

Name and Surname: _____

Degree and Group: _____

NIA and Signature: _____

Question:	1	2	3	4	5	Total
Points:	10	10	15	15	10	60
Score:						

Answers to Exercise 1 (T / F): (a) ____; (b) ____; (c) ____; (d) ____;

Guidelines:

- The exam consists of 5 exercises. The points assigned to each exercise are indicated next to its statement and in the table above. The first exercise consists of four test questions, all of them worth the same points (2.5/100 points); each wrong answer will incur a penalization equivalent to the points assigned to the question. The minimum score in this first exercise will be 0.
- Please, provide your answers to the test questions in this page. Answer the rest of the exercises in different pages; every exercise will be handed in separately. Please, do not forget to write your name, surname and group in all pages you hand in.
- The exam must be solved without using any notes, solved exercises collections or calculators. It will be just accepted a handwritten sheet (both sides).
- The time assigned to the exam is **3 hours**.

Notation:

- \hat{S}_{MMSE} : Minimum Mean Square Error estimator.
- \hat{S}_{MAD} : Minimum Mean Absolute Deviation Error estimator.
- \hat{S}_{MAP} : Maximum *a posteriori* estimator.
- \hat{S}_{ML} : Maximum Likelihood estimator.
- \hat{S}_{LMSE} : Linear Minimum Mean Square Error estimator.
- ML decider: Maximum Likelihood decider $[\phi_{\text{ML}}(\mathbf{x})]$.
- MAP decider: Maximum *a posteriori* decider $[\phi_{\text{MAP}}(\mathbf{x})]$.
- LRT: Likelihood Ratio Test.
- P_e : Probability of error.
- P_{FA} : Probability of false alarm.
- P_{M} : Probability of missing.
- P_{D} : Probability of detection.
- ROC curve: Operating Characteristics curve.

1. Primera parte

- (10 %) 1. Let S and X be two random variables with joint p.d.f.

$$p_{S,X}(s, x) = \frac{1}{x}, \quad 0 \leq s \leq x; \quad 0 \leq x \leq 1.$$

We want to estimate s based on the observation of x using an estimator \hat{s} that minimizes the following cost function:

$$c(s, \hat{s}) = \frac{1}{3}\hat{s}^3 - \frac{1}{2}s\hat{s}^2, \quad s, \hat{s} \in [0, 1].$$

- (a) Find the linear estimator with analytical shape $\hat{s}_L(x) = wx$ which provides the smallest mean cost.
- (b) Now consider the squared error function. Compute the MMSE estimator $\hat{s}_{\text{MMSE}}(x)$.

- (10 %) 2. Consider the decision problem given by equally probable hypotheses and likelihoods

$$p_{X|H}(x|1) = \frac{1}{2}(2-x), \quad 0 < x < 2$$

$$p_{X|H}(x|0) = 1, \quad 0 < x < 1$$

Consider also the family of detectors

$$\begin{array}{c} D = 1 \\ (x-1)^2 \geq \mu^2 \\ D = 0 \end{array}$$

with $0 < \mu < 1$.

- (a) Determine the decision regions as a function of μ .
- (b) Determine the probability of error of the decider as a function of μ .
- (c) Find the decider belonging to the family of deciders under consideration, which incurs in a minimum probability of error.

2. Segunda parte

- (15 %) 3. Consider a one-dimensional stochastic function $f(x) : \mathbb{R} \rightarrow \mathbb{R}$, which follows a Gaussian process $f(x) \sim \mathcal{GP}(0, k(x, x'))$, with covariance function $k(x, x') = \exp(-(x - x')^2 / (2\ell^2))$.

- (a) Plot a typical random sample from this GP between -2 and 2 for length-scales $\ell \rightarrow 0$, $\ell \rightarrow \infty$, $\ell = 1$ (i.e., three plots). Identify clearly which drawing corresponds to which length-scale.
- (b) Considering the case in which $\ell \rightarrow 0$ and having observed the following two input-output pairs of a single realization of the stochastic process

$$\mathcal{D} \equiv \{f(-1) = -1, f(+1) = +1\},$$

compute the predictive mean at $x = 0$, i.e., $p(f(0)|\mathcal{D})$ for the same realization of the stochastic process.

