$$\max_{\vec{w}} \ \vec{w}^T(\vec{m}_2 - \vec{m}_1)$$

$$\text{s.t.} \ \vec{w}^T \cdot \vec{w} = 1$$

Lagrangian function

$L(\vec{w}, \lambda) = \vec{w}^T(\vec{m}_2 - \vec{m}_1) - \lambda\,(\vec{w}^T\vec{w} - 1)$, $\lambda \neq 0$ is the Lagrange multiplier

$$\frac{\partial L(\vec{w}, \lambda)}{\partial \vec{w}} = \vec{m}_2 - \vec{m}_1 - \lambda \cdot 2\vec{w} = \vec{0}$$

$$\Rightarrow \ 2\lambda\vec{w} = \vec{m}_2 - \vec{m}_1$$

$$\vec{w} \propto (\vec{m}_2 - \vec{m}_1)$$

LDA : Two - Class.

$$J(\vec{w}) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}$$

$$m_2 = \vec{w}^T \cdot \vec{m_2} \quad , \quad m_1 = \vec{w}^T \cdot \vec{m_1}$$

$$s_1^2 = \sum_{n \in C_1} (y_n - m_1)^2 \quad , \quad y_n = \vec{w}^T \cdot \vec{x_n}$$

$$= \sum_{n \in C_1} \left(\vec{w}^T \vec{x_n} - \vec{w}^T \cdot \vec{m_1} \right)^2$$

$$= \sum_{n \in C_1} \vec{w}^T (\vec{x_n} - \vec{m_1}) \cdot (\vec{x_n} - \vec{m_1})^T \cdot \vec{w}$$

$$= \vec{w}^T \underbrace{\sum_{n \in C_1} (\vec{x_n} - \vec{m_1}) \cdot (\vec{x_n} - \vec{m_1})^T}_{\text{Class 1 covariance matrix}} \cdot \vec{w}$$

similarly,

$$s_2^2 = \vec{w}^T \underbrace{\sum_{n \in C_2} (\vec{x_n} - \vec{m_2}) \cdot (\vec{x_n} - \vec{m_2})^T}_{\text{Class 2 covariance matrix.}} \cdot \vec{w}$$

$$s_1^2 + s_2^2 = \vec{w}^T \cdot \underbrace{\left[\sum_{n \in C_1} (\vec{x_n} - \vec{m_1}) \cdot (\vec{x_n} - \vec{m_1})^T + \sum_{n \in C_2} (\vec{x_n} - \vec{m_2}) \cdot (\vec{x_n} - \vec{m_2})^T \right]}_{S_W \text{ : total within-class covariance matrix.}} \cdot \vec{w}$$

$$(m_2 - m_1)^2 = \left(\vec{w}^T \cdot \vec{m}_2 - \vec{w}^T \cdot \vec{m}_1\right)^2$$

$$= \left|\vec{w}^T (\vec{m}_2 - \vec{m}_1)\right|^2$$

$$= \vec{w}^T \cdot \underbrace{(\vec{m}_2 - \vec{m}_1) \cdot (\vec{m}_2 - \vec{m}_1)^T}_{S_B \text{ : between-class covariance matrix}} \cdot \vec{w}$$

$$J(\vec{w}) = \frac{\vec{w}^T \cdot S_B \cdot \vec{w}}{\vec{w}^T \cdot S_W \cdot \vec{w}}$$

$$\frac{\partial J(\vec{w})}{\partial \vec{w}} = \frac{2 S_B \cdot \vec{w} \cdot (\vec{w}^T \cdot S_W \cdot \vec{w}) - (\vec{w}^T \cdot S_B \cdot \vec{w})\, 2 S_W \cdot \vec{w}}{(\vec{w}^T \cdot S_W \cdot \vec{w})^2} = \vec{0}$$

$$(\vec{w}^T \cdot S_B \cdot \vec{w}) \cdot S_W \cdot \vec{w} = (\vec{w}^T \cdot S_W \cdot \vec{w}) \cdot S_B \cdot \vec{w} \qquad (*)$$

Because $S_B \triangleq (\vec{m}_2 - \vec{m}_1) \cdot (\vec{m}_2 - \vec{m}_1)^T$

We have. $S_B \cdot \vec{w} = (\vec{m}_2 - \vec{m}_1) \cdot \underbrace{(\vec{m}_2 - \vec{m}_1)^T \cdot \vec{w}}_{\text{scalar}}$

Hence, $S_B \cdot \vec{w}$ is always in the direction of $(\vec{m}_2 - \vec{m}_1)$

