1. (10 puntos) Complete las demostraciones de las reglas de asignación de la presentación de la semana 7.

Regla de asignación

En el análisis discriminante se asigna $X = x$ a la clase con mayor $\delta_k(x)$ donde:

$$\delta_k(x) = x\frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k)$$

Teorema de Bayes: $P(B_i \mid A) = \dfrac{P(B_i \cap A)}{\sum_{\ell=1}^{K} P(B_\ell \cap A)} = \dfrac{P(B_i)P(A \mid B_i)}{\sum_{\ell=1}^{K} P(B_\ell)P(A \mid B_\ell)}$

Aplicación del teorema de Bayes: $P(Y=k \mid X=x) = \dfrac{f_k(x)\pi_k}{\sum_{\ell=1}^{K} f_\ell(x)\pi_\ell}$

$X \sim N(\mu, \sigma^2) \Rightarrow P(Y=k \mid X=x) = \dfrac{\pi_k \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum_{s=1}^{K} \pi_s \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)}$

Note que el denominador no depende de k, por lo que se puede omitir para la comparación de clases. Al aplicar $\log(\pi_k \cdot f_k(x))$ se obtiene:

$$\begin{aligned}
\log(\pi_k \cdot f_k(x)) &= \log(\pi_k) + \log\left(\frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left(\frac{-(x-\mu_k)^2}{2\sigma^2}\right)\right) \\
&= \log(\pi_k) + \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) + \log\left(\exp\left(\frac{-(x-\mu_k)^2}{2\sigma^2}\right)\right) \\
&= \log(\pi_k) + \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{(x-\mu_k)^2}{2\sigma^2} \\
&= \log(\pi_k) + \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{(x^2 - 2x\mu_k + \mu_k^2)}{2\sigma^2} \\
&= \log(\pi_k) + \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{x^2}{2\sigma^2} + \frac{\cancel{2}x\mu_k}{\cancel{2}\sigma^2} - \frac{\mu_k^2}{2\sigma^2}
\end{aligned}$$

Nuevamente se omiten los términos que no dependen de k:

$$= \log(\pi_k) + \frac{x\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2}$$

$$\therefore \delta_k(x) = x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k) \;//$$

Regla de asignación

El Análisis Discriminante asigna $X = x$ a la clase con mayor $\hat{\delta}_k(x)$ donde

$$\hat{\delta}_k(x) = x\frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\sigma^2} + \log\hat{\pi}_k$$

La prueba es análoga a la anterior, $\hat{f}_k(x)$ se define como:

$$\hat{f}_k(x) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} \exp\left(-\frac{(x-\hat{\mu}_k)^2}{2\hat{\sigma}^2}\right)$$

Se aplica $\log(\hat{\pi}_k \cdot \hat{f}_k(x))$:

$$\begin{aligned}
\log(\hat{\pi}_k \cdot \hat{f}_k(x)) &= \log(\hat{\pi}_k) + \log\left(\frac{1}{\sqrt{2\pi}\,\hat{\sigma}} \cdot \exp\left(\frac{-(x-\hat{\mu}_k)^2}{2\hat{\sigma}^2}\right)\right) \\
&= \log(\hat{\pi}_k) + \log\left(\frac{1}{\sqrt{2\pi}\,\hat{\sigma}}\right) + \log\left(\exp\left(\frac{-(x-\hat{\mu}_k)^2}{2\hat{\sigma}^2}\right)\right) \\
&= \log(\hat{\pi}_k) + \log\left(\frac{1}{\sqrt{2\pi}\,\hat{\sigma}}\right) - \frac{(x-\hat{\mu}_k)^2}{2\hat{\sigma}^2} \\
&= \log(\hat{\pi}_k) + \log\left(\frac{1}{\sqrt{2\pi}\,\hat{\sigma}}\right) - \frac{(x^2 - 2x\hat{\mu}_k + \hat{\mu}_k^2)}{2\hat{\sigma}^2} \\
&= \log(\hat{\pi}_k) + \log\left(\frac{1}{\sqrt{2\pi}\,\hat{\sigma}}\right) - \frac{x^2}{2\hat{\sigma}^2} + \frac{\cancel{2}x\hat{\mu}_k}{\cancel{2}\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2}
\end{aligned}$$

Nuevamente se omiten los términos que no dependen de k:

$$= \log(\hat{\pi}_k) + \frac{x\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2}$$

$$\therefore \delta_k(x) = x \cdot \frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + \log(\hat{\pi}_k) \;//$$

En el caso de $p > 1$ predictores, el clasificador LDA asume que las individuos en la clase k sigue una distribución Gaussiana Multivariada con media μ_k, y Σ es la matriz de covarianzas igual para todas las clases.