- (c) Has \mathcal{D} been informative about $f(0)$? Provide an interpretation of the previous result.

- (15 %) 4. Consider the data set $\{(x_i, y_i), i = 1, \dots, N\}$ of independent and identically distributed samples. Binary variables $y_i \in \{0, 1\}$ have been generated according to the probabilistic model given by

$$p(y|x, w) = F(\tilde{y}(x - w)),$$

where $\tilde{y} = 2y - 1$ is a representation of the class label in $\{-1, 1\}$, F is the normalized Gaussian cumulative distribution function, that is,

$$F(x) = \int_{-\infty}^x \mathcal{N}(\alpha|0, 1) d\alpha,$$

and $\mathcal{N}(\alpha|0, 1)$ denotes the Gaussian function with zero mean and unit variance. We would like to estimate parameter w minimizing the non-negative likelihood

$$\text{NLL}(w) = -\ln(p(y_1, \dots, y_N | x_1, \dots, x_N, w)) \quad (1)$$

- (a) Find the learning rule of the steepest descent algorithm with step size η .
- (b) Find the learning rule corresponding to the Newton's method

HINT: in order to simplify the final expressions, you might find useful defining

$$r_i(w) = \frac{\mathcal{N}(x_i|w, 1)}{F(\tilde{y}_i(x_i - w))}$$

and using it to express the learning rules.

- (10%) 5. Consider a one-dimensional non-supervised learning problem, with a training set given by the 6 following samples: $x_1 = 1.8$, $x_2 = -0.2$, $x_3 = -1.6$, $x_4 = 2.6$, $x_5 = 1.6$ y $x_6 = -1.8$.
- (a) The k -means algorithm with $k = 2$ groups is used to cluster the data, using the following initialization for the group means: $\mu_1^{(0)} = x_3$ y $\mu_2^{(0)} = x_2$. How many iterations are necessary for the convergence of the algorithm? Find the means of each group after the convergence of the algorithm.
 - (b) Justify if there exists any other group representatives different from the previous ones that provide a smaller total quadratic distortion:

$$J = \sum_{x_k \in \mathcal{C}_1} (x_k - \mu_1)^2 + \sum_{x_k \in \mathcal{C}_2} (x_k - \mu_2)^2$$

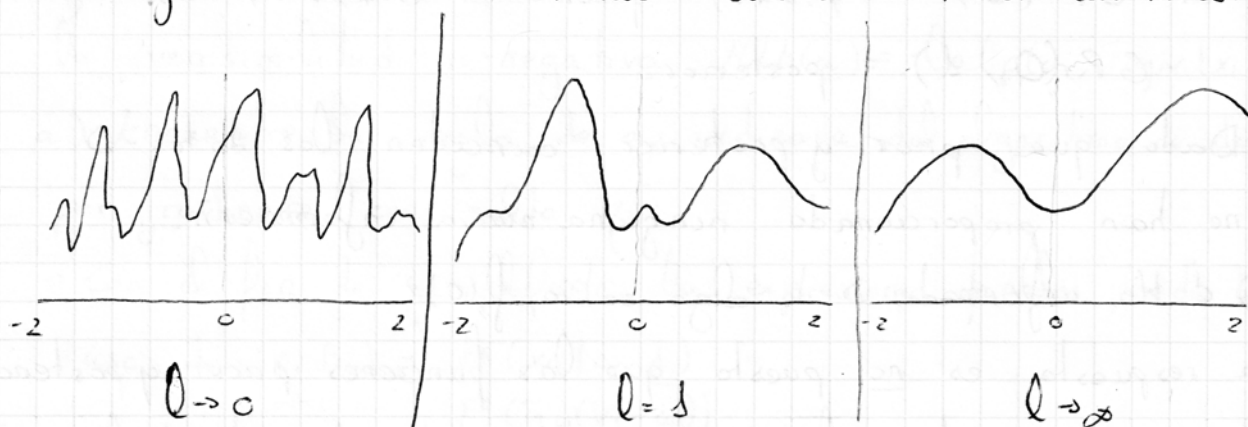
- (c) Using the assignment criterion of k -means, find the sample estimations of the *a priori* probabilities and of the variance of each group. Approximate your results with two decimal digits.
- (d) Assuming that the probability density function that has generated the data in each group is Gaussian, provide the analytic expression that allows to calculate the log-likelihood of the Gaussian Mixture Model that uses the parameters obtained in the previous subsections.

