مروارید لعل نور 9623097

گزارش تمرین سری چهارم درس یادگیری ماشین

سوال اول :

در ابتدا با ساختن ماتریس داده های یادگیری که اولین درایه در سطر های آن مشخص کننده r در مختصات قطبی است و دومین درایه از هر سطر مشخص کننده تتا در مختصات قطبی است . پس از ساختن داده ها باید آنها را لیبل گذاری کنیم که این کار توسط تابع define-y انجام میشود به این صورت که داده های با مقدار کمتراز 1 ، 1- و داده های بزگتر ، 1 لیبل میخورند .

سپس یادگیر svm را تعریف میکنیم و داده های ساخته شده را به آن فیت میکنیم در این قسمت میزان صحت یادکیری بروی داده های تست و ترین 98% و 95% است که با توجه به نمودار رسم شده مقدار درستی نیست زیرا طبقه بندی درست انجام نشده و تنها پیرامون خط نشان داده شده داده ها تقریبا درست تشخیص داده شده اند .

نتیجه ای که از این قسمت میگیریم آن است که برای طبقه بندی اینگونه داده ها به چیزی بیشتر از یک خط نیاز داریم

ب : در این قسمت با مقایسه مقدار صحت بدست آمده بر حسب c در مدل استفاده شده و با رسم نمودار به این نتیجه رسیدیم که 1 بهترین مقدار برای این مقدار ثابت است و به ازای استفاده از این مقدار در svm نرم خواهیم دید که مقدار صخت بع 100 % میرسد اما باز هم این مقدار درست نیست و با توجه به تصویر مشاهده میکنیم که قسمت عمده ای از دادها نادیده گرفته شده و تنها در قسمتی که خط مرزی رسم شده این تقسیم بندی صحیح صورت گرفته است .

ج: در این قسمت به سراغ استفاده از کرنل ها میرویم با توحه به کد (و کامنت گداری ها) متوجه میشویم که در این داده به خصوص هر دو کرنل rbf و چند جمله ای درجه 3 میزان صحت بالایی داند . در اینجا کرنل چند جمله ای را به عنوان کرنل بهتر انتخاب کرده و میزان صحت و نمودار تشخیص آنرا مشاهده میکنید.

سوال دوم :

ابتدا با یک یادگیر svm شروع میکنیم و دادگان را اموزش میدهیم صخت بدست آمده و ماتریس confusion و تعداد ساپورت وکتور ها در کد ( با پسوند .ipynp ) مشخص شده اند که دراینجا 87 % صحت عملکرد روی دادگان تست داریم و 88% رو validation .

ب : در این قسمت با استفاده از همان روش سوال قبل با آزمایش آموزش با c های مختلف به این نتیجه میرسیم که مقدار 1 بهترین مقدار است زیرا در این مقدار میزان صحت 95% است و هنوز overfit رخ نداده است . نتایح نهایی این آموزش نیز در کد مشخص شده اند ، 100% روی validation و 96% روی تست .

ج: در این قسمت از کرنل poly با درجه 8 استفاده کرده ایم که باعث بهبودی یادگیری شده است زیرا میزان صحت عملکرد همان 96% مانده اما تعداد ساپورت وکتور ها و خطا ها با توجه به ماتریس confusion کاهش یافته اند اما زمانی که از rbf استفاده میکنیم نتایج از این هم بهتر شده و صحت به 100% میرسد . پس کرنل دوم را انتخاب کرده و برای دادگان تست نیز آنرا بررسی میکنیم که همانطور که نشان داده شده صحت در این قسمت 98% است .

د: در این قسمت با ترکیب rbf از قسمت قبل و ثابت c که با توجه به نمودار صحت برحسب c ، 0.5 مقدار دهی شده است آموزش را دوباره انجام میدهیم و مشاهده میکنیم در هردو قسمت validation و تست صحت عملکرد 96% است ولی ما در قسما های قبل به مقادیر بالاتری هم دست یافته بودیم پس نتیجه میگیریم این نوع ترکیبی بری این دادگان عملکرد ضغیف تری نسبت به زمانی که از خود rbf به تنهایی استفاده کردیم دارد و مناسب تر است از آن نوع اموزش استفاده شود.

( تمامی نتایج و نمودار ها درون کد قابل مشاهده هستند و کامنت گذاری نیز انجام شده است)