Regla de asignación

El análisis discriminante asigna $X = x$ a la clase con mayor $\delta_k(x)$ donde:

$$\delta_k(x) = x^t\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^t\Sigma^{-1}\mu_k + \log(\pi_k)$$

La Normal Multivariada se define como:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\cdot(x-\mu_k)^t\Sigma^{-1}(x-\mu_k)\right)$$

Aplicando $\log(\pi_k \cdot f_k(x))$:

$$\log(\pi_k \cdot f_k(x)) = \log(\pi_k) + \frac{-1}{2}\cdot(x-\mu_k)^t\Sigma^{-1}(x-\mu_k)$$

$$= \log(\pi_k) - \frac{\Sigma^{-1}}{2}[x^t\cdot x - x^t\cdot\mu_k - x\mu_k^t + \mu_k^t\mu_k]$$

$$= \log(\pi_k) + \frac{1}{2}\cdot x^t\Sigma^{-1}\mu_k + \frac{1}{2}\cdot\mu_k^t\Sigma^{-1}x - \frac{1}{2}\cdot\mu_k^t\Sigma^{-1}\mu_k$$

Note que $x^t\Sigma^{-1}\mu_k = \mu_k^t\Sigma^{-1}x$, por tanto:

$$= \log(\pi_k) + x^t\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^t\Sigma^{-1}\mu_k$$

$$\delta_k(x) = x^t\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^t\Sigma^{-1}\mu_k + \log(\pi_k)$$

Regla de asignación

El QDA asigna $X = x$ a la clase con mayor $\delta_k(x)$ donde:

$$\delta_k(x) = -\frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}(x-\mu_k)^T\boldsymbol{\Sigma}_k^{-1}(x-\mu_k) + \log\pi_k.$$

$$= -\frac{1}{2}\mathbf{x}^t\Sigma_k^{-1}\mathbf{x} + \mathbf{x}^t\Sigma_k^{-1}\mu_k - \frac{1}{2}\mu_k^t\Sigma_k^{-1}\mu_k + \log(\pi_k) - \frac{1}{2}\log(|\Sigma_k|)$$

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}}\exp\left(-\frac{1}{2}\cdot(x-\mu)^t\Sigma_k^{-1}(x-\mu_k)\right)$$

Aplicando $\log(\pi_k\cdot f_k(x))$:

$$\log(\pi_k\cdot f_k(x)) = \log(\pi_k) - \frac{1}{2}\cdot(x-\mu_k)^t\Sigma_k^{-1}(x-\mu_k) - \log(|\Sigma_k|)^{1/2}$$

$$= \log(\pi_k) - \frac{1}{2}\cdot\Sigma_k^{-1}(x^t\cdot x - x^t\mu_k - \mu_k^t x + \mu_k^t\mu_k) - \frac{1}{2}\log(|\Sigma_k|)$$

$$= \log(\pi_k) - \frac{1}{2}x^t\Sigma_k^{-1}x + \frac{1}{2}x^t\Sigma_k^{-1}\mu_k + \frac{1}{2}\mu_k^t\Sigma_k^{-1}x - \frac{1}{2}\mu_k^t\Sigma_k^{-1}\mu_k - \frac{1}{2}\log(|\Sigma_k|)$$

Note que $x^t\Sigma_k^{-1}\mu_k = \mu_k^t\Sigma_k^{-1}x$, por tanto:

$$= \log(\pi_k) - \frac{1}{2}x^t\Sigma_k^{-1}x + x^t\Sigma_k^{-1}\mu_k - \frac{1}{2}\mu_k^t\Sigma_k^{-1}\mu_k - \frac{1}{2}\log(|\Sigma_k|)$$

$$\therefore\ \delta_k(x) = -\frac{1}{2}\log(|\Sigma_k|) - \frac{1}{2}x^t\Sigma_k^{-1}x + x^t\Sigma_k^{-1}\mu_k - \frac{1}{2}\mu_k^t\Sigma_k^{-1}\mu_k + \log(\pi_k)\ //$$