1] Considere una función estocástica unidimensional $f(x): \mathbb{R} \rightarrow \mathbb{R}$ que sigue un proceso Gaussiano $f(x) \sim GP(0, K(x, x'))$ con función de covarianza $K(x, x') = \exp(-(x-x')^2/(2l^2))$

a) Dibuje una muestra aleatoria típica del GP entre -2 y 2 para valores $l \rightarrow 0$, $l=1$ y $l \rightarrow \infty$ (tres gráficas). Identifique claramente a qué valor de l corresponde cada dibujo.

→ Las muestras son más "ruidosas" (tienen más picos, juntos entre sí) cuando $l \rightarrow 0$ que cuando $l \rightarrow \infty$.

↳ Conforme aumenta l , más "suaves" se hacen las muestras.



b) Considere ahora el caso para el que $l \rightarrow 0$ y habiendo observado el siguiente par de entradas-salidas para una única realización del proceso estocástico $D = \{f(-1) = -1, f(1) = +1\}$

Calcule la media predictiva en $x=0$ es $(p(f(0)|D))$ para la misma realización del proceso estocástico.

→ La media y la varianza de $p(f|D)$ viene dada por

$$p(y|D) = (f|_D + K(X_*, X) K(X, X)^{-1} y, \quad K(X_*, X_*) - K(X_*, X) K(X, X)^{-1} K(X, X_*)$$

$$K(X_*, X_*) - K(X_*, X) K(X, X)^{-1} K(X, X_*)$$

El Kernel $K(X, X)$ da lugar a una matriz identidad 2×2

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

El Kernel $K(x_*, x)$ da lugar a una matriz $1 \times 2 \neq 0$
 $= (x \cdot 10^{-21} \quad x \cdot 10^{-21})$ (por ejemplo)

El Kernel $K(x_*, x_*)$ da lugar a una matriz $1 \times 1 = 1$

Aplicamos la fórmula de $p(\theta|D)$

Media: $(x \cdot 10^{-21} \quad x \cdot 10^{-21}) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} y = (x \cdot 10^{-21} \quad x \cdot 10^{-21}) y \approx 0$

Varianza: $1 - (x \cdot 10^{-21} \quad x \cdot 10^{-21}) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (x \cdot 10^{-21} \quad x \cdot 10^{-21})^T = 1 - 0 = 1$

La media en $x=0$ es 0, y lo es independientemente del valor de los datos, pues el prior de $f(c)$ es un GP $(0, K(x_*, x_*))$ y como $K(x_*, x_*) = 1$, $GP(0, 1) = \text{posterior}$.

Dado que prior y posterior coinciden los datos D no han proporcionado ninguna nueva información.

c) ¿Ha informado D de algo sobre $f(c)$?

La respuesta es no puesto que las funciones prior y posterior en $x=0$ coinciden, luego el posterior que depende de los datos D , no ha ofrecido información más útil.

