



Classifying data using Support Vector Machines(SVMs) in Python

[Read](#)[Discuss](#)[Courses](#)[Practice](#)

Introduction to SVMs: In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. A Support Vector Machine (SVM) is a discriminative 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.

What is Support Vector Machine?

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification, implicitly mapping their inputs into high-dimensional feature spaces.

What does SVM do?



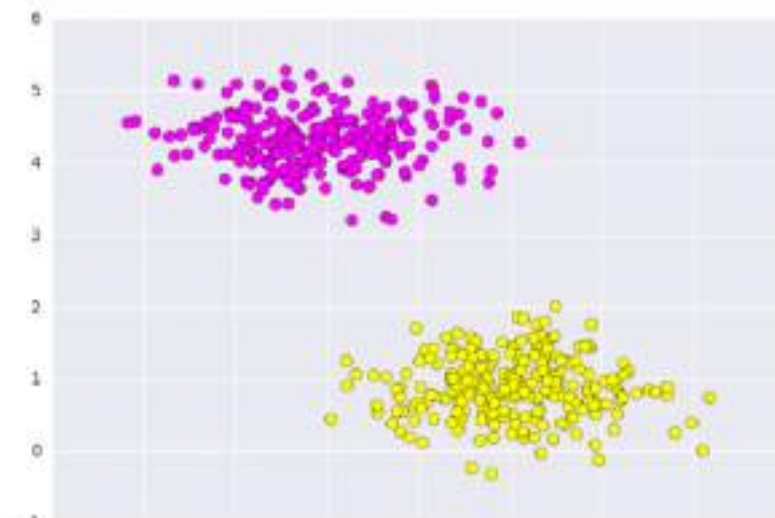
Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. Let you have basic understandings from this [article](#) before you proceed further. Here I'll discuss an example about SVM classification of [cancer](#) UCI datasets using machine learning tools i.e. scikit-learn compatible with Python. **Pre-requisites:** [Numpy](#), [Pandas](#), [matplotlib](#), [scikit-learn](#) Let's have a quick example of support vector classification. First we need to create a dataset:

python3

```
# importing scikit learn with make_blobs
from sklearn.datasets import make_blobs

# creating datasets X containing n_samples
# Y containing two classes
X, Y = make_blobs(n_samples=500, centers=2,
                  random_state=0, cluster_std=0.40)
import matplotlib.pyplot as plt
# plotting scatters
plt.scatter(X[:, 0], X[:, 1], c=Y, s=50, cmap='spring');
plt.show()
```

Output:



What Support vector machines do, is to not only draw a line between two classes here, but consider a region about the line of some given width. Here's an example of what it can look like:

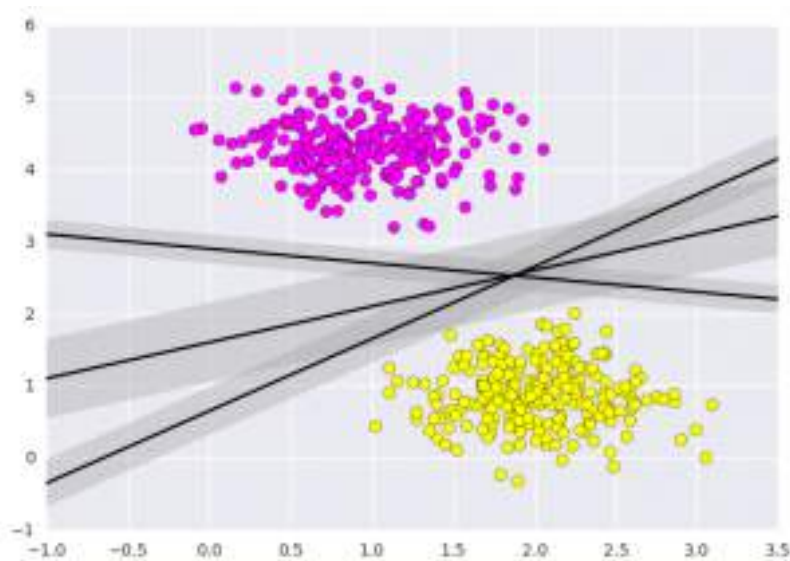
python3

```
# creating linspace between -1 to 3.5
xfit = np.linspace(-1, 3.5)

# plotting scatter
plt.scatter(X[:, 0], X[:, 1], c=Y, s=50, cmap='spring')

# plot a line between the different sets of data
for m, b, d in [(1, 0.65, 0.33), (0.5, 1.6, 0.55), (-0.2, 2.9, 0.2)]:
    yfit = m * xfit + b
    plt.plot(xfit, yfit, '-k')
    plt.fill_between(xfit, yfit - d, yfit + d, edgecolor='none',
                    color='#AAAAAA', alpha=0.4)

plt.xlim(-1, 3.5);
plt.show()
```





