LINEAR DISCRIMINANT ANALYSIS

- $x \in \mathbb{R}^d$; x IS A COLUMN VECTOR (DATA)

  TOTAL NUMBER OF CLASSES OF DATA POINTS $= K \ (\geq 2)$

  THE DIFFERENT CLASSES ARE DENOTED BY:

  $C_1, C_2, \ldots, C_K$.

  $\mathcal{X}$ = SET OF DATA POINTS

  $= \{ (x^t, r^t) \mid x^t \in \mathbb{R}^d$; r^t IS A 0-1 VECTOR OF SIZE K;

  $r_i^t = 1$, IF $x^t \in C_i$; AND 0 OTHERWISE;

  $1 \leq i \leq K$; $1 \leq t \leq n \}$

  $r^t = (\underset{1}{0}, \underset{2}{0}, \ldots, \underset{i}{1}, 0, \ldots, \underset{K}{0}) \Rightarrow$ INDICATES THAT DATA POINT x^t BELONGS TO CLASS C_i

- ASSUME THAT $K = 2$.

  THAT IS THERE ARE TWO CLASSES OF DATA.

- THE MEAN OF DATA POINTS IN CLASS C_i BE $m_i \in \mathbb{R}^d$

$$m_i = \frac{\sum_{t=1}^{n} r_i^t x^t}{\sum_{t=1}^{n} r_i^t}$$

- LET w BE A UNIT VECTOR. THAT IS: $w^T w = 1$.

  ORTHOGONAL PROJECTION OF VECTOR x^t ON TO THE UNIT VECTOR w IS:

$$x_{\parallel}^t = (w^T x^t)\, w = a^t w \in \mathbb{R}^d$$

$$a^t \triangleq (w^T x^t) \quad \text{(SCALAR)} \ ; \ a^t \in \mathbb{R}$$

THUS THE SET OF n SCALARS $\{a^1, a^2, \ldots, a^n\}$ REPRESENTS THE MAPPING FROM $\mathbb{R}^d$ TO $\mathbb{R}$. THAT IS, FROM THE ORIGINAL d-DIMENSIONAL SPACE TO A ONE-DIMENSIONAL SPACE ALONG THE UNIT VECTOR w.

- LET $\tilde{m}_i = w^T m_i \; ; \; i = 1, 2$

  THIS IS THE MEAN FROM SAMPLES AFTER PROJECTION.

  NOTE THAT ; $m_i \in \mathbb{R}^d$, AND $\tilde{m}_i \in \mathbb{R}$.

- TO DISCRIMINATE BETWEEN POINTS THAT BELONG TO THE TWO CLASSES

  - MAXIMIZE THE SEPARATION BETWEEN THE CLASSES.

    THAT IS, MAXIMIZE THE SEPARATION BETWEEN THE PROJECTED MEANS $|\tilde{m}_1 - \tilde{m}_2|$.

  - VARIANCE OF THE PROJECTED POINTS FOR EACH CLASS SHOULD ALSO NOT BE TOO LARGE.

    LARGE VARIANCE WOULD LEAD TO POSSIBLE OVERLAPS AMONG THE POINTS OF THE TWO CLASSES.

    THAT IS, SCATTER FOR PROJECTED POINTS WITHIN EACH CLASS IS SMALL.

    SCATTER IS DEFINED AS

    $$s_i^2 = \sum_{t=1}^{n} (a^t - \tilde{m}_i)^2 r_i^t \quad ; \; i = 1, 2$$

    THIS IS DIFFERENT THAN THE VARIANCE σ_i^2 OF CLASS C_i.

    IF n_i = # OF DATA POINTS OF CLASS C_i ; THEN

    $$s_i^2 = (n_i - 1)\,\sigma_i^2$$

    WE WANT THE NET SCATTER $(s_1^2 + s_2^2)$ TO BE SMALL.

- IN LDA, $J(w)$ IS MAXIMIZED WHERE

$$J(w) = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{s_1^2 + s_2^2}$$

$J(w)$ IS CALLED THE FISHER LDA OBJECTIVE

- $$(\tilde{m}_1 - \tilde{m}_2)^2 = (w^T m_1 - w^T m_2)^2$$
$$= (w^T(m_1 - m_2))^2$$
$$= w^T(m_1 - m_2)(m_1 - m_2)^T w$$
$$= w^T S_B w$$

WHERE $S_B \triangleq (m_1 - m_2)(m_1 - m_2)^T$ = BETWEEN CLASS SCATTER MATRIX

- CONSIDER THE SCATTER OF CLASS-1 DATA POINTS

$$s_1^2 = \sum_{t=1}^{n} (a^t - \tilde{m}_1)^2 r_1^t$$
$$= \sum_{t=1}^{n} (w^T x^t - w^T m_1)^2 r_1^t$$
$$= \sum_{t=1}^{n} w^T (x^t - m_1)(x^t - m_1)^T w \, r_1^t$$
$$= w^T S_1 w$$

WHERE $S_1 = \sum_{t=1}^{n} r_1^t (x^t - m_1)(x^t - m_1)^T$

THIS MATRIX IS CALLED THE WITHIN-CLASS SCATTER MATRIX FOR C_1

S_1/m IS THE ESTIMATOR FOR Σ_1, THE COVARIANCE MATRIX FOR CLASS C_1 DATA POINTS.