$$\begin{aligned} \omega_{ML} &= \arg\max_{\omega} \prod_{k=1}^N p(y^{(k)} | x^{(k)}, \omega) = \arg\max_{\omega} \prod_{k=1}^N p(y^{(k)} | x^{(k)}, \omega) \prod_{k=1}^N p(x^{(k)}) \\ &= \arg\max_{\omega} \prod_{k=1}^N p(y^{(k)} | x^{(k)}, \omega) \stackrel{\arg\max(x) = \arg\min(-x)}{=} \arg\min_{\omega} -\ln \prod_{k=1}^N p(y^{(k)} | x^{(k)}, \omega) = \\ &= \arg\min_{\omega} \sum_{k=1}^N -\ln(p(y^{(k)} | x^{(k)}, \omega)) = \underbrace{\arg\min_{\omega} \sum_{k=1}^N -\ln F(\hat{y}(x-\omega))}_{NLL} \end{aligned}$$

Para el algoritmo de optimización por gradiente, la regla de aprendizaje es la siguiente:

$$\omega^{(k+1)} = \omega^{(k)} - \eta \nabla_{\omega} NLL(\omega)$$

$$\nabla_{\omega} NLL(\omega) = \nabla_{\omega} \left[-\sum_{k=1}^N \ln F(\hat{y}(x-\omega)) \right] = -\sum_{k=1}^N \frac{1}{F(\hat{y}(x-\omega))} \cdot F'(\hat{y}(x-\omega)) y$$

12) Considere el conjunto de datos $\{(x_i, y_i), i=1, \dots, N\}$ de muestras distribuidas de forma idéntica e independiente. Las variables binarias $y_i \in \{0, 1\}$ han sido generadas según el modelo probabilístico dado por $p(y|x, w) = F(\hat{y}(x-w))$ donde $\hat{y} = 2y-1$ es una representación de las etiquetas de clase $\{-1, 1\}$ y F es la CDF Gaussiana normalizada:

$$F(x) = \int_{-\infty}^x N(\alpha|0,1) d\alpha$$

y $N(\alpha|0,1)$ se refiere a una función Gaussiana de media 0 y varianza 1. Queremos estimar el parámetro w minimizando la verosimilitud no-negativa $NLL(w) = -\ln(p(y_1, \dots, y_N | x_1, \dots, x_N, w))$

a) Encuentre la regla de aprendizaje del algoritmo de descenso por gradiente con salto η .

→ Con el fin de simplificar las expresiones finales puede hacer uso de $r_i(w) = \frac{N(x_i|w, 1)}{F(\hat{y}_i(x_i-w))}$

$$\begin{aligned} w_{ML} &= \arg\max_w \prod_{k=1}^N p(x^{(k)}, y^{(k)} | w) = \arg\max_w \prod_{k=1}^N p(y^{(k)} | x^{(k)}, w) \prod_{k=1}^N p(x^{(k)}) \\ &= \arg\max_w \prod_{k=1}^N p(y^{(k)} | x^{(k)}, w) \stackrel{\arg\max(w) = \arg\min(-x)}{=} \arg\min_w -\ln \prod_{k=1}^N p(y^{(k)} | x^{(k)}, w) = \\ &= \arg\min_w \sum_{k=1}^N -\ln p(y^{(k)} | x^{(k)}, w) = \arg\min_w \underbrace{\sum_{k=1}^N \ln F(\hat{y}_k(x-w))}_{NLL} \end{aligned}$$

No depende de w

En el algoritmo de optimización por gradiente, la regla de aprendizaje es: $w^{(k+1)} = w^{(k)} - \eta \nabla_w NLL(w)$.

$$\nabla_w NLL(w) = \nabla_w [-\sum_{k=1}^N \ln F(\hat{y}_k(x-w))] = -\sum_{k=1}^N \frac{1}{F(\hat{y}_k(x-w))} F'(\hat{y}_k(x-w)) (-\hat{y}_k)$$

$$F'(\hat{y}(x-w)) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(\hat{y}(x-w))^2}{2}} = e^{-\tilde{y}^2} \cdot N(x|w, 1)$$

$$\hookrightarrow \nabla_w NLL(w) = +\sum_{k=1}^N \hat{y}_k e^{-\tilde{y}_k^2} r_k(w)$$

Así, la regla de aprendizaje queda

$$\boxed{\omega^{(k+1)} = \omega^{(k)} - \eta \sum_{k=1}^N \tilde{y}^{(k)} \tilde{y}^{(k)2} r_k(\omega)}$$

b) Encuentra la regla de aprendizaje correspondiente al método de Newton.

