Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio 2013

Apellidos: ______ **Nombre:** ______

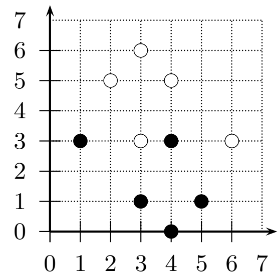
Profesor: ☐ **Carlos Martínez - Jorge Civera** ☐ **Roberto Paredes**

Cuestiones (3 puntos, 30 minutos, sin apuntes)

C | Los clasificadores basados en distancias

A) aprenden las distribuciones de probabilidad sin tener que memorizar los prototipos
B) requieren definir métricas
C) son clasificadores basados en la similitud a ejemplos dados o prototipos
D) inicialmente no requieren datos supervisados

D | La figura de la derecha muestra prototipos bidimensionales de 2 clases: $x_0 = (4, 3, \bullet)$, $x_1 = (3, 3, \circ)$, $x_2 = (1, 3, \bullet)$, $x_3 = (6, 3, \circ)$, $x_4 = (3, 1, \bullet)$, $x_5 = (4, 5, \circ)$, $x_6 = (4, 0, \bullet)$, $x_7 = (3, 6, \circ)$, $x_8 = (5, 1, \bullet)$, $x_9 = (2, 5, \circ)$. ¿Cuántos prototipos se eliminan al aplicar el algoritmo de edición de Wilson por orden de índice utilizando los 3 vecinos más próximos en distancia Euclídea?



A) 1
B) 2
C) 3
D) 4

B | Se puede asegurar que la probabilidad de error del clasificador por los k vecinos más próximos siendo N el número de muestras de aprendizaje tiende al error de Bayes cuando:

A) $k > 1$ y $N \to \infty$

B) $k = logN$ y $N \to \infty$

C) $N = \log k$ y $k \to \infty$

D) $N > 1$ y $k \to \infty$

A | El clasificador de Bernoulli se aplica a objetos representados mediante:

A) Vectores de bits.
B) Vectores de contadores (enteros no negativos) cuya suma es constante.
C) Vectores de reales.
D) Cadenas de símbolos.

D | ¿Cuál es el objetivo de las técnicas de suavizado?

A) Evitar probabilidades nulas.
B) Generalizar una distribución de probabilidad aprendida sobre una muestra.
C) Mejorar la tasa de acierto de un clasificador.
D) Todas las anteriores son correctas.

C | En una distribución gaussiana multidimensional, una matriz de covarianzas diagonal:

A) es la única posibilidad.
B) es una matriz de covarianzas incorrecta.
C) indica que las componentes carecen de correlación lineal entre sí.
D) indica que las componentes poseen una correlación lineal entre sí.

$\boxed{\text{A}}$ 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

$\boxed{\text{B}}$ El kernel polinomial presenta la forma:

A) $K(\mathbf{x},\mathbf{y}) = (\mathbf{x}+\mathbf{y})^d$
B) $K(\mathbf{x},\mathbf{y}) = (\mathbf{x}\cdot\mathbf{y}+c)^d$
C) $K(\mathbf{x},\mathbf{y}) = (\mathbf{x}\cdot\mathbf{y})+c^d$
D) $K(\mathbf{x},\mathbf{y}) = (\mathbf{x}^d+\mathbf{y}^c)$

$\boxed{\text{D}}$ La realimentación no determinista suele modelarse a partir del modelo estadístico:

A) $\hat{h} \approx \arg\max_h P(f|d)P(d|h')P(x|h)P(h|h',d)$
B) $(\hat{h},\hat{d}) \approx \arg\max_{h,d} P(d|h')P(x|h)$
C) $(\hat{h},\hat{d}) \approx \arg\max_{h,d} P(f|d)P(x|h)$
D) $(\hat{h},\hat{d}) \approx \arg\max_{h,d} P(f|d)P(d|h')P(x|h)P(h|h',d)$

$\boxed{\text{A}}$ La medida de la calidad de los sistemas interactivos se basa en:

A) el esfuerzo del operador humano.
B) el tiempo de cómputo del aprendizaje adaptativo.
C) la tasa de error final del sistema.
D) el número de acciones de supervisión propuesto por el sistema, sin tener en cuenta las correcciones del operador.

