

بسمه تعالی



تمرین سوم

## ماشین بردار پشتیبان<sup>۱</sup>

استاد درس: دکتر آرش عبدی هجراندوست

تدریسار آموزشی: آرمین توکلی

✓ نکات تمرین :

❖ مهلت تمویل ۱۴۰۱/۰۹/۰۴ ساعت ۲۳:۵۵

❖ مهلت ارسال به هیچ وجه قابل تغییر نیست .

❖ مواردی که بعد از تاریخ فوق ارسال شوند قابل قبول نبوده و نمره ای نخواهند داشت .

---

<sup>1</sup> Support Vector Machine (SVM)

❖ انجام تمرین تک نفره است. لطفا به تنهایی انجام شود، در غیر اینصورت نمره منفی در نظر گرفته خواهد شد.

❖ کل مفتوای ارسالی زیپ شود و نام فایل زیپ ارسالی HW3\_StudentNumber\_YourName باشد.

❖ مفتوای ارسالی دارای راهنما (read me) جهت تسهیل اجرا باشد.

❖ زبان برنامه نویسی دلفواه است. (پیشنهاد : پایتون)

❖ در صورت استفاده از زبان پایتون فایل کد ترجیحا به فرمت ipynb. بوده و فایل کد حتما به

صورت اجرا شده آپلود گردد و از وجود فروبی سلول ها اطمینان حاصل نماید.

❖ موارد ارسال شده در تاریخی که بعدا مشخص خواهد شد و متعاقبا اعلام می گردد به صورت آنلاین نیز تعمیل گرفته خواهند شد (صرفا آنچه در LMS طبق تاریخ فوق تعویل داده شده است بعدا به صورت حضوری تست شده و توضیح داده می شود).

❖ تنها تکالیفی که به LMS و قبل از مهلت ارسال، فرستاده می شوند بررسی خواهند شد.

❖ در صورت داشتن هر گونه سوال می توانید سوال خود را در گروه تلگرامی درس مطرح کنید.

❖ حداقل یک ساعت قبل از مهلت ارسال را احتیاطا هدف قرار دهید، تا مشکلات غیرقابل پیش

بینی مانند موارد زیر باعث عدم آپلود پاسخ ها در LMS و ارسال آنها از طریق ایمیل نشوند :

(قطعی اینترنت، تنظیم نبودن دقیق ساعت سایت با ساعت کرنویچ، کرش سیستم عامل و نیاز به فرمت، بارش زیبای شهاب سنگ از آسمان و ...)

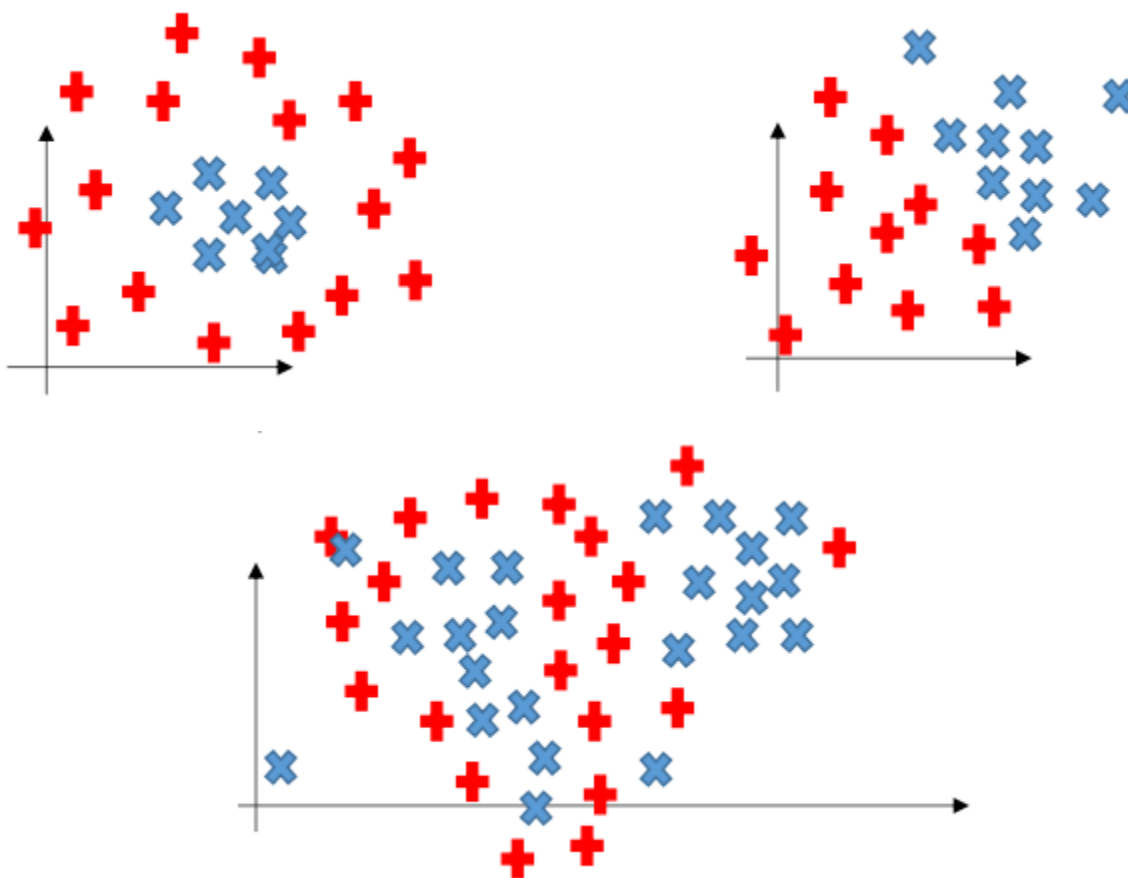
---

شرح تمرین :

