Linear Discriminant Analysis

K classes, $x \in \mathbb{R}^d$

$\downarrow$

$y_i \in \{1, 2, \cdots k\}$

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

w: projection vector

S_B: between-class scatter matrix

S_W: within-class scatter matrix

$$S_W = \sum_{k=1} \sum_{i \in C_k} (x_i - \mu_k)(x_i - \mu_k)^T$$

$$S_B = \sum_{k=1} \#|C_k| \cdot (\mu_k - \mu)(\mu_k - \mu)^T$$

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_n$$

then let's using w for maximize

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

set $w^T S_W w = 1$ as a constraint

then $\mathcal{L}(w, \lambda) = w^T S_B w - \lambda(w^T S_W w - 1)$

trying to maximize λ here

$$\Rightarrow \quad \frac{\partial \mathcal{L}}{\partial w} = 0 \qquad \frac{\mathcal{L}(w+dw,\lambda) - \mathcal{L}(w,\lambda)}{dw}$$

$$\Leftrightarrow \frac{\left[(w+dw)^T S_B (w+dw) - \lambda((w+dw)^T S_W (w+dw) - 1) \right. \left. - w^T S_B w + \lambda(w^T S_W w - 1) \right.}{dw}$$

scalar

$$\Leftrightarrow \frac{dw^T S_B w + w^T S_B dw - \lambda\, dw^T S_W w - \lambda\, w^T S_W dw}{dw}$$

$$\Leftrightarrow \frac{w^T S_B^T dw + w^T S_B dw - \lambda\, w^T S_W^T dw - \lambda\, w^T S_W dw}{dw}$$

$$\Leftrightarrow (w^T S_B^T + w^T S_B) - \lambda w^T S_W^T - \lambda w^T S_W$$

$$w^T (S_B^T - \lambda S_W^T) + w^T (S_B - \lambda S_W) = \vec{0}$$

$\Leftrightarrow \quad w^T(S_B^T + S_B - \lambda S_w^T - \lambda S_w) = \vec{0}$.

$\Leftrightarrow \quad w^T(S_B^T + S_B - \lambda S_w^T - \lambda S_w)w = 0$

$\Leftrightarrow \quad w^T S_B^T w + w^T S_B w - \lambda w^T S_w^T w - \lambda w^T S_w w = 0$

since $w^T S_w w = 1$ also $w^T S_w^T w = 1$.

$\Leftrightarrow \quad w^T S_B w = \lambda w^T S_w w \quad \ldots$

$\Leftrightarrow \quad w^T(S_B w - \lambda S_w w) = 0$

only when $S_B w = \lambda S_w w$

then $\lambda = S_B w (S_w w)^{-1}$

$= S_B w w^{-1} S_w^{-1}$

$= S_B S_w^{-1}$ weird!

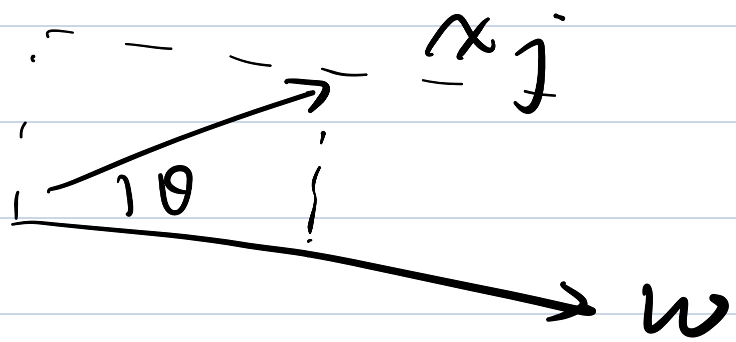
If we see: $S_B S_w^{-1}$ as A:

then $\lambda w = A w = S_B S_w^{-1} w$

then compute the eigenvector, and do projections.

lemma: let $x_j \in \mathbb{R}^d$ from class 1

$w \in \mathbb{R}^d$, the projection vector:



$$\frac{x_j \cdot w}{w \cdot w} w_j$$

$$\cos\theta = \frac{x_j \cdot w}{\|x_j\| \, \|w\|}$$

$$\|x_j\| \cos\theta \frac{w}{\|w\|} = \frac{x_j \cdot w}{\|w\| \cdot \|w\|} w$$

The length is $w^T x_j = z_j$

$$Var(z) = \frac{1}{n} \sum (z_j - \mu_z)^2$$

$$\mu_z = \frac{1}{n} \sum z_j$$

$$= \frac{1}{n} w^T (x_1 + \cdots + x_n)$$

$$= w^T \mu_x$$

$$\Rightarrow Var(z) = \frac{1}{n} \sum (w^T x_j - w^T \mu_x)^2$$

$$= \frac{1}{n} \sum (w^T (x_j - \mu_x)^2$$

$$= \frac{1}{n} \sum \left[(w^T (x_j - \mu_x))^T w^T (x_j - \mu_x) \right]$$

$$= \frac{1}{n} \sum \left[w^T (x_j - \mu_x)(x_j - \mu_x)^T w \right]$$

$$= w^T \left[\frac{1}{n} \sum (x_j - \mu_x)(x_j - \mu_x)^T \right] w$$

$$= w^T \left[\hat{P}_1 \hat{\sigma}_{x_1}^2 + \cdots + \hat{P}_2 \hat{\sigma}_{x_2}^2 \right] w$$

