$$\max_{\|u\|=1} \frac{\|\tilde{\mu}^{(1)} - \tilde{\mu}^{(2)}\|^2}{n_1 \tilde{C}^{(1)} + n_2 \tilde{C}^{(2)}} \quad \underline{\underline{\text{last lecture}}} \quad \max_{\|u\|=1} \frac{u^T S_B u}{u^T S_W u} \qquad (1)$$

where $S_B \stackrel{\text{def}}{=} (\mu^{(1)} - \mu^{(2)})(\mu^{(1)} - \mu^{(2)})^T$ is a sym, pos semi-def matrix; known as the <u>between-class scatter matrix</u>;

$S_W \stackrel{\text{def}}{=} n_1 C_1 + n_2 C_2$ is sym, pos. semi-def

(since C_1, C_2 are sym, pos semi-def) known as the <u>within-class scatter matrix</u>

- If $\exists$ unit u such that $u^T S_W u = 0$, then

$$u^T \underbrace{(n_1 C_1 + n_2 C_2)}_{S_W} u = n_1 \underbrace{u^T C_1 u}_{\geq 0} + n_2 \underbrace{u^T C_2 u}_{\geq 0} = 0$$

$$\Rightarrow \quad u^T C_1 u = u^T C_2 u = 0$$

This means no variance in the projection.
We may choose u as the direction.
(This case is unlikely to happen)

- If $\nexists$ unit u such that $u^T S_W u = 0$,
then for any $v \neq 0$, $v = \|v\| \frac{v}{\|v\|}$

$$v^T S_W v = \|v\|^2 \left(\frac{v}{\|v\|}\right)^T S_W \left(\frac{v}{\|v\|}\right) \neq 0.$$

This means S_W is pos def, thus invertible.
It can be decomposed as (from the EVD)

$$S_W = L^T L \qquad (2)$$

with L invertible.

Notice $$\max_{\|u\|=1} \frac{u^T S_B u}{u^T S_W u} = \max_{v \neq 0} \frac{v^T S_B v}{v^T S_W v} \qquad (3)$$

(since if v^* maximizes the RHS, then

$u^* \overset{\text{def}}{=} \frac{v^*}{\|v^*\|}$ maximizes the **LHS**; if u^* maximizes the LHS, then $v^* \overset{\text{def}}{=} u^*$ maximizes the RHS)

Insert ② into ③ (and rename v as u)

$$\max_{u \neq 0} \frac{u^T S_B u}{u^T S_W u} \overset{S_W = L^T L}{=} \max_{u \neq 0} \frac{u^T S_B u}{u^T L^T L u} \overset{\text{set } w = Lu}{\underset{u = L^{-1} w}{=\!=}}$$

$$\max_{L^{-1} w \neq 0} \frac{(L^{-1} w)^T S_B L^{-1} w}{w^T w} = \max_{w \neq 0} \frac{w^T (L^{-T} S_B L^{-1}) w}{\underbrace{w^T w}_{\|w\|^2}}$$

$$= \max_{w \neq 0} \left(\frac{w}{\|w\|}\right)^T L^{-T} S_B L^{-1} \left(\frac{w}{\|w\|}\right)$$

$$= \max_{\|w\| = 1} w^T (L^{-T} S_B L^{-1}) w$$ (Rayleigh quotient)

which is achieved when w is an eigenvector associated to the largest eigenvalue of $L^{-T} S_B L^{-1}$.

To find u, notice that

λ is an eigenvalue of $L^{-T} S_B L^{-1}$

$\Leftrightarrow$ $L^{-T} S_B L^{-1} w = \lambda w$ for some $w \neq 0$

L invertible

$\Leftrightarrow$ $S_B \underbrace{L^{-1} w}_{= u} = \lambda \underbrace{L^{T} w}_{= L^T L u = S_W u}$ for some $w \neq 0$

$\Leftrightarrow$ $S_B u = \lambda S_W u$ for some $u \neq 0$

S_W invertible

$\Leftrightarrow$ $S_W^{-1} S_B u = \lambda u$ for some $u \neq 0$

$\Leftrightarrow$ λ is an eigenvalue of $S_W^{-1} S_B$.

The argument shows if w is an eigenvector associated to the largest eigenvalue of $L^{-T} S_B L^{-1}$, then u is an eigenvector associated to the largest eigenvalue of $S_W^{-1} S_B$.