Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politècnica de València, Junio de 2017

Apellidos: ______ **Nombre:** ______

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

Cuestiones (2 puntos, 30 minutos, sin apuntes)

D Dado un clasificador definido por $c(x) = \arg\max_{c=1,\ldots,C} g_c(x)$. Indica cuál de las siguientes definiciones de $g_c(x)$ hace que **no** se corresponda a un clasificador de mínimo error:

A) $g_c(x) = e^{P(x,c)}$
B) $g_c(x) = P(x,c)^3$
C) $g_c(x) = \sqrt{P(x,c)}$
D) $g_c(x) = \cos P(x,c)$

C En Active Learning se dispone de dos conjuntos de muestras, uno de ellos etiquetado T y otro no etiquetado U. ¿Cuál de las siguientes afirmaciones **no** es correcta?

A) El conjunto de muestras no etiquetado U es mayor que el etiquetado T.
B) Se selecciona un subconjunto de U para ser etiquetado.
C) La composición del subconjunto de U seleccionado para ser etiquetado es indiferente.
D) El error del modelo entrenado con T y el subconjunto de U etiquetado debe ser menor que solamente con T.

B Se está resolviendo un problema de clasificación de imágenes de tamaño 50×50 píxeles. Se propone implementar sistemas usando características globales o características locales sobre ventanas de 10×10 píxeles. Indicar qué afirmación de las siguientes se dará siempre en los sistemas:

A) El número de vectores de características con la representación global será igual al número de puntos de interés determinado en la imagen
B) Las dimensiones de un vector de representación global será 2500 y las de representación local 100, suponiendo que el valor de cada píxel se toma como una característica
C) La clasificación por características locales será peor que la clasificación por representación global
D) Un vector de características de la representación global ocupará menos memoria que el conjunto de los de la representación local

C Disponemos de una señal acústica muestreada a 44kHz y 16 bits. ¿Cuánto espacio en memoria requiere la representación en frames (marcos) de 1 segundo de señal con un tamaño de ventana (W) de 100ms y un desplazamiento (S) de 50ms? 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 600 Kbytes.

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

A) Es ampliamente utilizada para el modelado de lenguaje.
B) Principalmente captura dependencias entre tokens consecutivos.
C) La representación bag-of-words se puede entender como una represetación mediante unigramas.
D) Su requerimiento de espacio en memoria es lineal 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) $\begin{pmatrix} 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}$
B) $\begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 0 \end{pmatrix}$
C) $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$
D) $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$

$\boxed{\text{B}}$ Dada la matriz de covarianzas de los datos originales $\Sigma_x \in \mathbb{R}^{D\times D}$, la matriz de proyección PCA $W \in \mathbb{R}^{k\times D}$ 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 la varianza residual (eliminada) al proyectar un conjunto de datos en $\mathbb{R}^D$ a $\mathbb{R}^k$?

A) $\sum_{j=1}^{D} \mathbf{w}_j\, \Sigma_x\, \mathbf{w}_j^t - \sum_{j=1}^{D} \Delta_{jj}$
B) $\sum_{j=1}^{D} \Delta_{jj} - \sum_{j=1}^{k} \mathbf{w}_j\, \Sigma_x\, \mathbf{w}_j^t$
C) $\sum_{j=1}^{k} \Delta_{jj} - \sum_{i=1}^{n} (x_i - \hat{x}_i)^2$ con $x_i \in \mathbb{R}^D$ y $\hat{x}_i \in \mathbb{R}^k$
D) $\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 - \sum_{j=1}^{k} \mathbf{w}_j\, \Sigma_x\, \mathbf{w}_j^t$ con $x_i \in \mathbb{R}^D$ y $\hat{x}_i \in \mathbb{R}^k$

$\boxed{\text{A}}$ Cuando aplicamos LDA, no tiene sentido proyectar a más de $C-1$ de dimensiones, ¿a qué se debe esta limitación?

