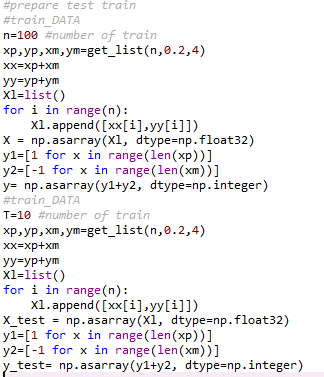
به نام خدا

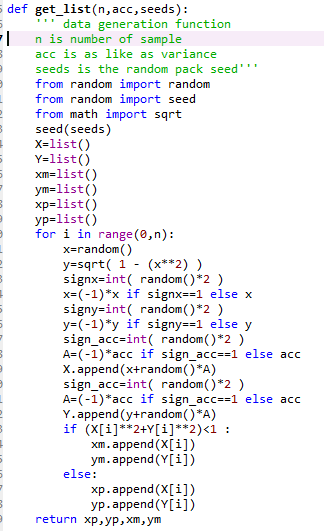
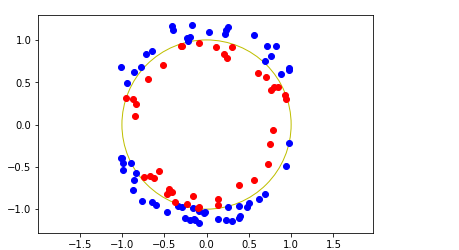
تمرین سری 4

یادگیری ماشین

سوال 1

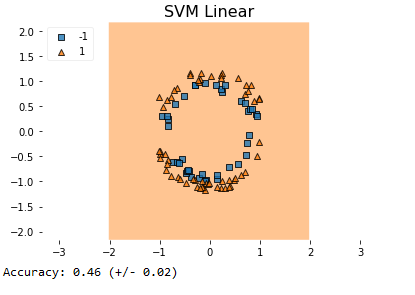
ابتدا اطلاعات را تخمین زده و وارد مسئله میکنیم

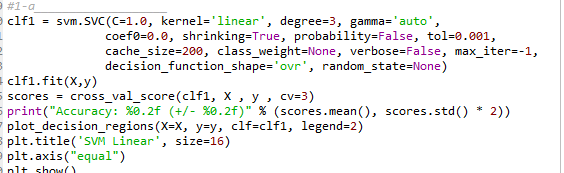
با acc میزان پخش شدن ، seed نوع عدد رندم ، n تعداد عدد عدد رندم



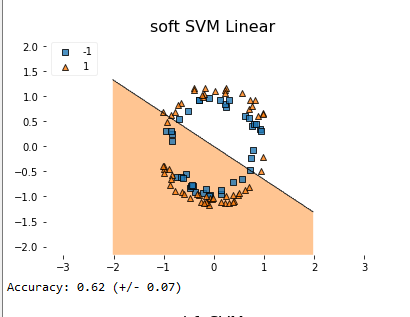
الف ) در بخش اول سوال svm خطی بدون ، خطا را میخواد ، پر واضح است که نمیتوان این کار را انجام داد که با تغییر ، پارامتر ها نتیجه تغییری نخواهد کرد ، با این حال cros\_validation انجام شد و دقت محاسبه شد.

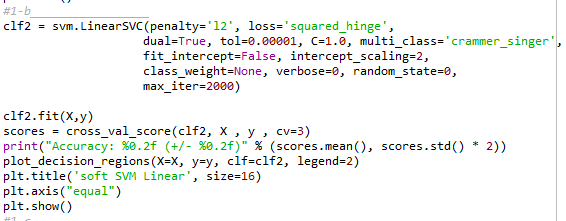
نتیجه:



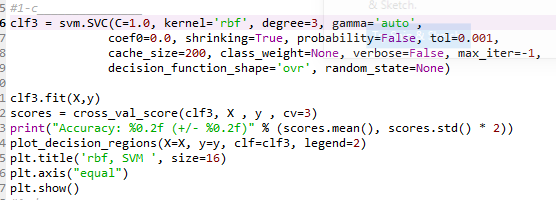


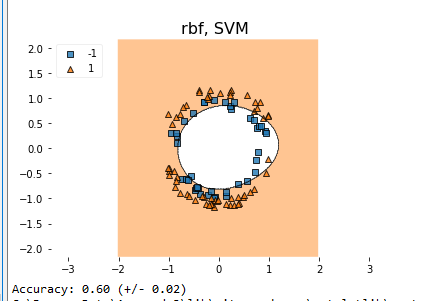
ب) در این بخش ، سوال از ما svm خطی به صورت soft را میخواهد به این معنی که خطا را میتوان پذیرش کرد ، برای یافتن بهرین پارامتر از روی داده یادگیری ، kFold\_cross validation با k=5 انجام شد و بهترین نتیجه 62 درصد حاصل شد.



