Examen de Teoría de Percepción - Primer Parcial

ETSINF, Universitat Politècnica de València, Abril de 2017

Apellidos: ☐ **Nombre:** ☐

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

Cuestiones (2 puntos, 30 minutos, sin apuntes)

C Dado un clasificador definido por $c(x) = \arg\max_{c=1,\dots,C} g_c(x)$. Indica cuál de las siguientes definiciones de $g_c(x)$ hace que se corresponda a un clasificador de mínimo error:

A) $g_c(x) = P(c)\,P(c|x)$
B) $g_c(x) = w_c x + w_{c0}$, con w_c y w_{c0} aprendidos por el algoritmo Perceptron
C) $g_c(x) = P(c|x)\,p(x)$
D) $g_c(x) = P(x|c)$

B Indica la afirmación correcta para un sistema de reconocimiento interactivo:

A) Se evalúan según el error final cometido
B) Emplean la realimentación del usuario
C) Siempre requieren una evaluación manual
D) Usan entrenamiento *on-line* en todas las interacciones del usuario

C Se está resolviendo un problema de clasificación de imágenes de tamaño 100 × 100 píxeles. Se propone implementar sistemas usando características globales o características locales con ventanas de 5 × 5 píxeles. Indicar qué afirmación de las siguientes se dará siempre en estos sistemas:

A) La clasificación por características locales será mejor que la clasificación por representación global
B) El vector de características de la representación global ocupará más memoria que el conjunto de características locales
C) La dimensión del vector de representación global será 10000 y la de representación local 25
D) El número de características locales extraídas será superior a 9000

D Disponemos de una señal acústica muestreada a 8kHz y 16 bits. ¿Cuánto espacio en memoria requiere la representación en *frames* (marcos) de 2 segundos de señal con una tamaño de ventana (W) de 50ms y un desplazamiento (S) de 5ms? Asume que cada componente de un *frame* se representa mediante un número en coma flotante (4 bytes).

A) Menos de 150 Kbytes
B) Entre 150 Kbytes y 300 Kbytes
C) Entre 300 Kbytes y 600 Kbytes
D) Más de 600 Kbytes.

$\boxed{\text{A}}$ ¿Cuál de las siguientes afirmaciones **no** es cierta sobre la representación de n-gramas?

A) Dado su grado de dispersión, no suele ser utilizada en la práctica
B) Captura el contexto en el que ocurre un token
C) Se suelen utilizar representaciones en memoria dispersas para compensar su dispersión
D) Su requerimiento de espacio en memoria es exponencial con el orden del n-grama

$\boxed{\text{A}}$ Sea un problema de clasificación en dos clases sobre vectores de $\mathbb{R}^3$, con las muestras $\{(0,0,1),\ (1,0,1),\ (0,1,1)\}$ para la clase 1 y las muestras {(1,1,1),(2,0,1),(0,1,2)} para la clase 2. Indicar qué matriz de proyección para $\mathbb{R}^2$ es la más apropiada para estas muestras.

A) $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$
B) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
C) $\begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 0 \end{pmatrix}$
D) $\begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$

$\boxed{\text{D}}$ Se quiere combinar PCA y LDA para la reducción de dimensión de una representación vectorial de $\mathbb{R}^D$ a $\mathbb{R}^k$; suponiendo que PCA reduce inicialmente a $\mathbb{R}^{k'}$, ¿qué condición de las siguientes es cierta con respecto a las tres dimensiones suponiendo que tenemos C clases?

A) $D \geq k'$ y $k' \geq k$ y $k' > C$
B) $D > k + k'$ y $k' \geq k$ y $C \geq k$
C) $D \geq k$ y $D \geq k'$ y $D \geq C$
D) $D \geq k'$ y $k' \geq k$ y $C > k$

$\boxed{\text{B}}$ ¿Cuál de las siguientes afirmaciones sobre PCA y LDA es cierta?

