Machine Learning

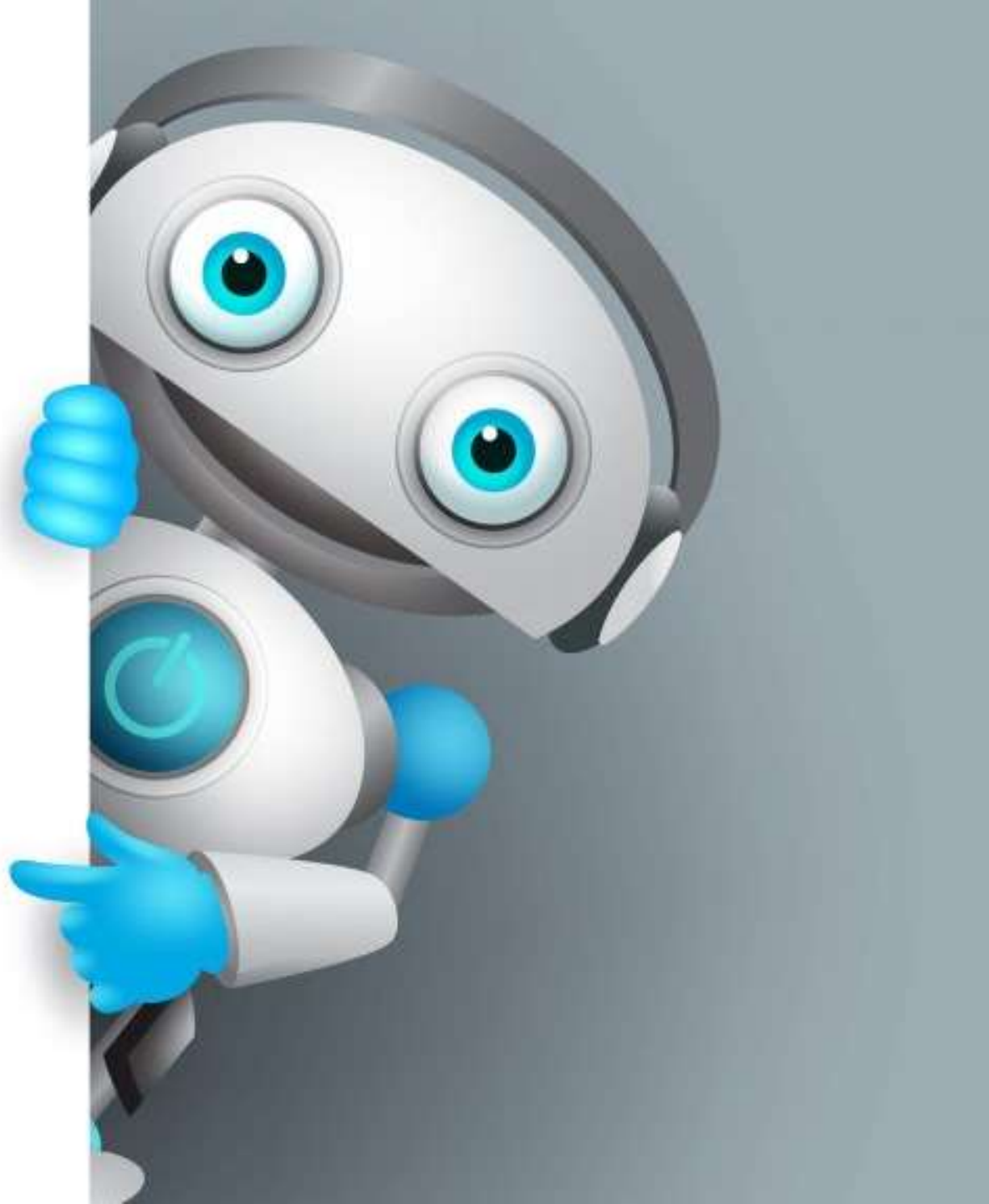
SUPPORT VECTOR MACHINE

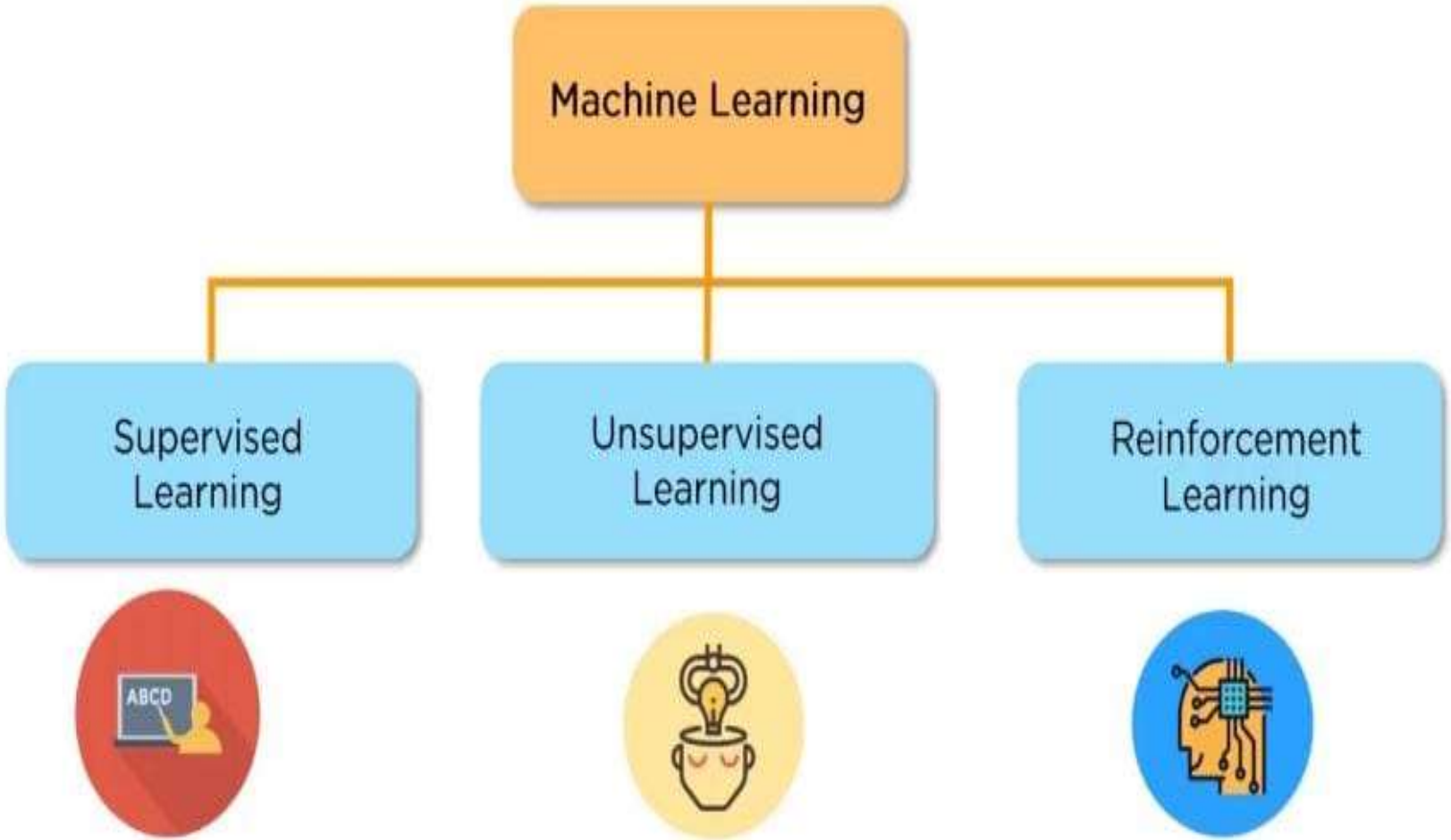
Agenda

- What is Machine Learning?
- Why Support Vector Machine?
- What is Support Vector Machine?
- Understanding Support Vector Machine
- Advantages of Support Vector Machine
- Use Case in Python

What is Machine Learning ?

