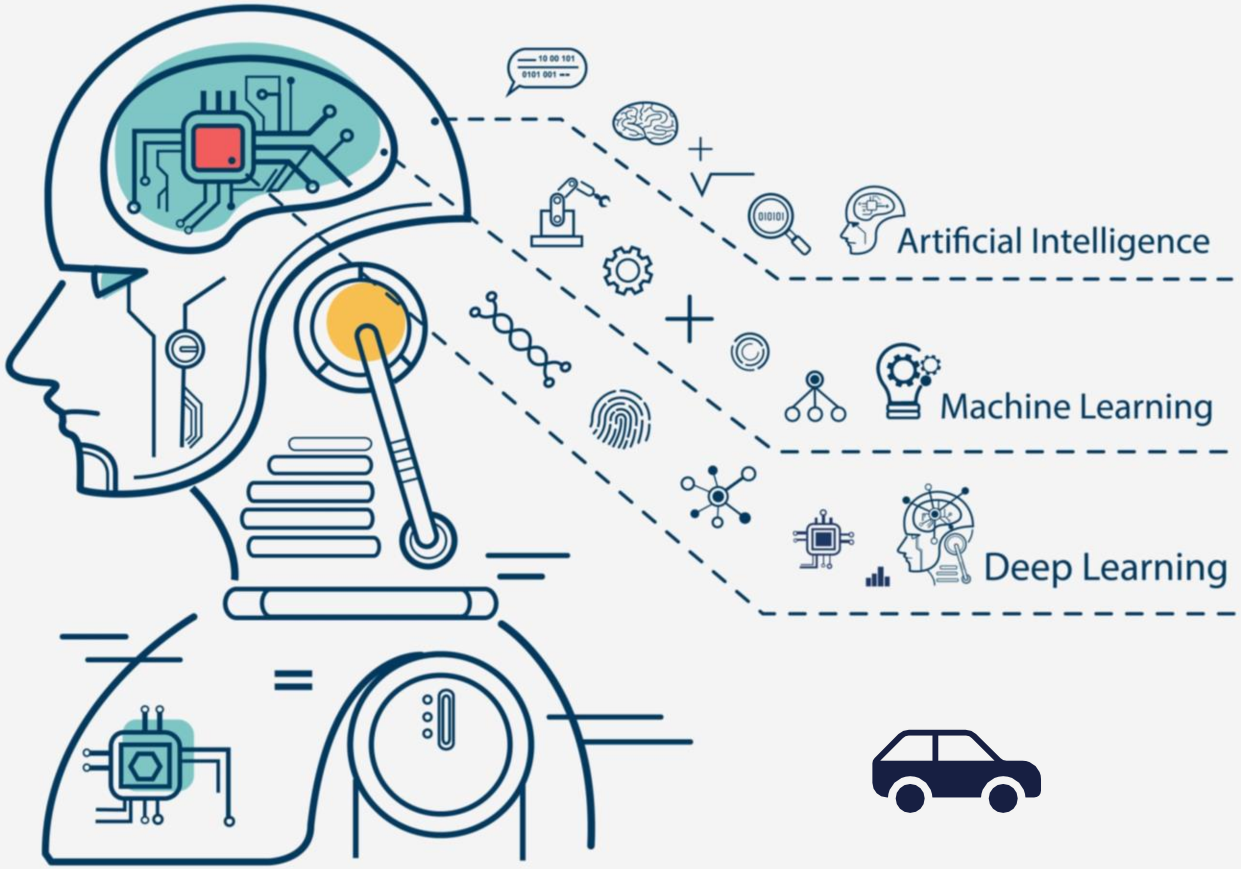
APSSDC

Andhra Pradesh State Skill Development Corporation

Skill AP
APSSDC



MACHINE LEARNING USING PYTHON

DAY 04 AGENDA

Logistic regression

Support Vector Machines

By Anil Kumar APSSDC

APSSDC

Andhra Pradesh State Skill Development Corporation

Logistic Regression for Machine Learning

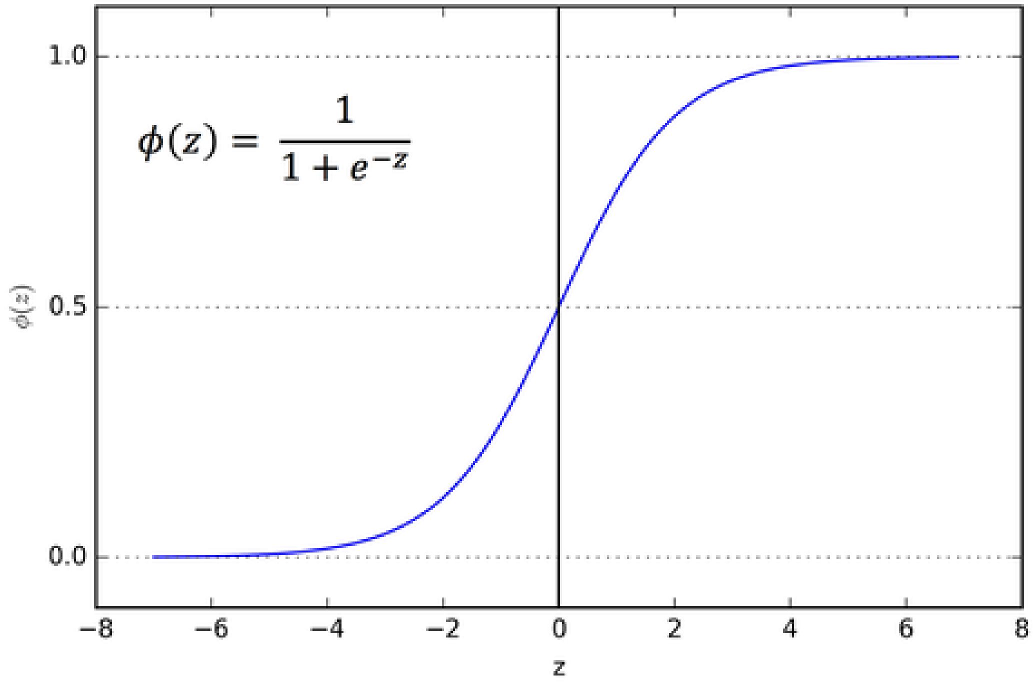
INTRODUCTION

- Supervised Learning Classification Problem
- When we classify only two classes it is called Binary Classification problem.
- Logistic Regression is a type of regression analysis used for predicting the outcome of dependent variable.
- Dependent variable(Y) is binary(0,1) and independent variables (X) are continuous in nature.
