

تکلیف هفتم درس شناسایی الگو

وحید ملکی

شماره دانشجویی: ۴۰۳۱۳۰۰۴

۱۱ آذر ۱۴۰۴

سؤال ۱

الف) اثبات فرمول بردار وزن فیشر:

هدف در تحلیل افتراقی فیشر، پیدا کردن بردار وزن w است که تابع هزینه زیر را بیشینه کند:

$$J_F(w) = \frac{|w^T(\hat{m}_1 - \hat{m}_2)|^2}{w^T \hat{\Sigma}_w w}$$

برای سادگی نمادگذاری، ماتریس پراکندگی بین کلاسی را $S_B = (\hat{m}_1 - \hat{m}_2)(\hat{m}_1 - \hat{m}_2)^T$ و ماتریس پراکندگی درون کلاسی را $S_W = \hat{\Sigma}_w$ می‌نامیم. تابع هزینه به فرم ریلی زیر بازنویسی می‌شود:

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

برای یافتن نقطه بیشینه، از $J(w)$ نسبت به w مشتق گرفته و برابر صفر قرار می‌دهیم. با استفاده از قاعده مشتق خارج قسمت $(\frac{u}{v})$:

$$\nabla_w J(w) = \frac{(2S_B w)(w^T S_W w) - (2S_W w)(w^T S_B w)}{(w^T S_W w)^2} = 0$$

چون مخرج کسر مخالف صفر است، صورت باید صفر شود:

$$(S_B w)(w^T S_W w) = (S_W w)(w^T S_B w)$$

عبارت $\lambda = \frac{w^T S_B w}{w^T S_W w}$ همان مقدار تابع هزینه (یک عدد اسکالر) است. پس می‌توان نوشت:

$$S_B w = \lambda S_W w$$

از طرفی می‌دانیم $S_B w = (\hat{m}_1 - \hat{m}_2)(\hat{m}_1 - \hat{m}_2)^T w$. عبارت $(\hat{m}_1 - \hat{m}_2)^T w$ نیز یک عدد اسکالر است. پس جهت بردار $S_B w$ هم‌راستا با بردار تفاضل میانگین‌ها $(\hat{m}_1 - \hat{m}_2)$ است. بنابراین:

$$\alpha(\hat{m}_1 - \hat{m}_2) = \lambda S_W w$$

که در آن α و λ ضرایب ثابت هستند. با ضرب معکوس ماتریس S_W از چپ:

$$w = \frac{\alpha}{\lambda} S_W^{-1}(\hat{m}_1 - \hat{m}_2)$$

از آنجا که در دسته‌بندی خطی تنها جهت بردار w اهمیت دارد، می‌توانیم از ضرایب ثابت صرف‌نظر کرده و بنویسیم:

$$w \propto \hat{\Sigma}_w^{-1}(\hat{m}_1 - \hat{m}_2)$$

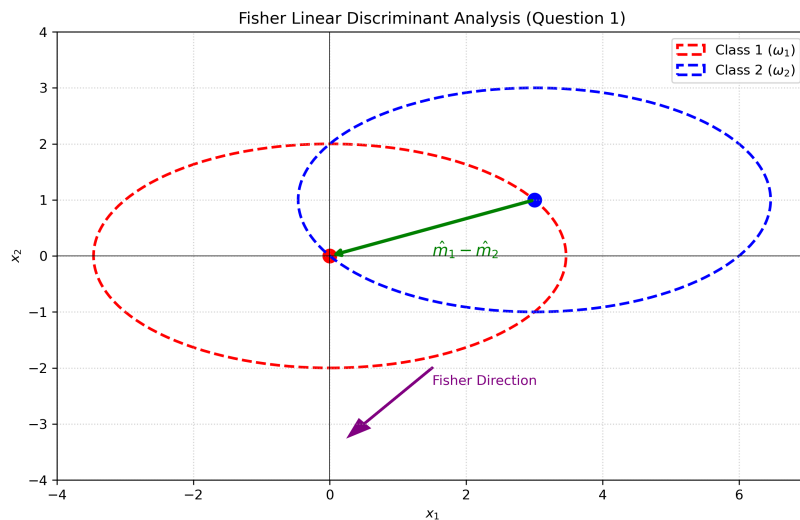
ب) نمایش هندسی و محاسبات:
داده‌های مسئله عبارتند از:

$$\hat{m}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \hat{m}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \hat{\Sigma}_1 = \hat{\Sigma}_2 = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

چون تعداد نمونه‌ها برابر است، ماتریس پراکندگی درون کلاسی برابر با کواریانس کلاس‌هاست: $\hat{\Sigma}_w = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$
بردار واصل مراکز برابر است با:

$$\hat{m}_1 - \hat{m}_2 = \begin{bmatrix} -3 \\ -1 \end{bmatrix}$$

شکل زیر مراکز ثقل دو کلاس، بیضی‌های هم‌چگال (که نشان‌دهنده توزیع گوسی هستند) و بردار واصل میانگین‌ها را نشان می‌دهد.



شکل ۱: نمایش هندسی دو کلاس و بردار واصل مراکز. بیضی‌ها نشان‌دهنده پراکندگی داده‌ها هستند.

ج) چرا بردار واصل مراکز انتخاب مناسبی نیست؟

همانطور که در شکل مشاهده می‌شود، بردار واصل مراکز $(\hat{m}_1 - \hat{m}_2)$ جهت تقریباً افقی دارد (مؤلفه x آن بزرگتر است). با این حال، بیضی‌های کواریانس نشان می‌دهند که پراکندگی داده‌ها در جهت افقی بسیار زیاد است (واریانس ۳)، در حالی که در جهت عمودی پراکندگی کمی دارند (واریانس ۱). اگر داده‌ها را روی بردار واصل مراکز تصویر کنیم، به دلیل واریانس زیاد در آن جهت، هم‌پوشانی (Overlap) زیادی بین دو کلاس رخ می‌دهد و خطای دسته‌بندی بالا می‌رود. بنابراین صرفاً توجه به فاصله مراکز کافی نیست و باید پراکندگی را نیز لحاظ کرد.

د) نقش ضرب ماتریس $\hat{\Sigma}_w^{-1}$:

ماتریس کواریانس داده شده قطری است، بنابراین مقادیر ویژه آن $\lambda_1 = 3$ (در راستای افقی) و $\lambda_2 = 1$ (در راستای عمودی) هستند. ماتریس معکوس کواریانس $\hat{\Sigma}_w^{-1}$ به صورت زیر است:

$$\hat{\Sigma}_w^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix}$$

وقتی این ماتریس در بردار واصل ضرب می‌شود:

$$w \propto \begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

ملاحظه می‌کنیم که مؤلفه افقی بردار (که در جهت واریانس زیاد بود) با ضریب $1/3$ تضعیف شده، اما مؤلفه عمودی (که در جهت واریانس کم بود) با ضریب 1 حفظ شده است. این ضرب باعث می‌شود جهت بردار وزن از حالت افقی به حالت مورب (45° درجه) تغییر کند که در آن راستا، هم فاصله مراکز مناسب است و هم پراکندگی داده‌ها کمتر است، و در نتیجه تفکیک بهتری حاصل می‌شود.