Examen de Teoría de Percepción

ETSINF, Universitat Politècnica de València, Abril de 2016

Apellidos: ______ **Nombre:** ______

Profesor: ☐ **Carlos Martínez** ☐ **Roberto Paredes**

Cuestiones (3 puntos, 30 minutos, sin apuntes)

C ¿En qué situación el clasificador por máxima verosimilitud puede expresarse por $c(x) = \arg\max p(x|c)$?

A) En cualquier situación
B) En ninguna situación
C) Cuando las probabilidades *a priori* $P(c)$ son iguales para todas las clases
D) Cuando hay independencia de $p(x)$

B Supongamos una imagen en escala de grises de tamaño 10×15 centímetros, escaneada con una resolución de 250 puntos por centímetro. ¿Qué tamaño ocuparía suponiendo que se cuantifican los niveles de gris en un único byte?

A) Menos de 5 Mbytes
B) Entre 5 y 10 Mbytes
C) Entre 10 y 15 Mbytes
D) Más de 15 Mbytes

C Tenemos un sistema de reconocimiento de imágenes de gran tamaño con restricciones temporales críticas. ¿Cuál de las siguientes técnicas de detección de puntos de interés para extracción de características locales es más apropiada?

A) Detector de contorno
B) Detector de esquinas complementado con extracción por rejilla
C) Extracción por rejilla complementada por extracción aleatoria
D) Cualquier técnica basada en información e invarianza

D Ante una señal de audio que recoge un rango de frecuencias hasta 16KHz, pero de las cuáles sólo son informativas las frecuencias hasta 8KHz, ¿qué procesos de adquisición son los más apropiados?

A) Muestreo directo a 8KHz
B) Muestreo directo a 16KHz
C) Filtrado paso bajo de 4KHz y muestreo a 8KHz
D) Filtrado paso bajo de 8KHz y muestreo a 16KHz

A Emplear una función global en la representación de *tokens* para clasificación de documentos es conveniente porque:

A) Se emplea información de la colección de documentos completa
B) Reduce el tamaño de los vectores de representación
C) Permite la introducción de contexto basado en secuencias de palabras
D) Permite una representación logarítmica de los documentos, dando mayor capacidad discriminativa

$\boxed{\text{C}}$ PCA se resuelve minimizando el error de reconstrucción. Al final se llega a una solución basada en eigenvectores de la matriz de convarianzas. Al aplicar la transformación lineal con dichos eigenvectores se tiene que:

A) El error de reconstrucción es 0
B) El error de reconstrucción es la suma de los eigenvalores de los eigenvectores empleados
C) La matriz de covarianzas en el espacio proyectado es diagonal
D) El error de reconstrucción es la suma de los eigenvectores empleados

$\boxed{\text{A}}$ ¿Cuál de estas afirmaciones sobre LDA es correcta?

A) LDA obtiene una matriz de proyección lineal optimizando una función objetivo que persigue maximizar las distancias interclase mientras se minimizan las intraclase
B) Es una proyección lineal donde no tiene sentido escoger más de $d-1$ eigenvectores siendo d la dimensionalidad de los datos originales
C) Es una proyección lineal que resulta del análisis de eigenvectores de la matriz de covarianzas Σ_x
D) LDA obtiene una matriz de proyección lineal optimizando una función objetivo que persigue maximizar las distancias intraclase mientras se minimizan las interclase

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

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

B) El uso de kernels es adecuado cuando el espacio de representación original no es linealmente separable

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

D) Las funciones kernel implican tener que proyectar los datos a un espacio de representación alternativo

$\boxed{\text{C}}$ Dado un problema de clasificación en 10 clases donde los objetos se representan en un espacio de 1000 dimensiones. Se desea obtener una representación en un espacio reducido de 2 dimensiones. En general, ¿cuál de las siguientes reducciones es la menos aconsejable desde el punto de vista de estabilidad numérica?

A) Proyectar primero con PCA a 100 dimensiones y luego con LDA a 2

B) Proyectar primero con PCA a 10 dimensiones y luego con LDA a 2

C) Proyectar con LDA a 2

D) Proyectar primero con PCA a 50 dimensiones y luego con LDA a 2

$\boxed{\text{B}}$ Dado un problema de clasificación en C clases donde los objetos se representan en un espacio de d dimensiones, se desea obtener una representación en un espacio reducido de k 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. Se debe de cumplir que:

A) $d' \leq C - 1'$ y $k \leq C - 1 \leq d$

B) $k \leq C - 1$ y $k \leq d' \leq d$

C) $k \leq C - 1 \leq d' \leq d$

D) $d' \leq C - 1$ y $k \leq d$

Examen de Teoría de Percepción

ETSINF, Universitat Politècnica de València, Abril de 2016

Apellidos: ____________ **Nombre:** ____________

Profesor: ☐ **Carlos Martínez** ☐ **Roberto Paredes**

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(2 puntos)** Se dispone del siguiente conjunto de datos sobre $\mathbb{R}^5$ para un problema de tres clases:

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

Se han realizado los procesos para obtener proyecciones PCA y LDA al espacio $\mathbb{R}^2$, dando los siguientes resultados:

$$\overline{\mathbf{x}} = \begin{bmatrix} 0.5 \\ 0.5 \\ -0.5 \\ 1 \\ 0.5 \end{bmatrix} \qquad W_{pca} = \begin{bmatrix} -0.3 & 0.6 \\ 0.6 & -0.2 \\ -0.7 & -0.1 \\ -0.1 & -0.6 \\ 0.3 & 0.5 \end{bmatrix} \qquad W_{lda} = \begin{bmatrix} 0.6 & -0.1 \\ 0.4 & -0.4 \\ 0.4 & -0.2 \\ -0.2 & 0.3 \\ 0.5 & 0.9 \end{bmatrix}$$