$$SNR(w) = \frac{(\hat{\mu}_{z_2} - \hat{\mu}_{z_1})^2}{(\hat{P}_1 \hat{\sigma}_{z_1}^2 + \hat{P}_2 \hat{\sigma}_{z_2}^2)}$$

for projection on w!

$$= \frac{(w^T(\hat{\mu}_{x_2} - \hat{\mu}_{x_1}))^2}{w^T \hat{S}_{x.avg} w}$$

$$W_{LDA} := \underset{w \in \mathbb{R}^d}{\operatorname{argmax}} \; SNR(w)$$

$$= \underset{w \in \mathbb{R}^d}{\operatorname{argmax}} \frac{(w^T(\hat{\mu}_{x_2} - \hat{\mu}_{x_1}))^2}{w^T \hat{S}_{x.avg} w}$$

solve for W_{LDA}

$$\frac{\partial \dfrac{(w^T(\hat{\mu}_{x_2} - \hat{\mu}_{x_1}))^2}{w^T \hat{S}_{x\,avg}\, w}}{\partial w} = 0$$

$\Rightarrow$ subject $w^T \hat{S}_{x\,avg}\, w = 1$

usng Lagrange multiplier:

$$\mathcal{L}(w, \lambda) = w^T(\mu_{x_2} - \mu_{x_1})(\mu_{x_2} - \mu_{x_1})^T w - \lambda\,(w^T \hat{S}_{x_{avg}}\, w - 1)$$

$$\frac{\partial \mathcal{L}(w, \lambda)}{\partial w} = \frac{\mathcal{L}(w + dw, \lambda) - \mathcal{L}(w, \lambda)}{dw}$$

$$= dw^T(\mu_{x_2} - \mu_{x_1})(\mu_{x_2} - \mu_{x_1})^T w$$

$$+ w^T(\mu_{x_2}-\mu_{x_1})(\mu_{x_2}-\mu_{x_1})^T dw$$

$$- \lambda\, dw^T \hat{S}_{x_{avg}} w - \lambda w^T \hat{S}_{x_{avg}} dw$$

$$\overline{\qquad\qquad\qquad\qquad\qquad\qquad}$$

$$dw$$

$$= \frac{w^T(\mu_{x_2}-\mu_{x_1})(\mu_{x_2}-\mu_{x_1})^T dw + w^T(\mu_{x_2}-\mu_{x_1})(\mu_{x_2}-\mu_{x_1})^T dw - \lambda w^T \hat{S}_{x_{avg}} dw - \lambda w^T \hat{S}_{x_{avg}} dw}{dw}$$

$$(=)\ w^T(\mu_{x_2}-\mu_{x_1})(\mu_{x_2}-\mu_{x_1})^T$$

$$= \lambda w^T \hat{S}_{x_{avg}}$$

$\Leftrightarrow (\mu_{x_2}-\mu_{x_1})(\mu_{x_2}-\mu_{x_1})^T w$

$= \lambda \hat{S}_{x_{avg}}^T w$

$\hat{S}_{x_{avg}}$ is a symmetric matix

also avg within-class-cv S_w

$\Leftrightarrow$

$(\mu_{x_2}-\mu_{x_1})\underbrace{(\mu_{x_2}-\mu_{x_1})^T w}_{\text{Scalar}} = \lambda S_{x_{avg}} w$

$\Rightarrow$ set $\alpha = (\mu_{x_2}-\mu_{x_1})^T w$

$\Rightarrow (\mu_{x_2}-\mu_{x_1})\underline{\underline{\alpha}} = \lambda S_{x_{\underline{\underline{avg}}}} w$

Scalar Constant matrix

$\Rightarrow$ w should be scalar of

$(\mu_{x_2} - \mu_{x_1})$

$$\Rightarrow w = S_{x_{avg}}^{-1} \frac{1}{\lambda} (\mu_{x_2} - \mu_{x_1}) \alpha$$

$\Rightarrow$ optimal direction for w

is $S_{x_{avg}}^{-1} (\mu_{x_2} - \mu_{x_1})$

$$W_{LDA} = S_{x_{avg}}^{-1} (\mu_{x_2} - \mu_{x_1})$$