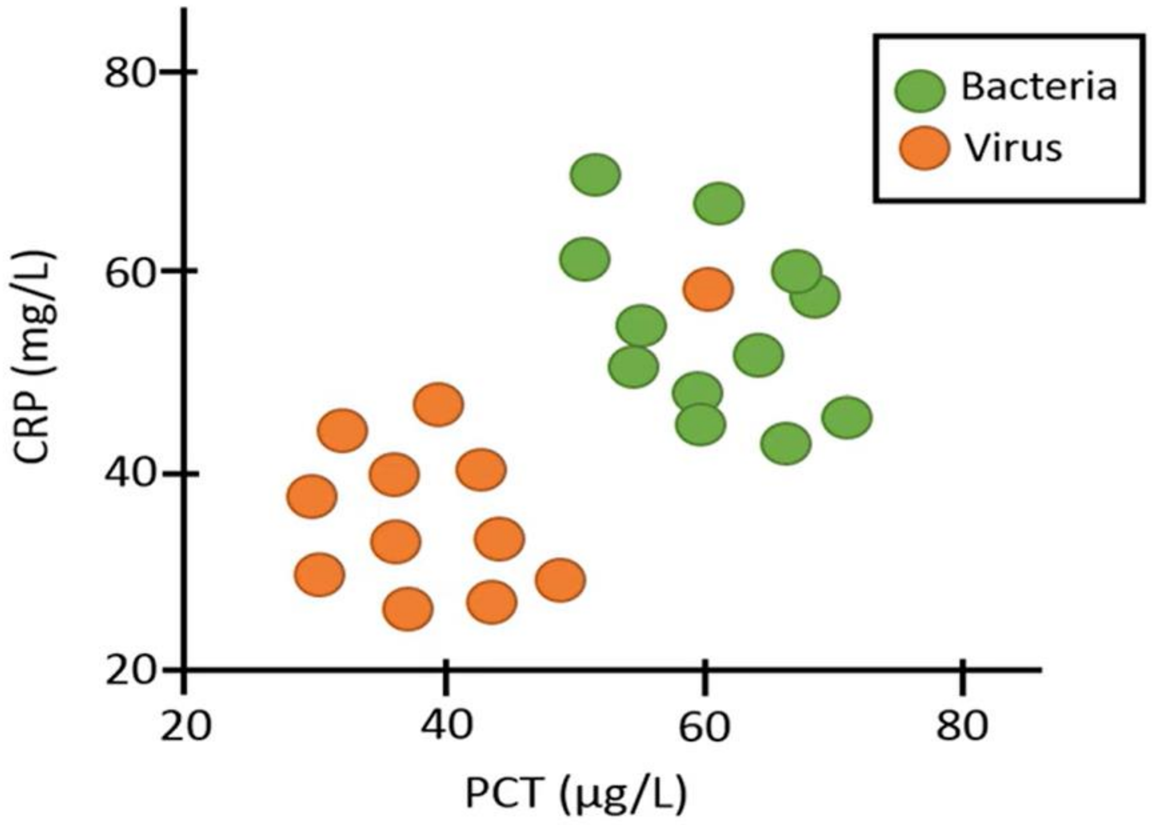
- The value of K should not be too high.
- The value of K should be an odd number.



Ref. https://www.youtube.com/watch?v=48RqX4HTtCE

Choosing optimal K-value

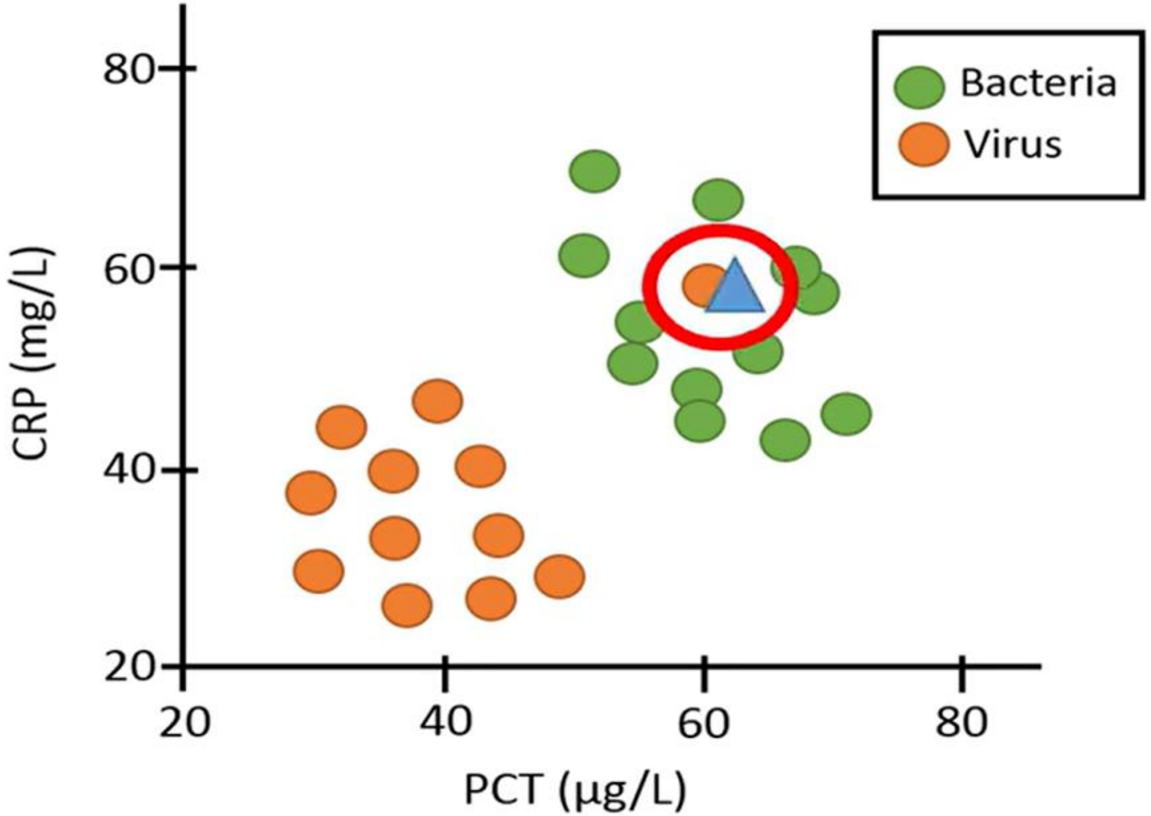
- The value of K should not be too high.
- The value of K should be an odd number.
- The value of K should not be equal to 1.



Ref. https://www.youtube.com/watch?v=48RqX4HTtCE

Choosing optimal K-value

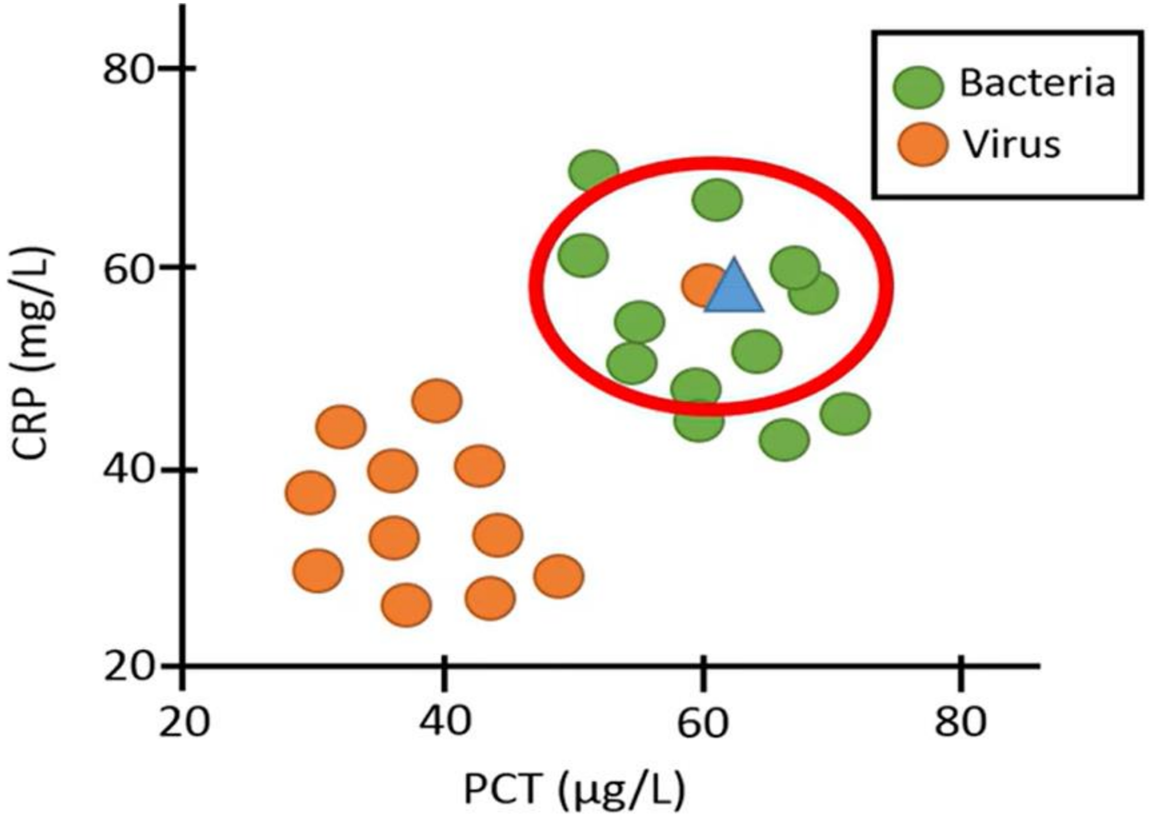
- The value of K should not be too high.
- The value of K should be an odd number.
- The value of K should not be equal to 1.



