

## الف) تأثیر حذف یک نمونه آموزشی بر مرز تصمیم

در SVM (حالت تفکیک پذیر خطی):

مرز تصمیم ثابت می ماند مگر اینکه نمونه حذف شده یک بردار پشتیبان باشد. در SVM، مرز تصمیم تنها توسط بردارهای پشتیبان تعیین می شود. اگر نمونه حذف شده بردار پشتیبان نباشد، مرز تغییر نمی کند. در غیر این صورت، مرز ممکن است تغییر کند

در Logistic Regression:

- مرز تصمیم تغییر می کند. Logistic Regression با بیشینه کردن درست نمایی کار می کند و همه نمونه ها در تعیین پارامترهای مدل نقش دارند. بنابراین، حذف هر نمونه حتی غیر مرزی بر مرز تأثیر می گذارد.

## ب) حد بالا برای تعداد نمونه های اشتباه طبقه بندی شده

- نمونه هایی که به درستی طبقه بندی نشده اند فرمول اصلی را نقض می کنند، یعنی  $\xi$  بزرگتر از یک است. تعداد این نمونه ها از جمع  $\xi$  ها کمتر است زیرا هر نمونه با  $\xi$  بزرگتر از یک حداقل ۱ واحد در جمع مشارکت می کند. پس کران بالا همان تعداد  $\xi$  ها است که از یک بزرگترند.

## پ) نقش ضریب C در SVM

- C وزن بین ماکزیمم کردن حاشیه و مینیمم کردن خطا را کنترل می کند.

- اگر C به صفر نزدیک شود اهمیت خطاها ناچیز است. مدل حاشیه بزرگی انتخاب می کند، حتی اگر خطای آموزشی زیاد باشد.

- اگر C به سمت بینهایت برود اهمیت خطاها بسیار زیاد است. مدل سعی می کند همه نمونه ها را درست طبقه بندی کند شبیه Hard SVM.

## ت) مقایسه Logistic Regression و Hard SVM در حالت تفکیک پذیر خطی

Logistic Regression با بیشینه کردن تابع درست نمایی کار می کند و مرز تصمیم آن تحت تأثیر تمامی نمونه های آموزشی حتی نقاط دور از مرز قرار می گیرد. این مدل احتمالات طبقه بندی را ارائه می دهد و برای تفسیرپذیری مناسب است. اما ممکن است به نقاط پرت حساس باشد.

Hard SVM تنها بر روی بردارهای پشتیبان (نقاط نزدیک به مرز) تمرکز می کند و مرز تصمیم آن با ماکزیمم کردن حاشیه مارجین تعیین می شود. این مدل نسبت به نقاط دور از مرز حساسیت کمتری دارد و معمولاً در داده های تفکیک پذیر خطی، تعمیم پذیری بهتری روی داده جدید نشان می دهد.

تفاوت کلیدی:

Logistic Regression برای تخمین احتمال و مدل‌های آماری مناسب است، در حالی که Hard SVM برای دستیابی به مرز بهینه با حاشیه زیاد طراحی شده است.

### ث) مقایسه Logistic Regression و Soft SVM در حالت غیر تفکیک‌پذیر خطی

Logistic Regression با افزودن یک جمله تنظیم سعی می‌کند از اورفیت شدن جلوگیری کند. خطاها در این مدل به صورت پیوسته از طریق تابع درست‌نمایی محاسبه می‌شوند و خروجی آن احتمال تعلق به هر کلاس است. این مدل انعطاف‌پذیری بالایی دارد اما ممکن است در حضور داده‌های نویزی یا همپوشانی زیاد، عملکرد ضعیف‌تری داشته باشد.

Soft SVM با معرفی متغیرهای  $\xi_i$  (خطاهای مجاز) و پارامتر  $C$ ، توازن بین حاشیه گسترده و تعداد خطاها ایجاد می‌کند. این مدل خطاها را به صورت گسسته به شکل  $\xi_i > 1$  محاسبه می‌کند و تمرکز اصلی آن روی تعمیم‌پذیری است. با تنظیم مناسب  $C$ ، می‌توان انعطاف‌پذیری مدل را کنترل کرد.

تفاوت کلیدی:

Logistic Regression احتمال‌محور است و برای تفسیرپذیری مناسب‌تر است، در حالی که Soft SVM با تمرکز بر حاشیه بهینه، برای داده‌های با نویز یا همپوشانی بالا کارآمدتر عمل می‌کند.