1

firstly use $a_1, a_2 \cdots a_m$ and $e_1 \cdots e_m$ to rebuild vector $\hat{x}$

$$x = \sum_{i=1}^{M} a_i e_i$$

$$J = \|x - \hat{x}\|^2 = \|x - \sum_{i=1}^{M} a_i e_i\|^2 = (x - \sum_{i=1}^{M} (a_i e_i))(x - \sum_{i=1}^{M} (a_i e_i))$$

$$= x \cdot x - 2x \cdot (\sum_{i=1}^{M} (a_i e_i)) + (\sum_{i=1}^{M} a_i e_i \cdot \sum_{i=1}^{M} a_i e_i)$$

for eigenvectors are orthonormal, so $e_i \cdot e_j = 0 \ (i \neq j)$ $e_i \cdot e_j = 1$

$\therefore \sum a_i e_i \cdot \sum a_i e_i = \sum a_i^2$

$\therefore J = x \cdot x - 2x \cdot \sum_{i=1}^{M} (a_i e_i) + \sum_{i=1}^{M} (a_i^2)$

Since minimizing of squared error of J with respect of $a_1 \cdots a_m$,

we only need to need to consider the second term, the other terms are not depend on a_i.

$$\frac{\partial J}{\partial a_i} = -2(x \cdot e_i) + 2a_i = 0$$

$$a_i = x e_i$$

$$\therefore a_i = e_i^T x, \quad i = 1, 2 \ldots M$$

② $\begin{cases} y = w^T x \\ \mu_i' = w^T \mu_i \\ \sigma_i' = w^T \Sigma_i w \end{cases}$

to Maximize the speration means, we want to maximize $(\mu_1' - \mu_2')^2$

to minimize the within class variance, we want to minimize $\sigma_1'^2 + \sigma_2'^2$

$$\frac{(\mu_1' - \mu_2')^2}{(\sigma_1'^2 + \sigma_2'^2)} = 1$$

take derivative of 'Lag' with respect to w and set it to 0

$$2(\mu_1' - \mu_2')(\sigma_1'^2 + \sigma_2'^2)w - 2(\mu_1' - \mu_2')(\sigma_1'^2 + \sigma_2'^2)w + 2\lambda w = 0$$

$\Rightarrow 2(\mu_1' - \mu_2')(\sigma_1'^2 + \sigma_2'^2 + \lambda)w = 0$

as for $w \neq 0$

we have $2(\mu_1' - \mu_2')(\sigma_1'^2 + \sigma_2'^2 + \lambda) = 0$

since $(\sigma_1'^2 + \sigma_2'^2 + \lambda) > 0$

$\therefore \mu_1' - \mu_2' = 0 \quad \mu_1' = \mu_2'$

$\therefore$ When $\mu_1' = \mu_2'$, $J_1(w)$ is maximized, which is also satisfied in $J_1(w) = \dfrac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$

(b.) incorporate the prior probabilities $P(w_i)$ into the criterion.

$$\mu_i' = w^T \mu_i$$

variance of projected density; $\sigma_i'^2 = w^T \Sigma_i w$

$\|w\| = 1$

we can get

$$2(u_1'-u_2')P(w_1)(\sigma_1'^2P(w_1)+\sigma_2'^2P(w_2))w-2(u_1'-u_2')P(w_1)(\sigma_1'^2P(w_1)+\sigma_2'^2P(w_2))w+2\lambda w=0$$

$$\Rightarrow 2(u_1'-u_2')P(w_1)(\sigma_1'^2P(w_1)+\sigma_2'^2P(w_2)+\lambda)w=0$$

$w \neq 0$ and $(\sigma_1'^2P(w_1)+\sigma_2'^2P(w_2)+\lambda$ is positive

$$\therefore (u_1'-u_2')P(w_1)=0$$

$$u_1'=u_2'$$

this also stands when $J_2(w)=\dfrac{(u_1-u_2)^2}{P(w_1)\sigma_1^2+P(w_2)\sigma_2^2}$

(c) $J(w_1)$ maximize the speration between the mean of the projected densities while considering the variance of each class.

$J(w_2)$ incorporates the prior probabilites of the classes in addition to the w and σ.

Fisher's LDA usually maximize $J(w_1)$ to find the optimal answer.

Therefore, $J(w_1)$ is closer.