

# SVM

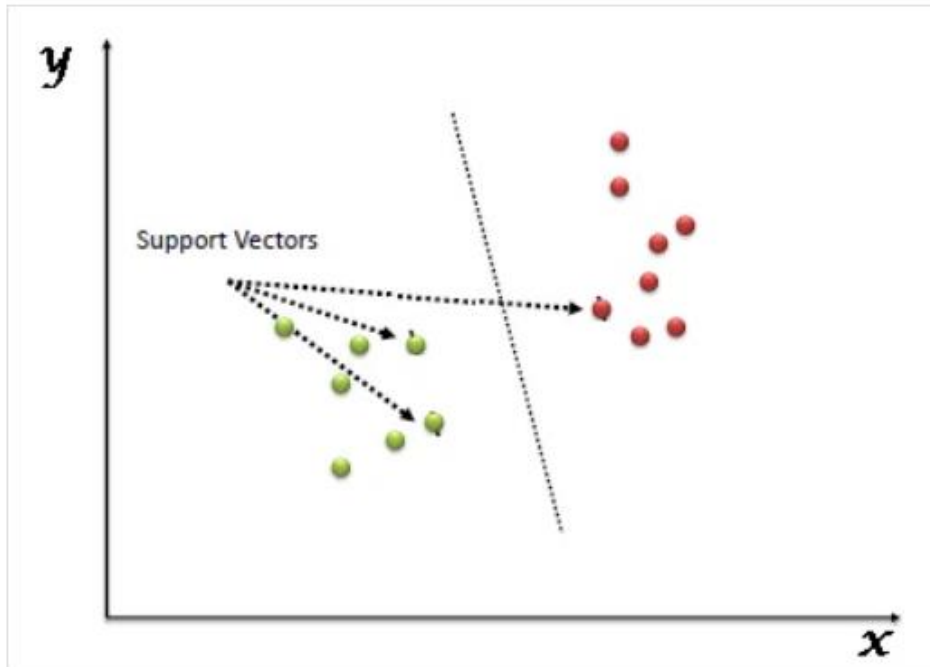
---

# فهرست مطالب

- ماشین بردار پشتیبان چیست؟
- الگوریتم ماشین بردار پشتیبان چگونه کار می کند؟
- پیاده سازی بردار پشتیبان در زبان پایتون
- مزایا و معایب
- نتیجه گیری

# ماشین بردار پشتیبان چیست؟

- Supervise learning
- Used for classification and regression (mostly used for classification)
- In SVM we represent any data according to its values into n-dimensional space



