

سوال ۱) در یک مدل SVM با kernel خطی یک خط جدا کننده

به صورت

$$w_1 a_{11} + w_2 a_{12} + b = 0$$

در صورتی که اثر جدا سازی مثبت فاصله آن ها از مرکز است یعنی آن به صورت  $r_i$  و  $(a_{1i} - a)^2 + (a_{2i} - b)^2$  است که  $r_i$  اثر کمتر از threshold داشته باشد به کلاس A و در غیر

این صورت به کلاس B مربوط است. پس باید یک فضاگشت داشته باشیم توان با

linear kernels هم به جواب رسید.

سوال ۲) دو تا انتخاب داریم یا kernel poly که یک چند جمله ای است

که با توجه به شکل با درجه ۳ می تواند تغییر در فضاگشت را داشته باشد

و  $rbf$  که به بعد هر یکی از بردار آن یک ضلعی صفحه در ابعاد یک

بعد کمتر از آن بعد می تواند آن را تغییر دهد و فضاگشت را تغییر دهد

ولی در یک اثر مثبت مهم است <sup>ترین</sup> یک kernel poly مربع وزن های مثبت می شود

اما اثر حقت مهم تر است  $rbf$  در یک بعد های بالاتر تاثیر بیشتری دارد

و ا تا حد خوبی یاد میگیرد.





Date: \_\_\_\_\_

Sa	Su	Mo	Tu	We	Th	Fr
----	----	----	----	----	----	----

Subject: \_\_\_\_\_

مسئله ۳) kernel rbf عملکرد بهتری داشت دقت بیشتری هم داشت  
به دلیل اینکه noise داده نسبتاً مشهود بود و مدل Poly عملکرد ضعیف  
تری دارد در این نمونه از داده چون rbf آن را به جبراً تقسیم می‌کرد و آن جبر  
feature های که اضافه می‌شوند نسبت به مدل که داده های این داده  
تمایز بیشتری قائل می‌شوند و می‌توانست برای داده ای که لزوماً اتوکامی  
ندارد در آن جبر تمایز قائل شود و باعث یادگیری Noise شود

پس در حالت عادی rbf دقت بیشتری داشته و می‌تواند اصلاح  
بالاتر است پس در حالتی که تقسیم پذیر مهم تر باشد احتمالاً Poly  
بهتر است

مسئله ۴) برای داده iris که خطی نمی‌باشد و به هم آمیخته و به دقت بالاتر  
ولی برای داده breast cancer خیر. در داده داده می‌تواند دو دسته  
اتوکام با Noise صحت دارد و دسته های گاهی تا حدی جدا شده اند ولی در  
نمای نهایی Noise ها بیشتر تمایز را ضعیف کرده اند ولی دلیل  
با مدل خطی که بیشتر به سمت یادگیری اتوکام رفت اصلاحاً لا دقت کمتری  
مدل هم ضعیف داشتیم منظره این که این مدل خطی under fit کرده است  
چرا که خیلی ساده شده و نتوانسته اتوکام را به خوبی یاد بگیرد و تقسیم  
داشتیم بهترین می‌توانیم یعنی میزان خطا را تا حدی ضرب تعلق کنیم





Date: \_\_\_\_\_

Sa	Su	Mo	Tu	We	Th	Fr
----	----	----	----	----	----	----

Subject: \_\_\_\_\_

سوال ۳) در ابعاد بالا تر به دلیل اینکه Feature ها زیاد می شوند ممکن است بعضی از Feature ها تأثیر ضعیف داشته باشند  $\text{kernel}^{rbf}$  از این نظر می تواند ضرب با هم کرده این تأثیر در Feature با هم می شود در بعد بالا تر جوابها را به صورت خطی بهتر شود ولی هزینه آن در زمان تدریس افزایش می یابد و در صورتی که  $\phi$  با داشتن یک Feature ضرب چن آن را چند جلدی و یک به نفع  $n$  بعضی مدل می تواند ممکن است آن Feature به ضرب  $\text{rbf}$  انویست را پیدا کند ولی با هزینه کمتری توان آن را انجام دهد و علاوه بر این هم بخشی از انتخاب است خصوصاً در بعد های بالا تر که یک Feature ضرب برای خطی می تواند با هزینه کمتر از  $\phi$  مدل ضرب ارائه دهد.