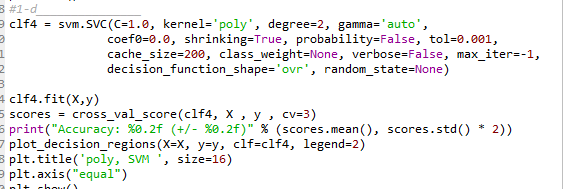


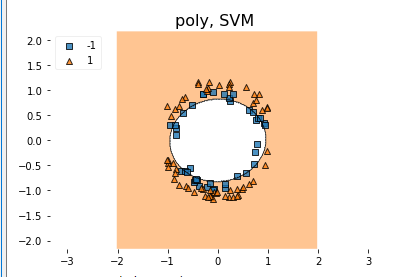
ج) در ادامه سوال پاسخ را برای تابع کرنل rbf ,poly درجه 2 و 3 میخواهد که به شرح زیر است:



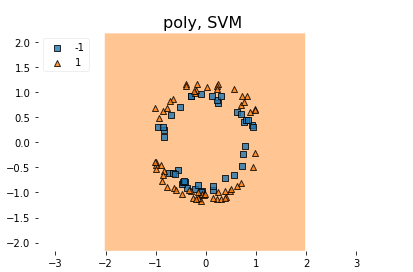


کرنل درجه 2





کرنل درجه 3

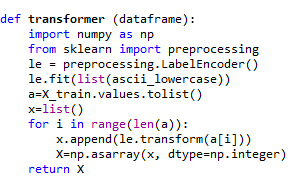


مشاهده میشود که نمیتواند محسابه کند چرا که ماهیت دایره تابعی درجه 2 است و با افزایش درجه خراب میشود.

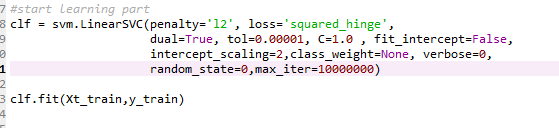
سوال2

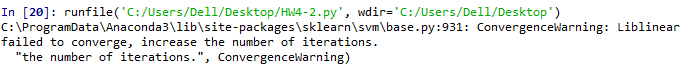
در تمرین شماره 2 نحوه تبدیل اطلاعات به csv توضیح داده شد. از همان دیتا فریم ساخته شده پانداس استفاده شد.

برای سوال 2 ابتدا به دلیل ماهیت categorical بودن اطلاعات باید توسط تابعی ، به اعداد ترجمه شوند تا بتوان از کتابخانه sklearn استفاده شود. برای این کار از مجموعه توابع preprocessing و توسط label enconding ، و تابع زیر ، نوع دیتا را عوض میکنیم

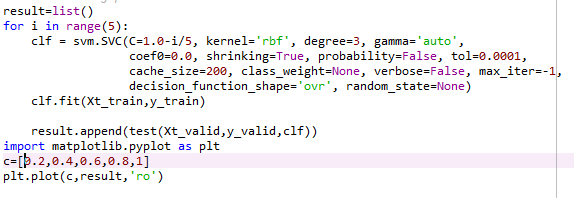


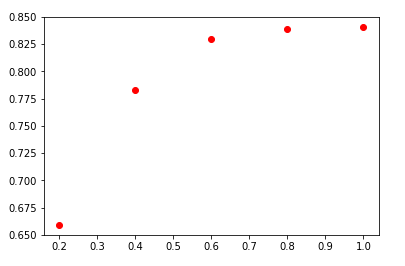
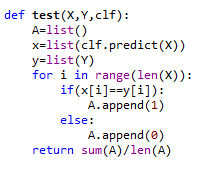
در بخش اول سوال 2 از ما استفاده از svm خطی با تخمین soft است. که با warning همگرا نشدن الگوریتم مواجه میشویم. که با افزایش iterations نیز مشکل حل نمیشود



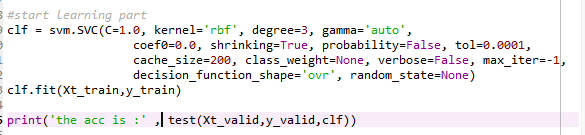


در بخش دوم میزان تغیرات c را در آموزش مد نظر دارد:

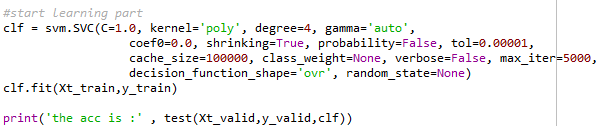
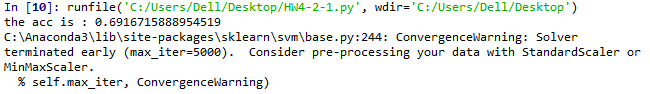




که با نزدیک شدن c به 1 افزایش میابد و میزان دقت روی داده validation به ازای c=1 ، 0.8405197873597164, می باشد و به ازای بررسی روی داده تست : 0.8524203069657615, می باشد.

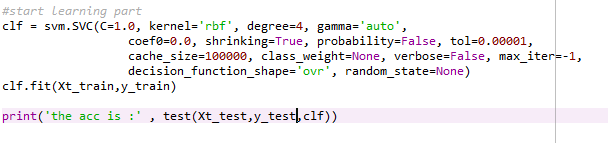
در بخش سوم ، تفاوت نتیجه به ازای چند جمله ای و rbf را میپرسد که دقت rbf روی داده validation برابر : 

و برای چند جمله ای در داده validation ، به طور کلی الگوریتیم همگرا نمیشود و باید ittration را محدود کنیم ، با بررسی چند حالت ، بهترین نتیجه را میتوان از ، درجه4 با max\_iter=5000 گرفت که مشاهده میشود:



نتیجه بهتر روی داده validation برای rbf است که نتیجه روی داده تست میدهد:





می باشد.

نکته جالب این است که این درصد ها در scale درصد های حاصل شده از درخت پس از یادگیری و حرص کردن است.