Examen de Teoría de Percepción - Recuperación 1^er parcial

ETSINF, Universitat Politècnica de València, Junio de 2015

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

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

Cuestiones (3 puntos, 30 minutos, sin apuntes)

$\boxed{D}$ Supongamos un problema de clasificación en el que se detecta el género de una película. ¿Cómo se expresaría en términos de la clasificación estadística $c(x) = \arg\max_c P(c|x)$?

A) x sería el género y c sería la película
B) x sería la película y c sería el género
C) Tanto x como c representarían la película
D) x sería una *representación* de la película y c una *etiqueta* asociada al género

$\boxed{B}$ Dado un problema de clasificación entre dos clases A y B, con funciones discriminantes asociadas g_A y g_B, la frontera de decisión entre A y B viene dada por las representaciones x que cumplen:

A) $\text{máx}_x\, g_A(x) = \text{máx}_x\, g_B(x)$
B) $g_A(x) = g_B(x)$
C) $\text{mín}_x\, g_A(x) = \text{máx}_x\, g_B(x)$
D) $g_A(x) > g_B(x)$

$\boxed{C}$ En un problema de reconocimiento de imágenes donde el detalle discriminativo mínimo es de 2 milímetros, ¿cual es la frecuencia de muestreo mínima que se debe aplicar (entre las enumeradas) para mantener ese nivel de detalle?

A) 512 muestras por metro
B) 768 muestras por metro
C) 1024 muestras por metro
D) 2048 muestras por metro

$\boxed{A}$ Dados los *codewords* { (a,(1,1)), (m,(3,-1)), (l,(-1,2)), (o,(3,3)) }, indicar la codificación por ese *codebook* de la secuencia (2,-1), (3,0), (2,1), (1,2), (0,2), (-1,1), (0,3), (2,3), (3,2)

A) mmaallloo
B) maaallllo
C) malo
D) mmaaoollo

$\boxed{B}$ La función *ldf* empleada en la clasificación de documentos se caracteriza por:

A) Disminuir la complejidad espacial de la representación *bag-of-words*
B) Atenuar los *tokens* con presencia en muchos documentos
C) Incluir contexto en la representación del documento
D) Evitar el proceso de *stemming*

X PCA se resuelve minimizando el error de reconstrucción. Al final de se llega a un un problema de minimización con restricciones que se resuelve mediante multiplicadores de Lagrange. El problema equivalente sería este:

A) $\mathbf{w}^* = \arg\min_{\mathbf{w}\in R^d} (\mathbf{w}\Sigma_{\mathbf{x}}\mathbf{w}^t + (1 - \mathbf{w}\mathbf{w}^t))$
B) $\mathbf{w}^* = \arg\min_{\mathbf{w}\in R^d} (\mathbf{w}\Sigma_{\mathbf{x}}\mathbf{w}^t + \lambda(1 - \mathbf{w}\mathbf{w}^t))$
C) $\mathbf{w}^* = \arg\min_{\mathbf{w}\in R^d} (\mathbf{w}\Sigma_{\mathbf{x}}\mathbf{w}^t + (1 - \mathbf{w}^t\mathbf{w}))$
D) Ninguno de los anteriores

Esta pregunta supone que los vectores son fila (no columna), no plantea la optimización de λ y cambia maximización por minimización. Por tanto, la opción dada inicialmente por correcta (C) no sería correcta.

B Dada la diagonalización de la matriz de covarianzas $\Sigma_{3\times3}$ en valores y vectores propios $\lambda_1 = 0.7$ con $\mathbf{w}_1 = (1\ 0\ 0)$, $\lambda_2 = 5.2$ con $\mathbf{w}_2 = (0\ 1\ 0)$, y $\lambda_3 = 2.7$ con $\mathbf{w}_3 = (0\ 0\ 1)$:

A) La proyección PCA de $\mathbb{R}^3$ a $\mathbb{R}^2$ se llevará a cabo con los vectores propios $\mathbf{w}_1$ y $\mathbf{w}_2$
B) La proyección PCA de $\mathbb{R}^3$ a $\mathbb{R}^2$ se llevará a cabo con los vectores propios $\mathbf{w}_2$ y $\mathbf{w}_3$
C) La proyección PCA de $\mathbb{R}^3$ a $\mathbb{R}^1$ se llevará a cabo con el vector propio $\mathbf{w}_1$
D) Ninguna de las anteriores dado que los eigenvectores no son ortonormales

D ¿Cuál de estas afirmaciones sobre LDA **NO** 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 $C - 1$ eigenvectores siendo C el número de clases
C) Es una proyección lineal que resulta del análisis de eigenvectores generalizados de dos matrices comúnmente expresadas como S_w y S_b
D) LDA obtiene una matriz de proyección lineal optimizando una función objetivo que persigue maximizar las distancias intraclase mientras se minimizan las interclase

A Se recomienda emplear funciones kernel cuando:

A) El espacio de representación original no es linealmente separable
B) El kernel escogido modela el producto escalar en un nuevo espacio de representación
C) El nuevo espacio de representación cabe en memoria
D) El espacio de representación original es linealmente separable

C Esencialmente, el algoritmo Kernel Perceptron lo que hace es:

A) Incrementar la importancia (peso) de las muestras correctamente clasificadas
B) Decrementar la importancia (peso) de las muestras correctamente clasificadas
C) Incrementar la importancia (peso) de las muestras incorrectamente clasificadas
D) Decrementar la importancia (peso) de las muestras incorrectamente clasificadas

Examen de Teoría de Percepción - Recuperación 1$^{\text{er}}$ parcial

ETSINF, Universitat Politècnica de València, Junio de 2015

Apellidos: ☐ **Nombre:** ☐

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

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(2 puntos)** Obtenida la siguiente matriz de proyección en un problema de dos clases:

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

   a) ¿Esta matriz de proyección proviene de realizar PCA? Razona la respuesta (**0.5 puntos**)

   b) ¿Esta matriz de proyección proviene de realizar LDA? Razona la respuesta (**0.5 puntos**)

   c) Obtener la proyección de los siguientes puntos (**0.5 puntos**)
   $\mathbf{x}_1 = (1, 0, -1, 1)$, $\mathbf{x}_2 = (1, 1, -2, 0)$, $\mathbf{x}_3 = (-1, 2, 2, -1)$, $\mathbf{x}_4 = (-1, 2, 2, -2)$

   d) Clasificar los puntos en el espacio proyectado mediante las siguientes funciones discriminantes (**0.5 puntos**):

$$\begin{array}{lll} g_A(\mathbf{x}) = \mathbf{w}_A \mathbf{x} & \text{con:} \quad \mathbf{w}_A = (1, 2, 1) & \text{*Notación compacta} \\ g_B(\mathbf{x}) = \mathbf{w}_B \mathbf{x} & \quad\quad\;\; \mathbf{w}_B = (1, -1, -1) & \end{array}$$

Solución:

a) No, no son ortonormales (ni ortogonales)

b) No, debería ser solo una fila (C-1)

c) Proyectar:

$$\begin{aligned} \mathbf{x}'_1 &= W\mathbf{x}_1 = (2, 1) \\ \mathbf{x}'_2 &= W\mathbf{x}_2 = (3, 2) \\ \mathbf{x}'_3 &= W\mathbf{x}_3 = (-1, 1) \\ \mathbf{x}'_4 &= W\mathbf{x}_4 = (-2, 1) \end{aligned}$$

c) Clasificar: Para clasificar hay que emplear la notación compacta:

$$\begin{aligned} \mathbf{x}'_1 &= (1, 2, 1) \\ \mathbf{x}'_2 &= (1, 3, 2) \\ \mathbf{x}'_3 &= (1, -1, 1) \\ \mathbf{x}'_4 &= (1, -2, 1) \end{aligned}$$

Y aplicar las funciones discriminantes de cada clase:

$$\begin{array}{rrl} g_A(\mathbf{x}'_1) = 6 & g_B(\mathbf{x}'_1) = -2 & \text{Clase } A \\ g_A(\mathbf{x}'_2) = 9 & g_B(\mathbf{x}'_2) = -4 & \text{Clase } A \\ g_A(\mathbf{x}'_3) = 0 & g_B(\mathbf{x}'_3) = 1 & \text{Clase } B \\ g_A(\mathbf{x}'_4) = -2 & g_B(\mathbf{x}'_4) = 2 & \text{Clase } B \end{array}$$

2. **(2 puntos)** Dada la función kernel

$$K(\mathbf{x},\mathbf{y}) = (|\mathbf{x}|^2 + |\mathbf{y}|^2)(\mathbf{x}\cdot\mathbf{y} + 2)$$

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

$$\mathbf{x}_1 = [-1\ 1], \mathbf{x}_2 = [0\ 1], \mathbf{x}_3 = [-1\ 0], \mathbf{x}_4 = [-1\ \ -1], \mathbf{x}_5 = [0\ \ -1]$$

se pide:

a) Obtener la matriz de kernel K asociada a X (**0.75 puntos**)

b) Realizar una iteración completa del algoritmo Kernel Perceptron (de x_1 a x_5) partiendo del conjunto de pesos $\alpha = (0,0,0,0,0)$ (**0.75 puntos**)

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

Solución:

a) K= $\begin{vmatrix} 16 & 9 & 9 & 8 & 3 \\ 9 & 6 & 4 & 3 & 2 \\ 9 & 4 & 6 & 9 & 4 \\ 8 & 3 & 9 & 16 & 9 \\ 3 & 2 & 4 & 9 & 6 \end{vmatrix}$

b) $\mathbf{x}_1$: $g(\mathbf{x}_1) = 0$, $c_1 g(\mathbf{x}_1) = 0 \leq 0$, $\alpha_1 = 1$, $\alpha = (1,0,0,0,0)$
$\mathbf{x}_2$: $g(\mathbf{x}_2) = 10$, $c_2 g(\mathbf{x}_2) = 10 > 0$
$\mathbf{x}_3$: $g(\mathbf{x}_3) = 10$, $c_3 g(\mathbf{x}_3) = -10 \leq 0$, $\alpha_3 = 1$, $\alpha = (1,0,1,0,0)$
$\mathbf{x}_4$: $g(\mathbf{x}_4) = -1$, $c_4 g(\mathbf{x}_4) = -1 \leq 0$, $\alpha_4 = 2$, $\alpha = (1,0,1,1,0)$
$\mathbf{x}_5$: $g(\mathbf{x}_5) = 9$, $c_5 g(\mathbf{x}_5) = -9 \leq 0$, $\alpha_5 = 1$, $\alpha = (1,0,1,1,1)$

c) $K(\mathbf{x}_1,\mathbf{y}) = 8$, $K(\mathbf{x}_2,\mathbf{y}) = 9$, $K(\mathbf{x}_3,\mathbf{y}) = 3$, $K(\mathbf{x}_4,\mathbf{y}) = 0$, $K(\mathbf{x}_5,\mathbf{y}) = 3$
Por tanto, $g(y) = 1\cdot1\cdot8+0\cdot1\cdot9+1\cdot(-1)\cdot3+1\cdot1\cdot0+1\cdot(-1)\cdot3+1\cdot1+0\cdot1+1\cdot(-1)+1\cdot1+1\cdot(-1) = 8-3-3 = 2$, $c(y) = +1$