

# حل تمرینات شناسایی الگو - سوال ۹ (ترم ۱۴۰۲)

وحید ملکی

۱۴۰۴ آذر ۲۲

## سوال ۹ (ترم ۱۴۰۲): ماشین بردار پشتیبان خطی

داده‌های مسئله به شرح زیر هستند:

- کلاس مثبت (+):  $x_1 = (1, 1), x_2 = (2, 2), x_3 = (2, 0)$
- کلاس منفی (-):  $x_4 = (0, 0), x_5 = (1, 0), x_6 = (0, 1)$

### الف) نمایش داده‌ها در فضای دو بعدی

با رسم نقاط در صفحه مختصات دکارتی، مشاهده می‌کنیم:

- نقاط کلاس منفی (دایره‌ها) یک مثلث در ناحیه پایین و چپ (نزدیک مبدأ) تشکیل می‌دهند.
- نقاط کلاس مثبت (ضربرها) در ناحیه بالا و راست قرار دارند.

### ب) تفکیک‌پذیری خطی

بله، این دو دسته به صورت خطی تفکیک‌پذیر هستند. اگر محدب‌ترین محیط (Convex Hull) هر دو دسته را رسم کنیم، هیچ تداخلی با هم ندارند و می‌توان یک خط مستقیم بین آن‌ها عبور داد (مثلاً خطی که از وسط فاصله نقاط  $(0, 1)$  و  $(1, 1)$  بگذرد).

### ج) محاسبه بردار وزن $w$ و بایاس $b$ (به صورت نظری)

هدف SVM یافتن عریض‌ترین باند بین دو دسته است. باید خطوط مرزی را بررسی کنیم:

۰.۱ اگر به مختصات دقت کنیم، مجموع مولفه‌ها  $(x_1 + x_2)$  الگوی خوبی بدست می‌دهد.

۰.۲ برای کلاس منفی:

$$(0, 0) \rightarrow 0, \quad (1, 0) \rightarrow 1, \quad (0, 1) \rightarrow 1$$

بیشترین مقدار مجموع برای کلاس منفی برابر ۱ است. پس خط مرزی کلاس منفی  $x_1 + x_2 = 1$  است.

۰.۳ برای کلاس مثبت:

$$(1, 1) \rightarrow 2, \quad (2, 2) \rightarrow 4, \quad (2, 0) \rightarrow 2$$

کمترین مقدار مجموع برای کلاس مثبت برابر ۲ است. پس خط مرزی کلاس مثبت  $x_1 + x_2 = 2$  است.

بنابراین مرز تصمیم‌گیری دقیقاً وسط این دو خط است ( $x_1 + x_2 = 1.5$ ). اما برای بدست آوردن  $w$  و  $b$  استاندارد (که در آن  $w^T x + b = \pm 1$ ):

$$w_1 x_1 + w_2 x_2 + b = +1 \quad \text{برای خط } x_1 + x_2 = 2$$

$$w_1 x_1 + w_2 x_2 + b = -1 \quad \text{برای خط } x_1 + x_2 = 1$$

از تفاضل دو معادله بالا داریم:

$$w_1(x_1 - x'_1) + w_2(x_2 - x'_2) = 2$$

با توجه به تقارن و شیب خطوط ( $x_1 + x_2 = cte$ ، شیب  $-1$ ، مولفه‌های وزن برابرند  $w_1 = w_2 = w$ ). تفاوت مقدار مجموع  $x_1 + x_2$  در دو خط برابر ۱ است، اما تفاوت خروجی باید ۲ باشد. پس باید ضرایب را در ۲ ضرب کنیم:

$$2(x_1 + x_2) + b' = \text{output}$$

بررسی می‌کنیم:

$$w = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

برای پیدا کردن  $b$ : برای بردار پشتیبان کلاس منفی  $(1, 0)$  و کلاس مثبت  $(0, 1)$ :