# الگوریتم ماشین بردار پشتیبان چگونه کار می کند؟

• سوال: چگونه میتوان یک خط راست مناسب تعیین نمود؟

• سناریو اول

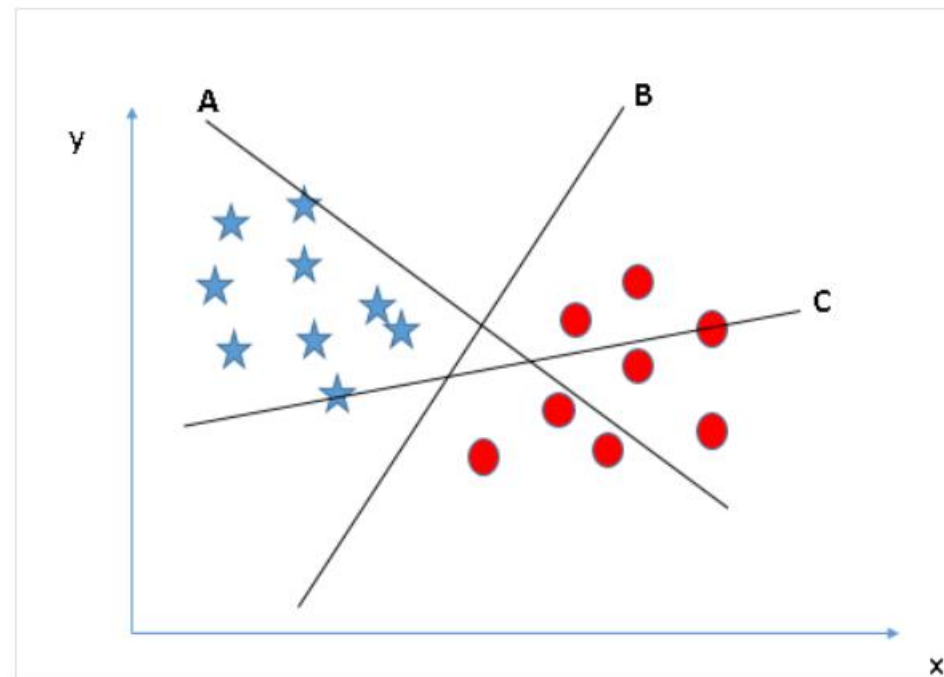
• سناریو دوم

• سناریو سوم

• سناریو چهارم

• سناریو پنجم

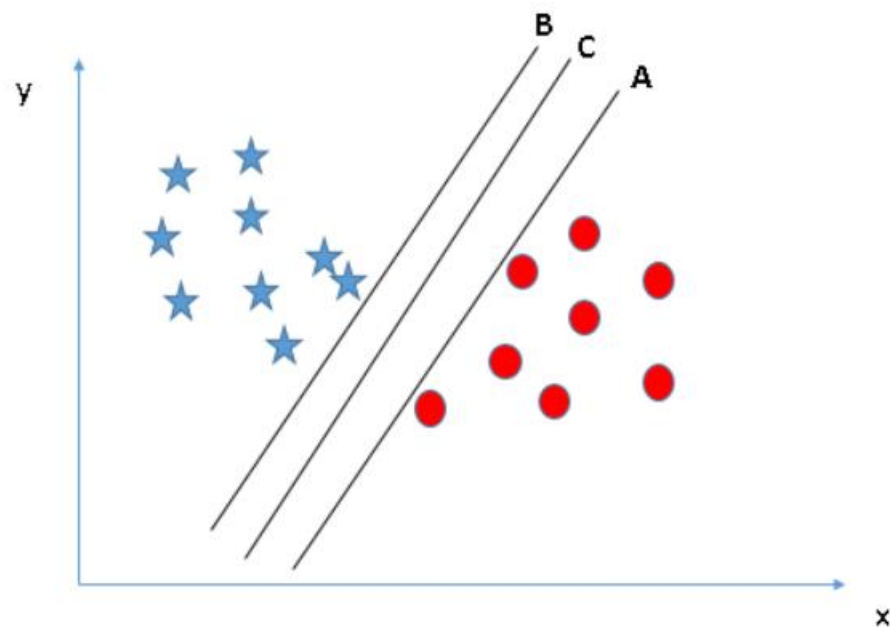
# یک



«خط راستی که دو دسته را به طور بهتری از یکدیگر جدا می‌کند، خطی است که باید انتخاب شود.»

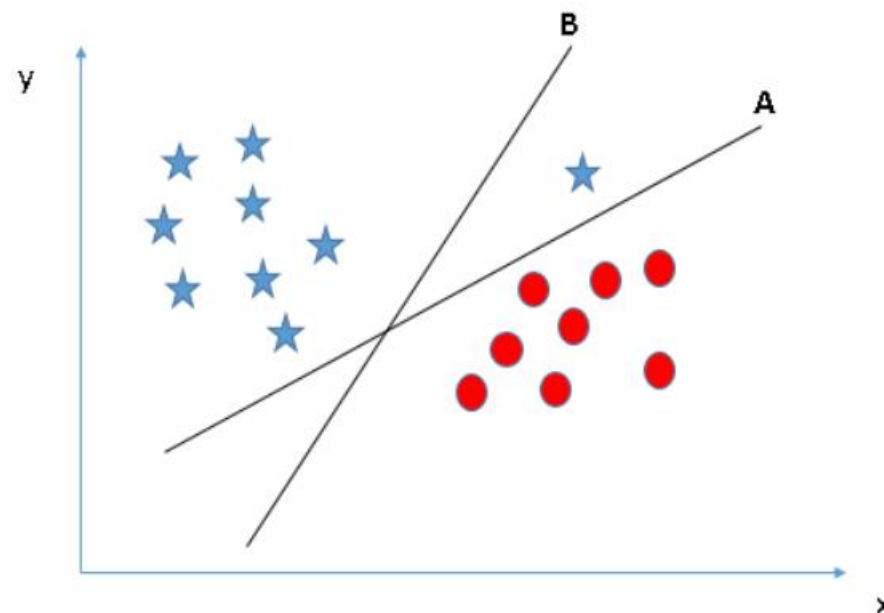
## دو

محاسبه فاصله نزدیک‌ترین نقطه داده (که از هر دسته‌ای می‌تواند باشد) از خط راست می‌تواند به انتخاب خط راست صحیح کمک کند. به این فاصله حاشیه گفته می‌شود.

