



گزارش تمرین SVM

علیرضا طباطبائی (9723052)

لینک CoLab

ابتدا تمامی دستورات دارای کامنت میباشند .

در اولین گام ، کتابخانه های مورد نیاز را به فایل اضافه کردیم .

سپس داده ها را خوانده و آنها را به 2 دسته آموزش و تست تقسیم بندی کردیم ( 70 و 30 درصد )

سپس هر بخش را به طور جداگانه در زیر توضیح میدهیم .

قسمت اول :

شبکه را با کرنل خطی آموزش دادیم و خروجی ها را پیشبینی کردیم .

سپس دقت را بر روی دادگان آموزش و تست محاسبه کردیم .

نتایج هر بخش در عکس زیر قابل مشاهده است :

```
Accuracy on Test Data in Part A is : 100.0

Confusion Matrix on Test Data in Part A is :

[[33.33  0.    0.   ]
 [ 0.    33.33  0.   ]
 [ 0.    0.    33.33]]

Number of Support Vectors in Part A in each Class is : [ 2 11 10]


Accuracy on Train Data in Part A is : 98.1

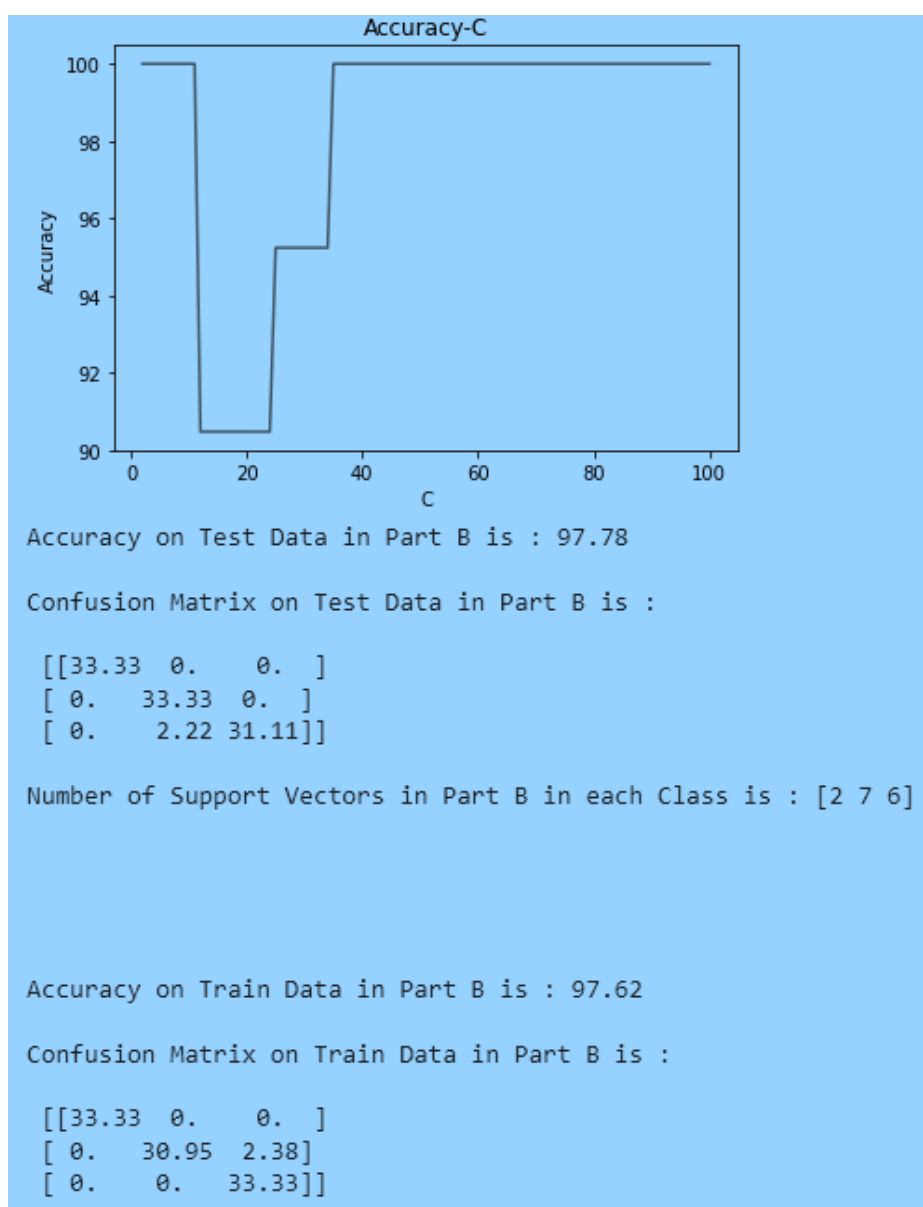
Confusion Matrix on Train Data in Part A is :

[[33.33  0.    0.   ]
 [ 0.    31.43  1.9  ]
 [ 0.    0.    33.33]]
```

