# الف) تأثیر حذف یک نمونه آموزشی بر مرز تصمیم

در SVM (حالت تفکیک پذیر خطی):

مرز تصمیم ثابت می‌ماند مگر اینکه نمونه حذف‌شده یک بردار پشتیبان باشد. در SVM، مرز تصمیم تنها توسط بردارهای پشتیبان تعیین می‌شود. اگر نمونه حذف‌شده بردار پشتیبان نباشد، مرز تغییر نمی‌کند. در غیر این صورت، مرز ممکن است تغییر کند

در Logistic Regression:

- مرز تصمیم تغییر می‌کند. Logistic Regression با بیشینه‌کردن درست‌نمایی کار می‌کند و همه نمونه‌ها در تعیین پارامترهای مدل نقش دارند. بنابراین، حذف هر نمونه حتی غیر مرزی بر مرز تأثیر می‌گذارد.

# ب) حد بالا برای تعداد نمونه‌های اشتباه طبقه‌بندی‌شده

- نمونه‌هایی که به‌درستی طبقه‌بندی نشده‌اندفرمول اصلی را نقض می‌کنند، یعنی ξi بزرگتر از یک است. تعداد این نمونه‌ها از جمع ها ξi ها کمتر است زیرا هر نمونه با ξi بزرگتر از یک حداقل ۱ واحد در جمع مشارکت می‌کند. پس کران بالا همان تعداد ξi هایی است که از یک بزرگترند.

# پ) نقش ضریب C در SVM

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

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

- اگر C به سمت بینهایت برود اهمیت خطاها بسیار زیاد است. مدل سعی می‌کند همه نمونه‌ها را درست طبقه‌بندی کند شبیه Hard SVM.

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

Logistic Regression با بیشینه‌کردن تابع درست‌نمایی کار می‌کند و مرز تصمیم آن تحت تأثیر تمامی نمونه‌های آموزشی حتی نقاط دور از مرز قرار می‌گیرد. این مدل احتمالات طبقه‌بندی را ارائه می‌دهد 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 می‌کند و مرز تصمیم آن با ماکزیمم کردن حاشیه مارجین تعیین می‌شود. این مدل نسبت به نقاط دور از مرز حساسیت کمتری دارد و معمولاً در داده‌های تفکیک‌پذیر خطی، تعمیم‌پذیری بهتری روی داده جدید نشان می‌دهد.

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 ξi​ (خطاهای مجاز) و پارامتر C، توازنی بین حاشیه گسترده و تعداد خطاها ایجاد می‌کند. این مدل خطاها را به صورت گسسته به شکلξi​>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.