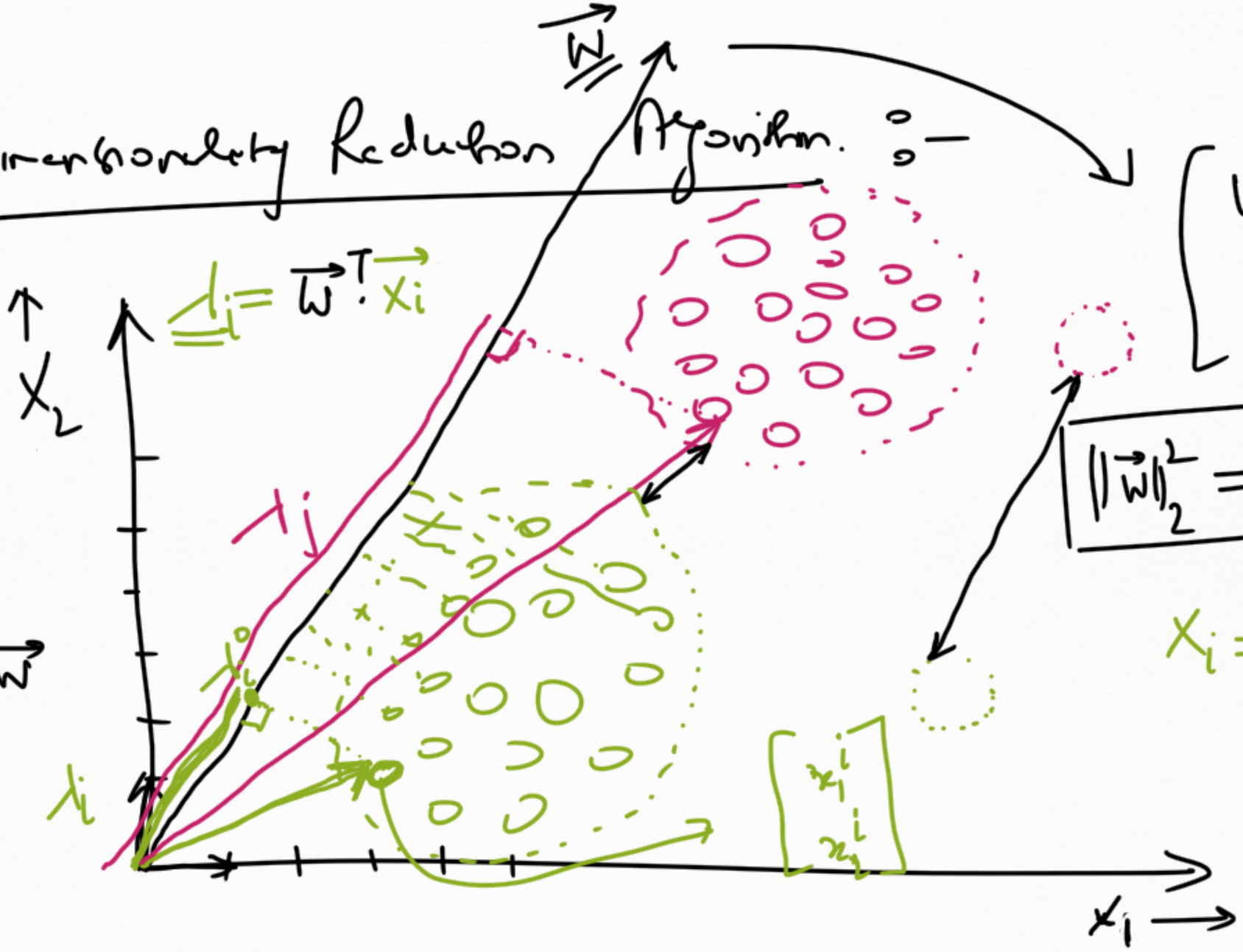
LDA as Dimensionality Reduction Algorithm :-

$$\lambda_i = \frac{\vec{W}^T \cdot \vec{X_i}}{\|\vec{W}\|_2^2}$$

$$\vec{P_i} = \lambda_i \vec{W}$$



$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\boxed{\|\vec{W}\|_2^2 = 1}$$

$$X_i = \begin{bmatrix} x_1^i \\ x_2^i \end{bmatrix}$$

So, we need to find such kind of unit vector $\vec{w}$ such that all points belonging to single class, when projected on that vector should give us minimum variance as possible but at same time, between class variance should be as maximum as possible

