Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politècnica de València, Junio de 2018

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

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

Cuestiones (2 puntos, 30 minutos, sin apuntes)

B 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 **no** se corresponda a un clasificador de mínimo error:

A) $g_c(x) = P(c|x)^2$
B) $g_c(x) = P(x|c)$
C) $g_c(x) = P(x,c)$ = P(x) * P(c | x)
D) $g_c(x) = P(c|x)$

A En *On-line Learning* se parte de un conjunto de muestras de entrenamiento iniciales T con el cual se entrena un modelo M y mediante la realimentación del usuario se genera un nuevo conjunto de entrenamiento T' que da lugar a un nuevo modelo M' que combina T y T'. Considera la tarea de clasificación de correos electrónicos en *spam* y *ham* donde T es el conjunto de correos TREC06 y T', un conjunto de correos propios que tú mismo has etiquetado, ¿qué combinación de T y T' pondera igualitariamente todas las muestras para estimar las probabilidades a priori de cada clase en el nuevo modelo M'?

A) Sumar el número de correos *spam* y *ham* en T y T', y normalizar.
B) Sumar las probabilidades a priori calculadas a partir de T y las calculadas a partir de T', y no normalizar
C) Sumar las probabilidades a priori calculadas a partir de T y las calculadas a partir de T', y normalizar
D) Sumar el número de correos *spam* y *ham* en T y T', y no normalizar.

C Se tiene un problema de clasificación de imágenes en niveles de gris donde se quiere emplear información global empleando la menor memoria posible. Teniendo en cuenta que se han definido 1024 niveles de gris, ¿qué tamaño mínimo de imagen de los presentados hace que la representación por histograma ocupe menos que la representación directa?

Histograma **Directa** **Histograma** **Directa**

L * ((log2(n+1)) / 8) < n * up_rounding((log2(L)) / 8) -> 1024 * up_rounding((log2(n + 1)) / 8) < 2n

A) 500 píxeles
B) 1000 píxeles
C) 1500 píxeles
D) 2000 píxeles

D Se quiere muestrear una señal de ancho de banda 3500 Hz y se sabe que puede existir ruido de alta frecuencia en el proceso de adquisición. ¿Qué proceso debe seguirse para garantizar una adquisición fiel de la señal?

A) Basta con muestrear a >7000 Hz
B) Aplicar un filtro para que pasen frecuencias ≤3500 Hz y muestrear a >3500 Hz
C) Aplicar un filtro para que pasen frecuencias en el rango $3500 \pm f_M$ Hz, donde f_M es la frecuencia del ruido
D) Aplicar un filtro para que pasen frecuencias ≤3500 Hz y muestrear a >7000 Hz

C Dado un token que aparece con una frecuencia constante $k > 0$ en todos los documentos de una colección, ¿qué función global le asignaría el menor valor?

A) La función Normal $G(t) = \left(\sum_d x_{dt}^2\right)^{-\frac{1}{2}}$

B) La función GfIdf $G(t) = \frac{\sum_d x_{dt}}{\sum_{d:x_{dt}>0} 1}$

C) La función Idf $G(t) = \log \frac{D}{\sum_{d:x_{dt}>0} 1}$ ya que, para este caso, sería: log(D/D) = 0.

D) Todas le asignan el mismo valor

$\boxed{\text{B}}$ ¿Cuántos valores propios distintos de cero se obtienen de la matriz $A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ -1 & 0 & -1 & 0 \end{pmatrix}$?

A) 1
B) 2
C) 3
D) 4

$\boxed{\text{D}}$ Dada la matriz de covarianzas de los datos originales $\Sigma_{\mathcal{X}} \in \mathbb{R}^{D\times D}$, la matriz de proyección PCA $W \in \mathbb{R}^{D\times k}$ donde $\mathbf{w}_j$ es el j-ésimo vector de proyección (de mayor a menor valor propio asociado) y la matriz de covarianzas diagonalizada de los datos originales $\Delta \in \mathbb{R}^{D\times D}$, ¿cuál de las siguientes expresiones caracteriza el error de reconstrucción de los datos de $\mathcal{X}$ al proyectarlos de $\mathbb{R}^D$ a $\mathbb{R}^k$ con W?