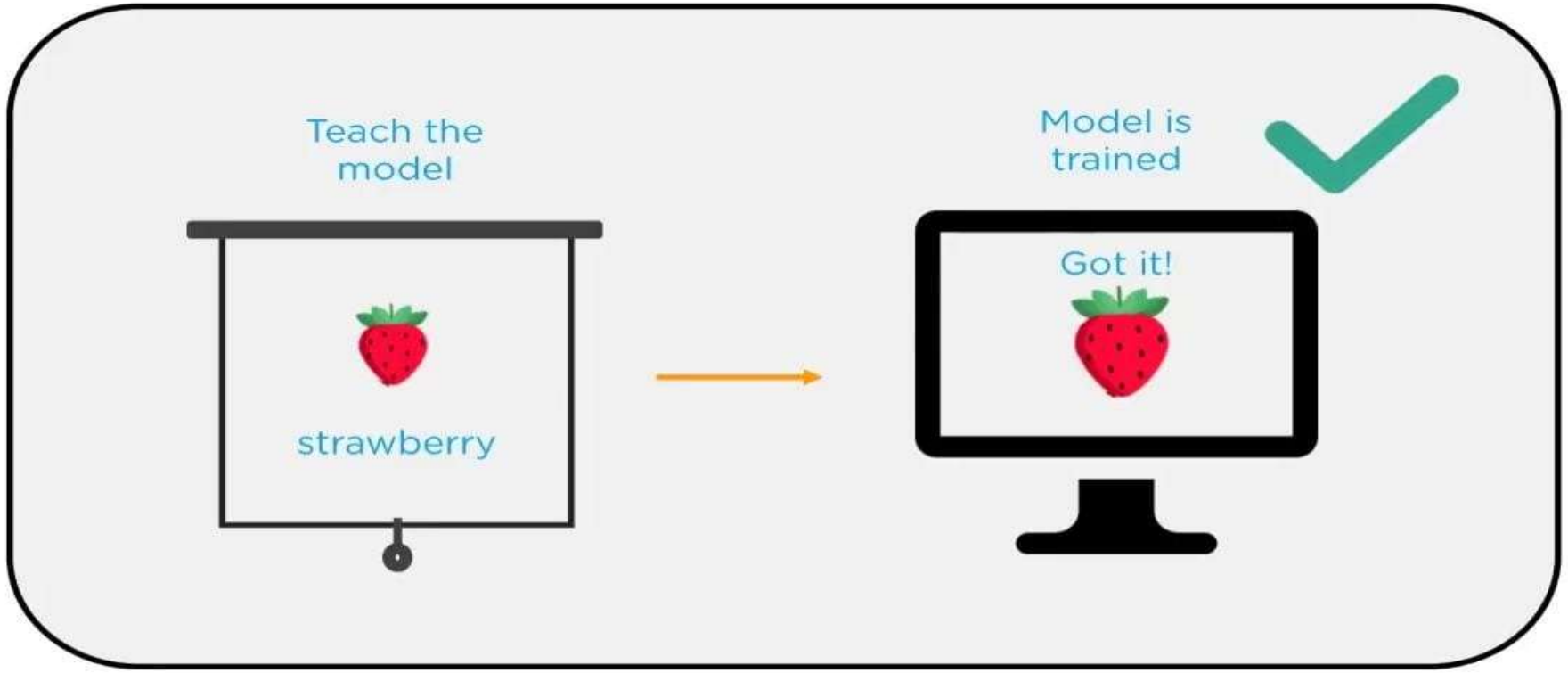
- Machine Learning is a subset of Artificial Intelligence .It focuses mainly on the designing of systems, thereby allowing them to learn and make predictions based on some experience which is data in case of machines.

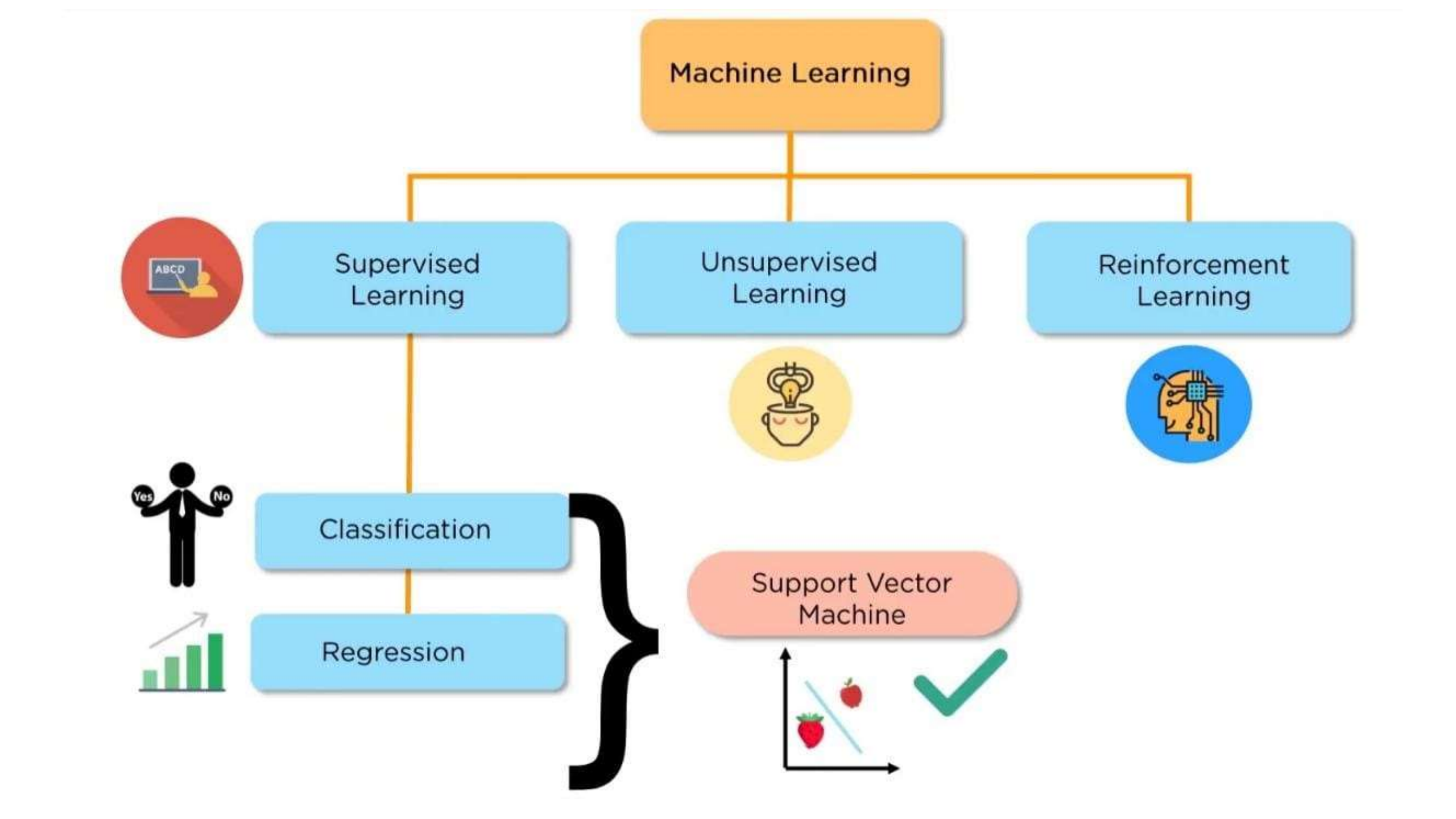



Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
ABCD

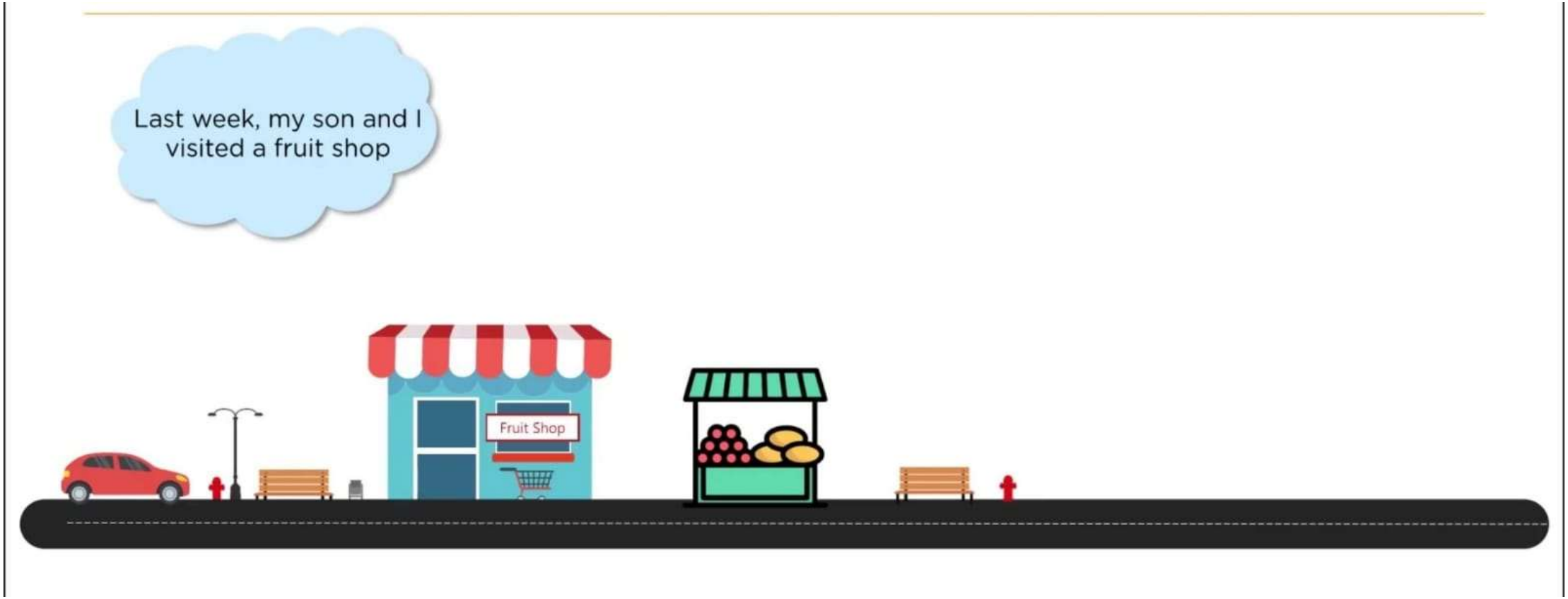
SUPERVISED LEARNING

Machine Learning model learns from the past input data and makes future prediction as output.




Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
ABCD
Yes
No
Classification
Regression
Support Vector Machine

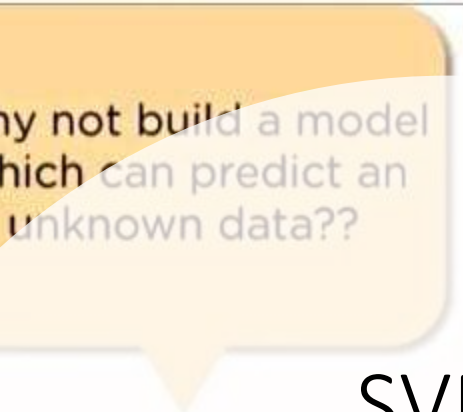
Why SVM & its Case Study?



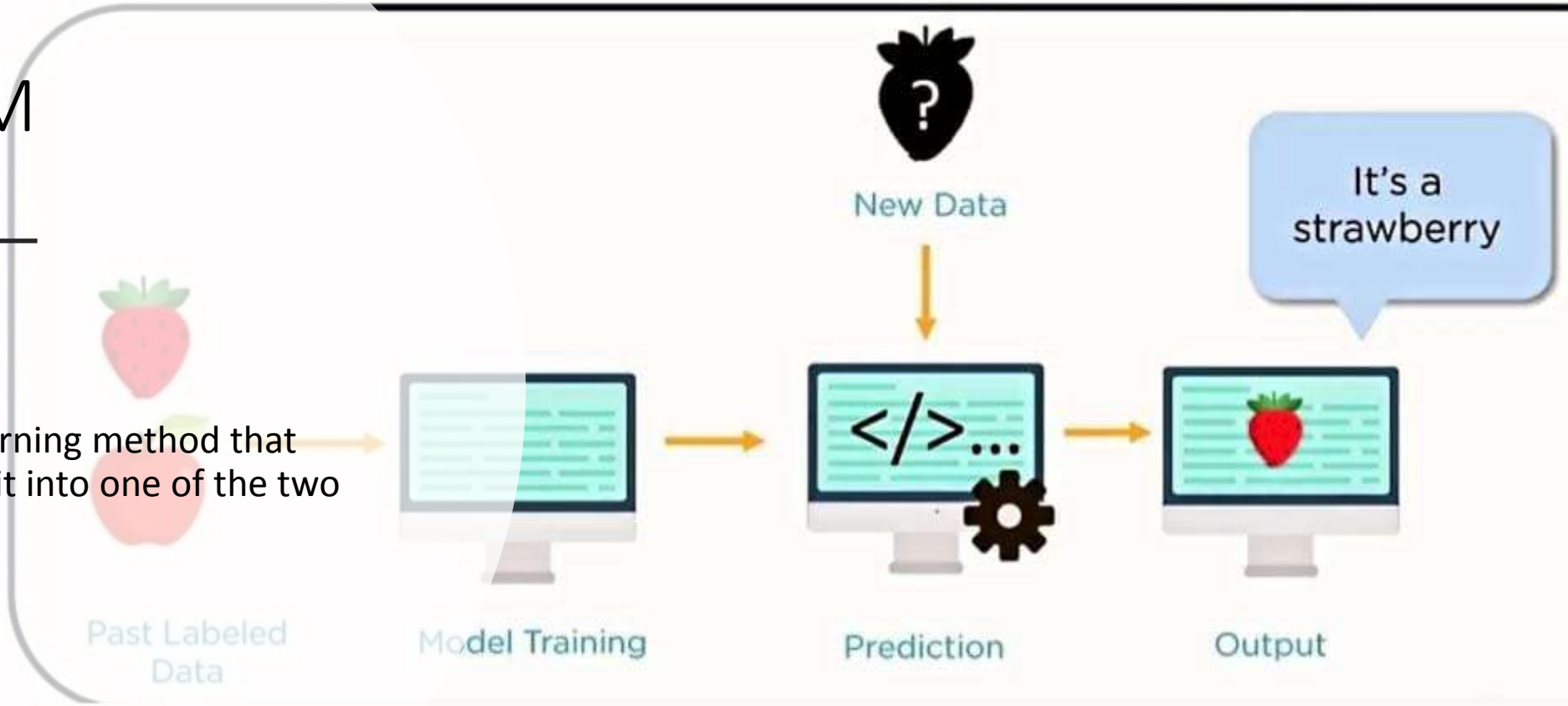

There, he found a fruit which was similar to both apple and strawberry
Dad, is that an apple or a strawberry?
?
Fruit Shop

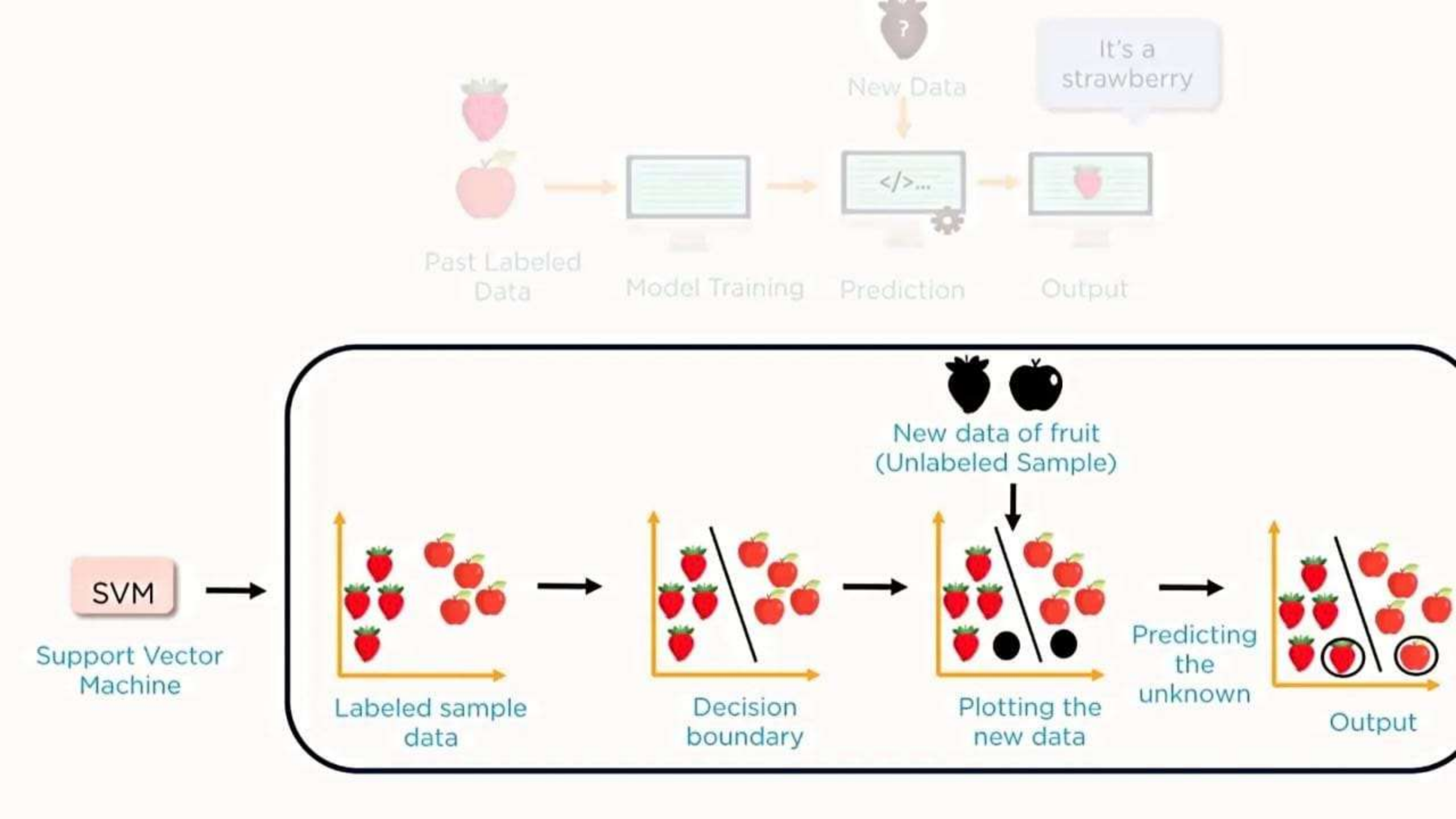

After a couple of seconds, he could figure out that it was a strawberry
It is a strawberry!
Fruit Shop

