It is natural to define this second term as a general between-class scatter matrix, so that the total scatter is the sum of the within-class scatter and the between-class scatter:

$$\mathbf{S}_B = \sum_{i=1}^{c} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t \tag{115}$$

and

$$\mathbf{S}_T = \mathbf{S}_W + \mathbf{S}_B. \tag{116}$$

If we check the two-class case, we find that the resulting between-class scatter matrix is $n_1 n_2/n$ times our previous definition.*

The projection from a $d$-dimensional space to a $(c-1)$-dimensional space is accomplished by $c-1$ discriminant functions

$$y_i = \mathbf{w}_i^t\mathbf{x} \qquad i = 1, \ldots, c-1. \tag{117}$$

If the $y_i$ are viewed as components of a vector $\mathbf{y}$ and the weight vectors $\mathbf{w}_i$ are viewed as the columns of a $d$-by-$(c-1)$ matrix $\mathbf{W}$, then the projection can be written as a single matrix equation

$$\mathbf{y} = \mathbf{W}^t\mathbf{x}. \tag{118}$$

The samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ project to a corresponding set of samples $\mathbf{y}_1, \ldots, \mathbf{y}_n$, which can be described by their own mean vectors and scatter matrices. Thus, if we define

$$\tilde{\mathbf{m}}_i = \frac{1}{n_i}\sum_{\mathbf{y}\in\mathcal{Y}_i}\mathbf{y} \tag{119}$$

$$\tilde{\mathbf{m}} = \frac{1}{n}\sum_{i=1}^{c} n_i\tilde{\mathbf{m}}_i \tag{120}$$

$$\tilde{\mathbf{S}}_W = \sum_{i=1}^{c}\sum_{\mathbf{y}\in\mathcal{Y}_i}(\mathbf{y} - \tilde{\mathbf{m}}_i)(\mathbf{y} - \tilde{\mathbf{m}}_i)^t \tag{121}$$

and

$$\tilde{\mathbf{S}}_B = \sum_{i=1}^{c} n_i(\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}})(\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}})^t, \tag{122}$$

it is a straightforward matter to show that

$$\tilde{\mathbf{S}}_W = \mathbf{W}^t\mathbf{S}_W\mathbf{W} \tag{123}$$

and

$$\tilde{\mathbf{S}}_B = \mathbf{W}^t\mathbf{S}_B\mathbf{W}. \tag{124}$$

*We could redefine $\mathbf{S}_B$ for the two-class case to obtain complete consistency, but there should be no misunderstanding of our usage.