- SIMILARLY $s_2^2 = w^T S_2 w$

$$S_2 = \sum_{t=1}^{n} r_2^t (x^t - m_2)(x^t - m_2)^T$$

THEREFORE $s_1^2 + s_2^2 = w^T (S_1 + S_2) w$

$$= w^T S_W w$$

WHERE $S_W = S_1 + S_2 =$ TOTAL WITHIN-CLASS SCATTER MATRIX

- $J(\omega) = \dfrac{\omega^T S_B \omega}{\omega^T S_W \omega}$

THE CONSTRAINT $\omega^T\omega = 1$ IS NOT NECESSARY BECAUSE ω OCCURS IN BOTH NUMERATOR AND DENOMINATOR.

THE ω THAT MAXIMIZES $J(\omega)$ IS IDENTICAL TO THE ω THAT MAXIMIZES $\ln J(\omega)$

$$\ln J(\omega) = \ln(\omega^T S_B \omega) - \ln(\omega^T S_W \omega)$$

$$\frac{\partial \ln J(\omega)}{\partial \omega} = \frac{2 S_B \omega}{\omega^T S_B \omega} - \frac{2 S_W \omega}{\omega^T S_W \omega} = 0$$

$$\Rightarrow S_B \omega = \frac{\omega^T S_B \omega}{\omega^T S_W \omega} \cdot S_W \omega$$

$$S_B \omega = J(\omega) S_W \omega$$

LET $\lambda = J(\omega)$, THEN

$$S_B \omega = \lambda S_W \omega$$

$\Rightarrow$ $S_W^{-1} S_B \omega = \lambda \omega$ (ASSUME S_W^{-1} EXISTS)

$\Rightarrow$ λ IS THE EIGENVALUE OF MATRIX $S_W^{-1} S_B$; ω IS THE CORRESPONDING EIGENVECTOR

AS $J(\omega) = \lambda$ HAS TO BE MAXIMIZED, PICK THE LARGEST EIGENVALUE OF $S_W^{-1} S_B$

THE CORRESPONDING DOMINANT VECTOR SPECIFIES THE BEST LINEAR DISCRIMINANT VECTOR ω.

- THE GOAL IS TO ACTUALLY DETERMINE THE DIRECTION OF THE VECTOR w IN THE EQUATION

$$w = \frac{1}{\lambda} S_W^{-1} S_B w$$

NOTE THAT:

$$S_B w = (m_1 - m_2)(m_1 - m_2)^T w$$

$(m_1 - m_2)^T \underset{\uparrow\, w}{w} = b$ IS A SCALAR

THEREFORE

$$w = \frac{1}{\lambda} S_W^{-1} (m_1 - m_2) b$$

LET $v = S_W^{-1}(m_1 - m_2) \Rightarrow w = \frac{b}{\lambda} v$

DIRECTION OF VECTOR w IS THE DIRECTION OF VECTOR v

THIS IS: $\cancel{w}\ \frac{v}{\|v\|} = w$

- IT CAN BE SHOWN THAT WHEN DATA OF CLASS C_i HAVE NORMAL DISTRIBUTION $N_d(\mu_i, \Sigma_i)$; $i = 1, 2$; THEN THE DISCRIMINANT IS LINEAR. CONSEQUENTLY, LINEAR DISCRIMINANT IS OPTIMAL IF THE DATA CLASSES ARE NORMALLY DISTRIBUTED.

ALGORITHM: LDA – 2 CLASSES

INPUT: $D = \{(x^t, r^t) \mid x^t \in \mathbb{R}^d$, r^t IS A 0–1 VECTOR OF LENGTH 2; $1 \le t \le n\}$

OUTPUT: UNIT VECTOR w

BEGIN

$D_i \leftarrow \{(x^t, r_i^t) \mid t \in \{1, 2, \ldots, n\}\}$; $i = 1, 2$

$n_i \leftarrow |D_i|$

$m_i \leftarrow \text{MEAN}(D_i)$; $i = 1, 2$ (CLASS MEANS)

$S_B \leftarrow (m_1 - m_2)(m_1 - m_2)^T$ (BETWEEN-CLASS SCATTER MATRIX)

$S_i \leftarrow \sum_{t=1}^{n} r_i^t (x^t - m_i)(x^t - m_i)^T$; $i = 1, 2$
(WITHIN-CLASS SCATTER MATRICES)

$S_W \leftarrow S_1 + S_2$ (TOTAL WITHIN-CLASS SCATTER MATRIX)

COMPUTE THE DOMINANT EIGENVALUE λ, AND THE CORRESPONDING EIGENVECTOR OF MATRIX $S_W^{-1} S_B$ (ASSUME S_W IS NONSINGULAR)

ALTERNATELY

COMPUTE: $v = S_W^{-1}(m_1 - m_2)$

$$w = \frac{v}{\|v\|}$$

END