Ref. https://www.youtube.com/watch?v=48RqX4HTtCE

Choosing optimal K-value

- The value of K should not be too high.
- The value of K should be an odd number.
- The value of K should not be equal to 1.

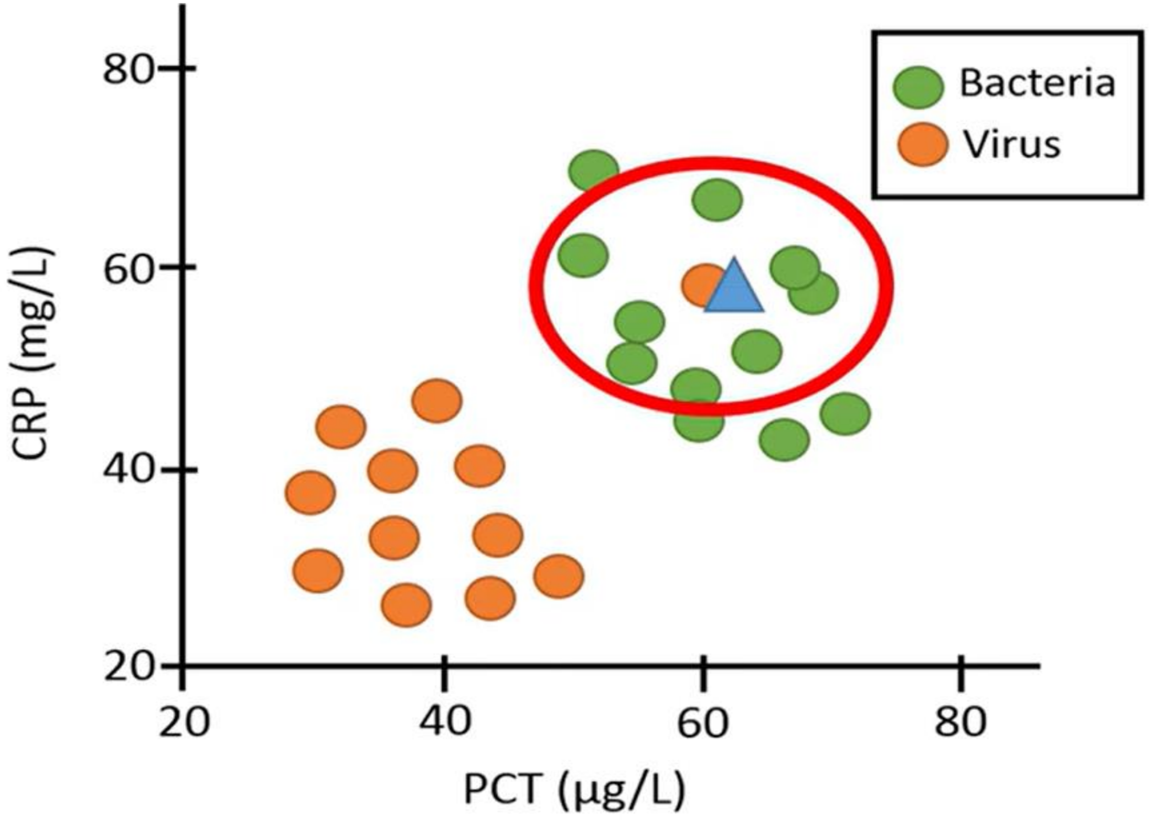


Ref. https://www.youtube.com/watch?v=48RqX4HTtCE

Choosing optimal K-value

- The value of K should not be too high.
- The value of K should be an odd number.
- The value of K should not be equal to 1.

The value of K should therefore not be too small or too large!



Ref. https://www.youtube.com/watch?v=48RqX4HTtCE

Choosing optimal K-value

Rule of thumb is to set K equal to: k = $\sqrt{n}$

- Apply cross validation technique for each k value to find optimal k value.

Advantages and Disadvantages of KNN

Advantages:

- Works well with smaller datasets with less number of features.

Advantages and Disadvantages of KNN

Advantages:

- Works well with smaller datasets with less number of features.
- Can be used for both classification and regression.

Advantages and Disadvantages of KNN

Advantages:

- Works well with smaller datasets with less number of features.
- Can be used for both classification and regression.
- Easy to implement for Multi-class classification problem.

Advantages and Disadvantages of KNN

Advantages:

- Works well with smaller datasets with less number of features.
- Can be used for both classification and regression.
- Easy to implement for Multi-class classification problem.
- Different distance criteria can be used.

Advantages and Disadvantages of KNN

Advantages:

- Works well with smaller datasets with less number of features.
- Can be used for both classification and regression.
- Easy to implement for Multi-class classification problem.
- Different distance criteria can be used.

Disadvantages:

- Choosing optimal “K” value.

Advantages and Disadvantages of KNN

Advantages:

- Works well with smaller datasets with less number of features.
- Can be used for both classification and regression.
- Easy to implement for Multi-class classification problem.
- Different distance criteria can be used.

Disadvantages:

- Choosing optimal “K” value.
- Less efficient with high dimensional data.

Advantages and Disadvantages of KNN

Advantages:

- Works well with smaller datasets with less number of features.
- Can be used for both classification and regression.
- Easy to implement for Multi-class classification problem.
- Different distance criteria can be used.

Disadvantages:

- Choosing optimal “K” value.
- Less efficient with high dimensional data.
- Doesn’t perform well on imbalanced dataset.

Advantages and Disadvantages of KNN

Advantages:

- Works well with smaller datasets with less number of features.
- Can be used for both classification and regression.
- Easy to implement for Multi-class classification problem.
- Different distance criteria can be used.

Disadvantages:

- Choosing optimal “K” value.
- Less efficient with high dimensional data.
- Doesn’t perform well on imbalanced dataset.
- Sensitive to Outliers.

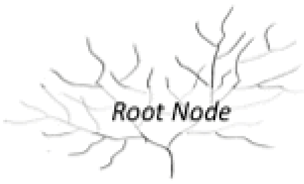
References

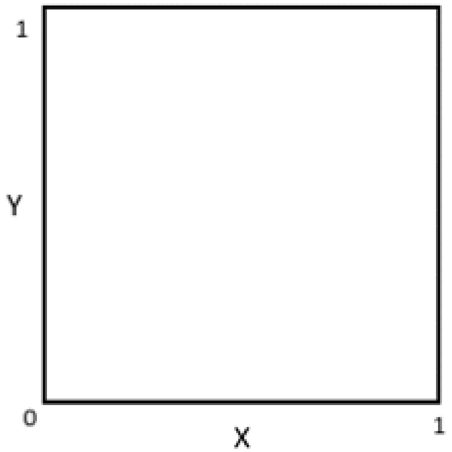
1. https://www.mygreatlearning.com/blog/knn-algorithm-introduction/.
2. https://towardsdatascience.com/how-to-find-the-optimal-value-of-k-in-knn35d936e554eb#:~:text=The%20optimal%20K%20value%20usually,be%20aware%20of%20the%20outliers.
3. https://medium.com/@luigi.fiori.lf0303/distance-metrics-and-k-nearest-neighbor-knn-1b840969c0f4.

Machine learning

Decision Tree

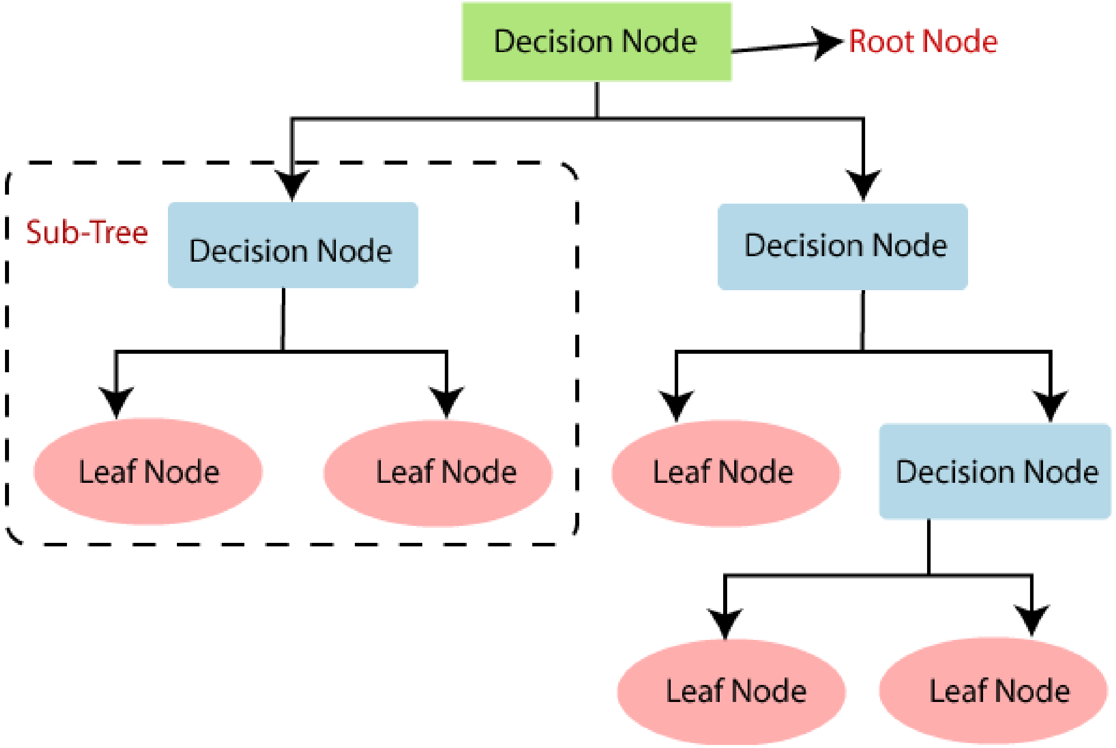
Decision Tree





For more tutorials: annalyzin.wordpress.com

Terminologies in Decision Tree



Root Node: represents entire population and further divided in two/ more homogeneous sets.

