Examen de Teoría de Percepción

ETSINF, Universitat Politècnica de València, Abril de 2014

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

Profesor: ☐ **Jorge Civera** ☐ **Roberto Paredes**

Cuestiones (3 puntos, 30 minutos, sin apuntes)

C ¿Cuándo podremos afirmar que la regla de clasificación estadística de mínimo error es $c(\mathbf{x}) = \arg\max_c p(\mathbf{x} \mid c)$?

A) Siempre que $P(c) = 1 \quad \forall c \in \{1 \cdots C\}$
B) Siempre que $P(c) = 0 \quad \forall c \in \{1 \cdots C\}$
C) Siempre que $P(c) = \frac{1}{C} \quad \forall c \in \{1 \cdots C\}$
D) Siempre que $P(c) \geq 0 \quad \forall c \in \{1 \cdots C\}$

C Dado un conjunto de muestras etiquetadas $X = \{(\mathbf{x}_1, c_1), \cdots, (\mathbf{x}_n, c_n)\}$ donde $\mathbf{x}_i \in \mathbb{R}^3$, podremos decir que la frontera de decisión entre dos clases es:

A) Un punto
B) Una recta
C) Un plano
D) Ninguna de las anteriores

A Dada una representación local con un tamaño de ventana de 11x11 de un conjunto de 100 imágenes de 33x33 píxeles en 65536 niveles de gris, ¿cuál sería el espacio máximo que ocuparía este conjunto con esta representación local?

A) Aproximadamente 12 Mbytes
B) Aproximadamente 6 Mbytes
C) Aproximadamente 125 Kbytes
D) Ninguna de las anteriores

Tamaño = n * s = 529 * 24200 B = 12.8 MB.
n = (33 - 11 + 1) * (33 - 11 + 1) = 529.
s = nd * up_rounding(log2(niveles de gris) / 8) = 100 * 11 * 11 * 2 = 24200.

B ¿Cuánto espacio como mínimo requiere el almacenamiento de una señal de 1 minuto con una frecuencia máxima de 15 KHz y correctamente adquirida mediante un sistema de sonido estéreo con muestras de 16 bits?

A) Aproximadamente 1.7 Mbytes
B) Aproximadamente 6.9 Mbytes
C) Aproximadamente 3.4 Mbytes
D) Ninguna de las anteriores

TAMAÑO (B) = DURACIÓN (s) * FRECUENCIA (Hz) * TAMAÑO MUESTRA (B) * nº de canales
Tamaño = 60 * 15000 * 2 * 2 * **2 (ya que es una banda, y no una señal muestreada)** = 7200000 B = 7.2 MB.

B ¿Cuál es el espacio requerido para almacenar una colección de 100 documentos representados (de forma densa) mediante la presencia o ausencia de sus trigramas (tripletas de tokens) sabiendo que el vocabulario es de 2000 tokens?

A) Aproximadamente 24 Kbytes
B) Aproximadamente 93 Gbytes
C) Aproximadamente 73 Kbytes
D) Ninguna de las anteriores

Tamaño = D * |V| * up_rounding(((log2(máximo nº de ocurrencias + 1) / 8)) = 100 * 2000^3 * 1 B = 100 * 2000^3 = 100 GB.

$\boxed{\text{B}}$ Dado un espacio de representación de d dimensiones se desea reducir a k dimensiones mediante PCA. Para ello se dispone de una matriz A_{dxn} compuesta por los n vectores de entrenamiento menos la media de dichos vectores. Entonces:

A) Escogeremos los k mayores eigenvectores (mayor eigenvalor asociado) de la matriz: $\frac{1}{n}\ A^t A$

B) Escogeremos los k mayores eigenvectores (mayor eigenvalor asociado) de la matriz: $\frac{1}{n}\ AA^t$

C) Escogeremos los k mayores eigenvectores (mayor eigenvalor asociado) de la matriz: $\frac{1}{n}\ A$

D) Escogeremos k mayores eigenvectores (mayor eigenvalor asociado) de la matriz: $\frac{1}{n}\ A^t$

$\boxed{\text{B}}$ Dado un problema de clasificiación en C clases donde los objetos se representan en un espacio de representación de d dimensiones. Se desea obtener una representación en un espacio reducido de $k < C - 1$ dimensiones. Mediante PCA se obtiene W como matriz de proyección a d' dimensiones y a partir de los datos una vez proyectados mediante PCA se obtiene V como la matriz de proyección mediante LDA, o sea: $\mathbf{x}' = VW\mathbf{x}$. Las dimensiones de dichas matrices son:

A) $W_{d' \times k}\ V_{d \times d'}$

B) $W_{d' \times d}\ V_{k \times d'}$

C) $W_{d \times d'}\ V_{d' \times k}$

D) $W_{d \times d'}\ V_{k \times d'}$

$\boxed{\text{A}}$ Cuál de las siguientes expresiones representa una solución al problema de optimización propuesto en LDA:

A) $W^* = \arg\max_W\ Tr(WS_bW^t)$ sujeto a $Tr(WS_wW^t) = 1$

B) $W^* = \arg\max_W\ Tr(WS_wW^t)$ sujeto a $Tr(WS_bW^t) = 1$

C) $W^* = \arg\max_W\ Tr(WS_w^-1W^t)$ sujeto a $Tr(WS_bW^t) = 1$

D) $W^* = \arg\max_W\ Tr(WS_bW^t)$ sujeto a $Tr(WS_w^-1W^t) = 1$

$\boxed{\text{C}}$ Cuál de las siguientes afirmaciones respecto a kernels es falsa:

A) Las funciones kernel modelan el producto escalar de dos vectores en un espacio de representación alternativo

B) No es necesario representar los vectores en el espacio de representación alternativo

C) El kernel polinomial es $K(\mathbf{x}, \mathbf{y}) = (\mathbf{xy} + c)^d$ con $c, d < 0$

D) El algorimo Kernel Perceptron acaba cuando todas las muestras de aprendizaje están bien clasificadas

$\boxed{\text{D}}$ Dado el conjunto de entrenamiento $X = \{(\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), \cdots, (\mathbf{x}_n, c_n)\}$. La función discriminante asociada al problema de clasificación binaria empleando kernels es:

A) $g(\mathbf{x}) = \sum_{i=1}^n \alpha_i c_i K(\mathbf{x}, \mathbf{x}_i)\ +\ \sum_{i=1}^n \alpha_i$

B) $g(\mathbf{x}) = \sum_{i=1}^n \alpha_i c_i K(\mathbf{x}, \mathbf{x}_i)\ +\ \sum_{i=1}^n c_i$

C) $g(\mathbf{x}) = \sum_{i=1}^n c_i K(\mathbf{x}, \mathbf{x}_i)\ +\ \sum_{i=1}^n \alpha_i c_i$

D) $g(\mathbf{x}) = \sum_{i=1}^n \alpha_i c_i K(\mathbf{x}, \mathbf{x}_i)\ +\ \sum_{i=1}^n \alpha_i c_i$

Examen de Teoría de Percepción

ETSINF, Universitat Politècnica de València, Abril de 2014

Apellidos: ________________ **Nombre:** ________

Profesor: ☐ **Jorge Civera** ☐ **Roberto Paredes**

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(2 puntos)** Sean las siguientes 6 muestras en un espacio $\mathcal{R}^4$:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix} \quad \mathbf{x}_4 = \begin{bmatrix} 2 \\ 1 \\ -1 \\ 2 \end{bmatrix} \quad \mathbf{x}_5 = \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix} \quad \mathbf{x}_6 = \begin{bmatrix} -1 \\ -2 \\ 0 \\ -1 \end{bmatrix}$$

asumiendo que los valores y vectores propios de la matrix de covarianza de los datos son:

$$\Lambda = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Nota: Los vectores propios en la matriz B son vectores columna.

a) Calcula la proyección PCA para todas las muestras de $\mathbb{R}^4$ a $\mathbb{R}^2$.

b) Dado que conocemos las etiquetas de clase de las muestras $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} \in A$ y $\{\mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6\} \in B$, calcula el error de clasificación del siguiente clasificador lineal para las muestras proyectadas en $\mathbb{R}^2$:

$$g_A(\mathbf{x}) = x - y + 1$$
$$g_B(\mathbf{x}) = -x + y - 1$$

Solución: En la solución original de este examen se obviaba, de forma errónea, la sustracción de la media a los vectores a proyectar

a) Las muestras dan un vector media $\mu = (\frac{1}{3}, \frac{1}{6}, -\frac{1}{3}, \frac{2}{3})^t$, de forma que los vectores a proyectar serán $\mathbf{x}_i - \mu$, es decir:

$$\mathbf{x}_1 - \mu = \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{6} \\ \frac{1}{3} \\ -\frac{5}{3} \end{bmatrix} \quad \mathbf{x}_2 - \mu = \begin{bmatrix} -\frac{4}{3} \\ -\frac{1}{6} \\ \frac{4}{3} \\ \frac{4}{3} \end{bmatrix} \quad \mathbf{x}_3 - \mu = \begin{bmatrix} \frac{5}{3} \\ -\frac{1}{6} \\ -\frac{5}{3} \\ \frac{1}{3} \end{bmatrix} \quad \mathbf{x}_4 - \mu = \begin{bmatrix} \frac{5}{3} \\ \frac{5}{6} \\ -\frac{2}{3} \\ \frac{4}{3} \end{bmatrix} \quad \mathbf{x}_5 - \mu = \begin{bmatrix} -\frac{4}{3} \\ \frac{11}{6} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} \quad \mathbf{x}_6 - \mu = \begin{bmatrix} -\frac{4}{3} \\ -\frac{13}{6} \\ \frac{1}{3} \\ -\frac{5}{3} \end{bmatrix}$$

