

به نام خدا



هوش مصنوعی و سیستم های خبره

پروژه چهارم

ماشین بردار پشتیبان SVM

دکتر آرش عبدی

پاییز ۱۴۰۲

طراح : محمد علی آژینی، نگین حقیقی

- در صورت وجود هر گونه ابهام تنها به طراح پیام دهید .
- با توجه به فشرده بودن برنامه ترم و پروژه های پیش رو و همچنین مکانیزم تاخیر 10 روزه امکان تمدید وجود ندارد.
- داک پروژه را واضح و مرتب بنویسید .
- در صورت تاخیر در ارسال پروژه و پر شدن 10 روز مجاز تاخیر نمره پروژه برای شما در نظر گرفته نخواهد شد.
- انجام تمرین تک نفره است. لطفاً به تنهایی انجام شود، در غیر اینصورت نمره منفی در نظر گرفته خواهد شد.
- زبان برنامه نویسی دلخواه است. (پیشنهاد: پایتون)
- موارد ارسال شده در تاریخی که بعداً مشخص میشود به صورت آنلاین نیز تحویل گرفته خواهند شد (صرفاً آنچه در کوئرا تحویل داده شده است بعداً به صورت آنلاین تست شده و توضیح داده میشود) .
- کل محتوای ارسالی را داخل فایل زیپ قرار داده و نام آن را شماره دانشجویی قرار دهید.

آیدی تلگرام طراحان :

@iAmMafhoot

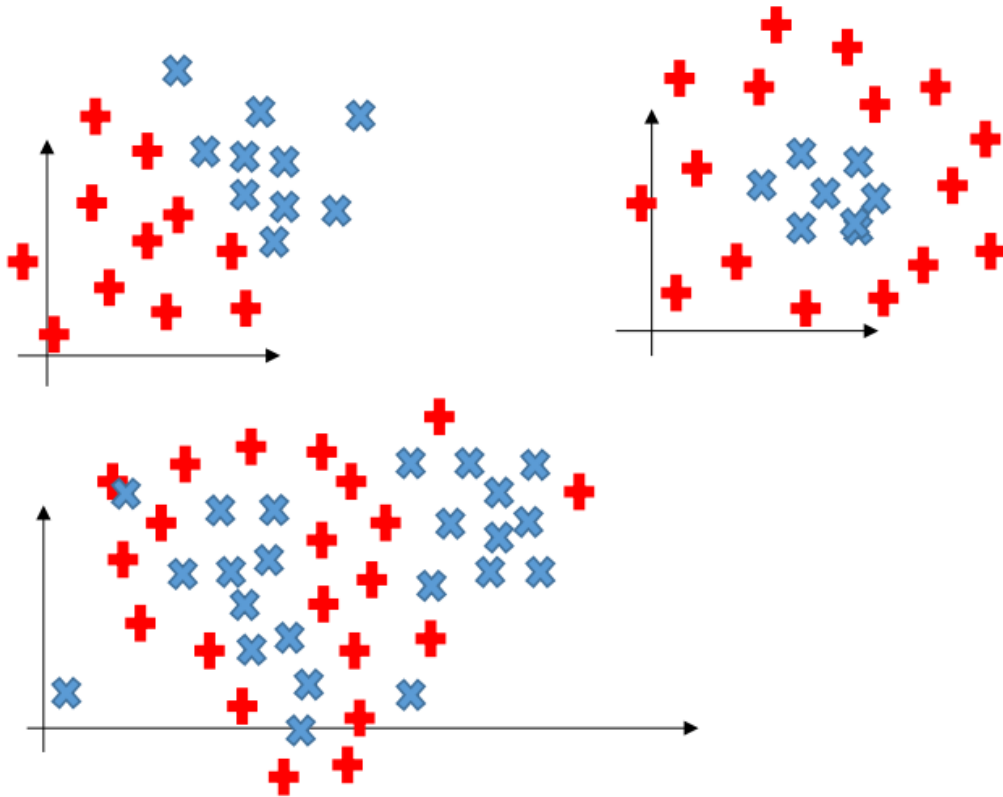
@Neg_hagh

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 100 تا 1000 نقطه) در فضای دو بعدی در نظر بگیرید و برخی از نقاط را عضو کلاس 1 و برخی دیگر را عضو کلاس 1- فرض کنید. نقاط را در یک نمودار دو بعدی رسم کنید. مثلا شکل های زیر میتوانند 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 های 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 میگذارند.

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

بخش ۱۲:

همان پایگاه داده ای که برای بخش پنجم پروژه شبکه عصبی به کار گرفته اید (مثلا USPS) را در این بخش با کمک SVM دسته بندی کنید. کدام هسته و با چه پارامترهایی جواب بهتری می دهد؟ آیا می توانید مقایسه ای در این پایگاه داده با شبکه عصبی داشته باشید؟²

بخش ۳:

ابتدا لازم است یک پایگاه داده (Data Set) برای خود انتخاب کنید. ترجیحا پایگاه داده های تصویری انتخاب نمایید.

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؟

در این بخش لازم است برای خود پایگاه داده (Data set) انتخاب کنید که ترجیحا دوتایی باشد. به عنوان مثلا عکس هایی از گربه ها و گل ها. برای انتخاب دیتاست خود، میتوانید در سایت هایی مانند [kaggle](https://www.kaggle.com/datasets) نیز سرچی کنید و با آنها دسته بندی خود را انجام دهید.

<https://www.kaggle.com/datasets>

پاورقی:

1- در این بخش و بخش های 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 Validation Cross یا جلوگیری از بیش برازش، 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!

2- 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 ستونهايش به اندازه طول و عرض تصویر است و مقدار هر درایه در ماتریس نیز معادل شدت روشنایی پیکسل مربوطه است. شدت روشنایی معمولاً عددی بین صفر تا 255 (یا بین صفر تا 1) 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 دریاورید که اندازه آن $m \times n$ خواهد بود. یعنی مثلاً 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 $m \times n$ بعدی خواهد بود و هر تصویر تبدیل به نقطه ای $m \times n$ بعدی در این فضا شده است. البته میتوانید قبل از دادن بردارها به SVM، اقدام به نرمال سازی آنها نیز نمایید. برای پروژه شبکه عصبی هم همین توضیحات صدق میکند، اگر پایگاه داده تصویری انتخاب کرده باشید.

آنچه تحویل داده میشود:

۱- کداجرایی برنامه با توضیحات لازم برای اجرا و احتمالاً Case Test هایی برای بخش اول

۲- گزارشی کامل از مسیر انجام کار، چالشهایی که احتمالاً 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 کرده اید.