A) $\sum_{j=1}^{k} \mathbf{w}_j^t \, \Sigma_{\mathcal{X}} \mathbf{w}_j - \sum_{j=1}^{k} \Lambda_{jj}$

B) $\sum_{j=1}^{k} \mathbf{w}_j^t \, \Lambda_{jj} \mathbf{w}_j - \sum_{j=1}^{D} \Lambda_{jj}$

C) $\sum_{j=1}^{D} \mathbf{w}_j^t \, \Lambda_{jj} \mathbf{w}_j - \sum_{j=1}^{D} \Lambda_{jj}$

D) $\sum_{j=1}^{D} \mathbf{w}_j^t \, \Sigma_{\mathcal{X}} \mathbf{w}_j - \sum_{j=1}^{k} \Lambda_{jj}$

$\boxed{\text{A}}$ Indica la característica de LDA que la distingue de PCA

A) Es una técnica de reducción de dimensionalidad supervisada.
B) Su resolución se basa en un problema de optimización.
C) Emplea las covarianzas de los datos.
D) Requiere el cálculo de vectores propios.

Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politècnica de València, Junio de 2018

Apellidos: ____________ **Nombre:** ____________

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

Problemas (4 puntos, 90 minutos, con apuntes)

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

   a) Representación global de una imagen en color RGB, con 256 niveles por cada color, de 45 × 45 píxeles, con representación directa. (**0.2 puntos**) Tamaño = 45 * 45 = 45^2 = 2025 -> R, G y B -> 2025 * 3 = 6075 B.

   b) Representación por características locales de una imagen de niveles de gris, con 256 niveles, de tamaño 45 × 45 con ventanas de 15 × 21 y desplazamiento unitario en ambas coordenadas, con representación por histograma. (**0.3 puntos**) n = (((Vy - C + 1) / Dv) * ((Vx - L + 1) / Dh)) = (45 - 15 + 1) * (45 - 21 + 1) = 775 B.

   c) Representación de una señal de audio de 5 minutos de alta fidelidad (muestreada a 44100Hz, muestras de 16 bits) en un sistema 2.1. (**0.2 puntos**) TAMAÑO (B) = 5 * 60 * 44100 * 2 * 3 = 79380000 B = 79.38 MB.

   d) Representación de una señal de audio monocanal con ancho de banda 5KHz, muestreada con la frecuencia mínima para mantener todas sus frecuencias representativas, de duración 1 minuto y representada por 1 byte por muestra. (**0.3 puntos**) TAMAÑO (B) = 60 * 5000 * 2 (ya que nos dan el ancho de banda) * 1 * 1 = 0.6 MB.

   Solución:

   a) 6075 bytes

   b) 396800 bytes S = 256 * 2 = 512 B. -> Tamaño final = n * s = 396800 B.

   c) 79380000 bytes TAMAÑO (B) = DURACIÓN (s) * FRECUENCIA (Hz) * TAMAÑO MUESTRA (B) * nº de canales

   d) 600000 bytes

2. (**1 punto**) Sea el siguiente conjunto de documentos de texto:

Doc	Texto
1	Pedro Sánchez se compromete a gobernar para los que no le votaron y para los que tampoco le votarán
2	Mariano Rajoy ya ha cambiado la dirección de la suscripción al Marca
3	Pedro Sánchez ha envejecido diez años desde que es presidente
4	Aparece a lápiz "P. Sánchez" en los papeles de Bárcenas
5	Las 25 frases más míticas de Mariano Rajoy durante su presidencia del Gobierno
6	Pedro Sánchez preguntará a las bases si debe dimitir por haberse ido a vivir a La Moncloa
7	Pedro Sánchez prometió su cargo sin Biblia y Dios sigue pensando que el presidente de España es Rajoy
8	Albert Rivera solo ve "españoles traidores" desde que hicieron a Pedro Sánchez presidente