SVM



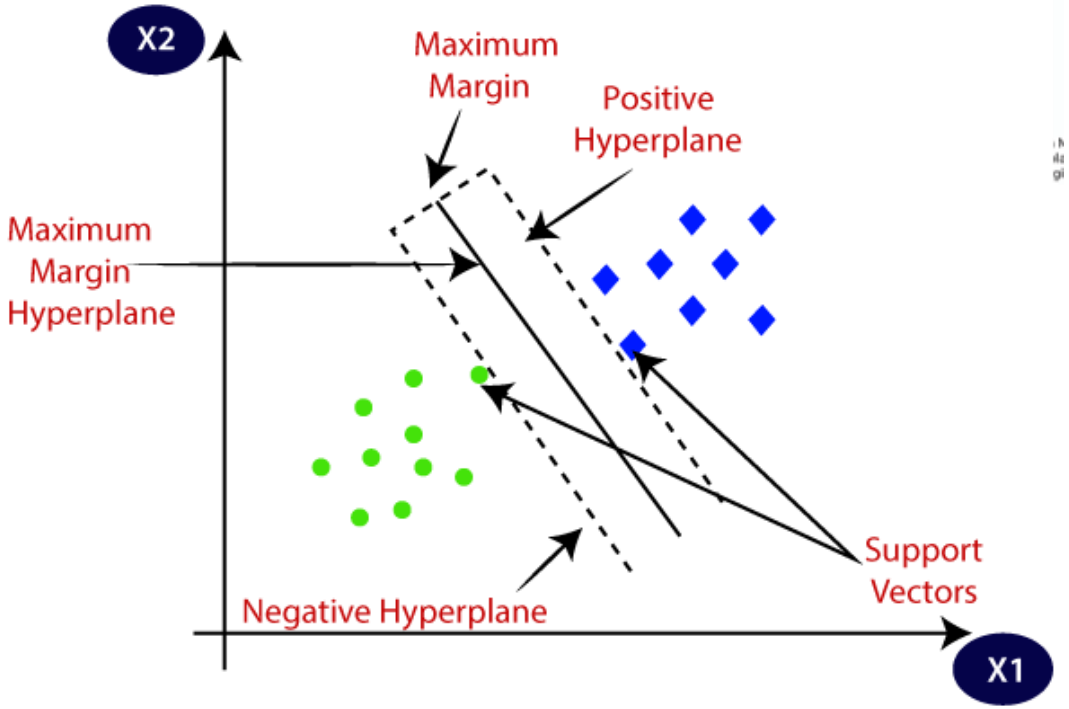
- SVM is a supervised learning method that looks at data and sorts it into one of the two categories.




New Data
It's a strawberry
Past Labeled Data
Model Training
Prediction
Output
New data of fruit (Unlabeled Sample)
SVM
Support Vector Machine
Labeled sample data
Decision boundary
Plotting the new data
Predicting the unknown
Output

UNDERSTANDING SVM ?

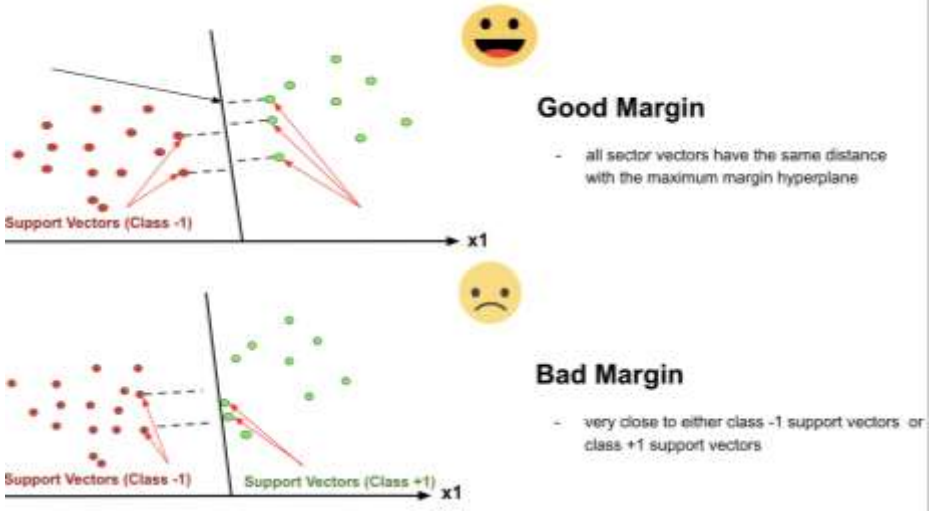
- Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.
- The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.
- SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

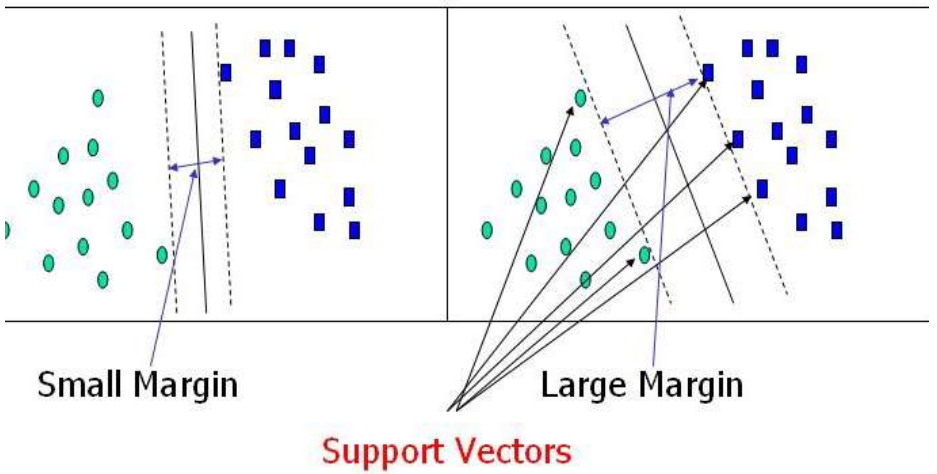


GOAL & it’s Keywords

- The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

- Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.



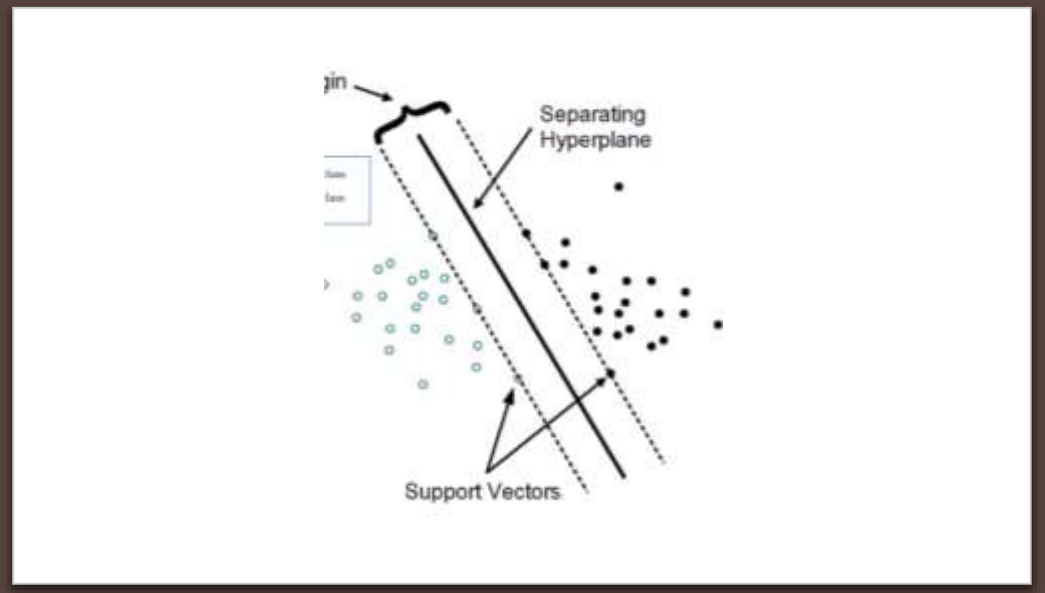


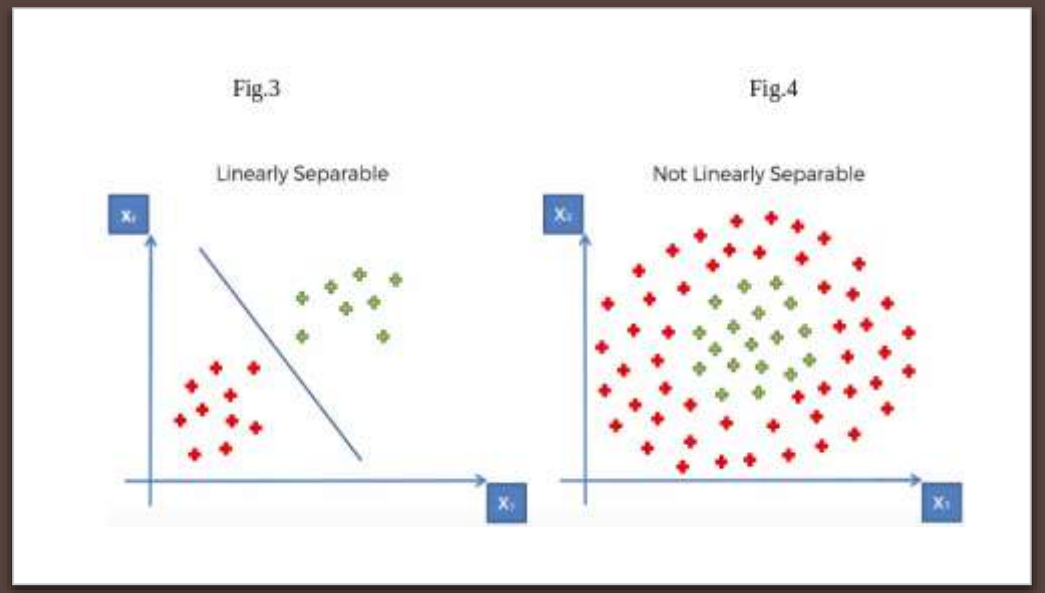
- **Margin is distance from the decision surface to the closest data point**

