

Diszkriminancia-analízis (2 csoportra)

Legyenek $X_1, \dots, X_p$ folytonos véletlen változók, Y egy csoportosító változó.

Legyenek $\bar{X}_1, \bar{X}_2$ a csoportonkénti átlagok és $\mathbf{D}_1, \mathbf{D}_2$ a kovariancia-mátrixok.

Továbbá legyen $\mathbf{A} = (\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2)^T$ és $\mathbf{D} = \mathbf{D}_1 + \mathbf{D}_2$.

Egy lineáris diszkriminancia-függvényt szeretnénk megadni. Ehhez olyan irányt keresünk, amelyre vett vetületet jól tudjuk szeparálni. A keresett irány az alábbi *Rayleigh-hányados* maximumhelye lesz:

$$R(\mathbf{v}) = \frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{\mathbf{v}^T \mathbf{D} \mathbf{v}}.$$

(A számlálóban a $\bar{X}_1 - \bar{X}_2$ vektor $\mathbf{v}$ irányba eső vetületének négyzete, a nevezőben $\mathbf{v}$ irányba eső szóródás szerepel.)

Legyen $\mathbf{D}^{-1}\mathbf{A}$ legnagyobb sajátértéke λ_1 és a hozzá tartozó sajátvektor $\mathbf{v}_1$. Belátható, hogy $R(\mathbf{v})$ maximumhelye $\mathbf{v}_1$. Ez alapján a *diszkriminancia-függvény* $f(\mathbf{x}) = \mathbf{v}_1^T \mathbf{x} + b_1$ alakú lesz. Ha $f(\mathbf{x}) > 0$, akkor $\mathbf{x}$ -et az első csoportba, különben a második csoportba soroljuk.

A szeparálás jóságát a *Wilks-féle* λ statisztikával jellemezzük: $\Lambda = \frac{1}{1+\lambda_1}$. A Λ értéke akkor kicsi, ha a csoportok jól szeparálhatók.