En este caso hay que calcular el valor de la matriz Hessiana que es $\nabla_{\omega}^2 \mathcal{H}(\omega) = \sum_{k=1}^N \tilde{y}^{(k)} \tilde{y}^{(k)2} r_k'(\omega)$

$$\begin{aligned} r_k'(\omega) &= \mathcal{N}'(x | \omega, 1) \cdot F(\tilde{y}(x - \omega)) - \mathcal{N}(x | \omega, 1) \cdot F'(\tilde{y}(x - \omega))(-\tilde{y}) \\ &= H(\omega) \end{aligned}$$

$$\vec{g} = \sum_{k=1}^N \tilde{y}^{(k)} \tilde{y}^{(k)2} r_k(\omega)$$

luego la regla de aprendizaje es $\boxed{\omega^{(k+1)} = \omega^{(k)} - \eta H^{-1}(\omega) \vec{g}}$

13) Considera un problema de aprendizaje no supervisado unidimensional con un conjunto de entrenamiento dado por las 6 siguientes muestras: $x_1 = 1.8$, $x_2 = -0.2$, $x_3 = -1.6$, $x_4 = 2.6$, $x_5 = 1.6$ y $x_6 = -1.8$

a) Se emplea el algoritmo K-means con $K=2$ grupos para agrupar los datos utilizando la siguiente inicialización para las medias grupales $\mu_1^{(0)} = x_3$ y $\mu_2^{(0)} = x_2$. ¿Cuántas iteraciones son necesarias para que el algoritmo converja? Calcula las medias grupales al alcanzar el estado de convergencia.

Inicialización: $\mu_1^{(0)} = -1.6$ $\mu_2^{(0)} = -0.2$

$$|K=1| \quad |x_1 - \mu_1^{(0)}|^2 = 11.56 \quad |x_2 - \mu_1^{(0)}|^2 = 1.96 \quad |x_3 - \mu_1^{(0)}|^2 = 0$$

$$|x_4 - \mu_1^{(0)}|^2 = 17.64 \quad |x_5 - \mu_1^{(0)}|^2 = 10.24 \quad |x_6 - \mu_1^{(0)}|^2 = 0.64$$

$$|K=2| \quad |x_1 - \mu_2^{(0)}|^2 = 4 \quad |x_2 - \mu_2^{(0)}|^2 = 0 \quad |x_3 - \mu_2^{(0)}|^2 = 1.96$$

$$|x_4 - \mu_2^{(0)}|^2 = 7.84 \quad |x_5 - \mu_2^{(0)}|^2 = 3.24 \quad |x_6 - \mu_2^{(0)}|^2 = 0.36$$

Grupo 1 = x_3, x_6

$$\mu_1^{(1)} = \frac{x_3 + x_6}{2} = -1.4$$

Grupo 2 = x_1, x_2, x_4, x_5

$$\mu_2^{(1)} = \frac{x_1 + x_2 + x_4 + x_5}{4} = 1.45$$

$$\text{Distorsión global} = (0 + 0.64) + (4 + 0 + 7.84 + 3.24) = 15.12$$

Iteración 1: $\mu_1^{(1)} = -1.4$ $\mu_2^{(1)} = 1.45$

$$|K=1| \quad |x_1 - \mu_1^{(1)}|^2 = 12.25 \quad |x_2 - \mu_1^{(1)}|^2 = 2.25 \quad |x_3 - \mu_1^{(1)}|^2 = 1.21$$

$$|x_4 - \mu_1^{(1)}|^2 = 18.49 \quad |x_5 - \mu_1^{(1)}|^2 = 10.89 \quad |x_6 - \mu_1^{(1)}|^2 = 0.01$$

$$|K=2| \quad |x_1 - \mu_2^{(1)}|^2 = 0.1225 \quad |x_2 - \mu_2^{(1)}|^2 = 2.72 \quad |x_3 - \mu_2^{(1)}|^2 = 9.13$$

$$|x_4 - \mu_2^{(1)}|^2 = 1.32 \quad |x_5 - \mu_2^{(1)}|^2 = 0.0225 \quad |x_6 - \mu_2^{(1)}|^2 = 10.56$$