A) Al rango de la matriz S_b
B) Al rango de la matriz S_w
C) Al rango de la matriz de covarianzas por clase Σ_c
D) Al rango de la matriz de covarianzas de los datos Σ_x

Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politècnica de València, Junio de 2017

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

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

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(1 punto)** Calcula el espacio requerido en memoria para las siguientes representaciones de imágenes y audio:

   a) Imagen de 1024×768 píxeles en color RGB, 256 niveles por color, por representación directa global. (**0.15 puntos**) $1024 \cdot 768 \cdot 3 = 2359296$ bytes

   b) Imagen de 1024×768 píxeles en escala de grises, 256 niveles, por representación por histograma. (**0.15 puntos**) $1024 \cdot 768$ píxeles necesitan 3 bytes, por tanto $256 \cdot 3 = 768$ bytes

   c) Imagen de 1024×768 píxeles en escala de grises, 256 niveles, por representación local, con ventanas de 33×33 píxeles extraídas cada 16 píxeles tanto en horizontal como en vertical. (**0.3 puntos**) Tamaño por ventana: $33 \cdot 33 \cdot 1 = 1089$ bytes; total ventanas posibles si fuera a cada píxel: $(1024 - 33 + 1)/16 = 62$ en horizontal, $(768 - 33 + 1)/16 = 46$ en vertical, total $62 \cdot 46 = 2852$ ventanas; en total, $2852 \cdot 1089 = 3105828$ bytes

   d) Señal de audio mono de 10 segundos de duración, con frecuencia máxima a reproducir de 8 KHz, 2 bytes por muestra. (**0.2 puntos**) $10 \cdot 16000 \cdot 2 = 320000$ bytes

   e) Señal de audio 5.1 de 1 minuto de duración, con frecuencia de muestreo de 44 KHz, 32 bits por muestra. (**0.2 puntos**) $60 \cdot 44000 \cdot 4 \cdot 6 = 63360000$ bytes

2. **(1 punto)** Tenemos los siguientes textos para 3 clases de documentos:

Número	Clase	Texto
1	Internacional	May gana sin mayoría absoluta y Corbyn pide su dimisión
2	Economía	El FROB vigiló al Popular durante cuatro días antes de la intervención
3	Política	El Constitucional falla que la amnistía de Montoro avaló el fraude fiscal
4	Internacional	Theresa May fracasa en su apuesta y pierde la mayoría absoluta
5	Economía	Los ganadores y perdedores en la quiebra de Popular
6	Política	El varapalo del Constitucional al Gobierno de Rajoy cuestiona futuras amnistías fiscales
7	Internacional	May pierde la mayoría absoluta y Corbyn pide su dimisión
8	Economía	El BCE rebajó las exigencias al Banco Popular hace sólo seis meses
9	Política	¿Qué consecuencias tiene el fallo que declara nula la amnistía de Montoro?

Se pide (en todos los casos, considera indistinto mayúsculas y minúsculas):

   a) Realizar la representación *bag-of-words* con las siguientes palabras: May, mayoría, absoluta, dimisión, Popular, intervención, quiebra, Constitucional, amnistía, Rajoy, Montoro, fraude. (**0.25 puntos**)

   b) Realizar la representación *bag-of-bigrams* con las siguientes secuencias de palabras: "Theresa May", "mayoría absoluta", "su dimisión", "Banco Popular", "la quiebra", "la intervención", "fraude fiscal", "amnistías fiscales", "el fallo". (**0.25 puntos**)

   c) Calcular el valor de las funciones globales GfIdf y Idf en el conjunto de documentos para los términos: y, mayoría, dimisión, de, la, intervención, el, fraude, en, quiebra. (**0.5 puntos**)

Solución:

a)

t/d	1	2	3	4	5	6	7	8	9
May	1	0	0	1	0	0	1	0	0
mayoría	1	0	0	1	0	0	1	0	0
absoluta	1	0	0	1	0	0	1	0	0
dimisión	1	0	0	0	0	0	1	0	0
Popular	0	1	0	0	1	0	0	1	0
intervención	0	1	0	0	0	0	0	0	0
quiebra	0	0	0	0	1	0	0	0	0
Constitucional	0	0	1	0	0	1	0	0	0
amnistía	0	0	1	0	0	0	0	0	1
Rajoy	0	0	0	0	0	1	0	0	0
Montoro	0	0	1	0	0	0	0	0	1
fraude	0	0	1	0	0	0	0	0	0

b)

t/d	1	2	3	4	5	6	7	8	9
“Theresa May”	0	0	0	1	0	0	0	0	0
“mayoría absoluta”	1	0	0	1	0	0	1	0	0
“su dimisión”	1	0	0	0	0	0	1	0	0
“Banco Popular”	0	0	0	0	0	0	0	1	0
“la quiebra”	0	0	0	0	1	0	0	0	0
“la intervención”	0	1	0	0	0	0	0	0	0
“fraude fiscal”	0	0	1	0	0	0	0	0	0
“amnistías fiscales”	0	0	0	0	0	1	0	0	0
“el fallo”	0	0	0	0	0	0	0	0	1

c)

