ΔΙΑΧΩΡΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

Κανόνας Fisher / Linear Discriminant Analysis – "lda" (R)

$H_0: X \in \Pi_1$ vs. $H_1: X \in \Pi_2$

$X \in \Pi_1$ αν

$$\underbrace{(\mu_1 - \mu_2)'}_{1\times k} \underbrace{\Sigma^{-1}}_{k \times k} \cdot \underbrace{x}_{k\times 1} > C \quad \text{ή}$$

$$\underbrace{x'}_{1\times k} \underbrace{\Sigma^{-1}}_{k\times k} \underbrace{(\mu_1 - \mu_2)}_{k\times 1} > C$$

Με βάση τον κανόνα ίσων σφαλμάτων ($e_{12} \equiv e_{21}$) & την υπόθεση κανονικότητας:

$$C = \frac{1}{2} \underbrace{(\mu_1 - \mu_2)'}_{1\times k} \underbrace{\Sigma^{-1}}_{k\times k} \underbrace{(\mu_1 + \mu_2)}_{k\times 1}$$

$$\text{ή} \quad \frac{1}{2} \underbrace{(\mu_1 + \mu_2)'}_{1\times k} \underbrace{\Sigma^{-1}}_{k\times k} \underbrace{(\mu_1 - \mu_2)}_{k\times 1}$$

Άρα $X \in \Pi_1$ αν

$$(\mu_1 - \mu_2)' \Sigma^{-1} x > \frac{1}{2} (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) \quad \text{ή}$$

$$(\mu_1 - \mu_2)' \Sigma^{-1} x \boxed{-} \frac{1}{2} (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) > \log(1) = 0 = \log\left(\frac{p_2}{p_1}\right)$$

p_1 = πιθ. ταξινόμησης στον Π_1

p_2 = πιθαν. ταξινόμησης στο Π_2

(Θεωρώ ότι $p_1 = p_2$)

Απόσταση X από τον Π_1

Απόσταση του X από το Π_2

Μετά από πράξεις

$\boxed{X \in \Pi_1}$ αν:

$$(\underset{\sim}{x}-\underset{\sim}{\mu}_1)'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}_1) < (\underset{\sim}{x}-\underset{\sim}{\mu}_2)'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}_2) \quad \text{ή}$$

$$\boxed{\Delta_1^2 < \Delta_2^2}$$

Απόσταση Π_1 από Π_2

$$\Delta^2 = (\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)'\Sigma^{-1}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2) = \text{Mahalanobis distance}$$

$\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \Sigma$ άγνωστα. Εκτίμηση με βάση data από Π_1 & Π_2.

Έστω 2 μόνο μεταβλητές:

Data Π_1

x_{111}	x_{121}
x_{112}	x_{122}
$\vdots$	$\vdots$
x_{11n}	x_{12n}

Data Π_2

x_{211}	x_{221}
x_{212}	x_{222}
$\vdots$	$\vdots$
x_{21n}	x_{22n}

$$\hat{\underset{\sim}{\mu}}_1 = \bar{\underset{\sim}{x}}_1 = (\bar{x}_{11}, \bar{x}_{12})' \qquad \hat{\underset{\sim}{\mu}}_2 = \bar{\underset{\sim}{x}}_2 = \begin{pmatrix} \bar{x}_{21} \\ \bar{x}_{22} \end{pmatrix}$$

$$\hat{\Sigma} = S = \begin{bmatrix} Var\, X_{.1} & Cov(X_{.1}, X_{.2}) \\ Cov(X_{.1}, X_{.2}) & Var\, X_{.2} \end{bmatrix} = \text{κοινή διασπορά όλα τα data (pooled)}$$

Κλασικοί Συμβολισμοί S_p^2, S^2_{pooled}

Υπεν.: $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var X \cdot Var Y}} = Corr(X,Y) \Rightarrow \boxed{Cov(X,Y) = \rho(X,Y)\sqrt{Var X\, Var Y}}$

Άρα $X \in \Pi_1$ αν

$$(\bar{x}_1 - \bar{x}_2)' S^{-1} x > \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' S^{-1}(\bar{x}_1 + \bar{x}_2)$$

$$> \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' S^{-1}\bar{x}_1 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' S^{-1}\bar{x}_2$$

ή

$$x' S^{-1}(\bar{x}_1 - \bar{x}_2) > \frac{1}{2}\bar{x}_1' S^{-1}(\bar{x}_1 - \bar{x}_2) + \frac{1}{2}\bar{x}_2' S^{-1}(\bar{x}_1 - \bar{x}_2)$$

Έστω $\hat{\ell}' = (\bar{x}_1 - \bar{x}_2)' S^{-1}$ ή $\hat{\ell} = S^{-1}(\bar{x}_1 - \bar{x}_2)'$

Άρα $X \in \Pi_1$ αν

$$\hat{\ell}' \cdot x > \frac{1}{2}(\hat{\ell}'\bar{x}_1 + \hat{\ell}'\bar{x}_2)$$

ή

$$x' \cdot \hat{\ell} > \frac{1}{2}(\bar{x}_1'\hat{\ell} + \bar{x}_2'\hat{\ell})$$ ή

$$x' \cdot \hat{\ell} - \frac{1}{2}(\bar{x}_1'\hat{\ell} + \bar{x}_2' \cdot \hat{\ell}) > 0$$

Επειδή από ίσα σφάλματα στα πιό πάνω, έχουμε και ίσες πιθανότητες τα γινόμενους: $P_1 \equiv P_2 \Rightarrow \log\left(\frac{P_2}{P_1}\right) = 0.$

