

# Convocatoria C2

## Fundamentos del Aprendizaje Automático

Grado en Ingeniería en Inteligencia Artificial  
Universidad de Alicante

Apellidos: \_\_\_\_\_

Nombre: \_\_\_\_\_

DNI: \_\_\_\_\_

### Instrucciones:

- Parte *test*:
  - Responda las preguntas tipo test eligiendo **solo una** opción.
  - Cada **respuesta correcta** suma **1 punto** y cada **respuesta incorrecta** resta **1/3** de puntuación.
  - Esta parte puntúa hasta un máximo de **4 puntos** de la nota final.
  - La duración máxima de esta parte es de **30 minutos**.
- Parte *desarrollo*:
  - Cada problema debe resolverse en una hoja separada.
  - Esta parte puntúa hasta un máximo de **6 puntos** de la nota final.
  - La duración máxima de esta parte es de **90 minutos**.

Duración del examen: **120 minutos**

## Parte I: Preguntas test

Rúbrica:	Respuesta correcta ( $R_c$ ): +1	Respuesta incorrecta ( $R_i$ ): -1/3	Nota (sobre 4): $4 \cdot (\sum R_c + \sum R_i)/20$
----------	----------------------------------	--------------------------------------	--

**Pregunta 1:** El aprendizaje automático:

- a) Requiere necesariamente datos etiquetados.
- b) Busca imitar la inteligencia humana.
- c) Infiere conocimiento a partir de datos.
- d) Ninguna de las anteriores es cierta.

**Pregunta 3:** La etapa *aprendizaje computacional* en un proyecto de aprendizaje automático:

- a) Busca extraer características adecuadas para la tarea objetivo.
- b) Explotar un modelo para obtener las predicciones dado un conjunto de datos.
- c) Extrae conocimiento de datos conocidos.
- d) Ninguna de las anteriores es correcta.

2

**Pregunta 5:** El error de Bayes:

- a) Representa el mínimo error insalvable de un clasificador.
- b) Se puede reducir utilizando un clasificador de  $k$  vecinos más cercanos.
- c) No depende del tipo de función de pérdida considerada.
- d) Ninguna de las opciones es correcta.

**Pregunta 2:** La optimización de un modelo de aprendizaje:

- a) Se realiza sobre la partición de *test*.
- b) Busca la memorización de los datos de *train*.
- c) Minimiza el error sobre la partición de *validation*.
- d) Todas las anteriores son ciertas.

**Pregunta 4:** Dado un modelo de aprendizaje automático:

- a) El error por sobreajuste (*overfit*) ocurre normalmente en modelos de alta complejidad.
- b) El error por sobreajuste (*overfit*) ocurre normalmente en modelos de baja complejidad.
- c) El error por infrajuste (*underfit*) ocurre normalmente en modelos de alta complejidad.
- d) Seleccionar una complejidad óptima elimina los errores de infrajuste (*underfit*) y sobreajuste (*overfit*).

**Pregunta 6:** La verosimilitud (*likelihood*) en la Teoría Bayesiana de la Decisión:

- a) Es suficiente para obtener un clasificador óptimo.
- b) Se puede aproximar mediante métodos como *Maximum Likelihood Estimation*.
- c) Es prescindible en tareas de regresión.
- d) Todas las anteriores son ciertas.

**Pregunta 7:** La validación cruzada basada en *leave-one-out*:

- a) Realiza un único particionado de los datos.
- b) Minimiza el error de sesgo (*bias*).
- c) Minimiza el error de varianza (*variance*).
- d) Ninguna de las anteriores es cierta.

**Pregunta 9:** La estimación no parámetrica de la verosimilitud:

- a) Asume una distribución normal de los datos.
- b) Considera procedimientos de estimación de la densidad de probabilidad.
- c) Es únicamente aplicable a casos de baja dimensionalidad (p.ej., menos de 10 características).
- d) Ninguna de las anteriores es cierta.

**Pregunta 11:** La regla del vecino más cercano:

- a) Permite estimar la probabilidad a posteriori en una tarea de clasificación.
- b) Tiene su error acotado por el riesgo de Bayes.
- c) Se puede utilizar en representaciones estructurales de datos.
- d) Todas las anteriores son ciertas.

**Pregunta 13:** El entrenamiento de un perceptrón multicapa:

- a) Considera únicamente la partición de *train* del conjunto de datos.
- b) Únicamente adapta los pesos de la capa oculta.
- c) Requiere de una activación *softmax* para tareas de regresión.
- d) Ninguna de las anteriores es cierta.