Splitting: Process of dividing node in two or more sub-nodes.

Decision Node: When a sub-node splits into further sub-nodes

Leaf / Terminal Node: Nodes do not split.

Pruning: When we remove sub-nodes of a decision node.

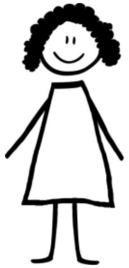
Branch/Sub-Tree: A subsection of the entire tree.

Parent and Child Node: A node, which is divided into sub-nodes is called a parent node of sub-nodes whereas sub-nodes are the child of a parent node.

Definition

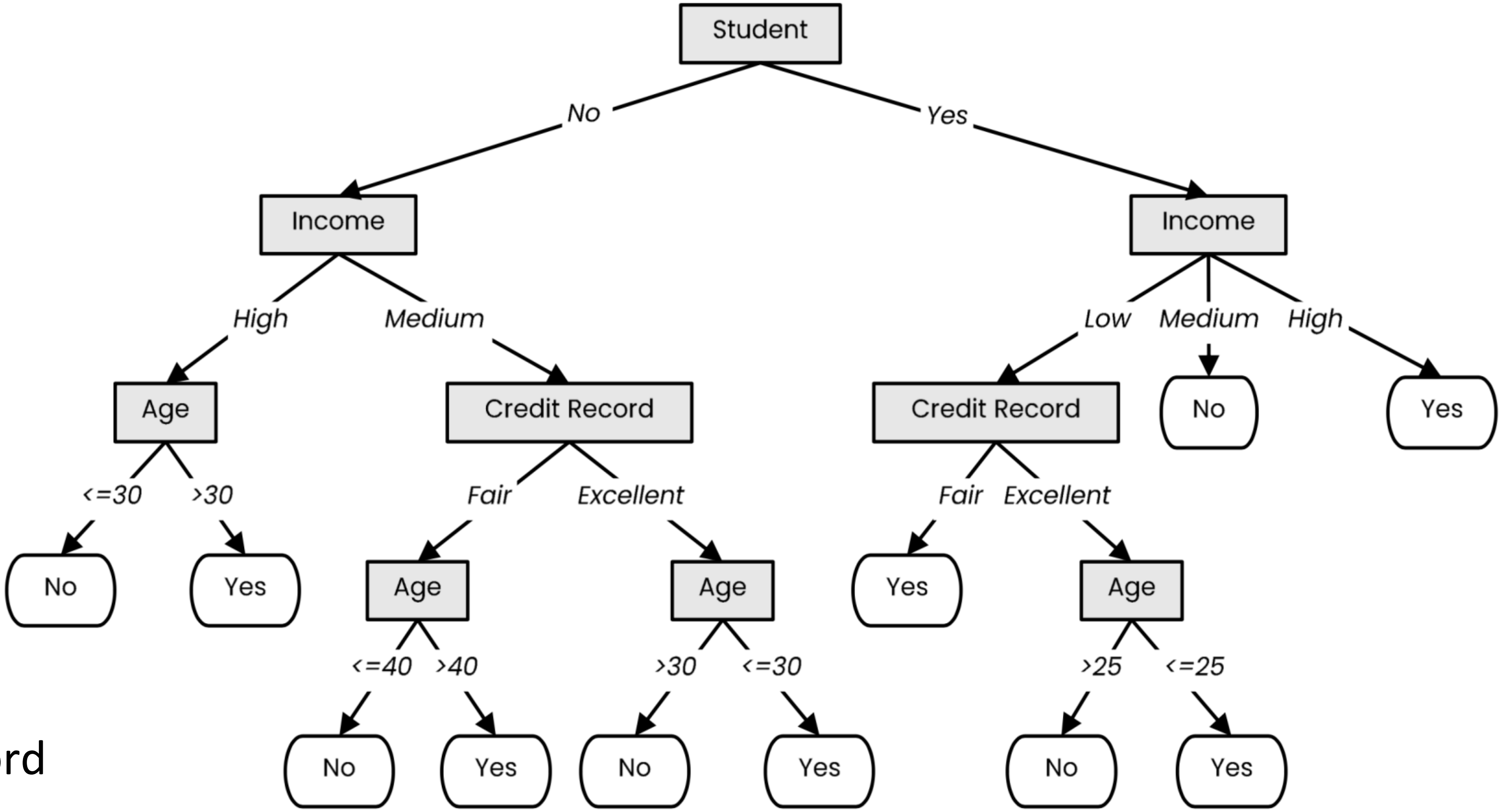
- A tree-like model that illustrates series of events leading to certain decisions
- Each node represents a test on an attribute and each branch is an outcome of that test

Who to loan?



- Not a student
- 45 years old
- Medium income
- Fair credit record

- Student
- 27 years old
- Low income
- Excellent credit record



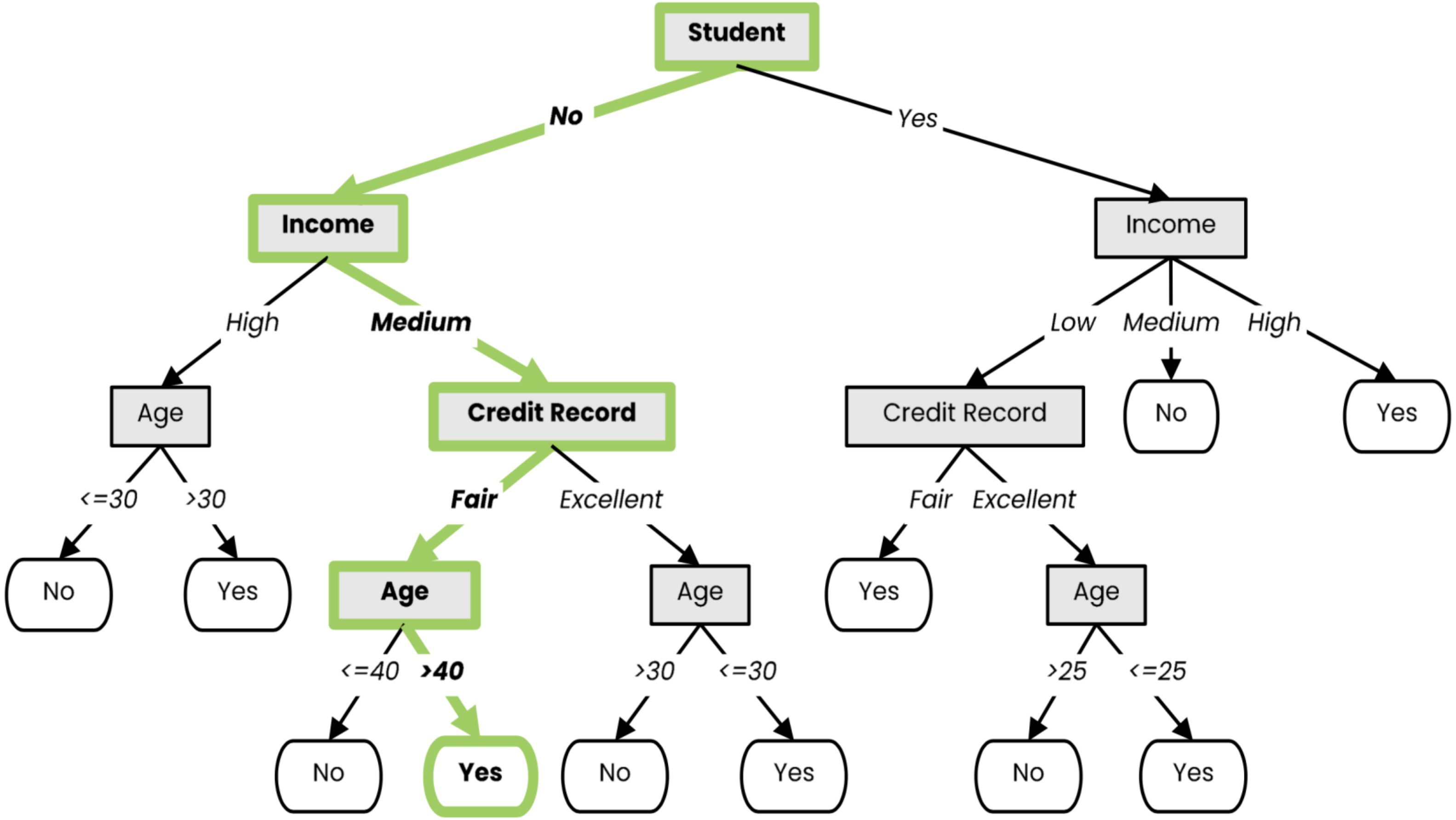
Definition

- A tree-like model that illustrates series of events leading to certain decisions
- Each node represents a test on an attribute and each branch is an outcome of that test

Who to loan?