- Dependent variable can take only two possible values such as “Yes or No”, “Default or No Default”, “Living or Dead”, “Responder or Non Responder”, “Yes or No” etc.

LOGISTIC REGRESSION

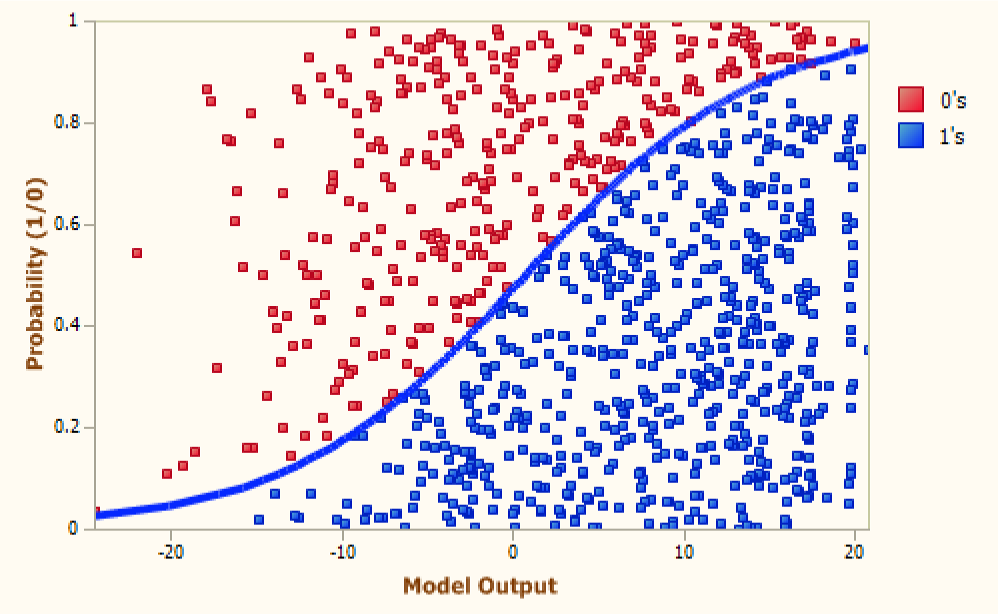
- The dependent variable **Y** must be categorical (1/0) and take binary value.
  - e.g.: if pass then Y = 1;else Y = 0
- We need a hypothesis function or we can call it a model.
- The output of this hypothesis should be between 0 and 1.
- We use sigmoid function as the hypothesis.
- Thus the goal of the Logistic Regression is to predict the likelihood that Y is equal to 1 (probability that Y = 1 rather than 0), given certain values of X.