Positive Hyperplane + Negative Hyperplane =Margin

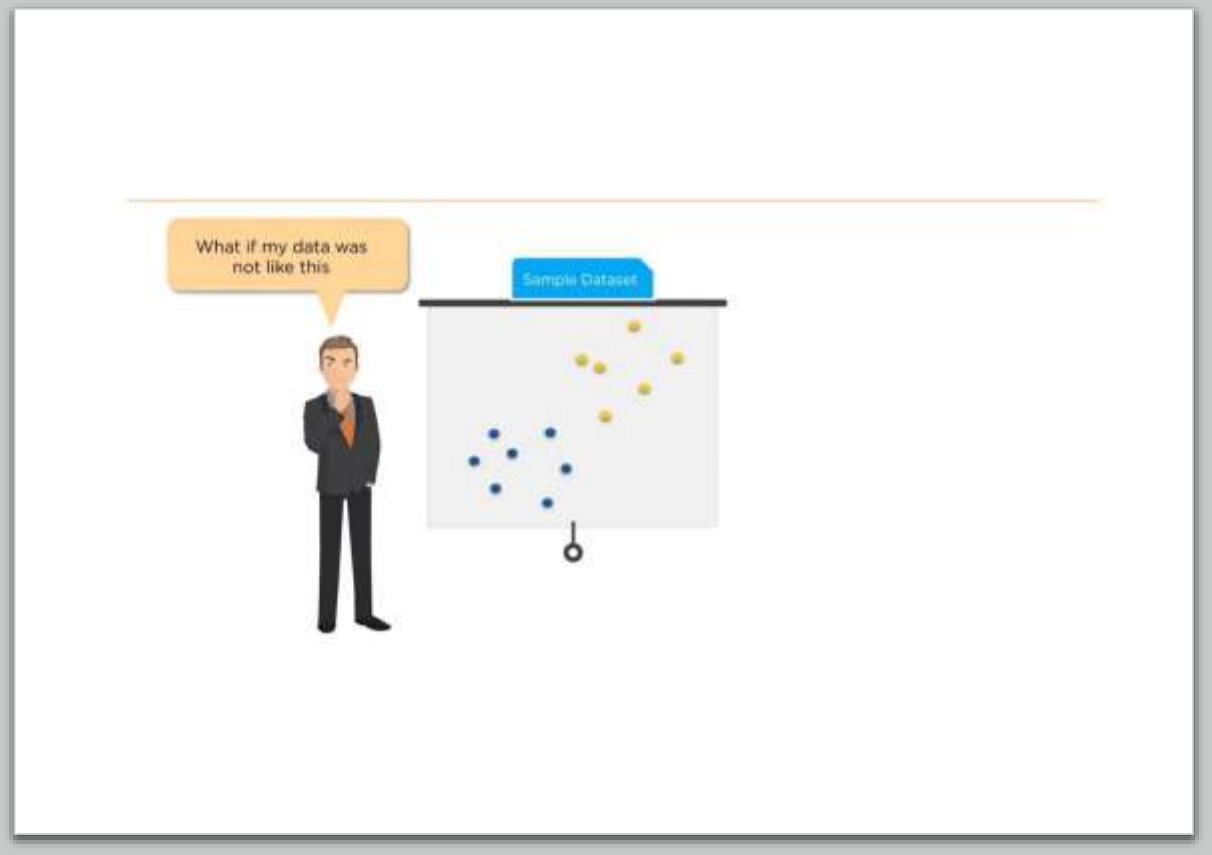
Suppose, D1+D2=M

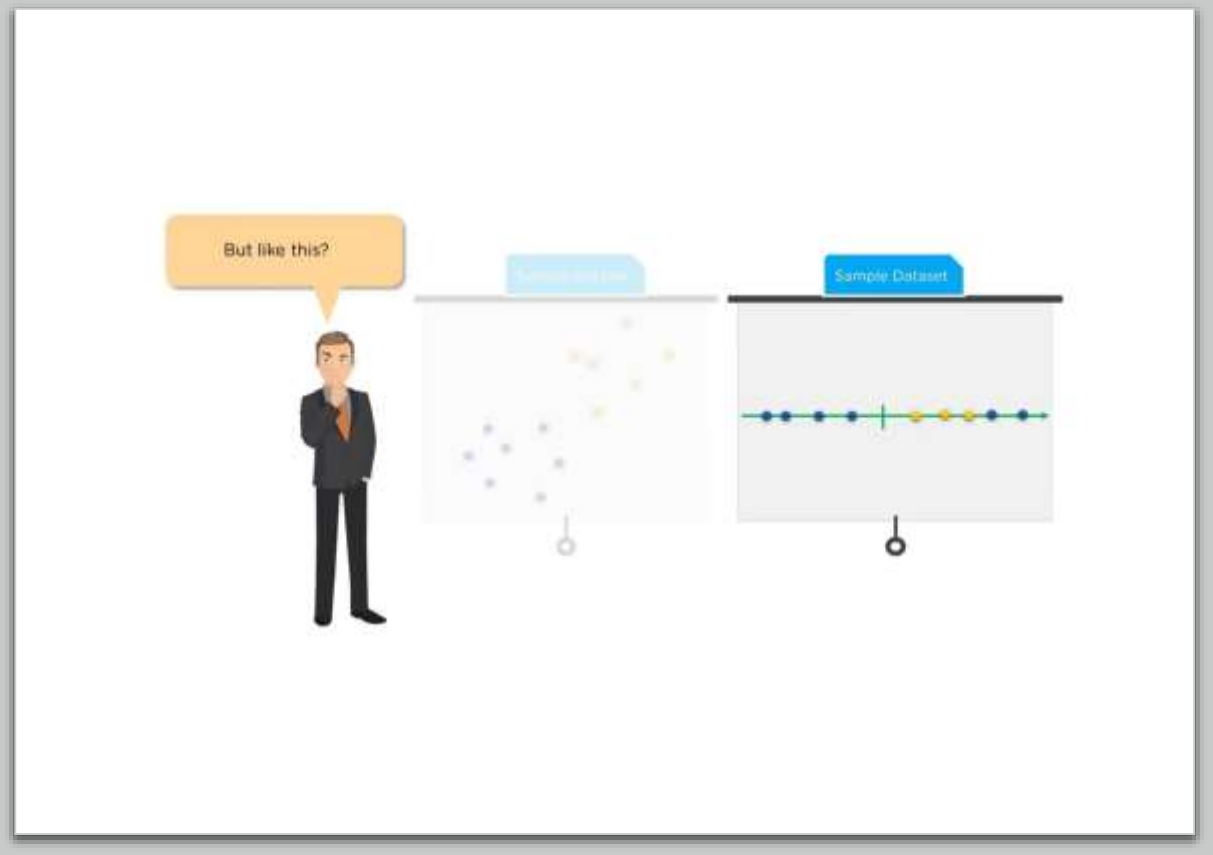
- **Linearly Separable**





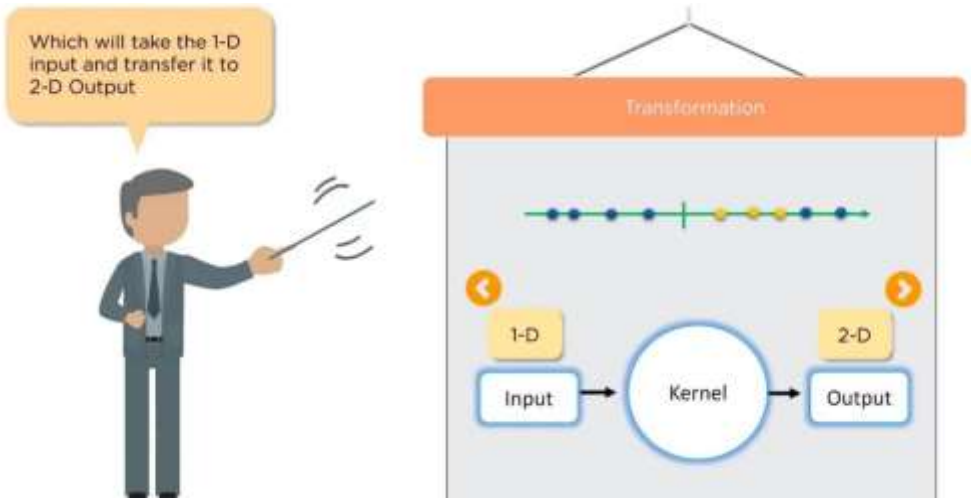
KERNAL

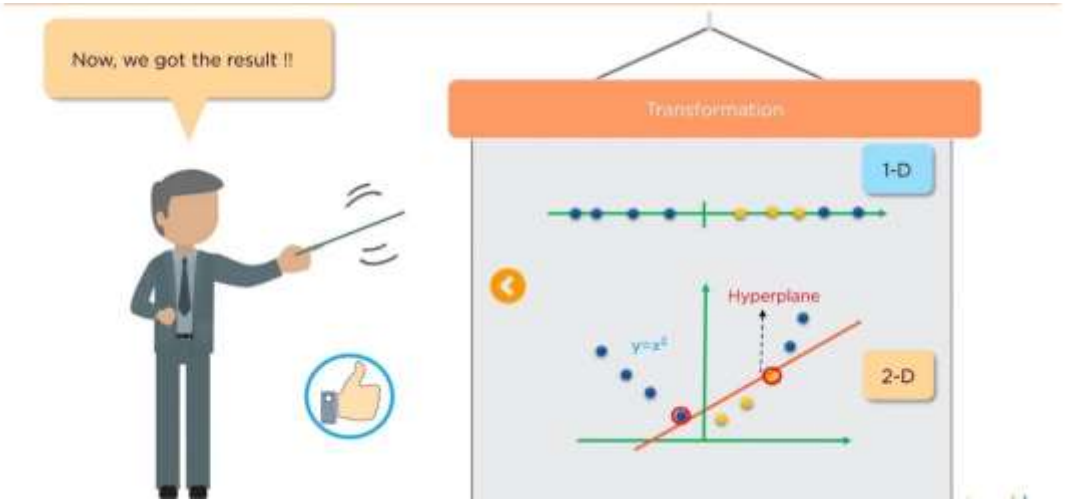

What if my data was not like this
Sample Dataset


But like this?
Sample Dataset

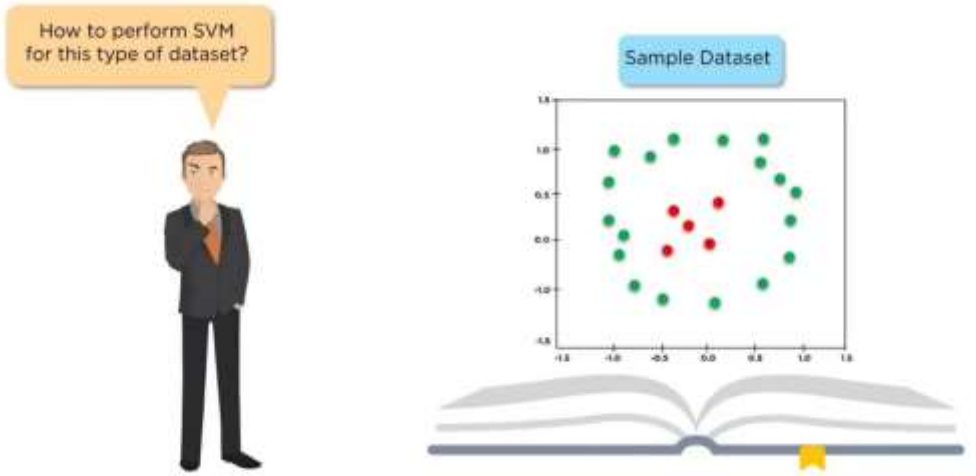
