

Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2021

Apellidos:

Nombre:

Profesor: ☐ Jorge Civera ☐ Carlos Martínez

Cuestiones (1.5 puntos, 30 minutos, sin apuntes)

☐ A) En una distribución gaussiana, la matriz de covarianzas puede ser:

- A) Diagonal
- B) Triangular superior con elementos no nulos fuera de la diagonal
- C) Triangular inferior con elementos no nulos fuera de la diagonal
- D) Sin elementos nulos y no simétrica

☐ D) Dado un problema de clasificación bidimensional en dos clases A y B equiprobables, donde las distribuciones condicionadas vienen dadas por multinomiales de parámetros $\mathbf{p}_A = \left(\frac{2}{3} \frac{1}{3}\right)$ y $\mathbf{p}_B = \left(\frac{3}{5} \frac{2}{5}\right)$, indica en qué clases se clasifican las muestras $\mathbf{x} = (1 \ 1)$ e $\mathbf{y} = (2 \ 2)$:

- A) \mathbf{x} en la clase A e \mathbf{y} en la clase A
- B) \mathbf{x} en la clase A e \mathbf{y} en la clase B
- C) \mathbf{x} en la clase B e \mathbf{y} en la clase A
- D) \mathbf{x} en la clase B e \mathbf{y} en la clase B

☐ C) Dado el siguiente conjunto de muestras bidimensionales sobre dos clases:

	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	\mathbf{x}_4	\mathbf{x}_5	\mathbf{x}_6
x_{n1}	0	1	-1	3	-1	1
x_{n2}	1	0	-1	2	0	1
c	A	A	A	B	B	B

la estimación de los parámetros de las correspondientes distribuciones gaussianas por máxima verosimilitud dan como resultado:

- A) $\mu_A = (0 \ 0)^t$, $\mu_B = (1 \ 1)^t$, $\Sigma_A = \Sigma_B = \frac{1}{6} \begin{pmatrix} 10 & 5 \\ 5 & 4 \end{pmatrix}$
- B) $\mu_A = \mu_B = \left(\frac{1}{2} \ \frac{1}{2}\right)^t$, $\Sigma_A = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, $\Sigma_B = \frac{1}{3} \begin{pmatrix} 8 & 4 \\ 4 & 2 \end{pmatrix}$
- C) $\mu_A = (0 \ 0)^t$, $\mu_B = (1 \ 1)^t$, $\Sigma_A = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, $\Sigma_B = \frac{1}{3} \begin{pmatrix} 8 & 4 \\ 4 & 2 \end{pmatrix}$
- D) $\mu_A = (0 \ 0)^t$, $\mu_B = (1 \ 1)^t$, $\Sigma_A = \frac{1}{3} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{pmatrix}$, $\Sigma_B = \frac{1}{3} \begin{pmatrix} 13 & -3 & 5 \\ -3 & 1 & -1 \\ 5 & -1 & 2 \end{pmatrix}$

- ☐ B El resultado de un suavizado de una distribución Bernoulli por truncamiento simple ha sido $\tilde{\mathbf{p}} = \left(\frac{1}{10} \frac{1}{2} \frac{19}{20}\right)$. ¿Qué se puede afirmar sobre el valor de la constante de truncamiento ϵ ?
- A) $\epsilon > 0.1$
 - B) $\epsilon < 0.1$
 - C) $\epsilon = 0.1$
 - D) $\epsilon \leq 0.1$
- ☐ A El algoritmo Kernel Perceptron:
- A) Suele emplear la matriz Gramm de los datos de entrenamiento
 - B) Incrementa los pesos de las muestras bien clasificadas
 - C) Tiene la convergencia garantizada
 - D) Se aplica principalmente sobre datos no vectoriales
- ☐ D Un kernel gaussiano se caracteriza por
- A) Proyectar explícitamente a una dimensión infinita
 - B) Emplear la función logaritmo
 - C) Usar como parámetros dos números equivalentes a media y varianza en la distribución gaussiana
 - D) Calcular la norma de un vector
- ☐ D Una característica que diferencia a LDA de PCA es que:
- A) No emplea vectores propios
 - B) Sólo usa la media por clase de los datos
 - C) Permite reducir a cualquier dimensión
 - D) Plantea un problema de optimización distinto
- ☐ B Una característica de la técnica de Bagging es que:
- A) Permite reducir el sesgo
 - B) Combina clasificadores fuertes
 - C) Se aplica a problemas de clasificación no binaria
 - D) No reduce la varianza

Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2021

Apellidos:

Nombre:

Profesor: ☐ Jorge Civera ☐ Carlos Martínez

Problemas (2 puntos, 90 minutos, con apuntes)

1. (1 punto) Se tiene el conjunto de datos siguiente sobre el vocabulario de símbolos $\Sigma = \{a, c, e, l, p\}$:

	cala pala alla	cepa pele lepe	leal peca ella
Clase	A	B	C

Se pide lo siguiente:

- Calcular todos los parámetros del clasificador multinomial por máxima verosimilitud para este conjunto de datos. **(0.5 puntos)**
- Realizar un suavizado con descuento absoluto por *backoff* con $\epsilon = \frac{1}{12}$ en los parámetros multinomiales estimados, usando distribución uniforme sobre todas aquellas componentes que queden a valor nulo. **(0.3 puntos)**
- Clasificar la palabra “lace” usando el clasificador con los parámetros originales y suavizados. **(0.2 puntos)**

Solución:

- a) En primer lugar, transformamos los datos en vectores de ocurrencias siguiendo el orden del alfabeto dado:

	\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
a	2	2	2	1	0	0	1	1	1
c	1	0	0	1	0	0	0	1	0
e	0	0	0	1	2	2	1	1	1
l	1	1	2	0	1	1	2	0	2
p	0	1	0	1	1	1	0	1	0
c_n	A	A	A	B	B	B	C	C	C

Por tanto, los parámetros son:

$$\hat{P}(A) = \hat{P}(B) = \hat{P}(C) = \frac{1}{3}$$

$$\hat{p}_A = \left(\frac{6}{12} \frac{1}{12} 0 \frac{4}{12} \frac{1}{12}\right) \hat{p}_B = \left(\frac{1}{12} \frac{1}{12} \frac{5}{12} \frac{2}{12} \frac{3}{12}\right) \hat{p}_C = \left(\frac{3}{12} \frac{1}{12} \frac{3}{12} \frac{4}{12} \frac{1}{12}\right)$$

- b) Aplicando el descuento correspondiente quedan los siguientes valores y descuentos totales:

$$\hat{p}_A = \left(\frac{5}{12} 0 0 \frac{3}{12} 0\right), \text{descuento } \frac{4}{12} \hat{p}_B = \left(0 0 \frac{4}{12} \frac{1}{12} \frac{2}{12}\right), \text{descuento } \frac{5}{12} \hat{p}_C = \left(\frac{2}{12} 0 \frac{2}{12} \frac{3}{12} 0\right), \text{descuento } \frac{5}{12}$$

Por tanto, aplicando el *backoff* correspondiente tendremos:

$$\tilde{p}_A = \left(\frac{5}{12} \frac{4}{36} \frac{4}{36} \frac{3}{12} \frac{4}{36}\right) = \left(\frac{15}{36} \frac{4}{36} \frac{4}{36} \frac{9}{36} \frac{4}{36}\right)$$

$$\tilde{p}_B = \left(\frac{5}{24} \frac{5}{24} \frac{4}{12} \frac{1}{12} \frac{2}{12}\right) = \left(\frac{5}{24} \frac{5}{24} \frac{8}{24} \frac{2}{24} \frac{4}{24}\right)$$

$$\tilde{p}_C = \left(\frac{2}{12} \frac{5}{24} \frac{2}{12} \frac{3}{12} \frac{5}{24}\right) = \left(\frac{4}{24} \frac{5}{24} \frac{4}{24} \frac{6}{24} \frac{5}{24}\right)$$