**Pregunta 8:** La métrica exhaustividad (*Recall*):

- a) Penaliza el haber estimado un elemento como negativo cuando realmente era positivo.
- b) Es una métrica definida para entornos multiclas.
- c) Se combina con *Accuracy* para obtener  $F_1$ .
- d) Se relaciona con *Precision* a través de *Accuracy*.

**Pregunta 10:** La estimación de densidad basada en  $k_n$  vecinos más cercanos:

- a) Considera un número de puntos fijo para estimar la densidad.
- b) Es costoso a nivel computacional.
- c) Tiene un tamaño de ventana variable.
- d) Todas las anteriores son correctas.

**Pregunta 12:** Dado un modelo perceptrón:

- a) Podemos modelar cualquier función real.
- b) Tiene su error acotado por el riesgo de Bayes.
- c) Se asume que tiene la función signo como activación.
- d) Ninguna de las anteriores es correcta.

**Pregunta 14:** El sobreajuste en un perceptrón multicapa:

- a) Rara vez ocurre.
- b) Se evita mediante métodos de validación cruzada.
- c) Se debe al principalmente al problema de *vanishing gradient*.
- d) Ninguna de las anteriores es cierta.

**Pregunta 15:** El algoritmo de clustering de *k-means*:

- a) Considera el algoritmo de Friedman para su entrenamiento.
- b) Incrementa la cantidad de *clusters* en cada iteración.
- c) Puede devolver un resultado diferente para ejecuciones sucesivas del método (no determinista).
- d) Ninguna de las anteriores es cierta.

**Pregunta 17:** En un test de contraste de hipótesis para múltiples clasificadores:

- a) La hipótesis nula es que los clasificadores tienen bondades significativamente diferentes.
- b) Deberemos siempre utilizar un test de tipo paramétrico.
- c) Es necesario un test de tipo post-hoc siempre que no se rechace la hipótesis nula.
- d) Ninguna de las anteriores es cierta.

4

**Pregunta 19:** Para una tarea de búsqueda deberé considerar:

- a) Un modelo basado en perceptrón multicapa.
- b) Un modelo basado en *k* vecinos más cercanos.
- c) No es posible realizar tareas de búsqueda mediante aprendizaje automático.
- d) Ninguna de las anteriores es correcta.

**Pregunta 16:** El método de *análisis por componentes principales*:

- a) Reparte de manera equitativa la varianza entre todos sus autovectores.
- b) Establece que los autovectores deben ser normales entre sí.
- c) Reduce la dimensionalidad siempre que la cantidad de autovectores sea mayor que la cantidad inicial de características.
- d) Todas las anteriores son ciertas.

**Pregunta 18:** Si he propuesto un nuevo modelo de clasificación y quiero compararlo con el que actualmente define el estado de la cuestión:

- a) Deberé utilizar un test de Friedman.
- b) Deberé utilizar un test de Wilcoxon.
- c) Deberé utilizar un test de Nemenyi.
- d) No puedo realizar la comparativa entre dos modelos.

**Pregunta 20:** Un modelo Naïve Bayes:

- a) Asume que los datos siguen una distribución estadística de Bernoulli.
- b) Requiere la estimación de la función verosimilitud.
- c) Prescinde de ciertas características (*features*) del conjunto de datos.
- d) Ninguna de las anteriores es correcta.

## Parte II: Problemas de desarrollo

**Problema 1** (3p): Considera un clasificador Bayesiano diseñado para un espacio de características  $\mathbb{R}^d$  con las siguientes condiciones:

- Binario con las clases  $\mathcal{W} = \{\omega_1, \omega_2\}$
- Probabilidades a priori  $P(\omega_1) = 0,2$  y  $P(\omega_2) = 0,8$ .
- Funciones de verosimilitud  $P(\mathbf{x}|\omega_1)$  y  $P(\mathbf{x}|\omega_2)$ , con  $\mathbf{x} \in \mathbb{R}^d$ .
- Dos posibles acciones  $\mathcal{A} = \{\alpha_1, \alpha_2\}$  correspondientes al etiquetado de un objeto  $\mathbf{x} \in \mathbb{R}^d$  como  $\omega_1$  o  $\omega_2$ , respectivamente.