Lets say that for any i^{th} example belonging to 0 class, $\underline{\underline{X_i^0}}$

when we project on the unit vector, we get the value,

λ_i^0

$$\underline{\underline{\lambda_i^0}} = W^T \cdot X_i^0$$

$$\underline{\underline{\lambda_i^1}} = W^T \cdot X_i^1$$

$$\underline{\underline{X_i^0}} = \begin{bmatrix} x_1^i \\ x_2^i \\ \vdots \\ x_N^i \end{bmatrix}$$

feature vector

$$\vec{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}$$

Within group variability :-

$$\mu_0 = \begin{bmatrix} \mu_1^0 \\ \mu_2^0 \\ \vdots \\ \mu_N^0 \end{bmatrix} \quad N \times 1$$

0 class variance after projection :-

$$y_1^0, y_2^0, y_3^0 \ldots \ldots y_{N_0}^0$$

$N_0 \rightarrow$ Number of rows in 0 class

$$= W^T \mu_0$$

$\mu_y^0 \longrightarrow$ Mean of the projected data of 0 class

$$\mu_y^0 = \left(\frac{y_1^0 + y_2^0 + \ldots \ldots + y_{N_0}^0}{N_0} \right) = \left(\frac{W^T \cdot x_1^0 + W^T x_2^0 + \ldots \ldots W^T x_{N_0}^0}{N_0} \right)$$

Therefore, $\mu_0 = \begin{bmatrix} \mu_1^0 \\ \mu_2^0 \\ \vdots \\ \mu_N^0 \end{bmatrix}$ $N \times 1$ vector $\rightarrow$ mean vector for 0 class

Variance of 0 class data after projection is given by -

$$Var(\lambda^0) = \frac{\sum_{i=1}^{N_0} (\lambda_i^0 - \mu_\lambda^0)^2}{N_0}$$

But we know that -

$$\lambda_i^0 = W^T \cdot X_i^0$$

and $\mu_\lambda^0 = W^T \cdot \mu_0$ ($N \times 1$)

Therefore, $$Var(\lambda^0) = \frac{\sum_{i=1}^{N_0} \left(\underline{w}^T \cdot \underline{x}_i^0 - \underline{w}^T \cdot \underline{\mu}_0\right)^2}{N_0} =$$

Lets say that $\underline{\underline{X}}^0$ is a matrix of 0 class data as –

$$X^0 = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ x_1^0 & x_2^0 & \cdots & x_{N_0}^0 \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \longrightarrow \underline{\underline{N}} \times N_0 \text{ matrix}$$

$$\mu_0 = \begin{bmatrix} \mu_1^0 \\ \mu_2^0 \\ \vdots \\ \mu_N^0 \end{bmatrix} \updownarrow \underline{\underline{N \times 1}}$$

$$(X^0 - \mu^0) = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ (X_1^0 - \mu^0) & (X_2^0 - \mu^0) & \cdots & (X_{N_0}^0 - \mu^0) \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \; N \times N_0$$

$W^T = 1 \times N$ vector

$$W^T \cdot (X^0 - \mu^0) = \begin{bmatrix} W^T \cdot (X_1^0 - \mu^0) & \cdots & W^T \cdot (X_{N_0}^0 - \mu^0) \end{bmatrix}$$

$1 \times N_0$ matrix

$$\underbrace{\begin{bmatrix} \underbrace{w^T \cdot (x_1^0 - \mu_0)} & \cdots & w^T \cdot (x_{N_0}^0 - \mu_0) \end{bmatrix}}_{w^T \cdot (X^0 - \mu^0)} \cdot \underbrace{\begin{bmatrix} w^T \cdot (x_1^0 - \mu_0) \\ w^T \cdot (x_2^0 - \mu_0) \\ \vdots \\ w^T \cdot (x_{N_0}^0 - \mu_0) \end{bmatrix}}_{[w^T \cdot (X^0 - \mu^0)]^T}$$