- Not a student
- 45 years old
- Medium income
- Fair credit record
- Yes

- Student
- 27 years old
- Low income
- Excellent credit record



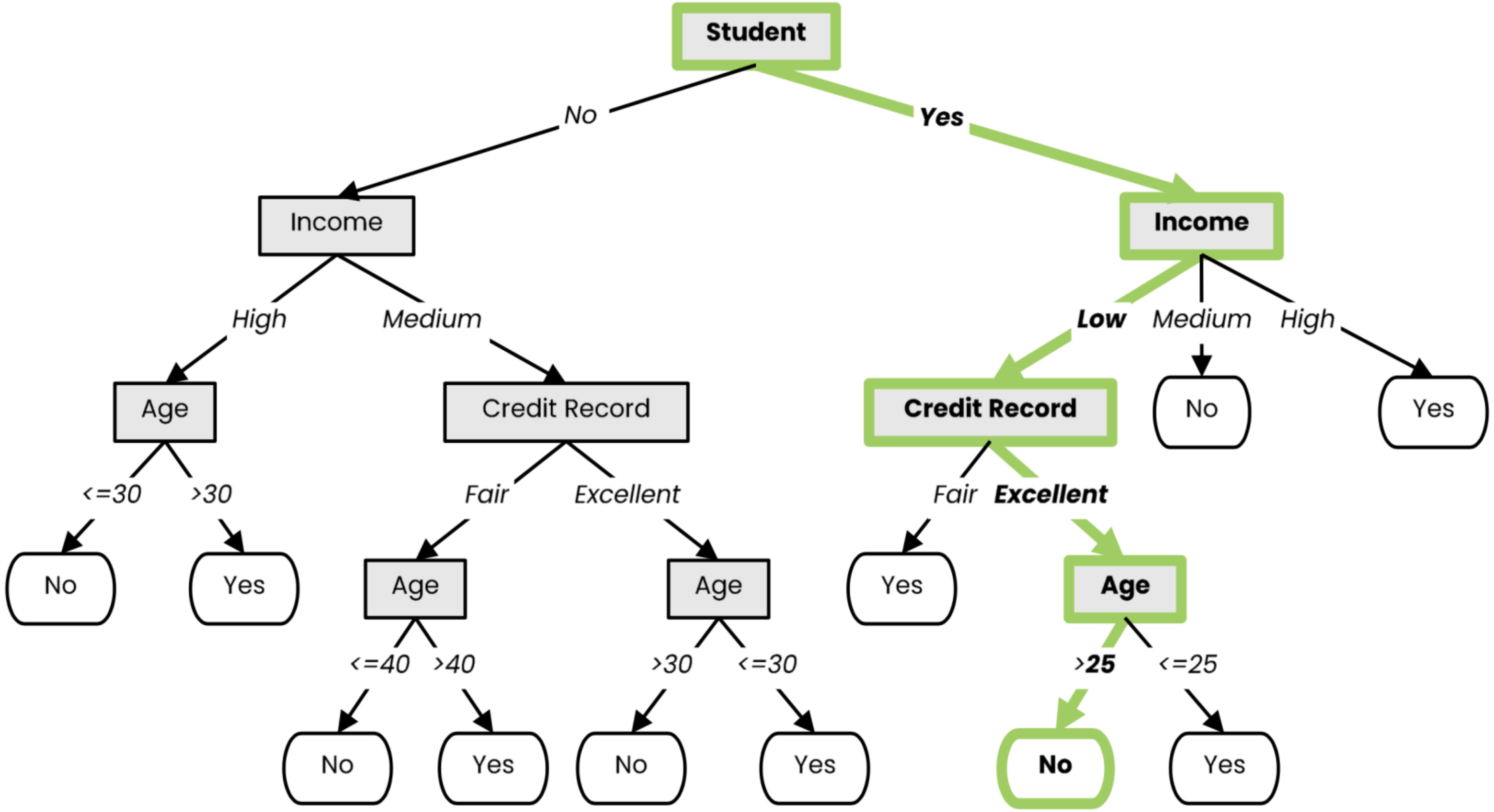
Definition

- A tree-like model that illustrates series of events leading to certain decisions
- Each node represents a test on an attribute and each branch is an outcome of that test

Who to loan?

- Not a student
- 45 years old
- Medium income
- Fair credit record
- Yes

- Student
- 27 years old
- Low income
- Excellent credit record



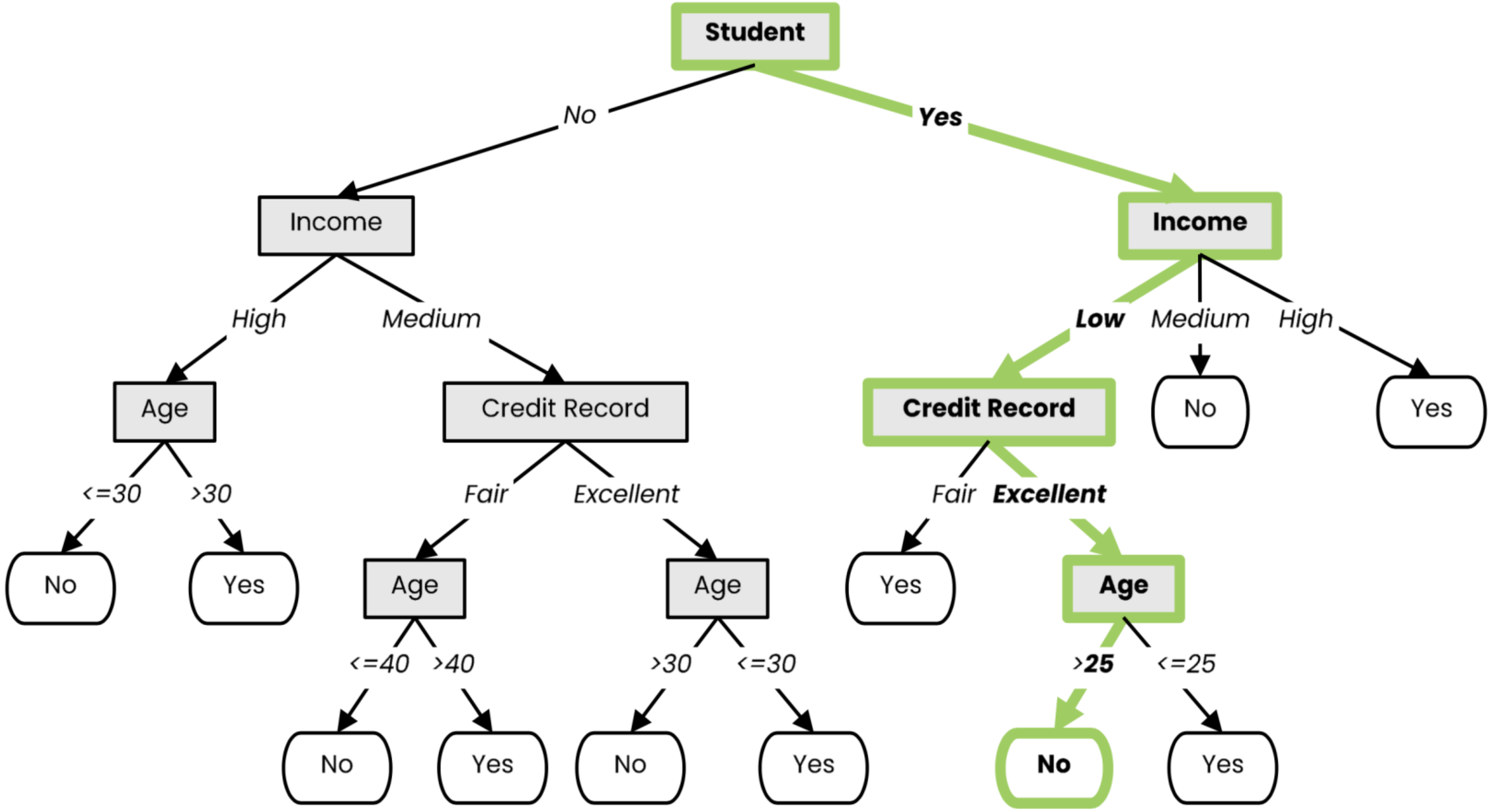
Definition

- A tree-like model that illustrates series of events leading to certain decisions
- Each node represents a test on an attribute and each branch is an outcome of that test

Who to loan?

