Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politècnica de València, Junio 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 $c(x) = \arg\max_c p(x)p(x|c)$?

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 $p(x)$

D Sean las funciones discriminantes sobre $\mathbb{R}^2$ $g_A(x) = 2x_1^2 - x_1 + x_2 - 5$, $g_B(x) = -x_1 + 3x_2^2 + 2$, $g_C(x) = 3x_1 - 2x_2 + 4$; indicar en qué clase se clasificaría la muestra $(1, 1)$:

A) En la clase A
B) En la clase B
C) Al azar entre B y C por dar el mismo valor sus funciones discriminantes
D) En la clase C

A ¿Cuál es el objetivo del paso de escalado durante el preproceso de OCR?

A) Uniformizar el tamaño final de las imágenes
B) Eliminar las partes no relevantes de la imagen
C) Igualar las dimensiones horizontal y vertical de la imagen
D) Conseguir una imagen sin transiciones abruptas

B Tenemos una señal de audio de 5 segundos de duración a la cuál se le aplica una ventana de Hamming de anchura $W = 20ms$ y desplazamiento $S = 10ms$. ¿Qué cantidad aproximada de marcos (*frames*) se puede esperar de este proceso?

A) 250
B) 500
C) 350
D) 700

C Indicar la afirmación **incorrecta** respecto a la representación *bag-of-words* de documentos de texto:

A) Cada documento se representa por un vector de tamaño igual al del vocabulario
B) Los vectores pueden ser binarios o de números naturales (*term frequency*)
C) El contenido de cada posición del vector de representación indica la frecuencia de una secuencia de $n > 1$ palabras
D) La colección de documentos se puede representar como una matriz de tamaño $V \times D$ (V tamaño del vocabulario, D número de documentos)

A Con respecto a PCA ¿cuál de las siguientes afirmaciones es falsa?

A) PCA preserva la separabilidad de las clases cuando se aplica después de LDA
B) PCA minimiza el error de reconstrucción de las muestras
C) Con PCA obtenemos una proyección lineal a partir de los k eigenvectores con mayor eigenvalor asociado
D) Con PCA deberemos restar la media de los vectores previamente a realizar la proyección

$\boxed{\text{D}}$ Sea un problema de clasificación en 10 clases donde las muestras se representan en 1000 dimensiones. En general, ¿cuál de las siguientes opciones de reducción de dimensionalidad es la menos adecuada?

A) Reducir con PCA a 10 dimensiones y con LDA a 9
B) Reducir con PCA a 100 dimensiones y con LDA a 9
C) Reducir con PCA a 100 dimensiones y con LDA a 2
D) Reducir con PCA a 10 dimensiones y con LDA a 2

$\boxed{\text{D}}$ Dado un problema de clasificación en C clases donde los objetos se representan en un espacio de representación de d dimensiones, se desea obtener una representación final en un espacio reducido de k dimensiones. Para ello se realizará primero una proyección mediante PCA a d' dimensiones con el fin de evitar singularidades, para posteriormente mediante LDA una proyección final a las k dimensiones. Por lo tanto se debe cumplir que, en general:

A) $d' <= C - 1$ y $k <= d$

B) $k <= C - 1$ y $d' <= d$

C) $d' <= min(C - 1, d)$ y $k <= d$

D) $k <= min(C - 1, d')$ y $d' <= d$

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

$\boxed{\text{B}}$ Sean $K_1(\mathbf{x}, \mathbf{y})$ y $K_2(\mathbf{x}, \mathbf{y})$ dos kernels, indica cuál de las siguientes expresiones *no* es un kernel

A) $(\mathbf{x}^t\mathbf{x}) \cdot K_1(\mathbf{x}, \mathbf{y}) \cdot (\mathbf{y}^t\mathbf{y})$

B) $aK_1(\mathbf{x}, \mathbf{y})^2 + bK_2(\mathbf{x}, \mathbf{y})$

C) $(c + K_1(\mathbf{x}, \mathbf{y}))^d \quad c, d > 0$

D) $\exp(K_1(\mathbf{x}, \mathbf{y}) + 1)$

Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politècnica de València, Junio de 2016

Apellidos: ______ **Nombre:** ______

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

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(2 puntos)**

   Dada una representación de las muestras en dos dimensiones, $\mathbf{x} \in \mathbb{R}^2$, se obtiene a partir de la matriz de covarianzas los siguientes eigenvectores y eigenvalores:

