

بسمه تعالی

## ماشین بردار پشتیبان

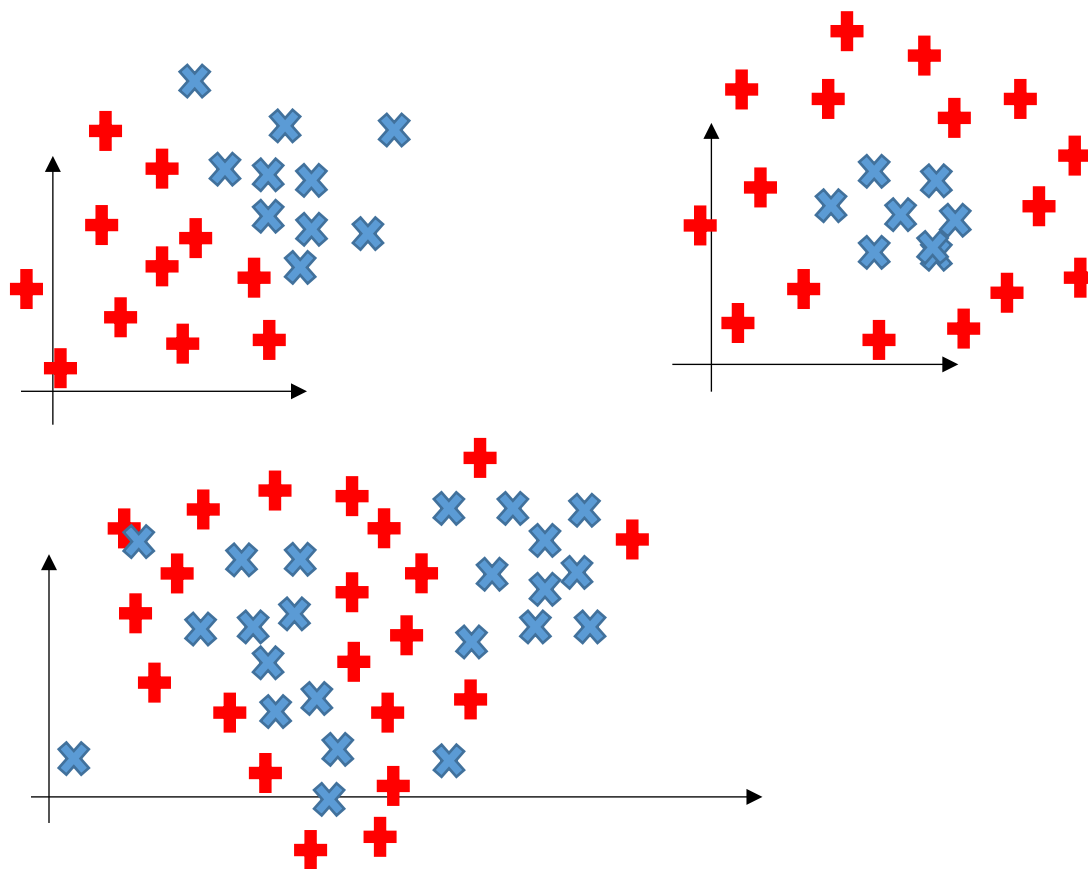
- مهلت تحویل ، ۴ بهمن ۱۴۰۱، ساعت ۲۳:۵۵
  - مهلت ارسال قابل تغییر نیست.
  - مواردی که بعد از تاریخ فوق ارسال شوند قابل قبول نبوده و نمره ای نخواهد داشت.
  - انجام تمرین تک نفره است.
  - کل محتوای ارسالی زیپ شود و نام فایل زیپ ارسالی HW6\_studentNumber باشد.
  - محتوای ارسالی دارای راهنما (read me) جهت تسهیل اجرا باشد.
  - زبان برنامه نویسی دلخواه است.
  - موارد ارسال شده در تاریخی که بعداً مشخص می‌شود به صورت مجازی نیز تحویل گرفته خواهند شد (صرفاً آنچه در lms طبق تاریخ های فوق تحویل داده شده است بعداً به صورت مجازی تست شده و توضیح داده می‌شود)
  - تنها تکالیفی که به lms و قبل از مهلت ارسال، فرستاده می‌شوند بررسی خواهند شد.
- 

## شرح:

در این تمرین از ابزارها و کتابخانه های آماده SVM برای آشنایی با قابلیت های دسته بندی SVM استفاده می کنیم.

