Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politécnica de Valéncia, Junio de 2019

Apellidos: | | **Nombre:** | |

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

Cuestiones (2 puntos, 30 minutos, sin apuntes)

$\boxed{\text{C}}$ Indica cuál de los siguientes es un clasificador de Bayes para un objeto x sobre un conjunto de clases $\mathbb{C}$:

A) $\arg\max_{c\in\mathbb{C}} p(x|c)$
B) $\arg\max_{c\in\mathbb{C}} P(c|x)^{-1}$
C) $\arg\min_{c\in\mathbb{C}} -\log P(c|x)$
D) $\arg\min_{c\in\mathbb{C}} \log P(c)p(x|c)$

$\boxed{\text{A}}$ La reformulación de la teoría de la decisión estadística en el caso interactivo se basa en incluir como factor condicionante, aparte del objeto a clasificar x:

A) La realimentación del usuario f
B) El modelo obtenido por entrenamiento *off-line* M
C) La clasificación no interactiva $\hat{c}(x)$
D) El objeto x en otra modalidad

$\boxed{\text{C}}$ Tenemos una extracción por características locales de 11×11 píxeles sobre una imagen de 64×128 píxeles. ¿Cuál es el máximo número de puntos de interés que podría haber?

A) Menos de 4000
B) Entre 4000 y 6000
C) Entre 6001 y 8000
D) Más 8000

n = (((Vy - C + 1) / Dv) * ((Vx - L + 1) / Dh))

n = (64 - 11 + 1) * (128 - 11 + 1) = 6372

$\boxed{\text{A}}$ Los modelos más usados actualmente en reconocimiento de habla continua son:

A) Los basados en *deep learning*
B) Los modelos ocultos de Markov (HMM) continuos con mixturas de gaussianas
C) Los modelos ocultos de Markov (HMM) discretos
D) Los modelos lineales que emplean segmentación de traza

$\boxed{\text{B}}$ Ante una colección de documentos donde hay un token que aparece una sola vez en uno solo de los documentos, ¿qué afirmación es cierta sobre las funciones globales aplicadas sobre ese token?

A) Las funciones globales normal, GfIdf e Idf valen 1
B) Las funciones globales normal y GfIdf valen 1, pero no necesariamente Idf
C) Las funciones globales normal e Idf valen 1, pero no necesariamente GdIdf
D) Las funciones globales GfIdf e Idf valen 1, pero no necesariamente la normal

$\boxed{\text{D}}$ ¿En cuál de las siguientes situaciones tiene menos sentido aplicar reducción de dimensionalidad?

A) La cantidad de parámetros del modelo puede desbordar la memoria
B) La dimensionalidad intrínseca es menor a la obtenida en la representación -> la de la representación debe ser menor.
C) Existen valores correlados en la representación
D) El número de muestras por clase es superior a la dimensión

$\boxed{\text{B}}$ ¿Qué características presenta la reducción de dimensión por PCA?

A) Preserva la continuidad, la discriminación y la invarianza
B) Minimiza el error de reconstrucción
C) Optimiza la cohesión intraclase
D) Minimiza el error de clasificación

$\boxed{\text{D}}$ ¿Cuál es el rango máximo de la matriz entre-clases S_b para un conjunto de muestras de $\mathbb{R}^D$ pertenecientes a C clases distintas?

A) D
B) $D - 1$
C) C
D) $C - 1$

Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politécnica de Valéncia, Junio de 2019

Apellidos: ______________________ **Nombre:** ____________

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

Problemas (4 puntos, 90 minutos, con apuntes)

1. (**1 punto**) Calcula el espacio ocupado en memoria por las siguientes representaciones:

   a) Representación global de una imagen de 64 niveles de gris, de 1280 × 1024 píxeles, por histograma (**0.2 puntos**)

   b) Representación por características locales obtenida por rejilla, con intervalos horizontal y vertical de 1 píxel y tamaño de ventana 15 × 21 (alto por ancho), en una imagen de 256 niveles de gris y tamaño 514 × 620 píxeles, y representación directa (**0.3 puntos**) n = (((Vy - C + 1) / Dv) * ((Vx - L + 1) / Dh)) = (514 - 15 + 1) * (620 - 21 + 1) = 300000.
   s = nd * ((log2(niveles de gris)) / 8) = 315.

   c) Representación de una señal de audio estéreo, con muestras de 8 bits y muestreada a 16KHz, con una duración de media hora (**0.25 puntos**) TAMAÑO (B) = DURACIÓN (s) * FRECUENCIA (Hz) * TAMAÑO MUESTRA (B) * nº de canales

   d) Representación de una señal de audio de 3 minutos de duración, para un sistema 5.1, con muestras de 32 bits y ancho de banda 18KHz, muestreada a la frecuencia mínima posible para su reproducción fiel (**0.25 puntos**)

   Solución: up_rounding((log2(n + 1)) / 8) = 2.6 (se redondea a 3).

   a) 192 bytes Tamaño = niveles de gris * up_rounding((log2(n + 1)) / 8) B = 64 * 3 B = 192 B.

   b) 9450000 bytes Tamaño = n * s = 300000 * 315 = 94500000 B.

   c) 57600000 bytes TAMAÑO (B) = 1800 * 16000 * 1 * 2 = 57600000 B.

   d) 155520000 bytes TAMAÑO (B) = 180 * 18000 * 4 * 6 * 2 = 155520000 B.

