

**MINI PROJECT SYNOPSIS**

**Group ID :** 27

**Project Title :** Leaf Classification using SVM

**Subject :** Machine Learning

**Project Guide :** Prof. Anand V. Kolapkar

**Technical Keywords:**

SVM (Support Vector Machine), SVC, Machine Learning, Classifier, Optimization, Prediction, Minuteness, Generalization Bias.

### 

### Problem Statement:

Apply the Support vector machine for classification on a data set obtained from UCI ML repository. For Example: Fruits Classification or Soil Classification or Leaf Disease Classification.

### Abstract:

A Support Vector Machine (SVM) is a discriminant classifier formally defined by a separating hyperplane. In other words, given labeled training data(supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples.

The leaf data set is taken from UCI ML repository. This data set consist in a collection of shape and texture features extracted from digital images of leaf specimens originating from total of 40 different plant species. Classifier model will be trained using this data set to classify the test data in one of the 36 predefined classes. The trained model will further help in accurate prediction of any leaf along with identification of diseases associated with it.

### Goals and Objectives :

* To load the whole data set into SVM with no errors.
* To train a model with highest possible accuracy using multiple values of Generalization Bias.
* To successfully predict the class of leaf from 36 possible outcomes.
* Maximize margin : we want to find a classifier whose decision boundary is furthest away from any data point.

### Mathematical Model :

* **Support Vector Machine (SVM) -**

The learning of the hyperplane in linear SVM is done by transforming the problem using some linear algebra. This is where the kernel plays role. For **linear kernel** the equation for prediction for a new input using the dot product between the input (x) and each support vector (xi) is calculated as follows:

f(x) = B(0) + sum(ai \* (x,xi))

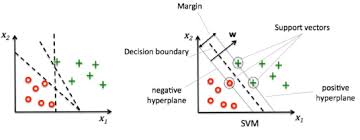
This is an equation that involves calculating the inner products of a new input vector (x) with all support vectors in training data. The coefficients B0 and ai (for each input) must be estimated from the training data by the learning algorithm.

The **polynomial kernel** can be written as

 K(x,xi) = 1 + sum(x \* xi)^d

and **exponential** as

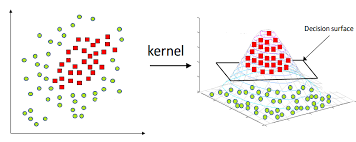
K(x,xi) = exp(-gamma \* sum((x — xi²)).



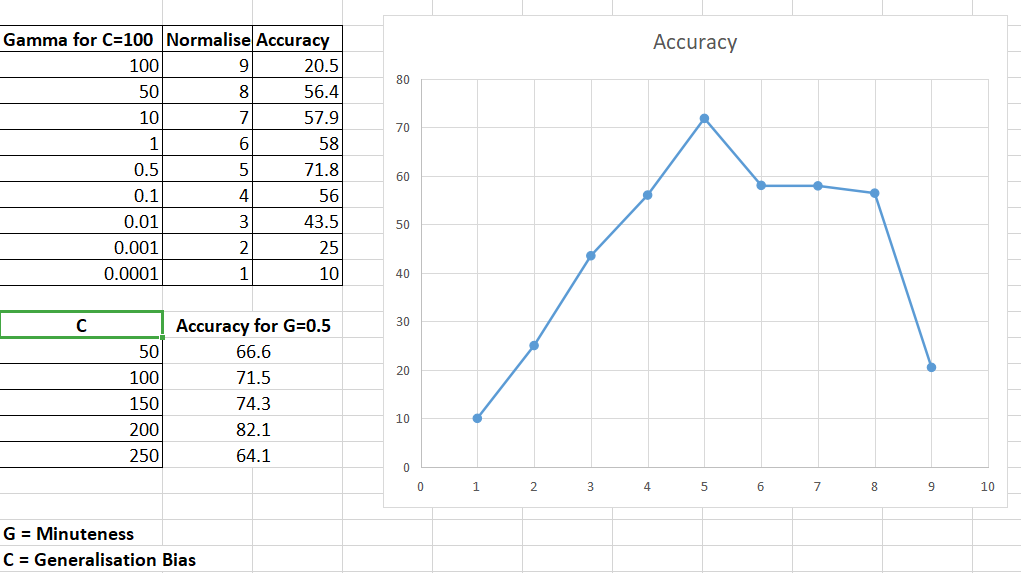
* **C and Gamma -**

C and Gamma are the parameters for a nonlinear support vector machine (SVM) with a Gaussian radial basis function kernel.

When we are using SVM, we are necessarily using one of the kernels: linear, polynomial or RBF=Radial Base Function (also called Gaussian Kernel).



* **Gamma Versus Accuracy -**

****