- Not a student
- 45 years old
- Medium income
- Fair credit record
- Yes

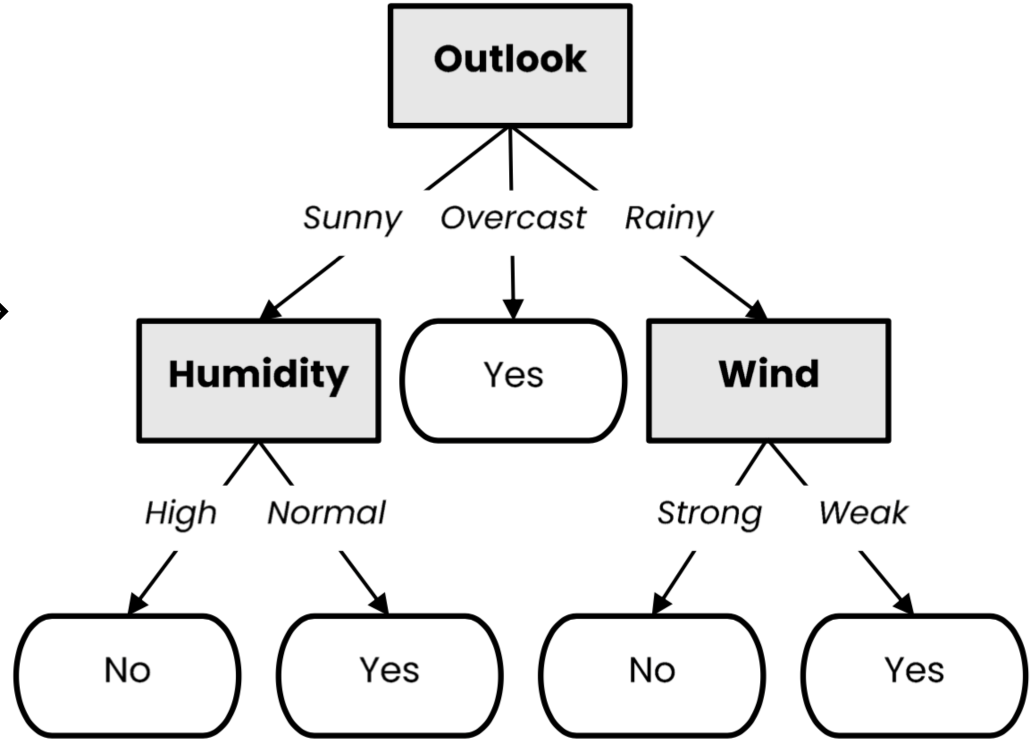
- Student
- 27 years old
- Low income
- Excellent credit record
- No



Decision Tree Learning

- We use labeled data to obtain a suitable decision tree for future predictions
- We want a decision tree that works well on unseen data, while asking as few questions as possible

Outlook	Temperature	Humidity	Wind	Play Tennis?
Sunny	Hot	High	Weak	**No**
Sunny	Hot	High	Strong	**No**
Overcast	Hot	High	Weak	**Yes**
Rainy	Mild	High	Weak	**Yes**
Rainy	Cool	Normal	Weak	**Yes**
Rainy	Cool	Normal	Strong	**No**
Overcast	Cool	Normal	Strong	**Yes**
Sunny	Mild	High	Weak	**No**
Sunny	Cool	Normal	Weak	**Yes**
Rainy	Mild	Normal	Weak	**Yes**
Sunny	Mild	Normal	Strong	**Yes**
Overcast	Mild	High	Strong	**Yes**
Overcast	Hot	Normal	Weak	**Yes**
Rainy	Mild	High	Strong	**No**



Decision Tree Learning

- Basic step: choose an attribute and, based on its values, split the data into smaller sets
  - Recursively repeat this step until we can surely decide the label

Outlook	Temperature	Humidity	Wind	Play Tennis?
Sunny	Hot	High	Weak	**No**
Sunny	Hot	High	Strong	**No**
Overcast	Hot	High	Weak	**Yes**
Rainy	Mild	High	Weak	**Yes**
Rainy	Cool	Normal	Weak	**Yes**
Rainy	Cool	Normal	Strong	**No**
Overcast	Cool	Normal	Strong	**Yes**
Sunny	Mild	High	Weak	**No**
Sunny	Cool	Normal	Weak	**Yes**
Rainy	Mild	Normal	Weak	**Yes**
Sunny	Mild	Normal	Strong	**Yes**
Overcast	Mild	High	Strong	**Yes**
Overcast	Hot	Normal	Weak	**Yes**
Rainy	Mild	High	Strong	**No**

Outlook

Decision Tree Learning

- Basic step: choose an attribute and, based on its values, split the data into smaller sets
  - Recursively repeat this step until we can surely decide the label

Outlook = *Sunny*

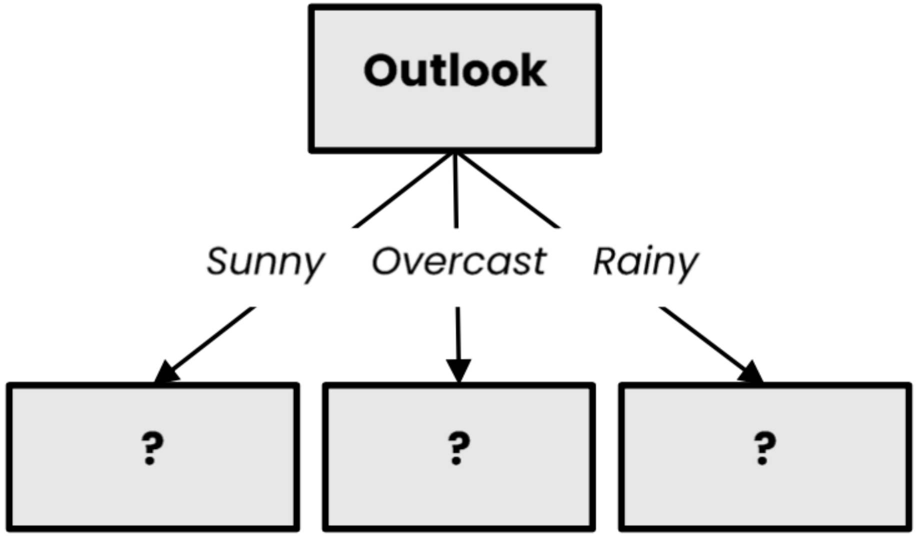
Temperature	Humidity	Wind	Play Tennis?
Hot	High	Weak	**No**
Hot	High	Strong	**No**
Mild	High	Weak	**No**
Cool	Normal	Weak	**Yes**
Mild	Normal	Strong	**Yes**

Outlook = *Overcast*

Temperature	Humidity	Wind	Play Tennis?
Hot	High	Weak	**Yes**
Cool	Normal	Strong	**Yes**
Mild	High	Strong	**Yes**
Hot	Normal	Weak	**Yes**

Outlook = *Rainy*

Temperature	Humidity	Wind	Play Tennis?
Mild	High	Weak	**Yes**
Cool	Normal	Weak	**Yes**
Cool	Normal	Strong	**No**
Mild	Normal	Weak	**Yes**
Mild	High	Strong	**No**



Decision Tree Learning

- Basic step: choose an attribute and, based on its values, split the data into smaller sets
  - Recursively repeat this step until we can surely decide the label

Outlook = *Sunny*

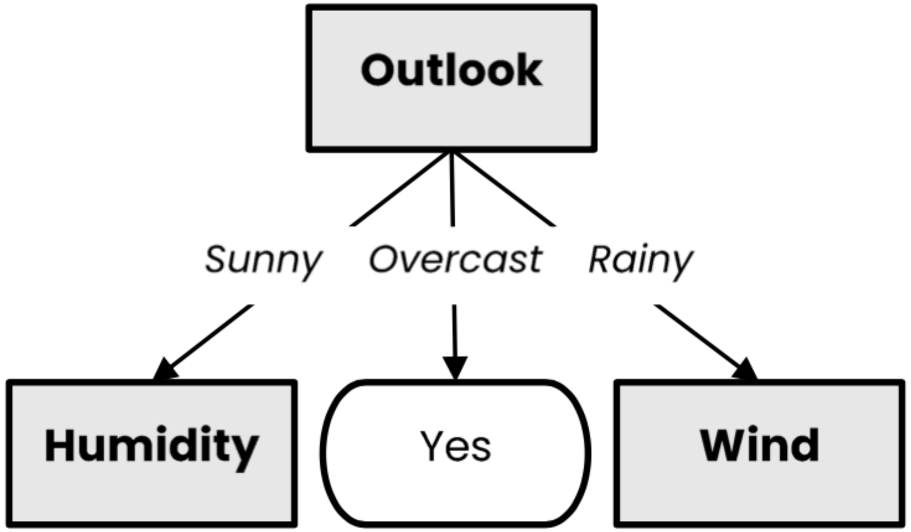
Temperature	Humidity	Wind	Play Tennis?
Hot	High	Weak	**No**
Hot	High	Strong	**No**
Mild	High	Weak	**No**
Cool	Normal	Weak	**Yes**
Mild	Normal	Strong	**Yes**

Outlook = *Overcast*

Temperature	Humidity	Wind	Play Tennis?
Hot	High	Weak	**Yes**
Cool	Normal	Strong	**Yes**
Mild	High	Strong	**Yes**
Hot	Normal	Weak	**Yes**

Outlook = *Rainy*

Temperature	Humidity	Wind	Play Tennis?
Mild	High	Weak	**Yes**
Cool	Normal	Weak	**Yes**
Cool	Normal	Strong	**No**
Mild	Normal	Weak	**Yes**
Mild	High	Strong	**No**



Decision Tree Learning

- Basic step: choose an attribute and, based on its values, split the data into smaller sets
  - Recursively repeat this step until we can surely decide the label