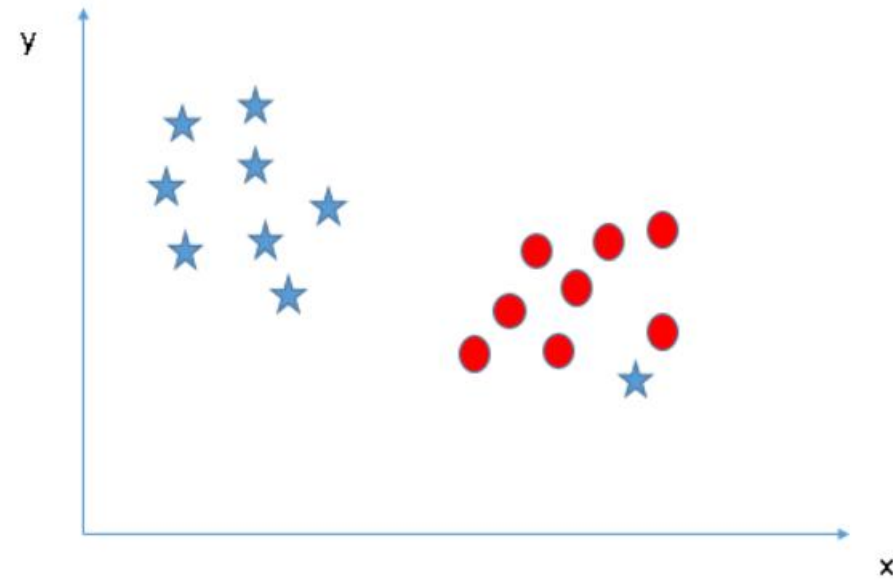


## سه

در الگوریتم ماشین بردار پشتیبان، خط راستی که دسته‌ها را به درستی تقسیم کند (صحت) بر خطی که حاشیه بیشتری دارد، دارای اولویت است.



