

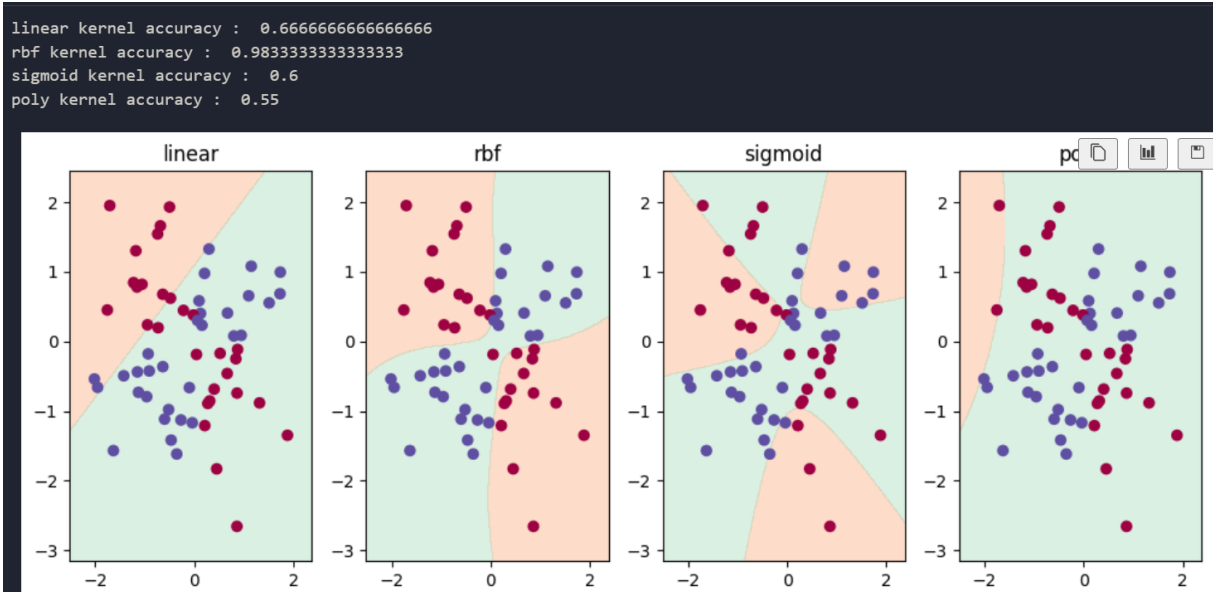
به نام خدا

گزارش پروژه Support Vector Machine

بخش 1 :

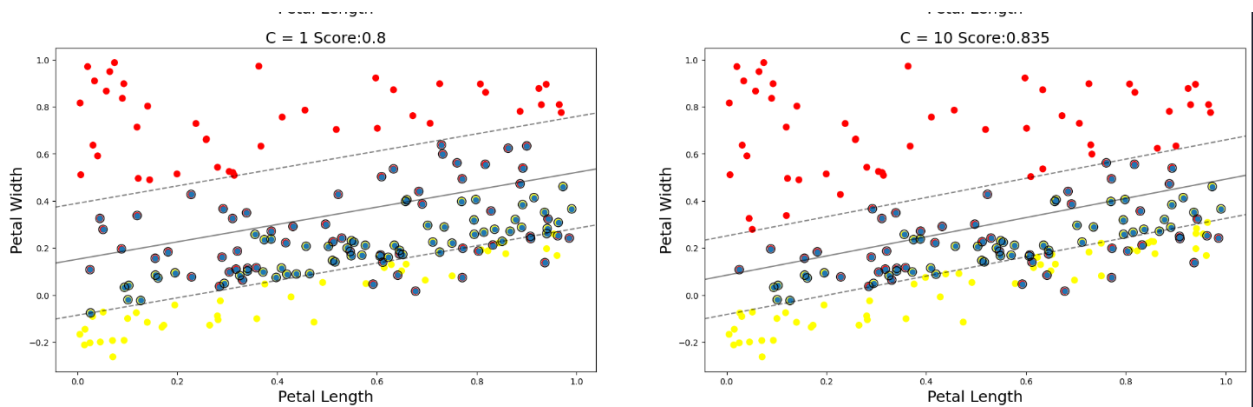
در این بخش 4 معیار مهم در کلاس بندی نقاط در SVM به همراه تاثیر آنها بر دقت اندازه گیری بررسی شد.