Grupo 1 = x_2, x_3, x_6

$$\mu_1^{(2)} = \frac{x_2 + x_3 + x_6}{3} = -1.2$$

Grupo 2 = x_1, x_4, x_5

$$\mu_2^{(2)} = \frac{x_1 + x_4 + x_5}{3} = \dots$$

$$\text{Distorsión global} = (0'1225 +$$

$$= (2'25 + 1'25 + 0'05) + (0'1225 + 1'32 + 0'0225) =$$

$$= 4'935$$

$$\text{Iteración 2: } \mu_1^{(2)} = -1'2 \quad \mu_2^{(2)} = 2$$

$$|K=1| \quad |x_1 - \mu_1^{(2)}|^2 = 9 \quad |x_2 - \mu_1^{(2)}|^2 = 1 \quad |x_3 - \mu_1^{(2)}|^2 = 0'16$$

$$|x_4 - \mu_1^{(2)}|^2 = 14'44 \quad |x_5 - \mu_1^{(2)}|^2 = 7'84 \quad |x_6 - \mu_1^{(2)}|^2 = 0'36$$

$$|K=2| \quad |x_1 - \mu_2^{(2)}|^2 = 0'04 \quad |x_2 - \mu_2^{(2)}|^2 = 4'84 \quad |x_3 - \mu_2^{(2)}|^2 = 12'96$$

$$|x_4 - \mu_2^{(2)}|^2 = 0'36 \quad |x_5 - \mu_2^{(2)}|^2 = 0'16 \quad |x_6 - \mu_2^{(2)}|^2 = 14'44$$

$$\text{Grupo 1} = x_2, x_3, x_6 \quad \mu_1^{(3)} = \frac{x_2 + x_3 + x_6}{3} = -1'2$$

$$\text{Grupo 2} = x_1, x_4, x_5 \quad \mu_2^{(3)} = \frac{x_1 + x_4 + x_5}{3} = 2$$

$$\text{Distorsión global} = (1 + 0'16 + 0'36) + (0'04 + 0'36 + 0'16) =$$

$$= 2'08$$

Como las medias obtenidas son las mismas que las anteriores, el algoritmo para aquí. Han sido necesarias 2 iteraciones para llegar a un estado de convergencia.

b) Justifica si existe algún otro representante de grupo distinto del apartado anterior que proporcione una menor distorsión cuadrática media.

$$J = \sum_{x_k \in G_1} (x_k - \mu_1)^2 + \sum_{x_k \in G_2} (x_k - \mu_2)^2$$

Dado que existen dos grupos, habiendo muestras positivas y negativas, lo mejor sería inicializar las medias grupales de manera que estén lo más alejadas entre sí, es decir

$$\mu_1^{(0)} = -1'8 = x_6 \quad \text{y} \quad \mu_2^{(0)} = 2'6 = x_4. \text{ Así:}$$

Inicialización:

$$K=1 \quad |x_1 - \mu_1^{(0)}|^2 = 12'96 \quad |x_2 - \mu_1^{(0)}|^2 = 2'56 \quad |x_3 - \mu_1^{(0)}|^2 = 0'04$$

$$|x_4 - \mu_1^{(0)}|^2 = 19'36 \quad |x_5 - \mu_1^{(0)}|^2 = 11'56 \quad |x_6 - \mu_1^{(0)}|^2 = 0$$

$$K=2 \quad |x_1 - \mu_2^{(0)}|^2 = 0'64 \quad |x_2 - \mu_2^{(0)}|^2 = 7'84 \quad |x_3 - \mu_2^{(0)}|^2 = 17'64 \\ |x_4 - \mu_2^{(0)}|^2 = 0 \quad |x_5 - \mu_2^{(0)}|^2 = 1 \quad |x_6 - \mu_2^{(0)}|^2 = 19'36$$