Outlook = *Sunny*

Humidity = High

Temperature	Wind	Play Tennis?
Hot	Weak	**No**
Hot	Strong	**No**
Mild	Weak	**No**

Humidity = Normal

Temperature	Wind	Play Tennis?
Cool	Weak	**Yes**
Mild	Strong	**Yes**

Outlook = *Overcast*

Temperature	Humidity	Wind	Play Tennis?
Hot	High	Weak	**Yes**
Cool	Normal	Strong	**Yes**
Mild	High	Strong	**Yes**
Hot	Normal	Weak	**Yes**

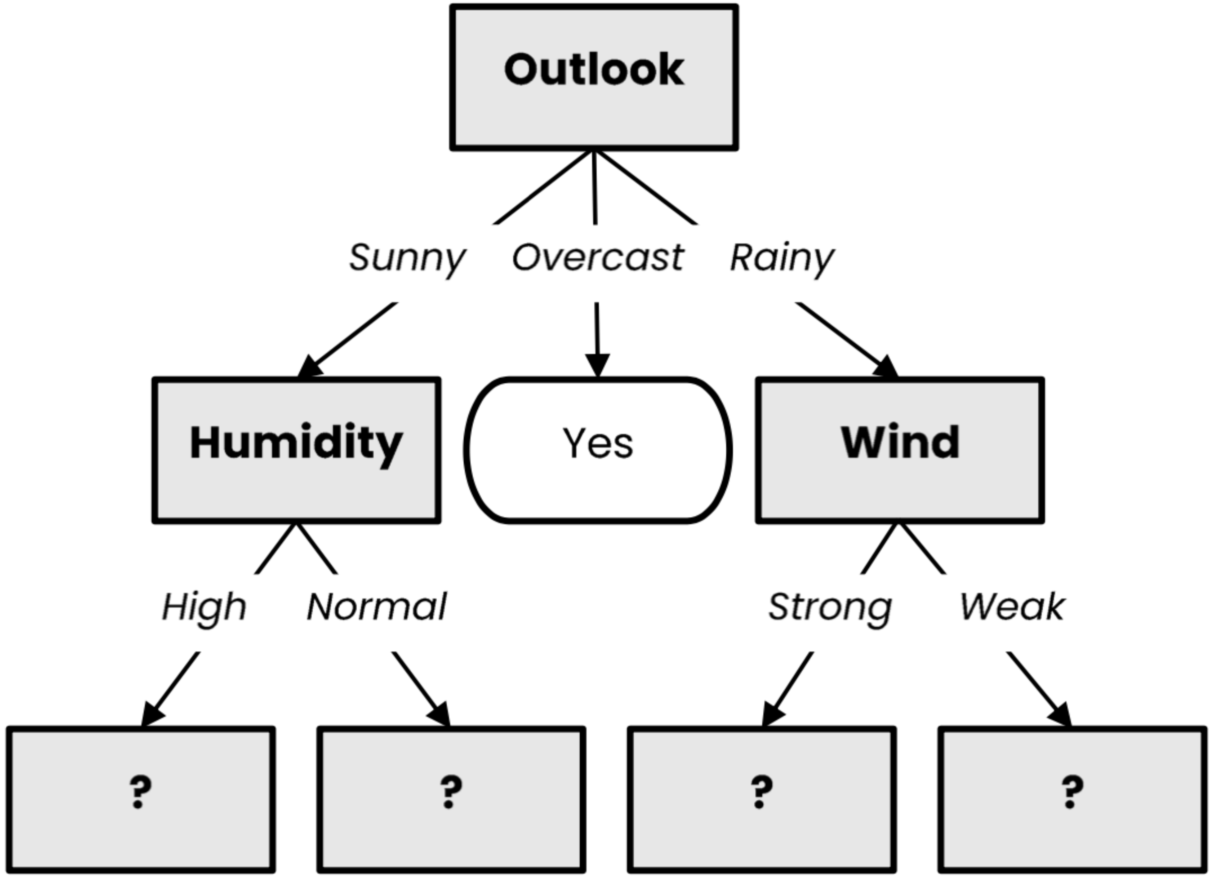
Outlook = *Rainy*

Wind = Strong

Temperature	Humidity	Play Tennis?
Cool	Normal	**No**
Mild	High	**No**

Wind = Weak

Temperature	Humidity	Play Tennis?
Mild	High	**Yes**
Cool	Normal	**Yes**
Mild	Normal	**Yes**



Decision Tree Learning

- Basic step: choose an attribute and, based on its values, split the data into smaller sets
  - Recursively repeat this step until we can surely decide the label

Outlook = *Sunny*

Humidity = High

Temperature	Wind	Play Tennis?
Hot	Weak	**No**
Hot	Strong	**No**
Mild	Weak	**No**

Humidity = Normal

Temperature	Wind	Play Tennis?
Cool	Weak	**Yes**
Mild	Strong	**Yes**

Outlook = *Overcast*

Temperature	Humidity	Wind	Play Tennis?
Hot	High	Weak	**Yes**
Cool	Normal	Strong	**Yes**
Mild	High	Strong	**Yes**
Hot	Normal	Weak	**Yes**

