7.5. FISHER'S METHOD

In d-space the line joining the means (i.e., the linear discriminant function) is not so easy to visualize, but in the case of two groups there always is one. Fisher [1936] designed a method in which points in d-space are projected onto that line (or a line parallel to it). He viewed separation and classification together and used an approach based on the distance between groups. One requirement is that the rule obtained be a linear rule. Hence we are interested in linear combinations $Y = \mathbf{l}^{\mathrm{T}}\mathbf{x}$. From our previous results, the mean of the combination is $\mu_Y = \mathbf{l}^{\mathrm{T}}\boldsymbol{\mu}_x$ while the variance of Y is $V(Y) = \mathbf{l}^{\mathrm{T}}\,\Sigma\mathbf{l}$. Using the linear combination changes a multivariate problem into a univariate problem because Y has dimension one. Fisher's idea was to pick $\mathbf{l}$ so as to maximize

$$\frac{(\mathbf{l}^{\mathrm{T}}\boldsymbol{\mu}_1 - \mathbf{l}^{\mathrm{T}}\,\boldsymbol{\mu}_2)^2}{\mathbf{l}^{\mathrm{T}}\,\Sigma\mathbf{l}} = \frac{\text{squared distance between the } Y \text{ means}}{\text{variance of } Y} \tag{7.22}$$