$1 \times N_0$

$$(\lambda_1^0 - \mu_\lambda^0) \quad \cdots \quad (\lambda_{N_0}^0 - \mu_\lambda^0)$$

$$(\lambda_1^0 - \mu_\lambda^0) \quad (\lambda_2^0 - \mu_\lambda^0) \quad \vdots \quad (\lambda_{N_0}^0 - \mu_\lambda^0)$$

$N_0 \times 1$

$$\frac{\sum_{i=1}^{N_0} (y_i^0 - \mu_y^0)^2}{N_0} = \frac{1}{N_0}\left[W^T \cdot (X^0 - \mu^0)\right] \cdot \underbrace{\left[W^T \cdot (X^0 - \mu^0)\right]^T}$$

$$= \frac{1}{N_0} W^T \cdot (X^0 - \mu^0) \cdot (X_0 - \mu_0)^T \cdot W$$

<u>For 1 class</u> :-

$$\frac{\sum_{i=1}^{N_1} (y_i^1 - \mu_y^1)^2}{N_1} = \frac{1}{N_0} W^T (X^1 - \mu^1) \cdot (X^1 - \mu^1)^T \cdot W$$

Within Class Variability :-

$$\sigma_W^2 = \frac{N_0 \sum_{i=1}^{N_0} \left(\lambda_i^0 - \mu_\lambda^0\right)^2 + N_1 \cdot \sum_{i=1}^{N_1} \left(\lambda_i^1 - \mu_\lambda^1\right)^2}{(N_0 + N_1) \rightarrow N_{train}}$$

$$\sigma_W^2 = \frac{N_0 \cdot W^T \cdot \underbrace{\overset{N \times N_0}{\left(X^0 - \mu_0\right)} \overset{N_0 \times N}{\left(X^0 - \mu_0\right)^T}}_{N \times N} \cdot W + N_1 \cdot W^T \cdot \underbrace{\overset{N \times N_1}{\left(X^1 - \mu_1\right)} \cdot \overset{N_1 \times N}{\left(X_1 - \mu_1\right)^T}}_{N \times N} \cdot W}{\underline{\underline{N_{train}}}}$$

$$S_W = \frac{\underbrace{N_0 \cdot (X^0 - \mu_0) \cdot (X^0 - \mu_0)^T}_{\underline{\underline{N \times N}}} + \underbrace{N_1 \cdot (X^1 - \mu_1) \cdot (X^1 - \mu_1)^T}_{\underline{\underline{N \times N}}}}{\underline{\underline{N_{train}}}}$$

Within class $\underline{\underline{\text{scatter}}}$ matrix $\rightarrow$ $\underline{\underline{N \times N}}$

$$\sigma_W^2 = W^T \cdot \underline{\underline{S_W}} \cdot W$$

$$GM_j = \frac{\mu_j^0 + \mu_j^1}{2} = \frac{W^T \cdot \underline{\mu_0} + W^T \cdot \underline{\mu_1}}{2} = W^T \cdot \left(\frac{\mu_0 + \mu_1}{2}\right)$$

$$= \underline{\underline{W^T \cdot GM}} \rightarrow \begin{bmatrix} GM_1 \\ GM_2 \\ \vdots \\ GM_N \end{bmatrix}$$

<u>Between Class Variability :-</u>

$$\sigma_B^2 = \frac{\sum_{i=0}^{1} \left(\mu_j^i - GM_j\right)^2}{2} = \frac{\sum_{i=0}^{1} \left(W^T \cdot \mu_i - W^T \cdot GM\right)^2}{2}$$

$$\underline{\underline{\mu}} = \begin{bmatrix} \uparrow & \uparrow \\ \mu_0 & \mu_1 \\ \downarrow & \downarrow \end{bmatrix}$$

$\underline{\underline{N \times 2}}$

$$\mathcal{GM} = \begin{bmatrix} \mathcal{GM}_1 \\ \mathcal{GM}_2 \\ \vdots \\ \mathcal{GM}_N \end{bmatrix} \updownarrow \; N \times 1 \; \underline{\underline{\text{vector}}}$$

$$\underline{\underline{\mu - \mathcal{GM}}} = \begin{bmatrix} \uparrow & \uparrow \\ (\mu_0 - \mathcal{GM}) & (\mu_0 - \mathcal{GM}) \\ \downarrow & \downarrow \end{bmatrix}$$