Importing datasets

This is the intuition of support vector machines, which optimize a linear discriminant model representing the perpendicular distance between the datasets. Now let's train the classifier using our training data. Before training, we need to import cancer datasets as csv file where we will train two features out of all features.

python3

```
# importing required libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# reading csv file and extracting class column to y.
x = pd.read_csv("C:\\...\\cancer.csv")
a = np.array(x)
y = a[:,30] # classes having 0 and 1

# extracting two features
x = np.column_stack((x.malignant,x.benign))

# 569 samples and 2 features
x.shape

print (x),(y)
```

```
[ [ 122.8  1001. ]
  [ 132.9  1326. ]
  [ 130.   1203. ]
  ...,
  [ 108.3   858.1 ]
  [ 140.1  1265. ]
  [  47.92  181.  ]]

array([ 0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,
        0.,  0.,  0.,  0.,  0.,  0.,  1.,  1.,  1.,  0.,  0.,  0.,  0.,
        0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  1.,  0.,
        0.,  0.,  0.,  0.,  0.,  0.,  0.,  1.,  0.,  1.,  1.,  1.,  1.,
        1.,  0.,  0.,  1.,  0.,  0.,  1.,  1.,  1.,  1.,  0.,  1.,
        ...,
        1.])
```

Fitting a Support Vector Machine

Now we'll fit a Support Vector Machine Classifier to these points. While the mathematical details of the likelihood model are interesting, we'll let read about those elsewhere. Instead, we'll just treat the scikit-learn algorithm as a black box which accomplishes the above task.

python3

```
# import support vector classifier
# "Support Vector Classifier"
from sklearn.svm import SVC
clf = SVC(kernel='linear')

# fitting x samples and y classes
clf.fit(x, y)
```

After being fitted, the model can then be used to predict new values:

python3

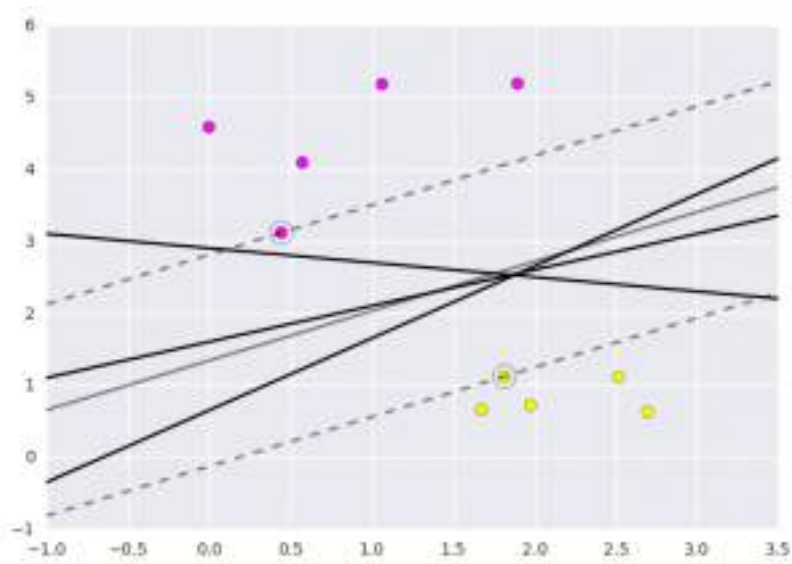
We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).

```
clf.predict([[85, 550]])
```

```
array([ 0.])
```

```
array([ 1.])
```

Let's have a look on the graph how does this show.





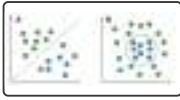







This is obtained by analyzing the data taken and pre-processing methods to make optimal hyperplanes using matplotlib func If you like GeeksforGeeks and would like to contribute, you can also write an article using [write.geeksforgeeks.org](https://www.geeksforgeeks.org/write-a-blog/) or mail your article to review-team@geeksforgeeks.org. See your article appearing on the GeeksforGeeks main page and help other Geeks. Please write comments if you find anything incorrect, or you want to share more information about the topic discussed above.

Whether you're preparing for your first job interview or aiming to upskill in this ever-evolving tech landscape, [GeeksforGeeks Courses](https://www.geeksforgeeks.org/courses/) are your key to success. We provide top-quality content at affordable prices, all geared towards accelerating your growth in a time-bound manner. Join the millions we've already empowered, and we're here to do the same for you. Don't miss out -






Last Updated : 01 Sep, 2023

28

Similar Reads

 Classifying data using Support Vector Machines(SVMs) in R	 Predicting Stock Price Direction using Support Vector Machines
 Introduction to Support Vector Machines (SVM)	 Hinge-loss & relationship with Support Vector Machines
 ML Classifying Data using an Auto-encoder	 Image classification using Support Vector Machine (SVM) in Python
 Classifying Clothing Images in Python	 Support Vector Regression (SVR) using Linear and Non-Linear Kernels in Scikit Learn
 Support vector machine in Machine Learning	 Train a Support Vector Machine to recognize facial features in C++

Related Tutorials

 OpenAI Python API - Complete Guide	 Computer Vision Tutorial
 Computer Science and Programming For Kids	 Pandas AI: The Generative AI Python Library
 Top Computer Vision Projects (2023)	

[Previous](#)[Next](#)[Support Vector Machine \(SVM\) Algorithm](#)[Support Vector Regression \(SVR\) using Linear and Non-Linear Kernels in Scikit](#)[Learn](#)

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).