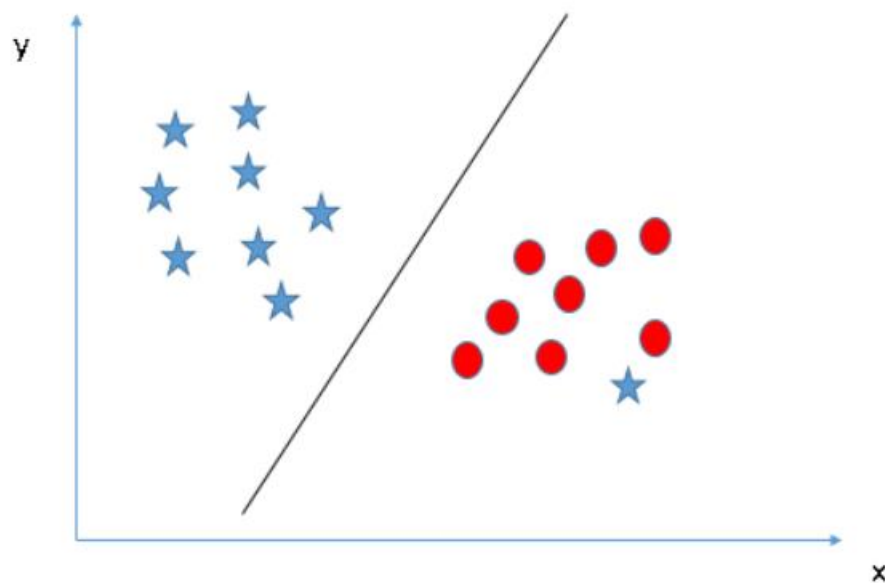
# چهار



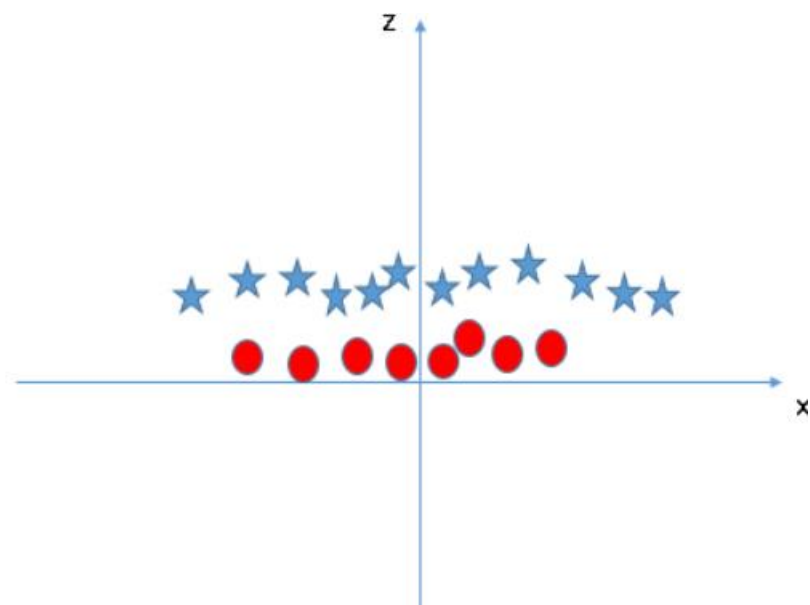
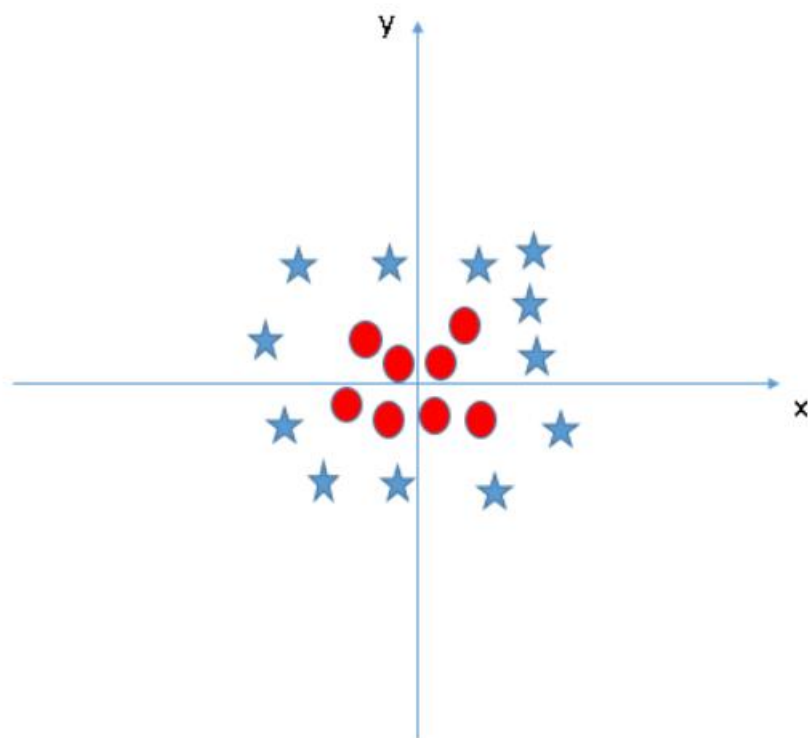


وجود یک ستاره آبی در قلمرو دسته دیگر به عنوان یک دورافتادگی برای دسته ستاره آبی محسوب می‌شود.

