$= \mathrm{Var}(\boldsymbol{l}^{\mathrm{T}}\mathbf{x}) = \boldsymbol{l}^{\mathrm{T}}\boldsymbol{l} = 1$. When this scaling is used, y is sometimes referred to as a "canonical variate" [Kleck80].

We have drawn the discriminant line in Fig. 7.13 with the data points projected onto this line. The slope is determined from Fisher's linear discriminant function. The other line in Fig. 7.12(a) defines the classification boundary. This line is perpendicular to the discriminant function and passes through the point associated with $\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$, thus separating the space into the regions Γ_1 and Γ_2. The effect of the prior probability is to shift the boundary line up or down the discriminant line. If, for example, $P(c_1)$ is larger than $P(c_2)$ we shift the boundary towards the second class, thus improving classification for the group with the higher prior. Thus for the case of two groups, Bayes' method represents a slight modification of Fisher's method. We emphasize that *in Fisher's method, the normality assumption is not used*, although it works best (i.e., is equivalent to Bayes' rule) when the assumption is valid. Indeed, when the optimal decision surface is far from linear it may give very bad results [Hand81].