NOT LINEARLY SEPARABLE

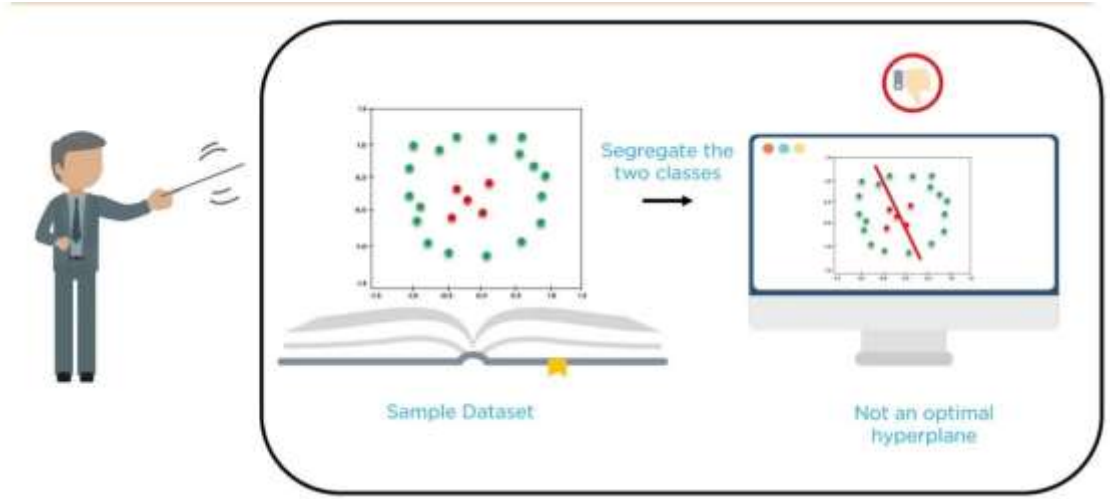
- 1-D to 2-D

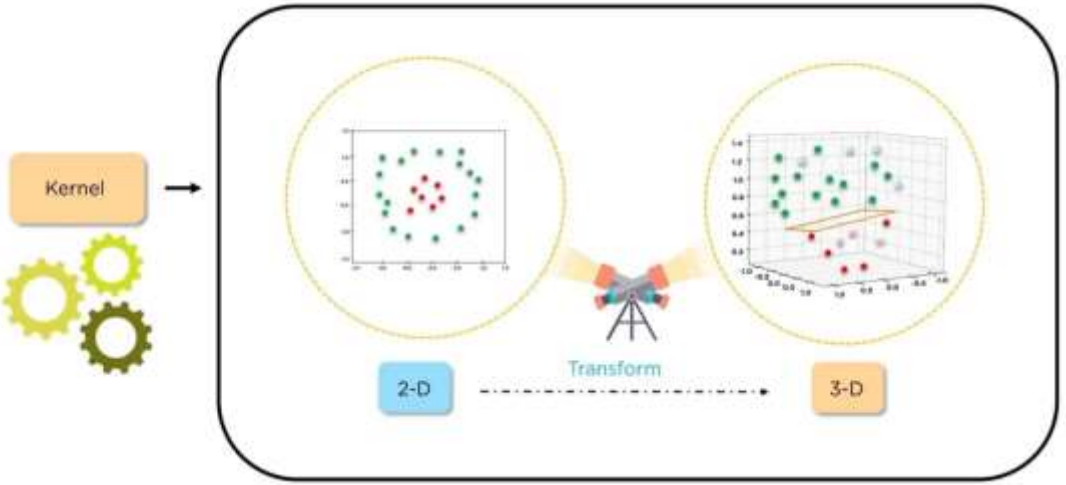


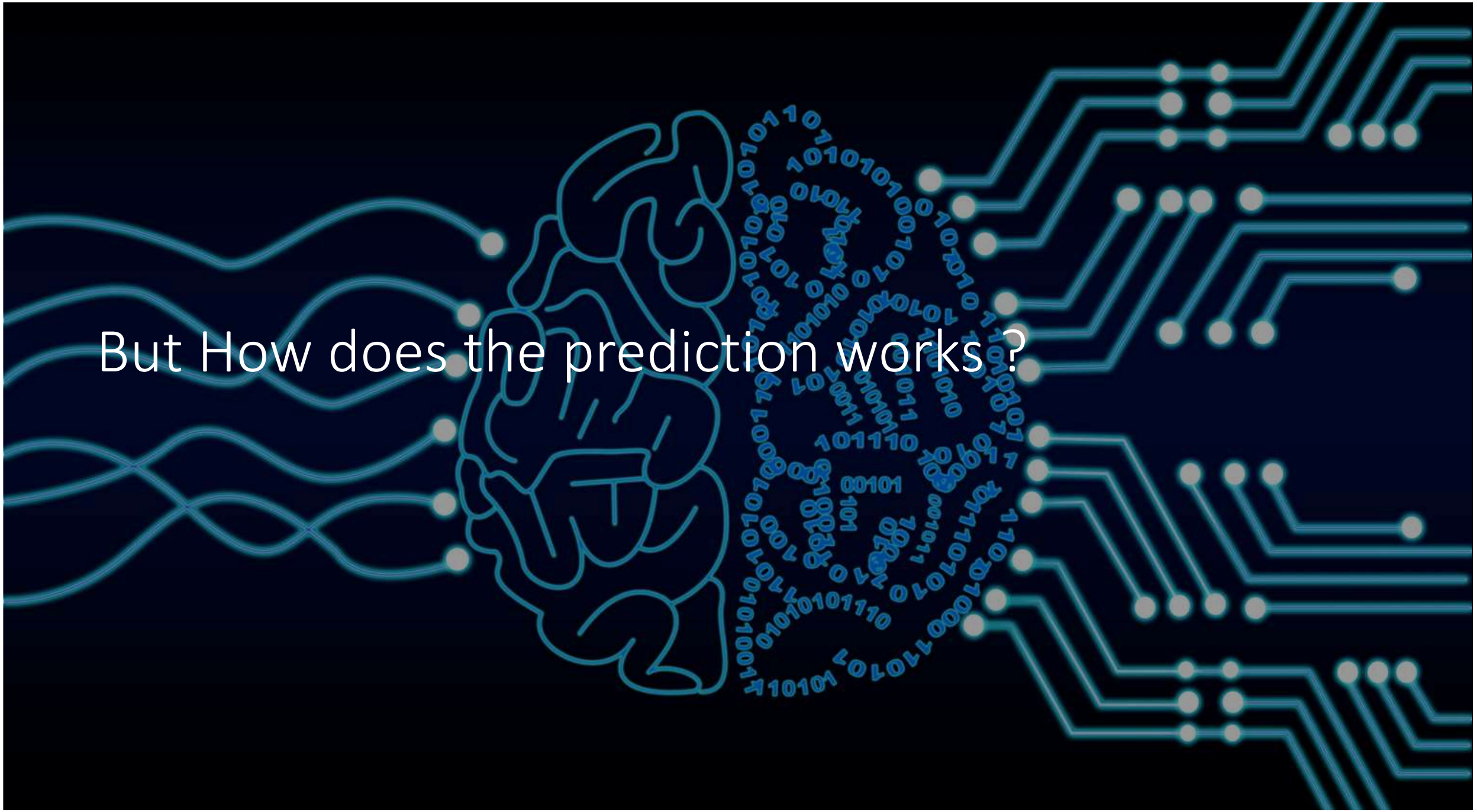


2-D to 3-D

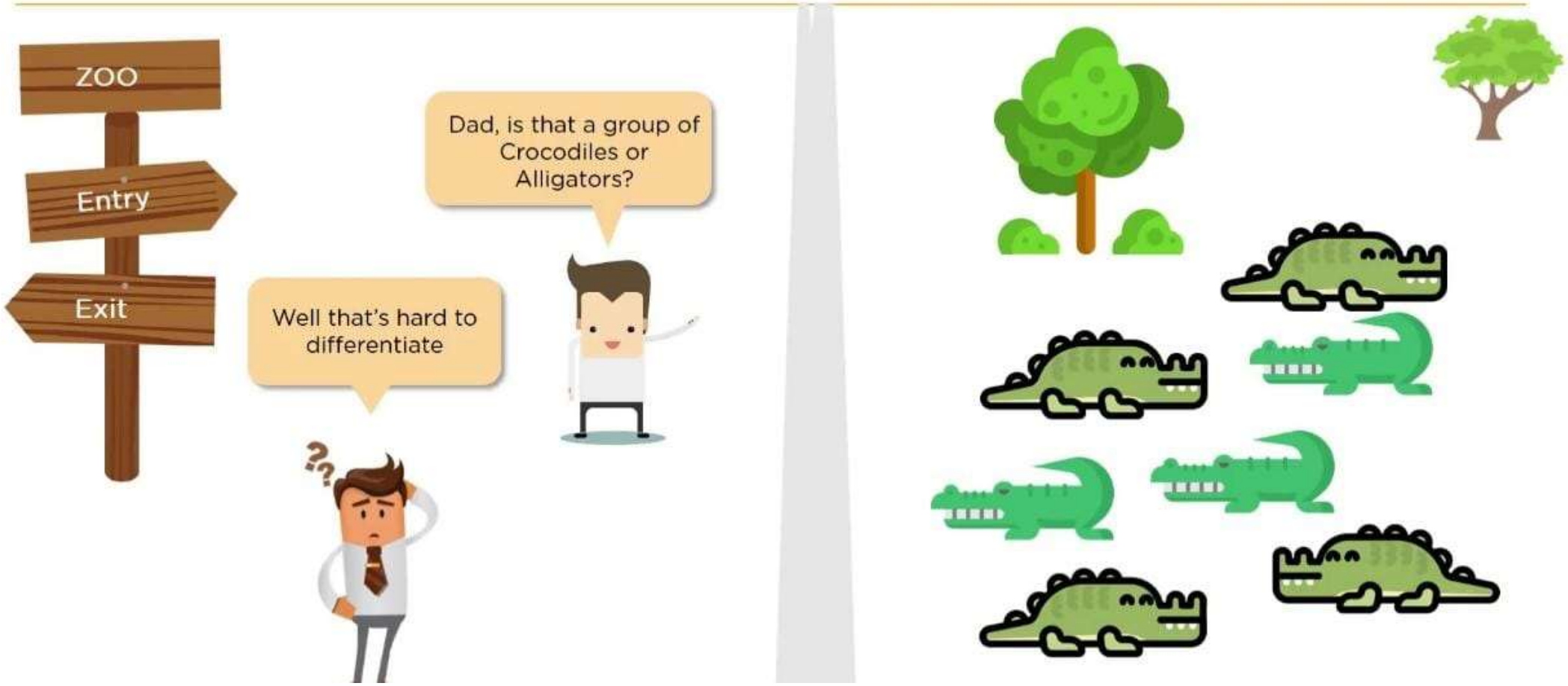

How to perform SVM for this type of dataset?
Sample Dataset


Segregate the two classes
Sample Dataset
Not an optimal hyperplane


Kernel
2-D
Transform
3-D

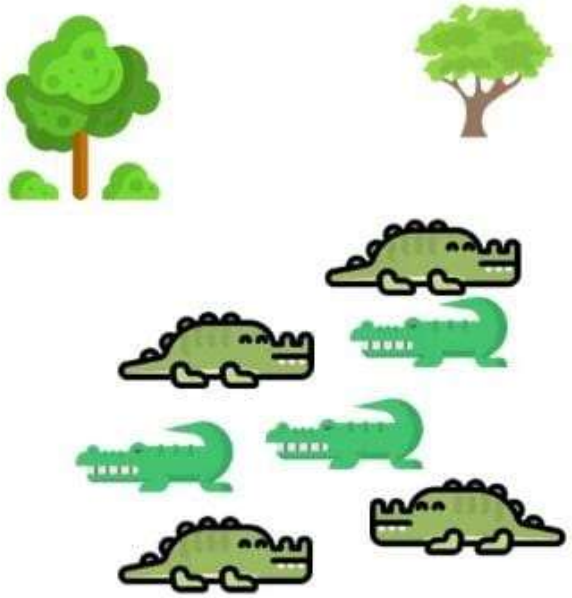

