تحلیل جداکننده‌ی خطی

Linear Discriminant Analysis

مجموعه داده‌ی $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ و تعداد k برچسب $y_i \in \{c_1, \dots, c_k\}$ متناظر هر نقطه $\boldsymbol{x}_i \in \mathbb{R}^d$ را در نظر می‌گیریم.

$$\boldsymbol{D}_i = \{\boldsymbol{x}_j^T | y_j = c_i\}, \qquad n_i = |\boldsymbol{D}_i|$$

برای سادگی فرض می‌کنیم $k = 2$ است و مجموعه داده‌ها تنها از دو دسته تشکیل شده‌اند. هدف پیدا کردن بردار $\boldsymbol{w}$ است که با تصویر کردن داده‌ها روی آن بیشترین جدایش بین دو دسته مشاهده شود.

همچنین طول بردار را واحد در نظر می‌گیریم $\boldsymbol{w}^T\boldsymbol{w} = 1$ و سپس بردار تصویر هر نقطه روی این بردار

$$\boldsymbol{x}_i' = \left(\frac{\boldsymbol{w}^T\boldsymbol{x}_i}{\boldsymbol{w}^T\boldsymbol{w}}\right)\boldsymbol{w} = (\boldsymbol{w}^T\boldsymbol{x}_i)\boldsymbol{w} = a_i\boldsymbol{w}$$

می‌شود. مقدار $a_i = \boldsymbol{w}^T\boldsymbol{x}_i$ طول سایه روی محور $\boldsymbol{w}$ می‌باشد. پس مجموعه‌ی $\{a_1, \dots, a_n\}$ نگاشت از $\mathbb{R}^d$ به $\mathbb{R}$ می‌باشد.

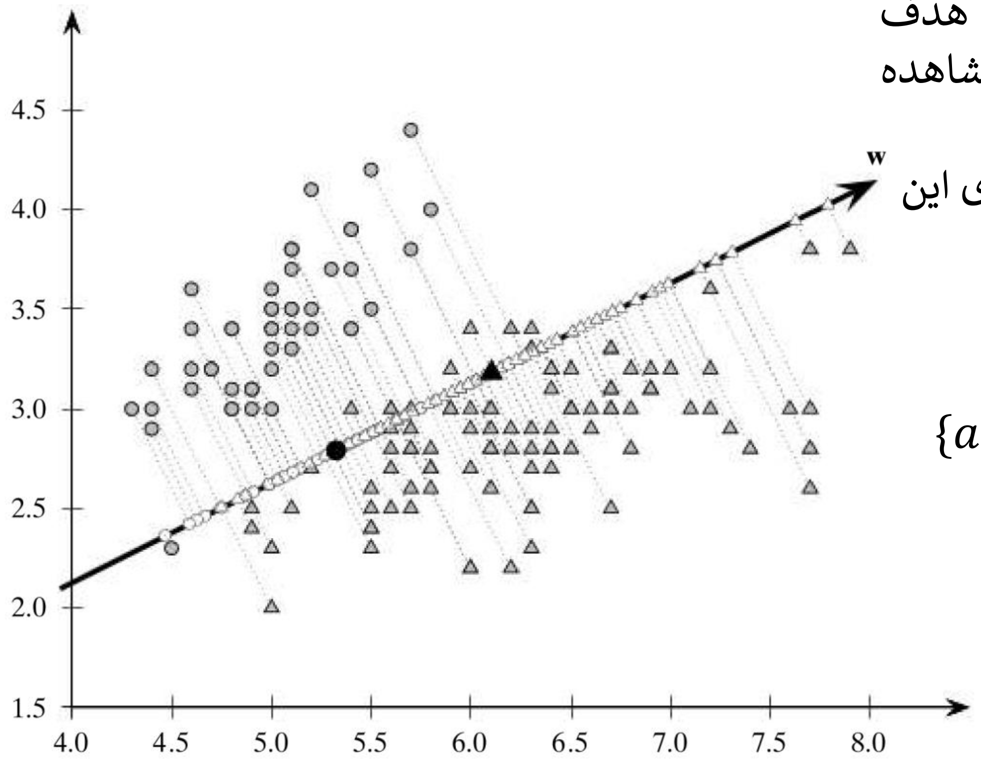

Figure 20.1. Projection onto w.