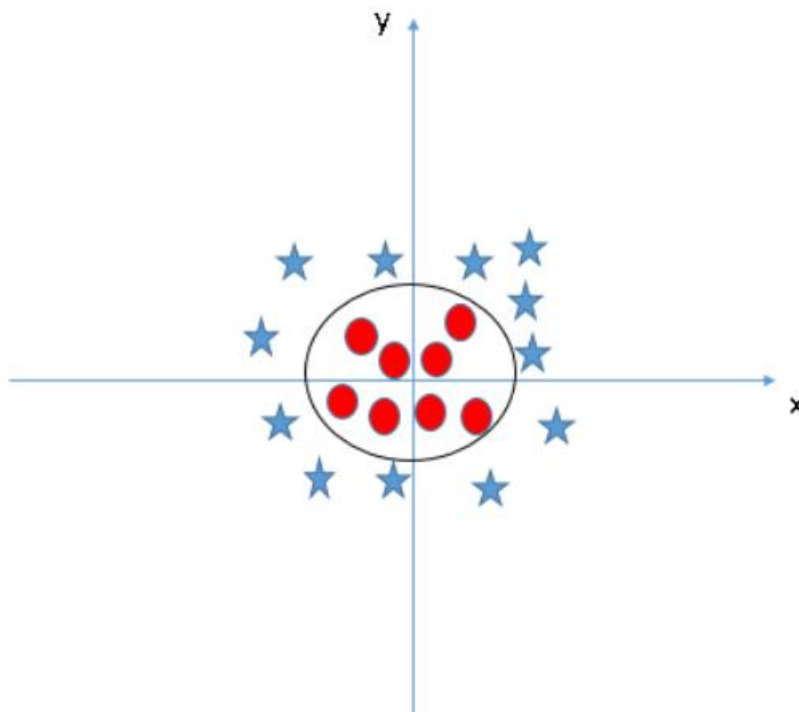
یکی از ویژگی‌های ماشین بردار پشتیبان آن است که دورافتادگی‌ها را **نادیده** گرفته و تنها خط راستی را که بیشترین حاشیه را با نقاط داده دسته‌ها دارد انتخاب می‌کند.



## پنج ( خط غیر راست)



- ماشین بردار پشتیبان از روشی که به آن ترفند هسته (کرنل) گفته می‌شود، استفاده می‌کند.
- در این روش در واقع توابعی وجود دارند که فضای ورودی بُعد پایین را دریافت کرده و آن را به فضای بُعد بالاتر تبدیل می‌کنند. این تبدیل، یک مسئله غیر قابل جداسازی را به مسئله قابل جداسازی مبدل می‌کند. به این توابع، تابع‌های هسته (کرنل) گفته می‌شود.