But How does the prediction works ?

USE CASE – PROBLEM STATEMENT

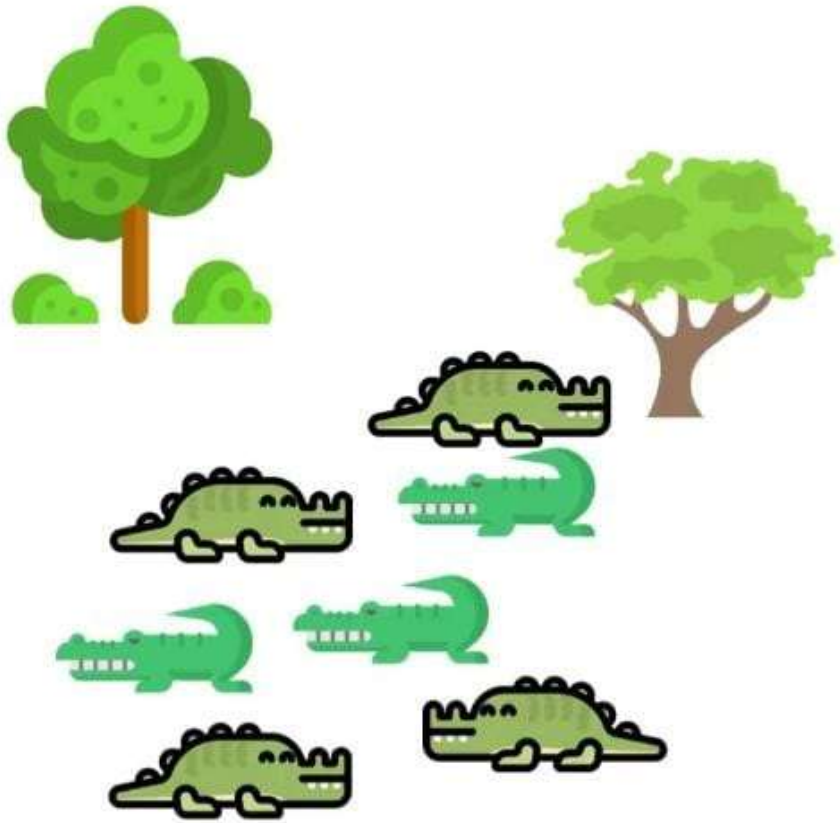


Difference

	Crocodiles	Alligators
Size	• Crocodiles are larger in size	• Alligators are smaller in size
Snout Width	• Crocodiles have narrow snout	• Alligators have wider snout




Let Support Vector Machine segregate the two groups
Dad, is that a group of Crocodiles or Alligators?



