Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2016

Apellidos: ☐ **Nombre:** ☐

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

Cuestiones (3 puntos, 30 minutos, sin apuntes)

$\boxed{\text{D}}$ ¿Cuál de los siguientes valores del parámetro $\mathbf{p}$ no define una distribución Bernoulli?

A) $\mathbf{p} = \left(\frac{3}{4}, \frac{1}{4}\right)^t$
B) $\mathbf{p} = \left(\frac{3}{4}, \frac{3}{4}\right)^t$
C) $\mathbf{p} = \left(\frac{1}{2}, \frac{1}{2}\right)^t$
D) Todos los valores anteriores del parámetro $\mathbf{p}$ definen una distribución Bernoulli.

$\boxed{\text{B}}$ ¿Qué tipo de fronteras de decisión define un clasificador basado en la distribución Bernoulli?

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

$\boxed{\text{D}}$ Dado el siguiente conjunto de vectores de contadores bidimensionales:

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

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

A) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{2}{5}, \frac{3}{5}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{2}{3}\right)^t$
B) $\hat{p}(1) = \frac{2}{3}$, $\hat{p}(2) = \frac{1}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{2}{3}, \frac{3}{5}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{2}{5}\right)^t$ sum(1) = 25
C) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{2}{3}, \frac{3}{5}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{2}{5}\right)^t$
D) $\hat{p}(1) = \frac{2}{3}$, $\hat{p}(2) = \frac{1}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{2}{5}, \frac{3}{5}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{2}{3}\right)^t$ sum(2) = 15

$\boxed{\text{B}}$ Una técnica de suavizado para la distribución multinomial es el descuento absoluto. En este suavizado se utiliza una distribución generalizada para repartir la masa de probabilidad descontada. ¿Qué tipo de distribución generalizada se podría utilizar en cualquier caso?

A) Una distribución Bernoulli.
B) Una distribución multinomial.
C) Una distribución Gaussiana con matriz de covarianzas diagonal.
D) Una distribución Gaussiana con matriz de covarianzas completa.

$\boxed{\text{B}}$ Un clasificador Gaussiano con matriz de covarianzas común es lineal, porque ...

A) el término cuadrático se hace cero.
B) el término cuadrático es constante.
C) el factor cuadrático W_c se convierte en la matriz identidad.
D) el factor cuadrático W_c tiene determinante nulo.

C En general, para todos dos puntos $\mathbf{x}$,$\mathbf{y}$ en un espacio vectorial se cumple la siguiente relación de distancias

A) $L_1(\mathbf{x}, \mathbf{y}) \leq L_2(\mathbf{x}, \mathbf{y})$
B) $L_1(\mathbf{x}, \mathbf{y}) \leq L_0(\mathbf{x}, \mathbf{y})$
C) $L_2(\mathbf{x}, \mathbf{y}) \leq L_1(\mathbf{x}, \mathbf{y})$
D) $L_2(\mathbf{x}, \mathbf{y}) \leq L_0(\mathbf{x}, \mathbf{y})$

B La frontera de decisión que se obtiene con el vecino más cercano es:

A) Lineal
B) Lineal a trozos
C) Lineal cuando hay más de dos clases
D) Ninguna de las anteriores

D Sea X un conjunto de n muestras de aprendizaje, C el número de clases y k el parámetro de los k-vecinos, entonces el clasificador por los k vecinos es equivalente a:

A) $c(\mathbf{x}) = c$ si $\mathbf{x}_{nn} \in X_c$ y $k > 3$
B) $c(\mathbf{x}) = c$ si $\mathbf{x}_{nn} \in X_c$ y $k > n$
C) $c(\mathbf{x}) = \arg\min\ P(c)$ si $k = n$ ó $1 - nn$ en caso de empate
D) $c(\mathbf{x}) = \arg\max\ P(c)$ si $k = n$ ó $1 - nn$ en caso de empate

C En general, ¿cuál de las siguientes afirmaciones es correcta?

A) El Bias aumenta al escoger clasificadores más fuertes
B) El Variance aumenta al incrementar el conjunto de muestras de aprendizaje
C) El Bias aumenta al escoger clasificadores más débiles
D) El Variance se reduce empleando Boosting

B Esencialmente en Bagging:

A) Se combinan diferentes clasificadores sobre el mismo conjunto de aprendizaje
B) Se combinan el mismo clasificador sobre diferentes conjuntos de aprendizaje
C) Se combinan diferentes clasificadores sobre diferentes conjuntos de aprendizaje
D) Ninguna de las anteriores

A En la teoría interactiva de la decisión, el criterio de decisión de una hipótesis h dada la señal x, la historia h' y la realimentación f es:

A) $\arg\max_h p(h|x, h', f)$
B) $\arg\max_h p(h, h', f|x)$
C) $\arg\max_h p(h, h', x|f)$
D) $\arg\max_h p(f|h, h', x)$

Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2016

Apellidos: ______ **Nombre:** ______

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

Problemas (4 puntos, 90 minutos, con apuntes)