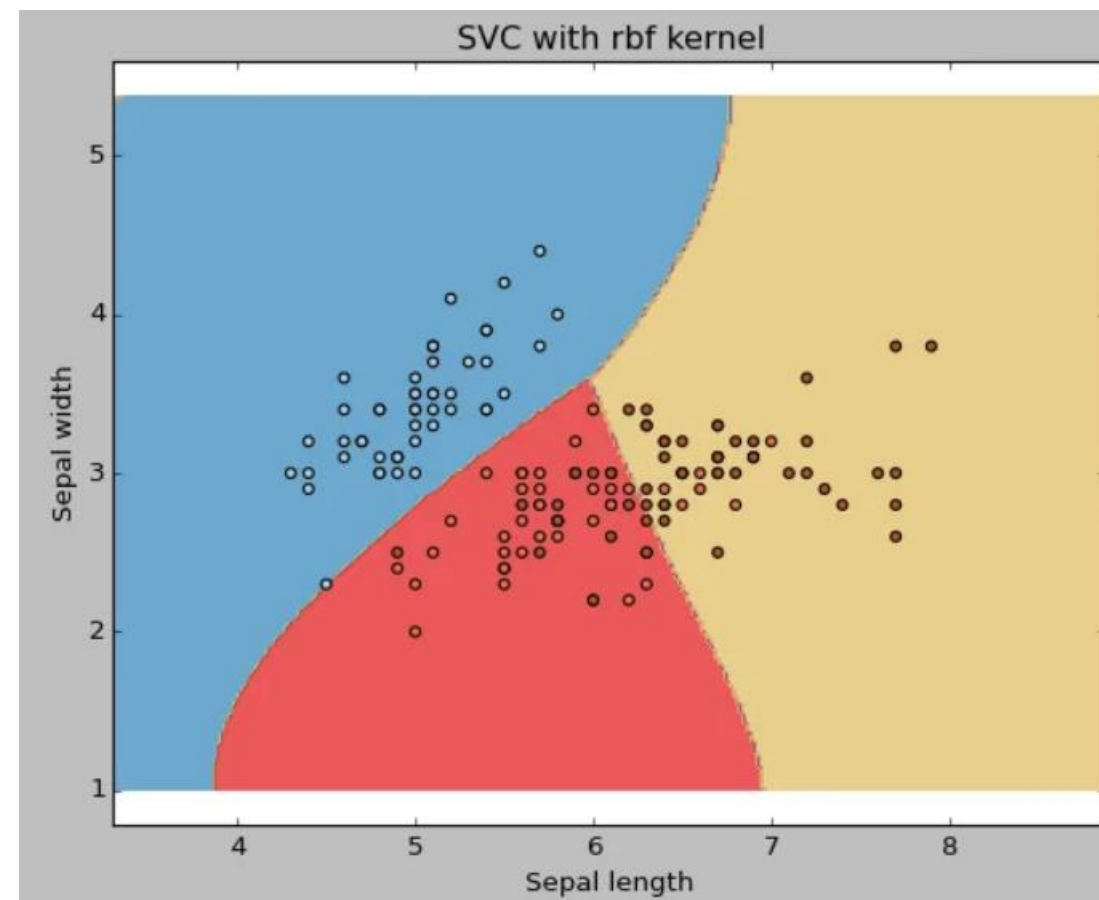
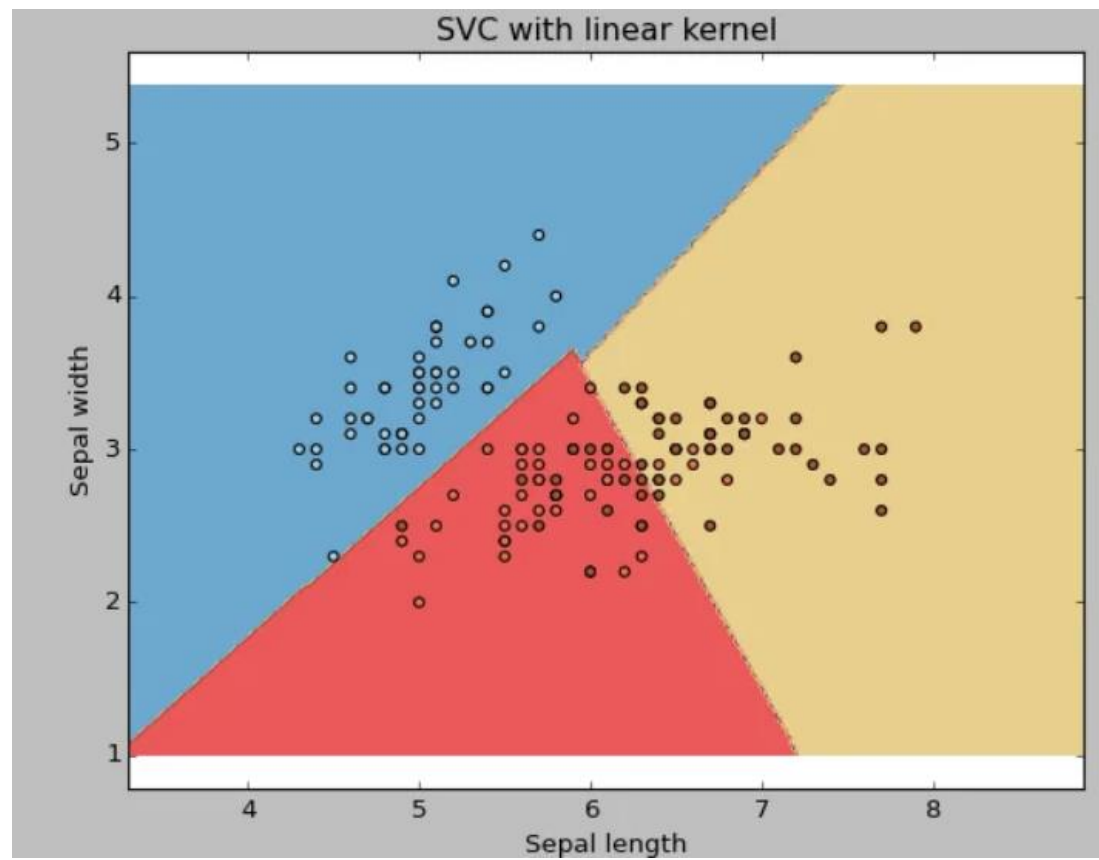
# پیاده سازی در پایتون