   Se pide:

   a) Calcular la representación *bag-of-words* para las palabras correspondientes a nombres propios: Pedro, Sánchez, Mariano, Rajoy, Marca, Bárcenas, Gobierno, Moncloa, Biblia, Dios, España, Albert, Rivera. (**0.3 puntos**)

   b) Calcular las funciones globales normal, GfIdf e Idf para los términos: para, los, la, que, a, en, de, más (**0.5 puntos**)

   c) A la vista de los resultados del apartado previo, ¿qué característica de Idf se confirma? (**0.2 puntos**)

   Solución:

   a)

t/d	1	2	3	4	5	6	7	8
Pedro	1	0	1	0	0	1	1	1
Sánchez	1	0	1	1	0	1	1	1
Mariano	0	1	0	0	1	0	0	0
Rajoy	0	1	0	0	1	0	1	0
Marca	0	1	0	0	0	0	0	0
Bárcenas	0	0	0	1	0	0	0	0
Gobierno	0	0	0	0	1	0	0	0
Moncloa	0	0	0	0	0	1	0	0
Biblia	0	0	0	0	0	0	1	0
Dios	0	0	0	0	0	0	1	0
España	0	0	0	0	0	0	1	0
Albert	0	0	0	0	0	0	0	1
Rivera	0	0	0	0	0	0	0	1

b)

Término	Normal	GfIdf	Idf
para	$\frac{1}{2}$	2	$\log 8$
los	$\frac{1}{\sqrt{5}}$	$\frac{3}{2}$	$\log 4$
la	$\frac{1}{2}$	2	$\log 8$
que	$\frac{1}{\sqrt{7}}$	$\frac{5}{4}$	$\log 2$
a	$\frac{1}{\sqrt{12}}$	$\frac{3}{2}$	$\log 2$
en	1	1	$\log 8$
de	$\frac{1}{2}$	1	$\log 2$
más	1	1	$\log 8$

c) Se confirma que Idf atenúa más los tokens que aparecen en más documentos; así, los tokens "que", "a" y "de", que aparecen en 4 documentos de los 8, presentan el menor valor de la colección.

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

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

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_3):

	$\mathrm{W}_{\mathrm{LDA}}$		
w_1	0	0	1
w_2	1	0	0
w_3	0	1	0

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 $\mathbb{R}^2$ (**0.4 puntos**).
d) ¿Qué proyección (PCA o LDA) consideras más adecuada para minimizar el error de clasificación? (**0.2 puntos**)

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:

$$\Sigma = \frac{1}{8}\begin{pmatrix} 4 & 4 & -2 & 2 & -2 & 2 & -4 & -4 \\ 4 & 4 & 2 & -2 & 2 & -2 & -4 & -4 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 & 4 & 1 \\ 4 & 4 & -1 \\ -2 & 2 & -1 \\ 2 & -2 & 1 \\ -2 & 2 & 1 \\ 2 & -2 & -1 \\ -4 & -4 & -1 \\ -4 & -4 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 6 & 0 \\ 6 & 10 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Calculamos los valores propios de la matriz de covarianzas

$$\left| \begin{pmatrix} 10-\lambda & 6 & 0 \\ 6 & 10-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{pmatrix} \right| = 0 \qquad \text{donde} \qquad \lambda_1 = 16, \quad \lambda_2 = 4 \quad \text{y} \quad \lambda_3 = 1.$$

Los vectores propios asociados son

$$\lambda_1 = 16 \quad \rightarrow \quad w_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \end{pmatrix}^t$$

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

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

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

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

c) Proyectamos sobre los dos vectores LDA

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

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