Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politécnica de Valéncia, Junio de 2021

Apellidos: [ ] **Nombre:** [ ]

Profesor: ☐ **Jorge Civera** ☐ **Carlos Martínez**

Cuestiones (1.5 puntos, 30 minutos, sin apuntes)

$\boxed{\text{B}}$ Indicar cuál de las siguientes fórmulas se corresponde a una distribución Bernoulli unidimensional de parámetro p siendo x una variable binaria:

A) $p(x) = p^{1-x}$
B) $p(x) = (1-p)^{1-x}p^x$
C) $p(x) = p^x(1-p)^x$
D) $p(x) = p^x(1-p)^{x-1}$

$\boxed{\text{C}}$ Dado un problema de clasificación en un espacio tridimensional en tres clases equiprobables, con probabilidades condicionadas gaussianas de parámetros $\mu_A = (-1\ 3\ -2)$, $\mu_B = (1\ -1\ 2)$ y $\mu_C = (0\ 0\ 1)$, con $\Sigma = I$ común, el vector $\mathbf{x} = (1\ 1\ 0)$ se clasificaría en la clase (Nota: $c^*(\mathbf{x}) = \arg\max_c \mu_c^t\Sigma^{-1}\mathbf{x} + \log P(c) - \frac{1}{2}\mu_c^t\Sigma^{-1}\mu_c$):

A) A
B) B
C) C
D) Hay empate

$\boxed{\text{A}}$ Dado el siguiente conjunto de muestras en $\{0,1\}^4$:

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$
x_{n1}	0	1	1	1	1	1
x_{n2}	1	0	1	1	0	1
x_{n3}	0	0	1	1	1	1
x_{n4}	1	0	1	1	1	1
c	A	A	A	B	B	B

la estimación de los parámetros por máxima verosimilitud de distribuciones Bernouilli daría como resultado:

A) $\mathbf{p}_A = (\frac{2}{3}\frac{2}{3}\frac{1}{3}\frac{2}{3})$, $\mathbf{p}_B = (\frac{3}{3}\frac{2}{3}\frac{3}{3}\frac{3}{3})$
B) $\mathbf{p}_A = (\frac{2}{4}\frac{1}{4}\frac{4}{4})$, $\mathbf{p}_B = (\frac{4}{4}\frac{3}{4}\frac{4}{4})$
C) $\mathbf{p}_A = (\frac{5}{6}\frac{4}{6}\frac{4}{6}\frac{5}{6})$, $\mathbf{p}_B = (\frac{1}{6}\frac{2}{6}\frac{2}{6}\frac{1}{6})$
D) $\mathbf{p}_A = (\frac{2}{7}\frac{2}{7}\frac{1}{7}\frac{2}{7})$, $\mathbf{p}_B = (\frac{3}{11}\frac{2}{11}\frac{3}{11}\frac{3}{11})$

$\boxed{\text{D}}$ Dada la distribución multinomial con parámetro $\mathbf{p} = \left(\frac{1}{10}\ \frac{7}{10}\ 0\ \frac{2}{10}\right)$, al aplicar suavizado por descuento absoluto y posterior *backoff* usando $\epsilon = \frac{1}{20}$ y la distribución uniforme, ¿qué afirmación es correcta respecto al parámetro suavizado resultante?

A) La componente que inicialmente tiene mayor valor no se ve alterada
B) Las componentes que originalmente no son nulas no se ven alteradas
C) La componente de menor valor sigue siendo la misma que inicialmente tenía menor valor
D) La componente de menor valor resultante no es la misma que en el original

$\boxed{\text{A}}$ Dada la función Kernel $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^t\mathbf{y})^2$, la matriz Gramm para el par de muestras $\mathbf{x} = (1\ 1)^t$ y $\mathbf{y} = (-1\ 1)^t$ es:

A) Una matriz diagonal
B) La matriz identidad
C) La matriz nula
D) Una matriz no simétrica

$\boxed{\text{C}}$ Si se tienen un par de kernels $K_1(\mathbf{x}, \mathbf{y})$ y $K_2(\mathbf{x}, \mathbf{y})$, ¿cuál de las siguientes combinaciones sería un kernel?