$\underline{\underline{N \times 2}}$

$$\frac{\sum_{i=0}^{1} \left(\mu_1^i - \mathcal{GM}_1\right)^2}{2} =$$

$$\frac{W^T \cdot \overbrace{(\mu - \mathcal{GM})}^{N \times 2} \overbrace{(\mu - \mathcal{GM})^T}^{2 \times N} \cdot W}{2}$$

So, $S_B = \underbrace{\frac{(\mu - \zeta M)\cdot(\mu - \zeta M)^T}{2}}_{N \times N \text{ matrix}}$

Between class scatter matrix

$\sigma_B^2 = W^T \cdot S_B \cdot W \longrightarrow$ Scalar Quantity

$$C = \frac{\text{Between Class Variability}}{\text{within class Variability}}$$

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$C(w) = \frac{\sigma_B^2}{\sigma_W^2} = \frac{w^T \cdot S_B \cdot w}{w^T \cdot S_W \cdot w}$$

$S_B \rightarrow 2 \times 2$, $w^T \cdot S_B \cdot w \rightarrow$ scalar

$S_W \rightarrow 2 \times 2$, $w^T \cdot S_W \cdot w \rightarrow$ scalar

$$\max_{w} C(w)$$

$$S_B = \begin{bmatrix} \sigma_{B1}^2 & 0 \\ 0 & \sigma_{B2}^2 \end{bmatrix} \qquad S_W = \begin{bmatrix} \sigma_{W1}^2 & 0 \\ 0 & \sigma_{W2}^2 \end{bmatrix}$$

$$w^T S_B \cdot w = \boxed{w_1^2 \cdot \sigma_{B1}^2 + w_2^2 \sigma_{B2}^2}$$

$$w^T \cdot S_W \cdot w = \boxed{w_1^2 \sigma_{W1}^2 + w_2^2 \cdot \sigma_{W2}^2}$$

$$\underline{\underline{c(w)}} = \underline{\underline{c(w_1, w_2)}} = \frac{w_1^2 \cdot \underline{\sigma_{B1}}^2 + w_2^2 \cdot \underline{\sigma_{B2}}^2}{w_1^2 \cdot \underline{\sigma_{W1}}^2 + w_2^2 \cdot \underline{\sigma_{W2}}^2} \Rightarrow \frac{f(w_1, w_2)}{\underline{\underline{g(w_1, w_2)}}}$$

$$\frac{\partial}{\partial w_1} \cdot c(w_1, w_2) = 0 \Rightarrow \frac{g(w_1, w_2) \cdot 2w_1 \cdot \sigma_{B1}^2 - 2w_1 \sigma_{W_1}^2 \cdot f(w_1, w_2)}{\left(w_1^2 \cdot \sigma_{W_1}^2 + w_2^2 \cdot \sigma_{W2}^2\right)^2} = 0$$

$$g(w_1, w_2) \cdot 2w_1 \cdot \sigma_{B1}^2 - 2w_1 \cdot \sigma_{W_1}^2 \cdot f(w_1, w_2) = 0 \longrightarrow (1)$$

$$\cancel{2} w_1 \cdot \sigma_{B_1}^2 - \cancel{2} w_1 \cdot \sigma_{w_1}^2 \cdot \frac{f(w_1, w_2)}{g(w_1, w_2)} = 0$$

$$\frac{f(w_1, w_2)}{g(w_1, w_2)} w_1 = \frac{\sigma_{B_1}^2}{\sigma_{w_1}^2} w_1 \Rightarrow \underline{\underline{C(W) \cdot w_1 =}} e_1 \cdot w_1 \rightarrow (2)$$

$$\frac{\partial}{\partial w_2} = 0 \Rightarrow \frac{g(w_1, w_2) \cdot 2w_2 \cdot \sigma_{B2}^2 - 2w_2 \cdot \sigma_{w2}^2 \cdot f(w_1, w_2)}{\left(w_1^2 \cdot \sigma_{w_1}^2 + w_2^2 \cdot \sigma_{w_2}^2\right)^2} = 0$$