A) Tanto PCA como LDA minimizan el error de clasificación
B) Ni PCA, ni LDA minimizan el error de clasificación
C) PCA no minimiza el error de clasificación, pero LDA sí al ser una técnica supervisada
D) LDA no minimiza el error de clasificación, pero PCA sí

Examen de Teoría de Percepción - Primer Parcial

ETSINF, Universitat Politècnica de València, Abril de 2017

Apellidos: **Nombre:**

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

Problemas (4 puntos, 90 minutos, con apuntes)

1. Calcula espacio requerido en memoria de las siguientes representaciones de imágenes:

   a) Representación directa global de una imagen de 1024×1080 píxeles en escala de grises a 65536 niveles (**0.25 puntos**). $1024 \cdot 1080 \cdot 2 = 2211840$ bytes

   b) Representación mediante histograma de una imagen RGB de 512×512 píxeles en escala de 256 niveles por color (**0.25 puntos**). 3 bytes por nivel $\cdot$ 256 niveles $\cdot$ 3 colores= 2304 bytes

   c) Representación de una imagen de 100×100 píxeles en escala de grises de 256 niveles mediante características locales de 25×25 píxeles extraídas una cada 5 píxeles (**0.25 puntos**). $((100-12-12)\cdot(100-12-12))/5 =$ 1155 características locales $\cdot 25 \cdot 25 \cdot 1 = 721875$ bytes

   d) Representación de una imagen de 80×80 píxeles en escala de grises de 512 niveles mediante características locales de 21×21 píxeles extraídas una cada 10 píxeles y representada cada una de ellas mediante histograma (**0.25 puntos**). $((80-10-10)\cdot(80-10-10))/10 \cdot 2 \cdot 512 = 368640$ bytes

2. Dada la siguiente matriz de *term frecuency* que representa la frecuencia de aparición de cada token en cada documento

x_{dt}	token 1	token 2	token 3	token 4	token 5	token 6
doc 1	10	1	10	1	0	0
doc 2	10	1	0	0	0	0
doc 3	10	1	10	1	10	1
doc 4	10	1	0	0	0	0
doc 5	10	1	10	1	0	0

   se pide:

   a) Calcular la función global *Normal* para todos los tokens (**0.20 puntos**).

   b) Calcular la función global *GfIdf* para todos los tokens (**0.20 puntos**).

   c) Calcular la función global *Idf* para todos los tokens utilizando logaritmo neperiano (**0.20 puntos**).

   d) En este caso, ¿cuál es la principal diferencia que observas entre la función global Normal y GfIdf?, ¿y entre GfIdf e Idf?(**0.40 puntos**).

   Solución:

$G(t)$	token 1	token 2	token 3	token 4	token 5	token 6
a) Normal	0.04	0.4	0.06	0.6	0.1	1
b) GfIdf	10	1	10	1	10	1
c) Idf (ln)	0	0	0.51	0.51	1.63	1.63

   d) La principal diferencia entre la función global Normal y GfIdf es que los tokens con frecuencias altas tienen un menor peso en la función Normal que en la GfIdf. La principal diferencia entre GfIdf e Idf es que la función Idf asigna el mismo peso a tokens con diferentes frecuencias (ya que solo considera la ocurrencia o no del token en un documento), mientras que GfIdf tiene en cuenta la frecuencia del token y le asigna un mayor peso.

3. Se tiene el siguiente conjunto de datos en $\mathbb{R}^5$ para tres clases distintas:

	Muestra					Clase
x_1	1	-1	-5	1	3	1
x_2	-1	0	3	0	-2	1
x_3	3	1	5	1	3	2
x_4	2	2	2	-1	-2	2
x_5	-4	-2	-3	1	-3	3
x_6	-1	0	4	-2	4	3

Se ha realizado el proceso para obtener proyecciones con PCA, dando el siguiente conjunto de vectores y valores propios (redondeados a dos decimales):

	$\mathrm{W}_{\mathrm{PCA}}$					λ
w_1	0.35	0.33	-0.78	0.01	-0.40	102.77
w_2	0.25	-0.07	-0.35	-0.36	0.83	49.92
w_3	0.83	-0.43	0.24	0.25	-0.08	22.76
w_4	-0.13	0.08	-0.20	0.90	0.35	6.02
w_5	0.32	0.83	0.42	0.01	0.15	0.00

De la misma forma, se ha obtenido la matriz de transformación para LDA para el máximo número de dimensiones posible:

	$\mathrm{W}_{\mathrm{LDA}}$				
w_1	0.85	0.20	0.04	0.37	-0.31
w_2	-0.38	0.82	-0.11	0.39	0.15

Se pide:

a) Si se quiere que el error de reconstrucción con PCA sea menor que 5, ¿a qué dimensión mínima habría que proyectar? ¿Y si sólo se requiriera un error de reconstrucción de 10? (**0.25 puntos**).

b) Realizar la proyección de las muestras $x_1, \ldots, x_6$ mediante PCA a $\mathbb{R}^2$ (**0.75 puntos**).

c) Realizar la proyección de las muestras $x_1, \ldots, x_6$ mediante LDA a $\mathbb{R}^2$ (**0.5 puntos**).

d) Representar gráficamente ambas proyecciones e indicar cuál sería la más apropiada respecto a la separabilidad de las clases (**0.25 puntos**).

e) ¿Sería razonable aplicar proyección LDA a una única dimensión? Razona la respuesta. (**0.25 puntos**)

Solución:

a) Para un error de reconstrucción menor que 5 sólo podría dejarse fuera el último vector propio (con valor propio de 0, que sería el error de reconstrucción), con lo que se debería proyectar a $k = 4$; para error menor que 10, se podría dejar también fuera el penúltimo vector propio (con valor propio 6.02, así que el error de reconstrucción total sería de $0 + 6.02 = 6.02$), con lo que se podría proyectar a $k = 3$.

b) En primer lugar hay que calcular la media y sustraerla a las muestras:

m	0	0	1	0	0.5	
	Muestra					Clase
$x_1 - m$	1	-1	-6	1	2.5	1
$x_2 - m$	-1	0	2	0	-2.5	1
$x_3 - m$	3	1	4	1	2.5	2
$x_4 - m$	2	2	1	-1	-2.5	2
$x_5 - m$	-4	-2	-4	1	-3.5	3
$x_6 - m$	-1	0	3	-2	3.5	3

El resultado de la proyección sería:

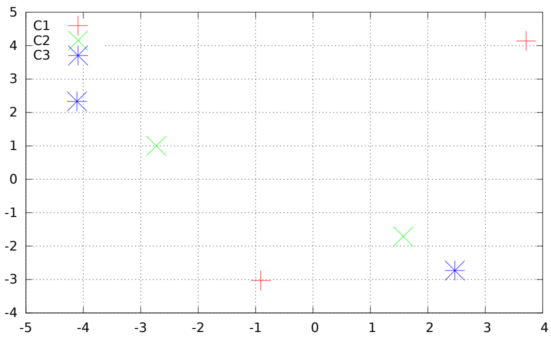
x'_1	3.71	4.14
x'_2	-0.91	-3.03
x'_3	-2.73	1.00
x'_4	1.57	-1.71
x'_5	2.47	-2.73
x'_6	-4.11	2.33

c) El resultado de la proyección sería:

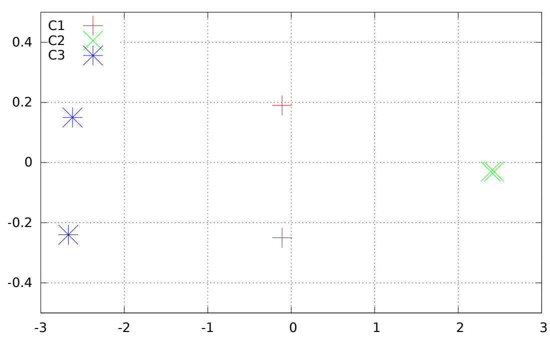
x'_1	-0.11	0.19
x'_2	-0.11	-0.25
x'_3	2.39	-0.03
x'_4	2.43	-0.03
x'_5	-2.62	0.15
x'_6	-2.67	-0.24

d) La representación gráfica sería:

PCA



LDA



Por tanto, es claro que LDA proporciona una mejor separabilidad de las muestras de cada clase.

e) Sí, es razonable, ya que en ese caso se tomaría sólo la primera componente del resultado de la proyección LDA, en la cual sigue habiendo una clara separabilidad entre las muestras de las distintas clases.