WHY SIGMOID??



By Anil Kumar APSSDC

BINARY CLASSIFICATION



By Anil Kumar APSSDC

DECISION BOUNDARY

- We deal with thresholds here.
- Suppose if we have any value at 0.8 then we define threshold and if the value is greater than threshold that will be rounded off to 1,less than threshold rounded off to 0,in order to find the output
- If hypothesis >= 0.5 pred = 1 else pred = 0
- Here Theta's play an important role in creating the decision boundary
- Once we have the decision boundary set we will not need any data to predict the output
- We can build complex boundaries if we use higher order H-functions

RELATIONSHIP OF LOGISTIC TO LINEAR REGRESSION

Logistic Function

$$P(x) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$

Log Odds

$$\log\left[\frac{P(x)}{1 - P(x)}\right] = \beta_0 + \beta_1$$

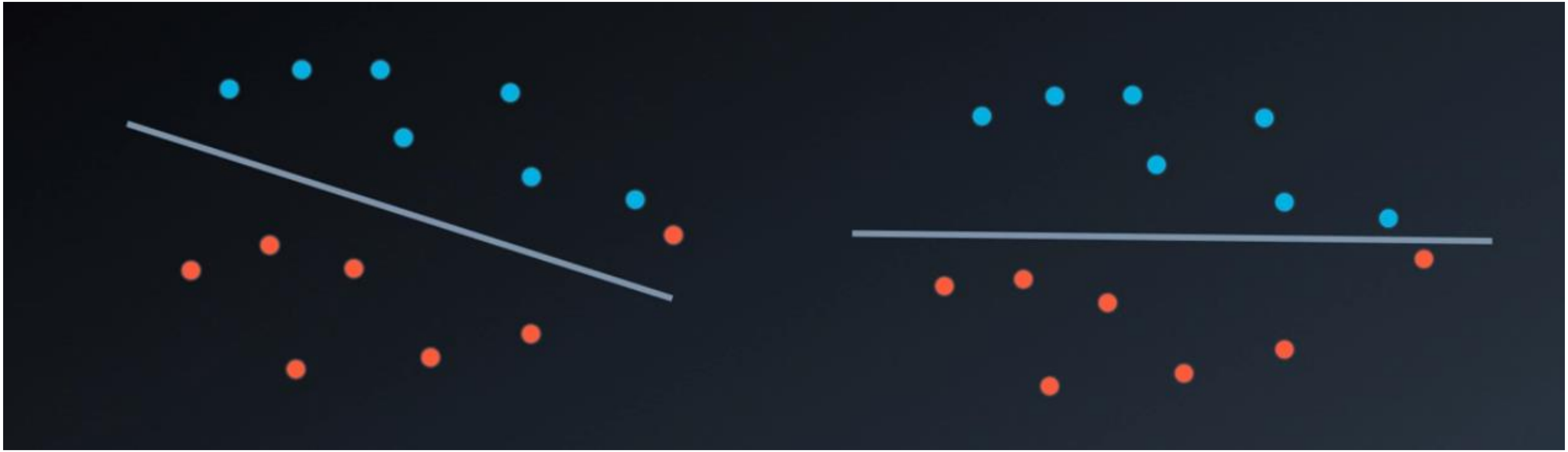
By Anil Kumar APSSDC

SUPPORT VECTOR MACHINES

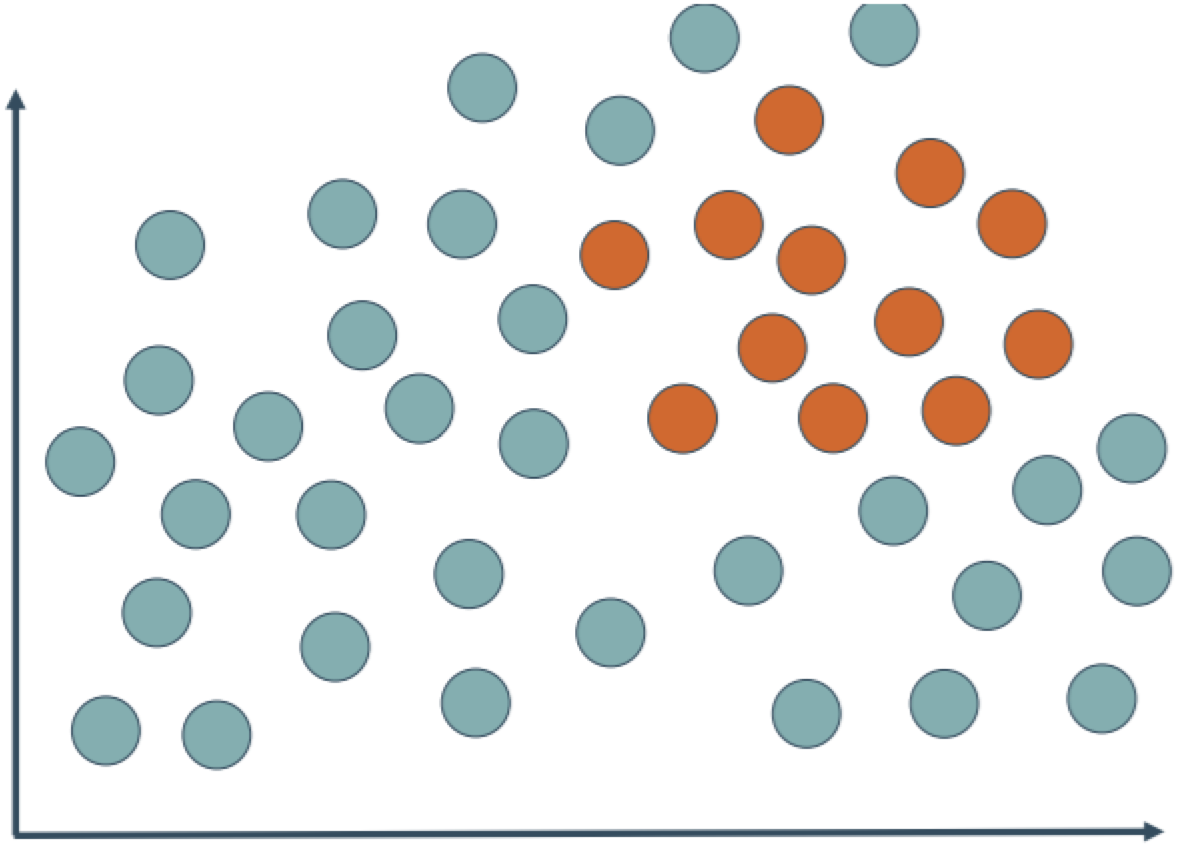
By Anil Kumar APSSDC