Αν $P_1 \neq P_2$ τότε $\boxed{X \in \Pi_1}$ αν:

$$\underset{\sim}{x}'\hat{\underset{\sim}{\ell}} - \frac{1}{2}\left(\bar{\underset{\sim}{x}}_1'\hat{\underset{\sim}{\ell}} + \bar{\underset{\sim}{x}}_2'\hat{\underset{\sim}{\ell}}\right) > \boxed{\log\left(\frac{p_2}{p_1}\right) \equiv m}$$

Αν $p_2 > p_1 \Rightarrow \boxed{m > 0}$ (πιο ΔΥΣΚΟΛΑ πάω Π_1)

Αν $p_2 < p_1 \Rightarrow \boxed{m < 0}$ (πιο ΕΥΚΟΛΑ πάω Π_1)

Έστω k πληθυσμοί. $\Pi_1, \Pi_2, \dots, \Pi_k$.

$x \in \Pi_i$ αν $\Delta_i^2 = \min\{\Delta_1^2, \Delta_2^2, \dots, \Delta_k^2\}$

Υπολογίζω:

$$\underset{\sim}{x}'\Sigma^{-1}(\underset{\sim}{\mu}_i - \underset{\sim}{\mu}_j) - \frac{1}{2}(\underset{\sim}{\mu}_i + \underset{\sim}{\mu}_j)'\Sigma^{-1}(\underset{\sim}{\mu}_i - \underset{\sim}{\mu}_j) > \log\left(\frac{p_j}{p_i}\right) \quad \left(\forall i, j,\ i \neq j\right)$$

Αν $\boxed{p_i = \frac{1}{k}}$ $\forall i \Rightarrow \log\left(\frac{p_j}{p_i}\right) = 0$

$x \in \Pi_i$ αν
$$\underset{\sim}{x}'\Sigma^{-1}(\underset{\sim}{\mu}_i - \underset{\sim}{\mu}_j)' - \frac{1}{2}(\underset{\sim}{\mu}_i + \underset{\sim}{\mu}_j)'\Sigma^{-1}(\underset{\sim}{\mu}_i - \underset{\sim}{\mu}_j) > \log\left(\frac{p_j}{p_i}\right), \ \forall j$$

Παράδειγμα $\boxed{k = 3}$ p_1, p_2, p_3 $\sum_{i=1}^{3} p_i = 1$

Αν ισοπίθανα $p_i = 0.3333$

$$(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)'\Sigma^{-1}\underset{\sim}{x} - \frac{1}{2}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)'\Sigma^{-1}(\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2) > \log\left(\frac{p_2}{p_1}\right)$$
← αν "<" $x \in \Pi_2$

$$(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_3)'\Sigma^{-1}\underset{\sim}{x} - \frac{1}{2}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_3)'\Sigma^{-1}(\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_3) > \log\left(\frac{p_3}{p_1}\right)$$
← αν "<" $x \in \Pi_3$

αν ">" $x \in \Pi_1$

αν όμως "<" $x \in \Pi_3$

~~" $x \in \Pi_3$~~

$$(\mu_2-\mu_3)'\Sigma^{-1}\underset{\sim}{x} - \frac{1}{2}(\mu_2-\mu_3)'\Sigma^{-1}(\mu_2+\mu_3) > \log\left(\frac{p_3}{p_2}\right)$$

αν "<" $x \in \Pi_3$

$x \notin \Pi_2$

$\boxed{\text{Αν ισχύει μία ανισότητα} \Rightarrow \underset{\sim}{x} \in \Pi \text{ με δείκτη του Παρονομαστή του log}}$

ALGORITHM

1 Calculate $\underset{\sim}{\bar{x}}_1$ δηλ. $\bar{x}_{11}, \bar{x}_{12}, \dots, \bar{x}_{1k}$ για Π_1

2. Calculate $\underset{\sim}{\bar{x}}_2$ δηλ. $\bar{x}_{21}, \bar{x}_{22}, \dots, \bar{x}_{2k}$ για Π_2

3. Calculate τον κοινό (pooled) εκτιμητή πίνακα 2×2 S

4. " S^{-1}

5. " $\underset{\sim}{\hat{\ell}}' = (\underset{\sim}{\bar{x}}_1-\underset{\sim}{\bar{x}}_2)'\cdot S^{-1}$ ή $\underset{\sim}{\hat{\ell}} = S^{-1}(\underset{\sim}{\bar{x}}_1-\underset{\sim}{\bar{x}}_2)'$

$\boxed{6}$ " $\underset{\sim}{\hat{\ell}}'\cdot\bar{x}_1$ ✓

$\boxed{7}$ " $\underset{\sim}{\hat{\ell}}'\cdot\bar{x}_2$ ✓

8 Provide $\underset{\sim}{x}$

$\boxed{9}$ Calculate $\underset{\sim}{\hat{\ell}}'\cdot\underset{\sim}{x}$ ✓

10. Compare $\underset{\sim}{\hat{\ell}}'\cdot\underset{\sim}{x}$ to $\frac{1}{2}(\underset{\sim}{\hat{\ell}}'\underset{\sim}{\bar{x}}_1 + \underset{\sim}{\hat{\ell}}'\cdot\underset{\sim}{\bar{x}}_2) = M$

if $\underset{\sim}{\hat{\ell}}'\cdot\underset{\sim}{x} > M \Rightarrow x \in \Pi_1$

else $\Rightarrow x \in \Pi_2$