1- بررسی kernel و نوع پراکندگی نقاط : 4 نوع kernel : linear, sigmoid, rbf, poly . در توابعی که نقاط به صورت خطی جدا پذیر هستند هر 4 تابع عملکرد خوبی دارند ولی در حالتی که پراکندگی به صورت خطی نمی باشد تابع linear قادر به جدا سازی داده ها نمی باشد و دقتی در حدود 50 درصد تا 60 درصد دارد .

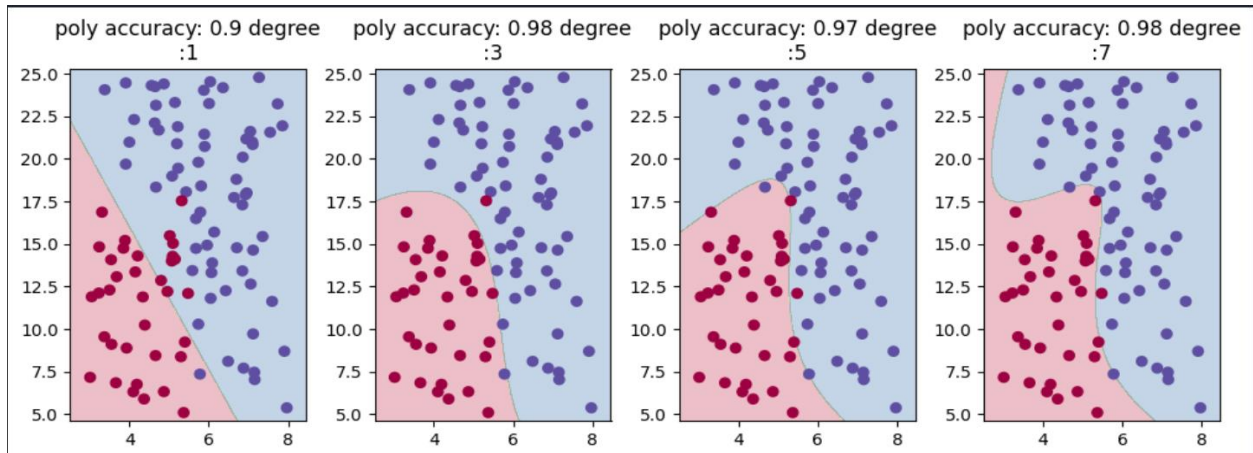


همانطور که مشاهده میشود مدل به هسته linear دارای کمترین دقت است.

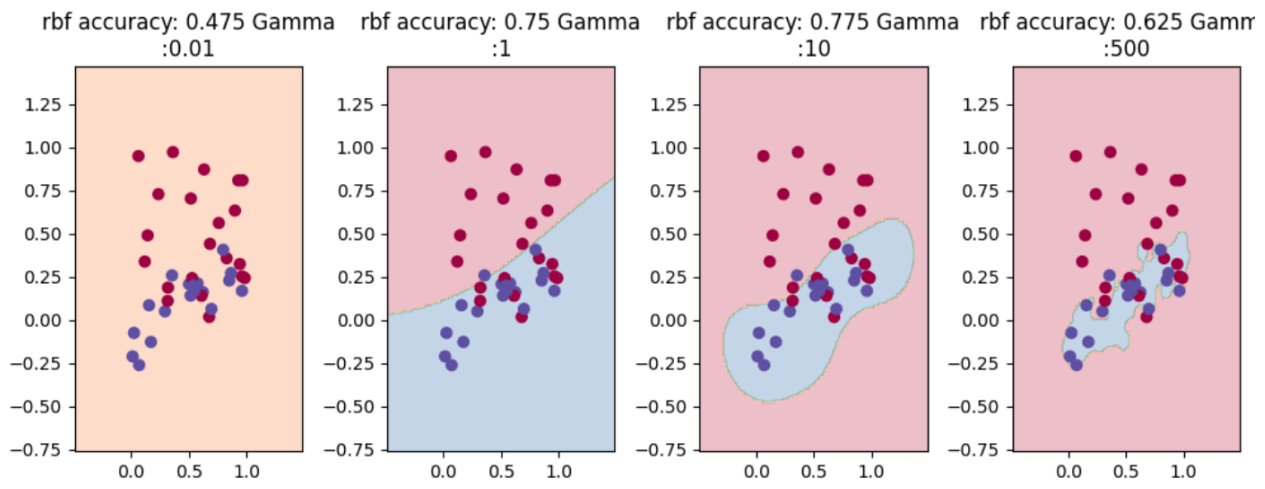
2- بررسی Regularization Parameter : هرچه میزان پارامتر C بزرگتر باشد بمعنی جا برای خطای کمتر و margin کوچکتر خواهیم داشت ، بالعکس در C کمتر margin بزرگتر خواهد بود .