HOW WILL YOU CLASSIFY THIS

CLASSIFICATION IN SVM

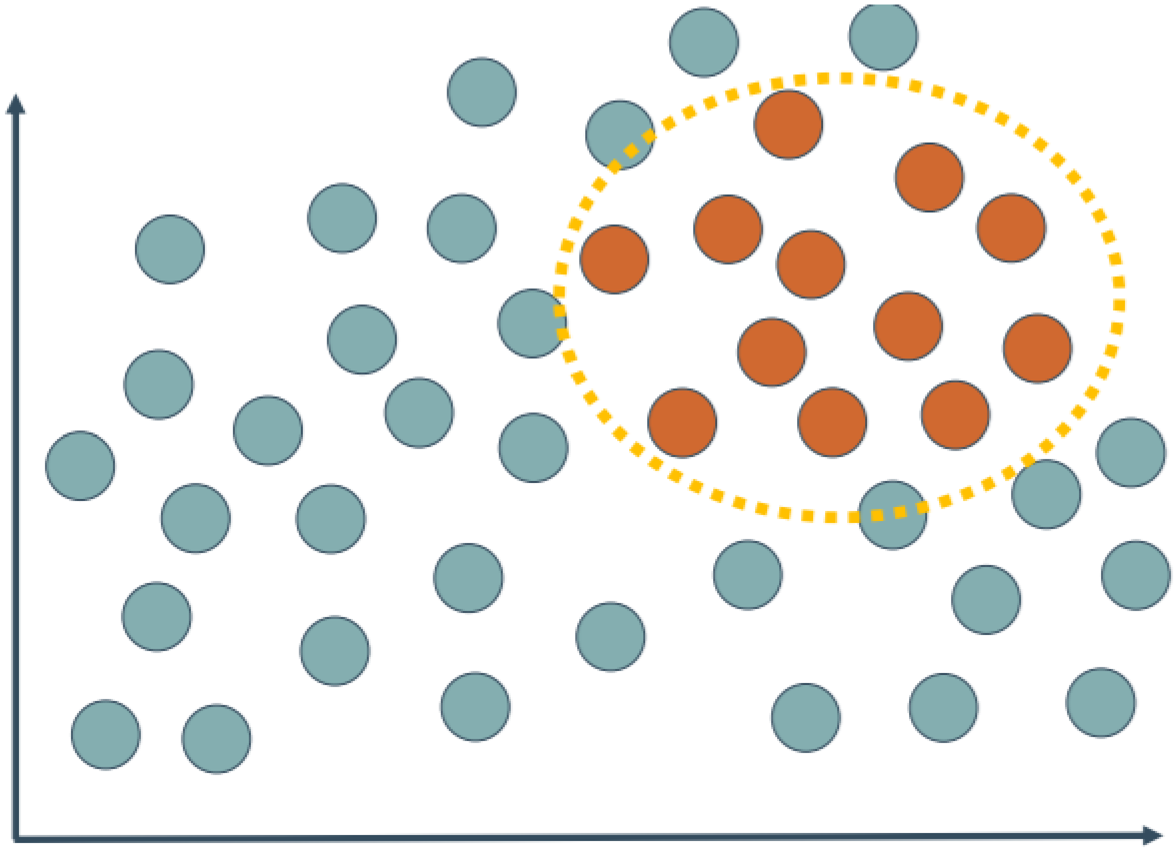


HOW WILL YOU CLASSIFY THIS



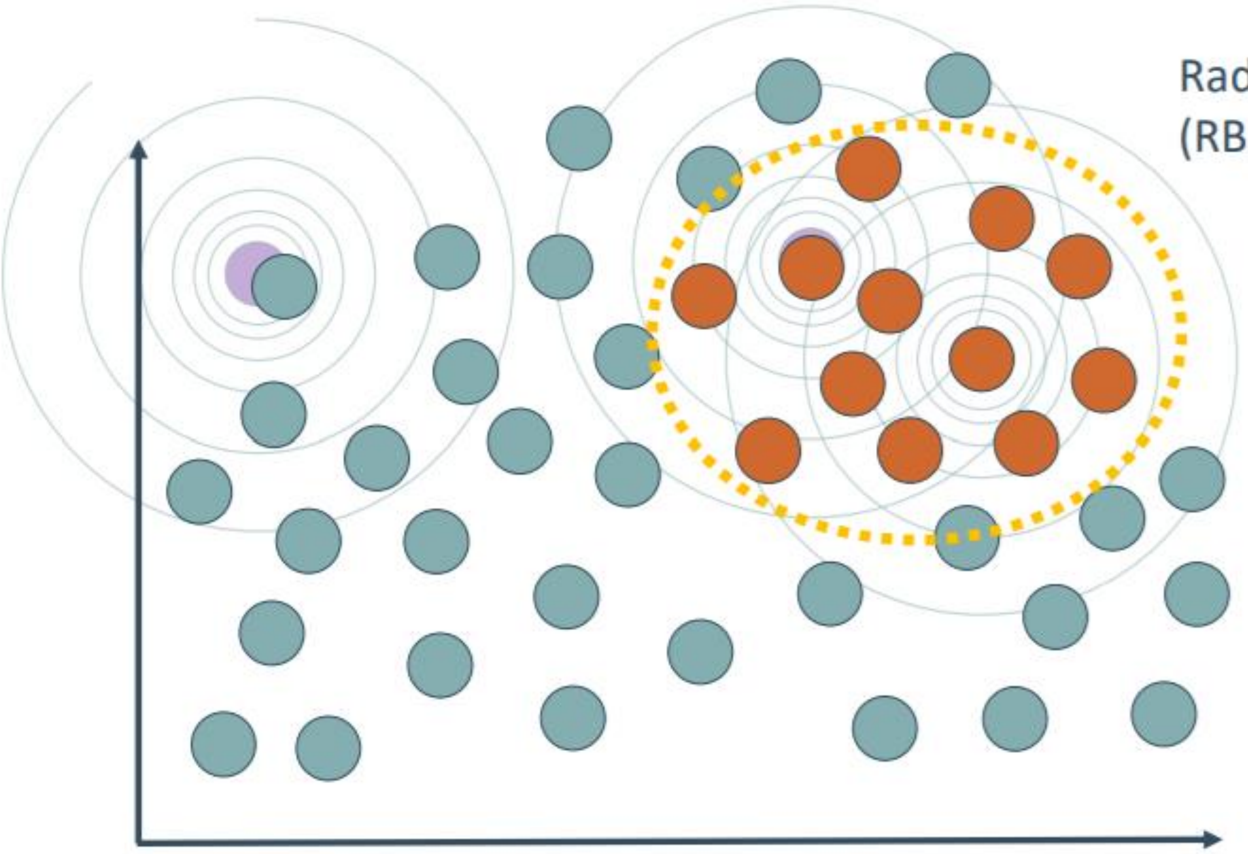
By Anil Kumar APSSDC

CLASSIFICATION IN SVC



By Anil Kumar APSSDC

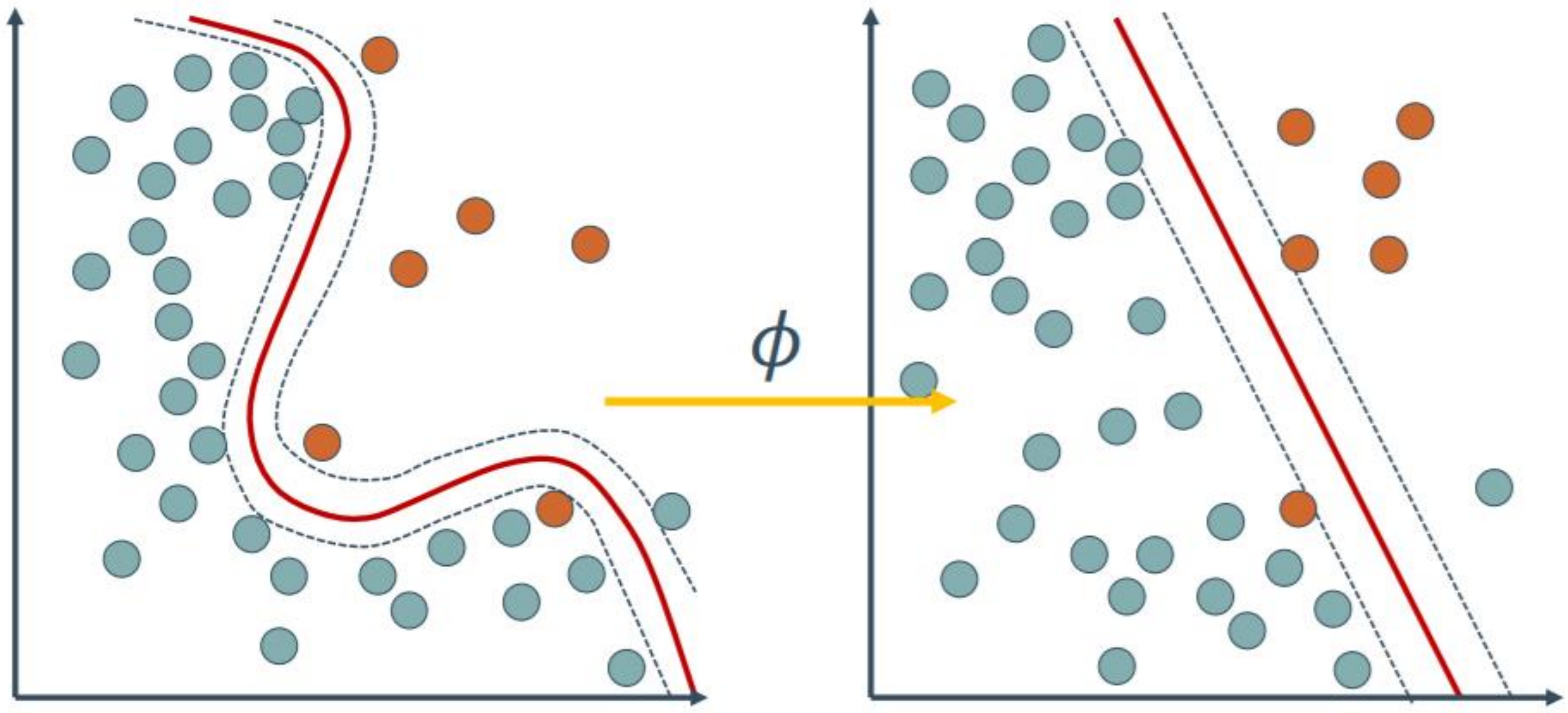
CLASSIFICATION USING SVC



Radial Basis Function (RBF) Kernel.

