$$\mathbf{S}_T = \sum_{n=1}^{N} (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T \tag{12.36}$$

$$= \sum_{\gamma=1}^{\Gamma} \sum_{\mathbf{x}\in\mathscr{X}_\gamma} \left(\mathbf{x} - \boldsymbol{\mu}_\gamma + \boldsymbol{\mu}_\gamma - \boldsymbol{\mu}\right)\left(\mathbf{x} - \boldsymbol{\mu}_\gamma + \boldsymbol{\mu}_\gamma - \boldsymbol{\mu}\right)^T \tag{12.37}$$

$$= \sum_{\gamma=1}^{\Gamma} \sum_{\mathbf{x}\in\mathscr{X}_\gamma} \left(\mathbf{x} - \boldsymbol{\mu}_\gamma\right)\left(\mathbf{x} - \boldsymbol{\mu}_\gamma\right)^T + \sum_{\gamma=1}^{\Gamma} N_\gamma \left(\boldsymbol{\mu}_\gamma - \boldsymbol{\mu}\right)\left(\boldsymbol{\mu}_\gamma - \boldsymbol{\mu}\right)^T \tag{12.38}$$

The first term in Equation 12.38 is identical to $\mathbf{S}_W$ of Equations 12.33. The second term seems to have the properties stated at the beginning of this section. In fact, we can call that the *between class* scatter matrix, giving us the following definition of S_B [8],

$$\mathbf{S}_B \stackrel{\Delta}{=} \sum_{\gamma=1}^{\Gamma} N_\gamma \left(\boldsymbol{\mu}_\gamma - \boldsymbol{\mu}\right)\left(\boldsymbol{\mu}_\gamma - \boldsymbol{\mu}\right)^T \tag{12.39}$$

Therefore, based on Equations 12.38, 12.33, and 12.39, the *total scatter* matrix may be written in terms of the *within class* and *between class* scatter matrices as follows,

$$\mathbf{S}_T = \mathbf{S}_W + \mathbf{S}_B \tag{12.40}$$

At this point, we have the between and within scatter matrices (hence covariances) of random variable, X, estimated based on the observation of N samples of X. The objective of linear discriminant analysis [5] is to maximize discriminability by performing a transformation ($T : \mathscr{R}^{(\Gamma-1)} \mapsto \mathscr{R}^D$) on X such that the transformed random variable, $Y : \mathbf{y} \in \mathscr{R}^{(\Gamma-1)}$ would have maximal ratio of the determinant of the *between class scatter matrix* to the determinant of the *within class scatter matrix*. The determinant is used since it is equal to the product of the Eigenvalues of the matrix and an indication of the *geometric mean* (Definition 6.79), of the Eigenvalues.

Let us define the following *linear transformation* for discriminant analysis of X,

$$\mathbf{y}_n = T(\mathbf{x}_n) \tag{12.41}$$

$$\stackrel{\Delta}{=} \mathbf{U}^T \mathbf{x}_n \tag{12.42}$$

where $\mathbf{U} : \mathscr{R}^{(\Gamma-1)} \mapsto \mathscr{R}^D$ is a rectangular matrix and $(\Gamma - 1) < D$. Therefore, the between class and within class scatter matrices of Y may be written in terms of those of X and the transformation matrix $\mathbf{U}$, as follows,

$$\mathbf{S}_B(Y) = \mathbf{U}^T \mathbf{S}_B(X)\mathbf{U} \tag{12.43}$$

$$\mathbf{S}_W(Y) = \mathbf{U}^T \mathbf{S}_W(X)\mathbf{U} \tag{12.44}$$

[5] Called *multiple discriminant analysis*[8] when there are more than two classes involved – $\Gamma > 2$.