3) **Demuestre lo siguiente:**

a) $V = V_B + V_W$

b) $\sum_{s=1}^{r} q_s g_s = 0$, por tanto rango$(C_g) \leq r - 1$

c) rango(C_g) = rango(V_B)

a) $V = V_B + V_W$

Definiciones: $V = X^t D X = \sum_{i=1}^{n} p_i x_i x_i^t = \sum_{s=1}^{r} \sum_{i \in C_s} p_i x_i x_i^t$ (matriz covarianza total)

$V_B = \sum_{s=1}^{r} q_s g_s g_s^t = C_g^t D_q C_g$ (matriz covarianza inter-clase)

$V_W = \sum_{s=1}^{r} q_s V_s = \sum_{s=1}^{r} \sum_{i \in C_s} p_i (x_i - g_s)(x_i - g_s)^t$ (matriz covarianza intra-clase)

$$V_B + V_W = \sum_{s=1}^{r} q_s g_s g_s^t + \sum_{s=1}^{r} \sum_{i \in C_s} p_i (x_i - g_s)(x_i - g_s)^t$$

$$= \sum_{s=1}^{r} q_s g_s g_s^t + \sum_{s=1}^{r} \sum_{i \in C_s} p_i (x_i - g_s)(x_i^t - g_s^t)$$

$$= \sum_{s=1}^{r} q_s g_s g_s^t + \sum_{s=1}^{r} \sum_{i \in C_s} p_i (x_i x_i^t - x_i g_s^t - x_i^t g_s + g_s g_s^t)$$

$$= \sum_{s=1}^{r} q_s g_s g_s^t + \sum_{s=1}^{r} \sum_{i \in C_s} p_i x_i x_i^t - p_i x_i g_s^t - p_i x_i^t g_s + p_i g_s g_s^t$$

Observe que $q_s = \sum_{i \in C_s} p_i$ y $g_s = \frac{1}{q_s} \sum_{i \in C_s} p_i x_i \Rightarrow \sum_{i = C_s} p_i x_i = q_s \cdot g_s$, por lo tanto:

$$= \sum_{s=1}^{r} \sum_{i \in C_s} p_i x_i x_i^t + \sum_{s=1}^{r} \cancel{q_s g_s g_s^t} - \cancel{q_s g_s g_s^t} - \cancel{q_s g_s^t g_s} + \cancel{q_s g_s g_s^t}$$

$$= \sum_{s=1}^{r} \sum_{i \in C_s} p_i x_i x_i^t$$

$$= V //$$

b) $\sum_{s=1}^{r} q_s g_s = 0$, por tanto rango$(C_g) \leq r - 1$

$$\sum_{s=1}^{r} q_s g_s = \sum_{s=1}^{r} \cancel{q_s} \cdot \frac{1}{\cancel{q_s}} \sum_{i \in C_s} p_i x_i$$

$$= \sum_{s=1}^{r} \sum_{i \in C_s} p_i x_i$$

$$= \sum_{i=1}^{n} p_i x_i$$

$= 0 \rightarrow$ Debido a que X está centrada

El rango es el número máximo de filas o columnas linealmente independientes de la matriz, como $\sum_{s=1}^{r} q_s g_s = 0$, entonces los g_s no son todos independientes. Por lo tanto,

$\text{rango}(C_g) \leq r-1$ //

c) $\text{rango}(C_g) = \text{rango}(V_B)$

Por definición, $V_B = C_g^t D_q C_g$. Si se toma $x \in \mathbb{R}^r$:

$V_B x = 0 \Leftrightarrow x^t C_g^t D_q C_g x = 0$

$\Rightarrow (C_g x)^t D_q (C_g x) = 0$

Sabemos que $C_g x$ es la matriz con filas $g_s^t x \geq 0$ y D_q la matriz $\text{diag}(q_j)$ donde $q_j > 0$. De forma que $(C_g x)^t D_q (C_g x) = 0$ solamente cuando $g_s^t x = 0 \Rightarrow C_g x = 0$. Tenemos que V_B y C_g tienen el mismo núcleo $\Rightarrow$ $\text{rango}(V_B) = \text{rango}(C_g)$:

$\text{rango}(V_B) = p - \dim(\text{Ker}(V_B)) = p - \dim(\text{Ker}(C_g)) = \text{rango}(C_g)$. //

debido al teorema de rango-nulidad (p = número columnas).