پاسخ نهایی:

$$\mathbf{w} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad b = -3$$

معادله خط جداکننده:  $.2x_1 + 2x_2 - 3 = 0$

#### د) بردارهای پشتیبان و مقدار حاشیه تمایز

بردارهای پشتیبان: نقاطی که روی خطوط مرزی  $w^T x + b = \pm 1$  قرار دارند:

- از کلاس مثبت:  $(1, 1)$  و  $(2, 0)$  (چون برای هر دو  $2x_1 + 2x_2 - 3 = +1$  می‌شود).
- از کلاس منفی:  $(0, 1)$  و  $(1, 0)$  (چون برای هر دو  $2x_1 + 2x_2 - 3 = -1$  می‌شود).

مقدار حاشیه ( $M$ ): فرمول حاشیه برابر است با  $\frac{2}{\|w\|}$ .

$$\|w\| = \sqrt{2^2 + 2^2} = \sqrt{8} = 2\sqrt{2}$$

$$M = \frac{2}{2\sqrt{2}} = \frac{1}{\sqrt{2}} \approx 0.707$$

#### ه) حذف یکی از بردارهای پشتیبان و تغییر حاشیه

خیر، حاشیه افزایش نمی‌یابد. چرا؟ زیرا در این مسئله خاص، بردارهای پشتیبان "افزونه" (Redundant) وجود دارد. خطوط مرزی توسط دو نقطه در هر طرف تعریف شده‌اند. مثلاً مرز کلاس مثبت ( $x_1 + x_2 = 2$ ) هم توسط  $(1, 1)$  و هم توسط  $(2, 0)$  پشتیبانی می‌شود. اگر  $(1, 1)$  را حذف کنیم، نقطه  $(2, 0)$  همچنان وجود دارد و اجازه نمی‌دهد خط مرزی عقب تر برود. همین شرایط برای کلاس منفی هم صادق است.

## و) حذف کدام بردار موجب افزایش حاشیه می شود؟

در این آرایش خاص داده‌ها، حذف هیچ‌یک از بردارهای پشتیبان به تنها موجب افزایش حاشیه نمی‌شود. برای افزایش حاشیه باید تمام نقاطی که روی یک خط مرزی قرار دارند (مثلاً هر دو نقطه  $(1, 1)$  و  $(2, 0)$ ) حذف شوند تا مرز بتواند جابجا شود.

## ز) اضافه کردن نمونه برچسب‌دار و کاهش حاشیه

بله، حاشیه کاهش می‌یابد. چرا؟ اگر نمونه جدیدی اضافه شود که در ناحیه حاشیه فعلی (ناحیه بین دو خط  $2x_1 + 2x_2 - 3 = \pm 1$ ) قرار گیرد، SVM مجبور می‌شود مرزهای جدیدی انتخاب کند تا آن نقطه را درست دسته‌بندی کند. این کار باعث می‌شود فاصله بین خطوط مرزی (حاشیه) کمتر شود.

## ح) تغییر در تعداد بردارهای پشتیبان با اضافه کردن داده

تعداد بردارهای پشتیبان ممکن است افزایش، کاهش یا بدون تغییر بماند (بسته به موقعیت داده).

- افزایش: اگر داده جدید دقیقاً روی مرز حاشیه فعلی بیفتد (مثلاً نقطه  $(1.5, 0.5)$  با کلاس مثبت)، به مجموعه بردارهای پشتیبان اضافه می‌شود.
- کاهش: اگر داده جدیدی درون حاشیه و بسیار نزدیک به مرز تصمیم‌گیری اضافه شود (مثلاً  $(1.1, 0.4)$  با کلاس منفی)، مرزها تغییر جهت می‌دهند. در مرز جدید، ممکن است برخی از بردارهای پشتیبان قبلی (مثل  $(2, 0)$ ) دیگر نزدیک ترین نقطه به مرز نباشند و از لیست پشتیبان‌ها خارج شوند.
- بدون تغییر: اگر داده جدید خیلی دور از مرزها باشد (مثلاً  $(10, 10)$ )، هیچ تأثیری ندارد.