By Anil Kumar APSSDC

CLASSIFICATION USING SVC



By Anil Kumar APSSDC

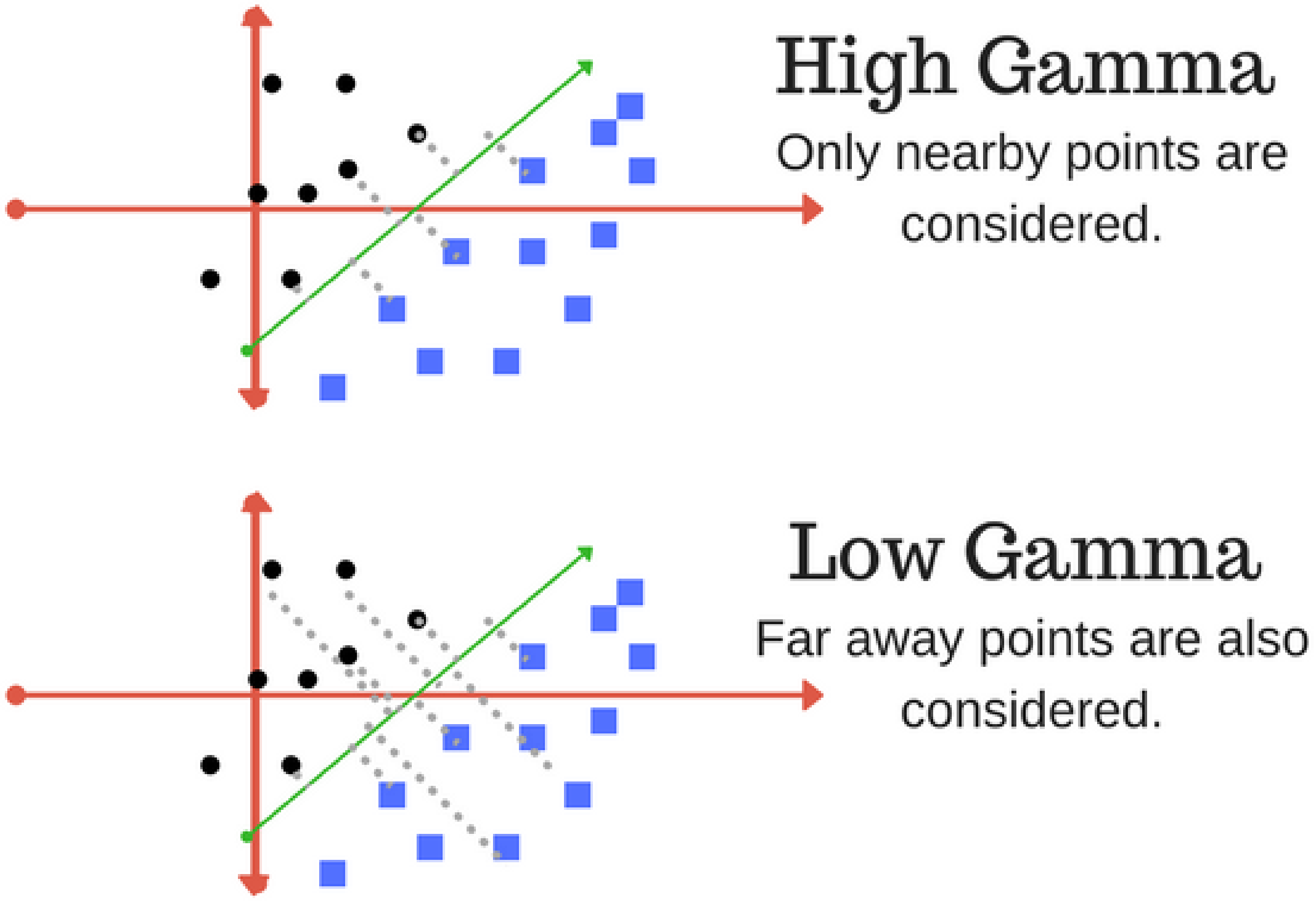
HYPERPARAMETERS

When we define the model, we can specify the hyperparameters. As we've seen in this section, the most common ones are

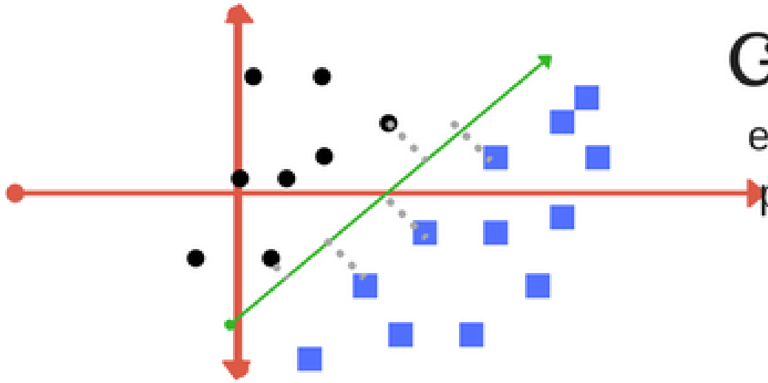
- **C**: The C parameter.
- **kernel**: The kernel. The most common ones are 'linear', 'poly', and 'rbf'.
- **degree**: If the kernel is polynomial, this is the maximum degree of the monomials in the kernel.
- **gamma** : If the kernel is rbf, this is the gamma parameter.

IMPACT OF GAMMA

The gamma parameter defines how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close’. In other words, wit low gamma, points far away from plausible separation line are considered in calculation for the separation line. Where as high gamma means the points close to plausible line are considered in calculation.

