Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2015

Apellidos: ☐ **Nombre:** ☐

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

Cuestiones (3 puntos, 30 minutos, sin apuntes)

$\boxed{\text{C}}$ El suavizado de una distribución de Bernoulli, ¿qué valores de los parámetros suele modificar?

A) Sólo valores cercanos a cero, pues el logaritmo de esos valores tiende a menos infinito.
B) Todos los valores en el rango de cero a uno.
C) Tanto valores cercanos a cero, como valores cercanos a uno.
D) No modifica ningún valor.

$\boxed{\text{B}}$ Dado el siguiente conjunto de vectores binarios bidimensionales:

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

¿Cuál es la estimación de los parámetros del clasificador Bernoulli más probable?

A) $\hat{p}(1) = \frac{1}{2}$, $\hat{p}(2) = \frac{1}{2}$, $\hat{\mathbf{p}}_1 = \left(\frac{3}{4}, \frac{1}{4}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{2}, \frac{1}{2}\right)^t$
B) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{3}{4}, \frac{1}{4}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{2}, \frac{1}{2}\right)^t$
C) $\hat{p}(1) = \frac{1}{2}$, $\hat{p}(2) = \frac{1}{2}$, $\hat{\mathbf{p}}_1 = \left(\frac{5}{6}, \frac{2}{6}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{1}{2}\right)^t$
D) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{5}{6}, \frac{2}{6}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{1}{2}\right)^t$

$\boxed{\text{B}}$ ¿Qué tipo de fronteras de decisión define un clasificador basado en distribuciones multinomiales?

A) Lineal definida a trozos.
B) Lineal.
C) Cuadrática.
D) Ninguna de las anteriores.

$\boxed{\text{A}}$ Sean A y B dos clases con igual prior y f.d.p. condicionales de clase gaussianas con los siguientes parámetros $\mu_A = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma_A = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$, $\mu_B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ y $\Sigma_B = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$, ¿qué tipo de frontera de decisión definen?

A) Lineal
B) Lineal definida a trozos
C) Cuadrática
D) Ninguna de las anteriores

$\boxed{\text{A}}$ Dada la muestra $\mathbf{x} = \begin{pmatrix} 0 \\ 0.25 \end{pmatrix}$, ¿a qué clase sería asignada por un clasificador gaussiano basado en los parámetros de la cuestión anterior?

A) A la clase A
B) A cualquiera de las dos clases
C) A la clase B
D) A ninguna de las dos clases

$\boxed{\text{C}}$ El algoritmo de condensado de Hart acaba cuando:

A) ninguna muestra se clasifica correctamente
B) cuando se vacía el conjunto GARBAGE
C) ninguna muestra se clasifica incorrectamente o se vacía el conjunto GARBAGE
D) ninguna muestra se clasifica incorrectamente

$\boxed{\text{A}}$ El algoritmo de edición de prototipos de Wilson:

A) el resultado puede depender del parámetro k con el que se realiza la clasificación
B) devuelve un conjunto con menos prototipos que el de entrada
C) las muestras ruidosas permanecen en el conjunto GARBAGE
D) se eliminan los prototipos cuya clasificación es correcta

$\boxed{\text{C}}$ Cuál de las siguientes afirmaciones es correcta:

A) En general el Bias aumenta al escoger clasificadores más fuertes
B) En general el Variance aumenta al aumentar el conjunto de aprendizaje
C) En general el Bias aumenta al escoger clasificadores más débiles
D) En general el Variance se reduce empleando boosting

$\boxed{\text{D}}$ En general cuál de las siguientes combinaciones de clasificadores funcionar mejor:

A) Combinar múltiples clasificadores k-nn con diferentes valores de k
B) Combinar múltiples clasificadores lineales con diferentes valores de $\mathbf{w}$
C) Combinar múltiples clasificadores k-nn con diferentes distancias
D) Combinar multiples clasificadores, lineales, k-nn, árboles de decisión etc.

$\boxed{\text{D}}$ En reinforcement learning:

A) Se trata de escoger $T' \subset T$ del menor tamaño posible para supervisar/etiquetar
B) En cada interacción se genera una muestra más que añadir al conjunto de aprendizaje
C) Se emplean los datos corregidos por el operador humano como nuevos datos para la adaptación de los modelos
D) Se calcula en término de operaciones de exploración y explotación

Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2015

Apellidos: ______ **Nombre:** ______

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

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(2 puntos)** Tenemos $N = 6$ cadenas de longitud $x_+ = 2$ procedentes de un vocabulario $V = \{a, b, c\}$ aleatoriamente extraídas de $C = 2$ distribuciones multinomiales independientes, donde las muestras de la clase 1 son $\mathcal{X}_1 = \{aa, bb, aa\}$ y las muestras de la clase 2 son $\mathcal{X}_2 = \{ab, bc, ac\}$. Se pide

   a) Calcula los parámetros del clasificador multinomial más probable respecto a estos datos. (0.5 puntos)

   b) Calcula el error global (Nota: $0^0 = 1$). (0.5 puntos)

   c) Suaviza los parámetros multinomiales de ambas clases aplicando Laplace con $\epsilon = \frac{1}{3}$. (0.5 puntos)

   d) Clasifica la cadena $y = cc$ con el clasificador multinomial **suavizado** del apartado anterior. (0.5 puntos)

Solución:

a) Representando nuestros datos como vectores de contadores tenemos

	aa	bb	aa	ab	bc	ac
x_a	2	0	2	1	0	1
x_b	0	2	0	1	1	0
x_c	0	0	0	0	1	1
c_n	1	1	1	2	2	2