## بفش اول :

ابتدا چند مساله دو کلاسه از فیلی ساده تا فیلی پیچیده طراحی کنید. به عنوان مثال، تعدادی نقطه (۵۰۰ تا ۵۰۰۰ نقطه) در فضای ۱۰ بعدی برای دو کلاس فرض کرده و نقاط را در نمودار دو بعدی رسم کنید.

مثال شکل های زیر می توانند چند مثال پیشنهادی باشند:



یکبار با استفاده و کمک از توابعی نقاطی را تولید کنید که هر بار بتوانید نقاطی و دسته هایی با ویژگی ها و شکلهای متمایز ایجاد و آزمایش کنید و یکبار نقاط را به صورت دستی تولید و آزمایش کنید.

سپس با ماشین بردار پشتیبان، اقدام به دسته بندی داده های سافتگی خود کنید. بدون هسته، با انواع هسته ها، پارامترهای مختلف برای هر هسته و ... را تست کنید و برای هر کدام از مجموعه داده ها بهترین پارامترها را پیدا کنید.

فقط جداکننده‌های که ماشین بردار پشتیبان یافته است را در کنار نقاط آموزشی به همراه فاصله **Margin** در یک نمودار رسم کنید برای این کار می‌توانید از کتابخانه‌های آماده بهره بگیرید.

- رفته رفته پیچیدگی داده‌ها را بیشتر کرده و عملیات بالا را روی داده‌ها تکرار کنید. پیچیدگی داده‌ها چه تاثیری در انتساب هسته و پارامترهای مربوطه خواهند داشت؟ به طور کامل شرح دهید.

- آیا می‌توان به طور قطع در مورد نوع هسته خاصی اظهار نظر کرد و مطمئن بود همواره بهترین خواهد بود؟ به طور کامل شرح دهید.

- در مورد پارامتر تنظیم<sup>۲</sup> روی داده‌ها توضیح دهید و نتایج مختلف را به سبب تغییر میزان آن به طور منظم به فرد به غیر از مابقی پارامترها به طور کامل شرح دهید.

### بخش دوم :

در این بخش با استفاده از این [لینک](#) داده‌گان مورد نظر را دانلود کنید. داده‌گان مورد نظر در مورد تشفیص وضعیت هوا هستند که در این مجموعه از چهار نوع وضعیت هوا تعدادی عکس از وجود دارد که شما می‌بایست با استفاده از ماشین بردار پشتیبان نوع وضعیت هوا را مشخص کنید.

لازم به ذکر است که مثل روال سابق می‌بایست مجموعه‌ای از داده‌ها را برای آموزش و مجموعه‌ای را برای آزمون در نظر بگیرید.

همگی عکس‌ها را در سطح فاکستری خوانده و سپس ابعاد تمامی تصاویر را یکی کرده (به انتساب خودتان) و را انجام دهید.<sup>۳</sup> سپس عملیات رده‌بندی

<sup>۲</sup> Regularization Term

<sup>۳</sup> Classification

در این بخش نیز با تمامی هسته‌ها و تغییر انواع پارامترها عملیات رده‌بندی را انجام داده و سپس نیز بهترین پارامترها را پیدا کنید.

همانند بخش قبل عملیات مصورسازی مرز تصمیم ماشین بردار پشتیبان و بردارهای یافته شده توسط این ماشین را نیز رسم کنید.

- آیا اگر مجموعه داده‌گان دارای عدم تعادل باشد (یعنی از هر کلاس به تعداد تقریباً یکسان داده موجود نباشد، برای یک کلاس تعدادی خیلی کم و برای دیگری تعدادی بسیار و یا تا حدی بیشتر)، آیا رده‌بندی ماشین بردار پشتیبان دچار مشکل می‌شود و رده‌بندی توسط آن تحت الشعاع قرار می‌گیرد؟ چرا؟. به طور کامل توضیح دهید.

- راه حل مشکل بالا یعنی عدم تعادل مجموعه داده‌گان چیست و تأثیرات کلی آن را روی عملیات رده‌بندی توسط هر رده‌بندی را توضیح دهید.

آنچه تمویل داده می‌شود:

- ۱) کد اجرایی برنامه با توضیحات لازم برای اجرا
- ۲) گزارش کاملی از مسیر انجام کار، چالش‌هایی که مواجه شده‌اید، اجراهایی که گرفتید و نتایجی که حاصل شده است. گزارش کار از اهمیت بالایی برخوردار است، مهم آن و فرمت استاندارد آن اهمیت ندارد، اما باید نشان دهنده مسیر انجام پروژه، چالش‌ها، راه حل‌ها و نتایج کار شما باشد