2. **(1 punto)** Sea una colección de documentos con un número de documentos D par y no nulo. En dicha colección se han detectado tres tipos de *tokens* básicos:

   - t_1: aparece una sola vez en todos los documentos
   - t_2: aparece dos veces en la mitad de documentos y ninguna vez en la otra mitad
   - t_3: aparece D veces en un único documento y ninguna en el resto

   Se pide:

   a) Indicar el *token* que da mayor valor para cada una de las funciones globales *Normal*, *GfIdf* e *Idf* suponiendo que la colección tiene más de dos documentos (**0.7 puntos**)

   b) ¿Qué ocurre con dichas funciones globales para esos *tokens* cuando $D = 2$? (**0.3 puntos**)

   Solución:

   a) Haciendo los cálculos de las funciones globales:

$$Normal(t_1) = \left(\sum_{d=1}^{D} 1^2\right)^{-\frac{1}{2}} = D^{-\frac{1}{2}} = \frac{1}{\sqrt{D}}$$

$$Normal(t_2) = \left(\sum_{d=1}^{D/2} 2^2\right)^{-\frac{1}{2}} = (2D)^{-\frac{1}{2}} = \sqrt{2D}$$

$$Normal(t_3) = (D^2)^{-\frac{1}{2}} = \frac{1}{D}$$

$$GfIdf(t_1) = \frac{D}{D} = 1 \qquad GfIdf(t_2) = \frac{D}{D/2} = 2 \qquad GfIdf(t_2) = \frac{D}{1} = D$$

$$Idf(t_1) = \log\frac{D}{D} = 0 \qquad Idf(t_2) = \log\frac{D}{D/2} = \log 2 \qquad Idf(t_3) = \log\frac{D}{1} = \log D$$

   Por tanto, los *tokens* con mayor valor para cada función son: *Normal* $= t_2$, *GfIdf* $= t_3$, *Idf* $= t_3$

   b) En el caso de $D = 2$, los valores de *GfIdf* e *Idf* para t_2 y t_3 se igualan

3. (**2 puntos**) Se dispone de un conjunto de muestras en $\mathbb{R}^3$ clasificadas en dos clases:

n	1	2	3	4	5	6	7	8
x_1	1	1	1	1	-1	-1	-1	-1
x_2	3	2	-2	-3	3	-2	2	-3
x_3	3	-2	2	-3	3	2	-2	-3
c	A	A	A	A	B	B	B	B

Por otra parte se ha calculado LDA, obteniéndose los siguientes vectores de proyección ordenados por valor propio generalizado de mayor (w_1) a menor (w_2):

	$\mathrm{W_{LDA}}$		
w_1	1	0	0
w_2	0	0	1

Se pide:

a) Calcula los vectores de proyección PCA del conjunto de muestras (**1 punto**).

b) Calcula la proyección de las muestras mediante PCA a $\mathbb{R}^2$ (**0.4 puntos**).

c) Calcula la proyección de las muestras mediante LDA a $\mathbb{R}$ (**0.4 puntos**).

d) ¿Qué proyección (PCA o LDA) consideras más adecuada para minimizar el error de clasificación? Justifica la respuesta (**0.2 puntos**).

Solución:

a) Para calcular los vectores de proyección PCA primero es necesario obtener la matriz de covarianzas de los datos. En este caso, como $\bar{\mathbf{x}} = (0\ 0\ 0)^t$, la matriz de covarianzas es:

$$\Sigma = \frac{1}{8}\left(\mathbf{x}_1 \cdot \mathbf{x}_1^t + \mathbf{x}_2 \cdot \mathbf{x}_2^t + \mathbf{x}_3 \cdot \mathbf{x}_3^t + \mathbf{x}_4 \cdot \mathbf{x}_4^t + \mathbf{x}_5 \cdot \mathbf{x}_5^t + \mathbf{x}_6 \cdot \mathbf{x}_6^t + \mathbf{x}_7 \cdot \mathbf{x}_7^t + \mathbf{x}_8 \cdot \mathbf{x}_8^t\right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6.5 & 2.5 \\ 0 & 2.5 & 6.5 \end{pmatrix}$$

Calculamos los valores propios de la matriz de covarianzas

$$\left|\begin{pmatrix} 1-\lambda & 0 & 0 \\ 0 & 6.5-\lambda & 2.5 \\ 0 & 2.5 & 6.5-\lambda \end{pmatrix}\right| = 0 \qquad \text{donde} \qquad \lambda_1 = 9, \quad \lambda_2 = 4 \quad \text{y} \quad \lambda_3 = 1.$$

[V, D] = eig(ex)

Los vectores propios asociados son

$$\lambda_1 = 9 \quad \rightarrow \quad w_1 = \begin{pmatrix} 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}^t$$

$$\lambda_2 = 4 \quad \rightarrow \quad w_2 = \begin{pmatrix} 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}^t$$

$$\lambda_3 = 1 \quad \rightarrow \quad w_3 = (1 \quad 0 \quad 0)^t.$$

b) Proyectamos sobre los dos vectores propios de mayor valor propio asociado

x' = Wt * x

Ordenar W de mayor a menor

n	1	2	3	4	5	6	7	8
x_1	$3\sqrt{2}$	0	0	$-3\sqrt{2}$	$3\sqrt{2}$	0	0	$-3\sqrt{2}$
x_2	0	$-2\sqrt{2}$	$2\sqrt{2}$	0	0	$2\sqrt{2}$	$-2\sqrt{2}$	0
c	A	A	A	A	B	B	B	B

c) Proyectamos sobre el vector LDA

n	1	2	3	4	5	6	7	8
x_1	1	1	1	1	-1	-1	-1	-1
c	A	A	A	A	B	B	B	B

d) A diferencia de la proyección PCA que asigna al mismo punto muestras de diferentes clases, la proyección LDA separa perfectamente las muestras de diferentes clases, y por tanto es más adecuada.