Se pide:

a) Obtener la proyección de los datos mediante el uso de PCA. (**0.75 puntos**)

b) Obtener la proyección de los datos mediante el uso de LDA. (**0.5 puntos**)

c) Realizar una representación gráfica de ambas proyecciones. (**0.25 puntos**)

d) A la vista de los resultados, ¿qué técnica de reducción de dimensionalidad parece más adecuada para este caso? Razona tu respuesta. (**0.25 puntos**)

e) ¿Aplicar LDA a una única dimensión daría buenos resultados? Razona tu respuesta. (**0.25 puntos**)

Solución:

a) Primer paso: restar $\overline{\mathbf{x}}$:

	Muestra					Clase
$x_1 - \overline{\mathbf{x}}$	0.5	0.5	1.5	0	0.5	1
$x_2 - \overline{\mathbf{x}}$	1.5	0.5	-0.5	-1	0.5	1
$x_3 - \overline{\mathbf{x}}$	-1.5	1.5	-0.5	0	-0.5	2
$x_4 - \overline{\mathbf{x}}$	0.5	-1.5	1.5	1	-1.5	2
$x_5 - \overline{\mathbf{x}}$	-0.5	-1.5	-0.5	-1	0.5	3
$x_6 - \overline{\mathbf{x}}$	-0.5	0.5	-1.5	1	0.5	3

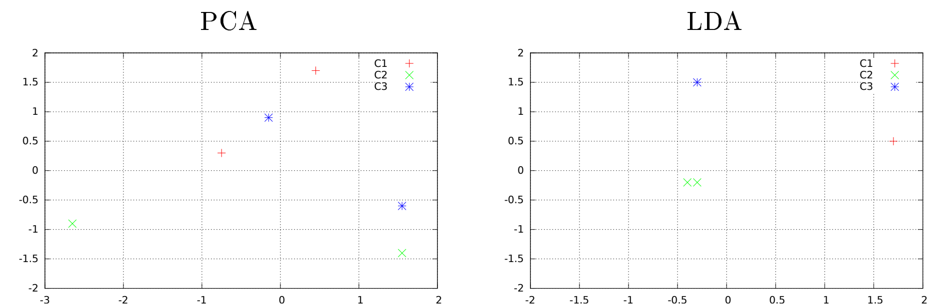
Segundo paso: proyección

	Muestra		Clase
$(x_1 - \overline{\mathbf{x}})W_{pca}$	-0.75	0.3	1
$(x_2 - \overline{\mathbf{x}})W_{pca}$	0.45	1.7	1
$(x_3 - \overline{\mathbf{x}})W_{pca}$	1.55	-1.4	2
$(x_4 - \overline{\mathbf{x}})W_{pca}$	-2.65	-0.9	2
$(x_5 - \overline{\mathbf{x}})W_{pca}$	-0.15	0.9	3
$(x_6 - \overline{\mathbf{x}})W_{pca}$	1.55	-0.6	3

b) Proyección:

	Muestra		Clase
$x_1 W_{lda}$	1.7	0.5	1
$x_2 W_{lda}$	1.7	0.5	1
$x_3 W_{lda}$	-0.4	-0.2	2
$x_4 W_{lda}$	-0.3	-0.2	2
$x_5 W_{lda}$	-0.3	1.5	3
$x_6 W_{lda}$	-0.3	1.5	3

c)



c) LDA permite una clara separabilidad lineal en $\mathbb{R}^2$, por lo que sería la opción más apropiada.

d) No, pues la proyección de la primera dimensión haría indistinguibles los elementos de las clases 2 y 3.

2. **(2 puntos)** Dada la función kernel $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 3)^2$ y el conjunto de aprendizaje en $\mathbb{R}^2$ $X = \{(\mathbf{x}_1, +1), (\mathbf{x}_2, +1), (\mathbf{x}_3, -1), (\mathbf{x}_4, -1), (\mathbf{x}_5, +1)\}$, siendo $\mathbf{x}_1 = [1\ -1]$, $\mathbf{x}_2 = [0\ 1]$, $\mathbf{x}_3 = [1\ 1]$, $\mathbf{x}_4 = [2\ -1]$, $\mathbf{x}_5 = [1\ 0]$, se pide:

   a) Obtener la matriz de kernel K asociada a X (**0.5 puntos**)

   b) Realizar una iteración del algoritmo Kernel Perceptron partiendo del conjunto de pesos $\alpha = (0, 0, 0, 0, 0)$ (**1 punto**)

   c) Clasificar la muestra $\mathbf{y} = [-2\ 1]$ de acuerdo a los pesos obtenidos en el apartado previo (**0.5 puntos**)

Solución:

a) Matriz kernel:

$$K = \begin{vmatrix} 25 & 4 & 9 & 36 & 16 \\ 4 & 16 & 16 & 4 & 9 \\ 9 & 16 & 25 & 16 & 16 \\ 36 & 4 & 16 & 64 & 25 \\ 16 & 9 & 16 & 25 & 16 \end{vmatrix}$$

b) Iteraciones algoritmo kernel perceptron

```
muestra 1, clase 1, fdl=0.000000  0.000000 --> error
muestra 2, clase 1, fdl=5.000000  5.000000 --> ok
muestra 3, clase -1, fdl=10.000000  -10.000000 --> error
muestra 4, clase -1, fdl=20.000000  -20.000000 --> error
muestra 5, clase 1, fdl=-26.000000  -26.000000 --> error

alfas:
1.000000 0.000000 1.000000 1.000000 1.000000
```

c) Clasificación de y. $g(y) = -7$ por lo tanto y pertenece a la clase -1