

- ۱

قسمت اول:

(a)

جواب بصورت زوج مرتب برگردانده شده که درایه دوم زوج مرتب ها نشاندهنده ی دفعات تکرار آن Item Set است

(['A'], 6)

(['B'], 8)

(['C'], 4)

(['D'], 5)

(['E'], 9)

(b)

(['A', 'B'], 4)

(['A', 'C'], 4)

(['A', 'E'], 5)

(['B', 'E'], 7)

(['C', 'E'], 4)

(['D', 'E'], 4)

(c)

همانطور که از قسمت قبل نیز پیداست فقط زوج ['B', 'E'] این خاصیت را دارد.

(d)

$C \implies A$

$C \implies E$

$C \implies A, E$

$A, C \implies E$

$C, E \implies A$

Total Rules Generated: 5

(e)

$B \implies E : 0.875$

قسمت دوم:

(a)

(['A'], 2)

(['B'], 3)

(['C'], 3)

(['E'], 3)

(['A', 'C'], 2)

(['B', 'C'], 2)

(['B', 'E'], 3)

(['C', 'E'], 2)

(['B', 'C', 'E'], 2)

(b)

A ==> C : 1.0

C ==> A : 0.6666666666666666

B ==> C : 0.6666666666666666

C ==> B : 0.6666666666666666

B ==> E : 1.0

E ==> B : 1.0

C ==> E : 0.6666666666666666

E ==> C : 0.6666666666666666

B ==> C,E : 0.6666666666666666

C ==> B,E : 0.6666666666666666

E ==> B,C : 0.6666666666666666

B,C ==> E : 1.0

B,E ==> C : 0.6666666666666666

C,E ==> B : 1.0

Total Rules Generated: 14

(c)

$A \implies C : 1.0$

$B \implies E : 1.0$

$E \implies B : 1.0$

$B, C \implies E : 1.0$

$C, E \implies B : 1.0$

Total Rules Generated: 5

(d)

همانطور که از قسمت b پیداست

$E \implies C : 0.6666666666666666$

(e)

بعبارتی مقدار support را برای ['B', 'C'] کافیت حساب کنیم که طبق قسمت a برابر ۲ میباشد.