Término	GfIdf	Idf
y	1	$\log\frac{9}{4}$
mayoría	1	$\log 3$
dimisión	1	$\log\frac{9}{2}$
de	1	$\log\frac{9}{5}$
la	1	$\log\frac{3}{2}$
intervención	1	$\log 9$
el	$\frac{6}{5}$	$\log\frac{9}{5}$
fraude	1	$\log 9$
en	1	$\log\frac{9}{2}$
quiebra	1	$\log 9$

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

n	1	2	3	4
x_1	2	-1	1	-2
x_2	2	1	-1	-2
c	0	1	1	0

Por otra parte se ha calculado LDA, obteniéndose el siguiente vector de proyección:

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

Se pide:

a) Calcula los vectores de proyección PCA del conjunto de muestras y represéntalos gráficamente (**0.75 puntos**).

b) Representa gráficamente la proyección de las muestras mediante PCA a $\mathbb{R}$ (**0.5 puntos**).

c) Representa gráficamente la proyección de las muestras mediante LDA a $\mathbb{R}$ (**0.5 puntos**).

d) ¿Qué proyección (PCA o LDA) consideras más adecuada para minimizar el error de clasificación? (**0.25 puntos**)

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

$$\Sigma = \frac{1}{4}\begin{pmatrix} 2 & -1 & 1 & -2 \\ 2 & 1 & -1 & -2 \end{pmatrix} \cdot \begin{pmatrix} 2 & 2 \\ -1 & 1 \\ 1 & -1 \\ -2 & -2 \end{pmatrix} = \begin{pmatrix} \frac{5}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{5}{2} \end{pmatrix}$$

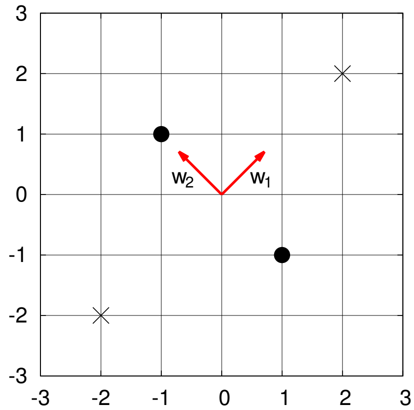
Calculamos los valores propios de la matriz de covarianzas

$$\left| \begin{pmatrix} \frac{5}{2} - \lambda & \frac{3}{2} \\ \frac{3}{2} & \frac{5}{2} - \lambda \end{pmatrix} \right| = 0 \qquad \text{donde} \qquad \lambda_1 = 4 \quad \text{y} \quad \lambda_2 = 1.$$

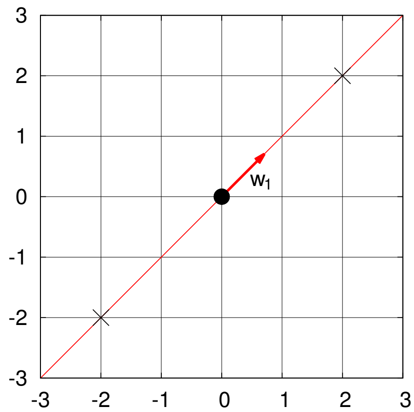
Los vectores propios asociados son

$$\lambda_1 = 4 \quad \rightarrow \quad w_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}^t$$

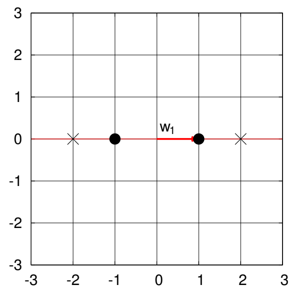
$$\lambda_2 = 1 \quad \rightarrow \quad w_2 = \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}^t.$$



b) Proyectamos sobre el vector propio de mayor valor propio asociado



c) Proyectamos sobre el único vector LDA



d) En este caso no hay una proyección que se pueda considerar más adecuada que la otra, porque en ambos casos es necesaria una frontera de decisión cuadrática para separar ambas clases.