Considerando este escenario, responde a las siguientes cuestiones:

**Apartado a)** Considerando una función de pérdida  $\lambda(\alpha_i|\omega_j)$ , define el *riesgo condicional*  $R(\alpha|\mathbf{x})$  para cada una de las acciones  $\alpha \in \mathcal{A}$ . (1,5p)

**Apartado b)** Si particularizamos la función de pérdida de la siguiente manera:

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i = j \\ \phi & i = 1 \wedge j = 2 \\ \phi^2 & i = 2 \wedge j = 1 \end{cases}$$

Particulariza las expresiones de *riesgo condicional*  $R(\alpha|\mathbf{x})$  para cada una de las acciones  $\alpha \in \mathcal{A}$  anteriormente obtenidas. (0,5p)

**Apartado c)** Obtén la expresión del ratio de verosimilitud y particulariza para los valores  $\phi \in \{1, 2, 4\}$ . (1p)

**Problema 2** (3p): Una empresa de muebles se plantea incorporar un asistente de ventas virtual que asista en la venta virtual de sus productos. Para ello considera que el canal voz es prioritario por su tipología de clientes, por lo que decide estudiar los sistemas de reconocimiento de voz existentes en el mercado y así poder elegir el más conveniente.

En su estudio de mercado comparan tres sistemas de reconocimiento del habla ( $H$ ) en diez diferentes conjuntos de grabaciones ( $\mathcal{D}$ ), obteniendo los resultados de la Tabla 1. Nótese que estos resultados reflejan la tasa de error de los diferentes métodos en los diferentes escenarios.

Cuadro 1: Comparativa de diferentes sistemas de reconocimiento del habla ( $H$ ) en diferentes conjuntos de grabaciones ( $\mathcal{D}$ ) en términos de error de transcripción.

Dataset	$H_1$	$H_2$	$H_3$
$\mathcal{D}_1$	32	9	31
$\mathcal{D}_2$	34	10	33
$\mathcal{D}_3$	30	8	32
$\mathcal{D}_4$	35	9.5	34
$\mathcal{D}_5$	33	11	32
$\mathcal{D}_6$	31	7	33
$\mathcal{D}_7$	36	12	35
$\mathcal{D}_8$	29	6	28
$\mathcal{D}_9$	37	13	36
$\mathcal{D}_{10}$	32	14	31

Considerando este escenario, responde a las siguientes cuestiones:

**Apartado a)** ¿Existe algún método que supere, de manera significativa ( $\alpha = 0,05$ ), al resto? Razona la respuesta. (1,5p)

**Apartado b)** ¿Y si ahora exigíramos unos resultados más exigentes y fijáramos el umbral de significancia en  $\alpha = 0,01$ ? Razona la respuesta. (0,5p)

**Apartado c)** Tras una serie de meses de análisis, la empresa que comercializa el producto  $H_2$  deja de proporcionar licencias. En esta coyuntura, la empresa de muebles nos plantea qué sistema de reconocimiento de los restantes nos interesaría utilizar. Razona tu respuesta. Considera, si es necesario, un nivel de significancia de  $\alpha = 0,05$ . (1p)

## Anexos

- Friedman:

$$\chi_F^2 = \frac{12 \cdot M}{C \cdot (C + 1)} \left[ \sum_{j=1}^C \bar{R}_j^2 \right] - 3 \cdot M \cdot (C + 1)$$

$C - 1$	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086
6	10.645	12.592	16.812
7	12.017	14.067	18.475
8	13.362	15.507	20.090
9	14.684	16.919	21.666
10	15.987	18.307	23.209

- Nemenyi:

$$CD = q_\alpha(C) \cdot \sqrt{\frac{C \cdot (C + 1)}{6 \cdot M}}$$

C	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
2	1.960	2.241	2.807
3	2.052	2.343	2.949
4	2.108	2.403	3.020
5	2.146	2.444	3.069
6	2.174	2.475	3.105
7	2.195	2.499	3.133
8	2.211	2.518	3.157
9	2.224	2.534	3.176
10	2.235	2.548	3.192

- Bonferroni-Dunn:

$$CD = q_{\alpha/C-1} \cdot \sqrt{\frac{C \cdot (C + 1)}{6 \cdot M}}$$

C	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
3	1.645	1.960	2.576
4	1.282	1.645	2.326
5	1.163	1.533	2.241
6	1.095	1.476	2.192
7	1.054	1.440	2.160
8	1.027	1.414	2.136
9	1.006	1.395	2.120
10	0.990	1.380	2.107