1. **(2 puntos)** Sea A y B dos clases con priors $p(A) = 1/2$ y $p(B) = 1/2$, y f.d.p. condicionales de clase gaussianas $p(\mathbf{x}\,|\,A) \sim N_2(\mu_A, \Sigma_A)$ y $p(\mathbf{x}\,|\,B) \sim N_2(\mu_B, \Sigma_B)$

$$\mu_A = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Sigma_A = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix} \quad \mu_B = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \Sigma_B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

   a) Calcula funciones discriminantes para A y B. (0.5 punto)

   b) Calcula la frontera de decisión entre las clases A y B. (0.5 puntos)

   c) Representa gráficamente la frontera y las regiones de decisión. (0.5 puntos)

   d) Se desea suavizar las matrices de covarianza de ambas clases mediante *flat smoothing*. Calcula el valor de α necesario para que el clasificador resultante sea lineal. (0.5 puntos)

Nota: $\Sigma_A^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$ y $\Sigma_B^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$

Solución:

a) En general, la función discriminante de un clasificador basado en la regla de Bayes se define como

$$g_c(\mathbf{x}) = p(c) \cdot p(\mathbf{x}|c)$$

o de forma equivalente

$$g_c(\mathbf{x}) = \log p(c) + \log p(\mathbf{x}|c)$$

siendo

$$p(\mathbf{x}\,|\,c) \sim N_2(\mu_c, \Sigma_c) = (2\pi)^{-1} \cdot \left|\Sigma_c\right|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_c)^t \Sigma_c^{-1} (\mathbf{x} - \mu_\mathbf{c})\right)$$

aplicamos logaritmo neperiano y operamos para simplificar la expresión resultante eliminando la constante $(2\pi)^{-1}$

$$g_c(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^t \Sigma_c^{-1}\mathbf{x} + \mu_c^t \Sigma_c^{-1}\mathbf{x} + \left(\log p(c) - \frac{1}{2}\log\left|\Sigma_c\right| - \frac{1}{2}\mu_c^t \Sigma_c^{-1}\mu_c\right)$$

Las funciones discriminantes de las clases serían:

$$\begin{aligned} g_A(\mathbf{x}) =& -\frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^t \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}^t \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &+ \left(\log\frac{1}{2} - \frac{1}{2}\log\left|\begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}\right| - \frac{1}{2}\begin{pmatrix} 0 \\ 0 \end{pmatrix}^t \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) \\ =& -x_1^2 - \frac{1}{4}x_2^2 - \log 2 \end{aligned}$$

$$\begin{aligned} g_B(\mathbf{x}) =& -\frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^t \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}^t \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &+ \left(\log\frac{1}{2} - \frac{1}{2}\log\left|\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\right| - \frac{1}{2}\begin{pmatrix} 0 \\ 1 \end{pmatrix}^t \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\ =& -\frac{1}{2}x_1^2 - \frac{1}{4}x_2^2 + \frac{1}{2}x_2 - \log 2 - \frac{1}{2}\log 2 - \frac{1}{4} \end{aligned}$$

b) La frontera de decisión entre las clases A y B se obtiene igualando sus respectivas funciones discriminantes

$$g_A(\mathbf{x}) = g_B(\mathbf{x})$$

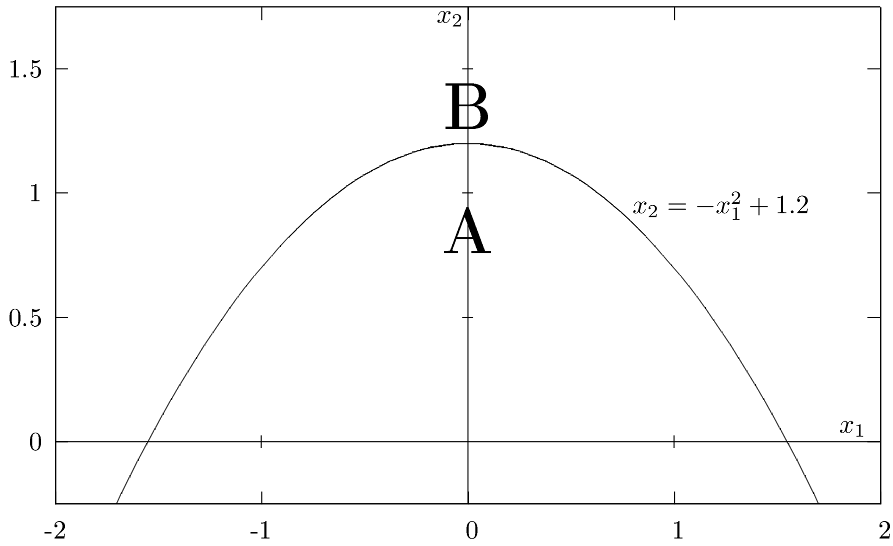
Por tanto,

$$-x_1^2 - \frac{1}{4}x_2^2 - \log 2 = -\frac{1}{2}x_1^2 - \frac{1}{4}x_2^2 + \frac{1}{2}x_2 - \log 2 - \frac{1}{2}\log 2 - \frac{1}{4}$$