سوال ۴) طبیعتاً بدون Noise دقت بیشتر است چرا که Noise آلودگی است و کمبود ترمی کندها را تعمیم پذیر را بیشتر

سوال ۵) با افزایش Noise آلودگی نهفته در داده کم می شود و ترمی شود طبیعتاً در مقادیر بالا آلودگی فاضی شده نمی شود پس Noise زیاد اثر منفی دارد تعمیم پذیر را زیاد ~~می کند~~ Noise به دقت بسیار به معنای نبود آلودگی و عدم قطعیت بسیار بالا است

سوال ۸) در حالت خطی چن برای خطی  $\phi$  متغیر نداشته ایم  $C$  چن در تدریس مدل جدید اما بین  $\phi$  و  $\text{rbf}$  توان گفت  $\text{rbf}$  بهتر کنار آمده است در بعد بالا تر آلودگی ها حاصل از Noise بهتر جوابها را می گویند مدل  $\phi$  هم عملکرد خوبی داشته است اما در Noise بیشتر اما عملکرد خوبی ندارد.

RASTA



سوال ۹) مدل عمکرد خطی داشته اند به دلیل اینکه  $class\_sep$  برای  $overlapping$  نسبتاً متغای اعمال می کنند اما برای  $class\_sep$  های پایین  $rbf$  عمکرد خطی بهتری دارد چون انوار حالت  $poly$  خارج می شود با خدایت کردن یا صیبه ابعاد بالا قدرستی یافتنی تر است.

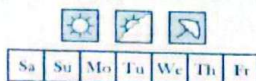
سوال ۱۰) جمله راجب  $class\_sep$  در سوال ۹ تر ضیح دادم اما در مورد تعداد  $feature$  های متغایر گفته عمکردیست است اما در صورتی افزایش هم خودیتر دیتایست و مشخص که با  $polynomial$  خرابتر و خدایت تر و  $rbf$  عمکرد بهتری دارد.

سوال ۱۱)  $C$  که بیشتر شود بیشتر خدایت به  $overfit$  و درک اتمو حتی اگر  $noise$  با مقدار دارد اما وقتی که  $C$  کمتر است  $underfit$  می شود مگر خطای ساده می بیند و شانس برای خدای خط و تعمیم در نظر نمی گیرد پس مقادیر  $C$  و  $\lambda$  عمکرد خطی داشته اند  $overfit$  و  $underfit$

سوال ۱۲) طبق  $heat\ map$  برای انواع مدل ها بیشترین خطای که در  $C$  های کم هستند اینست که داده ها را در یک دسته قرار داده است و به خطای کمتری که داده ها کمتر دارد اهمیت داده است یعنی  $۲۳$  دلیل به کمترین حدتعلق دارند ولی همه آن ها را  $۱$  پیمایشی کرده است و به خطای کمتری که برای حل آن با افزایش  $C$  تا حدی که به  $overfit$  نرسیم می توانیم این خطا را حل کنیم مثلاً  $[۵۰۰۰۰۰]$



Date: \_\_\_\_\_



Subject: \_\_\_\_\_

سوال ۱۲) فرض کنید که داده‌ها از هر یک از دو دسته به دست می‌آیند

۲۴ Sample به دست می‌آید که ۲ دسته به دست می‌آید و ۱۲ دسته به دست می‌آید  
 داده‌ها به ۲ دسته به دست می‌آید و ۱۲ دسته به دست می‌آید  
 به دست می‌آید و ۱۲ دسته به دست می‌آید

سوال ۱۲) مدل تا حدی  $VA\%$   $accuracy$  را  $test$  توانسته است که  
 جدا کند و در مورد  $kernel$  ها اثر  $C$  را متداری در نظر بگیریم  
 مدل به دست می‌آید  $linear$  با  $accuracy$  و اثر  $C$  را افزایش  
 دهیم با  $VA\%$  تا  $VA\%$  دقت بدون  $overfit$  می‌تواند