

# به نام خدا

پاسخ بخش تئوری تمرین دوم درس یادگیری ماشین

استاد درس: دکتر یاسمون برشبان

علی افشار دکتری ۱۴۰۴۱۲۲۶۱۵۱۰۱

سوال اول: تابع زیان

الف) مزیت‌های استفاده از تابع هزینه Logistic Regression چیست؟ چرا این تابع نسبت به توابع ساده‌تر مناسب‌تر است؟

پاسخ: به این علت که هر چه مقدار پیشیبینی مدل به واقعیت نزدیک‌تر باشد، مقدار تابع هزینه logistic regression به عدد صفر نزدیک‌تر می‌شود، و هرچه دورتر باشد، از صفر دورتر شده و به سمت بینهایت میل پیدا می‌کند. از طرفی از آنجایی که شامل لگاریتم است، ولگاریتم یک تابع تماماً مشتق پذیر و اکیدا Convex است، پس برای الگوریتم‌های بهینه‌سازی ای مثل کاهش گرادیان، عملکرد خوبی دارد و تضمین می‌شود که به جواب بهینه در global minima برسد.

نسبت به توابع ساده‌تر، به عنوان مثال نسبت به تابع هزینه روش دسته بندی پرسپترون، از آنجایی که پرسپترون، فقط برای نمونه‌هایی که غلط دسته بندی شده بودند، هزینه را حساب می‌کرد، پس سوگیری یا bias به سمت این نوع داده‌ها داشت، ولی logistic regression برای تمامی نمونه‌ها مقدار زیان را حساب می‌کند و این امر باعث رفع مشکل سوگیری می‌شود، به تعبیری ما اینجا برای تمام نمونه‌ها احتمالی را به دست می‌آوریم که آن نمونه به کلاس مد نظر ما

تعلق دارد، و هدف بیشینه کردن این احتمالات است، که می شود کمینه کردن تابع زیان (زیرا تابع زیان ضرب یک منفی در لگاریتم احتمال بدست آمده است)

ب) مزیت های استفاده از Hinge Loss در روش های مبتنی بر SVM چیست؟

پاسخ: این امکان را به نمونه ها می دهیم که درون margin قرار بگیرند یا حتی به اشتباه shonnd، ولی برای آنها یک جریمه یا خطا در نظر می گیریم، و سعی می کنیم که مجموع این جریمه یا خطاهای کمینه شوند (به کمک یک ضریبی). این روش باعث می شود که همیشه SVM به پاسخ برسد، همچنین نسبت به outlier و noise ها مقاوم تر شده و کمتر حساسیت نشان دهد.

چرا SVM در بسیاری از شرایط عملکرد بهتری از مدل های خطی دیگر دارد؟

پاسخ: سعی می کند بهترین خط را بین خطوط ممکن که همگی ممکن است توسط معیار هایی مثل precision و recall و accuracy یکسان عمل بکنند، در نظر بگیرد، از آن جهت که اندازه margin آن خط بیشترین مقدار ممکن باشد. همچنین در مقایسه با روش perceptron که فقط نمونه های اشتباه classify شده را در نظر می گرفت، تمام نمونه ها را برای پیدا کردن بهترین خط و همچنین support vector ها در نظر می گیرد.

چه ویژگی هایی این تابع هزینه را برای دسته بندی قدرتمند می کند؟

پاسخ: این تابع کمک می کند که ما همیشه به پاسخ برسیم، بر عکس hard-margin که تضمینی برای این اتفاق نداشت، همچنین نسبت به outlier و noise ها مقاوم تر است، و می تواند generalization یا تعمیم پذیری بهتری داشته باشد. همچنین اگر بطور کلی خود تابع هزینه SVM ممکن نباشد، می توان به کمک kernel های فضای را تغییر داد تا بتوان داده هایی که خطی تفکیک پذیر نیستند، بصورت غیر خطی و با همین SVM تفکیک شوند.

## سوال دوم: SVM

الف) توضیح دهید در رابطه با Soft Margin هر یک از حالات زیر به چه معنا است؟

پاسخ:

$\xi_i = 0 \Rightarrow$  درست classify شده و روی support vector کلاس درست یا بیرون

کلاس درست قرار دارد.

$\xi_i = 1 \Rightarrow$  دقیقاً روی خط وسط SVM قرار گرفته است.

$0 < \xi_i < 1 \Rightarrow$  درست classify شده ولی داخل margin کلاس درست قرار گرفته است.

$\xi_i > 1 \Rightarrow$  اشتباه classify شده است، البته که در margin کلاس غلط می‌تواند باشد.

ب) تاثیر پارامتر C در مساله Margin Soft به چه صورت است؟

پاسخ: هر چه C مقدار بیشتری داشته باشد، سختگیرانه‌تر عمل می‌کند، و اگر بنهایت در نظر گرفته شود با hard margin یکی می‌شود. هر چه مقدار کمتری داشته باشد، امکان خطای overlap نمونه‌های کلاس‌های مختلف با یکدیگر بیشتر شده و تابه soft تر عمل می‌کند.

ج) چرا کرنل در فرم دوگان SVM بسیار کارا عمل می‌کند؟

پاسخ: به این علت که در فرم دوگان، فقط ضرب خارجی نمونه‌ها با یکدیگر در تابع هزینه در نظر گرفته می‌شود، و پیچیدگی مساله بهینه سازی افزایش پیدا نمی‌کند. بخاطر این است که می‌توان بطور مستقیم حاصل ضرب داخلی تغییر شکل یافته را محاسبه کرد، به جای اینکه ابتدا تک تک نمونه‌ها را تغییر شکل داد و سپس ضرب داخلی را اعمال کرد، که اینگونه سریع‌تر می‌شود. همچنین اگر از فرم اصلی استفاده نمی‌شد، به این علت که ما به کمک کرنل افزایش بعد انجام میدهیم، که ممکن است این افزایش زیاد هم باشد، در آن حالت ضرب به تعداد ابعاد وابسته

بود که نسبت به فرم دوگان عملکرد ضعیف تری داشت، زیرا پیچیدگی فرم دوگان به تعداد ابعاد وابسته نیست و میتوان از ابتکاری که برای تغییر ابعاد به کمک کرنل گفته شد، استفاده کرد.  
منون از توجهتون