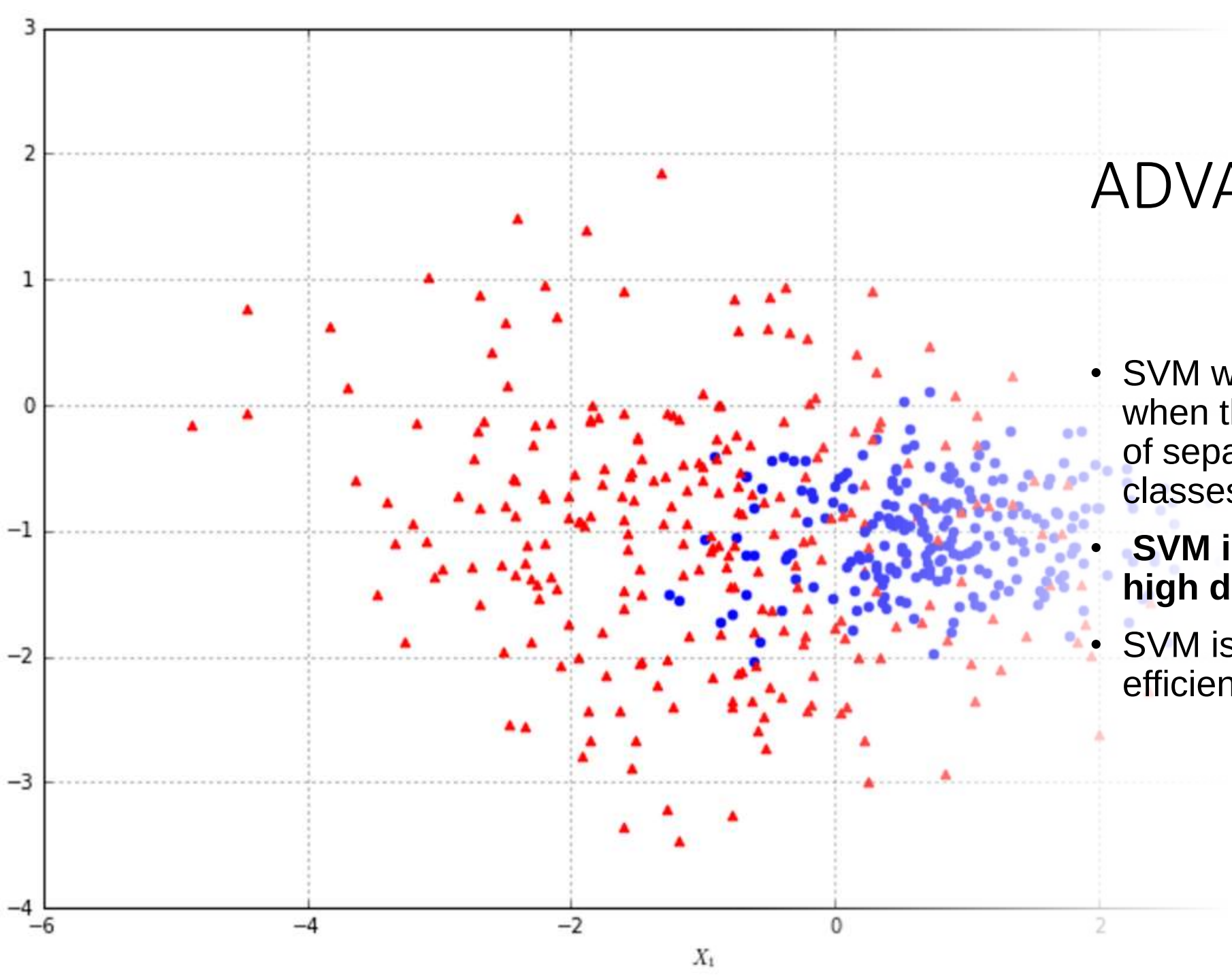
SVM Use Cases

Face detection

Text and hypertext categorization

Classification of images

Bioinformatics



ADVANTAGES

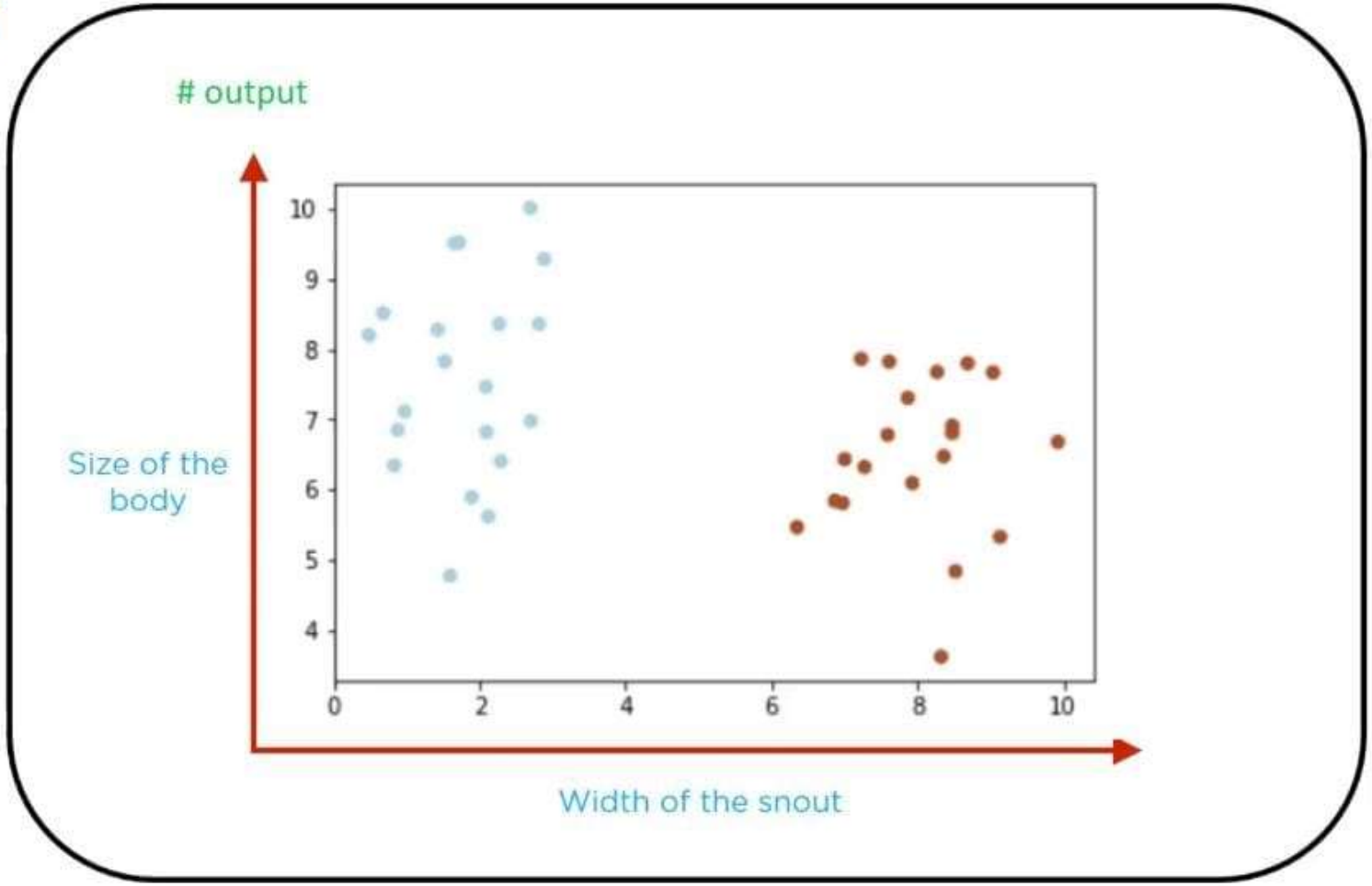
- SVM works relatively well when there is a clear margin of separation between classes.
- **SVM is more effective in high dimensional spaces**.
- SVM is relatively memory efficient.

Use Case in PYTHON ?

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn import svm
from sklearn.datasets.samples_generator import make_blobs
# we create 40 separable points
X, y = make_blobs(n_samples=40, centers=2, random_state=20)

# fit the model, don't regularize for illustration purposes
clf = svm.SVC(kernel='linear', C=1000)
clf.fit(X, y)

plt.scatter(X[:, 0], X[:, 1], c=y, s=30, cmap=plt.cm.Paired)
```


output
Size of the body
Width of the snout
10
9
8
7
6
5
4
0
2
4
6
8
10

THANK YOU