   $\lambda_1 = -0.56155$, $\mathbf{w}_1 = [0.61541 \quad -0.78821]$
   $\lambda_2 = 3.56155$, $\mathbf{w}_2 = [-0.78821 \quad -0.61541]$

   Sea la media de las muestras $\mu = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ y $\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$. Se pide:

   a) La proyección a una dimensión con menor error de reconstrucción (**0.5 puntos**)
   b) El error de reconstrucción de dicha proyección (**0.5 puntos**)
   c) La proyección a dos dimensiones (**0.5 puntos**)
   b) El error de reconstrucción de dicha proyección a dos dimensiones (**0.5 puntos**)

   Solución: Ejecutar en octave:

```
w1=[0.61541 -0.78821]
w2=[-0.78821, -0.61541]
x=[2;0]
mu=[1;0]
```

   a) Para obtener el menor error de reconstrucción se escoge el eigenvector de mayor eigenvalor asociado y se proyecta como $W * (\mathbf{x} - \mu)$

```
w2*(x-mu)
ans = -0.78821
```

   Por lo tanto la proyección a una dimensión es $x' = -0.78821$

   b) La reconstrucción se calcula como $\mu + W^T * (W * (\mathbf{x} - \mu))$, en octave sería:

```
mu+(w2'*(w2*(x-mu)))
ans =
   1.62128
   0.48507
```

   Por lo que $\hat{x} = [1.62128 \ \ 0.48507]^T$ y el error sería $(\mathbf{x} * \hat{\mathbf{x}})^2$, en octave:

```
xh=mu+(w2'*(w2*(x-mu)))
(x-xh)'*(x-xh)
ans =  0.37873
```

   Por lo tanto el error de reconstrucción es 0.37873

   c) Para la proyección a dos dimensiones se conforma la matriz de proyección W con los eigenvectores en filas y se multiplica por $(\mathbf{x} - \mu)$:

```
W=[w2;w1]
W*(x-mu)
ans =
  -0.78821
  0.61541
```

   Por lo tanto la proyección a dos dimensiones es $x' = [-0.78821 \ \ 0.61541]^T$

   d) La reconstrucción se calcula como $\mu + W^T * (W * (\mathbf{x} - \mu))$, en octave sería:

```
mu+(W'*(W*(x-mu)))
ans =
   2.00000
   0.00000
```

   Por lo que $\hat{x} = [2.0 \ \ 0.0]^T$ exactamente igual a $\mathbf{x}$ como era de esperar, error de reconstrucción 0.0

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

   a) Obtener la matriz de kernel K asociada a X (**0.25 puntos**)

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

   c) Clasificar las muestras $\mathbf{y_1} = [1\ \ 2]$ y $\mathbf{y_2} = [0\ \ 0]$ de acuerdo a los pesos obtenidos en el apartado previo (**0.5 puntos**)

   d) ¿Crees que en este problema de clasificación era necesario emplear kernels? Razona la respuesta. (**0.25 puntos**)

Solución:

a) Matriz kernel:

$$K= \begin{vmatrix} 25 & 1 & 9 & 9 \\ 1 & 4 & 0 & 1 \\ 9 & 0 & 9 & 4 \\ 9 & 1 & 4 & 4 \end{vmatrix}$$

b) Iteraciones algoritmo kernel perceptron

```
muestra 1, clase 1, fdl=0.000000  0.000000 --> error
muestra 2, clase -1, fdl=2.000000  -2.000000 --> error
muestra 3, clase 1, fdl=9.000000  9.000000 --> ok
muestra 4, clase -1, fdl=8.000000  -8.000000 --> error

alfas:
1.000000 1.000000 0.000000 1.000000
```

c) Clasificación de y_1:
$g(y_1) = 25 - 0 + 0 - 9 + 1 - 1 + 0 - 1 = 15$ por lo tanto y_1 pertenece a la clase $+1$
Clasificación de y_2:
$g(y_2) = 1 - 1 + 0 - 1 + 1 - 1 + 0 - 1 = -2$ por lo tanto y_2 pertenece a la clase -1

d) No, no es necesario puesto que las muestras de entrenamiento en el espacio original ya son linealmente separables.