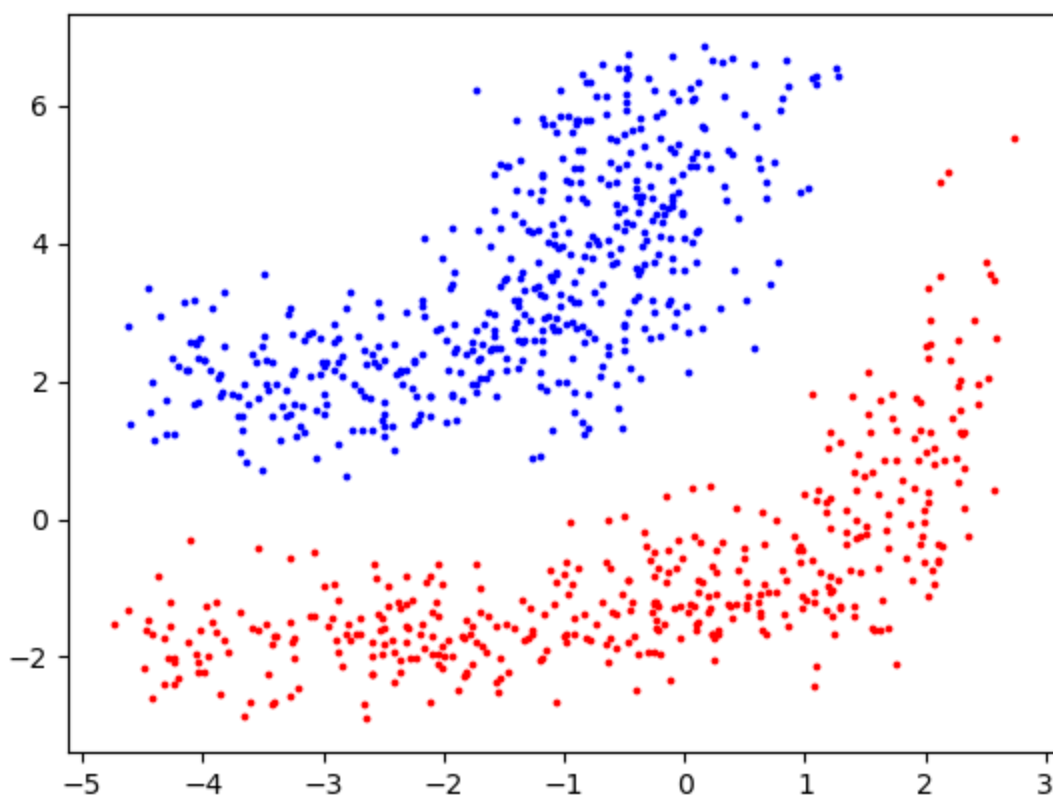
-۲

با توجه به این که برای split کردن dataset اصلی از تابعی استفاده شده که هر بار خروجی مختلفی میدهد (مثلا ۰,۲ از dataset اصلی را بصورت تصادفی اسپلیت می کند در صورت نیاز به مجموعه اعتبارسنجی همین کار را یک بار دیگر هم انجام میدهد که در کد قابل مشاهده است) تمامی قسمت ها پشت سر هم و در یک فایل پیاده سازی شده اند. ولی با توجه به این که برای تحویل سؤالات هر بخش نیاز به فایلی مجزا داشته همان تک فایل مورد استفاده در تمامی بخش ها کپی شده است.

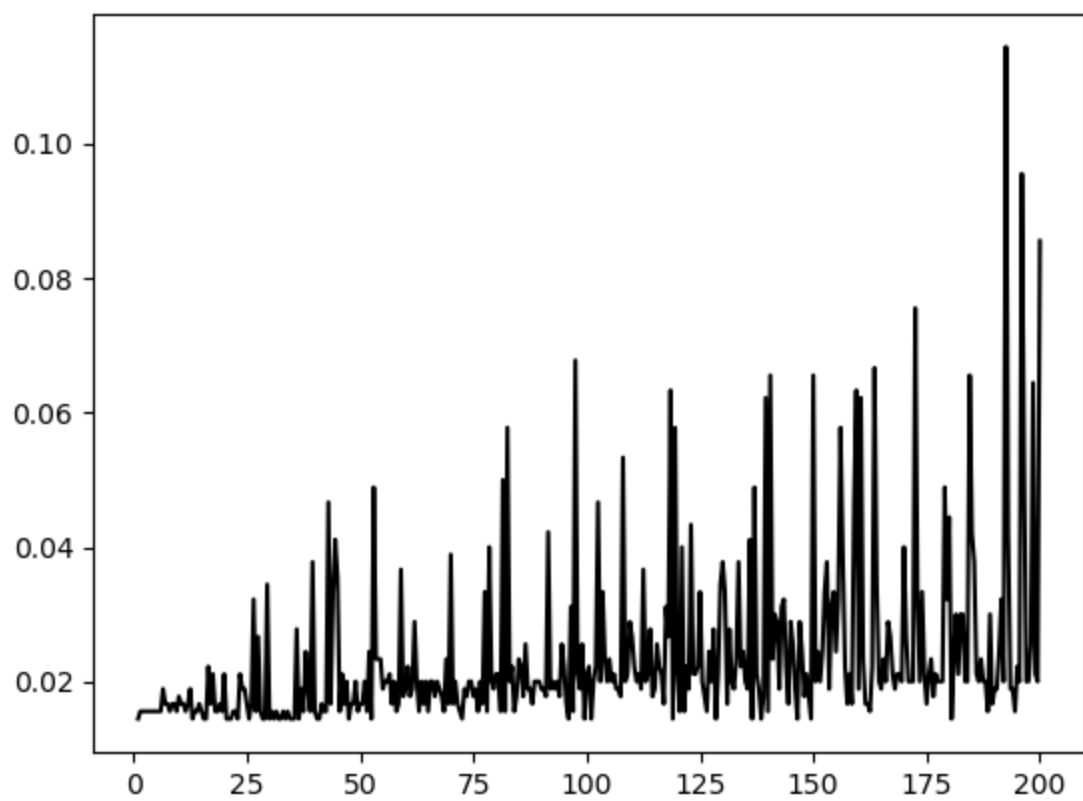
این مطلب در هر دو قسمت اول و دوم قابل مشاهده است.