## پارامترهای الگوریتم ماشین بردار پشتیبان چگونه تنظیم می‌شوند؟

```
sklearn.svm.SVC(C=1.0, kernel='rbf', degree=3, gamma=0.0, coef0=0.0, shrinking=True,  
                probability=False, tol=0.001, cache_size=200,  
                class_weight=None, verbose=False, max_iter=-1, random_state=None)
```

## کرنل:

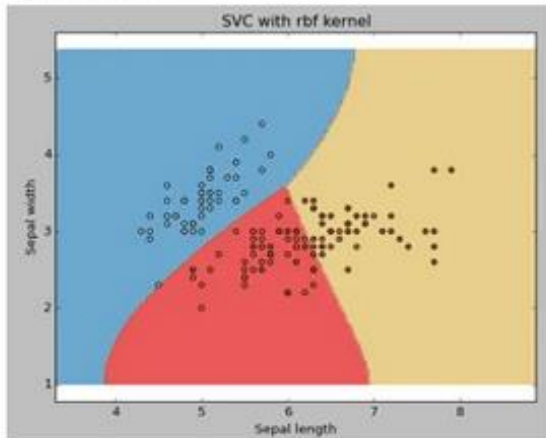
این پارامتر پیش از این مورد بررسی قرار گرفت. گزینه‌های گوناگونی شامل «linear»، «rbf»، «poly» برای کرنل وجود دارند و در حالت پیش فرض کرنل روی پارامتر rbf قرار دارد. rbf و poly برای خط جداساز غیر راست مفید هستند. در مثال زیر، از یک کرنل خطی linear بر روی دو ویژگی از مجموعه داده معروف iris برای طبقه‌بندی گونه‌های آن استفاده شده است.



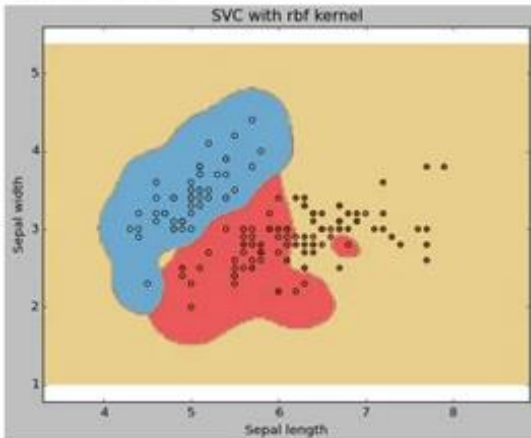
# گاما :

ضریب کرنل برای rbf، poly و sigmoid است. هرچه مقدار گاما بیشتر باشد، الگوریتم تلاش می‌کند برازش را دقیقاً بر اساس مجموعه داده‌های تمرینی انجام دهد و این امر موجب تعمیم یافتن خطا و وقوع مشکل بیش برازش (Over-Fitting) می‌شود.