Ejercicio 6

Se tiene la observación $t=(\text{Isabel}, F, 4, ?)$ por ende se quiere calcular que:

$$P(\text{Clase}=c \mid \text{Género}=F, \text{Altura}=4)$$

En los datos de entrenamiento tenemos que hay un 14 y por clase se tiene:

$P:3 \wedge M:7 \wedge A:4$ personas

Por en se tienen las probabilidades a priori:

$Pr(P)=\frac{4}{15}$, $Pr(M)=\frac{8}{15} \wedge Pr(A)=\frac{3}{15}$

Ahora veamos la verosimilitudes condicionales:

- Para P:

$$Pr(F \mid P)=\frac{3}{4} \quad \wedge \quad Pr(\text{Altura}=4 \mid P)=\frac{0}{4}$$

$$\Rightarrow P(F,4 \mid M)\cdot P(M)=\frac{3}{4}\cdot 0 \cdot \frac{4}{15}=0 \;//$$

- Para M:

$$Pr(F \mid M)=\frac{6}{8} \quad \wedge \quad Pr(\text{Altura}=4 \mid M)=\frac{3}{8}$$

$$\Rightarrow Pr(F,4 \mid M)\cdot Pr(M)=\frac{6}{8}\cdot\frac{3}{8}\cdot\frac{8}{15}=\boxed{\frac{144}{960}}$$

- Para A:

$$Pr(F \mid A)=\frac{0}{3} \quad \wedge \quad Pr(\text{Altura}=4 \mid A)=\frac{0}{3}$$

$$\Rightarrow Pr(F,4 \mid A)=0 \;//$$

Por lo tanto se se sigue:

$$Pr(M \mid F,4)=\frac{0{,}15}{0{,}15+0+0}=1 \;//$$

Se concluye que Isabel será clasificada en la clase M con una proba 1 por medio del método de Naive Bayes.

7. (20 puntos) Para la siguiente tabla, se tiene una nueva fila: $12 = (1, 3, 2, 4, ?)$. Prediga manualmente si el individuo es buen pagador o mal pagador.

Id	Monto Crédito	Ingreso Neto	Monto Cuota	Grado Académico	Buen Pagador
1	2	4	1	4	Sí
2	2	3	1	4	Sí
3	4	1	4	2	No
4	1	4	1	4	Sí
5	3	3	3	2	No
6	3	4	1	4	Sí
7	4	2	3	2	No
8	4	1	3	2	No
9	3	4	1	3	Sí
10	1	3	2	4	Sí
11	1	4	2	4	Sí

Se utiliza el método de Bayes, de forma que:

$$P(Y=y \mid X=x) = P(Y=k \mid X=x)$$
$$= \frac{P(X=x \mid Y=y)P(Y=k)}{\sum_{i=1}^{k} P(X=x \mid Y=i)P(Y=i)}$$

En este caso, $P(BP=1 \mid X=x) = \dfrac{P(X=x \mid BP=1)\,P(BP=1)}{P(X=x \mid BP=1)P(BP=1) + P(X=x \mid BP=0)\,P(BP=0)}$

donde $X = 1,3,2,4$, $P(BP=1) = \frac{7}{11} \approx 0.6364$, $P(BP=0) = \frac{4}{11} \approx 0.3636$

Caso $BP = Sí = 1$:

$P(\text{Monto Crédito} = 1 \mid BP=1) = \frac{3}{7}$
$P(\text{Ingreso Neto} = 3 \mid BP=1) = \frac{2}{7}$
$P(\text{Monto Cuota} = 2 \mid BP=1) = \frac{2}{7}$
$P(\text{Grado Académico} = 4 \mid BP=1) = \frac{6}{7}$

$\Rightarrow \frac{3}{7} \cdot \frac{2}{7} \cdot \frac{2}{7} \cdot \frac{6}{7} \approx 0{,}02999$

Caso $BP = No = 0$:

$P(\text{Monto Crédito} = 1 \mid BP=0) = \frac{0}{5}$
$P(\text{Ingreso Neto} = 3 \mid BP=0) = \frac{1}{5}$
$P(\text{Monto Cuota} = 2 \mid BP=0) = \frac{0}{5}$
$P(\text{Grado Académico} = 4 \mid BP=0) = \frac{0}{5}$

$\Rightarrow 0$

Por lo tanto, $P(BP=1 \mid X=x) = \dfrac{0.2999 \cdot 0.6364}{0.2999 \cdot 0.6364 + 0 \cdot 0.3636}$
$= 1$

$\therefore$ El individuo es buen pagador. //