A) $K_1(\mathbf{x}, \mathbf{y}) - K_2(\mathbf{x}, \mathbf{y})$
B) $K_1(\mathbf{x}, \mathbf{y}) + K_2(\mathbf{x}, \mathbf{y})^{-1}$
C) $K_1(\mathbf{x}, \mathbf{y}) \cdot K_2(\mathbf{x}, \mathbf{y})$
D) $K_1(\mathbf{x}, \mathbf{y}) \cdot K_2(\mathbf{x}, \mathbf{y})^{-1}$

$\boxed{\text{B}}$ A la hora de combinar reducciones de dimensión por PCA seguida por LDA hay que tener en cuenta que:

A) PCA tiene su dimensión destino limitada por el número de muestras
B) La dimensión final está restringida por el número de clases
C) Sólo debe hacerse cuando las clases son originalmente linealmente separables
D) La dimensión intermedia está restringida por el número de clases

$\boxed{\text{D}}$ Las fuentes de error en clasificación sobre las que puede actuar la combinación de clasificadores son:

A) Ruido y varianza
B) Sesgo y ruido
C) Sesgo, ruido y varianza
D) Sesgo y varianza

Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politécnica de Valéncia, Junio de 2021

Apellidos: ______ **Nombre:** ______

Profesor: ☐ **Jorge Civera** ☐ **Carlos Martínez**

Problemas (2 puntos, 90 minutos, con apuntes)

1. **(1 punto)** Se tiene el conjunto de datos siguiente:

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$	$\mathbf{x}_8$
x_{n1}	$\frac{1}{2}$	0	$-\frac{1}{4}$	$-\frac{1}{4}$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
x_{n2}	$\frac{1}{2}$	$-\frac{3}{2}$	1	0	$\frac{3}{2}$	$-\frac{1}{2}$	2	1
x_{n3}	$\frac{1}{7}$	$-\frac{3}{7}$	$-\frac{5}{7}$	1	$-\frac{2}{7}$	$-\frac{6}{7}$	$-\frac{8}{7}$	$\frac{4}{7}$
c_n	A	A	A	A	B	B	B	B

Se pide lo siguiente:

a) Calcular todos los parámetros del clasificador gaussiano por máxima verosimilitud para ese conjunto de datos. **(0.5 puntos)**

b) Establecer la frontera de decisión entre las dos clases. **(0.3 puntos)**

c) Clasificar el punto $\mathbf{y} = (1\ \frac{1}{2}\ -1)^t$. **(0.2 puntos)**

Solución:

a) $P(A) = P(B) = 0.5$, $\mu_A = (0\ 0\ 0)$, $\mu_B = \left(\frac{1}{2}\ 1\ -\frac{3}{7}\right)$

$\Sigma_A = \Sigma_B = \begin{pmatrix} \frac{3}{32} & 0 & 0 \\ 0 & \frac{7}{8} & 0 \\ 0 & 0 & \frac{3}{7} \end{pmatrix}$

Como se puede ver, la matriz de covarianzas es común para ambas clases

b) Para matriz de covarianzas común y al ser las clases equiprobables, las funciones discriminantes son:

$$g_c(\mathbf{x}) = \mu_c^t \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mu_c^t \Sigma^{-1} \mu_c$$

Por tanto:

$$g_A(\mathbf{x}) = 0 \qquad g_B(\mathbf{x}) = \frac{16}{3} x_1 + \frac{8}{7} x_2 - x_3 - \frac{89}{42}$$

Por tanto, la frontera de decisión es: $\frac{16}{3} x_1 + \frac{8}{7} x_2 - x_3 - \frac{89}{42} = 0$

c) Calculamos el valor de las funciones discriminantes para el punto dado:

$$g_A(\mathbf{y}) = 0 \qquad g_B(\mathbf{y}) = \frac{201}{42}$$

Con lo cual se clasifica en la clase B

2. **(0.5 puntos)** Dado el conjunto de datos siguiente:

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$
x_{n1}	0	2	-1	-2	3	2
x_{n2}	1	2	-2	0	0	-2
c_n	A	A	B	B	C	C

Se pide:

a) Calcular las matrices S_b y S_w asociadas a los mismos. (**0.4 puntos**)

b) ¿Es necesario aplicar una reducción de dimensión por LDA a una única dimensión para mejorar una clasificación basada en clasificadores lineales? (**0.1 puntos**)