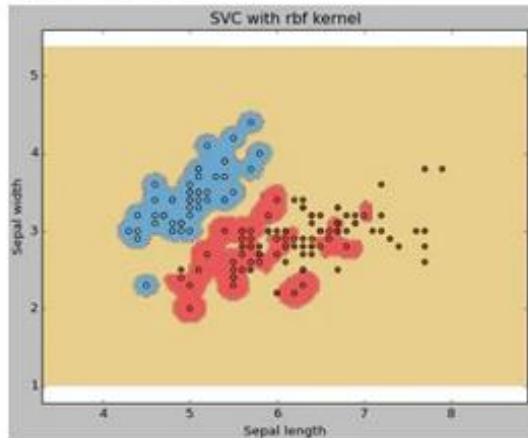
gamma = 0



gamma = 10



gamma = 100

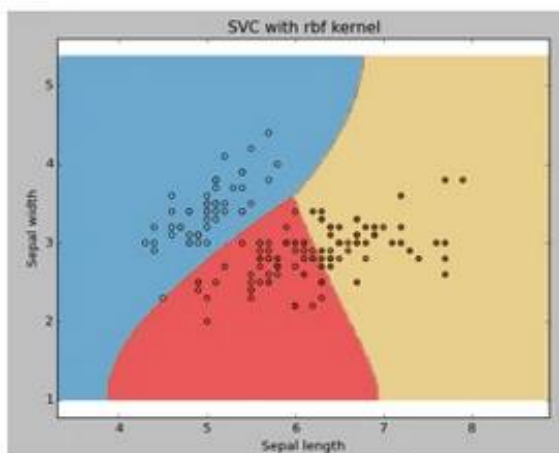




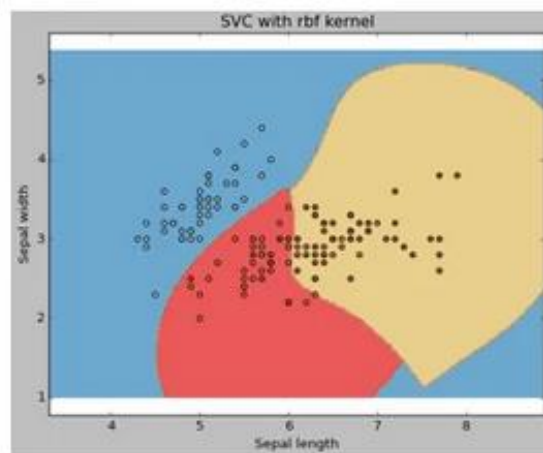
C:

پارامتر جریمه  $C$ ، برای جمله خطا است. این پارامتر همچنین برقراری تعادل بین مرزهای تصمیم‌گیری هموار و طبقه‌بندی نقاط داده تمرینی را کنترل می‌کند.

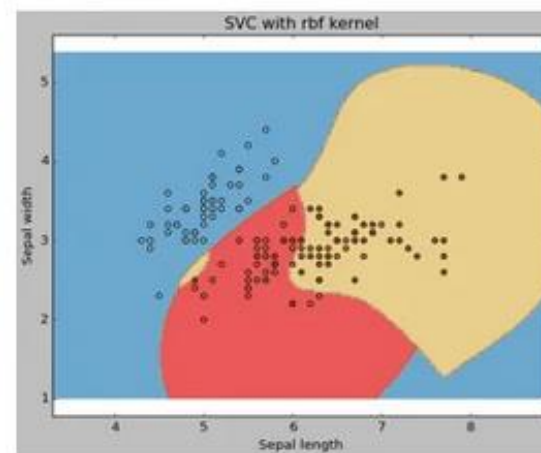
$c=1$



$C=100$



$c=1000$



# مزایا و معایب الگوریتم ماشین بردار پشتیبان

## • مزایا

- حاشیه جداسازی برای دسته‌های مختلف کاملاً واضح است.
- در فضاهای با ابعاد بالاتر کارایی بیشتری دارد.
- در شرایطی که تعداد ابعاد بیش از تعداد نمونه‌ها باشد نیز کار می‌کند.
- یک زیر مجموعه از نقاط تمرینی را در تابع تصمیم‌گیری استفاده می‌کند (که به آن‌ها بردارهای پشتیبان گفته می‌شود)، بنابراین در مصرف حافظه نیز به صورت بهینه عمل می‌کند.

## • معایب

- هنگامی که مجموعه داده‌ها بسیار بزرگ باشد، عملکرد خوبی ندارد، زیرا نیازمند زمان آموزش بسیار زیاد است.
- هنگامی که مجموعه داده نوفه (نویز) زیادی داشته باشد، عملکرد خوبی ندارد و کلاس‌های هدف دچار همپوشانی می‌شوند.
- ماشین بردار پشتیبان به طور مستقیم تخمین‌های احتمالاتی را فراهم نمی‌کند و این موارد با استفاده از یک اعتبارسنجی متقابل (Cross Validation) پرهزینه پنج‌گانه انجام می‌شوند. این امر با روش SVC موجود در کتابخانه scikit-learn پایتون، مرتبط است.

پایان