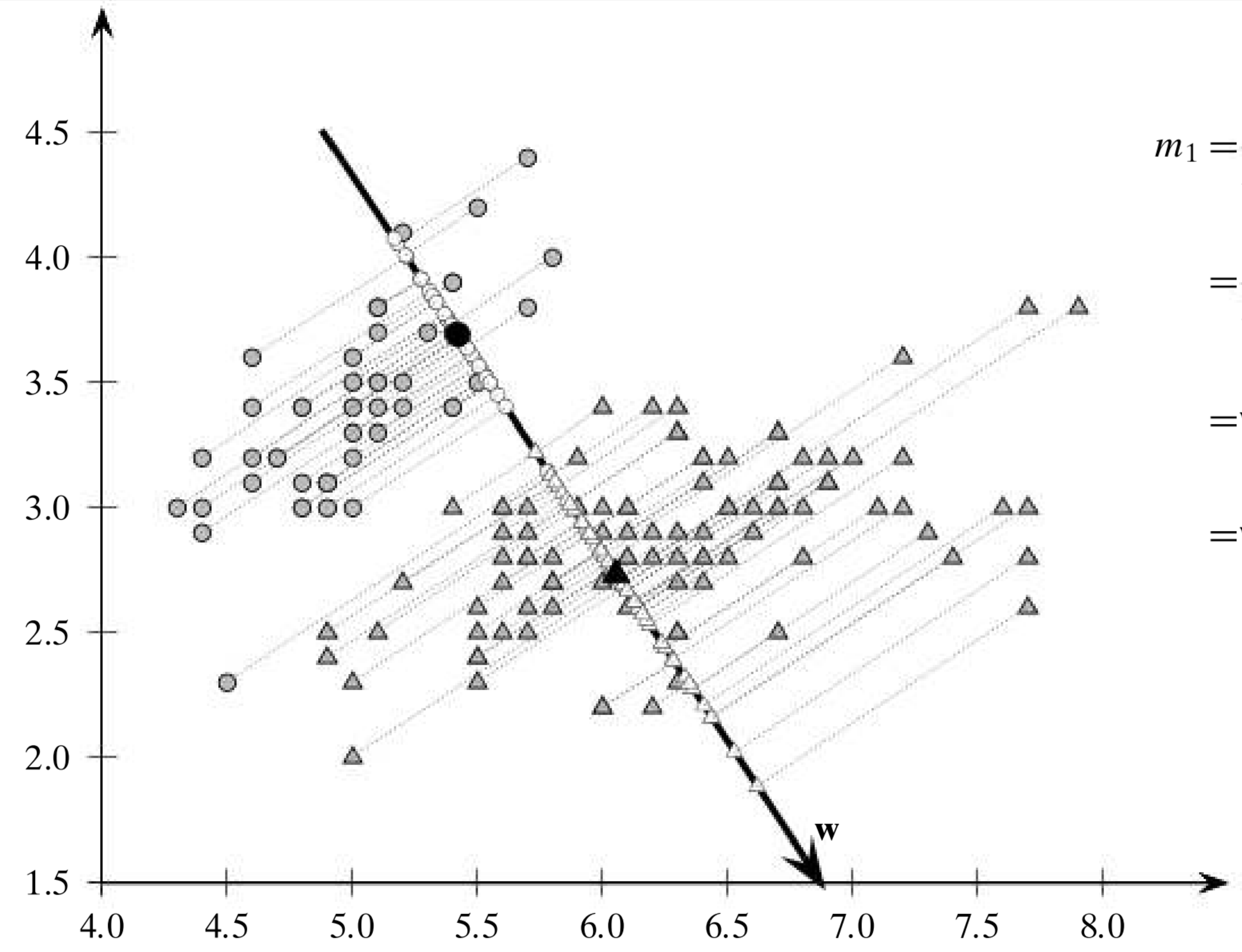

Figure 20.2. Linear discriminant direction w.