قسمت ب :

ابتدا دادگان آموزش را به 2 دسته آموزش و ولید تقسیم کردیم ( 80 و 20 درصد )  
سپس شبکه را به ازای مقادیر مختلف C بین 2 تا 100 آموزش دادیم و نمودار صحت آن را رسم کردیم .

سپس بهینه ترین مقدار C را انتخاب کردیم (  $C = 2$  معمولا ) و شبکه را با آن آموزش دادیم و نتایج بر روی دادگان آموزش و تست به شرح زیر است :



## قسمت پ :

در این بخش ، کرنل غیرخطی RBF و Poly را برای آموزش استفاده کردیم و خطا را بر روی دادگان ولیدیشن حساب کردیم و کرنل همینه که Poly بود را برای مراحل بعدی استفاده کردیم . دقت چند جمله ای بر روی مجموعه آموزش و تست به صورت زیر است :

Accuracy on Test Data in Part C is : 97.78

Confusion Matrix on Test Data in Part C is :

```
[[33.33  0.    0.   ]
 [ 0.    33.33  0.   ]
 [ 0.     2.22 31.11]]
```

Number of Support Vectors in Part C in each Class is : [3 7 5]

Accuracy on Train Data in Part C is : 97.62

Confusion Matrix on Train Data in Part C is :

```
[[33.33  0.    0.   ]
 [ 0.    30.95  2.38]
 [ 0.     0.    33.33]]
```

قسمت ت :

در این قسمت ، کرنل همینه قسمت قبل را انتخاب کرده و با روش قسمت ب ، مقدار C همینه را پیدا میکنیم .

سپس نتایج به صورت زیر بدست می آید :

```
Accuracy on Test Data in Part D is : 95.56
```

```
Confusion Matrix on Test Data in Part D is :
```

```
[[33.33  0.    0.   ]
 [ 0.    33.33  0.   ]
 [ 0.     4.44 28.89]]
```

```
Number of Support Vectors in Part D in each Class is : [3 5 5]
```

```
Accuracy on Train Data in Part D is : 97.62
```

```
Confusion Matrix on Train Data in Part D is :
```

```
[[33.33  0.    0.   ]
 [ 0.    30.95  2.38]
 [ 0.     0.    33.33]]
```

مقایسه نتایج :

برای این دسته از دادگان ، بهترین روش ، استفاده از کرنل خطی بود که البته استفاده از سایر روش ها نیز مشکلی ندارد ولی در حد بسیار کمی ، دقت پایین تری دارد .  
استفاده از SVM غیر خطی ، در شرایطی میتواند بهینه باشد ولی در اینجا تاثیر خاصی ندارد .

مقایسه نتایج با روش پزین :

استفاده از این روش ، تفاوت چندانی با پزین ندارد ولی روش SVM در حد یک درصد ، دقت بهتری دارد . اما در روش پزین ، نیاز است که توزیع پیشین را بدانیم یا حداقل درست حدس بزنیم ولی در اینجا نیاز نیست .