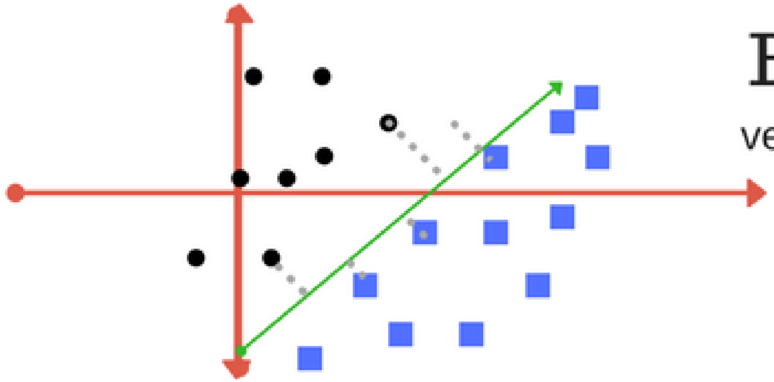


MARGIN



Good margin

equidistant as as far as possible for both side.



Bad margin

very close to blue class.

STEPS FOR APPLYING SVM ALGORITHM

1.Build a support vector machine model

- Create a support vector machine classification model using scikit-learn's SVC and assign it to the variablemodel.

2. Fit the model to the data

- If necessary, specify some of the **hyperparameters.** The goal is to obtain an accuracy of 100% in the dataset. Hint: Not every kernel will work well.

3. Predict using the model

- Predict the labels for the training set, and assign this list to the variable y_pred.

4. Calculate the accuracy of the model

- For this, use the function sklearn function accuracy_score.
- When you hit Test Run, you'll be able to see the boundary region of your model, which will help you tune the correct parameters, in case you need them.