que resulta en una parábola vertical

$$x_2 = -x_1^2 + \log 2 + \frac{1}{2} \approx -x_1^2 + 1.2$$

c)



d) El suavizado por *flat smoothing* de un matriz de covarianza $\hat{\Sigma}_c$ se calcula como una combinación lineal de dicha matriz y la matriz de identidad I, de forma que la matriz de covarianzas resultante es:

$$\tilde{\Sigma}_c = \alpha\,\hat{\Sigma}_c + (1-\alpha)\,I \quad \forall c$$

En nuestro caso

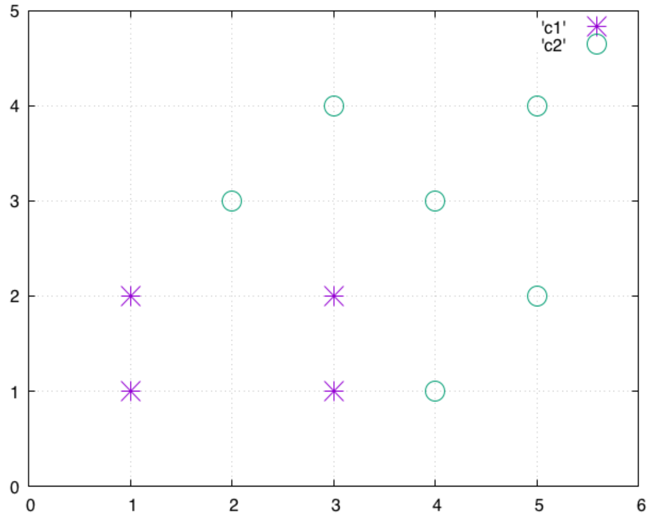
$$\tilde{\Sigma}_A = \alpha \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\alpha}{2}+1 & 0 \\ 0 & \alpha+1 \end{pmatrix}$$
$$\tilde{\Sigma}_B = \alpha \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \alpha+1 \end{pmatrix}$$

Por otra parte, sabemos que un clasificador Gaussiano es lineal si la matriz de covarianzas es la misma para todas las clases. Si igualamos $\tilde{\Sigma}_A$ y $\tilde{\Sigma}_B$ para los elementos no nulos, tenemos

$$\frac{\alpha}{2} + 1 = 1$$
$$\alpha + 1 = \alpha + 1$$

Si resolvemos este sistema de ecuaciones, vemos que la solución es $\alpha = 0$, es decir, obviamente si $\tilde{\Sigma}_A = \tilde{\Sigma}_B = I$.

2. **(2 puntos)** Dado el conjunto de aprendizaje $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8, \mathbf{x}_9, \mathbf{x}_{10}\}$ 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

$$\mathbf{x}_1 = (1, 1)$$
$$\mathbf{x}_3 = (1, 2)$$
$$\mathbf{x}_5 = (3, 2)$$
$$\mathbf{x}_{10} = (3, 1)$$

Clase B

$$\mathbf{x}_2 = (4, 1)$$
$$\mathbf{x}_4 = (5, 2)$$
$$\mathbf{x}_6 = (2, 3)$$
$$\mathbf{x}_7 = (4, 3)$$
$$\mathbf{x}_8 = (3, 4)$$
$$\mathbf{x}_9 = (5, 4)$$

El orden de recorrido de los prototipos en el algoritmo de Hart es **creciente** con el índice de los mismo: $\mathbf{x}_1 \ldots \mathbf{x}_{10}$.

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

Solución:

Primera parte, construcción del conjunto S y G:

$$\begin{aligned}
&\mathbf{x}_1 \rightarrow S\\
&\mathbf{x}_2, \text{Error} \rightarrow S\\
&\mathbf{x}_3, \text{Acierto} \rightarrow G\\
&\mathbf{x}_4, \text{Acierto} \rightarrow G\\
&\mathbf{x}_5, \text{Error} \rightarrow S\\
&\mathbf{x}_6, \text{Error} \rightarrow S\\
&\mathbf{x}_7, \text{Error} \rightarrow S\\
&\mathbf{x}_8, \text{Acierto} \rightarrow G\\
&\mathbf{x}_9, \text{Acierto} \rightarrow G\\
&\mathbf{x}_{10}, \text{Error} \rightarrow S
\end{aligned}$$

$S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_{10}\}$ y $G = \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_8, \mathbf{x}_9\}$

Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_3, \text{Acierto, no mover a } S\\
&\mathbf{x}_4, \text{Acierto, no mover a } S\\
&\mathbf{x}_8, \text{Acierto, no mover a } S\\
&\mathbf{x}_9, \text{Acierto, no mover a } S\\
&error = 0 \rightarrow \text{Acabar}
\end{aligned}$$

Acaba con el conjunto reducido: $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_{10}\}$