Proyectamos las muestras utilizando los dos vectores propios asociados a los dos valores propios de mayor magnitud

$$W = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Esto se reduce a quedarnos con la tercera y segunda componentes de cada una de las muestras:

$$\mathbf{x}'_1 = \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{6} \end{bmatrix} \quad \mathbf{x}'_2 = \begin{bmatrix} \frac{4}{3} \\ -\frac{1}{6} \end{bmatrix} \quad \mathbf{x}'_3 = \begin{bmatrix} -\frac{5}{3} \\ -\frac{1}{6} \end{bmatrix} \quad \mathbf{x}'_4 = \begin{bmatrix} -\frac{2}{3} \\ \frac{5}{6} \end{bmatrix} \quad \mathbf{x}'_5 = \begin{bmatrix} \frac{1}{3} \\ \frac{11}{6} \end{bmatrix} \quad \mathbf{x}'_6 = \begin{bmatrix} \frac{1}{3} \\ -\frac{13}{6} \end{bmatrix}$$

b) Calculamos para cada muestra la función discriminante de cada clase, etiqueta de clase estimada y comparamos con la etiqueta real para ver si se ha producido un error de clasificación:

Muestra	$g_A(\mathbf{x})$	$g_B(\mathbf{x})$	Estimada	Real	Error
$\mathbf{x}'_1$	$\frac{3}{2}$	-$\frac{3}{2}$	A	A	No
$\mathbf{x}'_2$	$\frac{5}{2}$	-$\frac{5}{2}$	A	A	No
$\mathbf{x}'_3$	-$\frac{1}{3}$	$\frac{1}{2}$	B	A	Sí
$\mathbf{x}'_4$	-$\frac{1}{2}$	$\frac{1}{2}$	B	B	No
$\mathbf{x}'_5$	-$\frac{1}{2}$	$\frac{1}{2}$	B	B	No
$\mathbf{x}'_6$	$\frac{7}{2}$	-$\frac{7}{2}$	A	B	Sí

En total se producen 2 errores de clasificación.

2. **(2 puntos)** Sea la siguiente función kernel, $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^2$ y sea el siguiente conjunto de aprendizaje $X = \{(\mathbf{x}_1, +1), (\mathbf{x}_2, +1), (\mathbf{x}_3, +1), (\mathbf{x}_4, -1), (\mathbf{x}_5, -1), (\mathbf{x}_6, -1)\}$ con:

$$\mathbf{x}_1 = \begin{bmatrix}1\\1\end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix}1\\0\end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix}0\\1\end{bmatrix} \quad \mathbf{x}_4 = \begin{bmatrix}2\\1\end{bmatrix} \quad \mathbf{x}_5 = \begin{bmatrix}1\\2\end{bmatrix} \quad \mathbf{x}_6 = \begin{bmatrix}0\\0\end{bmatrix}$$

Se pide:

a) Obtén la matriz kernel K de las muestras de entrenamiento

b) Realiza una iteración (desde $\mathbf{x}_1$ hasta $\mathbf{x}_6$) del algoritmo Kernel Perceptron

c) Clasifica la muestra de test $\mathbf{x} = (0 \quad 2)$ con el valor de los pesos α obtenido en el apartado anterior

Solución:

a) $K = \begin{bmatrix} 9 & 4 & 4 & 16 & 16 & 1 \\ 4 & 4 & 1 & 9 & 4 & 1 \\ 4 & 1 & 4 & 4 & 9 & 1 \\ 16 & 9 & 4 & 36 & 25 & 1 \\ 16 & 4 & 9 & 25 & 36 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$

b) $x_1 \rightarrow \text{error} \Rightarrow \alpha = (1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0)$
$x_4 \rightarrow \text{error} \Rightarrow \alpha = (1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0)$
$x_6 \rightarrow \text{error} \Rightarrow \alpha = (1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1)$

Final: $\alpha = (1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1)$

c) $g(\mathbf{x}) = 9 + 1 - 9 - 1 - 1 - 1 = -2 \Rightarrow c(\mathbf{x}) = -1$