میانگین (m_i) و پراکندگی (s_i^2) سایه‌ها روی بردار $\boldsymbol{w}$ را محاسبه می‌کنیم.

$$m_1 = \frac{1}{n_1}\sum_{\mathbf{x}_i \in \mathbf{D}_1} a_i = \frac{1}{n_1}\sum_{\mathbf{x}_i \in \mathbf{D}_1} \mathbf{w}^T\mathbf{x}_i = \mathbf{w}^T\left(\frac{1}{n_1}\sum_{\mathbf{x}_i \in \mathbf{D}_1} \mathbf{x}_i\right) = \mathbf{w}^T\boldsymbol{\mu}_1$$

$$m_2 = \mathbf{w}^T\boldsymbol{\mu}_2$$

$$s_i^2 = \sum_{\mathbf{x}_j \in \mathbf{D}_i} (a_j - m_i)^2 = n_i\sigma_i^2$$

بردار مورد نظر راستایی است که میانگین کلاس‌ها بیشترین فاصله از هم و پراکندگی کلاس‌ها کمترین مقدار باشد.

$$\boxed{\max_{\mathbf{w}} \; J(\mathbf{w}) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}}$$

با دو روش بردار $\boldsymbol{w}$ را می‌توانیم محاسبه کنیم.

$$\boxed{\max_{\mathbf{w}} \; J(\mathbf{w}) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}}$$

$$\begin{aligned}(m_1 - m_2)^2 &= \left(\mathbf{w}^T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right)^2 \\ &= \mathbf{w}^T\left((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\right)\mathbf{w} \\ &= \mathbf{w}^T\mathbf{B}\mathbf{w}\end{aligned}$$

$$\mathbf{B} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T$$

$$\begin{aligned}s_1^2 &= \sum_{\mathbf{x}_i \in \mathbf{D}_1} (a_i - m_1)^2 \\ &= \sum_{\mathbf{x}_i \in \mathbf{D}_1} (\mathbf{w}^T\mathbf{x}_i - \mathbf{w}^T\boldsymbol{\mu}_1)^2 \\ &= \sum_{\mathbf{x}_i \in \mathbf{D}_1} \left(\mathbf{w}^T(\mathbf{x}_i - \boldsymbol{\mu}_1)\right)^2 \\ &= \mathbf{w}^T\left(\sum_{\mathbf{x}_i \in \mathbf{D}_1} (\mathbf{x}_i - \boldsymbol{\mu}_1)(\mathbf{x}_i - \boldsymbol{\mu}_1)^T\right)\mathbf{w} \\ &= \mathbf{w}^T\mathbf{S}_1\mathbf{w}\end{aligned}$$

$$\mathbf{S}_i = n_i\,\boldsymbol{\Sigma}_i$$

$$\boxed{\max_{\mathbf{w}} \; J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{B}\mathbf{w}}{\mathbf{w}^T\mathbf{S}\mathbf{w}}}$$

$$\frac{d}{d\mathbf{w}}J(\mathbf{w}) = \frac{2\mathbf{B}\mathbf{w}(\mathbf{w}^T\mathbf{S}\mathbf{w}) - 2\mathbf{S}\mathbf{w}(\mathbf{w}^T\mathbf{B}\mathbf{w})}{(\mathbf{w}^T\mathbf{S}\mathbf{w})^2} = \mathbf{0}$$

$$\mathbf{B}\,\mathbf{w} = J(\mathbf{w})\mathbf{S}\mathbf{w}$$

$$\mathbf{B}\mathbf{w} = \lambda\mathbf{S}\mathbf{w}$$

If $\mathbf{S}$ is *nonsingular*, that is, if $\mathbf{S}^{-1}$ exists

$$\boxed{(\mathbf{S}^{-1}\mathbf{B})\mathbf{w} = \lambda\mathbf{w}}$$

$$\begin{aligned}\mathbf{B}\mathbf{w} &= \left((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\right)\mathbf{w} \\ &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\left((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\mathbf{w}\right) \\ &= b(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\end{aligned}$$

$$\mathbf{B}\mathbf{w} = \lambda\mathbf{S}\mathbf{w}$$

$$b(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \lambda\mathbf{S}\mathbf{w}$$

$$\mathbf{w} = \frac{b}{\lambda}\mathbf{S}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

$$\boxed{\mathbf{w} = \mathbf{S}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}$$