$$-\cancel{2}w_2 \cdot \sigma_{w_2}^2 \cdot \frac{f(w_1, w_2)}{g(w_1, w_2)} = -\cancel{2}w_2 \cdot \sigma_{B2}^2 \Rightarrow C(w) \cdot w_2 = \frac{\overset{e_2}{\sigma_{B2}^2}}{\sigma_{w2}^2} \cdot w_2$$

$$C(W)\cdot W_1 = \frac{\sigma_{B1}^2}{\sigma_{W1}^2}\cdot W_1$$

$$C(W)\cdot W_2 = \frac{\sigma_{B2}^2}{\sigma_{W2}^2}\cdot W_2$$

$$\begin{bmatrix} \frac{\sigma_{B1}^2}{\sigma_{W1}^2} & 0 \\ 0 & \frac{\sigma_{B2}^2}{\sigma_{W2}^2} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = C(W) \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}$$

$$\longrightarrow \text{(+)}$$

$$\frac{\sigma_{B1}^2}{\sigma_{W1}^2} W_1 = C(W)\cdot W_1$$

$$\frac{\sigma_{B2}^2}{\sigma_{W2}^2} W_2 = C(W)\cdot W_2$$

$$\Rightarrow \begin{bmatrix} \frac{\sigma_{B1}^2}{\sigma_{W1}^2} & 0 \\ 0 & \frac{\sigma_{B2}^2}{\sigma_{W2}^2} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = C(W) \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}$$

$$S_B = \begin{bmatrix} \sigma_{B1}^2 & \underline{0} \\ \underline{0} & \sigma_{B2}^2 \end{bmatrix} \qquad S_B^{-1} = \begin{bmatrix} \frac{1}{\sigma_{B1}^2} & 0 \\ 0 & \frac{1}{\sigma_{B2}^2} \end{bmatrix}$$

$$S_W = \begin{bmatrix} \sigma_{W1}^2 & \underline{0} \\ \underline{0} & \sigma_{W2}^2 \end{bmatrix} \qquad S_W^{-1} = \begin{bmatrix} \frac{1}{\sigma_{W1}^2} & 0 \\ 0 & \frac{1}{\sigma_{W2}^2} \end{bmatrix}$$

$$S_B \cdot S_W^{-1} = \begin{bmatrix} \sigma_{B1}^2 & 0 \\ 0 & \sigma_{B2}^2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_{W1}^2} & 0 \\ 0 & \frac{1}{\sigma_{W2}^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma_{B1}^2}{\sigma_{W1}^2} & 0 \\ 0 & \frac{\sigma_{B2}^2}{\sigma_{W2}^2} \end{bmatrix} \rightarrow \text{Positive Definite}$$

$$= \rightarrow 3$$

PSD

By ③ and ④

Eigen vectors of $S_B \cdot S_W^{-1}$

$$N^* = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = W$$

$$\underbrace{(S_B \cdot S_W^{-1})}_{2\times 2} \cdot \underset{2\times 1}{w} = c\,w \cdot \underset{2\times 1}{w}$$

Simply apply

Apply linear svd $(S_B \cdot S_W^{-1})$

$$\underset{2\times 2}{A} \cdot \underset{2\times 1}{X} = \lambda X$$

$$Q \cdot \Lambda \cdot Q^T$$

$$\underset{2\times 1}{X} \quad \begin{bmatrix} \uparrow & \uparrow \\ W^* & \\ \downarrow & \downarrow \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \leftarrow w_1 \rightarrow \\ \leftarrow w_2 \rightarrow \end{bmatrix}$$

$$w^* = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}$$

$$\begin{bmatrix} w^* & \uparrow & \uparrow \\ \downarrow & \downarrow & \downarrow \end{bmatrix} \longrightarrow N \times 1$$

$$\underbrace{S_B \cdot S_W^{-1}}_{\substack{N \times N \\ A}} \cdot \underset{\substack{N \times 1 \\ x}}{w} = \underset{\lambda}{C(w)} \cdot \underset{x}{w}$$

$$\xrightarrow{\text{svd}} \Rightarrow \underset{N \times N}{Q} \cdot \Lambda \cdot Q^T$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{bmatrix}$$