EJERCICIO 1:

Para demostrar $\longrightarrow$ que el valor de a que maximiza al estadístico en $a = \Sigma - 1X, X(\mu_1 - \mu_2)$, utilizaremos multiplicadores de Lagrange

1ro) escribimos la función objetivo

$$f(a) = \frac{(E[a_1^x] - E[a^t X_2])^2}{Var(a^t X_1 - a^t X_2)}$$

Utilizando las propiedades de la esperanza y varianza, nos queda:

$$f(a) = \frac{(a\mu_1 - at\mu_2)^2}{a^2 Var(X_1) + t^2 Var(X_2) - 2at \underbrace{Cov(X_1, X_2)}_{\text{es la cov. entre } X_1 \text{ y } X_2}}$$

Utilizando multiplicadores de lagrange, buscamos el valor de a que maximiza f(a) sujeto a:

$$g(a) = a - \lambda(a\,2Var(X_1) + t^2 Var(X_2) - 2at\, Cov(X_1, X_2))$$

Derivando f(a) e igualando a cero $\longrightarrow$ encontramos los puntos críticos

$$f'(a) = 2\mu_1(\mu_1 - t\mu_2)(a\,2Var(X_1) + t^2 Var(X_2) - 2at\,Cov(X_1, X_2))\,2 - 2\mu_2(a\mu_1 - \mu_2 t)(a\,2\,Var(X_1)$$
$$+ t^2 Var(X_2) - 2at\,Cov(X_1, X_2))\,2 = \emptyset$$

Resolvemos para a:

$$a(\mu_1 - \mu_2 t) = Cov(X_1, X_2)t$$
$$a = Cov(X_1, X_2)t\mu_1 - \mu_2 t \quad \ldots \circledast$$

Resplazando $\circledast$ en g(a) y derivamos con respecto a λ:

$$g'(a) = 1 - 2\lambda a(Var(X_1) + t^2 Var(X_2) - 2t\,Cov(X_1, X_2)) = \emptyset$$

Resolviendo para λ:

$$\lambda = \frac{1}{2a(Var(X_1) + t^2 Var(X_2) + 2t\,Cov(X_1, X_2))}$$

Sustituyendo los valores de a y λ en f(a), obtenemos:

$$f(a) = \frac{(Cov(X_1, X_2)(\mu_1 - \mu_2 t))^2}{(Var(X_1) + t^2 Var(X_2) - 2t\,Cov(X_1, X_2))(Var(X_1)(\mu_1 - \mu_2 t)^2 + t^2 Var(X_2)(\mu_2 - \mu_1 t)^2)}$$

Para maximizar esta función, debemos buscar el valor de t _ue maximiza f(a)

$$\frac{d}{dt} f(a) = \frac{-2Cov(X_1, X_2)(\mu_1 - \mu_2 t)(Var(X_1) - Var(X_2)t)}{(Var(X_1) + t^2 Var(X_2) - 2t\,Cov(X_1, X_2))^2(\mu_1 - \mu_2 t)^2(\mu_2 - \mu_1 t)^2}$$

Implica que:

$$Cov(X_1, X_2)(Var(X_1) - Var(X_2)t) = \emptyset$$

lo cual :

$$t = \frac{Var(X_1)}{Var(X_2)} \qquad \text{sí} \qquad Cov(X_1, X_2) \neq \varnothing \qquad \ldots$$

Sustituimos a t en la expresion para a:

$$a = \frac{Cov(X_1, X_2)\, Var(X_1 - \mu_1 + \frac{\mu_2}{Var(X_2)} \cdot Cov(X_1, X_2)}{Var(X_1) - 2Cov(X_1, X_2)\frac{Var(X_1)}{Var(X_2)} + t^2\, Var(X_2)}$$

Simplicando y sustituyendo

$$a = \frac{Cov(X_1, X_2)}{Var(X_1) + Var(X_2)} \underbrace{\left(\frac{\mu_1}{Var(X_1)} - \frac{\mu_2}{Var(X_2)}\right)}$$

es la diferencia $\mu_1 - \mu_2$ escalada por los inversos de las Var.

Por lo que se puede escribir

$$a = \alpha(\mu_1 - \mu_2)$$

$$\Longrightarrow \quad \alpha = \frac{Cov(X_1, X_2)}{Var(X_1) + Var(X_2)}$$

y como α no depende ni de μ_1 ni de μ_2, entonces el valor de a que maximiza al estadistico es proporcional a la diferencia $\mu_1 - \mu_2$ y en particular:

$a = \Sigma^{-1}_{X,X}(\mu_1 - \mu_2)$ maximiza la funcion objetivo

■

EJERCICIO 2:

1. LDA se espera que funcione mejor en el conjunto de entrenamiento, ya que asume que las matrices de covarianza de las clases son iguales, lo que implica un límite de decisión lineal. Como el límite de decisión de Bayes también es lineal, LDA debería ajustarse mejor al modelo subyacente.

2. QDA se espera que funcione mejor en el conjunto de entrenamiento, ya que permite modelar límites no lineales al permitir que las matrices de covarianza de las clases sean diferentes. Si el límite de Bayes no es lineal, QDA tiene la capacidad de ajustarse mejor a la complejidad del límite de decisión.

3. A medida que aumenta el tamaño de la muestra N, se espera que la precisión de la prueba de QDA en relación con LDA mejore. Con un tamaño de muestra más grande, QDA tiene más datos para estimar las matrices de covarianza de cada clase de manera más precisa, lo que puede conducir a una mejora en su precisión en comparación con LDA, especialmente cuando las covarianzas de las clases pueden ser diferentes.

4. La afirmación es falsa. Si el límite de decisión de Bayes es lineal, LDA es más adecuado porque asume que las matrices de covarianza de las clases son iguales, lo que permite un límite de decisión lineal. QDA, al permitir diferentes matrices de covarianza para cada clase, podría introducir flexibilidad innecesaria y resultar en un sobreajuste. En este caso, se espera que LDA tenga un mejor desempeño, ya que está mejor alineado con la verdadera estructura del problema.

EJERCICIO 8 :

Y: v.a. si empresa emite dividendos ó no $\begin{cases} 1 & \text{si emite} \\ \emptyset & \text{no emite} \end{cases}$

X: % de beneficio año pasado

$X \sim$ Normal

$$P(Y|X) = \frac{P(X|Y) * P(Y)}{P(X)}$$

Los que emiten dividendos	Los que no emiten dividendos
$\mu_y = 10$	$\mu_z = \emptyset$
$\sigma^2_y = 36$	$\mu_z = 36$

Ahora :

$$P(Y|X=4) = \frac{P(X=4|Y) * P(Y)}{P(X=4)}$$

Entonces :

$$P(X=4|Y) = \frac{1}{\sqrt{2\pi(36)}} e^{-\frac{(4-10)^2}{2(36)}} = \frac{1}{6\sqrt{\pi}} e^{-\frac{1}{2}}$$

$$= 0.1065$$

$$P(X=4|Z) = \frac{1}{\sqrt{2\pi(36)}} e^{-\frac{(4-\emptyset)^2}{2(36)}} = 0.01295$$

$$P(X=4) = P(X=4|4) * P(Y) + P(X=4|Z) * P(Z)$$

$$= 0.1065(.8) + 0.01295(0.2)$$

$$= 0.1114$$

Por ultimo :

$$P(Y|X=4) = \frac{P(X=4|Y) * P(Y)}{P(X=4)} = \frac{0.1065(0.8)}{0.1114} = 0.76481$$