Examen de Teoría de Percepción - Recuperación Segundo Parcial

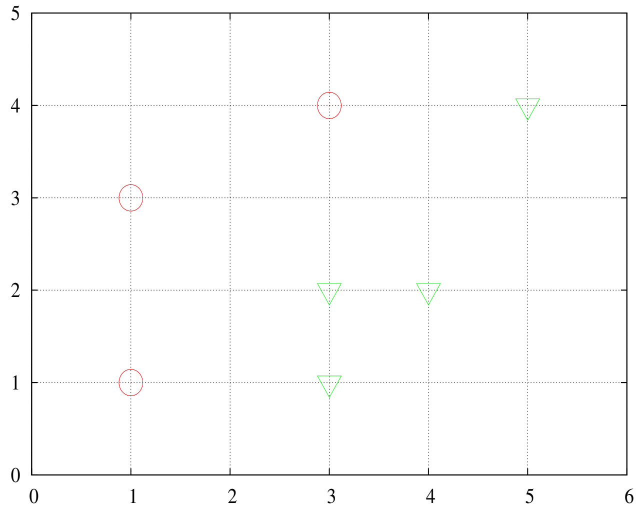
ETSINF, Universitat Politècnica de València, Junio 2013

Apellidos: **Nombre:**

Profesor: ☐ **Carlos Martínez - Jorge Civera** ☐ **Roberto Paredes**

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(1 punto)** Dado el conjunto de aprendizaje $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7\}$ con la distribución de clases que muestra la figura. Realiza una ejecución del algoritmo de condensado de Hart.



Clase A

$\mathbf{x}_1 = (1, 1)$

$\mathbf{x}_3 = (1, 3)$

$\mathbf{x}_5 = (3, 4)$

Clase B

$\mathbf{x}_2 = (3, 1)$

$\mathbf{x}_4 = (4, 2)$

$\mathbf{x}_6 = (3, 2)$

$\mathbf{x}_7 = (5, 4)$

El orden de recorrido de los prototipos en el algoritmo de Hart es el del índice de los mismo: $\mathbf{x}_1 \dots \mathbf{x}_7$

Solución:

Primera parte, construcción del conjunto S y G:

$$\begin{aligned}
&\mathbf{x}_1 \to S \\
&\mathbf{x}_2, \text{Error} \to S \\
&\mathbf{x}_3, \text{Acierto} \to G \\
&\mathbf{x}_4, \text{Acierto} \to G \\
&\mathbf{x}_5, \text{Error} \to S \\
&\mathbf{x}_6, \text{Acierto} \to G \\
&\mathbf{x}_7, \text{Error} \to S
\end{aligned}$$

$S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_7\}$ y $G = \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6\}$

Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_3, \text{Acierto, no mover a } S \\
&\mathbf{x}_4, \text{Acierto, no mover a } S \\
&\mathbf{x}_6, \text{Acierto, no mover a } S \\
&error = 0 \to \text{Acabar}
\end{aligned}$$

Acaba con el conjunto reducido: $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_7\}$

2. **(1 punto)** Tenemos $N = 12$ vectores binarios tridimensionales aleatoriamente extraídas de $C = 3$ distros Bernoulli independientes:

n	1	2	3	4	5	6	7	8	9	10	11	12
x_{n1}	1	1	1	1	0	0	0	0	0	1	1	0
x_{n2}	0	0	0	0	1	1	0	1	1	1	1	1
x_{n3}	1	0	1	0	1	1	1	1	0	0	0	0
c_n	1	1	1	1	2	2	2	2	3	3	3	3

a) Estima todos los parámetros del clasificador Bernoulli más probable.

b) Suaviza los prototipos Bernoulli con truncamiento simple ($\epsilon = 1/8$) y clasifica la muestra $y = (0\ 0\ 1)$.

Solución:

a)

$$p(1) = p(2) = p(3) = \frac{1}{3}$$

$$\hat{\mathbf{p}}_1 = \frac{1}{4}\begin{pmatrix} 1+1+1+1 \\ 0+0+0+0 \\ 1+0+1+0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1/2 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \frac{1}{4}\begin{pmatrix} 0+0+0+0 \\ 1+1+0+1 \\ 1+1+1+1 \end{pmatrix} = \begin{pmatrix} 0 \\ 3/4 \\ 1 \end{pmatrix}$$

$$\hat{\mathbf{p}}_3 = \frac{1}{4}\begin{pmatrix} 0+1+1+0 \\ 1+1+1+1 \\ 0+0+0+0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1 \\ 0 \end{pmatrix}$$

b)

$$\hat{\mathbf{p}}_1 = \begin{pmatrix} 1 \\ 0 \\ 1/2 \end{pmatrix} \to \hat{\mathbf{p}}_1 = \begin{pmatrix} 7/8 \\ 1/8 \\ 1/2 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \begin{pmatrix} 0 \\ 3/4 \\ 1 \end{pmatrix} \to \hat{\mathbf{p}}_2 = \begin{pmatrix} 1/8 \\ 3/4 \\ 7/8 \end{pmatrix}$$

$$\hat{\mathbf{p}}_3 = \begin{pmatrix} 1/2 \\ 1 \\ 0 \end{pmatrix} \to \hat{\mathbf{p}}_3 = \begin{pmatrix} 1/2 \\ 7/8 \\ 1/8 \end{pmatrix}$$