- c) La palabra “lace” se codifica como $\mathbf{x} = (1 \ 1 \ 1 \ 1 \ 0)^t$

Aplicando el clasificador original, y teniendo en cuenta que las clases son equiprobables, tendremos:

$$g_A(\mathbf{x}) = \left(\frac{6}{12}\right)^{x_1} \left(\frac{1}{12}\right)^{x_2} 0^{x_3} \left(\frac{4}{12}\right)^{x_4} \left(\frac{1}{12}\right)^{x_5}$$

$$g_B(\mathbf{x}) = \left(\frac{1}{12}\right)^{x_1} \left(\frac{1}{12}\right)^{x_2} \left(\frac{5}{12}\right)^{x_3} \left(\frac{2}{12}\right)^{x_4} \left(\frac{3}{12}\right)^{x_5}$$

$$g_C(\mathbf{x}) = \left(\frac{3}{12}\right)^{x_1} \left(\frac{1}{12}\right)^{x_2} \left(\frac{3}{12}\right)^{x_3} \left(\frac{4}{12}\right)^{x_4} \left(\frac{1}{12}\right)^{x_5}$$

Por tanto: $g_A(\mathbf{x}) = 0$, $g_B(\mathbf{x}) = \frac{10}{12^4}$, $g_C(\mathbf{x}) = \frac{27}{12^4}$, y se clasificaría en la clase C

Para el clasificador suavizado, tendríamos:

$$g_A(\mathbf{x}) = \left(\frac{15}{36}\right)^{x_1} \left(\frac{4}{36}\right)^{x_2} \left(\frac{4}{36}\right)^{x_3} \left(\frac{9}{36}\right)^{x_4} \left(\frac{4}{36}\right)^{x_5}$$

$$g_B(\mathbf{x}) = \left(\frac{5}{24}\right)^{x_1} \left(\frac{5}{24}\right)^{x_2} \left(\frac{8}{24}\right)^{x_3} \left(\frac{2}{24}\right)^{x_4} \left(\frac{4}{24}\right)^{x_5}$$

$$g_C(\mathbf{x}) = \left(\frac{4}{24}\right)^{x_1} \left(\frac{5}{24}\right)^{x_2} \left(\frac{4}{24}\right)^{x_3} \left(\frac{6}{24}\right)^{x_4} \left(\frac{5}{24}\right)^{x_5}$$

Por tanto: $g_A(\mathbf{x}) = \frac{2160}{36^4}$, $g_B(\mathbf{x}) = \frac{400}{24^4}$, $g_C(\mathbf{x}) = \frac{480}{24^4}$, y se clasificaría también en la clase C

2. **(0.5 puntos)** Dada la función Kernel $K(\mathbf{x}, \mathbf{y}) = \mathbf{x}^t \cdot \mathbf{y} + 1$ y el conjunto de datos siguiente:

	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	\mathbf{x}_4
x_{n1}	0	2	-1	-2
x_{n2}	1	2	-2	0
c_n	-1	+1	-1	+1

Se pide realizar una iteración completa del algoritmo Kernel Perceptron, mostrando sus pasos y el conjunto de pesos resultantes, partiendo del conjunto de pesos iniciales $\alpha = \mathbf{0}$.

Solución:

$$\mathbf{x}_1: g(\mathbf{x}_1) = 0 \quad c_1 g(\mathbf{x}_1) = 0 \leq 0 \rightarrow \text{Error} \rightarrow \alpha = (1 \ 0 \ 0 \ 0)$$

$$\mathbf{x}_2: g(\mathbf{x}_2) = \alpha_1 c_1 K(\mathbf{x}_2, \mathbf{x}_1) + \alpha_1 c_1 = (-1)3 + (-1) = -4 \quad c_2 g(\mathbf{x}_2) = -4 \leq 0 \rightarrow \text{Error} \rightarrow \alpha = (1 \ 1 \ 0 \ 0)$$