## بخش ۱:

ابتدا چند مساله دو کلاسه از خیلی ساده تا خیلی پیچیده طراحی کنید. به عنوان مثال، تعدادی نقطه (مثلا ۱۰۰ تا ۱۰۰۰) در فضای دو بعدی در نظر بگیرید و برخی از نقاط را عضو کلاس ۱ و برخی دیگر را عضو کلاس ۱- فرض کنید. نقاط را در یک نمودار دو بعدی رسم کنید. مثلا شکل های زیر می توانند چند مثال پیشنهادی باشند:



سعی کنید در تولید نقاطی مانند نقاط فوق، دستی کار نکنید، یعنی با کمک توابعی نقاطی را تولید کنید که هر بار بتوانید نقاطی و دسته هایی با ویژگی ها و شکلهای متمایز ایجاد و آزمایش کنید.

سپس با SVM اقدام به دسته بندی داده های ساختگی خود کنید. بدون هسته، با انواع هسته ها، پارامترهای مختلف برای هر هسته و ... را تست کنید.

خط جداکننده ای که SVM یافته است را در کنار نقاط آموزشی در یک نمودار رسم کنید. ترجیحا خط مربوط به Margin را هم رسم کنید. برای رسم خط جدا کننده و Margin، کتابخانه های متداول معمولا تابعی در اختیار می گذارند.

اگر پیچیدگی داده‌ها را بیشتر کنید، چه تاثیری در انتخاب هسته و چه تاثیری در پارامترهای آن هسته خواهد داشت؟ قدری تَفَلُّسْ کنید!

## بخش دوم<sup>۱</sup>:

ابتدا لازم است یک پایگاه داده (Data Set) برای خود انتخاب کنید. ترجیحا پایگاه داده‌ای تصویری انتخاب نمایید. مثلا به عنوان پیشنهاد، می‌توانید پایگاه داده‌های مربوط به شناسایی ارقام دستنویس انگلیسی نظیر USPS یا MNIST (یا اگر برای ارقام دستنویس فارسی پایگاه داده‌ای پیدا کردید) را در نظر بگیرید<sup>۲</sup>.

پایگاه داده انتخاب شده به دو بخش train و test تقسیم کنید و فاز آموزش را روی train set انجام دهید و سپس دقت دسته بندی با SVM را در داده های تست به دست بیاورید. کدام هسته و با چه پارامترهایی جواب بهتری می‌دهد؟<sup>۳</sup>

آنچه تحویل داده می‌شود:

- ۱- کداجرایی برنامه با توضیحات لازم برای اجرا و احتمالا Test Case هایی برای بخش اول
- ۲- گزارشی مختصر از مسیر انجام کار، چالش‌هایی که احتمالا مواجه شدید، اجراهایی که گرفتید و نتایجی که حاصل شده است. دقت در داده های آموزشی و آزمایشی چقدر بوده و چقدر تفاوت داشته؟ آیا بیش برآزش داشته اید؟ و سایر آنالیزهایی که کرده‌اید.

---

<sup>۱</sup> - در این بخش و بخشهای بعدی، نکات و اصول کلی که در یادگیری ماشین باید رعایت شود و در پروژه های دیگر مانند درخت تصمیم و ... تجربه کرده اید را رعایت کنید. مثلا در مواقع لازم Cross Validation یا جلوگیری از بیش برآزش، توجه به اهمیت تعداد داده آموزشی و آزمایشی، مقایسه دقت در مجموعه آموزشی و آزمایشی به طور جداگانه و سایر مواردی که روح کلی یادگیری ماشین اقتضا میکند را رعایت کنید، البته اگر اعتقادی به روح دارید!

<sup>۲</sup> - فایل USPS\_images.rar یکی از این دیتاست هاست که میتوانید روی همان کار کنید.

<sup>۳</sup> - تصاویر را لود کنید، اگر به صورت رنگی (RGB) لود شده اند، آنها را خاکستری (Grayscale) کنید. بدین ترتیب هر تصویر تبدیل به ماتریسی میشود که تعداد سطرها و ستونهاش به اندازه طول و عرض تصویر است و مقدار هر درایه در ماتریس نیز معادل شدت روشنایی پیکسل مربوطه است. شدت روشنایی معمولا عددی بین صفر تا ۲۵۵ (یا بین صفر تا ۱) است. بیشترین مقدار شدت روشنایی متناظر با رنگ سفید مطلق و مقدار صفر آن متناظر با رنگ سیاه مطلق است. حال کافی است این ماتریس n در m را به صورت برداری تک بعدی در بیاورید که اندازه آن n×m خواهد بود. یعنی مثلا سطر به سطر (یا ستون به ستون) آن را پشت سر هم قرار دهید. بردار حاصل شده همان چیزی است که باید به عنوان نقطه ورودی به SVM بدهید. یعنی فضای مساله n×m بعدی خواهد بود و هر تصویر تبدیل به نقطه ای n×m بعدی در این فضا شده است. البته می‌توانید قبل از دادن بردارها به SVM، اقدام به نرمال سازی آنها نیز نمایید.