En nuestro caso, dado que las priors son idénticas, la regla de clasificación se reduce a: $\hat{c}(y) = \arg\max_c p(y \mid c)$

$$p(y = (0\ 0\ 1) \mid c = 1) = (7/8)^0(1-7/8)^{(1-0)}(1/8)^0(1-1/8)^{(1-0)}(1/2)^1(1-1/2)^{(1-1)} = 1/8 \cdot 7/8 \cdot 1/2 = \frac{7}{128}$$

$$p(y = (0\ 0\ 1) \mid c = 2) = (1/8)^0(1-1/8)^{(1-0)}(3/4)^0(1-3/4)^{(1-0)}(7/8)^1(1-7/8)^{(1-1)} = 7/8 \cdot 1/4 \cdot 7/8 = \frac{49}{256}$$

$$p(y = (0\ 0\ 1) \mid c = 3) = (1/2)^0(1-1/2)^{(1-0)}(7/8)^0(1-7/8)^{(1-0)}(1/8)^1(1-1/8)^{(1-1)} = 1/2 \cdot 1/8 \cdot 1/8 = \frac{1}{128}$$

La muestra y se clasifica en la clase 2.

3. **(1 punto)** Se definen las clases A y B en un espacio bidimensional; cada clase está modelada respectivamente por una gaussiana, de forma que sus parámetros son:

$$\mu_A = (\ 1 \quad 1\ )^t \quad \Sigma_A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\mu_B = (\ 0 \quad -1\ )^t \quad \Sigma_B = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$

Sabiendo que las probabilidades *a priori* de cada clase son $P(A) = 0.4$ y $P(B) = 0.6$, clasifica la muestra $\mathbf{x} = (\ 1 \quad -1\ )^t$.

Solución:

Usando clasificación por máxima verosimilitud y regla de Bayes, el problema se reduce a calcular

$$c^* = \arg\max_{c \in \{A,B\}} P(c)p(x|c)$$

Por tanto:

$$p(x|A) = \frac{1}{|\Sigma_A|^{\frac{1}{2}} 2\pi} \cdot e^{-\frac{1}{2}(\mathbf{x}-\mu_A)^t \Sigma_A^{-1} (\mathbf{x}-\mu_A)} = \frac{1}{2\sqrt{2}\pi} \cdot e^{-\frac{1}{2}(\ 0 \quad -2\ ) \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 \\ -2 \end{pmatrix}} = \frac{1}{2\sqrt{2}\pi} \cdot e^{-1}$$

$$p(x|B) = \frac{1}{|\Sigma_B|^{\frac{1}{2}} 2\pi} \cdot e^{-\frac{1}{2}(\mathbf{x}-\mu_B)^t \Sigma_B^{-1} (\mathbf{x}-\mu_B)} \frac{1}{2\sqrt{3}\pi} \cdot e^{-\frac{1}{2}(\ 1 \quad 0\ ) \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}} = \frac{1}{2\sqrt{3}\pi} \cdot e^{-\frac{1}{6}}$$

$$P(A|x) = P(A)p(x|A) = 0.4 \cdot \frac{1}{2\sqrt{2}\pi} \cdot e^{-1} \approx 0.0166$$

$$P(B|x) = P(B)p(x|B) = 0.6 \cdot \frac{1}{2\sqrt{3}\pi} \cdot e^{-\frac{1}{6}} \approx 0.0467$$

Por tanto, se clasifica en la clase **B**.

4. **(1 punto)** Tenemos la muestra de entrenamiento $X = \{(\mathbf{x}_1, +1), (\mathbf{x}_2, -1), (\mathbf{x}_3, +1), (\mathbf{x}_4, -1)\}$, y la matriz de kernel:

$$K = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{2} & 1 & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{1}{3} & 1 \end{pmatrix}$$

Obtén el valor de las α_i que daría el algoritmo Kernel-Perceptrón para este conjunto de datos.

Solución:

Partimos de $\alpha = \{0, 0, 0, 0\}$, usando la función a optimizar $g(\mathbf{x}) = \sum_{i=1}^{4} \alpha_i \cdot c_i \cdot K(\mathbf{x}, \mathbf{x}_i) + \sum_{i=1}^{4} \alpha_i \cdot c_i$.

