Name : Premashish Mukherjee
G-Number : G01390485

1.21.

To prove

$$a \leq (ab)^{1/2} \rightarrow ①$$

Given

$a \leq b \rightarrow ②$ & a, b are non-negative hence $a \geq 0$ and $b \geq 0$ $\hookrightarrow ③$

Let's write 1 as

$(ab)^{1/2} - a \geq 0$ by substracting a on both side of inequality

Now we need to prove

$$\Rightarrow a^{1/2}\left(b^{1/2} - a^{1/2}\right) \geq 0$$

from eqⁿ ② we have $a \leq b$, hence $b^{1/2} \geq a^{1/2}$, hence $b^{1/2} - a^{1/2} \geq 0$

from eqⁿ ③ we have $a \geq 0$, hence $a^{1/2} \geq 0$, hence

$$a^{1/2} * \left(b^{1/2} - a^{1/2}\right) \geq 0.$$

hence proved.

hence

$$a \leq (ab)^{1/2} \hookrightarrow ④$$

Now in Decision theory, we know that to have less misclassification error, we choose the class with high joint probability which in turn means high posterior probability

i.e. $P(x, C_1) > P(x, C_2)$ for Region R_1 — ⑤

$P(x, C_2) > P(x, C_1)$ for Region R_2 — ⑥

we have

$$P(\text{mistake}) = P(x \in R_1, C_2) + P(x \in R_2, C_1)$$

$$= \int_{R_1} P(x, C_2)\, dx + \int_{R_2} P(x, C_1)\, dx \quad \text{(7)}$$

Now from 5, we have for R_1

$$P(x, C_1) > P(x, C_2)$$

$$P(x, C_2) < P(x, C_1)$$

Also from eqn 4 i.e. $a \leq (ab)^{1/2}$ we can write.

$$\underline{P(x, C_2) < \left(P(x, C_2)\, P(x, C_1)\right)^{1/2}}$$

hence for

$$\underline{\int_{R_1} P(x, C_2)\, dx \leq \int_{R_1} \left(P(x, C_2)\, P(x, C_1)\right)^{1/2} dx \quad \text{(8)}}$$

Similarly using eqn 6 & eqn 4 we have for R_2

$$\underline{\int_{R_2} P(x, C_1)\, dx \leq \int_{R_2} \left(P(x, C_2)\, P(x, C_1)\right)^{1/2} dx \quad \text{(9)}}$$

Combining 8 & 9 we can transform eqn 7 into.

$$P(\text{mistake}) \leq \int \left(P(x, C_2)\, P(x, C_1)\right)^{1/2} dx$$

hence proved

4.4 constraint $\Rightarrow w^T w = 1$

Lagrange multiplier

$$L(x, \lambda) = f(x) + \lambda g(x)$$

here we have a function of w, hence

$$L(w, \lambda) = f(w) + \lambda g(w)$$

$g(w)$ = equality constraint $= w^T w - 1 = 0$

$$f(w) = w^T m_K \rightarrow (4.23) = w^T (m_2 - m_1)$$

hence

$$L(w, \lambda) = w^T m_K + \lambda (w^T w - 1)$$

Taking the derivative w.r.t w for maximization

$$\frac{\partial L(w, \lambda)}{\partial w} = m_K + 2\lambda w$$

setting it to 0 for finding w for maximization

$$m_K + 2\lambda w = 0$$

$$w = \frac{-m_K}{2\lambda} = \frac{-(m_2 - m_1)}{2\lambda}$$

hence $w \propto (m_2 - m_1)$ result of maximization

hence proved

4.5. from $J(w) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}$ → ① to prove $J(w) = \frac{W^T S_B W}{W^T S_W W}$

here $m_2 = W^T M_2$ → sample mean for the projected points → ②

$m_1 = W^T M_1$ → ③

$S_i = \sum_{x \in C_i} (x - M_i)(x - M_i)^T$ → ④

$S_W = S_1 + S_2$ → ⑤

$s_i^2 = \sum_{x \in C_i} (W^T x - m_i)^2$ → ⑥

Now let's take eqⁿ ① denominator $s_1^2 + s_2^2$

Now

$$s_1^2 = \sum_{x \in C_1} (W^T x - m_1)^2$$

from 3 we have

$$= \sum_{x \in C_1} (W^T x - W^T M_1)^2$$

$$= \sum_{x \in C_1} W^T (x - M_1)(x - M_1)^T W$$

from eqⁿ 4 we have

$$= W^T S_1 W \rightarrow ⑦$$

Similarly

$$s_2^2 = \sum_{x \in C_2} (W^T x - m_2)^2 = \sum_{x \in C_2} (W^T x - W^T M_2)^2 = W^T S_2 W$$ → ⑧

from eq^{n} 7 & 8 we have $s_1^2 = W^T S_1 W$

$$s_2^2 = W^T S_2 W$$

~~from eq^{n} 5 we can write~~

$$\cancel{S_W = S_1 + S_2}$$

Now the denominator $s_1^2 + s_2^2 = W^T S_1 W + W^T S_2 W$

$$= W^T (S_1 + S_2) W$$

from eq^{n} 5

$$= W^T (S_W) W \to (9)$$

Now let's take the numerator.

$$\Rightarrow (m_2 - m_1)^2$$

where $m_i = W^T m_i$

hence numerator

$$\Rightarrow (W^T m_2 - W^T m_1)^2 = W^T (m_2 - m_1)(m_2 - m_1)^T W$$

We know from eq^{n} 4.27 of the book

$$S_B = (m_2 - m_1)(m_2 - m_1)^T$$

(between class covariance)

Let substitute S_B in above eq^{n}

$$(m_2 - m_1)^2 = (W^T m_2 - W^T m_1)^2 = W^T (m_2 - m_1)(m_2 - m_1)^T W$$

$$= W^T S_B W \to (10)$$

Now Subsitute 10 & 9 in numerator and denominator.

$$J(w) = \frac{w^T S_B w}{w^T S_w w}$$

hence proved