Example 20.2 (Linear Discriminant Analysis). Consider the 2-dimensional Iris data (with attributes `sepal length` and `sepal width`) shown in Example 20.1. Class c_1, corresponding to `iris-setosa`, has $n_1 = 50$ points, whereas the other class c_2 has $n_2 = 100$ points. The means for the two classes c_1 and c_2, and their difference is given as

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 5.01 \\ 3.42 \end{pmatrix} \qquad \boldsymbol{\mu}_2 = \begin{pmatrix} 6.26 \\ 2.87 \end{pmatrix} \qquad \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \begin{pmatrix} -1.256 \\ 0.546 \end{pmatrix}$$

The between-class scatter matrix is

$$\mathbf{B} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T = \begin{pmatrix} -1.256 \\ 0.546 \end{pmatrix} \begin{pmatrix} -1.256 & 0.546 \end{pmatrix} = \begin{pmatrix} 1.587 & -0.693 \\ -0.693 & 0.303 \end{pmatrix}$$

and the within-class scatter matrix is

$$\mathbf{S}_1 = \begin{pmatrix} 6.09 & 4.91 \\ 4.91 & 7.11 \end{pmatrix} \qquad \mathbf{S}_2 = \begin{pmatrix} 43.5 & 12.09 \\ 12.09 & 10.96 \end{pmatrix} \qquad \mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2 = \begin{pmatrix} 49.58 & 17.01 \\ 17.01 & 18.08 \end{pmatrix}$$

$\mathbf{S}$ is nonsingular, with its inverse given as

$$\mathbf{S}^{-1} = \begin{pmatrix} 0.0298 & -0.028 \\ -0.028 & 0.0817 \end{pmatrix}$$

Therefore, we have

$$\mathbf{S}^{-1}\mathbf{B} = \begin{pmatrix} 0.0298 & -0.028 \\ -0.028 & 0.0817 \end{pmatrix} \begin{pmatrix} 1.587 & -0.693 \\ -0.693 & 0.303 \end{pmatrix} = \begin{pmatrix} 0.066 & -0.029 \\ -0.100 & 0.044 \end{pmatrix}$$

The direction of most separation between c_1 and c_2 is the dominant eigenvector corresponding to the largest eigenvalue of the matrix $\mathbf{S}^{-1}\mathbf{B}$. The solution is

$$J(\mathbf{w}) = \lambda_1 = 0.11$$

$$\mathbf{w} = \begin{pmatrix} 0.551 \\ -0.834 \end{pmatrix}$$

Example 20.3. Continuing Example 20.2, we can directly compute $\mathbf{w}$ as follows:

$$\mathbf{w} = \mathbf{S}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$
$$= \begin{pmatrix} 0.066 & -0.029 \\ -0.100 & 0.044 \end{pmatrix} \begin{pmatrix} -1.246 \\ 0.546 \end{pmatrix} = \begin{pmatrix} -0.0527 \\ 0.0798 \end{pmatrix}$$

After normalizing, we have

$$\mathbf{w} = \frac{\mathbf{w}}{\|\mathbf{w}\|} = \frac{1}{0.0956} \begin{pmatrix} -0.0527 \\ 0.0798 \end{pmatrix} = \begin{pmatrix} -0.551 \\ 0.834 \end{pmatrix}$$

Note that even though the sign is reversed for $\mathbf{w}$, compared to that in Example 20.2, they represent the same direction; only the scalar multiplier is different.