1^a iteración

- $g(\mathbf{x}_1) = 0 \rightarrow c_1 g(\mathbf{x}_1) = 0 \leq 0 \rightarrow \alpha_1 = 1 \rightarrow \alpha = \{1, 0, 0, 0\}$
- $g(\mathbf{x}_2) = K(\mathbf{x}_2, \mathbf{x}_1) + \alpha_1 = \frac{1}{2} + 1 = \frac{3}{2} \rightarrow c_2 g(\mathbf{x}_2) = -\frac{3}{2} \leq 0 \rightarrow \alpha_2 = 1 \rightarrow \alpha = \{1, 1, 0, 0\}$
- $g(\mathbf{x}_3) = K(\mathbf{x}_3, \mathbf{x}_1) - K(\mathbf{x}_3, \mathbf{x}_2) + \alpha_1 - \alpha_2 = \frac{1}{3} - \frac{1}{2} + 1 - 1 = -\frac{1}{6} \rightarrow c_3 g(\mathbf{x}_3) = -\frac{1}{6} \leq 0 \rightarrow \alpha_3 = 1 \rightarrow \alpha = \{1, 1, 1, 0\}$
- $g(\mathbf{x}_4) = K(\mathbf{x}_4, \mathbf{x}_1) - K(\mathbf{x}_4, \mathbf{x}_2) + K(\mathbf{x}_4, \mathbf{x}_3) + \alpha_1 - \alpha_2 + \alpha_3 = \frac{2}{3} - \frac{1}{3} + \frac{1}{3} + 1 - 1 + 1 = \frac{5}{3} \rightarrow c_4 g(\mathbf{x}_4) = -\frac{5}{3} \leq 0 \rightarrow \alpha_4 = 1 \rightarrow \alpha = \{1, 1, 1, 1\}$

2^a iteración

- $g(\mathbf{x}_1) = K(\mathbf{x}_1, \mathbf{x}_1) - K(\mathbf{x}_1, \mathbf{x}_2) + K(\mathbf{x}_1, \mathbf{x}_3) - K(\mathbf{x}_1, \mathbf{x}_4) + \alpha_1 - \alpha_2 + \alpha_3 - \alpha_4 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{2}{3} + 1 - 1 + 1 - 1 = \frac{1}{6} \rightarrow c_1 g(\mathbf{x}_1) = \frac{1}{6} > 0 \rightarrow \alpha = \{1, 1, 1, 1\}$
- $g(\mathbf{x}_2) = K(\mathbf{x}_2, \mathbf{x}_1) - K(\mathbf{x}_2, \mathbf{x}_2) + K(\mathbf{x}_2, \mathbf{x}_3) - K(\mathbf{x}_2, \mathbf{x}_4) + \alpha_1 - \alpha_2 + \alpha_3 - \alpha_4 = \frac{1}{2} - 1 + \frac{1}{2} - \frac{1}{3} + 1 - 1 + 1 - 1 = -\frac{1}{3} \rightarrow c_2 g(\mathbf{x}_2) = \frac{1}{3} > 0 \rightarrow \alpha = \{1, 1, 1, 1\}$
- $g(\mathbf{x}_3) = K(\mathbf{x}_3, \mathbf{x}_1) - K(\mathbf{x}_3, \mathbf{x}_2) + K(\mathbf{x}_3, \mathbf{x}_3) - K(\mathbf{x}_3, \mathbf{x}_4) + \alpha_1 - \alpha_2 + \alpha_3 - \alpha_4 = \frac{1}{3} - \frac{1}{2} + 1 - \frac{1}{3} + 1 - 1 + 1 - 1 = \frac{1}{2} \rightarrow c_3 g(\mathbf{x}_3) = \frac{1}{2} > 0 \rightarrow \alpha = \{1, 1, 1, 1\}$
- $g(\mathbf{x}_4) = K(\mathbf{x}_4, \mathbf{x}_1) - K(\mathbf{x}_4, \mathbf{x}_2) + K(\mathbf{x}_4, \mathbf{x}_3) - K(\mathbf{x}_4, \mathbf{x}_4) + \alpha_1 - \alpha_2 + \alpha_3 - \alpha_4 = \frac{2}{3} - \frac{1}{3} + \frac{1}{3} - 1 + 1 - 1 + 1 - 1 = -\frac{1}{3} \rightarrow c_4 g(\mathbf{x}_4) = \frac{1}{3} > 0 \rightarrow \alpha = \{1, 1, 1, 1\}$

Al no haber cambios en esta segunda iteración, se tiene finalmente que $\alpha = \{1, 1, 1, 1\}$.