From $(*)$, we obtain:

$$S_W \cdot \vec{w} \propto \vec{m}_2 - \vec{m}_1$$

That is, $\vec{w} \propto S_W^{-1} \cdot (\vec{m}_2 - \vec{m}_1)$ $\quad$ S_W need to be full rank

LDA: Multiple-Class

- The within-class variance in the projection space.

$$\sum_{k=1}^{K} \sum_{n \in C_k} (y_n - m_k)^2$$

$$= \sum_{k=1}^{K} \sum_{n \in C_k} \left(\vec{w}^T \cdot \vec{x}_n - \vec{w}^T \cdot \vec{m}_k\right)^2 = \sum_{k=1}^{K} \sum_{n \in C_k} \vec{w}^T \cdot (\vec{x}_n - \vec{m}_k) \cdot (\vec{x}_n - \vec{m}_k)^T \cdot \vec{w}$$

$$= \vec{w}^T \cdot \underbrace{\sum_{k=1}^{K} \sum_{n \in C_k} (\vec{x}_n - \vec{m}_k) \cdot (\vec{x}_n - \vec{m}_k)^T}_{S_W \text{ : within-class covariance matrix}} \cdot \vec{w} = \vec{w}^T \cdot S_W \cdot \vec{w}$$

- The Between-class variance in the projection space.

$$\sum_{k=1}^{K} N_k (m_k - m)^2 = \sum_{k=1}^{K} N_k \left(\vec{w}^T \cdot \vec{m}_k - \vec{w}^T \cdot \vec{m}\right)^2$$

$$= \sum_{k=1}^{K} N_k \, \vec{w}^T (\vec{m}_k - \vec{m}) \cdot (\vec{m}_k - \vec{m})^T \cdot \vec{w}$$

$$= \vec{w}^T \cdot \underbrace{\sum_{k=1}^{K} N_k \cdot (\vec{m}_k - \vec{m}) \cdot (\vec{m}_k - \vec{m})^T}_{S_B \text{ : between-class covariance matrix}} \cdot \vec{w} = \vec{w}^T \cdot S_B \cdot \vec{w}$$

Our problem: maximize the following objective function w.r.t. $\vec{w}$

$$J(\vec{w}) = \frac{\vec{w}^T \cdot S_B \cdot \vec{w}}{\vec{w}^T \cdot S_W \cdot \vec{w}}$$

Since we can always scale $\vec{w}$ such that $\vec{w}^T \cdot S_W \cdot \vec{w} = 1$, and this will not
change the value of the objective function $J(\vec{w})$.

Then: the problem becomes:

$$\max_{\vec{w}} \ \vec{w}^T \cdot S_B \cdot \vec{w}$$

subject to $\vec{w}^T \cdot S_W \cdot \vec{w} = 1$.

$$L(\vec{w}, \lambda) = \vec{w}^T \cdot S_B \cdot \vec{w} - \lambda \left(\vec{w}^T \cdot S_W \cdot \vec{w} - 1 \right)$$

$$\frac{\partial L(\vec{w}, \lambda)}{\partial \vec{w}} = \vec{0}$$

$$\Rightarrow \cancel{2} \cdot S_B \cdot \vec{w} - \lambda \cdot \cancel{2} S_W \cdot \vec{w} = \vec{0}$$

$\vec{w}$ is an eigenvector of $S_W^{-1} \cdot S_B$, and λ is the associated eigenvalue.

$$S_B \cdot \vec{w} = \lambda \cdot S_W \cdot \vec{w}, \quad \Rightarrow \quad S_W^{-1} \cdot S_B \vec{w} = \lambda \cdot \vec{w}.$$

$$\frac{\partial L(\vec{w}, \lambda)}{\partial \lambda} = -\vec{w}^T \cdot S_W \cdot \vec{w} + 1 = 0 \quad \Rightarrow \quad \vec{w}^T \cdot S_W \cdot \vec{w} = 1$$

$$\Rightarrow \quad \lambda \cdot \vec{w}^T \cdot S_W \cdot \vec{w} = \vec{w}^T \cdot S_B \cdot \vec{w}$$

$$\Rightarrow \quad \lambda = \vec{w}^T \cdot S_B \cdot \vec{w}$$

Because we want to maximize $\vec{w}^T \cdot S_B \cdot \vec{w} = \lambda$.

λ must be the largest eigenvalue of $S_W^{-1} S_B$.

Thus, the solution of $\vec{w}$ is the eigenvector of $S_W^{-1} \cdot S_B$ corresponding to the largest eigenvalue.