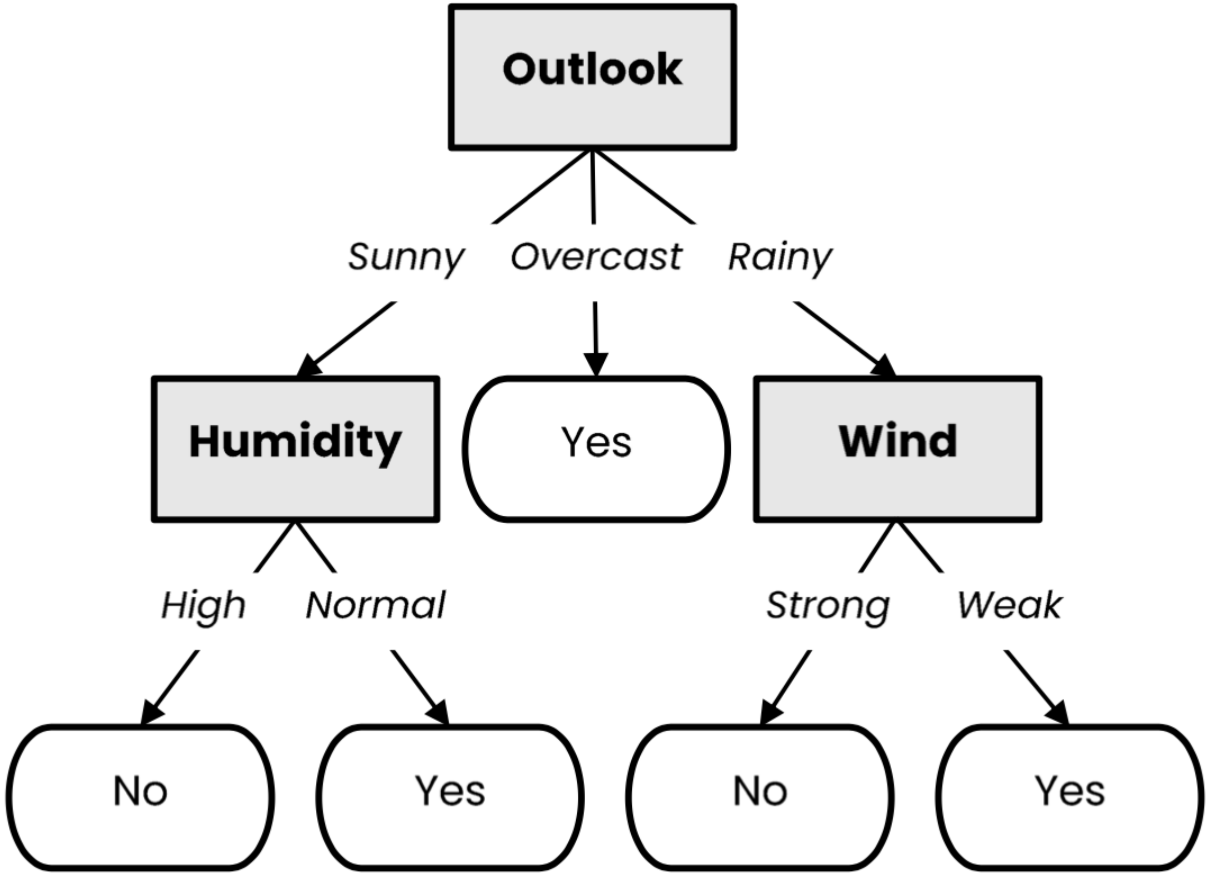
Outlook = *Rainy*

Wind = Strong

Temperature	Humidity	Play Tennis?
Cool	Normal	**No**
Mild	High	**No**

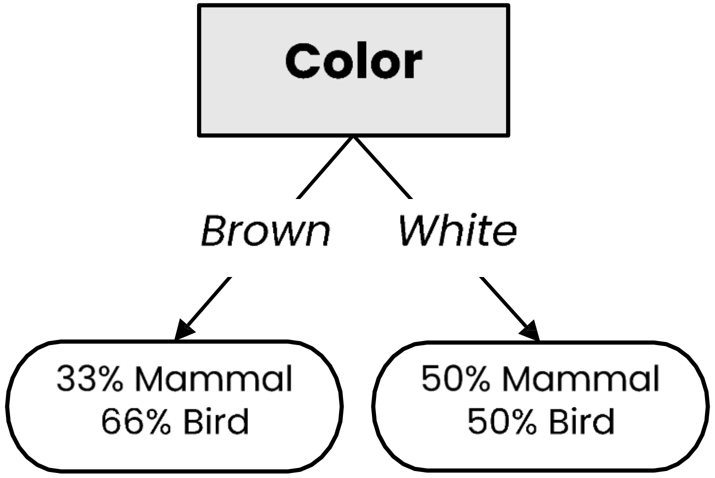
Wind = Weak

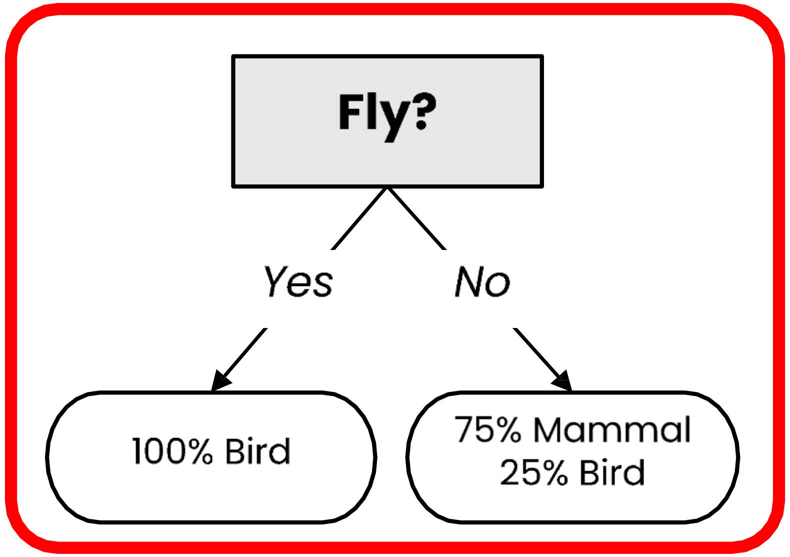
Temperature	Humidity	Play Tennis?
Mild	High	**Yes**
Cool	Normal	**Yes**
Mild	Normal	**Yes**



What is a good attribute?

Does it fly?	Color	Class
No	Brown	**Mammal**
No	White	**Mammal**
Yes	Brown	**Bird**
Yes	White	**Bird**
No	White	**Mammal**
No	Brown	**Bird**
Yes	White	**Bird**





- Which attribute provides better splitting?
- Why?
  - Because the resulting subsets are more pure
  - Knowing the value of this attribute gives us more information about the label (the entropy of the subsets is lower)

When To Stop Building the Tree?

- When all the leaf nodes are pure (leaf nodes have data that belong to one class)
- When a certain criteria is met (Such as the height of the tree exceeds certain limit, In which case the leaf nodes might not be pure but might output probability of a class instead of the class itself)
- When all the features are used.

Summary

- Decision tree: A tree-like model that illustrates series of events leading to certain decisions
- A good attribute selection leads to the better splitting of the data
- Stopping the building of tree is also important to prevent the model complexity and overfitting

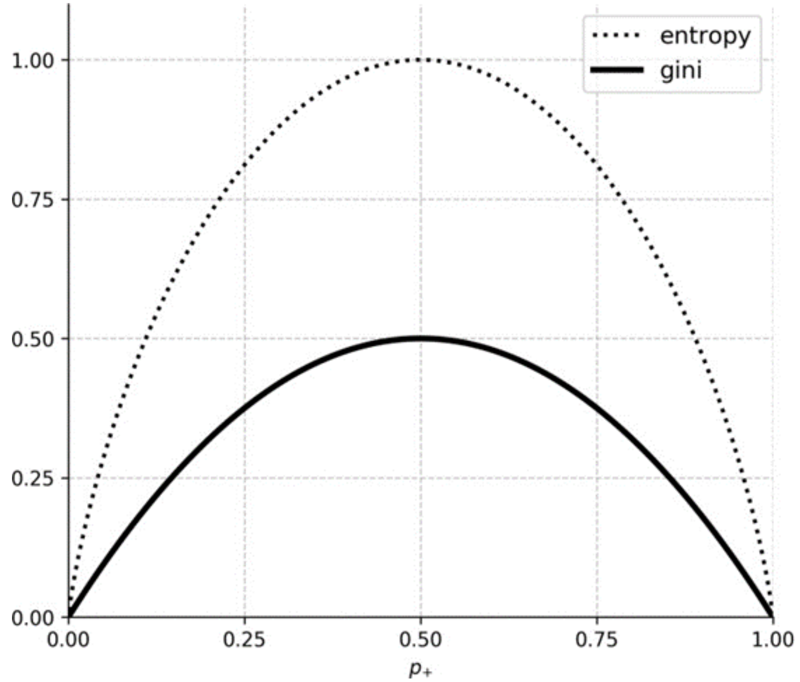
Machine learning

- Course Code:
- Unit 4

  Supervised Learning: Decision Trees and Ensemble Learning
- Lecture 3

  Ginni Index/Ginni Impurity

Gini Index/Gini Impurity



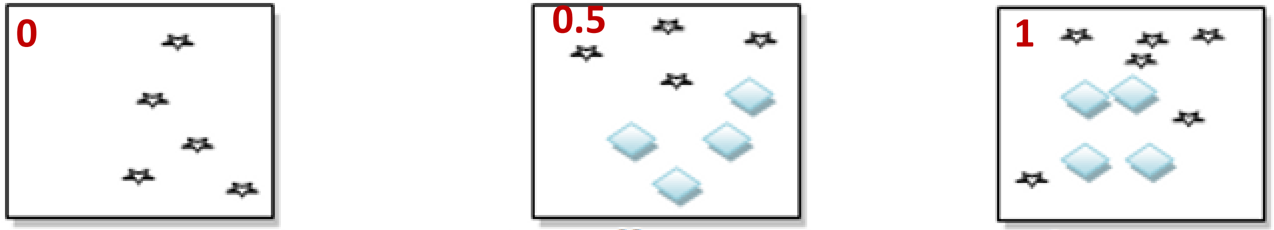
- Gini impurity measures how often a randomly chosen example would be incorrectly labeled if it was randomly labeled according to the label distribution

Gini Index/Gini Impurity

- It is calculated by subtracting the sum of the squared probabilities of each class from one.
- It favors larger partitions and easy to implement whereas information gain favors smaller partitions with distinct values.
- Gini Index works with the categorical target variable “Success” or “Failure”. It performs only Binary splits.
- Calculate Gini index using: Gini Index= $1 - \sum_j P_j^2$

Probability of each class

- If all the elements are linked with a single class then it is called pure. It ranges from 0-1



pure | Equally distributed | Impure

Note: An attribute with a lower Gini index should be preferred.

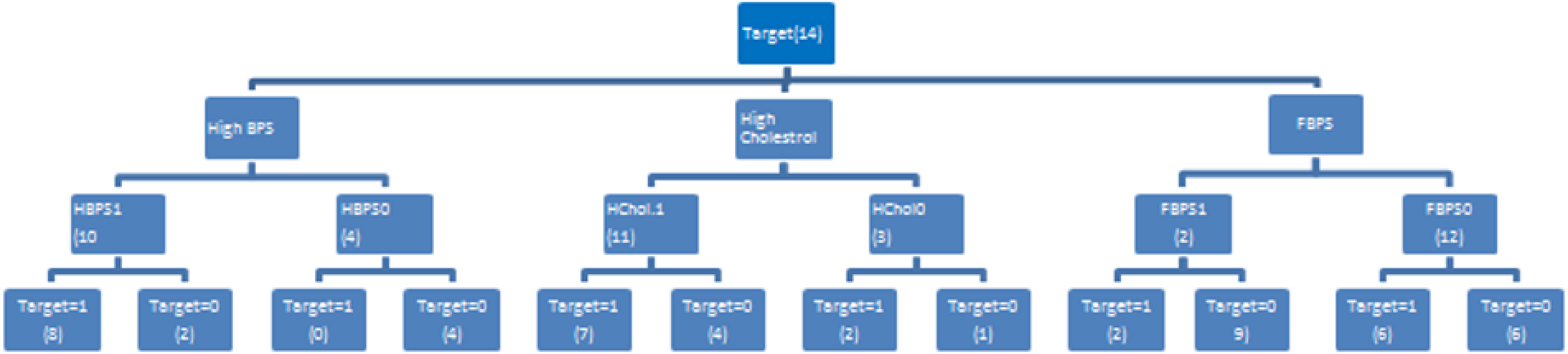
Problem: Identify the root node and calculate the gini index of each attribute?