La estimación de los parámetros del clasificador multinomial es

$$p(1) = p(2) = \frac{3}{6} = \frac{1}{2}$$

$$\hat{\mathbf{p}}_1 = \frac{1}{6}\begin{pmatrix} 2+0+2 \\ 0+2+0 \\ 0+0+0 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 4 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ 0 \end{pmatrix}$$

$$\hat{\mathbf{p}}_2 = \frac{1}{6}\begin{pmatrix} 1+0+1 \\ 1+1+0 \\ 0+1+1 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$$

b) En general, el error global $p(e)$ se calcula como

$$p(e) = \sum_x p(x, e) = \sum_x p(x)\, p(e \mid x) = \sum_x p(x)\, \left(1 - \max_c p(c \mid x)\right).$$

Como sólo tenemos dos clases con igual prior y aplicando Bayes

$$\begin{aligned}
p(e) &= \sum_x p(x)\, \min_c p(c \mid x) \\
&= \sum_x p(x)\, \min_c \frac{p(c)\, p(x \mid c)}{p(x)} = \sum_x \min_c p(c)\, p(x \mid c) \\
&= \frac{1}{2} \sum_x \min_c p(x \mid c) \\
&= \frac{1}{2} \sum_x \frac{x_+!}{x_1!\, x_2!\, x_3!} \min\left(p_{1a}^{x_a} \cdot p_{1b}^{x_b} \cdot p_{1c}^{x_c},\, p_{2a}^{x_a} \cdot p_{2b}^{x_b} \cdot p_{2c}^{x_c}\right) \\
&= \frac{1}{2} \sum_x \frac{x_+!}{x_1!\, x_2!\, x_3!} \min\left(\left(\frac{2}{3}\right)^{x_a} \cdot \left(\frac{1}{3}\right)^{x_b} \cdot 0^{x_c},\, \left(\frac{1}{3}\right)^{x_a} \cdot \left(\frac{1}{3}\right)^{x_b} \cdot \left(\frac{1}{3}\right)^{x_c}\right)
\end{aligned}$$

$\mathbf{x}$			$\frac{x_+!}{x_1!\,x_2!\,x_3!}$	$\prod_d p_{cd}^{x_d}$		
x_a	x_b	x_c		$c=1$	$c=2$	mín_c
0	0	2	1	0	$\frac{1}{9}$	0
0	2	0	1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
2	0	0	1	$\frac{4}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
0	1	1	2	0	$\frac{1}{9}$	0
1	0	1	2	0	$\frac{1}{9}$	0
1	1	0	2	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

Sumando para cada x:

$$p(e) = \frac{1}{2}\left(1\cdot 0 + 1\cdot\frac{1}{9} + 1\cdot\frac{1}{9} + 2\cdot 0 + 2\cdot 0 + 2\cdot\frac{1}{9}\right) = \frac{2}{9}$$

c)

$$\hat{\mathbf{p}}_1 = \frac{1}{2}\begin{pmatrix} \frac{2}{3}+\frac{1}{3} \\ \frac{1}{3}+\frac{1}{3} \\ 0+\frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{6} \end{pmatrix}$$
$$\hat{\mathbf{p}}_2 = \frac{1}{2}\begin{pmatrix} \frac{1}{3}+\frac{1}{3} \\ \frac{1}{3}+\frac{1}{3} \\ \frac{1}{3}+\frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$$

d)

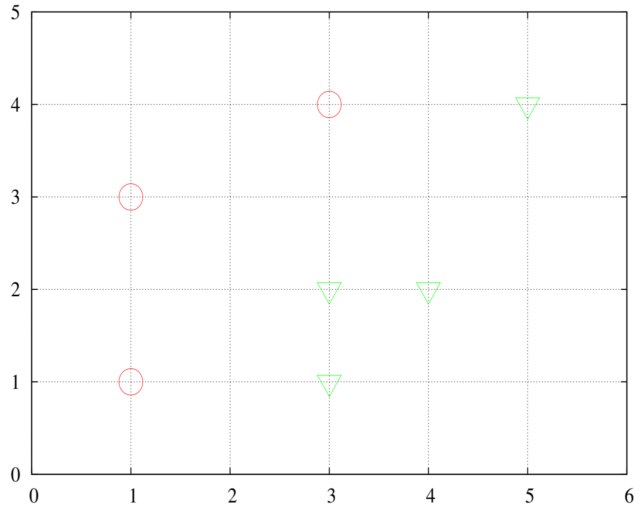
En nuestro caso, dado que las priors son idénticas, la regla de clasificación se reduce a:

$$\hat{c}(y) = \arg\max_c p(y \mid c) = \arg\max_c \prod_d p_{cd}^{x_d}$$

$$p(y=(0\;0\;2)^t \mid c=1) = \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{3}\right)^0 \cdot \left(\frac{1}{6}\right)^2 = \frac{1}{36}$$
$$p(y=(0\;0\;2)^t \mid c=2) = \left(\frac{1}{3}\right)^0 \cdot \left(\frac{1}{3}\right)^0 \cdot \left(\frac{1}{3}\right)^2 = \frac{1}{9}$$

La muestra y se clasifica en la clase 2.

2. **(2 puntos)** 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 con distancia euclídea y $k = 1$.



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 **decreciente** con el índice de los mismo: $\mathbf{x}_7 \ldots \mathbf{x}_1$.

En caso de **empate** de distancias se clasifica a la clase incorrecta.

Solución:

Primera parte, construcción del conjunto S y G:

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

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

Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_6, \text{Acierto, no mover a } S\\
&\mathbf{x}_3, \text{Acierto, no mover a } S\\
&\mathbf{x}_2, \text{Acierto, no mover a } S\\
&error = 0 \rightarrow \text{Acabar}
\end{aligned}$$

Acaba con el conjunto reducido: $S = \{\mathbf{x}_7, \mathbf{x}_5, \mathbf{x}_4, \mathbf{x}_1\}$