قسمت اول:

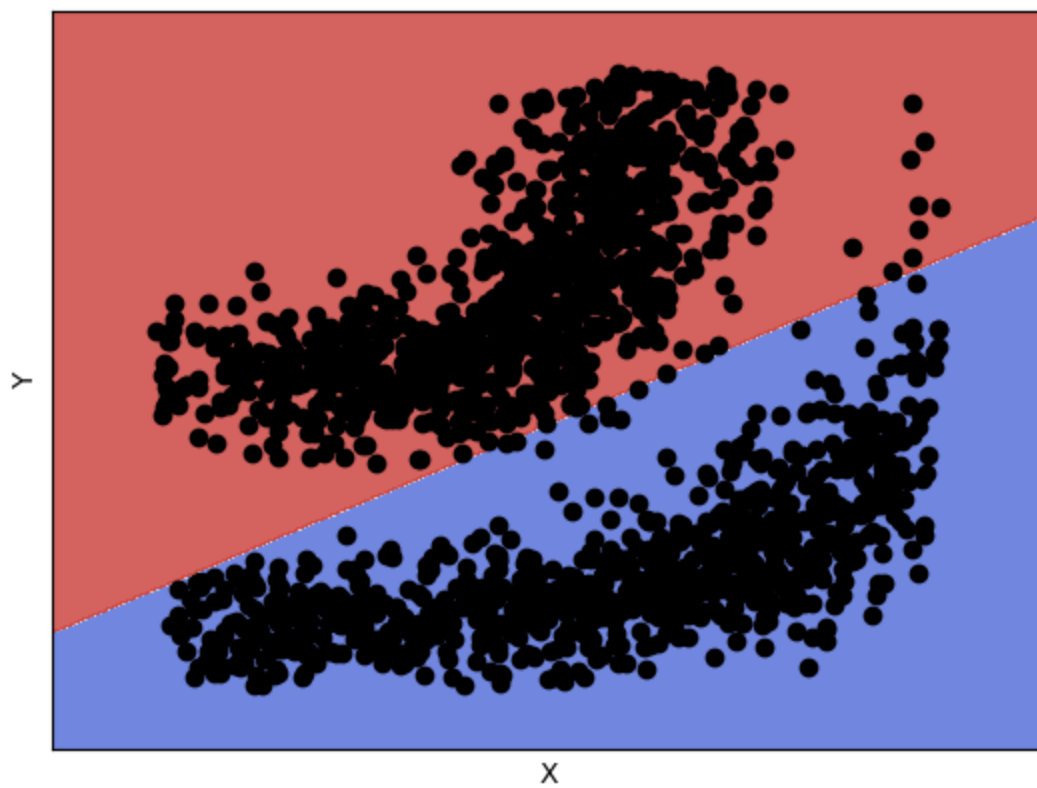
(a)



(b)

