Proof of Exercise 9 Demo

Show that the solution for the problem:

$$\max_a \left\{ \frac{a^\top B a}{a^\top W a} \right\},$$

is obtained by setting a equal to the eigenvector of $W^{-1}B$ that corresponds to the largest eigenvalue.

W = measure of group dispersions $\qquad$ B = dispersion between groups

Since the vector a can be scaled arbitrarily withouth affecting the ratio, we can formulate the problem as follows:

$$\max_a \left\{ a^\top B a \right\} \qquad \text{s.t.} \quad a^\top W a = 1.$$

Let $W^{1/2}$ be the symmetric square root of W. Note that W is symmetric. Let $z = W^{1/2}a$ and $a = W^{-1/2}z$. Then

$$\begin{aligned} a^\top B a &= \left(W^{-1/2}z\right)^\top B \left(W^{-1/2}z\right) = z^\top W^{-1/2} B W^{-1/2} z, \\ a^\top W a &= \left(W^{-1/2}z\right)^\top W \left(W^{-1/2}z\right) = z^\top z. \end{aligned}$$

Note that $W^{-1/2}BW^{-1/2}$ is symmetric and hereby the spectral decomposition exists $\left(W^{-1/2}BW^{-1/2} = \Gamma\Lambda\Gamma^\top \text{ and } \Gamma^\top\Gamma = I\right)$,

$$\begin{aligned} z^\top W^{-1/2} B W^{-1/2} z &= z^\top \Gamma\Lambda\Gamma^\top z = w^\top \Lambda w, \qquad \left(w = \Gamma^\top z\right) \\ z^\top z &= z^\top \Gamma\Gamma^\top z = w^\top w. \end{aligned}$$

Now we can reformulate the problem as

$$\max_w \left\{ w^\top \Lambda w \right\} = \max_w \left\{ \sum_{i=1}^{p} \lambda_i w_i^2 \right\} \quad \text{s.t. } w^\top w = 1.$$

Since $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$, we choose the first element of w to be one and the rest to be 0. This means that $z = \Gamma w = \gamma_1$, where γ_1 is the first eigenvector of $W^{-1/2}BW^{-1/2}$ and $a = W^{-1/2}z = W^{-1/2}\gamma_1$.

Note that for any two matrices $A \in \mathbb{R}^{n\times p}$ and $C \in \mathbb{R}^{p\times n}$, the non-zero eigenvalues of AC and CA are the same and have the same multiplicity (Theorem A.6.2 of Mardia, Kent and Bibby). Now let $A = W^{-1/2}B$ and $C = W^{-1/2}$, this means that the non-zero eigenvalues of $CA = W^{-1}B$ are the same as $AC = W^{-1/2}BW^{-1/2}$. Hence, λ_1 is the largest eigenvalue of $W^{-1}B$. Since γ_1 is the eigenvector corresponding to the largest eigenvalue λ_1 of $W^{-1/2}BW^{-1/2}$, we have that

$$\begin{aligned} W^{-1}B\left(W^{-1/2}\gamma_1\right) &= W^{-1/2}\left(W^{-1/2}BW^{-1/2}\gamma_1\right) = W^{-1/2}\lambda_1\gamma_1 \\ &= \lambda_1\left(W^{-1/2}\gamma_1\right). \end{aligned}$$

This shows that $a = W^{-1/2}\gamma_1$ is the eigenvector of $W^{-1}B$ corresponding to its largest eigenvalue λ_1.