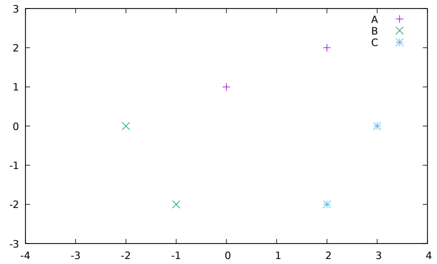
Solución:

a) $\mu = \left(\frac{2}{3}\ -\frac{1}{6}\right)$, $\mu_A = \left(1\ \frac{3}{2}\right)$, $\mu_B = \left(-\frac{3}{2}\ -1\right)$, $\mu_C = \left(\frac{5}{2}\ -1\right)$,

$$S_b = 2(\mu_A - \mu)(\mu_A - \mu)^t + 2(\mu_B - \mu)(\mu_B - \mu)^t + 2(\mu_C - \mu)(\mu_C - \mu)^t = \begin{pmatrix} \frac{49}{3} & \frac{5}{3} \\ \frac{5}{3} & \frac{25}{3} \end{pmatrix}$$

$$\Sigma_A = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} \end{pmatrix},\ \Sigma_B = \begin{pmatrix} \frac{1}{4} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix},\ \Sigma_C = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix},\ S_w = \Sigma_A + \Sigma_B + \Sigma_C = \begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{9}{4} \end{pmatrix}$$

b) A la vista de la representación gráfica de los datos en la dimensionalidad original, ya existe una separación lineal entre las clases, y por lo tanto no sería necesario.



3. **(0.5 puntos)** Se tiene el siguiente conjunto de datos y clasificadores lineales:

$$\mathbf{x}_1 = ((0,0,0),-1), \mathbf{x}_2 = ((1,1,1),+1), \mathbf{x}_3 = ((-1,0,1),-1), \mathbf{x}_4 = ((-1,-1,1),+1)$$

$$g_1(\mathbf{z}) = \begin{cases} +1 & z_1 > 0 \\ -1 & z_1 \leq 0 \end{cases} \quad g_2(\mathbf{z}) = \begin{cases} +1 & z_2 > -1 \\ -1 & z_2 \leq -1 \end{cases} \quad g_3(\mathbf{z}) = \begin{cases} +1 & z_3 > 1 \\ -1 & z_3 \leq 1 \end{cases} \quad g_4(\mathbf{z}) = \begin{cases} +1 & z_1 + z_2 + z_3 \geq 0 \\ -1 & z_1 + z_2 + z_3 < 0 \end{cases}$$

Se pide realizar una primera iteración de AdaBoost sobre estos datos y clasificadores, indicando la tabla de acierto y fallo por clasificador, el clasificador escogido, el error en primera iteración (ϵ_1), el peso del clasificador escogido (α_1) y los pesos de las muestras en la siguiente iteración ($\mathbf{w}^{(2)}$).

Solución:

Tabla acierto/fallo:

	g_1	g_2	g_3	g_4
x_1	✓	X	✓	X
x_2	✓	✓	X	✓
x_3	✓	X	✓	X
x_4	X	X	X	X

Vector de pesos de muestras inicial: $\mathbf{w}^{(1)} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$

Error ponderado por clasificador: $g_1 : \frac{1}{4}$, $g_2 : \frac{3}{4}$, $g_3 : \frac{2}{4}$, $g_4 : \frac{3}{4}$

Clasificador escogido: g_1

Error de clasificación: $\epsilon_1 = \frac{1}{4}$

Peso del clasificador: $\alpha_1 = \frac{1}{2}\log 3$

Pesos de las muestras en la siguiente iteración:

	$w_n^{(1)} \exp(-c_n \alpha_1 C_1(\mathbf{x}_n)$	$\mathbf{w}^{(2)}$
$\mathbf{x}_1$	$\frac{1}{4}\frac{1}{\sqrt{3}}$	$\frac{1}{6}$
$\mathbf{x}_2$	$\frac{1}{4}\frac{1}{\sqrt{3}}$	$\frac{1}{6}$
$\mathbf{x}_3$	$\frac{1}{4}\frac{1}{\sqrt{3}}$	$\frac{1}{6}$
$\mathbf{x}_4$	$\frac{1}{4}\sqrt{3}$	$\frac{1}{2}$
Suma	$\frac{3}{4}\frac{1}{\sqrt{3}} + \frac{1}{4}\sqrt{3} = \frac{\sqrt{3}}{2}$	