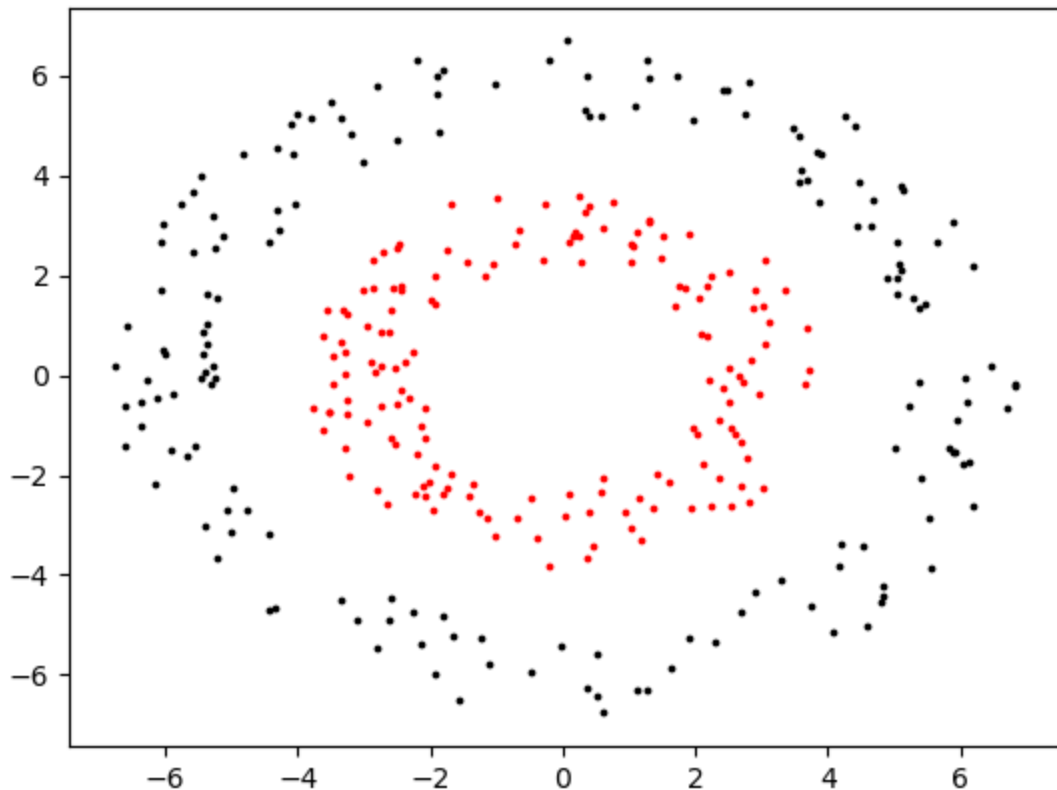


(c)



قسمت دوم:

(a)



با توجه به این که مرز دو بعدی دایره است به نظر می رسد همان تبدیل اسلاید ۵۳ برای این کار مناسب باشد:

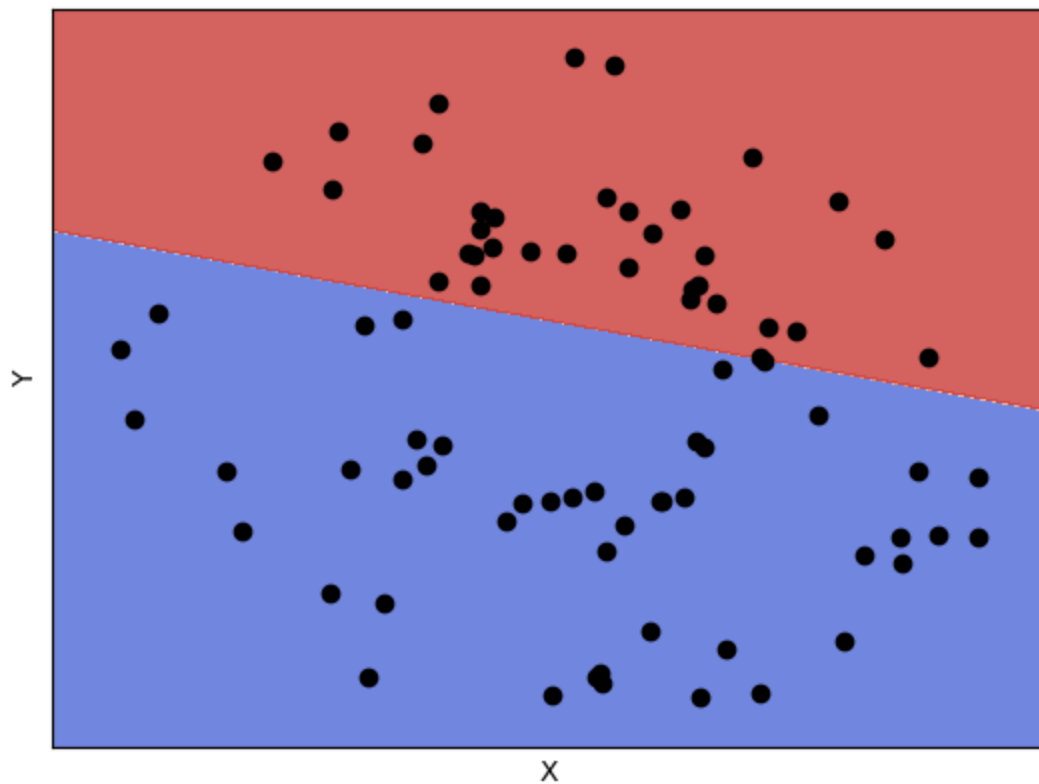
$$\Phi : (x_1, x_2) \longrightarrow (x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, 1).$$

(b)

مختصات جدید ۵ بعدی است. منظور از نمایش در مختصات جدید را متوجه نشدم یا تبدیل سراسر تری وجود دارد که صرفاً از دو بعد مختصات ها را به سه بعد می برد که نتوانستم چنین تبدیلی پیدا کنم.