3- بررسی پارامتر degree در مدل با هسته poly : هر چه بیشتر باشد انعطاف خط جدا کننده بیشتر میشود . اما اگر خیلی زیاد شود باعث بیش پردازش میشود.



بررسی معیار γ : بیشتر شدن این معیار ارتباط مستقیم با fit شدن بیشتر خط جدا کننده با داده ها دارد. البته از یه حدی بیشتر سبب بیش پردازش میشود.



همانطور که مشخص است در اولین عکس از سمت راست مدل دچار بیش پردازش شده.

بخش 2 :

در این بخش دیتاست mnist که ارقام دست نویس هست را اینبار با روش SVM کلاس بندی کردیم. همانند پروژه شبکه عصبی ابتدا دیتاست را آماده میکنیم و به دو دسته آموزش و آزمایش تقسیم میکنیم. مدل یادگیری دارای هسته rbf است که مناسب ترین هسته به خاطر انعطاف پذیری آن است.

دقت در پروژه SVM :

accuracy in test set : 0.9676

accuracy in train set : 0.9525

دقت در پروژه شبکه عصبی :

Accuracy in test set : 0.96670001745224

Accuracy in train set: 0.9790

به علت ساده بودن دیتاست تفاوت چندانی میان نتایج مشاهده نمیشود 😊

بخش سوم :

برای این بخش دیتاست دوتایی Cat & Flower انتخاب کردم که از ساییت Kaggle دانلود کردم .

چالش داشتم با آماده سازی داده ها برای مدل یادگیری بسیار چالش داشتم !!!!!!!! 😞 😊

ابتدا تمامی عکس ها را به صورت gray scale در آوردم و سایز همگی را 64x64 کردم . همچنین shape داده ها را به دو بعد تبدیل کردم. تمامی پیکسل های عکس خاکستری شده به صورت یک وکتور در می آید که برای مدل یادگیری قابل پردازش باشد .



مدل یادگیری دارای هسته rbf هست . چون تعداد داده های آموزش به حدود 13 هزار میرسید و مدت زمان آموزش مدل هم در گوگل کولب و هم در سیستم خودم بسیار طولانی بود، تعداد داده ها را به 6000 هزار کاهش دادم (مجبوریه دیگه 😊)

دقت مدل در مرحله آموزش در حدود 68 درصد و برای داده های آزمایش 51 درصد بود . علت کم بودن دقت در روش svm مقایسه با CNN این است که بطور کلی روش های deep learning برای image classification مناسب ترند .

Image classification یک کار پیچیده است که به توانایی یادگیری روابط فضایی بین پیکسل ها و ویژگی های سلسله مراتبی نیاز دارد. یادگیری عمیق بهتر از یادگیری ماشینی قادر به یادگیری این الگوهای پیچیده است، به همین دلیل است که برای کارهای طبقه بندی تصاویر ترجیح داده می شود.