روش svm یکی از روش هایی است که بر اساس کرنل های متفاوت عمل میکند و هر کرنل نحوه نگاشت خطی است که برای تفکیک داده ها مورد استفاده قرار میگیرد.به طور کلی این که کدام کرنل مناسب تر باشد بستگی دارد به ویژگی ها و شکل دیتاست ، یک دیتا میتواند به گونه ای باشد که با کرنل های متفاوت خوب عمل کند و یا برعکس با یک کرنل بهتر از دیگری نتیجه دهد.

|  |  |  |  |
| --- | --- | --- | --- |
| SVM Kernels | | |  |
| index | name | accuracy |  |
| 1 | Rbf | 0.9716 |  |
| 2 | Linear | 0.9666 |  |
| 3 | polynomial | 0.9783 |  |
| 4 | sigmoid | 0.8216 |  |
| 5 | precomputed | \_ |  |

همانطور که مشخص است به صورت default از حالت rbf استفاده میشود که در این دیتاست بهترین نتیجه را ندارد اما در بسیاری موارد کرنل موفقی است همچنین کرنل precomputed به علت ماهیت دیتای مربعی که میخواهد بر روی دیتاست ما قابل اعمال نیست.

3- در مقایسه soft margin با hard margin باید گفت که اولی بهتر است چرا که در زمانی که از hard margin استفاده میکنیم یک داده ی پرت میتواند مرز های طبقه بندی در مدل را جا به جا کند و این اتفاق باعث میشود تا مدل به سمت over fitting نزدیک شود.

5- این دیتاست ویژگی برای one hot encoding ندارد چرا که تمام دیتاست به صورت مقادیر عددی پیوسته یا گسسته است و ویژگی در آن نیست که categorical باشد.

قسمت امتیازی: در تمام مدل های اجرا شده از استاندارد سازی استفاده شده است و علت آن این است که مدل های svm چون از خاصیت برداری استفاده میکنند لذا استاندارد سازی تاثیر بسیار مطلوبی روی عملکرد مدل خواهد گذاشت.

7- در نهایت اعمال تغییرات روی دیتاست و اجرای مدل باعث شده است دقت آن کاهش یابد در رابطه با اضافه کردن ستون مساحت بدیهی است که به یک سری از ویژگی ها توجه بیشتری شده است و در واقع این ویژگی میتواند زائد باشد چرا که ابعاد موبایل دوبار در نظر گرفته شده است و در ارتباط با binning نیز طبیعتا دقت مدل کاهش یافته است.