$$\text{Grupo 1} = x_3, x_6, x_2 \quad \mu_1^{(1)} = \frac{x_2 + x_3 + x_6}{3} = -1'2$$

$$\text{Grupo 2} = x_1, x_4, x_5 \quad \mu_2^{(1)} = \frac{x_1 + x_4 + x_5}{3} = 2$$

$$\text{Distorsión global} = (0'04 + 4 + 0) + (0'64 + 0 + 1) = 5'68$$

⇒ Considerablemente menor que 15'12 en el apartado a.

Como las medias grupales nuevas son las mismas que las medias finales del apartado a, en la siguiente iteración el algoritmo convergerá con distorsión global de 2'08.

→ En este caso solo ha hecho falta una iteración para converger.

c) Utilizando el criterio de asignación del K-medias, encuentra las probabilidades a priori de las muestras, así como la varianza de cada grupo.

→ Dado que cada grupo tiene 3 muestras asignadas y hay 6 muestras en total, se pueden calcular las probabilidades a priori como $|3/6 = 1/2|$, por lo que las probabilidades a priori de cada grupo son equiprobables.

Para el caso de la varianza, aplicamos la siguiente fórmula:

$$\text{Var-grupo} = \frac{1}{n^{\circ} \text{ elementos/grupo}} \sum_{k=1}^N (x\text{-grupo} - \text{media-grupo})^2 \text{ para } N \text{ elementos en el grupo}$$

Así:

$$\text{Var}_1 = \frac{1}{3} [(x_2 - (-1'2))^2 + (x_3 - (-1'2))^2 + (x_6 - (-1'2))^2] = \frac{1'52}{3} =$$

$$= 0'51$$

$$\text{Var}_2 = \frac{1}{3} [(x_1 - (2))^2 + (x_4 - 2)^2 + (x_5 - 2)^2] = 0'187$$

d) Asumiendo que la función de densidad que ha generado los datos en cada grupo, es Gaussiana, proporciona la expresión analítica que permite calcular la log-verosimilitud del modelo mixto Gaussiano, que utiliza los parámetros obtenidos en las secciones anteriores.

$$\left. \begin{array}{l} \mu_1 = -1/2 \quad \text{Var}_1 = 0.51 \\ p_1 = 1/2 \end{array} \right\} \begin{array}{l} \mu_2 = 2, \quad \text{Var}_2 = 2 \\ p_2 = 1/2 \end{array}$$

$$p(x_k | \mu_1, \mu_2, \omega) = \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi} \cdot 0.51} \cdot e^{-\frac{(x+1/2)^2}{2 \cdot 0.51}} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi} \cdot 2} \cdot e^{-\frac{(x-2)^2}{4}}$$

La verosimilitud viene dada por $\prod_{k=1}^6 p(x_k | \mu_1, \mu_2, \omega)$
y la log verosimilitud es:

$$\ln \prod_{k=1}^6 p(x_k | \mu_1, \mu_2, \omega) = \sum_{k=1}^6 \ln p(x_k | \mu_1, \mu_2, \omega) =$$

$$= \ln \left(\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi} \cdot 0.51} \right) + \ln \left(e^{-\frac{(x+1/2)^2}{2 \cdot 0.51}} \right) + \ln \left(\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi} \cdot 2} \right) + \ln \left(e^{-\frac{(x-2)^2}{4}} \right) =$$

$$= -\frac{(x+1/2)^2}{2 \cdot 0.51} - \frac{(x-2)^2}{4} + 0.938 - 2.3 =$$

$$= -x^2 - 2.4x + 1.44 - \frac{x^2 + 2x - 4}{4} - 3.238;$$

$$= \frac{-4x^2 - 9.6x - 5.46 - x^2 + 2x - 4 - 12.952}{4} =$$

$$= \boxed{\frac{-5x^2 - 7.6x - 22.41}{4}}$$