- 最佳变换向量 $\boldsymbol{w}^*$的求取

为求使 $J_{\mathrm{F}}(\boldsymbol{w})=\dfrac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{w}}$ 取极大值时的 $\boldsymbol{w}^*$，可以采用 Lagrange 乘数法求解。令分母等于非零常数，即：

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{w}=c\neq 0$$

定义 Lagrange 函数为：

$$L(\boldsymbol{w},\lambda)=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{w}-\lambda(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{w}-c)$$

其中 λ 为 Lagrange 乘子。将上式对 $\boldsymbol{w}$ 求偏导数，可得：

$$\frac{\partial L(\boldsymbol{w},\lambda)}{\partial \boldsymbol{w}}=\boldsymbol{S}_b\boldsymbol{w}+\boldsymbol{S}_b^T\boldsymbol{w}-\lambda(\boldsymbol{S}_w\boldsymbol{w}+\boldsymbol{S}_w^T\boldsymbol{w})=2\boldsymbol{S}_b\boldsymbol{w}-\lambda 2\boldsymbol{S}_w\boldsymbol{w}$$

令偏导数为零，有:

$$\boldsymbol{S}_b\boldsymbol{w}^*-\lambda\boldsymbol{S}_w\boldsymbol{w}^*=0$$

注：

y是标量，$\boldsymbol{x}=\begin{bmatrix}x_1 & x_2 & \cdots & x_d\end{bmatrix}^T$是$d$维列向量,则$\dfrac{\partial y}{\partial \boldsymbol{x}}=\begin{bmatrix}\dfrac{\partial y}{\partial x_1} & \dfrac{\partial y}{\partial x_2} & \cdots & \dfrac{\partial y}{\partial x_d}\end{bmatrix}^T$

常用求导公式：

$$\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{A}^T$$

$$\frac{\partial \boldsymbol{x}^T\boldsymbol{x}}{\partial \boldsymbol{x}}=2\boldsymbol{x}$$

$$\frac{\partial \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}^T\boldsymbol{x}$$

即

$$\boldsymbol{S}_b\boldsymbol{w}^*=\lambda\boldsymbol{S}_w\boldsymbol{w}^*$$

其中 $\boldsymbol{w}^*$就是 $J_{\mathrm{F}}(\boldsymbol{w})$的极值解。因为 $\boldsymbol{S}_{\mathrm{w}}$ 非奇异，将上式两边左乘 $\boldsymbol{S}_w^{-1}$，可得：

$$\boldsymbol{S}_w^{-1}\boldsymbol{S}_b\boldsymbol{w}^*=\lambda\boldsymbol{w}^*$$

上式为求一般矩阵$\boldsymbol{S}_w^{-1}\boldsymbol{S}_b$的特征值问题。利用$\boldsymbol{S}_b=(\boldsymbol{m}_1-\boldsymbol{m}_2)(\boldsymbol{m}_1-\boldsymbol{m}_2)^T$的定义，将上式左边的$\boldsymbol{S}_b\boldsymbol{w}^*$写成：

$$\boldsymbol{S}_b\boldsymbol{w}^*=(\boldsymbol{m}_1-\boldsymbol{m}_2)(\boldsymbol{m}_1-\boldsymbol{m}_2)^T\boldsymbol{w}^*=(\boldsymbol{m}_1-\boldsymbol{m}_2)R$$

其中$R=(\boldsymbol{m}_1-\boldsymbol{m}_2)^T\boldsymbol{w}^*$为一标量，所以$\boldsymbol{S}_b\boldsymbol{w}^*$总是在向量$(\boldsymbol{m}_1-\boldsymbol{m}_2)$的方向上。因此$\lambda\boldsymbol{w}^*$可写成：

$$\lambda\boldsymbol{w}^*=\boldsymbol{S}_w^{-1}\boldsymbol{S}_b\boldsymbol{w}^*=\boldsymbol{S}_w^{-1}(\boldsymbol{m}_1-\boldsymbol{m}_2)R$$

从而可得：

$$\boldsymbol{w}^*=\frac{R}{\lambda}\boldsymbol{S}_w^{-1}(\boldsymbol{m}_1-\boldsymbol{m}_2)$$

由于我们的目的是寻找最佳的投影方向，$\boldsymbol{w}^*$的比例因子对此并无影响，因此可忽略比例因子$\frac{R}{\lambda}$，有：

$$\boldsymbol{w}^*=\boldsymbol{S}_w^{-1}(\boldsymbol{m}_1-\boldsymbol{m}_2)$$