Article Contributed By :

A Afzal Ansari

Vote for difficulty

Current difficulty : [Easy](#)

Easy

Normal

Medium

Hard

Expert

Improved By : [skull tone](#), [rkbhola5](#), [gwb67hdeem4u87y6oocpvhdxccr2dmlta3eqedqp](#)Article Tags : [Computer Subject](#) , [Machine Learning](#) , [Python](#)Practice Tags : [Machine Learning](#), [python](#)[Improve Article](#)[Report Issue](#)

A-143, 9th Floor, Sovereign Corporate
Tower, Sector-136, Noida, Uttar Pradesh -
201305

feedback@geeksforgeeks.org

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).



Company

About Us
Legal
Terms & Conditions
Careers
In Media
Contact Us
Advertise with us
GFG Corporate Solution
Placement Training Program
Apply for Mentor

Languages

Python
Java
C++
PHP
GoLang
SQL
R Language
Android Tutorial

DSA Roadmaps

DSA for Beginners
Basic DSA Coding Problems
DSA Roadmap by Sandeep Jain
DSA with JavaScript
Top 100 DSA Interview Problems
All Cheat Sheets

Explore

Job-A-Thon Hiring Challenge
Hack-A-Thon
GfG Weekly Contest
Offline Classes (Delhi/NCR)
DSA in JAVA/C++
Master System Design
Master CP
GeeksforGeeks Videos

DSA Concepts

Data Structures
Arrays
Strings
Linked List
Algorithms
Searching
Sorting
Mathematical
Dynamic Programming

Web Development

HTML
CSS
JavaScript
Bootstrap
ReactJS
AngularJS

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).

[Express.js](#)[Lodash](#)

Computer Science

[GATE CS Notes](#)[Operating Systems](#)[Computer Network](#)[Database Management System](#)[Software Engineering](#)[Digital Logic Design](#)[Engineering Maths](#)

Data Science & ML

[Data Science With Python](#)[Data Science For Beginner](#)[Machine Learning Tutorial](#)[Maths For Machine Learning](#)[Pandas Tutorial](#)[NumPy Tutorial](#)[NLP Tutorial](#)[Deep Learning Tutorial](#)

Python

[Python Programming Examples](#)[Django Tutorial](#)[Python Projects](#)[Python Tkinter](#)[OpenCV Python Tutorial](#)[Python Interview Question](#)

DevOps

[Git](#)[AWS](#)[Docker](#)[Kubernetes](#)[Azure](#)[GCP](#)

Competitive Programming

[Top DSA for CP](#)[Top 50 Tree Problems](#)[Top 50 Graph Problems](#)[Top 50 Array Problems](#)[Top 50 String Problems](#)[Top 50 DP Problems](#)[Top 15 Websites for CP](#)

System Design

[What is System Design](#)[Monolithic and Distributed SD](#)[Scalability in SD](#)[Databases in SD](#)[High Level Design or HLD](#)[Low Level Design or LLD](#)[Crack System Design Round](#)[System Design Interview Questions](#)

Interview Corner

GfG School

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).

[Experienced Interviews](#)[CBSE Notes for Class 10](#)[Internship Interviews](#)[CBSE Notes for Class 11](#)[Competitive Programming](#)[CBSE Notes for Class 12](#)[Aptitude Preparation](#)[English Grammar](#)

Commerce

UPSC

[Accountancy](#)[Polity Notes](#)[Business Studies](#)[Geography Notes](#)[Economics](#)[History Notes](#)[Human Resource Management \(HRM\)](#)[Science and Technology Notes](#)[Management](#)[Economics Notes](#)[Income Tax](#)[Important Topics in Ethics](#)[Finance](#)[UPSC Previous Year Papers](#)[Statistics for Economics](#)

SSC/ BANKING

Write & Earn

[SSC CGL Syllabus](#)[Write an Article](#)[SBI PO Syllabus](#)[Improve an Article](#)[SBI Clerk Syllabus](#)[Pick Topics to Write](#)[IBPS PO Syllabus](#)[Share your Experiences](#)[IBPS Clerk Syllabus](#)[Internships](#)[Aptitude Questions](#)[SSC CGL Practice Papers](#)

@GeeksforGeeks, Sanchhaya Education Private Limited, All rights reserved