(c)

با پارامتر های پیشفرض یک SVM خطی آموزش داده ایم که نتیجه به این صورت میباشد:

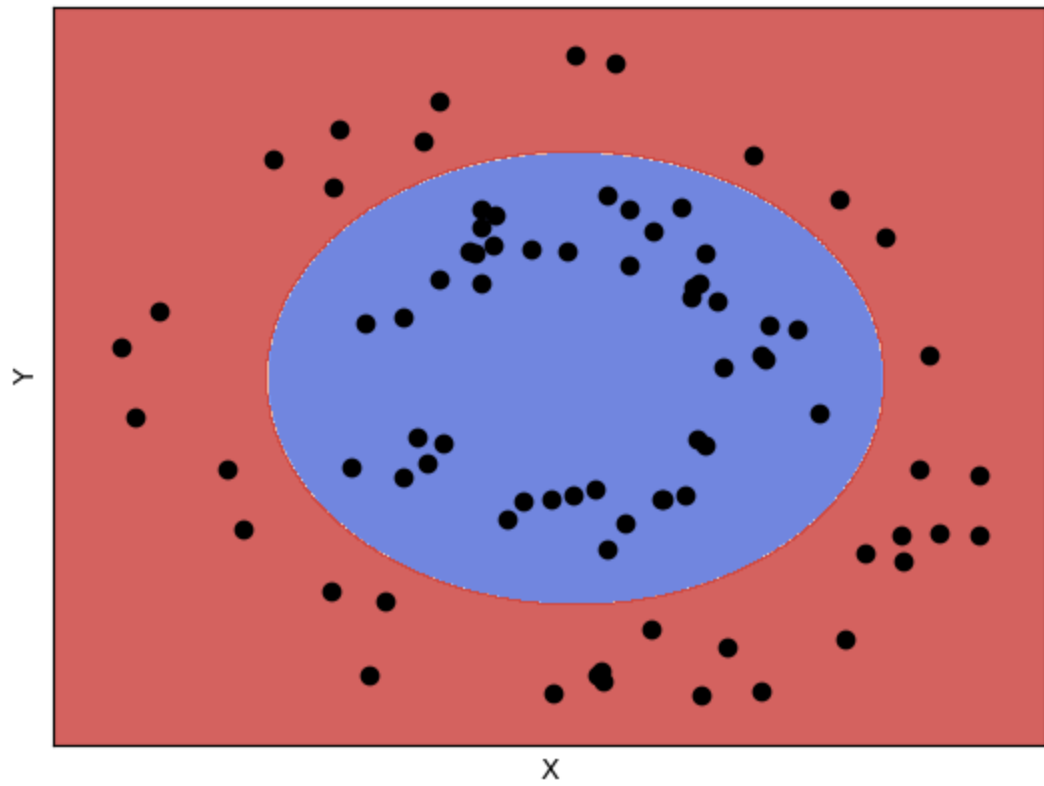


همانطور که در شکل قابل مشاهده است classifier خطی نیمی از داده ها را در یک طرف و نیمی دیگر را در طرفی دیگر کلاس بندی کرده با توجه به شکل دایره ای و مرز دایره ای بین کلاس ها که در قسمت قبل گفته شد پیش بینی می شود نیمی از داده ها اشتباه predict شوند:

The classification error on y_{test} using linear kernel SVM : 0.5875

(d)

برای دسته بندی مناسب باید یک SVM با polynomial kernel بسازیم. polynomial kernel و جزئیات ساخت آن در کد بصورت comment گفته شده است و این جا فقط به نتیجه تصویری بسنده می کنیم:



همانطور که میبینید این classifier بسیار بهتر از classifier قبلی عمل می کند:

The classification error on y_{test} using poly kernel SVM : 0.0