$$\mathbf{x}_3: g(\mathbf{x}_3) = \alpha_1 c_1 K(\mathbf{x}_3, \mathbf{x}_1) + \alpha_2 c_2 K(\mathbf{x}_3, \mathbf{x}_2) + \alpha_1 c_1 + \alpha_2 c_2 = (-1)(-1) + (+1)(-5) + (-1) + (+1) = -4 \quad c_3 g(\mathbf{x}_3) = 4 > 0 \rightarrow \text{Acuerdo} \rightarrow \alpha = (1 \ 1 \ 0 \ 0)$$

$$\mathbf{x}_4: g(\mathbf{x}_4) = \alpha_1 c_1 K(\mathbf{x}_4, \mathbf{x}_1) + \alpha_2 c_2 K(\mathbf{x}_4, \mathbf{x}_2) + \alpha_1 c_1 + \alpha_2 c_2 = (-1)(1) + (+1)(-3) + (-1) + (+1) = -4 \quad c_4 g(\mathbf{x}_4) = -4 \leq 0 \rightarrow \text{Error} \rightarrow \alpha = (1 \ 1 \ 0 \ 1)$$

Conjunto de pesos final: $\alpha = (1 \ 1 \ 0 \ 1)$

3. **(0.5 puntos)** Se tiene el siguiente conjunto de datos y clasificadores lineales:

$$\mathbf{x}_1 = ((0, 0), +1), \mathbf{x}_2 = ((1, 1), +1), \mathbf{x}_3 = ((-1, 1), +1), \mathbf{x}_4 = ((-1, -1), -1)$$

$$g_1(\mathbf{z}) = \begin{cases} +1 & z_1 \geq 1 \\ -1 & z_1 < 1 \end{cases} \quad g_2(\mathbf{z}) = \begin{cases} +1 & z_2 > 0 \\ -1 & z_2 \leq 0 \end{cases} \quad g_3(\mathbf{z}) = \begin{cases} +1 & z_1 + z_2 \leq 1 \\ -1 & z_1 + z_2 > 1 \end{cases} \quad g_4(\mathbf{z}) = \begin{cases} +1 & z_1 - z_2 \geq 0 \\ -1 & z_1 - z_2 < 0 \end{cases}$$

Se pide realizar una primera iteración de AdaBoost sobre estos datos y clasificadores, indicando la tabla de acierto y fallo por clasificador, el clasificador escogido, el error en primera iteración (ϵ_1), el peso del clasificador escogido (α_1) y los pesos de las muestras en la siguiente iteración ($\mathbf{w}^{(2)}$).

Solución:

Tabla acierto/fallo:

	g_1	g_2	g_3	g_4
x_1	X	X	✓	✓
x_2	✓	✓	X	✓
x_3	X	✓	✓	X
x_4	✓	✓	X	X

Vector de pesos de muestras inicial: $\mathbf{w}^{(1)} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$

Error ponderado por clasificador: $g_1 : \frac{2}{4}, g_2 : \frac{1}{4}, g_3 : \frac{2}{4}, g_4 : \frac{2}{4}$

Clasificador escogido: g_2

Error de clasificación: $\epsilon_1 = \frac{1}{4}$

Peso del clasificador: $\alpha_1 = \frac{1}{2} \log 3$

Pesos de las muestras en la siguiente iteración:

	$w_n^{(1)} \exp(-c_n \alpha_1 C_1(\mathbf{x}_n))$	$\mathbf{w}^{(2)}$
\mathbf{x}_1	$\frac{1}{4} \sqrt{3}$	$\frac{1}{2}$
\mathbf{x}_2	$\frac{1}{4} \frac{1}{\sqrt{3}}$	$\frac{1}{6}$
\mathbf{x}_3	$\frac{1}{4} \sqrt{3}$	$\frac{1}{6}$
\mathbf{x}_4	$\frac{1}{4} \frac{1}{\sqrt{3}}$	$\frac{1}{6}$
Suma	$\frac{3}{4} \frac{1}{\sqrt{3}} + \frac{1}{4} \sqrt{3} = \frac{\sqrt{3}}{2}$	