SNO	HighBps	HighChol	Fbs	target
1	1	1	1	1
2	1	1	0	1
3	1	1	0	1
4	0	1	0	0
5	0	1	0	0
6	1	0	0	0
7	1	1	0	1
8	0	1	0	0
9	1	0	1	1
10	1	0	0	0
11	1	1	0	1
12	1	1	0	1
13	1	1	0	1
14	0	1	0	0

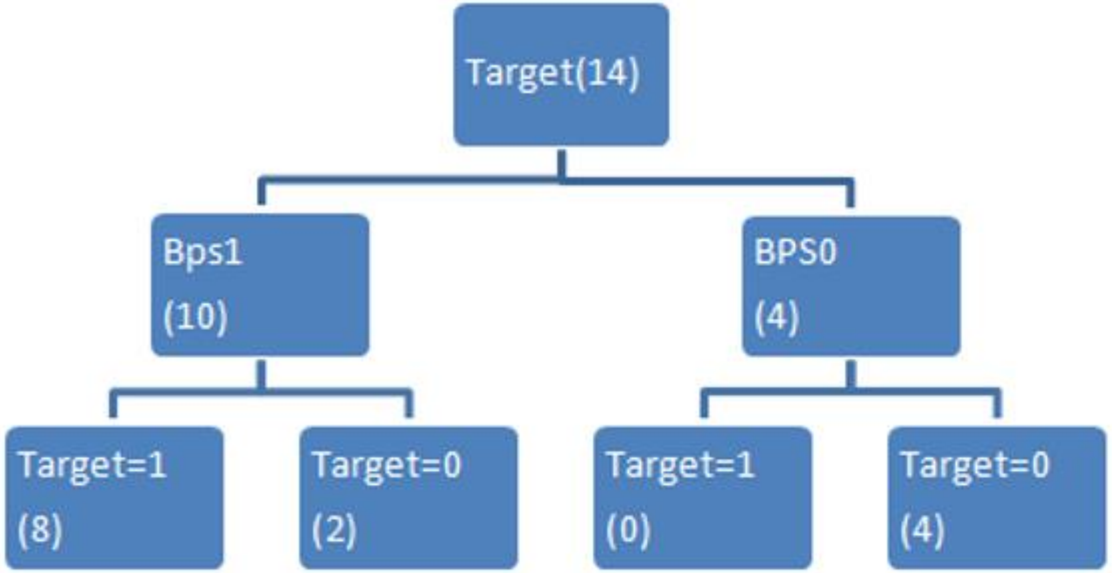
The given table represents on which factor heart disease occurs in person(target) dependent variable depending on HighBP, Highcholestrol, FBS(fasting blood sugar).

Note: The original values are converted into 1 and 0 which depict numeric classification

Decision tree for above table



- Gini index for High Bps



If (Bps=1 and target =1)=8/10

if(Bps=1 and target=0)=2/10

Gini index PBPS= $1-\left((PBps1)^2+(PBps0)^2\right)$

= 1-{{8/10)2+(2/10)2}

=0.32

Probability for parent node:

P0= 10/14

P1=4/14

Now we calculate for child node :

i) For BPS=1

for bps =1 this is the table

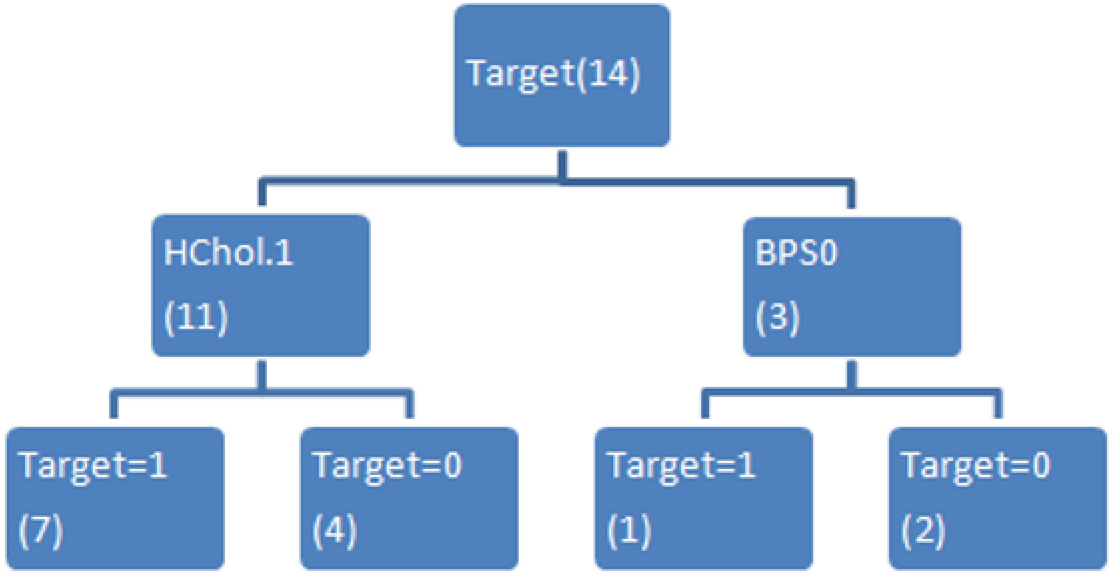
HighBps	HighChol	Fbs	Target
1	1	1	1
1	1	0	1
1	1	0	1
1	0	0	0
1	1	0	1
1	0	1	1
1	0	0	0
1	1	0	1
1	1	0	1
1	1	0	1

2) if BPS=0,

HighBps	HighChol	Fbs	Target
0	1	0	0
0	1	0	0
0	1	0	0
0	1	0	0

If (BPS=0 and target=0)=4/4=1
If (BPS=0 and target=1)=0
Gini index PBPS0=1-{(1)-(0)}
= 1-1
=0
Weighted Gini index
w.g =P0*GBPS0+ P1*GBPS1
= 4/14*0 + 10/14*0.32
= 0.229

- Gini index for High Cholestrol:



If (Hchol.=1 and target=1)=7/11
If (HChol.=1 and target=0)=4/11

Gini index = 1-[(7/11)2+(4/11)2]

= 0.46

Probability of parent node
P1=11/14
P0=3/13
1) For HChol.=1

HighBps	HighChol	FBS	Target
1	1	1	1
1	1	0	1
1	1	0	1
0	1	0	0
0	1	0	0
1	1	0	1
0	1	0	0
1	1	0	1
1	1	0	1
1	1	0	1
0	1	0	0

2) If HChol.=0

HighBps	HighChol	FBS	Target
1	0	0	0
1	0	1	1
1	0	0	0

If (Hchol.=0 and target=1)=1/3

If (HChol.=0 and target=0)=2/3

Gini index= 1-[(1/3)2+(2/3)2]

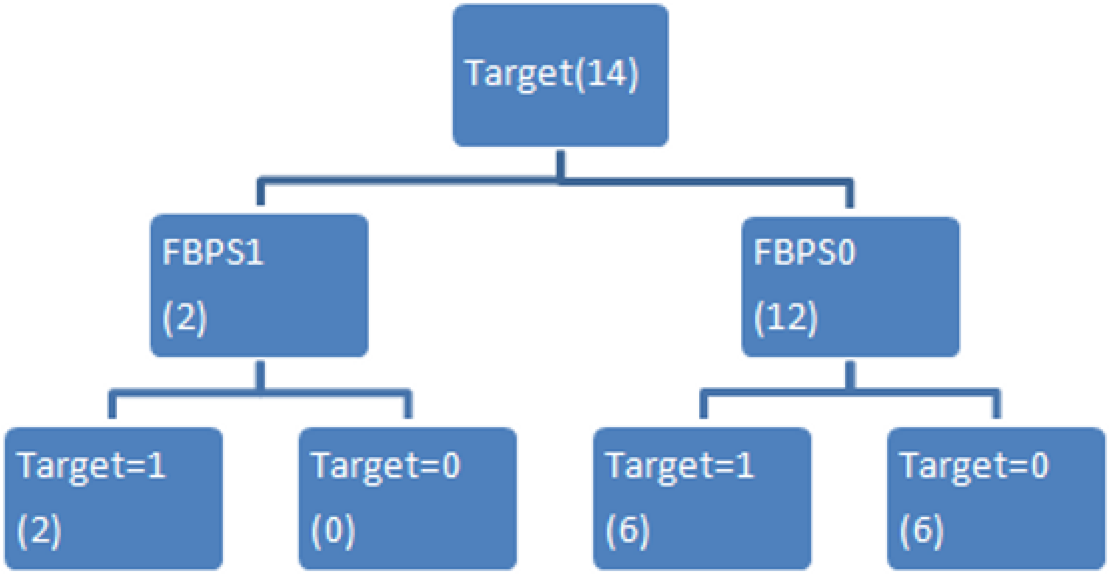
= 0.55

Weighted Gini index = P0*GHChol.0+P1*GHChol.1

= 3/14*055+11/14*0.46

= 0.47

- Gini index for FBPS:



Probability of parent node

P1=2/14

P0=12/14

1) for FBPS=1

HighBps	HighChol	Fbs	Target
1	1	1	1
1	0	1	1

If (FBps=1 and target =1)=2/2

if(FBps=1 and target=0)=0

Gini index PFBPS1=1-{(PFBPS1)2+(PFBPS0)2

= 1- [(1)2+0]

=1-1=0

2) for FBPS=0,

HighBps	HighChol	Fbs	Target
1	1	0	1
1	1	0	1
0	1	0	0
0	1	0	0
1	0	0	0
1	1	0	1
0	1	0	0
1	0	0	0
1	1	0	1
1	1	0	1
1	1	0	1
0	1	0	0

If (FBps=0 and target =1)=6/12=0.5
if(FBps=0 and target=0)=6/12=0.5

Gini index PFBPS0=1-{(PFBPS1)2+(PFBPS0)2]
= 1-[(0.5)2+(0.5)2]

= 0.5

Weighted Gini index = P0*GFBPS0+ P1*GFBPS1
= 6/7*0.5+1/7*0

=0.42

Comparing Gini Index:

Attribute	HighBPS	Hchol.	FBPS
Gini index	0.22	0.47	0.42

As *HighBPS* is less it is the winner

Conclusion: *HighBPS* is used as the *root node* for constructing of Decision Tree and the further tree is built.