For the dimensionality reduction, we first consider the case of two classes, and then we can consider the generalization of the Linear discriminant for more than two classes.
Suppose we take the D-dimensional input vector X and project it down to one dimension using

$$y = W^T X$$

If we classify $y > -w_0$ as C_1, and otherwise as class C_2, then we obtain the standard linear classifier. We can, by adjusting the components of the weight vector W, select a projection that maximizes the class separation.
Consider that, there are N_1 examples of class C_1 and N_2 examples of class C_2, so the mean vectors of the two classes are given by

$$\mathbf{m_1} = \frac{\mathbf{1}}{\mathbf{N_1}} \sum_{\mathbf{n} \in C_1} \mathbf{X_n}, \qquad \mathbf{m_1} = \frac{\mathbf{1}}{\mathbf{N_2}} \sum_{\mathbf{n} \in C_2} \mathbf{X_n}$$

One way to measure the separation of classes is find the separation of the projected class means, when projected onto W. This implies we choose W so as to maximize

$$m_2 - m_1 = W^T(\mathbf{m_2} - \mathbf{m_1})$$

where,

$$m_k = W^T \mathbf{m_k}$$

is the mean of the projected data from class C_k. However, this expression can be made large by increasing the magnitude of W. To solve this problem, we could constrain W to to have unit length, so that $\sum_i w_i^2 = 1$.